AD-777 743

PRELIMINARY REPORTS, MEMORANDA AND TECHNICAL NOTES
OF THE MATERIALS RESEARCH COUNCIL SUMMER CONFERENCE
HELD AT LA JOLLA, CALIFORNIA, JULY 1973
VOLUME I

MICHIGAN UNIVERSITY

PREPARED FOR
ADVANCED RESEARCH PROJECTS AGENCY

JULY 1973

AD777743

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

La Jolla, California

Vol. I

July 1973

Department of Materials and Metallurgical Engineering

D D C

APR 19 1974

D

PRELIMINARY REPORTS, MEMORANDA AND TECHNICAL NOTES

of the

MATERIALS RESEARCH COUNCIL SUMMER CONFERENCE

La Jolla, California

July 1973

VOLUME I
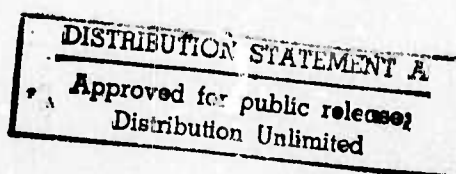
VOLUME I   TABLE OF CONTENTS

    I. Foreword

   II. Steering Committee

  III. Participants

   IV. Consultants

    V. Preliminary Reports, Memoranda and Technical Notes

The following papers fall into two categories; (1) papers
in a state ready for publication, and (2) reports and
memoranda for limited distribution representing work in
progress. The former category is available for general
distribution and in some cases are in the process of
publication in the appropriate technical journals. The
limited distribution reports and memoranda represent
initial ideas, problem suggestions, position papers, and
status reports and are aimed primarily to stimulate dis-
cussion with the Council. However, they are available
subject to the author's release by request to the Project
Director.

vi

VOLUME II   TABLE OF CONTENTS

TITLE

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.

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 M. F. Hawthorne
Department of Chemistry
University of California
Los Angeles, California 90024

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

Dr. George H. Vineyard
Brookhaven National Laboratory
Upton, Long Island, New York 11973

ARPA MATERIALS RESEARCH COUNCIL
1973 GUEST CONSULTANT LIST

R. A. Andrews
Naval Research Laboratory
Washington, D.C. 20375

C. B. Archembeau
California Inst. of Technology
Seismological Laboratory
Pasadena, California 91109

R. E. Balzhiser
Electric Power Research Inst.
Palo Alto, California

Frank Baratta
Army Materials & Mechanics
    Research Center
Watertown, Massachusetts 02172

M. R. Beasley
Harvard University
Pierce Hall
Cambridge, Massachusetts 02138

A. L. Bement
Massachusetts Inst. of Tech.
Cambridge, Massachusetts 02139

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

J. C. Bokros
Gulf Energy & Environmental Sys.
P.O. Box 81608
San Diego, California 92112

M. G. Bowman
Los Alamos Scientific Laboratory
P.O. Box 1663
Los Alamos, New Mexico 87544

W. L. Bradley
Colorado School of Mines
Golden, Colorado 80401

R. G. Brandt
Office of Naval Research
1030 Green Street
Pasadena, California 91106

B. F. Brown
Chemistry Department
The American University
Washington, D.C. 20016

F. C. Brown
Xerox Palo Alto Research Center
3180 Porter Drive
Palo Alto, California 94304

E. J. Cairns
General Motors Research Labs.
Electrochemistry Department
12 Mile & Mound
Warren, Michigan 48090

E. Catalano
Lawrence Livermore Laboratory
P.O. Box 808
Livermore, California 94550

W. H. Cook
Stanford Research Institute
Menlo Park, California 94025

D. O. Cowan
Johns Hopkins University
Chemistry Department
Baltimore, Maryland 21218

A. M. Diness
Office of Naval Research
Arlington, Virginia 22217

Morris Eisenberg
Electrochimica Corporation
2485 Charlston Road
Mt. View, California

Norbert Elsner
Gulf General Atomic
Box 608
San Diego, California 92112

A. G. Evans
National Bureau of Standards
Inorganic Materials Division
Washington, D.C. 20234

E. A. Fisher
Ford Motor Company
Turbine Research Department
Scientific Laboratory
Dearborn, Michigan 48121

A. L. Giorgi
Los Alamos Scientific Laboratory
Los Alamos, New Mexico 87544

A. J. Glass
Lawrence Livermore Laboratory
P.O. Box 818
Livermore, California 94550

William Goldberg
Air Force Materials Lab.
Wright-Patterson AFB
Ohio 45433

D. H. Gurinsky
Brookhaven National Laboratory
Metallurgy Department
Upton, Long Island, New York

Marvin Hass
U.S. Naval Research Laboratory
Washington, D.C. 20024

A. J. Heeger
University of Pennsylvania
David Rittenhouse Laboratory
Philadelphia, Pennsylvania 19174

P. W. Hoff
Lawrence Livermore Laboratory
P.O. Box 808
Livermore, California 94550

J. P. Howe
Gulf General Atomic
P.O. Box 608
San Diego, California 92112

J. R. Huff
U.S. Army Mobility Equipment
Research & Development Center
Electrotechnology Department
Electrochemical Division
Ft. Belvoir, Virginia 22060

J. K. Hulm
Westinghouse Electric Company
Research Labs.
Pittsburgh, Pennsylvania 13235

W. R. Hunter
U.S. Naval Research Laboratory
Washington, D.C. 20375

J. C. Hurt
Army Research Office
Box CM Duke Station
Durham, North Carolina 27706

Robert Jaffee
Battelle
505 King Avenue
Columbus, Ohio 43201

H. H. Johnson
Cornell University
Bard Hall
Ithaca, New York 14850

Al Joseph
Rockwell International, Inc.
Research Center
Thousand Oaks, California

Alfred Kahan
Air Force Cambridge Research Lab.
Hanscom Field
Bedford, Massachusetts 01730

Roy Kaplow
Massachusetts Inst. of Tech.
Cambridge, Massachusetts 02139

P. L. Kelley
Massachusetts Inst. of Tech.
Lincoln Laboratory
Lexington, Massachusetts 02173

G. C. Kugler
ESB Inc.
Tech. Center
Yardley, Pennsylvania 19067

F. F. Lange
Westinghouse Research and
    Development Center
Churchill Borough
Pittsburgh, Pennsylvania 15235

A. C. Lawson
University of California
Physics Department
La Jolla, California 92038

E. M. Lenoe
Army Materials & Mechanics
    Research Center
Arsenal Street
Watertown, Massachusetts 02172

C. C. Liang
P. R. Mallory
Burlington, Massachusetts 01803

J. C. Light
University of Chicago
Chicago, Illinois 60637

W. A. Little
Stanford University
Physics Department
Stanford, California 94305

J. J. Loferski
Brown University
Div. of Engineering
Providence, Rhode Island 02912

Margaret MacVicar
Massachusetts Inst. of Tech.
Cambridge, Massachusetts 02139

B. T. Matthias
University of California
Mayer Building
La Jolla, California 92037

Robert Mehrabian
Massachusetts Inst. of Tech.
Cambridge, Massachusetts 02139

R. W. Minck
Ford Motor Company
Science Research Center
Dearborn, Michigan 48121

S. P. Mitoff
General Electric Corporation
    Research & Development
P.O. Box 8
Schenectady, New York 12301

Amos Nur
Stanford University
Dept. of Geophysics
Stanford, California 94305

R. J. O'Connell
Harvard University
Dept. of Geology
Cambridge, Massachusetts 02138

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

Ward Plummer
University of Pennsylvania
Physics Department
Philadelphia, Pennsylvania

G. M. Pound
Stanford University
Materials Science Dept.
Stanford, California 94305

C. B. Raleigh
National Center for Earthquake
    Research
U.S. Geological Survey
345 Middlefield Road
Menlo Park, California 94025

Victor Rehn
Naval Weapons Center
China Lake, California 93555

Roy Rice
Naval Research Laboratory
Washington, D.C. 20375

Burton Richter
Stanford University
S.L.A.C.
Stanford, California 94305

L. J. Rogers
U.S. Army Electronics Command
Ft. Monmouth, New Jersey 07703

Stanley Ruby
Advanced Projects Res. Agency
1400 Wilson Blvd.
Arlington, Virginia 22209

J. R. Schrieffer
University of Pennsylvania
Dept. of Physics
Philadelphia, Pennsylvania 19104

J. J. Schuldies
Ford Motor Company
Scientific Research Lab.
Dearborn, Michigan 48121

Sigmund Schuldiner
Naval Research Laboratory
Washington, D.C. 20375

Mitchel Shen
University of California
Gilman Hall
Berkeley, California 94720

P. Shewmon
Argonne National Laboratory
Argonne, Illinois 60439

R. R. Smith
Air Force Materials Lab.
Metal Composites Branch
Wright-Patterson AFB
Ohio 45433

Marshall Sparks
Xonics Inc.
6837 Havyenhurst Avenue
Van Nuys, California 91406

W. E. Spicer
Stanford Electronic Lab.
Stanford, California 94305

R. W. Staehle
Ohio State University
Dept. of Metallurgical Engineering
Columbus, Ohio 43210

J. L. Stanford
Naval Weapons Center
China Lake, California 93555

C. M. Stickley
Director, Materials Sciences
   Office
Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, Virginia 22209

Myron Strongin
Brookhaven National Laboratory
Physics Department
Upton, Long Island, New York 11973

Masaki Suenaga
Brookhaven National Laboratory
Upton, Long Island, New York 11973

Max Swerdlow
Air Force Office of Scientific
   Research
1400 Wilson Blvd.
Arlington, Virginia 22209

Alex Tachmindji
Deputy Director
Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, Virginia 22209

