AD-758 571

PRELIMINARY REPORTS, MEMORANDA AND TECHNICAL NOTES OF THE MATERIALS RESEARCH COUNCIL SUMMER CONFERENCE, CENTERVILLE, MASSACHUSETTS

Edward E. Hucke

Michigan University

Prepared for:

Advanced Research Projects Agency

July 1972

Preliminary Reports, Memoranda and Technical Notes of the Materials Research Council Summer Conference

Centerville, Massachusetts

July 1972

Sponsored by
Advanced Research Projects Agency
ARPA Order No. ~~0236~~ 2341

Department of Materials and Metallurgical Engineering

PRELIMINARY REPORTS, MEMORANDA AND TECHNICAL NOTES

of the

MATERIALS RESEARCH COUNCIL SUMMER CONFERENCE

Centerville, Massachusetts

July 1972

ARPA Order Number: ~~0236~~ 2341
Program Code Number: 1D10
Contractor: The Regents of The University of Michigan
Effective Date of Contract: 1 May 72
Contract Expiration Date: 30 June 73
Amount of Contract: $237,690
Contract Number: DAHC15-71-C-0253
Principal Investigator: Professor Edward E. Hucke
Department of Materials & Metallurgical Engineering
The University of Michigan
Ann Arbor, Michigan 48104
(313) 764-3302

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the U.S. Government.

TABLE OF CONTENTS

TITLE PAGE

TITLE PAGE

Foreword

This collection of papers does not constitute a formal reporting of the activities of the ARPA Materials Research Council Summer Conference. Each report, memorandum or technical note is a draft of the author or authors and is their work alone. The Steering Committee, in conjunction with the authors, will decide how this material can best be presented as a formal report to ARPA.

Steering Committee

Dr. George H. Vineyard
Secretary of the Steering Committee
Brookhaven National Laboratory
Upton, Long Island, New York 11973

Professor Willis H. Flygare
Noyes Chemical Laboratory
University of Illinois
Urbana, Illinois 61801

Professor John P. Hirth
Metallurgical Engineering Department
Ohio State University
Columbus, Ohio 43201

Professor James A. Krumhansl
Department of Physics
Clark Hall of Science
Cornell University
Ithaca, New York 14850

Professor Frank A. McClintock
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Professor Howard Reiss
Department of Chemistry
University of California
Los Angeles, California 90024

Professor James Rice
Division of Engineering
Brown University
Providence, Rhode Island 02912

Dr. Robb M. Thomson
National Bureau of Standards
Institute of Applied Technology
Washington, D.C. 20234

Professor Michael Tinkham
Department of Physics
Harvard University
Cambridge, Massachusetts 02138

Project Director

Professor Edward E. Hucke
Department of Materials and Metallurgical Engineering
The University of Michigan
Ann Arbor, Michigan 48104

Participants

Professor Michael B. Bever
Department of Metallurgy & Materials Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Professor A. Bienenstock
Materials Science Department
Stanford University
Stanford, California 94305

Professor Nico Bloembergen
Division of Engineering & Applied Physics
Harvard University
Cambridge, Massachusetts 02138

Professor Bernard Budiansky
Division of Engineering & Applied Science
Harvard University
Cambridge, Massachusetts 02138

Professor Robert Coble
Materials Science Department
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Professor Morris Cohen
Department of Metallurgy & Materials Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Dean Daniel C. Drucker
Engineering College
University of Illinois
Urbana, Illinois 61801

Professor Pol E. Duwez
W. M. Keck Laboratory of Engineering Materials
California Institute of Technology
Pasadena, California 91109

Professor H. Ehrenreich
Pierce Hall
Harvard University
Cambridge, Massachusetts 02138

Professor John D. Ferry
Department of Chemistry
University of Wisconsin
Madison, Wisconsin 53706

Dr. John J. Gilman, Director
Materials Research Center
Allied Chemical Corporation
Morristown, New Jersey 07960

Professor Robert Gomer
James Franck Institute
University of Chicago
Chicago, Illinois 60637

Professor Robert A. Huggins
Center for Materials Research
Stanford University
Stanford, California 94305

Professor Walter Kohn
Department of Physics
University of California
La Jolla, California 92037

Professor Erastus H. Lee
Department of Applied Mechanics
Stanford University
Stanford, California 94305

Professor Donald J. Lyman
Materials Science & Engineering
The University of Utah
Salt Lake City, Utah 84112

Professor John L. Margrave
Department of Chemistry
Rice University
Houston, Texas 77001

Professor Elliott W. Montroll
Department of Physics & Astronomy
University of Rochester
Rochester, New York 14534

Professor Paul L. Richards
Department of Physics
University of California
Berkeley, California 94720

Professor Albert J. Sievers
Laboratory of Atomic & Solid State Physics
Cornell University
Ithaca, New York 14850

ARPA MATERIALS RESEARCH COUNCIL
1972 GUEST CONSULTANT LIST

Richard A. Alliegro
Norton Company
ICD R&D Department
1 New Bond Street
Worchester, Massachusetts 01606

Alton F. Armington
U.S. Air Force Cambridge
Research Laboratory
Solid State Science Dept.
L. G. Hanscom Field
Bedford, Massachusetts 01730

Richard E. Balzhiser
Office of Science & Technology
Exec. Office of the President
New Exec. Office Building
Washington, D.C. 20202

Francis I. Baratta
Army Mat. & Mech. Res. Ctr.
Applied Mechanics Laboratory
Watertown, Massachusetts 02172

Michael Bass
Raytheon Company
Research Division
28 Seyon Street
Waltham, Massachusetts 02154

H. Robert Baumgartner
Norton Company
Industrial Ceramics Division
1 New Bond Street
Worchester, Massachusetts 01606

Peter Beardmore
Ford Motor Company
Metallurgy Department
Dearborn, Michigan 48121

Arden L. Bement
Massachusetts Inst. of Technology
Metallurgy & Materials Science
Cambridge, Massachusetts 02139

H. E. Bennett
Michelson Laboratory, Code 6018
Naval Weapons Center
China Lake, California 93555

Morris Berg
AC Sparkplug Division
General Motors Corporation
1300 Dort Highway
Flint, Michigan 48556

G. H. Bishop
Metals Division
Army Mat. & Mech. Res. Ctr.
Watertown, Massachusetts 02172

Rudolph A. Black
Advanced Research Projects Agency
Nuclear Monitoring Research Center
1400 Wilson Boulevard
Arlington, Virginia 22308

Joseph I. Bluhm
Army Mat. & Mech. Res. Ctr.
Mechanics Research Laboratory
Watertown, Massachusetts 02172

J. C. Bokros
Gulf Energy & Environmental Systems
Medical Products
P.O. Box 608
San Diego, California 92112

Norman L. Boling
Owens Illinois
C&TP Department
1700 N. Westwood Avenue
Toledo, Ohio 43600

Raymond T. Bratton
Westinghouse Electric Company
Materials Science Department
Pittsburgh, Pennsylvania 15235

Morris Braunstein
Hughes Research Laboratories
Chemical Physics Department
3011 S. Malibu Canyon Road
Malibu, California 90265

B. F. Brown
4430 King Avenue S.W.
Washington, D.C. 20032

Spencer H. Bush
Battelle-Northwest
P.O. Box 999
Richland, Washington 99352

J. E. Burke
General Electric Research Lab.
P.O. Box 8
Schenectady, New York 12301

Brice Carnahan
Chemical Engineering Dept.
University of Michigan
Ann Arbor, Michigan 48104

John M. Carpenter
Nuclear Engineering
University of Michigan
300 Auto Laboratory
Ann Arbor, Michigan 48104

R. J. Charles
General Electric Company
Research & Development Center
Schenectady, New York 12301

Edward V. Clougherty
Man Laboratories, Inc.
21 Erie Street
Cambridge, Massachusetts 02139

John W. Coburn
IBM K05/281
Monterey & Cottle Roads
San Jose, California 95114

Herbert T. Corten
Theoretical & Applied Mechanics
University of Illinois
Urbana, Illinois 61801

K. H. Cotter
USAF/ASD-ENFSA
C-5 Independent Structural
Review Team
Lockheed Georgia Company
Dept. 72-27, Zone 457
Marietta, Georgia 30060

Frank Dachille
Pennsylvania State University
Materials Science Laboratory
Materials Research Laboratory
University Park, Pennsylvania 16802

J. deKlerk
Westinghouse Electric Company
Research & Development Center
Churchill Borough
Pittsburgh, Pennsylvania 15235

Thomas Deutsch
Raytheon
Research Division
28 Seyon Street
Waltham, Massachusetts 02154

Derek B. Dove
Department of Metallurgy
University of Florida
Gainesville, Florida 32601

Wilfred H. Dukes
Bell Aerospace Company
P.O. Box 1
Buffalo, New York

Sabri Ergun
U.S. Bureau of Mines
Interior Department
4800 Forbes Avenue
Pittsburgh, Pennsylvania 15213

Martin A. Erickson
Ford Tractor Operations
Product Evaluation & Test
25 E. Maple Road
Troy, Michigan 48084

Anthony G. Evans
University of California
Materials Department
Los Angeles, California 90024

P. L. Farnsworth
Battelle-Northwest
Ceramics & Graphite Section
Battelle Boulevard
Richland, Washington 99352

William A. Fate
Ford Motor Company
Physics Dept., Scientific Lab.
P.O. Box 2053
Dearborn, Michigan 48121

C. E. Feltner
Ford Motor Company
Scientific Research Staff
P.O. Box 2053
Dearborn, Michigan 48121

John R. Fenter
Air Force Materials Laboratory
Materials Physics Div. (AFML/LP)
Wright-Patterson Air Force Base
Ohio 45433

Robert C. Fitzgerald
Balto Gas & Electric
Power Production
G & E Building
Baltimore, Maryland

M. C. Flemings
Department of Materials Science
Massachusetts Inst. of Technology
Cambridge, Massachusetts 02139

George W. Ford
Physics Department
University of Michigan
Ann Arbor, Michigan 48104

William C. Frazier
Westinghouse Research &
Development
Materials Science Department
Churchill Borough
Pittsburgh, Pennsylvania 15235

Karl Freed
University of Chicago
Chemistry Department
James Franck Institute
Chicago, Illinois 60637

Otto R. Gericke
Army Mat. & Mech. Res. Ctr.
Nondestructive Test Dev. Br.
Watertown, Massachusetts 02172

Concetto R. Guiliano
Hughes Research Labs.
Laser Department
Malibu, California 90265

David Goldstein
U.S. Naval Ordnance Laboratory
Code 211
White Oak, Maryland 20910

A. E. Gorum, Director
Army Mat. & Mech. Res. Ctr.
Watertown, Massachusetts 02172

Charles Grosskreutz
National Bureau of Standards
U.S. Department of Commerce
Washington, D.C. 20234

Gary L. Haller
Yale University
Engineering & Applied Science
Mason Laboratory
9 Hillhouse Avenue
New Haven, Connecticut 06520

William J. Harris, Jr.
Assn. of American Railroads
Research & Test Department
1920 L Street
Washington, D.C. 20030

Robert Hills, Jr.
Itek Corporation
Optical Systems Division
10 Maguire Road
Lexington, Massachusetts 02173

Richard W. Hoffman
Case Western Reserve University
Department of Physics
Rockefeller Building
University Circle
Cleveland, Ohio 44106

A. N. Holden
Westinghouse Electric Company
Gas Turbine Systems
Box 9175
Lester, Pennsylvania 19113

D. Holmes
Air Force Weapons Laboratory
Kirtland Air Force Base
New Mexico 87117

H. H. Johnson
Cornell University
Materials Science & Engineering
Bard Hall
Ithaca, New York 14850

R. Kammereck
Pennsylvania State University
Materials Sciences
University Park, Penn. 16802

Roy Kaplow
Massachusetts Inst. of Technology
Metallurgy & Materials Science
Cambridge, Massachusetts 02139

R. Nathan Katz
Army Mat. & Mech. Res. Ctr.
Ceramics Division
Watertown, Massachusetts 02172

Jhan M. Khan
Lawrence Livermore Laboratory
Surface Technology Laboratory
L-503
P.O. Box 808
Livermore, California 94550

Ram Kossowsky
Westinghouse Research
Material Sciences Department
Churchill Borough
Pittsburgh, Pennsylvania 15235

Jerome Kruger
National Bureau of Standards
Met. Div., Corrosion Section
B-254 - Materials
Washington, D.C. 20234

Fred F. Lange
Westinghouse R&D
Materials Science Department
Churchill Borough
Pittsburgh, Pennsylvania 15235

E. M. Lenoe
Army Mat. & Mech. Res. Ctr.
Mechanics Research Laboratory
Arsenal Street
Watertown, Massachusetts 02172

David R. Lide
National Bureau of Standards
Office of Standard Reference Data
Washington, D.C. 20234

H. A. Lipsitt
Wright-Patterson Air Force Base
Ohio 45433

Capt. John Loomis
Air Force Weapons Laboratory
Laser Division/LRO
Kirtland Air Force Base
New Mexico 87117

Harold Lurie
Director, Research and Development
New England Electric System
Westboro, Massachusetts 01581

Harold M. Manasevit
North American Rockwell
Physical Sciences Department
P.O. Box 4173 - HA 20
Anaheim, California 92803

A. F. McLean
Ford Motor Company
Turbine Research
20000 Rotunda Drive
Dearborn, Michigan 48121

Eric Mendel
IBM, CD
D/360 Crystal Growth & Process Development
Bldg. 310-52
Hopewell Junction, New York 12533

David F. R. Mildner
University of Michigan
Nuclear Engineering Dept.
300 Automotive Laboratory
Ann Arbor, Michigan 48105

Donald G. Miller
Westinghouse R&D
Material Sciences Dept.
Churchill Borough
Pittsburgh, Pennsylvania 15235

Paul C. Paris
Del Research Corporation
427 Main Street
Hellertown, Pennsylvania 18055

Elio Passaglia
National Bureau of Standards
Metallurgy Division
Washington, D.C. 20234

H. W. Paxton
National Science Foundation
Materials Research Dept.
Washington, D.C. 20550

C. D. Pears
Southern Research
2000 9th Avenue South
Birmingham, Alabama 35205

Janet S. Perkins
Army Mat. & Mech. Res. Ctr.
Watertown, Massachusetts 02172

Richard L. Pober
Man Laboratories, Inc.
21 Erie Street
Cambridge, Massachusetts 02139

James O. Porteus
Naval Weapons Center
Research Department
Code 6017
China Lake, California 93555

Harold Posen
USAF Cambridge Research Labs.
Solid State Sciences
L. G. Hanscom Field
Bedford, Massachusetts 01730

E. Neville Pugh
University of Illinois
Metallurgy & Mining Engineering
Urbana, Illinois 61801

G. D. Quinn
Army Mat. & Mech. Res. Ctr.
Watertown, Massachusetts 02172

S. Victor Radcliffe
Case Western Reserve University
Metallurgy & Materials Science
University Circle
Cleveland, Ohio 44106

Roy W. Rice
U.S. Naval Research Laboratory
Chemistry Department
Washington, D.C. 20390

David W. Richerson
Norton Company
Research & Development
1 New Bond Street
Worcester, Massachusetts 01606

John J. Schuldies
Ford Motor Company
Turbine Research
20000 Rotunda Drive
Dearborn, Michigan 48121

Mitchel Shen
University of California
Chemical Engineering Dept.
Berkeley, California 94720

Chester T. Sims
General Electric Company
Materials & Processes Lab.
Schenectady, New York 12345

George Sines
University of California
Materials Department
Engineering School
Los Angeles, California 90024

M. J. Sinnott, Director
Materials Sciences Office
Advanced Research Projects
Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

Gerald Smolinsky
Bell Telephone Laboratories
Murray Hill, New Jersey 07974

Marion J. Soileau, Jr.
Air Force Weapons Laboratory/LRO
Kirtland Air Force Base
New Mexico 87117

Roger W. Staehle
Ohio State University
Metallurgical Engineering Dept.
116 W. 19th Avenue
Columbus, Ohio 43210

Ferdi B. Stern, Jr.
Magnaflux Corporation
25 W. 43rd Street
New York, New York 10036

Eli Sternberg
California Inst. of Technology
Div. of Eng. & Applied Science
Pasadena, California 91109

S. S. Sternstein
Rensselaer Polytechnic Institute
Materials Division
Troy, New York 12181

C. Martin Stickley
Deputy Director
Materials Sciences Office
Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

Jacob J. Stiglich, Jr.
Boride Products, Inc.
2879 Aero Park Drive
Traverse City, Michigan 49684

D. H. Stone
Assn. of American Railroads
Research Department
3140 S. Federal Street
Chicago, Illinois 60616

Karsten H. Styhr
Ford Motor Company
Turbine Research
20000 Rotunda Drive
Dearborn, Michigan 48121

Alan S. Tetelman
University of California
Materials Department
6531 Boelter Hall
Los Angeles, California 90024

Donald O. Thompson
North American Rockwell
Science Center
Materials Department
1049 Camino Dos Rios
Thousand Oaks, California 91360

Garth L. Tingey
Battelle Northwest Laboratory
Fuels & Materials Department
P.O. Box 999
Richland, Washington 99352

H. H. Uhlig
Massachusetts Inst. of Technology
Metallurgy & Materials Science
Cambridge, Massachusetts 02139

E. C. van Reuth
Naval Ship Res. & Dev. Lab.
Propulsion Department
Annapolis, Maryland 21402

Francis L. VerSnyder
United Aircraft Corporation
Pratt & Whitney Aircraft Div.
Materials Engr. & Res.
400 Main Street
E. Hartford, Connecticut

Louis H. Von Ohlsen
Bell Telephone Laboratories
Dept. 2143
2525 N. 11th Street
Reading, Pennsylvania 19604

John L. Vossen
RCA Laboratories
Process Research Laboratory
Princeton, New Jersey 08540

John E. Wachtman, Jr.
National Bureau of Standards
Inorganic Materials Division
Washington, D.C. 20234

Richard Weiss
Army Mat. & Mech. Res. Ctr.
Watertown, Massachusetts 02172

E. T. Wessel
Westinghouse Electric Company
Research and Development Center
Churchill Borough
Pittsburgh, Pennsylvania 15235

Thomas J. Whalen
Ford Motor Company
Scientific Research Staff
P.O. Box 2053
Dearborn, Michigan 48121

Sheldon Wiederhorn
National Bureau of Standards
Building 223, Room A359
Washington, D.C. 20234

James O. Wilkes
University of Michigan
Chemical Engineering
Ann Arbor, Michigan 48104

Frits Zernike
Perkin-Elmer
Research Department
Norwalk, Connecticut 06856

THE LINE TENSION OF A CRACK LINE, AND ITS APPLICATION TO KINKS ON CRACKS

R. M. Thomson

Abstract

We investigate the definition of an effective line tension for a crack line similar to that used for dislocations. Although cracks have a distinct tendency to wipe out regions of higher curvature relative to straighter configurations, the effective line tension for a crack is not easy to separate from the crack extension force, which is area dependent. We are able to define a tension term, however, which must be used with discretion, and which is dependent on the size of the curved region.

THE LINE TENSION OF A CRACK LINE, AND ITS APPLICATION TO KINKS ON CRACKS

R. M. Thomson

Introduction

Previous work[1,2] has introduced the concept that in a brittle material where plasticity effects at a crack tip can be neglected, and where the cohesive region at the crack tip is small, a crack will experience trapping by the lattice. The effect is analogous to the Peierls energy of the dislocation. It has been shown[2] that in cases of this sort there is a region about the Griffith crack depth where thermal creep of the crack should be observable. In this creep region, the thermally activated process will depend upon forming kinks on the crack line, Fig. 1, whose energy will be the dominant physical parameter. The shape of a single kink has been studied in a computer calculation which simulates the iron lattice[3], but we would like to develop a qualitative model of kinks built on the sort of quasi-continuum considerations which are familiar for dislocations.

The model we adopt for the calculation is a quasi-continuum in which the effect of the discrete lattice is taken over into a continuum description in the form of an oscillating term in the surface energy of period equal to the lattice

constant.[1,2] This oscillating form gives rise to a washboard potential representing the atom rows. We shall further simplify the picture by assuming as in Fig. 1, that the washboard is sensitive to the atom rows lying in the y-direction, but not in the x-direction.

Physically, the kink shape will be determined by a balance struck between minimizing the additional surface energy of the portion of the crack which lies on top of the substrate washboard potential and the elastic energy of the kink. We shall first examine the elastic energy of small fluctuations on an otherwise straight line, and from this derive an expression for the line tension of the crack. From this, the energy of the kink will be estimated and the variable activation energy will be directly demonstrated which has been suggested on heuristic grounds.[2]

Energy Release of a Curvilinear Crack

We must begin by writing down a prescription for calculating the total elastic energy change of a curvilinear crack as it is deformed into a new shape. The result is a simple generalization of Irwin's energy release formula for two dimentions as reviewed, say by Sih and Liebowitz.[4] We envision a small element of crack line, and set up a local coordinate system as shown in Fig. 2. Then the force of the line or energy release rate is given by

$$G_I dx = \lim_{\delta \to o} \frac{1}{\delta} dx \int_o^{\delta} \tfrac{1}{2}\sigma_z\,(\delta-B,O)\; u_z\,(O,B,O) \qquad (1)$$

In the local coordinate system, we can write

$$u_z(x,y,z) = u_z(x,r,\theta) = \frac{K}{\mu}\left(\frac{2r}{\pi}\right)^{\frac{1}{2}}(1-\nu)f(\theta)$$

$$\sigma_z(x,y,z) = \sigma_z(x,r,\theta) = \frac{K}{(2\pi r)^{\frac{1}{2}}}\, g(\theta) \qquad (2)$$

$$f(\theta=0) = 1$$

$$g(\theta=\pi) = 1$$

Substituting (2) into (1), we get the familiar form

$$G_1 = \frac{(1-\nu)^2}{E} K_1^{\,2} \qquad (3)$$

The assumption made in writing (1) is that no displacement but u_z is generated, which is strictly true for the plane elliptical crack, and for infinitesimal displacement at the origin of x. Although lacking a rigorous proof for other cracks, we rely on St. Venants principle to generate a stress field with the general functional form of a straight crack in the local vicinity of the crack line (though with a different K). For the work done in changing the shape of the original line, the elastic energy change ΔE_{el} must be integrated over the entire line

$$\Delta E_{el} = -\int G dS \qquad (4)$$

where S is the area generated by the crack line as it moves

from its original configuration to its final configuration, Fig. 3. In using (4), one must of course take account of the functional form of G, as it depends upon the shape of the total crack line, and upon its spatial variation within S. G is thus not a simple point function, but also a functional of the line shape because the stress intensity factor depends upon the total shape of the line, and other external geometric factors. It is this functional quality of G and K which makes our task difficult.

Application to the Griffith Crack

We now apply (4) to the case of a straight crack line lying in one of the atomic troughs of Fig. 1, as it is deformed into a kinked line with the nose of the kinked lying in the adjacent trough. To calculate the total energy change as the kink is formed, we must add surface energy to (4).

$$\Delta E_{total} = -\int G dS + 2\int \gamma_0 dS + \int a\gamma_1 (1-\cos 2\pi y/a) d\ell \qquad (5)$$

The surface energy in (5) is assumed to be made up of two parts, a normal Griffith term strictly proportional to the new surface generated, and a second term which reflects the washboard potential on which the crack sits. For a straight crack, one can define[1] a surface energy as a function of the crack position which takes the form

$$S(y) = 2\gamma_0 y + P(y) \qquad (6)$$

where P(y) is a periodic function in the lattice. We interpret

(6) to mean that for a segment $d\ell$ of a crack lying at position y on the washboard, then an amount $P(y)d\ell$ is added to the energy. (We appreciate that the similtude of this quasi-continuum model to the true discrete model is qualitative, but believe that the quasi-continuum model possesses enough of the physical truth to serve as a mental crutch in predicting the actual kink behavior.)

A zeroth approximation to the kink which has been suggested by Eshelby[5] is to use the value for G in (5) appropriate to the straight crack line. Then Eq. (5) becomes

$$E_{total} = (2\gamma_o - G) \int dS + \int \gamma_1 a(1-\cos 2\pi y/z)d\ell \qquad (7)$$

The boundary conditions for finding the minimum of E_{total} in (7) are that all shapes of line are investigated where the tip of the kink is pegged at the bottom of the trough $y_o + a$ in Fig. 1. Obviously, in this approximation, the minimum configuration is that where the kink is an abrupt spike of zero area.

The result is unphysical, because the Eshelby approximation says that all planar crack shapes which have the same area also have the same energy. We know that cracks tend to smooth out small irregularities, and we now investigate a case where rigorous solutions exist in order to establish the next higher order approximation to (5) and (7).

Line Tension of Cracks

Exact solutions exist for the three dimensional stress and displacements of the elliptical planar crack[2], and we use these solutions to establish the part of the line energy of a crack which depends on the line length as well as the cracked area.

We begin with a circular planar crack (so called penny shaped crack), and elongate it so that it becomes an ellipse with major and minor axes a and b respectively. The minor axis, b will remain the same as the original radius of the circle as shown in Fig. 4. We then compare the energy change calculated by Eq. (4) with the energy to expand the original penny shaped crack to a larger circle whose area is the same as the ellipse. Since the cracked areas are the same, any difference in energy must be associated with the length of the line, and can be converted into an effective line tension of the elliptical crack line.

In terms of the stress intensity factor for mode I, (4) becomes

$$G = \frac{1-\nu^2}{Y} \int K^2 dS \qquad (8)$$

Y is Young's modulus, ν the Poissons ratio, and K is the mode I stress intensity factor. K is given by

$$K_{ellipse} = \frac{\sigma}{\Phi(k)} \sqrt{\frac{\pi b}{a}} \left(a^2\sin^2\theta + b^2\cos^2\theta\right)^{\frac{1}{4}}$$

$$\Phi(k) = \int_0^{\pi/2} (1-k^2\sin^2\beta)^{\frac{1}{2}} d\beta \tag{9}$$

$\Phi(k)$ is the complete elliptic integral, σ is the uniform tensile stress at ∞. We now assume that the ellipticity is small, and expand K in terms of the small parameter, ε,

$$\varepsilon = \frac{a-b}{a} \tag{10}$$

Then

$$K_{ellipse} = 2\sigma\sqrt{\frac{b}{\pi}} \left[1 + \frac{\varepsilon}{2}(1-\cos^2\theta)\right] \tag{11}$$

To second order in ε,

$$\Delta E_{el} = -\int G dS = -\frac{4\sigma^2 b^3 (1-\nu^2)}{Y} \varepsilon\left(1 + \frac{\varepsilon}{2}\right) \tag{12}$$

The first order term in ε is precisely the Eshelby approximation, as one finds by substituting the expression $G = 4\sigma^2(1-\nu^2)b/\pi Y$ into the elastic part of (7) and noting that the increment in area from the initial circle to the larger ellipse is $\pi b^2\varepsilon$ to first order in ε.

The radius, r_1, of an expanded circle whose area is equal to the ellipse is given by

$$r_1 = \frac{b}{\sqrt{1-\varepsilon}} \tag{13}$$

and from (9), the stress intensity factor of the circle is

$$K_{circle} = 2\sigma\sqrt{\frac{r}{\pi}} \tag{14}$$

With substitution of (14) in (4), we get for the energy change of the circle,

$$\Delta E_{el} = -\int GdS = -\frac{4\sigma^2 b^3 (1-\nu^2)}{Y} \varepsilon\left(1 + \frac{5}{4}\varepsilon\right) \tag{15}$$

Comparing (15) with (12) we have

$$\Delta E_{ellipse} - \Delta E_{circle} = \frac{3}{2} \frac{\sigma^2 b^3 \varepsilon^2 (1-\nu^2)}{Y} \tag{16}$$

The strain energy released by the ellipse is smaller than that of the circle for the same area of crack, showing that some of the stored energy is retained in the additional line length of the ellipse.

Writing the excess line length in the ellipse as compared with the circle of equal area,

$$\Delta \ell = \ell_{ellipse} - \ell = \frac{3}{16} \pi b \varepsilon^2 \tag{17}$$

We note that both the excess energy and excess line length depend upon the square of ε, and we can thus define a line tension,

$$\mathcal{E} = \frac{\Delta E_{ellipse} - \Delta E_{circle}}{\Delta \ell} = \frac{8\sigma^2 b^2 (1-\nu^2)}{\pi Y} \tag{18}$$

This result must be interpreted before it can be applied to the kink problem because one does not know in that case what value to write for b. However, the results, (18), are understandable on the basis of another argument which, though more qualitative, allows one to transfer (18) to the kink problem directly.

If we have two crack configurations in which one configuration differs only in a local region from the other, Fig. 5, the far fields of the two configurations will be the same. However, the measure of the difference in elastic energies of the two can be estimated by comparing the integrated strain energies over the regions where they differ. Thus (Fig. 5),

$$\Delta E = \int W_1 dV - \int W_2 dV$$

$$\simeq \frac{1}{2\mu} \int \frac{K_1^2}{2\pi r} 2\pi r \, dr \, d\ell_1 - \frac{1}{\mu} \int \frac{K_2^2}{2\pi r} 2\pi r \, dr \, d\ell_2$$

$$\simeq \frac{1}{2\mu} K^2 \ell_k \, \Delta\ell_k$$

$$\simeq \frac{2}{\pi\mu} \sigma^2 R \ell_k \Delta\ell_k \qquad (19)$$

In this expression, ℓ_k is the length of the kink, which is also the radius over which the volume integral is taken around the crack line, $\Delta\ell_k$ is the difference in line length between the two crack configurations, and R is the total linear dimension of the cracked region. We have used the K appropriate to a penny shaped crack line of radius R. The crack line tension is then roughly

$$\mathcal{E} = \frac{\Delta E}{\Delta \ell} \simeq \frac{2\sigma^2 R \ell_k}{\pi\mu} = \frac{4\sigma^2 R \ell_k (1+\nu)}{\pi Y} \qquad (20)$$

Equations (20) and (18) are seen to be essentially the same if we make the physically reasonable assumption in (18) that the dimension of the kink is the diameter of the circle in Fig. 4,

$\ell_k = 2b$. Eq. (20) is now an equation which can be applied to general kink configurations.

Kink Energy

We now wish to estimate the kink energy in the terms of the last section. We write the elastic energy to form the kink from the straight line from Eq. (5) in the form

$$(\Delta E_{el})_{kink} = -\int G \, dS \tag{21}$$

where the integral is over the surface generated when the kink is formed from the straight crack. We have already noted that (21) is a functional of the total shape of the kink line and the external geometry. We have also remarked how the Eshelby approximation leads to unphysical results, and that the crack line shows some of the attributes of a line tension. We are thus led to postulate an approximation to the complicated form (21) by

$$(\Delta E_{el})_{kink} = -G_o \int dS + \mathcal{E} \left[\int d\ell\right]^{kink}_{straight\ crack} \tag{22}$$

G_o is the strain energy release function for the straight crack, and $\mathcal{E}$ is an appropriate line tension.

We find, however, that this approximation is very tenuous. It can lead to inaccuracies when both area and length change are of the same order of magnitude in some geometrical parameter. For example, for the case of the circle which is increased in radius from r to $r(1 + \varepsilon)$, the correct result does not involve

a length change contribution in the first order of ε. However, when configurations are being compared which have the same area, line tension-like forces come into play in higher order which straighten out the crack. In our use of (22) we shall be careful to keep the effects of first order area changes and higher order length changes separated.

The total crack energy change with a kink is from (5) now written

$$\Delta E_{kink} = (2\gamma_o - G_o)\delta A + \mathcal{E}\delta\ell + \int \gamma_1 (1-\cos 2\pi y/a)\, d\ell \qquad (23)$$

Because our estimate already contains a considerable degree of approximation in the use of a line tension, we shall consider only saw-toothed kinks because of the simplicity of the calculations, as shown in Fig. 6. We shall also assume that the crack is near its Griffith equilibrium depth, y_g, where

$$G_o = 2\gamma_o + g \qquad (24)$$

and where g represents a small increase in K over the continuum equilibrium value. In this case, we are once more dealing with a situation where the line tension is the major term, and we shall use the line tension value calculated in the last section, Eq. (20). Equation (23) becomes

$$\Delta E = -g\delta A + \mathcal{E}\delta\ell + \int a\gamma_1 \left(1-\cos\frac{2\pi y}{a}\right) d\ell \qquad (25)$$

$$\delta A = a^2 \cot\theta$$

$$\delta\ell = 2a\,\frac{\sin^2\theta/2}{\sin^2\theta}$$

in terms of the parameters shown in Fig. 6. Also,

$$\ell = \frac{4\sigma^2 R(1+\nu)}{\pi Y} \frac{a}{\sin\theta} \simeq \frac{\gamma_o a}{\sin\theta} \tag{26}$$

where we have related the crack size to the surface energy γ_o by the Griffith relation. Then

$$\Delta E = -ga^2 \cot\theta + 4\gamma_o a^2 \frac{\sin^2\theta/2}{\sin^2\theta} + \frac{2\gamma_1 a^2}{\sin\theta} \tag{27}$$

Finding the minimum kink configuration, $\frac{d}{d\theta}(\Delta E) = 0$ gives

$$\frac{\gamma_1}{\gamma_o} = \frac{\sin\theta}{\cos\theta} - 4\frac{\sin^2\theta/2}{\sin\theta} + \frac{g}{2\gamma_o \cos\theta} \tag{28}$$

The expressions for ΔE and $\frac{\gamma_1}{\gamma_o}$ both contain the external stress in the form g, and thus give the dependence of the activation energy on the external stress. For small angles, and in the limit $g << \gamma_1$ we have

$$\frac{\gamma_1}{\gamma_o} = \frac{\theta^3}{4} + \frac{g}{2\gamma_o} \tag{29}$$

$$\Delta E \sim 4\gamma_o a^2 (1 + \frac{\theta^2}{4}) - \frac{ga^2}{\theta}$$

For large angles, of course, the limit becomes

$$\Delta E \sim \gamma_1 a^2 - ga^2 \tag{30}$$

In the limit of small external stresses, the expected linear behavior on stress is obtained with an effective activation volume as given by (29). For larger stresses near the point where the activation energy for kink creation is cancelled

by the stress activation, the stress dependence is quite complicated. This is of course also just the region where the line tension approximation breaks down.

The expression for the kink energy then derived does not vary significantly from the value

$$\Delta E \simeq \gamma_1 a^2 \tag{31}$$

as the equilibrium shape of the crack varies from smooth to abrupt.

These results indicate that crack creep should be observable in the cases where lattice trapping is important, and qualitatively confirm the earlier picture that during the creep regime, the activation energy of the creeping crack decreases as the crack moves through the trapping region giving rise to an acceleration of the crack as a function of time.

Conclusions

We have shown that the concept of line tension as applied to cracks in solids is not as useful as it has been in the theory of dislocations. Although cracks do have a distinct tendency to wipe out regions of high curvature relative to straighter regions, the effective line tension is not easy to separate from the crack extension force term which is an area dependent term. Nevertheless, we have found an approximate and intuitively handy line tension expression to use in cases of near equilibrium where the area terms are small. In this case, the crack line tension force

is proportional to the extent of the kink region itself in an intuitively reasonable way.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253, with the University of Michigan.

References

1. R. Thomson, C. Hsieh, and R. Rana, J. Appl. Phys. 42 3154 (1971).

2. C. Hsieh, R. Thomson, *Lattice Theory of a Fracture Crack and Crack Creep*, to be published.

3. M. F. Kanninen and P. Gehlen, Intern. J. Fract. Mech. 7, 471 (1971).

4. G. C. Sih and H. Liebowitz, *Fracture, Vol. II*, p. 67 (1968) by Liebowitz, Academic Press, New York.

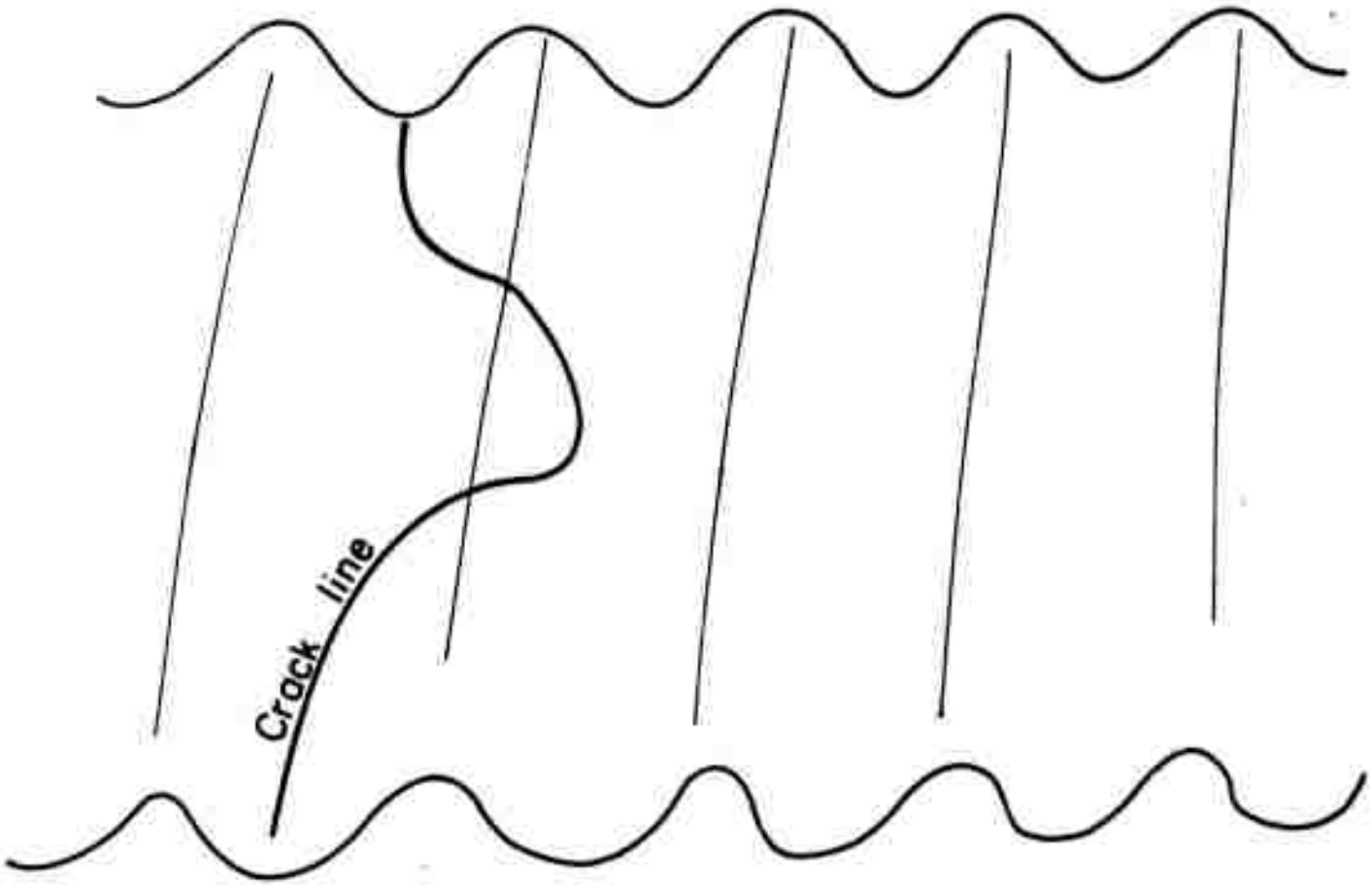


Figure 1. A kink in a crack line on an atomic substrate.

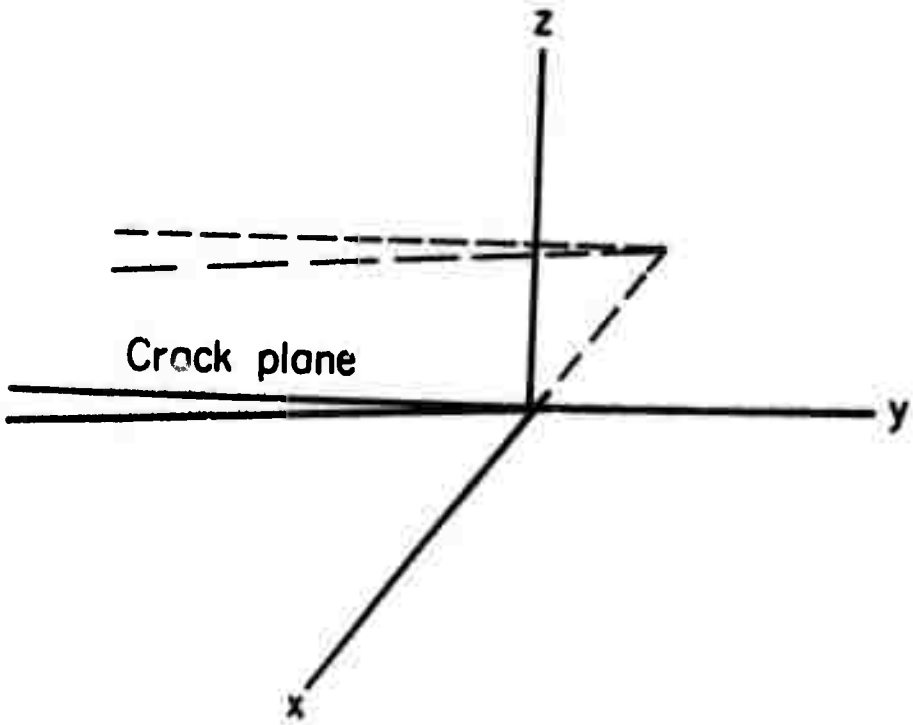


Figure 2. Coordinate system used for describing the expansion of a crack.

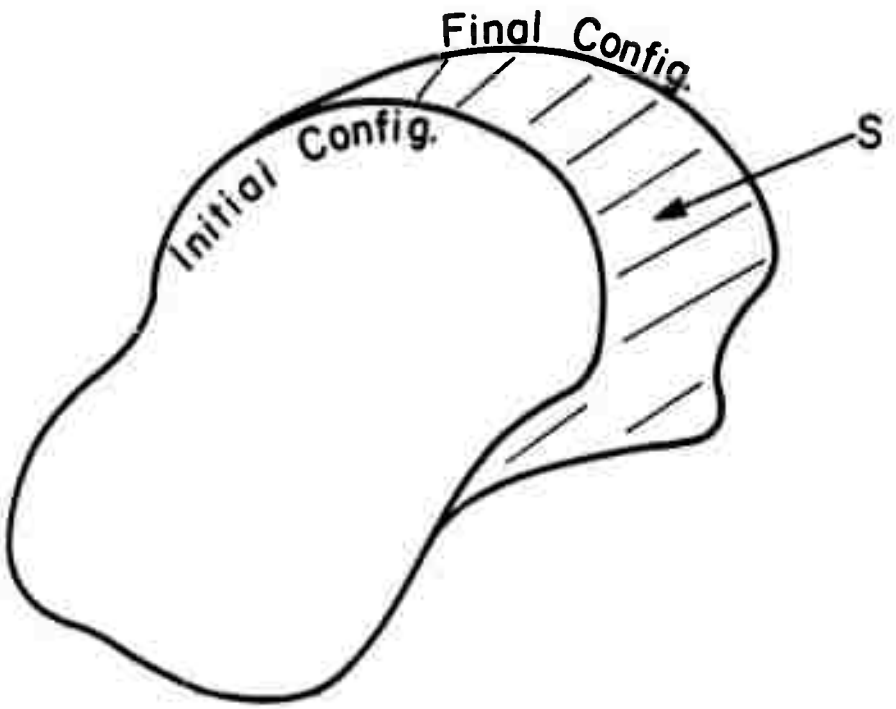


Figure 3. Initial and final crack configurations for Eq. (4).

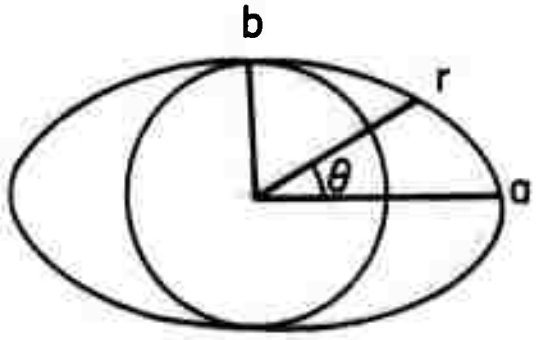


Figure 4. Expansion of a circular crack into an elliptical crack whose minor axis equals the radius of the original circle.

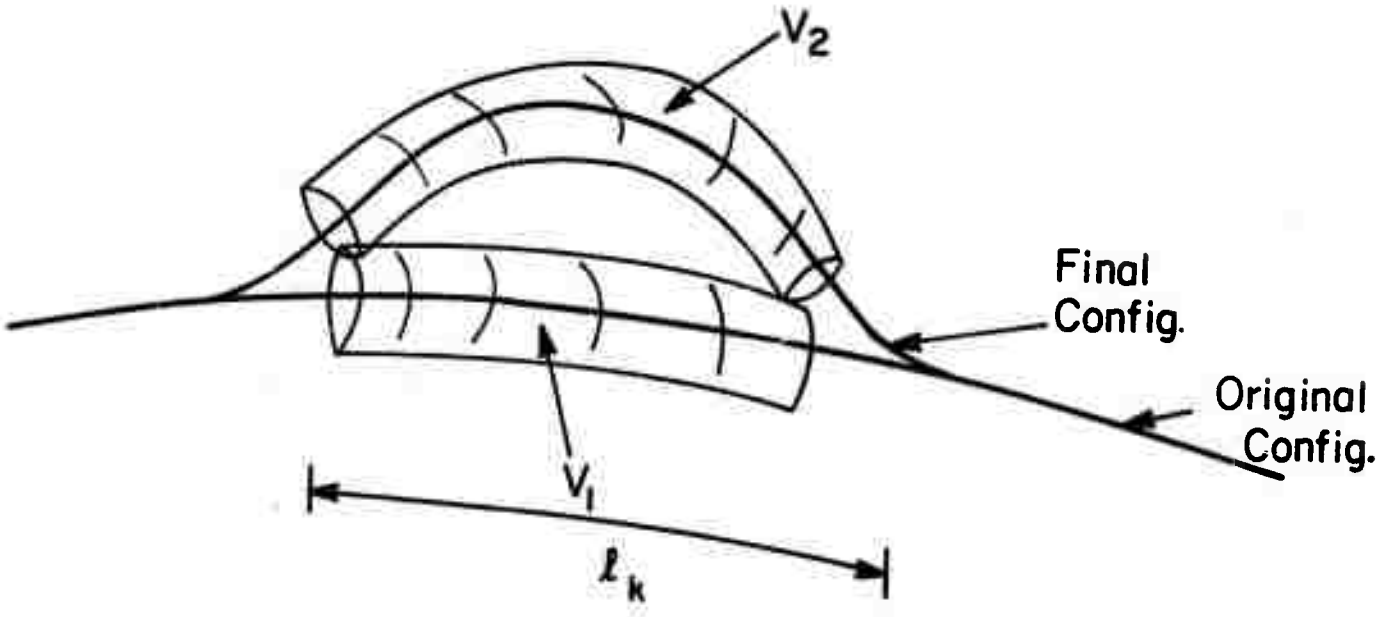


Figure 5. Local excursion of a crack with integration volumes indicated referring to Eq. (19).

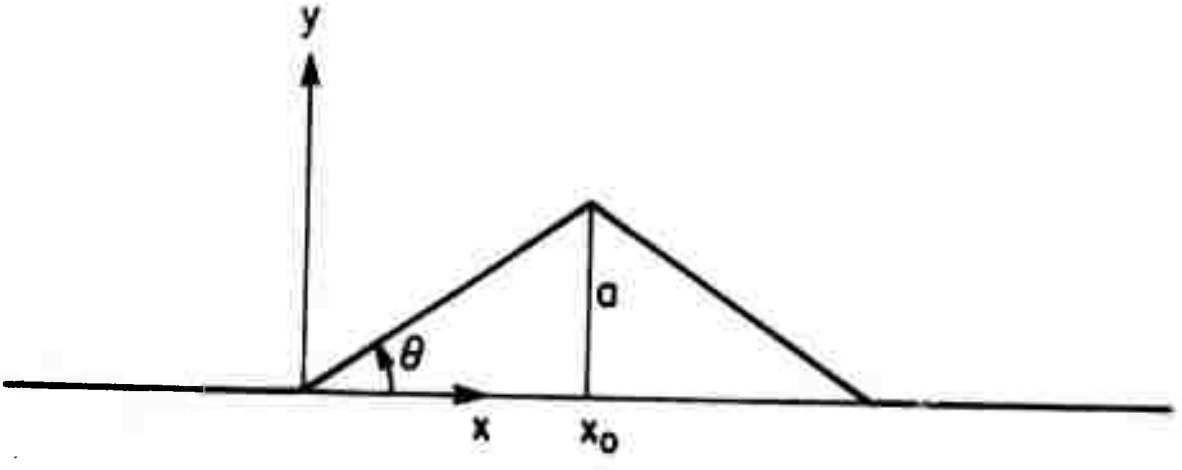


Figure 6. Schematic sawtooth kink with parameters as used in text.

RANDOMNESS AND WAVE PROPAGATION IN INHOMOGENEOUS MEDIA

J. A. Krumhansl

Abstract

The study of elastic wave propagation in composite media has led to the development and application of various methods of analysis appropriate to periodically inhomogeneous materials. In general these methods cannot be extended exactly to randomly inhomogeneous media. Nonetheless, some progress has been made during recent years in the analysis of electromagnetic waves in random media, and in solid state physics in the description of disordered materials. It appears that these methods can be used for disordered composites; this paper will survey the methods and indicate a few applications.

RANDOMNESS AND WAVE PROPAGATION IN INHOMOGENEOUS MEDIA

J. A. Krumhansl

I. INTRODUCTION

The study of composite materials has provided an opportunity for participants from many disciplinary backgrounds. The problem of wave propagation in the elastic limit is a case in point. The solid state physicist has dealt extensively with wave propagation in periodically homogeneous media (this is the essence of describing electron motion in crystal lattices) and has developed powerful formal and approximate methods to deal with such problems. This background has been of advantage in the study of elastic wave propagation. However, one is soon reminded that the vector fields of elasticity require considerably more care mathematically than is characteristically true of scalar Schrödinger waves, and finds that the continuum mechanician faces formidable difficulties when attempting to develop computationally or descriptively practical methods to deal with real situations.

Fortunately, in periodically inhomogeneous elastic media reasonable progress has been made during the last few years toward development of methods similar to those used in solid state physics.

Thus encouraged, and with the concurrent activity by physicists on disordered (either substitutionally or structurally) materials, it seems timely to consider whether methods found useful for describing electron and vibrational waves in disordered media can be applied to elastic wave propagation in disordered composites.

This paper suggests how this approach might be so developed. It is only a beginning; we hope that it will lead to future applications. It is likely that certain general features of wave propagation in random alloys - at least for homogeneous randomness - will also hold true in disordered composites. References for background are to be found in mechanics,[1] mathematical physics,[2] and solid state physics.[3,4]

There are two different physical limits, depending on whether the effective wave length of the "average wave" one studies is much larger than the scale (i.e., correlation length) of inhomogeneities, or vice versa. In the former case a well defined average wave propagates, with damping; in the latter case, a local ray approximation can be expected to lead to diffusive scattering.

Two types of disorder may be encountered: The medium may have periodic geometric structure but the material parameters (e.g., fibers) vary randomly from one unit cell to the next. More generally, also much less tractable, the geometric structure may not have long range order, and is only readily characterizable by a heirarchy of correlation functions. To the writer's

knowledge, at least in the literature of solid state physics, only recently have any partially successful analyses[5,6] been carried out on this latter case (frequently called "amorphous disorder" to distinguish from "substitutional disorder").

Finally, one must remember that the case of disordered composites requires an extension of the discussion of weak fluctuations from a uniform medium; that is, one is already starting out with significant periodic variation of material parameters in the ordered composite, and randomness is then added to the problem. Thus, for substitutionally disordered composites we substitute Fourier-Floquet methods for Fourier plane wave methods. No such simple extension is possible for amorphous disorder which is the principal reason for the difficulty already noted. In the present paper, formal considerations will be confined to substitutionally disordered composites, based largely on methods used in solid state for substitutionally disordered alloys.

II. OUTLINE OF GENERAL FORMALISM

The statement of the dynamics of atom motion in solids, linear elastic waves, electromagnetic waves, and electron waves can by suitable representation be cast in the form

$$\hat{L}\phi = \hat{T}\phi + \hat{F}$$

where $\hat{L}$ and $\hat{T}$ are linear operators, spatial and temporal respectively, and ϕ is a field; the field may be scalar or vector,

$\hat{L}$ and $\hat{T}$ will be appropriate dyads; and $\hat{F}$ is a generalized applied force. The description of physical phenomena is obtained from a knowledge of the fields ϕ, and the "spectrum" of the operators. The operator manipulations will be defined later, in the context of applications at hand; let it be presumed at this point that standard operator algebra is legitimate.

The operator $\hat{T}$ usually has a simple form (e.g., $\rho(x)\partial^2/\partial^2 t^2$), so we take advantage of the linearity to construct solutions

$$\phi(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{i\omega t}\, \Psi(\omega,t) \tag{1}$$

whence, typically, one obtains the transformed equation

$$[\hat{L} + \rho(x)\omega^2]\Psi(\omega,x) = \hat{f}(\omega,x,t=0) \tag{2}$$

where f depends on applied forces and initial conditions.

Formally this problem is solved by the Green's function (operator) $[\hat{L} + \rho(x)\omega^2]^{-1} = \hat{G}$, which satisfies

$$[\hat{L} + \rho(x)\omega^2]G = \delta(x-x') \tag{3}$$

where δ is the Dirac function in a continuum, or Kronecker on a lattice. The solution is obtained by

$$\Psi(\omega,x) = \hat{G}\hat{F} \tag{4}$$

followed by application of equation (1).

The practical problem, of course, is to construct the

Green's function. The eigenfunctions of L, enumerated by index k, and the spectrum can be found of such a form that

$$\hat{L}\, u(k,x) = -\lambda_k \rho(x)\, u(k,x) \tag{5}$$

For a uniform medium $u(k,x) = e^{ikx}$ (plane waves). For a perfectly periodic medium $u(k,x) = e^{ikx} v_\alpha(k,x)$ where $v_\alpha(k,x)$ has the period of the medium, i.e., Floquet solutions, which fall into bands (index α) through which the eigenvalues vary continuously with k covering a "Brillouin zone." The $\lambda_{k,\alpha}$ are real, and $u_\alpha(k,x)$ may be orthonormalized:

$$\int dx \rho(x)\; u_\alpha^*(k,x) u_{\alpha'}(k',x) = \delta_{\alpha\alpha'} \delta(k-k')$$

$$\sum_\alpha \int_{B.Z.} dk \rho(x)\; u_\alpha^*(k,x) u_\alpha(k,x') = \delta(x-x') \tag{6}$$

where $\delta_{\alpha\alpha'}$ and $\delta(k-k')$, $\delta(x-x')$ are Kronecker or Dirac delta-functions. From this property transform relations are developed

$$f(x) = \sum_\alpha \int dk\; F_\alpha(k)\; u_\alpha(k,x)$$

$$F_\alpha(k) = \int dk\; \rho(x)\; u_\alpha^*(k,x) f(x) \tag{7}$$

which have been applied by Lee et al. in the discussion of transients in periodic composites. The Floquet solutions also provide a specific basis for representation of the Green's operator in x-space:

$$\hat{G}(x,x';\omega^2) = \sum_\alpha \int dk\; u_\alpha(k,x)\, (\omega^2 - \lambda_{\alpha,k})^{-1} u_\alpha^*(k,x') \tag{8}$$

Although $\hat{G}$ is useful to obtain the displacement fields from initial conditions and applied forces, it also contains useful information in itself. That is, considered as a function of ω, $\hat{G}$ has poles at all of the resonances $\lambda_{\alpha,k}^{\frac{1}{2}}$. This fact is of use even in a randomly perturbed system since the average $\hat{G}$, denoted by $\langle\hat{G}\rangle$, by virtue of the linear operation of averaging, still contains the spectrum. Many of the solid state applications of this method to both periodic and randomly perturbed periodic systems have been carried out in recent years, aided particularly by modern computer capability.

III. APPLICATION TO RANDOMLY PERTURBED COMPOSITES

We wish to find $\hat{G}$ in a system whose parameters vary in a random manner from the periodic composite. For definiteness one may envisage in the following development that we have a set of fibers (or layers) whose elastic parameters or density vary randomly about some average value; thus

$$[\hat{L}+\rho(x)\omega^2] = [\hat{L}_o+\rho_o(x)\omega^2] + [\Delta\hat{L}+\omega^2\Delta\rho(x)]$$

$$= [\hat{L}_o+\rho_o(x)\omega^2] - \hat{D} \qquad (9)$$

The operator $\hat{D}$, describing the random deviation from periodicity, would have the form $\hat{D} = \sum_i \hat{D}_i(x)$ with i indicating the ith fiber in our illustrative example. If the variations are statistically independent from fiber to fiber then the distribution of $D_i(x)$ would be given by some single site probability function $p_i(D_i)$; if for some processing reason the deviations of fibers were

correlated, a joint distribution $p_{ij}(D_i,D_j)$ would characterize the structure, and so on for higher order correlations. We will restrict considerations to simple statistical distributions. A single site average is then computed by

$$\langle D_i \rangle = \sum_{D_i} p_r(D_i) D_i \tag{10}$$

A formal operator algebraic solution for G may be found in terms of $\hat{G}_o = [\hat{L}_o + \rho_o \omega^2]^{-1}$. From Eqs. (3), (7), and $\hat{G}_o$ defined, it follows that

$$\hat{G} = \hat{G}_o + \hat{G}_o \hat{D} \hat{G} \tag{11}$$

This equation has the formal solution

$$G = \frac{1}{(1-\hat{G}_o \hat{D})} \hat{G}_o \tag{12}$$

with the series expansion

$$\hat{G} = \hat{G}_o + \hat{G}_o \hat{D} \hat{G}_o + \hat{G}_o \hat{D} \hat{G}_o \hat{D} \hat{G}_o + \ldots . \tag{13}$$

Equations (12) and (13) are formally exact solutions, depending on convergence (which can be established only by going to a definite representation, such as the Floquet basis).

There are two distinct ingredients remaining in carrying on the problem from here: First, ways of solving (approximately in almost all cases) the operator equations; second, the determination of statistically averaged forms of the Green's function. Both of these are needed in actual application, and a series of

increasingly more realistic approximations have been developed for treating random alloys. The lower order approximations and averages can be found as follows.

From (11) the lowest order statistical approximation to

$$\langle G \rangle = \hat{G}_o + \hat{G}_o \langle \hat{D}\hat{G} \rangle \tag{14}$$

is to neglect the statistical non-independence of $\hat{D}$ and $\hat{G}$ (motivated by noting that many interactions usually occur in $\hat{G}$ before coming back to a previous site) and thus decoupling

$$\langle \hat{D}\hat{G} \rangle \simeq \langle \hat{D} \rangle \langle \hat{G} \rangle \tag{15a}$$

whence

$$\langle \hat{G} \rangle \simeq G_o + G_o \langle D \rangle \langle \hat{G} \rangle \tag{15b}$$

A next higher approximation may be found by collecting some more terms in the series Equation (13)

$$\hat{G} + \hat{G}_o + \hat{G}_o [\hat{D} + \hat{D}\hat{G}_o\hat{D}]\hat{G} \tag{16a}$$

with the statistical "decoupling"

$$\langle \hat{G} \rangle \simeq \hat{G}_o + \hat{G}_o [\langle \hat{D} \rangle + \langle \hat{D}\hat{G}_o\hat{D} \rangle] \langle \hat{G} \rangle \tag{16b}$$

It is intuitively apparent that while the averaging $\langle D \rangle$ demands the single fiber probability distribution, the next quantity $\langle DG_oD \rangle$ will generally require the two fiber probability distribution (correlation function).

These are the lowest order of a heirarchy of possible

"averaging-decoupling" schemes, and correspond closely to the average wave approximation studied by Karal and Keller.[2] Depending on the particular problem, approximations exact to much higher order[3,4] have been developed to discuss the lattice vibrations of disordered alloys or electronic properties of liquid metals; however, it is not apparent that these specialized techniques are needed at this present evolution of the analysis of composites. Therefore, let us conclude by putting the above in specific algebraic form. We can generally make one obvious reduction; for statistically homogeneous media $\langle D \rangle$ is independent of site, say $\langle D \rangle = D_o$, so one might just as well include D_o directly in the redefined quantities $[L_o + \rho_o\omega^2 - D_o] = (G_o)^{-1}$, $D \to (D-D_o)$. Then $\langle D-D_o \rangle = 0$, so that Equation (16b) becomes (with the redefined G_o and D)

$$\langle G \rangle \simeq G_o + G_o[0 + \langle (D-D_o)G_o(D-D_o \rangle] \langle G \rangle \tag{16c}$$

IV. ALGEBRAIC REDUCTION OF OPERATOR EQUATIONS

We continue the example of a composite whose fibers are regularly spaced but have randomly varying material parameters. Algebraically the operators are generally non-diagonal in x, that is, $G \equiv G(x,x')$. (However, we have made the assumption that material operators such as L, ρ, D are local). Similarly, operator products are generally integral operations, e.g.,

$$G_oDG = \int dx'' G_o(x,x'')D(x'')G(x'',x') \tag{17}$$

whence the operator calculations generally involve integral

equations, particularly in x-coordinate space, and the manipulation of the various equations is not easy.

If, however, we transform from x-space to α, k space by the recipe that an operator A(x,x') transforms

$$\iint dx dx' \rho_o(x)\, u^*_\alpha(k,x) A(x,x') u_{\alpha'}(k',x') \rho_o(x') = A_{\alpha,\alpha';k,k'} \tag{18}$$

Then, as in Fourier methods, we can sometimes convert the integral relations to simple algebraic equations. For example, from the definition of G_o, Equations (7) and (8) it follows by orthonormality that

$$G_{o\alpha\alpha';kk'} = \delta_{\alpha\alpha'}\delta(k-k')(\omega^2-\lambda^o_{\alpha,k})^{-1} \tag{19}$$

This is the usual result - the Green's function in eigenfunction representation is "diagonal".

Next consider <D> in α,k representation. This is written out as (D is diagonal in x)

$$\langle D\rangle = \sum_{\{D_i\}} P_r(D_1,D_2,\ldots)\int dx \rho_o(x) u^*_\alpha(k,x)\left(\sum_i D_i(x)\right) u_{\alpha'}(k',x)\rho_o(x) \tag{18}$$

where the $\{D_i\}$ means the set of all values of the random variables D_1, D_2, ... and $P_r(D_1,D_2,\ldots)$ is the appropriate configuration probability. In this model a particular $D_i(x)$ is non-zero only in the ith cell; then because of the quasi-periodicity of the of the Floquet solutions Equation (20) may be reduced to a sum over cell centers X_i,

$$\langle D\rangle_{\alpha\alpha',kk'} = \langle \sum_i e^{i(k-k')X_i} \eta_{i,\alpha\alpha',kk'} \rangle \tag{21a}$$

$$= \sum_i e^{i(k-k')X_i} \langle \eta_{i,\alpha\alpha',kk'} \rangle \tag{21b}$$

where $\eta_{i,\alpha\alpha',kk'}$ denotes one of the (random magnitude) integrals appearing in Equation (20). But if the material is homogeneously random all expectations are the same and the phase sum adds to $2\pi\,\delta(k-k')$, whence

$$\langle D\rangle_{\alpha\alpha',kk'} = \delta(k-k') d_{\alpha\alpha',k} \tag{22}$$

where

$$d_{\alpha\alpha',k} = 2\pi \sum_{\{D_1\}} p_r(D_1) \int dx \rho_o(x) u^*_\alpha(k,x) D_1(x) u_{\alpha'}(k,x) \rho_o(x) \tag{23}$$

Experience in similar problems has shown that "interband" terms $\alpha \neq \alpha'$ are usually much smaller than for $\alpha = \alpha'$, so that Eq. (22) may be approximated

$$\langle D\rangle_{\alpha\alpha',kk'} \simeq \delta_{\alpha\alpha'}\delta(k-k') d_{\alpha\alpha,k} \tag{24}$$

With this change of representation Equation (15b), taken in $\alpha\alpha'$, kk' representation becomes diagonal, therefore algebraic, and each diagonal element $\alpha = \alpha'$, $k = k'$ satisfies the relation $\langle G\rangle_{\alpha\alpha,kk} = G_{o,\alpha\alpha,kk} + G_{o,\alpha\alpha,kk} d_{\alpha\alpha,k} \langle G\rangle_{\alpha\alpha,kk}$ which, finally, may be manipulated to the form

$$\langle G\rangle_{\alpha\alpha,kk} = (\omega^2 - \lambda^o_{\alpha,k} - d_{\alpha\alpha,k})^{-1} \tag{25}$$

The physical interpretation is as follows: First,

because of the statistical homogeneity the average wave still has well defined k, as expressed by the term $\delta(k-k')$ in $\langle G\rangle$; second, to lowest order average the eigenfrequency of the average wave is shifted from $(\lambda^{o}_{\alpha k})^{\frac{1}{2}}$ to $(\lambda^{o}_{\alpha k} + d_{\alpha\alpha,k})^{\frac{1}{2}}$, which is a renormalization of the dispersion and therefore of group velocity; third, to this order of averaging $d_{\alpha\alpha,k}$ is necessarily a real quantity so that the average wave is not damped. In the specific case of long elastic waves, because of the form of D, the effective elastic modulus does not involve $\Delta\rho\omega^2$ as $\omega \to 0$, so depends only on elastic moduli; similarly the ω^2 dependence leads to an effective mass depending only on $\Delta\rho$. Also, note that the averages are not simple volume averages but are weighted by the Floquet amplitudes.

If the same procedure is applied to the next higher order approximation, Equation (16c), together with certain material assumptions (homogeneous randomness, no correlation of randomness between fibers, "interband" terms small) a next higher order correction to $\lambda^{o}_{\alpha,k}$, denoted by $\Delta^{(2)}_{\alpha,k}$ is found. With these assumptions

$$\Delta^{(2)}_{\alpha,k} \simeq \sum_{\{D_1\}} p_r(D_1) \int dx \rho_o (D_1 - \langle D\rangle)^2 G_o(x;x;\omega^2) \rho_o u^*_\alpha(k,x) u_\alpha(k,x) \qquad (26)$$

The new renormalized dispersion relation is

$$\langle G\rangle_{\alpha\alpha,kk} = (\omega^2 - \lambda^{o}_{\alpha,k} - d_{\alpha\alpha,k} - \Delta^{(2)}_{\alpha,k})^{-1} \qquad (27)$$

Now, however, a more complete examination of the analytic properties of the renormalization is necessary, because the presence of $G_o(x,x;\omega^2)$ in the definition of $\Delta^{(2)}_{\alpha,k}$ introduces a collection

of poles. This analytic question is intrinsic to perturbation theory in a continuous spectrum, and has been discussed in the physics literature.[7] Here we quote only the necessary result, it being noted that to insure time causality in the Fourier integral inversion on ω it will be necessary to introduce a small imaginary part of ω, i.e., $\omega \to \omega + i\varepsilon$; then it may be shown that $G_o(xx;\omega + i\varepsilon)$ can be treated as having analytic real and imaginary parts, being effective weightings of the actual poles in G_o. In particular, the imaginary part is just proportional to the total number of states having frequency ω. Thus, we now find that

$$\Delta^{(2)}_{\alpha,k} = \beta^{(2)}_{\alpha,k} + i\ \gamma^{(2)}_{\alpha,k} \tag{28}$$

This means that although spatially there may still be an average wave of well defined k, it is no longer a stationary state (with real $\omega_{\alpha,k}$). Rather, that particular mode will damp out in time due to $\gamma^{(2)}_{\alpha,k}$, whose physical origin is simply scattering into other modes of the same frequency. The damping in this approximation is proportional to the mean square fluctuation of the material parameters from the average, and is necessarily non-zero. Thus in this and all higher approximations the Floquet wave will be damped.

VI. SUMMARY AND INTERPRETATION

This has been an outline of how formalism used to analyze random alloys might be applied to the analogous case of random variations in the material properties of fiber reinforced

or layered composites. As a methodology it is a perturbation method, which can make use of Floquet solutions already developed for an average medium. This procedure has been much more productive in solid state physics than attempting to analyze differential equations with random coefficients, or to make statistical approximations to a free energy functional which is then to be used in a variational statement.

We have not made numerical calculations on a model random composite as yet, but from experience in the random mass alloy expect several results:

(1) There will be well defined average Floquet waves characterized by wave vector k, but in general they will be damped in time.

(2) The low frequency average wave will be damped least since there are fewer states of the same frequency into which it can scatter.

(3) The dispersion relations of the periodic composite are shifted. On closer examination of the perturbation series it is found that convergence is slowest near the unperturbed band edge, so the most prominent changes in the dispersion occur at the band edges - they lose their "sharpness."

(4) The main point to be stressed is that all in all, moderate homogeneous randomness does not destroy Floquet properties in the average waves. Indeed, in analogous problems in disordered alloys a band structure is retained, albeit shifted and "fuzzy" at former band edges.

(5) However, some possibly needed information is lost by the

procedure outlined. First, the fluctuations from the average field and the scattered field have not been discussed much; they are important in conserving energy, of course. Second, the detailed local variations of displacement, stress, and strain are not examined explicitly when studying the average Green's function; to the extent that these may determine peak stresses, plastic flow, and failure, further information is needed. To lowest order one might consider the average wave as an initial condition and look in detail at the stresses within one perturbed cell; interactions between material variations between cells are likely to be much smaller.

(6) Lastly, the formalism above reduces in the case of a randomly perturbed homogeneous medium to simple Fourier representation of the Green's functions and operators. As such, the low order approximations discussed above are equivalent to those of Karal and Keller.

ACKNOWLEDGEMENT

This work was supported by the Advanced Research Projects Agency, Materials Research Council, University of Michigan Contract No. DAHC15-71-C-0253.

REFERENCES

1. M. J. Beran, Phys. Stat. Sol. (a) 6, 365 (1971); a review with many previous references.

2. F. Karal and J. B. Keller, J. Math. Phys. 5, 537 (1964).

3. D. W. Taylor, Phys. Rev. 156, 1017 (1967).

4. L. Schwartz and H. Ehrenreich. Annals of Phys. (N.Y.) 64, 100 (1971).

5. S. Y. Wu and P. L. Taylor, Phys. Rev. B2, 1752 (1970).

6. K. S. Dy and S. Y. Wu, Phys. Rev. B4, 1173 (1971).

7. A. A. Maradudin, in Solid State Physics, Ed. Ehrenreich, Seitz, Turnbull, Suppl. 3, (second edition, Academic Press, 1971), see p. 384ff.

THREEFOLD COORDINATED MODEL STRUCTURE OF AMORPHOUS GeS, GeSe and GeTe

A. Bienenstock

Abstract

The black P structure is presented as a model for the structures of amorphous GeS, GeSe and GeTe. It is shown that the short interatomic distances, low near neighbor coordinations and high covalencies of the amorphous materials, relative to the crystalline, can be rationalized with the model. When scaled to the near neighbor interatomic distances in the amorphous materials, the model yields satisfactory agreement with the observed position and area of the second neighbor X-ray radial distribution function peaks. The model predicts: (a) A first neighbor peak area for GeS which is significantly different from that predicted by the random covalent model, (b) phase separation in certain composition regions which, for the Ge-S system, should be observable by means of transmission electron microscopy and (c) differences between the valence band densities of states associated with this and the random covalent model which should be observable with photoemission experiments.

THREEFOLD COORDINATED MODEL STRUCTURE OF AMORPHOUS GeS, GeSe and GeTe

A. Bienenstock

I. Introduction

Considerable attention[1-4] has been directed recently towards understanding the atomic arrangements in the amorphous compounds GeS, GeSe and GeTe. Both the interest and the difficulty arise from the fact that the short range order in these amorphous compounds is quite different from that in the corresponding crystals.

Radial distribution studies show that the nearest neighbor distances in the amorphous materials are 0.2 to 0.3Å less, and the coordination numbers are significantly lower than those in the crystals. The crystals are all commonly described as distortions of the rock salt structure. As a result of the distortions, each atom has three neighbors which are separated by a distance which is slightly less than or equal to the sum of the ionic radii. Three further nearest neighbors are separated by distances which are a few tenths of an Å longer. In the amorphous materials, the nearest neighbor interatomic distances are well described by covalent radii.

The more covalent nature of the amorphous materials is also shown by X-ray photoemission (ESCA) studies[5] of the

crystalline compounds and some amorphous alloys in the Ge-Te system, including amorphous GeTe. These studies show that the core levels of germanium are 1 to 2 eV deeper in the crystalline compounds than in the amorphous materials. This result has been interpreted as corroborating the larger ionicity of the crystalline materials.

The difference in bonding is also demonstrated in the fundamental band gaps[6] of crystalline and amorphous GeTe. In the former, the gap is of the order of 0.2 eV while in the amorphous compound it is approximately 0.8 eV.

Because of the above mentioned dissimilarities between the crystal and amorphous phases, structural studies based on radial distribution techniques have remained ambiguous. Two types of models have been proposed. In the first, every germanium is surrounded by three chalcogens and each chalcogen is surrounded by three germanium atoms. This coordination is consistent with the area of the first radial distribution function peak. Another model which yields consistency with that area is the random covalent model[1]. In this model, each germanium is coordinated by four atoms and each chalcogen with two, in accordance with the 8-N rule. The degree of chemical ordering beyond this structural ordering is assumed to be zero. That is, the system is assumed to be a random alloy.

The random covalent model has enjoyed a great deal of popularity since its inception. It has been shown to predict

with great accuracy the peak areas of radial distribution functions for alloys in all three of the germanium-chalcogen systems. It also provides structural justification for the Mott model which explains the small impurity effects in many semiconductors. That is, the conductivities of many amorphous semiconductors are extremely insensitive, relative to those of crystalline semiconductors, to the presence of many impurities. Mott proposed that these impurities are coordinated such that their covalent bonding requirements are satisfied. As a result, they do not contribute donor and acceptor states and do not, therefore, appreciably influence the conductivity. The random covalent model is then merely an extension of Mott's picture to concentrated alloys.

Another feature of the model which is attractive is that it provides a structural picture for homogeneous amorphous alloys over the entire germanium-chalcogen composition range. For reasons discussed below, one would anticipate phase separation in certain amorphous alloys if the threefold coordinated model is appropriate for the compounds. Such phase separation in Ge-Te alloys has been searched for by a few groups without success.

Attempts to interpret existing data in terms of the threefold coordinated model have met some problems. Betts[7], for instance, noted that the As-like crystalline GeTe structure could be distorted further from the rock salt structure to obtain a threefold coordination similar to that in the amorphous material. Areas of second neighbor radial distribution function peaks cal-

culated from this model were, however, very much lower than those measured. In addition, if one views the crystalline materials as already being threefold coordinated, it was difficult to understand why the nearest neighbor distances in the amorphous materials are so much shorter than those in the crystalline and the amorphous materials appear so much more covalent than do the crystalline. Nevertheless, it should be noted that the first neighbor peak areas associated with radial distribution functions of alloys containing between 33 1/3 and 50 percent germanium are always consistant with a model in which the alloy is phase separated into a dichalcogenide with the SiO_2 atomic arrangement and a threefold coordinated monochalcogenide.

The purpose of this paper is to present a threefold coordinated model which is consistent with the data published thus far. The paper begins with a detailed study of the crystalline GeS and GeSe structures. An attempt is made to explain the unusual atomic arrangements in these crystals. Then, a threefold coordinated model of the amorphous materials, which uses the first neighbor peak positions and then predicts second neighbor peak positions and areas is presented. In addition, a rationale for the shorter interatomic distances in the amorphous materials is presented. One consequence of this model is that one would anticipate phase separation of alloys containing between 33 1/3 and 50 atomic percent germanium. A discussion of why this phase separation cannot be observed in germanium-tellurium alloys, as well as a suggestion of a better system to study, are presented.

In addition, it is shown that careful radial distribution studies of the amorphous germanium sulfide should also serve to distinguish between the two models. Finally, it is suggested that one would anticipate quite different valence electron energy structures for the random covalent and threefold coordinated models. These are analyzed in some detail. It is shown that studies of valence electron photoemission should serve to distinguish between the two models.

II. The Crystal Structures of Black P, GeS and GeSe

Although the structures of GeS and GeSe are commonly described as distortions of rocksalt, they are considered here, with SnS and SnSe, to be distortions of the black P structure, as noted by Rawson[8]. All are built upon tetragonal lattices whose unit cell dimensions are listed in Table 1. It should be noted that the b and c axes of all five materials hardly differ, while the a axis increases steadily with increasing average atomic size. The dimensions of the isomorphic SnS and SnSe structures are also shown.

Figure 1a shows the structure of black phosphorus in projection along the a axis, as determined by Hultgren et al[9]. The x coordinate of each atom is also shown in the figure. The structure consists of double layers stacked along, and in pairs of planes which are perpendicular to, the c axis. A single double layer is represented by the atoms denoted 1-2-3-4-5-6. In keeping with the 8-N rule, each atom is threefold coordinated, with an

average separation of 2.18Å. Atom 4, for instance, is coordinated by atom 3 and by two atoms, with x equal to 0 and 1, denoted by 5. The single and double lines connecting atoms are used to represent the single and double coordinations, respectively.

The adjacent double layer is represented by atoms 7-8-9-10-11-12. The interlayer bonding is rather weak, as is evidenced by the shortest interlayer interatomic separation of 3.68Å.

Despite the large interlayer separation, black P is a high pressure structure. Hultgren et al. address the question of why this high pressure form does not take on the As structure. They state, "A good reason is that the black phosphorus structure is closer packed than the arsenic structure and so is favored by the high pressure under which black phosphorus is formed. If phosphorus assumed the arsenic structure, retaining the bond distances and angles of black phosphorus and also the closest distance of approach between atoms in different layers, it would have a density of only 2.44 instead of 2.69." This fact becomes important below when the large areas of the amorphous chalcogenide second neighbor radial distribution peaks are considered.

Figure 1b shows the same projection of the GeSe structure, as determined by Okazaki[10]. This structure is similar, but not identical, to the black P structure. The double layers are distorted so that the atoms no longer sit on pairs of planes perpendicular to the c axis. While the threefold coordination of, say, Ge atom 4 is still evident, the distortion tends to reduce the intralayer bond angles so that the two Se atoms denoted by 1 are

significantly closer to, and the two Ge atoms denoted by 2 are significantly further from, Ge atom 4 than they would be if the simpler black P structure were maintained. This distortion of the layers also has the effect of moving Se atom 12 closer to Ge atom 4 than it would be in the black P structure. As a result, there are three short near-neighbor separations of approximately 2.57Å and three long near-neighbor separations of approximately 3.33Å. The corresponding separations for the other crystals with this structure are summarized in Table 1. The net effect is to distort the black P structure into one which is closer to rocksalt. Nevertheless, the basic threefold coordination associated with an average of five electrons per atom is apparent from the ratio of short to long near-neighbor separations. Evidence in support of the position that the structure should be viewed as closer to that of black P than NaCl is contained in the fact that the two "second neighbor" Ge atoms represented by 2 are closer to Ge atom 4 than is "first neighbor" Se atom 12.

On the other hand, there is considerable evidence that the bonding has an appreciable ionic component. The short near-neighbor distances are very close to the sum of the ionic radii, as shown in Table 1, and are 0.2-0.3Å longer than the sum of the covalent radii. The ESCA study[5] mentioned above also indicates a higher ionicity than that found in the amorphous materials. One can see the origin of this ionicity by starting GeSe in the black P structure and allowing it to distort. We present this

exercise because of the potential insight it yields for understanding amorphous structures.

To construct GeSe with a black P structure, an average of five electrons per atom must be obtained. This means the creation of Ge^- and Se^+. As a result, the p^3 bonding is appropriate and the black P structure can be obtained. Because, however, of the stronger attractive potential of the Se, the center of gravity of the bonding electron cloud is closer to the Se than the Ge, so that the Se is effectively negatively ionized, as would be expected from simple chemical arguments and is demonstrated by the ESCA studies. This situation is, of course, quite analogous to that of III-V compounds with the zinc blende structure.

With the effective ionization, however, the Madelung contribution to the cohesive energy is increased if the number of oppositely charged near neighbors increases, and the distance to similarly charged near neighbors is increased. Hence, the distortion is obtained. With this distortion, however, we can expect a further movement of the valence electrons towards the chalcogens and away from the Ge atoms to obtain a further increase of the Madelung contribution to the cohesive energy, and correspondingly a decrease in the distance between atoms 1 and 4. The distance between atoms 4 and 5, on the other hand, increases to something closer to the sum of the ionic radii.

In support of this picture, it should be noted that the $\underline{a}$ lattice parameter of GeS and GeSe is almost exactly the appropriate ionic chalcogen diameter, which would be expected from the

ionic picture of the structure, since this axial length is determined by anion-anion contact. In the Sn salts, however, a is larger.

III. Model Structure for the Amorphous Materials

As indicated above, the radial distribution functions of the amorphous materials differ significantly from those anticipated from a microcrystalline model. The features which must be explained can be summarized as follows:

a) The nearest neighbor distances are 0.2-0.3Å shorter than those in the crystal and are well described by covalent radii.

b) If it is assumed that every Ge is surrounded by chalcogen only, and every chalcogen by Ge atoms only, the coordination number is 3.

c) The crystalline long near-neighbor separation is always close to a minimum in the amorphous rdf. That is, a very few atoms in the amorphous materials are separated by that distance. Instead, the second maximum in the rdf occurs for separations which are about 1Å larger.

d) The area of the second rdf peak is large, indicating high "second neighbor" coordination.

To explain these features with a model which is based on the crystalline structure and threefold coordination, we assume that the basic double layer structure is maintained in the amorphous materials. Given this assumption, it is then necessary to explain why the same basic coordination leads to ionic bond

distances in the crystalline materials and covalent in the amorphous.

We have noted, in Section II, that the coordination must be fairly high to support ionic bonding. That is, if the coordination number were only 3, we would anticipate a primarily covalent p^3 bonding scheme. The further ionicity arises because atoms like number 1 of Fig. 1b are long near neighbors of atoms like number 4. In terms of the basic layer-like covalent bonding scheme, however, atom number 1 is a third neighbor of atom number 4. Even if we did not have the radial distributions of these amorphous compounds, experience with vitreous SiO_2 and Ge would tend to indicate that it would be extremely difficult to maintain such a correlation of third neighbor distances in the amorphous materials. The absence of the long near neighbor peak in the rdf's of the germanium monochalcogenides indicates that this third neighbor correlation is not maintained. As a result, the near neighbor covalent bonding predominates the bond distances are short with respect to the crystal. Hence, the basic double layer structure is capable of dealing with points (a), (b) and (c) of the first paragraph of this section.

This leaves the rdf second neighbor peak positions and areas to be explained. In a system of this complexity, where the crystal cannot give reliable guidance, there is considerable arbitrariness about any detailed model. It seems appropriate to this author, therefore, to take the simplest possible model

and determine the extent to which it is consistent with observations. It is reasonable, therefore, to determine whether the double layer arrangement of black P can account for most of the atomic correlations in the second neighbor peak.

To determine if the basic double layer structure of black P can account for the second neighbor rdf peaks of the compounds, we have assumed that the bond angles in the latter are identical to those in black P, so that all intralayer distances scale like the nearest neighbor distance.

Table 2, therefore, presents a list of all interatomic distances in black P of less than 7Å, in its first column. Each of these distances has been multiplied by a constant for each compound to obtain the corresponding distances in the model of the compound. That constant is fixed to yield the correct average nearest neighbor distance presented by the rdf's. The distances so obtained are presented in the second through fourth columns. The fifth column indicates through the signs, = and ≠, whether the associated pairs of atoms in the compounds are of the same or different atomic species, respectively. The sixth column presents the number of atomic pairs, for a single fixed central atom, at the associated distance. Finally, the seventh through ninth columns present the contribution of that pair to an X-ray diffraction rdf peak area.

The intralayer distances in black P fall into relatively distinct groups which are, therefore, also characteristic of the model compounds. Three neighbors form the first rdf peak at 2.19Å.

Then, there are sets of closely spaced distances from 3.31 to 4.38Å, consisting of 13 pairs. The remaining pairs are relatively evenly spaced, starting at 5.16Å and extending to 6.62Å. All other pairs have separations over 7Å. The corresponding scaled results for GeTe may be compared directly with the rdf of Betts et al.[1], which shows a peak at 2.6Å with an area of 5100±500 electrons[2]. The distance has, of course, been scaled to be equal, and the calculated area of 4992 is in good agreement with experiment. The rdf also shows a second peak which extends from approximately 3.3 to 5.5Å, has a maximum at 4.2Å and an area of 27,500 ±1500 electrons[2]. The calculated areas from the pairs which extend from 3.93 to 5.20Å is 23,232 electrons[2]. At first inspection, this agreement appears too poor to allow further consideration of the model. It should be noted, however, that this range is precisely that in which one would anticipate interlayer contributions from the next double layer on the side of the central atom. The area which would have to be accounted for from such interlayer contributions is 4268±1500 electrons[2], which could be contributed by less than three such neighbors in the range. Since this is a quite reasonable number, the area agreement is not bad. The distribution of distances is a bit more troubling. The rdf shows a smooth maximum at 4.2Å, while the calculation predicts one peak, with 8 pairs involved, centered at approximately 4.03Å and another, with five pairs involved, centered at approximately 4.96Å. There is no trace of a resolution of the two peaks in the rdf. This

shows clearly the inadequacy of such a simple model. While it is giving reasonable agreement for the total area, it does correctly predict the details of the interatomic distances. This is, of course, to be expected, since the model has completely neglected the distortions of the layer structure which are likely to take place because there is a partial ionicity necessarily present and because the layers are part of an amorphous structure. In addition, the model gives no detailed information about the interlayer contributions.

Nevertheless, the work presented thus far does place the threefold coordinated model on at least an equal footing with the random covalent model. Hence, one must search out methods of distinguishing the two models. Some suggestions are presented below.

IV. Radial Distribution Studies

It has been shown by Betts et al.[11] that it would be virtually impossible to distinguish between the random covalent and threefold coordinated models for amorphous GeTe on the basis of the near neighbor X-ray diffraction rdf peak area because of the extremely high accuracy required. This statement is also true for amorphous GeSe. These authors note, however, that a neutron diffraction rdf on GeTe could succeed. Unfortunately, it would be quite difficult to produce enough sample for the experiment. It should be noted, however, that the situation is somewhat different for amorphous GeS. Here, the random covalent and threefold coordinated models predict X-ray rdf areas of 1920

and 1536 electrons^2, respectively. These should be distinguishable. Unfortunately, so such rdf has yet been published.

The closest thing to it is an X-ray rdf[4] on a sample of composition $Ge_{.42}S_{.58}$, for which the peak area is 1647 electrons^2. A phase separated model, in which the two phases are assumed to be the threefold coordinated GeS and a GeS_2 phase with the vitreous SiO_2 structure, predicts an area of 1456 electrons^2. Hence, that work would tend to indicate that the random covalent model is more appropriate. It should be noted, however, that the negation of the threefold model depends on an assumption about the nature of the phase separated species. It would be more desirable to have an rdf of the pure compound.

V. Phase Separation

The random covalent and threefold coordinated models appear quite different in their predictions with respect to phase separation. The random covalent model is able to accommodate all compositions with ease. The threefold coordinated model, though, depends on having an average of five valence electrons per atom. Hence, one would anticipate small solubilities of either Ge or the chalcogens in the amorphous compounds, and phase separation for appreciable deviations from stoichiometry.

Verhelle and Bienenstock[12] have searched unsuccessfully for such phase separated in amorphous films of the composition $Ge_{.46}Te_{.54}$. These studies involved transmission electron microscopy studies of sputtered films in both the unannealed and

annealed states. As a result of more recent studies[13] of the densities of such films, however, it can be shown that the electron densities of amorphous GeTe and $GeTe_2$ differ by less than 1.3%. As a result, little contrast would be expected in transmission, even if such separation were present. A similar situation is anticipated for the Ge-Se system. For the Ge-S system, however, the situation is quite different. The densities of amorphous GeS and GeS_2 are[10] 1.624 and 1.26 gm/cc, respectively. These mass densities imply electron densities which differ by 25% of their average. Hence, the contrast associated with phase separation should be quite apparent. Unfortunately, no such studies have been performed, to my knowledge.

It should also be pointed out, however, that replica studies of etched, annealed samples of Ge-Se glasses in the 40 at. % Ge range have been performed by Mortyn and Bienenstock[14], as well as Feltz et al.[15] In both cases, no separation was observed. The failure to observe separation with replicas must, however, be considered inconclusive.

In summary, then, it must be concluded that phase separation studies performed thus far do not show such separation, but that the systems and techniques employed thus far are not the most sensitive. Further work should be performed on the Ge-S system using transmission electron microscopy.

VI. Electronic Structure and Photoemission

One technique which might distinguish between the two

types of coordination is photoemission. Let us consider the differences one would anticipate in the general form of the valence band structure for the random covalent and threefold coordinated models.

In the random covalent model, the Ge is fourfold coordinated, so that sp^3 bonding is appropriate for it. One consequence of such bonding is that the atomic gap between the outer s and p states essentially disappears. Kastner[16], in his analysis of bonding states in chalcogenide glasses, indicates that Ge-Ge and Te-Te bonding states are at virtually the same energy with respect to the vacuum level. One may anticipate that Ge-Te bonding states also lie in this energy regime. Hence, it seems reasonable to assume that all the bonding states will contribute one broad macimum to the density of states, forming the lower energy end of the valence band. The upper end of the valence band would be formed by the lone pair Te states. Finally, the Te s states would lie below the valence band, separated by an energy gap.

In amorphous threefold coordinated GeTe, however, the situation is quite different. There are no lone pair Te states. Similarly, there is no sp^3 hybridization of the Ge valence electrons. Hence, one would anticipate a broad valence band representing the chemical p^3 bonding. Then, isolated and below that would be a Ge s band. Then, well below that, there would be an isolated Te s band.

Such differences in the general forms of the densities of

states could be observed by means of X-ray or, possibly ultraviolet photoemission.

VII. Conclusions

The rationalization of the short nearest neighbor distances, low coordination and low ionicity in amorphous GeSe, GeSe and GeTe, relative to the crystals, on the basis of a threefold coordinated model places that model on an equal footing with the random covalent model. If the threefold coordinated model is valid, then virtually no evidence that highly concentrated, disordered, amorphous chalcogenide alloys exist. Their absence is not surprising, since they demand near neighbors which show appreciable electronegativity differences. Hence, there would be an appreciable number of similarly charged nearest neighbors.

Even this argument, though, must be accepted with caution. The transmission electron microscopy work of Chaudhari and Herd[17] indicates that there is no phase separation in amorphous $Ge_{15}Te_{85}$. If this is the case, one would anticipate that there are appreciable numbers of Te-Te pairs in which each Te is also bonded to a Ge and is, therefore, slightly negatively charged. Hence, one cannot rule out homogeneity on the basis of this positive Coulomb energy. On the other hand, at these low Ge concentrations it would, presumably, be more than cancelled by the negative Coulomb energy associated with the Ge-Te pairs.

At any rate, the lack of conclusive evidence for the existence of homogeneous disordered alloys indicates that phase

separation or segregation, with complex bonding schemes, rather than the simple 8-N bonding, may be much more prevalent in the polyatomic chalcogenide amorphous materials than has been assumed in most theoretical discussions of their physical properties.

Acknowledgement

This research was supported by the Advanced Research Po Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 and under Contract Number DAHC04-70-C-0044, which was monitored by U.S. Army Research Office-Durham.

References

1. F. Betts, A. Bienenstock and S. R. Ovshinsky, J. Non-Cryst. Solids 4, 554 (1970).

2. D. B. Dove, M. B. Heritage, K. L. Chopra and S. K. Bahl, Appl. Phys. Letters 16, 138 (1970).

3. R. W. Fawcett, C. N. J. Wagner and G. S. Cargill, III, to be published in the Proceedings of the Fourth International Conference on Amorphous and Liquid Semiconductors, Ann Arbor, 1971.

4. S. C. Rowland, S. Narasimhan and A. Bienenstock, J. Appl. Phys. 43, 2714 (1972).

5. F. Betts, A. Bienenstock and C. Bates, see ref. 3.

6. R. Tsu, W. E. Howard and L. Esaki, J. Non-Cryst. Solids 4, 322 (1970).

7. F. Betts, Ph.D. Thesis, Stanford University, 1972. Unpublished.

8. H. Rawson, *Inorganic Glass-Forming Systems*, Academic Press, New York, 1967, p. 253.

9. R. Hultgren, N. S. Gingrich and B. E. Warren, J. Chem. Phys. 3, 351 (1935).

10. A. Okazaki, J. Phys. Soc. Jap. 13, 1151 (1958).

11. F. Betts, A. Bienenstock, D. T. Keating and J. deNeufville, to be published in J. Non-Cryst. Solids.

12. Y. Verhelle and A. Bienenstock, unpublished work.

13. Energy Conversion Devices, Inc. Third Semi-Annual Technical Report on Contract DAHC15-70-C-0187 with the Advanced Research Projects Agency.

14. F. Mortyn and A. Bienenstock, unpublished work.

15. A Feltz, J. J. Buettner, F. J. Lippmann and W. Maul, private communication and ref. 3.

16. M. Kastner, Phys. Rev. Letters 28, 355 (1972).

17. P. Chaudhari and S. R. Herd, see ref. 3.

TABLE 1. Crystalline Lattice Parameters and Interatomic Separations

	a	b	c	r_1	r_1'	r_2
P	3.31	4.38	10.50	2.18		
GeS	3.64	4.29	10.42	2.58	2.57	2.97
GeSe	3.82	4.38	10.79	2.57	2.64	3.33
SnS	3.98	4.33	11.18	2.66	2.77	3.31
SnSe	4.19	4.46	11.57	2.80	2.84	3.39
GeTe*				2.84	2.84	3.16

*GeTe has a rhombohedral, rather than tetragonal, lattice.

r_1 is the average short near neighbor separation.

r_1' is the sum of the divalent radii.

r_2 is the average long neighbor separation.

TABLE 2. Coordination Distances and Numbers in the Model Threefold Coordinated Structure

r_P	r_{GeS}	r_{GeSe}	r_{GeTe}		No.	GeS	GeSe	GeTe
2.18	2.34	2.40	2.59	≠	2	1024	2176	3328
2.20	2.36	2.41	2.61	≠	1	512	1088	1664
3.31	3.55	3.63	3.93	=	2	1280	2180	3728
3.41	3.66	3.74	4.05	≠	2	1024	2176	3328
3.43	3.68	3.76	4.07	=	4	2560	4306	7456
3.97	4.26	4.35	4.71	≠	2	1024	2176	3328
4.15	4.45	4.55	4.93	≠	1	512	1088	1664
4.38	4.70	4.80	5.20	=	2	1280	2180	3728
5.16	5.54	5.65	6.13	≠	2	1024	2176	3328
5.30	5.69	5.81	6.29	≠	2	1024	2176	3328
5.49	5.89	6.02	6.52	=	4	2560	4306	7456
5.56	5.97	6.09	6.60	≠	1	512	1088	1664
5.78	6.20	6.33	6.86	≠	2	1024	2176	3328
5.80	6.22	6.36	6.88	=	2	1280	2180	3728
5.81	6.23	6.37	6.90	=	2	1280	2180	3728
6.01	6.45	6.59	7.14	≠	2	1024	2176	3328
6.46	6.93	7.08	7.67	≠	2	1024	2176	3328
6.62	7.10	7.25	7.86	=	2	1280	2180	3728

In Column 5, the symbols - and ≠ are used to indicate that the atoms are of similar or dissimilar species, respectively. Column 6 lists the coordination numbers associated with each distance. Columns 7, 8 and 9 list the contributions to the area of an X-ray diffraction radial distribution associated with each interatomic separation.

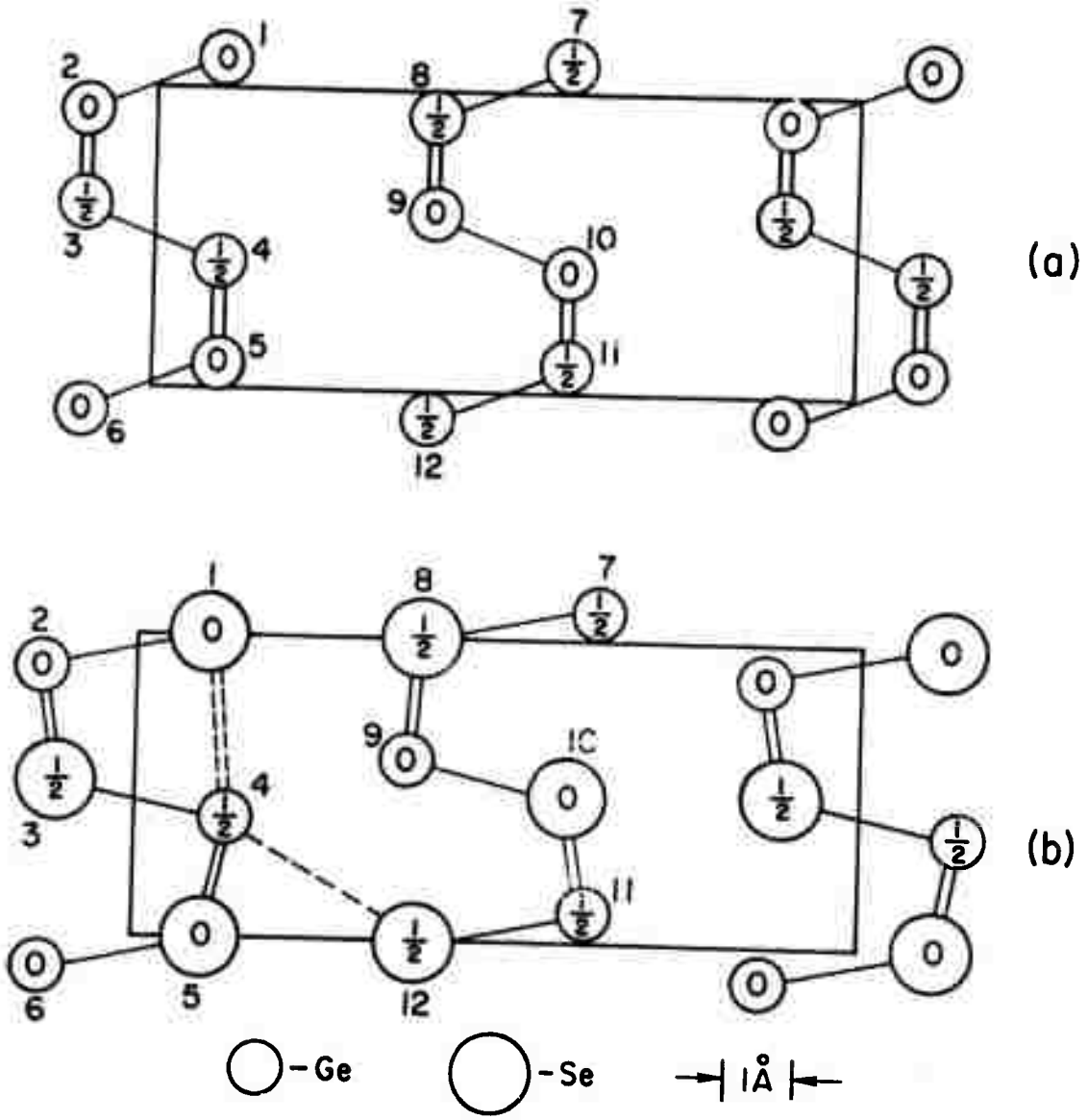


Figure 1. Projections along the c-axis of the (a) black P and (b) GeSe structures. Solid lines denote nearest neighbors. Dashed lines denote the further neighbors of the GeSe structure. Double lines indicate two near neighbors on layers c/2 above and c/2 below the atoms to which they are attached. The numbers 0 and 1/2 denote the coordinates along the c-axis of each atom.