M. R. Tek
University of Michigan
Dept. of Chemical Engineering
Ann Arbor, Michigan 48104

Alan Tetelman
University of California
6531 Boelter Hall
Los Angeles, California 90024

G. L. Tingey
Battelle Northwest Labs.
Box 999
Richland, Washington 99352

C. W. Tobias
University of California
101 Gilman Hall
Berkeley, California 94720

C. C. Tsuei
California Institute of
    Technology
W. M. Keck Laboratory
Pasadena, California 91109

E. C. van Reuth
Materials Sciences Office
Advanced Research Projects
    Agency
1400 Wilson Blvd.
Arlington, Virginia 22209

P. L. Walker, Jr.
Pennsylvania State University
Mineral Industries Bldg.
University Park, Pennsylvania

R. E. Watson
Brookhaven National Laboratory
Upton, Long Island, New York

C. H. Wells
Pratt & Whitney Aircraft
Materials Engineering & Research
    Laboratory
Middletown, Connecticut

Sheldon Wiederhorn
National Bureau of Standards
Washington, D.C. 20234

Mark Wilkins
Lawrence Livermore Laboratory
P.O. Box 808
Livermore, California 94550

H. G. F. Wilsdorf
University of Virginia
Dept. of Materials Science
Charlottesville, Virginia 22903

Clarence Zener
Carnegie-Mellon University
Science Hall
Pittsburgh, Pennsylvania 15213

STRESS CORROSION CRACKING AND HYDROGEN EMBRITTLEMENT
FROM THE VIEWPOINT OF THE DEFECT SOLID STATE

J. P. Hirth

Abstract

A Conference on SCC and HE was held in June 1973 at
Firminy, France, under the auspices of the U.S. National
Association of Corrosion Engineers and of the French Company,
Creusot-Loire. Eight guest assessors from outside disciplines
were invited to attend, to present a critique of the Conference,
and to present a viewpoint of the SCC-HE field. This paper is
my critique and assessment, presented from the viewpoint of the
defects in solids, dislocation theory area. The first four
pages include a number of suggested critical experiments. The
remainder is a more detailed commentary on SCC-HE.

SCC AND HE FROM THE
VIEWPOINT OF THE DEFECT SOLID STATE

J. P. Hirth

1. Introduction

In attempting to comment upon a meeting as broad as
the one discussed here on stress corrosion cracking and hydrogen
embrittlement, it is impossible to discuss the entire spectrum
of material presented. Instead, a limited number of issues
are pinpointed where alternate interpretations are possible,
where critical experiments seem feasible, or where phenomena
from other fields which have not received attention in the SCC
area might warrent study. Making a presentation in decreasing
order of importance, we first present a list of suggested ex-
periments, and next discuss the issues derived from the confer-
ence proceedings.

2. Suggested Experiments

A. Films and Adsorbates at Crack Tips I

Since as shown by Markworth et al.[1] and as discussed
below, the atomic configuration at a crack tip differs markedly
from that on a flat, unstressed surface, film formation and
adsorption kinetics may differ there very much also. That is,
the high strain of the atomic bonds most likely causes a change
in the adsorption energies of foreign atoms and molecules,

which, in turn, affects the aforementioned kinetics. While the extreme strains of 40 percent exhibited in the above tips[1] cannot be achieved for bulk surfaces, some indication of the effect of large strains on adsorption and film formation could be achieved by studying such kinetic processes on whiskers elastically strained to 1 to 5 percent.[2]

B. Films and Adsorbates at Crack Tips II

An alternative procedure to test the same phenomena as mentioned in A would be to study adsorption in computer simulation experiments. Such experiments have yielded good estimates of free surface energies and stresses[3]. Evidently, by methods analogous to those used by Markworth et al.[1], one could measure adsorption energies on unstressed free surfaces, on free surfaces stressed to the theoretical strength ($\sim$15 percent strain) and, less simply, on stressed crack tips in a simulated crystal.

C. HVEM Study of Advancing Crack

In view of the continuing controversy[4-6] about the site (interior vs. near surface) where hydrogen damages metal, it seems most important to actually observe crack propagation. Such observations have been performed on thin films in vacuum for 304 stainless steel and for silver in high voltage electron microscopy[7]. The resulting fracture surfaces were correlated with those on thicker, plane strain specimens and showed similar structures. Moreover, Wilsdorf[8] states that his microscope could

be immediately adapted to in situ gaseous environmental studies. Support of such work is strongly urged.

D.   Multiple Stress State Studies

With well established techniques for permeation studies[4,9,10] and with the determination of the pressure coefficient in uniaxial tension, more detailed studies would appear to be in order. Tests on oriented single crystals under biaxial and triaxial loading could establish the complete strain tensor associated with interstitial hydrogen in iron (see discussion of ref. 9). Of course, tests on such crystals after deformation on a single slip system would provide information on the role of dislocations in diffusion and in trapping.

E.   Adsorption to Defects

Another area where computer simulation would be valuable is in studies of adsorption to defects. Grain boundaries, dislocations, and stacking faults have all been suggested as either strong or weak adsorption sites[11-14]. Valuable quidelines to real system behavior could be provided by simulation studies, which are now within the capability of advanced computer techniques[1]. In particular, interactions of point defects with the nonlinear elastic fields of one and surface defects, not studied heretofore, could be investigated in this manner.

F.   Bicrystal Studies

A possibility which has been studied[15] but which has

-4-

received little attention in the SCC field, although it has been documented in the general mechanical property field[16] is that of incompatibility effects at interfaces. Because of the crystallographic nature of slip and of elastic anisotropy, incompatibility stresses can arise at interfaces which can both change the effective applied stress and induce additional stress components near interfaces. These stresses could be important in general in intergranular SCC modes and in grain size effects: in particular, they could influence the transition from inter-granular to transgranular failure. Studies of oriented bicrystals are needed. It has been shown that plastic incompatibility effects influence the intergranular fracture of aluminum-zinc alloys in various media. Elastic compatibility effects have not been studied in corrosive environments although mechanical properties have been shown to be affected[17]. In this connection, joint studies of molybdenum and tungsten or of aluminum and nickel at equivalent homologous temperatures and electrochemical conditions would be of interest, since the second member (but not the first) of each pair is essentially elastically isotropic (anisotropy factor ~ 1) while the members of each pair have identical slip systems and similar stacking fault energies.

3.   Commentary Upon Issues

   3.1 Education

   The paramount need in the SCC-HE area, superceding the important requirements for research and development, is that

for the education of system designers in the engineering principles of corrosion. This need was indicated by the introductory papers at the conference, in particular the series of case studies presented by Bush and Dillon[18]. Interestingly, this conclusion reinforces the major finding of an interdisciplinary conference on environmental degradation of materials[20], where the same problem was pinpointed, as a consequence of which it was proposed that a series of handbooks on corrosion cracking be produced for use by system designers. The funding of such a project should be a primary objective of the SCC scientific community and of funding agencies.

3.2 Crack Tip Stress State and Microstructural Scale

A crucial question with regard to mechanisms of SCC-HE is the state of stress at the crack tip, which depends critically on the detailed structure in the tip region and which remains uncertain in most cases. This point has been discussed by McClintock[19]; some detailed aspects of the situation are added here. The various elastic-plastic solutions, including work-hardening solutions, provide a good representation of the state of the material down to a size scale of the order of the dislocation cell spacing[19-21]. This cell spacing has been shown to be inversely proportional to the local effective stress[22-23] σ and universally related to the true strain.[22-23] The relation with stress is

$$d \simeq 3b \ (\mu/\sigma)$$

where d is the cell diameter, b is the magnitude of the dis-
location Burgers vector and μ is the shear modulus.

Typical dislocation cell spacings near the crack tip
are thus of the order of 200 to 1000Å or 40 to 200 b, where b
is the dislocation burgers vector magnitude, usually equal to
the atomic diameter. Within this spacing, the activation and
operation of discrete dislocation slip sources are important.
Of particular importance, these sources can operate in a manner
to either increase or decrease the critical stress intensity
for crack propagation relative to the well-defined stress in-
tensity K_{CG} to propagate a pure Griffiths crack. The works of
Rice and Thomson[24] and of Gilman[25] have shown* that with or
without the presence of a cracking environment, a crack can
propagate either with or without the emission if dislocations
from the crack tip depending on the relative surface energies
of the crack and line tensions of the dislocations; with the
actual parameters for different materials both situations are
predicted for different materials with iron a marginal case. In
the former case the critical stress intensity would exceed K_{CG}
while in the later case it would either equal or exceed K_{CG} if
no other dissipative process accompanied crack propagation.

Let us now present two subcases of the latter case which

_____

*The former paper included nonlinear elasticity via the use of a
Peierls model for the emitted dislocation while neither paper
included nonlinear elasticity for the crack tip field. As dis-
cussed in the next section, the work of Markworth et al.[1] indi-
cates that nonlinear effects are important; nevertheless the works
cited above provide suggestive guidelines.

lead to overall K_C values greater and less than K_{CG}. Consider the situation in Fig. 1 and suppose that the only possible dislocation sources are in a cell wall at a distance L from the crack tip. The stress field of the crack tip[19,20] is such that edge dislocations emitted from the wall and of the signs shown would tend to move to positions A or B. At site B, the dislocation will relax the normal stress tending to extend the crack. Hence, operation of sources to produce dislocations or pileups of type B will concentrate a stress opposed to the applied stress and raise the critical stress intensity leading to a possible case $K_C > K_{CG}$. At site A, the dislocation will enhance the normal stress tending to extend the crack[21], and a pileup could lead to a possible case $K_C < K_{CG}$. Mnemonically, it is useful in this case to recall that with an Eshelby-type pileup[26], a large grain size and a limited number of slip systems, brittle cracking can be induced at a critical stress approaching zero because the stress concentration produced by the pileup multplies the elastic stress by a large number (the number of dislocations in the pileup); an analogous situation applies for the crack, although scale is still important in that pileup stresses can be relaxed by the motion of other defects. A more detailed analysis of the situation is presented in Appendix I, where slip coplanarity is also considered, leading to what is tantamount to a mechanistic version of the model presented by Tetelman and Kunz[27].

With dislocation emission from the crack tip and pileup at a cell boundary, microcrack nucleation ahead to the macrocrack

would be possible. This phenomenon has been observed in single crystals[7] and has been shown to be consistent with discrete slip system operation. The mechanistic details, however, remain unresolved.

With regard to the actual stress at the crack tip, let us follow McClintock[19] for the model of Fig. 1. If the stress will increase toward the crack tip according to the Mode I elastic solution $\sigma \propto K/\sqrt{r}$, with r the distance from the tip. Hence, at a cutoff distance $\sim b$ near the tip (introduced to account for nonlinear elasticity analogous to the core cutoff of dislocations) the stress at the tip will be enhanced and will reach a value $Y\sqrt{L/b}$. For the case of hardened alloy steel with L = 100b this enhancement is sufficient to give a stress equal to the theoretical strength of a perfect crystal at the crack tip.

3.3 The Atomistic Level at the Crack Tip

The result of Markworth et al.[1] is quite significant with regard to the critical stress to propagate a crack tip and with respect to environmental cracking mechanisms. The crack tip regions of their Figs. 3-7 are highly nonlinear elastically. Nonlinear elastic effects become marked in simple tension at strains $\gtrsim 2$ percent. The elastic bond strains in their crack extend up to 40 percent with one case of 100 percent. Moreover, the nonlinear elastic region extends 5 to 10 atomic diameters away from the crack. These facts have several implications:

(a) Adsorption and absorption (or entry) kinetics for
such highly strained bonds will be markedly different from those
of a flat unstrained or mildly strained surface. Film formation,
at least in the early stages, will also differ. Hence, crack
tip calculations using bulk equilibrium data may require modi-
fication[25]. Also, the extension of bulk polarization rates and
film formation kinetics and stability regimes to crack tip
regions, the basis for much of the justification of the film rup-
ture theory of SCC, may be questionable, as may the use of linear
elastic results. Hence the need for experiment A, the study of
strained metal whiskers.

(b) In consideration of second-order interaction effects,
for example the coupling of electron concentration and stress
with subsequent effect on local hydrogen concentration[28], third
and higher order elastic interaction terms may be equally important.

(c) The nonlinear elastic region will have effective
elastic constants lower than the unstrained bulk material.

(d) The result for iron confirms the Rice-Thomson pre-
diction that iron is marginal with respect to dislocation punching
versus crack extension, since both processes occur (Markworth et al.
Figs. 4 and 5) at about the same critical stress intensity. How-
ever, the mode of dislocation entry is unusual. In their Fig. 4,
dislocations have formed but remain fixed in position within two
atom layers of the crack surface. This may be associated with
the lower elastic moduli in the nonlinear crack region which

-10-

would lead to a "modulus" type elastic interaction with the dislocation (in the precipitate-solute atom sense) tending to bind the dislocation to the crack tip region. This could further account in part for the fact that the critical stress intensity is only 2.59 multiples of K_{CG}, even though plastic deformation is induced to the limited extent discussed above.

(e) A question that remains open is whether reversible crack extension at $K_C = K_{CG}$ is ever possible. If the bond breaks reversibly at the crack tip, such reversible extension is possible. However, if the bond passes through an activated state, bond "pinging" will occur in addition to creating new surface, some energy will be dissipated as heat and K_C will exceed K_{CG}. More detailed work of the type performed by Markworth et al.[1] is required to settle this issue. As a preliminary guideline, calculations are presented in Appendix II which indicate that the pinging effect is small in metals: similar effects are likely for ionic crystals but not for covalent crystals where rearrangement could lead to highly irreversible effects.

3.4 Interaction of Hydrogen and the Applied Stress

Since hydrogen interstitials in bcc iron produce a tetragonal distortion field, the effect of stress on solubility will not be a simple hydrostatic one as in the first order treatments of the effect[9,29]. Referred to cube axes the strain field will be $\varepsilon_{11} > \varepsilon_{22} = \varepsilon_{33}$, $\varepsilon_{12} = \varepsilon_{13} = \varepsilon_{23} = 0$. Referred to a molar volume of iron and to the specimen axes, the strain field of the

defect will be ε'_{ij}, the applied stress field σ'_{ij} and the inter-
action energy

$$W = \sum \sigma'_{ij} \, \varepsilon'_{ij} \, V \tag{1}$$

where V is the molar volume. For simple tension the applied
stress is σ'_{11} and can be resolved into the cube axis coordinates
by

$$\sigma_{ij} = a_{i1} \, a_{j1} \, \sigma'_{11} \tag{2}$$

where a_{ij} are the appropriate direction cosines. Hence,

$$W = [a_{11}^2 \, \varepsilon_{11} + (a_{21}^2 + a_{31}^2) \, \varepsilon_{22}] \, \sigma'_{11} \, V$$

$$= 3[a_{11}^2 \, \varepsilon_{11} + (a_{21}^2 + a_{31}^2) \, \varepsilon_{22}] \, \sigma'_h \, V \tag{3}$$

where σ'_h is the hydrostatic component of the stress. For a
general polycrystal, the interaction energy will be the integral
over all orientations of interstitials. In eulerian coordinates
for the cube axes relative to the tensile axis, the bracketed
term in Eq. (3) is simply $[\cos^2\theta \, \varepsilon_{11} + \sin^2\theta \, \varepsilon_{22}]$ where $\cos \theta$
$= a_{11}$. If all orientations appeared, as for a polycrystal, and
all were equally likely, the integral of (3) would yield the
average value

$$\bar{W} = \frac{3\sigma'_h V}{2\pi} \int_0^{2\pi} \int_0^{\pi} [\cos^2\theta \, \varepsilon_{11} + \sin^2\theta \, \varepsilon_{22}] \, \sin\theta d\theta d\phi$$

$$= 3\sigma'_h \, V[0.333 \, \varepsilon_{11} + 0.667 \, \varepsilon_{22}] \tag{4}$$

-12-

For the isotropic defect approximation employed by others[9,10],
$\varepsilon_{11} = \varepsilon_{22} = e/3$, where e is the dilatational strain, i.e.,
$\bar{V}_H/V$, with $\bar{V}_H$ the partial molar volume of hydrogen, and Eq. (4)
reduces to the standard result

$$\bar{W} = \sigma_h' \bar{V}_H \qquad (5)$$

With an asymmetric interstitial defect, however, one
expects $\varepsilon_{11} > \varepsilon_{22}$. Moreover, while all orientations are expected
to be present in a polycrystal, with varying solubility in a
stress field one also expects the interstitials to align them-
selves in the most favorable direction energetically[30,31].
Let us assume complete alignment. Then, for tension, ε_{11}, a
compressive strain, will align with the tensile axis to give
the largest negative interaction energy, i.e., the coordinate
x_1 will align as closely as possible to the tensile axis, and
orientations will vary from $\theta = 0$, perfect alignment, to $\theta = 35.3°$,
the angle between [100] and [111] directions, which represents
the maximum possible deviation of the tensile axis from the
favored direction. Then,

$$\bar{W} = 3\sigma_h' V[0.042 \ \varepsilon_{11} + 0.625 \ \varepsilon_{22}] \qquad (7)$$

In the actual case, perfect alignment would not be obtained and
there would be a Boltzmann distribution of atoms aligned in
unfavorable directions (neglect of this is the reason that
Eqs. (6) and (7) do not reduce to Eq. (4) in the limit $\varepsilon_{11} \rightarrow \varepsilon_{22}$).

However, at low temperatures, perfect alignment should be an excellent approximation.

The result of Eqs. (6) and (7) shows explicitly the point made in the discussion of Ref. (9): If data corresponding to these relations are _phenomenologically_ fitted to Eq. (5), the resulting parameter is only an apparent value of the molar volume of Hydrogen $\bar{V}_{Happ}$ which will differ from the true value Ve. As an example, suppose that $\varepsilon_{11} = 2 \varepsilon_{22}$, so that $e = 2 \varepsilon_{11}/3$. Then Eq. (6) gives

$$\bar{V}_{Happ} = 1.56 \, Ve = 1.56 \, \bar{V}_H \tag{8}$$

Equation (7) gives $\bar{W} = 1.065 \, \sigma_h' \, V \, \varepsilon_{11}$, which yields

$$\bar{V}_{Happ} = 1.60 \, Ve = 1.60 \, \bar{V}_H \tag{9}$$

With other choices of ε_{11} and ε_{22} the difference in results could be larger: e.g., with $\varepsilon_{22} = 0$, Eq. (8) gives $\bar{V}_{Happ} = 2.42 \, \bar{V}_H$, while Eq. (9) gives $\bar{V}_{Happ} = 0.38 \, \bar{V}_H$.

The above results show that the phenomenological molar volumes of hydrogen resulting from a fit of Eq. (5) to data[9,29] in general should differ from the true value. Moreover, the above results predict, except for the special case $\varepsilon_{22} = 0.482$ ε_{11}, that $\bar{V}_{Happ}$ in tension will differ from that in compression for a polycrystal. Of course, Eqs. (8) and (9) would be altered somewhat in the anisotropic elastic case or with incompatibility stresses present or with preferred orientation of the polycrystals

-14-

so no great significance should be placed on the value $\varepsilon_{22} = 0.482\ \varepsilon_{11}$. The above result does suggest possible detailed analysis of tension-compression data[9] to ascertain whether it is conceivable that $\bar{V}_{Happ}$ is different in tension and compression. The results to date[9,29], coupled with the above analysis, suggest that $\varepsilon_{22} \sim 0.5\ \varepsilon_{11}$, unlike the case of carbon in iron where $\varepsilon_{11} = 0.38$, $\varepsilon_{22} = -0.02$. More importantly, the result suggests the need for oriented single crystal work.