PLASTIC RELAXATION VIA TWIST DISCLINATION MOTION IN POLYMERS

J. J. Gilman

Abstract

Measurements of the internal friction in polymers exhibit a spectrum of peaks at low temperatures that are related to the plastic behavior at higher temperatures. This manuscript deals with a particular supramolecular defect that behaves in a manner consistent with the observations. The defect is a twist disclination consisting of two molecules that cross over one another.

The energies of formation and motion for the twist disclination are calculated in terms of the width of the twisted region, the displacement that is produced by the twist, the molecular radius, and the elastic stiffnesses of the molecules. The total energy is minimized to find the optimum width and the formation energy.

For motion to occur dilatations caused by contour modulations of the molecules must be overcome. This effect yields expressions for the motion activation energy, and the stress for non-activated motion.

It is shown that molecular bending occurs more readily by elastic deformation than by conformation-change deformation unless the radius of curvature is less than a critical value (of the order of atomic dimensions).

PLASTIC RELAXATION VIA TWIST DISCLINATION MOTION IN POLYMERS

J. J. Gilman

I. INTRODUCTION

A number of solid polymers exhibit one or more discrete internal friction peaks at low temperatures (below about 50°K).[1] The temperatures at which these peaks occur indicate that the processes causing them have low activation energies. The discrete forms of the peaks indicate a relatively small range of relaxation times for these processes. Therefore, specific geometric entities must be associated ith the relaxations rather than regions of general disarray.

A specific mechanical relaxation process is discussed in this manuscript whose properties are consistent with the requirements. This is the twist disclination that was previously described in the elastic approximation by Gilman and Li[2]. In this paper its properties will be considered from a molecular viewpoint.

Consider two molecules lying side by side as in Figure 1. As indicated by the end view, the molecules lie on a plane that is inclined at 45° to the plane of the paper. The action of a shearing stress parallel to the plane of the paper and perpendicular to the axes of the molecules has caused a segment of the uppermost molecule to shear into the conjugate inclined position

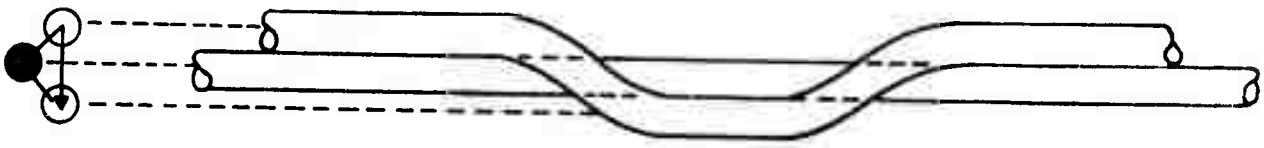


Figure 1. Partial shearing of one molecular chain relative to another. This has created two twist disclinations of opposite signs.

as sketched in the drawing. This has created a pair of kinks of opposite signs in the molecule. Motion of these kinks along the length of the molecular pair is accompanied by plastic shearing. Oscillations of the kinks up and down the two molecules can cause cyclic relaxation of an applied cyclic stress.

Figure 2A shows a single kink configuration. A displacement downward of the right-hand half of the configuration yields the more symmetric disposition of Figure 2B. This can be described in terms of a twist disclination loop that lies in the vertical mid-plane of the twist. Motion of the disclination loop parallel to the molecular pair axis causes plastic shearing to occur.

A purpose of this paper is to give expressions for the energy of formation and the activation energy for the motion of the disclination loop. The starting point is to assume an analytic shape for the kinks and then adjust its parameters to minimize the energy. The activation energy for motion is obtained by considering the energy fluctuations that will result from the "bumpiness" of the contours of the molecules.

The analytic shape that is assumed is shown in Figure 3A and written:

$$y(x) = \frac{\delta}{\pi} \tan^{-1} \left(\frac{x}{w}\right)$$

where δ is the displacement and w is the width of the transition region where most of the displacement occurs. The slope as shown in Figure 3B is:

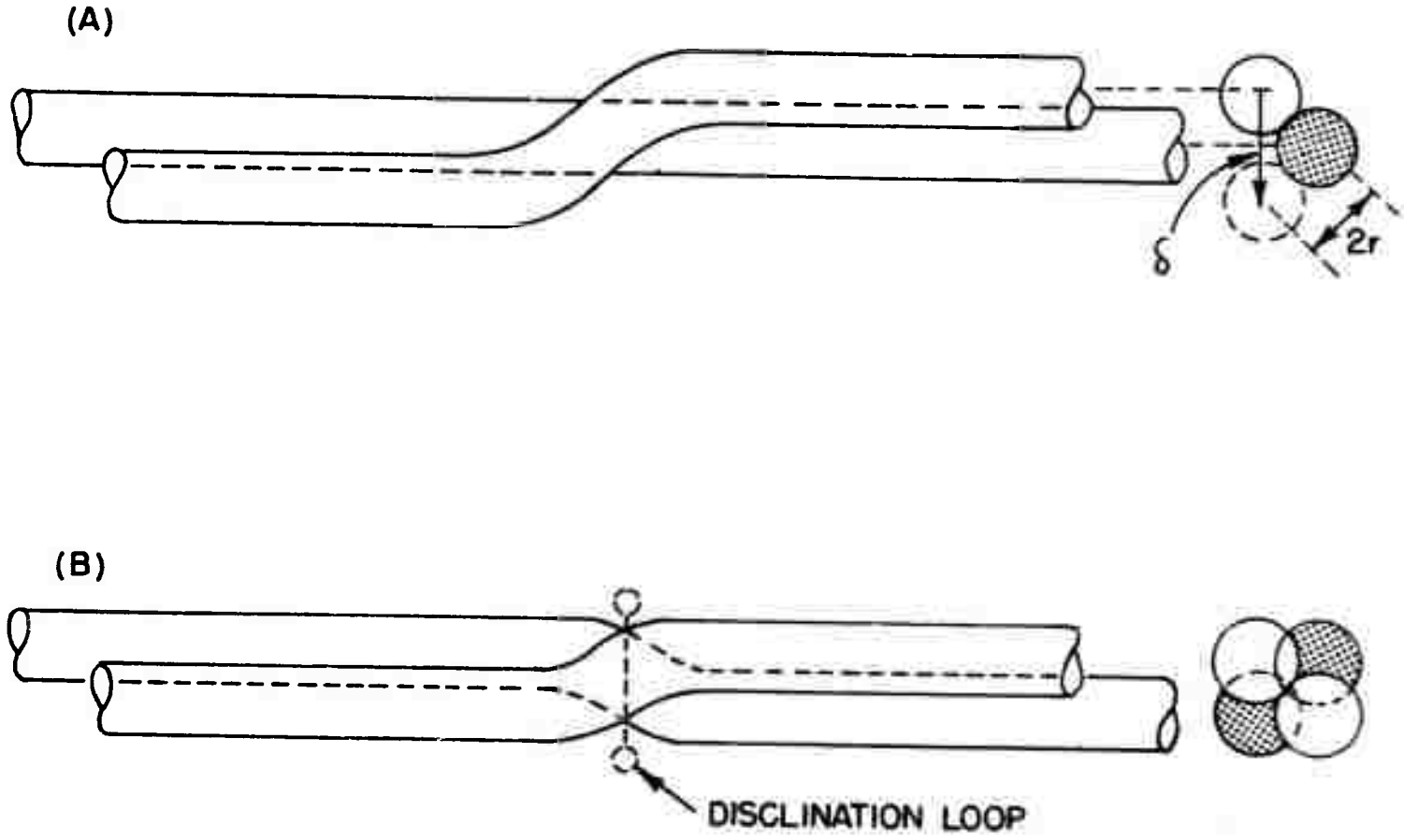


Figure 2. Schematic twist disclination in a polymer: (A) one bent molecule lying on a straight molecule; (B) symmetrically bent molecules derived from the above configuration.

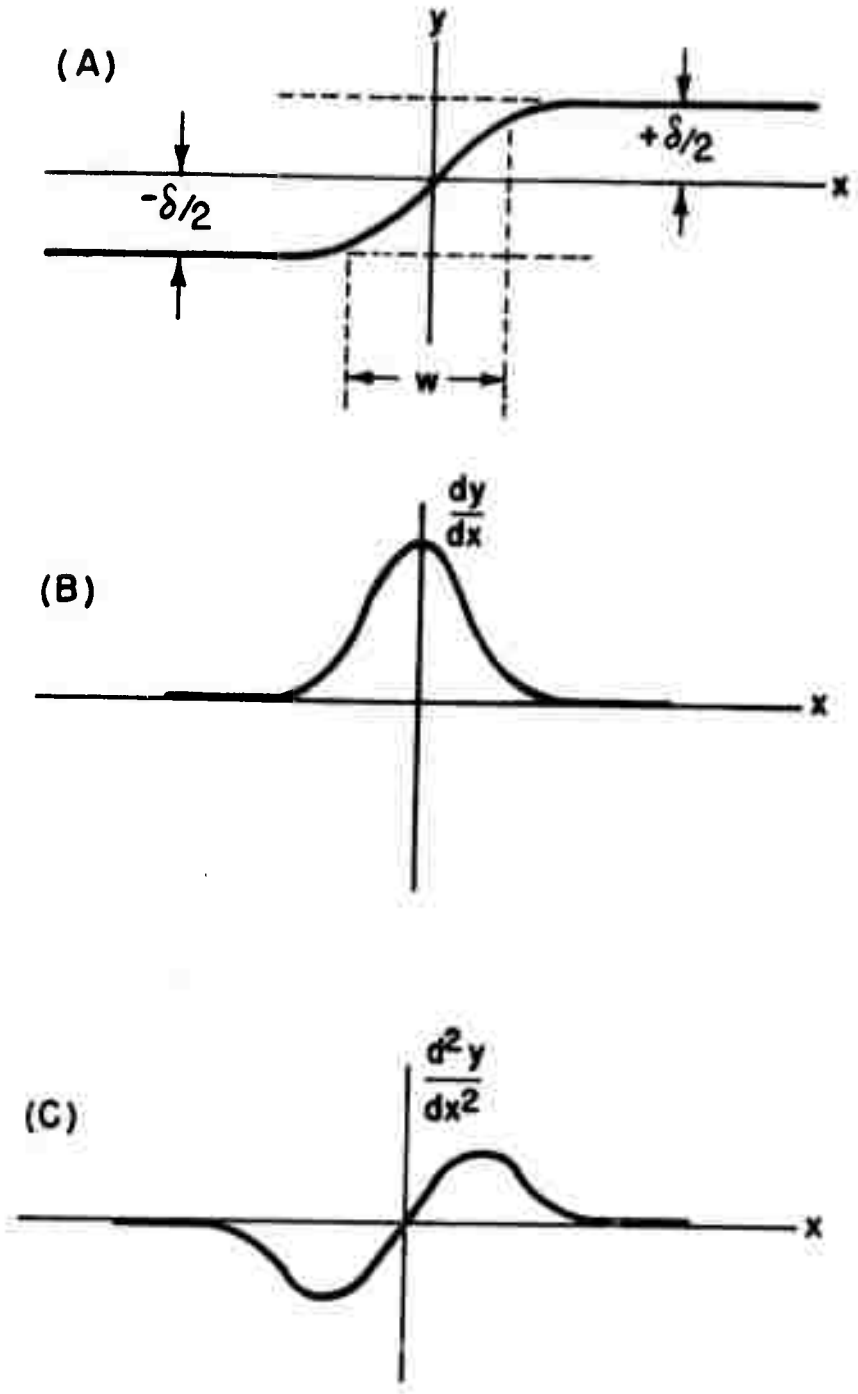


Figure 3. Analytic description of bend molecules in twist disclinations: (A) displacement function and definition of width; (B) slope which describes angle between molecules; (C) curvature function.

$$y'(x) = \frac{\delta w}{\pi}\left(\frac{1}{x^2+w^2}\right)$$

and the curvature as in Figure 3C is:

$$y'(x) = \frac{2\delta w}{\pi}\left[\frac{x}{(x^2+x^2)^2}\right]$$

Even when $w = \delta$ this curvature is only $1/2\pi w$ and therefore relatively small.

The energy of a kink can be resolved into various terms which are at least partially independent of each other and can be approximately calculated. The terms that will be considered here are:

a. elastic strain energy of bending (or the bending associated with molecular conformation changes).

b. strain energy induced by the length increase that occurs in the transition region.

c. orientation energy associated with the effects of changing the relative orientations of the two molecules
   1. molecular interaction changes
   2. dilatation changes

d. energy that depends on the position of the disclination relative to the monomer units of the molecules
   1. dilatations
   2. chemical interactions

II. STRAIN ENERGIES

First, consider the elastic strain energy of flexture, U_f, if the bending moment is M, the Young's modulus is E, and

the section modulus is $I = \pi r^4/4$ for a round rod, then:

$$U_f = \int_{-\infty}^{+\infty} \frac{M^2}{2EI}\, dx$$

and since the radius of curvature, R is given by:

$$R = \frac{EI}{M} \simeq \left(\frac{d^2y}{dx^2}\right)^{-1}$$

the strain energy becomes:

$$U_f = 2\int_0^{\infty} dU_f = \int_0^{\infty} \frac{EI}{R^2}\, dx = \frac{4\delta^2 w^2 EI}{\pi^2} \int_0^{\infty} \frac{x^2 dx}{(x^2+w^2)^4}$$

but the integral equals $\pi/32w^5$, so after substituting for I the energy is:

$$U_f = \frac{E\delta^2 r^4}{32w^3}$$

Next, consider the strain energy caused by the change of length that is needed to kink an initially straight molecule. The longitudinal strain ε, in an element of length is: $(\frac{dl}{dx} - 1)$; but $dl^2 = dx^2 + dy^2$ so:

$$\varepsilon = 1 + \left[\left(\frac{dy}{dx}\right)^2\right]^{\frac{1}{2}} - 1$$

The strain energy density is:

$$1/2\ E\varepsilon^2$$

so the strain energy in an element of length dl is:

$$1/2\ E\varepsilon^2(\pi r^2 dl)$$

and the total strain energy in kinked molecule is:

$$U_s = \frac{\pi E r^2}{2} \int_{-\infty}^{+\infty} \left\{ \left[1 + \left(\frac{dy}{dx}\right)^2 \right]^{\frac{1}{2}} - 1 \right\}^2 \left[1 + \left(\frac{dy}{dx}\right)^2 \right]^{\frac{1}{2}} dx$$

For $W >> \delta$, (dy/dx) is always small so the term in curly brackets is approximately, $1/2\ (dy/dx)^2$ and the term in square brackets is approximately, $1 + 1/2\ (dy/dx)^2$. Therefore:

$$U_s \simeq \frac{\pi E r^2}{4} \int_{-\infty}^{+\infty} \left(\frac{dy}{dx}\right)^2 \left[1 + 1/2 \left(\frac{dy}{dx}\right)^2 \right] dx$$

neglecting the quartic term this is:

$$U_s \simeq \frac{E(r\delta w)^2}{2\pi} \int_{-0}^{+\infty} \frac{dx}{(x^2+w^2)^2} \simeq \frac{3E(r\delta)^2}{8w}$$

In order to evaluate the relative importances of these two strain energy terms, their ratio is formed:

$$\frac{U_f}{U_s} = \frac{1}{12} \frac{(r^2)}{(w^2)}$$

Since w must exceed r both from the physical viewpoint and from the approximations of the analysis, it is concluded that the flexural energy can be neglected relative to the stretch energy.

III. ELASTIC VS. CONFORMATIONAL BENDING

If relative rotations can occur between segments of molecules about axes that do not lie parallel to the axes of both adjacent segments then a sequence of discrete rotations can result in a bent overall shape of the molecule.

The point at which a rotation occurs will be called a

node, and N is taken to be the number of nodes per unit length. Let the projection (onto the plane of bending) of the misorientation angle between two adjacent segments be ϕ. Then the radius of curvature will be:

$$R = \phi/N$$

Now, if the excess energy per rotational node is μ, then the energy per unit length of conformational bending is:

$$U_{cb} = \frac{\mu\phi}{R}$$

On the other hand the energy per unit length for an elastic bend is:

$$U_{eb} = \frac{EI}{R^2} = \frac{\pi E r^4}{4R^2}$$

The two energies are equal for a particular radius of curvature:

$$R^* = \frac{EI}{\mu\phi}$$

so for $R > R^*$ the elastic energy is smaller than the conformational energy. An illustration of the magnitude of R^* may be obtained by choosing a reasonable set of parameters:

$$E = 10^{11} \text{ d/cm}^2$$

$$r = 10^{-8} \text{ cm}$$

$$\mu = 2 \text{ kcal/mol} = 0.13 \times 10^{-12} \text{ ergs}$$

$$\phi = \pi/6$$

yielding: $R^* = 1.2 \times 10^{-8}$ cm

Thus the elastic bending energy may be used to describe the behavior down to very small radii of curvature.

IV. ORIENTATION ENERGY

The low energy state of a pair of molecules obtains when they are in contact with their axes parallel. The maximum misorientation and hence the maximum interaction energy obtains at the plane of the disclination loop in Figure 2B. Here the misorientation angle is:

$$2\left(\frac{dy}{dx}\right)_{x=0} = \frac{2\delta}{w\pi}$$

If ψ_m is taken to be the maximum misorientation energy (per unit length) associated with the maximum misorientation angle, then:

$$\psi(x) = \psi_m\left(\frac{w^2}{x^2+w^2}\right)$$

Misorientation energy may be considered to consist of two principal terms. One is the change in the chemical interaction energy between the molecules that results from misorientating them. The other is the strain energy associated with the local dilatation that occurs as a result of one molecule crossing the other. For the configuration of Figure 2B the maximum dilatational strain at x = 0 is:

$$\varepsilon_{yy} = (\sqrt{2} - 1) = 0.41$$

The total misorientation energy may be obtained by inte-

grating along the length of the configuration:

$$U_o = \int_{+\infty}^{-\infty} \psi(x)\, dx$$

$$= 2\psi_m \int_0^{\infty} \frac{w^2 dx}{(x^2+w^2)} = \pi w \psi_m$$

V. ENERGY OF FORMATION

Assuming that the twist disclination interacts only weakly with the surrounding medium, except for whatever dilatation it causes, its energy of formation can be found by minimizing the sum of the strain and the orientation energies.

The total energy may be written:

$$U_t = 2U_s + U_o = \frac{A}{w} + Bw$$

with: $A = \frac{3E}{4}(r\delta)^2$; $B = \pi\psi_m$. Differentiating and setting the derivative equal to zero yields an expression for the width, W* at which the energy of formation is a minimum:

$$W^* = (A/B)^{\frac{1}{2}} = \frac{r\delta}{2}\left(\frac{3E}{\pi\psi_m}\right)^{\frac{1}{2}}$$

To estimate the magnitude of this, ψ_m can be taken as $\simeq Er^2/12\pi$ to that $W^* \simeq 3\delta$. This indicates that atomically sharp twist disclinations can indeed be expected to exist.

The energy of formation, U_f is found by substituting W* into the expression for the total energy:

$$U_f = 2(AB)^{\frac{1}{2}} \simeq 3r\delta(E\psi_m)^{\frac{1}{2}}$$

which is approximately, $Er^3/\sqrt{2}$ or about 0.05 e.v. ≃ 1 kcal/mol. Therefore these defects can be formed readily at moderate temperatures.

If the disclination interacts with the matrix strongly, then the elastic energy of the disclination loop must be added to $U_s + U_o$ to get the total energy. This will tend to increase U_f as the temperature is decreased.

VI. MOBILITY

Polymer molecules are not smooth in shape but are "bumpy". Therefore the volume of a twist disclination will depend on the position of its center. If a coordinate, z defines this position then its volume change may be expressed:

$$\Delta V = \xi \sin \frac{2\pi z}{a}$$

where a and ξ are the period and amplitude of the volume change, respectively.

The strain energy, U_ξ, associated with a volume change depends on the size of the volume, V_o and the shear modulus of the medium in which it occurs:

$$U_\xi = \frac{G(\Delta V)^2}{2\pi V_o} = \left(\frac{G}{2\pi V_o}\right) \xi^2 \sin^2 \left(\frac{2\pi z}{\alpha}\right)$$

The difference between the minimum and maximum values of this energy gives the activation energy, U_m^* for the motion of the disclination:

$$U_m^* = \frac{G\xi^2}{2\pi V_o}$$

In the absence of thermal or stress activation a twist disclination can be moved by an applied shear stress that equals or exceeds the quasi-static stress that resists its motion. This can be obtained by differentiating its energy with respect to its position. The force acting on it is:

$$f_\xi = \frac{dU_\xi}{dz} = \frac{2G\xi^2}{aV_o}\left[\sin\left(\frac{2\pi z}{a}\right)\cos\left(\frac{2\pi z}{a}\right)\right]$$

and its maximum value occurs when the term in brackets equals one-half:

$$f_\xi^m \simeq \frac{G\xi}{aV_o}$$

The area associated with this force is approximately δw, so the maximum reactive stress is:

$$\sigma_\xi^m \simeq \frac{G\xi^2}{aw\delta V_o}$$

or since $aw\delta \simeq V_o$:

$$\sigma_\xi^m \simeq G\left(\frac{\xi}{V_o}\right)^2$$

VII. DISCUSSION

In Section V it was pointed out that the energy of formation of a twist disclination loop is small and comparable with the thermal energy at 40°K so no strong barrier to their formation exists and applied stresses will aid their formation. Therefore, it is believed that the activation energy associated

with the observed internal friction peaks should be associated with the motion of a loop.

From Section VI, the volume change associated with motion is:

$$\xi = \frac{2\pi U_m}{G\varepsilon}$$

Using the experimentally observed value for the activation energy: U_m = 4 kcal/mol = 1.7×10^{-13} ergs together with G = 10^{12} d/cm^2 for a typical molecule; and letting the strain be 0.23 (the amount associated with a sphere moving from a close-packed position to a saddle point) - the volume change is about $3.6A^3$ or one-fourth the volume of a polyethylene monomer unit. This is a quite reasonable value and therefore tends to support the present theory.

The model described here is consistent with the characteristics of this relaxation phenomenon as summarized by Hiltner and Baer[4]. They state that for three linear polymers (polyethylene, polyethylene terephthalate, and polyoxymethylene) the relaxation peak is small or absent in specimens quenched from the melt, but is enhanced by prior deformation or annealing. Its activation energy is less than 4 kcal/mole; the temperature of its maximum point is independent of pretreatment; and its width is nearly consistent with a single relaxation time.

Hiltner and Baer[4] have proposed a qualitative model based on kinks along dislocation lines, but they have not given the quantitative features such as the activation energy. Their model

is related to the present one in as much as the dislocation lines that they discuss could be constructed by combining twist disclinations into sets, or by activating the appropriate glide processes. However, because of interactions between the individual disclinations such collections of twist disclinations would be unlikely to yield relaxation peaks that would be as sharply defined as the observed ones.

ACKNOWLEDGEMENT

This work was supported in part by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with the University of Michigan.

REFERENCES

1. A Hiltner and E. Baer, "Relaxation Processes at Cryogenic Temperatures", CRC Critical Reviews in Macromolecular Science, April (1972).

2. J. J. Gilman and J. C. M. Li, J. Appl. Phys. 41, 4248 (1970).

3. C. D. Armeniades, E. Baer, and J. K. Rieke, "Cryogenic Relaxations of Polystyrene and Polyfluorinated Styrenes", J. Appl. Polymer Science 14, 2635 (1970).

4. A. Hiltner and E. Baer, "A Dislocation Mechanism for Cryogenic Relaxations in Crystalline Polymers", Polymer Journal 3, 378 (1972).

STATISTICAL MECHANICS OF POLYMER NETWORKS

H. Reiss

Abstract

The conventional theory of rubber elasticity is reviewed critically and discussed in terms of a simple gaussian network. An approximate theory, based on a variation principle, is then introduced and compared with the conventional theory. It is shown that this theory leads to the standard results of rubber elasticity when the restrictive assumptions, inherent in the conventional theory, are imposed on it. However, the new theory can be more simply extended to nongaussian networks, and several methods for achieving this extension are derived and discussed. In future work, specific applications of the new method to non-gaussian networks will be attempted.

STATISTICAL MECHANICS OF POLYMER NETWORKS

H. Reiss

I. Introduction

The statistical thermodynamics of polymer networks has been studied for many years, beginning with the theory of rubber elasticity. This theory has been advanced in several variant, although similar, forms by Kuhn,[1] James and Guth,[2-5] and Wall.[6] Many authors have contributed in the interim. These theories have been largely restricted to networks containing "chains" whose end-to-end distances are gaussian distributed. Although the James and Guth approach has the beginnings of greater generality, all of the methods are usually subjected to certain restrictive assumptions. These are:

(1) The nonconsideration of intra- (except for short range connectivity) and interchain potential energy effects.

(2) The assumption of an equilibrium distribution of chain end-to-end distances in the undeformed state of the rubber. This implies that the network is formed by putting together independent chains whose configurations are initially equilibrium-distributed.

(3) The assumption of affine deformation. This implies that the vectors connecting the junction points of the network deform in the same way as the vectors defining the local

continuum strain field.

(4) The arbitrary imposition of some macroscopic constraint such as maintenance, by the rubber, of constant volume during deformation.

As a result of these assumptions, rubber elasticity is reduced in its essentials to the theory of the deformation of a single polymer chain; the properties of the network appearing only through a factor which accounts for the plurality of chains, and because of the assumption of constant volume. For homogeneous tensile strain, the stress proves to be related to strain by

$$\tau = \frac{kT\nu_e}{V_o}\left(\alpha - \frac{1}{\alpha^2}\right) \quad . \tag{1}$$

Here α is the relative tensile extension

$$\alpha = L/L_o \quad , \tag{2}$$

where L_o is the undeformed length of the rubber, and L the length under deformation. ν_e represents the number of (effective) chains per cubic centimeter, V_o is the volume of the sample, while k is Boltzmann's constant and T the temperature.

More recent technology has found use for polymer networks which are definitely nonrubber-like. These extend all the way from polymeric glasses, through composites, to biological media such as fibrin clots. The individual network chains are non-gaussian, involve the effects of potential energy in an important way, and are usually affected by interchain effects. Obviously, it would be desirable to be able to treat such networks on a

molecular basis. On the other hand, the difficulties already encountered with the simple case of the gaussian network indicates that, for this purpose, a rigorous theory would be out of the question.

S. F. Edwards and K. F. Freed, in a series of joint and individual papers,[7] have introduced a fresh viewpoint by adopting a "self-consistent field" (SCF) approach coupled with abundant use of field theoretic techniques. They seem to be able to generate all the old results and promise to generate new ones in which many of the above-mentioned restrictive assumptions are unnecessary. Nevertheless, application of their elegant method to a real system is bound to be ponderous. For that reason, it is worthwhile investigating cruder techniques (in which the approximations are still more well defined than in the usual theory of rubber elasticity) which allow an easy treatment of important nongaussian networks. Such techniques are investigated below.

Before proceeding to this development, it should be mentioned that Sternstein[8] has developed a quasi-macroscopic treatment of network structures in which the mechanical properties of the individual network elements (chains - in the case of molecular elements). Sternstein's treatment may meet the molecular treatment at some intermediate point, and the union of both methods may prove valuable. However, we make no attempt in this direction in the present study.

II. Exact Analysis of a Gaussian Network

It is convenient to deal first with a "small" gaussian network whose thermodynamic properties can be evaluated exactly. Not only will the results clarify the difficulties associated with a full theory of networks, but we shall compare them with those obtained from an approximate theory which may be more easily extended to more complex systems. Because we will want to diagram some results in a plane, it is convenient, at first, to limit consideration to a two-dimensional system. No increase in difficulty occurs on passing to a three-dimensional network.

If $\vec{r}_i$ locates one end of a chain in our network while $\vec{r}_j$ locates the other, we shall assume that the probability of an end-to-end distance $\vec{r}_j - \vec{r}_i$, such $\vec{r}_i$ is fixed and $\vec{r}_j$ lies in $d\vec{r}_j$ is given by

$$P_{ij}(|\vec{r}_j-\vec{r}_i|)d\vec{r}_j = \frac{d\vec{r}_j}{\pi n_{ij} a^2} e^{-|\vec{r}_j-\vec{r}_i|^2/na^2} \quad . \qquad (3)$$

In this equation, it is assumed that the ij chain consists of n_{ij} links, each of which has a gaussian end-to-end distribution with a^2 the mean square end-to-end distance. It can be shown that the Helmholtz free energy of such a chain, with its ends <u>fixed</u> at $\vec{r}_i$ and $\vec{r}_j$, respectively, may, except for an uninteresting additive constant (which may, however, depend on temperature), be represented by

$$A_{ij}(|\vec{r}_j-\vec{r}_i|) = kT\{\frac{|\vec{r}_j-\vec{r}_i|^2}{n_{ij}a^2} + \ell n(\pi n_{ij} a^2)\} \quad . \qquad (4)$$

We shall treat the simple network shown in Figure 1,

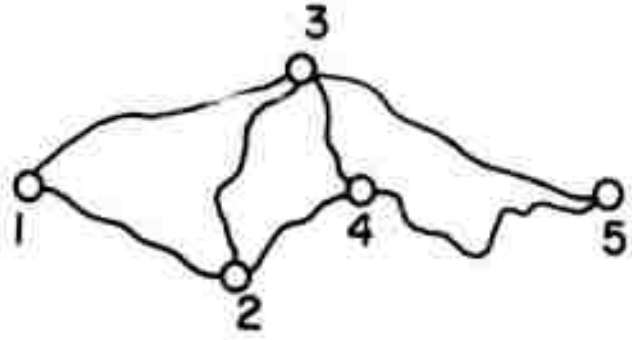


Figure 1

in which there are seven chains connected by five numbered junction points whose coordinates are $\vec{r}_1$, $\vec{r}_2$, $\vec{r}_3$, $\vec{r}_4$, $\vec{r}_5$. The seven chains have, as respective numbers of links, n_{12}, n_{23}, n_{24}, n_{13}, n_{35}, n_{34}, and n_{45}. The network lies in the xy-plane so that the various r's have x and y components.

The probability of a configuration such that the junction points lie in the volume element $d\vec{r}_1 d\vec{r}_2 d\vec{r}_3 d\vec{r}_4 d\vec{r}_5$ is given by

$$KP_{12}P_{13}P_{23}P_{24}P_{34}P_{35}P_{45}d\vec{r}_1 d\vec{r}_2 d\vec{r}_3 d\vec{r}_4 d\vec{r}_5 \quad , \qquad (5)$$

where K is a normalization constant and the P_{ij} are given by Eq. (3). In fact, substitution of Eq. (3) into Eq. (5) produces, as an exponent, a sum of two quadratic forms, one in the x's and one in the y's. Since both quadratic forms are identical except for the substitution of y by x or vice versa, we concentrate on the manipulation of the x-form. This is

$$\frac{(x_2-x_1)^2}{n_{12}} + \frac{(x_3-x_1)^2}{n_{13}} + \frac{(x_3-x_2)^2}{n_{23}} + \frac{(x_4-x_2)^2}{n_{24}}$$

$$+ \frac{(x_4-x_3)^2}{n_{34}} + \frac{(x_5-x_3)^2}{n_{35}} + \frac{(x_5-x_4)^2}{n_{45}} \quad ; \qquad (6)$$

which may be expanded and written in the alternative form

$$\{(\frac{1}{n_{12}} + \frac{1}{n_{13}})x_1^2 + (\frac{1}{n_{35}} + \frac{1}{n_{45}})x_5^2\}$$

$$+ \{[\frac{1}{n_{12}} + \frac{1}{n_{23}} + \frac{1}{n_{24}}]x_2^2 + [\frac{1}{n_{13}} + \frac{1}{n_{23}} + \frac{1}{n_{34}} + \frac{1}{n_{35}}]x_3^2$$

$$+ [\frac{1}{n_{24}} + \frac{1}{n_{34}} + \frac{1}{n_{45}}]x_4^2 - \frac{2}{n_{23}} x_2x_3 - \frac{2}{n_{24}} x_2x_4 - \frac{2}{n_{34}} x_3x_4\}$$

$$+ \{- \frac{2x_1}{n_{12}} x_2 - [\frac{2x_1}{n_{13}} + \frac{2x_5}{n_{35}}]x_3 - \frac{2x_5}{n_{45}} x_4\} \qquad (7)$$

We assume that junctions 1 and 5 are held at fixed positions (boundary conditions). Thus, x_1 and x_5 in Eq. (7) are to be treated as constants. Then the terms in the first curly bracket form a constant, Ω. We define the matrix Ç whose elements are

$$C_{22} = [\frac{1}{n_{12}} + \frac{1}{n_{23}} + \frac{1}{n_{24}}]$$

$$C_{33} = [\frac{1}{n_{13}} + \frac{1}{n_{23}} + \frac{1}{n_{34}} + \frac{1}{n_{35}}]$$

$$C_{44} = [\frac{1}{n_{24}} + \frac{1}{n_{34}} + \frac{1}{n_{45}}]$$

$$C_{23} = C_{32} = - \frac{1}{n_{23}}$$

$$C_{24} = C_{42} = -\frac{1}{n_{24}}$$

$$C_{34} = C_{43} = -\frac{1}{n_{43}} \tag{8}$$

and the vector $\vec{v}$ with components

$$v_2 = -\frac{2x_1}{n_{12}}$$

$$v_3 = -\left(\frac{2x_1}{n_{13}} + \frac{2x_5}{n_{35}}\right)$$

$$v_4 = -\frac{2x}{n_{45}} \quad . \tag{9}$$

Further, if by the vector $\vec{x}$ we mean (x_1, x_2, x_3), Eq. (7) may be expressed as

$$\Omega + \vec{x}^T \cdot \underset{\sim}{C} \cdot x + \vec{v}^T \cdot \vec{x} \quad , \tag{10}$$

where $\vec{x}^T$ and $\vec{v}^T$ are symbols for the transpose vectors. Introducing a unitary transformation, represented by the matrix $\underset{\sim}{Q}$, $\underset{\sim}{C}$ may be diagonalized to yield $\underset{\sim}{\Lambda}$,

$$\underset{\sim}{Q}^{-1} \cdot \underset{\sim}{C} \cdot \underset{\sim}{Q} = \underset{\sim}{\Lambda} \quad . \tag{11}$$

We also define the transformed vectors $\vec{\xi}$ and $\vec{u}$ by

$$\vec{x} = Q \cdot \vec{\xi} \quad , \tag{12}$$

$$\vec{v} = Q \cdot \vec{u} \quad . \tag{13}$$

Then it follows that

$$\vec{x}^T \cdot C \cdot \vec{x} = \vec{\xi}^T \cdot \underset{\sim}{\Lambda} \cdot \vec{\xi} = \sum_j x_j \xi_j^2 \tag{14}$$

and

$$\vec{v}^T \cdot \vec{x} = \sum_j u_j \xi_j \quad . \tag{15}$$

Since $\underset{\sim}{Q}$ is unitary, the Jacobian of the transformation is unity. The exponential in Eq. (5) may now be expressed in diagonalized form. With $\vec{r}_1$ and $\vec{r}_5$ held fixed,

$$\begin{aligned}
\frac{1}{K} = \frac{1}{K_{15}} &= \int_{-\infty}^{\infty} d\vec{r}_2 \int_{-\infty}^{\infty} d\vec{r}_3 \int_{-\infty}^{\infty} d\vec{r}_4 \; P_{12}P_{13}P_{23}P_{24}P_{34}P_{35}P_{45} \\
&= \left(\frac{1}{a^2}\right)^7 \frac{1}{n_{12}} \frac{1}{n_{13}} \frac{1}{n_{23}} \frac{1}{n_{24}} \frac{1}{n_{34}} \frac{1}{n_{35}} \frac{1}{n_{45}} \\
&\cdot \exp\left\{-\frac{1}{a^2}\left[\left(\frac{1}{n_{12}} + \frac{1}{n_{13}}\right)x_1^2 + \left(\frac{1}{n_{35}} + \frac{1}{n_{45}}\right)x_5^2\right]\right\} \\
&\cdot \prod_j \left\{\int_{\infty}^{\infty} e^{-\frac{1}{a^2}\left[\lambda_j \xi_j^2 + u_j \xi_j\right]} d\xi_j\right\} \\
&\cdot \text{(similar exponential factor for the y components).}
\end{aligned} \tag{16}$$

The integration in Eq. (16) may be performed immediately with the result

$$\int_{-\infty}^{\infty} e^{-\frac{1}{a^2}\left[\lambda_j \xi_j^2 + u_j \xi_j\right]} d\xi_j = a\left(\frac{\pi}{\lambda_j}\right)^{\frac{1}{2}} e^{u_j^2/4a^2\lambda_j} \quad ; \tag{17}$$

so that the product of integrals becomes

$$(a^2\pi)^{3/2}\left\{\prod_j \lambda_j^{-\frac{1}{2}}\right\} e^{\frac{1}{4a^2}\sum_j \frac{u_j^2}{\lambda_j}} \quad . \tag{18}$$

Now,

$$\sum_j \frac{u_j}{\lambda_j} = \vec{u}^T \cdot \underset{\sim}{\Lambda}^{-1} \cdot \vec{u} = \vec{v}^T \cdot \underset{\sim}{Q} \cdot \underset{\sim}{\Lambda}^{-1} \cdot \underset{\sim}{Q}^{-1} \cdot \vec{v} = \vec{v}^T \, \underset{\sim}{C}^{-1} \cdot \vec{v}$$

$$= \sum_{k\ell} \frac{C^{k\ell}}{|\underset{\sim}{C}|} v_k v_\ell \quad , \tag{19}$$

where $|\underset{\sim}{C}|$ is the determinant of $\underset{\sim}{C}$ and $C^{k\ell}$ is the cofactor of the $k\ell^{th}$ element. Furthermore,

$$\pi_j \lambda_j^{-\frac{1}{2}} = |\underset{\sim}{C}|^{-\frac{1}{2}} \quad . \tag{20}$$

Substituting Eqs. (19) and (20) into (18) and the result into (16) yields (after treating the factors belonging to the y components in the same way),

$$\frac{1}{K} = \left(\frac{1}{a^2}\right)^7 \frac{1}{n_{12}n_{13}n_{23}n_{24}n_{34}n_{35}n_{45}}$$

$$\exp\left\{-\frac{1}{a^2}\left[\frac{1}{n_{12}} + \frac{1}{n_{13}}\right]x_1^2 - \frac{1}{a^2}\left[\frac{1}{n_{35}} + \frac{1}{n_{45}}\right]x_5^2\right\} \tag{21}$$

$$\exp\left\{-\frac{1}{4a^2|\underset{\sim}{C}|}\sum_{k\ell} C^{k\ell}(v_{kx}v_{\ell x} + v_{ky}v_{\ell y})\right\} \quad ,$$

where v_{kx} and v_{ky}, etc., are components of the v vectors containing x's and y's, respectively.

The probability that junction 3 is at $\vec{r}_3$ when $\vec{r}_1$ and $\vec{r}_5$ are held fixed, is prescribed by the product of K_{15} by the integral appearing in Eq. (16) with the proviso that the integration over $\vec{r}_3$ is not performed. The remaining integral can be

performed in exactly the same manner used to arrive at Eq. (21), and the result (substituting for K_{1s} from Eq. (16)) is

$$P(\vec{r}_3) = \frac{|\underset{\sim}{C}|}{|C^*|}\frac{1}{a^2\pi}\exp\{-\frac{1}{a^2}[C_{33}(x_3^2+y_3^2) + (v_{3x}+v_{3y})x_3]\}$$

$$\exp\{\frac{1}{4a^2|\underset{\sim}{C}|}\sum_{k\ell} C^{k\ell}(v_{kx}v_{\ell x}+v_{ky}v_{\ell y}) \qquad (22)$$

$$-\frac{1}{4a^2|C^*|}\sum_{k\ell} C^{*k\ell}(v^*_{kx}v^*_{\ell x}+v_{ky}+v_{\ell y})\} \quad ,$$

where $\underset{\sim}{C}^*$ is the two-by-two matrix whose elements are

$$\begin{aligned} C^*_{22} &= C_{22} \\ C^*_{44} &= C_{44} \\ C^*_{24} &= C^*_{42} = C_{24} = C_{42} \quad ; \end{aligned} \qquad (23)$$

and $\vec{v}^*_x$, $\vec{v}^*_y$ are the vectors with components

$$\begin{aligned} v_{2x} &= -(\frac{2x_1}{n_{12}} + \frac{2x_3}{n_{23}}) \quad , \\ v_{4x} &= -(\frac{2x_3}{n_{34}} + \frac{2x_5}{n_{45}}) \quad , \\ v_{2y} &= -(\frac{2y_1}{n_{12}} + \frac{2y_3}{n_{23}}) \quad , \\ v_{4y} &= -(\frac{2y_3}{n_{34}} + \frac{2y_3}{n_{45}}) \quad . \end{aligned} \qquad (24)$$

After considerable algebraic manipulation, the exponential, dependent on $\vec{r}_3$ in Eq. (22), may be rewritten as

$$e^{-|\vec{r}_3-\vec{\alpha}_3|^2/\omega_3^2} \quad , \qquad (25)$$

in which

$$\omega_3^2 = a^2/C_{33} = a^2/\left(\frac{1}{n_{13}} + \frac{1}{n_{23}} + \frac{1}{n_{34}} + \frac{1}{n_{35}}\right) \quad , \tag{26}$$

and $\vec{\alpha}_3$ is a constant vector with components

$$\alpha_{3x} = -\frac{\begin{vmatrix} C_{22} & \frac{v_{2x}}{2} & C_{24} \\ C_{32} & \frac{v_{3x}}{2} & C_{34} \\ C_{42} & \frac{v_{4x}}{2} & C_{44} \end{vmatrix}}{|\underset{\sim}{C}|} \tag{27}$$

and

$$\alpha_{3y} = -\frac{\begin{vmatrix} C_{22} & \frac{v_{2y}}{2} & C_{24} \\ C_{33} & \frac{v_{3y}}{2} & C_{34} \\ C_{42} & \frac{v_{4y}}{2} & C_{44} \end{vmatrix}}{|\underset{\sim}{C}|}$$

Equation (25) shows several things. First, the junction point (in this case junction 3) is distributed in a <u>gaussian</u> manner about an average center of fluctuation specified by $\vec{\alpha}_3$ which itself depends only on the coordinates of those junctions which are held fixed. Furthermore, the "width" of the fluctuation is measured by ω_3 which depends <u>only</u> on the lengths of the chains connected to the junction in question. If the n's measuring

these lengths are large, then ω_3 is large (the network is soft in the neighborhood of junction 3).

These results are essentially the same as those of James and Guth[5] but specialized to the system of Figure 1. James and Guth show quite generally that the junction point distributions of a gaussian network are gaussian about a center whose location depends only on the fixed junctions and that the $\vec{\alpha}$ vectors are determined by the following set of linear equations, one for each nonfixed junction,

$$\sum_j \frac{(\alpha_{xj}-\alpha_{xi})}{n_{ij}} = 0 \quad , \tag{29}$$

$$\sum_j \frac{(\alpha_{yj}-\alpha_{yi})}{n_{ij}} = 0 \quad , \tag{30}$$

where again the sums over j extend over those junctions connected to the i^{th}. In fact, Eqs. (27) and (28) can be shown to be solutions of Eq. (30) for the system of Figure 1.

In anticipation of the approximate method to be introduced in the next section, we draw attention to the fact that $P(\vec{r}_i)$, as we have defined it, represents the distribution of the i^{th} junction averaged over all possible positions of the other junctions. This joint distribution of junctions <u>cannot</u> be represented as

$$P_{joint}(\vec{r}_1,\vec{r}_2,\ldots) = \prod_i P_i(\vec{r}_i) \quad , \tag{31}$$

where the product on the right goes over all nonfixed junctions. However, the "superposition approximation" inherent in Eq. (31) may still be a good one.

III. Junction Distribution from a Variation Principle

In the present section, we begin the development of the approximate (hopefully more tractable) theory which forms the goal of this investigation. We base it on a variation principle constructed during the 1971 ARPA Materials Conference and described in the Proceedings[9] of that Conference. The variation principle is as follows. Suppose $A(\{\vec{r}\})$ represents the Helmholtz free energy of a network when its junctions (whose collective free coordinates are symbolized by $\{\vec{r}\}$). Then it may be shown that the following functional of $P^*(\{\vec{r}\})$, A^*, represents an upper bound to the true free energy of the system:

$$A^* = \int P^*(\{\vec{r}\})A(\{\vec{r}\})d\{\vec{r}\} + kT \int P^*(\{\vec{r}\})\ell nP^*(\{\vec{r}\})d\{\vec{r}\}. \quad (32)$$

In this equation, the integration extends over the full space of the free junctions. A^* is obviously a function of the fixed junctions.

If we are concerned with a network in which there are no interchain interactions, then we may write

$$A(\{\vec{r}\}) = \sum_{ij} A_{ij}(\vec{r}_i,\vec{r}_j) \quad ,$$

where $\vec{r}_i$ and $\vec{r}_j$ are the coordinates of the i^{th} and j^{th} junctions, respectively, and $A_{ij}(\vec{r}_i,\vec{r}_j)$ is the free energy of the chain connecting these junctions when its ends are at $\vec{r}_i$ and $\vec{r}_j$, respectively. The sum in Eq. (33) goes over all chains ij. In most cases of interest, and certainly, in the case of gaussian chains, we may write

$$A_{ij}(\vec{r}_i,\vec{r}_j) = A_{ij}(|\vec{r}_j-\vec{r}_i|) \quad . \tag{34}$$

If, in the cases to which Eq. (33) or (34) applies, we choose as a trial function

$$P^*(\{\vec{r}\}) = \prod_i \phi_i(\vec{r}_i) \tag{35}$$

with the normalization requirement

$$\int \phi_i(\vec{r}_i)d\vec{r}_i = 1 \quad , \tag{36}$$

the $\phi_i(r_i)$ which minimize A* in Eq. (32) assume an especially simple form. Before proceeding to the investigation of this form, we note that P* in Eq. (35) is actually of the form exhibited in Eq. (31), so that it cannot be exact.

Substituting Eqs. (34) and (35) into (32) yields (using Eq. (36)),

$$\begin{aligned} A^* = &\sum_{ij} \int \phi_i(\vec{r}_i)\phi_j(\vec{r}_j)A_{ij}(|\vec{r}_j-\vec{r}_i|)d\vec{r}_i d\vec{r}_j \\ &+ kT \sum_i \int \phi_i(\vec{r}_i)\ell n\phi_i(\vec{r}_i)d\vec{r}_i \end{aligned} \tag{37}$$

where the ij sum goes over all chains. Taking the variation with respect to ϕ_i and subject to Eq. (36) yields for the form which extremalizes A*,

$$\phi_i(\vec{r}_i) = K_i \exp\{-\frac{1}{kT}\sum_j \int \phi_j(\vec{r}_j)A_{ij}(|\vec{r}_j-\vec{r}_i|)dr_j\} \tag{38}$$

in which K_i is a normalization constant the the $\sum_j$ in the exponent goes over all j junctions connected to the i^{th}.

Equation (38) represents a set of simultaneous nonlinear integral equations for the determination of the various ϕ_i. In general, they possess no analytical solution. However, for the special case that the network is gaussian, an analytical solution does exist!

We may regard a gaussian chain as itself made of links, each of whose lengths is gaussian distributed with a^2 representing the mean square end-to-end distance of a link.[10] Then if we restrict consideration to two dimensions, the probability that the ij^{th} chain has end-to-end distance $|\vec{r}_j-\vec{r}_i|$ is given by Eq. (3) and the free energy of the chain by Eq. (4). If we consider the three-dimensional case, we have

$$P_{ij}(|\vec{r}_j-\vec{r}_i|) = \left(\frac{3}{2\pi n a^2}\right)^{3/2} e^{-3|\vec{r}_j-\vec{r}_i|^2/2na^2} \tag{39}$$

and

$$A_{ij}(|\vec{r}_j-\vec{r}_i|) = \frac{3}{2}\,kT\left\{\frac{|\vec{r}_j-\vec{r}_i|^2}{n_{ij}a^2} - \ell n\,\frac{3}{2\pi n_{ij}a^2}\right\} \quad . \tag{40}$$

Equation (40) is substituted into Eq. (38). Now assume that the i^{th} solution of the resulting set of equations is specified by

$$\phi_i(\vec{r}_i) = \left(\frac{3}{2\pi\lambda_i^2}\right)^{3/2} e^{-3|\vec{r}_i-\vec{\alpha}_i|^2/2\lambda_i^2} \quad , \tag{41}$$

where $\vec{\alpha}_i$ is a constant vector and λ_i is a constant measuring the width of the gaussian distribution which Eq. (41) represents. If we substitute Eq. (41) into the set of Eq. (38) containing Eq. (40), it is found that Eq. (41) is indeed a solution of the set, provided that

$$\lambda_i^2 = a^2 / \sum_j \frac{1}{n_{ij}} \quad ; \tag{42}$$

and that

$$\sum_j \frac{(\alpha_{xj} - \alpha_{xi})}{n_{ij}} = 0 \quad ,$$

$$\sum_j \frac{(\alpha_{yj} - \alpha_{yi})}{n_{ij}} = 0 \quad , \tag{43}$$

$$\sum_j \frac{(\alpha_{zj} - \alpha_{zi})}{n_{ij}} = 0 \quad .$$

In these equations, the sums over j extend over the junctions connected to the i^{th} junction. Note that $\phi_i(\vec{r}_i)$ in Eq. (41) is normalized.

If we had considered a two-dimensional system, Eq. (4) would have been substituted into Eq. (38) and the normalized solution would be

$$\phi_i(\vec{r}_i) = \frac{1}{\pi \omega_i^2} e^{|\vec{r}_i - \vec{\alpha}_i|^2 \omega_i^2} \quad , \tag{44}$$

where again $\vec{\alpha}_i$ is a constant vector and ω_i a constant measuring the width of the distribution. These quantities are given by

$$\omega_i^2 = a^2 / \sum_j \frac{1}{n_{ij}} \tag{45}$$

and

$$\sum_j \frac{(\alpha_{xj} - \alpha_{xi})}{n_{ij}} = 0 \quad ,$$

$$\sum_j \frac{(\alpha_{yj} - \alpha_{yi})}{n_{ij}} = 0 \quad .$$

Equations (45) and (46) are identical, respectively, with Eqs. (26), (29) and (30). Thus, Eq. (44) is identical with

Eq. (25) (which is written for i = 3). As a result, we have

$$\phi_i(\vec{r}_i) = P_i(\vec{r}_i) \quad , \tag{47}$$

and P* is P_{joint} of Eq. (31). Thus, the "superposition approximation" to the joint distribution function specified by Eq. (31) is the "best" trail function which can be chosen as a product of one-junction functions, referred to the variation principle!

If we are interested in the distribution of the i^{th} junction averaged over all possible positions of the other junctions. Thus, at least for gaussian networks, the variation principle gives an excellent result not very far from exact.

IV. A Simple Example

For illustrative purposes, we work out the example of the network exhibited in Figure 1, subject to the restrictions that

$$\alpha_{x_1} = \alpha_{y_1} = 0 \quad ,$$

$$\alpha_{x_5} = 4a, \alpha_{y_5} = 0 \quad ; \tag{49}$$

and

$$\begin{aligned} n_{12} &= 2 \quad , \\ n_{13} &= 3 \quad , \\ n_{23} &= 1 \quad , \\ n_{24} &= 4 \quad , \\ n_{34} &= 5 \quad , \\ n_{35} &= 6 \quad , \\ n_{45} &= 7 \quad . \end{aligned} \tag{50}$$

Solving Eqs. (30) and using Eq. (45), the centers of the distributions of the free junctions (Nos. 2, 3, and 4) are found to be

$$\begin{aligned} \alpha_{2x} &= 0.272a, \quad \alpha_{2y} = 0 \quad , \\ \alpha_{3x} &= 1.067a, \quad \alpha_{3y} = 0 \quad , \\ \alpha_{4x} &= 1.729a, \quad \alpha_{4y} = 0 \quad . \end{aligned} \tag{51}$$

and the widths of the distributions obtained from Eq. (45) are

$$\begin{aligned} \omega_2 &= 0.756a \quad , \\ \omega_3 &= 0.772a \quad , \\ \omega_4 &= 1.299a \quad . \end{aligned} \tag{52}$$

These results are diagramed in Figure 2, where the circles indicate the widths of the distributions; and their centers, the centers of the distributions. Clearly, environment of junction 4 is much softer than that of junctions 2 and 3. Because of the boundary conditions, all junctions lie on the x-axis.

V. Application to Rubber Elasticity

The standard result of rubber elasticity, Eq. (1), is derived from the James and Guth theory by making the assumptions listed in Section I. However, the assumptions of affine deformation need only be made for the surface of the rubber specimen, since it may be shown that Eqs. (29) and (30) guarantee that all centers of junction point distributions will deform affinely if

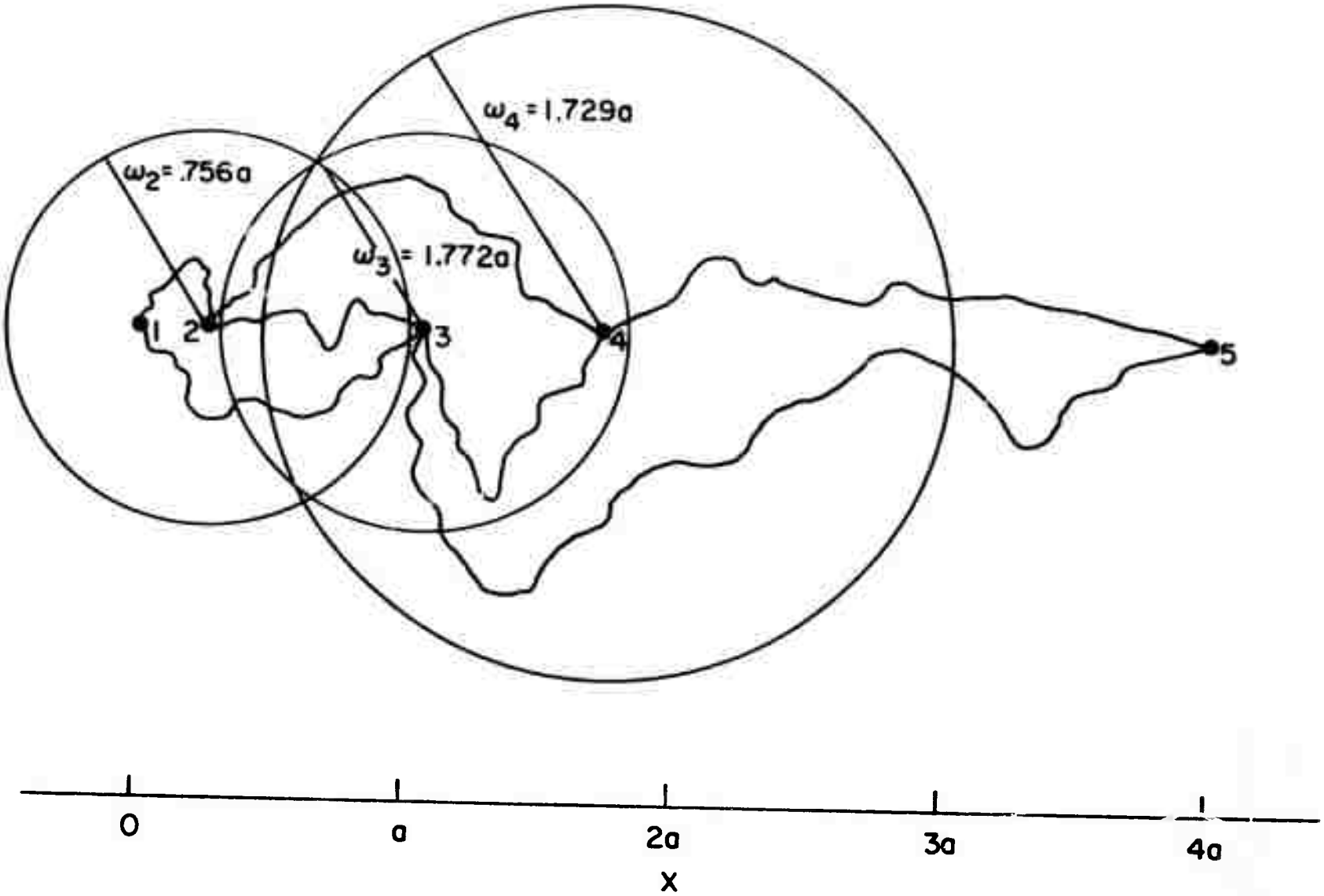


Figure 2

only the centers on the surface deform in this manner.

The theory of rubber elasticity, based on the variation principle discussed in Section III, will be shown below to yield the standard theory when the assumptions (Section I) of the standard theory are invoked. Thus, since $P_i(\vec{r}_i)$ of the James and Guth theory is identical with $\phi_i(\vec{r}_i)$ derived from the variation principle, and James and Guth have in effect used the superposition approximation, it appears as if the James and Guth theory is in fact the approximate one, derivable from the variation principle.

Before proceeding to the further development of the latter theory, we should note once again that the methods of Edwards and Freed[7] may not be subject to some of these restrictive assumptions unless they have been made implicitly.

For simplicity, we restrict attention to an incompressible specimen of rubber subjected to uniform tensile deformation. If L is the length and W the width of the specimen while L_o and W_o are the corresponding quantities in the undeformed state, the condition of incompressibility may be stated as

$$W^2L = W_o^2L_o = V_o \tag{53}$$

where V_o is the constant volume of the rubber. Then the requirement that the centers of gaussian distributions deform affinely takes the form

$$\frac{\alpha_{xj}-\alpha_{xi}}{\alpha^{o}_{xj}-\alpha^{o}_{xi}} = \frac{L}{L_o} = \alpha \quad , \tag{54}$$

$$\frac{\alpha_{yj}-\alpha_{yi}}{\alpha^{o}_{yj}-\alpha^{o}_{yi}} = \frac{\alpha_{zj}-\alpha_{zi}}{\alpha^{o}_{zj}-\alpha^{o}_{zi}} = \frac{W}{W_o} = \left(\frac{L_o}{L}\right)^{\frac{1}{2}} = \left(\frac{1}{\alpha}\right)^{\frac{1}{2}} \quad , \tag{55}$$

where α is the extension ratio (in the x-direction) as in Eq. (1) and should not be confused with the center of a gaussian distribution.

We proceed by adopting A* of Eq. (37) as the Helmholtz free energy of the network (it is actually only an upper bound). If we substitute ϕ_i from Eq. (41) into (37), we obtain

$$A^* = -\frac{3}{2}\,kT \sum_{ij} \ell n \frac{3}{2\pi n_{ij} a^2}$$

$$+ \frac{3}{2}\,kT \sum_{ij} \frac{1}{n_{ij}}\left[\lambda_i^2+\lambda_j^2+(\alpha_{jx}-\alpha_{ix})^2+(\alpha_{jy}-\alpha_{iy})^2+(\alpha_{jz}-\alpha_{iz})^2\right] \tag{56}$$

$$+ \frac{3}{2}\,kT \sum_{i} \ell n \frac{3}{2\pi\lambda_i^2} - \frac{3}{2}\,kT\,N_c \quad .$$

In this equation, the sums over ij go over all chains, while the sum over i extends over all free junctions. N_c is the number of chains in the network. Now the tensile stress in the x-direction is given by

$$\tau = \frac{1}{W_o^2}\frac{\partial A}{\partial L} \simeq \frac{1}{W_o^2}\frac{\partial A^*}{\partial L} \quad . \tag{57}$$

Substituting Eqs. (54) and (55) into (56), and the latter into Eq. (57), then yields

$$\tau = \frac{1}{W_o^2}\frac{\partial A^*}{\partial L} = \frac{3kT}{2a^2V_o}\sum_{ij}\left\{\frac{2(\alpha_{xj}^o-\alpha_{xi}^o)^2}{n_{ij}}\alpha - \frac{[(\alpha_{yj}^o-\alpha_{yi}^o)^2+(\alpha_{zj}^o-\alpha_{zi}^o)^2]}{n_{ij}}\frac{1}{\alpha^2}\right. \quad (58)$$

It is now assumed that the network is constructed in the undeformed state in such a manner that all chains of size n_{ij} have their equilibrium gaussian end-to-end distributions. Then for the sum over the N_n chains of size n, we have

$$\sum_{n_{ij}=n}(\alpha_{xj}^o-\alpha_{xi}^o)^2 = \sum_{n_{ij}=n}(\alpha_{yj}^o-\alpha_{yi}^o)^2 = \sum_{n_{ij}=n}(\alpha_{zj}^o-\alpha_{zi}^o)^2 = \frac{1}{3}N_n a^2 \quad (59)$$

Summing over all n, as required in Eq. (58), and substituting in that equation, yields

$$\tau = \frac{kTN_c}{a^2V_o}\left\{\alpha - \frac{1}{\alpha^2}\right. \quad , \quad (60)$$

which is identical with Eq. (1) if N_c is interpreted as ν_e, the effective number of chains. The effective number merely accounts for the fact that some chains within the network have free ends. This is a detail which we do not pursue in the present development.

Thus, we have shown that the variation principle leads to the conventional theory of rubber elasticity when the con-

ventional assumptions are made. It even leads to the correct singlet junction point distribution in the gaussian limit. It therefore appears to be a viable approximation. We now turn to the main purpose of this study, the development of an approximate theory for nongaussian networks, and investigate what the variation principle offers in this respect.

VI. Extension to Nongaussian Networks

Equation (38) still holds in the case of nongaussian networks provided that interchain interactions are excluded from consideration. However, if A_{ij} is no longer quadratic in $|\vec{r}_j-\vec{r}_i|$, an exact solution of this set of equations is no longer possible.

As a crude approximate solution, we might assume that the ϕ_j distributions are still spherical with respect to some center $\vec{\alpha}_j$, and in fact take ϕ_j to be the δ function

$$\phi_j(|\vec{r}_j-\vec{\alpha}_j|) = \frac{1}{4\pi|\vec{r}_j-\vec{\alpha}_j|^2}\,\delta\,(|\vec{r}_j-\vec{\alpha}_j|) \quad . \tag{61}$$

Then Eq. (38) yields

$$\phi_j(r_i) = K_i e^{-\frac{1}{kT}\sum_j A_{ij}(|\vec{\alpha}_j-\vec{r}_i|)} \quad , \tag{62}$$

which may be substituted into Eq. (37), and a theory of network elasticity developed in much the same manner as in Section V.

A somewhat more sophisticated procedure would involve expanding A_{ij} in a power series in some parameter μ. Denoting

the leading quadratic term by A_{ij}^{o}.

$$A_{ij} = A_{ij}^{o} + \mu A_{ij}^{(1)} + \mu^2 A_{ij}^{(2)} \quad . \tag{63}$$

Expanding ϕ_j in the same manner,

$$\phi_j = \phi_j^{o} + \mu\phi_j^{(1)} + \mu^2\phi_j^{(2)} + \ldots \quad ; \tag{64}$$

and substituting Eqs. (63) and (64) into (38) yields

$$\begin{aligned} \phi_i^{o} + \mu\phi_i^{(1)} = K_i e^{-\frac{1}{kT}\sum_j \int \phi_j^{o} A_{ij}^{o} d\vec{r}_j} \\ - \frac{K_i}{kT} e^{-\frac{1}{kT}\sum_j \int \phi_j^{o} A_{ij}^{o} d\vec{r}_j} \{\sum_j \int \phi_j^{(o)} A_{ij}^{(o)} d\vec{r}_j \\ + \sum_j \int \phi_j^{(1)} A_{ij}^{(o)} dr_j\} + \mu^2 \{ \quad \} + \ldots \quad . \end{aligned} \tag{65}$$

Equating coefficients of equal powers of μ yields

$$\phi_i^{(o)} = K_i e^{-\frac{1}{kT}\sum_j \int \phi_j^{o} A_{ij}^{o} d\vec{r}_j} \quad , \tag{66}$$

$$\phi_i^{(1)} = \frac{\phi_i^{(o)}}{kT} \{\sum_j \int \phi_j^{(o)} A_{ij}^{(1)} d\vec{r}_j + \sum_j \int \phi_j^{(1)} A_{ij}^{(o)} dr_j\} \quad . \tag{67}$$

Equation (66), which is identical with Eq. (38), shows that $\phi_i^{(o)}$ is given by Eq. (41). Then Eq. (67) represents a set of linear integral equations for the determination of $\phi_i^{(1)}$, and are presumably more easily soluble than the set of nonlinear integral equations represented by Eq. (38). Clearly, one can generate

linear sets of equations for the higher order perturbations.

All of the methods for nongaussian networks thus far presented have in common that they yield the singlet junction point distribution function in a form convenient for developing a theory of network elasticity patterned after the standard theory of rubber elasticity. As approximations, however, they suffer from the fact that the zeroth order solutions are themselves approximate. We therefore consider a method in which the zeroth order solutions are exact, but in which it may not be possible to pattern the theory of network elasticity after the more conventional approaches.

The method once again involves expanding $A = \sum A_{ij}$ in terms of some bookkeeping parameter μ (eventually to be set equal to unity) such that the zeroth order term is quadratic. Thus, we write, to the first order,

$$A = A^{(0)} + A^{(1)} = \sum_{ij} A_{ij}^{(0)} + \mu \sum_{ij} A_{ij}^{(1)} \quad , \tag{68}$$

where the sum extends over all chains. Corresponding to $A^{(0)}$, in this expansion, there are a set of normal coordinates $\vec{\xi}_x$, $\vec{\xi}_y$, and $\vec{\xi}_z$ specified by the transformation, Eq. (11), such that

$$\begin{aligned} \vec{x} &= \underset{\sim}{Q} \cdot \vec{\xi}_x \quad , \\ \vec{y} &= \underset{\sim}{Q} \cdot \vec{\xi}_y \quad , \\ \vec{z} &= \underset{\sim}{Q} \cdot \vec{\xi}_z \quad . \end{aligned} \tag{69}$$

These coordinates diagonalize $A^{(0)}$ so that Eq. (68) may be

expressed, in terms of them, as

$$A(\{\vec{\xi}=\vec{\xi}_x,\vec{\xi}_y,\vec{\xi}_z) = \sum_i A_i^{(o)}(\vec{\xi}_i=\vec{\xi}_{x_i},\vec{\xi}_{y_i},\vec{\xi}_{z_i}) + \mu A^{(1)}(\{\vec{\xi}\}). \quad (70)$$

We now return to Eq. (32), substitute Eq. (70), and choose for $P^*(\{\vec{\xi}\})$,

$$P^*(\{\vec{\xi}\}) = \prod_i \theta_i(\xi_i) \quad . \quad (71)$$

We note that when A is given by $A^{(o)}$ alone (the network is gaussian), the trial function, $P^*(\{\vec{\xi}\})$, given by Eq. (71) is capable of representing the <u>exact</u> solution, since the normal coordinates are indeed independent of one another.

Since the Jacobian of the transformation represented by $\underset{\sim}{Q}$ is unity, we may now express Eq. (32) as

$$A^* = \int P^*(\{\vec{\xi}\})A(\{\vec{\xi}\})d\{\vec{\xi}\} + kT \int P^*(\{\vec{\xi}\})d\{\vec{\xi}\} \quad (72)$$

and perform the variation with respect to $P^*(\{\vec{\xi}\})$.

Returning to $\{\vec{r}\}$ space for the moment, we recall that certain of the junctions are to be held fixed. The coordinates of these therefore contribute certain constant terms to A, which we denote by Γ. We could, of course, utilize Eq. (68) with higher order terms in μ, but suppose we retain only the linear term. Then we can write

$$A(\{\vec{r}\}) = \{A'(\{\vec{r}\}) + \Gamma\} + \mu A^{(1)}(\{\vec{r}\}) \quad , \quad (73)$$

in which

$$A' + \Gamma = A^{(o)}(\{\vec{r}\}) \quad ; \quad (74)$$

and note that it is the transform of A' which is diagonalized to $A^{(o)}(\{\vec{\xi}\})$,

$$A^{(o)}(\{\vec{\xi}\}) = A^{(o)}(\{\vec{r}\}) - \Gamma \tag{75}$$

or

$$A^{o}(\vec{r}) = \sum_i A_i^{(o)}(\vec{\xi}_i) + \Gamma \quad . \tag{76}$$

We also assume that θ_i can be represented by a function linear in μ,

$$\theta_i = \theta_i^{(o)} + \mu\theta_i^{(1)} \quad . \tag{77}$$

Substitution of Eq. (73), together with

$$A(\{\vec{\xi}\}) = \sum_i A_i^{(o)}(\vec{\xi}_i) + \Gamma + \mu A^{(1)}(\{\vec{\xi}\}) \tag{78}$$

into Eq. (72) and performing the variation with respect to the functions θ_i, subject to the condition

$$\int \theta_i d\vec{\xi}_i = 1 \tag{79}$$

or

$$\int \theta_i^{(o)} d\vec{\xi}_i = 1, \quad \int \theta_i^{(1)} d\vec{\xi}_i = 0 \quad , \tag{80}$$

leads, upon equating the coefficients of equal powers of μ, to

$$\int \prod_i \theta_i^{(o)} \{\sum_t A_t^{(o)} + \Gamma\} \frac{d(\{\vec{\xi}\})}{d\vec{\xi}_j} + kT\, \theta_j^{(o)}\, \ell n\, \theta_j^{(o)} = K_j' \theta_j^{(o)} \tag{81}$$

and

$$\theta_j^{(1)} = -\frac{1}{kT}\{B_j\theta_j^{(o)} + \int A^{(1)} \prod_i \theta_i^{(o)} \frac{d(\{\vec{\xi}\})}{d\vec{\xi}_j}\} \quad , \tag{82}$$

in which both K_j' and B_j are constants. In deriving Eqs. (80) and (81), use has been made of Eq. (79). The derivation is straightforward but somewhat tedious; terms must be collected rather carefully since many of them cancel.

Applying Eq. (79) again to (80), and performing some of the indicated integrations, yields

$$\theta_j^{(o)} = \frac{e^{-A_j^{(o)}/kT}}{\int e^{-A_j^{(o)}/kT} d\vec{\xi}_j} \quad , \tag{83}$$

while application of Eq. (80) to Eq. (82) determines

$$\begin{aligned} B_j &= -\int A^{(1)} \prod_i \theta_i^{(o)} d\{\vec{\xi}\} \\ &= B \text{ independent of } j. \end{aligned} \tag{84}$$

As indicated earlier, the result for $\theta_j^{(o)}(\vec{\xi}_j)$ is exact. These solutions can be transformed back to $\{\vec{r}\}$-space with relative ease. It is advantageous to perform this transformation since then $A^{(1)} = \sum_{ij} A_{ij}^{(1)}$, and is represented as a sum of individual chain functions, whereas this is not true in $\{\vec{\xi}\}$-space.

For the development of a theory of network elasticity, A* should be expressed in terms of the θ_i. For this purpose, we substitute Eq. (71), together with Eq. (77), into Eq. (72), obtaining,

$$A^* = \sum_i \int \theta_i A_i^{(o)} d\vec{\xi}_i + \Gamma + \mu \int (\pi \theta_i) A^{(1)} d\{\vec{\xi}\} + kT \sum_i \int \theta_i \ell n \theta_i d\vec{\xi}_j \quad , \tag{85}$$

where Eq. (80) has been used. Dropping terms beyond the linear and using Eq. (77), we get

$$A^* = \sum_i \int \theta_i^o A_i^{(o)} d\vec{\xi}_i + \Gamma + kT \sum_i \int \theta_i^{(o)} \ell n \theta_i^{(o)} d\vec{\xi}_i + \mu \{ \sum_i \int \theta_i^{(1)} A_i^{(o)} d\vec{\xi}_i$$

$$+ \int (\pi \theta_i^{(o)}) A^{(1)} d\{\vec{\xi}\} + kT \sum_i \int \theta_i^{(1)} \ell n \theta_i^{(o)} d\vec{\xi}_i \} \quad , \tag{86}$$

where, again, Eq. (80) has been used. Combining Eq. (85) with (84) gives

$$\theta_i^{(1)} = \frac{1}{kT} \{ \theta_i^{(o)} \int (\pi \theta_i^{(o)}) A^{(1)} d\{\vec{\xi}\} - \int (\pi \theta_j^{(o)}) A^{(1)} \frac{d\{\vec{\xi}\}}{d\vec{\xi}_i} \} \quad . \tag{87}$$

It is convenient to introduce certain definitions and to list a number of relations. Thus, we write

$$\pi \theta_j^{(o)} = P_o^*(\{\vec{\xi}\}) \quad . \tag{88}$$

Since the Jacobian of the transformation is unity, we find

$$P_o^*(\{\vec{\xi}\}) = P_o^*(\{\vec{r}\}) \quad . \tag{89}$$

We also recall (making use of Eqs. (81) and (88)) that

$$\sum_i \int \theta_i^{(o)} \ell n \theta_i^{(o)} d\vec{\xi}_i = \int P_o^*(\{\vec{\xi}\}) \ell n P_o^*(\{\vec{\xi}\}) d\{\vec{\xi}\};$$

$$\int \theta_i^{(o)} f(\vec{\xi}_i) d\vec{\xi}_i = \int P_o^*(\{\vec{\xi}\}) f(\vec{\xi}_i) d\{\vec{\xi}\} \quad . \tag{90}$$

Furthermore, we define the average of a function F over the zeroth order distribution as

$$\langle F\rangle_0 = \int P_0^*(\{\vec{\xi}\})F(\{\vec{\xi}\})d\{\vec{\xi}\} = \int P_0^*(\{\vec{r}\})F\{\vec{r}\})d\{\vec{r}\} \quad . \quad (91)$$

Substituting Eq. (87) into (86), and using Eqs. (88) through (91), we obtain

$$\begin{aligned} A^* = {} & \langle A^{(o)}(\{\vec{\xi}\})+\Gamma\rangle_0 + \langle \mu A^{(1)}(\{\vec{\xi}\})\rangle_0 + kT\langle \ell n P_0^*(\{\vec{\xi}\})\rangle_0 \\ & + \frac{\mu}{kT}\{\langle A^{(o)}(\{\vec{\xi}\})\rangle_0\langle A^{(1)}(\{\vec{\xi}\})\rangle_0 - \langle A^{(o)}(\{\vec{\xi}\})A^{(1)}(\{\vec{\xi}\})\rangle_0 \\ & + kT\,\langle \ell n P_0^*(\{\vec{\xi}\})\rangle_0\langle A^{(1)}(\{\vec{\xi}\})\rangle_0 \\ & - kT\,\langle \ell n P_0^*(\{\vec{\xi}\})A^{(1)}(\{\vec{\xi}\})\rangle_0\} \quad . \end{aligned} \quad (92)$$

Transforming to $\{\vec{r}\}$-space, noting the second of Eqs. (91) and using Eq. (77), gives

$$\begin{aligned} A^* = {} & \langle A^{(o)} + \mu A^{(1)}\rangle_0 + kT\langle \ell n P_0^*\rangle_0 \\ & + \frac{\mu}{kT}\{\langle A^{(o)}\rangle_0\langle A^{(1)}\rangle_0 - \langle A^{(o)}A^{(1)}\rangle_0 \\ & - kT\,\langle \ell n P_0^*\rangle\langle A^{(1)}\rangle_0 - kT\,\langle (\ell n P_0^*)A^{(1)}\rangle_0\} \quad . \end{aligned} \quad (93)$$

Since now $A^o = \sum_{ij} A_{ij}^{(o)}$, $A^{(1)} = \sum_{s\ell} A_{s\ell}^{(1)}$, if we make the approximation

$$P_0^*(\{\vec{r}\}) = \pi_i \phi_i^{(o)}(\vec{r}_i) \quad , \quad (94)$$

Eq. (94) becomes

$$\begin{aligned} A^* = {} & \langle A^{(o)}+\mu A^{(1)}\rangle_0 + kT\langle \ell n P_0^*\rangle_0 \\ & + \frac{\mu}{kT}\{\sum_{ij}\sum_{s\ell}[\langle A_{ij}^{(o)}\rangle_0\langle A_{ij}^{(1)}\rangle_0 - \langle A_{ij}^{(o)}A_{s\ell}^{(1)}\rangle_0] \\ & + kT\sum_i\sum_{s\ell}[\langle \ell n\phi_i^o\rangle_0 A_{s\ell}^{(1)}\rangle_0 - \langle A_s^{(1)}\ell n\phi_i^o\rangle_0]\} \quad . \end{aligned} \quad (95)$$

This equation could also have been obtained through the manipulation of Eqs. (66) and (67). In fact, if we represent $\phi_i^{(o)}$ by Eq. (66), it may be demonstrated that the second term in curly brackets in Eq. (95) is exactly twice the negative of the first term. Thus, Eq. (95) may be simplified to

$$A^* = \langle A^{(o)} + \mu A^{(1)} \rangle_o + kT \langle \ell n P_o^* \rangle_o$$
$$- \frac{\mu}{kT} \sum_{ij} \sum_{s\ell} [\langle A_{ij}^{(o)} \rangle_o \langle A_{s\ell}^{(1)} \rangle_o - \langle A_{ij}^{(o)} A_{s\ell}^{(1)} \rangle_o] \quad . \tag{96}$$

The last term in this equation is a fluctuation expression measuring the correlation between $A^{(o)}$ and $A^{(1)}$. If $P_o^*(\{\vec{r}\})$ is given by Eq. (94), it is clear that the sum will only contain terms such that i,j,s,ℓ are not all different; i.e., terms for which the chains ij and sℓ are the same chains, or, at least, have a junction in common. In all other cases,

$$\langle A_{ij}^{(o)} A_{s\ell}^{(1)} \rangle_o = \langle A_{ij}^{(o)} \rangle_o \langle A_{s\ell}^{(1)} \rangle_o \quad , \tag{97}$$

so that these terms will not survive in Eq. (96). If N_c represents the total number of chains, the fluctuation term will thus be of the order of N_c. Since $A^{(1)}$ is also of this order, the fluctuation cannot be neglected in the first order correction to A*. When ϕ_i^o corresponds to a fairly narrow distribution, however, it may not be large.

Equation (94) is of course identical with Eq. (35), except that the zero subscript and superscript are used to

denote the zeroth order solution corresponding to a gaussian network (as does the function in Eq. (35)). Nevertheless, as indicated in Section III, Eq. (35) is an excellent approximation.

The first two terms in Eq. (93) may be expressed as

$$A^* = \int P_0^*(\vec{r})A(\{\vec{r}\})d\{\vec{r}\} + kT \int P_0^*(\{\vec{r}\})\ell nP^*(\{\vec{r}\})d\{\vec{r}\} \quad . \quad (98)$$

This is the free energy which the system would have if it were constrained to have the zeroth order distribution, but the actual free energy corresponding to the configuration $\{\vec{r}\}$ remained exact. This suggests an interesting method for achieving an approximation capable of representing the system out to all orders.* The function $\phi_i^0(\vec{r})$ in Eq. (95) is given by Eq. (41), in which the λ_i are fixed quantities! Suppose we assume the general validity of Eq. (98), but allow the λ_i's to be undetermined parameters which can be adjusted so that P_0^* is no longer the zeroth order solution but is altered, in some way, to account for the higher order perturbations not included when P_0^* is restricted to the proper zeroth order function. Equation (98) will still be an upper bound for the true free energy since $A(\{\vec{r}\})$ is still exact; it is the same equation as Eq. (32) except that P_0^* appears where P^* appeared in the latter. Thus, the method of adjustment of the λ's is apparent. Perform the integrations in Eq. (94) and then find the minimum of A^* with respect to the various λ_i. The value of A^* corresponding to

*Suggested by K. F. Freed in private communication to the author.

this minimum can then be used, together with the assumption of affine deformation to achieve a theory of network elasticity.

In practice, one might simplify the process by restricting all λ_i to be the same and equal to λ. A_{ij} can always be expanded in a power series in $|\vec{r}_j-\vec{r}_i|$. Since the ϕ_i are now gaussians, the integrals of the form

$$\int \phi_i(\vec{r}_i-\vec{\alpha}_i)\phi_j(\vec{r}_j-\vec{\alpha}_j)|\vec{r}_j-\vec{r}_i|^n d\vec{r}_i d\vec{r}_j \qquad (99)$$

can always be done so that A* will once again appear as a function of the various $|\vec{\alpha}_j-\vec{\alpha}_i|$, and the assumption of affine deformation therefore conveniently used. We plan to explore this method in subsequent investigations.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with the University of Michigan.

REFERENCES

1. W. Kuhn, Kolloid Z. 68, 2 (1934).
2. E. Guth and H. M. James, Ind. Eng. Chem. 33, 1060 (1941).
3. H. M. James and E. Guth, Phys. Rev. 59, 111 (1941).
4. H. M. James and E. Buth, Ind. Eng. Chem. 34, 1365 (1942).
5. H. M. James and E. Guth, J. Chem. Phys. 11, 455 (1943).
6. F. T. Wall, J. Chem. Phys. 10, 132 (1942).
7. S. F. Edwards, Proc. Phys. Soc. (London) 91, 513 (1967); J. Phys. A1, 15 (1968); Proc. Phys. Soc. (London) 92, 9 (1967); J. Phys. C3, 1 (1969); Discussions, Faraday Soc. 49, 43 (1970); S. F. Edwards and K. F. Freed, J. Phys. A2, 145 (1969); J. Phys. C3, 739 (1970); J. Phys. C3, 750 (1970); J. Phys. C3, 760 (1970); K. F. Freed, J. Chem. Phys. 54, 1453 (1971); J. Chem. Phys. 55, 5588 (1971).
8. S. S. Sternstein and G. M. Lederle, *Polymer Networks: Structural and Mechanical Properties* (Plenum, New York, 1971), ed. by A. J. Chompff and S. Newman, p. 57; S. S. Sternstein, *Cellulose and Cellulose derivatives* (Wiley, New York, 1971), ed. by N. Bikales and L. Segal, Vol. 5, part 4, p. 649.
9. H. Reiss, "Preliminary Reports, Memoranda, and Technical Notes of ARPA Materials Summer Conference," Woods Hole (1971), p. 351.
10. H. Yamakawa, *Modern Theory of Polymer Solutions* (Harper and Row, New York, 1971), p. 16.

STRESS AVERAGING IN THE DISLOCATION MICROMECHANICS ANALYSIS OF DEFORMATION

J. P. Hirth

Abstract

The strain-rate of a deforming crystal is related to a sum over dislocation segments of isolated thermally activated and viscously damped motions. Appropriate averaging methods are suggested to yield a relation between strain rate and the macroscopic variables of stress and temperature. Several specific examples are analyzed in detail.

STRESS AVERAGING IN THE DISLOCATION MICROMECHANICS ANALYSIS OF DEFORMATION

J. P. Hirth

Introduction

Dorn and his coworkers have contributed significantly to the development of constitutive relations relating strain rate to experimental variables for a deforming specimen, particularly for high temperature creep, e.g., ref. 1 and 2. An important topic relating to the development of such constitutive relations from theories for motion of individual dislocations or dislocation segments is that of stress and strain averaging over the geometry of a deforming crystal.

The concept of averaging the resolved shear stress over slip systems and then relating this average value to a critical value for yield or flow is a familiar one in macroscopic crystal plasticity.[3,4] This concept has been extended to the interpretation of polycrystal deformation in terms of dislocation models in what is essentially the athermal limit where deformation is completely driven by the applied stress.[5,6] In the latter case, the average value of the resolved shear stress over the five most highly stressed systems, the Taylor value[3] of crystal plasticity, is related to the critical resolved shear stress for dislocation motion. Other mechanistic

interpretations of dislocation motion have been proposed, however, and there is considerable experimental support for them in terms of such parameters as the measured thermal activation energy. The latter mechanisms include several forms of stress-assisted, thermally-activated dislocation motion and of viscously damped motion. In the present paper, we propose appropriate simplified averaging methods to include stress resolution approximately in these latter models.

Dislocation Micromechanics Analysis of Deformation

In many cases, the deformation of metal crystals is analyzed in terms of theories of activated motion of dislocations, reviewed in ref. 7, sometimes referred to as dislocation micromechanics. Figure 1 is a prototype illustration of a dislocation moving by thermally activated breakaway from pinning points. The dislocation is stationary while pinned and moves at nominally constant velocity v_D under some viscous damping control mechanism while free. Let us consider the time required for it to move from point 4 to point 5 sweeping out an area A, or distance A per unit length of line.

$$t = \frac{A}{v_D} + \frac{1}{\nu} \tag{1}$$

where ν is the frequency with which pins are broken. Then the mean velocity of the dislocation is

$$v = \frac{A}{t} = \frac{A\nu v_D}{A\nu + v_D} \tag{2}$$

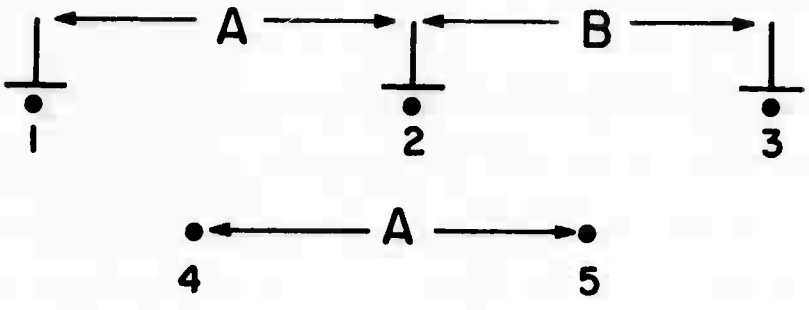


Figure 1

Unit length of dislocation moving from pinning point to pinning point (1,2,3) by thermally activated flow. Reference marks 4 and 5 are also shown

The terms A, ν and v_D in general vary for segments throughout the system for several reasons. These terms depend on the local resolved stress during the process, on the local defect distribution which influences both the internal interaction stress between the defects and the segment in question and the pin spacing, and on the screw-edge character of the dislocation segment which varies from segment to segment as a function of dislocation orientation.

In the limit of weak pins or high temperatures $A\nu >> v_D$ and the motion becomes damping controlled

$$v = v_D \tag{3}$$

In the low temperature limit $v_D >> A\nu$ and the motion becomes activation controlled

$$v = A\nu \tag{4}$$

Usually, Eq. (3) or (4) is used to test experimental data. Indeed, for pure metals, the damping constants are such[8] that Eq. (4) should be a good approximation for many experimental situations. However, cases exist, such as the drag of Cottrell, Snoek or core atmospheres of solute atoms,[9] where the condition $v_D \simeq A\nu$ can obtain. Functional fitting to Eq. (2) should be done for self-consistency in the latter cases, and, since such fitting is of little more difficulty than the use of Eqs. (3) or (4), is recommended in general.

Except under shock-loading conditions, the form of v_D is a linear viscous form[10]

$$v_D = B\sigma b \tag{5}$$

where B is the damping coefficient, σ is the local effective resolved shear stress on the glide plane and b is the dislocation Burgers vector. The factor B varies with screw-edge character, temperature and solute concentration. The effective stress σ, the sum of the resolved applied stress and the resolved internal stress, varies from point to point both because of changes in slip system orientation and because the internal stress varies with the local defect concentration. There is no general agreement on the form for ν. The various forms proposed for this term include the following

$$\nu = \nu_o \exp(-Q/kT) \exp(\sigma bA/kT) \tag{6}$$

which is the form for highly irreversible, stress-assisted, thermally activated flow.[1,11] Here, Q is the activation energy, k is Boltzmann's constant, T is absolute temperature and ν_o is a pre-exponential frequency factor.

$$\nu = \nu_o \exp(-Q/kT) \sinh(\sigma bA/kT) \tag{7}$$

which is the form for nearly reversible, stress-assisted, thermally activated flow.[7,12,13]

$$\nu = \nu_o \sigma^n \exp(-Q/kT) \tag{8}$$

an alternate thermally activated form.[14]

$$\nu = \nu_0 \exp(-C/\sigma) \tag{9}$$

which is both a tunneling form[15] and, when combined with a term exp (-Q/kT), another alternate thermally activated form.[7] Mixed forms are also used[16,17]: for example if ν_0 in Eq. (6) or (7) contained a factor σ^n, the form would be a mixed form of Eq. (6) or (7) together with (8).

With the various expressions for the terms in Eq. (2), the velocity of individual dislocations is related to the overall strain rate $\dot{\varepsilon}$ by

$$\dot{\varepsilon} = \sum_i m_i L_i b v_i / V \tag{10}$$

where m_i is a strain resolution factor, L_i is the segment length, all for the i^{th} segment, and V is the crystal volume. The usual assumption (see the discussion in refs. 7, 15 and 18) is that v_i is the same value v for all segments and that the sum of $m_i L_i / V$ is simply the average Schmid factor $m = \overline{m_i}$, multiplied by the mobile dislocation density $\sum L_i / V = \rho$. Then Eq. (10) reduces to

$$\dot{\varepsilon} = m\rho\, bv \tag{11}$$

With Eq. (6), say, Eq. (11) is typically further reduced to the form

$$\dot{\varepsilon} = \dot{\varepsilon}_0 \exp(-Q/kT) \exp(\sigma bA/kT) \tag{12}$$

In some cases,[15,19] models for the variation of ρ and v with plastic strain as a result of dislocation multiplication are incorporated in Eq. (11) so that in forms such as Eq. (12) an explicit dependence of $\dot{\varepsilon}_o$ and A on strain, stress and temperature results. More often, however, $\dot{\varepsilon}_o$ is assumed to be constant. In almost all cases, though, the only account taken of averaging is via the factor m which is implicitly assumed to be a direct geometrical average of m_i.

Evidently, since the effective resolved shear stress σ varies from segment to segment, and since parameters such as A, B and ν_o vary from point to point as discussed previously, the assumption that geometrical averaging only is necessary for m_i is a poor one. While it is recognized that the assumption is poor, the assumption is generally made anyway because the general averaging problem (over $\sim 10^{15}$ entities) is inordinately complicated. Mura[20,21] has presented the general tensor form of Eq. (10) which can in principal be summed. However, even with present day computers, the general problem is not tractable.

As an interim between the above roughest approximation and the general solution, we propose simplified averaging procedures to include the averaging of all parameters in Eq. (10) which explicitly depend on σ. While the forms are somewhat complex, they can be handled readily with modern computer techniques.

Stress Resolution Modifications

For example, we treat the simple tension test, the most common testing form in mechanistic studies. Then the strain and stress resolution factor (Schmid factor) is

$$m_i = \cos\theta \cos\beta \tag{13}$$

where θ is the angle between the glide plane normal and the tensile axis and β is the angle between the glide direction and the tensile axis. Eq. (10) can then be modified with the inclusion of any of the forms for v_i by replacing σ by

$$\sigma = \sigma_a m_i + \sigma_I$$

where σ_a is the applied stress and σ_I is the internal stress. In general, σ_I will vary with σ_a, T, and m_i and in such a case must then be included in the subsequent integrals over m_i. In the following examples, since little data is available for σ_I, we assume it to be negligible with respect to $\sigma_a m_i$, an approximation thought to be valid at low temperatures.

The factor L_i also may vary with m_i because of preferred orientation in single crystals or textured polycrystals. The sum in Eq. (10) is then performed over all orientations. In many cases, however, for a polycrystal, L_i is random and the sum in Eq. (10) can be replaced by an integral over all orientations. This procedure should yield a reasonable approximation even for a single crystal when the crystal system in question has a multiplicity of glide systems as for cubic

crystals. The simplest case is that of damping control, Eqs. (3) and (5), which, together with Eq. (10) yield

$$\dot{\varepsilon} = b^2 B \sigma_a \int m_i^2 \, (L_i/V) dm_i \tag{14}$$

In carrying out the integral over m_i, it is convenient to introduce both Eulerian and spherical coordinates, Fig. 2. If the segments L_i are uniformly distributed in the glide plane, if the glide plane poles are distributed uniformly over solid angle increment $\sin\theta \, d\theta \, \phi$ in spherical coordinates and if for each glide plane orientation, the glide directions are uniformly distributed over the Eulerian coordinate increment $d\kappa$, then

$$(L_i/V) dm_i = \rho \sin\theta \, d\theta \, d\varphi \, d\kappa \tag{15}$$

In these coordinates $\cos\beta = \sin\theta \sin\kappa$, and Eq. (14) becomes

$$\dot{\varepsilon} = \frac{\rho b^2 B \sigma_a}{8\pi^2} \int_0^{2\pi} \int_0^{2\pi} \int_0^{\pi} (\cos\theta \sin\theta \sin\kappa)^2 \sin\theta d\theta d\kappa d\phi \tag{16}$$

The integral over $d\phi$ is trivial. The integrals over ϕ and κ much be performed with care, however, The dislocations move to contribute positively to $\dot{\varepsilon}$ independent of the sign of m_i. Thus, the integrals over θ and κ must be performed over absolute values of the sin and cos functions. Thus,

$$\dot{\varepsilon} = \frac{2\rho b^2 B \sigma_a}{\pi} \int_0^{\pi/2} \int_0^{\pi/2} \cos^2\theta \sin^3\theta \sin^2\kappa d\theta d\kappa = \rho b^2 B \sigma_a / 15 \tag{17}$$

In the analysis of data fitted to expressions of the above type,

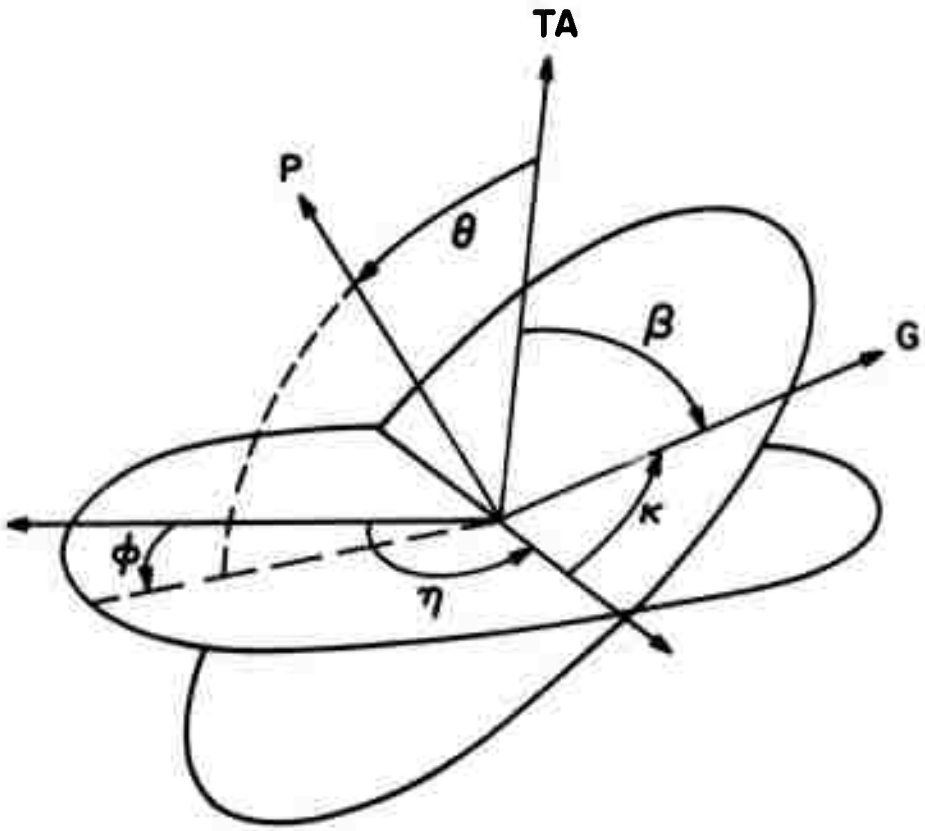


Figure 2

Spherical coordinates θ, ϕ and Eulerian coordinates η, κ, θ in relation to glide plane coordinates θ and β. The tensile axis, TA, glide plane pole, P, and glide direction, G, are depicted.

there is some concern about separation of the dislocation density ρ into a mobile and an immobile fraction. With the present averaging scheme, in order to be consistent with the initial premise, that motion is strictly damping controlled, one should insert the <u>total</u> dislocation density into Eq. (17) to extract data on the mobility B as a function of $\dot{\varepsilon}$ and σ_a. Related to the total dislocation density usage, the random distribution average value m_A of m_i over all orientations is $m_A = 0.212$ in contrast to the Taylor value[3,4] $m_T = 0.327$ for face-centered-cubic crystals. The Taylor value is obtained with various possible assumptions,[3,4] but they are equivalent to the assumptions that only the five most highly stressed systems in cubic crystals contribute, that each of these satisfies a critical resolved shear stress criterion, and that the crystals are randomly distributed as in the present case. The reason for the difference in these m values is obvious since the seven systems not averaged in the latter case have m_i values lower or equal to those averaged. We emphasize that we do not propose that m_A replace m_T for a critical-resolved-shear-stress plasticity theory. The Taylor value arises from the compatibility requirement that five independent slip (deformation) systems are required for an arbitrary strain. In the context of the present paper, the compatibility conditions would appear as a dependence of the internal stress σ_I on position in the vicinity of a grain boundary. Such compatibility effects have been demonstrated on the microscale for both isotropic[22,23] and anisotropic elastic[24]

crystals, and some discussion of the implications of such effects on macroscopic plastic behavior has been presented.[4] Here, we are focussing on the method of averaging, given a model, and the above compatibility effects could be included if an explicit model for the internal stress variation were developed.