For emphasis, we repeat the point made by Oriani[4] that the above analysis leads to the expectation of interaction of hydrogen with both edge and screw dislocations and additionally with the stress fields of both Mode I or II and Mode III cracks.

3.5 Hydrogen Diffusivity

Hydrogen permeation is becoming well established in view of the excellent work in the field, reviewed in refs. 6, 9 and 10 at the conference. Several questions remain, however. The first is the result of Bockris and coworkers[9,29] that the dependence of solubility on the applied stress tensor accounts completely for the stress dependence of the hydrogen permeation rate, i.e., the hydrogen diffusivity is independent of the stress tensor. In the common thermal activation viewpoint, the activated state involves a configuration where the diffusing atom strains the surrounding lattice to a degree different from that in the ground state. This implies an activation volume, or, in the context of the previous section, an activated strain

tensor. Such activation volumes have been observed in a number of cases[32]; in particular, they have been predicted theoretically[33,34] and observed experimentally[33] for interstitials in body-centered-cubic metals. Hence, if hydrogen strains the lattice in the ground state to the degree suggested by the measurements of $\bar{V}_H$, the above viewpoint would certainly predict an activation volume for diffusion of the same order of magnitude, so the experimental result is most surprising. We do note that there is one other such result[35] which created similar puzzlement[35], that of carbon in iron, where the true activation volume was shown to be zero (the activation strain was not studied): perhaps all interstitials in iron fall into an anomalous activation volume class.

Two further thoughts arise in the above connection. The activation volume would be zero if the diffusing atom acquired sufficient energy to pass over the barrier independent of the configuration of the iron atoms in the activated state. This situation does not seem likely, but, if it were so, the heat of transport of hydrogen in iron would equal the activation enthalpy for diffusion in the simple mechanistic model of thermal transport. Thus, experiments on thermal transport of hydrogen in iron would appear to be interesting. Alternatively, is it possible that the applied stress affects either the entry kinetics or the concentration or energy level of trap sites in a way that compensates for an effect on hydrogen mobility? Such an exact compensation would seem fortuitous, but it is known that trap-point defect

-16-

interaction energies can be changed by applied stress[36] and
that transient effects caused by interface constraints can occur
in diffusion[37]. Finally, the effect being sought on D is a small
one, so it seems conceivable that it is hidden in the experimental
scatter in the transient buildup or through-time tests, but not
in the steady-state permeability case.

A second point that raises interesting questions is the
isotope effect found by Oriani[4] on the threshold stress for
hydrogen-deuterium cracking of iron. He has suggested that the
difference is related to the effect on cohesive energy or specific
volume rather than that on transport properties, although he has
noted that a negligible difference has been found between hydrogen
and deuterium solubilities in α-iron[38] which would rule out the
specific volume effect and would at least bring the cohesive argu-
ment into question. The estimates of the isotope effect for
hydrogen-deuterium in steel[39] are $D_{H_2} \sim 2.6 \ D_{D_2}$ which is almost
twice the value required to explain the effect found by Oriani
on the basis of a change in permeability caused by a difference
in D values.

Oriani suggests that kinetics can be ruled out on the
basis that 2.5×10^2 to 2.5×10^4 sec would be required for hydrogen
to penetrate the calculated plastic zone of 5×10^{-3} cm. However,
one could argue that the hydrogen plays its role at the sub-
structure scale of 200 to 1000 $\overset{\circ}{A}$ as discussed in section 3.2.
Then, with diffusion from either a planar or a line source into
a semi-infinite medium, which should be an appropriate approxi-

-17-

mation, one can calculate the time required for the concentration C at a distance X from the crack at a time t to reach 99 percent of its value C_∞ at infinite time t_∞. With $D = 10^{-8}$ cm^2/sec and $X = 1000$ Å, the time required is $t \simeq 3\times10^{-3}$ sec. It is perhaps more realistic to take for D the value of Armco iron since the region is defect free in the cases of section 3.2. Then $D = 5\times10^{-5}$ cm^2/sec and at $X = 1000$ Å, the time is $t \simeq 6\times10^{-7}$ sec. A smaller X would make the times even shorter. Thus, if one postulates that the hydrogen is effective at some characteristic distance which is a fraction of the plastic zone thickness in size, it would appear possible to explain the result on the basis of transport kinetics. The predicted isotope effect on D is in the right direction and is within a factor of two of the result which would give exact agreement with experiment.

3.6 Adsorption of Hydrogen to Defects

Miodownik[11] has noted that the calculation of the elastic contribution to the heat of solution by Bockris and coworkers[9,29] is about an order of magnitude greater than the heat of solution measured experimentally and has suggested a compensating negative electronic correction*. Here, we show that the derivation by Bockris et al.[24] is in error by a factor of about 8 because of an inadvertent mistake in their derivation and because of the neglect of image relaxations. In between their Eqs. (13) and (14),

---

*Were this so, the interaction energy treated later would also be high by such a factor.

Bokris et al. correctly note that the volume increase of an iron lattice with all interstitial sites occupied is $3 \bar{V}_H$ per hydrogen atom. However, they then incorrectly set the dilatational strain (our e, their ε) equal to $3 \bar{V}_H/V$, whereas, as discussed in section 3.4, $e = \bar{V}_H/V$. The error is that with all interstitial sites occupied, 3 gram-atoms of hydrogen are added, so in fact the strain is one-third of their value or simply $\bar{V}_H/V$.

Secondly, as first shown by Eshelby[40], the external value change δV produced by an interstitial atom exceeds the internal volume change δv because of image relaxations at a free surface. In the case of iron, Eq. (2-99) of ref. 31 gives

$$\delta V = (1 + 4\mu/3B)\delta v = 1.64\delta v \tag{10}$$

where $B = 1.78 \times 10^{12}$ dynes/cm^2 is the bulk modulus and $\mu = 8.6 \times 10^{11}$ dynes/cm^2 is the shear modulus[31]. Now, in the case of an external stress as in section 3.4, the appropriate volume term giving rise to interations with external pressure is δV. However, in considering the formation energy, neglecting a very small image stress, the appropriate volume change associated with the self stress field (Bockris et al.[29], Eq. (12)) produced by the defect is δv. Let us set $\delta V = \bar{V}_H$. Then, the corrected value of Bockris et al.[29], Eq. (15), the energy of solution is

$$W = \frac{B}{2V} (\delta v)^2 = \frac{B}{2(1.64)^2 V} (\bar{V}_H)^2 \tag{11}$$

This gives a result of $W = 7.3/N_A$ Kcal/mole versus the results[29] of $59/N_A$ Kcal/mole. If $\bar{V}_H = \bar{V}_{Happ}/1.6$ as suggested by Eqs. (8)

-19-

and (9), the value of W would be further reduced for the hydro-static interaction. However, in this case a shear term would have to be added, but the individual ε_{11}, ε_{22} value to permit this are unavailable so we do not pursue the matter. The result does show that a self-consistent interpretation of $\bar{V}_{Happ}$ vs. $\bar{V}_H$ gives a solution energy close to the experimental value.

Turning to adsorption to a dislocation, we note that Eq. (26) of Bockris et al.[29] for the concentration of hydrogen in the stress field of a dislocation is formally correct in the Boltzmann approximation of their Eq. (18) and with their stated assumptions. However, four terms neglected in going from Eq. (24) to (25) in their work are of the order of their final result for their analysis and of opposite sign, so their result is an over-estimate by two orders of magnitude as shown in Appendix IV. A more rapidly converging series which avoids these problems can be obtained by the method of ref. 31, p. 464, with the inclusion of higher-order terms than the conventional first order estimate presented there*, see Appendix IV. However, one should be cautious in applying the higher-order terms except for very dilute solutions. The reason is that except for very dilute solutions, site satur-ation occurs in the region where the higher-order terms are significant and then one must use a Fermi-Dirac distribution instead of a Boltzmann distribution (e.g., pp. 465-467 in ref. 31). For the parameters used in ref. 29 the value of r where saturation

_____

*For interstitials with their large δV's: for substitutionals the first-order case suffices.

-20-

occurs is about 2b and is on the tension side of the dislocation only, so the neglect of higher-order terms is reasonable for the hydrostatic elastic interaction; again, the inclusion of shear terms could change the situation. Expressed as N/L, the excess number of hydrogen atoms adsorbed per unit length, the result of Appendix IV gives $N/L \simeq 10^9$ cm^{-1} or about 20 hydrogen atoms per plane threaded by the dislocation. Recent computer simulations[41] indicate that the region within a radius $r \sim 10\text{Å}$ is a highly non-linear elastic one with effectively lower elastic constants. This suggests that a large "modulus"-type interaction is possible which could contribute roughly an additional 20 hydrogen atoms per plane threaded by the dislocation or $N/L = 10^9$ cm^{-1}. The estimate by Bockris et al.[29] for the core concentration is a gross overestimate and shows the need for using the Fermi-Dirac distribution. Use of their values[29] in their Eq. (29) gives a core concentration of about 100 hydrogen atoms per iron atom along the core (with the value of the equilibrium hydrogen concentration taken as 2×10^{-5} mole fraction[10]). This is impossible, since the Fermi-Dirac saturation limit is about 1 to 3 hydrogen atoms per iron atom, or $N/L = 5 \times 10^7$ cm^{-1}.

Miodownik[11] has suggested that adsorption to grain boundaries is likely to be small. However, we note that many boundaries in all but very well annealed pure metal crystals are likely to contain grain boundary dislocations with long range strain fields[42]. These would adsorb hydrogen in the same manner as the lattice dislocations discussed above. Also, of course,

-21-

the nonlinear boundary region would be a source for adsorption of about a monolayer of hydrogen per unit area.

Finally, for a Mode I crack, the stress field has a hydrostatic component

$$\sigma_h = \frac{(1+\nu)}{3} \frac{K_1}{(2r)^{\frac{1}{2}}} \cos \frac{\theta}{2} \tag{12}$$

This will give rise to an adsorption field analogous to that given by Eq. (22) in ref. 29. The resulting expression is again in the form of an infinite series as given by Eq. (23) in Appendix IV. Evaluation of the series shows that it gives a number of hydrogen atoms per unit length of crack of the same order as the number per unit length of the dislocation. The former result can greatly exceed the dislocation result for cases of small K and large r_1, with the opposite holding true for large K and small r_1. This result suggests that huge adsorption fields are possible in the stress fields of cracks and of dislocation pileups. A site-saturation, Fermi-Dirac atmosphere at the crack tip would tend to support the HE mechanism suggested by Johnson[6].

3.7 Crack Initiation

It seems that crack initiation is more-or-less being neglected in comparison to measurements of crack propagation as a function of stress intensity, whereas the incubation time could be a significant fraction of the life in many cases of cracking. Sufficient data exists in some cases to estimate the incubation time. Suppose that one knows the crack velocity v as a function

-22-

of stress intensity K for a given system. Then suppose that
one observes crack initiation and propagation in a smooth speci-
men, so that, once the crack starts, one can estimate the crack
length as a function of time t and hence estimate K(t). Then
the initiation time t_i and the propagation time t_p can be related
to the time to failure t_f by

$$t_i = t_f - \int_{t_o}^{t_f} v(K) \frac{\partial K}{\partial t} dt$$

Here t_o is the time at which, by back extrapolation in time, K
is so low that v has reached effectively zero velocity, e.g.,
$v = 10^{-10}$ cm/sec.

3.8 Cracking Mechanisms

The above discussion suggests our view that more detailed
studies are needed to fully establish the HE cracking mechanism
in iron and in steels. Questions remain regarding the role of
diffusion, the effect of stress on diffusion, and the degree of
adsorption to defect sites. The asymmetry of the strain field of
hydrogen in bcc metals can have important implications; in par-
ticular, maximum adsorption will likely occur other than in the
region of maximum hydrostatic tension ahead of a Mode I crack.

With regard to SCC, the preponderance of the results
presented at the conference strongly favored the initiation of
SCC by the film-rupture mechanism. In this connection, however,
we note some reservation with regard to the situation at the

-23-

crack tip. Also, it appears that considerable work is needed involving the microscopic scale around the crack tip considered by McClintock[19] in order to more completely describe detailed mechanisms of crack propagation in SCC by the film rupture model. In the most advanced effort along these lines, Vermilyea[43] was forced to leave t_c, the time for strain to build up and restart the crack, as an unknown parameter. Micromechanics or mechanics models are required to give estimates of parameters such as t_c.

Acknowledgements

APPENDIX I

A Detailed Model for Dislocation Effects at Crack Tips

As shown by Rice and McClintock[19], there are various scales within which different viewpoints prevail in describing the situation about a crack tip. Here, we wish to focus on the penultimate fine scale; that of the dislocation structure. To place this in a concrete context, let us consider a planar crack. At the grossest scale the solution is the elastic K_I solution, next the small scale yielding solution of Rice[44], or various elastic-plastic solutions reviewed elsewhere[18,19] next a region of size nominally related to discrete dislocation spacings, and finally the atomic configuration at the crack tip.

There is substantial evidence that coplanarity of slip enhances SCC sensitivity, e.g., the discussions at this conference. On the other hand, the mechanism of enhancement remains obscure. One possibility is that of the requirement for a large slip step to initiate the slip dissolution model, or the variant of requiring a large slip event to rupture a film followed by dissolution as in the slip-dissolution case. However, Latanision and Staehle[45] found that the presence of a protective film on nickel led to finer slip in cases where SCC was enhanced. The other notion is related to the increased likelihood of large stress concentrations with coplanar arrays, but a detailed mechanism has not been put forward. Here we shall attempt to

indicate qualitatively that coplanar slip together with a fixed microstructure scale can influence cracking susceptibility. Since no detailed observations are available at such a scale, however, we first briefly consider the case of slip around a nondeforming particle for which observations are available and then extend the arguments to the case of the crack.

Consider a spherical, nondeforming particle in a single crystal deformed by simple shear on a single slip system and suppose that dislocations not captured by the particle pass through the region in question. The elastic field of the captured dislocations can be determined by a method due to Eshelby[40], in which the particle is removed, the medium sheared uniformly, and the particle then reinserted, deforming the medium as shown in Fig. 2.

The resultant field is equivalent to a set of closely spaced Orowan dislocation loops in the interface of the particle[46] and corresponds reasonably with expectation from a continuum mechanics calculation[47] for a strain-hardening elastic-plastic case, which is the two dimensional analog of Fig. 2. On observing such particle microstructures, however, relaxation is observed over a scale corresponding to the interparticle spacing λ, occurring by the formation of punched-out loops and helices, as illustrated in Fig. 3. Thus, only on a scale of size $> \lambda$ will the discrete relaxed dislocation result and the continuum result coincide.

Now consider a Mode I crack tip, Fig. 4, on the scale of the dislocation microstructure denoted by S, the source spacing, which could be particle spacing, dislocation cell wall spacing, or grown in dislocation spacing, as microstructural scale parameters.

To make the point, suppose that the sources are such as to produce glide on a plane and in a direction both normal to the crack plane. In configuration (a) the stress field of the crack will be such as to operate S_1 to produce blunting of the crack as in (b). Let a new crack now start to grow on an atomically sharp basis as in (c), with some relaxation continuing at source S_1. With the crack in configuration (d), the stress field will be such as to operate source S_2. In this case, however, the dislocation motion, while relaxing the total strain energy of the crystal, will increase the local normal stress at the crack tip. One possible result would be that the local critical stress is exceeded so that the crack will grow rapidly until it reproduces configuration (a), whereupon the cycle repeats.

The picture is obviously speculative, but it does show a possible role of structure on the discrete dislocation scale. An alternative would be for the crack to emit dislocations rather than activating sources. Also, sources could operate on a variety of systems inclined to the crack plane, to cite two different possibilities.

Consider now the role of coplanarity in the model of Fig. 4. In materials where slip is coplanar (fcc, low stacking

fault energy; hcp, basal slip), the situation of Fig. 4 would be favored. The intrinsic slip pattern would be one of multi-polar arrays of dislocations of one slip system which could act as the sources in Fig. 4. Moreover, the low stacking fault energy (fcc or hcp) or low Peierls stress (hcp) would tend to restrict slip to a single plane, as shown in the figure. With non-coplanar slip, Fig. 5, the dislocation arrangement will tend to be one of cell walls containing dislocations of several glide systems. Further, systems of this type coincide with those in which cross-slip is a relatively easy process. Thus, both from the viewpoint of more slip systems and cross-slip, the situation of Fig. 4(c) is less likely to occur. Instead, as indicated in Fig. 5, continual blunting of the crack would occur.

APPENDIX II

A Preliminary Calculation of Atomic Relaxation
at a Crack Tip

Of importance in evolving cracking mechanisms as noted
in the main text, is the question of whether the bond relaxation
in cracking is nearly reversible or highly irreversible at the
crack tip as the crack propagates. In order to provide some
insight in this question, we present a simplified representation
of the crack tip. The model is shown in Fig. 6. Lateral dis-
placements are ignored, the medium is regarded as continuous ex-
cept for three bonds, represented by Morse "springs", and the
central bond is supposed to break at the unstable equilibrium
position. The constraint of the surrounding elastic medium is
introduced by supposing the total length of the three bonds to
remain constant as the middle bond breaks. This is, of course,
an over-estimate of the medium compliance, and will lead to an
under-estimate of both the bond relaxation and the energy of
relaxation.

The Morse potential energy as a function of bond length
r is

$$W = W_o \{\exp[-2a(r_o - r_o)] - 2 \exp[-a(r-r_o)]\} \tag{13}$$

where W_o is the energy at the rest equilibrium position r_o and
a is a parameter. A typical value for metals[48] is $a = 5/r_o$.
We further define the reduced variable $Z = (r-r_o)/r_o$, i.e., Z

is the bond strain. With the above values the unstable equilibrium is at $Z = 0.14$.

From Eq. (13), the total energy of the three bonds is then in the rest state $Z = 0$,

$$W_O^o = -3\ W_O \tag{14}$$

in the unstable equilibrium state $Z = 0.14$,

$$W_1 = -2.24\ W_O \tag{15}$$

In the state corresponding to the broken bond of Fig. 6, two bonds will relax to new equilibrium positions Z_2 and the third to a position $Z_3 = 0.42 - 2Z_2$. The values Z_2 and Z_3 are given by setting $\partial W/\partial Z_2 = 0$. The result is $Z_2 = 0.041$, $Z_3 = 0.34$, with a corresponding energy

$$W_2 = -2.27\ W_O \tag{16}$$

Hence, with the Morse potential, the displacement relaxation is relatively large, the displacement of the broken bond being $0.34\ r_O$ versus $0.14\ r_O$ in the unstable equilibrium position. However, the energy of relaxation is relatively small $(W_1 - W_2)/W_O^o = 0.01$ relative to the net energy change $(W_2 = W_O^o)/W_O^o = 0.24$. Thus, the Morse potential calculation tends to indicate that the process approaches a reversible one.