The other forms of v_i, when inserted into Eq. (10), lead to analytically intractable forms. However, they can be solved readily by numerical methods. As perhaps the most used thermal activation expression, we consider as an example the case of Eqs. (4) and (6), assuming that ν_o and A are not functions of σ_a, θ, κ and ϕ. Again the distribution of segments is assumed to be uniform so that Eq. (15) applies. Then

$$\dot{\varepsilon} = \sum m_i L_i b v_i / V$$

$$= \sum (m_i L_i b/V \; A\nu_o \exp(-Q/kT) \exp(\sigma_a m_i bA/kT)$$

$$= \frac{A\nu_o \rho b}{8\pi^2} \exp - \frac{Q}{kT} \frac{4}{\pi^2} \int_0^{\pi} \int_0^{2\pi} \int_0^{2\pi} \sin^2\theta\cos\theta\sin\kappa \exp$$

$$\times \left(\frac{\sigma_a bA \sin\theta\cos\theta\sin\kappa}{kT}\right) d\theta d\kappa d\phi \qquad (18)$$

$$= \frac{\pi A\nu_o \rho b}{2} \exp - \frac{Q}{kT} \frac{4}{\pi^2} \int_0^{\pi/2} \int_0^{\pi/2} \sin^2\theta\cos\theta\sin\kappa \exp$$

$$\times (D\sin\theta\cos\theta\sin\kappa) \; d\theta d\kappa$$

$$= \frac{\pi A\nu_o \rho b}{2} \exp - \frac{Q}{kT} f_2(D)$$

Here the abbreviation $D = \sigma_a bA/kT$ is involved and the factor $(2/\pi)^2$ is carried along with the integral because of the form of the latter as a Bessel function or a related function. The first remaining integral in Eq. (18) has the solution

$$f_1(E) = \frac{2}{\pi} \int_0^{\pi/2} \sin\kappa \exp(E \sin\kappa) d\kappa \qquad (19)$$

$$= \frac{2}{\pi} + I_1(E) + L_1(E)$$

where $E = D\sin\theta\cos\theta$, I_1 is the modified Bessel function of the first kind and L_1 is the modified Struve function. The quantity $f_1(E)$ is tabulated in Table 1 and shown graphically in Fig. 3. With $f_1(E)$ known the desired factor in Eq. (18) is given by

$$f_2(D) = \frac{2}{\pi} \int_0^{\pi/2} \sin^2\theta\cos\theta \; f_1(D\sin\theta\cos\theta) d\theta \qquad (20)$$

This function, evaluated numerically, is tabulated in Table 2 and presented graphically in Fig. 4. Since $\dot{\varepsilon}(D)$ is directly proportional to $f_1(D)$, Fig. 4 provides the test for a correlation of experimental data of $\dot{\varepsilon}$ as a function of σ_a with the model of Eq. (6). With data as a function of temperature, it is obvious that Fig. 4 could also be used to test reduced strain rate data $\dot{\varepsilon} \exp(Q/kT)$ as a function of the reduced variable σ_a/T. Evidently the correlation of the average function of Fig. 4 differs markedly from that of the segment function of

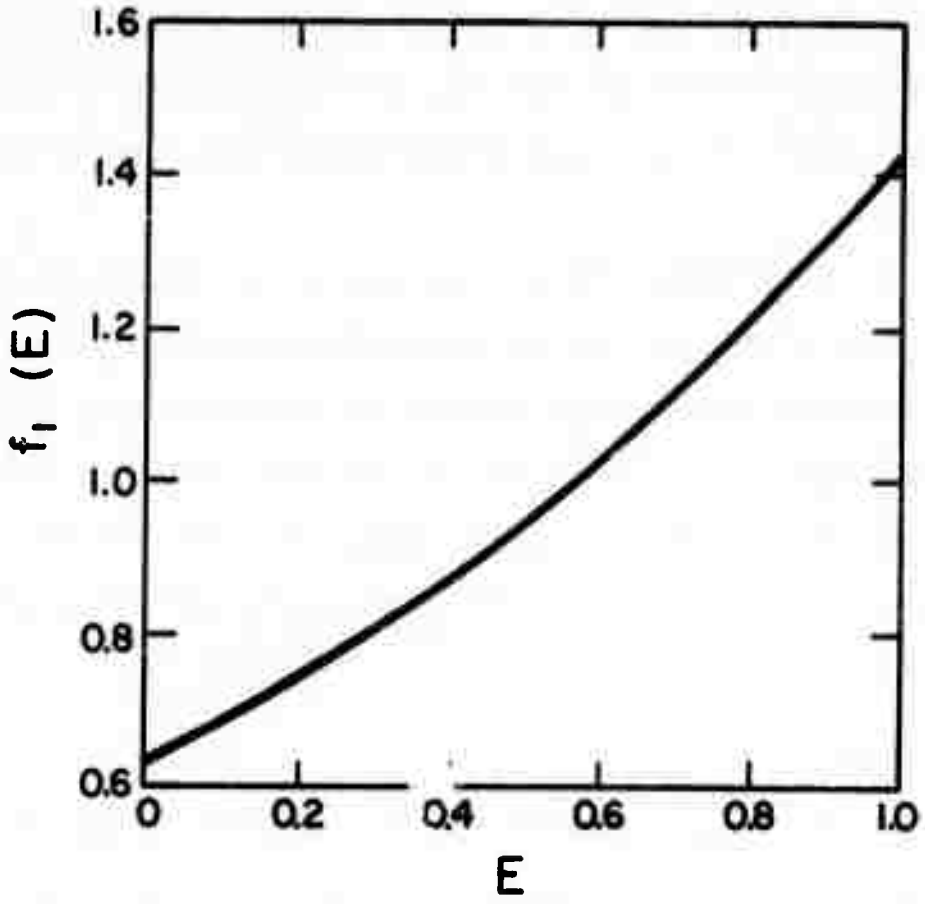


Figure 3

Plot of $f_1(E)$ vs. $E = \sigma_a bA \sin\theta \cos\theta/kT$.

Eq. (6) with σ_a.* Hence, when forms such as Eq. (6) are assumed for segment velocities, averages such as the above one, or other averages with appropriately modified values of $L_1(\theta,\kappa,\phi)$, $\nu_o(\theta,\kappa,\phi)$, $A(\theta,\kappa,\phi)$ and $\sigma_I(\theta,\kappa,\phi)$, when dislocation distributions are not random, should be used for self-consistency instead of the source function for v_i in testing data for correlation with theory.

The other forms for v_i, including the mixed activation-damping form of Eq. (2) can be handled analogously with a minimal cost of computer time. Five seconds of computer time were expended to develop Fig. 4. A useful integral solution which includes that of Eq. (14) but which also includes the requisite integrals for a variety of other forms of v_i is item 3.387, number 5, in ref. 25.

Acknowledgements

The author is grateful for the support of this work, initiated under the ARPA Materials Research Council Contract with the University of Michigan (DAHC15-71-C-0253) and continued under NSF Grant GK-18770. Useful comments on this research by J. J. Gilman, E. Montroll and R. Speiser and aid with the computer programs by B. Carnahan, H. Geary and J. Wilkes are also gratefully acknowledged.

*If one were to fit the actual $f_2(D)$ function of Fig. 4 with the exponential function of Eq. (6), the apparent activation area A would be a factor of 20 smaller than the true activation area of the $f_2(D)$ function.

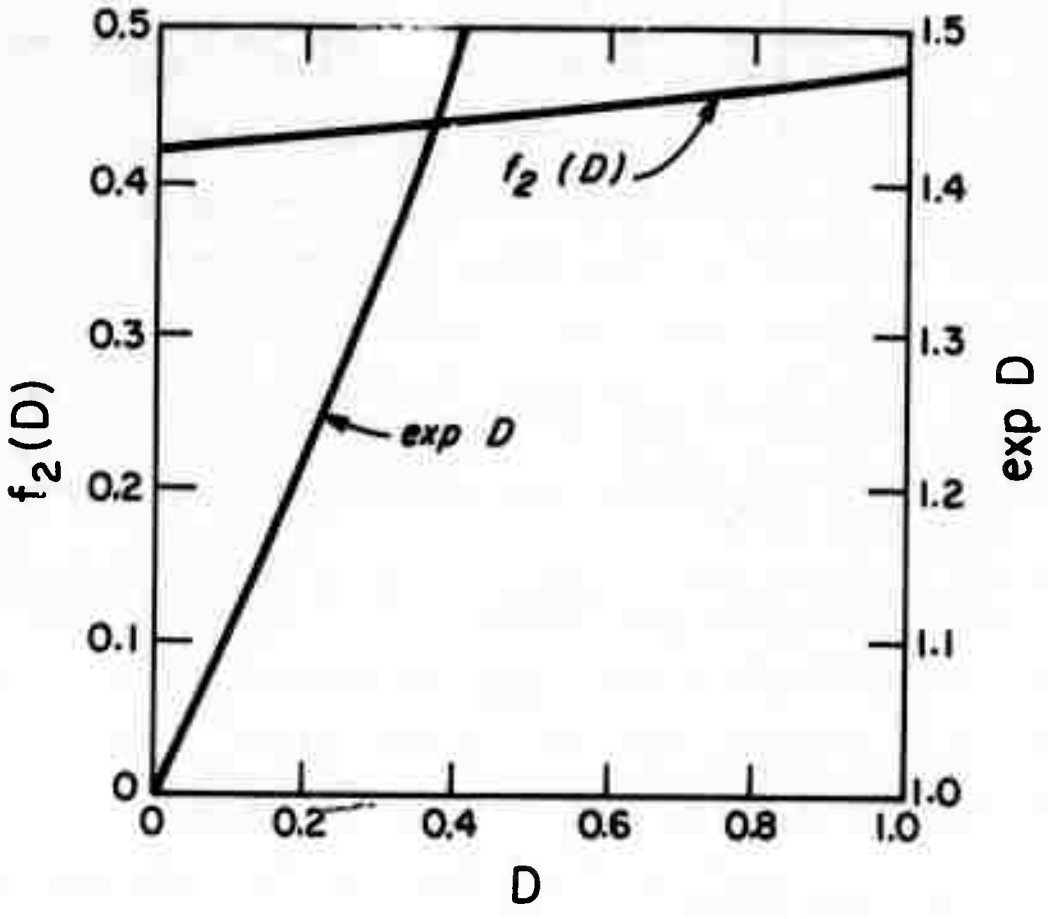


Figure 4

Plot of f (D) vs. D = $_a$bA/kT, solid line, and of the generating function of Eq. (6), exp D, vs. D, dashed line.

References

1. O. D. Sherby, J. L. Lytton and J. E. Dorn, Acta Met. 5, 219 (1957).

2. J. E. Dorn, in *Creep Fracture of Metals at High Temperatures*, H. M. Stationery Office, London, 1956, p. 89.

3. G. I. Taylor, J. Inst. Met. 62, 307 (1938).

4. J. F. W. Bishop and R. Hill, Phil. Mag. 42, 414, 1298 (1951).

5. R. W. Armstrong, Met. Trans. 1, 1169 (1970).

6. U. F. Kocks, Met. Trnas. 1, 1121 (1970).

7. J. C. M. Li, in *Dislocation Dynamics*, eds. A. R. Rosenfield et al., McGraw-Hill, New York, 1968, p. 87.

8. V. R. Parameswaran, N Urabe and J. Weertman, J. Appl. Phys. 43, 2982 (1972).

9. J. P. Hirth and J. Lothe, *Theory of Dislocation*, McGraw-Hill, New York, 1968, pp. 594-623.

10. G. Leibfried, Z. Phys. 127, 344 (1950).

11. Z. S. Basinski, Acta Met. 5, 684 (1957).

12. G. Schoeck, Phys. Stat. Sol. 8, 499 (1965).

13. J. W. Christian and B. C. Masters, Proc. Roy. Soc. (London) A281, 240 (1964).

14. C. R. Barrett and O. D. Sherby, Trans. AIME, 233, 1116 (1965).

15. J. J. Gilman, in *Dislocation Dynamics*, eds. A. R. Rosenfield et al., McGraw-Hill, New York, 1968, p. 3.

16. O. D. Sherby and P. M. Burke, Prog. Materials Sci., 13, 325 (1968).

17. J. Weertman, Trans. Quart. ASM 61, 681 (1968).

18. J. P. Hirth and W. D. Nix, Phys. Stat. Sol. 35, 177 (1969).

19. P. P. Gillis and J. J. Gilman, J. Appl. Phys. 36, 3370 (1965).

20. T. Mura, Phil. Mag. 8, 843 (1963).

21. T. Mura, Adv. Materials Res. 3, 1 (1968).

22. J. J. Hauser and B. Chalmers, Acta Met. 9, 802 (1961).

23. J. D. Livingston and B. Chalmers, Acta Met. 5, 322 (1957).

24. R. E. Hook and J. P. Hirth, Trans. Jap. Inst. Met., Suppl., 9, 778 (1968).

25. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals and Products*, Academic Press, New York, p. 322.

TABLE 1

The Quantity $f_1(E)$ as a Function of E

E	$f_1(E)$
0.00	0.63662
0.01	0.64164
0.02	0.64670
0.03	0.65181
0.04	0.65696
0.05	0.66216
0.06	0.66740
0.07	0.67268
0.08	0.67801
0.09	0.68338
0.10	0.68880
0.20	0.74563
0.30	0.80753
0.40	0.87496
0.50	0.94845
0.60	1.02857
0.70	1.11592
0.80	1.21120
0.90	1.31513
1.00	1.42854

TABLE 2

The Quantity $f_2(D)$ as a Function of D

D	$f_2(D)$
0.00	0.42441
0.01	0.42484
0.02	0.42527
0.03	0.42569
0.04	0.42612
0.05	0.42655
0.06	0.42699
0.07	0.42742
0.08	0.42786
0.09	0.42830
0.10	0.42874
0.20	0.43322
0.30	0.43787
0.40	0.44269
0.50	0.44770
0.60	0.45289
0.70	0.45828
0.80	0.46387
0.90	0.46967
1.00	0.47570

AC LOSSES IN SUPERCONDUCTING MAGNET SUSPENSIONS FOR HIGH-SPEED TRANSPORTATION

M. Tinkham

Abstract

A rather general relation is derived between vertical accelerations of the train and cryogenic power dissipation due to AC currents induced in the superconducting suspension magnets. Our theoretical results give a good account of the rather large losses observed in the Fuji Electric test vehicle, which used less than state-of-the-art conductors. Even with parameters estimated to represent the best commercially available materials, however, it appears that these losses will be comparable with the total heat leak due to all other causes if the accelerations are as large as permitted by subjective human ride quality considerations. Empirical data on losses under appropriate conditions of B, $\dot{B}$, and I/I_c may be needed to ascertain whether improved materials will be required to avoid serious design constraints by cryogenic heating.

AC LOSSES IN SUPERCONDUCTING MAGNET SUSPENSIONS FOR HIGH-SPEED TRANSPORTATION

M. Tinkham

I. Introduction

Although considerable analysis of the performance of high-speed trains using superconducting magnets for levitation has appeared[1], relatively little attention seems to have been given to the quantitative aspects of the problem of AC losses in the superconductors. These arise because of current changes induced in the magnets by the time-varying environment seen by the magnets as the train moves over any irregularities in the quideway (or, more severely, over a structured track made up of conducting loops). Even with an ideally smooth guideway, there will inevitably be a certain amount of "heaving" motion at the natural frequency of the suspension due to wind buffeting, etc. Although material parameters are subject to uncertainty, the analysis presented here shows that the power dissipation caused by the resulting AC currents will probably be comparable with the cryogenic heat load due to all other heat leaks unless either: (1) the vertical accelerations can be held below the level required for passenger comfort, or (2) technological improvements in performance of presently available twisted composite superconducting wires can be made. These estimates indicate that

detailed consideration of this problem should be made in evaluating any specific design for superconducting levitation to be sure that the AC losses in operation will be at an acceptable level.

II. Constant Flux Magnets

When a superconducting magnet is operated in persistent current mode, the total flux threading it

$$\Phi_1 = L_1 I_1 + M I_2 \tag{1}$$

remains constant, so long as it remains fully superconducting. In this expression, L_1 is the inductance of the magnet, M the mutual inductance to its environment (predominantly the guideway), I_1 is the magnet current, and I_2 is the current in the guideway. (In writing this, we have replaced the total environment rather schematically by a single lumped constant circuit, but this is adequate to bring out the essential features.) The current I_2 will be zero when the train is at rest over a normal metal guideway, but at full speed it will be essentially the same as if the guideway were superconducting. In the latter case, I_2 will be such as to cancel the flux in the guideway due to the train magnets, and we can write

$$\Phi_2 = L_2 I_2 + M I_1 = 0 \tag{2}$$

Eliminating I_2 between these two relations, we have

$$I_1 = \frac{\Phi_1/L_1}{1-k_{12}^2} \tag{3}$$

where

$$k_{12}^2 = M^2/L_1L_2 \tag{3a}$$

characterizes the strength of the coupling between the magnets and the guideway. Thus, the persistent current is not a constant current, but one which varies with the environment.

Variations of Magnet Current in Moving Train

From (3) it follows that when the train is brought to high speed, the current I_1 in the superconducting magnet will increase by a factor of $(1-k_{12}^2)^{-1}$ compared to its value when the train is at rest. This factor is of significant size, since k_{12}^2 must be reasonably large ($\sim$0.3) to give substantial lift. Moreover, when the train goes over "bumps" in the track, k_{12}^2 will change, inducing AC currents.

These relationships can be made more quantitative by noting that at high speeds the lift force on the magnet can be written as

$$F = I_1I_2|dM/dh| = L_1I_1^2k_{12}^2M^{-1}|dM/dh| \tag{4a}$$

$$= \frac{(\Phi_1^2/L_1)k_{12}^2M^{-1}|dM/dh|}{(1-k_{12}^2)^2} \tag{4b}$$

[If the force law (4a) is not familiar, the result may be derived by noting that the total magnetic energy is $E = 1/2\ L_1I_1^2 + MI_1I_2 + 1/2\ L_2I_2^2 = (\Phi_1^2/2L_1)(1-k_{12}^2)^{-1}$, using (2) and (3). Then $F = |dE/dh|$ yields (4b).] In these expressions, dM/dh is the rate of change of M with the height h of the magnet above

the guideway. This will depend on the detailed geometry. But since the flux density due to a long current-carrying wire falls as 1/r, typically $M(h) \simeq M(h_o)(h_o/h)^n$, with $n \simeq 1$; in which case, $M^{-1}\, dM/dh = nh^{-1} \simeq h^{-1}$. Note that increasing the coupling k_{12}^2 by decreasing h not only increases the lift per unit magnet current, as indicated by (4a); it also causes that current to increase because of the condition Φ_1 = constant, as reflected by the factor $(1-k_{12}^2)^{-2}$ in (4b).

Actually, since a given magnet will be characterized by a maximum permissible current, namely $I_1 = I_c$, (4a) is the appropriate formula for finding the maximum lift force. However, in "charging" the magnets, it must be borne in mind that I_1 will increase as given by (3) when the train is moving, and will increase further on "bumps" in which k_{12}^2 is momentarily increased. Thus, the initial value of I_1 must be set far enough below its maximum permissible value to allow for any imaginable increase of k_{12}. If that is done, there should be no "run-down" of the magnets from use.

Effect of Structured Track

With a periodic structured track, consisting of either a series of loops or a "ladder" configuration, the train magnets effectively see a periodically modulated mutual inductance coupling them to the track. For example, if the track loops were of the same length as the train loops (an extremely unfavorable case), then the track loop current would vary between zero, when a track

loop equally overlapped two successive oppositely polarized train magnets, and a maximum value when the two sets of loops were in register. In this case, M effectively varies between zero and some maximum value M_o comparable to that for a solid continuous track. As a result, the lift force would undergo 100% modulation and the magnet current I_1 would also incur a large degree of modulation ($\sim M_o^2/L_1L_2$), with associated intolerable levels of dissipation in the superconducting magnets.

This situation is greatly improved by taking the track loops to be shorter than the train loops by some fraction $N'/N = 1/n_R$, where N' and N are the number of train loops and track loops in the length of the train. In that case, the modulation of M and of the lift force are typically reduced by a factor $\sim n_R^{-p}$; the exponent $p \simeq 2$ will depend on such parameters as the length/width ratio of the train magnet coils, their separation, and their height, which determine the waveform of the flux density at the track. For example, if the waveform of $B_z(x)$ at the track were a square wave, the fractional modulation would be reduced by a factor $1/n_R$ (i.e., $p = 1$), since in the extreme cases n_R and (n_R-1) coils would be effective in giving lift. However, realistic field patterns are much smoother than a square wave (the higher Fourier components of $B_z(x)$ attenuate more rapidly), so the track loops at either end of a train loop contribute less than those in the middle. Hence the modulation depth will fall faster than n_R^{-1} (i.e., $p > 1$). In fact by suitable choice of winding configuration it should be possible to cancel the dominant

Fourier component for a given value of n_R, greatly reducing the size of the modulation.

The numerical calculations of Yamada and Iwamoto[2] for a specific track configuration display the qualitative features described here. The fractional modulation is found to be reduced by a factor of $\sim n_R^{-2} = 10^{-2}$ for $n_R = 10$; and $n_R = 6$ is found to give anomalously low modulation ($\sim 3 \times 10^{-3}$), presumably due to a near cancellation of the 6th Fourier component.

This reduction in the modulation depth of the coupling by large values of n_R not only reduces the ride roughness; it also reduces the modulation amplitude of I_1 in the superconducting magnet, in the same proportion. The associated AC losses are also reduced, but with one less power of n_R, because the hysteretic power dissipation is proportional to $\dot{I}_1 = \omega \delta I_1 \propto n_R \delta I_1$ for fixed train magnet length.

A further reduction in AC magnet currents can be achieved by introducing a normal conducting sheet between the track and the train magnets to screen out the time-varying fields at the magnets. If thin compared to the skin depth, such a screen reduces the AC currents by a factor of $(1+\omega^2\tau^2)^{-\frac{1}{2}}$, where τ is the L/R decay time of the currents in the screen. Since this shield plate also increases the damping of the vertical motion of the train, it will probably be advantageous to include it even for unstructured track systems. Maximum damping will occur when $\omega\tau \simeq 1$; on the other hand, good screening requires that $\omega\tau >> 1$. For example, a 1 cm sheet of aluminum will give maximum damping at a frequency

near the natural frequency of the suspension, $\omega_o/2\pi \simeq 1$ Hz, and screening by a factor of $\sim 10^{-2}$ at 100 Hz, a typical frequency associated with motion over a structured track.

III. AC Losses in Superconducting Magnets

As discussed above, when there is a change in the environment seen by a superconducting magnet in persistent current mode, the current in the magnet changes to hold the total flux constant. For thin extended loop magnets of the sort considered for levitation purposes, the local flux density B in the winding itself is dominated by the local current and changes in proportion to it, even though the total flux is constant.

To illustrate this point by a concrete example, consider a 300,000 ampere-turn superconducting loop with cross-sectional radius a = 1.5 cm at a height h = 30 cm above an ideal ground plane. The local field at the surface of the superconductor is 40,000 gauss; the image field reflecting the presence of the ground plane is 1,000 gauss. A height decrease of 1% will cause the image field to increase by ~1%, and the magnet current to increase by an amount smaller by a factor of order $\ell n\ h/a \simeq 3$. Evidently, the resulting ~0.3% increase in the 40,000 gauss field will dominate. In general, the local field effect will be stronger by a factor of the order of the ratio of the height to the conductor radius; this ratio will always be quite large, as in this example.

If the superconductor were in the form of infinitely fine

insulated wire, there would be no significant dissipation of energy due to changing currents. In practice, magnets are would of wires of sufficient diameter that each can carry substantial currents, and the conductors are stabilized against flux jumps by making them a composite in which the superconductor is in contact with a good normal conductor, such as copper. The best performance has been obtained by using multifilamentary wire, containing ∿100 superconducting filaments of diameter ∿10 microns imbedded in copper, and subsequently twisted to "decouple" the separate filaments.

Origin and Order of Magnitude of Losses

There are two main sources of energy dissipation in the conductor: hysteresis loss in the superconductor and eddy current loss in the copper. If there were no superconducting filaments present to modify current patterns, the eddy current loss per unit volume would be of the order of

$$P = (1/12)\dot{B}^2 d'^2/\rho \qquad \text{(emu)} \qquad (5)$$

where d' is the wire diameter and ρ is its resistivity[3]. (The exact coefficient will depend on the shape of the cross section of the wire.) For $\rho = 2 \times 10^{-8}$ ohm-cm = 20 emu (copper at low temperature), $P \simeq 4 \times 10^{-10} d'^2 \dot{B}^2$ watts/cm^3. Thus, for the representative values d' = 0.025 cm, $\dot{B}$ = 1000 gauss/sec, and V = 10^5 cm^3 as the volume of material for the magnets in one vehicle, this amounts to only ∿0.02 watt; this is negligible

compared to the static heat leak into the cryostat, which is estimated to be ∿50 watts per car. We conclude that with good design, this simple eddy current loss would pose no serious heating problem.

We now consider the hysteresis loss in an isolated thin superconducting filament of thickness d and critical current density J_c (typically 1.5×10^5 amperes/cm^2). So long as $2\pi J_c d << \Delta B$, where ΔB is the total change in B in a cycle, the filament can do little screening of the field, and one can take $\dot{B}$ to have the same value in the filament as outside. The electric field induced along the wire varies as $E = \dot{B}x$, where x is measured from the center of the filament. The energy dissipated per unit volume is $J_c E$, where J_c is the critical current density, taken constant as in the Bean[4] model. Averaged over the thickness d, this gives the conventional result

$$P = J_c|\dot{B}|d/4 \qquad \text{(emu)} \tag{6}$$

per unit volume of superconductor. Unlike the eddy current loss, this hysteresis loss depends on the <u>first</u> power of $\dot{B}$. For making numerical estimates, it is more convenient to rewrite (6) as

$$P = I_c|\dot{B}|d/4 \tag{6a}$$

per unit <u>length</u>. With the typical values $I_c = 4 \times 10^5$ amperes (for the whole cable of wires, each containing many filaments), $d = 10^{-3}$ cm, and a length of 10^4 cm per car, we obtain $P = 10^{-2}\dot{B}$

watts. For $\dot{B} = 10^3$ gauss/sec, this leads to an estimated dissipation of $P \simeq 10$ watts. Since this is of the same order as the anticipated total cryogenic heat loading, it is clear that one will be required to hold this loss to a minimum.

Unfortunately, the estimate based on (6) forms a lower bound to the dissipation for any given composite conductor since it neglects the copper matrix. Currents will be induced which flow through the copper between filaments, and the presence of the superconducting paths greatly enhances the dissipation due to these eddy currents because the induced emf is dropped over a shorter distance in the copper. As discussed in detail by Wilson, et al[5], this extra loss can be reduced by twisting the wire with a pitch $p = 4\ell$ such that ℓ is much less than a critical length ℓ_c determined by

$$\ell_c^2 = 2J_c d\rho/|\dot{B}| \tag{7}$$

For typical values, ℓ_c is of the order of a centimeter. Without the twist, the dissipation turns out to be essentially the same as given by (6), but with d increased to the diameter d' of the whole bundle of filaments, and the advantage of the fine filaments is lost. It should be noted that (7) was derived by considering a simple model of two superconducting sheets of thickness d separated by copper. For an actual composite wire containing many layers of superconducting filaments, the twist pitch at which the loss reaches this upper limit is given approximately by

$$\ell_c'^2 \simeq J_c d'\rho/2|\dot{B}| = (d'/4d)\ell_c^2 \qquad (7a)$$

This follows from (7) upon replacing d by d'/4, corresponding to assuming a filling factor of 1/2 in a thickness (d'/2) on either side of center. With the twist, the extra loss falls as ℓ^2 for given $\dot{B}$, so that (6) is replaced by

$$P = J_c|\dot{B}|d_{eff}/4 \qquad (8)$$

where

$$d_{eff} \simeq d' \qquad (\ell>\ell_c') \qquad (9)$$

and

$$d_{eff} \simeq d(1+4\ell^2/\ell_c^2) = d+2\ell^2\dot{B}/J_c\rho \qquad (\ell<\ell_c') \qquad (10)$$

Note that increasing $\dot{B}$ not only increases the loss (6) for an isolated filament; it also decreases the critical length ℓ_c', and therefore increases the extra loss. Thus, although (6) varies as $\dot{B}$, the <u>extra</u> loss varies as $\dot{B}^2$, so that typically the observed loss varies more rapidly than linearly with $\dot{B}$. From (10) it is evident that it does little good to reduce d below $\ell^2\dot{B}/J_c\rho$; ℓ^2 must be reduced in proportion to d to keep $\ell<\ell_c'$, for given values of $\dot{B}$, J_c, and ρ.

From this discussion it is clear that even with twisting, it is hard to hold d_{eff} right down to d, especially for small values of d. If performance specifications are critical, one must rely on empirical test data to determine d_{eff}, but the

formula (10) should be adequate for semi-quantitative estimates. For more detailed results, the report[5] of the Rutherford Laboratory group is particularly useful.

Dependence on Design Parameters

Having seen that these AC losses will be a serious design concern, let us set up a reasonably general model for estimating their size and dependence on design parameters. It is convenient to relate the AC losses to $\ddot{z}$, the associated vertical acceleration of the train. Taking the derivative of (4b) with respect to k_{12}^2 (and neglecting the smaller associated change of $M^{-1}dM/dh$), we have

$$\frac{\ddot{z}}{g} = \frac{\delta F}{F} = \frac{1+k_{12}^2}{1-k_{12}^2} \frac{\delta k_{12}^2}{k_{12}^2} \tag{11a}$$

Similarly, from (3) we have

$$\frac{\delta I_1}{I_1} = \frac{\delta k_{12}^2}{1-k_{12}^2} \tag{11b}$$

if no screening plate is in use. Combining these relations,

$$\frac{\delta I_1}{I_1} = \frac{k_{12}^2}{1+k_{12}^2} \frac{\ddot{z}}{g} \tag{11c}$$

For a cylindrical conductor of radius a, carrying current I_1, the spatial average value of B in the conductor is $\bar{B} = 4I_1/3a$. Thus, if δI_1 is varying at ω,

$$\dot{\bar{B}} = 4\omega\delta I_1/3a \tag{12}$$

and from (6a) and (8) we have for the power dissipation per unit

length

$$P = \frac{\omega I_c \delta I_1 d_{eff}}{3a} = \frac{\omega I_c I_1 d_{eff}}{3a} \frac{k_{12}^2}{1+k_{12}^2} \frac{\ddot{z}}{g} \tag{13}$$

Since the total weight of the train mg must be sustained by the force (4a),

$$mg = L_1 I_1^2 k_{12}^2 M^{-1} |dM/dh| \tag{14}$$

Combining this relation with (13), and including a factor $(1+\omega^2\tau^2)^{-\frac{1}{2}}$ for the effect of a screening plate, the <u>total</u> power dissipation in supporting the train weight can be written as

$$P = \frac{\omega mg}{(1+\omega^2\tau^2)^{\frac{1}{2}}} \left(\frac{\ddot{z}}{g}\right) \left(\frac{I_c}{I_1}\right) \left(\frac{d_{eff}}{3a}\right) \frac{M}{|dM/dh|} \frac{1}{1+k_{12}^2} \frac{1}{L_1'} \tag{15}$$

In this, $L_1' \simeq 2\ell n(w/a) \simeq 8$ is the dimensionless inductance (in emu) per unit length of the magnet conductor, w being the separation of the conductors. Also, typically $|dM/dh| \simeq M/h$ and $k_{12}^2 \simeq 1/3$. With these substitutions, (15) simplifies to

$$P = \frac{1}{32} \frac{d_{eff}}{a} \frac{I_c}{I_1} \frac{\omega hmg}{(1+\omega^2\tau^2)^{\frac{1}{2}}} \frac{\ddot{z}}{g} \tag{16}$$

This result should be quite general, and, apart from small changes in the numerical coefficient, give a good approximation in most practical configurations. If a screening plate is used, the frequency dependence in (16) drops out for $\omega > 1/\tau$. If we further assume the ;onable value $I_1/I_c = 1/2$, then (16) simplifies to

$$P = \frac{1}{16} \frac{d_{eff}}{a} \frac{hm\ddot{z}}{\tau} \qquad (\omega\tau >> 1) \tag{17}$$

which brings out the critical factors in a particularly concise way. Finally, if we reinstate the frequency dependence of (16) and insert the reasonable values h = 30 cm, m = 5×10^4kg, $d_{eff} = 10^{-3}$ cm, and a = 1.5 cm, we have

$$P \simeq \frac{6\omega(\ddot{z}/g)}{(1+\omega^2\tau^2)^{\frac{1}{2}}} \quad \text{(watts)} \qquad (18)$$

Applying this result to the heaving motion at $\omega_o/2\pi \simeq 1$ Hz, where $\omega_o\tau \lesssim 1$, we have $P \simeq 25\ (\ddot{z}/g)$ watts. Now, the maximum acceleration at 1 Hz which gives satisfactory subjective ride quality[6] is about 0.15g, for which $P \simeq 4$ watts. Thus, if the above estimate is accurate, one can conclude that if the ride is smooth enough for passenger comfort, it will also be satisfactory from the point of view of cryogenic heating.

In evaluating the estimate above, it is important to note that it represents a sort of lower limit to the heating. The true value might well be an order of magnitude greater, enough to double the estimated static heat leak from all other sources. For one thing, we have considered only the lift force. The dissipation associated with guidance and propulsion forces might be comparable in size. As discussed above, however, the major source of underestimation of loss is our assumption of perfect decoupling of 10μ filaments, so that $d_{eff} = d = 10^{-3}$ cm. For the assumptions of our example computation, it turns out that $\dot{B} = 6 \times 10^3$ gauss/sec, so that $\ell_c \sim 3$ mm. Thus, a twist pitch of a few millimeters would be required (according to (10)) to approach

within a factor of two of the fully decoupled loss rate. Moreover, simple twisting is not enough; full transposition is required for ideal performance. That is, all conductors must share equally in exterior and interior positions in the wire, and eventually in the cable of wires. To achieve decoupling with reasonable twist pitches at these rapid sweep rates and small values of d, it may be necessary to resort to a 3-material composite. In such a composite, each superconducting filament is surrounded not only by copper, but also by a thin layer of relatively high-resistance alloy material to insulate it partially from its neighbor filaments. In this way, one can combine the thermal stabilization of the pure copper with the low AC loss of the alloy, at the expense of an increase in the complexity of the composite. Such materials have already been tested by the Rutherford Laboratory group[5].

Despite our reservations about the exact numerical results presented here, this calculation should be made quite accurate by our simple expedient of using a suitable $d_{eff} \geq d$ instead of d in (13), (15), and (16). This d_{eff} can be estimated using (10), or better, determined empirically by using (8) and the measured loss in a test sample operated at the same values of B, $\dot{B}$, and I/I_c as would be found in the actual application.

Let us now consider the case of a structured track, with track loops of length $\sim$1 meter. At full speed, this will correspond to a frequency of $\sim$100 Hz, and (18) leads to $P \simeq 4000\ (\ddot{z}/g)$

watts, if one ignores the effect of the screen. Subjective ride quality[6] would again allow $\ddot{z} \simeq 0.15g$ at this rather high frequency, in which case $P \simeq 600$ watts, clearly unacceptably high. None-the-less, there is no serious problem, because, as noted above, one can hold $\ddot{z} \lesssim 0.01g$ with a reasonable choice of a track loop length small compared to the train loop length. In that case, P would be only $\sim$40 watts, which is comparable with other expected heat inputs. Moreover, this high-frequency field can be reduced by a factor $(1+\omega^2\tau^2)^{-\frac{1}{2}} \sim 1/100$ by inserting a thick normal conducting sheet between magnets and track. For example, for typical coil sizes a 1 cm aluminum sheet at room temperature would give satisfactory shielding ($\omega\tau \simeq 100$), while increasing the weight of the train by less than 5%, even if it covered the entire area of the train. Given such a screen, the additional AC loss due to structured track should present no serious difficulty.

IV. Comparison with Experiment

It is of considerable interest to compare the results of the above theoretical analysis with the experimental results of Hirai, et al[7] on the Fuji Electric test vehicle, the only relevant data available to the author at this time. This vehicle, weighing 650 kg, was levitated at a height of about 25 cm over a structured track moving at a speed of 100 km/hr. The magnet conductor weighed 29.5 kg, carried 855 amperes giving 2×10^5 ampere-turns, and produced a maximum field of 23,000 gauss; the energy stored was 4.5×10^4J. The cross section of the conductor was 1.8×3.2mm,

and it contained 161 filaments of NbTi, each 80μ in diameter, embedded in 6x the weight of OFHC copper, and twisted with a pitch of 10 cm. Eighteen layers of such conductors were used, each containing 13 turns. As sketched in Fig. 1, the two magnet coils were crescent shaped to conform to the circular test track, with dimensions of ∿110 cm along the circumference and ∿25 cm between the two sides of the loop. The structured track contained 6 loops, so that the ratio of loop sizes was $n_R = 3$, and it could be rotated at up to 600 rpm. There was a 3.2 mm thick aluminum screening plate between magnets and track, cooled by conduction from liquid N_2.

We may compute k_{12}^2 from (4a) using the quoted weight and stored energy $1/2\ L_1 I_1^2$, and by assuming $M^{-1}|dM/dh| = 1/h$. The result is $k_{12}^2 = 0.018$. We can see from (3) that the current should increase by a similar amount when the magnet is levitated over the moving track. The observed increase is quoted as 1-2%, certainly in excellent agreement. This coupling value is an order of magnitude less than would be desirable for efficient levitation of a real train. The low value is a consequence of the small width of the coils compared to their height above the track.

At full speed of 10 rps, the 6 track loops modulate the environment of the magnets at a frequency of 60 Hz. Search coils installed below the magnets revealed an AC magnetic field of up to 50 gauss at this frequency. This value can be accounted for roughly by noting that the maximum field at the coil is B ≃ 20,000

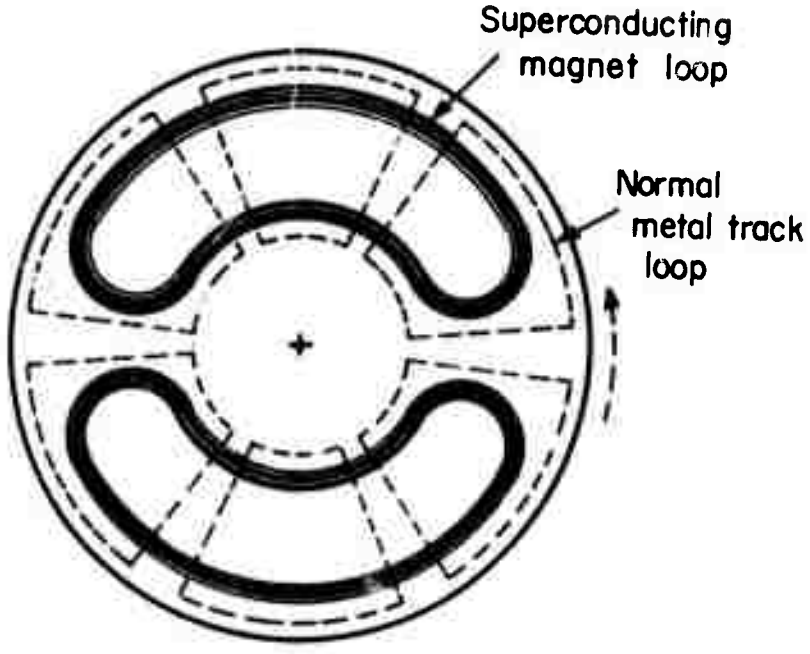


Figure 1. Schematic diagram of magnet and track configuration in the test of the Fuji Electric vehicle. (See ref. 7).

gauss, modulated by a fraction $\sim n_R^{-2} k_{12}^2 \simeq (1/9)(0.018) = 0.002$, so that $\delta B \simeq 40$ gauss is a reasonable estimate. This AC field is attenuated by a factor $(1+\omega^2\tau^2)^{-\frac{1}{2}} \simeq 1/60$, for $\omega = 120\pi$ and $\tau = L/R \simeq 1/6$ second. (The inductance of a single turn of the size and shape of the magnet coil is about 1μH, and the resistance of a similar size strip of a sheet of 3mm aluminum at 77°K, where $\rho \simeq 0.2\mu$ ohm-cm, is about 6μ ohm. Since the ratio L/R is determined by the narrow dimension of the loop, it would be about the same for a loop of the size of the track loop as for the magnet loop.) To emphasize that this screening effect is an impedance effect and not a skin depth effect, we note that under these conditions the skin depth δ would be about 3mm, the thickness of the screening plate. Yet the attenuation is by a factor of 60, not by e. Thus, the AC field applied to the magnet is only ∼1 gauss, so that $\dot{B} = \omega\delta B \simeq 300$ gauss/sec.

For this value of $\dot{B}$ and the parameters of the conductor listed above, the characteristic length ℓ_c defined by (7) is about 5 cm. Thus, the 10 cm twist pitch is too long to be fully effective in reducing losses, and the effective filament diameter d_{eff} will be at least 0.2 mm. With the volume of the superconductor being only ∼1/6 of the total conductor, we estimate $V \simeq 700$ cm^3. With $J_c \simeq 200{,}000$ amps/cm^2, (8) then leads to a total dissipated power $P \simeq 2$ watts. An alternate estimate can be made using (15), taking $\ddot{z}/g = 1/n_R^2 = 1/9$, $I_c/I_1 = 2200/855 = 2.6$, $L_1' \simeq 2\ \ell n\ 10 = 4.6$, and other numerical values appropriate to the system; this

leads to P ≃ 2 watts, within the accuracy of the estimates used. For comparison, the value determined by measuring the extra helium boil-off was 18-27 watts, an order of magnitude larger than these estimates. Since the loss due to imperfect decoupling varies as $\dot{B}^2$, this discrepancy could be accounted for if, because of its finite extent, the screening plate reduced $\dot{B}$ by only a factor of 20 instead of 60, as we estimated. Or, the extra loss may stem from an underestimate of d_{eff} due to the absence of complete transposition in the winding. In any case, we are reminded that our estimate (10) is generally a lower bound, and that empirical tests are needed to see exactly how large these losses are. It is interesting to note that the computed eddy current loss in the copper (5) <u>if there were no superconducting filaments</u> is only ∿0.01 watts.

A more relevant comparison is with the other cryogenic heat leaks. These were only ∿10 watts, of which Hirai, et al attribute a major portion to conduction down two short current leads (which could be eliminated in a practical design for a train). They estimate the other heat inputs as only ∿3 watts. Scaled up in proportion to weight from the test vehicle of 650 kg to a practical vehicle of 50,000 kg, this 3 watts becomes 230 watts, but presumably improved engineering design in the full scale version could bring this down to our estimated 50 watts. On the other hand, the AC loss of ∿20 watts, although it caused no problem in the test vehicle, would scale up to 1500 watts, if no improvements

were made; this would be an excessive burden on the cryogenic system of a high-speed train.

One can get an appreciation of the practical significance of such dissipation levels by recalling that 1 watt-hour of energy will boil off about 1.3 liters of liquid helium to form about 0.9 cubic meters of gas at STP. Thus, in an 8-hour shift, a 50 watt heat leak will boil off some 500 liters of liquid to form 3.6 m^3 of gas, after compression to 100 atmospheres. On the other hand, a 1500 watt heat leak would boil off 15 cubic meters of liquid helium to form 110 m^3 of gas at 100 atm. On-board storage of the larger amount of helium, either as liquid or as gas, would be impractical. If instead one used a refrigeration system on the train, the power requirement to remove 1 watt at 4°K would be about 75 watts for full Carnot efficiency, and probably ∿300 watts for even a rather efficient real machine. Thus, while a 50 watt heat leak scales to only ∿15 kw, 1500 watts scales up to ∿450 kw. The latter value would obviously involve unreasonably massive refrigeration machinery as well as obvious problems in transferring energy to the rapidly moving train. [The much larger propulsive power (∿6MW) need not be transferred if an active-track linear synchronous motor drive is used.] We conclude that neither cryogenic system could comfortably tolerate a heat leak much greater than our nominal 50 watts without adding excessive weight burden to the train.

As can be seen from (15), the excessive dissipation rate of this model vehicle could be reduced somewhat by going to a

larger value of n_R (to reduce $\ddot{z}/g \simeq 1/n_R^2$), by working with I_1 closer to I_c, and by using a thicker or larger screening plate to attenuate the AC field. However, the major means of improving the performance will be by use of a conductor with finer filaments and better decoupling, the latter to be achieved by decreasing ℓ and/or by increasing ρ. In this way, one could hope to reduce d_{eff} from $\sim$0.2 mm to $\sim$0.02 mm, which would reduce the dissipation by a factor of 10, to a level smaller than other heat leaks.

V. Conclusion

Although the test data just discussed were dominated by the effects of the high-frequency currents due to a structured track, these can be controlled by increasing n_R or using a thicker or more inclusive screening plate. We conclude that the most serious AC loss problem in a real train will be that associated with the low-frequency heaving motion of the train at frequencies near the natural frequency of the suspension ω_o, for which a screening plate is relatively ineffective (except for damping the motion). For this motion, the requirement of acceptable cryogenic heating appears to place a limit on vertical acceleration which is comparable in severity with that imposed by subjective ride comfort. Because of the uncertainties in our numerical estimates of the conditions to be encountered in a train magnet, it would be necessary to obtain empirical data at appropriate field values and sweep rates in order to make a more precise statement. It is possible that economical commercial 3-component composite con-

ductors will need to be developed to keep the AC losses from imposing the most demanding specification on acceptable vertical accelerations, with all the difficulty and expense that implies for track perfection and electronic feedback systems to maintain the ride quality.

VI. Acknowledgements

This study was initiated under the auspices of the ARPA Materials Research Council Contract with the University of Michigan (DAHC15-71-C-0253) during the July 1972 meeting of the Council at Centerville, Massachusetts. The author is happy to acknowledge many stimulating and helpful discussions with Professor P. L. Richards during the above meeting on the subject of this paper and many other related topics. Research was supported in part by the Office of Naval Research, the Advanced Research Projects Agency, and the National Science Foundation. Thanks are also due to Profes[illegible] M. R. Beasley for a critical reading of this manuscript. Finally, the author thanks Dr. L. C. Davis for drawing his attention to an error in a preprint version of this paper.

References

1. See, for example, P. L. Richards and M. Tinkham, J. Appl. Phys. 43, 2680 (1972), and references cited therein.

2. T. Yamada and M. Iwamoto, Trans. IEEJ 92-B, 19 (1971).

3. Throughout this paper, general formulae are quoted in emu; numerical results are expressed in practical units.

4. C. P. Bean, Phys. Rev. Letters 8, 250 (1962).

5. M. N. Wilson, C. R. Walters, J. D. Lewin, and P. F. Smith, J. Phys. D3, 1518 (1970); A. H. Spurway, J. D. Lewin, and P. F. Smith, J. Phys. D3, 1572 (1970).

6. "Technical Extracts from TACV System Study and Analysis Report", TRW Report No. 06818-6008-R000 for Dept. of Transportation, Oct. 1967, unpublished. The acceptable level of acceleration has a broad minimum near 10 Hz, where it is about 0.03 g.

7. K. Hirai, S. Akiyama, H. Fujino, and K. Onodera, "A Lightweight Superconducting Magnet for a Test Facility of Magnetic Suspension for Vehicles", Proceedings of 4th International Conference on Magnet Technology, Brookhaven National Laboratory, 1972, ed. Y. Winterbottom, to appear.

COMMENTS ON THE PROSPECTS FOR MAJOR IMPROVEMENTS IN RECHARGEABLE BATTERIES TO OPERATE AT AMBIENT TEMPERATURES

R. A. Huggins

Abstract

The important parameters relating to high performance batteries are described. The factors controlling the specific energy are pointed out, and it is shown that, in view of recent progress on solid electrolytes, there are no fundamental limitations to the development of high specific energy cells to operate at ambient temperatures. The achievement of high values of specific power at such temperatures will be a much more difficult problem, and the important factors are described. The primary limitations will be in the cathode system, and several different approaches to circumventing them are described.

COMMENTS ON THE PROSPECTS FOR MAJOR IMPROVEMENTS IN RECHARGEABLE BATTERIES TO OPERATE AT AMBIENT TEMPERATURES

R. A. Huggins

Introduction

Because of the importance of power sources for a wide range of engineering applications, attention has been given to the development of electrochemical devices for the conversion and storage of energy for many years and several different types of such devices have been developed and serve in a variety of applications.

Increased attention has been given to this matter in the last several years for two major reasons. The first of these has to do with the desire to find alternatives to the internal combustion engine which is presently used for vehicular propulsion. One obvious alternative would be to develop electrically powered vehicles, but it is widely recognized that this would be impractical at the present time because of the limitations in the performance of currently available electric power sources.

The second major reason for the sudden increase in interest in this area is that it appears that a major breakthrough has occurred which could have a large impact upon

battery (and perhaps fuel cell) technology. As a result of recent work we are now aware of a number of solids in which ionic transport can be about as fast as that normally found in liquids. Some of these solids are electronic insulators, and can thus be used as solid electrolytes. As will be mentioned later, there are several obvious potential advantages in the use of solid electrolytes rather than traditional liquid electrolytes. In addition it has also been recently shown that there may be important advantages in the use of certain types of mixed conductors (in which both ionic and electronic species are mobile) in connection with solid electrolyte systems.

As a result of the emergence of solid electrolytes, it is now possible to reconsider a number of aspects of battery and fuel cell design so that radical changes can also be expected here as well. Not the least of these new possibilities is the opportunity for the development of miniature all-solid battery systems which might be integrated into electronic circuitry.

It is not the intent here to discuss these recent developments in detail, as a number of reviews of progress in this area can be found elsewhere[1]. Instead, since most of the developmental efforts oriented toward non-miniature battery systems require operation at elevated temperatures, attention will be focused here upon possibilities for the development of high performance battery systems to operate at ambient temperatures.

Fundamental Parameters

In addition to the obvious important practical matters such as cost, mechanical ruggedness, and safety, there are four fundamental parameters of primary concern in any battery system. The first of these is the specific energy, which specifies the amount of energy that can be stored per unit weight of the total battery system. The second is the specific power or the amount of power available per unit weight. The third is the storage life, which is determined by the rate of self discharge of the battery system with no external load. The fourth is the operating life, generally expressed in terms of the number of charge-discharge cycles that can be sustained.

Approximate values of the specific energy and specific power of various battery systems, as well as the general range of performance obtained from fuel cells, are presented in Fig. 1. In general, the better fuel cell systems which involve the oxidation of gaseous or liquid fuels tend to have good values of specific energy, so that a great deal of energy can be stored per unit weight. However, they are typically limited by kinetic considerations so that their values of specific power are relatively low. Battery systems, on the other hand, are generally less effective in terms of energy storage but are typically more attractive than fuel systems in applications involving large current drain because of their greater values of specific power.

To put the data shown in this figure in perspective, it has been estimated[2] that rechargeable batteries having specific

energy and specific power values of about 100 watt-hrs. per pound and 100 watts per pound will be needed for proposed designs of an electrically powered automobile weighing about 3000 pounds with a range of 200 miles at a constant speed of 55 mph (or about 165 miles at a speed of 70 mph). This includes the power required for air conditioning, heating, and typical accessories, as well as propulsion. Batteries which meet these criteria also would readily meet the federal suburban driving cycle requirement. Estimates have also been made[3] of the energy and power requirements for a smaller urban vehicle. In that case, both the specific energy and specific power of the battery system could be substantially reduced. The latter to a value of about 45 watts per pound.

Specific Energy

The amount of energy that can be stored in any electrochemical system has a definite theoretical limit. This value is determined by the free energy change in the net cell reaction, and if it is divided by the equivalent weight of the reactants alone, one can obtain the maximum theoretical specific energy for any given battery system.

One can, of course, never reach this limit in practical systems because of irreversible processes as well as the fact that additional weight is always present in the electrolyte and the container, as well as in electrical leads, connectors, etc. Nevertheless, the value of the maximum theoretical specific energy provides guidance with respect to which types of electro-

chemical systems are most worth exploring. Figure 2 presents data for a number of systems plotted versus the gram equivalent weight of the reactants themselves. Noting that values of the specific energy are plotted on a logarithmic scale, one readily sees that they vary over a wide range. This figure also illustrates the great importance of the atomic weight of the constituents. Most important, however, is the fact that the traditional battery couples are way down at the bottom of the figure, so that it is theoretically possible to find electrochemical systems in which the specific energy might be much greater than that which is possible in traditional battery systems.

Since the free energy change ΔG can be expressed as

$$\Delta G = \Delta H - T\Delta S \quad , \tag{1}$$

we see that the value of the theoretical maximum specific energy is actually greater the lower the temperature. Therefore, it does not itself constitute an impediment to the development of high performance batteries to operate at ambient temperatures.

It is also necessary that an electrolyte be available that not only has very high ionic conductivity and very low electronic conductivity but also has sufficient chemical stability that it does not deteriorate under conditions ranging from very reducing on the anode side to very oxidizing on the cathode side. This also does not provide a fundamental limitation upon

the possibility of reducing the operating temperature. Actually, it is a favorable factor, as most reactions involved in corrosion or chemical attack proceed at lower rates the lower the temperature.

To achieve high values of specific energy in actual batteries it is also necessary to have a very weight-efficient design utilizing practical materials for the container and current leads which will not be attacked by the chemical components of the system and which can be readily fabricated and of low cost. This is obviously not a factor which would limit the possibility of reducing the operating temperature of such a battery.

Thus, we see that there are no inherent restrictions relating to the specific energy that become especially important as the operating temperature is lowered.

Specific Power

Although there are no basic reasons why low temperature battery systems might be disadvantageous from the standpoint of specific energy, the situation is quite different with regard to specific power, for here one is concerned primarily with kinetic phenomena which typically have a strong temperature dependence. In order to look at the various factors which may be involved, let us consider the simple cell illustrated schematically in Fig. 3, in which it is assumed that the relevant overall chemical reaction is

$$M + X = MX \qquad (2)$$

If we take an operational view of this simple hypothetical cell, we see that the anode system on the left has three functions. M atoms must be supplied to the anode-electrolyte interface, these atoms must give off electrons to become M^+ ions, and the electrons must be transported away from the interface into the external circuit.

The function of the electrolyte is simply to transport these M^+ ions from the anode system side to its interface with the cathode system. It is, of course, important that no other species be transported through the electrolyte to any appreciable extent.

One of the functions of the cathode system is to supply electrons to combine with the M^+ ions which arrive at the cathode-electrolyte interface by transport through the electrolyte. X atoms must also be supplied to react with the M to form the reaction product MX. It is also important that this reaction product be transported away from the interface so that it does not impede further reaction.

Therefore, we see that an important aspect of each of the sections, or compartments, in this simple cell involves the transport of either atomic or ionic species. Although, in many aqueous electrochemical systems one finds that the ionization or reduction steps at the interfaces themselves may be rate controlling, this is typically not the case in high temperature

systems involving solid electrolytes, and is also probably not important in most such systems at ambient temperatures. Instead, the primary concern with regard to kinetic phenomena must be focused upon mass transport of species to and from the interfaces in the electrode systems, as well as through the electrolyte.

Materials with Very Rapid Mass Transport

In recent years there has been a great deal of progress in the identification of materials in which such mass transport is much more rapid than was earlier assumed to be possible in solids. A number of these (sometimes called superionic conductors) have diffusion rates which are immensely greater than is the case in the conventional ionic materials with which we are much more familiar. Some of these materials with unusually rapid mass transport are primarily ionic conductors (superionic conductors), whereas others are mixed conductors in which the electronic conductivity is also appreciable. Data on the ionic conductivity of some superionic conductors as well as several of the more conventional ionic solids are shown in Fig. 4, which also illustrates how these values vary with temperature. The most important point to be seen from this figure is the truly immense difference in behavior at low temperatures. To illustrate this point, extrapolation of the data for pure KCl, a common alkali halide, to 25°C would give a value of the ionic conductivity that is 10^{28} times lower than that for one of the best presently known superionic conductors, $RbAg_4I_5$. Another important thing that should be noted from these data is the relatively small

temperature dependence of the ionic conductivity in the superionic conductor group. This means that the properties of the solid electrolyte will not vary a great deal if the operating temperature of the battery is reduced.

Some of the superionic materials undergo phase transformations at relatively low temperatures which cause a sudden change in the value of the ionic conductivity. Typically, the structure above the transition temperature is less close packed and has a much higher conductivity than the low temperature phase. Thus, one must beware of this situation in certain cases. Data relating to these structural changes for some important ionic conductors are included in Table 1.

This problem does not occur in some superionic conductors, however. A prominant example is the beta alumina family. Fig. 5 shows the temperature dependence of the ionic conductivity and chemical diffusion coefficient of the solid electrolyte sodium beta alumina (nominally $NaAl_{11}O_{17}$). These data were obtained by the use of three different experimental techniques, and the linearity of this plot indicates that a single physical mechanism is involved over an extremely wide range of temperature, from 800°C all the way down to -190°C. Obviously, there is no problem of structural instability in this case.

The beta alumina structure is particularly important, for it has been shown[6] that a large number of monovalent cations can be introduced into it. Several of these ions have rather high values of ionic mobility. Careful measurements have been

made of the ionic conductivity of five of these[7] using appropriate mixed conductors as reversible solid electrodes. These data are shown in Fig. 6.

It is now recognized that the unusually high ionic mobility in these superionic conductors is related to the presence of ionic-sized tunnels within their crystal structures. In the case of beta alumina, the structure can be visualized as consisting of blocks of close-packed gamma alumina separated by bridging layers which contain tunnels which are about half filled with the mobile monovalent cations. This structure is shown schematically in Fig. 7. As a result of this layered structure, the ionic conductivity is extremely anisotropic, with rapid motion in the plane of the bridging layers, but essentially no transport in the other direction.

Limitations at Lower Temperatures

If we now return to the question of the features which might limit the possibility of the operation of high performance batteries at ambient temperatures, it is readily apparent that if one uses a solid electrolyte that does not go through a phase transformation, the only disadvantage with operation at lower temperatures, so far as the electrolyte part of the cell is concerned, has to do only with the temperature dependence of the ionic conductivity. Since this is a reasonably small factor with the better solid electrolytes, it is quite apparent that it does not constitute a fundamental limitation.

This does not mean, however, that the electrolyte question has yet been adequately resolved, for we are not yet aware of really satisfactory solid electrolytes for the transport of either lithium or potassium, two of the more interesting possible anode constituents. Nevertheless, if one can assume a reasonable degree of optimism, it becomes apparent that internal resistance due to the electrolyte is not the fundamental stumbling block to ambient temperature high performance batteries.

There are several other possible features that may provide a more serious limitation upon the specific power that can be obtained from an ambient temperature high specific energy battery system. The rate of charge flow through the battery may be controlled by slow interfacial reactions at the electrolyte-electrode contact. However, as was mentioned before, this is an important problem in many aqueous battery systems but is not expected to be limiting in non-aqueous systems of the type most apt to be used with solid electrolytes.

Limitations due to mass transport within the anode system are probably also not going to be particularly important in most cases. Present high specific power batteries involve liquid metal anodes, so that mass transport of the metallic element in the anode is not a problem. If alkali metal anodes are used at temperatures below the melting point of the alkali metal itself (lithium 179°C, sodium 97.5°C, potassium 62.3°C) difficulty might arise at very high specific current densities due to the formation of voids at the interface between the anode and the

electrolyte. However, ambient temperatures are sufficiently close to the melting point of the alkali metals that the rate of diffusion of vacancies is very high. As a result, any vacancies that form at the interface can diffuse away into the bulk of the metal rapidly, rather than accumulating to form porosity and a reduced area of contact to the electrolyte. As a matter of fact, small cells involving the use of solid sodium anodes with beta alumina have been shown to operate satisfactorily at room temperature with moderate values of interfacial currents.

On the other hand, transport processes within the cathode system are apt to be very important. The present approach to this problem is to use cathodes in which the reaction product is a liquid (e.g., Na_2S_5, which is liquid at 300°C). Diffusion within the liquid as well as mass flow of the bulk liquid serve to transport the reaction product away from the interface, so that the reaction can continue unhindered. If this does not happen rapidly enough, a blocking layer will form, and the whole process will be drastically slowed down.

This mass transport process must be accompanied at the electrolyte-cathode interface by reduction of the transporting M^+ ions by electrons supplied through an electronic conductor. In the case of the sodium-sulphur batteries now under development, this function is served by the use of graphite felt, which is pressed against the interface. It is, of course, desirable to have as good contact as possible between the graphite and the interface. However, the better the contact at this three-

phase interface, the more interference there is with diffusion and mass flow within the liquid. This is a source of considerable practical difficulty in some of the battery systems presently being developed.

Different Approaches to the Cathode Problem

A different approach to this problem would be to use a mixed conductor to supply the electrons for the reduction reaction. For example, if a thin layer of a mixed conductor were deposited upon the cathode surface of the electrolyte, electrons could be furnished across the whole electrolyte surface, and since the mixed conductor is essentially transparent to the ionic species, the area of contact with the liquid in the cathode compartment would not be reduced. In addition, there would be no geometric hinderence to diffusion or mass flow within the liquid.

By use of such a mixed conductor configuration one should be able to make some improvement in the overall kinetic performance of a liquid cathode system, and thus reduce the temperature of operation somewhat.

There are, however, other schemes that can be used to allow operation at even lower temperatures. One method would be to have a three-phase cathode system in which, in addition to the electronic conductor, a reactant phase is dispersed within an electrolyte. In this manner the mass transport throughout the whole cathode system is accomplished by means of the electrolyte, rather than requiring transport of the reaction product

away from the interface. By this means, the cathode operating temperature requirement can be substantially reduced, although there will be a weight penalty due to the presence of the electrolyte within the cathode compartment. This technique has been used to permit low temperature operation of sodium-conducting systems with inorganic cathode reactants such as $CuCl_2$[8], as well as several types of organic reactants[9].

A somewhat different approach would be the use of a solid mixed conductor to perform both functions at once, with the reactant finely divided within a mixed conductor matrix. In such a case, there is no need to sacrifice additional weight and volume for a separate electronic conductor such as graphite, as the mixed conductor can serve that function as well. Here again, however, one has to pay a weight penalty.

It has been recognized for some time[10] that the chemical diffusion coefficient in some mixed conductors can be extremely high. As a result of the coupled interaction between the electronic and ionic fluxes in such materials, diffusion coefficients of the order of 10^{-3} $cm^2/sec.$ can be observed in some cases. Although there have been relatively few measurements reported to date on mixed conductors containing mobile cations, it seems reasonable to expect that a number of useful materials of this type will be found.

These various different cathode structure possibilities are illustrated schematically in Fig. 8.

Storage Life

There are two types of factors that provide the primary limitations upon the storage life of any battery system. One of these has to do with the ability of the various constituent materials present to withstand the various kinds of corrosion and chemical attack that arise because of the presence of active chemical species, both oxidizing and reducing, in such systems. As a general rule, these problems become more severe the higher the temperature, so that they do not pose any inherent limitation upon the possibilities for reducing the operational temperature of presently conceived high performance battery systems.

The second, and perhaps more fundamental, matter relates to self discharge. In batteries and fuel cells with liquid electrolytes there are two general mechanisms by which self discharge can take place. Uncharged (generally molecular) species can migrate from one electrode compartment to the other. This typically occurs by diffusion and/or convection through the electrolyte or through an adjacent gas phase. In addition, electronic conductivity through the electrolyte allows compensating internal ionic conductivity, and thus self discharge.

One of the inherent advantages of solid electrolytes is that the transport of uncharged species can be virtually prohibited. In addition, it has been shown[11] that, at least in some solid electrolytes, the electronic conductivity at ambient temperatures can be many orders of magnitude lower than typically found in liquid electrolytes. As a result, the storage, or shelf,

life of solid electrolyte batteries is one of their great potential advantages, and this is certainly not a limiting feature to their ambient temperature use.

Operating Life

Unfortunately, very little can presently be said about the prospects for the operating life of ambient temperature cells. The primary reason for this is that there is as yet very little understanding of the reasons for deterioration of present cells at elevated temperature. Until the pertinent mechanisms can be identified, it seems useless to speculate about this matter, despite its obvious importance.

Summary

The important parameters relating to the performance of advanced solid electrolyte battery systems and the factors that limit their values have been discussed. A review of the controlling mechanisms has indicated that there should be no fundamental difficulty in achieving high values of specific energy at ambient temperatures.

The major bottleneck that is to be expected in the quest for better battery systems to operate at low temperatures lies in the problem of achieving high specific power values. Although more work has to be done on the electrolyte problem, it seems clear that the primary limitation will be in the cathode system. Two new approaches to cathode system design involving the use of solid mixed conductors have been suggested.

The characteristics of solid electrolytes lead to the expectation that batteries with appropriate solid electrolytes should have a storage life vastly superior to present experience with systems utilizing liquid or paste electrolytes.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with the University of Michigan.

References

1a. R. T. Foley, J. Electrochem. Soc. 116, 161 (1969).

b. M. N. Hull, Energy Conversion 10, 215 (1970).

c. J. N. Mrgudich, "Solid Electrolyte Batteries", *Encyclopedia of Electrochemistry*, Reinhold, New York, 1964, p. 84.

d. T. Takahashi, Denki Kagaku 36, 402 (1968).

e. T. Takahashi, Denki Kagaku 36, 481 (1968).

f. B. B. Owens, Advances in Electrochemistry and Electrochemical Engineering 8, 1 (1971).

g. E. J. Cairns and H. Shimotake, Science 164, 1347 (1969).

h. R. Jasinski, *High Energy Batteries*, Plenum Press (1967).

i. U.S. Department of Commerce Report, "The Automobile and Air Pollution", (1967).

j. J. L. Sudworth and I. Dugdale, Power Sources 2, 547 (1968).

k. J. L. Sudworth and M. D. Hames, Power Sources 4, 1 (1970).

l. M. N. Hull, Proc. 22nd Annual Power Sources Conference (1968), p. 106.

m. L. Heyne, Electrochim. Acta 15, 1251 (1970).

n. D. O. Raleigh, to be published in Adv. Electroanal. Chem. (1972).

2a. U.S. Department of Commerce Report, "The Automobile and Air Pollution", (1967).

b. J. T. Salihi, "Impact of Electric Cars on Environment and the Economy - Battery Development for the Electric Car", to be published in Spectrum (June, July, 1972).

3. F. G. Will, "Large Scale Energy Storage and Solid Electrolyte Batteries", presented at the Electrochemical Society meeting, Houston (May, 1972).

4a. B. B. Owens, Advances in Electrochemistry and Electrochemical Engineering 8, 1 (1971).

4b. S. Geller and B. B. Owens, "Silver Ion Site-Distribution, Structure and Conductivity of the Solid Electrolyte Pyridinium Hexaiodopentaargentate, $[C_5H_5NH]\ Ag_5I_6$, Between -30 and 125°C", to be published.

5a. M. S. Whittingham and R. A. Huggins, J. Chem. Phys. 54, 414 (1971).

b. M. S. Whittingham and R. A. Huggins, "Beta Alumina - Prelude to a Revolution in Solid State Electrochemistry", to be published in National Bureau of Standards Pub. No. 364, ed. by R. S. Roth and S. J. Schneider (1972).

6a. Y. F. Y. Yao and J. T. Kummer, J. Inorg. Nucl. Chem. 29, 2453 (1967).

b. H. Saalfeld, H. Matthies, and S. K. Datta, Ber. Deutsche. Keram. Ges. 45, 212 (1968).

c. M. S. Whittingham, R. W. Helliwell and R. A. Huggins, U.S. Government Res. and Devel. Rept. 69, 158 (1969).

d. R. H. Radzilowski and J. T. Kummer, Inorg. Chem. 8, 2531 (1969).

7a. M. S. Whittingham and R. A. Huggins, J. Electrochem. Soc. 118, 1 (1971).

b. M. S. Whittingham and R. A. Huggins, J. Chem. Phys. 54, 414 (1971).

c. M. S. Whittingham and R. A. Huggins, "Beta Alumina - Prelude to a Revolution in Solid State Electrochemistry", to be published in National Bureau of Standards Publ. No. 364, Ed by R. S. Roth and S. J. Schneider (1972).

8. K. O. Hever, cited by J. T. Kummer, MAP-PENNTAP Seminar on Materials Problems and Opportunities in Energy Storage, Philadelphia (1971).

9. Work at TRW, H. P. Silverman, private communication.

10. C. Wagner, J. Chem. Phys. 21, 1819 (1953).

11. M. S. Whittingham and R. A. Huggins, J. Electrochem. Soc. 118, 1 (1971).

TABLE 1.

Examples of Solid Ionic Conductors that have Phase Transformations at Relatively Low Temperatures[4]

Material	Transition Temperature
CuI	400°C
Ag_3SI	235°C
AgI	144°C
Cu_2HgI_4	67°C
Ag_2HgI_4	50°C
$[C_5H_5NH]Ag_5I_6$	50°C
KAg_4I_5	-136°C
$RbAg_4I_5$	-155°C

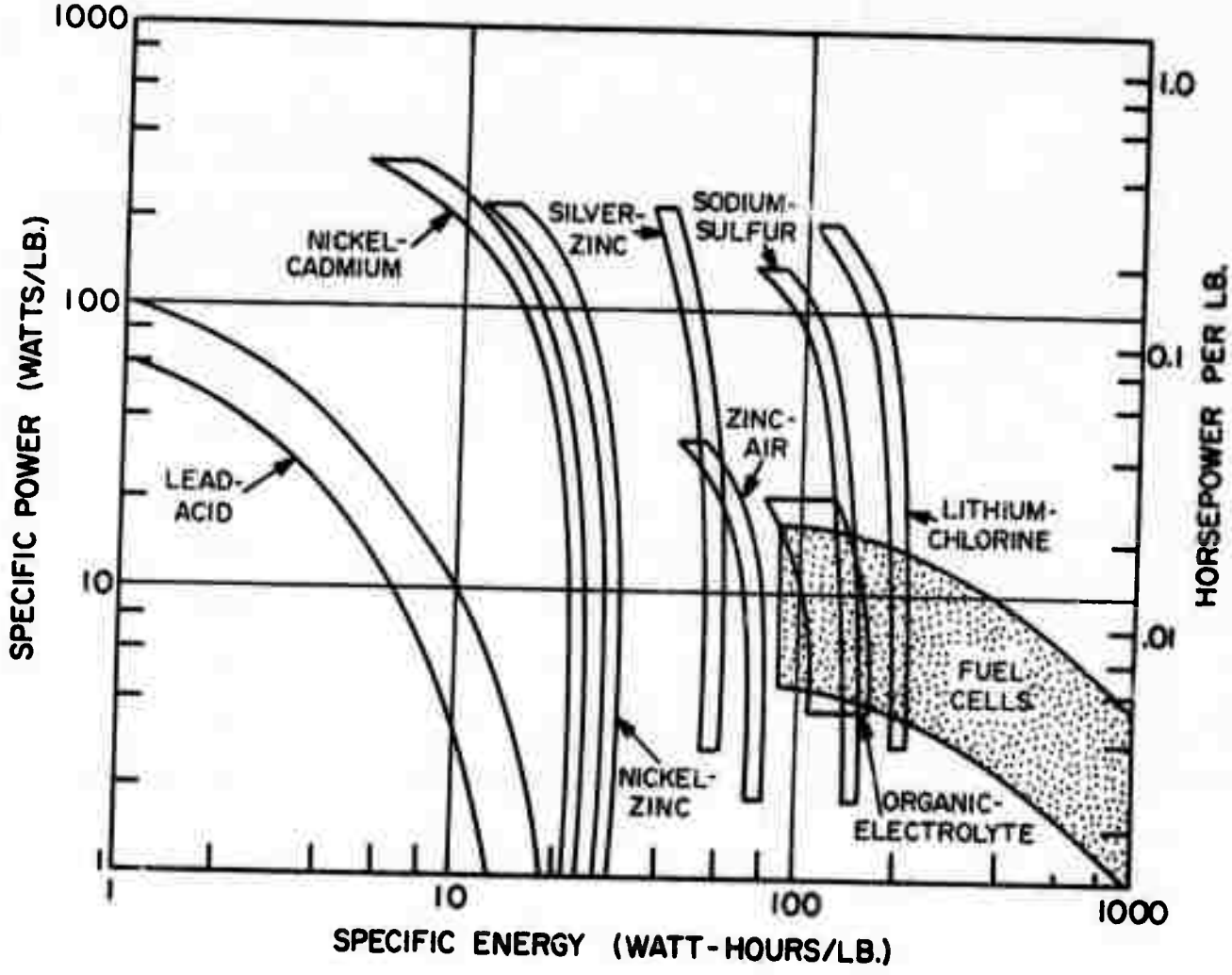


Figure 1. Values of specific power and specific energy for various types of battery and fuel cell systems considered for vehicle propulsion applications. (From The Automobile and Air Pollution, U.S. Department of Commerce, 1967 (1.i.)).

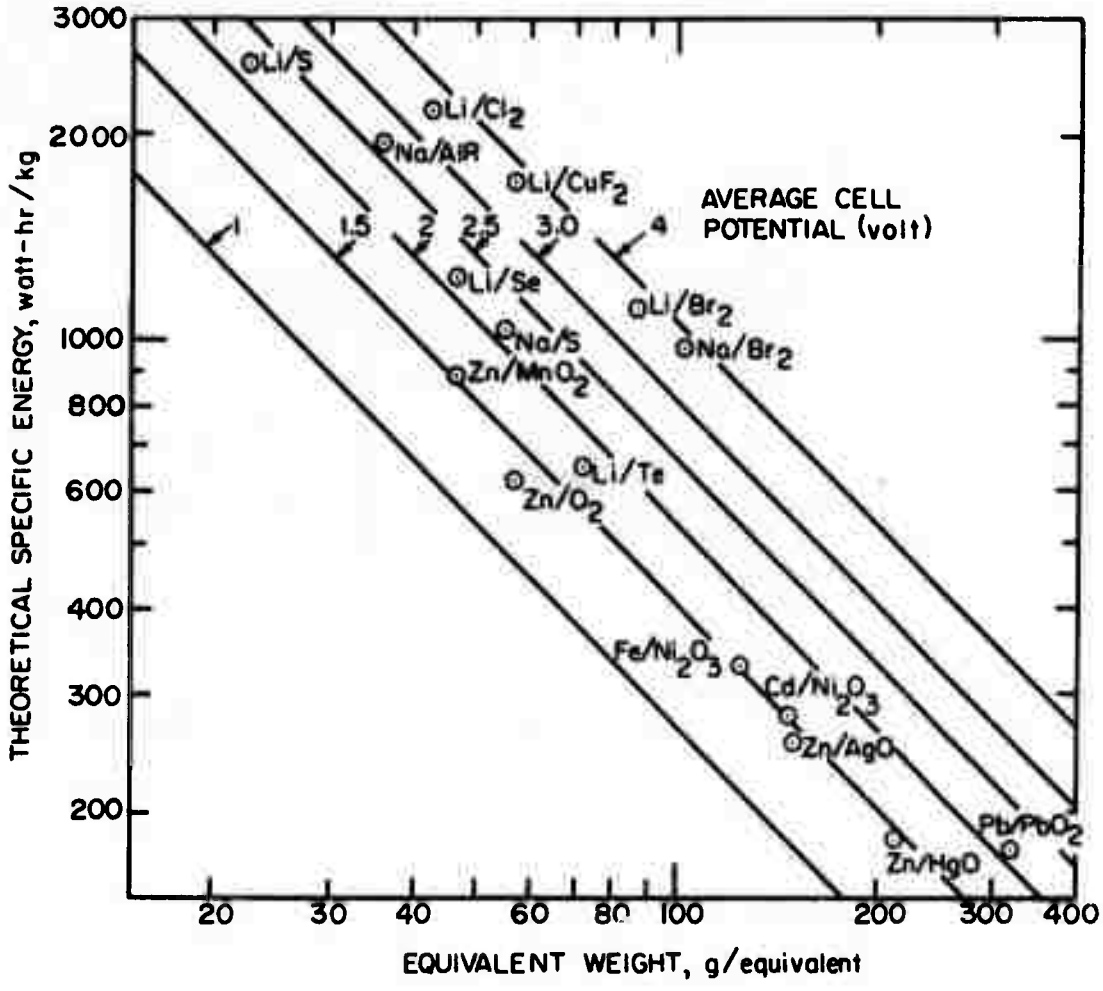


Figure 2. Maximum theoretical specific energy for various electrochemical reactant systems plotted versus their gram equivalent weights. This illustrates the importance of both the electron gravity difference and the atomic weights of the constituents of battery systems. (From Cairns and Shimotake, 1969 (l.g.)).

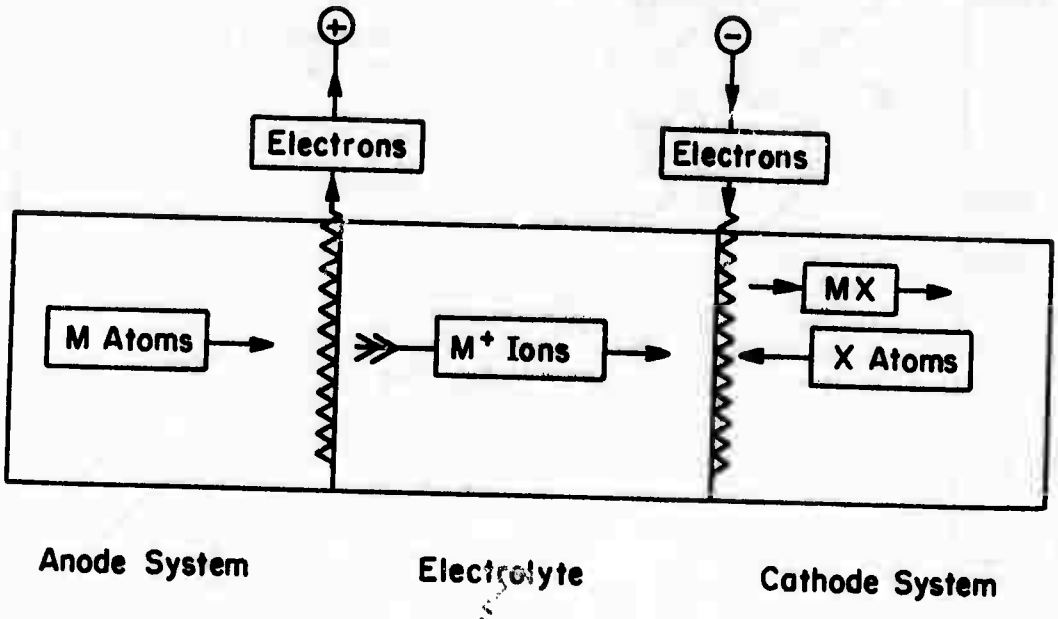


Figure 3. Schematic representation of elementary electrochemical cell involving transport of M^+ ions through the electrolyte from the anode system to the cathode system.

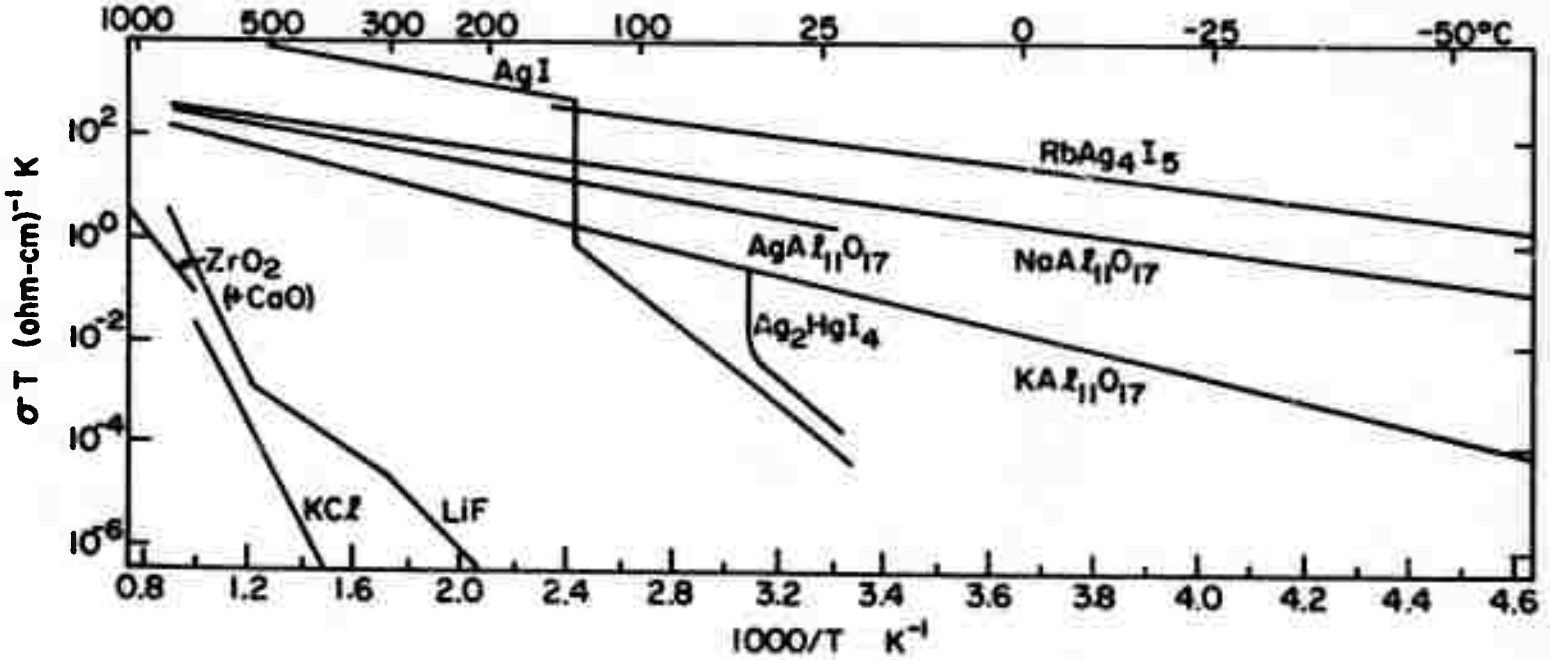


Figure 4. Conductivity data for various ionic conductors compared over a wide range of temperature. Note the tremendous difference at low temperatures between normal close-packed materials like KCl and LiF and the superionic conductors such as $RbAg_4I_5$ and the beta aluminas.

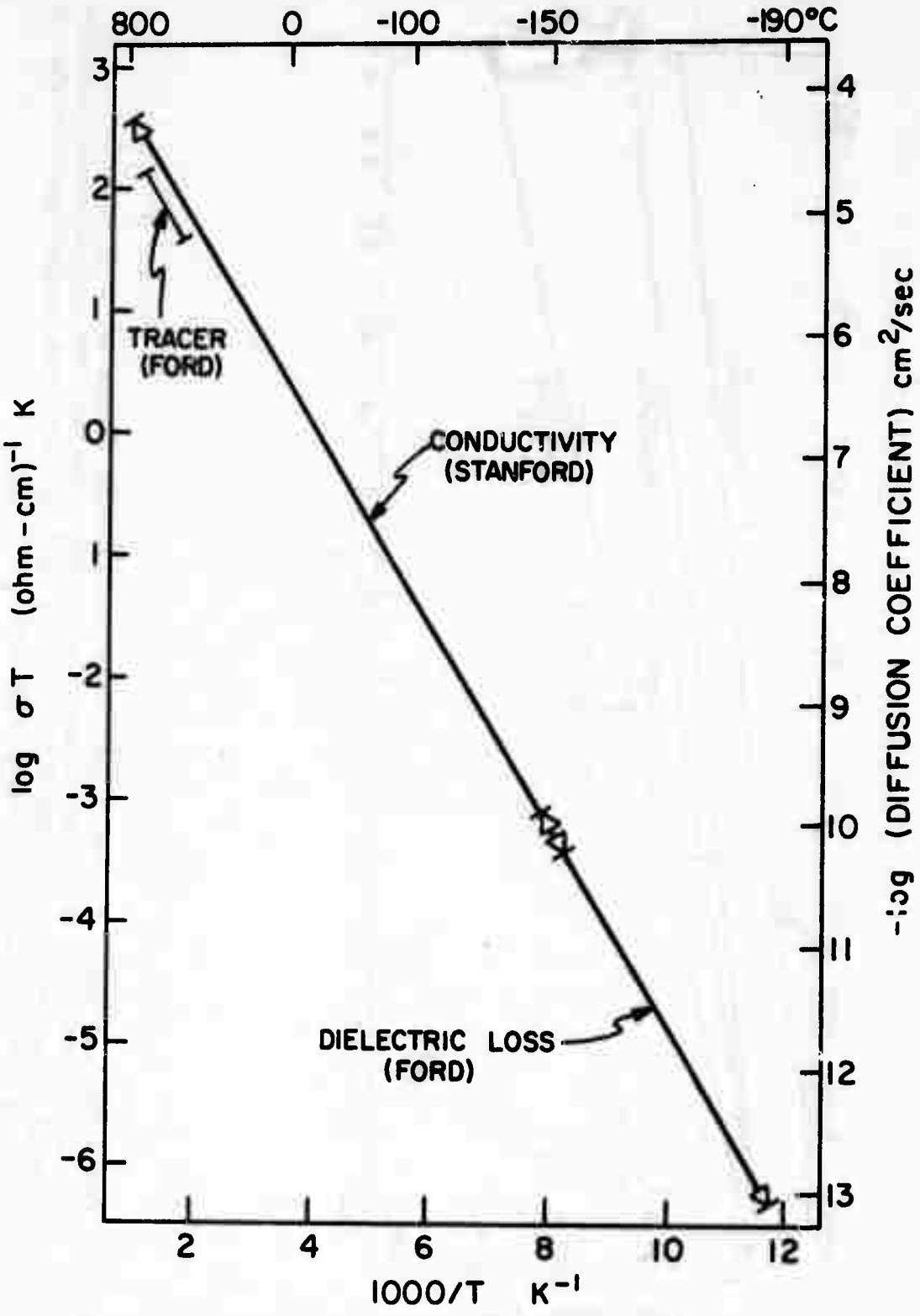


Figure 5. Ionic conductivity, radiotracer diffusion, and dielectric loss data for sodium beta alumina. (From Whittingham and Huggins, 1972 (5)).

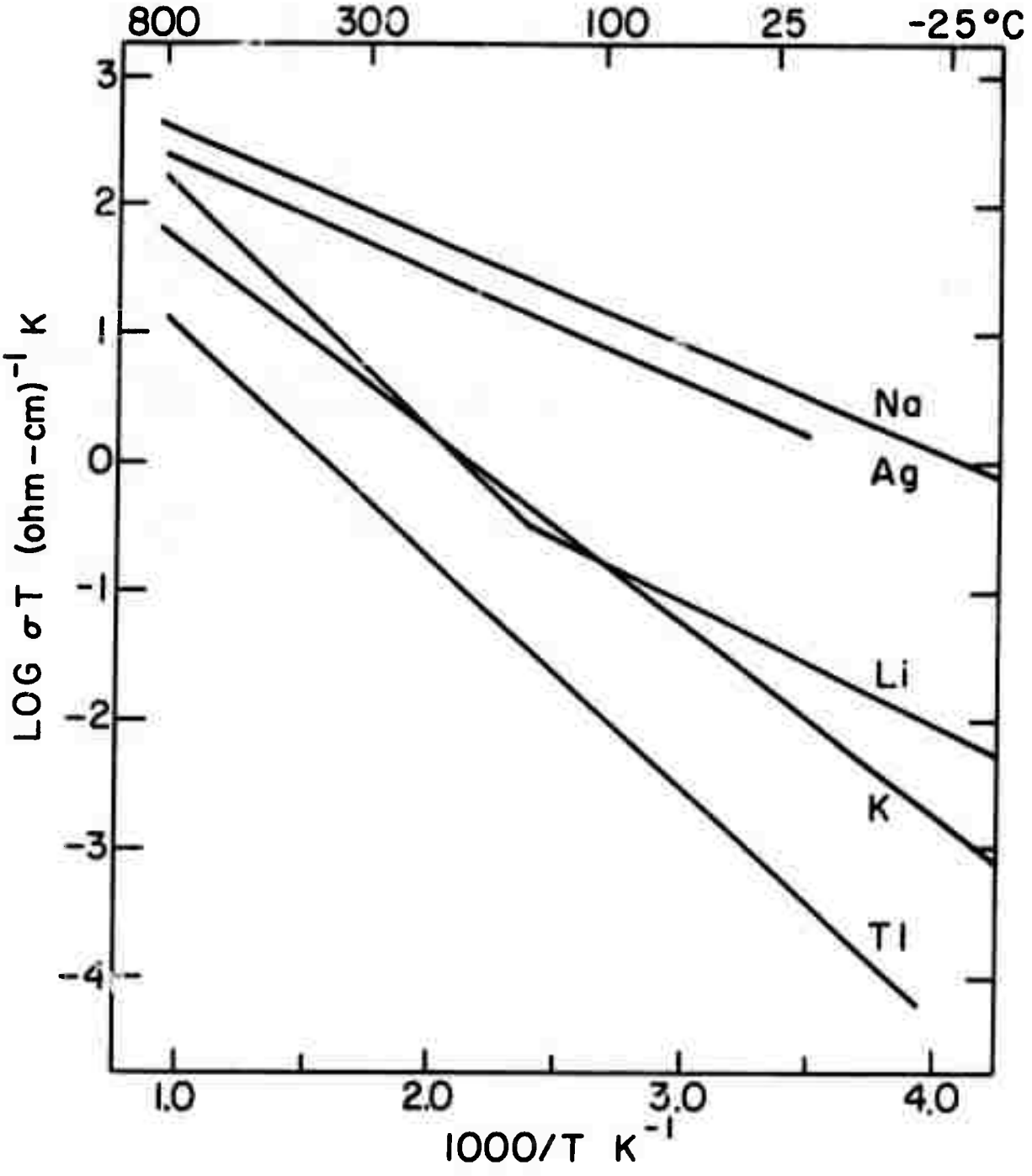


Figure 6. Conductivity of five different ions in beta alumina, $MAl_{11}O_{17}$. (From Whittingham and Huggins, 1972).

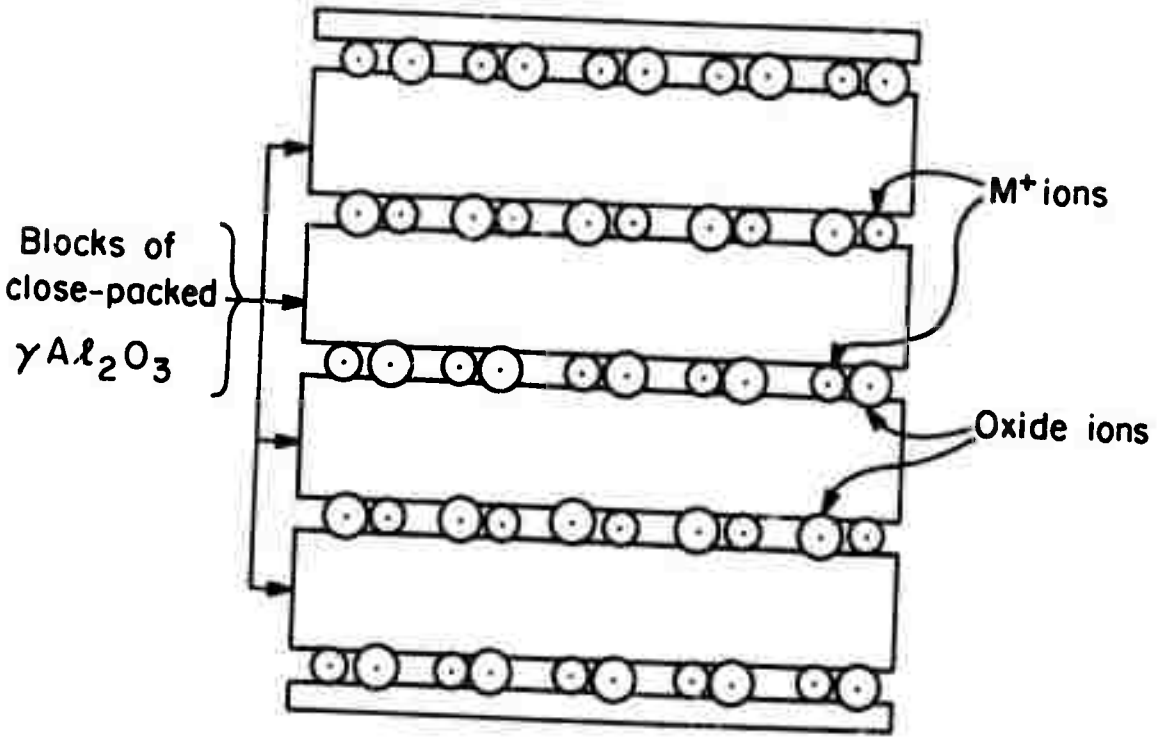
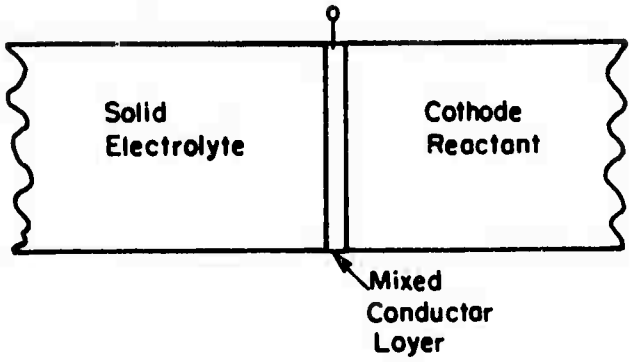
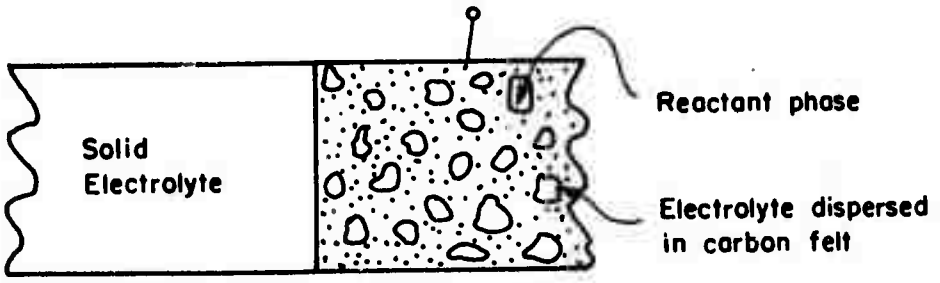


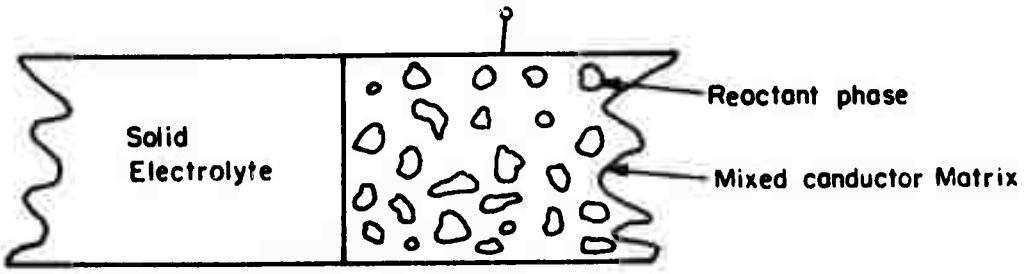
Figure 7. Schematic representation of the structure of beta alumina.



a. Mixed conductor layer between solid electrolyte and cathode reactant.



b. Three-phase system with reactant dispersed within electrolyte inside carbon felt.



c. Two-phase system with reactant dispersed within mixed conductor.

Figure 8. Various different possible types of cathode structures.

HARDNESS OF PURE ALKALI HALIDES

J. J. Gilman

Abstract

An explicit expression is derived for the indentation hardness numbers of alkali halide crystals in terms of ionic binding, combined with the theory of plastic identation. The resistance to indentation is caused by the electrostatic faults that exist at the cores of {100} $\langle \bar{1}10 \rangle$ dislocations.

The theoretical hardness is given by (c.g.s. units):

$$H_o = 0.096 \frac{e^2}{\varepsilon b^4} = 1.2 \times 10^{-2} C_{44}$$

where e = electron charge; ε = static dielectric constant; b = Burgers displacement; and C_{44} = elastic shear stiffness. The theoretical hardness for NaCl is 17 kg/mm^2 compared with the observed value of 16.7 kg/mm^2.

HARDNESS OF PURE ALKALI HALIDES

J. J. Gilman

Recent data obtained by Chin, Van Uitert, Green and Zydzik (1972) exhibit a simple relationship between the indentation hardnesses of pure salts and their elastic stiffnesses (Fig. 1). The purpose of this paper is to show that their result is consistent with a simple quantitative model of the resistance to dislocation motion on secondary glide planes in these crystals.

The authors mentioned above measured the hardnesses of sodium and potassium halide crystals that contained various amounts of deliberately included impurities. They obtained linear correlations and by extrapolating to zero impurity contents, they obtained values for the hardnesses of the pure salts which they designed, H_o. Next, they found a linear correlation between H_o and the Young's moduli in the cube diagonal directions, E_{111}. The author has found an equally good correlation with the stiffness constant, C_{44} as shown in Fig. 1. The correlation is:

$$H_o = 1.4 \times 10^{-3} C_{44} \ (kg/mm^2) \qquad (1)$$

and calculation of the coefficient is the purpose here.

I. Hardness and Flow Stress

Pure salts flow on their primary glide planes at stresses at least as low as $10 g/mm^3$ which is about 10^3 smaller than the observed H_o values. However, the plastic strains produced by a hardness indenter cannot be accomodated completely by primary glide so it is necessary for secondary systems to become active during indentation. An analysis of the effect of this on the plastic constraint factor has been given by J. R. Rice and reviewed by Gilman (1972).

The primary glide system in alkali halide crystals is $\{1\bar{1}0\}\langle\bar{1}10\rangle$ and the secondary system is $\{001\}\langle\bar{1}10\rangle$. Since the flow stress for the secondary system is much higher than for the primary system, it limits the behavior. If its magnitude is τ_s it is related to the hardness number by:

$$\tau_s = H/2\sqrt{2} \qquad (2)$$

according to the model of Rice; and it exerts a force, $\tau_s b$ per unit length of secondary dislocation line.

The author has pointed out elsewhere (Gilman, 1961) that secondary flow in crystals with the rock-salt structure is resisted by the fact that it requires ions to become juxtaposed in a high energy configuration when the structure is half-sheared. That is, when a relative translation of b/2 has occurred on the glide plane (b is the Burgers displacement).

If the incremental energy per unit length of the dislocation line for this configuration is U_m then the force, $b\tau_s$

provided by the applied stress must do enough work to supply this amount of energy. The distance through which the force acts is b/2 so:

$$U_m = \frac{b^2 \tau_s}{2} \left(\frac{ergs}{cm}\right) \tag{3}$$

and U_m is related to H_o through Eq. (2):

$$H_o = \frac{4\sqrt{2}\, U_m}{b^2} \tag{4}$$

This reduces the present problem to the calculation of U_m.

II. The Mid-Glide Energy of a {001} <$\bar{1}$10> Dislocation in the Rock-Salt Structure

Juxtaposition of ions of the same charge sign creates a line in the glide plane along which there are repulsive electrostatic forces acting as illustrated in Fig. 2 taken from Gilman (1961).

The electrostatic situation is illustrated with further detail in Fig. 3. The overall configuration consists of an infinite set of planes spaced b/2 apart. Each plane contains three ions of the same sign, and the sign alternates from one plane to the next. The ions lie at the corners of isoceles triangles alternately pointing up and down. They have bases and heights of lengths b and $a = b/\sqrt{2}$, respectively.

To obtain an expression for the net repulsive force per unit length along the set of planes, a point is chosen at the center of one of the triangles and the radii forces that act on it are calculated (by symmetry the forces normal to the

planes cancel).

Consider plane 1 in isolation first. Its configuration is shown in Fig. 4 where the force center is chosen such that the net radial force acting through it is zero. Then any one force will be representative. Take the vertical force which results from the repulsions between ions 1 and 2 and between 2 and 3. Each repulsive force is $q_i q_j / \varepsilon \ell^2$ where $q_i q_j$ are the ionic charges, ℓ is the interionic distance and ε is the dielectric constant. In order to estimate the screening effects of the surrounding polarizable medium, partial charges on the ions are used. These partial charges are chosen in proportion to the fractions of the spherical ions that lie within the appropriate boxes in Fig. 3. In the case of Fig. 4, for example, $q_1 = e/2$ and $q_{2,3} = e/4$. The net vertical repulsive force is:

$$f_1^r = \frac{2}{E}\left(\frac{q^2}{l^2}\right) \cos\theta = \frac{ae^2}{\varepsilon b^3}\left(\frac{1}{3}\right)^{3/2} \qquad (5)$$

In Fig. 5 the configuration for calculating the radical force that acts through the force center from the planes labeled +2 and -2 is shown. By symmetry, only one ion pair needs to be considered; namely, pair (2,4). (The interactions of 1 with 6 and of 3 with 5 are neglected). The result, remembering that there are two similar planes is:

$$f_2^a = \frac{ae}{\varepsilon b^3}\left(\frac{1}{3}\right)^{3/2} \qquad (6)$$

Thus, the first and second terms cancel as the author has previously suggested on intuitive grounds (Gilman, 1961). However,

a second order effect remains because the subsequent terms f_3^r, f_4^a, etc., yield a net repulsion. The first six terms of the series are:

$$f = \frac{ae^2}{\varepsilon b^3}\left(3^{-3/2} - 3^{-3/2} + 7^{-3/2} - 11^{-3/2} + 19^{-3/2} - 27^{-3/2}\right)$$

and the series converges to ~ 0.035 so:

$$f \simeq 0.035\ \frac{ae^2}{\varepsilon b^3} \tag{7}$$

This line of forces can be viewed as producing a pressure on the inside of a cylindrical hole of radius, r, where r is to be chosen to make the electrostatic and elastic fields self-consistent. However, r will not be determined here because the purpose is simply to show that the elastic strain energy is small compared with the electrostatic energy. For this purpose it is sufficient to let $r = a = b/\sqrt{2}$.

The electrostatic force inside r produces an elastic strain field around the hole that exerts a counter-vailing pressure of opposite sign to contain the force line. Expressed per unit length and in terms of r, the force is:

$$F = \frac{f}{b} \simeq \frac{0.017}{\varepsilon}\left(\frac{a}{b^2}\right)\frac{e^2}{r^2} \simeq \frac{A}{r^2} \tag{8}$$

The pressure to be contained is:

$$P = \frac{F}{2\pi r} \tag{9}$$

and the energy of the strain field that it creates can be found by integrating the work done by the pressure during the creation

of the field. The integral of the surface tractions times the displacements over the surface of the cylinder gives this work (in polar coordinates):

$$U_s = \frac{1}{2} \oint (\sigma_r u_r + \tau_{r\theta} u_\theta) ds \tag{10}$$

Now the angular displacements are zero in this particular case so this integral retains only its leading term. Using results from Love (1927, p. 213) the radial displacements, u_r, and the stress, σ_{rr}, in the presence of a pressure, P, are:

$$u_r = \frac{Pr}{2\mu} \; ; \qquad \sigma_{rr} = P \tag{11}$$

where μ = shear modulus. Then the strain energy is:

$$U_s = \frac{P^2 r}{4\mu} \int_0^{2\pi} r d\theta \tag{12}$$

Since

$$P = \frac{A}{2\pi r^3}$$

the expression for the strain-energy becomes:

$$Us = \frac{+A^2}{8\pi\mu r^4} \tag{13}$$

Integrating the electrostatic force from $r = \infty$ to r yields the following for the electrostatic energy:

$$U_e = \frac{A}{r} \tag{14}$$

which is about 200 times larger than the elastic energy so the latter can be neglected.

III. Comparison of Theory and Experiment

Combining Eqs. (4) and (14) yields:

$$H_o = 0.096 \frac{e^2}{\varepsilon b^4} \tag{15}$$

and electrostatic theory gives the following expression for the elastic stiffness (Krishnan and Roy, 1952):

$$C_{44} = 0.35 \frac{e^2}{r_o^4} \tag{16}$$

Recalling that $b = \sqrt{2}\, r_o$ and substituting Eq. (16) into (15) and assuming that ε scales with r_o^4 gives:

$$H_o \text{ (calc.)} = 1.2 \times 10^{-2} C_{44} (d/cm^2) \tag{17}$$

which gives excellent agreement with the experimental Eq. (1).

Furthermore, Eq. (15) can be used to calculate particular values of hardness. For example, NaCl has the parameters: $e = 4.8 \times 10^{-10}$ esu; ε static $= 5.6$; and $b = 3.9 \times 10^{-8}$ cm.; so its calculated hardness is:

$$H_o^{calc} = 17 \text{ kg/mm}^2$$

compared with the measured value of $16.7^{kg}/mm^2$. Again, the agreement is excellent.

Because of the many assumptions of this theory the agreement with experiment should not be overly emphasized. Nonetheless, the fact that in this case hardness can be calculated in terms of fundamental quantities is quite remarkable and supports the qualitative features of the model.

IV. Summary

An explicit expression is derived for the indentation hardness numbers of alkali halide crystals in terms of ionic binding, combined with the theory of plastic indentation. The resistance to indentation is caused by the electrostatic faults that exist at the cores of {100}<110> dislocations.

The theoretical hardness is given by (c.g.s. units):

$$H_o = 0.096 \frac{e^2}{\varepsilon b^4} = 1.2 \times 10^{-2} C_{44}$$

where e = electron charge; ε = static dielectric constant; b = Burgers displacement; and C_{44} = elastic shear stiffness. The theoretical hardness for NaCl is 17 kg/mm^2 compared with the observed value of 16.7 kg/mm^2.

Acknowledgement

This research was supported in part by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with the University of Michigan.

References

G. Chin, L. Van Uitert, M. Green, and G. Zydzik, "Hardness, Yield Strength and Young's Modulus in Halide Crystals", Scripta Met. 6, 475 (1972).

J. J. Gilman, "Hardness - A Strength Microprobe", to be published by ASM (1972).

A. E. Love, *A Treatise on the Mathematical Theory of Elasticity*, 4th Ed., Cambridge University Press (1927).

J. J. Gilman, *Progress in Ceramic Science*, Vol. 1, Ed. J. E. Burke, Pergamon Press, London (1961) p. 157.

K. S. Krishnan and S. K. Roy, "Proc. Roy. Soc.", 210A, 481 (1952).

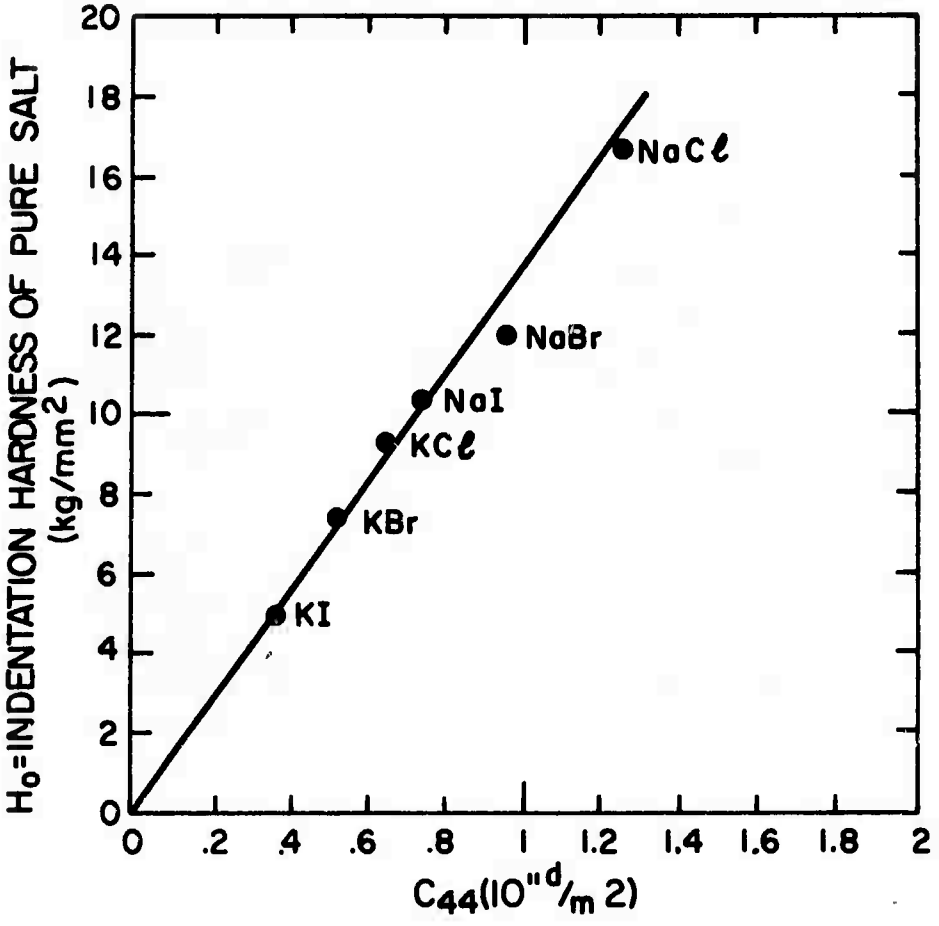


Figure 1. Correlation of Hardness and Elastic Shear Stiffness for Pure Alkali Halides.

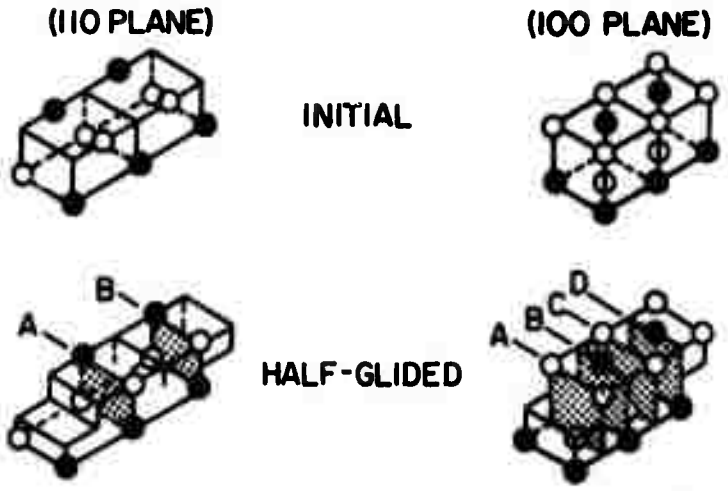


Figure 2. Ion Configurations at Glide Planes in Rock Salt Structure Comparing Initial and Holy-glided Positions for Two Planes.

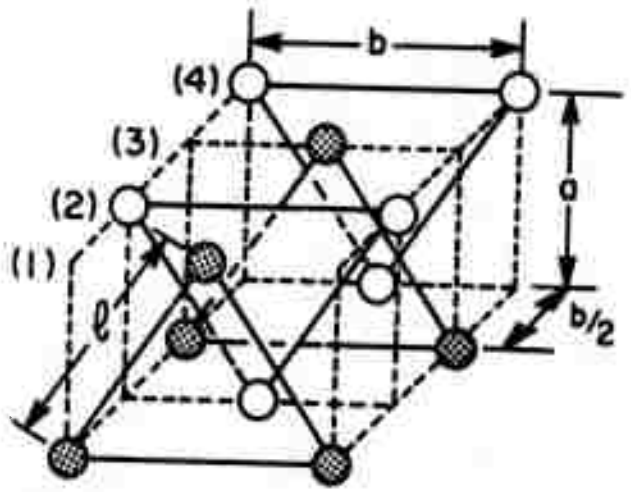


Figure 3. Configuration of the Electrostatic Fault for the Holy-Glided Position on the {100} Plane.

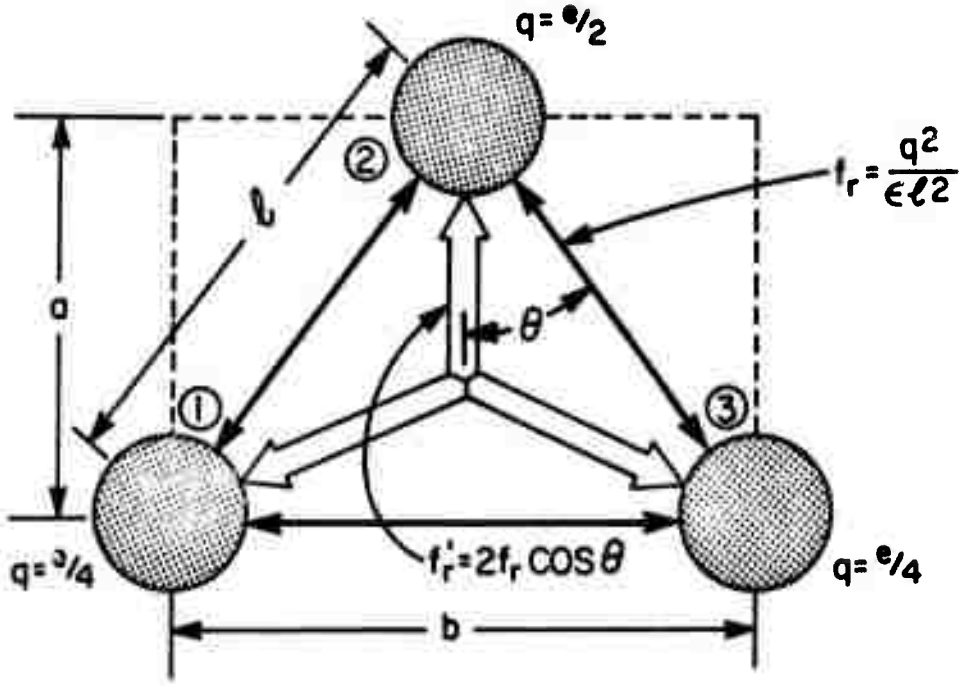


Figure 4. Detail and Force Diagram for Plane (1) of Figure 3.

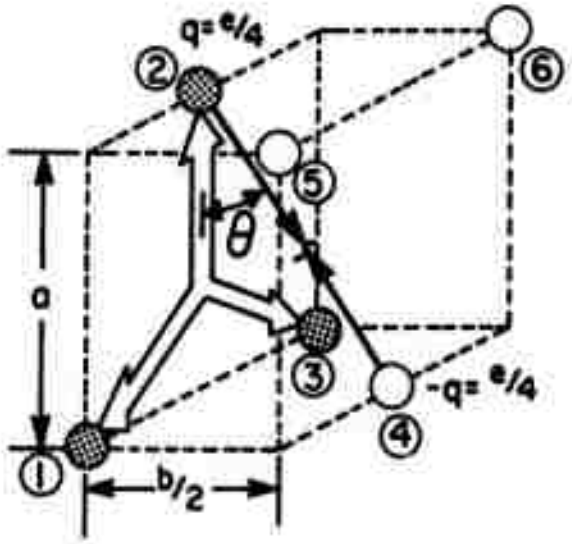


Figure 5. Force Diagram for Planes (1) and (2) of Figure 3.

COMPUTER EXPERIMENTS ON ATOMIC MODELS OF CRACKS: THOUGHTS ABOUT PROBLEMS AND OPPORTUNITIES

G. H. Vineyard

Abstract

Although the process of the formation and propagation of cracks in brittle fracture has been under investigation for a long time, new opportunities for improved understanding are foreseen in computer experiments on atomic models of the crack tip and its surrounding region. Exploratory work has been reported recently by Gehlen, Kanninen, Chang, and others. Possible atomic models for extensions of this work are considered, and a number of problems inherent in such computations are discussed. The atomistic calculations are divided into two classes, static and dynamic. A point of view is recommended which goes beyond the usual description of crack fronts and related energies by dealing with extremal points of the potential energy in configuration space. This viewpoint provides a unified and more nearly rigorous framework for discussion and is naturally related to the results of static computer experiments. Problems that may be illuminated if not completely solved by computer experiments are discussed. In particular, the role of jogs and kinks in crack propagation and the question of the formation of dislocations by advancing cracks are seen as suitable for more concentrated attack.

COMPUTER EXPERIMENTS ON ATOMIC MODELS OF CRACKS: THOUGHTS ABOUT PROBLEMS AND OPPORTUNITIES

G. H. Vineyard

I. Introduction

Most theoretical treatments of the process of crack formation and propagation are based on continuum models, and the atomic details are imported, where necessary, in ad hoc fashion. Recently exploratory computer experiments on atomic models have been made[1-7], and this approach offers interesting new possibilities for learning more about the true state of affairs at the tip of a crack, and perhaps about fracture in general.

It should be admitted that it is conceivable that knowing a great deal more about atomic configurations and atomic processes at the tips of cracks would not contribute substantially to understanding the complex problems of fracture. Thus the scale on which fracture phenomena are critically determined might be larger than atomic, which is to say that fracture processes might be very insensitive to atomic details. In this event, computing more about such details would be only "scientific voyeurism". However, I do not believe this is the case, and the following discussion is based on that assumption.

II. Elements of the Models

A model of the type under discussion consists of a finite set of mass points (called the explicit set) which interact with assumed forces and represent the atoms near the crack tip. The explicit set is surrounded in some artificial manner that simulates a very much larger set of atoms than can be treated in a computer (ideally, an infinite set), and in which the explicit set can be thought to be imbedded The boundaries of the larger set are supplied with stresses or strains of arbitrary magnitude capable of producing fracture processes within the explicit set.

In the simplest procedure, all the atoms in two opposite faces of the explicit set itself are subjected to outward forces, F, simulating uniform tensile stresses imposed on a small specimen. Alternatively, shearing stresses can be imposed, or uniform displacements on the bounding faces. An infinitely long straight crack can be modeled to advantage by applying periodic boundary conditions to the atom in boundary planes normal to the crack edge (or edges). Cracks of more complex shape can be modeled in this way if a larger repeat distance is used.

In general, the stress field around a crack falls off more slowly with distance than the field around point defects*, and this imposes serious difficulties in the choice of boundary conditions, similar to (but worse than) the difficulties with

*The stress field around a linear crack falls off at large distances like $1/\sqrt{r}$ where r is the distance from the crack edge, while the stress field around a "penny-shaped" crack (or around a point defect) falls off as $1/r^3$.

atomic models of dislocations. With one dimensional cracks the difficulty is most acute, and work to date indicates that qualitatively wrong results can be obtained if fixed displacements are imposed on the boundaries of the explicit set in the case of a linear crack[7]. Gehlen and Kanninen[7] have imbedded the explicit set in an elastic continuum, and in a manner similar to that employed earlier by Johnson and others for point defect calculations, have relaxed the two systems by successive approximations so as to obtain a self consistent imbedding. This method avoids the artifacts associated with rigid boundary conditions for linear cracks.

III. The Computations

Two distinct types of computations can be made. The first, and simpler, are the essentially static calculations, which explore the potential energies of various atomic configurations; the second type are the dynamic calculations in which atomic motions are computed, using Newtonian dynamics, starting from an assumed set of atomic coordinates and velocities. Various relaxation processes and quasidynamic processes that are used to find extremals of the potential energy are considered in this terminology to be static calculations. It is important to distinguish the kinds of information available from the two types of calculations. Since the static calculations are so much simpler, much of the work should be of that type. The static results form a background from which dynamic events can be selected, classified, and pro-

perly understood.

A. Static calculations

A unified point of view with respect to the static calculations can be expressed as follows: Let the (3-vector) coordinates of the N atoms in the explicit set be $\vec{r}_1$, $\vec{r}_2$, ... $\vec{r}_N$, and the M coordinates employed to describe the displacements in the imbedding medium be s_1, s_2, ... s_M. Altogether the problem has 3N + M degrees of freedom. A configuration of the system is a particular set of values of all these coordinates, abbreviated $\vec{R}$. It can be conveniently thought of as a point in a space of 3N + M dimensions. Let the forces applied at the boundaries of the entire system be denoted F (which may be a single number or a complex of numbers). The case of specified displacements at the boundary is similar, and, alternatively, F could represent these displacements. The model possesses a potential energy which depends on the configuration $\vec{R}$ and also on F; the latter is conveniently treated as a parameter. Denote this potential energy $\Phi_F(\vec{R})$. The computer evaluates $\Phi_F(\vec{R})$, and by means of various possible algorithms searches out its extremals. The essence of the static problem, then, is the finding of extremal configurations, the evaluation of the energy of those configurations, and the determination of the dependence of both energies and configurations on F.

Extremals may be classified further, as minima, maxima, and saddle points. The minima are points of equilibrium, either

stable or metastable. In the vicinity of minima there will be saddle points, and passage from one equilibrium configuration of the system to another by atomic displacements (e.g., opening or closing of the crack, movement of jogs, etc.) will be accomplished most easily by the traversal of at least one saddle point. "Accomplished most easily" means with the least increase of potential energy during the traversal. "Peierls-Nabarro" type barriers to the advance of a crack are represented quite generally by the occurrence of such saddle points between minimum points in $\Phi_F(\vec{R})$. The activation energy for thermally assisted jumping from one crack position to the next is the potential energy of the intervening saddle point minus the potential energy of the minimum point. The entropy of activation can be calculated from the curvatures of the potential energy surface at the minimum point and at the saddle point[8].

Increases of the force parameter F can cause a minimum point to disappear. The least value of F that does this is the critical stress (or strain) for forcing the crack out of that configuration. The critical F for each type of minimum point is of interest. In a similar way the dependence on F of the free energy of activation for each thermally assisted process can be computed*.

*Usually the computation of activation free energy and its use with absolute rate theory is a much more expeditious way of determining the influence of temperature on crack motion than the direct introduction of thermal vibrations into a dynamic computer model, although the latter approach may have interest as an aid to visualization, and has been applied occasionally in an effort to check the assumptions on which rate theory is based.

By discussing the extremal points of $\Phi_F(\vec{R})$ one has a quite general approach to the equilibrium forms of the crack, the influence of applied stress, etc. Concepts such as the shape and position of the crack, which are essential in the continuum approach, cannot be precisely defined on an atomic scale. Knowledge of the structure of the function $\Phi_F(\vec{R})$ in configuration space makes it unnecessary to attempt to define them. Arbitrary conventions can be set up for convenient description of the successive minima of $\Phi_F(\vec{R})$ in terms of an effective crack shape and position. The arbitrariness in the conventions then becomes clear. In particular, it can now be seen that it is not possible to define unambiguously a surface energy of a crack which varies with the crack position on an atomic scale of fineness. Moreover, knowledge of $\Phi_F(\vec{R})$ renders such a definition unnecessary. It should also be remarked that all of the foregoing discussion is valid without regard to the range of interatomic forces (except that interatomic forces of too long a range will require values of N that are unmanageably large for practical computations). Thus the question of which bonds at the crack tip have been broken and which have not, which cannot be answered clearly when forces are longer ranged than near-neighbors, is also seen not to require an answer.

Computer programs that locate extremals usually find local maxima or minima, but not saddle points. Various quasi-dynamic procedures, in particular, all have this property. After two near-by minima have been located, the intervening saddle point

can be located by adding an adjustable constraint such that for some value of the adjustment parameter the system can be in the saddle point; because of the constraint the saddle point then becomes a minimum point. In the case of a point defect calculation in which an atom can jump between two equilibrium positions, a suitable constraint is provided by confining the jumping atom to a plane approximately normal to the path it prefers in jumping between the two positions. The adjustment parameter is the position of the plane. For each of a succession of values of this parameter the minimum of energy can be determined by the standard methods. The maximum of these minima is then the energy of the saddle point. Similar constraints may be found in problems of crack propagation. Moreover, useful approximations to the saddle point energy which are simpler to calculate may sometimes be seen after such a constraint has been devised. Thus, by holding neighboring atoms in fixed positions, rather than letting them relax at each stage, the work to move the constraining plane may be calculated readily, and provides an upper bound to the saddle point energy.

B. Dynamic calculations

If the displacements in the imbedding medium s_1, s_2, ... s_M are kept fixed, the Newtonian equations of motion for the N atoms in the explicit set can be solved by amplifying the computer programs in familiar ways. This is one possible class of dynamical computations which, subject to the rather restrictive boundary

conditions, show how cracks might advance (or retreat). For short times the simulation should be adequate, but for longer times the rigidity of the boundaries will exert a perturbation of growing significance. A rough lower limit on when the boundary perturbation can be significant is the time of transit of sonic waves across the explicit set. Such problems may be started by assuming a nonequilibrium configuration and releasing it from the rest or with various chosen initial atomic motions. It may also be started by changing the boundary conditions in a programmed way to represent increasing applied stress.

A rigorous accounting of the influence of the surrounding region can be made, in principle, by the method of Green's functions. Because of practical limitations on computer memory the method has not yet been employed in any defect problem to my knowledge, but further consideration should be given to the possibility of using it in problems of crack propagation. Alternatively, the method may suggest other approximations which make more reasonable demands on the computer. For a crack finite in all dimensions the explicit set can be chosen large enough that the displacements at its boundaries remain small, the first order vibration theory may be employed beyond the boundaries. Let the force exerted by the i^{th} atom of the explicit set on the n^{th} atom of the surrounding set be $\vec{f}_{in}(t)$. Define the Green's function for the surrounding set $\overleftrightarrow{G}_{nm}(t)$ which is a second rank tensor, as the displacement of atom m of the surrounding set at time t caused by a unit force applied to atom n of the surrounding set for unit

time interval at time 0. In this statement the term "caused by" is to be understood as the displacement that results when all atoms were at rest in their equilibrium positions at time 0, and the only disturbance is the application of a unit force on atom n at time 0 for time dt, divided by dt. In the crack problem we have in addition the forces $\vec{F}_n(t)$ applied to the outer boundary of the surrounding set (forces supplied by the tensile test machine, or whatever).

The (vector) displacements $\vec{s}_m(t)$ in the surrounding set can now be written

$$\vec{s}_m(t) = \sum_i \sum_n \int_{-\infty}^{t} \vec{f}_{in}(t') \cdot \overleftrightarrow{G}_{nm}(t-t')dt' + \sum_n \int_{-\infty}^{t} \vec{F}_n(t') \cdot \overleftrightarrow{G}_{nm}(t-t')dt' \quad . \qquad (1)$$

Here sums over i run over the atoms of the explicit set, sums over n run over the atoms of the surrounding set. In practice the first term above runs only over near pairs spanning the boundary between the two sets, and the second term runs only over the outer faces to which forces are applied.

In principle, the application of this expression in a machine computation might be as follows: The forces $\vec{f}_{in}(t)$ depend on $\vec{r}_i(t)$ and $\vec{s}_n(t)$ near the boundary of the explicit set. Working in finite differences, and given diaplacements at $t - \Delta t$ and at all previous time steps, the time integrals on the right hand side of (1) can be evaluated as sums which involve the displacement at

all previous time steps. The displacements $\vec{s}_m(t)$ near the boundary of the explicit set can thus be evaluated, which allows the forces $\vec{f}_{in}(t)$ to be evaluated, and which in turn can be used with the equations of motion for the explicit set to update the displacements in the explicit set.

Practical computational difficulties are two-fold. First, the Green's function must be evaluated. This may be done by Fourier transforms, although the computation may be tedious and the resulting function, being dependent on three parameters (n, m, and t) is voluminous, in numerical terms. Second, the displacements near the boundary for all previous times must be stored. This presents a grave and perhaps insuperable demand on the computer memory.

Variations and refinements on Eq. (1) can be made. The surrounding set could be approximated as a continuum, instead of a lattice, with a finite number of degrees of freedom assigned. For the crack problem, a second consideration is that a linear crack produces finite displacements at infinity, violating the assumptions implicit in the foregoing discussion that displacements in the surrounding set need only be treated to first order. The first remedy which suggests itself is to solve the problem with identical geometry in first order continuum elasticity approximation and use the correspondingly distorted lattice as zero order positions for the surrounding set, upon which small motions can be superposed. The small motions could be treated by the Green's function machinery, where the Green's function for

the distorted lattice would be needed. Without having taken a close look at this idea as applied to any particular case, its practicality cannot be asserted.

A second improvement to allow for finite strains at infinity is to define the Green's function with respect to a surrounding set which is on a perfect lattice that is undergoing a steady change of strain (or stress). Thus, suppose a perfect lattice is being homogeneously deformed, starting from the undeformed lattice at time 0 and deforming at a constant rate to correspond to steadily increasing strain applied at the extremeties of a large specimen. Taking an origin at the center of the specimen, the position of the n^{th} atom at time t is

$$\vec{r}_n(t) = \vec{r}_n(0) + \overleftrightarrow{\varepsilon}\cdot\vec{r}_n(0)t \qquad (2)$$

where $\overleftrightarrow{\varepsilon}$ is the strain rate tensor. For to not too large (thus limiting consideration to the first order strains) this displacement set leaves each atom in equilibrium*. Small vibrations of the atoms may be superimposed on these displacements, and the first order equation of motions continue to apply to such vibrations. Thus a Green's function still exists which relates the vibrations to their sources. Equation (1) needs only to be modified by the addition of the term $\overleftrightarrow{\varepsilon}\cdot\vec{r}_n t$ (or $\overleftrightarrow{\varepsilon}\cdot\vec{s}_n t$) to its right hand side.

---

*At large t this expression does not in general leave each atom in equilibrium, because Poisson's ratio may change with strain.

IV. Questions to be Investigated

The role of jogs and kinks in crack propagation should be investigated in much closer detail than has yet been done. If, as seems likely from the evidence in hand, the critical stress for propagation is substantially lower for a jogged crack, the next question to be answered is how jogs are created and renewed. In particular, the intersections of cracks with dislocations* may provide sufficient renewal of jogs in many cases. On the other hand the interaction of stress fields may tend to displace the dislocation ahead of the moving crack, thus preventing intersection. Thermally activated processes may be important, and still other mechanisms can be imagined.

The critical stress for propagation in various crystallographic directions can also be investigated. It is conceivable that some cracks propagating under fully dynamic conditions may prefer crystallographic directions which are not the weakest against quasi static propagation.

The influence of various crystal lattices and the effect of different interatomic interactions have been investigated only in rudimentary ways. The tendency of a crack to seek other crystallographic planes for its propagation is of interest, as well as the possible tendency in any circumstances for a brittle crack to become blunt. The tendency of a crack to be a source

*When a crack cuts through a dislocation, a step will be introduced in the crack edge if the Burger's vector of the dislocation has a component in the plane of the crack and perpendicular to the edge.

for dislocations is of great importance, and may not be understood except with the aid of sophisticated calculations of this type. Indeed the number of mechanistic problems which may be susceptible to investigation by computer methods as outlined above is so vast that the practical question is to select fruitful and reasonably easy questions for early attack.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253, with the University of Michigan.

References

1. P. C. Gehlen and M. F. Kanninen, *Inelastic Behavior of Solids*, M. F. Kanninen et al, Eds., McGraw-Hill, New York, p. 587 (1970)

2. M. I. Kanninen and P. C. Gehlen, *Interatomic Potentials and Simulation of Lattice Defects*, P. C. Gehlen et al, Eds., Plenum Press, New York (1972).

3. R. Chang, Int. J. Fracture Mechanics 6, 111 (1970).

4. W. R. Tyson and L. C. R. Alfred, Presentation at the Corrosion Fatigue Conference, University of Conn., June 14-18, 1971.

5. J. N. Goodier and M. F. Kanninen, Tech. Report 165, Division of Engineering Mechanics, Stanford University (1966).

6. J. N. Goodier, *Treatise on Fracture*, H. Liebowitz, Ed., Vol. 2, p. 1, Academic Press, New York (1968).

7. P. C. Gehlen, private communication.

8. G. H. Vineyard, J. Phys. Chem. Solids 3, 121 (1957).

SIMULTANEOUS DETERMINATION OF LONG-CHAIN BRANCHING AND MOLECULAR WEIGHT DISTRIBUTION IN POLYMERS

J. D. Ferry

Abstract

A proposed procedure is outlined to combine gel permeation chromatography data with linear viscoelastic properties measured in dilute solution and extrapolated to infinite dilution, in order to derive simultaneously the molecular weight distribution and branching index of an unfractionated polymer sample with long-chain branching.

SIMULTANEOUS DETERMINATION OF LONG-CHAIN BRANCHING AND MOLECULAR WEIGHT DISTRIBUTION IN POLYMERS

J. D. Ferry

I. Introduction

The presence of long-chain branching in polymers strongly affects certain of their material properties, including viscosity,[1,2] steady-state compliance,[2] and behavior in rupture.[3] The degree of branching is usually difficult to measure, however, or even to detect if it is small, and quantitative studies of this important feature are sparse partly for this reason. Branching is usually accompanied by enhanced molecular weight distribution, and the two characteristics jointly influence physical properties in a complicated manner.

Proposals have been made recently to combine gel chromatography (GPC) data with intrinsic viscosity[4-6] or sedimentation coefficient in dilute solution[7] to evaluate long-chain branching and molecular weight distribution simultaneously. The primary goal has been a correct evaluation of molecular weight distribution, and in one scheme[6] the branching is not even calculated explicitly.

Another dilute solution property, namely, dynamic viscoelasticity at frequencies below the reciprocal of the terminal relaxation time, has been found to be considerably more sensitive

to long-chain branching than the other hydrodynamic properties mentioned above.[8-11] If the primary interest is in the branching, viscoelastic measurements might be exploited to gauge it, in combination with gel chromatography data. The present report explores this possibility.

II. Method of Miltz and Ram Combining GPC and Intrinsic Viscosity

The proposed method will parallel in some respects one devised by Miltz and Ram[5] which utilizes as input the gel chromatrogram and the intrinsic viscosity of the unknown sample together with the following basic relations:

(a) The GPC calibration for linear samples of the polymer. (This depends on the hydrodynamic volume of a species of molecular weight M_i, proportional to $M_i[\eta]_i$, where $[\eta]_i$ is the intrinsic viscosity of the species.)

(b) The Mark-Houwink equation for the linear polymer, $[\eta]_{il} = KM_i^a$, where K and a are constants for a particular solvent and temperature.

(c) The relation of Zimm and Stockmayer[12] between g_i and m_i, where g_i is the ratio of radii of gyration of a branched molecule to that of a linear molecule of the same molecular weight, and m_i is the number of trifunctional branch branch points on the molecule.

(d) The assumption that $g_i' = g_i^{\frac{1}{2}}$, where g_i' is the ratio of branched to linear intrinsic viscosities for the given molecular weight, viz., $[\eta]_i/[\eta]_{il}$.

(e) The assumption that $m_i = \alpha M_i$ (or, in a modification that we shall not adopt, $m_i = \alpha'(M_i - M_o)$.

(f) Combination of (b), (c), (d), and (e) to give

$$[\eta] = \Sigma\, w_i [\eta]_{il} g_i' \qquad (1)$$

where w_i is the weight fraction of the i^{th} species. The procedure for calculation is as follows:

(1) The chromatogram is compared with the calibration curve assuming that the hydrodynamic volume is $M_i[\eta]_{il} = KM_i^{a+1}$, to obtain a preliminary set of M_i for arbitrarily subdivided w_i.

(2) These M_i and w_i are substituted into Eq. (1), with adjustment of the constant α (see (e) above) to give the correct value of $[\eta]$.

(3) With this value of α, the chromatogram is re-interpreted taking the hydrodynamic volume as $g_i' KM_i^{a+1}$, and a new set of M_i is obtained.

(4) Steps (2) and (3) are repeated until there is no further change.

III. Linear Viscoelastic Properties at Low Frequencies

The dynamic viscoelastic behavior in shear, extrapolated to infinite dilution, is described at low frequencies (i.e., a logarithmic decade or so below the reciprocal of the longest relaxation time) by two coefficients:

$$[A_G] = \lim_{\omega \to 0} [G']/\omega^2 \qquad (2)$$

$$[\eta] = \lim_{\omega \to 0} [G"]/\omega\eta_s \qquad (3)$$

where

$$[G'] = \lim_{c \to 0} G'/c \qquad (4)$$

$$[G"] = \lim_{c \to 0} (G"-\omega\eta_s) \qquad (5)$$

Here ω is the radian frequency, c the polymer concentration (g./cc.), η_s the solvent viscosity, and G' and G" the storage and loss shear moduli. The coefficient $[A_G]$ decreases rapidly with branching; unfortunately, it increases rapidly with broadening molecular weight distribution, but the two effects may be separable by combination with GPC data. In this case, both $[A_G]_i$ and $[\eta]_i$ must be considered for the various species.

IV. Proposed Method Combining GPC and Low-Frequency Viscoelastic Properties

From sinusoidal measurements[12] at sufficiently low frequencies, [G'] and [G"] are obtained by extrapolations to zero concentration, and the coefficients $[A_G]$ and $[\eta]$ are determined. (Of course, $[\eta]$ could alternatively be determined simply from capillary viscometry.) The actual frequency range depends on the solvent viscosity and the polymer molecular weight and branching; the solvent viscosity may be adjusted so that it falls within the frequency limitations of the experimental method.

There is no theoretical relation between $[A]_i/[A]_{il}$ for randomly branched polymers corresponding to the factor g_i' for the analogous intrinsic viscosity ratio. (Here, for convenience, the subscript G is omitted.) However, Osaki[11] has calculated $[A]_i/[A]_{il}$ and $[\eta]_i/[\eta]_{il}$ for comb polymers with certain regular geometries, viz., with arms of equal length equally spaced along the backbone. A specific geometry is characterized by the number of arms, f, and the fraction of mass in the backbone, λ. It is proposed to match a randomly branched molecule of m_i branch points with corresponding comb by identifying m with f and choosing the value of λ_i which makes $[\eta]_i/[\eta]_{il}$ for the two models identical. Then $[A_i]/[A]_{il}$ is calculated for the corresponding comb by the procedure of Osaki.

It may be remarked that the ratios calculated by Osaki's procedure depend on the value chosen for the hydrodynamic interaction parameter h*. Usually measurements will be made in a good solvent with a relatively low value of h*. The Osaki calculation may actually be better for a matched randomly branched molecule than for a real comb, because of problems arising from high segment densities in the interior of the latter.

The input for the analysis then includes that listed under II above and also:

(g) The relation between $[A]_i/[A]_{il}$ and m_i, λ_i obtained as just described.

(h) The value of $[A]_{il}$ calculated as $([\eta]_{il}\eta_s)^2(M_i/RT)(S_2/S_1^2)$, where S_2 and S_1 are numbers of the order of unity which

depend on the value of h* and are available from the Lodge-Wu evaluation[13,14] of the Zimm theory.[15]

(i) The summation relation $[A_G] = \Sigma\, w_i[A]_i$. (2)

The procedure for calculation is as follows:

(1) The chromatogram is compared with the calibration curve assuming that the hydrodynamic volume is $M_i[\eta]_{il} = KM_i^{a+1}$, to obtain a set of M_i for arbitrarily divided w_i.

(2) These M_i and w_i are substituted into Eq. (2), with adjustment of the constant α (see IIe) which determines m_i and therefore $[A]_i/[A]_{il}$, to give the correct value of $[A]_G$.

(3) With this value of α, the chromatogram is re-interpreted taking the hydrodynamic volume as $g_i' KM_i^{a+1}$, and a new set of M_i is obtained.

(4) Steps (2) and (3) are repeated until there is no further change.

(5) As an additional check, $[\eta]$ is calculated from Eq. (1) and compared with the experimental value.

V. Proposed Experimental Tests

Randomly branched samples of polystyrene,[13] polybutadiene,[4] polyethylene,[5] and various copolymers have been described in the literature. The procedure outlined above will be tested on several of these if possible. Depending on the mode of synthesis, the relation appropriate to trifunctional branch points referred to in IIc may be replaced by one for tetrafunctional branch points.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253.

References

1. G. Kraus and J. T. Gruber, J. Polymer Sci., A-2, 3, 681 (1965).
2. T. Fujimoto, T. Narukawa, and M. Nagasawa, Macromol., 3, 57 (1970).
3. J. Meissner, Rheol. Acta, 10, 230 (1971).
4. G. Kraus and C. J. Stacy, J. Polymer Sci., A-2, 10, 657 (1972).
5. J. Miltz and A. Ram, Polymer, 12, 685 (1971).
6. A. Ram and J. Miltz, J. Appl. Polymer Sci., 15, 2639 (1971).
7. L. H. Tung, J. Polymer Sci., A-2, 7, 47 (1969); 9, 759 (1971).
8. K. Osaki, Y. Mitsuda, R. M. Johnson, J. L. Schrag, and J. D. Ferry, Macromol., 5, 17 (1972).
9. Y. Mitsuda, K. Osaki, J. L. Schrag, and J. D. Ferry, Polymer J. (Japan), 4, 00 (1973).
10. K. Osaki and J. L. Schrag, J. Polymer Sci., submitted.
11. K. Osaki, unpublished calculations.
12. J. L. Schrag and R. M. Johnson, Rev. Sci. Inst., 42, 224 (1971).
13. T. Masuda, Y. Nakagawa, Y. Ohta, and S. Onogi, Polymer J. (Japan), 3, 92 (1972).

CURRENT FLUCTUATIONS FROM SMALL REGIONS OF ADSORBATE COVERED FIELD EMITTERS. A METHOD FOR DETERMINING DIFFUSION COEFFICIENTS ON SINGLE CRYSTAL PLANES.

Robert Gomer

Abstract

A novel method is presented for determining surface diffusion coefficients of adsorbates on single crystal planes of field emitters in terms of the time correlation function of current fluctuations. The method is based on the fact that current fluctuations are related to adsorbate density fluctuations, whose time correlation is governed by relaxation times simply related to diffusion coefficients. A general formalism is presented, some idealized cases are worked out analytically, and the case of a circular aperture is treated in detail numerically. Switching between different adsorption states, in addition to but not affected by diffusion, is also included. The forms of the correlation function are shown to be complicated, with a 1/t tail at large t, and cannot be represented by simple exponential decay.

CURRENT FLUCTUATIONS FROM SMALL REGIONS OF ADSORBATE COVERED FIELD EMITTERS. A METHOD FOR DETERMINING DIFFUSION COEFFICIENTS ON SINGLE CRYSTAL PLANES.

Robert Gomer

Introduction

The development of the field emission microscope by E. W. Mueller[1] made possible the direct observation of diffusion processes on clean metal surfaces.[2] In most investigations the emitter was shadowed with adsorbate from one side, and the spreading of this non-uniform layer observed as a function of time, temperature and coverage. While a great deal of semi-quantitative information has been obtained in this way,[2] the method suffers from the defect that diffusion observed on a given crystal plane also depends on transport over adjacent regions, so that the unequivocal correlation between diffusion constants and substrate structure is often difficult. In addition, visual observations are possible only on highly emitting regions. To some extent this can be overcome by the use of a small probe-hole in the screen of the tube, through which a small portion of the electron beam is allowed to pass. In this way emission from very small regions ($\gtrsim$ 10 Å in linear dimension) even of high work function, planes can be measured. The arrival of adsorbate is signalled by changes in current.[3]

Nevertheless, even this refinement does not resolve the uncertainties pointed out above. For adsorbates visible in the field ion microscope direct observation on individual planes is possible, as shown by the experiments of Ehrlich.[4] Unfortunately, even the introduction of low field image gases does not extend the class of visible adsorbates beyond metallic ones,[5] although it does of course permit the observation of metal atoms with lower ionization potentials. The reason for this is almost certainly the absence of appreciable local density of states just above the Fermi level for most covalently bonded adsorbates. For such adsorbates one is therefore restricted to field emission. During experiments on field emission from single planes carried out by T. Engel and the author, we noticed that heating of emitters covered with equilibrated CO layers led to fluctuations in probe hole current. It seemed plausible that these were caused by concentration fluctuations over the small region probed. Since such concentration fluctuations build up and decay by surface diffusion, their time correlation function must contain the local surface diffusion coefficient D through some characteristic decay time $\tau_0 \sim r_0^2/4D$, where r_0 corresponds to the dimension of the region probed. An explicit expression for the correlation function of current fluctuations thus provides a means of determining diffusion coefficients on single crystal planes.

The fact that current fluctuations in field emission may be related to surface diffusion has also been noted by Kleint,[6] who attempted to adapt a model by Bess[7] for 1/f noise in semi-

conductors. This model treats switching between two adsorption states, mediated by surface diffusion. It does not consider at all the build-up and decay of concentration fluctuations over small regions, which is central to the present work, and thus addresses itself to quite a different situation both experimentally and theoretically. In the following, the correlation function for current fluctuations from small regions is derived on the basis of diffusion limited concentration fluctuations.

Relation between Emission Current and Adsorbate Density

In the absence of adsorbate the current density of field emission j is given by the well-known Fowler-Nordheim relation, which we may write as

$$\ell nj = \ell nB - (6.8\times10^{7}\ v/F)\phi_o^{3/2} \qquad (1)$$

where B is a field-insensitive term, F the field in volts/cm, ϕ_o the work function of the clean surface in eV, and v an image correction term.[1] The presence of an adsorbate modifies emission by changing ϕ and in some cases B. The change in ϕ is due to the dipole moment associated with each ad-particle. This consists of a zero field component and one induced by the applied field. Since the latter is proportional to applied field, a simple analysis[8] shows that the work function change resulting from polarization of the adsorbate by the applied field appears as a field-independent term, i.e., seems to modify B. For present purposes, however, we are not interested in the variation of

current with field, and the separation into intrinsic and induced dipoles need not be made. While the apparent changes in B observed in plots of $\ell ni/F^2$ vs. 1/F can largely be explained by polarization, it may also happen that an adsorbate produces anti-resonances near the Fermi level[9] (in slightly different language, a marked decrease in local density of states at the Fermi level[10]). Since in most cases virtually all emission comes from within 0.5 eV of the Fermi level, electronic effects outside this energy range cannot affect total current markedly. This phenomenon, first pointed out in connection with field emission by Duke and Alferieff,[9] can lead to a real change in B. Experimentally it is usually found that log B is proportional to coverage, so that we may write quite generally for the total current i from a region area A

$$\ell ni = \ell nAB_o + c_1 c - (6.8\times10^7/F)\phi^{3/2} v \qquad (2)$$

where c_1 is a constant (positive, negative or 0), c the density of adparticles in the region and ϕ the effective work function in A at field F and for a given ad-particle distribution. Writing

$$\phi = \bar{\phi} + \delta\phi \qquad (3)$$

where $\bar{\phi}$ is the time-averaged work function and $\delta\phi$ its (small) fluctuation, we see that

$$\phi^{3/2} \simeq (\bar{\phi})^{3/2}(1 + \frac{3}{2}\,\delta\phi/\bar{\phi}) = (\bar{\phi})^{3/2} + \frac{3}{2}\sqrt{\bar{\phi}}\,\delta\phi \qquad (4)$$

so that

$$\delta \ell ni \equiv \ell ni - \overline{\ell ni} = c_1' \delta n + c_2 \delta \phi \tag{5}$$

where $\delta n = n - \bar{n}$, $c_2 = -1.02\ 10^8 v\bar{\phi}^{\frac{1}{2}}/F$, and $\overline{\ell ni}$ the log of the current when $n = \bar{n}$ and $\phi = \phi$, and $c_1' = c_1/A$.

To complete the relation between current and adsorbate density we must express ϕ, or alternately $\bar{\phi}$ and $\delta\phi$, in terms of n. It will be useful to work in terms of the work function increment caused by adsorption

$$\Delta\phi = \phi - \phi_o \tag{6}$$

For sufficiently large areas and for uniform adsorbate distributions

$$\Delta\phi = 2\pi Pn/A \tag{7}$$

where P is the dipole moment of the ad-particles, defined in such a way that the plane of zero potential bisects it. Thus the contribution per ad-particle, W becomes, in this limit, $2\pi P/A$, and we may think of this as the average over A of the potential contributed by an adparticle. This suggests that we look for $\Delta\phi$ as the sum over all particles i in the system (not merely in A) of the quantity

$$W_i(\vec{x},d) = \frac{1}{A}\int_A \frac{P_1 d\ dx'}{[(\vec{x}-\vec{x}')^2+d^2]^{3/2}} \tag{8}$$

where d is the distance above the surface where the potential

is to be evaluated and $\vec{x}$ denotes the position of the dipole (within or without A) with respect to an origin, conveniently chosen at the center of the probed area A. It is easily verified that for a circular area of radius r_o and a dipole placed at its center

$$W(0,d) = \frac{2\pi Pd}{\pi r_o^2}\left(\frac{1}{d} - \frac{1}{\sqrt{r_o^2+d^2}}\right) \tag{9}$$

which, for $d \leq r_o$ approaches the limit $2\pi P/A$, and for $r_o \to 0$ the limit P/d^2. In field emission (unlike thermionic emission) the potential at $d \leq 10$ Å from the surface determines current, since the barrier width is $\sim$10 Å at the Fermi level, from whose vicinity most of the emission comes. Thus, it suffices to calculate W for d $\sim$5 Å.

Equation (8) can be integrated analytically for an arbitrary location of the dipole for the case of a square probed region of length 2a. The result is

$$W(x,y) = \frac{P}{4a^2}\left[\tan^{-1}\left(\frac{(1-x)(1-y)}{z\sqrt{(1-y)^2+(1-x)^2+z^2}}\right)+\tan^{-1}\left(\frac{(1-x)(1+y)}{z\sqrt{(1+y)^2+(1-x)^2+z^2}}\right)\right.$$
$$\left.+\tan^{-1}\left(\frac{(1+x)(1-y)}{z\sqrt{(1-y)^2+(1+x)^2+z^2}}\right)+\tan^{-1}\left(\frac{(1+x)(1+y)}{z\sqrt{(1+x)^2+(1+y)^2+z^2}}\right)\right] \tag{10}$$

where all quantities are expressed in units of a and z = d/a. It can readily be shown that, for a dipole at (0,0), the same limits result as in the circular case. Thus, the approximation that a dipole makes its full contribution $2\pi P/A$ to the work

function when it is within A and zero otherwise is a reasonably good one for $z \ll 1$.

General Expression for the Correlation Function of Current Fluctuations

We can now express $\delta \ell ni \equiv \ell ni - \overline{\ell ni}$ in terms of sums over the particle positions in the entire system by combining Eqs. (5) and (8):

$$\delta \ell ni = c_1' \left(\sum_i \delta(\vec{x}_i) - \overline{\sum_i \delta(\vec{x}_i)} \right) + c_2 \left(\sum_i W(\vec{x}_i) - \overline{\sum_i W(\vec{x}_i)} \right) \tag{11}$$

The symbol $\delta(\vec{x}_i)$ has the value unity when the coordinates $\vec{x}_i$ of particle i lie in A and zero otherwise, and the bars denote time (or ensemble) averages. It follows that the time correlation function $f(\tau)$ of the quantity $\delta \ell ni$ is given by

$$f(\tau) \equiv \overline{\delta \ell ni(0)\, \delta \ell ni(\tau)} = \overline{[c_1' \sum_i \delta_i(0) + c_2 \sum_i W_i(0)][c_1' \sum_j \delta_j(\tau) + c_2 \sum_j W_j(\tau)]}$$

$$- \overline{[\sum_i (c_1' \delta_i + c_2 W_i)]^2} \tag{12}$$

with $W_i(\tau) \equiv W(\vec{x}_i, \tau)$, etc., since

$$\overline{[\sum_i (c_1' \delta_i(0) + c_2 W_i(0))]} \; \overline{[\sum_j (c_1' \delta_j(\tau) + c_2 W_j(\tau))]} = \overline{[\sum_i (c_1' \delta_i + c_2 W_i)]^2} \tag{13}$$

etc. The last term in Eq. (12) is the square of the quantity

$$\sum_i (c_1' \delta_i + c_2 W_i) = N \int_{A_{tot}} \frac{d\vec{x}}{A_{tot}} (c_1' \delta(\vec{x}) + c_2 W(\vec{x}))$$

$$= N \left(\frac{A}{A_{tot}} c_1' + c_2 \bar{W} \right)$$

$$= c_1'\bar{n} + c_2\overline{\Delta\phi} \tag{14}$$

where N is the total number of ad-particles in the entire system, A_{tot} the total area,

We next separate the first term in Eq. (12) into a sum over i = j, and a sum over i ≠ j:

$$\overline{\sum_i [c_1'\delta_i(0)+c_2W_i(0)][c_1'\delta_i(\tau)+c_2W_i(\tau)]}$$

$$+ \overline{\sum_{i\neq j} [c_1'\delta_i(0)+c_2W_i(0)][c_1'\delta_j(\tau)+c_2W_j(\tau)]}$$

We now assume that the displacements of different particles are not time-correlated at all, so that the ensemble average of the double sum becomes

$$(N^2-N)\left[\int_{A_{tot}} \frac{d\vec{x}}{A_{tot}} (c_1'\delta(\vec{x})+c_2W(\vec{x}))\right]^2$$

Since N is very large, $N^2-N \simeq N^2$, and thus the double sum is cancelled by the term evaluated in Eq. (14). Hence,

$$f(\tau) = \overline{\sum_i (c_1'\delta_i(0)+c_2W_i(0))(c_1'\delta_i(\tau)+c_2W_i(\tau))} \tag{15}$$

It might be thought at first that the neglected term

$$N\left[\int_{A_{tot}} \frac{d\vec{x}}{A_{tot}} (c_1'\delta+c_2W)\right]^2$$

is of some significance. However, Eq. (14) shows that it is of the form $\frac{1}{N}(c_1'\bar{n} + c_2\overline{\Delta\phi})^2$. We shall see shortly that it is of

order A/A_{tot}, relative to the term retained in Eq. (15).

The evaluation of the correlation function now becomes simple in principle. We write

$$f(\tau) = N\int_{A_{tot}} \frac{d\vec{x}}{A_{tot}}(c_1'\delta(\vec{x})+c_2W(\vec{x}))\,\mathcal{P}(x|x',\tau)d\vec{x}'\,(c_1'\delta(\vec{x}')+c_2W(\vec{x}')) \quad (16)$$

Here $d\vec{x}/A_{tot}$ is the absolute probability that a given particle is in $d\vec{x}$ at $t = 0$, and

$$\mathcal{P}(\vec{x}|\vec{x}',\tau)d\vec{x}' = \frac{d\vec{x}'}{4\pi D\tau}\exp - (|\vec{x}-\vec{x}'|^2/4D\tau) \quad (17)$$

the conditional (normalized) probability that, if it was in $d\vec{x}$ at $t = 0$, it will be in $d\vec{x}'$ at $t = \tau$. Expanding Eq. (16), we have

$$f(\tau) = \bar{n}[(\frac{c_1}{A})^2 g_1(\tau)+(\frac{2\pi Pc_2}{A})^2 g_2(\tau)+2(\frac{c_1c_2 2\pi P}{A_2})g_3(\tau)] \quad (18)$$

with

$$g_1(\tau) = \frac{1}{A}\int_A \frac{d\vec{x}\,d\vec{x}'}{4\pi D\tau}\,e^{-|\vec{x}-\vec{x}'|^2/4D\tau} \quad (19a)$$

$$g_2(\tau) = \frac{1}{A}\int_{A_{tot}} \frac{W'(\vec{x})W'(\vec{x}')}{4\pi D\tau}\,e^{-|\vec{x}-\vec{x}'|^2/4D\tau}\,d\vec{x}\,d\vec{x}' \quad (19b)$$

$$g_3(\tau) = \frac{1}{A}\int_{A_{tot}} \frac{W'(\vec{x})}{4\pi D\tau}\,e^{-|\vec{x}-\vec{x}'|^2/4D\tau}\,\delta(\vec{x}')d\vec{x}\,d\vec{x}' \quad (19c)$$

where we have written $W' = W/(2\pi P/A)$. As it stands, Eq. (16) can only be integrated numerically. To show the structure of $f(\tau)$ we now make the assumption that $W' = \delta$, i.e., is unity in A and zero outside, so that $f(\tau)$ reduces to

$$f(\tau) = \bar{c}\left(\frac{c_1+2\pi P c_2}{A}\right)^2 \int_{A_t} \frac{d\vec{x}\, d\vec{x}'}{4\pi D\tau}\, e^{-|\vec{x}-\vec{x}'|/4D\tau} \tag{20}$$

The quadratures in Eq. (20) can be carried out for a square area, $A = 4a^2$ and yield, for that case

$$f(\tau) = \bar{n}\left(\frac{c_1+2\pi P c_2}{A}\right)^2 g(\tau) \tag{21}$$

with

$$g(\tau) = \left[\mathrm{erf}(\tau_0/\tau)^{\frac{1}{2}} + \sqrt{\frac{\tau}{\pi\tau_0}}\,(e^{-\tau_0/\tau}-1)\right]^2 \tag{22}$$

$\bar{n}$ is the average number of ad-particles in A and

$$\tau_0 = a^2/D \tag{23}$$

We may write equivalently

$$f(\tau) = \frac{(\bar{c}c_1+c_2\overline{\Delta\phi})^2}{\bar{n}}\, g(\tau) \tag{24}$$

For $\tau/\tau_0 \ll 1$ $g(\tau)$ reduces to

$$g(\tau) \simeq (1 - \pi^{-\frac{1}{2}}(\tau/\tau_0)^{\frac{1}{2}})^2 \tag{25a}$$

and for $\tau/\tau_0 \gg 1$

$$g(\tau) \simeq (\tau/\pi\tau_0)(e^{-\tau_0/\tau}-1)^2 \tag{25b}$$

so that $f(\tau)$ decays as $1/\tau$ at large τ.

Equation (22) or (25a) shows that

$$f(0) = \bar{n}\left(\frac{c_1+c_2 2\pi P}{A}\right)^2 = (\bar{c}\, c_1+c_2\overline{\Delta\phi})^2/\bar{c}A \tag{26}$$

which corresponds to the mean square "strength" of a fluctuation, and is proportional to $\bar{n}$ as it should be.

We are now able to verify that the term neglected in our derivation, $(1/\bar{c}\ A_{tot})(c_1' \bar{n} + c_2 \overline{\Delta\phi})^2$, is of order A/A_{tot} relative to $f(o)$ by comparison with Eq. (26).

$f(\tau)$ from Decay of Density Fluctuations

It is interesting that Eqs. (20) and (22) can also be derived from a slightly different viewpoint, by using Onsager's hypothesis that fluctuations decay according to macroscopic laws, i.e., by studying the decay of a fluctuation according to a macroscopic diffusion equation. We can do this by postulating a delta function fluctuation at $(\vec{x},0)$, with $\vec{x}$ somewhere in A. At time τ the concentration profile will be

$$c(\vec{x}',\tau) = \frac{e^{-|\vec{x}-\vec{x}'|/4D\tau}}{4\pi D\tau} \qquad (27)$$

and the total fraction of the original strength of the fluctuation left in A after time τ is found by integrating with respect to $\vec{x}'$ over A. To get the ensemble average we must now average over the initial position $\vec{x}$ as well and thus obtain, using the mean square strength of fluctuations, Eq. (20) and the subsequent development. The equivalence of these two methods is no accident of course, but stems from the relation between D and the random walk assumption inherent in Eq. (17). Finally, it should be pointed out that the averaging over the initial positions of delta functions, followed (or preceded) by integration of the concentration profiles over A,

is equivalent to postulating a uniform, step-like concentration fluctuation over A and following its time development. (The arbitrariness of this fluctuation profile comes from the weighting function $W' = \delta$; a more realistic weighting function would lead to a less abrupt profile). It is possible to treat this diffusion problem in the standard way and to obtain solutions of the form

$$c(r,\tau) = \sum_j A_j J_0(a_j r)\, e^{-a_j^2 D\tau} \tag{28}$$

(for circular probe areas), where the J_0's are Bessel functions of order 0, with roots a_j; chosen to make $J_0(a_j r_0) = 0$. The A_j's are the appropriate coefficients for the expansion of $c(r,0)$ in the orthonormalized set of Bessel functions. The roots are given approximately by

$$a_j \simeq \pi(j-\tfrac{1}{4})/r_0 \tag{29}$$

so that each term in Eq. (28) is multiplied by a time factor of the form $e^{-\tau/\tau_j}$, with $\tau_j = r_0^2/[\pi^2(j-\frac{1}{4})^2 D]$. Integration of $c(r,\tau)$ over A would then give $g(\tau)$. The point to note is that only for very special assumptions of the initial fluctuation profile, namely $c(r,0)$ equal to the lowest order Bessel function, would $g(\tau)$ show an exponential decay. The statistically correct assumption about the fluctuation leads to the more complicated relation Eq. (22); this can be expressed, however, in the form of Eq. (28), or more correctly as an integral over A of Eq. (28) (which leaves the τ dependence intact), i.e., as a series of

different exponential decay terms. It is interesting that, even for $\tau_o/\tau \ll 1$, Eq. (25b) does not go over into a single exponential decay term, as Eq. (28) might suggest. The reason is that the random positions of the individual fluctuations over A assign too small a weight to the A_u term to ever let it predominate, despite the fact that it decays more slowly than any other term. As we shall see, the fact that Eqs. (21)-(25) have been derived for a square probe area does not alter these conclusions, since the long time limit of $g(\tau)$ for a circular aperture will be shown to be of the same form as Eq. (25b).

Circular Aperture

The preceding discussion has concentrated on a square aperture and an assumed step function form for W, to show analytically the salient features of $f(\tau)$. We treat here the general case of a circular aperture, starting with $W(\vec{x})$. Changing to polar coordinates gives

$$W'(\rho) = \frac{W(\rho)}{(2\pi P/\pi r_o^2)} = \frac{z}{2\pi} \int_0^{2\pi} \int_0^1 \frac{\rho' d\rho' d\theta}{(\rho^2+\rho'^2+z^2-2\rho\rho'\cos\theta)^{3/2}} \qquad (30)$$

where $\rho = r/r_o$, $z = d/r_o$, with r_o the radius of the probed region. Integration over ρ' and partial integration over θ gives

$$W'(\rho) = 1 - \frac{z}{\pi} \int_0^{\pi} \frac{(\rho^2+z^2-\rho\cos\theta)d\theta}{(\rho^2\sin^2\theta+z^2)(\rho^2+z^2+1-2\rho\cos\theta)^{\frac{1}{2}}} \qquad (31)$$

These integrals were evaluated for a number of values of z on an IBM 360 computer. The results, as well as graphs of $W'(\rho)\rho$ are

shown in Figs. 1 and 2. It is seen that for small z $W'(\rho)$ approaches a step function $\delta(\rho)$ 1 for $\rho \leq 1$, 0 for $\rho > 1$.

The integrals appearing in Eq. (19) expressed in polar coordinates are all of the form

$$g_k(\tau) = (\tau_o/\tau)\int d\rho d\rho' d\theta d\theta' R(\rho)R'(\rho')\rho\rho' e^{-(\rho^2+\rho'^2-2\rho\rho'\cos(\theta-\theta'))(\tau_o/\tau)} \tag{32}$$

where

$$\tau_o = r_o^2/4D \tag{33}$$

and $\rho = \tau/\tau_o$, $\rho' = r'/r_o$, and R and R' are either W' or δ. The angular dependence appears only as the difference of θ and θ', and thus can depend only on $\psi = \theta-\theta'$. Thus we may use the equivalent angle variables $\theta-\theta'$ and $\theta+\theta'$, and integrate at once over the latter obtaining a factor of 2π. The integral over ψ is of the form

$$2\int_0^\pi e^{2\rho\rho'(\tau_o/\tau)\cos\psi} d\psi = 2\pi J_o(i\rho\rho'(\tau_o/\tau)) = 2\pi I_o(2\rho\rho'(\tau_o/\tau)) \tag{34}$$

where I_o is the Bessel function of order zero and imaginary argument ix, given by

$$J_o(ix) \equiv I_o(x) = \sum_{m=o}^{\infty} \frac{(x/2)^{2m}}{(m!)^2} \tag{35}$$

Thus the integrals in Eq. (19) reduce to double integrals over ρ and ρ':

$$g_1(\tau) = 4(\tau_o/\tau)\int_0^1\int_0^1 I_o(2\rho\rho'\tau_o/\tau)e^{-(\rho^2+\rho'^2)\tau_o/\tau}\,\rho\rho'd\rho d\rho' \qquad (36a)$$

$$g_2(\tau) = 4(\tau_o/\tau)\int_0^\infty\int_0^\infty W'(\rho)W'(\rho')I_o(2\rho\rho'\tau_o/\tau)e^{-(\rho^2+\rho'^2)\tau_o/\tau}\,\rho\rho'd\rho d\rho' \qquad (36b)$$

$$g_3(\tau) = 4(\tau_o/\tau)\int_{\rho=o}^\infty\int_{\rho'=o}^1 W'(\rho)I_o(2\rho\rho'\tau_o/\tau)e^{-(\rho^2+\rho'^2)\tau_o/\tau}\,\rho\rho'd\rho d\rho' \qquad (36c)$$

For large τ, $I_o \simeq 1$, and with the further assumption that $W'(\rho) = W'(\rho') = \delta$, we immediately obtain

$$g(\tau) = (\tau/\tau_o)(1-e^{-\tau_o/\tau})^2 \quad ; \quad \tau_o/\tau << 1 \qquad (37)$$

which is of the same form as Eq. (25b) for a square aperture. It is interesting that in terms of D we thus obtain (using Eqs. (23) and (33))

$$\begin{aligned} g(\tau) &\to (a^2/\pi D)\tau^{-1} \qquad \text{square of half-width } \underline{a} \\ g(\tau) &\to (r_o^2/4D)\tau^{-1} \qquad \text{circle of radius } r_o \end{aligned} \qquad (38a)$$

as $\tau \to \infty$.

The integrals in Eq. (36) were evaluated numerically on an IBM 360 computer for various values of z and $\tau = \tau/\tau_o$. In those integrations with infinite upper limits, approximate upper limits ρ_u were chosen as follows: The average work function increment $\overline{\Delta\phi}$ can be expressed as an integral of W(r):

$$\overline{\Delta\phi} = 2\pi P\bar{c} = 2\pi\int_0^\infty \bar{c}W(r)rdr \qquad (39)$$

so that

$$2\int_0^\infty W'(\rho)\rho d\rho = 1 \tag{40}$$

The upper limits ρ_u were chosen so that

$$1 - 2\int_0^{\rho_u} W'(\rho)\rho d\rho \leq 0.01 \tag{41}$$

Since it is clear from Eq. (8) that for large ρ, $W(\rho) \propto \rho^{-3}$, it can readily be shown that condition (41) is equivalent to

$$2\rho_u^2 W'(\rho_u) \leq 0.01 \tag{42}$$

The following values of ρ_u were used in accordance with this criterion: For $z = 1.0$, $\rho_u = 52.2$; for $z = 0.5$, $\rho_u = 25.3$; for $z = 0.2$, $\rho_u = 11.4$; for $z = 0.1$, $\rho_u = 7.5$.

In order to evaluate the integrals at $t = 0$, it is simplest to start with the definition of $f(0)$. (We show explicitly only the case for $c_1 = 0$. The extension to finite c_1 is obvious.)

$$f(0) = c_2^2 \overline{\sum_i W(\vec{r}_i,0)W(\vec{r}_i,0)}$$

$$= \bar{c}\, c_2^2 \left(\frac{2\pi P}{A}\right)^2 \int W'(r)W'(r')rdrd\theta d\theta' \delta(\theta-\theta')\delta(r-r')dr' \tag{43}$$

$$= \bar{n}\left(\frac{2\pi P c_2}{A}\right)^2 2\int_0 (W'(\rho))^2 \rho d\rho$$

so that

$$g_2(0) = 2\int_0^\infty (W'(\rho))^2 \rho d\rho \tag{44}$$

For $z \neq 0$ both $g_2(0)$ and $g_3(0)$ are less than unity, and $g_2(0) < g_3(0)$. The physical reason for these facts is the following. It will be shown in a later section that $f(0)$ is approximately the relative mean square current fluctuation. For $z > 0$, dipoles beyond the probed area contribute to emission in it, thus effectively increasing its size and consequently decreasing the relative fluctuations. Since g_3 represents a cross term between g_1 and g_2, it represents a smaller effective area than g_2 and hence a larger fluctuation.

The results of the computer calculations are shown in Figs. 3 and 4 respectively as functions of $t = \tau/\tau_o$ for various values of z. The case $z = 0$ corresponds to $W' = \delta$, and applies when $c_2\ 2\pi\ P << c_1$. The other curves in Fig. 3 apply directly for the appropriate z when $c_2\ 2\pi\ P >> c_1$. For intermediate cases c_1 and P would have to be known separately in order to combine the appropriate curves of Figs. 3 and 4. In general, this may be rather difficult but, as pointed out, c_1 is in fact generally small. At large t, $g_2(t)$ and $g_3(t) \to 1/t$. It is easy to see from Eqs. (36) and (40) that this must be the case. At small t the curves do <u>not</u> behave like e^{-t}, except over very narrow ranges.

The dependence of g_2 and g_3 on z requires some discussion. For a fixed τ_o, i.e., for a given r_o, g_2 and g_3 vary most rapidly

with time as z decreases, as we should expect. It must be remembered, however, that d is fixed at ∿5A, so that $z \propto 1/r_o$, and hence $z^2 \propto 1/\tau_o$. Thus, a better idea of the real time behavior of $f(\tau)$ can be gained by plotting $g_2(t)$ vs. $t/z^2 = \frac{4D}{25}\tau$. Since an experiment would look at the relative change in $f(\tau)$ as a function of τ, we show $g_2(t)/g_2(0)$ and $g_3(t)/g_3(0)$ vs. $\log_{10}(t/z^2)$ in Figs. 5 and 6.

This mode of presenting the results shows two interesting features. First, as τ_o decreases, so does the real time required for the decay of the correlation function, as we should expect. Second, and more important, the shape of the normalized curves for different z is almost identical. This is true both for g_2 and g_3 and for the comparison of g_2 with g_3. The only major differences are displacements along the log t/z^2 axis. It now becomes an obvious extension to compare the curves with the corresponding z = 0 case. We may do this, for instance, by comparing t for $g_2(z,t)/g_2(z,0) = .5$ with t for $g_1(t) = .5$. These ratios can be thought of as increases in the effective values of r_o^2. We show in Fig. 7 the factors β_k (k = 2 or 3) defined by

$$\beta = r_o(\text{effective})/r_o = (t(g)/t(g_1))^{\frac{1}{2}} \tag{46}$$

Both β_2 and β_3 vary linearly with increasing z, i.e., decreasing r_o. As expected, $\beta_3 < \beta_2$ because one limit of integration for g_3 extends only to $\rho' = 1$.

Thus we arrive at the intuitively satisfying result that the smearing out of $W'(\rho)$ beyond $\rho = 1$ does not change the form of the normalized correlation functions, but leads to a correction in τ_0 which can be expressed as an increase in effective r_0. For cases in which $c_1 \neq 0$, the admixture of g_1 and g_3 to g_2 will have the effect of reducing the overall β. Since the curves are all of comparable shape, it would take detailed knowledge of all the parameters involved to decide on the exact contributions of each g. A hypothetical mixture is shown in Fig. 5. In practice, the importance of these considerations can be minimized by choosing a reasonably large r_0 and hence small z, for which $\beta - 1$ is small in any case.

Relation of r_0 to Single Plane Parameters

The last section has touched on the effective radius within which time correlated events contribute to the measured correlation function. For the idealized, infinite plane case this is of academic or computational interest only. For a real field emitter of radius 1000-5000 Å planes of given crystallographic orientation are of 50-200 Å in radius, and it is therefore important to see if the effective r_0 can be made smaller than these dimensions, since we wish to measure D for single planes. It is clear from Fig. 7 that this is so. Another criterion, somewhat more arbitrary, would be to pick the value of ρ, at which $2\int_0^{\rho} \rho' W'(\rho') d\rho'$ had reached some fixed value, say 0.80 or 0.90. It should be pointed out, however, that the curvature of

the emitter imposes an absolute limit on the effective size of the probed region, since the horizon distance r_m for a point d Angstrom above the surface of a sphere or radius r_t is

$$r_m = \sqrt{2r_t d} \tag{47}$$

For d = 5 Å and r_t = 1000 Å, r_m = 100 Å. It may now be asked to what extent the curvature of the surface affects W' and the subsequent calculations, which were carried out for a plane surface. For all cases considered the effect is very small, and hence we may assume the validity of the calculations for the curved emitter with confidence.

Extension to Interactions between Ad-particles

We have assumed so far that there is no interaction between diffusing particles. In many systems, however, this is unlikely to be true. A detailed calculation taking interaction into account would be very difficult. However, it is known that even for biased random walks the form of Eq. (17) is not changed, although the value of D will be affected. Thus our results will most probably remain valid, with D a function of average concentration.

Spectral Density Representation

The results obtained here have been presented as correlation functions. It would be possible, of course to represent them as spectral densities by taking their Fourier transforms. Since experimental correlation functions can be obtained directly, this step is omitted here. It is worth pointing out, however, that the

power spectrum will not be fitted with any accuracy by functions of the form

$$S(\omega) = \frac{\tau_o^{-1}}{\omega^2+\tau_o^{-2}} \tag{48}$$

because the correlation function is so poorly approximated by an exponential.

Inclusions of Flip-Flop Processes

For certain adsorbates, for instance CO on bcc transition metals, different adsorption states characterized by different dipole moments can exist on a given crystal plane.[11] It is possible that under rather special conditions a switching back and forth could occur, although in most cases the change from a given ad-state to another seems irreversible. It is fairly easy, with some restrictions, to extend the present calculation to include flip-flop. The treatment we shall use is a simple adaptation of one worked out by Machlup[12] to treat current noise in semiconductors.

Let state A be characterized by a dipole moment P_A and let its mean life with respect to flipping to state B be τ_A. Let state B be similarly characterized by P_B and τ_B. Let $f_A(\tau)$ and $f_B(\tau)$ be the probabilities that an ad-particle is in state A and B respectively. Then

$$\Delta\phi = \left(\frac{2\pi}{A}\right) \sum_{i=1}^{N} W'(\vec{x}_i)\,(f_A P_A + f_B P_B) \tag{49}$$

and

$$\overline{\Delta\phi(0)\Delta\phi(\tau)} = (2\pi/A)^2 \sum_{i,j} W'(x_i,0)W'(x_j,\tau)\overline{(f_{iA}(0)P_A+f_{iB}(0)P_B)}$$

$$\overline{(f_{jA}(\tau)P_A+f_{jB}(\tau)P_B)} \qquad (50)$$

To proceed further we must now make two key assumptions: 1) we assume that the flipping of different ad-particles is wholly uncorrelated, and 2) we assume that an average diffusion coefficient $\bar{D}$ can be defined as

$$\bar{D} = D_A\bar{f}_A + D_B\bar{f}_B \qquad (51)$$

The latter assumption is certainly valid when flipping times are short relative to diffusion times. When the opposite is true, i.e., if there are two populations each with its own diffusion coefficient but with no switching from state A to B, we are in fact dealing with a different problem whose solution will be indicated also. It should also be pointed out that our model assumes that the switching times have nothing to do with diffusion. Kleint's and Bess'[7] model, on the other hand, assumes that switching times are <u>determined</u> by diffusion, since an ad-particle in state A can find a B site only by migrating there.

To return to our model, we have by analogy with our previous results

$$\delta\phi\delta\phi(\tau) = [(2\pi/A)^2 \bar{c}\int \frac{W'(\vec{x})W'(\vec{x}')}{4\pi\,\bar{D}\tau} e^{-|\vec{x}-\vec{x}'|^2/4\bar{D}\tau} d\vec{x}\, d\vec{x}']$$

$$\times \overline{(f_A(0)P_A+f_B(0)P_B)(f_A(\tau)P_A+f_B(\tau)P_B)} \qquad (52)$$

The term in brackets in Eq. (52), involving f_A and f_B, is independent of the integral over coordinates and thus appears as a modulating factor of the previous results. To evaluate it we note that

$$\bar{f}_A = \tau_A/(\tau_A+\tau_B) \tag{53a}$$

$$\bar{f}_B = \tau_B/(\tau_A+\tau_B) \tag{53b}$$

and that

$$\bar{f}_{AA} \equiv \overline{f_A(0)f_A(\tau)} = \bar{f}_A S_{AA}(\tau) \tag{54a}$$

$$\bar{f}_{AB} \equiv \overline{f_A(0)f_B(\tau)} = \bar{f}_A S_{AB}(\tau) \tag{54b}$$

where $S_{AA}(\tau)$ is the probability of an even number of flips in time τ for a particle starting in state A at $t = 0$, and so on. It is shown by Machlup[12] that

$$\bar{f}_{AA} = (\bar{f}_A)^2 + \bar{f}_A\bar{f}_B\, e^{-\tau/\tau_1} \tag{55a}$$

where $1/\tau_1 = \frac{1}{\tau_A} + \frac{1}{\tau_B}$. It is easy to show that

$$\bar{f}_{BB} = (\bar{f}_B)^2 + \bar{f}_A\bar{f}_B\, e^{-\tau/\tau_1} \tag{55b}$$

$$\bar{f}_{AB} = \bar{f}_{BA} = \bar{f}_A\bar{f}_B(1-e^{-\tau/\tau_1}) \tag{55c}$$

With these results we obtain for $f(\tau)$, assuming $c_1(A) = c_1(B) = 0$

$$f(\tau) = \bar{n}\left(\frac{2\pi c_2}{A}\right)^2 g_2(\tau)\left[(\bar{f}_A P_A+\bar{f}_B P_B)^2 + e^{-\tau/\tau_1}\bar{f}_A\bar{f}_B(P_A-P_B)^2\right] \tag{56}$$

with $\bar{f}_A$, $\bar{f}_B$ given by Eqs. (53a) and (53b) and $g_2(\tau)$ by Eq. (19b), with $\tau_0 = r_0^2/4D$ and $c_1 = 0$. Equation (56) has the form of the previous result, with an average dipole moment, $\bar{P} = (\bar{f}_A P_A + \bar{f}_B P_B)$ substituted for P, plus a term which multiplies $g(\tau)$ by a decaying exponential. If $P_A = P_B$, or if $\tau_1 << \tau$, the previous results apply with P replaced by $\bar{P}$, i.e., the exponential term drops out. The explanation when $P_A = P_B$ is obvious. The situation for $\tau_1 << \tau$ can be thought of in two ways. If the emphasis is on diffusion, i.e., if $\tau \sim \tau_0$, it simply means that flip-flop is so rapid relative to diffusion that all flip-flop correlations have washed out in times of order τ_0, so that the mean fractional populations of A and B are maintained. On the other hand, $\tau_1 << \tau_0$ also means that, for $\tau \sim \tau_1$, $g(\tau)$ remains constant, so that pure flip-flop dominates the correlation function for $\tau << \tau_0$. This can be expressed by formally excluding diffusion; in that case $g(\tau)$ must be replaced by $2\int_0^\infty (W'(\rho))^2 \rho d\rho$ in Eq. (56), with the result that the correlation function then consists of a constant term plus a decaying exponential.

For $\tau_1 >> \tau \sim \tau_0$ (but $\tau_A \sim \tau_B$) (i.e., the presence of states A and B but no flip-flop in times of order τ_0), Eq. (55) gives the pure diffusion result, but with P^2 replaced by

$$P_A^2(\bar{f}_A^2+\bar{f}_A\bar{f}_B) + P_B(\bar{f}_B^2+\bar{f}_A\bar{f}_B) = P_A^2\bar{f}_A^2+P_B^2\bar{f}_B^2 \qquad (57)$$

However, in this limit the assumption $D = \bar{f}_A D_A + \bar{f}_B D_B$ breaks down because a molecule in state A remains in state A over times of the order of τ_0, and hence the problem must be

reformulated. If it is assumed that the probabilities $S_{AB} = S_{AB} = 0$, a straightforward extension of Eq. (15) then leads to

$$f(\tau) = \bar{n}[\bar{f}_A^2\left(\frac{2\pi P_A c_2}{A}\right)^2 g_A(\tau) + \bar{f}_B^2\left(\frac{2\pi P_B c_2}{A}\right)^2 g_B(\tau)] \qquad (58)$$

where g_A and g_B contain the diffusion coefficients D_A and D_B respectively.

Finally, Eq. (56) can be generalized to include both different dipole moments and Fowler-Nordheim pre-exponentials for states A and B. The result, whose derivations are completely straightforward, is

$$f(\tau) = \bar{n}\{g_1(\tau)[(\bar{f}_A c'_{1A}+\bar{f}_B c'_{1B})^2+e^{-\tau/\tau_1}\bar{f}_A\bar{f}_B(c'_{1A}-c'_{1B})^2]$$

$$+ g_2(\tau)\left(\frac{2\pi c_2}{A}\right)^2[(\bar{f}_A P_A + \bar{f}_B P_B)^2 + e^{-\tau/\tau_1}\bar{f}_A\bar{f}_B(P_A-P_B)^2]$$

$$+ g_3(\tau)2\left(\frac{2\pi c_2}{A}\right)[\bar{f}_{AA}c'_{1A}P_A+\bar{f}_{BB}c'_{1B}P_B+\bar{f}_{AB}(c'_{1A}P_B+c'_{1B}P_A)]\} \qquad (59)$$

where $c_{1A} = c_1$ for state A, etc.

Correlation Function in Terms of Field Emitted Current

We give for completeness some elementary transformations of the measurable forms of the correlation function. According to Eq. (12), the quantity we have calculated is

$$\overline{\ln(i(0)/\bar{i})\ln(i(\tau)/\bar{i})} = \overline{\ln i(0)\ln i(\tau)} - \ln\bar{i}^2 \qquad (60)$$

with $\bar{i}$ defined as $\bar{B}\, e^{-\frac{6.8\times10^7(\bar{\phi})^{3/2}v}{F}}$. If the fluctuations

are small the logarithm can be expanded so that

$$f(\tau) \simeq \overline{(i(0)i(\tau)} - (\bar{i})^2)/(\bar{i})^2 \qquad (61)$$

Thus, at $\tau = 0$ we have

$$\langle \Delta i^2 \rangle \equiv \overline{i^2} - (\bar{i})^2 \equiv (\bar{i})^2 f(0) = (\bar{i})^2 \bar{n} \left(\frac{c_1 + c_2 2\pi P}{A}\right)^2 g(0) \qquad (62)$$

The mean square current fluctuation seems to be proportional to the square of the mean current. It must be remembered, however, that c_2 is proportional to $\ell n \bar{i}$. Thus the strict proportionality between the mean square current fluctuation and $\bar{i}^2$ should occur only when c_1 predominates, i.e., when pre-exponential changes dominate over work function changes in adsorption.

It may be worth pointing out that, when the logarithm $\ell n(i/\bar{i})$ can be expanded, the entire treatment presented in this paper becomes valid for regions of arbitrary size without the implicit restriction that the (instantaneous) work function must be uniform over the probed region. For we may then obtain the total current fluctuation over the region of interest by summing over arbitrarily small area elements j the quantities: $c_1' \delta n_j + c_2 \delta \phi_j$ to obtain Eq. (5) without implicit restrictions.

Estimate of Fluctuation Size

Equation (62) provides a convenient way to estimate the relative root mean square current fluctuation. For simplicity, we assume that $c_1 = 0$, so that

$$\langle \Delta i^2 \rangle^{\frac{1}{2}} / \bar{i} = (\bar{n})^{\frac{1}{2}} c_2 \frac{2\pi P}{A} \sqrt{g_2(0)} = \frac{1}{\sqrt{\bar{n}}} c_2 \overline{\Delta\phi} \tag{63}$$

writing

$$\bar{n} = \pi r_o^2 \bar{c}_{max} \theta \tag{64}$$

$$\overline{\Delta\phi} \simeq \overline{\Delta\phi}_{max} \theta \tag{65}$$

with $\theta = \bar{c}/\bar{c}_{max}$, where $\bar{c}_{max}$ is the maximum adsorbate concentration and $\overline{\Delta\phi}_{max}$ the maximum ϕ change, we have

$$\langle \Delta i^2 \rangle^{\frac{1}{2}} / \bar{i} = \frac{c_2}{\sqrt{\pi}} \frac{\overline{\Delta\phi}_{max}}{r_o} \left(\frac{\theta}{\bar{c}_{max}}\right)^{\frac{1}{2}} \sqrt{g_2(0)}$$

$$= \frac{6.8\times10^7 v}{\sqrt{\pi}\ F} \left(\frac{3}{2}(\bar{\phi})^{\frac{1}{2}}\right) \frac{\overline{\Delta\phi}_{max}}{r_o} \left(\frac{\theta}{\bar{c}_{max}}\right)^{\frac{1}{2}} \sqrt{g_2(0)} \tag{66}$$

Thus we see that the relative root mean square current fluctuation is inversely proportional to r_o, directly proportional to the maximum work function increment, almost proportional to the square root of relative coverage (since $\sqrt{\bar{\phi}}$ is almost θ independent), and inversely proportional to the square root of maximum adsorbate concentration. By contrast, the concentration fluctuation alone has the form

$$\langle \Delta n^2 \rangle^{\frac{1}{2}} / \bar{n} = \frac{1}{\sqrt{\bar{n}}} = \frac{1}{\sqrt{\pi}} \frac{1}{r_o} \left(\frac{1}{\theta \bar{c}_{max}}\right)^{\frac{1}{2}} \tag{67}$$

The difference between Eqs. (66) and (67) results from the fact that the relative current fluctuation is proportional to the absolute density fluctuation, $(\bar{n}^{\frac{1}{2}})$, not the relative one, $(\bar{n})^{-\frac{1}{2}}$.

Although values of coverage and $\Delta\phi$ vary from adsorbate to adsorbate and from plane to plane, we may gain some idea of $\langle \Delta i^2 \rangle^{\frac{1}{2}}/\bar{i}$ by picking $\Delta\phi = 0.5$ eV, $\bar{c}_m = 10^{15}$ molecules/cm^2, $\bar{\phi} = 6$ eV, and $F = 4\times10^7$ v/cm. With these assumptions we obtain $\langle \Delta i^2 \rangle^{\frac{1}{2}}/\bar{i} = 5.5(\theta^{\frac{1}{2}}/r_o)\sqrt{g(0)}$, with r_o in Angstrom. Thus for $r_o = 50$ Å and $\theta = 0.5$ we obtain a 7% root mean square fluctuation. This indicates that a substantial effect should be observable under most conditions. In fact, for alkali metal or inert gas adsorption on tungsten where concentrations are much lower and dipole moments higher than those chosen in our example, the fluctuations may be so large that the logarithmic form of the current fluctuations in Eq. (12) must be used.

Effect of Finite Resolution

The resolution δ of the field emission microscope is of the order of $\sim$20 Å and varies with $r_t^{\frac{1}{2}}$, r_t being the tip radius.[1,2] Thus there is a probability that an electron originating in the probed region but near its boundary will not be recorded, and there is also a probability that electrons originating near but outside the boundary will be recorded. (The net result must, of course, be an average current equal to that obtained with infinitely sharp resolution.) Consequently, the effective radius of the probed region will appear larger by roughly $\delta/2$. An exact calculation is not difficult in principle under the conditions of validity of Eq. (60). In the simplest case where only g_1 need be considered it amounts to a double convolution of the resolution

function (for ρ and ρ') with the previous result. Even this becomes a very time consuming computer calculation, and was therefore omitted.

Acknowledgements

It is a pleasure to thank Professor Walter Kohn for valuable help and discussions. The computer calculations were carried out by Mr. Robert Fleischaker, NSF Undergraduate Research Participant, Summer 1972, and by Mr. J. R. Chen. This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253, with the University of Michigan, and by National Science Foundation Grant GP-29388.

References

1. R. H. Good and E. W. Müller, Handbuch d. Physik 21, 176 (1956).
2. R. Gomer, *Field Emission and Field Ionization*, Harvard University Press (1961).
3. T. Engel and R. Gomer, J. Chem. Phys. 50, 2428 (1969); ibid 52, 1832.
4. G. Ehrlich and F. G. Hudda, J. Chem. Phys. 44, 1039 (1966).
5. R. Lewis and R. Gomer, Surf. Sci. 26, 197 (1971).
6. Ch. Kleint, Surf. Sci. 25, 394 (1971).
7. L. Bess, Phys. Rev. 103, 72 (1956).
8. L. D. Schmidt and R. Gomer, J. Chem. Phys. 42, 3573 (1965).
9. C. B. Duke and M. E. Alferieff, J. Chem. Phys. 46, 923 (1967).
10. D. Penn, R. Gomer and M. H. Cohen, Phys. Rev. B5, 768 (1972).
11. C. Kohrt and R. Gomer, Surf. Sci. 24, 77 (1971).
12. S. Machlup, J. Appl. Phys. 25, 341 (1954).

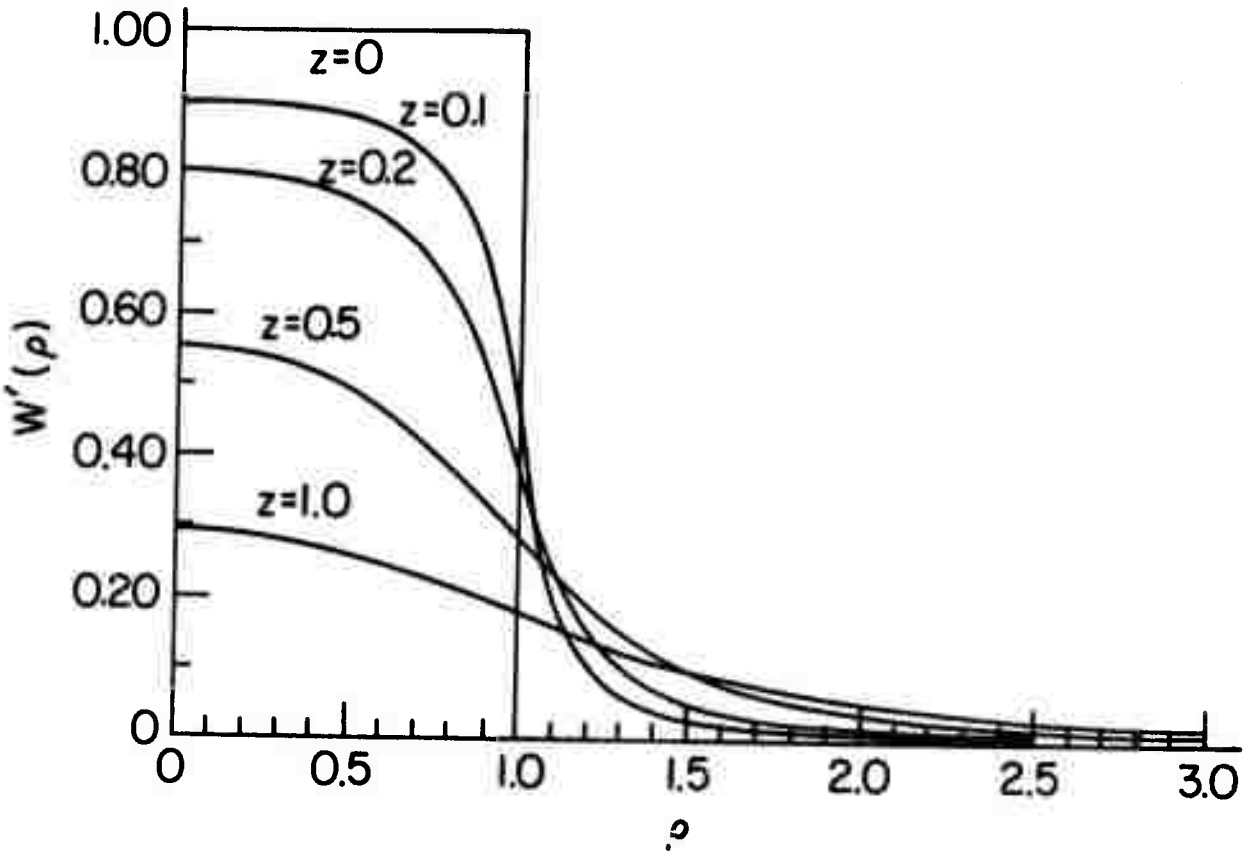


Figure 1. Plots of $W'(\rho)$ vs. ρ for various values of $z = d/r_o$, with r_o the radius of the circular probed area. $z = 0$ refers to $W'(\rho) = \delta(\rho)$.

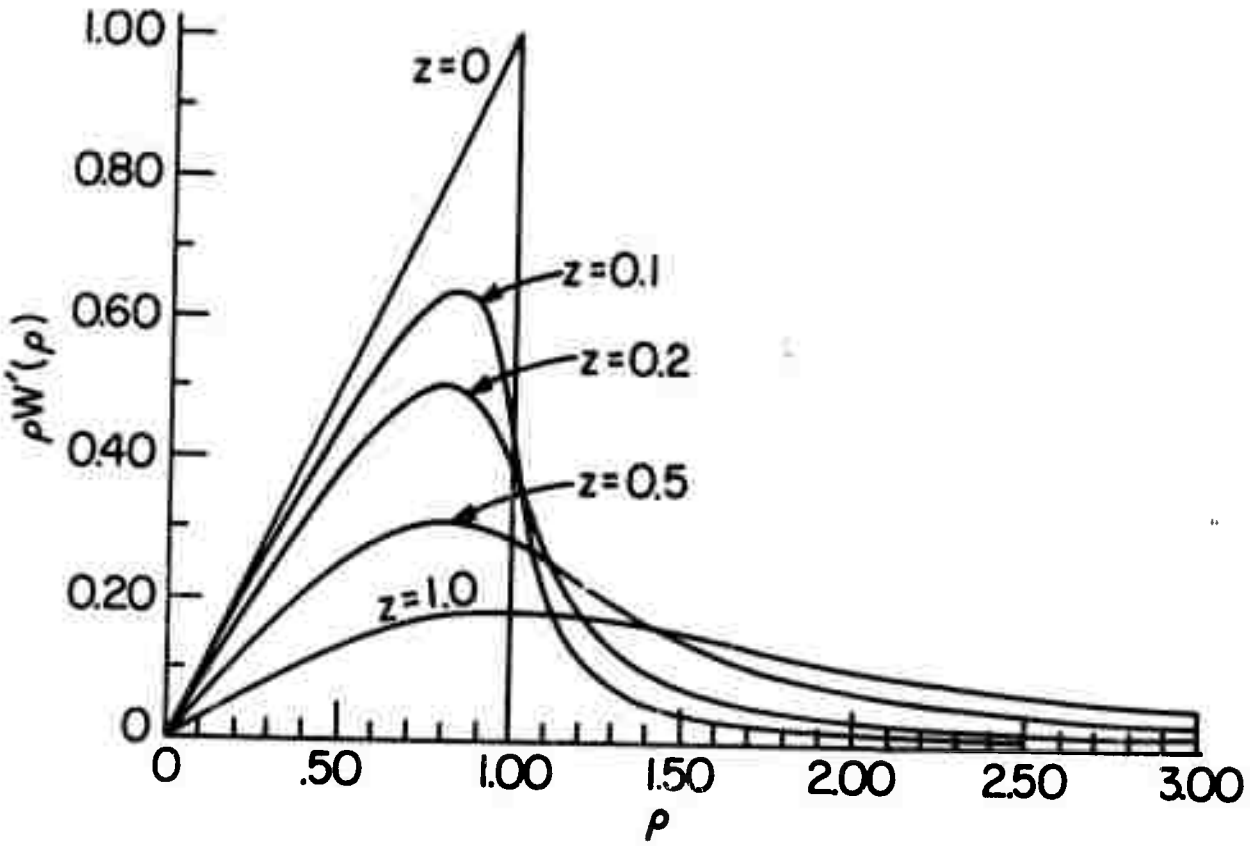


Figure 2. Plots of $\rho W'(\rho)$ vs. ρ for circular probed areas.

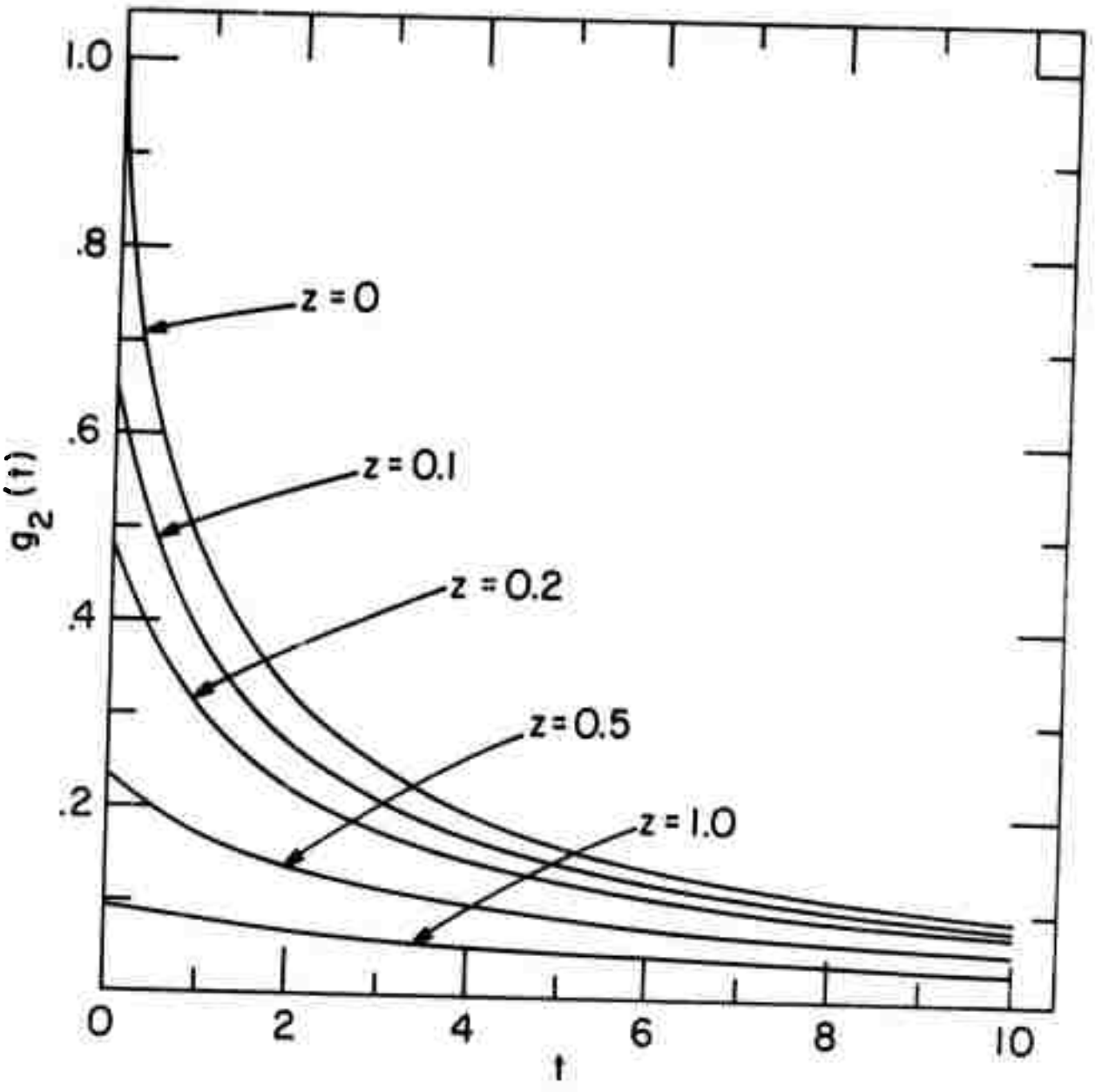


Figure 3. $g_2(t)$ vs. t for various values of $z = r/r_0$. The curve marked z = 0 corresponds to $g_1(t)$. $t = \tau/\tau_0$, with $\tau_0 = r_0^2/4D$. g_1 and g_2 are defined by Eqs. (36a) and (36b) respectively.