A more definitive calculation is expected to shift the above result further in the direction of reversibility. As shown by McClintock and O'Day[49], the results of Fuchs' atomic

calculations for copper, which should be fairly accurate, indicate a maximum bond strain of $Z = 0.045$ versus the above value of $Z = 0.14$ for a Morse potential, $Z = 0.12$ to 0.22 for Born-Mayer potentials, and $Z = 0.25$ for a sinusoidal potential. Thus, the above Morse potential is too "soft" for copper, and probably for other metals. A more realistic potential would probably shift the situation toward reversibility with lower broken bond displacements and a lower energy of relaxation.

On the other hand, as noted above, the compliance of the surrounding medium is underestimated, a factor which would tend to compensate for the potential law and shift the situation toward that of a pinging type relaxation. With the already highly simplified model of Fig. 6, however, it does not appear worthwhile to attempt more elaborate, and perforce lengthier, calculations. Rather, we await the results of further computer simulations[1]. The above calculation does serve to show that the question reversibility in the crack tip propagation remains open.

APPENDIX III

The Eulerian coordinates for the compression problem
are shown in Fig. 7. As noted in the text the X_1 axis can take
on values with θ from 35.3° to 90°. However, in order that the
X_1 axis remain that furthest away from the compression axis, only
value of X_1 such that the angle $\beta < \theta$ are allowed and similarly
for the angle between the compression axis and a line normal to
P and at an angle $\kappa + \pi/2$. For this geometry $\cos \beta = \sin\theta \sin\kappa$.
Thus, κ can vary between $\kappa_1 = \sin^{-1}(\cot \theta)$ and $\kappa_2 = \cos^{-1}(\cot \theta)$.
The integral equivalent to Eq. (4) then becomes

$$\bar{W} = \frac{3\sigma_n'V}{\pi^2} \int_0^{2\pi} \int_{\kappa_1}^{\kappa_2} \int_{35.3°}^{\pi/2} [\cos^2\theta\ \varepsilon_{11} + \sin^2\theta\ \varepsilon_{22}]\ \sin\theta d\theta d\kappa d\phi$$

$$= \frac{3\sigma_n'V}{\pi} \int_{35.3°}^{\pi/2} [(\cos^2\theta\ \sin\theta\ \varepsilon_{11} + \sin^3\theta\ \varepsilon_{22})$$

$$\times\ [\cos^{-1}(\cot\theta) - \sin^{-1}(\cot\theta)]d\theta \tag{17}$$

since the integrals over ϕ and κ are trivial. Equation (17)
can be solved by the transformation of variables $\cot\theta = \cos\phi$
giving

$$\bar{W} = \frac{6\sigma_n'V}{\pi} \int_{\pi/4}^{\pi/2} \frac{[(\pi/2)-2\phi]}{(1+\cos^2\phi)^{5/2}}\ (\cos^2\phi\ \sin\phi\ \varepsilon_{11} + \sin^3\phi\ \varepsilon_{22})d\phi \tag{18}$$

-32-

This expression can be solved by integration by parts to get rid of the factor ϕ, followed by explicit integration using expressions 2.271 and 2.272 in ref. 50 to give Eq. (7) of the text.

The basic equation for the number of solute atoms adsorbed per unit length in the stress field of an edge dislocation is (ref. 29, p. 1215; ref. 31, p. 464)

$$\frac{N}{L} = C_O \int_0^{2\pi} \int_{r_0}^{r_1} [\exp\left(\frac{K \sin \theta}{r}\right) - 1] dr \, d\theta \tag{19}$$

where $(K \sin \theta)/r$ is $\bar{W}/kT$, Eq. (5), with the hydrostatic stress field of the dislocation substituted, C_O is the bulk equilibrium concentration of hydrogen and r_0 and r_1 are core and outer cutoff radii in the stress field of the dislocation. Bockris et al.[29] correctly evaluated the indefinite integral over dr but then neglected several terms in r_0 and r_1 on inserting the limits. Keeping these terms one finds that their Eq. (27) acquires additional terms

$$C_{add} = C_O \left[-\frac{r_0^2}{r_1^2} I_0(K) - \frac{Kr_0}{r_1^2} I_1(K) + I_0\left(\frac{Kr_0}{r_1}\right) \right.$$
$$\left. + \frac{K}{r_1} I_1\left(\frac{Kr_0}{r_1}\right) - 1 \right] \tag{20}$$

where I_0 and I_1 are modified Bessel functions of order zero and one, respectively. The integrals leading to Eq. (20) are of the form

-34-

$$\int\limits_{0}^{2\pi} \exp\,(K\,\sin\,\theta/r_o)\,d\theta \quad \text{and} \quad \int\limits_{0}^{2\pi} \sin\theta\,\exp\,(K\,\sin\theta/r_o)\,d\theta$$

The first is given by item 3, 387.1 in ref. 50 and the second by item 8, 431.5.

For the case treated by Bockris et al.[29], i.e., $K = 17.2\ r_o$, $r_1 = 1000\ r_o$, their Eq. (27) gives the result $C = 51.9478\ C_o$. Equation (20) for the same parameters gives the result $C_{add} = -50.5196 C_o$. The correct total result is thus $C = 1.4282 C_o$ which is considerably smaller and gives a result for N/L a factor of 121 times smaller than that of ref. 29.

A simpler approach is simply to extend the series expansion on p. 464 of ref. 31, which gives the result to replace Eq. (27) of ref. 29 of

$$C = C_o[1 + \frac{K^2}{2r_1^2}\ \ell n\ \frac{r_1}{r_o}$$

$$+ \sum_{n=1}^{\infty}\ (\frac{K}{2})^{2n+2}\ \frac{1}{r_1^{\,2}n[(n+1)!]^2}\ (\frac{1}{r_o^{\,2n}} - \frac{1}{r_1^{\,2n}})\,] \tag{21}$$

This result converges somewhat more rapidly than the combination of Eq. (20) and Bockris et al.[29] Eq. (27), directly giving $C = 1.4282 C_o$. These values of C can be converted to number of atoms adsorbed per unit length of dislocation by

$$\frac{N}{L} = (C-C_o)\ \pi(r_1^{\,2}-r_o^{\,2}) \tag{22}$$

Also of interest with respect to the text is the analogous expression for adsorption in the stress field of a Mode I crack, Eq. (12). The result is

$$\frac{N}{L} = \frac{\pi C_0 A^2}{2} \left[\frac{(r_1 - r_0)}{r_0} + \frac{1}{16} \ln \frac{r_1}{r_0} \right.$$

$$\left. + \sum_{n=1}^{\infty} \frac{A^n}{2^{2n+2} n[(n+2)!]^2} \left(\frac{1}{r_0^n} - \frac{1}{r_1^n} \right) \right] \tag{23}$$