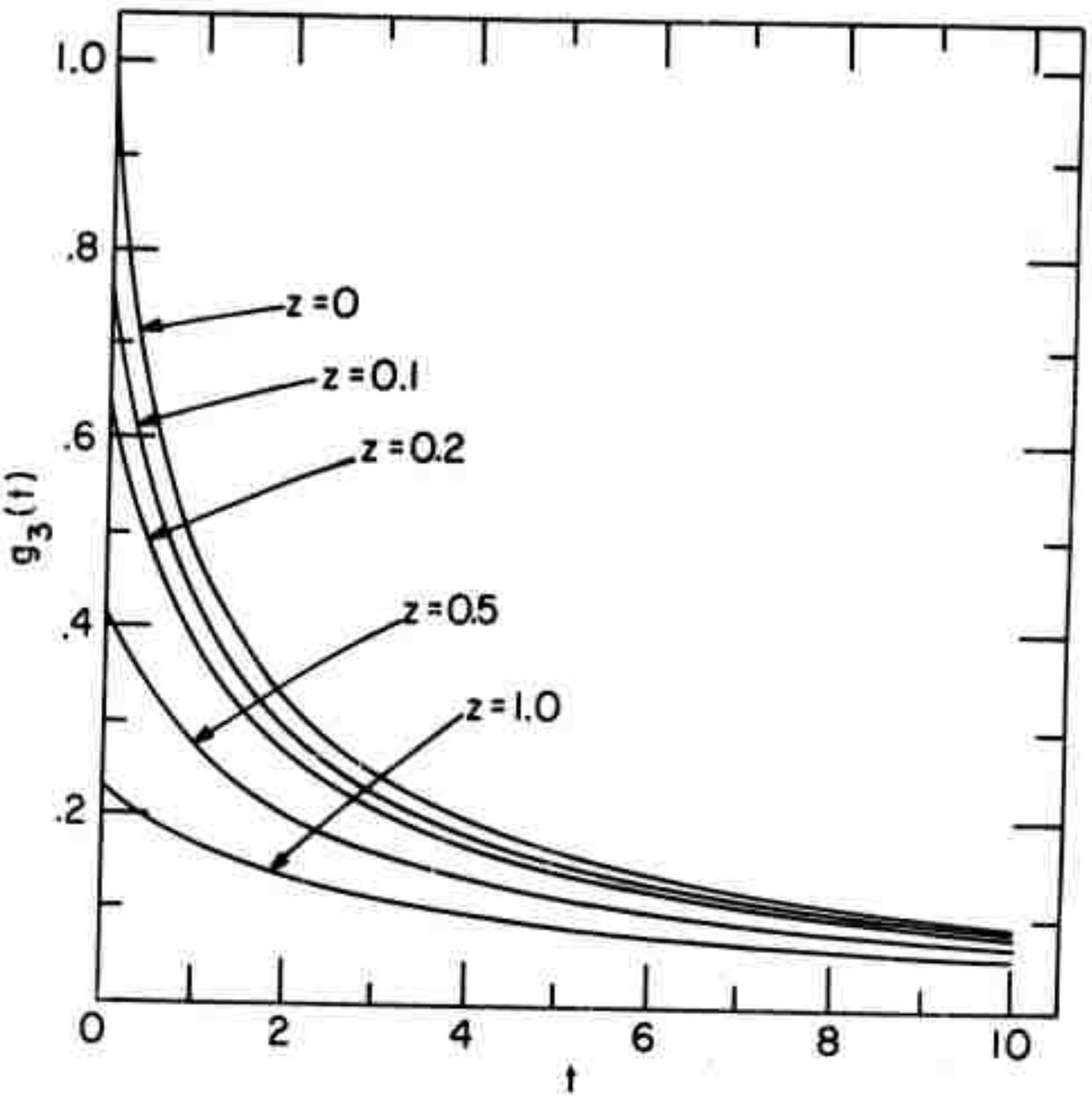


Figure 4. $g_3(t)$ vs. t for various values of z. g_3 is defined by Eq. (36c), all other notation as in Figure 3.

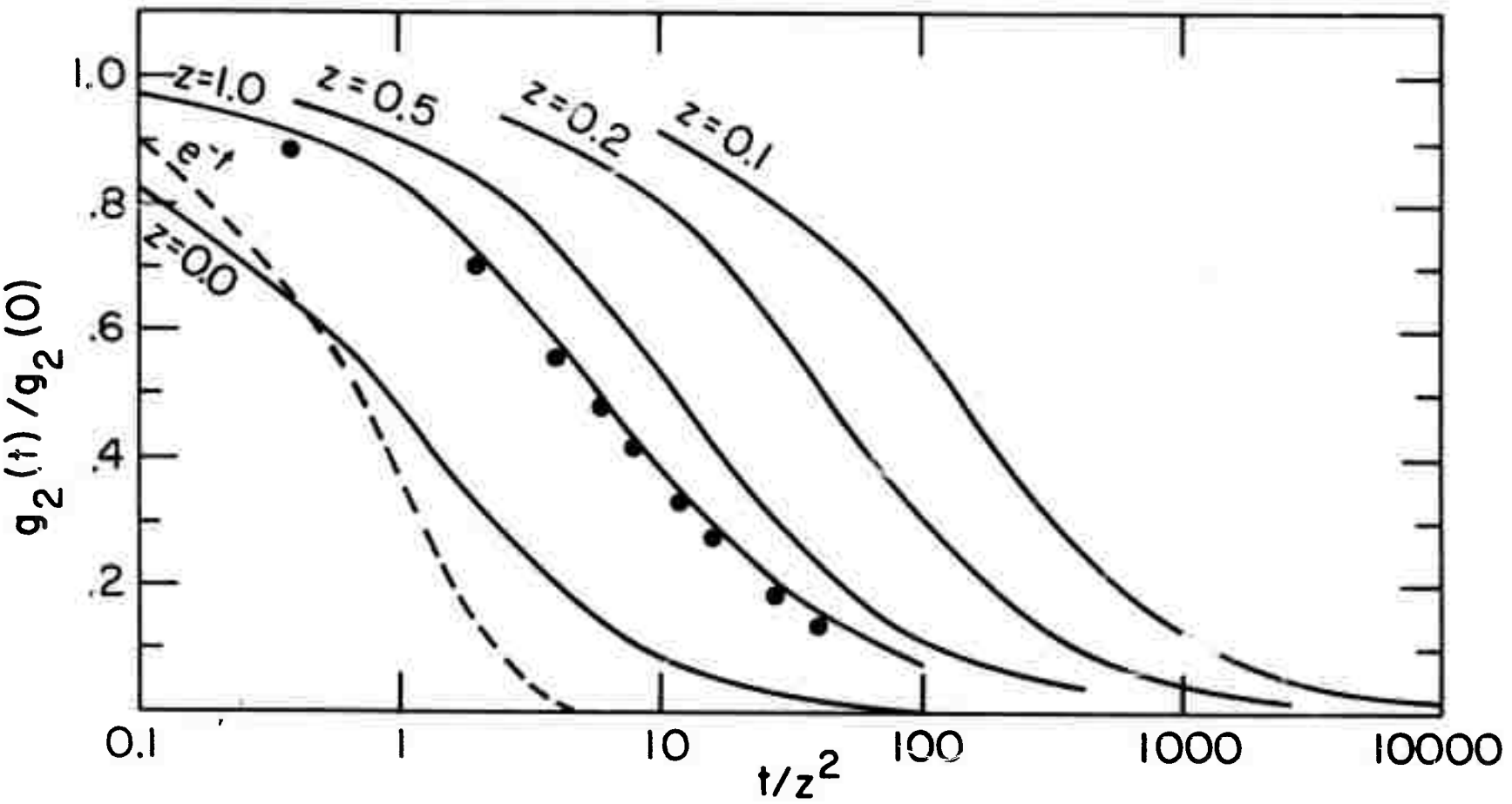


Figure 5. Plots of $g_2(t)/g_2(0)$ vs. $\log_{10}(t/z^2)$. Also shown as z = 0 curve if $g_1(t)$ vs. t. Dashed curve is e^{-t} vs. t. Solid points represent a hypothetical mixture: $g(t) = 0.25g_1(t) + 0.25g_2(t) + 0.5g_3(t)$ for z = 0.5, plotted against log(t/0.25). It is seen that the curve is shifted to the left, as expected.

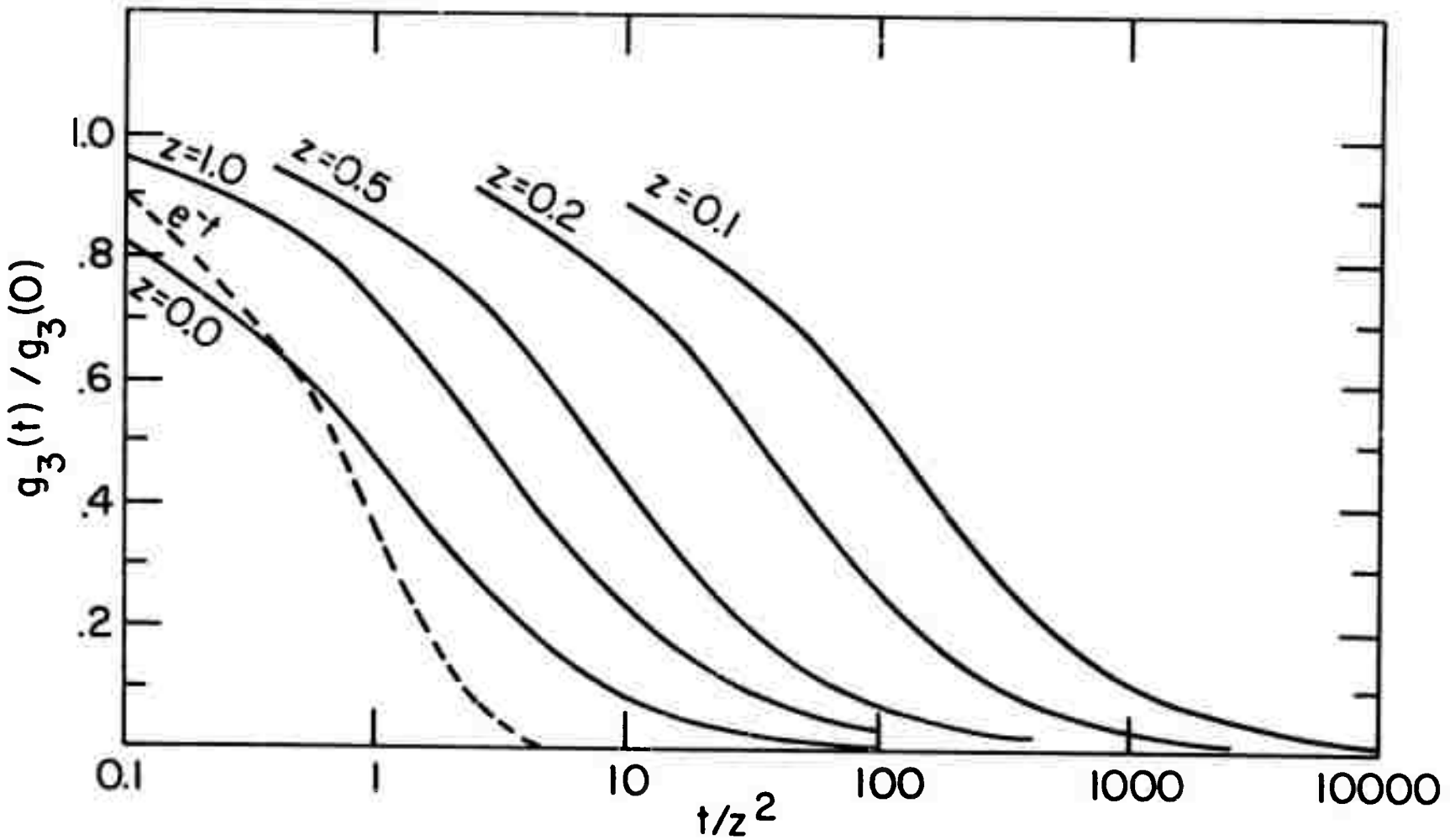


Figure 6. Plots of $g_3(t)/g_3(0)$ vs. $\log_{10}(t/z^2)$ for various values of z. z = 0 curve refers to plot of $g_1(t)$ vs. log t. e^{-t} vs. log t shown as dashed curve.

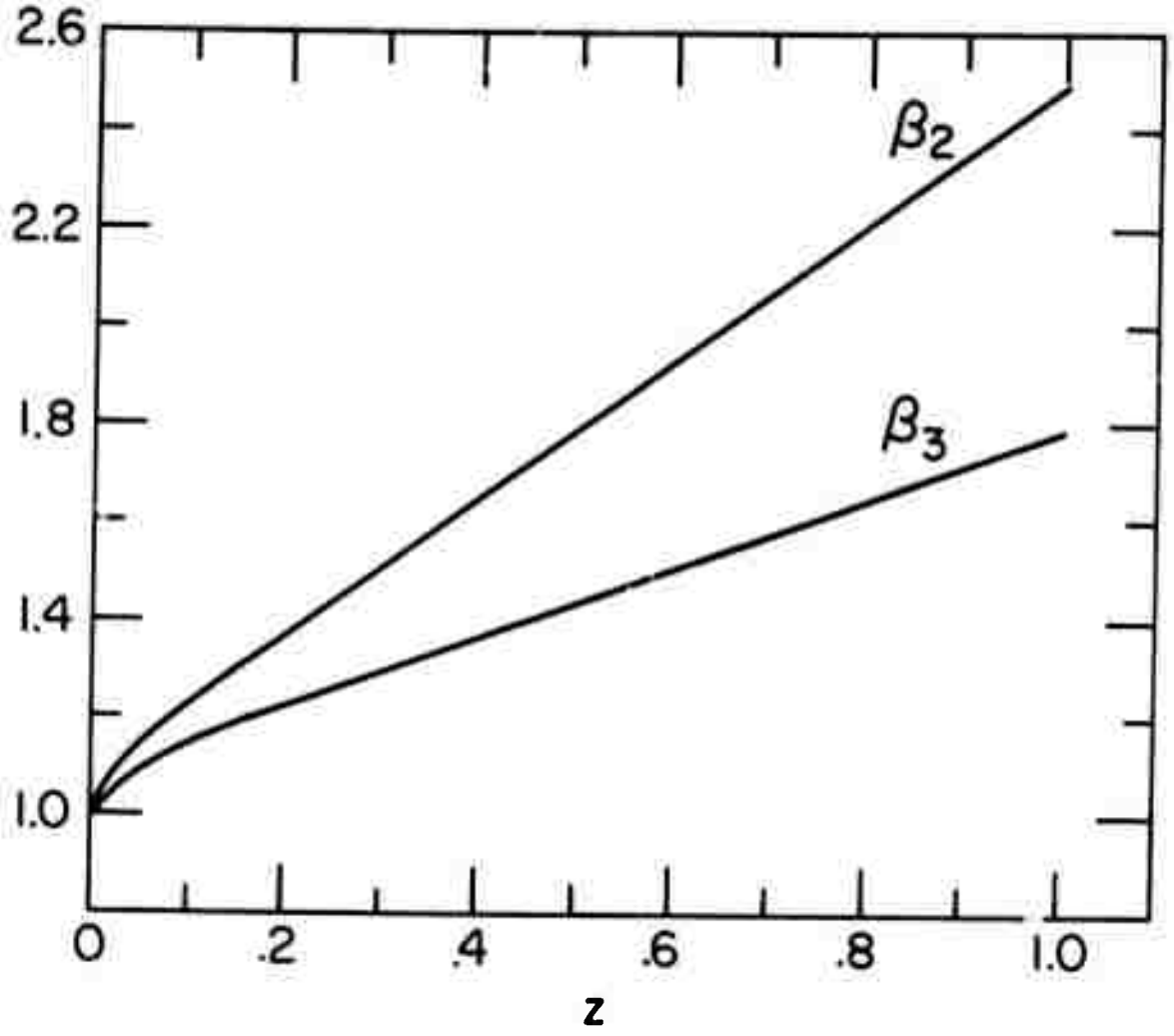


Figure 7. Plots of $\beta_2 = (t_2/t_1)^{\frac{1}{2}}$ and $\beta_3 = (t_3/t_1)^{\frac{1}{2}}$ vs/z. t_1, t_2 and t_3 are defined as the times required for $g_1(t)$, $g_2(t)/g_2(0)$, and $g_3(t)/g_3(0)$ respectively to become 0.5. βr_0 is the <u>effective</u> radius of the probed region.

CRYSTALLIZATION RATE OF AMORPHOUS ALLOYS

P. E. Duwez

Abstract

A summary is given of the various modes of crystallization of amorphous alloys. All transformations reported so far depend on both temperature and time, and hence there is no unique way of defining a crystallization temperature. In some cases the transformation takes place in one step, that is, directly from the amorphous structure to the equilibrium crystalline structure. In most cases, however, and depending on both time and temperature, intermediate metastable crystalline phases with apparently complex crystal structures are found before final equilibrium is reached. The various methods used to study the transformation kinetics are reviewed.

CRYSTALLIZATION RATE OF AMORPHOUS ALLOYS

P. E. Duwez

I. Introduction

Many studies have been published of the rate of crystallization of amorphous alloys. Both isochromal and isothermal methods have been used. The progress of crystallization can be followed more or less quantitatively by X-ray or electron diffraction, optical or electron microscopy, thermal analysis, electrical resistivity, mechanical properties and in some cases, magnetic susceptibility or Mossbauer spectroscopy. In all these methods, the difficulty is to correlate the measured variable with the actual concentration of the crystalline phase (or phases) dispersed in the amorphous matrix. A general conclusion of these studies is that the amorphous to crystalline transformation generally proceeds through one or several intermediate steps in which metastable phases nucleate and grow, and these phases ultimately disappear either at higher temperatures or after longer times of annealing.

II. Microscopic Methods

Microscopic methods provide a direct observation of the nucleation and growth of the crystalline phases. Obviously the electron microscope is more useful than the optical microscope

since the interesting features are much smaller than one micron. The first electromicroscopic observations of the growth of crystals in an amorphous alloy was reported by Willens (1962). The alloy was a gun-quenched foil of Te_{85}-Ge_{15} which was thin enough for transmission microscopy with 100 kv electrons. The specimen was heated by intermittant pulses in the intensity of the electron beam and pictures were taken at regular intervals. Since both temperature, and time at temperature were not known, the results were purely qualitative. It was clearly concluded however that the growth of the crystalline phases (which in this case are the equilibrium phases Te and TeGe) was typically dendritic. The dendrite formation was tentatively attributed to the low thermal conductivity of the amorphous alloy, which is almost an electrical insulator. Low thermal conductivity however is not a necessary condition since it was found later that well defined dendrites also appear in the early stage of crystallization of a really metallic Fe_{75}-P_{15}-C_{10} amorphous alloy (Rostogi and Duwez, 1970).

Dendritic growth is not always observed in the crystallization of amorphous alloys. In fact, it seems to be an exception rather than a rule. Crewdson (1966) investigated many binary amorphous alloys of Pd, Pt, Rh with Si, Ge and Sb and in all cases the crystalline phases appear to grow more or less equiaxed. A striking example of this type of morphology in a Pt_{77}-Si_{23} alloy was reproduced in a paper by Duwez (1967). Since then several electromicroscopy studies of partially crystallized amorphous alloys have been reported (Chen and Turnbull 1969, Masumoto and

Maddin, 1971, Srivastava et. al., 1972, Revcolevschi and Grant, 1972).

Hot stage electron microscopy is probably the most useful and perhaps the only technique to study the morphology of phase separation and crystal growth in amorphous alloys. Quantitative determinations of the various phases present at a given time might also be possible, based on techniques developed in the field of quantitative metallography. For such studies to be meaningful, a very accurate control of the temperature of the specimen itself (an not only of the stage) is absolutely necessary because of other techniques (such as electrical resistivity) have shown that the rate of crystallization is generally a very steep function of temperature.

III. Calorimetric Methods

Thermal analysis and differential thermal analysis are probably the simplest of all calorimetric methods. A small sample of amorphous alloy (or few mm^2) is spot welded to fine thermocouple wires (0.005 in. or less) and heated at an approximately constant rate. With rates of heating from 100°C/sec or more, a sharp rise in temperature occurs when the sample crystallizes very rapidly. For a typical amorphous alloy such as Pd_{80}-Si_{20}, this temperature is about 420°C and the increase in temperature corresponds to a latent heat of crystallization of about 1.0 kcal/mole. In spite of its limitations, the thermal analysis method is useful in rapidly determining the maximum temperature an

amorphous alloy can be subjected to before catastrophic crystallization takes place. This temperature could be called the "burst crystallization" temperature. The structure of a sample heated just above this temperature is not the equilibrium structure, and its X-ray diffraction pattern shows a series of poorly resolved peaks, but very little evidence for the presence of an amorphous phase.

Measuring the specific heat as a function of temperature provides more significant information and permits the determination of a glass transition temperature T_g which constitutes one of the more convincing proofs of the glassy nature of amorphous alloys. At T_g, a very sharp increase in the specific heat occurs as shown by Chen and Turnbull (1969). Like the crystallization temperature, the glass transition temperature is not unique and depends on the rate of heating. Most of the reported values for T_g were obtained with a heating rate of about 20°C/min. In general, T_g is very close to the temperature T_c (crystallization temperature) measured at the same heating rate. The difference T_c-T_g is a measure of the "stability" of the glassy state which increases with the increase in T_c-T_g. By measuring this difference, Chen and Turnbull (1969) have shown that ternary amorphous alloys Pd-X-Si in which X is Cu, Ag or Au are more stable than binary Pd-Si alloys.

IV. X-Ray Diffraction and Electrical Resistivity Methods

These two techniques are discussed simultaneously because

they can be performed in parallel, on identical specimens and their results can therefore be correlated.

The X-ray diffraction method is based on the intensity of the crystalline reflections as a function of time and temperature. A diffractometer is generally used for this purpose, mostly because a relatively large area of the specimen (about 1 cm^2) is involved. The measured intensity is not exactly proportional to the amount of a given phase in the amorphous matrix, but it is an approximate number for kinetics measurements.

Electrical resistivity measurements are well suited to study amorphous to crystalline transformations because of the large decrease in resistivity. Most of the results reported so far were obtained at a constant rate of heating. When comparing the results of several investigators on a given alloy system, it is important to keep in mind that the rate of heating can greatly affect the shape of the resistivity-temperature curve, and hence the conclusions of the authors. Most of the studies were made with heating rates of 1 or 1.5°C/min (Duwez 1967, Maitrepierre 1970, Lin 1969, Sinha 1970). The results on Pd_{80}-Si_{20} reported by Chen and Turnbull (1969) were obtained at a heating rate of 10°C/min. The heating rate used by Revcolevschi and Grant for Cu_3Zr_2 is not mentioned in their paper.

The shape of the resistivity-temperature curve is approximately the same for all amorphous metallic alloys. At low temperatures, the relationship is linear, except for the presence of a resistivity minimum (Kondo effect) in alloys containing a magnetic

atom. The resistivity either increases or decreases with temperature but in all cases the temperature coefficient of resistivity is much smaller than in crystalline alloys. This is due to the fact that the important scattering mechanism is that due to disorder in the structure and the phonon scattering contribution to the resistivity is relatively small. At a given temperature, which may be around room temperature for Au_{80}-Si_{20} and as high as 400 to 500°C for Pd-base alloys, the resistivity decreases very rapidly with increasing temperature due to the nucleation and growth of one or several crystalline phases. After reaching a minimum, the resistivity starts increasing linearly with temperature as the structure approaches equilibrium. In almost all cases, intermediate non-equilibrium phases are observed by X-ray diffraction. In some cases these phases may be quite stable within certain intervals of temperatures and this is clearly shown by the presence of steps in the resistivity-temperature curve (see for example Maitrepierre 1970).

At a rate of heating of 1°C/min it is possible to correlate the resistivity curve with the structural changes in the alloy observed by X-ray diffraction. In this case, a specimen large enough for diffractometer measurements is heated for 20 minutes at successive temperatures with 20°C intervals and a diffraction pattern is taken after each step. This is a time-consuming method, because to detect small structural changes, the rate of scanning must be small enough to assure accurate counting statistics, but it appears to be the only reliable method to explain the origin

of the various steps observed in the resistivity-temperature curves.

There is no doubt that in order to obtain meaningful quantitative kinetics data, it is necessary to use the isothermal annealing method. Only very sketchy results have been reported so far. Srivastava et al. (1972) studied the transformation of gun-quenched Pt_{66}-Sb_{34}, but their data are restricted to only four temperatures in a narrow range (212 to 240°C) and the longest annealing time was only 5 hours. Within these ranges they found an activation energy of about 49.5 to 55.5 kcal/mole. It is doubtful that a single activation energy can be assigned to a process as complicated as the amorphous to crystalline transformation when it takes place by overlapping steps as described above. Because of the very low diffusion rates involved in the early stages of crystallization (in the low temperature range) detailed kinetics studies may require at least several months and may be several years of measurements. Such an effort involving a few typical and well-chosen alloys would probably be very rewarding.

A combination of X-ray diffraction and electrical resistivity methods has been used by Duwez and Tsuei (1970) in a kinetics study of the crystallization of amorphous Te_{75}-Cu_{25}-Au_5 alloys obtained by gun-quenching from the liquid state. These amorphous alloys are semiconductors, but are metallic conductors in their equilibrium state. Isothermal annealing temperatures range from 45 to 89°C with total times for complete transformation

from 1 hour to 1 month. The transformation was followed by resistivity measurements in the high-temperature range, and X-ray diffraction became more practical for the slow transformations at lower temperature. In these alloys the transformation seems to proceed directly to the equilibrium phase (Te + TeCu) without intermediate metastable crystalline phases. The intensity of the strongest crystalline Te line versus log-time at a given temperature increases according to an S curve typical of a nucleation and growth process. A plot of log of half-reaction time versus reciprocal temperature (linear) leads to a heat of activation of 1.73 ± 0.04 eV. The time for half-reaction is about 20 min. at 89°C and 700 hours at 42°C. Extrapolating these data to room temperature indicates a half-reaction time of approximately four years.

V. Mechanical Properties Method

Measuring mechanical properties of an amorphous alloy at various stages of crystallization is not by itself a satisfactory method to obtain kinetics data. When performed in parallel with X-ray diffraction, however, it can lead to interesting correlations. Such a correlation has been described by Masumoto and Maddin (1971) for Pd_{80}-Si_{20} amorphous alloys. Their X-ray work confirms the previous observation of Crewdson (1966) that the first phase to crystallize out of the amorphous matrix is Pd or a metastable Pd-Si solid solution. This is followed by the crystallization of another metastable phase (for which they give a list of spacings)

and finally the equilibrium state is reached. According to the authors the stable phase is Pd_3Si, although the alloy contains only 20 at.% Si instead of 25 at.%. These three stages in the crystallization process are reflected by observable changes in fracture and yield stresses and in Young's modulus. The total elongation after fracture, however, seems to be insensitive to the structural changes, except for a detectable increase when crystallization is almost complete. Viscous flow was observed in tensile tests at 300 to 400°C.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

References

1. Chen, H. S. and Turnbull, D., J. Chem. Phys. 48, 2560 (1968).
2. Chen, H. S. And Turnbull, D., Acta Met. 17, 1021 (1969).
3. Crewdson, R. C., Ph.D. Thesis, California Institute of Technology, Pasadena, California (1966).
4. Duwez, P. E., Willens, R. H. and Crewdson, R. C., J. Appl. Phys., 36 2267 (1965).
5. Duwez, P. E., in *Phase Stability in Metals and Alloys*, P. S. Rudman, J. Stringer and R. I. Jaffee Eds., McGraw-Hill Book Co., New York, 1967, p. 523.
6. Duwez, P. E. and Tsuei, C. C., J. Non-Cryst. Sol. 4, 345 (1970). This is a short report. A more extensive paper is in preparation.
7. Giessen, B. C., and Wagner, C. N. J., in *Liquid Metals-Chemistry and Physics*, S. Z. Beer ed., Marcel Dekker, New York, 1972, pp. 633-695.
8. Lin, S. C. H., J. Appl. Phys. 5, 2173 (1969).
9. Maitrepierre, P., J. Appl. Phys. 41, 498 (1970).
10. Masumoto, T. and Maddin, R. H., Acta Met. 19, 725 (1971).
11. Pond, R., Jr. and Maddin, R. H., Trans. AIME 245, 2475 (1969).
12. Rastogi, P. K. and Duwez, P., J. Non-Cryst. Sol. 5, 1 (1970).
13. Revcolevschi, A. and Grant, N. J., Met. Trans. 3, 1514 (1972).
14. Sinha, A. K., Phys. Rev. B. 1, 4541 (1970).
15. Srivastava, P. K., Giessen, B. D. and Grant, N. J., Met. Trans. 3, 977 (1972).
16. Tsuei, C. C., Phys. Rev. 170, 775 (1968).
17. Willens, R. H., J. Appl. Phys. 33, 3269 (1962).

THE ROLE OF CRACKS, PORES AND ABSORBING INCLUSIONS ON LASER INDUCED DAMAGE THRESHOLD AT SURFACES OF TRANSPARENT DIELECTRICS

N. Bloembergen

Abstract

The concentration of the electric field strength in the neighborhood of micropores and cracks may lower the nominal external intensity for electric avalanche breakdown by a factor two to one hundred depending on the geometry of the crack and the dielectric constant.

The presence of absorbing inclusions at the edge of microcracks will often be the dominant mechanism giving the lowest surface damage threshold. Inclusions and cracks with characteristic dimensions less than about 10^{-6} cm will not lower the breakdown threshold appreciably.

THE ROLE OF CRACKS, PORES AND ABSORBING INCLUSIONS ON LASER INDUCED DAMAGE THRESHOLD AT SURFACES OF TRANSPARENT DIELECTRICS

N. Bloembergen

I. Introduction

It is well established that electric avalanche breakdown induced by laser pulses of very high-power density is the cause of damage in the intrinsic, perfect lattice of large band gap transparent insulators.[1,2] If the bulk material contains absorbing inclusions, the damage threshold is often determined by the thermal stresses induced by the heating of the inclusion in the laser beam. The case of platinum particles in glass has been studied in detail both experimentally and theoretically.[3]

It has been found that surfaces of such bulk materials, even those that are scrupulously cleaned or that have been obtained by cleavage in vacuum, have a lower damage threshold than the bulk. For laser glass and sapphire (Al_2O_3) the best obtainable surface damage threshold is a factor four to five lower than in the bulk.[4,5,6] If the surface is contaminated with additional absorbing dirt, the damage threshold may be lowered much more.

One might expect that for a perfect intrinsic surface the mechanism for electric breakdown would be the same as that for the bulk. Although there would be changes in the intrinsic

absorption of surface states, surface excitations and vibrations, such effects are much too small to influence the observed damage threshold, especially for light frequencies in the middle of the bulk transparent region. Even a monolayer of strongly absorbing foreign atoms or molecular groups such as OH at 10.6 μm is not a determining factor for the surface damage threshold.

It is the purpose of this note to show that the presence of sub-microscopic cracks and pores will cause local variations in the electric field strengths, which can readily account for an apparent lowering of the surface breakdown intensity by a factor 4 or 5 compared with that of the bulk.

The same factor will be operative if the damage mechanism is initiated by absorbing inclusions. This absorption mechanism is more effective for inclusions near the edge of microcracks and crevasses than for the same inclusions in the bulk lattice.

II. Examples of Local Electric Field Values Determining Breakdown and Damage

The initiation of an electric avalanche breakdown is a local phenomenon and may occur on a length scale of one micron or less.[6] A typical example of this is afforded by the filamentary fossile damage tracks after self-focusing has occurred in glasses. The size of absorbing inclusions initiating damage is usually also small compared to the wave length. Larger inclusions can readily be determined by optical means and they are assumed to have been eliminated.

If a standing wave interference pattern is present in the bulk material, damage will occur as soon as the field strength in the antinodes exceeds the critical value. This phenomenon was carefully observed and analyzed by Crisp et.al.[4], who observed that the nominal external intensity for damage at the rear surface of a plane parallel glass window is about a factor 1.5 less than for the front surface. The reason is that the Fresnel reflections for normal incidence set up a field distribution such that the electric field at the front surface is 0.8 E_{inc} and the electric field at the rear surface is 0.96 E_{inc}, where E_{inc} is the amplitude of the incident wave. The breakdown occurs if the field at either surface reaches the threshold value E_{Br}. Thus, the nominal breakdown intensity of the incident wave is $(0.96/0.8)^2$ higher for the front surface.

By the same type of argument the nominal breakdown intensity of a dielectric layer, more than $(1/4)\lambda$ thick, covering a metal mirror, will be about four times lower than for the same layer covering a low index dielectric material. The standing wave pattern in front of the mirror enhances the maximum electric field in the antinode by a factor two over the incident wave.

As a corollary, a thin protective coating over a metal surface will be at an electric field node and is not likely to be damaged. By the same argument no damage would be expected at a dielectric surface placed $(1/2)\lambda$ away from a mirror.

The same type of argument may be applied to an ideal anti-reflective coating, $(1/4)\lambda$ thick with index n, where n is the

index of the bulk substrate. The field strength in this case will be largest at the front surface, where it will be equal to the incident amplitude E_{inc}. The field strength is smallest at the interface with the substrate where it equals $E_{inc}\, n^{-\frac{1}{2}}$. Thus breakdown threshold in the anti-reflective coating is lowest at the front surface. If the rear of the window is also anti-reflection coated, its exit face has the same breakdown threshold.

III. Local Fields near Micropores and Cracks

The same idea may now be applied to the local field configuration near small pores, crevasses, grooves, scratches and cracks. It will be assumed that all characteristic dimensions are small compared to the wave length. The electric field configurations, and possible internal field enhancements may then be calculated by the methods of electrostatics.[7] The problem has been solved exactly in closed form for cavities of ellipsoidal shape.[8] Here we shall consider explicitly three representative geometries which permit an estimate of the enhancement factors. Figure 1 shows a linearly polarized wave normally incident on a dielectric with dielectric constant $\varepsilon = n^2$. We consider a spherical pore, a cylindrical groove and a crack in the shape of an oblate ellipsoid and ask for the field strengths at points A, B and C, respectively. The questionable assumption will be made that the field at B is the same as for a cylinder completely embedded in the dielectric. The assumption of a complete ellipsoidal cacity will be quite good in the case of geometry C.

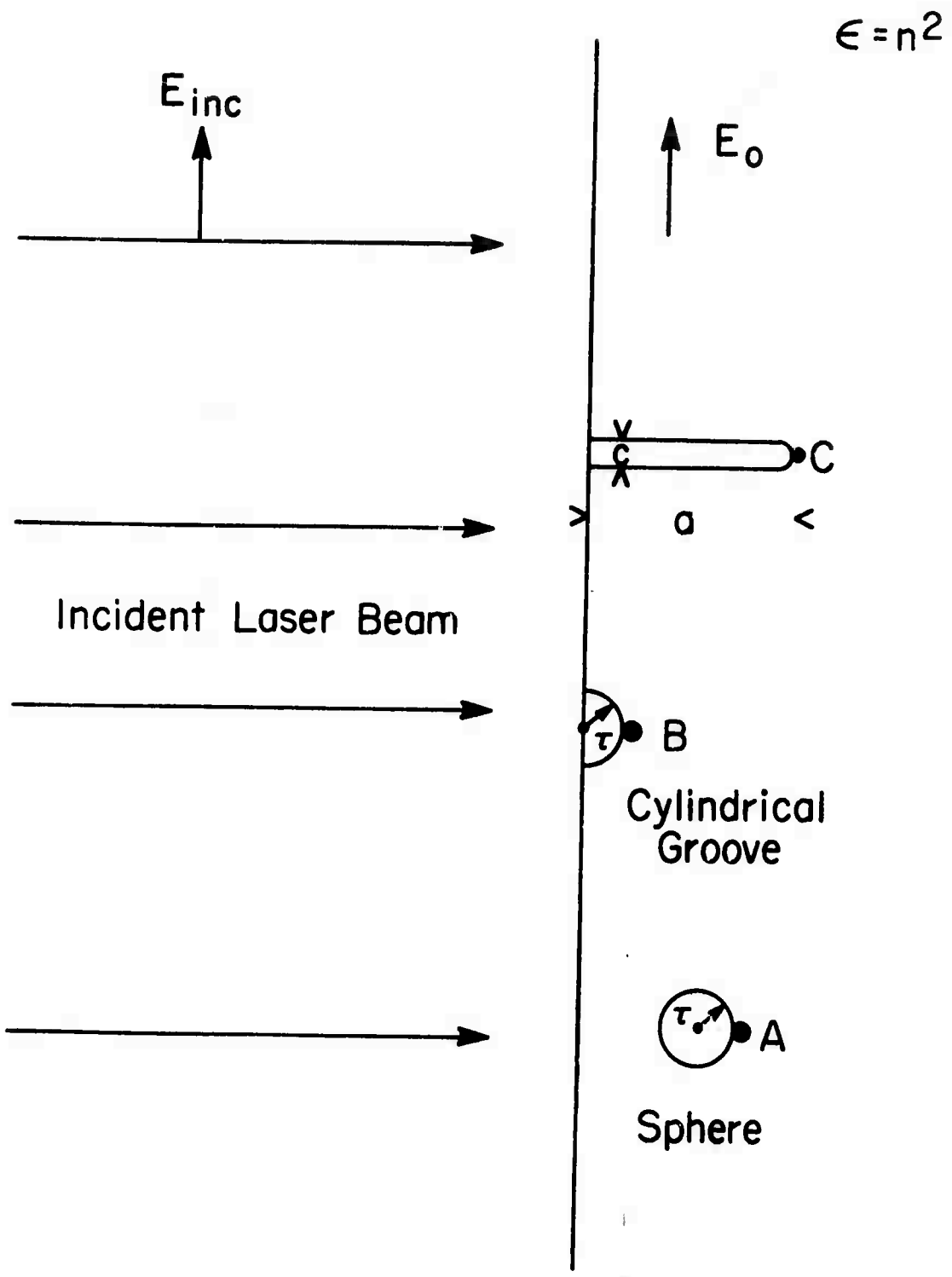


Figure 1. Representative geometries for electric field enchancement near pores, scratches and incipient cracks. Typical dimensions are: τ = 0.1-1μm, c = 0.1μm and a = 1μm.

The field inside the cavities is given in terms of the field E_o inside the uniform dielectric as follows

$$E_{ins} = \frac{1}{1 + \frac{1-\varepsilon}{\varepsilon} L} E_o$$

where L is the appropriate depolarization factor. Note that the field at the points A, B and C in the dielectric is equal to the field in the pore.

A. For the sphere one has $L = 1/3$, and

$$E_A = \frac{3\varepsilon}{2\varepsilon + 1} E_o$$

B. For the cylinder, with axis perpendicular to the paper, $L = 1/2$ gives

$$E_B = \frac{2\varepsilon}{\varepsilon + 1} E_o$$

The maximum enhancement factor is 2 for $\varepsilon \to \infty$. For NaCl with $n = 1.5$ and $\varepsilon = 2.25$ at 10.6 μm one finds a field enchancement factor 1.4. This type of groove lowers the apparent threshold intensity for surface damage by a factor 2 in NaCl and nearly by a factor 4 in high index materials such as CdTe and ZnSe.

C. In this most important and interesting case

$$L = 1 - \frac{\pi}{2}\frac{c}{a} \quad \text{or} \quad L \simeq 1 \text{ for } c/a << 1/\varepsilon < 1$$

In this limit the field at point C is given by

$$E_C = \varepsilon E_o$$

The apparent breakdown intensity is decreased by a factor ε^2, or 5 times for NaCl and by about a factor 100 for CdTe and ZnSe.

If the damage threshold is determined by absorbing inclusions, the same factors are valid if the inclusions are situated at locations A, B or C, respectively, and have a size approximately equal to the solid dots in the figure and if they have a real index equal to the index of the bulk window material and a loss tangent $\varepsilon''/\varepsilon'' << 1$. It is believed that these situations correspond rather closely to real conditions. It is most likely that inclusions collect at the bottom of a groove as in B, or that a crack ends at an inclusion as at C.

Thus the lowering of the apparent threshold intensity for damage by surface cracks and grooves may indeed be expected to lie between two and five for low index materials such as the alkali halides, glass and sapphire. The factor may be considerably larger for high index materials such as ZnSe, CdTe and chalcogenide glasses.

Similar factors will also enhance the surface absorptivity at low power levels, if the absorption centers are concentrated at locations of electric field enhancements. In this connection it is worthwhile to mention the experimental results of Deutsch,[9] who found that the surface absorptivity of cleaved alkali halide crystal surface was increased by a factor 4. Presumably, these surfaces have more incipient cracks and sharp edges.

All these factors are multiplicative to the factors due to standing wave patterns discussed in the preceding section.

It is clearly possible to extend the calculations to other, more complex, geometries where the index of the inclusion near the pore is different from the bulk, or to the case of a metallic inclusion. If the crack is approximated by a sharp wedge, electrostatic solutions may be obtained by the method of conformal mapping. The electric field at the tip of the edge would then diverge as $d^{-\frac{1}{2}}$ where d is the distance from the infinitely sharp edge. This same dependence occurs also in the famous Sommerfeld solution for diffraction at a conducting half plane.[10] Maxwell[11] already observed that the infinite electric field would cause breakdown. Therefore, the tip of the wedge would be rounded. In the next section it will be shown that small cracks and inclusions, say, with diameters less than 0.1 μm, are much less effective in producing damage than those with diameters between 0.2 μm and 1 μm. Thus the solution for the case of a sharp wedge has only academic interest.

IV. Damage by Thermal Stress from Absorbing Inclusions

The problem of thermal stress around an absorbing inclusion heated by a laser beam has been discussed by Hopper and Uhlmann.[3] Their attention was primarily directed at platinum inclusions in glass exposed to a laser pulse of 3×10^{-8} sec duration and they found that the heating effect went through a maximum as the size of the metallic particles was increased. For platinum the particle diameter is large compared to the optical thickness, which is about 0.01 μm. The energy absorbed

during the laser pulse is proportional to the cross section a^2, the temperature rise of the metallic volume is proportional to a^{-1}, as heat conduction losses may be ignored during the duration of the laser pulse. For very small sizes (<< 0.2 μm), by contrast, the inclusion can lose energy due to heat conduction to the surrounding glass suring the laser pulse. Consequently, for a given beam power, pulse duration and opaque particles, the thermal stress goes through a maximum. It occurs at 0.2 um particle size for Pt in Nd laser glass with a 30×10^{-9} sec pulse duration.

For dielectric or semiconducting inclusions with a smaller specific absorptivity the size of particles with the lowest damage threshold will be larger. The interest for the case of dielectric surfaces is in the damage threshold of inclusions as a function of their radius in the regime where this radius is small compared to the optical absorption depth. Only a crude qualitative analysis of the expected temperature rise of the inclusion will be given here. A more quantitative formulation along the same lines as the theory of Hopper and Uhlmann could readily be developed.

The initial rate of heating in the absorbing inclusion causes the temperature to rise according to:

$$\Delta T = \frac{\omega \chi'' |E|^2}{c_v} t$$

where ω is 2π times the light frequency, $\chi'' = \varepsilon/4\pi$ is the imaginary part of the susceptibility and c_v is the specific heat per unit volume. The optical absorption depth (Beers' length) α^{-1}, is the

inverse of the absorption coefficient,

$$\alpha = 8\pi^2 \chi''/\lambda n$$

where λ is the wave length in vacuum and n the index of refraction. For very short laser pulses of duration t_p we should use the equation for ΔT with $t = t_p$. For many situations of interest the time t is, however, determined by the characteristic time t_c for heat conduction away from the inclusion. A steady state temperature rise is obtained with $t = t_o = a^2/D_T$, where D_T is the thermal diffusivity. The order of magnitude of D_T for many solids near room temperature is about 10^{-2} cm^2/sec, so that t_c is about one microsecond for a particle of one micron diameter. For inclusions smaller than 0.1 μm radius and laser pulse durations larger than 10^{-8} sec, we are in the steady state limit, and the approximate temperature rise at the inclusion is

$$\Delta T_{ss} \simeq \frac{\omega \chi'' |E|^2 a^2}{D_T C_v}$$

A representative numerical example is as follows: $\omega \simeq 6\pi \times 10^{13}$ corresponding to $\lambda = 10$ μm, $\chi'' \simeq 10^{-2}$, corresponding to an optical absorption depth of 2×10^{-3} cm at 10.6 μm. Typically such an absorption would occur in a dirty semiconductor with about 10^{18} conduction electrons per cm^3. We take $c_v = 0.1$ cal/cc, $D_T = 10^{-2}$ and $a = 0.1$ μm. A catastrophic temperature rise, arbitrarily taken to be 8000°C, is achieved for a field strength $E = 1.2 \times 10^5$ volts/cm, acting at the site of the inclusion. For

the same inclusion with a radius of 0.01 μm (100Å), the field strength required for a catastrophic temperature rise would be 1.2×10^6 volts/cm. This field strength is, however, comparable to the field strength for intrinsic avalanche dielectric breakdown of the pure bulk material. For laser pulses with $t_p \geq 10^{-8}$ sec, this critical field strength[6] lies near 10^6 volts/cm. It is therefore concluded that inclusions of sizes of 0.01 μm or less are not harmful.

Inclusions with sizes of 1 μm or more can readily be detected and avoided. The most harmful inclusions are consequently those with sizes between 0.1 and 1 μm, and optical absorption depths between 10^{-3} and 10^{-1} cm.

In our numerical example, the catastrophic heating field $E = 1.2 \times 10^5$ volts/cm corresponds to an external power flux density, after a reduction factor of four due to geometric field enhancement near a crack is taken into account, of 10 megawatts/cm^2. Without the presence of a crack, the breakdown power density would be 40 megawatts/cm^2. This is clearly in reasonable accord with experimental findings, taking into account the considerable spread in size of inclusions and their loss tangents.

Finally the question should be raised whether small cracks are damaging in materials absolutely free from absorbing inclusions. The geometric enhancement factor of the electric field is present even for cracks as narrow as 5Å, because down to this scale electrostatic continuum theory remains valid. Electric avalanche break-

down fields are, however, also dependent on the size of the region over which they are applied. This dependence has been demonstrated experimentally for dc breakdown in thin alkali halide crystals. The electron drift or transit time between the dc electrodes is here a crucial quantity.[6]

For light induced breakdown, the loss of hot carriers due to diffusion outside the discharge volume must be considered. A crude estimate indicates that the effective containment time of the carriers is about 10^{-8} sec in a high light intensity region with a characteristic dimension of 0.5 μm or more. For a region with dimensions of 0.01 μm or less the containment time is decreased to the picosecond regime. Here the critical field for avalanche breakdown is increased by an order of magnitude,[6] more than off-setting the geometric factor for threshold lowering. It is therefore concluded that not only inclusions, but also pores and cracks with characteristic dimensions smaller than 0.01 μm may be ignored.

V. Conclusions

The order of magnitude considerations in this paper could be refined considerably by a quantitative analysis of the electric field, thermal field and stress field configuration near cracks and inclusions.

However, in view of our lack of knowledge of the precise geometry and nature of the cracks and impurities, such more refined calculations, although desirable, are not essential. The

following conclusions may be drawn from our qualitative analysis.

1. Incipient sub-microscopic cracks and pores at surfaces or in optical coatings lead to local enhancement of the electric field strength in laser beams.

2. This enhancement causes a decrease in the nominal damage threshold intensity. The decrease is estimated to lie between two and five for low index materials such as alkali halides, glass and sapphire. It may be considerably larger in high index materials such as ZnSe and CdTe.

3. This decrease in damage threshold due to cracks is present not only for inclusion-free materials, where the damage mechanism is electron avalanche breakdown, but also for materials with bulk inclusions where the damage is due to thermal stress around the absorbing inclusions.

4. The prediction of a factor two to five reduction in surface damage threshold compared to the bulk is in good agreement with observations on carefully cleaned surfaces of glass and Al_2O_3.

5. During surface and coating preparation the deposition of additional absorbing inclusions should be carefully avoided. Inclusions, pores and cracks of characteristic dimensions smaller than 10^{-6} cm (100Å) are not harmful. The presence of absorbing inclusions, grooves, scratches and incipient cracks larger than 0.1 μm should be scrupulously avoided. Super-polishing and chemical polishing techniques will be beneficial in high power window surface preparation.

6. It will be extremely difficult to produce anti-reflective coatings for $\lambda = 10.6$ μm, with a thickness of several microns which have the required absence of cracks, pores and absorbing inclusions. The simultaneous concentration of electric field, stress field, and inclusions at the edges of cracks presents a very powerful conspiracy to lower damage thresholds.

Acknowledgement

This work was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with the University of Michigan.

References

1. E. Yablonovitch, Appl. Phys. Letters 19, 495 (1971).

2. D. Fradin, E. Yablonovitch and M. Bass, Paper presented at the Conference on Fundamentals of Damage in Laser Materials, Boulder, Colorado, June, 1972.

3. R. W. Hopper and D. R. Uhlmann, J. Appl. Phys. 41, 4023 (1970).

4. M. D. Crisp, N. L. Boling and G. Dubé, Appl. Phys. Letters, to be published (1972).

5. C. R. Guiliano, Appl. Phys. Letters 21, 39 (1972) and private communication.

6. E. Yablonovitch and N. Bloembergen, submitted to Phys. Rev. Letters.

7. See, for example, L. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Pergamon Press, London, 1960.

8. J. A. Stratton, *Electromagnetic Theory*, pp. 211-214, McGraw-Hill, New York, 1940.

9. F. A. Horrigan and T. F. Deutsch, Raytheon Research Report, July 1972, and private communication.

10. See, for example, M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, London (1960).

11. J. C. Maxwell, *A Treatise on Electricity*, Vol. I, p. 294, Oxford Clarendon Press, 1892.

PROPAGATION OF LOW FREQUENCY ELASTIC DISTURBANCES IN A COMPOSITE MATERIAL

W. Kohn

Abstract

In the limit of low frequencies the displacement $u(x,t)$ in a one-dimensional composite can be written in the form of an operator acting on a slowly varying envelope function, $U(x,t)$: $u(x,t) = (1 + v_1(x)\frac{\partial}{\partial x} + \ldots)\, U(x,t)$. $U(x,t)$ itself describes the overall long wavelength displacement field. It satisfies a wave equation with constant, i.e., x-independent, coefficients, obtainable from the dispersion relation $\omega = \omega(k)$ of the lowest band of eigenmodes: $(\partial^2/\partial t^2 - \bar{c}^2\partial^2/\partial x^2 - \beta\partial^2/\partial x^4 + \ldots)U(x,t) = 0$. Information about the local strain, on the micro-scale of the composite laminae, is contained in the function $v_1(x)$, explicitly expressible in terms of the periodic stiffness function, $\eta(x)$, of the composite. Appropriate Green's functions are constructed in terms of Airy functions. Among applications of this method is the structure of the so-called head of a propagating pulse.

PROPAGATION OF LOW FREQUENCY ELASTIC DISTURBANCES IN A COMPOSITE MATERIAL

W. Kohn

1. Introduction

The propagation of an elastic disturbance in a composite materials is, in general, a very complicated phenomenon due to the multiple reflections and refractions taking place at the interfaces of the constituents. This complexity is borne out by measurements and by computer calculations.[1] Under these circumstances it is obviously most desirable to extract some simple features of such disturbances which can be understood in general terms, without recourse to an experimental measurement or a computer calculation in each new case.

A significant contribution of this kind was made by Peck and Gurtman.[2] These authors studied the stress wave produced in a laminated composite half-space due to pulsed loading on the bounding face. Using an analysis in terms of the waveguide-like modes of the system, they obtained an explicit expression for this stress wave in the form of an infinite series of Fourier integrals. For points far from the source of the disturbance they evaluated these integrals by means of the saddle point method and obtained the leading term of the solution in the form of an integral over an Airy function. Explicit use of Bloch-Floquet

functions was made by J. A. Krumhansl.[3] It was found that, whereas for a homogeneous medium a square wave pressure pulse propagates as an unattenuated disturbance with a sharp discontinuity at its front, in a composite the disturbance is dispersed, the maximum stress greatly reduced, and the front smoothed out. The solutions obtained by Peck and Gurtman apply after long times near the front edge of the disturbance and are called head of the pulse solutions. An illustration is shown in Fig. 1.

An inspection of the Peck and Gurtman solution in the region of the head of the pulse shows that the disturbance there is composed of very long wavelength normal modes. This leads one to look for a general theory of the propagation of long wavelength disturbances in composite materials.

In recent years there have been considerable efforts devoted to the development of approximate long wavelength theories for elastic waves in composites.[4,5,6] Here we shall put forward a compact and formally exact form of such a theory. The present paper deals with a one dimensional composite, a sequel will deal with the three-dimensional case.

The results are quite easily stated, and a direct extension of the theory of wave propagation in a uniform medium. In the latter case, in one dimension, the displacement, u(x,t), satisfies the wave equation

$$\left(\frac{\partial^2}{\partial t} - c \frac{\partial^2}{\partial x^2}\right) u(x,t) = 0, \qquad (1.1)$$

where c is the sound velocity.

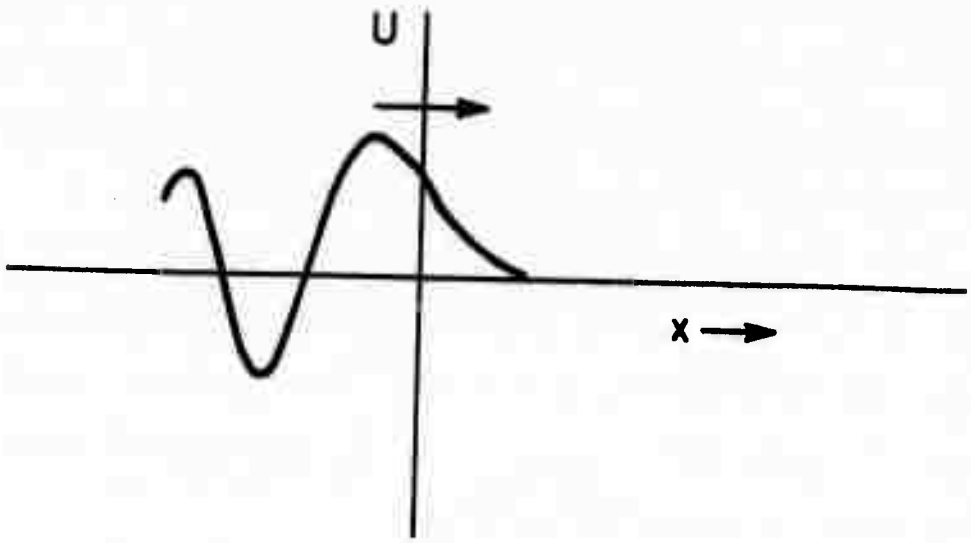


Figure 1. Schematic representation of the displacement in the head of a pulse. The characteristic length in this pattern is much longer than the length of the unit cell, provided the front is far from the origin of the pulse.

Calling the stiffness η, the strain and stress are given by

$$\varepsilon = \frac{\partial u}{\partial x}, \qquad \sigma = \eta \frac{\partial u}{\partial x} \quad . \tag{1.2}$$

A long wavelength disturbance is a composite is described as follows: The displacement at a point is given by

$$u(x,t) = (1 + v_1(x) \frac{\partial}{\partial x} + \ldots)U(x,t); \tag{1.3}$$

here the "envelope function" $U(x,t)$ satisfies the wave equation

$$(\frac{\partial^2}{\partial t^2} - \bar{c}^2 \frac{\partial^2}{\partial x^2} - \beta \frac{\partial^4}{\partial x^4} + \ldots)U(x,t) = 0 \tag{1.4}$$

where $\bar{c}$ is the long wavelength sound velocity of the composite and β is the fourth order expansion coefficient of the square of the frequency of the lowest band of eigenmodes:

$$\omega_1^2(k) = \bar{c}^2 k^2 - \beta k^4 + \ldots \tag{1.5}$$

If the composite has a unit cell length a and a (periodic) stiffness function $\eta(x)$, the function $v_1(x)$ is given by

$$v_1(x) = \int_{a/2}^{x} \left(\frac{\eta(x)^{-1}}{\overline{\eta^{-1}}} - 1 \right) dx \tag{1.6}$$

where

$$\overline{\eta^{-1}} = \frac{1}{a} \int_0^a \eta^{-1}(x)\, dx \tag{1.7}$$

The local strain is given by

$$\frac{\partial u(x,t)}{\partial x} = \frac{\partial U(x,t)}{\partial x} \cdot \frac{\eta(x)^{-1}}{\overline{\eta^{-1}}} + \ldots \tag{1.8}$$

and the local stress is

$$\sigma = \eta(x) \frac{\partial u}{\partial x} = \frac{\partial U(x,t)}{\partial r} \cdot \frac{1}{\eta^{-1}} + \ldots \qquad (1.9)$$

$v_1(x)$ is a periodic function of x and vanishes at the points $(n + 1/2)$ a. Therefore, $U(x,t)$ describes the displacement at the unit cell mid-points $(n + 1/2)$ a. In order for the expansions in (1.3) and (1.4) to converge well, U must be a slowly varying function of n; this is the long wavelength (or low frequency) assumption of the present theory.

We note the following characteristic features of the solution (1.3). It is the product of two terms. One of these, $U(x,t)$, is a slowly varying function of x and t. It satisfies a generalized wave equation, (1.4), whose coefficients are independent of x and t and derived entirely from the dispersion relation $\omega_1(k)$ of the lowest band of normal modes. $\omega_1(k)$ can be obtained experimentally, from a study of normal modes, or by variational or other theoretical methods.[7] The wave equation (1.4) must be solved, subject to whatever the initial and boundary conditions of the problem at hand are. The second factor, $(1 + v_1(x) \frac{\partial}{\partial x} \ldots)$ carries the information about strains and stresses on the scale of the microstructure of the composite. The function $v_1(x)$, defined in (1.6) and (1.7), depends entirely on the static elastic behavior of the material.

Thus, to obtain detailed information about the local displacements, strains and stresses associated with a low frequency

disturbance two steps are necessary: (1) the homogeneous wave equation (1.4) must be solved; (2) information about the static elastic behavior of the material must be introduced (either from theory or experiment). This convenient division, which is here demonstrated for one-dimensional composites, remains valid in three dimensions.

With the present theory any low frequency phenomenon in composites can be discussed. The limitation is given by $ka << \pi$, or equivalently, by $\omega << \pi\bar{c}/a$.

In Sec. 2 the theory just summarized will be developed. In Sec. 3 it is applied, by way of example, to the propagation of a pulse in a semi-infinite composite.

2. The Low Frequency Expansion

We consider an infinite one-dimensional composite, of periodically varying density, $\rho(x)$, and stiffness, $\eta(x)$, with spatial periodicity a. The wave equation for the displacement $u'x,t)$ in an elastic disturbance is

$$\rho(x) \frac{\partial^2 u(x,t)}{\partial t^2} - \frac{\partial}{\partial x}\left[\eta(x) \frac{\partial u(x,t)}{\partial x}\right] = 0 \quad . \tag{2.1}$$

The normal modes of vibration have the form

$$u(x,t) = u(x,k)e^{-i\omega t} \quad . \tag{2.2}$$

The function $u(x)$, is quasi-periodic (or Floquet-Bloch like), i.e.,

$$u(x + a,k) = u(x,k)e^{ika} \quad , \tag{2.3}$$

where k is the wave number. It can therefore also be written in the form

$$u(x,k) = v(x,k)e^{ikx} \tag{2.4}$$

where v(x,k) is strictly periodic,

$$v(x + a,k) = v(x + a) \quad . \tag{2.5}$$

Thus we have finally

$$u(x,t) = v(x,k)e^{i(kx-\omega t)} \quad . \tag{2.6}$$

Substituting this form in (2.1) gives the following equation for v(c,k)

$$\left[\left(\eta \frac{d^2}{dx^2} + \frac{d\eta}{dx}\frac{d}{dx}\right) + ik\left(2\eta \frac{d}{dx} + 2\frac{d\eta}{dx}\right) + (ik)^2\eta\right]v(x,k)$$

$$+ \rho\omega(k)^2 v(x,k) = 0 \quad , \tag{2.7}$$

which together with the periodic boundary conditions completely determines v(x,k) (to within a factor) and $\omega(k)^2$. For small k we now make the following power series expansions:[8])

$$v(x,k) = 1 + (ik)\, v_1(x) + (ik)^2 v_2(x) + \ldots \tag{2.8}$$

and

$$\omega(k)^2 = \bar{c}^2k^2 - \beta k^4 + \ldots \tag{2.9}$$

Substitution into (2.7) and collection of powers of $(ik)^2$ leads to the following equations

$$(\eta \frac{d^2}{dx^2} + \frac{d\eta}{dx}\frac{d}{dx})\ 1 = 0 \tag{2.10}$$

$$(\eta \frac{d^2}{dx^2} + \frac{d\eta}{dx}\frac{d}{dx})\ v_1(x) + \frac{d\eta}{dx} = 0 \tag{2.11}$$

$$(\eta \frac{d^2}{dx^2} + \frac{d\eta}{dx}\frac{d}{dx})\ v_2(x) + (2\eta \frac{d}{dx} + \frac{d\eta}{dx})\ v_1 + \eta(x) + \eta(x)\bar{c}^2 = 0$$

etc. (2.12)

An arbitrary low frequency disturbance can be written as

$$u(x,t) = u^+(x,t) + u^-(x,t) \tag{2.13}$$

where

$$u^{\pm}(x,t) = \frac{1}{2\pi}\int_{\infty}^{\infty} dk A^{\pm}(k) v(x,k) e^{i(kx \pm \omega(k))t} \tag{2.14}$$

and

$$\omega(k) = +(\omega(k)^2)^{\frac{1}{2}} . \tag{2.15}$$

The function A(k) is assumed to be significant only for small values of k. In this case we can use the expansion (2.8) and write

$$u^{\pm}(x,t) = \frac{1}{2\pi}\int dk A^{\pm}(k)(1+(ik)v_1(x) + \ldots)e^{i(kx\pm\omega(k)t)}$$

$$= (1+v_1(x)\frac{\partial}{\partial x} + \ldots)U^{\pm}(x,t) \tag{2.16}$$

where

$$U^{\pm}(x,t) = \frac{1}{2\pi}\int A^{\pm}(k) e^{i(kx\pm\omega(k)t)} \tag{2.17}$$

Using the expansion (2.9) for $\omega(k)^2$, we see immediately that U^+ and U^- satisfy the wave equation

$$\frac{\partial^2 U^{\pm}(x,t)}{\partial t^2} = -\omega^2 \left(\frac{1}{i}\frac{\partial}{\partial x}\right) U^{\pm}(x,t)$$

$$= \left(\bar{c}^2 \frac{\partial^2}{\partial x} - \beta \frac{\partial^4}{\partial x^4} + \ldots\right) U^{\pm}(x,t) \quad . \tag{2.18}$$

Hence the displacement in a low frequency disturbance has the form

$$u(x,t) = \left(1+v_1(x)\frac{\partial}{\partial x} + \ldots\right) U(x,t) \quad . \tag{2.19}$$

Here $U(x,t)$ satisfies the same equation (2.18) as $U^{\pm}(x,t)$. The function $v_1(x)$, by Eq. (2.11) satisfies

$$\frac{d}{dx}\left[\eta \frac{dv_1}{dx} + \eta\right] = 0 \tag{2.20}$$

which, together with the periodicity property

$$v_1(x+a) = v_1(x) \tag{2.21}$$

results in

$$v_1(x) = \int_{a/2}^{x} \left(\frac{\eta^{-1}(x)}{\overline{\eta^{-1}}} - 1\right) dx, \tag{2.22}$$

where

$$\overline{\eta^{-1}} \equiv \int_0^a \eta^{-1}(x)\, dx/a \tag{2.23}$$

Since $v\,((n + \tfrac{1}{2})a) = 0$, it follows by (2.19) that in the middle of each unit cell, to this order of approximation,

$$u(x,t) = U(x,t) \qquad \text{at } x = (n + \tfrac{1}{2})a \quad . \tag{2.24}$$

From the form (2.19) we can now immediately determine local displacements, strains and stresses. The leading term of the displacement is simply

$$u(x,t) = U(x,t) + \ldots \tag{2.25}$$

The leading term of the local strain is

$$\frac{\partial u}{\partial n} = (1 + \frac{dv_1}{dx}) \frac{\partial U}{\partial x} + \ldots \tag{2.26}$$

$$= \frac{\eta^{-1}(x)}{\overline{\eta^{-1}}} \frac{\partial U}{\partial x} + \ldots \tag{2.27}$$

i.e., the local strain is proportional to the strain of the envelope function U and inversely proportional to the local stiffness. It is a rapidly varying function of x. The local stress is given by

$$\sigma = \eta(x) \frac{\partial u}{\partial x} = \frac{1}{\overline{\eta^{-1}}} \frac{\partial U}{\partial x} \tag{2.28}$$

and is a slowly varying function of x. However, the absence of stress variations on a microscale is perculiar to one dimension where a macroscopic compression produces a uniform stress.

3. An Illustration: Pulse Propagation in a Half Space

Consider long wavelength disturbances in a one dimensional semi-infinite ($x \geq 0$) composite. The end is considered as free, hence we have the zero stress boundary condition (see Eq. (2.27)),

$$\frac{\partial U(x,t)}{\partial x}\Big|_{x=0} = 0 \quad . \tag{3.1}$$

In addition we specify the initial conditions, at $t = 0$,

$$U(x,0) = F_1(x) \qquad x \geq 0 \tag{3.2}$$

and

$$\frac{\partial U(x,0)}{\partial t} = F_2(x) \qquad x \geq 0 \quad . \tag{3.3}$$

We now look for the disturbance $U(x,t)$ at subsequent times.

This half space problem is easily converted to a problem over the entire space by adding to (3.2) and (3.3) the supplementary initial conditions

$$U(x,0) = F_1(-x) \qquad x \leq 0 \tag{3.2'}$$

and

$$\frac{\partial U(x,0)}{\partial t} = F_2(-x) \qquad x \leq 0 \quad . \tag{3.3'}$$

As we shall verify, the boundary condition (3.1) is then automatically satisfied.

To solve the new problem in the whole space, we now introduce two Green's functions, $G_1(x-x';t)$ and $G_2(x-x';t)$ which satisfy the wave equation

$$\left[\frac{\partial^2}{\partial t^2} - \omega^2\left(\frac{\partial}{\partial x}\right)\right] G_i(x-x';t) = 0, \; i = 1,2; t > 0 \tag{3.4}$$

and the following initial conditions:

$$G_1(x-x',0) = \delta(x-x') \tag{3.5a}$$

$$\frac{\partial G_1(x-x',t)}{\partial t}\Big|_{t=0} = 0, \tag{3.5b}$$

and

$$G_2(x-x',0) = 0 \tag{3.6a}$$

$$\frac{\partial G_2(x-x',t)}{\partial t}\Big|_{t=0} = \delta(x-x') \quad . \tag{3.6b}$$

Then the solution of the original problem, in the half space $x \geq 0$, is given by

$$U(x,t) = \int_{-\infty}^{\infty} G_1(x-x';t)\, F_1(x')dx'$$

$$+ \int_{-\infty}^{\infty} G_1(x-x';t)\, F_2(x')dx' \tag{3.7}$$

as may be immediately verified.

We obtain the Green's functions by Fourier analysis. Beginning with $G_1(x-x',k)$ we write

$$G_1(x;t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk[A_1^{+}(k)e^{i(kx+\omega(k)t)}$$

$$+ A_1^{-}(k)e^{i(kx-\omega(k)t}] \quad . \tag{3.8}$$

The initial condition (3.5b) gives

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} dk\ \omega(k)(A_1^{+}(k) - A_1^{-}(k))e^{ikx} = 0 \quad , \tag{3.9}$$

so that

$$A_1^+(k) = A_1^-(k); \tag{3.10}$$

and (3.5a) given

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} dk\,(A_1^+(k) + A_1^-(k)e^{ikn} = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\; e^{ikx} \quad , \tag{3.11}$$

resulting in

$$A_1^+(k) + A_1^-(k) = 1 \quad . \tag{3.12}$$

Combining (3.10) and (3.12) yields

$$A_1^+(k) = A_1^-(k) = \frac{1}{2} \tag{3.12}$$

and

$$G_1(x;t) = \frac{1}{4\pi}\int_{-\infty}^{\infty} dk[e^{i(kx+\omega(k)t)} + e^{i(kx-\omega(k)t)}] \; . \tag{3.13}$$

For long wavelength problems, it is evidently sufficient to expand $\omega(k)$ giving

$$G_1(x;t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\; \cos[(x-\bar{c}t)k+(\frac{\omega'''}{6}t)k^3]$$

$$+ \cos[-(x+\bar{c}t)k+(\frac{\omega'''}{6}t)k^3] \tag{3.14}$$

$$= \frac{1}{2(\frac{\omega'''t}{2})^{1/3}}\left\{Ai[\frac{x-\bar{c}t}{(\frac{\omega'''t}{2})^{1/3}}] + Ai[\frac{-(x+\bar{c}t)}{(\frac{\omega'''t}{2})^{1/3}}]\right\}$$

where Ai is the Airy function,[11] defined as

$$Ai\left(\frac{x}{(3a)^{1/3}}\right) = \frac{(3a)^{1/3}}{\pi}\int_0^{\infty} \cos(at^3 + xt)dt \quad . \tag{3.15}$$

The calculation of G_2 is entirely analogous. We expand

$$G_2(x;t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk[A_2^{+}(k)\ e^{i(kx+\omega(k)t)} + A_2^{-}(k)\ e^{i(kx-\omega(k)t)}] \qquad (3.16)$$

and from the initial conditions (3.7a) and (3.7b) we find

$$G(x;t) = \int_0^t dt\ \frac{1}{2(\frac{\omega'''t}{2})^{1/3}} \left\{ Ai[\frac{x-\bar{c}t}{(\frac{\omega'''t}{2})^{1/3}}] - Ai[\frac{-(x+\bar{c}t)}{(\frac{\omega'''t}{2})^{1/3}}] \right\} . \qquad (3.17)$$

With these Green's functions any initial value problem for a composite half space with free boundary is now explicitly solved by Eq. (3.7). We would like to emphasize that even if in the initial distributions $F_1(x)$ and $F_2(x)$ (Eqs. (3.2) and (3.3)) short wavelength Fourier components, which violate the condition $ka << \pi$, are included, the solution (3.7) calculated with the approximate Green's functions (3.14) and (3.17), will - because of the linearity of the problem - correctly describe the long wavelength components of the solution. Thus, for example, if $F_1(x)$ and $F_2(x)$ have step function character, the head of the resulting pulse is still correctly described by the present long wavelength theory.

Acknowledgements

This research was supported by the Advanced Research

Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253. I would like to thank Professors J. A. Krumhansl and E. H. Lee for many stimulating and helpful discussions.

References

1. Two recent general reviews, containing extensive lists of references, are (a) J. C. Peck, in "Shock Waves and the Mechanical Properties of Matter", J. J. Burke and V. Weiss, Eds., Syracuse University Press, 1971; and (b) J. C. Peck, "Dynamics of Composite Materials, American Soc. of Mech. Engineers, 1972; a very recent experimental paper, including theoretical analysis and back references in (c) C. D. Lindergren and D. S. Drumheller, in ref. 1b.

2. J. C. Peck and G. A. Gurtman, J. Appl. Mech. 36, 478 (1969).

3. J. A. Krumhansl, Prelim. Reports, Memoranda and Technical Notes of ARPA Materials Summer Conference, Dept. of Chem. and Met. Engineering, University of Michigan, p. 176, 1970.

4. C. T. Sun et al., J. Appl. Mech. 35, 467 (1968).

5. J. D. Achenback and C. T. Sun, in ref. 1b.

6. G. A. Hegemier, in Ref. 1b.

7. For the system of one-dimensional slabs the exact dispersion relation has been given by C. T. Sun et al., J. Appl. Mech. 35, Trans. ASME 90, Series E, p. 467 (1968). General variational methods are discussed by W. Kohn et al., J. Appl. Mech. (to be published) and S. Nemat-Nasser (to be published).

8. See, for example, J. Callaway, *Energy Band Theory*, Ch. 2.7, Academic Press, New York and London, 1964.

9. Actually the lower limit in the integral (2.22) is arbitrary. Different choices of this limit are automatically compensated for, in the low frequency limit, by a slightly different choice of U.

10. M. Abramovitz and I. A. Stegun, Dover Publications, Inc., New York (1964), *Handbook of Mathematical Functions*, p. 446; L. D. Landau and E. M. Lifshitz, *Quantum Mechanics*, p. 491, Pergamon Press Ltd., 1958.

IT'S A RANDOM WORLD

J. A. Krumhansl

Abstract

Random semiconductors, alloys, and magnetic materials have recently received considerable attention by solid state physicists, and a variety of theoretical approaches have been developed. However, these developments are only part of an extensive history of randomness in physical situations. This paper is a report on some of the related history in other fields where randomness has been an ingredient; a basic bibliography is provided.

IT'S A RANDOM WORLD

J. A. Krumhansl

I. Introduction

It is a random world; in almost any physical, chemical, or biological system -- indeed even geographical and sociological interactions -- one can find that randomness in either the parameters, or in the structure, is an essential element of important situations.

Physicists have recently devoted considerable attention to the properties of random semiconductors[1], alloys[2], and magnetic[3] materials -- primarily from the quantum mechanical point of view. The properties of liquid metals have been discussed theoretically[4].

However, the list of statistical applications which one finds in the literature is truly impressive. Several major reviews have recently been written; the book, "Statisitical Continuum Theories," by M. J. Beran[5], and the review article, "Wave Propagation in Random Media," by U. Frisch[6] will give the reader extensive bibliographical compilations of studies of the properties and dynamics in random systems, while the review, "Some Fundamental Ideas in Topology and their Application to Problems in Metallography" by Lida K. Barrett and C. S. Yust[7]

will introduce the reader to an extensive literature on structural questions in random arrays.

I have not yet found any way to neatly categorize the structure of the subject of randomness, although a few regularly recurring themes will be discussed in the next sections. But to give some idea of the range of problems which have been looked at, here are some examples*: In geophysical applications -- the thermal, mechanical, and electrical properties or randomly polycrystalline rocks and minerals; the pore and fissure structure; the seismological characteristics of waves in geological matter; wave motion in the sea, and its power spectrum; underwater sound propagation in water with random density and temperature variations; turbulence in the sea, atmosphere, and ionosphere; the twinkling of stars; radiowave propagation in the above random media; radiative and convective transfer in astrophysical applications. In materials science -- the structure and properties of heterogeneous materials, such as polycrystalline commercial alloys or sintered ceramics; flow through randomly porous media; the properties of reinforced composite materials; wood, polymers, paper.

In biological science I have not made any extensive survey, but at the very least the structure of plant and animal tissue immediately provides the same kinds of situations encountered above in materials and geophysical situations; in addition, the

*Detailed references are not given except where references are not to be found in the reviews cited above.

biochemical and genetic processing in such structures must inevitably reflect consequences of randomness; similarly, while it is one thing to model biological or human population situations with averaged parameters, since many of these descriptions are non-linear the introduction of randomness can lead to entirely new qualitative behavior[8]. In modern communications technology, stochastic problems abound; the response of linear and non-linear electrical networks to random noise signals is a central issue, which has led to an extensive mathematical development; of course, fading and the like in radiowave propagation has long been recognized as having essential stochastic characteristics. In reactor technology, neutron transport is a stochastic diffusion process and is therefore a subject which has produced an extensive literature on transport in stochastic systems. Partial coherence in either acoustic or optical wave applications (holography, etc.) is of considerable current interest. And then there is the vast subject of equilibrium and non-equilibrium statistical mechanics. Undoubtedly this list is far from complete, but it should give some idea of the ever widening spiral in which the writer has been led by an early summer's fancy that it might be a good idea to browse the world of randomness.

II

There does seem to be one methodology which has been useful in a wide variety of instances, and which is remarkable in regard to the number of independent rediscoveries of essent-

ially the same approach. In solid state physics, the most recent examples[2] have been "Coherent Potential Approximation" (CPA) or "Self Consistent" methods in alloys. Previous or parallel developments in other contexts of this method are the subject of this section.

It is frequently possible to divide the physical problem in a random system into two parts -- an average behavior, plus the fluctuations from it. An exact treatment must keep both, and when the latter become comparable to the former, the separation is meaningless. Fortunately, there are many situations in which the separation is both useful and meaningful. In this case the "self consistent local field theory" is quite useful for describing the average field; however, to complete sum rules (energy conservation, for example) the fluctuations must be included in principle.

Mean field theories have usually been invented as needed for the application at hand. A. Einstein[9] was concerned with the properties of a suspension of small particles in his classic Brownian motion studies. Lord Rayleigh[10] and many others[11] have since addressed the question of the average permativity of a heterogeneous medium, both in the static and wave propagation regime.

The essence of the standard problem is relatively simply stated. The field (electromagnetic, elastic, Schrödinger, etc.) denoted by u, obeys a linear equation of motion characterized by an operator $\hat{L}$, which in turn is parameterized by specification

of the properties of the medium. In a homogeneous medium, $\hat{L}$ depends only on a few parameters, while in a random medium these parameters (conductivity, potential, index) vary from place to place according to some probability distribution; for brevity we denote the random parameter as ε in $\hat{L}(\varepsilon)$. The equation of motion is

$$\hat{L}(\varepsilon)\, u = 0 \tag{1}$$

The operator $\hat{L}$ may also depend on other important parameters which are not random (e.g., E in $\hat{H}$ - E for Schrödinger waves, ω^2 in acoustic and electromagnetic waves). The essence of the averaging methods is to define $\hat{L} = \langle\hat{L}\rangle + \hat{L}_1$, and attempt to calculate $\langle u\rangle$ or other quantities such as $\langle\varepsilon u\rangle$, and from them effective parameters $\varepsilon^* = [\langle\varepsilon u\rangle/\langle u\rangle]$. This may be done in various approximations, both self-consistent or not. When stated in this way a variety of methods[6] from the literature of perturbation theory in applied mathematics, as well as from field theoretical methods in physics may be applied. It is feasible to obtain the "self consistent local field theory" as the best low order local approximation to a resummed perturbation series; it also can be obtained by choosing a local field such that scattering vanishes to first order in the "concentration" of defects, i.e., all repeated scatterings by $\hat{L}_1$ at a particular position. One may then compute an average ε^*, as above, to obtain an effective medium parameter.

Neither time nor space allow a detailed comparison of various formalisms to be made here, but we may illustrate by the calculation of the dielectric constant of a composite medium consisting of small spheres of dielectric ε_2 imbedded in ε_1 which are assumed to be randomly located, and non-overlapping. The "equation of motion" is

$$[(\underline{\nabla} \cdot \varepsilon(\underline{r}))\ \underline{E}(r)] = 0 \tag{2}$$

in the absence of free charges. The field $u \to \underline{E}(\underline{r})$ and $\hat{L} \to \underline{\nabla} \cdot \varepsilon(\underline{r})$. If an applied charge $\rho(r)$ is imposed then it is useful in solving the problem to have the Green's function $\hat{G} = (\underline{\nabla} \cdot \varepsilon(\underline{r}))^{-1}$ which satisfies $\hat{L}\,\hat{G} = \delta(\underline{r} - \underline{r}')$. Indeed, a systematic method for solving may now be based on writing $\hat{L} = \langle\hat{L}\rangle + \hat{L}_1 = \hat{L}_0 + \hat{L}_1$, whence the operator relation holds that

$$\hat{G} = \hat{G}_0 + \hat{G}_0\,\hat{L}_1\,\hat{G} \tag{3}$$

Further there is an exact scattering representation for fields

$$\underline{E} = \underline{E}_0 - \hat{G}\hat{L}_1\,\underline{E}_0 = \underline{E}_0 + \underline{E}_{scat.} \tag{4}$$

where $\underline{E}_0$ is any field which satisfies $\langle\hat{L}\rangle\underline{E}_0 = 0$. Equation (4) can be used to define an ε^*. In the example at hand $\varepsilon = \varepsilon_2$ within the spheres, ε_1 in the host; thus $\hat{L}_i = \sum_i (\varepsilon_2 - \varepsilon_1)$ index i running over spheres.

The specific practical problem for a dielectric function is for some direction, say x: $\varepsilon^*_{xx} = [\langle\varepsilon E_x\rangle/\langle E_x\rangle]$. Here, the

averaging is volume averaging.

The basic scattering problem is to compute the scattering around a sphere in the presence of some local field, assumed to be dominated by the volume averaged uniform field. In fact, it is tedious but not difficult to solve this problem using a Green's function G_0; on the other hand the exact fields around a single sphere or ellipsoid can easily be calculated using spherical harmonics. When this is done, if we take spheres to scatter independently except for each to contribute to a volume average field, then

$$\frac{E_x}{E_{xo}} = 1 + \frac{\varepsilon_1 - \varepsilon_2}{2\varepsilon_1 + \varepsilon_2}\frac{a^2}{r^3}(1-3\cos^2\theta) \qquad \text{outside a sphere}$$

$$= 1 + \frac{\varepsilon_1 - \varepsilon_2}{2\varepsilon_1 + \varepsilon_2} \qquad \text{inside a sphere} \qquad (5)$$

$$= 1 + (E_{scat.}/E_{ox})$$

If a volume average of $\langle \varepsilon E_x \rangle$ $\langle E_x \rangle$ is now taken, and f_2 is the volume fraction occupied by ε_2, the integrations yield

$$\varepsilon^*_{xx} = \frac{\langle \varepsilon \rangle + f_2 \varepsilon_2 \left(\frac{\varepsilon_1 - \varepsilon_2}{2\varepsilon_1 + \varepsilon_2}\right)}{1 + f_1 \left(\frac{\varepsilon_1 - \varepsilon_2}{2\varepsilon_1 + \varepsilon_2}\right)} \qquad (6)$$

This expression is not self consistent, since the field incident on the spheres was assumed to be E_0 in a medium of $\varepsilon = \varepsilon_1$. For the moment, however, expanding (6) to lowest order in $\varepsilon_2 - \varepsilon_1$ which would apply to small dielectric fluctuations one obtains

an expression obtained by various means[11,12]:

$$\varepsilon^*_{xx} \simeq \langle\varepsilon\rangle - \frac{f_2(\varepsilon_1 - \langle\varepsilon\rangle)^2}{3\langle\varepsilon\rangle} + \qquad (7)$$

A study of the methods shows that (6) is equivalent to the "Average T-matrix Approximation" (ATA) used in random alloys[2], while (7) is a low order correction to the virtual crystal approximation.

In order to make the method totally symmetric in ε_1 and ε_2, as well as self consistent, we take the medium to be describable by an average ε^* and then, when in medium 1 scatter by $\varepsilon_1 - \varepsilon^*$, when in 2 by $\varepsilon_2 - \varepsilon^*$. With the assumptions statistical independence and volume averaging one obtains the implicit equation for ε^*_{xx} (after much algebra):

$$1 = \sum_i \frac{3f_i}{2 + (\varepsilon_i/\varepsilon^*_{xx})} \qquad (8)$$

On the other hand, taking volume average x components only, the scattering problem (neglecting interparticle scatterings) in an effective medium ε^*_{xx} looks like:

$$\langle E_x\rangle = \langle E_{ox}\rangle + \left(\sum_i \frac{\varepsilon^*_{xx} - \varepsilon_i}{2\varepsilon^* + \varepsilon_i} f_i\right)\langle E_{ox}\rangle + \qquad (9)$$

If we take what at first sounds like a physically different condition from the previous, which is that the first order scattering shall vanish on the average

$$\sum_i \left(\frac{\varepsilon^*_{xx} - \varepsilon_i}{2\varepsilon^*_{xx} + \varepsilon_i}\right) f_i = 0 \qquad (10)$$

then since $\sum f_i = 1$, it may be found that (10) and (8) are exactly the same. Thus, the self consistent local field is equivalent to the choice of an effective medium such that the averaged single site scattering vanishes. This is identical to the CPA method in alloys[2], and we now pass on to cite a number of historically parallel developments in different contexts. However, to summarize, the essence of the method is to solve for the exact field around a representative element of the medium, which is taken to be imbedded in an effective medium determined in turn by requiring that the mean of the scattering by the random elements, taken as statistically independent, shall vanish.

Here are a number of examples where the same philosophy has been developed in one formalism or another:

Yonezana[13] and Leath[14] provide a diagrammatic basis for the self-consistent method (CPA) as applied to the quantum mechanics of random alloys.

Landauer[15], in 1952, analyzed the conductivity of a random mixture of different materials, and derived an expression which is simply Equation (10) above, with conductivities σ substituted for ε. Comparisons were made with experiment.

The thermal properties of random composites have received considerable theoretical and experimental attention. Important contributions to the subject were made by Kerner[16], and more recently by Budiansky[17]. A brief review of experimental and theoretical work has been given.[18]

The understanding of elastic properties of a random

(polycrystalline) aggregate is a long standing subject of concern in many subjects -- geology, agronomy, civil engineering, continuum mechanics, and ceramics, to mention a few. Here again the "self-consistent" method developed, with original contributions to the subject by Budiansky[17], by Hill[18], and by Kröner[19]; indeed the similarity of Kröner's formalism to that developed for studying electronic properties of alloys is remarkable.

One should, of course, recognize that the elastic fields are tensor quantities so the problem is considerably more complicated than that of Schrödinger scalar waves in random systems, and thus the idealizations of the self-consistent method may be more serious. None-the-less, Thomsen[20] has recently applied Kröner's method to the elasticity of polycrystals and rocks, with some success in comparison with experiment.

It might also be mentioned in passing that an entirely different approach to the properties of heterogeneous materials uses variational methods[21] to set bounds of effective parameters; to the extent that I have checked several cases the self-consistent results fall between the expected upper and lower bounds (e.g., $\langle\varepsilon\rangle$ vs. $\langle\varepsilon^{-1}\rangle^{-1}$ for ε^* in the case of dielectrics).

Another area in which the same philosophy has developed is that of determining an effective propagation constant for the average wave in a random medium. In the context of both acoustic wave propagation and electromagnetic wave propagation a tremendous amount has been done, and reference 6 contains a bibliography with 154 entries up to 1968. Noteworthy in physics, the

work of Foldy (1945)[22] and Lax (1951, 1953)[23] set out the approach which has since evolved into self-consistent methods in various wave propagation applications. While at first it may not seem that the mean static parameters discussed above are closely related to wave propagation, once the wave equation is Fourier transformed, the technical aspects of the formalism are essentially the same, i.e., self-consistent local field theory.

It is difficult to know how to conclude this section, since it is certain that this list of applications is not nearly complete. Perhaps the best message to the reader is that he should also go out and shop, with the above list as a beginning.

III.

In this last section I wish to make a few comments on special topics in the theory of random systems.

First of all, the self-consistent local field theory is suited to random systems only in circumstances where the average (sometimes called "coherent") field is a useful, meaningful quantity. On the other hand, sometimes we have "localized" resonant phenomena, or effects due to clustering, in which case the approach of a self-consistent field has limited value. Is it possible to extend the idea to 2, 3, ... n, site self-consistent local fields? Perhaps, but there has not yet developed a clear basis on which to proceed (although several recent attempts have been made). So I am taking the course of not commenting on them at this writing. The essence of the difficulty is that for

strong scattering there is no suitable parameter to use in collecting the various terms in approximate series expressions. It seems clear that beyond the self-consistent local field approximation, higher order theories have not yet emerged in suitably effective form.

What about the more general problem of the complete description of the behavior of random systems? One sobering fundamental view of the outlook for progress may serve as useful perspective in closing this paper. The model problem discussed above is so easily stated, and indeed a stochastic operator L which depends linearly on random parameters is not all that complicated statistically, so what is the trouble? Probably many workers have recognized the simplistic nature of this view, but Kraichnan[24] and Beran[5] particularly dra attention to the perfectly apparent fact that <u>although the equation of motion may depend only linearly on the random parameters</u>, <u>the solutions are almost invariably highly non-linear functionals</u> of the stochastic quantities -- for whose description there does not exist straightforward methodology. Indeed the general p blem shares the same intractabilities as appear in strong turbulence, strong interaction field theory, and many non-linear systems. Thus, it is likely that we must be content with solutions of special examples for guidance when we wish to go much beyond discussing average fields in random systems.

As I began, let me repeat in closing that it's a random physical world, which on the one hand presents similar problems

in widely different context, but on the other hand offers challenges beyond our common mathematical and conceptual tools. To paraphrase a common expression -- that's good news, or bad news, depending on your taste.

Acknowledgement

This work has received support from the Atomic Energy Commission, the Materials Research Council of the Advanced Research Projects Agency, and the National Science Foundation through the Materials Science Center, Cornell.

References

1. Many aspects of Random Semiconductors are regularly discussed in J. Non-Cryst. Solids.

2. Soven, P., (1967) Phys. Rev. 156, 809; Ibid (1969) 178, 1136. Kirkpatrick, S., Velicky, B., and Ehrenreich, H. (1970) Phys. Rev. B1, 3250.

3. Proc. Intern. Symposium on Amorphous Semiconductors, (1973) to be published, J. Non-Cryst. Solids.

4. Schwarta, L. and Ehrenreich, H. (1971) Annals of Physics 64, 100.

5. Beran, M. J. (1968), *Statistical Continuum Theories*, Interscience, New York; also, (1971) Physica Solidii Stati A6, 364.

6. Frisch, U., (1968), "Wave Propagation in Random Media", in *Probabilistic Methods in Applied Mathematics*, Ed. A. T. Bharucha-Reid, Academic Press, New York.

7. Barrett, Lida K. and Yust, C. S. (1970) Metallography 3, 1.

8. Goel, N. S., Maitra, S. C., Montroll, E. W. (1971), Rev. Mod. Phys. 43, 231.

9. Einstein, A. (1906), *Investigations on the Theory of Brownian Motion*, Dover reprint, New York, 1956.

10. Lord Rayleigh, (1892) Phil. Mag. Ser. 5, 34, 481; Keller, J. B., (1964) J. Math. Phys. 5, 548.

11. Brown, W. ;. (1955), J. Chem. Phys. 23, 1514, Reviews much previous work; see also ref. 5.

12. Landau, L. and Lifshitz, E., *Electrodynamics of Continuous Media*, p. 45, Addison Wesley, New York.

13. Yonezawa, F., (1968), Prog. Theor. Phys. 40, 734.

14. Leath, P. L. (1970), Phys. Rev. B2, 3078.

15. Landauer, R. (1952), J. Appl. Phys. 23, 779.

16. Kerner, E. H. (1956), Proc. Phys. Soc. 69B, 808.

17. Budiansky, B. (1970), J. Composite Materials 4, 286.

18. Hill, R. (1967), J. Mech. Phys. Solids 15, 79.

19. Kröner, E. (1967), J. Mech. Phys. Solids 15, 319.

20. Thomsen, L. (1972), J. Geophys. Res. 77, 315.

21. Hashin, Z. (1964), Appl. Mech. Rev. 17, 1.

22. Foldy, L. L. (1945), Re. Mod. Phys. 23, 287: (1952), Phys. Rev. 85, 621.

23. Kraichnan, R. H. (1962) Roc. Symp. Appl. Math. 13, 199; American Math. Soc.

THEORY OF IONIC TRANSPORT IN CRYSTALLOGRAPHIC TUNNELS

W. H. Flygare and R. A. Huggins

Abstract

A model has been developed for the treatment of the motion of ions through crystallographic tunnels, as are found in materials that are interesting solid electrolytes. Consideration of both point charge and higher order attractive terms as well as overlap repulsion effects allows the calculation of the minimum energy positions of mobile ions and the activation energy barrier that they must surmount to move through the tunnel in the lattice. Calculations have been made for ions of different sizes in the AgI lattice which show that there is a set of minimum energy paths which do not follow the centerline of the tunnel, but deviate periodically, with both direction and magnitude depending upon the cationic size. Also, in accordance with experimental observations, the activation energy for motion is smallest for cations of intermediate size, where the Coulombic, polarization, and repulsive contributions to the total energy are best balanced.

THEORY OF IONIC TRANSPORT IN CRYSTALLOGRAPHIC TUNNELS

W. H. Flygare and R. A. Huggins

Introduction

Many experiments have shown that unusually large values of diffusion coefficient and ionic conductivity are found in a number of materials which have crystal structures that are characterized by the existence of linear or nearly linear tunnels. Under proper conditions certain ions (e.g., silver, copper, and the alkali ions) can readily move along such tunnels under the influence of chemical or electrical forces. The magnitude of the resulting transport fluxes in some electronic insulators makes them particularly interesting because of their potential use as solid electrolytes in new types of batteries or fuel cells.

Several groups of materials are now known in which ionic motion is especially rapid. These include two classes of cubic structures, those in which the anion arrangement is bcc (e.g., α AgI, α Ag_2S, α-Ag_3SI, and $(Na,Li)_2$ SO_4), and some in which the anions are arrayed in an fcc packing (e.g., α-CuI, α-Ag_2HgI_4, and α-Li_2SO_4). There are also several types of linear and layer structures. One important family of the latter are the beta-aluminas, whose nominal formula is $MAl_{11}O_{17}$, where M can be any of a number of different monovalent cations. Several reviews

describing these materials, sometimes called "superionic-conductors", recently appeared[1-3].

One of the interesting aspects of the available data is that there appears to be an optimum size for the ion translating through the tunnel. Both ions significantly smaller than the transverse dimensions of the tunnel, and much larger ions appear to have lower values of mobility than ions of intermediate size. This has recently been illustrated by experiments on the conductivity of a series of ions in the beta alumina structure[4].

The purpose of this paper is to present a model for the calculation of the energy as a function of position of a mobile ion in the tunnel region of a rigid crystal lattice. This model predicts the minimum energy path for the mobile ion, and the activation energy for translational motion is also obtained. The importance of the size of the transporting ion upon both its path and the activation energy for motion are clearly demonstrated. Although the approach that we present here is generally applicable to any of the tunnel-type lattices mentioned above, it will be illustrated by application to the transport of monovalent cations in the AgI system.

The Model

The energy of an ion in a lattice is described by a sum of electrostatic and overlap repulsion terms. The electrostatic term includes the sum over point charge interactions (Madelung sum) and the higher order terms which include the point charge-

induced dipole, point charge-induced quadrupole, dipole-dipole (point charge induced), and dipole-dipole (dispersion) terms. The general summations used in alkali halide lattices have been examined by Quigley and Das[5] to show that the Li^+ substitutional impurity in KCl has an energy minimum off-center along both the <111> and <100> axes as demonstrated earlier by experiment. The diffusion of interstitial lithium, copper, and silver impurities in the diamond lattice has also been discussed by Weiser[6]. Both of these calculations used techniques developed by others[7].

In the present work we will use the following electrostatic energy for the i^{th} point charge ion in a lattice:

$$E = e^2 \sum_j \frac{q_i q_j}{r_{ij}} - \frac{e^2}{2} \sum_j \frac{\alpha_j}{r_{ij}^4} q_i \qquad (1)$$

The sum over j is over all lattice ions, q_i and q_j are the fractions of charge of the mobile ion and fixed lattice charges respectively, α_j is the dipole polarizability of the j^{th} fixed lattice atom (I^-) and r_{ij} is the distance from the mobile ion to the j^{th} lattice ion. The second term in Eq. (1) is the polarization of self-energy of a non-polarizable mobile cation in the lattice of fixed polarizable anions (I^-). We are assuming that the polarizability of the mobile ion is much less than that of the ions in the host lattice and therefore contributions from these terms are ignored.

The orientational dipole-dipole terms,

$$E_{ij} = \frac{\underset{\sim}{\mu}_i \cdot \underset{\sim}{\mu}_j}{(r_{ij})^3} - \frac{3(\underset{\sim}{\mu}_i \cdot \underset{\sim}{r}_{ij})(\underset{\sim}{\mu}_j \cdot \underset{\sim}{r}_{ij})}{(r_{ij})^5} \qquad (2)$$

are neglected. These terms are normally smaller ($\mu = E\alpha$, $E = eq_i/r^2$) than the corresponding scalar self-energies (see Eq. (1)) by a factor of $1/r^2$ for the dominant near-neighbor terms and will therefore also be neglected. This approximation is consistent with the method used by Weiser[6].

The overlap repulsion term between closed-shell ions is given by[8]

$$E_{rep_{ij}} = \beta_{ij} e^{(r_i + r_j - r_{ij})/\rho} \qquad (3)$$

where r_i and r_j are the ionic radii of ions i and j, r_{ij} is the interionic distance, and β_{ij} is a multiplicative factor depending on the nature of the interacting ions[9]. The final energy of the i^{th} mobile ion is

$$E = e^2 \sum_j \frac{q_i q_j}{r_{ij}} - \frac{e^2}{2} \sum_j \frac{\alpha_j q_i}{r_{ij}^4} + \sum_j \beta_{ij} e^{(r_i + r_j - r_{ij})/\rho} \qquad (4)$$

Eq. (4) diverges as $r_{ij} \to 0$ when the cation radius is very small. This is sometimes referred to as the "polarization catastrophy". The energy can be made to converge by making α_j dependent on the interionic distance becoming lower as the ions approach each other or by converting the exponential repulsive term to a $(\frac{1}{r_{ij}})^{12}$ term when the ions are closer than the sum of ionic radii. These refinements are easy to apply, but Quigley and Das[5] have shown that the minimum positions for the substitutional Li^+ im-

purity in KCl were adequately predicted with Eq. (4). Thus, we will use Eq. (4) directly without correcting for the polarization catastrophy at small r_{ij}. This analysis will be valid for $r_i+r_j < r_{ij}$.

We will now proceed to apply this model to a description of the energy of a mobile positive ion in the AgI lattice as a function of its position.

The AgI Structure

Ag^+ ions have unusually high mobilities and diffusion constants in the AgI lattice[10]. The crystal structure of AgI has been given by Strock[11] and is illustrated in Fig. 1. According to Strock, the crystal structure is essentially a body-centered cubic arrangement of I^- ions which are in contact. There are several near-equivalent available positions for the Ag^+ ions, as indicated in the diagram. The large number of energetically near-equivalent and vacant sites for Ag^+ in AgI can be described in terms of locations along the transport path through the crystallographic tunnels. The unusual ionic mobility and large Debye-Waller factors indicate that tunnels of relatively constant potential energy apparently occur throughout the lattice.

The two positions of the I^- ions in the bcc unit cell (see Fig. 1) lead to the following sets of interionic distances, assuming the cation to reside at xyz:

$$R_1 = ((x-Ia)^2 + (y-Ma)^2 + (z-Na)^2)^{\frac{1}{2}}$$

$$R_2 = ((x-a/2-Ia)^2 + (y-a/2-Ma)^2 + (z-a/2-Na)^2)^{\frac{1}{2}}$$

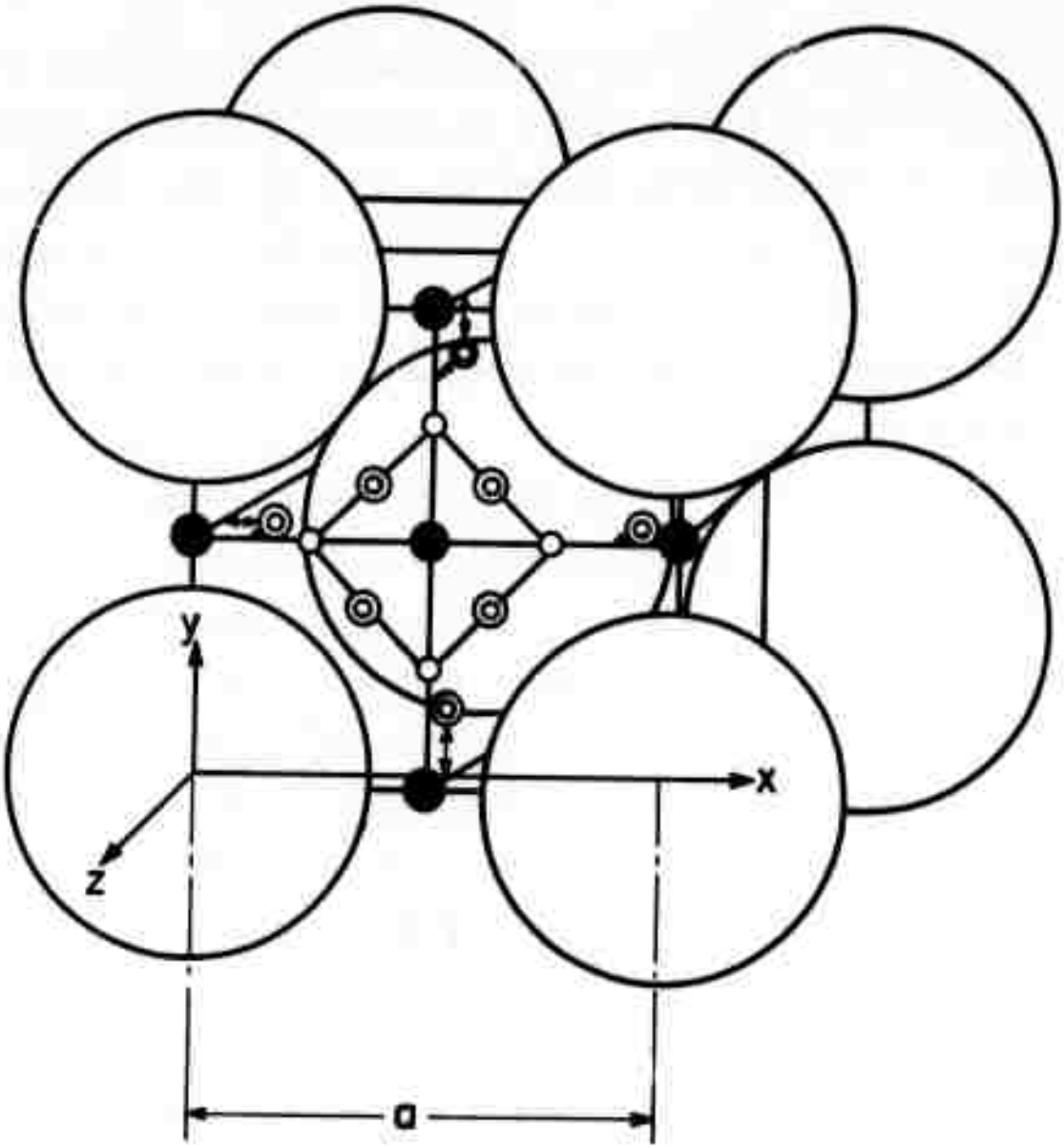


Figure 1: The AgI lattice (10,11). The large ions are I^- and the small ions represent Ag^+ ion positions. There are 2 I^- ions, 6 ● Ag^+ ion positions with two-fold coordination, 12 o Ag^+ ion positions with four-fold coordination, and 24 ◎ Ag^+ ion positions with three-fold coordination in the unit cell.

The energy of an Ag^+ ion was calculated with a digital computer as a function of x, y, and z in the host I^- lattice according to Eq. (4), the R_i values in Eq. (5), $q_i = q_z$ = the electronic charge, $\alpha(I^-) = 6.4\times10^{-24}$ cm^3, $\beta_\pm = 1.0$, $\rho = 0.333$[12] and the repulsion radius for the iodine ion of $r(I) = 1.75\mathring{A}$[13]. I, M, and N are the multipliers of the lattice distance (a = 5.034) along the x, y and z axes.

The calculated energies of a positive ion of radius r_+ as a function of position in the I^- ion bcc lattice for successively larger extensions of the lattice were examined first. We found that dE/dr from a center point in the lattice tunnel (x=0, y=0, z=a/2, $r^2=x^2 + y^2 + z^2$) was constant to 5-6 significant figures if a distance of more than 5 unit cell dimensions were summed along the ±x, ±y, and ±z directions. Of course, the value of E increases with increasing lattice size but dE/dr is constant within the above limits.

The next problem was to account for the presence of the average positive charge distribution caused by the mobile Ag^+ ion distribution in the AgI lattice. Adding in the effects of the average positive charge distribution is equivalent to subtracting the Coulomb contribution in the xy plane for any given value of z from the result obtained using Eq. (4) for a finite lattice. This merely acknowledges that the position of an ion in a lattice is more sensitive to the $1/r^4$ attractive and exponential repulsive terms in Eq. (4) than the corresponding Coulombic terms. This is also in agreement with the results of Quigley and Das[5].

Results

Using the above methods the energy was computed as a function of cation position (and size) along a tunnel parallel to the cube edge in the x-direction passing through the point $x = 0$, $y = 0$ and $z = a/2$. Deviations along the transport path were also probed for various values of y and z near the $y = 0$ and $z = a/2$ centerline. These computations were examined for the minimum energy paths in the crystal as a function of the cation size.

In all cases, the minimum energy paths deviate considerably from the centerline of the tunnel. Both the magnitude and the direction of deviation are dependent upon cation size.

A cation moving down the tunnel passes between pairs of anions with centerlines alternately rotated 90°. In the case of small cations ($r_+ < 0.8\text{Å}$) the polarization term in Eq. (4) dominates the potential energy, causing the cation minimum energy route to split into two equivalent paths which deviate toward the alternating anion pairs. On the other hand the repulsive terms are more important for the larger cations ($r_+ > 0.8\text{Å}$) causing path deviations in a plane normal to the anion-anion pair axes. These effects are shown schematically in Fig. 2.

By finding the lowest energy path in the x-direction, the variation of potential energy during translation through the tunnel can be obtained. Data for a series of cations of different radii are plotted in Fig. 3. It is seen that the periodic variation in the energy has an opposite phase for relatively small and relatively

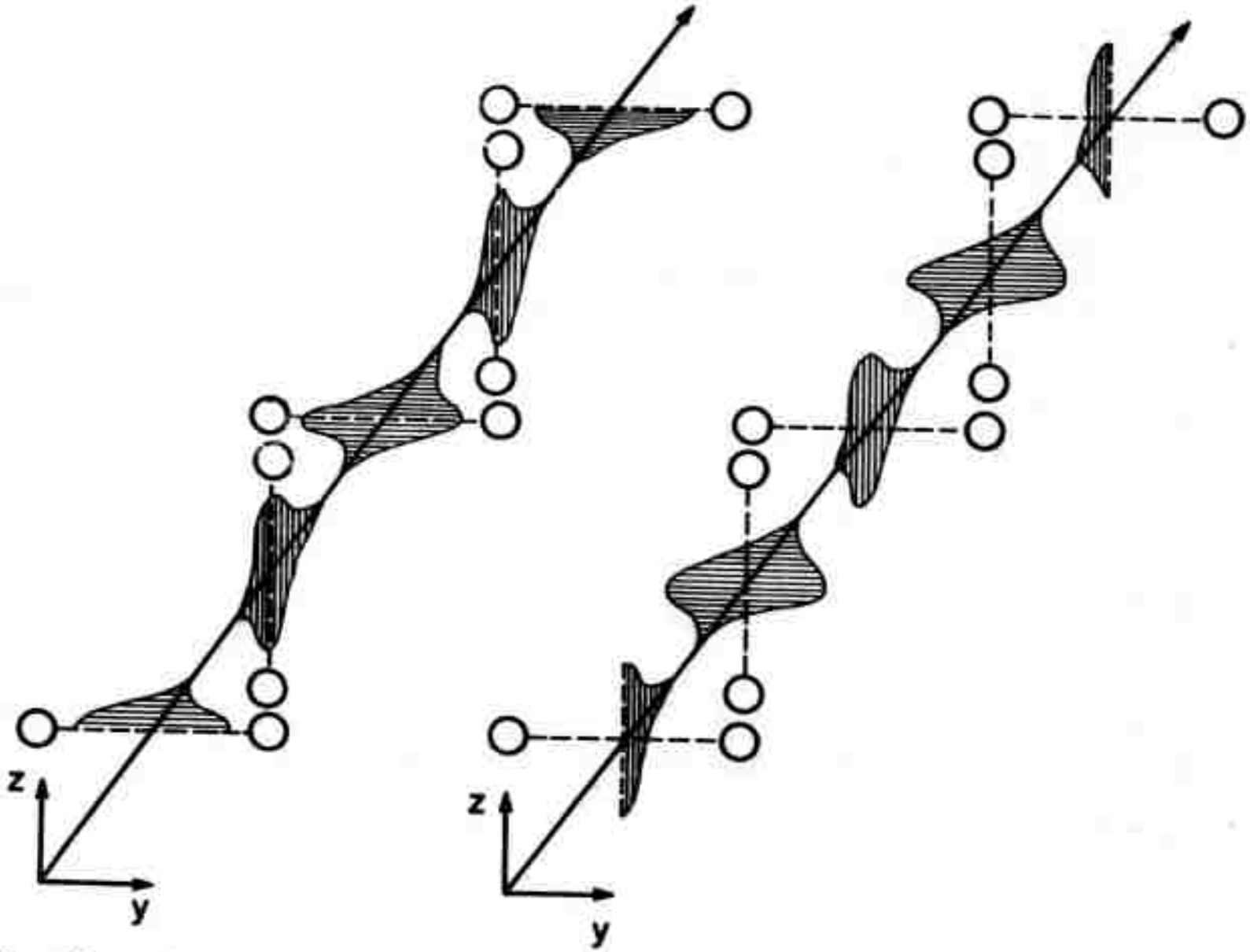


Figure 2: Schematic drawing of loci of minimum energy paths for cations of different sizes; on the left, small ions are attracted toward pairs of I^- ions (indicated by circles), on the right, paths of larger ions are orthogonal to the anion-anion axis.

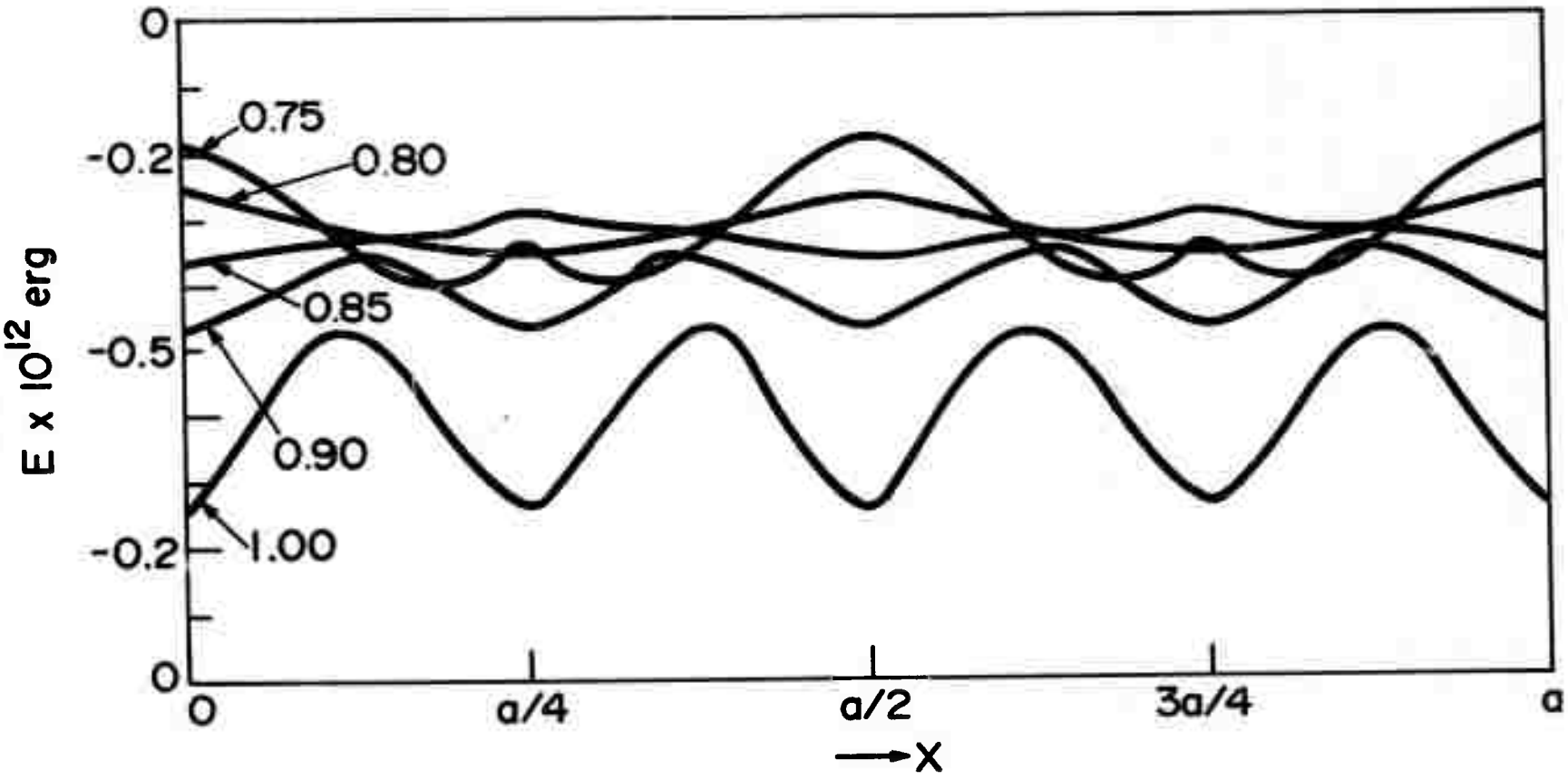


Figure 3: Energy variation as cations of variable radii follow minimum energy paths along the x axis in AgI.

large cations. The amplitudes also increase as the cationic size becomes more extreme.

Both small and large cations must go over substantial energy barriers, due to the dominance of one or the other of the terms in the total energy. In the case of cations of intermediate size, these effects largely balance each other, and the energy varies much less as the ion moves through the tunnel. If one interprets the difference between maxima and minima in energy along the transport path as an activation energy for motion, it is clear why ions of intermediate size have the highest mobility in materials with crystallographic structures containing tunnels of this type.

Values of activation energy calculated for hypothetical cations of several sizes are presented in Fig. 4. It is seen that there is a sharp variation with size, the lowest value being obtained for a cation radius of about 0.83Å.

The influence of the particular values of the constants used in the calculation was also investigated. It was found that, except in the case of very small cations, the results are relatively insensitive to small changes in the anion polarizability. Likewise, only for very large ions are the results influenced substantially by changes in the repulsion constants. Fortunately, ions of intermediate size are of most interest here, so that this is not a significant problem.

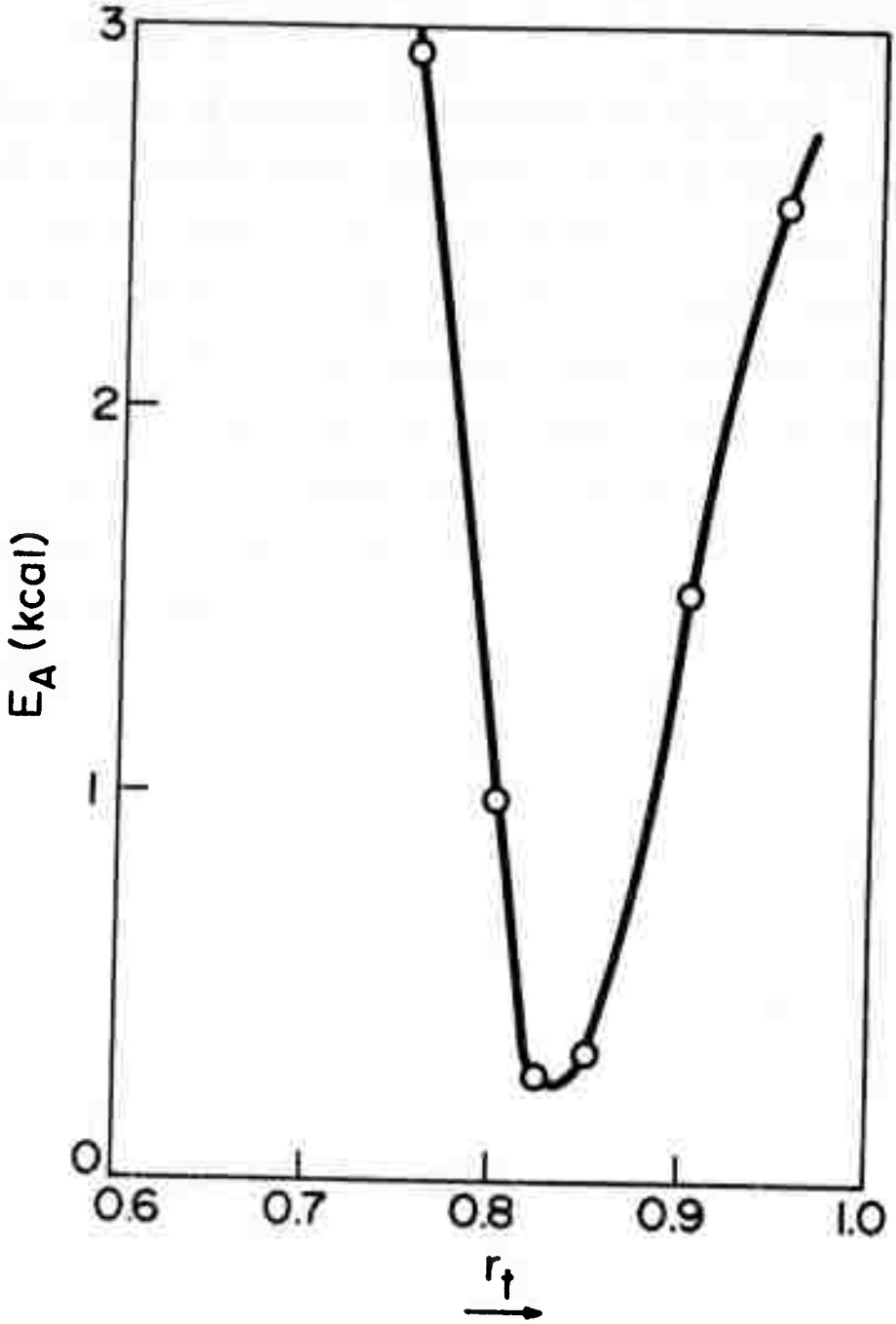


Figure 4: Activation energy for cation transport in AgI as a function of cation radius.

Discussion

The model presented here represents a new approach to the explanation of the unusually large values of diffusion coefficients and ionic conductivities found in certain classes of solids. Materials of this type are all characterized by crystallographic tunnels through which ions can move. It has been shown that the mobility of such ions should be highly sensitive to their size. The minimum energy barrier for transport occurs for intermediate sizes, where the effects of the Coulombic, polarization, and repulsion terms are almost balanced.

In the particular example chosen for calculation, this model predicts that a cation with a repulsion radius between 0.80 and 0.85Å should have the greatest mobility. This may be compared to the experimental observation that silver ions diffuse most readily in this (B 23) structure. Ionic radii vary substantially with coordination in solids[14], and the effective ionic size of silver in materials of this structure is now known, but present ideas[10] lead to a value of ionic radius somewhere between lithium (0.60Å) and sodium (0.95Å).

The values of the activation energy and the large influence of cation radius that result from these calculations compare favorably with the unusually low experimental values for materials with tunnel structures. Examples are 1.48 kcal/mole for AgI[15], 2.8 kcal/mole for the closely related $RbAg_4I_5$[16], and 3.8 kcal/mole for Na^+ and 4.0 kcal/mole for Ag^+ in the beta alumina structure[17,18]. The minimum in activation energy of about 0.25 kcal/mole at r=0.83Å

from Fig. 3, if taken literally, indicates that ions of this size will have an extremely high ionic mobility. Applying 1.48 kcal/mole to Fig. 3 indicates an Ag^+ ion radius near either 0.79 or 0.90 in AgI. In view of the close correspondence between the ionic conductivity of silver and sodium in beta alumina[4], the latter seems a very reasonable value.

Of course we have ignored the non-negligible cation polarizability and a more complete calculation may shift both the position and magnitude of the minimum found in Fig. 3 slightly. However, the general trends observed here will be preserved.

Our results also bear on the question of the preferred site for cations of different sizes in structures in which there are several types of sites that apparently have nearly equivalent values of potential energy.

It is generally assumed that there are three types of possible cation sites in the AgI structure. These occur at the points of two-fold, three-fold, and four-fold coordination as shown in Fig. 1. By comparing Fig. 1 with the positional variation of the energy shown in Fig. 3, it is apparent that smaller cations will tend to have lowest values of energy near the a/8 three-fold coordinated positions. Larger cations will be at lowest energy at the a/4 four-fold coordinated positions.

As has been pointed out elsewhere[10], thermodynamic data make it reasonable to assume that in materials such as $(Na,Li)_2SO_4$, which also have the AgI structure, the two different types of cations do not mix randomly. In light of the results of these

calculations, one would expect that mixtures of cations of different sizes would definitely tend to become ordered among the different sets of sites, due to the influence of ionic size upon preferred location. Related variations in site preference for different monovalent cations have also been found in the beta alumina structure.

Acknowledgement

This research was supported by the Advanced Research Projects Agency under Contract No. DAHC15-71-C-0253 with the University of Michigan. The friendly assistance of Professor J. O. Wilkes with the computer portion of the work is gratefully acknowledged.

References

1. L. Heyne, Electrochimica Acta 15, 1251 (1970).

2. W. Baukal, in *From Electrocatalysis to Fuel Cells*, ed. by G. Sandstede, Univ. of Wasington Press, Seattle (1972).

3. M. S. Whittingham and R. A. Huggins, in Proc. 7th Int. Symp. on the Reactivity of Solids, Chapman and Hall, London (1972).

4. M. S. Whittingham and R. A. Huggins, in *Solid State Chemistry*, Nat. Bur. of Standards Special Pub. No. 364, ed. by R. S. Roth and S. J. Schneider (1972).

5. R. J. Quigley and T. P. Das, Solid State Comm. 5, 487 (1967).

6. K. Weiser, J. Phys. Chem. Solids 17, 149 (1960), and Phys. Rev. 126, 1427 (1962).

7. C. J. F. Bottcher, *Theory of Electric Polarization* (Elsevier Publishing Co., Inc., New York, 1952); N. F. Mott and M. J. Littleton, Trans. Faraday Soc. 34, 485 (1938); E. S. Rittner, R. A. Hutner, and F. K. dePre, J. Chem. Phys. 17, 198 (1949); M. Born and J. E. Mayer, Z. Physik 75, 1 (1932); M. F. C. Ladd and W. H. Lee, Prog. Solid State Chem. 1, 37 (1964).

8. M. Born and K. Huang, *Dynamical Theory of Crystal Lattices* (Oxford Press, 1954).

9. L. Pauling, Zeit. f. Krist. 67, 377 (1928).

10. J. Krogh-Moe, in *Selected Topics in High-Temperature Chemistry*, ed. by M. Forland, et. al., Universitetsforlaget, Oslo (1966).

11. L. W. Strock, Zeit. f. Phys. Chem. 25, 441 (1934).

12. W. H. Flygare and D. W. Hafemeister, J. Chem. Phys. 45, 372 (1966).

13. M. L. Huggins, J. Chem. Phys. 5, 143 (1937); M. L. Huggins, in *Phase Transformations in Solids*, ed. by Smoluchowski et. al., John Wiley and Sons, New York (1951).

14. R. D. Shannon and C. T. Prewitt, Acta Cryst. B25, 925 (1969).

15. W. Biermann and W. Jos[illegible] phys. chem. NF 25, 139 (1960).

16. G. G. Bentle, J. Appl Phys. 39, 4036 (1968).

17. M. S. Whittingham and R. A. Huggins, J. Chem. Phys. 54, 414 (1971).

18. M. S. Whittingham and R. A. Huggins, J. Electrochem. Soc. 118, 1 (1971).

THEORETICAL CALCULATION OF THERMODYNAMIC PROPERTIES OF IRON-CARBON AUSTENITES

S. K. Das and E. E. Hucke

Abstract

A calculation of all the partial and integral thermodynamic properties of iron-carbon austenite is presented. The configurational entropy has been calculated using Boltzman's relation. The partial vibrational entropy of iron has been calculated using Einstein's model for the specific heat of a crystalline solid, together with available lattice parameter data of austenite as a function of temperature and composition. The resulting total entropy has been integrated over temperature to obtain the partial Gibb's free energy of iron; and then partial properties of carbon have been obtained by Gibb's-Duhem integrations. The integration constants have been evaluated using the available thermodynamic data for the phases in equilibrium with the austenite phase. The results obtained in the present model are in remarkable agreement with the available experimental data.

Introduction

The fcc solid solution is the heart of the iron-carbon binary system and its thermodynamic properties and boundary lines are rather well-known. Its thermodynamic properties have been studied by many authors employing a variety of techniques, i.e., gas equilibration, emf technique, and vapor pressure methods. However, the most relaible data on activities of carbon and iron in austenite as functions of temperature and composition come through studies of the equilibria:

$$\underline{C}(\text{dissolved in iron}) + 2H_2(g) = CH_4(g) \qquad (1)$$

$$\underline{C}(\text{dissolved in iron}) + CO_2(g) = 2CO(g) \qquad (2)$$

The thermodynamic data for austenite, along with data on α phase, liquid phase and other phases of the iron-carbon phase diagram have recently been reviewed by Chipman.[1] Hultgren et al[2] have also tabulated selected data. An excellent review of all the theoretical models proposed for the thermodynamic properties of austenite has recently been published by Mclellan and Chraska.[3] These models can be classified into two groups: geometric exclusion models and atomic interaction models. The geometric exclusion models are not realistic since they imply that a C-atom nearest-neighbor pair can have a zero or infinite potential energy depending upon its orientation. Furthermore, these models are generally athermal in nature, i.e., all the deviation from regular behavior is due to the non-ideal configurational entropy and the partial enthalpy is made independent of composition. The atomic interaction models for austenite are the simplest form of statistical interaction model for interstitial solid solutions where

zeroth order approximation has been used. The configurational entropy and internal energy are those of a random solution, the internal energy computed as a pairwise summation over first nearest-neighbor interaction energies.

In these models C-C interaction energy has either been estimated or deduced from available experimental data for the activity of carbon in austenite. These estimates have varied from -1.05 to -2.15 Kcal/mole. To obtain an interpolation of the activity of carbon in austenite as a function of temperature, Poirier[4] has assumed a linear temperature dependence of interaction energies of C-C and C-Fe pairs. The resulting four constants have been evaluated to best fit the experimental data for the activity of carbon in austenite. The constants so determined are then used to determine the activity of carbon in austenite as a function of temperature and composition.

Apart from the criticism that the first nearest neighbor quasi-chemical models do not adequately describe the energetics of the metallic solution, all the models described so far have taken into account only configurational entropy and enthalpy, which are independent of temperature. In contrast, the present method attempts to include vibrational terms and hence accounts for temperature dependence of entropy and enthalpy.

In the proposed model the configurational entropy of austenite has been calculated from the knowledge of its structure using Boltzman's definition of configurational entropy. Einstein proposed a model of solid as a simple harmonic oscillator. Using his model of the solid and the available lattice parameters of

austenite as a function of temperature and composition, the partial vibrational entropy of iron has been calculated. The total (configurational plus vibrational) partial entropy of iron thus obtained has been integrated with respect to temperature to obtain the partial Gibb's free energy of iron. The subsequent Gibb's-Duhem integration of the partial Gibb's free energy of iron gives the partial Gibb's free energy of carbon in solution, hence all the partial and integral thermodynamic properties of austenite. The integration constants have been evaluated using the available thermodynamic data of the phases in equilibrium with the austenite phase. All the relevant steps have been shown in Figure 1.

Configurational Entropy

The structure of austenite in iron-carbon alloys is well characterized.[5] Figure 2 schematically shows the structure of fcc austenite. In this structure, the carbon atoms occupy the mid-points of cube edges and the cube centers. These are equivalent positions, for in each case a carbon atom finds itself located between two iron atoms along any <100> direction. In the drawing (Figure 2), the positions that carbon atoms occupy are shown by black dots. It must be recognized, however, that

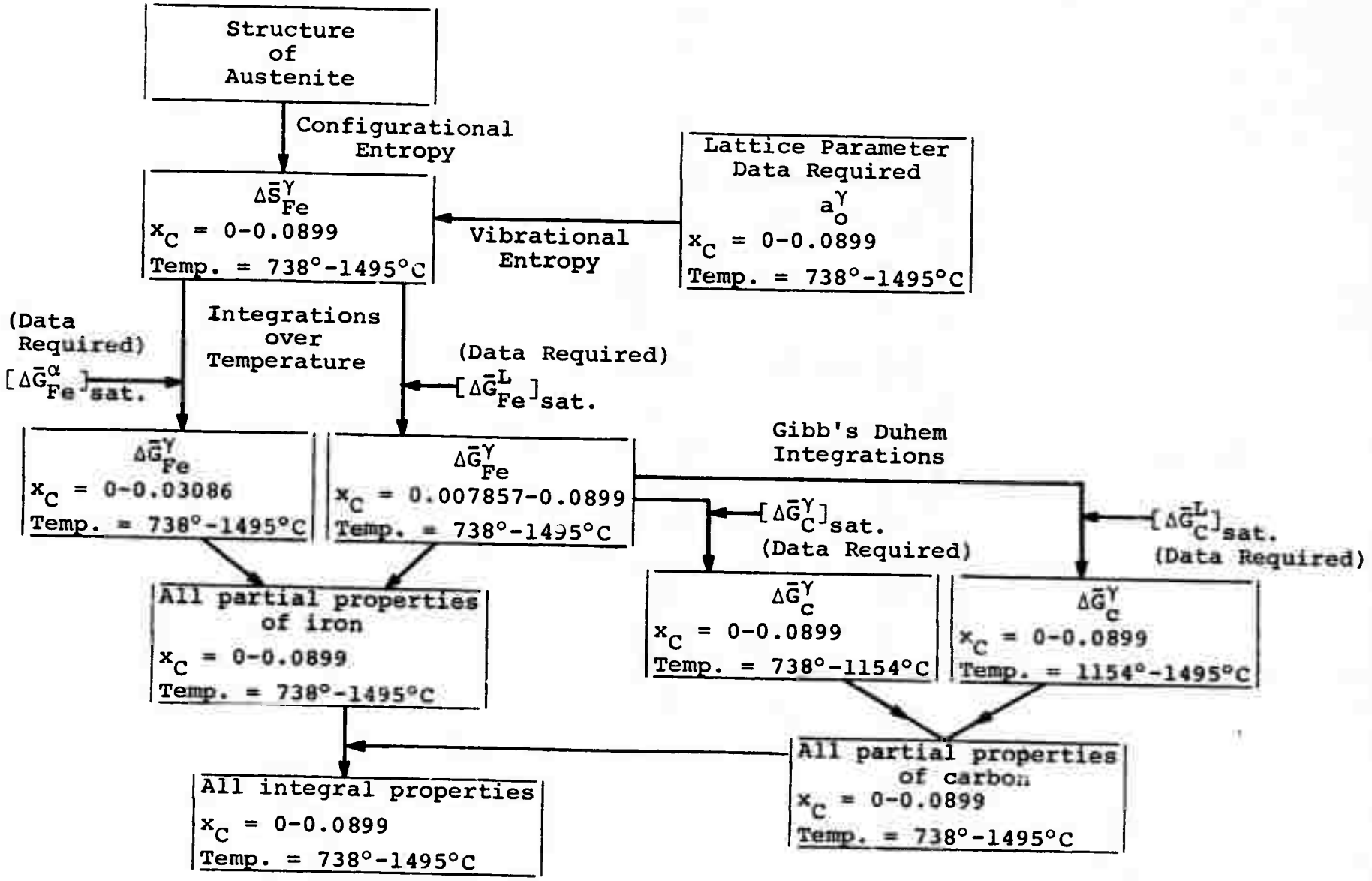


Figure 1: Layout of the Steps

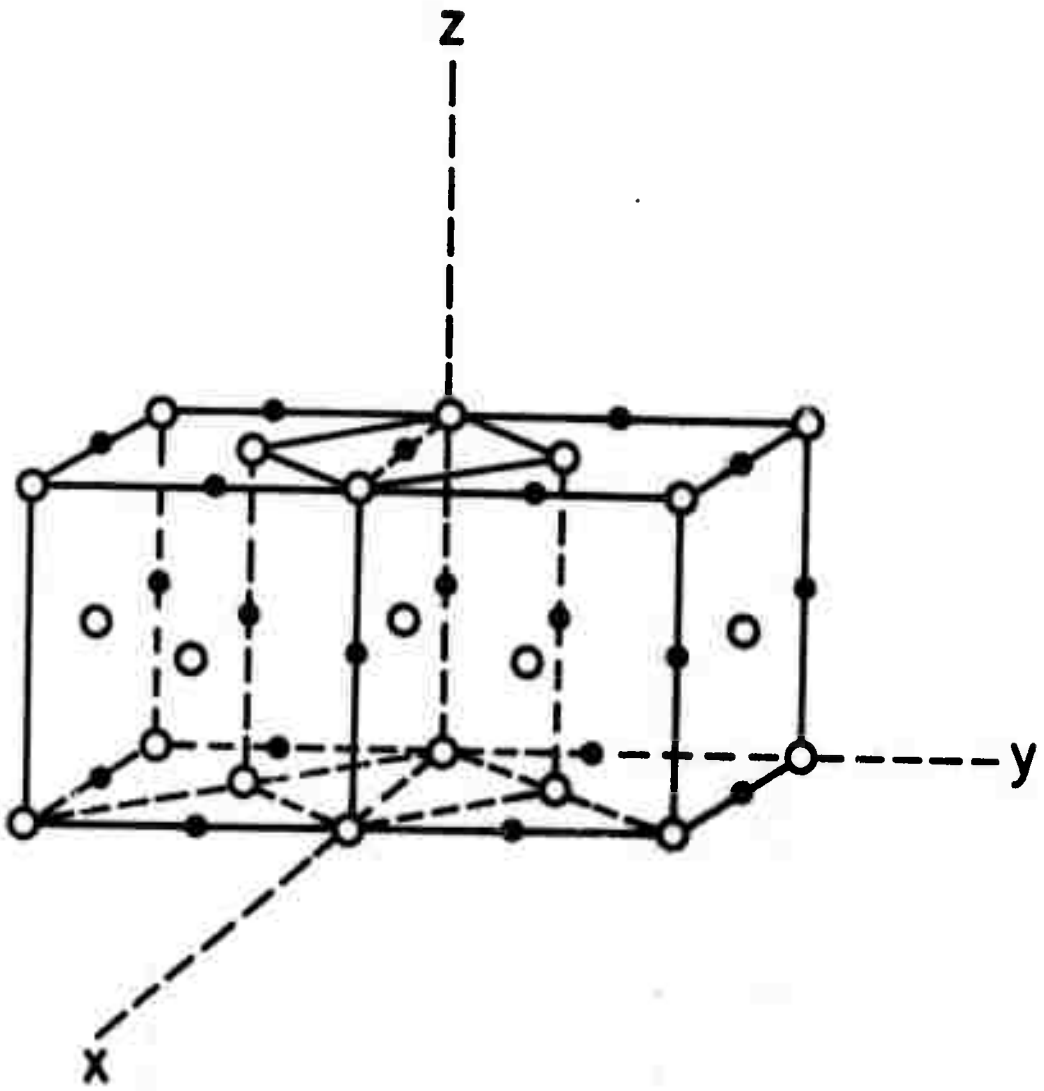


Figure 2. Face-centered cubic austenite (From: Physical Metallurgy Principles, by R. E. Reed-Hill.)

●: Carbon atoms

O: Iron atoms

actually in any given steel specimen only a very small percentage of the possible positions is ever filled. In the face-centered cubic structure there are as many possible positions for carbon atoms as there are for iron atoms. This means that if all the positions were filled, the alloy would have a composition containing 50 atomic percent carbon. The maximum actually observed is 8.9 atom percent (2.08 weight percent).

Since pure face centered cubic iron and graphite have an ordered structure and are presumably in complete internal equilibrium, there is a unique way of arranging iron and carbon atoms in fcc and graphite lattice, respectively. Therefore, the configurational entropy of reactants in Eq. 3 are zero individually, and thus the integral configurational entropy change of reaction 3 is the configurational entropy of austenite.

$$(1-x_C)Fe_{(\gamma)} + x_C C_{(gr)} = Fe_{1-x_C}\,C_{x_C} \quad (\gamma\text{-solid solution}) \qquad (3)$$

Consider an ensemble of N atoms, of which there are Nx_C carbon atoms and $N(1-x_C)$ iron atoms. The number of available sites in the lattice which carbon atoms can occupy is $N(1-x_C)$. Hence, the total number of distinguishable ways of arranging $N(1-x_C)$ available carbon sites into Nx_C occupied carbon sites and $N(1-2x_C)$ unoccupied carbon sites is

$$W = \frac{N(1-x_C)!}{Nx_C!\ N(1-2x_C)!} \qquad (4)$$

The filling of the iron sites is not considered since all such sites are assumed occupied by iron atoms, i.e., no point defects or vacancies are presently considered. Equation (4) is valid

because all the possible carbon sites are considered equivalent. For large values of N sterling's approximation for the logarithms of the factorials can be used. The configurational entropy of austenite then reduces to

$$\bar{S}^{\gamma}con. = \Delta\bar{S}^{\gamma}con. = k\ell nW = R\left[(1-x_C)\ell n\left(\frac{1-x_C}{1-2x_C}\right) + x_C\ \ell n\left(\frac{1-2x_C}{x_C}\right)\right] \qquad (5)$$

Hence the partial configurational entropy of iron in the solution is:

$$\Delta\bar{S}^{\gamma}_{Fe,\ con.} = R\ \ell n\left(\frac{1-x_C}{1-2x_C}\right) \qquad (6)$$

Partial Vibrational Entropy of Iron

In a crystal each atom is vibrating about a given position and hence there is a randomness associated with a given atom at any time. If we consider that each atom is moving in a cell, we should expect the vibrational randomness to be related to the volume of the cell. The larger the volume of the cell, the larger will be the randomness and the larger will be the vibrational entropy. Moreover, the lattice parameters of the cell would be expected to be proportional to the amplitude of the vibration of the atoms.[6] In the face centered cubic structure of iron the holes are barely large enough to accommodate the carbon atoms in the interstices, and the crowding introduces considerable strain in the structure. For these reasons the vibrational contribution of the carbon to the total entropy appears to be relatively small.

At elevated temperatures where most frequencies are close to a single characterisitc frequency, ν, the Einstein's approach

seems to be a good approximation to estimate the vibrational entropy contribution to a crystal. In the Einstein's theory, each atom is a three-dimensional harmonic oscillator, and the thermodynamic functions for a crystal are 3N times the thermodynamic functions for a single oscillator. Suppose that the characteristic frequency of pure iron changes from ν to ν' due to the introduction of carbon atoms. The partial vibrational entropy of iron is then

$$\Delta\bar{S}^{\gamma}_{Fe,\ vib.} = 3R\ \ell n\left(\frac{\nu}{\nu'}\right) \tag{7}$$

Assuming that the characteristic frequency is inversely proportional to the square root of the amplitude, which is directly proportional to the lattice parameter, one gets

$$\Delta\bar{S}^{\gamma}_{Fe,\ vib.} = \frac{3}{2}\ R\ \ell n\left(\frac{a^{\gamma}_{o}}{a^{Fe}_{o}}\right) \tag{8}$$

Ridley and Stuart[7] have experimentally measured the lattice parameters of austenite as a function of temperature and weight percent carbon. It is important to emphasize th . they measured the lattice parameters of equilibrium austenite, and not retained austenite as reported by Roberts.[9]

Ridley and Stuart[7,8] have given the slopes and intercepts of the least-square fitted straight lines of the lattice parameters versus carbon weight percent at three temperatures (298°, 1000°, 1473°K).

Temperature	Slope	Intercept
298°K	0.0316	3.5735
1000°K	0.0296	3.6300
1473°K	0.0282	3.6681

A least-square analysis on both slopes and intercepts, individually, yields the following equations:

$$\text{Slope} = 0.03247 - 2.8902 \times 10^{-6}T$$

$$\text{Intercept} = 3.5497 + 8.0376 \times 10^{-5}T$$

Hence, a single equation which describes the variation of the lattice parameters of austenite as a function of temperature and carbon weight percent is:

$$a_o^{\gamma}(A^\circ) = m + nT$$

where $m = a + c \cdot X_C$, $n = b + d \cdot X_C$, $a = 3.5497$, $b = 8.0376 \times 10^{-5}$, $c = 0.03247$, $d = -2.8902 \times 10^{-6}$.

Using the above relations, the partial vibrational entropy of iron can be written as

$$\Delta\bar{S}^{\gamma}_{Fe,\ vib.} = \frac{3}{2} R\ \ell n\left(\frac{m+nT}{a+bT}\right) \qquad (9)$$

As X_C tends to zero, $\Delta\bar{S}^{\gamma}_{Fe,\ vib.} \to 0$ from Eq. (9), and when $T \to 0$, $\Delta\bar{S}^{\gamma}_{Fe,\ vib.}$ tends to zero by definition which is also embedded in Eq. (7). Adding Eqs. (6) and (9), we can write the total (configurational + vibrational) partial entropy of iron in austenite as

$$\Delta\bar{S}^{\gamma}_{Fe} = R\ \ell n\left(\frac{1-x_C}{1-2x_C}\right) + \frac{3}{2} R\ \ell n\left(\frac{m+nT}{a+bT}\right) \qquad (10)$$

Partial Thermodynamic Properties of Iron

The partial Gibb's free energy of iron, hence all other partial properties of iron in the solution can be obtained by integrating Eq. (10). In order to integrate the entropy with respect to temperature, one integration constant at every compo-

sition must be specified. This may be accomplished for every composition, if one temperature where the partial Gibb's free energy of iron in the solution is known is specified. For this purpose either the thermodynamic properties of the liquid phase at the liquid-austenite and austenite phase boundary or the thermodynamic properties of the α-phase at α-γ and austenite phase boundary can be used.

Writing the equation using the thermodynamic properties of the liquid phase, one gets:

$$[\Delta\bar{G}^{\gamma}_{Fe}]_{x_C,T} - [\Delta\bar{G}^{\gamma}_{Fe}]^{\gamma,\gamma+L}_{x_C,T_1(X_C)} = -\int_{T_1(X_C)}^{T} (\Delta\bar{S}^{\gamma}_{Fe})\,dT \qquad (11)$$

where T_1 is the temperature corresponding to the composition X_C on the austenite-liquid and austenite boundary line AB in Fig. 3. Line AB (temperature °C versus carbon weight percent) has been assumed to be a straight line whose least-square fitted equation is $T_1 = 1804.2192 - 182.4010\ X_C$ with a correlation coefficient of 0.9994 which, by any standard is an excellent fit.

The expression for the partial Gibb's free energy of iron in liquid iron has been given by Chipman.[1] Since at temperature T_1 and composition X_C, the equilibrium relation yields;

$$[\Delta\bar{G}^{\gamma}_{Fe}]^{\gamma,\gamma+L}_{x_C,T_1(X_C)} = [\Delta\bar{G}^{L}_{Fe}]^{\gamma,\gamma+L}_{x_C,T_1(X_C)} \qquad (12)$$

Chipman's expression can be written as:

$$[\Delta\bar{G}^{\gamma}_{Fe}]^{\gamma,\gamma+L}_{x_C,T_1(X_C)} = RT_1[-2.3026\ y_C^2(0.36 + \frac{1700}{T_1}) + \ln(1-y_C)] \qquad (13)$$

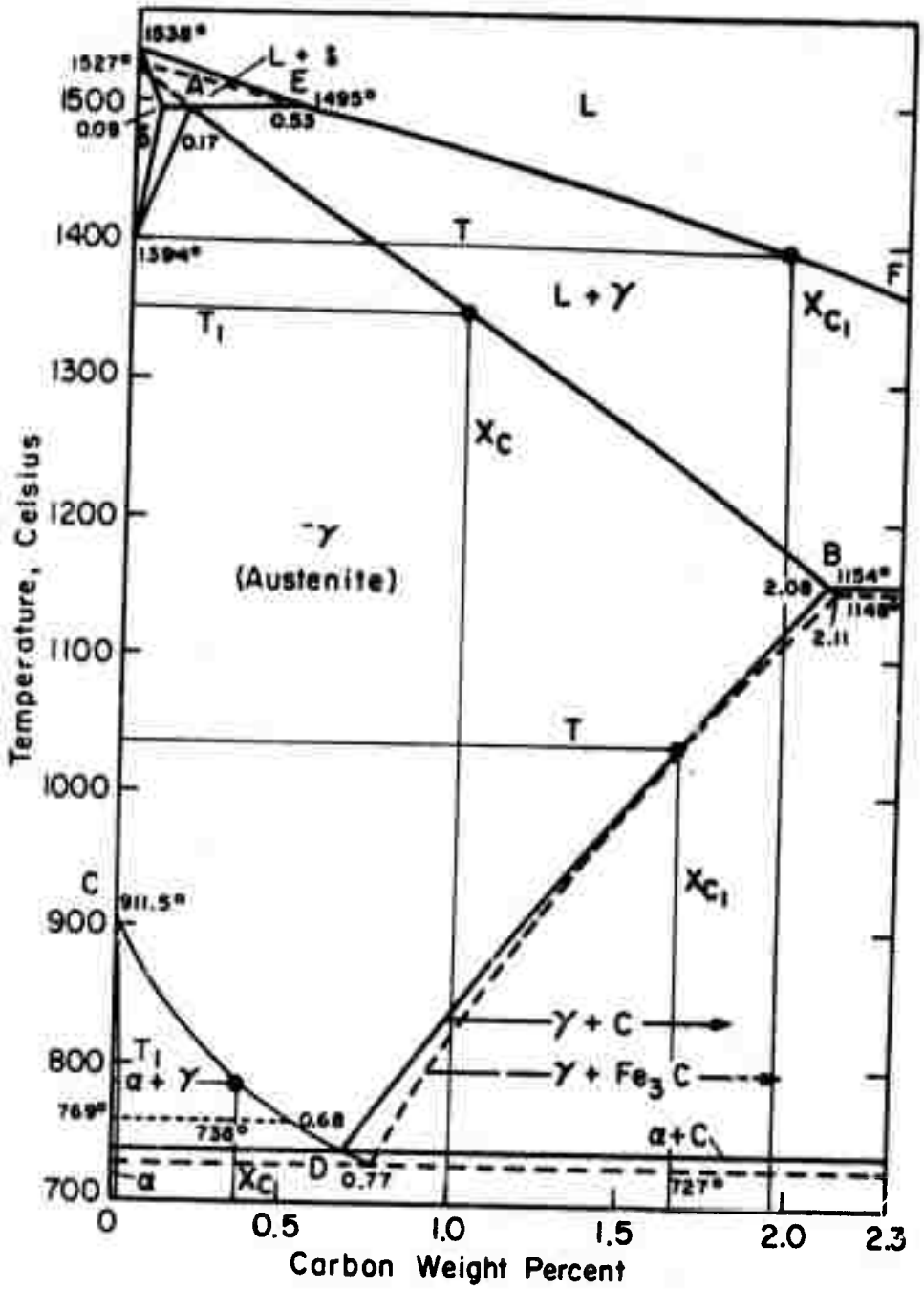


Figure 3. Portion of the phase diagram Fe-C. Metastable γ-range and system Fe-Fe_3C shown by dashed lines. Curie temperature dotted.

where $y_C = \frac{x_C}{1-x_C}$.

Integrating Eq. (11) and substituting the integration constant from Eq. (13), one can write

$$\Delta\bar{G}^{\gamma}_{Fe} = -R(T-T_1)\ln\left|\frac{1-x_C}{1-2x_C}\right| + \frac{3}{2}R[(T+a/b)\ln(a+bT) - (T_1+a/b)\ln(a+bT_1) - (T+m/n)\ln(m+nT) + (T_1+m/n)\ln(m+nT_1)] \quad (14)$$

$$+RT_1[-2.3026\ y_C^2(0.36 + \frac{1700}{T_1}) + \ln(1-y_C)]$$

The above equation is valid for x_C = 0.007857 - 0.0899 and temperature 738° - 1495°C.

Similarly, another equation valid for x_C = 0 - 0.03086 and temperature 738° - 1495°C can be written using the partial Gibb's free energy of iron in the α-phase at the α, α+γ phase boundary. The partial Gibb's free energy of carbon in the α-phase has been given by Chipman[1] as:

$$[\Delta\bar{G}^{\gamma}_{C}]^{\alpha,\alpha+\gamma}_{x_C,T_1(X_C)} = RT_1[2.3026(\frac{5550}{T_1} - 2.49) + \ln y_C] \quad (15)$$

The expression for the partial Gibb's free energy of iron in the α-phase can be obtained by Gibb's-Duhem integration as:

$$[\Delta\bar{G}^{\gamma}_{Fe}]^{\alpha,\alpha+\gamma}_{x_C,T_1(X_C)} = -RT_1y_C \quad (16)$$

T_1 in this case is the temperature on the α,α+γ boundary line CD in Fig. 3 at a composition X_C.

All other partial properties of iron in austenite can now be calculated using standard thermodynamic relations.

Partial Thermodynamic Properties of Carbon

The partial Gibb's free energy of carbon in the solution can be obtained by Gibb's-Duhem integration of the partial Gibb's free energy of iron. Here also it is necessary to evaluate integration constants, which in this case would be a composition X_C at every temperature where the partial Gibb's free energy of carbon in austenite is known. In the temperature region of 738°-1154°C, the partial Gibb's free energy of carbon along the boundary line of γ and γ+C phases is known, since at every temperature (between 738°-1154°C) a composition, X_C, is known where austenite is saturated with carbon and hence its partial Gibb's free energy relative to graphite is zero. The expression for $\Delta\bar{G}_C^{\gamma}$ is:

$$[\Delta\bar{G}_C^{\gamma}]_{X_C,T} - [\Delta\bar{G}_C^{\gamma}]^{\gamma,\gamma+C}_{X_{C_1}(T),T} = \int_{X_C=X_{C_1}}^{X_C=X_C} -\left(\frac{12.01}{55.85}\right)\left(\frac{100-X_C}{X_C}\right) d[\Delta\bar{G}_{Fe}^{\gamma}] \qquad (17)$$

where the expression for $d[\Delta\bar{G}_{Fe}^{\gamma}]$ can be obtained by differentiating Eq. (14)

$$d[\Delta\bar{G}_{Fe}^{\gamma}] = -\frac{RT\,dx_C}{(1-x_C)(1-2x_C)} + \frac{3}{2}R\left[182.4010\ \ell n\left(\frac{a+bT_1}{m+nT_1}\right) - \frac{d(T-T_1)}{n} + \frac{(bc-ad)}{n^2}\ \ell n\ \left(\frac{m+nT_1}{m+nT}\right)\right]dX_C \qquad (18)$$

$$+ R(151.1988)y_C^2 dX_C - R(2.3026)y_C(0.72T_1+3400)dy_C$$

where $$dy_C = \frac{dx_C}{(1-x_C)^2}$$

$$dx_C = \frac{(55.85-43.84x_C)^2}{(12.01)(100)(55.85)}\,dX_C$$

$$x_C = \frac{X_C/12.01}{\left[\frac{X_C}{12.01} + \frac{100-X_C}{55.85}\right]}$$

All other terms have been defined previously. Line BD (temperature °C versus carbon weight percent) in Fig. 3 has been assumed to be a straight line whose least-square fitted equation is $X_{C_1}(T) = -2.7980 + 0.0034107T$ with a co-relation coefficient of 0.9998 which represents an excellent fit. X_{C_1} is the intersection of line BD (boundary between the γ and γ+C phases) at any temperature T. For the entire region for which Eq. (17) is valid (x_C = 0-0.0899, temperature 738°-1154°C),

$$[\Delta\bar{G}_C^{\gamma}]^{\gamma,\gamma+C}_{X_{C_1}(T),T} = 0 \tag{19}$$

The expression for $d[\Delta\bar{G}_{Fe}^{\gamma}]$ expressed in T, X_C and dX_C from Eq. (18) has been substituted in Eq. (17), and the resulting integration has been performed numerically using a Gauss-Legendre quadrature exact up to 95th degree polynomial, i.e., 48 point formula.

Similarly, another equation for $\Delta\bar{G}_C^{\gamma}$ valid for x_C = 0-0.0899, temperature 1154°-1495°C can be written using the following expression given by Chipman[1] for $\Delta\bar{G}_C^{L}$ at the boundary of γ and γ+L phases.

$$[\Delta\bar{G}_C]^{L}_{X_{C_1}(T),T} = [\Delta\bar{G}_C^{\gamma}]^{\gamma,\gamma+L}_{x_C,T} = RT\left[2.3026\left(\frac{1180}{T} - 0.870 + (0.72 + \frac{3400}{T})y_{C_1}\right) + \ell n\left(\frac{y_{C_1}}{1-y_{C_1}}\right)\right] \tag{20}$$

where $y_{C_1} = \frac{x_{C_1}}{(1-x_{C_1})}$ and x_{C_1} is the intersection on line EF of Fig. 3 at any temperature T. In other words, in this case it is

possible to evaluate the constants of integration using the thermodynamic properties of the liquid phase.

After obtaining the values of $\Delta\bar{G}_C^\gamma$ from Eqs. (17) and (18), the expression for $\Delta\bar{S}_C^\gamma$ can be readily obtained by differentiating Eq. (17) with respect to T. Equation (17) can be written in the following form:

$$\Delta\bar{G}_C^\gamma = -\left[\frac{12.01}{55.85}\right]\int_{X_C=X_{C_1}(T)}^{X_C=X_C} f(X_C,T)\,dX_C \tag{21}$$

Using Eqs. (17)-(24), one can develop an expression for $\Delta\bar{S}_C^\gamma$.

$$\Delta\bar{S}_C^\gamma = -\frac{d}{dT}[\Delta\bar{G}_C^\gamma] \tag{22}$$

$$\frac{d}{dT}\int_{a(T)}^{b(T)} f(X_C,T)\,dX_C = f[b(T),T]b'(T) - f[a(T),T]a'(T) + \int_{a(T)}^{b(T)} \frac{\partial}{\partial T} f(X_C,T)\,dX_C \tag{23}$$

$$\frac{\partial f}{\partial T}(X_C,T) = -\left(\frac{100-X_C}{X_C}\right)\left[\frac{R(5585)}{(100-X_C)(1201-67.86X_C)} + \frac{3}{2}R\left\{\frac{d}{n} + \frac{(bc-ad)}{n(m+nT)}\right\}\right] \tag{24}$$

and $b'(T) = 0$, $a'(T) = 0.0034107$. After obtaining $\Delta\bar{G}_C^\gamma$ and $\Delta\bar{S}_C^\gamma$, all other partial properties of carbon can be obtained.

Results and Discussion

The calculated results for partial properties of iron, partial properties of carbon and finally the integral properties of the solution are shown and compared with the experimentally obtained values listed by Hultgren et al[2] in Tables 1, 2, and 3,

TABLE 1: Computed and Experimental Values of the Partial Properties of Iron in Solution at 1426°K

Standard State: Pure iron at 1426°K for iron

x_{Fe}	$\Delta\bar{S}^{Y}_{Fe}$ cal/mole °K		$\Delta\bar{G}^{Y}_{Fe}$ cal/mole		$\Delta\bar{H}^{Y}_{Fe}$ cal/mole		a^{Y}_{Fe}		γ^{Y}_{Fe}		$\Delta\bar{G}^{xsY}_{Fe}$ cal/mole		$\Delta\bar{S}^{xsY}_{Fe}$ cal/mole°K	
	Present Model	Experimental	Present Model	Experimental	Present Model	Experimental	Present Model	Experimental	Present Model	Experimental	Present Model	Experimental	Present Model	Experimental
0.99	0.025	0.020	- 28	- 30	8	- 1	0.990	0.989	1.000	0.999	0	- 1	0.006	0.000
0.98	0.051	0.041	- 60	- 62	13	- 4	0.979	0.978	0.999	0.998	- 3	- 5	0.011	0.001
0.97	0.078	0.063	- 95	- 98	16	- 8	0.967	0.966	0.997	0.996	- 8	- 11	0.017	0.002
0.96	0.105	0.085	-134	-136	16	-15	0.954	0.953	0.994	0.993	- 17	- 21	0.023	0.004
0.95	0.133	0.108 (±0.020)	-178	-178 (±20)	12	-24 (±20)	0.939	0.939 (±0.008)	0.988	0.988 (±0.008)	- 34	- 33 (±20)	0.032	0.006 (±0.020)
0.94	0.162	0.131	-225	-223	6	-36	0.924	0.924	0.983	0.983	- 48	- 49	0.038	0.009
0.93	0.192	0.156	-277	-272	- 3	-50	0.907	0.908	0.975	0.976	- 72	- 67	0.048	0.012
0.92	0.223	0.182	-334	-326	-16	-66	0.889	0.891	0.966	0.969	- 98	- 89	0.058	0.016
0.91*	0.255	0.208	-394	-383	-30	-86	0.870	0.874	0.956	0.960	-127	-116	0.068	0.021

*Phase Boundary

TABLE 2: Experimental and Computed Values of the Partial Properties of Carbon in Solution at 1426°K

Standard State: Pure Graphite at 1426°K for carbon

x_C	$\Delta\bar{G}^{Y}_{C}$ cal/mole		$a_C{}^{Y}$		$\gamma_C{}^{Y}$		$\Delta\bar{G}^{xsY}_{C}$ cal/mole		$\Delta\bar{S}^{Y}_{C}$ cal/mole °K		$\Delta\bar{H}^{Y}_{C}$ cal/mole		$\Delta\bar{S}^{xsY}_{C}$ cal/mole°K	
	Present Model	Experimental	Present Model	Experimental	Present Model	Experimental	Present Model	Experimental	Present Model	Experimental	Present Model	Experimental	Present Model	Experimental
0.01	-8463	-8312	0.050	0.053	5.030	5.323	4563	4738	14.668	13.353	12463	10729	5.536	4.201
0.02	-6314	-6105	0.108	0.116	5.430	5.800	4782	4981	12.904	11.932	12100	10910	5.128	4.158
0.03	-4920	-4710	0.176	0.190	5.867	6.323	5017	5226	11.853	11.083	11994	11094	4.889	4.115
0.04	-3834	-3643	0.259	0.276	6.475	6.911	5296	5478	11.094	10.466	11997	11282	4.696	4.070
0.05	-2915	-2753 (±200)	0.358	0.378 (±0.028)	7.160	7.570 (±0.570)	5582	5736 (±200)	10.494	9.977 (±0.380)	12060	11474 (±500)	4.540	4.024 (±0.380)
0.06	-2180	-1973	0.477	0.499	7.950	8.309	5878	6000	9.995	9.568	12163	11671	4.404	3.977
0.07	-1354	-1267	0.620	0.638	8.857	9.121	6185	6268	9.565	9.213	12295	11871	4.282	3.929
0.08	- 659	- 614	0.793	0.805	9.912	10.064	6504	6543	9.184	8.899	12446	12076	4.164	3.880
0.09*	0	0	1.000	1.000	11.111	11.111	6828	6823	8.842	8.615	12618	12285	4.057	3.830

*Phase Boundary

TABLE 3: Experimental and Computed Values of the Integral Properties of the Solution at 1426°K

x_C	$\Delta\bar{G}^{Y}$ cal/mole		$\Delta\bar{H}^{Y}$ cal/mole		$\Delta\bar{S}^{Y}$ cal/mole°K		$\Delta\bar{G}^{XSY}$ cal/mole		$\Delta\bar{S}^{XSY}$ cal/mole°K	
	Present Model	Experimental	Present Model	Experimental	Present Model	Experimental	Present Model	Experimental	Present Model	Experimental
0.01	-112	-113	132	106	0.171	0.153	46	46	0.060	0.042
0.02	-185	-183	254	214	0.308	0.279	93	95	0.113	0.084
0.03	-240	-236	375	325	0.431	0.394	143	146	0.162	0.125
0.04	-282	-276	494	437	0.544	0.500	196	199	0.210	0.167
0.05	-315	-307 (±30)	614	551 (±50)	0.651	.601 (±0.040)	247	255 (±30)	0.257	0.207 (±0.040)
0.06	-338	-328	735	666	0.752	0.697	308	314	0.299	0.247
0.07	-352	-342	858	784	0.848	0.790	366	376	0.345	0.286
0.08	-360	-349	981	905	0.940	0.879	430	442	0.386	0.325
0.09*	-360	-349	1106	1027	1.028	0.965	499	509	0.425	0.364

*Phase Boundary

respectively. It can be seen that the agreement is remarkable.

Hultgren et al[2] have given estimates of the accuracy of the selected thermodynamic properties at $X_C = 0.05$. It can be seen in Tables 1 to 3 that all the values of partial and integral Gibb's free energies and activity and activity coefficients of both iron and carbon are well within the estimated accuracy limits of the experimental data. For example, at $x_C = 0.05$, the $\Delta\bar{G}^{\gamma}_{Fe}$, and hence a^{γ}_{Fe} and γ^{γ}_{Fe} values are identical with the experimental data. The values of $\Delta\bar{G}^{xs\,\gamma}_{Fe}$ differ by only 1 cal/mole. Similarly, the calculated values of $\Delta\bar{G}^{\gamma}_{C}$, a^{γ}_{C}, γ^{γ}_{C}, $\Delta\bar{G}^{xs\,\gamma}_{C}$ and $\Delta\bar{G}^{\gamma}$, $\Delta\bar{G}^{xs\,\gamma}$ lie well within the uncertainty of the reported experimental data at $x_C = 0.05$. The calculated values of $\Delta\bar{S}^{\gamma}_{Fe}$, $\Delta\bar{S}^{\gamma}_{C}$, $\Delta\bar{S}^{\gamma}$, $\Delta\bar{S}^{xs\,\gamma}_{C}$, $\Delta\bar{S}^{xs\,\gamma}$, $\Delta\bar{H}^{\gamma}_{C}$ and $\Delta\bar{H}^{\gamma}$ are respectively about 4%, 1%, 2%, 3%, 4%, 1% and 2% higher than the reported uncertainty band at $x_C = 0.05$. In the case of $\Delta\bar{H}^{\gamma}_{Fe}$ and $\Delta\bar{S}^{xs\,\gamma}_{Fe}$, the experimental uncertainty is about 100% and 200%, and the match in these cases with calculated values is extremely poor.

This calculation scheme appears very promising for application in other alloy systems and can be extended to ternary or multicomponent systems. The most promising systems are probably those with interstitual solid solutions where the thermodynamic properties of the interstitial compounds are known. The required lattice parameter data should be much easier to measure than the thermodynamic properties of the solid solutions. The real power of this model is its ability of predicting the partial heat capacity of constituents in an interstitial solid solution at elevated temperatures as a function of temperature as well as

composition. All that is needed is a knowledge of the structure of the phases and lattice parameter data. Unfortunately, the data on the direct measurements on the heat capacity of austenite as a function of temperature and composition are not available for comparison in this case.

Nomenclature

x: mole fraction

X: weight percent

T: absolute temperature

G: Gibb's free energy, cal./gm-mole

S: entropy, cal./°K-gm-mole

H: enthalpy, cal./gm-mole

a: relative activity; γ: activity coefficient

W: thermodynamic probability

N: total number of atoms

k: Boltzman's constant

R: universal gas constant

ν: frequency

a_o: lattice parameter (A°)

Subscripts

C: carbon

Fe: iron

con.: configurational contribution

vib.: vibrational contribution

Superscripts

γ: austenite phase

α: α-phase

L: liquid phase

xs: excess properties

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract Number DAHC15-71-C-0283.

References

1. J. Chipman: Met. Trans., 1972, vol. 3, p. 55.
2. R. Hultgren, R. L. Orr, P. D. Anderson and K. K. Kelley: Selected Values of Thermodynamic Properties of Metals and Alloys, Supplement, John Wiley & Sons, Inc., Oct. 1971.
3. R. B. Mclellan and P. Chraska: Mat. Sci. & Eng., 1971, vol. 7, p. 305.
4. D. R. Poirier: Trans. AIME, 1968, vol. 242, p. 685.
5. R. E. Reed-Hill: Physical Metallurgy Principles, D. Van Nostrand Company, Inc., 1964.
6. R. A. Swalin: Thermodynamics of Solids, John Wiley & Sons, Inc., 1970.
7. N. Ridley and H. Stuart: Metal Sci. J., 1970, vol. 4, p. 219.
8. N. Ridley, H. Stuart and L. Zwell: Trans. AIME, 1969, vol. 245, p. 1835.
9. C. S. Roberts: Trans. AIME, 1953, vol. 191, p. 203.

RESEARCH NEEDS AND TECHNICAL OPPORTUNITIES FOR A PROGRAM ON THE RELIABILITY OF BRITTLE MATERIALS

A. G. Evans* and R. L. Coble

Abstract

A Symposium on Mechanical Properties of Brittle Materials was held on July 17, 18, 19, 1972 at Centerville, Massachusetts. The Symposium reviewed the progress in the ARPA Brittle Materials Program with additional papers by others from industry and universities, thus in toto, constituting a review of the present technical situation relative to the reliability of brittle materials. Mechanical and fracture properties, materials development, statistical variation, and non-destructive testing in brittle materials were all reviewed. Recommendations for further work involve:

1. Work in critical stress intensities factor for small inherent flaws
2. Work on second phase materials
3. Slow crack growth at elevated temperatures
4. Work on the investigation of the effects of multiaxial stress
5. Work on statistical strength variations
6. Additional work on flaw detection procedures.

*Materials Department, School of Engineering and Applied Science, University of California, Los Angeles.

RESEARCH NEEDS AND TECHNICAL OPPORTUNITIES FOR A PROGRAM ON THE RELIABILITY OF BRITTLE MATERIALS

A. G. Evans and R. L. Coble

I. INTRODUCTION

The substantial economic benefits that can be accrued from the incorporation of ceramic components in gas turbines is now well established.[1,2] This provides the impetus for the current development of ceramic parts for gas turbines. It is apparent that considerable progress has been made during this program in defining and solving the problems involved in the structural exploitation of ceramic materials. It is intended in this report to define the problems that still require definition and solution before the program can reach an entirely successful conclusion.

A program investigating structural applications for ceramic materials contains three primary components; design, materials development and materials assessment. The techniques or the scientific principles involved in each of these are summarized in Fig. 1. Much of the initial effort was devoted to design and materials development, so that most of the progress has occurred in these areas. The techniques for design are well developed and it is not expected that any insurmontable problems

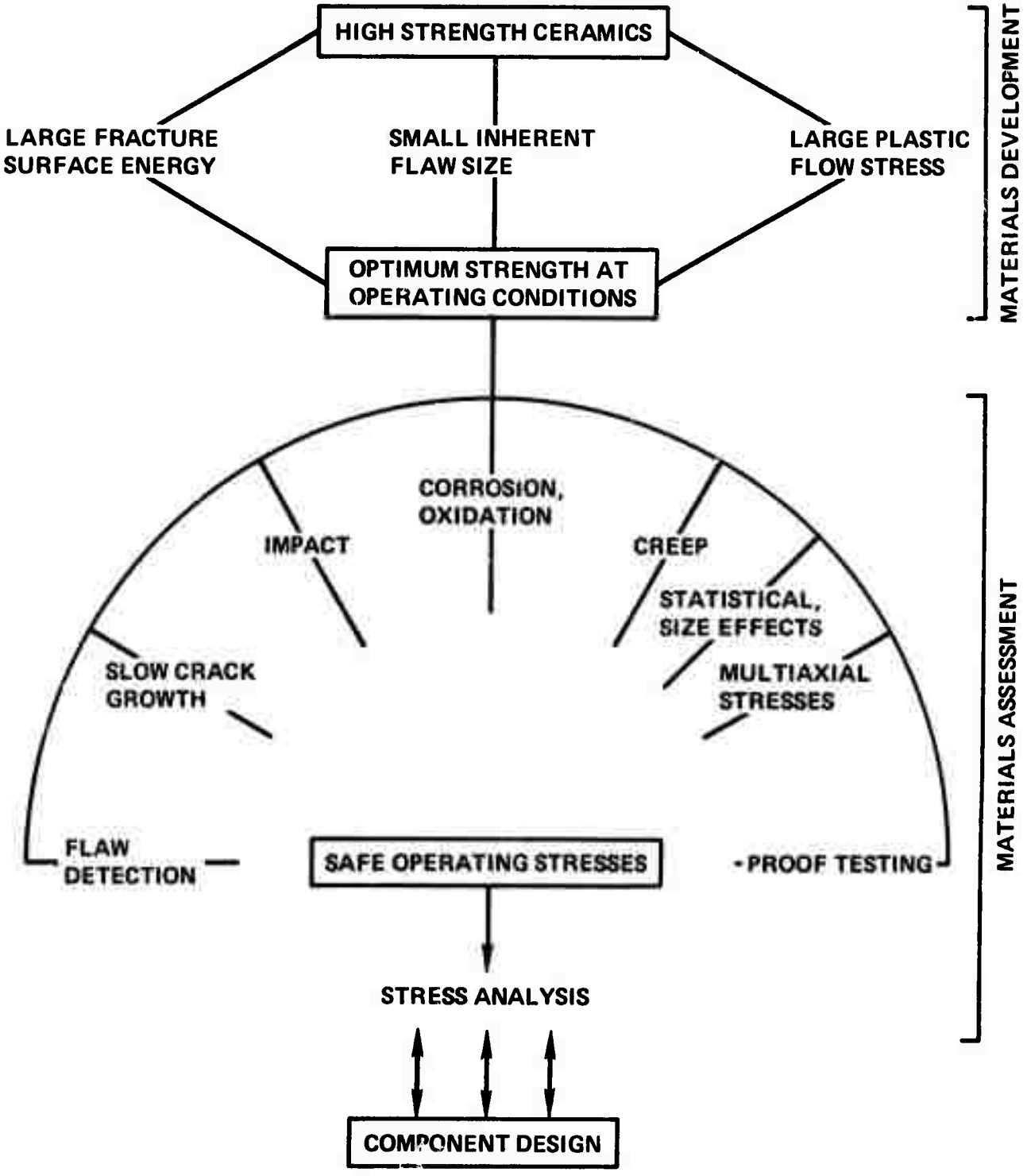


Figure 1.

Development procedure for structural ceramics.

will arise in subsequent design calculations. There is, however, some merit in devoting an additional effort to innovative turbine designs that are better suited to these very brittle materials than the conventional designs. Substantial advances have also been made in materials development, but additional improvements still appear to be required--particularly with regard to materials for rotor blades and discs. Materials assessment has only recently received some attention and much additional work is required in order that realistic long-term component strengths can be predicted. Then, rational quantitative decisions regarding further design iterations and materials developments can be made where required.

II. DESIGN CONSIDERATIONS

Several novel designs for turbines have been suggested as more appropriate for the incorporation of ceramic components than the conventional designs.[3,4] These are aimed primarily at reducing the mechanical stresses in the components subjected to the maximum tensile stress during operation, i.e., the rotor. For the small turbine--which requires that the rotor should be made from a ceramic material (to establish a substantial economic advantage compared to superalloy turbines)--preliminary stress calculations based on alternative designs may show that ceramic rotors for these turbines are subjected to stresses low enough to enable available materials to be used.

III. MATERIALS DEVELOPMENT

Considerable improvements in the strength of structural

ceramic materials have been achieved since the inception of this program[5] (Fig. 2). The latest developments have produced materials which may be adequate for most of the turbine components, excluding, perhaps, the rotor. Much additional work is needed, however, to establish the potential for further strength improvements if materials assessment indicates the available materials are not adequate.

The inadequacies of the latest high strength Si_3N_4 and SiC materials lie primarily in their performance at temperatures above 1000°C, where strength starts to diminish rapidly. Methods for improving strength in this temperature regime are thus eagerly sought. Most of the available information has been obtained for Si_3N_4 and the discussion related primarily to this material, but the principles developed can also be applied to other ceramic materials. The analysis of failure in hot-pressed silicon nitride has shown that an environmentally independent slow crack-growth precedes catastrophic fracture at temperatures $\gtrsim$ 1000°C.[6] There are also good indications that the slow growth proceeds directly from pre-existing flaws[7]* so that flaw initiation is not required.

---

*Fracture has been observed to initiate at a constant stress intensity factor,[7] primarily from impurity particles, at the critical low strength extreme of the strength distribution. A flaw initiation stage may be required at the high strength extreme, but this extreme is of little importance in design considerations.

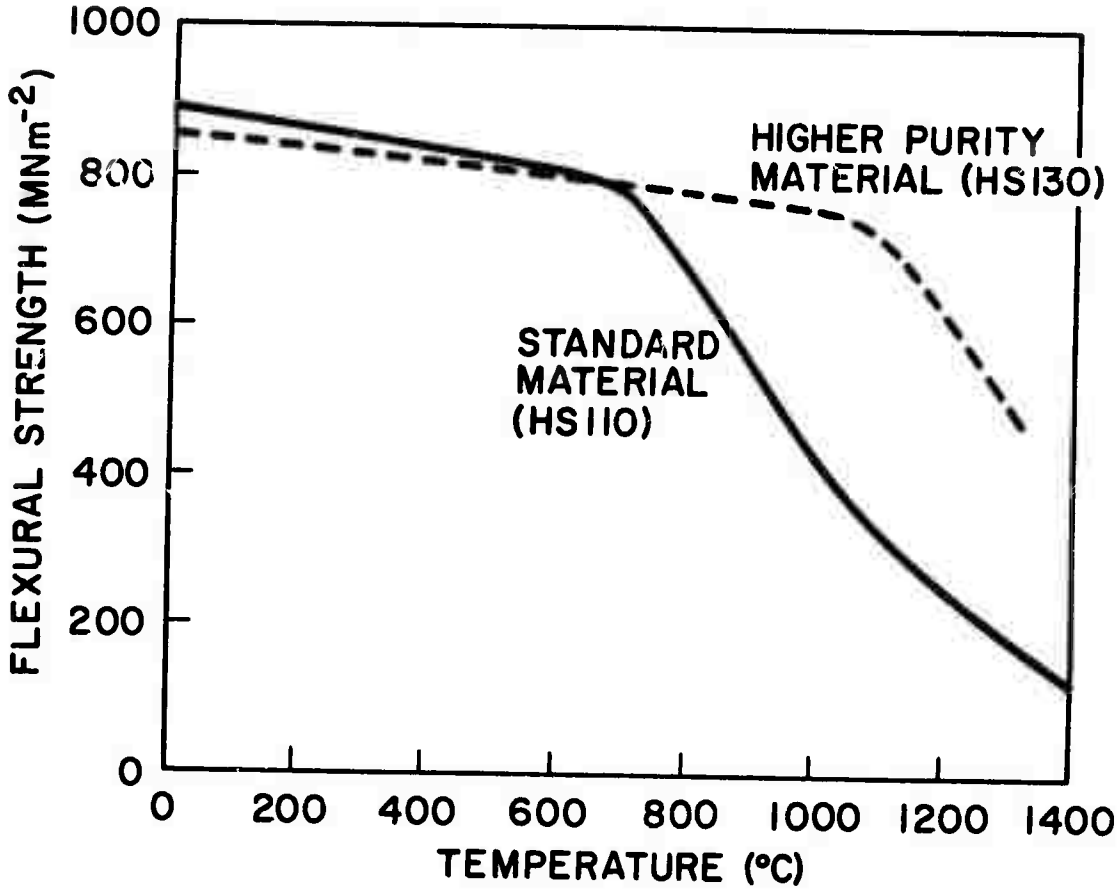


Figure 2.

The temperature dependence of strength of hot pressed silicon nitrides produced by Norton Company.

This suggests two potentially effective approaches* to strength enhancement, (a) a reduction in flaw size to produce a strength increase at all temperatures, (b) a microstructural modification to reduce the rates of slow crack growth and creep. These approaches are considered separately.

3.1 Flaw Size Reduction

A reduction in the size of the flaws, C, that control strength at the low strength extreme will result in a strength increase approximately proportional to $C^{\frac{1}{2}}$.[8] There is therefore a substantial benefit to be obtained by studies of relatively large flaws, leading to suggestions for the fabrication controls needed to effect a strength enhancement. This type of study can be performed most effectively by correlating microscopic flaw observations with flaw size calculations based on a fracture mechanics analysis. Then some predictions can be made concerning the flaw size reductions needed to achieve the requisite strength increase.

3.1.1 Application of Fracture Mechanics

It must be first established that the crack extension conditions for the flaws that control strength are similar to those for the large through cracks used to measure K_{Ic}.

There are many conditions which would preclude a correspondence between K_I for small flaws and K_{Ic}. These are presented

*The approach that optimizes K_{Ic} by microstructural modifications has already received extensive attention, especially for Si_3N_4,[6] and there are only limited possibilities for further improvements using this approach.

first as a basis for discussion.

A. The single crystal, or grain boundary, fracture surface energy is substantially smaller than the value measured for polycrystalline material. Flaws smaller than the grain size will extend to the first grain boundary with K_I equal to the single crystal value. For subsequent extension across the adjacent grain boundaries ($\sim$4 of these) K_I must increase (Fig. 3). Eventually it will increase to K_{Ic} measured for a large through crack. The number of grains needed at the crack front for K_I to reach K_{Ic} has not been established due to the inherent difficulties in measurements of the requisite type. Experiments on bicrystals and coarse-grained polycrystals of cubic material could resolve this uncertainty.

B. In elastically or thermally anisotropic materials, internal strains adjacent to the grain boundaries will develop and these will locally reduce the stress needed for crack extension along the first grain. It is unlikely, however, that the number of grains at the crack front needed for an equivalence between K_I and K_{Ic} will be very different from the isotropic case, because the grain orientation relationships are essentially the same.

C. When slow crack growth occurs, the subcritical crack extension is relatively larger (compared to the initial flaw size) for small flaws than for large flaws. The underestimate incurred in K_I evaluations, based on the initial flaw size, C_o ($K_I = \sigma\sqrt{\pi C_o}$), will thus be larger for small pre-existing flaws than for large through cracks.[9] The occurrence of slow crack growth can be

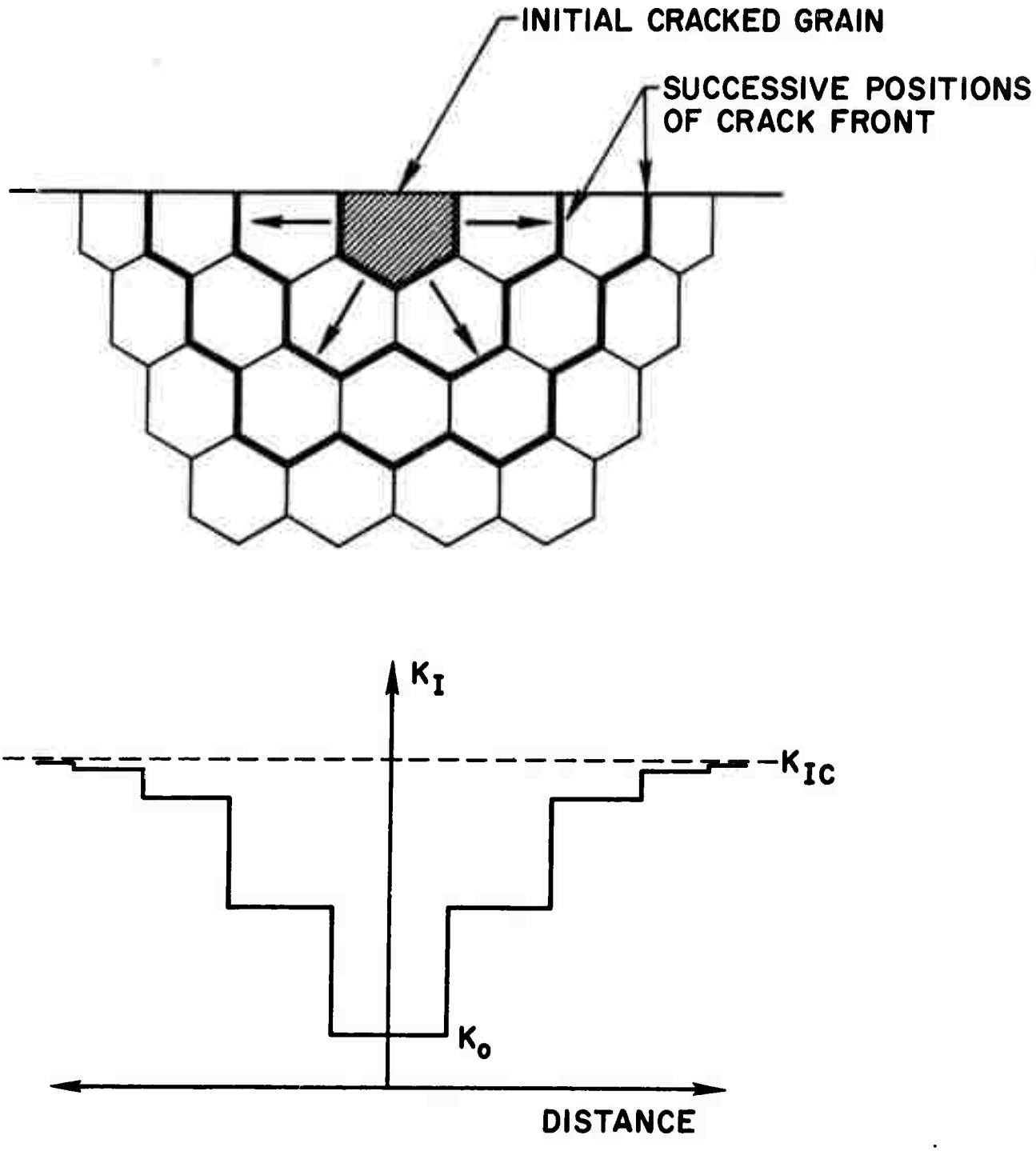


Figure 3.

The effect of crack length on K_I for cracks similar in size to the grain diameter.

detected and analyzed, however, with recently developed techniques,[10] so that flaw size calculations can incorporate the subcritical crack extension. This does not, therefore, limit the application of a fracture mechanics analysis to small flaws.

D. When microstructural inhomogeneities which affect K_{Ic} are present, the crack front must be large enough to intersect several of these so that crack extension will duplicate the extension of a large through crack. Calculations[11] and measurements[12,13] show that the effect of second phase dispersions on K_{Ic} is dependent on the ratio of the diameter to the spacing of the obstacle and the obstacle penetrability. For obstacles with low penetrability--such as well-bonded particles with large fracture strain[11]-- the effect is large and there will be a difference between K_{Ic} for large through cracks and K_I for flaws less than $\sim$3 times the obstacle spacing. For high penetrability obstacles--such as coarse intergranular porosity--the effect is small[14,15] and this is no longer an important consideration.

E. At elevated temperatures where general plastic flow can occur at stresses comparable to the fracture stress, several processes occur which lead to differences between K_I for small flaws and K_{Ic} for large flaws: (i) the plastic zone around small flaws interacts with the surface giving values of $K_I < K_{Ic}$--this has been observed for polycrystalline MgO[16] where $K_I < K_{Ic}$ for flaws $\lesssim 10G$ (where G is the grain size); (ii) subcritical extension can occur due either to the generation of flaws in the plastic zone[17] or to the stress enhancement at the crack tip associated with

dislocations (of the appropriate sign) generated outside the plastic zone which propagate towards the crack.[18]

F. Small inherent flaws such as impurity particles and pores are unlikely to have flaw tip condigurations equivalent to the sharp cracks used to measure K_{Ic}. Mechanisms for generating sharp cracks from the relatively blunt inherent flaws--at relatively low stress levels--have been proposed,[8,19] but these are largely qualitative and can only be verified by experiment.

It is apparent that the theoretical development of models for the extension of small flaws is very limited. It is necessary, therefore, to rely primarily on experimental correlations; although discrepancies due to slow crack growth and microstructural inhomogeneities can be allowed for with available models. K_I for small inherent flaws can only be <u>larger</u> than K_{Ic} if the inherent flaws do not form sharp cracks, as described in (F). K_I values <u>less</u> than I_{Ic} could be due to (A), (B) or (C). It follows, therefore, that at low temperatures where plasticity effects are negligible, K_I for flaws significantly larger than the microstructural features <u>cannot</u> be less than K_{1c}, once the extension due to slow crack growth has been incorporated. Experimental measurements of K_I for small flaws in Si_3N_4 are examined in terms of these deductions, in an attempt to assess the merits of a fracture mechanics approach to materials development.

Three independent measurements of $K_I(\gamma)$ have been made for hot pressed silicon nitride.[7,20,21] These are summarized in

Table I. The values obtained may be compared with $K_{Ic}(\gamma_i)$ values,[22] plotted in Fig. 4.* There is a very close correlation between K_I values obtained for the impurity particles and K_{Ic}, at both 25°C and 1300°C, indicating that $K_I \simeq K_{Ic}$ for flaws $\gtrsim$ 100μm in diameter (grain size ∿3μm). These data suggest that fracture mechanics may be used as a quantitative technique in materials development (for flaws larger than ∿30G), to produce materials with elevated temperature strengths at the low strength extreme $\gtrsim$ 60,000 psi-- equivalent to a flaw size of 100μm. (The results on the indentation flaws introduce a discordant note, however. These flaws are larger than the inherent flaws and larger than the microstructural inhomogeneities and should therefore give a good correspondence between K_I and K_{Ic}. Additional experiments of this type are needed to determine whether the discrepancy between K_I and K_{Ic} can be satisfactorily resolved.)

3.1.2 Chemistry and Processing

The flaws that control the strength of the best available hot pressed silicon nitrides have been identified primarily as impurity particles. Several analyses of these particles have indicated that many of them are iron (often found associated with tungsten) or silicon. Methods for reducing the size of the iron and silicon impurity particles should thus lead to strength en-

*It is apparent that K_I is never > K_{Ic} indicating that the inherent flaws--impurity particles--are converted into sharp cracks prior to fracture at relatively low stresses, as described in (F); this is consistent with observations made with several aluminum oxides.[15]

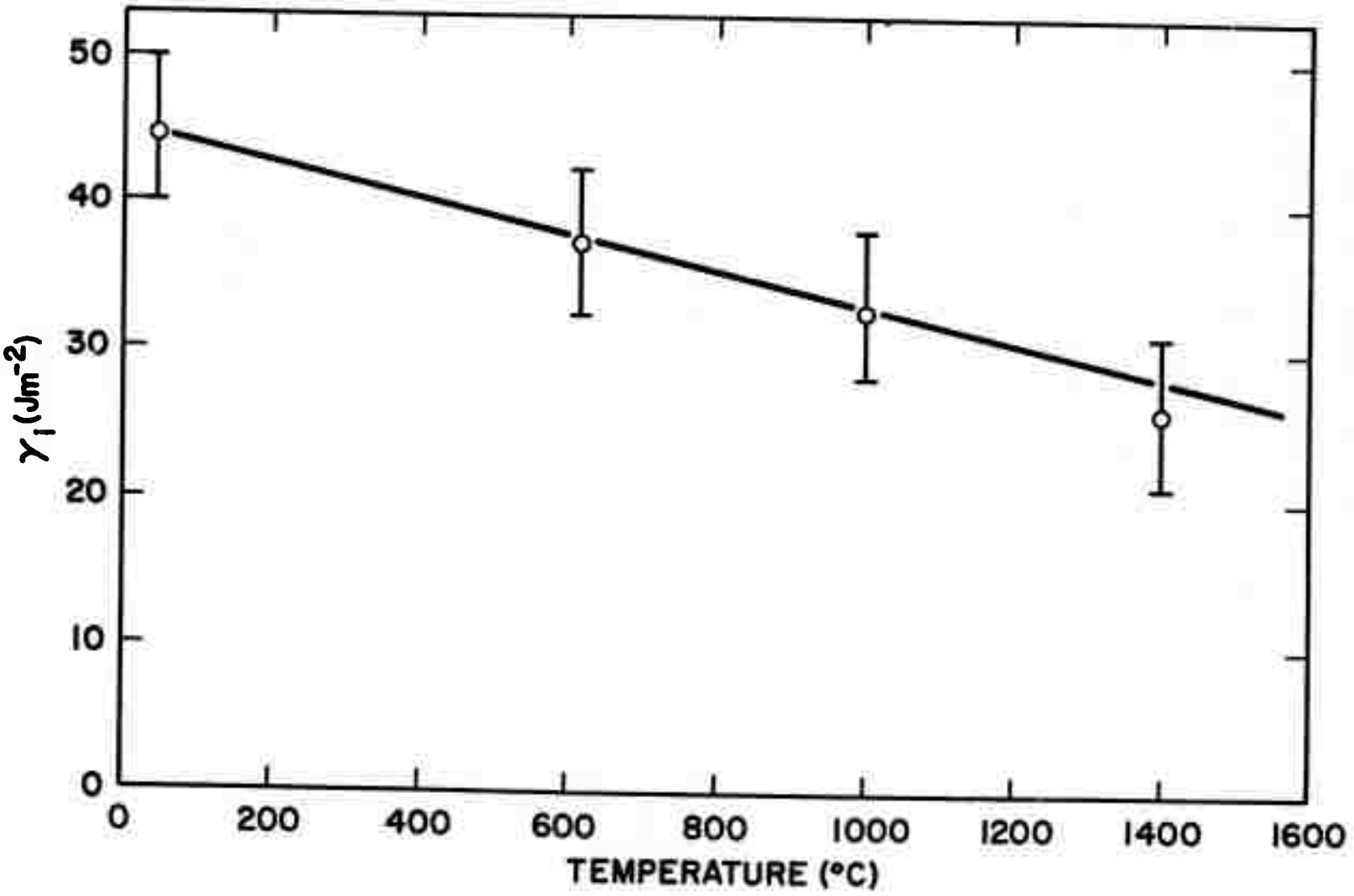


Figure 4.

The effect of temperature on $\gamma_i(K_{Ic}^{\frac{1}{2}}/2E)$ for hot pressed silicon nitride.

hancements, particularly at the low strength extreme.

Removal of metallic inclusions is best achieved by leaching after milling, perhaps through the sequential use of organic acids/bases--if the use of aqueous acids/bases is undesirable, due, for example, to the formation of $Si(OH)_4$. Also, milling with copper--rather than iron-slugs is often an advantage because of the relative ease with which copper can be dissolved (one wash in HNO_3 is sufficient, whereas removal of iron requires multiple treatments). Presently, MgO is added to the Si_3N_4 powder prior to grinding. The purpose of this addition is to facilitate densification in hot pressing. MgO (or other additives) could be added to the mix after clean-up by short-time milling, or by the addition of aqueous $Mg(NO_3)_2$--which decomposes to MgO upon heating.

3.2 Slow Crack Growth and Creep

Hot-pressed silicon nitride deforms at high temperatures by a combination of slow crack growth and creep. These effects have been attributed to the presence of a small amount of residual glassy phase at the grain boundaries. There is no defective evidence for this, but there are circumstantial indications, i.e., some glass has been observed in transmission electron microscopy,[7,22] and a reduction in the CaO and Al_2O_3, content has substantially reduced the slow crack growth and creep rates. It would appear therefore that the elimination of this residual glassy phase should substantially improve the high temperature mechanical properties of hot pressed silicon nitride. A dis-

cussion of hot pressing in the presence of a liquid phase is thus presented, suggesting various possibilities for the elimination of a glassy phase in the final product.

The results to date suggest that densification of Si_3N_4 and SiC during hot pressing at low pressures (< 10,000 psi) only takes place when a liquid is present. The models for liquid phase sintering establish the following conditions as criteria for operability:[23]

The liquid must completely wet the solid, the solubility of the solid in the liquid must be fintie, that of the liquid in the solid should be small. These quantities cannot be quantatively specified for operability independently; they are, however, important in governing the volume fraction of liquid (V_ℓ) which will exist at equilibrium for a given quantity of second component (X_o) added to the primary phase. The mass balance of the second component is: $X_o = X_\ell V_\ell + X_s(1-V_\ell)$. Where X_ℓ and X_s are the mole fractions of the second component in the liquid and solid, respectively. The ratio of these quantities is called the distribution coefficient: $X_s/X = k_o$. Rewriting the above expression using this definition gives:

$$X_o = X_\ell(V_\ell(1-k_o) + k_o)$$

which, for systems with limited solubility ($k_o << 1$) can be abbreviated to read:

$$X_o \simeq X_\ell V_\ell$$

When the volume fraction of liquid is less than the volume of the space not occupied by the packing of the solid grains, densification of the compact requires that the crystals of the solid phase undergo change in shape by some mechanism of atom transport. When a liquid phase is present it is presumed that the solid dissolves into the liquid from regions blocking consolidation (and therefore under stress), diffuses through the liquid and re-deposits at locations on the crystal surfaces which are approximately stress-free. In hot pressing, a compressive (or shear) load is imposed on the area blocking consolidation, the dominant mode of which may be sliding of the crystals relative to one another to re-pack into a more dense assembly. Flow is then controlled by dissolution of asperities/blocking grains. The important point is that some solubility of the solid in the liquid is essential, and it must also be stress dependent. In detail, the overall rate of consolidation can be controlled either by the rate of dissolution of the solid in the liquid at the solid/liquid interface, or by the rate of diffusion of the constituents of the primary phase in the liquid.[24]*

In the above framework the equilibrium relation was used. This would approximately apply to powder systems which have been pre-equilibrated by heat treatment at the sintering temperature, followed by quenching, grinding the product, fabricating compacts

*Dissolution rate appears to be the rate-controlling process in sintering of WC-Co, diffusion in the liquid is presumed to be rate controlling in silicate-based ceramics. Definitive proof is lacking in both cases.

and finally sintering/hot pressing. In powder metallurgy practice, this is referred to as use of "pre-alloyed" powders. However, most systems are not pre-equilibrated. Even those which are will probably undergo some segregation due to cooling and may not re-equilibrate until some time has elapsed after reheating to the sintering temperature. At the opposite end of the scale are the systems furthest removed from equilibrium as the sintering temperature is reached. These are comprised of mixtures of the components in pure forms: no A in B, and no B in A. At temperatures above the eutectic, the constituents react to form liquid and assuming $\alpha_{(A)}$ is the solid phase in equilibrium with the liquid, B diffuses into solid α_A at a rate governed by lattice diffusion of B in α. The end state depends on whether $X_o < X_s$ or $X_o > X_s$. In the former case the liquid will eventually disappear whereas in the latter an equilibrium V_ℓ will remain at the sintering temperature. The former case provides a possibility of utilizing liquid phase sintering (which is attractive because of the faster rates than solid state sintering) while ending up with a single phase solid at the end of the process. The material might be processed in this manner if the solubility of B in α does not decrease too rapidly with decreasing temperature, (or if the α phase field is broad enough) so that it will be in the single phase region at the temperature of service.

Because the phase equilibria are not known, it is impossible to assert just what the probability of success would be for utilizing this scheme for Si_3N_4 with variable O_2 contents and

MgO additions. Nevertheless, there are several findings which suggest that it may work.

(1) O_2 free Si_3N_4 is nominally the β form, whereas the α form is denoted as $Si_{11.5}N_{15}O_{0.5}$. This difference in oxygen content suggests that silica films could be absorbed in the crystal phase. The starting powder should be oxygen free (in the lattice).

(2) MgO additions enhance the sinterability of both α and β Si_3N_4, starting with ground powders, but does not work with nitrided silicon. A tentative conclusion is that MgO has entered the lattice and is not available to form a liquid during hot pressing. It is noteworthy that these observations have not been duplicated. In view of the long-range plans for the utilization of these materials, multicomponent phase equilibria should be determined for the systems from which the empirical data have shown great promise.

Si_3N_4 w.r.t. O_2, Al_2O_3, Mgo--and the prominent elements these materials will encounter in service from the fuels, ash, and other salts. In fact, advantage might be taken of nominally aggressive phases which might form from the above set to form the liquid phase during sintering. This might also be advantageous in that the samples can be partially pre-equilibrated with respect to the environment in which they must operate.