where r_0 and r_1 are inner and outer cutoff radii centered on the crack tip and

$$A = \frac{\sigma_h \bar{V}_H r^{\frac{1}{2}}}{kT \cos(\theta/2)}$$

with σ_h given by Eq. (12).
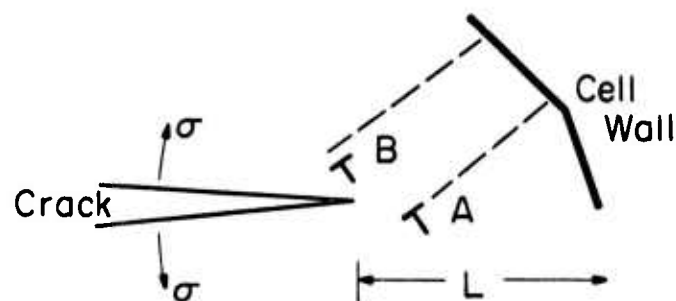
Figure 1. Model for emission of edge dislocations from a cell
wall under the influence of the stress field of a
crack.
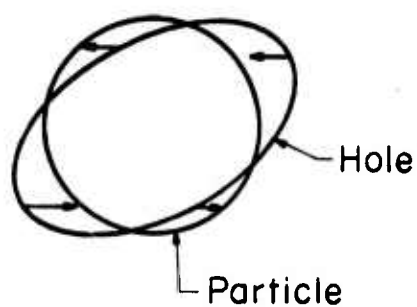
Figure 2. Local deformation field associated with the re-
insertion of a rigid ball into a strained hole.
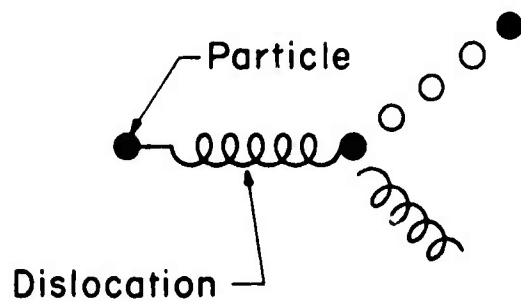
Figure 3. Dislocation loop arrangements formed by dislocations
punched out from particles.
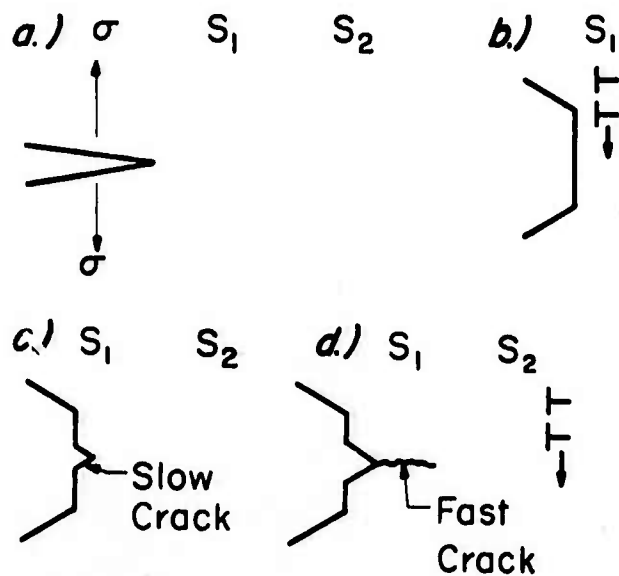
Figure 4. Stages in intermittent growth of a Mode I crack
under the influence of edge dislocations emitted
from dislocation sources S_1 and S_2.
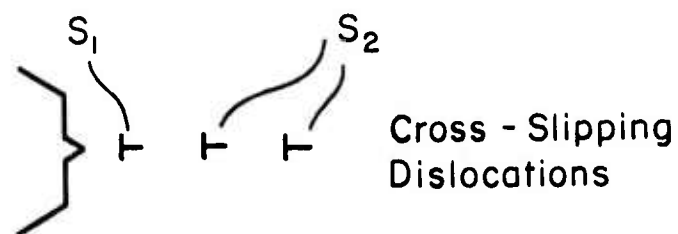
Figure 5. Crack configuration of Fig. 4c, with cross-slipping dislocations.
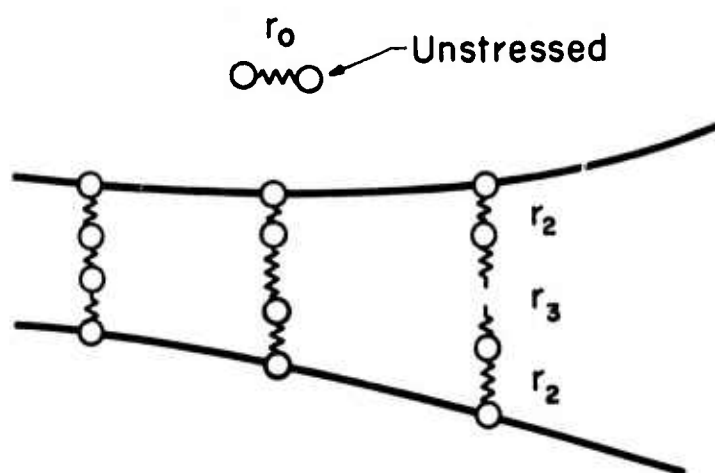
Figure 6. Simple model of a crack tip.
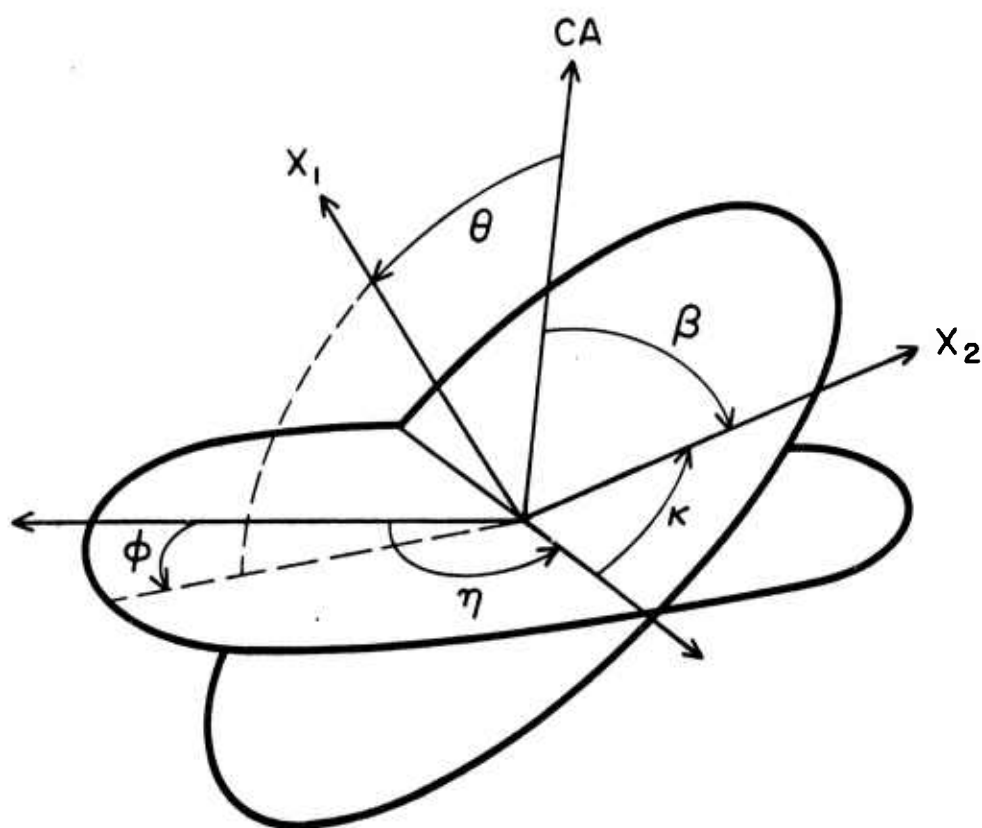
Figure 7. Eulerian coordinates θ, η, κ and cylindrical coordinates θ, ϕ for cube axes X_1, X_2, X_3 relative to the compressive axis CA.

References

1. Stress Corrosion Cracking and Hydrogen Embrittlement of Iron-Based Alloys, Vol. 5, International Conference Series, National Association of Corrosion Engineers, Houston, 1974. Article by A. J. Markworth, M. F. Kanninen and P. C. Gehlen.

2. S. S. Brenner, in R. H. Doremus et al. (eds.) "Growth and Perfection of Crystals," Wiley, New York, 1958, p. 3.

3. C. W. Price, Battelle Columbus Laboratories, Columbus, Ohio, 43201, research in progress.

4. R. A. Oriani, in ref. 1.

5. C. D. Beacham, in ref. 1.

6. H. H. Johnson, in ref. 1.

7. R. W. Bauer and H. G. F. Wilsdorf, Report No. MS-3533-106-72U, University of Virginia, Charlottsville, Virginia, August 1972.

8. H. G. F. Wilsdorf, private communication, July 1973.

9. J. O'M. Bockris, in ref. 1.

10. L. Nanis and T. K. G. Namboodhiri, in ref. 1.

11. A. P. Miodownik, in ref. 1.

12. C. J. McMahon, H. C. Seng and K. Yoshino, in ref. 1.

13. P. Lacombe, M. Aucouturier, J. P. Laurent and G. Lapasset, in ref. 1.

14. R. Gibala, in ref. 1.

15. W. J. Kovacs and J. R. Low, Jr., Met. Trans. $\underline{2}$, 3385 (1971).

16. J. P. Hirth, Met. Trans. $\underline{3}$, 3047 (1972).

17. R. E. Hook and J. P. Hirth, Acta Met. 15, 535 (1967).

18. S. H. Bush and R. L. Dillon, in ref. 1.

19. F. A. McClintock, in ref. 1.

20. F. A. McClintock and J. R. Rice, in "Environmental Degradation of Stressed Materials," Rept. of the ARPA Materials

Research Council Summer Conference, University of Michigan, Ann Arbor, Michigan, July 1971, p. 202.

21. R. M. Thomson and J. P. Hirth, in ref. 20, p. 182.

22. A. S. Keh and S. Weissman, in G. Thomas and J. Washburn (eds.), "Electron Microscopy and Strength of Crystals," Interscience, New York, 1963, p. 131.

23. G. Langford and M. Cohen, Trans. ASM 62, 623 (1969).

24. J. R. Rice and R. M. Thomson, Phil. Mag., in press, July 1973.

25. J. J. Gilman, in ref. 1.

26. J. D. Eshelby, F. C. Frank and F. R. N. Nabarro, Phil. Mag. 42, 351 (1951).

27. A. S. Tetelmen and S. Kunz, in ref. 1.

28. W. A. Tiller, in ref. 1.

29. J. O'M. Bockris, W. Beck, M. A. Genshaw, P. K. Subramanyan and F. S. Williams, Acta Met. 19, 1209 (1971).

30. A. W. Cochardt, G. Schoeck and H. Wiedersich, Acta Met. 3, 533 (1955).

31. J. P. Hirth and L. Lothe, "Theory of Dislocations," McGraw-Hill, New York 1968, p. 470.

32. J. R. Manning, "Diffusion Kinetics for Atoms in Crystals," Van Nostrand, London, 1968, p. 34.

33. G. W. Tichelaar, R. V. Coleman and D. Lazarus, Phys. Rev. 127, 485 (1962).

34. R. A. Johnson, In "Diffusion in Body-Centered-Cubic Metals," Am. Soc. Met., Cleveland, 1965, p. 357.

35. J. Bass and D. Lazarus, J. Phys. Chem. Solids 23, 1820 (1962).

36. J. P. Hirth and M. Cohen, Met. Trans. 1, 3 (1970).

37. M. L. Swanson, G. M. Pound, R. F. Mehl and J. P. Hirth, Trans. AIME 224, 742 (1962).

38. T. Heumann and D. Primas, Z. f. Naturforsch. 21a, 260 (1966).

39. J. R. Manning, "Diffusion Kinetics for Atoms in Crystals," Van Nostrand, London, 1968, p. 132.

40. J. D. Eshelby, Solid State Phys. $\underline{3}$, 79 (1956).

41. R. G. Hoagland, Ph.D. Thesis, Ohio State University, Columbus, Ohio, June 1973.

42. J. P. Hirth and R. W. Balluffi, Acta Met. $\underline{21}$, 929 (1973).

43. D. A. Vermilyea, in ref. 1.

44. N. Levy, R. V. Marcal, W. J. Ostergren and J. R. Rice $\underline{7}$, 143 (1971).

45. R. M. Latanision and R. W. Staehle, in "Fundamental Aspects of Stress Corrosion Cracking," Nat. Assn. of Corr. Eng., Houston, 1969, p. 214.