The reasoning which leads to the hypothesis that "disappearing liquid phase sintering" can be expected to work originates from the different time dependence, size effects, and

diffusivities or reaction rate coefficients which pertain to separate phenomena taking place.

(A) Diffusive homogenization of the second component into the particles of the solid phase depend on the diffusivity of B in $\alpha(D_\alpha^B)$ and the particle size of the solid ($\bar{d}$). To first approximation, the homogenization time (t) is given by:[25]

$$(\bar{d})^2 \simeq D_\alpha^B t \qquad (1)$$

(B) During densification by grain rearrangement due to sliding over the liquid films, the densification rate ($d\rho/dt$) is proportional to the liquid viscosity (η), the applied pressure (P_a), and the particle size ($\bar{d}$). To first approximation:

$$d\rho/dt \propto P_a/\bar{d}\eta \qquad (2)$$

(C) During the stage of densification when significant grain shape change is required, the process can be controlled by diffusion of A in the liquid (D_ℓ^A) or by the dissolution rate of A into the liquid by a surface-limited process $k_{A\to\ell}$.

(i) In the case when diffusion in the liquid is rate controlling the shrinkage ($\Delta L/L_o$) models give the dependence on size, etc.

$$(\Delta L/L_o)^3 = C(D_\ell^A P_a/\bar{d}^3)t \qquad (3)$$

where C is a constant.

(ii) In the case when the surface reaction is controlling, the shrinkage rate is

$$(\Delta L/L_o)^2 = D(k_{A\rightarrow \ell} P_a)\, t/\bar{d}^2 \qquad (4)$$

Although we do not have the desired data for the particular systems of interest, the models can be applied qualitatively: Since the time dependence of models, A, B, C_I and C_{II} are different and the respective rates are nominally adjustable by changing the particle size, it would appear to be generally possible to adjust the size, and X_o to allow for completion of densification by a liquid process within some time interval less than that required for homogenization--case A.

Also, it is noteworthy that the difference in magnitudes of D_α^B and D_ℓ^A ($< 10^{-10}$ and $< 10^{-5}$cm^2/sec., respectively) give very different completion times for the respective processes (A and C_I). However, these values are very dependent on the specific system compositions, and on the structure of composition of the liquid. Since data are lacking for the systems of interest, no generalizations can be drawn at this time.

3.3 Promising Materials

Much of the preceding discussion has related to hot pressed silicon nitride. There are several other SiC and Si_3N_4 materials with equal, or greater, potential for structural applications. These are considered briefly here.

A. Hot Pressed Silicon Carbide

The hot pressing of silicon carbide can only be conducted at economically feasible temperatures and pressures in the presence of certain additives. Two independent additives are being used,

aluminum oxide* and boron†. Material prepared using an aluminum oxide additive, with careful control of the maximum particle size and impurity content, can be produced with a room temperature flexural strength of 135,000 psi (corresponding to a maximum particle size of ∿6μm) and a strength at 1375°C of 80,000 psi. The boron containing material shows equally promising strengths; material prepared from 1μm diameter SiC powder, using 0.4% boron has a room temperature flexural strength of 106,000 psi and a strength at 1500°C of 65,000 psi. One potential disadvantage of the boron additive is the notorious effect of boron in oxidation enhancement, due to the formation of low melting point oxide glasses. The overall boron content appears to be too small to substantially affect the oxidation rate, but if segregation of the boron into the oxide occurs--as observed in several other materials--the effect could be substantial. It is evidently important to examine the oxidation performance of the boron material prior to extensive fabrication developments.

B. Pyrolytic Silicon Carbide

Pyrolytic silicon carbide can be deposited in relatively thin film form (≲ 200μm) with remarkably high strengths, up to 250,000 psi;[25] this strength is retained up to > 1400°C. There are substantial practical difficulties, however, involved in the fabrication of bulk component by chemical vapor deposition--largely due to the exponential dependence of the deposition rate

*Material produced by Norton Company
†Material produced by General Electric

on substrate temperature, which leads to the formation of nodular structures. Much improved microstructural control can be achieved by physical vapor deposition--using an activated reactive evaporation process-- where deposition is independent of the substrate temperature.[27] This should be considered as an alternative approach to the production of bulk components by vapor deposition.

If the problems involved in the fabrication of bulk components prove intractable, relatively thin coatings on lower strength base materials should be considered. This approach has already demonstrated significant strength improvements, by increasing the strength of the surface--where most fractures originate.

IV. MATERIALS ASSESSMENT

4.1 Slow Crack Growth

Techniques for studying slow crack growth are now available.[10] These entail the use of load relaxation at constant displacement--using a double torsion specimen--to obtain the relationship between crack velocity V and stress intensity factor, K (Fig. 5). The K, V diagrams generated can then be used to predict the important time dependent failure parameters, such as time to failure at constant stress, effects of strain rate on strength, etc. For example, the time to failure at constant stress, ψ, is given by:[10]

$$\psi = \frac{2}{\sigma^2 Y^2} \int \frac{K}{V} \, dK$$

where Y is a geometrical constant. The predicted effects are

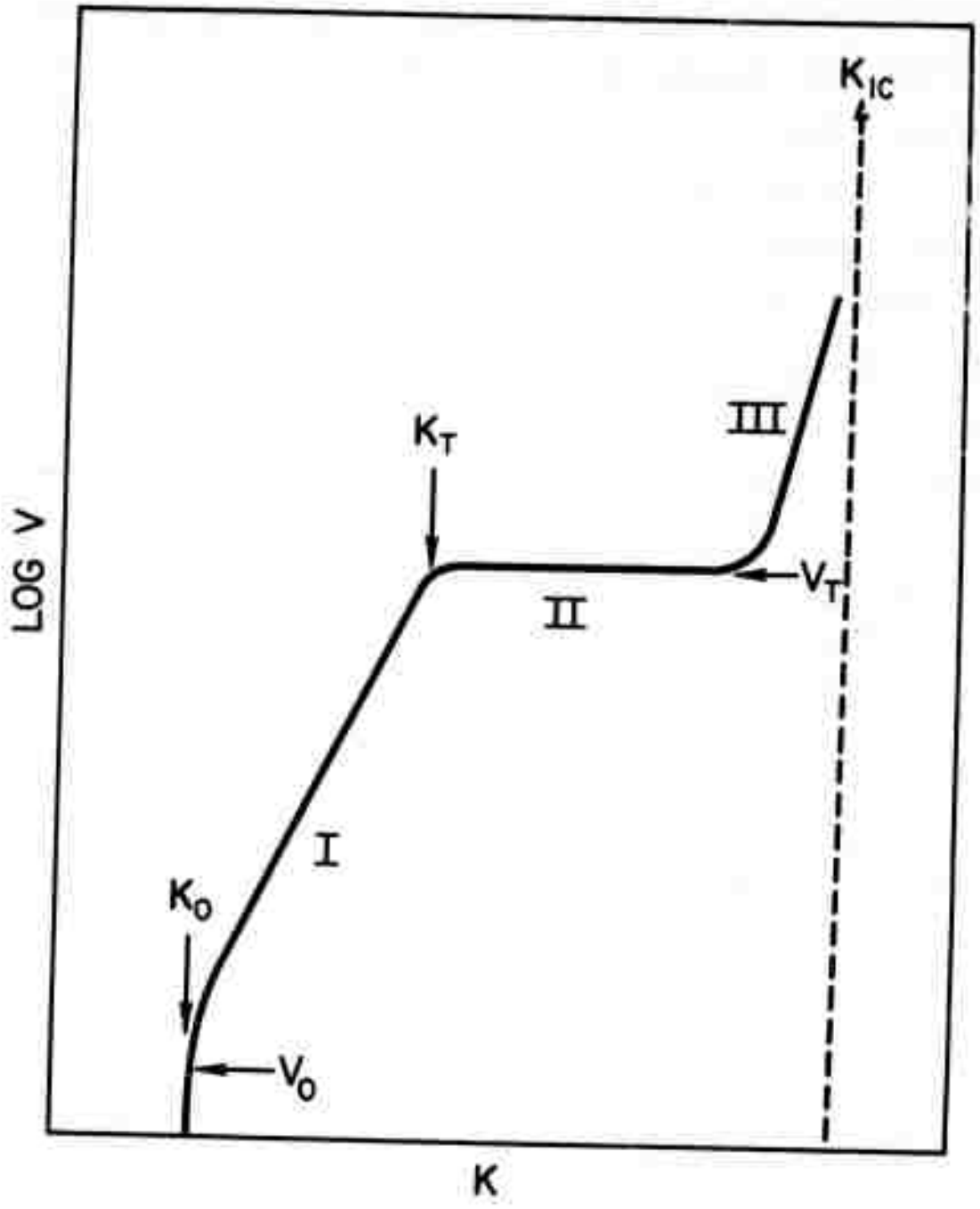


Figure 5.

The schematic variation of crack velocity with stress intensity factor during slow crack growth.

found to be in excellent agreement with time to failure data for ceramic materials because failure in most structure ceramics is exclusively propagation controlled, i.e., there is no flaw initiation stage.

Studies of mechanisms leading to slow crack growth in glass have resulted in a sound understanding of the slow crack growth process.[28] Similar studies for structural ceramics could be equally enlightening and may lead to suggestions for reducing the rate of slow crack growth. These studies should, therefore, proceed in conjunction with the phase diagram/chemistry investigations proposed in Section 3.2.

It is considered, therefore, that the role of slow crack growth in the high temperature fracture of structural ceramics can be assessed* without substantial difficulty, once the appropriate experiments are initiated. Then the relative importance of the effects of slow crack growth and creep processes on the long-term strength may be evaluated, and suggestions for reducing the magnitude of these effects may follow.

4.2 Multiaxial Stresses

The strength of a ceramic component is dependent upon the state of stress in that component.[29] For ceramic turbine components strengths in the biaxial tension and tension/compression

*It is worth noting here that the existence of a slow crack growth limit if of considerable benefit in long-term failure predictions, and the evaluation of this limit (if it exists) should constitute an important part of the slow crack growth investigation.

quadrants are of primary interest. There are very little reliable data in these quadrants and it is apparent that good data on structural ceramics is required. The major problems here relate to the design of the test specimens and fixtures required to maintain a constant volume of material under constant stress of the requisite type, for all stress ratios (this is required to avert problems associated with the effects of volume and stress gradients on strength--see Section 4.3). A sound mechanical analysis of the specimens and fixtures is needed to obtain satisfactory data,[30] with all data obtained preferably on the same type of specimen.

The satisfactory interpretation of the data requires measurements of K_{Ic} for multiaxial stress, in addition to strength measurements. Then the separate effects on strength of the statistical distribution of flaws and K_{Ic} can be assessed. Measurements of this type have not previously been obtained for ceramic materials.

4.3 Statistical and Size Effects

One of the major requirements for the structural application of ceramic materials is the prediction of the strength of a component at a predetermined failure probability. The simplest solution to this is to ensure that the flaws in the component are small so that the stress to extend the largest flaws is less than the maximum stress on the component. The size of the largest permissable flaw can be computated using a fracture mechanics analysis (see Section 3.1) incorporating a slow crack growth analysis where required. This approach requires, of course, that

a fracture mechanics analysis can be applied to small inherent flaws (see Section 3.1), and that flaws of the requisite size can be detected either non-destructively or by proof testing. The potential for this approach is considered in detail in the subsequent section. When this method cannot be used, the only solution lies in a thorough understanding of the statistical aspect of strength.

The basis for using statistical analyses to predict the strength of components from tests on small laboratory specimens is the exact definition of the distribution function at the low strength extreme. The simplest function that is consistent with extreme value statistics is the Weibull distribution function.[31] This gives a relationship for the "risk of rupture," B;

$$B = \int_V \left(\frac{\sigma-\sigma_u}{\sigma_o}\right)^m dV$$

where σ_u is the stress below which there is zero probability of failure, σ_o and m are constants for a given material and dV is a volume element; the integration is conducted over all volume elements containing flaws that may lead to failure. This function is only expected to apply if there is a random distribution of flaws of the same type (i.e., the surface condition and fabrication history of the test materials are identical) and fracture occurs directly from these flaws without sub-critical flaw extension. Even under these conditions, there is no fundamental reason why this distribution function should describe the statis-

tical strength variation and its use must be verified empirically.

For hot pressed silicon nitride preliminary experimental evidence suggests that the Weibull distribution function describes the room temperature strength variations quite closely.[32] The integrations have been conducted over the total volume or total surface area only, whereas fracture mechanics suggests that some weighting factor for surface flaws should be incorporated in a volume integration (if the volume and surface flaws are similar). This gives the risk of rupture, as

$$B = \int_{A\Delta t} \left(\frac{1.4\ \sigma-\sigma_u}{\sigma_o}\right)^m dV + \int_{V_o-A\Delta t} \left(\frac{\sigma-\sigma_u}{\sigma_o}\right)^m dV$$

where A is the surface area, Δt is the depth of the largest surface flaw, V_o is the volume; the factor 1.4 is obtained from fracture mechanics and is the ratio of the stress to extend an internal flaw to the stress to extend a surface flaw of the same size. This modified integration gives an exact prediction of the effects of size on strength and it may be concluded that the Weibull distribution is a precise description of the strength variations in silicon nitride at room temperature. Equally good stregnth predictions have been obtained for pyrolytic SiC at room temperature[26] --another example of a material which fractures directly from pre-existing flaws.

The prediction of the room temperature strength variations are, of course, not the ultimate objective of the analysis. The strengths at the operating temperatures are of primary significance.

If the strengths at these temperatures are still in the regime where fracture occurs directly from inherent flaws (as for pyrolytic SiC) then the room temperature distribution should still apply in principle--because the flaw size distribution is unaffected--with the strengths reduced by the ratio of K_{Ic} values at the two temperatures. When another strength regime is entered, as for Si_3N_4 above ~1000°C, usually the strength distribution will be quite different because the failure mechanism has changed. If plastic flow initiated fracture is the origin of the strength decrease, then the basic criteria for the application of the Weibull distribution no longer apply and it is unlikely that the analysis will give correct predictions. If the strength decrease is due primarily to slow crack growth, however, the initial flaw sizes are still given by the room temperature distribution, so that this distribution in conjunction with an independent slow crack growth analysis should, in principle, enable strength predictions to be made as a function of strain-rate. (This also implies that in a slow crack growth situation, the statistical parameters can only be correlated if tests are conducted at the same strain-rate). It may be concluded therefore that an analysis of failure mechanisms conducted in conjunction with the high temperature statistical analysis may lead to a rational interpretation of the statistical data.

Finally, if statistical analyses are used for component strength predictions, the analysis must be performed for each new batch of material because fabrication history usually affects

the statistical parameters. This introduces the question about the number of tests that need to be performed to obtain valid statistical parameters. This problem has been addressed by an analysis of extreme value statistics, and computer programs have been developed which enable the parameters to be specified with the associated confidence limits.[33]

4.4 Flaw Detection

A very important prerequisite for the successful exploitation of ceramic materials for structural applications is the incorporation of a flaw detection procedure. In principle, this may then be approached in two ways; (a) under load (e.g., during proof testing) using acoustic emission and (b) using techniques that rely on the wave perturbations created by inhomogeneities, e.g., radiography, acoustics, holography. These approaches are considered separately, although both may need to be used cooperatively in a practical flaw detection procedure.

It has been established that acoustic emissions can be obtained due to subcritical crack extension in several ceramic materials--lithium alumino silicate, sandstone and limestone[34] (type I), and vitreous carbon[35] and glass[36] (type II). In type I, materials microcracks are created ahead of the primary crack, prior to fracture, due to microstructural inhomogeneities, etc., and the acoustic emission is associated with microcrack formation. In type II materials slow crack growth precedes fracture and low-level emissions have been detected during the flow growth process. It is established, therefore, that emissions can be obtained in

ceramic materials if catastrophic failure is preceded by processes that lead to subcritical cracking. On this basis it is expected that acoustic emissions will be obtained from hot pressed silicon nitride tested at temperatures > 1000°C, where substantial slow crack growth precedes failure.* It is not evident, however, that emissions will be expected during stressing at room temperature, and it is apparent that a substantial experimental program is required to evaluate the acoustic emissions from structural ceramic materials. Then, it will be possible to assess whether the monitoring of acoustic emissions during proof testing will be a viable flaw detection procedure.

The other flaw detection techniques have the advantage that they do not require component loading, but this is usually counteracted by the disadvantages associated with a more extensive flaw detection procedure, if a thorough screening of flaws is to be achieved. There are numerous techniques that could be exploited for flaw detection; the ones selected for the initial studies are the more conventional techniques, since substantial expertise is available. These are radiographic and acoustic techniques. It seems reasonable to explore the potential of these before devoting a substantial effort to alternative methods. The preliminary experiments[7] indicate that a combination of radiography and

*It is worth noting at this stage that it may be necessary to distinguish between emission due to cracking and emission due to plastic flow, especially if acoustic emission is used in conjunction with high temperature proof testing. Some preliminary research into analytical procedures for achieving this distinction--such as emission summations--are merited.

acoustics is capable of detection internal flaws > 200μm in diameter. Refinements of the acoustic technique should enable smaller flaws to be detected, ≳ 10 times the largest microstructural features, i.e., ≳ 100μm for Si_3N_4. If flaws of this size can be detected consistently, short-term strengths > 70,000 psi at 1300°C can be guaranteed, which would be entirely adequate for most turbine applications. More detailed studies using these techniques could, therefore, establish them as satisfactory flaw detection procedures provided that they can be effectively adapted to routine studies on turbine components.

It remains therefore to detect surface flaws with similar precision. Dye penetrants and surface wave techniques are being tried in this contest, but it is premature to conclude anything about their success.

If it is found that the flaw detection techniques are not capable of detecting flaws of the requisite magnitude, i.e., those that lead to failure at stresses below the operational stress, it may still be possible to achieve reliable failure predictions by combining flaw detection with a statistical analysis. In fact, incorporation of a flaw detection procedure defines quite precisely the minimum strength for the statistical analysis, thereby enabling fewer specimens to be used to predict the working stress at a predetermined failure probability, within specified confidence limits. Ultimately, this may prove to be the most satisfactory approach to failure control.

4.5 High Strain Rate Effects

The thermal strains generated in the turbine components develop quite rapidly. If flaw detection and fracture mechanics are to be used successfully to screen turbine components for operational stresses, the critical stress intensity factor at these high strain rates is required. These can be obtained from Charpy impact experiments, if the machine is instrumented to measure the maximum load during impact. It is suggested that experiments of this type should be conducted for structural ceramic materials. Preliminary data for alumina[37] indicate that the dynamic stress intensity factor is similar to K_{Ic}. (If this is confirmed for other ceramic materials, this test could then be used as a simple routine test technique to obtain K_{Ic}.)

V. RECOMMENDATIONS

A. Additional work aimed at comparing the critical stress intensity factor for the extension of small inherent flaws with K_{Ic} are required. This will establish the flaw size regime in which fracture mechanics can be used for strength predictions; thereby qualifying both failure control and certain materials development programs.

B. Strength enhancement in hot pressed silicon nitride may be achieved by a thorough evaluation of the role of second phase materials. The generation of phase equilibria, with respect to the expected environmental conditions in use, would be of substantial assistance in evaluating the potential for removal of

the phases that lead to strength reductions at elevated temperatures. Also, the elimination of impurities that limit the strength--by acting as fracture origins--is of primary importance; this may be achieved by employing various chemical treatments prior to hot pressing.

C. A study of slow crack growth in hot pressed silicon nitride at elevated temperatures should be initiated for both strength prediction in available materials and to assist in developments aimed at minimizing the slow crack growth. The evaluation of the slow crack growth limit (if it exists) should constitute an important part of this investigation.

D. An investigation of the effects of multiaxial stress on the strength of structural ceramics is required. This should incorporate a specimen design and test procedure that enables the requisite stress to be of constant magnitude over a fixed volume of material, for all stress ratios.

E. Work should continue on statistical strength variations, particularly in relation to establishing the conditions for application of the Weibull distribution function.

F. Flaw detection procedures in ceramic materials require substantial development. Extensive work on detection of acoustic emissions is indicated, and refinements of acoustic techniques of flaw detection should be purused. Finally, work on the detection of surface flaws should proceed, exploiting in particular the use of surface wave techniques.

TABLE I

The Critical Fracture Surface Energy to Propagate Small Flaws in Hot Pressed Silicon Nitride

INVESTIGATOR	FLAW TYPE	TEMP(°C)	$\gamma (Jm^{-2})$
AMMRC[21]	Cracks, Impurity particles (50 - 200μm diameter)	25	42 ± 12
WESTINGHOUSE[7]	Impurity particles (140 - 290μm diameter)	1300*	26 ± 4
NRL[20]	Indentation Flaws	25	∿10

*Tested at a large enough displacement rate to avert significant slow crack growth.

REFERENCES

1. A. F. McLean, E. A. Fisher and D. E. Harrison, "Brittle Materials Design, High Temperature Gas Turbine," AMMRC CTR 72-3, 1972.

2. A. F. McLean, "The Application of Ceramics to the Small Gas Turbine," ASME Publication No. 70-GT-105, 1970.

3. G. A. Sines, private communication.

4. E. Ryshkewitch, unpublished work.

5. Data supplied by Norton Company on their materials.

6. F. F. Lange, unpublished work at Westinghouse Research Laboratories.

7. R. Kossowsky, unpublished work at Westinghouse Research Laboratories.

8. R. W. Davidge and A. G. Evans, Mater. Sci. Eng. 6 (1970) 281.

9. A. G. Evans, Phil. Mag., to be published.

10. A. G. Evans, Inter. J. Fracture Mechanics, to be published.

11. A. G. Evans, Phil. Mag., to be published.

12. F. F. Lange, J. Amer. Ceramics Soc. 54 (1971) 614.

13. F. F. Lange and K. C. Radford, J. Mat. Sci. 6 (1971) 1197.

14. L. E. Simpson, J. Amer. Ceramics Soc., to be published.

15. A. G. Evans and G. Tappin, Proc. Brit. Ceram. Soc. 20 (1972) 275.

16. A. G. Evans and R. V. Davidge, Phil. Mag. 20 (1969) 373.

17. A. G. Evans, D. Gilling, and R. W. Davidge, J. Mat. Sci. 5 (1970) 187.

18. F. J. P. Clark, R. A. J. Sambell and H. G. Tattersall, Phil. Mag. 7 (1962) 393.

19. R. W. Davidge and T. J. Green, J. Mat. Sci. 3 (1968) 629.

20. R. W. Rice, unpublished work at NRL.

21. F. Barrata, unpublished work at AMMRC.

22. A. G. Evans, unpublished work at Harwell.

23. W. D. Kingery, *Introduction to Ceramics*, John Wiley, New York, 1960, Chap. 12.

24. J. Brophy et al., Trans. AIME 230 (1964) 769.

25. W. D. Kingery, *Introduction to Ceramics*, John Wiley, New York, 1960, Chap. 8.

26. A. G. Evans, C. Padget and R. W. Davidge, J. Amer. Ceramics Soc., to be published.

27. R. F. Bunshah, UCLA Report UCLA-ENG-7227, April 1972.

28. S. M. Wiederhorn, *International Conference on Corrosion Fatigue, Stress, Corrosion*, June 1971, to be published (NBS Report 10 565, April 1971).

29. See for example, H. W. Babel and G. Sines, J. Basic Eng., p. 285, June 1968.

30. H. W. Babel and G. Sines, ASTM Journal of Materials 3 (1968) 134.

31. W. Weibull, J. Appl. Mechanics 18 (1951) 293.

32. W. Fraser, unpublished work at Westinghouse Research Laboratories.

33. F. A. McClintock, to be published.

34. R. Hoagland, G. T. Hahn, A. R. Rosenfeld, R. Simon and C. D. Nicholson, ARPA Report 1579, 1971.

35. G. L. Tingey, ARPA report 1830 (1971-1972).

36. G. Alers, Report for Naval Air Contract, No. N00019-71-C-0344.

37. R. Bertolotti, to be published.

A WORKSHOP ON FRACTURE DATA HELD AT CENTERVILLE, MASSACHUSETTS July 20, 21, 1972

Report Prepared by Charles Grosskreutz
National Bureau of Standards

Abstract

A workshop organized by the National Bureau of Standards was held under the auspices of the Materials Research Council to review needs for compilations of fracture data. The recommendations will be used to design an appropriate activity in Fracture Data at the Bureau of Standards. Participants from three basic industries were invited: automobile and construction equipment, railroad, and pressure vessel and electrical generating equipment. In addition, professionals from the academic community and the National Bureau of Standards attended. Recommendations for the development of two types of fracture data compilations were made, one for a simplified introductory compilation for day-to-day design purposes. This handbook is intended to alleviate the paucity of modern design handbooks oriented to fracture design control. A second, more complete fracture data compilation again directed at the design community is also needed, and the specific subjects are outlined in the report.

A WORKSHOP ON FRACTURE DATA HELD AT CENTERVILLE, MASSACHUSETTS

Report Prepared by Charles Grosskreutz
National Bureau of Standards

FOREWORD

The National Bureau of Standards has traditionally provided evaluated data compilations to the scientific and technical community of the United States. Since 1963 the focus of this activity has been the Office of Standard Reference Data, which manages a coordinated program (NSRDS) for compiling, evaluating, and distributing important property data. Although the original charter of the NSRDS program included data on mechanical properties, funding restrictions have so far limited activity to a smaller class of physical and chemical properties (thermodynamic, transport, atomic and molecular, etc.). In keeping with its mandate to assure maximum application of physical and engineering sciences to the advancement of technology in industry and commerce, the Bureau is now seeking to extend its active data program to include mechanical properties needed by design and materials engineers in industry. To assist in the definition of the Bureau's proper role in this activity and to identify the most effective means for communicating data for use in the avoidance of failures, a Fracture Data Workshop was held

on July 20 and 21 at Centerville, Massachusetts, under the auspices of the ARPA Materials Research Council.

This Workshop was organized around the fracture data needs of three basic industries: automobile and construction equipment, railroad, and pressure vessel and electrical generating equipment. People representing each of these industries were invited to attend, along with professionals from the academic community and NBS whose interest and expertise lay in the fracture of materials. (A list of attendees and their affiliations is given in an Appendix to this report.) The first day of the Workshop was devoted to an overview of fracture control practices and fracture data usage in the three different industries represented. The second day was given over to general discussions which were directed toward defining the nature and types of fracture data compilations which would be of greatest use to the industries represented.

This report, which is in the nature of a "position paper", is intended to record the highlights of these two days' activities and present in concise form the conclusions and recommendations of the workshop participants.

I. General Usage of Fracture Data in Industrial Design

Workshop members heard from representatives of each of the three industries (see Foreword) who gave an overall view of the fracture control practices as they saw them. A brief description of some of the highlights of these presentations are

given in the following paragraphs.

A. Automobile and Construction Equipment Industry

Fatigue is the most prevalent failure mode which is encountered in the automotive industry. In the preliminary design stage, fatigue life predictions are made which are based largely on cumulative damage procedures which make use of selected fatigue and fracture properties of the materials. Materials selection, especially for critical components, is influenced to some extent by the fracture properties of these materials. Fatigue life predictions are also made for prototype hardware, based on known fracture properties of the material and test track load inputs.

In the construction equipment industry, fatigue is again the dominant mode of failure. Smooth specimen, constant load fatigue data is used for the design of engines and box beams used in the construction of heavy earth-moving equipment. In addition to the use of smooth specimen fatigue data for design, the construction industry is using cyclic stress-strain behavior of materials in making life estimations. One of the chief concerns in this industry is that of evaluating life of welded structures.

Fracture mechanics has not been used extensively due primarily to the use of low carbon materials of relatively thin cross-sections in the equipment structures. Furthermore, no mechanism is available in the hands of the user for evaluating crack size from which rate of growth could be analyzed. Within the past year the need for application of fracture mechanics has

been made somewhat more relevant in safety frames which must meet certain legislated strength requirements down to a temperature of zero degrees Fahrenheit. Such requirements involve the transition temperature characteristics of low carbon materials.

B. Railroads

The design practices in this industry are very old, and very little usage is made of modern fracture control concepts. However, failures and accidents are kept at a low rate by the use of sophisticated, nondestructive testing (NDT) programs. For example, incipient cracks in rails are detected by a magnetic device carried by an instrumented inspection car. The sensitivity of this device is such that flaws can be detected before they reach the critical size which would lead to catastrophic failure. Reliability of operation and service is of increasing concern to this industry, and fracture control programs which are based on modern design practices are being introduced at present both in the design and in the operation of equipment.

C. Pressure Vessels for Nuclear Power Generation

Fatigue and stress corrosion cracking (SCC) represent the two largest modes of failure in this industry. Because of high public visibility and the potential for catastrophic failure in nuclear power generating plants, fracture control is a most important ingredient in design procedures in this industry. Extensive use is made of fracture mechanics and NDT to define and detect critical flaw sizes. The industry has generated extensive

data compilations which are needed in these design procedures. Moreover, the industry is highly regulated by codes which specify quality assurance and inspection both during construction and operation of electrical generating plants.

D. Summary

These three industries reflect a wide spectrum of procedures and degree of sophistication in the use of fracture data for failure avoidance both in design and operation. However, in all cases, failure avoidance and reliability are prime motivating factors in their design departments. For those components which occupy a critical position in a given vehicle or structure, the availability of valid fracture data is extremely important in the design and operation of that equipment. To a lesser degree, fracture data plays a role in the selection of materials. Other contributing factors in materials selection are: formability, machineability, and availability. One additional major point of interest which emerged from these presentations was that there does not exist in the United States a useful collection of failure data and case histories for these industries. This lack is due in part to a fear on the part of the industries that their competitive position would be threatened by the release of such information. On the other hand, a resolution of many failure problems could be enhanced and expedited if such information were readily available. In countries such as the United Kingdom and West Germany where various industries are nationalized, rather

good collections of failure data are available*.

II. Fracture Data Compilations in Industry

A. General Impressions

It became clear during the Workshop discussions that large, knowledgeable industries (automotive, pressure vessel) are for the most part self-sufficient with respect to fracture data. That is, these companies will always perform verification tests and measure the fracture properties of interest in their own laboratories, especially for critical components. Moreover, companies in these industries often maintain their own development laboratories in which new design and fracture control procedures are being developed. If these procedures require new or different fracture property data, then these laboratories will make their own measurements and compilations. Therefore, it was the consensus of the Workshop participants that the present need for fracture data compilations was marginal for such large, knowledgeable industries and that NBS could not count on them for enthusiastic support of such a project.

However, the respresentatives from these industries at the Workshop all agreed that such a data compilation would be extremely valuable for preliminary design purposes. Moreover,

---

*The airline industry (a user, rather than a manufacturer of products) in the United States has an excellent program for exchanging information on structural and mechanical failure in their operating fleet of aircraft. This situation is due primarily to the advantage which accrues to each airline to obtain this information quickly to prevent accidents and loss of life.

many smaller companies exist which do not maintain the measurement and test laboratories to generate the requisite data. It was, therefore, concluded that in spite of the self-sufficiency of many of our larger industries, that a need did, in fact, exist for fracture data compilations for the reasons just cited. It was further pointed out that as the data compilation grows and the number of contributers increases, that in five-ten years even the larger industries would find the compilation useful and important to their design and fracture control operations.

It was generally agreed that the most useful fracture data compilation would display the whole range of fracture data, from sophisticated K_{Ic} and da/dN vs. ΔK curves to the traditional parameters like yield stress and reduction in area. With such a wide spectrum of data there is a much better chance that any given designer will use it and that he may be led to the more sophisticated concepts in the long run.

B. Some Specific Needs

Workshop participants raised a number of specific needs which should be met in a fracture data compilation. These items are listed below with a short, explanatory sentence following each.

1. Cast Materials Data. The fracture properties of castings are becoming more important as industry seeks to reduce the costs of machining operations. At present, there is very little data on these properties which can be easily found.

2. Effect of Manufacturing Processes, such as Forging, Stamping, Welding, etc., on Fracture Properties. Although the fracture properties of materials are often available for various pre-treatments, design engineers must know what these properties will be after a component has been fabricated from the parent metal. Obtaining this data will more than likely require an extensive laboratory test program.

3. The Effects on Residual Stresses on Fracture Properties. Uncertainties in the residual stress pattern must be taken into account in the application of the principles of fracture mechanics in the design of components.

4. Fracture Properties of Joints. Almost all products contain joints of one kind or another, e.g., welded, bolted, adhesive, etc. In many instances, failure occurs at joints, but very few data compilations include their fracture properties. It was the consensus of the Workshop that the fracture properties of welded joints should receive first priority in this category.

5. The Effects of Abrasion and Wear on Fracture Properties of Metal.

6. Nondestructive Testing Data. The application of fracture mechanics to design would be greatly enhanced if a tabulation of the minimum flaw size detectable by a given NDT technique were available. Such a tabulation would probably take the form of curves which show the probability of finding a flaw of a given size using a particular NDT method and the materials to which it is applicable.

C. The Need for Introductory Material in Fracture Data Compilations

The Workshop was unanimous in its opinion that the possibility of fracture data misuse required that carefully written introductory or prefactory material be included in any compilation. The introductory material would stress the variability to be expected in fracture data for a given material, and state explicitly the conditions and limitations on the use of fracture data for design purposes. The fracture problem would be discussed quite incisively. Such questions as: "How do you test?", "Why do you test?", "What is the meaning of the test results?", and "How are the results used to design against fracture or to predict life to fracture?", would be included. This concern expressed by the Workshop reflects the paucity of modern design handbooks which are oriented toward fracture control. Additional details about this part of the data compilation are discussed more fully in the following section.

III. Format and Content of Fracture Data Compilations

A. General Philosophy

There was unanimous agreement that handbooks are the best format for communicating fracture data to the user. The concept of a National Fracture Data Center with computer access to data and/or literature references was not viewed as a viable, workable entity. The Workshop agreed that both the format and the content of a fracture data handbook should be tightly coupled

to current and projected industrial design procedures, so that the data will, in fact, be used at the operating level. Such coupling presupposes that the people who compile the handbook are knowledgeable concerning current and projected design procedures. It was strongly suggested that alternative methods of data presentation be circulated to potential industrial users early in the project so that a useful format can be determined. Other suggestions for achieving the desired coupling included the formation of an industry-dominated advisory panel and placing selected NBS personnel in the industrial environment for limited periods during the project.

The problem of fracture data variability and criteria for selecting "best" values for the parameters was met in the following way by the Workshop. Because most of the data variability is a result of differences in microstructure, prior mechanical history, composition, heat treatment, test method, and specimen geometry, it was agreed, insofar as possible, that all valid data for a given material be included in the compilation and that the material and test methods be specified for each datum. Thus, instead of attempting to give a "best value" with an uncertainty range, the handbook would provide the entire spectrum of data so that the design or materials engineer could look at this data and select that which is most relevant to his particular problem.

It was further agreed that the contents of the compilation be limited to brittle, ductile, or fatigue fracture modes,

as distinguished from such time dependent modes as stress corrosion, or stress-rupture.

The important question on how to educate the user of the fracture data compilation is solved by the introductory material mentioned in a previous section. Not every design engineer can attend a short course nor does every company have the resources to reeducate its engineers formally from time to time. Thus, the prefatory material for a fracture data compilation takes on added significance as it is required to perform the important function of education for data use. As one member of the Workshop put it, we are in the position to influence design practice both now and in the future by the data which we make easily available and useable.

B. Specific Suggestions on Format and Content

It was unanimously recommended that the fracture data compilation be prepared in two volumes. The first volume would contain the introductory material which has already been discussed. During the Workshop discussions, this volume was variously referred to as prefatory, introductory, or a missionary volume. During the brief time available for discussion, there was not complete agreement on the content of this first volume, but several valuable suggestions were made which are summarized below. The second volume is envisioned as an extensive compilation of fracture data for a large number of materials.

1. Volume One, The "Missionary Volume". This volume should be tutorial both in its objective and in its makeup. Furthermore, it should be property oriented and might contain simple one-dimensional diagrams to show the range of such fracture properties as ultimate tensile strength, K_{Ic}, reduction in area, etc., for representative materials. The book's primary mission would be to show how and where to use fracture data and some of the pitfalls which may await the inexperienced user. For example, the effects of specimen size of fracture toughness measurements would be discussed. It was the unanimous consensus of the Workshop that a specimen metal should be chosen, such as 4340 steel, which is widely used and for which much data exists and that the handbook should illustrate the use of fracture data for design with this material. Several design procedures for fatigue life prediction might be illustrated, and calculation of critical flaw size for a given specimen geometry could be illustrated. In each of the examples, pitfalls such as environmental sensitivity or changes in properties due to subsequent heat treatments or manufacturing operations would be pointed out. A section should appear to illustrate how simple stress analyses are made for fillets and other standard component configurations so that K_{Ic} data can be effectively utilized. The origins of data variability should be discussed, and test procedures and sample geometries should be described which yield valid fracture data.

Workshop participants were divided on the scope of the contents of Volume One. One group argued for keeping the volume simple and tutorial, with only a few examples to illustrate the correct usage of fracture data and to point out some precautions and hazards in its use. The other group wanted to go further and add some fracture data compilations for representative materials.

2. Volume Two--Extended Fracture Data Compilation. This volume should be material oriented. That is, the book would be divided into sections, each of which is given over to a material or class of materials for which all the known fracture properties are compiled. In this sense, it would follow the format of the Aerospace Structural Metals Handbook, which is published by the Mechanical Properties Data Center. This volume should be an extended data compilation offering a wide spectrum of choices to the design engineer. It was strongly suggested that data for each material be given in matrix form so that dependence of the fracture property on composition, metallurgical structure, heat treatment, specimen size, test method, etc., would be clearly indicated. Moreover, if the dependence of the fracture property on any of these is unknown, a blank space should appear so that the user is aware of this gap in the data.

C. Effects of Environment, Temperature, and Time

Considerable discussion was given to the problem of displaying the important effects of environment, temperature and time on fracture of materials. The question of temperature

dependence is probably dealt with most easily. In many cases, systematic measurements of fracture properties have been made as a function of temperature and can simply be included in the above two volumes. However, the synergistic interaction of environment, temperature and time raises a more fundamental problem. The problem is partially allieviated by the decision to restrict the data compilation to such fracture modes as brittle, ductile, and fatigue fracture; nevertheless, there is still an obligation to record, where possible, the effects of these parameters on fracture data. While there was no agreement on how to achieve this, the final consensus was that environmental influences should be included only when the results are clear, and that variations with temperature be included only when they are well-known. Nevertheless, clear statements should be made both in Volumes One and Two to the effect that unknown environmental or temperature effects may exist which could vitiate fracture data for some applications. A stress corrosion cracking handbook is now in the planning stage (see the Report of ARPA Materials Research Council dated 7-10-72 of the Stress Corrosion Group) and could be used as a cross reference.

It was suggested that a general warning be given in both volumes to the effect that possible influences on fracture behavior can be exerted by working and recrystallization textures, residual impurities and inclusions.

IV. Specific Materials and Properties for Fracture Data Compilation

In the short time which was available, the Workshop undertook to suggest specific materials and specific fracture properties which might be included in the data compilation (Volume Two). To limit the options somewhat, the NBS Organizing Committee had restricted the discussions of this Workshop to structural ferrous materials, and these are the materials from which the Workshop made its specific recommendations. However, it was the unanimous opinion of Workshop participants that other materials, particularly Al and its alloys, be included even in the first editions of the data compilation.

A. Materials

Material recommendations are listed in Table I and are grouped according to the interests of the three industries represented at the Workshop. Those materials which are preceded by an asterisk are those which are believed to have been most widely investigated for fracture properties. These materials would then represent the logical starting point for data collection.

B. Fracture Properties

There was general agreement that traditional mechanical properties, such as UTS and yield strength, should be included in the compilation for the sake of completeness. There was considerable diversity of opinion over which of the more sophisticated, modern fracture properties should be included. The

TABLE I

Representative Structural Ferrous Materials

Ground Transportation Industry

Nodular Irons

*Grey Cast Iron (30,000 UTS)

Hot Rolled Steels
- 1020 (A515)
- *1010
- TC 128B

*High Strength Low Alloy
- NAX
- MAR-TEN
- COR-TEN

Boron Steels
- 10B62

Gear Steels
- 8620

Heat Treatable - Medium Strength
- *1045
- 1041
- 1080
- 5160

Die Steels
- *H-11

Stainless
- 200 series
- *300 series
- 400 series

Cast Steels

Pressure Vessel & Electrical Equipment Industry

*Ni-Cr-Mo-V Rotor
Ni-Mo-V Steels (forged)

517 - Large Pressure Vessels
*542 - Petro Chemical Vessels
*533
508 Nuclear Vessels

Cr-Mo-V Turbine Steels
*4140
4340 Electrical Equipment

543 - HY Series - Pressure Vessels for Submarines

Marageing Steels

*Denotes those materials for which extensive data exists and which could be used in a limited effort.

properties listed in Table II are indicative of the spectrum of properties suggested by the Workshop. No attempt was made to select one property over another one for inclusion. And, in fact, it was specifically recommended by one group that all of these properties should be included in the compilation.

V. Execution of the Project

A number of valuable recommendations were made for optimum use of time and money in the collection of fracture data.

A. Time Table

It was the unanimous opinion that Volume One of the compilation should be completed within one year's time. No time limit was placed on Volume Two, although it was suggested that the collection of data for Volume Two be begun concurrently with the work on Volume One.

B. Level of Effort

The effort to complete Volume One was estimated to be between one and two man-years.

C. Data Sources

The major sources of fracture data are suppliers (or vendors), the "knowledgeable" industries (e.g., automotive, pressure vessel industry), DoD experience on various aircraft, (e.g., F-111 experience), technical literature and the Metals Property Council.

TABLE II

Representative Fracture Properties

Common Properties

UTS

σ_y (for designated offset)

E

Hardness (Bhn)

R.A.

elongation (for 1", 2" and 8" gage)

True stress-true strain properties:

- n - Strain hardening exponent
- ε_f - true fracture ductility
- σ_f - true fracture strength

Endurance Limit at cycles

S-N Curves

Brittle fracture Mode:

- K_{Ic}
- $A_{crit.}$ at $\sigma_y/2$
- K_c or R-curves
- NDT Temperature
  - Drop-weight
  - Dynamic Tear
  - Pre-cracked, instrumented Charpy
- Charpy V-notch curves
- Charpy Keyhole curves
- Pre-cracked (slow bend) Bend data

Fatigue Mode:

- ε_{tot} vs. N curves
  - intercept values,
    - ε_f' - fatigue ductility coefficient
    - σ_f' - fatigue strength coefficient
  - slopes of elastic & plastic lines
- da/dN vs. ΔK curves
- Spectrum loading da/dN
- Cyclic stress-strain curve
- K_t - K_f relations

D. The Role of NBS

Several Workshop participants expressed the feeling that NBS should be in this type of activity as part of its future thrust and need to interact more closely with the Nation's industry. However, it was the consensus of the group that NBS did not now have the personnel necessary to achieve the recommendations of the Workshop. It was suggested that either a new man be hired to do this or that an external person be named to work with current NBS personnel. It was deemed important that one man (presumably at NBS) plan and manage the project. The hope was expressed that the external man would be an industry representative and not one of the "knowledgeable" people from the university community.

E. Problems and Barriers for NBS

Nearly everyone expressed the thought that difficulties might arise for NBS, as a government agency, to extract the relevant data from private industry. Suspicion and fear of possible regulation or interference would prevent many industries from cooperating fully. One tactic was suggested (Feltner, Ford Motor Company) that NBS might send an intern to an industrial plant to gain experience in what industry needs and would use in the way of a fracture data compilation. Another possible tactic would be to work through societies, such as the SAE, in which private industry has a considerable stake and to which they contribute time and data for common purposes.

VI. Summary of Conclusions and Recommendations of the Workshop

Although many opinions, suggestions and recommendations were offered by individual Workshop members, many of which are documented in the body of this report, there did emerge a consensus on a few important questions. These consensus conclusions and recommendations are listed below.

1. Although large, knowledgeable industries will probably always generate their own data and data compilations, a definite need exists for a fracture data compilation for the purposes of preliminary design exercises and those smaller companies that cannot afford to make their own verification tests.

2. Because of these needs, the Workshop participants recommend that a fracture data compilation should be made.

3. The best format for communicating fracture data is the handbook. Specifically, such a book should be divided into two volumes. Volume One to contain prefatory material which would educate the user in simple terms and by example on the use of fracture data; and Volume Two which would be an extended compilation of fracture data for representative materials to include at least structural ferrous materials and the aluminum alloys. The scope of the data to be included should be limited to brittle, ductile, or fatigue fracture modes as distinguished from such time-dependent modes as stress corrosion, or stress rupture.

4. A wide spectrum of fracture properties and material conditions should be included in Volume Two so that the user can choose his own "best" value for his particular application. These properties and materials should be coupled closely to current and forecasted design practices in industry.

5. The effects of temperature, environment and time should be included only when the results are clear and the variation of properties with these parameters are well-known. However, cautionary statements should be made that fracture properties are often strongly effected by these parameters.

6. Volume One should be completed within one year's time.

7. Volume Two will require considerably longer time to compile. Sample data displays should be sent out to various industrial users at an early date in the compilation to obtain suggestions for the best format.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with the University of Michigan.

Appendix

List of Attendees
Fracture Data Workshop

Dr. S. H. Bush
Battelle Northwest Laboratories
P.O. Box 999
Richland, Washington 99352

Professor H. Corten
Dept. of Theoretical and
Applied Mechanics
University of Illinois
Urbana, Illinois 61801

Dr. K. H. Cotter
Dept. of the Air Force
Headquarters Aeronautical
Systems Division
Wright-Patterson Air Force Base
Ohio 45433

Professor Daniel Drucker
Dean of Engineering
University of Illinois
Urbana, Illinois 61801

Mr. Martin A. Erickson
Tractor Division
Ford Motor Company
Dearborn, Michigan 48023

Dr. C. E. Feltner
Ford Motor Company
Scientific Laboratory
20000 Rotunda Drive
Dearborn, Michigan 48124

Dr. J. C. Grosskreutz
National Bureau of Standards
B120 Materials Building
Washington, D.C. 20234

Dr. W. J. Harris, Jr.
Vice President
Research and Test Department
Association of American Railroads
1920 L Street, N.W.
Washington, D.C. 20036

Professor John Hirth
Dept. of Mechanical Engineering
Ohio State University
Columbus, Ohio 43210

Dr. D. R. Lide, Jr.
National Bureau of Standards
A537 Administration Building
Washington, D.C. 20234

Professor Frank McClintock
Dept. of Mechanical Engineering
Massachusetts Inst. of Technology
Cambridge, Massachusetts 02139

Dr. Paul C. Paris
Del Research Corporation
P.O. Box 222
Bethlehem, Pennsylvania 18016

Dr. Elio Passaglia
National Bureau of Standards
B264 Materials Building
Washington, D.C. 20234

Dr. Harold Paxton
Division Director of Materials
Resources
National Science Foundation
1800 G Street, N.W.
Washington, D.C. 20550

Professor Roger Staehle
Dept. of Metallurgical Engineering
Ohio State University
116 W. 19th Avenue
Columbus, Ohio 43210

Mr. Daniel Stone
Association of American Railroads
AAR Research Center
3140 South Federal Street
Chicago, Illinois 60616

Professor Alan S. Tetelman
Chairman, Materials Department
School of Engineering
6531 Boelter Hall
University of California
Los Angeles, California 90024

Dr. R. M. Thomson
National Bureau of Standards
B109 Technology Building
Washington, D.C. 20234

ARPA MATERIALS RESEARCH COUNCIL
Centerville, Massachusetts
10-12 July 1972

SOME PERSPECTIVES AND RECOMMENDATIONS ON STRESS-CORROSION CRACKING

Reported by

Morris Cohen and H. H. Johnson

January 1973

TABLE OF CONTENTS

Some Perspectives and Recommendations on Stress-Corrosion Cracking

Morris Cohen* and H. H. Johnson**

ABSTRACT

The Proceedings of the July 1971 ARPA Materials Research Conference on "Environmental Degradation of Stressed Materials" have been reviewed to provide a basis for reaching up-dated conclusions and recommendations concerning important areas in stress corrosion. Attention is also drawn to various unresolved issues in the field, both general and specific. In addition to the research opportunities thus highlighted, there is an overriding need to make the available information and experience on stress corrosion more readily accessible to the engineering community; a handbook series for this purpose is proposed and delineated.

---

* Department of Metallurgy and Materials Science, Massachusetts Institute of Technology, Cambridge, Mass. 02139

** Department of Materials Science and Engineering, Cornell University, Ithaca, New York 14850.

1. Introduction

A Conference on "Environmental Degradation of Stressed Materials" was organized by Professors J. P. Hirth and F. A. McClintock at Woods Hole, Massachusetts on 21-23 July 1971 under the auspices of the ARPA Materials Research Council. Most of the emphasis was directed to stress-corrosion cracking (SCC) phenomena. The Proceedings of the Conference were published in the form of thirteen papers and a Summary Report, as Volume II of the Preliminary Reports, Memoranda and Technical Notes of the ARPA Materials Summer Conference (July 1971).

A year later, on 10-12 July 1972, a meeting was held in Centerville, Massachusetts, again under the aegis of the ARPA Materials Research Council, in order: (a) To assess the Proceedings of the above Conference as a basis for reaching a set of conclusions and recommendations on important areas in stress corrosion, and (2) To examine the feasibility of preparing a Handbook on Stress Corrosion, and suggest ways of implementing such a project.

These objectives were attained as reported herein.

2. Participants

The task group taking part in the July 1972 study consisted of:

B. F. Brown, Formerly Naval Research Laboratory, now American University
M. Cohen, Massachusetts Institute of Technology
J. P. Hirth, Ohio State University
H. H. Johnson, Cornell University

J. Kruger, National Bureau of Standards
F. A. McClintock, Massachusetts Institute of Technology
H. W. Paxton, National Science Foundation
E. N. Pugh, University of Illinois
H. H. Uhlig, Massachusetts Institute of Technology

3. Conclusions and Recommendations Based on the July 1971 Conference

A. General Points Concerning the 1971 Conference

A. 1 It was the consensus of the participants that the Summary Report of the 1971 Woods Hole Conference on "Environmental Degradation of Stressed Materials" was a valuable summary of existing knowledge and problems. The proposed classification of stress-corrosion cracking mechanisms (Table I) was accepted as sufficiently comprehensive for present purposes.

A. 2 Existing knowledge of stress corrosion is not adequately appreciated by the engineering community. As a consequence, structures are frequently designed and fabricated in ignorance of the stress-corrosion possibilities, and occasional catastrophes occur which could have been avoided.

A. 3 It appears that in many instances stress-corrosion cracking (SCC) is inherently a low-probability event, and very real questions arise as to the cost-benefit balance if the most stringent protective measures are undertaken. There has been no attempt yet to analyze SCC and preventive measures on a cost-benefit basis; if it could be carried out, the analysis would probably lead to quite different pro-

TABLE I

Suggested Classification of Candidate Stress-Corrosion Cracking Mechanisms

1. Dissolution Mechanisms

1.1 Film-rupture
Crack propagates by local dissolution of metal at crack tip due to prevention of passivation there by plastic deformation.

1.2 Strain-accelerated dissolution
Crack propagates by localized anodic dissolution. Lattice defects introduced by plastic deformation are thought to accelerate the dissolution process.

2. Mechanical Mechanisms

2.1 Hydrogen embrittlement
Hydrogen accumulates within metal in the crack-tip region, leading to localized weakening either by void formation or lowering of cohesive strength. Crack propagates by mechanical fracture of weakened region.

2.2 Adsorption
Surface-active species adsorb and interact with strained bonds at the crack tip, causing reduction in bond strength and leading to crack propagation.

3. Mixed Mechanisms

3.1 Brittle film
Crack propagates by repeated formation and rupture of a brittle film which grows into metal at crack tip.

3.2 Tunnel model
Crack propagates by formation of deep pits or tunnels via dissolution followed by linking of these pits or tunnels by ductile rupture.

tective strategies for different types of engineering application. Good statistical data and detailed descriptions of SCC events will be required for such an analysis.

A. 4 It is recommended that a strong effort be undertaken to collect and systematize SCC knowledge, and that the resulting documents be widely publicized in the engineering community. This program might be undertaken in three steps. The first would be a compilation of case histories of important stress-corrosion failures, with an accompanying text describing the important lessons and principles derived from each case history. The second would be a compilation of qualitative stress-corrosion information on major alloy systems in the more common environments, with the information presented in tabular form by both alloy system and environment. Finally, a full handbook summary of quantitative engineering data on SCC is recommended. In all three instances, the formats should be selected to be of maximum usefulness to design, operating and materials engineers.

B. General Technical Points

B. 1 SCC is a complicated problem with aspects of metallurgy, surface and electrochemistry, defect physics, and mechanics, including both crack-tip phenomena and mechanical design. Consequently, in a given SCC situation, prevention or amelioration of SCC may be attempted by any of several widely different techniques, which must be evaluated both technically and economically:

i) replace susceptible alloy by one more resistant;

ii) cladding or plating;

iii) design change, e. g., to mitigate stress concentrations or sites for localized environmental concentration;

iv) improvement in fabrication techniques or procedures, e. g. shot peening;

v) improved housekeeping (cleaning, fit-up practices, etc.) during fabrication;

vi) improved inspection procedures to detect potentially damaging flaws;

vii) environmental control - such as deaeration, anionic inhibitors, galvanic coupling and cathodic protection.

The use of one or more of these techniques constitutes an SCC-control plan. It seems probable that the concept of such control plans, where several techniques are adopted in harmony to prevent SCC, will be increasingly used for high-risk structures.

B. 2 Because SCC is so complicated, basic research has not yet led to guidelines for predicting possible SCC susceptibility in untried combinations of materials and environments. Basic research has suggested several conceptual mechanisms which, in the future, may be sufficiently developed for predictive purposes. However, for unproved performance in engineering systems, parametric research with systematic control over experimental variables will be required in the

foreseeable future. This broad-focus approach may well indicate the optimum directions for basic research. It has been successful in improving the SCC resistance of aluminum alloys.

B. 3 It is recommended that at least one SCC system be studied in multidimensional-variable space, to suggest possible empirical correlations for other systems. Appropriate variables would include pH, temperature, potential, microstructural parameters, anion and cation species, etc. Of particular interest would be the determination of cracking regions in potential-pH space for both smooth and pre-cracked specimens.

B. 4 For situations where crack propagation is life-limiting, experimental techniques based on fracture mechanics are valid for at least a semi-quantitative assessment of SCC susceptibility. The parameters required are K_{ISCC} and the V-K plot. This approach has been most widely used for the higher strength materials, but is potentially applicable at lower strength levels. However, crack branching and the resultant uncertainty in K are more commonly encountered at lower strength levels; this requires further theoretical and experimental attention.

B. 5 It was agreed that present understanding of corrosion fatigue as well as its relation to SCC is unsatisfactory. The data-base should be enlarged and, if possible, critical experiments devised. Since even conceptual models are largely lacking, parametric research will probably be the most productive approach for the time being.

B. 6 Crack-tip chemistry experiments suggest strongly that all deep aqueous-media stress-corrosion cracks propagate in local environments suitable for the reduction of hydrogen ions, although hydrogen may not necessarily be the causative agent in the cracking process. The crack-tip chemistry becomes independent of the bulk environment. Pourbaix diagrams provide an essential framework for this interpretation.

C. Progress and Problems Concerning Some Major Alloy Systems

C. 1 Aluminum Alloys

Substantial improvement in the SCC resistance of high-strength aluminum alloys has been obtained in recent years at little or no cost in strength. Semi-empirical or parametric procedures were used, in which systematic variations in composition, processing, and heat treatment, followed by laboratory corrosion testing, have been the guiding factors. For example, elimination of iron in the melting stock removed a massive precipitate in the eventual service microstructure which was associated with SCC susceptibility. This approach has lengthened the laboratory SCC-failure times by factors up to 10^7, thus guaranteeing satisfactory performance for many applications, even though the threshhold stress intensity for crack growth may remain low.

This has led to substantial reduction in the incidence of SCC failures, which were commonly due to intergranular crack propagation especially in the short-transverse direction of aluminum-alloy forgings. At the same time, improved design procedures decreased the stresses in the

short-transverse direction; no doubt this has also contributed to more favorable failure statistics.

There remains a lack of basic understanding of the grain-boundary failure mechanism, and some controversy over the role of aging in relation to SCC susceptibility. It appears that critical experiments could be devised to evaluate the role of aging and the associated factor of the degree of precipitate/matrix coherence. This might suggest the directions for further practical improvements.

C. 2 High-Strength Steels (σ_y > 180 ksi)

SCC of high-strength martensitic steels in aqueous environments remains a troublesome problem. At strength levels of approximately 235 ksi and above, all high-strength steels in water exhibit K_{ISCC} values in the low range of 10 - 20 ksi-$(in)^{1/2}$. Substantial efforts to improve performance by variations in processing and thermal treatments have not yet been successful; some increase in failure time has been obtained, but not in sufficient magnitude to be of practical value, and the threshhold stress intensities have not been raised. At strength levels between 180 and 235 ksi, the different high-strength steels exhibit widely scattered performance levels in water; this can be attributed to metallurgical factors which, if understood and brought under control, could result in appreciable improvement. Bainitic structures would be of interest in this connection.

At these high-strength levels, the problem of flaw detection is

formidable. For a strength level of 200 ksi and a K_{ISCC} of 10 ksi-$(in)^{1/2}$, the critical flaw size is well under 10^{-3} in, which calls for extreme capability in non-destructive testing techniques. At present, the very high-strength steels are useable only with the most careful processing, plating, and inspection procedures.

There is much circumstantial evidence that hydrogen is the cause of water-induced cracking of high-strength steels, although the specific mechanism remains controversial. The hydrogen concept is supported by crack-growth rates, hydrogen-permeation studies, and crack-tip electrochemistry experiments.

It is of considerable interest that the SCC resistance of mild and medium-carbon steels is substantially improved by severe cold working. This improvement may well result from the very fine, high dislocation-density microstructure produced by cold working; such a microstructure could hinder the passage of hydrogen. Further research along these lines would be worthwhile, including some delineation between the effects of cold working on crack initiation vs. crack propagation under stress-corrosion conditions.

Prior austenitic grain boundaries are a promising area for research with respect to aqueous-media SCC cracking of quenched and tempered steels. Water-induced cracks often propagate along these boundaries; if they could be strengthened or "cleaned up," perhaps by directional solidification or electroslag remelting, substantial improvement in

performance might result.

C. 3 Titanium Alloys

Service SCC-failure experience with titanium alloys is somewhat limited, but substantial laboratory data are available. Both intergranular and transgranular cracking modes have been observed, and there is evidence that the intergranular mode, which occurs primarily in halogen-containing organic liquids, can be categorized as strain-accelerated dissolution.

The mechanism of the transgranular mode of failure is perhaps more controversial, with questions centering around the possible role of hydrogen. The rather high crack velocities observed seem not to be compatible with a dissolution mechanism.

Two mechanisms have been proposed to explain titanium-alloy failures in the rather unique environment of nitrogen tetroxide. On the basis of metallographic evidence, a brittle-film model has been proposed, but more recently an electrochemical-dissolution mechanism involving the formation of a $TiO(NO_3)_2$ complex has also been suggested.

In recent years there seems to have been general improvement in the SCC resistance of titanium alloys as measured by the safe strength levels; this appears attributable to closer control of the α - β microstructure as produced by variations in composition and thermal treatment. It is possible that further improvements could be obtained by this route.

C. 4 Austenitic Stainless Steels

Recent problems with nuclear-power systems have again highlighted the problem of stress-corrosion cracking of austenitic stainless steels, especially in aqueous media containing chloride and fluoride ions. This system is not easy to study realistically, since the service-environmental factors are difficult to duplicate in the laboratory. Some effort has been devoted to improvements in SCC resistance by variations in composition and processing; this area is also intertwined with economic and fabrication factors.

At the moment, mechanistic controversies center on the relative probabilities that hydrogen, stress sorption, or slip-step dissolution are the key features of the SCC mechanism. Experimental evidence also suggests that film dissolution and formation kinetics are important, but film considerations can be incorporated into any of the above three concepts. The major problem for the hydrogen idea is that the mere presence of hydrogen does not at all guarantee that it is the causative agent, while it is not yet clear that the slip-step dissolution model is compatible with the brittle SCC fracture surfaces commonly observed. * The stress-sorption concept is consistent with the critical potentials measured for austenitic stainless steels and with the beneficial effect

---

* The operative dissolution may be on too fine a scale to be detected.

of inhibiti[illegible] ions, but it is probable that the other concepts are also in accord with the critical-potential idea. It is evident that new experiments must be devised to evaluate these concepts.

D. Specific Technical Concepts and Issues

D. 1 With most alloys the specific-ion concept becomes less significant at higher strength levels. Nominally mild environments can become aggressive at increasing strength levels, and the same environment, e. g., water, may then be harmful to widely different alloy systems. It is not wholly clear why this should be so.

D. 2 The operative role of hydrogen as a general SCC causative agent for several alloy systems is not adequately explored. It may be possible to devise some unequivocal experiments with respect to the role of hydrogen, in which a thin specimen is (a) stress cracked from one side in an appropriate, well-controlled electrochemical cell and (b) hydrogen charged from the other side in a separate, independently-controlled electrochemical cell.

D. 3 There is need for quantitative physical models to account for the brittleness induced by hydrogen.

D. 4 The concept of a "critical potential" for SCC-initiation susceptibility should be examined for a larger number of alloy systems.

D. 5 Experimental criteria for SCC susceptibility have been suggested under the various titles of "border-line passivity," "critical potential," and "critical ratio of film-formation to metal-dissolution

rates." The degree of compatibility and utility of these concepts should be established.

D. 6 In situations where stress-corrosion cracks emanate from pits, it is believed that the occluded volume of the pit leads to local electrochemical changes, e. g., acidification, which then causes cracking. However, the abrupt transition from a rounded pit to a sharp crack is poorly understood. In general, pitting is neither necessary not sufficient as a precursor for SCC; situations are well-known where SCC is not preceded by pitting, and in which pits do not lead to cracks.

D. 7 Concentration profiles near grain boundaries should be studied for intergranular SCC systems by such techniques as Auger spectrocopy, ESCA, energy-loss analysis, etc.

D. 8 Stress-corrosion media in service do not usually cause extensive general corrosion. Is this significant, or does it simply follow from the fact that alloys are often selected on the basis of good general corrosion resistance? Attempts should be made to model the stability/instability conditions of a corroding interface.

D. 9 There appears to be only meager understanding of the mechanisms whereby anion inhibitors operate. This area could be of substantial importance in the development of preventive methods. The interaction of anions with imperfections such as emerging dislocations and flaw tips should be studied, if techniques can be devised.

D. 10 Questions concerning the mechanical/chemical and continuous/discontinuous nature of SCC crack growth in many systems remain unsettled. This is of considerable mechanistic significance, particularly with respect to the slip/dissolution model.

D. 11 The material and mechanical description near a moving crack tip remains disturbingly approximate. Physical models and analytic techniques are needed on the dislocation and lattice scales. It would be desirable to describe quantitatively the interaction between a strained discrete lattice and specific adsorbed ions.

D. 12 One wonders whether the improved SCC resistance of cold-worked steel has more general validity. A systematic investigation of the role of cold deformation in SCC resistance would be highly desirable.

D. 13 Attempts to produce SCC-resistant microstructures by thermal cycling through a phase-transformation range should be undertaken for steels and titanium alloys. This could lead to finer microstructures with newly-formed boundaries, which might well have improved resistance to SCC.

D. 14 Carefully designed bicrystal studies have many interesting possibilities for revealing the interplay of phenomena at grain boundaries under conditions of stress corrosion.

D. 15 The influence of systematic and random stress variations on crack growth should be studied under SCC conditions. This line of research may help explain why service-failure times are often longer than indicated by laboratory tests.

4. Proposed Handbook Series on Stress Corrosion

There is urgent need to make information and experience on stress corrosion easily accessible to design and operating engineers as well as to materials engineers. SCC failures constitute a serious and costly problem in many technologies, and the potential dangers will become even more intense. Hence, it is essential that the hazards of SCC be fully recognized by the engineering community, and that ways of avoiding, or coping with, this threat be published in a readily useful form. As recommended in A. 4 of the previous section, we have concluded that a three-part handbook series would be an effective and realistic vehicle for the information-transfer process being proposed.

A. Part I. Documented Case Histories Involving Stress Corrosion

It is intended that this part of the handbook series will be a compilation of those stress-corrosion failures which have maximum educational value. The circumstances and cause(s) of each failure will be described in a standardized format, highlighting such factors as the design and application, materials and environment involved, operating conditions, critical circumstances, corrective measures taken, lessons learned, guiding statements and precautions for other applications. The case histories to be selected should cover a wide range of materials and environmental combinations, typical of general engineering practice. Wherever helpful, illustrations (photographs, line drawings, micrographs) should be included.

The list of 22 case histories given below is indicative only. These instances of stress-corrosion failures are known and presumably can be documented. A more thorough search would undoubtedly reveal many other cases of importance for this compilation.

Stress-Corrosion Failure	Source of Information
1. Dresden Reactor: failure of 17-4 PH control rod, wrong aging temperature used for SCC resistance	S. H. Bush, paper delivered at 1971 Conference, Woods Hole, Mass.
2. Swedish Agesta Reactor: failure of inconel 600 inspection tubes	"
3. 304 Stainless Steel Primary Reactor: from radiolysis of freon	"
4. Shippingsport Reactor: caustic cracking of 304 steam generator tubes	"
5. Small Steam Turbine: 302 stainless moisture shield	"
6. Savannah River Reactor: failure of 304 stainless heat-exchanger tubes	"
7. Oyster Creek Reactor: pressure-vessel control-rod system, 304 stainless stubs welded to ferritic steel vessel	"
8. Saturn: Al 356 casting failed, Hg thermometer broke, Hg attacked aluminum alloy catastrophically	Mentioned by Roger Staehle
9. Bone-Brace Corrosion Fatigue	Mentioned by Roger Staehle; John Wulff might have information on biomaterial failures

10.	Silver Bridge over Ohio River	Mentioned by Roger Staehle
11.	Apollo Oxidizer Tanks: Ti 6-4 alloy in contact with methanol and N_2O_4	"
12.	F-111 Main Support Member: failure of Si-modified 4340	"
13.	Comet Aircraft: poor design, poor fatigue properties in moist air, corrosion fatigue	"
14.	Cracking in Bell Telephone Relays in Los Angeles: nickel-brass attacked by nitrates in air	Mentioned by H. H. Uhlig, written up somewhere in ASTM
15.	John Hancock Building in Boston: attack of brass in air-conditioning fixtures by ammonium nitrates produced by spark-electrostatic dust eliminators	Mentioned by H. H. Uhlig
16.	Failure of D6A Steel Hydraulic Actuator moist atmosphere	Tiffany and Masters article in ASTM 381
17.	LEM: Al 7075 - T651 Tapered Pin Fittings	Mentioned by Roger Staehle
18.	Ti Rivets Clad with Cadmium; failures in 747 aircraft	"
19.	Tank Cars for Transport of Anhydrous Ammonia; quenched and tempered steels	Mentioned by Roger Staehle; Phelps and Loginow article in Corrosion
20.	J-2 Engine: hydride formation in Ti-5Al-2.5Sn	Mentioned by Roger Staehle; H. Reiss of UCLA should have further information
21.	H_2S Problems in Oil Industry; some wells were capped when even 9Ni failed	Mentioned by Roger Staehle. see articles in Corrosion over past several years

22. Radioactive Nitrates in Steel Tanks — M. Holtzworth et al., Matls. Protection 7 (1968) 36

Other possible SCC case histories from which lessons can be learned are:

1. Aluminum Alloy Forgings: failures in short-transverse direction.
2. Cathodically Protected Gas Pipelines: caustic embrittlement.
3. Galvanic Couples: martensitic stainless nails used on aluminum roofing.
4. High-Strength Fasteners: hydrogen embrittlement.
5. Copper-Base Alloys: intergranular cracking.

B. Part II. Guidelines for Protection Against Stress Corrosion

This part of the handbook series will consist of tabulated qualitative information concerning stress corrosion, as distilled from the case histories and other available sources. There will be two types of displays:

1. Tabulations by alloy systems

Steels	Al Alloys	Ti Alloys	Stainless Steels
4340 - 250 ksi	2024	6-4	304 stainless
4340 - 180/200 ksi	5040	8-1-1	316 stainless
HY80	7075 - T6	6-6-2	410 stainless
1020	Casting alloy	B120VCA	Ferritic stainless
A533		βIII	17-4 PH
Maraging			Inconel 600
			Incalloy

Cu Alloys	Other Alloys
70-30	Monel
60-40	Mg alloys
Cu-Be	
Bronze	

For each of the above alloys, the known SCC dangers or conditions to be avoided will be listed with respect to (a) the material itself (impurities, microstructure, state of stress, surface quality, etc.), (b) the assembly (welds, crevices, galvanic couples, etc.), and (c) the environment (pH, temperature, oxygen concentration, halide ions, etc.).

In addition, for each alloy system, there will be a list of known methods for effectively avoiding stress corrosion, such as plating or cladding, galvanic protection, oxygen scavenging, inhibitors, pH control, and temperature limits.

2. Tabulations by environment

Atmosphere	Liquid sodium
Pure water	Liquid mercury
Tap water	Liquid lithium
Sea water	Liquid NaK
Carbonic acid	
Other acids	
Caustic solutions	
Nitrates	
Chlorides	
Ammonium solutions	
Sulfates	
Sulfides	
Cyanides	

For each of the above media and for given alloys studied in such media, the following designations will be posted:

A = no SCC failures found in laboratory tests, or in service.

B = SCC failures found in laboratory tests, but not in service.

C = SCC failures found in laboratory tests, and in service

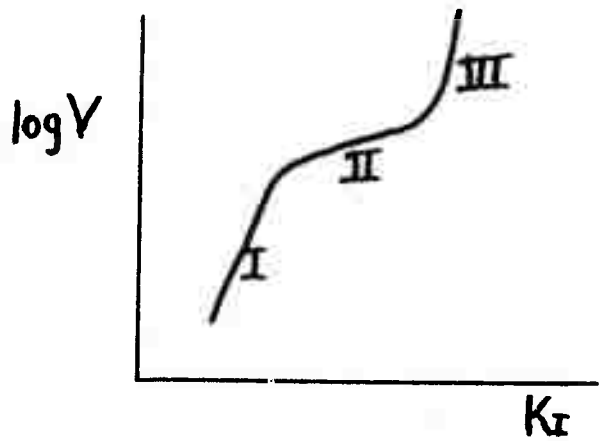
Careful attention should be paid to the tabular or matrix form to be adopted for presenting the above compilations; the choice of format will depend on the amount and nature of the collected SCC information.

Wherever possible, references are to be given in the tabulations for further information and for keying back to the case histories in Part I.

C. Part III. Stress-Corrosion Data Relating to Design and Failure Analysis

Quantitative stress-corrosion data of engineering utility are to be assembled in this part of the handbook series. The selected alloy/environment systems will match, to the extent possible, those tabulated in Part II, and specific information will be included on the prevailing conditions for which the data are relevant (temperature, ionic concentrations, applied potential, critical potential, etc.).

Crack-propagation data will be based on the relation between crack velocity (V) and the operating stress intensity (K_I). The general pattern shown below is representative of many SCC tests with precracked specimens, although the full curve has been determined in relatively few



cases. For practical applications, regime I and the plateau or inflection region (regime II) are of particular interest, since the crack velocity in regime III is far beyond anything that can be tolerated in service.

For some systems, especially with high-strength titanium alloys and steels, the slope in regime I is essentially vertical, and the corresponding value of K_I is K_{ISCC}, representing the threshhold stress sensitivity below which safe performance is assured. However, for aluminum alloys and probably for lower-strength titanium alloys and steels, a definite K_{ISCC} threshhold cannot be fixed because the slope in regime I is not zero. It is then useful to take K_{ISCC} at an arbitrary crack-growth rate, say 10^{-7} cm/sec, which is sufficiently slow to guarantee satisfactory performance for the intended service.

The (V, K_I) values associated with the inflection point are also of interest, inasmuch as conditions beyond this point are clearly in an unsafe range for the alloy and environment under consideration.

The suggested stress-corrosion data may be divided into two categories: static and cyclic loading. Wherever appropriate, special remarks, warnings and references should be given.

Category 1 Data on Static Loading

A. Tabulate stress for chosen SCC initiation times with smooth specimens (1 hr., 1 day, 1 yr.)

B. Tabulate K_I for chosen SCC initiation times with precracked specimens (1 hr., 1 day, 1 yr.).

C. Tabulate K_I ($\sim K_{ISCC}$) for a selected low rate of crack growth with precracked specimens ($\sim 10^{-7}$ cm/sec).

D. Tabulate d log V/dK_I for regime I (precracked specimens).

E. Tabulate V and K_I corresponding to inflection point in log V - K_I curve (precracked specimens).

Category 2 Data on Cyclic Loading

F. Tabulate stress for lifetime of 10^8 cycles; specify mean stress and frequency (smooth specimens).

G. Tabulate ΔK_I for lifetime of 10^6 cycles; specify mean ΔK_I and frequency (precracked specimens).

H. Tabulate parameters c and n for corrosion-fatigue crack data represented by $\frac{d\ (\text{crack length})}{d\ (\text{no. of cycles})} = c\Delta K_I^n$; specify mean ΔK_I, range of ΔK_I, and frequency (precracked specimens).

D. Implementation of the Handbook Series on Stress Corrosion

We believe this data-collection and handbook-publication program to be of sufficient importance to warrant funding by some combination of NSF, ARPA, the Electric Power Research Institute, and a number of industrial organizations. Under such auspices, a Steering Committee of authorities in the field of stress corrosion would be appointed to plan, budget, implement and coordinate the three parts of the handbook series.

The following suggestions are intended to be illustrative of a plan for action. The Steering Committee could be initiated and chaired by Dr. H. W. Paxton. This Committee would prepare a formal proposal and

budget, seek funds, and appoint editors for Parts I, II and III. It is estimated that, aside from committee and panel participation, Part I will require a 1/2 man-year effort, Part II 1 - 1 1/2 man-years, and Part III about 3 man-years. Institutions with data-collection and distribution expertise, such as the National Bureau of Standards or Battelle Memorial Institute, could prove quite valuable in this program, expecially in connection with Part III.

It is not advisable to decide at this point whether the three parts of the handbook should be published in separate volumes or in some other combination. We do see the need, however, for publishing Part I as soon as possible. All such decisions would fall within the prerogatives of the Steering Committee.

WORKSHOP ON MATERIALS FOR ENERGY CONVERSION

A. L. Bement and R. Kaplow

I. Introduction

A two-day workshop on materials for energy conversion was held at Centerville, Massachusetts, on July 20-21, 1972. The attendees are listed in Appendix A. During the first day of the workshop brief presentations were given to highlight the general characteristics of selected energy conversion technologies and materials problems associated with these technologies. The presentations given are listed below:

Topic	Presenter
1. Enumeration of Proposed New Energy Schemes	A. L. Bement
2. Solar	R. Kaplow
3. Fast Breeder Reactors	A. L. Bement
4. Fission Reactors	S. H. Bush
5. Fusion Reactors	A. L. Bement
6. Magnetohydrodynamics	P. L. Farnsworth
7. Steam and Gas Turbines	C. T. Sims
8. Review of Current Materials Processing Issues	M. Flemings

The second day was devoted to the development of opinions and projections of materials problems in two of the above technologies; namely, gas turbines and fusion reactors. The intention

was to select one technology that is fairly far along in development and another at an early stage. For this exercise, not only the probable occurrence of "needs" was considered, but also such factors as relative importance, the time of probable impact on planning and execution processes, the segment of materials technology to which the problem is directed, the sort of R & D required, and the cost of the necessary work were discussed. To provide a structure for some of these points, a list of elements (or functions) of materials technology (Appendix B) and a sequence of steps in the process of planning and executing a new energy technology (Appendix C) were developed. An attempt was made to organize the discussion in terms of filling out a "matrix" in which columns corresponded to energy conversion schemes, rows to the separate elements of materials technology, and the "boxes" segmented to highlight the existence, severity, impact time, and cost of potential problems.

II. Summary

The two-day workshop on materials for energy conversion proved to be a useful and stimulating exercise for reviewing some of the more important (and more obvious) materials problems associated with advanced electrical generation technology. The short time available and the limited representation restricted the depth of examination and study, and provided only a hasty survey of a few of the important elements of materials technology for two selected systems. From the survey presentations and workshop

discussions the following general statements related to materials problems can be made:

a. Nearly all advanced electrical generation technologies have materials problems that will have a significant impact on design limits and on operating or maintenance costs. In some systems (possibly high-performance gas turbines, MHD generators, and fusion reactors) materials problems and limitations may jeopardize technical or economic feasibility.

b. Materials properties limits will have a greater controlling factor on new technology development and will very likely be the pacing factor in improving the thermal efficiency and plant factor of existing systems.

c. There is an urgent need for increasing pressure for the industrial scale-up of new materials processes to respond to a growing emphasis on quality assurance and to allow for the substitution of more advanced or less strategic materials.

d. Materials in advanced energy conversion systems will be submitted to increasingly aggressive and stringent environments. This will heighten the need for improved materials characterization methods, protective measures, and reliability assessment. A close interaction between the designer and materials engineer is absolutely essential.

e. Materials problems causing serious maintenance difficulties and loss of plant availability in present systems are occurring in the more conventional components that have not benefited from the main thrust of the engineering and materials

R & D in support of critical system components. The origin of many of these problems stems from those elements of materials technology (fabrication, codes and standards, inspection and testing methods) that are usually overlooked in system development programs and are left to equipment suppliers to solve during or subsequent to vendor qualification and procurement stages.

f. In-depth materials studies in support of advanced energy conversion, transmission, and utilization programs are needed in order to facilitate a more meaningful assessment of technical and economic feasibility and to provide adequate lead time to anticipate and solve potentially limiting materials problems.

g. The ARPA Materials Research Council could make substantial contributions toward identifying and outlining solutions to many of the materials problems in energy conversion technology.

III. Points Made During Short Presentations

A. Enumeration of Proposed New Energy Schemes

1. Technology Options

In the fall of 1970 the Electric Research Council appointed a Research and Development Goals Task Force with the assigned responsibility of charting the utility industry's future research programs in the areas of energy conversion, transmission, distribution, environmental control, energy utilization, industry growth, system development and fundamental research. In the area of energy conversion technologies the priorities defined below were

established:

Priority 1 - Critically important: projects having an indispensable effect on all goals, which by their nature must receive first attention.

- present methods
- fission breeder reactors (mid 1980's)
- fusion reactors (mid 1990's)

Priority 2 - Very important: projects having a somewhat less intense impace, but which nevertheless must be included in any meaningful R & D program.

- open-cycle MHD (mid 1980's)
- fuel cells (10-20 MW size) - (late 1970's)
- bulk storage batteries

Priority 3 - Important: projects of significance to future planning and continuing operations.

- unconventional cycles (K/steam, CO_2)

Priority 4 - Desirable: other projects which are useful to accomplish stated goals.

- closed-cycle MHD
- solar
- thermionic

Other - (generally restricted in application, limited in geographical siting, or of marginal promise for large-scale electrical generation).

- geothermal, tidal, air storage, rocket engine, thermo-electric radiation converters, radioisotrope

2. General Comments

Greater attention should be given to the better utilization of current energy supply [examples: (1) reduction of line losses through the use of better transformer materials, electrical insulators, and conductors; (2) development of improved inexpensive space thermal insulation materials.].

B. Solar

1. Materials Problems

a. Photo-voltaic: The cell cost must be greatly reduced; a high carrier mobility must be achieved and stably maintained; the carrier trapping rate (defect concentration) must be low; junction composition must be carefully controlled.

b. Space-solar photovoltaic: The cell materials must be resistant to surface and bulk damage due to various forms of radiation and particle bombardment.

c. Thermal energy generation: The lens (or mirror) materials must be structurally rigid, erosion resistant, and mass producible at low cost. The thin film filters must be stable at operating temperatures and resistant to interdiffusion among complex layer structures. Corrosion, mass transport, and purity control with respect to the liquid metal (or molten salt) heat transfer fluid may be problems due to the large distances involved and the high surface-to-volume ratio of the conduit network. System design may have to accommodate rotation to face the sun.

d. Thermal space heating: Inexpensive, weather-resistant solar panels, a manufacturing capability for small-

and medium-size units and an inexpensive, effective thermal storage medium are the principal requirements.

2. General Comments

Materials will have a significant effect on economic feasibility of solar energy utilization. Independent evaluation of cost of achieving economical solar energy production should be made. Present national effort is unreasonably small based on potential benefit. Environmental impacts, such as climate modifications in the near atmosphere above large-scale collectors and microwave transmission dangers from space solar systems, must also be considered.

C. Fast Breeder Reactors

1. Materials Problems

a. Radiation effects: These include volume swelling of structural materials and fuel cladding due to radiation-induced void formation; in-reactor creep and growth; swelling and cracking of fuel and neutron absorber (B_4C) materials; and helium embrittlement of fuel cladding.

b. Compatibility: These include fission product intergranular attack of fuel cladding; sodium corrosion and mass transport; chloride- and caustic-induced stress corrosion cracking of steam generator tubing; composition changes in austenitic stainless steel due to carbon transport (Type 316 stainless steel is the present cladding alloy due principally to available reactor experience; stabilized stainless steels are available, but weld cracking is a problem that requires better welder qualification

and process control procedures).

c. Mechanical behavior: These include fuel/clad mechanical interactions; combined creep and low-cycle fatigue; and the need to maintain core restraint (maintain fixed core geometry to control heat transfer and neutron reactivity under the influences of radiation-induced dimensional changes).

2. General Comments

Fast breeder reactors require the heavy use of strategic materials (chromium and nickel). There is a need to consider possible substitutes for chromium and nickel and the use of lower grade alloys for some applications where use of the highest-grade alloy may not be warranted. It may be difficult to find enough sites for nuclear parks. The projected number of annual installations may be beyond the capabilities of the industries involved (nuclear systems and fuel suppliers, banking/finance, construction, utility, insurance, etc.). Safeguards against the covert diversion of plutonium for weapon use will become difficult to enforce due to increasing dispersion.

D. Fission Reactors

1. Materials Problems

Light-water reactors employ considerable quantities of stainless steel for pressure vessel liners and primary circuit piping. These materials are susceptible to stress corrosion cracking, especially when oxygen is present in the water coolant and when the stainless steel is in the sensitized condition. Similarly, Inconel 600, which is currently used in pressurized-

water reactors for steam generator piping, is very sensitive to water quality (crevice corrosion at the steam/water interface). The possibility exists of using ferritic materials if better (cleaner) water quality can be achieved.

Low-hardenability steel (A533B) is currently used in very heavy sections for pressure vessels. The heavier the vessel wall thickness the greater is the possibility that large flaws will be undetected during inspection and testing. Although there are additional realizable economies of scale, plant size is limited by the ingot size capability of the steel industry. There are potential benefits in going to higher-strength steels to allow for larger plant size; however, these may aggravate problems due to hydrogen embrittlement, temper embrittlement, and reduced fracture toughness. An estimated 100 man years of research and development would be required to successfully qualify a new pressure vessel steel.

The potential for serious materials problems seems lessened in high-temperature, gas-cooled reactors because of their use of prestressed concrete vessels, coated-particle fuel, a graphite core, and low-pressure coolant. However, due to the present lack of experience it is not possible to adequately assess long-term materials performance and reliability for these systems.

2. General Comments

The increasing frequency and volume of radioactive fuel and waste shipments enhances the likelihood of major radioactive spill. The significant amounts of stored energy and useful

isotropes in these wastes should be factored into future resource planning. The heavy use of strategic materials, such as stainless steels and nickel-base alloys, often represents a "strategic retreat" based on lack of adequate knowledge of environmental effects on structural materials.

E. Fusion Reactors

1. Materials Problems

a. Toroidal magnetic confinement systems: The more serious problems are related to surface erosion, thermal loading, and radiation damage of the first (vacuum) wall. Surface erosion involves vaporization, sputtering, blistering, and exfoliation due to intense radiation wall loadings. Radiation damage includes void swelling, helium embrittlement, radiation-induced creep, and increases in the ductile-brittle transition temperature (bcc alloys). Other problems include possible lack of long-term alloy stability at wall thermal loads, tritium transport, corrosion by lithium or Li-Be fluoride salts, differential magnetic loads, and cyclic stresses. There is also some concern for possible radiation effects to magnet superconducting coils.

b. Laser pellet systems: The loading of the containment wall by radiation and particle fluxes is a major uncertainty affecting technical feasibility of these systems. Also, windows and mirrors for the high-intensity lasers will represent a major extension of present materials technology. Other problems include helium injection in the wall materials and lithium corrosion.

F. Magnetohydrodynamics

1. Materials Problems

a. Ceramic materials: Problems include contamination, erosion and corrosion of ceramic insulators and electrodes from coal ash and seed products; thermal shock and spallation; anion transport and polarization in ionically conducting electrodes; insulator breakdown; and creep of preheater refractories.

b. Metallic materials: Problems include hot corrosion of boiler tubes with sulfur-bearing seed; creep/fatigue and relaxation/fatigue; high-temperature oxidation; cracking and rejection of protective coatings (Al_2O_3, Cr_2O_3, and NiAl) due to thermal fatigue; corrosion in stack gas cleanup and seed recovery systems; and thermal stress distortion of water-cooled components (such as pre-heater valves).

c. Slag reactions: There is the possibility that stable slag films will form on MHD duct inner surfaces (conductivity and viscosity are relevant); the reactions between insulation oxides (MgO especially) and slag can lead to excessive erosion rates.

2. General Comment

Electrical transport at high temperatures in ceramic materials must be actively pursued to support the development of stable, long-lived electronic conductors. The large number of design parameters for MHD generators provide a challenge in optimizing performance and reliability with regard to matched materials properties.

G. Steam and Gas Turbines and Generators

1. Steam Supply Systems

Limitations of available commercial materials have prevented improvement of operating limits (example: fossil-fueled boilers designed for supercritical steam operation have been de-rated because of the high frequency of stainless steel tube failures resulting from a combination of creep and fatigue). Stronger materials and a better limitation of materials limitations are needed.

2. Gas Turbines

Operating performance is limited by the stress rupture life and the hot corrosion and oxidation resistance of first-stage turbine blades. In comparison with ~1700F operating temperature of present turbines, first-stage bucket surface temperatures of 2300F (near-term goal for ceramic-bladed turbines), 2900F (stoichiometric burning of coal gas), and 4300F (hydrogen combustion temperature) are future targets. The surface stability of Ni- and Co-based alloys to hot corrosion from sulfur- and vanadium-containing fuels is a major problem. This is presently partially overcome by coatings, which are not likely to provide an answer at much higher temperatures. Questions exist regarding the surface stability of ceramic blades (Si_3N_4,SiC), but improvements in processing have led to greater confidence. Dispersion-strengthened materials, single crystals, and advanced methods of air cooling are being explored to extend the temperature range

of blades. Ni and Mo are being discussed, but are not seen feasible for stoichiometric burning. Furthermore, they degrade in helium atmosphere (because of impurities in the gas). There is as yet no successful coating for these refractory metals. Coatings for Ni, Fe, and Co alloys employ Al, Cr, Ti, and Si, but these coatings are not stable above 2300°F because of diffusion, oxidation, and vaporization.

3. Steam Generators, Boilers, Heat Exchangers

These components require greater attention to materials and fabrication technology than they are currently receiving. Failure of these components is usually the key factor limiting the operating efficiency of current power plants (especially nuclear). The amount of attention being given to advanced materials and assembly procedures by the (usually small) component suppliers is inadequate.

H. Materials Processing

The metals processing industry has gone through growing pains as requirements for materials have increased. The principal problems are lack of adequate process control, absence of sensible specifications, processing bottlenecks, and the demand for new materials through process innovations. In some instances, the technology being applied is four decades old; in other instances the primary process involves extensive re-working because of inadequate process control (example: casting processes that require costly X-ray radiography and weld repairs which lead to

long and expensive process lags). Advanced casting technology has been available for some time, but has not been adopted by industry. There is an insufficient transfer of technology down to the mill and shop level, especially within the steel industry. These problems need to be studied and solved to the point where there is an obvious improvement in production. There should be sufficient freedom in product specifications to provide incentives for applying new process developments. Welding is another of the process bottleneck areas. Training procedures for welders may need to be altered to include more difficult materials as part of initial qualification testing. Technology advancements are providing more latitude for designing processes to yield new materials or material combinations for specific needs (example: metal-ceramic eutectic composites produced by directional solidification).

IV. Workshop Sessions

Two workshop sessions were conducted to predict materials requirements and problems for advanced energy conversion systems. The two systems selected for discussion were the gas turbine and the magnetically confined, toroidal (Tokamak), fusion reactor.

A. Gas Turbine Specifications

Nominal projected specifications for an electrical power-tenerating gas turbine, circa 1990, were given by C. T. Sims as shown below:

1. Power: 500 MW

2. Temperature for first stage partition, 3000°F (to represent a major step improvement compared with current rate of increase).
3. Fuel: Coal gas (low Btu)
4. Cord length of first stage bucket: 25 inches
5. Temperature of leading edge of blades: 3000°F
6. Blade life: 100,000 hours (replaceable)
7. Thermal efficiency: 55-60%
8. Thermal rating: 10,000 Btu/KWh_e
9. a. Blade root stress: (75,000 psi est. for heavy-element blades)

   b. Blade creep rupture stress: 20,000 psi; 100,000 hours
10. General characteristics: Multi-stage turbine; coupled system; combined-cycle; base-load; back-pressure turbine
11. Cooling: air (gas) cooling around bucket shaft
12. Wheel: austenitic stainless steel operating at ∿2200°F (cooled)

B. Impact of Materials Technology on Turbine Development

The discussion was organized so as to consider several selected items in detail from the list of materials technology elements given in Appendix B. The specific points identified in the discussion were ranked in accordance with the following levels of impact:

* Affects technical feasibility, but can be resolved within available lead time

** Will have a significant impact on design limits and on operating or maintenance costs

*** May jeopardize technical or economic feasibility

1. Materials Selection Against Design Requirements (Under Item 2, Appendix B)

* Carbon

Advantages: high-temperature strength; good compatibility (with reducing gas); low cost; good thermal stability; available; good thermal shock resistance; light weight; high modulus; high strength-to-weight ratio; high thermal conductivity.

Disadvantages: requires close control of combustion stoichiometry (reducing); may require two-stage combustion (high fuel to air ratio in first stage plus an after burner in the second stage); possible COS formation; highly frangible material; carbon deposition in reducing atmosphere.

* Oxides (examples: Al_2O_3, zirconia (ZrO_2), aluminum-magnesium spinels)

Advantages: good oxidation resistance

Disadvantages: limited strength, volatility of suboxides above 1900°C

** Carbides (hafmium, tantalum, silicon; silicon carbide might be adequate in a reducing atmosphere)

** Cooled Metals (examples: Ni-base alloys, refractory metals such as molybdenum- or niobium-based alloys; Ta and W are too dense, too expensive, and relatively unavailable.)

Disadvantages: surface incompatibility; limited strength at high temperature (creep); possible unavailability of strategic materials; higher density increases stress levels.

Composites

Advantages: possibility of going to laminated blade with a tough inner core and a temperature-resistant, thermally-insulating outer layer.

2. Materials Processing (Under Item 5, Appendix B)

Oxides

Specific Process Objectives (Many of these apply in general):

** microstructure control

** Toughness-crack arrest capability

* shape control

* machinability

** impurity control and characterization

* thermal shock resistance (metal-phase structures)

** resistance to erosion due to particulate transport

Specific Processes:

** sintering polycrystals with liquid-free boundaries

** chemical vapor phase deposition

* directionally-solidified multiphase structures (metal phase structures, for example)

*** fusion casting

** hot pressurized sintering

Carbon

Process Objectives:

** control of oxygen potential

** directional properties

Processes:

* machining

* pre-polymer processing

3. Reliability Assessment (Under Item 10, Appendix B)

** need to develop failure analysis and pre-service testing techniques for refractory ceramics:

nondestructive testing

proof-testing

failure prediction

definition of failure mechanism

** testing: there is a need to establish the economic impact of proof-testing and to develop methods for making projections from test data to life predictions

** fail-safe designs are required to avoid catastrophic failures

4. Code Specifications and Standards (Under Item 8, Appendix B)

* fracture control (design codes, process and product specifications, and standards for testing and quality control)

* specifications for process control need to be based on purity, microstructure and flaw detection (examination capability for < 1μm details)

5. R & D Projections and Cost Estimates

   a. Basic: gas turbine materials for 3000°F "uncooled" operation (rotor cooled, blades not cooled; ceramic materials)

   b. R & D cost summary

Hot stages	$12.0 × 10^6
Wheels	4.0 × 10^6
Other parts	4.0 × 10^6
Total Materials	$20.0 × 10^6

   (Note: The R & D effort costed above must be "balanced" with a design effort of considerable magnitude.)

   c. R & D cost breakdown

Element	Hot Stage Parts $12.0 × 10^6		Wheels $4.0 × 10^6	Other Components $4.0 × 10^6
General research	10 yrs / $400 K/yr	A		
Ore reduction and refinement				
Basic processing and primary fabrication	8 yrs / $520 K/yr			
Properties	8 yrs / $200 K/yr	A		
Codes, specifications, standards	5 yrs / $60 k/yr	B		

Secondary-level fabrication	5 yrs $200 K/yr	A
Performance testing	6 yrs $120 K/yr	B
Maintenance, inspection and repair (nondestructive testing of ceramics)	8 yrs $60 K/hr	B
Documentation	2 yrs $40 K/yr	B
Education	2 yrs $40 K/yr	B

A - principal effort
B - supporting effort

d. Comments

(1) a speed-up of the time scale to perhaps five years should be considered

(2) turbine temperature should be as high as possible to maximize fuel efficiency (aim for maximum combustion temperature)

(3) the cost estimates are too low

(4) the materials effort should be closely coupled with design

C. Magnetically Confined Toroidal Fustion Reactor Specifications

Nominal projected specifications for a "Tokamak" fusion reactor, circa 2000, were given by A. L. Bement as shown below:

1. First wall dimensions:

toroid diameter: 9.4-21 meters

inside diameter: 2-7 meters

thickness: 0.25-2.5 cm

2. Operating environment:

plasma temperature: 10^8 °C

vacuum wall temperature: 400-1000°C

gas pressure adjacent to wall: 10^{-2}-10^{-2} torr

wall neutron loading: ~10 MW/m^2

neutron flux (40% > 1 MeV): 4×10^{15} n/cm^2-sec

ion current to wall: 10^{14}-10^{16} #/cm^2-sec

wall coolant: Li or $LiF-BeF_2$

radiation loading: (unknown)

3. Target wall service life: 5-10 years

D. Impact of Materials Technology on Fustion Reactor Development

1. Materials Selection Against Design Objectives (Under Item 2, Appendix B)

a. candidates for first-wall materials:

* Nb-1Zr

** Mo-based

** Va-based (Ti, Ti-Cr)

** Ni-based

b. relative advantages of Nb-1Zr:

low vapor pressure

high compatibility with liquid Li

high strength

high ductility (coupled with low ductile/
brittle transition temperature)
good fabricability and weldability
medium cost
reasonable availability
low magnetic susceptibility

c. disadvantages of Nb-1Zr:
high nuclear decay heat
high affinity for oxygen, nitrogen, hydrogen
high permeability for tritium and hydrogen
susceptible to hydrogen embrittlement
medium susceptibility to void formation
and swelling

d. general comments:

(1) quartz (high He permeability, low thermal expansion) may be a useful inner coating. It may also be useful in reducing neutron loading on the metal wall (see also liquid surface below).

(2) a porous surface with interconnected pores at the inner wall may be useful in allowing embedded helium to escape.

(3) a replaceable inner liner should be considered.

(4) develop a surface which liquifies during pulses of heat (possibly solidifying between pulses)--

must be compatible with inner wall and have high surface tension, liquid metal embrittlement might be a problem.

2. Materials Processing (Under Item 5, Appendix B)

   a. Objectives

      Develop a highly refined substructure (fine grain size, high dislocation density, highly dispersed precipitates) to impart toughness and minimize void swelling

      Prepare first wall surfaces to impart coolant compatibility and to minimize wall erosion.

   b. Processes

   *** on-site welding may be required (grain coarsening could be a problem in fusion welding)--consider upset welding; diffusion bonding, explosive welding.

   * hot isostatic powder metallurgy products (for limiters)

   * Jaffee metal process (for coolant loops); honeycomb-type structures.

3. Fabrication and Construction (Under Item 9, Appendix B)

   The assembly of large, thin-wall sections into a toroidal shell will probably require new fabrication and joining techniques.

V. General Discussion

A. ARPA Ceramic Turbine Effort

This was given as an example of a focused research effort having the characteristics of high risk, quick return, and large payoff. The objective is to develop ceramic-bladed gas turbines operating at 2500°F, to include one 30 MWe power unit and one 400 Hp propulsion unit. The design of these turbines calls for the use of SiC and/or Si_3N_4 blades, and 25% of the $11M which ARPA is providing to the $35-40M combined effort (Ford and Westinghouse) is going to materials development. In order to adopt this technology for optimal commercial power production it will be necessary to modify present objectives to include the use of coal gas, to add front-end fuel processing or conversion, and to achieve higher combustion flame temperature (~3500°F).

B. Pulsed Laser Fusion

1. Description of the Process

The target fuel is a frozen D,T pellet (radius 0.4 mm, ~60 μgrams), containing high-Z elements to amplify power by bremsstrahlung radiation. The pellet is exposed to intense laser light beams (60K joules), and as a result of the ablation of a portion of the material the remaining fuel is densified to a pressure of $\sim 10^{12}$ atm (matter is Fermi degenerate). Fusion then occurs, yielding ~1800K joules of energy.

2. Materials Problems

The total loading on the wall may be severe; it can vary

greatly depending on design requirements (e.g., microexplosion rate, sphere diameter, blast absorption schemes--such as a liquid lithium inner coating). Fundamental uncertainties are perhaps too large to allow a sharp focus on the specific mechanical behavior of the inner containment sphere wall, although fatigue and nonuniform loading are virtually certain to be difficulties. Other materials problems are associated with laser windows and mirrors, irradiation damage, lithium corrosion, and thermal shock.

C. Superconducting Materials (particularly for totating equipment-generators)

Some of the materials problems discussed were the following:

1. The stability of the superconducting state under vibration is uncertain.
2. There are virtually no engineering data at 4°K for stainless steel, ferritic materials, and plastics.
3. There is a need for low-cost processes for fabricating Nb_3Al and Nb_3Sn superconducting cables which have useful ductility. (Some recent developments at Cal Tech and Brookhaven National Laboratory look promising.)

D. Steam Generators for Pressurized Water Reactors

Steam generators are a current problem in PWR's because of frequent rubing failures (Inconel 600). The origin of the failures is the buildup of corrosion deposits on the secondary

side of the steam generator. These deposits promote crevice corrosion at the water evaporation line and also cause leakage at rolled joints. The consequence of this leakage is the transport of radioactivity from the primary to the secondary system, which greatly aggravates plant maintenance and necessitates elaborate water treatment in the secondary system. An expanded program of corrosion research and fabrication development is required to minimize susceptibility of the steam generator tubing to crevice corrosion. The situation is complicated by fuel failures which contaminate the primary system. The recent increase in fuel element failures is making nuclear plants a less desirable option for base-load power generation.

Steam generator failures have also occurred at the Fort St. Vrain Reactor (an HTGR), apparently due to stress corrosion cracking. In general existing knowledge of stress corrosion cracking is not fully utilized in the design and manufacture of steam supply equipment.

E. General Comments on Future Directions

Severe problems currently exist for the utilities; and there is some question whether they can surmont these problems during the next four to five years. There are uncertainties relating to the overall reliability of nuclear power plants, and environmental problems exist with conventional fossil plants (technology does not yet exist for stack gas cleanup). Some of the problems are institutional in nature. Technology programs

and expertise are highly dispersed witnin government. This is especially true in fossil fuel programs which are not yet well organized. Utilities still depend on equipment suppliers for hardware technology, and consequently they are presently ill-equipped to assume the role of major R & D prime movers. There are over 400 standard setting organizations within the country whose efforts require close coordination. Finally, there is no one in OST with specific expertise in materials, but there should be.

In-depth studies of materials problems in electrical power generation and distribution are required; superficial studies are insufficient -- the problem is too large and too much is at stake. A proper representation is required, to include materials experts and designers from metal producers and fabricators, equipment suppliers, government, utilities, and research institutions. It is important that procedures be established for the timely implementation of recommendations produced by these studies. The ARPA Materials Research Council could make substantial contributions to many of the critical materials problem areas.

Acknowledgement

This study was supported by the Advanced Research Projects Agency under Contract No. DAHC15-71-C-0253 with the University of Michigan.

APPENDIX A

Participants at the Workshop on Materials for Energy Conversion

Panel

Arden L. Bement
Spencer H. Bush
Philip L. Farnsworth
Robert Fitzgerald
Merton C. Flemings
Roy Kaplow
Chester T. Sims

Council Members and Guests

Richard E. Balzhiser
Michael B. Bever
Rudolph A. Black
Robert Coble
Morris Cohen
Pol E. Duwez
Robert Gomer
Edward E. Hucke
Walter Kohn
Paul L. Richards
Michael Tinkham
John E. Wachtman, Jr.

APPENDIX B

Elements of Materials Technology

1. General research (new materials, new processes, new understanding).
2. Materials selection against design requirements.
3. Supply and procurement of raw materials.
4. Ore reduction and refinement.
5. Basic processing to useful material.
6. Primary fabrication.
7. Properties measurements.
8. Codes, specifications, standards.
9. Secondary-level fabrication and construction (established procedures).
10. Performance testing, reliability assessment, failure analysis.
11. Maintenance, inspection and repair.
12. Documentation, information supply and retrieval.
13. Education (public, technical, professional).

APPENDIX C

Steps in the Planning and Execution of New Energy Technology Programs

1. Pre-design Studies
2. Conceptual Design (Title I)
3. Applied Materials Research and Process Development
4. Intermediate Design (Title II)
5. Design Testing
6. Manufacturing Methods Development
7. Detailed Design (Title III)
8. Procurement
9. Component Inspection and Testing
10. Installation and Inspection
11. Plant Acceptance Testing
12. Post-Operational Testing and Refitting

All or part of this cycle may be repeated for the sequential stages of development: prototype, demonstration and production facilities.

A SURVEY OF VARIATIONAL METHODS FOR ELASTIC WAVE PROPAGATION ANALYSIS IN COMPOSITES WITH PERIODIC STRUCTURES*

E. H. Lee

Abstract

The propagation of harmonic waves through composite media with periodic structures is analyzed using Floquet or Bloch theory common in crystal lattice studies. Variational principles in the form of integrals over a single cell of the composite are developed, and provide a means of determining phase velocities and stress distributions in Floquet waves which travel through the composite unchanged in form from cell to cell. The variational principles apply to three-dimensional lattices, but applications to one-dimensional lattices are emphasized since, for this case, the exact solution is available to assess the accuracy of Rayleigh-Ritz variational procedures with the objective of suggestion techniques for application to higher dimensional lattices. Both strain energy and complementary energy principles are utilized. Because of the sharp change in elastic constants from the stiff reinforcing filaments to the softer matrix, the analysis must accommodate discontinuities in strain across the interfaces. The Floquet wave solutions form a complete set of functions over the while of space and thus provide a means of expressing general transient motions. The motion caused by an impulsive pressure variation on the surface of a half space is obtained in this way.

---

*See "Dynamics of Composites", ASME, 1972.

ENTANGLEMENT NETWORKS CROSSLINKED IN STRAINED STATES*

J. D. Ferry

Abstract

Linear 1,2-polybutadience is cross-linked to 0°C by γ irradiation while strained in simple extension with extension ratios from 1.3 to 2.0. During irradiation times up to several hours, entanglement slippage is slight, since the temperature is only slightly above the glass transition. Subsequently, samples are released and reach their equilibrium states of ease at room temperature. From the extension ratio at state of ease, the ratio of $\nu_{\underline{X}}$ (effective network strands terminated by cross-links introduced) to $\nu_{\underline{N}}$ (effective network strands terminated by entanglements) is calculated by composite network theories of Flory and others; and from the extension ratios together with the modulus, measured at small extensions, $\nu_{\underline{N}}$ is calculated explicitly. It appears that $\nu_{\underline{X}}$ increases approximately independent of irradiation, and it corresponds to a molecular weight between effective entanglement loci of about 13,000. This figure, however, which is larger than that deduced from rheological properties of the uncross-linked polymer, is subject to future downward correction for partial entrapment of the entanglements and other refinements.

*See Proceedings of the National Academy of Sciences, 1972.

ENZYME CASCADES AND THEIR CONTROL IN BLOOD PLASMA*

E. W. Montroll

Abstract

Since the uninterrupted, well regulated flow of blood is vital to the life of higher animals it is essential that mechanisms exist to respond to fluctuations induced by accidents and external agents. In the course of the evolution of those species which have survived, no outside expert was available to make quick repairs when fluctuations became too large. Hence the appearance of an assortment of molecules capable of inducing the chemical reactions and participating in the physical mechanisms basic to the spontaneous repair processes had to be concurrent with the development of the complex animals.

Any leaks which appear in the blood conducting tubes have to be plugged before too much of the valuable fluid is lost. Invading organisms which penetrate into the blood stream must be destroyed before their population multiplies to a level which endangers the life of the host. The blood pressure must be controlled and the system should be capable of adjusting to the variations in the elasticity and of the cross section of the tubes of the circulatory system which result from aging.

The response mechanisms should be well tuned so that they are not activated by false signals. Spontaneous blood coagulation

in arteries without leaks is as dangerous as no coagulation to seal lesions. It seems likely that the large numbers of steps which precede the activation of the final repair process in enzyme cascades has evolved to distinguish signal from noise. If each of the preliminary steps follows some clue from the agent which demands the completion of the cascade, their large number would require a precise characterization of the agent. Then the probability that the normal harmless fluctuations in the system would mimic the agent in more than a small number of ways at a given time would be very small. The development strategy of such a control system would be analogous to the design strategy of a sophisticated sea mine which is constructed to explode under a special kind of ship such as an aircraft carrier and which is also supposed to be difficult for the enemy to sweep. A mine sweeper with a noise maker which simulates the noise of the carrier would only activate one component of the response system; a long electrical conductor towed by the sweeper might mimic the magnetic field of the hull of the carrier and activate the next component. However, unless the sweeper could also generate the special pressure field developed by such a long streamlined object in rapid motion in the water the mine would eventually return to its inactive state without exploding.

The volume of clinical and biochemical literature of enzyme cascades, especially blood coagulation probably exceeds that of almost any other specialized technical subject. Some of the important basic ideas have resulted from curious old obser-

vations which one would hardly have expected to have the same status as those from carefully planned experiments. Clever mechanisms which are still of interest were proposed long before one had any idea of the structure of the molecules involved. There is even a certain charm to a subject of such great activity which can still have important process factors named after patients who lacked them - Christmas, Hageman, Stuart - rather than after the physicians who discovered them. When the Mendelian laws of heredity became generally known at the turn of the century one of the first applications to human genetics was the analyses of the inheritance of hemophilia. This was, of course, facilitated by known records of victims of the disease in royal families whose geneologies were documented.

The process of hemostasis and the formation of the hemostatic plug involves (i) highly specialized cells of the blood stream, the platelets, which have an afinity for sites in the blood vessel where damage has occured, (ii) blood coagulation, the phase transition of monomer fibrinogen into a polymeric fibrin network in and around the aggregated platelets, stabilizing the plug, and (iii) the dynamics of the blood vessel which leads to these processes and which finally must perform the healing process.

An important part of the extensive literature on hemostasis is to be found in the proceedings of many conferences on the subject. Some of the conferences such as the Hamburger Symposium Über Blutgerinnung and the Wayne State Conferences are held at regular intervals as are those of various international

committees on blood such as the Committee on Blood Clotting Factors and the Committee on Haemostasis and Thrombosis. One of the main journals on the subject, Thrombosis et Diathesis Haemorrhagica has a supplement which contains the proceeds of many of the important conferences. The proceedings of the conference of the British Royal Society (organized by R. G. MacFarlane) on triggered enzyme systems in blood plasma (1 July 1969 issue of Proc. Roy. Soc. B) has been very valuable to the author in preparing this review.

Since the main intent of the review is to present a guide to the literature to physicists and engineers who wish to learn something of biological control mechanisms, it will be made more self contained by starting with a few remarks on nature of proteins and classical enzyme kinetics. An attempt has been made to use a minimum of medical terminology. If some has crept in with insufficient definition the uninitiated can find good explanations in one of the standard books such as Best and Taylor, Physiological Basis of Medical Practice (Williams and Wilkins). Two excellent books have recently appeared on blood clotting, one edited by W. H. Seegers and entitled Blood Clotting Enzymology (Acad. Press 1967) and the other, edited by K. Laki, entitled Fibrinogen (Dekker 1968).

---

*To appear in Annual Reviews of Biophysics and Bioengineering Vol. II

THEORETICAL MODELS AND EXPERIMENTAL PROPERTIES OF LIQUID METALS

J. L. Margrave

Abstract

There are now available a great many measurements of various properties of liquid metals over wide ranges of temperature and pressure and it is informative to examine these properties in the light of the various available theories as proposed by Mott, MacDonald, Bernal, Ziman, Faber, Edwards, Knight, Heine and others. Of special interest in this report are the structures deduced for liquid metals as well as the absolute values, and perhaps more crucially, the temperature coefficients of these values for electrical conductivity, thermal conductivity, Hall coefficients, Knight shifts, specific heats and optical properties.

The studies of high melting metals by the technique of levitation calorimetry are being continued and new data for Ni and Ta are evaluated.[1] The general problems of correction for heat losses due to radiation and thermal conduction to the gaseous atmosphere have been resolved and all data adjusted appropriately.[2] Favorable intercomparisons have been made between our work and Russian studies of liquid vanadium[3] and between our work and studies at Sandia (Albuquerque) or liquid copper.[4]

---

1. D. W. Bonnell, A. J. Valerga and J. L. Margrave, to be published.
2. D. W. Bonnell, A. J. Valerga and J. L. Margrave, to be published.
3. L. Gurvich, private communication, 1972.
4. H. P. Stephens, presented at the 27th Calorimetry Conference, Park City, Utah, July 22, 1972.

ENERGETICS OF STRAINED ORGANIC MOLECULES AND OF VARIOUS "CARBON" SAMPLES BY COMBUSTION CALORIMETRY

J. L. Margrave

Abstract

High-precision combustion calorimetry for pure organic compounds can lead to important energetic information about bond energies and strain energies in unusual configurations. For example, various propellanes (synthesized by Prof. P. Eaton, University of Chicago) have been burned and strain energies evaluated.[1] Also, a series of diazo-organic compounds has been burned to provide a reliable value for the energy contribution of the group $\vdash N=N \dashv$.[2]

Studies of the heats of combustion of various "carbon" samples, including some of the materials prepared by polymer pyrolyses by Prof. E. Hucke, are in progress in an effort to establish quantitative enthalpy differences and to ink these to structural characteristics.

---

1. J. L. Wood and J. L. Margrave, to be published.
2. J. L. Wood, P. S. Engel and J. L. Margrave, presented at 27th Calorimetry Conference, Park City, Utah, July 20, 1972.

STRUCTURAL STUDIES AND CHEMICAL SYNTHESES IN LOW-TEMPERATURE MATRICES

J. L. Margrave

Abstract

New techniques for structural studies and for synthetic chemistry are explored in which potentially reactive (unstable) species generated either in arcs, discharges or furnaces, or by pyrolysis are co-condensed in either inert or reactive matrices at low temperatures and then allowed to warm up and react.

Structural studies of current interest include:

(1) the establishment of the structure LiNC as the preferred formula for indicating the proper order in "lithium cyanide". NaCN and KCN are normal cyanides.[1]

(2) the determination of bond angles for the species Li_2S, Al_2S, Ga_2S, In_2S, Tl_2S and for $SiCl_2$ and $SiBr_2$.[2]

(3) the determination of the symmetry (C_{3v}) and other parameters for matrix-isolated CCl_3-radicals by esr spectroscopy.[3]

In the synthetic area, interest is directed to the reaction of atoms and molecules at low concentrations in inert matrices,[4] e.g.,

(a) $(Cr + 2\ C_6H_6)_{\substack{Ar\ or\ Ne \\ Matrix}} = Cr(C_6H_6)$

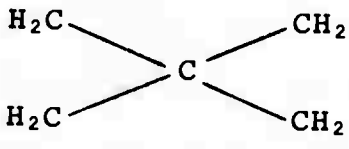
(b) $(Li + O_2)_{\substack{Ar\ or\ Ne \\ Matrix}} = LiO_2$

(c) $(SiF_2 + BF_3)_{\substack{Ar,\ Ne\ or\ N_2 \\ Matrix}} = SiF_3\text{-}(SiF_2)_n\text{-}BF_2$

and also to the reactions of atoms and molecules directly with reactive matrices,[5] e.g.,

(d) $(CH_4 + F/F_2)_{F_2\text{-Matrix}} = CF_4,\ CF_3H,\ CF_2H_2,\ CFH_3$

(e) $(BF + BF_3)_{BF_3\text{-Matrix}} = B_2F_4,\ B(BF_2)_3$, etc.

(f) $(C + C_2H_4)_{C_2H_4\text{-Matrix}} =$  etc.

(g) $(U + CO)_{CO\text{-Matrix}} = U(CO)_x$ where x = 1,2,3,4; etc.

One of the major advances in this field has been the development of low-temperature devices and techniques which allow synthesis on the multi-gram scale and thus bring this approach to synthetic chemistry into direct competition with traditional laboratory methods.

---

1. Z. K. Ismail, R. H. Hauge and J. L. Margrave, to be published.
2. R. H. Hauge and J. L. Margrave, to be published.
3. G. Maass, R. H. Hauge and J. L. Margrave, Z. anorg. u. allgem. Chemie, 1972, in press.
4. See for example the work of P. L. Timms, Bristol; Lester Andrews, Univ. of Virginia; D. Milligan, National Bureau of Standards; and J. L. Margrave, Rice University.
5. See for example the work of P. L. Timms, Bristol; P. Skell, Penn. State University; W. Weltner, Univ. of Florida; A. Streitweiser, Univ. of California, Berkeley; and J. L. Margrave, Rice University.

SYNTHESES, STRUCTURES AND THERMODYNAMIC PROPERTIES OF PERFLUOROCARBONS

J. L. Margrave

Abstract

The increasing interest in solid perfluorocarbons as practical lubricants, as battery constituents, as chemical reactants in diamond syntheses and for other applications makes it especially important to have workable syntheses and reliable structural and thermodynamic properties. Our patent covering both certain compositions of matter as well as broad synthetic approaches for the syntheses of super-stoichiometric CFX^R has recently been issued.[1] Other preparative approaches to pure fluorocarbons and to the preparation of "gradient polymers" in which perfluorocarbon surfaces are formed on hydrocarbon substrates are described in other patent applications. Thermal stabilities, gas permeabilities, chemical corrosion resistance and other properties are under investigation.

Among the structural studies in progress are (1) nmr studies of solid CFX which gives information on the number and types of groups in the solid (-CF ; $\rangle$CF ; $\rightarrow$CF).

---

1. R. J. Lagow, R. B. Badachhape, J. L. Wood and J. L. Margrave, U.S. Patent No. 3,674,432, issued July 4, 1972.

(2) ESCA studies on CFX to identify characteristic energy levels for various groups.

(3) low-temperature C_p-measurements to provide the absolute entropy at 298°K for use in electrochemical calculations for the Li/CFX cell.

(4) heats of formation of C_4F, $CF_{0.7}$, $CF_{0.9}$ and $CF_{1.2}$ to establish the enthalpies as a function of composition.

(5) studies similar to the preceding for compounds like $C_{10}F_{18}$, $C_{14}F_{24}$, $C_{18}F_{30}$, etc.

SOLAR ENERGY, A NATURAL RESOURCE FOR EVERYONE

J. L. Margrave

Abstract

Three new factors now on the scene in the United States and throughout the world may finally make the exploitation of solar energy, which has long been practical in a scientific and engineering sense, economically feasible and politically desirable: The factors are

(1) the tremendous surge of interest in non-polluting energy sources.

(2) the growing recognition that fossil fuel resources are really finite and the realization that this dire prediction for the future is not so far away as shortages of natural gas, fuel oil and electric power lead to "brown-outs", "black-outs", and to price increases and restrictions on usage.

(3) the societal concern about better living standards for all people which creates an impetus for better housing and new construction with minimal operating costs. Here is an opportunity to develop and install on an extended basis solar water heaters and house heating units which are efficient and attractive both to the resident from the viewpoint of economics and to society as a whole from the two viewpoints outlined above.

Scientists, engineers and architects can work together with governmental agencies to facilitate the development and use for the people of the world this free, non-polluting energy resource.

HYDROGEN AND HYDRIDES--CHEMICAL ENERGY CARRIERS

J. L. Margrave

Abstract

As a part of the solution to the long-range energy problem, one must devise a way to store and transport--probably in chemical form--the energy produced by burning fossil fuels or by nuclear fission or fusion reactors.[1] From an economic point of view, some of the most attractive energy carriers are elemental hydrogen and the various simple hydrides. Gaseous hydrogen can be distributed under pressure through the same systems now being used for natural gas, and the binary hydrides are liquids or solids, often with reasonable properties of the same sort as gasoline.

Among the sources of hydrogen currently available are the pyrolysis of hydrocarbons, and of coal and oil shales. On a long-range basis, the production of hydrogen will probably be electrolytic--from water, or KHF_2 or molten metal hydrides. One must consider the economics and the scientific needs in choosing an electrolysis process and produce both hydrogen (a reducing agent) and probably either fluorine or oxygen (oxidizing agents). There

1. See series of articles, "Hydrogen, Likely Fuel of the Future", Chemical and Engineering News, p. 14, June 26; p. 16, July 3 p. 27, July 10, 1972.

is some advantage in choosing to produce fluorine because of its greater chemical versatility.

The metal hydrides themselves are interesting solid energy carriers, capable of yielding hydrogen either on heating or on hydrolysis. Light-element hydrides are especially attractive on an energy/gram basis.

POLYCHROMATIC X-RAY DIFFRACTION. A RAPID AND VERSATILE TECHNIQUE FOR THE STUDY OF SOLIDS UNDER HIGH PRESSURES AND HIGH TEMPERATURES

L. M. Albritton and J. L. Margrave

Abstract

The use of polychromatic X-ray diffraction with multichannel energy analysis offers a rapid and versatile new approach to the problem of X-ray diffraction of solids, especially under conditions of high temperatures and high pressures.[1] Phase transitions at pressures up to ∿50,000 atm. and temperatures up to ∿500°C have been observed for alkali halides, PbF_2 and other systems.

---

1. L. M. Albritton and J. L. Margrave, High Temperatures-High Pressures, 1972, in press.

CONTINUUM DESCRIPTIONS OF DEFORMATION

E. H. Lee

The stress-deformation relations for polymers fall into several categories, which depend on the chemical structure and physical variables (e.g., temperature) of the specimen subject to testing. The structure of the constitutive relations differs appreciably for small and finite strain situations. In the latter case, care must be exercised in including finite rotations and the change in the shape of an element on the definition of stress in the deformed state, whereas in the former, infinitesimal strain theory permits the undeformed geometry or the deformed geometry to be used interchangeably in defining stress, thus providing a measure of uncoupling between stress and deformation. The coupling associated with finite deformation is termed geometrical nonlinearity. Nonlinearity can arise in the stress-deformation relations independently of this geometrical effect and is termed material nonlinearity, which, for certain materials, can be important even at small strains.

What are the ranges of satisfactory representation of constitutive relations, e.g.,

glassy and crystalline states: linear or nonlinear small strain theory.

rubbery state: linear and nonlinear small strain theory, finite strain theory.

Can we characterize the circumstances in which material non-linearity can be expected to be important? What forms of operator relationship are most convenient from the standpoint of a) analytical work, b) introduction of measured material characteristics, for either infinitesimal strain or finite strain situations?

Since the influence of temperature on stress-deformation properties of polymers can be so marked, and thermo-mechanical coupling can constitute appreciable sources of heat energy, what are the presently available most satisfactory forms of thermo-mechanical constitutive relations? Low thermal conductivity magnifies the thermal influence, and can make these effects dominant in applications. Under what circumstances can the temperature shift concept (thermo-rheologically simple material behavior) provide an adequate basis for simplifying the structure of thermo-mechanical relations?

Thermo-mechanical relations can be based on irreversible thermodynamic approaches using hidden variables, or by direct use of functional forms of the observed mechanical and thermodynamic variables, (e.g., dependent variables depending on the history of variation of the independent variables through integral or rate terms). Is our present knowledge sufficient to suggest appropriate forms of such laws, and what experiments have been done, or need to be done, to provide the required knowledge of the physical behavior to permit such laws to be formulated? Is it appropriate to use functional forms for such quantities as internal energy,

specific heat and thermal expansion, in addition to the mechanical functional laws of classical viscoelasticity?

Acknowledgement

This study was supported by the Advanced Research Projects Agency under Contract No. DAHC15-71-C-0253 with the University of Michigan.

GRADIENT MATERIALS

R. L. Coble

The superior performance of gradient or graded materials over materials with uniform properties for armor has been known for a long time--current results show further improvements but leave unanswered several questions about design--"how complex a basic structure is needed, how many layers should be assembled, of what properties, in what sequence, with what internal phase distribution, and what pre-stress conditions are desirable? The theory of ballistics response seems either to be inadequate or has not been satisfactorily tested to resolve simple questions such as whether continuous gradients in properties are needed or whether step function variations would be preferred. A coordinated model studies program should be launched to resolve this question. The feasibility of producing continuous gradient structures by powder metallurgy techniques has been demonstrated. An alternate approach by diffusion of a component to give continuous property variation has not been attempted. Comparison with glued/brazed assemblies with controlled states of pre-stress should be conducted on materials specially selected to answer the design questions without concern about current fabrication feasibility into useful hardware.

The various microstructures which can be produced, with variable connectivity and volume fractions of interpenetrating ductile and brittle phases should be studied for variations in plastic and elastic constraint, shock wave alternation, and for crack stopping ability independently and in addition to the ballistic performance to contribute to understanding the latter.

Acknowledgement

This study was supported by the Advanced Research Projects Agency under Contract No. DAHC15-71-C-0253 with the University of Michigan.

BIAXIAL STRESS RELAXATION IN GLASSY POLYMERS

S. S. Sternstein

Experimental evidence was presented to demonstrate that strain field can alter stress relaxation kinetics in the linear viscoelastic regime. The origin of this kinetic coupling was shown to be due to the isotropic component of the strain tensor and is due to two processes: (1) the isotropic strain accelerates the long time spontaneous relaxation of the metastable reference state for the glassy polymer. This leads to an excess spontaneous densification that may explain yield point recovery and stress aging in glassy polymers. (2) The Viscoelastic relaxations are shifted to lower timescales in a manner suggested by Ferry. These two process phenomenon explain the observation that tension relaxation is slower than torsion relaxation. (In press, J. Applied Phys., Oct., 1972) Classical linear viscoelasticity theory must be revised for application to glassy polymers.

In terms of a molecular basis, it is suggested that the dilatational strain component expands a Cohen-Turnbull type cage surrounding a segmental oscillation, of the chain backbone thereby allowing rotational isomerism of the chain to a lower free energy state. Both inter- and intramolecular restrictions to

isomerism are important but principally it is the intermolecular barrier that is most sensitive to dilatational strain. A rotational jump results in a local free volume excess that increases the probability that the next jumps will occur adjacent to the previous jump. Thus, an accelerating rate process is observed that may ultimately lead to a precipitous local chain motion that results in a shear band or craze of highly oriented chain segments imbedded within a glassy matrix. It is suggested that an Ising type model may be used to describe the phenomenon provided that at least a two-dimensional model is used, one coupling energy set along the chain, one isotropic set transverse to the chain to describe the intermolecular energies.

Acknowledgement

This study was supported by the Advanced Research Projects Agency under Contract No. DAHC15-71-C-0253 with the University of Michigan.

ELASTIC NETWORK THEORY

S. S. Sternstein

A theory has been derived previously for predicting global elastic properties of a discretely bonded array of conservative "elements" whose deformational characteristics are presumed to be known. Basically, integral equations are derived for the macroscopic components of the stress tensor and the work function in terms of unknown elemental deformation functions. These functions are determined such that the work function is minimized subject to constraints imposed by the applied (macroscopic) stresses. The problem is solved by the calculus of variations and is of the isoperimetric type. The elemental deformation functions form a set whose functionality is determined by the particular elastic behavior of the network's constituent elements.

The theory is being applied to protein fiber networks whose mechanical properties are being studied by Ferry. Here, the principal problem is to develop a suitable description of the elastic behavior of the individual fibers and then to solve the network equations for appropriate boundary conditions.

A second application is to crosslinked elastomer networks (the rubber elasticity problem). For non-Gaussian polymer molecules certain operational difficulties preclude solution for

the crosslink deformations unless an artificial orientation energy term is introduced into the elemental free energy expression. However, the treatment of Reiss circumvents this difficulty and allows both treatments to be combined in a way that simultaneously minimizes the network free energy and maximizes the probability of the calculated crosslink deformation states.

Acknowledgement

This study was supported by the Advanced Research Projects Agency under Contract No. DAHC15-71-C-0253 with the University of Michigan.

COMMENTS ON PROTEIN NETWORKS

J. D. Ferry

The current theoretical treatments of macromolecular networks should be valuable in elucidating some problems of protein structures. A summary of some pertinent facts follows.

1. Fibrinogen (a ternodular rod about 600Å long), after chemical modification by thrombin, polymerizes spontaneously to form long, thin fiber elements whose length and width increase with time. There is a broad distribution of final lengths and widths and it can be controlled by the clotting conditions. The fibers are mostly remarkably straight but branch here and there to form a network. The coarser the network (low pH, low ionic strength) the greater its deformability without rupture; paradoxically, the coarser the fibers the more flexible they seem to be.

2. Depending on whether fibrinoligase enzyme and calcium ion are present or absent, the clot is "ligated" with primary chemical bonds (amide bonds) and insoluble, or is "unligated" and can be dissolved in concentrated NaBr or other reagents to give units with the same size and shape as the original fibrinogen, then repolymerized reversible, etc. Yet, the ligated and the unligated clot look essentially the same in the electron microscope.

3. The shear modulus of unligated clots, measured in a time scale of the order of 10^{-2} sec., is roughly proportional to the square of the volume fraction of fibrinogen, with a proportionality coefficient of the order of 10^6 dyne/cm^2, but both coefficient and exponent are of somewhat doubtful reproducibility; better controlled experiments are needed. The shear modulus of ligated clots under comparable conditions is somewhat higher (perhaps factor of 2).

4. The storage modulus (as function of frequency) or relaxation modulus (as function of time) of ligated clots is nearly constant over a remarkably wide range of time or frequency scale (several decades), indicating no relaxation processes. Unligated clots, by contrast, show stress relaxation within periods of a minute or more. When the stress has relaxed nearly to zero at constant strain and then the strain is brought to its original zero point, a reverse stress is observed, indicating that the structure has rearranged to a new rest position. The reverse stress in turn relaxes with time. (The influence of network coarseness on this contrast between ligated and unligated clots has not yet been fully explored.)

5. Coarse, unligated clots with fibrin volume fraction about 0.004 can be compacted with expulsion of most of the solvent (buffered salt solution) to form a two-dimensional network structure ("fibrin film") with a fibrin volume fraction of the order of 0.2. Such a film can be extended in its plane in

simple elongation at least 100% and its behavior mimics rubber-like elasticity though with some retardation of response and a permanent set of the order of 20%.

Collagen[6,7]

Natural collagen structures such as tendons have complicated supramolecular morphology which makes their mechanical properties more difficult to interpret. However, films of reconstituted collagen can be prepared, with protein weight fractions of 0.4 to 1.0, in which the largest unit is the tropocollagen molecule, about 15Å in diameter and 2800Å long. Such films apparently have very high moduli, at least in the range of 0.6 to 1.0 weight fraction of protein, but creep without any elastic recovery as though the fibrous elements slip past each other with an effective viscosity which decreases rapidly with increasing proportion of diluent. As a result, the apparent modulus measured at a fixed time interval drops precipitously with increasing diluent. More detailed viscoelastic data are not yet available.

The relation of the mechanical properties of the three-dimensional clot networks and the two-dimensional film networks to the mechanical and molecular characteristics of the fibers of which they are composed is an important unsolved problem.

Acknowledgement

This study was supported by the Advanced Research Projects Agency under Contract No. DAHC15-71-C-0253 with the University of Michigan.

References

1. K. Laki, ed., *Fibrinogen*, Dekker, New York, 1968.

2. J. D. Ferry and P. R. Morrison, J. Amer. Chem. Soc., 69. 388, 400 (1947).

3. J. D. Ferry, M. Miller, and S. Shulman, Arch. Biochem. Biophys. 34, 424 (1951).

4. W. W. Roberts, Ph.D. Thesis, Northwestern University, 1971, and 2 papers in preparation (with E. Mokros).

5. W. W. Roberts and J. D. Ferry, unpublished experiments.

6. I. V. Yannas, Rev. Macromol. Chem., C7, 49 (1972).

7. I. V. Yannas and C. Huang, Macromol. 5, 99 (1972).

POLYMERIC ENTANGLEMENT NETWORKS CROSS-LINKED IN STATES OF STRAIN

J. D. Ferry and S. S. Sternstein

Abstract

The results of recent experiments on states of ease of amorphous highly entangled polymers which have been cross-linked while strained in simple extension are reviewed, and deviations from the predictions of an earlier report on this subject are examined. In particular, the approach to the state of ease following release of stress after cross-linking is unexpectedly slow, and therefore the kinetics of this process have been treated. With the approximation of linear viscoelasticity, it is found that the sum of the relaxation moduli of the engantlement and cross-link networks can be calculated from observations of sample dimensions as a function of time during approach to equilibrium. For neo-Hookean viscoelasticity, the relaxation moduli of the individual networks can in principle be calculated. Further experiments designed to clarify the phenomenon are outlined.

ON THE ANALYSIS OF MATERIALS RECYCLING

Michael B. Bever

Abstract

This memorandum reviews the general characteristics of the recycling of materials and summarizes pertinent definitions, parameters and functional relations. The analysis of recycling is then considered in terms of engineering, statistics, economics, pollution control, thermodynamics, short-range forecasting and long-range considerations. Published analytical treatments are surveyed systematically and in some respects the analysis of recycling is developed further. Inherent in this approach is the interpretation of recycling of all types of materials from a generalized and unified viewpoint.

ON THE ANALYSIS OF MATERIALS RECYCLING

Michael B. Bever

1. Introduction

Recycling, salvage for reuse, substitution and synthesis of materials and materials-oriented design are industrial practices which stretch the supply of materials and contribute to the conservation of materials.[1,2] Among these practices recycling has probably attracted the greatest amount of attention. The technical, economic and ecological literature, as well as trade publications, have dealt with various aspects of recycling, but the approach has been predominantly qualitative. The quantitative and analytical aspects are only beginning to receive thorough consideration.

This memorandum will deal with the analysis of the recycling of materials. The emphasis will be on a systematic survey of essential features. Inherent in this approach is the analysis of the recycling of all types of materials from a general and unified viewpoint.

2. General Characteristics of Materials Recycling

(2.1) Recycling can perform two functions: the conservation of materials and the disposal of waste. Materials differ in the extent to which their recycling fulfills these functions. The list of recycled materials, therefore, forms a spectrum which ranges from "all conservation of materials" to "all waste disposal".

(2.2) Two forces may drive recycling: market forces and environmental forces. The market forces are predominant in the established secondary materials industries. They also result in the conservation of materials and in the disposal of at least some waste but this is incidental rather than deliberate. The environmental forces are primarily directed at waste disposal.

(2.3) A generic flow sheet representing the structure of established industrial recycling can easily be visualized. Such a flow sheet will be made up of three loops - one loop each for home scrap, industrial scrap and old scrap. (These terms apply to metals. Other materials have their own terms, such as "cullet" for recycled glass and "paperstock" for waste paper.) The three types of recycling differ in the ratio of recycled to recyclable material, as discussed in Section 4.3, the relevant time period and the elasticity of supply. The recycling of home scrap and industrial scrap is important for two reasons: it affects the costs of primary production and it uses up the sometimes limited capacity of primary production processes for absorbing secondary materials.

(2.4) The ratio of recycled to recyclable material differs for different materials. It decreases roughly in the order: precious metals, non-ferrous metals such as copper, ferrous metals, certain other non-ferrous metals such as zinc, paper, rubber, glass and plastics.

(2.5) Two effects are typically associated with recycling: a saving in energy and a reduction in pollution relative to primary production. These effects, however, are not inherent in recycling and may be absent.

(2.6) Recycling helps in the disposal of waste, but a distinction must be made between products which are segregated and the general waste stream. At least a fraction of discarded automobiles and containers - both steel and aluminum - are in the category of segregated post-user residues. These residues are recycled by the established secondary materials industries. On the other hand, recycling is only beginning to play a role in the reduction of the general waste stream. It will require the development and use of new sorting and reclamation technologies and the development of new markets for the materials produced.

(2.7) It should be emphasized that the various classes of materials - nonmetals as well as metals - fall into the same general pattern described in the foregoing. Many details, however, differ. One major difference, even among metals, is the type of plant which processes the recycled materials. At one extreme the secondary material, for example steel scrap, is processed along with the virgin raw material; at the other extreme, the secondary material, for example aluminum scrap, is processed in separate plants; some materials, for example, copper scrap, are processed either in primary or in separate secondary plants.

The foregoing points are discussed in detail in a recent review of recycling.[3] A thorough study of the major secondary materials industries is now available.[4] Other general references are cited in later sections of this memorandum.

3. Definitions and Parameters

The analysis of recycling involves various definitions and parameters, the most important of which will be covered in this section.

(3.1) The amount recycled, usually expressed in units of weight, for a given period and economy, is the fundamental quantity of recycling. It presents problems of measurement, but not of definition.

(3.2) The scrap generation coefficient is the ratio of scrap generated by manufacturing industries to their total consumption of a given material. It thus refers only to industrial scrap. Values of this coefficient have been collected for the year 1954 in a study conducted by the Department of Commerce.[5]

(3.3) The life cycle is the average life of a given material in manufactured goods (usually averaged by weight). It governs the amount of post-user scrap which becomes available during any time period. The life cycle is difficult to estimate even for a single class of products. Examples of estimates may be found in the literature.[4,6] The life cycle has no significance for home scrap and for industrial scrap; it is zero for the former and may usually be neglected for the latter, although when the volume of production changes there is a delay period for industrial scrap (an "impedance" effect).

(3.4) A solid waste generation rate has been defined by Randers and Meadows[7] as follows: if the number of products (understood as units of a product) in use is P, and the average life time of the products L, then the number of products discarded per year is P/L. If the average amount of waste from each product is w, the solid-waste generation rate in the steady state is Pw/L.

(3.5) Secondary reserves comprise the total inventory (or pool) of a material which is potentially available for recycling. In analogy to the terminology of economic geology the total inventory may be classified as "secondary resources". The value of the secondary reserves depends on these resources and the recovery coefficient applicable to them.

(3.6) The recycling potential is the amount that could be recycled under ideal (or real) conditions in any given time period. It is related to secondary reserves.

4. Functional Relations

Parameters characterizing recycling can be related to other variables or to each other, resulting in various functional relations. Several such relations will be discussed in this section.

(4.1) The ratio of recycled material to current demand has obvious meaning and has been used extensively in published statistics. This ratio has also been called the "recycling level" and defined as "percent old scrap in total supply".[8] Primary resources tend to be conserved as the value of this ratio, applied to post-user scrap only, increases. The amount of post-user scrap that becomes available for recycling depends on the life cycle and the rate of consumption at the beginning of the life cycle.

A change in the rate of the demand for a material affects the ratio of recycled material to current demand. If the average life cycle of products is long, and the rate of demand increases rapidly, the amount of recycled material will be a small fraction of current demand. This has been and continues to be the case for aluminum. If the rate of

demand growth is exponential, the ratio of recycled material to current demand is constant.[8] It should be noted that if the life cycle is short, and the amount recycled (in unit time) remains essentially constant, the ratio considered here becomes more meaningful. This is the case for recycled paper.

(4.2) The reuse ratio is the input of reused material as a fraction of the total materials input.[9] It is the microanalytical equivalent of the ratio discussed in 4.1.

(4.3) The ratio of recycled material to recycling potential is a measure of the recycling efficiency. This ratio is invariably unity for home scrap, tends to approach unity for industrial scrap, and varies for post-user scrap. It has also been called "recycling rate" and defined as "percent material recycled after useful life".[8]

5. Technical Information and Analysis

The literature on the processing of secondary materials is extensive. It covers both current industrial practice and ongoing research and development, largely Government supported. A few representative items are listed in the References.[10,11,12,13]

6. Flow Sheets

Flow sheets are a useful device for analyzing and presenting technical information on the processing of recycled materials. Such flow sheets have been published for metals and other materials.[4,13] Flow sheets are particularly well suited for showing the relation between primary and secondary sources in the production of a material.[4,13,14] The literature contains several generalized flow sheets with varying emphasis on technical and economic aspects.[6,13]

7. Statistics and Industry Studies

Statistics on industrial recycling are collected and published by the U.S. Bureau of Mines,[15] the Bureau of the Census, other U.S. and foreign Government agencies and by trade associations. The data covered include the tonnages and prices of secondary materials and other information on sources and applications of secondary materials. Statistical information on structural features of the secondary materials industries, such as the size and location of companies active in this field, is also available. On the other hand, published data on the costs of recycling various materials are limited.[16]

Statistical information for specific periods can be presented in the form of flow sheets. The Bureau of Mines publishes economic data in this form.[15] A detailed flow sheet of this kind has been worked out for copper.[17]

Studies of the industries producing the major materials are typically concerned with sources of supply, including secondary materials, the structure of the industry, and its economic characteristics, such as price movements and markets. Such studies include technological trends and their interaction with economic developments. Examples are given in the references;[18,19] publications dealing exclusively with the economics of industrial recycling are also listed.[4,14,20,21]

8. Economic Analysis

The amount of attention which economists have given to industrial recycling activities has been remarkably limited. One reason for this probably is the difference between industrial recycling and ecologically

motivated recycling, which is still more of an objective than a current practice. This dichotomy calls for different analytical treatments: analysis of the former by conventional economics and analysis of the latter by welfare economics. Economic aspects of recycling have been considered by engineers.(22)

Economic analysis can contribute to the understanding of the established recycling industries. The following brief list of economic concepts suggests obvious applications to recycling: price elasticity, economies of scale, labor-intensive and capital-intensive industries, competition between materials, concentration and integration of industries.

The more general economic approach to recycling considers all waste, including recoverable scrap, as residuals of production or consumption. The role of these residuals has been analyzed in terms of welfare economics, including such concepts as externalities and residuals management models.(9,23,24,25)

Economic analysis of recycling must be concerned with its optimization. This has led to the recognition that maximum recycling in general does not represent the optimum because it would require an allocation of excessive resources, specifically energy and manpower. In principle, an optimum recycling ratio can be determined for any given case.(9)

9. Analysis of Pollution Effects

Recycling reduces land pollution inasmuch as it disposes of waste. Recycling processes, however, cause their own pollution, particularly air pollution. In general, these pollution effects tend to be smaller than those of primary production but this is not necessarily so.

Various aspects of the pollution caused by secondary materials production have been investigated.[26,27] Much further work, including work concerned with comparing the pollution effects of primary and secondary production, still needs to be done.

10. Thermodynamic Analysis

Thermodynamic concepts have been applied to the analysis of recycling. Rose, Gibbons and Fulkerson tentatively proposed a "minimum energy path" for waste management.[28] The concept of entropy has obvious applications to the dispersion of materials in connection with recycling.[29] Berry has considered recycling in terms of thermodynamic potentials and has applied this analysis to the manufacture of automobiles with and without the use of recycled materials.[30]

In a more restricted technical sense the energy requirements of production from secondary sources have been compared to the energy requirements of primary production. A report of the Oak Ridge National Laboratory is concerned with calculated free energies for the production of metals from different minerals and scrap.[31] These energies are theoretical thermodynamic values, such as the energy of reduction of an oxide or the energy required for melting. The calculations assume 100 percent efficiency (except in the conversion to equivalent coal energies) and do not take into account the energies expended in auxiliary operations, such as transportation and the collection and physical processing of scrap. The energy requirements of production from secondary sources, including comparisons with the requirements of primary production, should be analyzed in detail.

11. Forecasts of Future Recycling

The First Annual Report of the Secretary of the Interior under the Mining and Minerals Policy Act of 1970 states that there is "no way to forecast" recycling in the year 2000.[32] In contrast to this some of the graphs in the Interim Report of the National Commission on a Materials Policy, dated April, 1972, include specific projections to the year 2000 of recycled materials.[33] These projections, however, do not seem to be intended as forecasts, but as indications of the tonnages recycling must contribute if expected future demands are to be met.

Quantities of various secondary materials which will be recovered by recycling in the future were forecast in the study by Landsberg, Fischman and Fisher.[6] In an analysis of the prospects for ferrous scrap Elliott has made a detailed forecast for a single industry.[34] This analysis is of particular interest because several countervailing dynamic factors, which are already known, will affect the future demand-supply relation for ferrous scrap. On the basis of these factors Elliott predicted an increase in scrap demand and suggested a possible shortage of supply. This prediction contrasted with the generally pessimistic attitude of the ferrous scrap industry toward the short-term prospects at the time the analysis was made (1971).

Forecasting of future recycling volumes can probably be improved in several respects. Industrial scrap of metals can be estimated from the scrap generation coefficients together with estimated annual future production levels in the manufacturing industries. The relative importance

of various industries probably will change in such a way that the overall scrap generation coefficient may be assumed to increase slightly as the economy continues to mature.

Methods for estimating the pool of post-user scrap available for recycling probably can be refined. Past volumes of consumption of different types of materials in the production of various types of goods are generally quite well known. The life cycles of major product groups must be estimated and applied to the products manufactured in past periods. The fractions recycled are probably most difficult to estimate and will be affected by such variables as future prices and labor costs. (A similar type of calculation has to be made for the determination of the ratio of actually recycled to potentially recycled materials discussed above in 4.3. An example is a calculation of the ratio for aluminum.[35])

12. Long-Range Analysis

The long-range prospects for recycling of a material depend on the future demand level and the relative contributions of primary and secondary production toward satisfying this demand. Perbix has analyzed the effects of changes in demand.[8] He assumed that a period of exponentially growing demand is superseded by a period of constant demand. Perbix interpreted the period of growing demand as a "primary production or exploitation economy" and the period of constant demand as a "recycling or conservation economy". These two major periods are separated by a transition period, during which primary production decreases and secondary production continues to increase. At the end of

the transition period both levels have attained constant levels. From this point on primary production makes up the difference between the recycling level and the total industrial demand level.

Perbix assumed that primary production will decline to accommodate an increasing volume of secondary production. This assumption ignores the competition which will occur between primary and secondary production. In spite of this limitation, the results of the analysis are of great interest. One conclusion is that the length of the transition period equals the average useful life of the material. Perbix also analyzed the case of linear demand growth superseding exponential growth.

Perbix's analysis is akin to systems analysis by the methods of industrial dynamics.[36] Renders and Meadows[7] have carried out such an analysis involving rates of solid waste generation, air and water pollution, the depletion of natural resources and the volume of products in use as affected by different rates of recycling. According to one of their conclusions a 50 percent tax on virgin materials and a 50 percent subsidy for recycled materials would be a "nearly successful" policy because the standard of living would decline less rapidly and the depletion of resources would be retarded compared to an economy without recycling, although pollution would still rise.

A sweeping projection for the future envisages a steady state in which all materials are recycled. If this were combined with a constant demand level, primary production would become unnecessary. It must be recognized, however, that total recycling is not physically possible and in general not optimal from the standpoint of the conservation of all resources. The physical limitations result from dissipative materials

uses, dilution in alloys, dispersion of materials in small parts or by geographical dispersion, and losses due to corrosion and shipwreck. While these effects cannot be completely eliminated, they can be reduced and the rate of recycling can be increased.

A steady state can be conceived in which only a small fraction of material is lost and can be made up by a small volume of primary production while most of the demand is satisfied by production from secondary sources. Such a model has been proposed by Goeller.[37] Brooks has pointed out that combined with a continuously growing level of material use even total recycling would eventually run up against the limits of available materials required for this growing perpetual inventory.[38]

Acknowledgments

The writer is greatly indebted to Dr. Robert W. Crandall, Associate Professor of Economics, Massachusetts Institute of Technology, for many clarifying and stimulating discussions. Support by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 is gratefully acknowledged.

References

1. National Materials Advisory Board, Elements of a National Materials Policy, Publication NMAB 294, Washington, D.C., August, 1972.
2. M. B. Bever, in: Panel Discussion on a National Policy on Resources, National Commission on Materials Policy, University Forum, Massachusetts Institute of Technology, May 30-June 2, 1972.
3. M. B. Bever, Recycling and the Supply of Materials, in: National Commission on Materials Policy, University Forum, Massachusetts Institute of Technology, May 30-June 2, 1972.
4. Battelle Memorial Institute, A Study to Identify Opportunities for Increased Solid Waste Utilization, Vols. 1-9, Columbus, Ohio, 1972.
5. U.S. Department of Commerce, Industrial Scrap Generation: Iron, Copper, Aluminum, Washington, D.C., 1957.
6. H. H. Landsberg, L. L. Fischman and J. L. Fisher, Resources in America's Future, Resources for the Future, 1963.
7. J. Randers and D. L. Meadows, The Dynamics of Solid Waste, Technology Review, March-April, 1972, 20.
8. G. W. Perbix, Perspectives in Recycling, Remarks Before Annual Meeting, Bureau International de la Récupération, June, 1972.
9. W. O. Spofford, Jr., Solid Residuals Management: Some Economic Considerations, National Resources Journal, Vol. 11, 1971, 561.
10. National Association of Secondary Materials Industries, Effective Technology for Recycling Metal, New York, 1971.

11. E. Scheuer, The Sorting of Scrap Metals and Alloys, Metallurgical Reviews, Vol. 1, 1956, 339.

12. U.S. Bureau of Mines, Bureau of Mines Research Programs on Recycling and Disposal of Mineral-Metal and Energy-Based Solid Wastes, Information Circular 8529, Washington, D.C., 1971.

13. Mineral Waste Utilization, Proceedings of the First, Second and Third Mineral Waste Utilization Symposium, 1968, 1970 and 1972 respectively, U.S. Bureau of Mines and IIT Research Institute.

14. D. L. Siebert, Impact of Technology on the Commercial Secondary Aluminum Industry, U.S. Bureau of Mines, IC 8445, Washington, D.C., 1970.

15. U.S. Bureau of Mines, Mineral Facts and Problems, 1970, Bureau of Mines Bulletin 650, Washington, D.C., 1970.

16. National Association of Secondary Material Industries, Industrial Profile and Cost Factors in Nonferrous Scrap Metal Processing, New York, 1969.

17. Copper Development Association, Annual Data 1970, New York, 1970.

18. A. D. McMahon, Copper, A Materials Survey, U.S. Bureau of Mines, IC 8225, Washington, D.C.

19. S. Brubaker, Trends in the World Aluminum Industry, Resources for the Future, 1967.

20. National Association of Secondary Material Industries, A Study of the Secondary Lead Market in the United States, New York, 1969.

21. J. J. Gray and P. McIlroy, A Survey of the Secondary Titanium Market, U.S. Bureau of Mines, IC 8532, Washington, D.C., 1971.

22. M. J. Spendlove, A Profile of the Nonferrous Secondary Metals Industry, in: Proceedings of the Second Mineral Waste Utilization Symposium, 1970, Page 7 (See Ref. 13.)

23. A. V. Kneese, R. U. Ayres and R. C. d'Arge, Economics and the Environment, Resources for the Future, 1970.

24. A. V. Kneese, Background for the Economic Analysis of Environmental Pollution, Swedish J. of Economics, Vol. 73, 1971, 1.

25. C. S. Russell, Models for Investigation of Industrial Response to Residuals Management Action, Swedish J. of Economics, Vol. 73, 1971, 134.

26. National Association of Secondary Material Industries, Studies of Dislocation Factors: II. The Secondary Material Industries and Environmental Problems, New York, 1968.

27. U.S. Department of Commerce, Economic Impact of Air Pollution Controls on the Secondary Nonferrous Metals Industry, Washington, D.C., 1969.

28. D. J. Rose, J. H. Gibbons and W. Fulkerson, Physics Looks at Waste Management, Physics Today, Feb., 1972, 32.

29. Ref. 24, Page 19.

30. R. S. Berry, Recycling, Thermodynamics and Environmental Thrift, Bulletin Atomic Scientists, May, 1972, 8.

31. J. C. Bravard and C. Portal, with Revisions by P. H. Wadia and J. T. Day: Energy Expenditures Associated with the Production and Recycle of Metals, Oak Ridge National Laboratory, Report ORNL-MIT-132, May 26, 1971.

32. First Annual Report of the Secretary of the Interior under the Mining and Minerals Policy Act of 1970, March, 1972.

33. National Commission on Materials Policy, Towards a National Materials Policy, Interim Report, Washington, D.C., April, 1972.

34. J. F. Elliott, Reclaimed Scrap and Solid Metallics for Steelmaking, Distributed at M.I.T. Industrial Liaison Symposium, Dec. 1, 1971.

35. Ref. 4, Table B-1.

36. J. W. Forrester, Industrial Dynamics, MIT Press, 1961.

37. H. E. Goeller, Oak Ridge National Laboratory.

38. Harvey Brooks, Materials in a Steady State World, Met. Trans., Vol. 3, 1972, 759.

ON THE MORPHOLOGY OF POLYMERIC ALLOYS

M. B. Bever and M. Shen

Abstract

Many unique polymeric materials can be prepared by combining two or more homopolymers or random copolymers. These multicomponent polymer systems, or polymeric alloys, are available as block copolymers, graft copolymers, or polyblends. Most of these polymeric alloys show microphase separation; the resulting morphology has been found to exert considerable influence on the properties of the material. A systematic discussion of these morphological features is given in this memorandum and possible applications of stereology to these materials are explored.

ON THE MORPHOLOGY OF POLYMERIC ALLOYS

M. B. Bever and M. Shen

1. Introduction

Polymeric alloys - or multicomponent polymers - are gaining in scientific interest and technical importance. For several decades polymer research has concentrated on the development of new monomers. As a result many useful polymeric materials can now be made from monomers. Extensive work on multicomponent polymers, however, has begun only recently.

The properties of polymeric alloys are greatly affected by their morphology. Recently discovered experimental techniques have made it possible to study the morphology of polymeric alloys in new ways. In particular, staining techniques are now being used for the examination of these alloys by electron microscopy.[1]

The interpretation of three-dimensional structures from two-dimensional representations has made great progress. In particular, quantitative stereology provides methods for precise morphological investigations.[A,B,C] These methods have been applied extensively to biological materials and metals but they have been applied to polymeric materials only in isolated instances.

This memorandum will systematically consider the morphological fractures of polymeric alloys and will explore possible

applications of stereological methods to such materials.

2. Polymers - General

2.1 Classification

Table I shows the major classes of polymeric materials. Homopolymers, which consist of identical repeating units or identical monomers, can be amorphous or crystalline. Bulk crystalline polymers usually contain some amorphous material, but it is possible, at least in principle, to reduce this material to the vanishing point, for instance by crystallizing from dilute solutions.[2]

Copolymers consist of two or more types of monomers. There are four classes of copolymers. (i) In random copolymers the comonomers are arranged in irregular sequences. (ii) In alternating copolymers, the comonomers occur in a regular way. (iii) Block copolymers are homopolymers in which the respective comonomers are chemically connected at the ends; they can form in different sequences: diblock, triblock, and segmented block polymers. (iv) In graft copolymers additional chains are connected onto random sites along the host polymer. Random and alternating copolymers are not generally considered as multicomponent polymers or polymeric alloys because their components are combined on the monomeric level. Block and graft copolymers are considered as polymeric alloys because their components exist on the polymeric level.

Another class of polymeric alloys are the polyblends. They are physical mixtures, rather than chemical combinations,

of two or more polymers. The component polymers may be homopolymers or copolymers.

2.2 Formation of Polymers

The monomers which serve as starting materials for polymers are usually organic molecules containing chemically reactive (or "functional") groups. Commonly used organic monomers are styrene, ethylene, methylmethacrylate and butadiene.[3,4] Organometallic and inorganic monomers are also gaining in importance.[5-7] An example of organometallic monomers is dimethyl siloxane; examples of inorganic monomers forming polymers are sulfur, selenium and sodium phosphate.

Polymeric alloys can be homogeneous or heterogeneous (single-phase or heterophase; "compatible" or incompatible"). Because of the large molecular weight of polymers there are relatively few macromolecules in a given volume. Hence the entropy of mixing of two polymeric species is negligible, and the free energy of mixing is determined by the enthalpy of mixing. Because this enthalpy is often positive, polymeric molecules are generally immiscible and most polymeric alloys are heterogeneous rather than homogeneous.[8,9]

Homopolymers and random and alternating copolymers are generally homogeneous. Most block copolymers, on the other hand, are immiscible or incompatible and the same is true for graft copolymers and for polyblends.

Polyblends are similar in structure and properties to incompatible block and graft copolymers but they tend to have

coarser morphologies. Polyblends are prepared either by mechanically mixing (for example, milling) the two component polymers or by dissolution of the components in a common solvent and subsequent evaporation of the latter.

3. Morphology

3.1 Morphological Phenomena

The morphology of polymeric alloys results from the spatial arrangement of assemblies of macromolecules on a supramolecular scale. The complexity of these structures increases rapidly in going from the amorphous homogeneous polymers to the various types of heterogeneous polymers. An overview of the basic morphological features of polymeric alloys is given in Table II.

Various morphological features can occur in homopolymers. In amorphous homopolymers there is some evidence for the occurrence of a nodular structure.[10,11] In crystalline homcpolymers, the size and orientation of spherulites are important.[12] The details of these features will be discussed in Section 4.

The characteristic features of heterogeneous polymeric alloys involve the mutual relation of their constituents. In two-phase alloys, both phases may be continuous, one may be continuous and the other discontinuous, or both phases may be discontinuous. In all three cases the particles of each phase have size, shape and orientation as their principal characteristics. Typical structures are spheres or rods distributed in a matrix, or structures consisting of alternate layers of two phases. The

details of these structures will be discussed in Section 5.

The scale of the morphological features of polymeric alloys extends over a wide range. For example, the nodular structure of homogeneous amorphous polymers is of the order of 100 A.U. The size of the spherulites in crystalline polymers ranges from a few hundred A.U. to millimeters. In heterophase polymers the domain structure is generally of the order of several hundred to 1000 A.U.

3.2 Methods of Observation

The most important methods of morphological investigation of polymers are electron microscopy (EM) and small-angle X-ray scattering (SAXS). Light scattering, electron diffraction and optical microscopy, including phase contrast microscopy, have also proved useful. Electron microscopy and small-angle scattering owe their importance, at least in part, to their high resolving power, which is usually required by the relatively fine scale of the morphological features of polymers.

The application of transmission electron microscopy to polymers had to wait for a staining technique which was discovered only a few years ago.[1] Transmission electron microscopy and small-angle X-ray scattering, which is an essentially indirect method, now complement each other in the investigation of polymeric alloys.

The replica method of electron microscopy has also been extensively used. For some applications, particularly the investigation of fracture surfaces, scanning electron microscopy has been useful. Optical microscopy with polarized light is

especially suitable for the investigation of crystalline polymers. Phase contrast microscopy has limited application, partly because of its relatively coarse resolution.

The scale of polymer morphology is much finer than the scale of many of the morphological features of metals. This explains in part a difference in the history of the morphological studies of metals and polymers. In the early stages of physical metallurgy light microscopy (metallography) played a dominant role. Only after subsequent developments of X-ray diffraction and electron microscopy could the atomic arrangements and fine-scale microstructure be investigated. In contrast to this, polymeric structures were first investigated on the molecular scale. Their investigation on the supramolecular scale by the techniques of light scattering and small-angle X-ray and electron scattering occurred next and was followed by electron microscopy.

3.3 Methods of Interpretation

Stereological methods are available for the interpretation of the morphology of polymeric alloys.[A,B,C] In the case of structures composed of only one kind of constituent, the size, size distribution and shape of the units of this constituent can be measured and any orientational features can be characterized. In the case of structures consisting of more than one kind of constituent, the proportions and arrangement of the units of the constituents can be determined and their sizes, shapes and orientational features can be characterized. (Details on these methods will be included in the final version of this memorandum.)

4. The Morphology of Homogeneous Polymeric Alloys

4.1 General

Multicomponent systems may be homogeneous or heterogeneous. Homogeneous phases - solid solutions ("mixed crystals") and compounds - occur in many metallic and other inorganic systems. In polymeric systems miscible pairs are relatively few, as already discussed in Subsection 2.1. Nevertheless, a number of important multicomponent polymeric systems are homogeneous or "compatible". Homogeneous multicomponent polymer phases may be amorphous or crystalline (See Table II).

4.2 Crystalline Homogeneous Polymeric Alloys

Crystalline homopolymers can exist as single crystals[2] or as polycrystalline aggregates.[12] Polymer single crystals form thin platelets with lateral dimensions of the order of one micron and a thickness of approximately 100 A.U. The crystallites in polycrystalline polymers - or spherulites - grow from the melt as spheres but ultimately attain polyhedral shapes because of mutual interference by impingement. In this regard they closely resemble polycrystalline metals and ceramics. Some amorphous material is always present in polycrystalline polymers. They are therefore more correctly referred to as semi-crystalline.

Homogeneous crystalline polymeric alloys can be formed in the system selenium-tellurium.[12a] These alloys may be considered as inorganic high polymers. Homogeneous crystalline alloys have not been observed in organic polymeric systems. Whether homogeneous crystalline phases of multicomponent polymeric systems

can exist in organic polymers deserves both theoretical and experimental investigation.

4.3 Amorphous Homogeneous Polymeric Alloys

For many years amorphous homogeneous polymers were believed not to have morphological features. Some evidence, however, has recently been reported for the existence of supramolecular structures in such systems.[10,11] This evidence consists of density variations observed by transmission electron microscopy. The denser regions, which have been referred to as "nodules", have been attributed to localized imperfect chain folding. (See Fig. 1).

5. The Morphology of Heterophase Polymeric Alloys

5.1 General

The morphology of heterophase systems depends on the arrangement of the constituent phases. Three possibilities exist in a two-phase system: (i) both phases are discontinuous, (ii) one phase is continuous and the other is discontinuous, and (iii) both phases are continuous. These three cases were considered for polymeric systems by Molau[13] and Matsuo and Sagaye[14]; these authors published schematic sketches, an example of which is reproduced in Fig. 2.

It is of interest that the continuous phase in a matrix-dispersoid system need not necessarily be present in the greater amount. A small amount of one phase can form continuous envelopes (a "network") surrounding large particles of the other phase, which thus becomes discontinuous.

The two phases of a heterogeneous polymeric system may in

principle be both amorphous or one may be amorphous and the other crystalline. In the case of one continuous and one discontinuous phase, the crystalline phase is more likely to be the discontinuous one.

5.2 Structures Consisting of Two Discontinuous Phases

The most general case of two discontinuous phases is represented by an aggregate of crystallites of two types. This case is encountered in inorganic systems, for example, in the alpha plus beta region of the copper-zinc system: the grains of the two phases exist in random aggregation without any tendency for clustering by grains of either phase. As a result, the connectivity of each of these phases is that resulting from a random arrangement and depends primarily on the volume fractions of the two phases.

The analogous case in polymers would require that spherulites of a crystalline phase A and spherulties of a crystalline phase B coexist in an aggregate without clustering of either type. Such a structure has never been observed. In principle it seems possible, however, to synthesize a block copolymer consisting of two crystallizable polymeric phases or blocks, each of which may form its own crystallites.

Another possible arrangement of two discontinuous phases is a lamellar structure. Such structures have been observed in triblock copolymers of styrene-isoprene-styrene[15,16] and styrene-butadiene-styrene.[14,17-19] A typical example is shown in Fig. 3. It should be noted that the micrograph in this figure was produced

by electron transmission microscopy after staining of the butadiene phase with osmium tetroxide. The thickness of each lamella is of the order of 500 A.U. This is smaller by a factor of fifty than typcial interlamellar distances in metallic eutectics and eutectoids. In polymers the thickness of the lamellae depends on the molecular weight of the constituent polymeric blocks: the higher the molecular weight the thicker the lamellae.[15,19]

The apparent spacings in a section depend not only on the true interlamellar spacing but also on the angle at which the lamellae are cut. In general the smallest apparent spacing is most nearly equal to the approximate true spacing if this is approximately constant.

The lamellar structures in polymeric alloys differ from those in metallic alloys in another respect: the lamellae in polymeric alloys are almost always curved. The thickness of the lamellae, however, appears to be fairly uniform in any given specimen.

In the interpretation of lamellar structures in polymers the nature of transmission electron microscopy must be kept in mind. In this type of microscopy features throughout the thickness of the specimen are projected into the plane of observation. Transmission electron micrographs of polymeric specimens, however, appear to approach a two-dimensional section rather than a projection over a volume. The probable explanation for this is as follows. The significant structure is generally revealed by staining with osmium tetroxide vapor and the staining reaction is

limited to the surface layer. This argument assumes that the staining occurs on only one side of the specimen. Otherwise, the electron transmission micrograph would reproduce both the front and rear surface layers; unless the section is very thin, this would interfere with the formation of a useful representation. In this connection it is important that the specimen thickness is very small because the ultramicrotome sections are only of the order of 350 to 1000 A.U. thick.

5.3 Structures Consisting of One Continuous and One Discontinuous Phase

In a typical structure consisting of one continuous and one discontinuous phase the continuous phase is present in excess and serves as matrix while the discontinuous phase is present as a dispersoid. The particles of the dispersed phase may differ in size, number, shape, and, if they are anisotropic, also in orientation. The size and number of the dispersed particles are not independent of each other and of the volume fraction. The particles may be polydisperse (they may have size distributions) or they may be monodisperse. Typical shapes of the particles are spheres, rods, or plates.

In polymeric alloys a crystalline dispersed phase is present in the form of spherulites.[20] An example is shown in Fig. 4. Each spherulite is an aggregate of crystallites. The constituent crystallites grow radially and finally impinge upon each other. They are approximately spherical. Since a spherulite is thus polycrystalline rather than a single crystal, the close

analogy between a spherulite and a metal grain which has been postulated in the literature is overstated. In crystalline polymers amorphous material is present at the boundaries of the crystallites.

In most two-phase polymeric alloys both the matrix and the dispersed phase are amorphous. Examples are shown in Figs. 5 and 6. Fig. 5 shows a phase contrast photomicrograph of a polyblend of polybutadiene and polystyrene.[21] In this structure, spheres of polybutadiene are imbedded in a matrix of polystyrene. The size of the spherical particles is of the order of one micron with an appreciable spread. This relatively large size of the dispersed particles is characteristic of polyblends.

Fig. 6 is an electron photomicrograph of a styrene-butadiene-styrene block polymer.[14] It illustrates a structure in which the particles of the dispersed phase have the shape of rods. Fig. 6a is oriented normal and Fig. 6b parallel to the long dimension of the rods. It should be noted that in limited regions the rods have the same orientation. The diameters of the rods are of the order of a few hundred A.U.; their length is of the order of microns. Such fine-scale structures are usually found in block and graft copolymers.

Instances in which the dispersed phase particles have the shape of platelets do not appear to have been reported in the literature.

Matrix-dispersoid systems in which the particles of the dispersed phase have the shape of spheres or rods illustrate

several stereological relations. In two-dimensional sections monodispersed spherical particles should produce a distribution of circular discs having different radii since the spheres are in general not cut in equivalent planes. Some published electron micrographs show such a distribution. In other micrographs the discs are of the same size. Discs of equal size cannot be generated by cutting monodispersed spheres unless the spheres are arranged in a regular manner on a lattice and the cut is parallel to a plane of this lattice. These two conditions are unlikely to be fulfilled, and discs of equal size cannot generally be explained in this manner. The only possible explanation is that the discs are sections through rods. The question then arises why the rods are cut in circular sections rather than in ellipsoidal sections. The answer appears to be that the electron micrographs lack the necessary precision.

The volume fractions of the phases, the distribution of polydisperse particle sizes and other quantitative morphological features of polymeric alloys can be found by well-known stereological methods.

Under certain conditions of formation the dispersed particles in polymeric systems are arranged in regular periodic patterns. In two-dimensional sections the traces of the particles appear to be arranged in regular hexagonal or square arrays. These structures have been referred to as macrolattices.[22-28] An example is shown in Fig. 7.

The macrolattices in polymeric alloys resemble periodic structures which have been found in some metallic eutectics. The periodic spacing in polymeric materials, however, is very much smaller; it is approximately one fiftieth of that in metal eutectics. [To be checked. Significance of factor of 1/50?] Both rod-shaped and spherical particles can form a macrolattice.

A periodicity of the arrangement of dispersoid particles in some polymeric specimens has also been found by small-angle X-ray scattering.[23,26-28] McIntire determined in triblock copolymers of styrene-butadiene-styrene an orthorombic unit cell with cell dimensions of 676, 676, 566 A.U. In a recent investigation Kim[28] analyzed experimental data for a similar system by Porod's two-phase theory and obtained essentially identical results. Specifically he found that styrene spheres with a radius of 132 A.U. formed a macrolattice which was hexagonal. Part of the difference between McIntire's and Kim's results may be attributed to the difference in molecular weights of the styrene blocks which in Kim's work was approximately half that in Mcintire's.

In matrix-dispersoid systems the dispersed particles may show orientation effects, in particular, if they have anisotropic shapes which may tend to be aligned in certain directions. An example of the alignment of rods in a polymeric material is a styrene-butadiene-styrene block copolymer prepared by extrusion.[14,24] (Fig. 6)

The macrolattice implies that a region has a high degree of periodicity and alignment. Neighboring regions may have a

macrolattice with different orientations. The boundaries between such macrolattice regions are analogous to grain boundaries in metals, as was pointed out first by Hoffmann.[25] (See Fig. 9) The absolute size of metal grains, however, is always much greater than that of the macrolattice domains of polymers. A similar argument can be made for regions with different alignments of dispersed particles.

In a structure consisting of one continuous and one discontinuous phase the continuous phase may be present in only a small amount but may envelope the particles of the discontinuous phase. This results in a three-dimensional network of the continuous phase. Structures approximating such networks are found in certain polyblends. For example, if a specimen consisting of over 90 percent plastic (polystyrene) and a few percent of rubber (polybutadiene) is prepared while being stirred, the rubber phase will envelope the particles of the plastic phase.[29] (Fig. 8) The network constituent in this case has often been referred to as the "rubber phase volume".

5.4 Structures Consisting of Two Continuous Phases

Two phases can form a structure in which both are continuous. This requires that they form lath-like interpenetrating networks. Some evidence of the existence of such structures has been reported for metallic alloys and also for glasses. (Ref.: J. W. Cahn)

Two independent networks can also occur in polymers. To prepare a specimen with such a structure one must allow monomers

to diffuse into a pre-existing homogeneous polymer network, which at the outset is itself continuous. If the added monomer is then polymerized, it forms a second continuous network. Since this polymer exists on a very fine scale, its morphological structure is not clearly resolvable even by electron microscopy. Evidence derived from mechanical property measurements, however, has confirmed the existence of a two-phase structure in such a case.[30,31]

5.5 Interface Layer

The existence of an interface layer between adjoining regions (or domains) of different composition is at present controversial. The formation of such a layer requires diffusion. Since its volume would be small, the layer would be very difficult to detect by electron microscopy.

It is noteworthy that Kim's analysis of small-angle X-ray diffraction data achieved satisfactory results without the assumption of an interfacial layer.[28] The interpretation involving the existence of an interfacial layer, on the other hand, gave satisfactory results in certain other cases.[23] In metallic systems two coexisting phases are known to be able to have sharp interphase interfaces. This is especially true of structures formed by eutectic solidification.

6. Effects of Variables on Morphology

6.1 General

Several variables involving the material for the process of preparation can have important effects on the morphology of polymeric alloys. These variables are the composition and

molecular weight of the polymers, the nature of the solvent, the rate of evaporation, and the addition of extra homopolymers in the preparation of the alloys.

6.2 Composition

The morphology of two-phase polymeric systems, for example, the block copolymer of styrene-butadiene-styrene, depends markedly on the proportion of the two phases. At high concentrations of styrene this phase is continuous and the isoprene is discontinuous; with increasing isoprene both phases become essentially discontinuous; at high isoprene concentrations the continuous phase is isoprene and the dispersed phase styrene.[15] An example of this inversion is shown in Fig. 10.

6.3 Method of Preparation

The manner in which a polymeric alloy is prepared can have a decisive effect on its morphology.[14,15] For example, when an alloy of styrene and isoprene is prepared by "casting" from methyl-ethyl ketone (MEK), the continuous phase is isoprene, as shown in Fig. 11a. When an alloy of the same composition is prepared by casting from carbon tetrachloride both phases develop a high degree of connectivity and each appears nearly continuous, as shown in Fig. 11b.

In the preparation of polymeric alloys by casting, the solvent is usually evaporated slowly. Fast evaporation tends to smear out the morphological features. By contrast the initial concentration of the polymers in the solvent appears to have little, if any, effect on the morphology.

In the preparation of polyblends, specifically the oil-in-oil emulsion of polybutadiene in styrene monomers, the use of stirring and the stirring rate exert important effects on the resulting morphology.[21,29,32] Without stirring or with very low rates of stirring the rubber phase tends to exist as discontinuous particles, while higher rates lead to a larger rubber phase volume which tends to envelop the polystyrene phase. (See Subsection 5.3)

Polymeric alloys can be prepared from solid starting materials, for example, pellets of a block copolymer of styrene-butadiene-styrene, by the application of heat and pressure (compression molding).[14] The resulting morphology is believed to approach the equilibrium structure predicted by statistical mechanics.[23] This structure may be different from that of the same polymeric alloy prepared by solvent casting.

6.4 Molecular Weight

The predominant effect of molecular weight on the morphology of polymeric alloys is exerted on the size of the constituent units. With increasing molecular weight the size of these units tends to increase. Such an increase has been observed for the interlamellar spacing in lamellar structures. (See Subsection 5.2; Fig. 3) In this case the thickness is of the order of the root mean square end-to-end distance of the polymeric chains.[15,19]

Similarly, the diameter of spherical domains has been found to increase with increasing molecular weight of the dispersed component.[15,28,34,25] Such an increase occurs, for example, in the block copolymer of butadiene-styr ne-butadiene. It has

been observed by transmission electron microscopy[15] and small-angle X-ray scattering[35] (Fig. 12). The polymer of higher molecular weight has a sharper X-ray diffraction peak, indicating that any interfacial region between the two polymers is sharper than when the molecular weights are lower. A corresponding increase in size may be expected for the cross sections of rods representing the dispersed phase.

6.5 Polyblends of Homopolymers and Block Copolymers

Blending of homopolymers with block copolymers can produce a special type of polymeric alloy. For example, homopolymers of styrene can be blended into a block copolymer of styrene-isoprene-styrene.[36] With increasing amounts of polystyrene the domain size of this phase increases; at high concentrations the structure becomes increasingly irregular as shown by the small angle-diffraction investigation of McIntire and Campos-Lopez.[35] Block copolymers can also be blended with homopolymers of both constituent polymers. For example, if the ratio of butadiene to styrene is maintained constant, the morphology tends to remain the same except that the size of the domain increases.

Acknowledgement

This study was supported by the Advanced Research Projects Agency under Contract No. DAHC15-71-C-0253 with the University of Michigan.

TABLE I.

Classification of Polymers

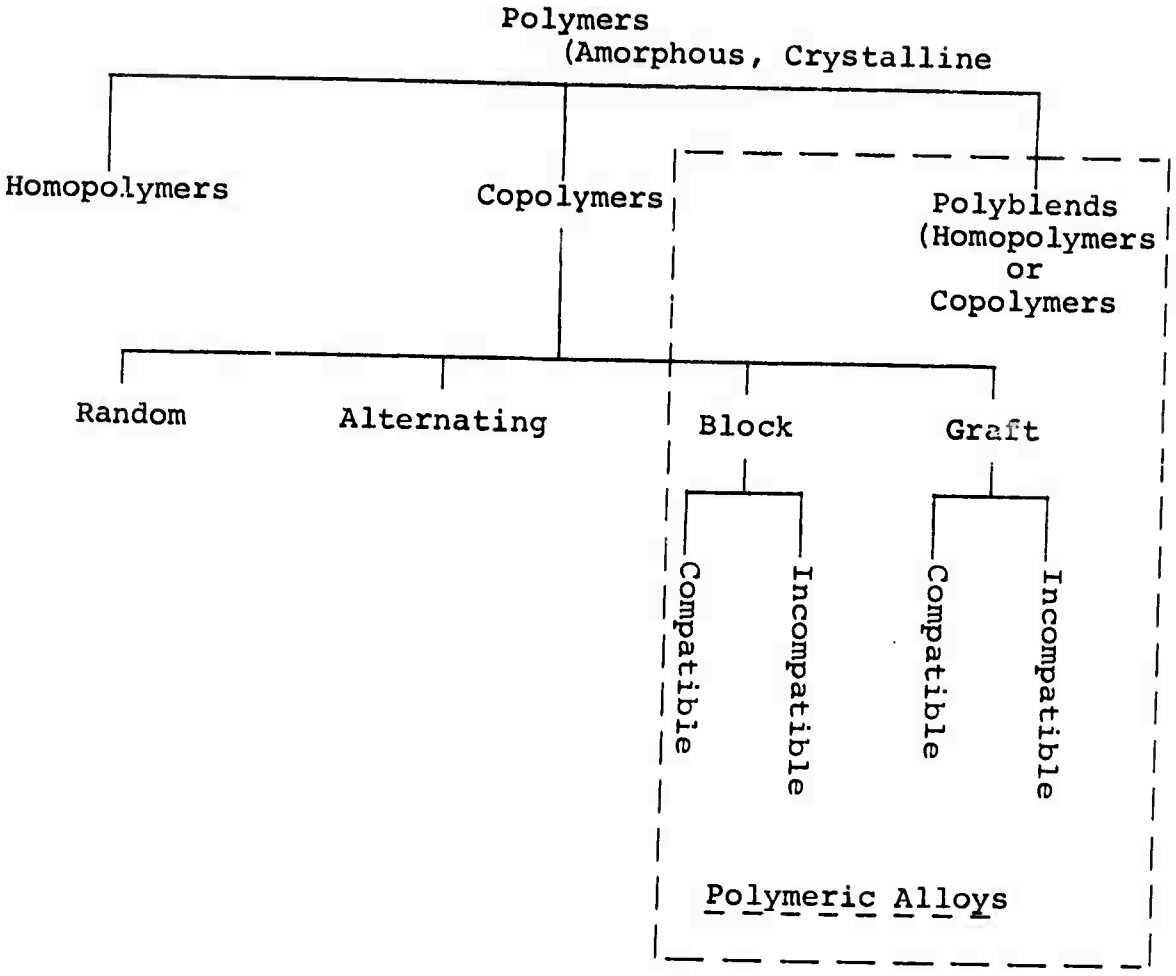


TABLE II.

Morphological Structures of Polymeric Alloys

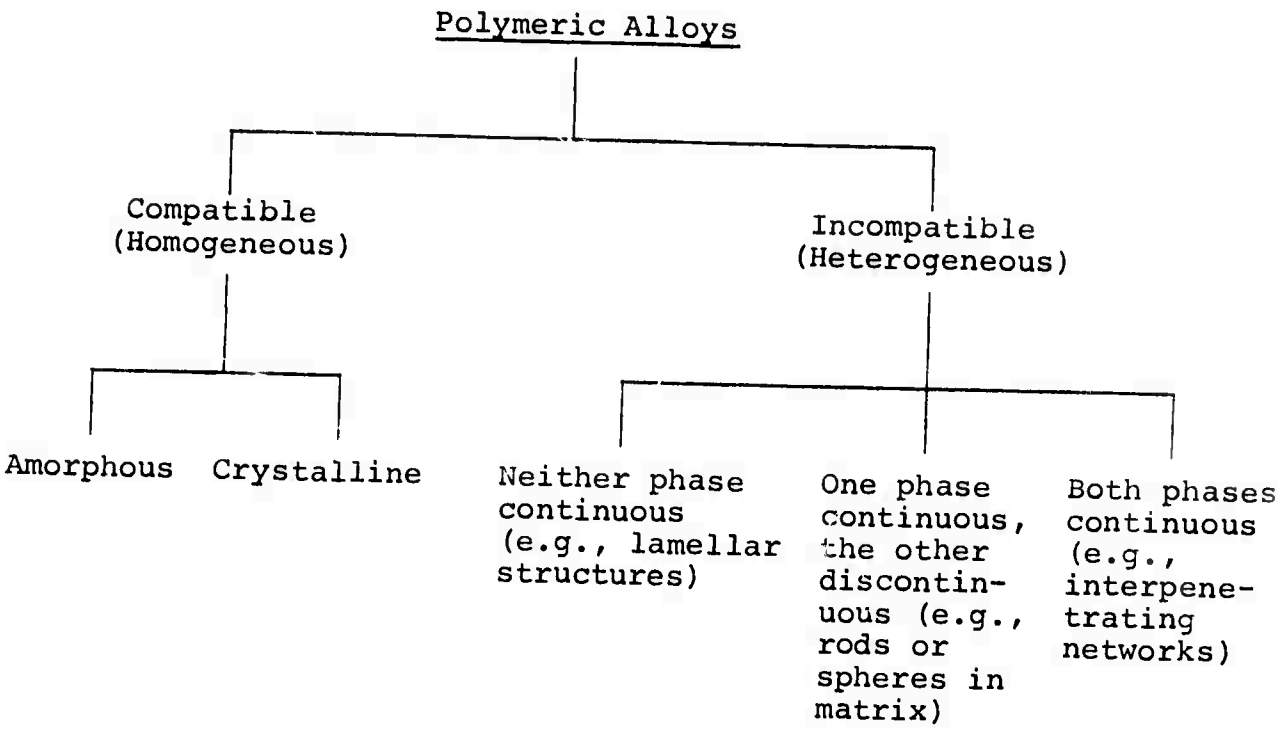


FIGURE CAPTIONS

Figure 1: Electron photomicrograph of unshadowed thin film of atactic polystyrene. Fig. 1 of Ref. 10.

Figure 2: Schematic illustration of various phase structures composed of plastic (P) and rubber (R) components. Rubber content increases from the left to the right. Fig. 14 of Ref. 13.

Figure 3: Electron photomicrograph of ultrathin sections of a styrene-butadiene-styrene block copolymer (40% styrene) cast from cyclohexane solution. Sections were cut normal (left) and parallel (right) to the film surface. Fig. 11 of Ref. 14.

Figure 4: Polarized optical photomicrograph of a block copolymer of styrene and ethylene oxide (19.6% styrene) cast from chloroform solution. Fig. 2 of Ref. 20.

Figure 5: Electron photomicrograph of ultrathin sections of a styrene-butadiene-styrene block copolymer (80% styrene) prepared by compression molding. Fig. 8 of Ref. 21.

Figure 6: Electron photomicrograph of ultrathin sections of a styrene-butadiene-styrene block copolymer (60% styrene) cast from cyclohexane solution. Sections were cut normal (left) and parallel (right) to the film surface. Fig. 10 of Ref. 14.

Figure 7: (a) Electron photomicrograph of an ultrathin section of a styrene-butadiene block copolymer (68% styrene)

cast from xylene solution (annealed at 110°C for one hour); (b) optical photomicrograph of an eutectic consisting of metal #1 fibers incorporated in metal #2 matrix. Fig. 15 of Ref. 25.

Figure 8: (a) Soap-bubble model of a mosaic crystal; (b) boundaries between regions of long-range order in a styrene-butadiene block copolymer with various orientations of the rod-shaped butadiene aggregates. Fig. 10 of Ref. 25.

Figure 9: Electron photomicrograph of a high impact polystyrene in which 6% of rubber (polybutadiene) was blended. The rubber phase volume in this polyblend is 78%. Fig. 2 of Ref. 21.

Figure 10: Electron photomicrographs of ultrathin sections of styrene-isoprene block copolymers varying in composition from 20 to 70% of styrene. Sections were cut normal to the surface of the film cast from toluene solution. Fig. 1 of Ref. 15.

Figure 11: Electron photomicrographs of ultrathin sections of a styrene-isoprene block copolymer (40% styrene) cast from (a) methyl ethyl ketone and (b) carbon tetrachloride. Sections were cut normal to the film surface. Fig. 2 of Ref. 15.

Figure 12: Low angle X-ray scattering from styrene-butadiene-styrene block copolymers having different compositions. Fig. 2 of Ref. 35.

References

1. K. Kato, Polymer Letters 4, 35 (1966).

2. P. H. Geil, *Polymer Single Crystals*, Interscience, New York 1963.

3. R. W. Lenz, *Organic Chemistry of Synthetic High Polymers*, Interscience, New York, 1967.

4. A. Ravve, *Organic Chemistry of Macromolecules*, Dekker, New York, 1967.

5. M. F. Lappert and G. J. Leich, eds., *Developments in Inorganic Polymer Chemistry*, Elsevier, Amsterdam, 1962.

6. F. G. A. Stone and W. A. G. Graham, eds., *Inorganic Polymers*, Academic Press, London, 1962.

7. A. V. Tobolsky and W. J. MacKnight, *Polymeric Sulfur and Related Polymers*, Interscience, New York, 1965.

8. L. Bohn, Kolloid Z. u. Z. f. Polymere 213, 55 (1966).

9. S. Krause, J. Macromol. Sci. C7, 251 (1972)

10. S. L. Lambert and G. S. Y. Yeh, Proc. Electron Microsc. Soc. Am., C. J. Arceneaux, Ed., Claitor's Publ., Baton Rouge, 1968, p. 414.

11. T. G. F. Schoon, Br. Polym. J. 2, 86 (1970).

12. Holland, in *Recrystallization, Grain Growth and Textures*, American Society for Metals, Metals Park, Ohio, 1966, p. 539.

12a. G. C. Das and M. B. Bever, *A Thermodynamic Investigation of Crystalline and Amorphous Selenium-Tellurium Alloys* (to be published).

13. G. E. Molau, in *Block Polymers*, S. Aggarwal, ed., Plenum, New York, 1970, p. 79.

14. M. Matsuo and S. Sagaye, in *Colloidal and Morphological Behavior of Block and Graft Copolymers*, G. E. Molau, ed., Plenum, New York, 1971, p. 1.

15. T. Inoue, T. Soen, T. Hashimoto and H. Kawai, J. Polymer Sci., Part A-2, 7, 1283 (1969).

16. T. Uchida, T. Soen, T. Inoue and H. Kawai, J. Polymer Sci., Part A-2, 10, 101 (1972).

17. E. Vanzo, J. Polymer Sci., Part A-1, 4, 1727 (1966).

18. H. Hendus, K. H. Illers and E. Ropte, Kolloid Z. u. Z. f. Polymere 216, 110 (1967).

19. E. B. Bradford and Vanzo, J. Polymer Sci., Part A-1, 6 1661 (1968).

20. R. G. Crystal, P. F. Erhardt and T. J. O'Malley, in *Block Polymers*, S. L. Aggarwal, ed., Plenum, New York, 1970.

21. H. Keskkula and P. A. Traylor, J. Appl. Polymer Sci. 11, 2361 (1967).

22. E. Fischer, J. Macromol. Sci.-Chem. A2, 1285 (1968).

23. D. McIntyre and E. Campos-Lopex, Macromol. 3, 322 (1970).

24. J. Dlugosz, A. Keller and E. Pedemonte, Kolloid Z. u. Z. f. Polymere 242, 1125 (1970).

25. G. Kampf, M. Hoffman and H. Kromer, Ber. Bunsenges. Physik. Chem. 74, 851, 859 (1970), J. Macromol. Sci.-Phys. B6, 167 (1972).

26. P. R. Lewis and C. Price, Polymer 12, 258 (1971); ibid., 13, 20 (1972).

27. P. S. Pillai, D. I. Livingston and J. D. Strang, Rubber Chem. Tech. 45, 241 (1972).

28. H. G. Kim, to be published.

29. E. R. Wagner Rubber Chem. Tech. 43, 1129 (1970).

30. L. H. Sperling and D. W. Friedman, J. Polymer Sci.. Part A-2, 7, 425 (1969).

31. D. Klempner, E. L. Frisch and K. C. Frisch, J. Polymer Sci. Part A-2, 8, 921 (1970).

32. G. E. Molau and H. Keskkula, J. Polymer Sci., Part A-1, 4, 1595 (1966).

33. D. J. Meier, J. Polymer Sci., C26, 81 (1969).

34. D. F. Leary and M. C. Williams, Polymer Letters 8, 335 (1970).

35. D. McIntyre and E. Campos-Lopez, in S. L. Aggarwal, op. cit., p. 19.

36. M. Moritani, T. Inoue, M. Motegi, H. Kawai and K. Kato, in G. E. Molau, op. cit., p. 33.

A. R. T. DeHoff and F. N. Rhines, *Qunatitative Microscopy*, 1968, McGraw-Hill.

B. E. E. Underwood, *Quantitative Stereology*, 1970, Addison-Wesley.

C. H. J. Rack and R. W. Newman, Chapter on Microstructures in: Cahn, R. W., ed., *Physical Metallurgy*, 2nd ed., American Elsevier, 1971, p. 705.

INCLUSION PATTERNS AND STRESS CRITERIA FOR QUASI-STATIC TO SPALL FRACTURES BY VOID COALESCENCE

D. C. Drucker

Abstract

A crude model is offered for the predictive calculation, as well as qualitative understanding, of the process of fracture through microscopic void coalescence in metallic alloys. Problems discussed include plane-strain crack initiation and dynamic propagation (K_{Ic} values), fracture of smooth and notched round tensile specimens, and spall caused by high velocity impact. The voids of diameter d and effective spacing D are assumed to be generated by the cracking of inclusions of diameter d and the failure of their bond to the ductile matrix. The effective initial D for the connected fracture surface is taken to be about 1/2 the average spacing in this highly simplified picture which is quasi-static on the microscale. The matrix is approximated as perfectly plastic with a yield stress σ_o appropriate for the level of strain rate and work hardening in each problem considered. A lower bound $2\sigma_o \ell n\ (D/d)$ and an upper bound $\frac{4\sigma_o}{3}\left[\ell n\frac{D}{d} + \frac{D}{3d} + \frac{1}{4}\right]$ for full lateral constraint, although not close, then provide a reasonable estimate of the void coalescence stress on the microscale for comparison with the macroscopic stress levels in cracks or necks and the time history of stress in spall.

INCLUSION PATTERNS AND STRESS CRITERIA FOR QUASI-STATIC TO SPALL FRACTURES BY VOID COALESCENCE

D. C. Drucker

Introduction to Model and General Approach

The model of material which is employed here, is based conceptually on a simplification of a spheroidized steel or an aluminum alloy. Spherical holes replace the large number of roughly equiaxed inclusions visible in the optical microscope. The assumption is that at the rather high values of stress of interest, these large inclusions will crack or lose their bond to the matrix at an early enough state of the plastic deformation. The ductile matrix which contains inclusions and defects on a submicron and atomic scale is replaced by an infinitely ductile material whose yield strength is in the range of structural metals and whose work-hardening rate in the plastic range is moderate. For clarity of description and results it will be idealized further as perfectly plastic with a yield stress σ_o appropriately adjusted for the relevant level of plastic strain (several percent), strain rate, and temperature.

Quasi-static growth or dynamic propagation of a crack in plane strain illustrates well the basic points to be made. If there are no voids present at the start and none of significance are generated, large plastic deformation around the crack tip

extends an appreciable distance above and below the plane of separation, and would be visible on the macroscale. Local tractions on the plane in the vicinity of the crack tip must reach and then are restricted to approximately $(1 + \frac{\pi}{2})\sigma_o$ on average.

If, on the contrary, the plane of separation contains many voids on the microscale whose spacing is very close, the stress level in the plastic links between the voids is restricted to σ_o and large plastic deformation is confined to the links except for the several percent of plastic strain needed in actuality to reach the yield strength of the idealized materials, no bulk plastic deformation occurs. All flow takes place in an extremely thin layer (a fraction of the hole diameter) bordering the surface of separation. The fracture would occur at small nominal stress σ_N and appear completely brittle on the macroscale.

As the picture is changed by increasing the distance D between the holes, the joining or void coalescence clearly involves a thicker layer of material and a higher local stress. However, the layer of large plastic deformation remains of order d until the spacing D is large enough to require a high enough stress to involve at least the beginning of macroscopic plastic deformation, about $(1 + \frac{\pi}{2})\sigma_o$. Lateral void growth and decrease in lateral spacing of the voids which accompany the plastic deformation finally lead to an effective D/d at which void coalescence takes over and the macroscopic flow ceases.

The transition from a high value of K_{Ic} to a very low one as D/d decreases requires the more detailed look at void patterns and void coalescence which follows in subsequent sections but the general results given by the model are clear. Moderately rapid loading or dynamic crack propagation alters the picture somewhat from the static. Strain-rate effects, to a first approximation, simply raise the value of σ_o and so would have little effect on the localization of plastic deformation, which is the key aspect of the model and the prototype as well. Inertia effects do enter and should raise the limit value of $(1 + \frac{\pi}{2})\sigma_o$ for macroscopic plastic deformation while hardly affecting the stress required to form the microscopic plastic links between the holes. It is in this sense that void coalescence on the microscale is quasi-static. Dynamic crack propagation therefore favors void coalescence over macroscopic flow and should lead to a transition to lower K_{Ic} values at somewhat larger values of D/d than for quasi-static crack growth.

A quite different effect appears for extremely high and short-time loading as in the problem of spall in macroscopically uniaxial strain produced by impact on flat plates. Maximum stresses are very high but may not produce separation because void coalescence requires sufficient stress at all stages of the process. It is not the maximum stress but a time integral of the stress level which determines whether or not full separation or even noticeable void growth will occur.

The necking of a smooth tension bar involves still another regime which also enters into void coalescence at the root of cracks following large local plastic strains in a limited region ahead of the crack. In necking, as at a plastic strain concentration, large macroscopic strains draw out the matrix of the material and effectively concentrate the voids. If, as is likely, the inclusions have cracked or decohered prior to large plastic strain, the neck will start at a "weak" cross-section already populated far more by equivalent voids than the average for the material. The added concentration caused by the lateral contraction accompanying large extensional strains further increases the effective void ratio (decreases effective D/d). It is easy to picture a material with average D/d = 5 having an effective initial D/d = 2.5 for the self-selected fracture surface which is reduced below 2 by a plastic strain in the neck of the order of 20%. At so low a value of effective D/d, void coalesence can occur over the cross-section of the tensile specimen despite the rather moderate triaxiality of tension associated with the geometry of the neck.

In the region of strain concentration at the root of a crack, triaxiality is large and significant void growth accompanies macroscopic (although small scale) plastic straining. Effective D/d values in the neighborhood of 2 can be visualized as emerging early even when starting from a rather clean material, initial average D/d of 10. Large strains, say 50% or more, would

be needed to reach void coalescence in the neck of a tension specimen of so clean a material. Very pure ductile material necks down to a point or a line without the void formation and coalescence which gives the familiar cup and cone fracture for almost all structural metals. The addition of hydrostatic pressure, sufficient to inhibit or even prevent void initiation and growth, permits ordinary structural metals to behave like pure materials.

Quantitative predictions for static and dynamic fracture initiation and propagation require good estimates of effective initial and subsequent D/d ratios for real materials in test specimens and in actual structures or machine parts. Such estimates and their consequences are developed in the sections which follow. The direct design problem for the materials has the converse objective of producing the desirable combination of microstructural properties to give maximum fracture toughness and strength within metallurgical constraints. The goal is not easy to achieve in an economical structural alloy with an appreciable volume fraction of impurities as well as desired phases. An almost fully ductile matrix of high flow strength can be achieved by solution and precipitation hardening with dimensions on the scale of less than 1000Å (0.1 μm) down to the atomic. On the other hand, particles above this range, including the many ordinarily visible in the optical microscope, should dominate the fracture process. They should have a minimum size which is as large as possible and as uniform as possible. For a given volume

fraction of inclusions the average D/d is independent of absolute size. The larger d, the fewer the number of particles, the larger the effective D for a random distribution, and, most important of all, the greater the tensile extension in the process of void coalescence. A bimodal distribution of particle sizes which skips the range of 0.1 - 10μm would have many advantages. These interesting questions involved in the design of metallic alloys will not be pursued further here. Quantitative development is needed, however, on details of the topics listed below:

Connectivity of Fracture Path and Effective D/d (like percolation theory)

Limit Loads, Critical Stress Levels, and Stress-Separation Relations During Void Coalescence (McClintock, Rice, Drucker,...)

Crack Initiation and Propagation in Plane Strain and Notched Tensile Bars

Spall (essentially a copy of my earlier unpublished note No. 30)

Necking of Smooth Tensile Specimens

Acknowledgement

This study was supported by the Advanced Research Projects Agency under Contract No. DAHC15-71-C-0253 with the University of Michigan.

SPALL FRACTURE BY HOLE GROWTH IN INCOMPRESSIBLE ELASTIC PLASTIC MATERIAL

F. A. McClintock

Abstract

Calculations are presented for spall fracture by hole growth as approximated by the spherical growth of holes in incompressible, elastic-plastic material. An order of magnitude estimate indicates that inertia effect, are more important than rate effect. Comparisons are made with the empirical fracture criterion of Tiller and Butcher, and with the more detailed numerical calculations by Wilkens.

CONFORMATION OF THE MODE AND WAVE FRONT APPROACH TO THE ANALYSIS OF WAVE PROPAGATION IN PERIODIC COMPOSITES

E. H. Lee

Abstract

The limiting phase velocity of Floquet waves at high frequencies should be consistent with the geometrical optics limit of a wave front travelling through each component at the appropriate wave speed. The resulting average wave speed should be in accordance with the phase velocity for high order bands of the frequency spectrum. By studying the structure of the mode shapes with increasing band number, a pattern becomes apparent, which permits determination of the relationship between band order and frequency. This is shown to be in agreement with the geometrical optics limit.

INFLUENCE OF PROPERTIES GRADIENTS ON STRESS WAVE PROPAGATION APPLICATIONS

E. H. Lee. B. Budiansky and D. C. Drucker

Abstract

Variation of elastic properties through a plate has been considered in order to increase its resistance to penetration. The relative merit of properties changes in a series of steps - graded material - or continuous variation - gradient material - is investigated.

It is found that for an elastic slab with elastic modulus increasing with depth, the stress wave front associated with an applied discontinuous surface pressure increases in proportion to $\sqrt{\rho(x)\tau(x)}$, where ρ is the density and τ the elastic wave speed for dilatational waves. This grows indefinitely with increasing $\tau(x)$. However, for a sudden change of properties at an interface, it is known that in the limit of change to a rigid body ($\tau \to \infty$), the stress magnitude only doubles. This paradox is explained by noting a singular approach to the limit in the continuously varying case. A boundary layer of high stress peak occurs for sharp changes of properties; which narrows in time towards zero duration as the gradient material approaches a graded one. The significance of this result to material damage under dynamic loading is assessed.

DETERMINATION OF STRESS PROFILES FOR WAVES IN PERIODIC COMPOSITES

L. Bevilacqua, J. A. Krumhansl and E. H. Lee

Abstract

Floquet or Bloch wave theory provides a convenient basic set of functions for representation of the propagation of transient elastic stress waves in periodic composites (Krumhansl, ARPA Mat. Res. Council Report, p. 175, 1970). Variational principles for computing dispersion relations and hence phase velocities generate a band structure of pass and no-pass frequency bands (Kohn, Krumhansl and Lee, ARPA Mat. Res. Council Report, Vol. I, Paper No. 2, 1969 and ASME Preprint 71-APMW-21, to appear in Jour. Appl. Mech.). Dispersion curves (frequency versus wave number) were accurately evaluated for laminar composites by using smooth Fourier series test functions for displacement in a Rayleigh-Ritz approximation procedure, but the corresponding stress profiles were unsatisfactory since the required continuity of stress at the inclusion-matrix inferface was ruled out by the use of the smooth test functions for displacement and corresponding continuous strain profiles.

In this paper exact stress profiles are calculated for waves propagated normally to the laminae, and satisfactory

approximations to these are generated with the extended variational principle which permits independent test functions to be used in the matrix and inclusions. An augmented plane wave approach in the Rayleigh-Ritz procedure was adopted in which exact solutions of the wave equation were used as component test functions in the filament. It was found that with this procedure, accuracy was essentially independent of the ratio of elastic moduli of the inclusion and matrix. The variational approach is applicable to two and three-dimensional composite configurations, which are not amenable to exact evaluation.