46. L. M. Brown and W. M. Stobbs, Phil. Mag. $\underline{23}$, 1185 (1971).

47. W. C. Huang, Rept. SM-45, Harvard University, Cambridge, Mass., October 1970.

48. L. A. Girifalco and V. G. Weizer, Phys. Rev. $\underline{114}$, 687 (1959).

49. F. A. McClintock and W. R. O'Day, Jr., "Proceedings of the First International Conference on Fracture," Sendai, 1966, p. 75.

50. I. S. Gradshteyn and I. M. Ryzhik, "Table of Integrals, Series and Products," Academic Press, New York, 1965.

EARTHQUAKES AND THE MECHANICAL
BEHAVIOR OF ROCK

D. C. Drucker
B. Budiansky
J. R. Rice

Abstract

Earthquakes, spectacular as they may be, are but a
minor manifestation of large and continuing motion of the
crustal plates and underlying mantle. This relative motion
remains essentially steady over times very large in comparison
with the intervals between earthquakes. Attention is directed
toward regions such as California in which the tectonic plates
primarily slide past each other horizontally (strike-slip),
as distinguished from thrust fault regions such as Japan, Alaska,
and South America, in which crustal slabs are moving deep down
into the mantle. Thrust faults do occur in the predominantly
strike-slip region as a consequence of geometric mismatch.

Among the ideas and questions explored are: tectonic
stresses; local and average stress drop during earthquakes;
differences in the extent of the deformed region in the direction
transverse to the fault between strike-slip and an associated
thrust fault; creep and small motions due to slip prior to and
subsequent to earthquakes; patterns of displacement and of stress

variation through the lower, middle, and upper portion of the crust; possible melting of rock during sliding, healing and weakening of warm rock under pressure in the presence of water; dilatancy in "solid" and jointed rock, its cause, its role as a significant stabilizing factor, and its influence on the speed of stress waves to provide an advance indication of an impending earthquake.

EARTHQUAKES AND THE MECHANICAL
BEHAVIOR OF ROCK

D. C. Drucker
B. Budiansky
J. R. Rice

Phenomenological Observations

A suitable explanation for the occurence and the be-
havior of earthquakes should cover a large number of somewhat
confusing phenomenological observations. Among these are the
indications of a shear stress build up of about 500 psi or
30 bars in 100 years. This number is consistent with a shear
strain increment of 10^{-6} per year[1], which multiplied by a shear
modulus of 5×10^6 psi gives 5 psi per year. Yet the value of
30 bars accumulated in 100 years and relieved by a major earth-
quake, according to seismic records[2], is very small in compar-
ison with the strength of rock in the crust or the frictional
resistance to the sliding of rock on rock even at moderate depth
as indicated by measurements in the laboratory on essentially
coherent specimens[3]. The following set of numbers[1] gives some
feeling for this disparity and the variation of properties
through the crust. At a depth of 10 km the lithospheric pressure
is approximately 3 kb (45000 psi) and the temperature is approx-
imately 200°C. At 5 km the lithospheric pressure is half as much,
the shear strength of dry rock would be above 20,000 psi, and the

-47-

dry sliding frictional resistance would be in excess of 10,000 psi. At a depth of 25 km the temperature rises to about 600°C.

Variations in local conditions along and transverse to a fault are enormously greater than might be implied from the observation that earthquakes occur with such small build-up and drop of average stress. After all, just 100 meters difference in elevation changes the lithospheric pressure and the strength of the rock in the upper portion of the crust by 30 bars. Quite apart from this overwhelming influence of pressure, considerable variation is likely in the strength of rock, the coefficients of static or sliding friction, stress concentration and allied piling-up or hanging-up of relative motion, pressure of water, permeability and all physical properties of so inhomogeneous a system as the crust of the earth.

In oil fields and in large regions of the upper crust without strong seismic activity, shear stresses or principal stress differences of the order of 100 or 200 bars (a few thousand psi) have been measured at depths of a kilometer or more.[4]

Although considerable variations in rock types and properties do exist, earthquakes appear to initiate in and propagate across sedimentary as well as igneous rocks without dramatic difference.

Earthquakes are very local on a global scale. The very largest involve slip over a small fraction of the periphery of the tectonic plates. Minor earthquakes are very local with

dimensions of the region of slip small compared with the depth of the upper portion of the crust. The exceptionally straight San Andreas fault is wavy, not absolutely plane, with waves of about 1 km amplitude and wavelength of 50 km or so.

A marked decrease in the speed of travel of body waves, V_p and V_s, and in their ratio V_p/V_s, has been observed prior to thrust fault earthquakes in the USSR, the Adirondacks of New York State, and most spectacularly in a large region in the vicinity of the San Fernando fault.[5] The time lag between the observed drop in V_p/V_s and the subsequent recovery just prior to the earthquake correlates conceptually with diffusion of water to the region of the fault. Rock dilates as it deforms toward failure, thus inducing suction in the pore water and possibly vapor formation in fissures which open wide. With vapor formation or the formation of new dry cracks the speed of body waves would drop.[6] The pressure drop, large or small, tends to lend stability to the rock against continuing deformation until pore water flow eliminates the dilatancy hardening.[6b,7] No similar changes in wave speeds have been observed across the San Andreas fault.

Rocks are dilatant in bulk as small cracks open under shear stress.[8] Rock masses with joints between them also are dilatant. The behavior of "solid" or jointed rock resembles that of a dense sand but with far lower porosity and permeability. Water pressure in all cases reduces the effective pressure or

normal compressive stress on each plane and so lowers the strength of the rock and the resistance to frictional sliding. Increasing the pore pressure by pumping water down a deep well has produced earthquakes.[9]

Tentative Conclusions

The tentative conclusions reached are that the tectonic shear stresses in the crust are many times greater than the average shear stress drop during an earthquake, rather than comparable as most often proposed. However, the motion just before, during, and just after a major strike-slip earthquake occurs is constrained by the lower 3/4 or so of the crust which does not slip appreciably in this time interval. It is helpful to think of the driving force for the relative motion as the somewhat spread out screw-type dislocation displacement pattern of the lower and middle regions of the crust, which builds up incrementally over 100 years, rather than the high tectonic shear stress, which always is present, Fig. 1. Based on laboratory experiments[3], initiation of slip does require consideration of the high tectonic shear stress as does the continuing slip in the presence of friction in the upper region of the crust.

Much of the strike-slip fault length and area in the upper as well as the middle and lower crust does slip or creep steadily or sporadically with little or no noticeable seismic activity.[10] Regions in which large earthquakes occur are regions in which relative motion is impeded or hung-up over large or

small areas of the fault. The resulting stress concentration in these pinning regions eventually becomes high enough to initiate sliding. Unless melting intervenes, sliding proceeds at an appreciable frictional stress but a stress below the breakaway stress. Propagation ceases when the boundary of the surface of sliding spreads out far enough into regions of lower shear stress or into regions of higher friction. The area of the newly slipped region of the fault may range from negligibly small to extremely large, most often not large enough to break through to the surface of the earth, but at the other extreme extending hundreds of kilometers along the fault.

With the incremental elastic displacement pattern in the middle portion of the crust controlling the relative motion across the fault in the upper portion of the crust, there still may be appreciable non-uniformity over local regions of the slipped fault. Aftershocks to equalize the relative motion across the fault may go either forward or backward, therefore, but the clear preference is for a continuation of the slip as local regions which barely hang on during the earthquake then release their hold, as deep regions creep rapidly, and the elastic tectonic shear reasserts itself. Dilation, pore pressure drop, and inflow of water to again raise the pore pressure would account for the sticking and then slipping of local regions.[7b,11]

Large effects of dilatancy on the pressure of water in saturated rock, or in open areas between joints in saturated

-51-

rock masses, can be explained by large nonlinear changes of crack or gap volume with small changes in applied stress. Hertz contact stress and displacement theory or block rotations make for repeatable volume changes from one earthquake to the next.

Fig. 2 provides a possible explanation for the detection of a large dilatancy effect over appreciable regions prior to an associated thrust fault earthquake when such effects would not show up transverse to a strike-slip one. Much larger transverse horizontal dimensions, much larger volumes of rock are shown to be disturbed in the thrust fault associated with an offset in the dominant strike-slip system of faults. A correlation of time delay between dilatation and earthquake with depth of focus of the earthquake is worth exploring along with present explanations based on the inflow of water from the boundary of the dilated region to its center.[5b] Conditions for localization of deformation in a band are, in general, strongly dependent on the ratios of principal stresses (plane strain vs. axially symmetric compression, etc.) and this too may contribute to the difference between strike-slip and thrust faults, the latter possibly being subjected to stress states consistent with extensive dilatant deformation before localization occurs.

The cyclic nature of earthquakes, or more accurately the the repetition of events in time, is described but not explained by the build-up and release of 30 bars of shear stress. A slow and partial chemical healing (sintering or cement-like reaction)

of fractured or jointed rock by water in the upper crust under rather high pressure but moderate heat might explain the 30 bar barrier and release. At the same time, the presence of water weakens the rock and permits fresh fracturing to take place at much lower applied stress than in dry rock.[12]

Suggested Experiments and Analysis

Several suggestions for field measurements follow from the qualitative picture of the dominantly strike-slip fault system typified by California. Probably most important from the overall geophysical viewpoint is the determination of the actual tectonic stresses in the region. Shear stress, or principal stress difference, in particular should be determined and then changes monitored at a sufficient number of locations in depth and plan along and away from the San Andreas. Techniques employed in mining research for the determination of tectonic stresses far ahead of the mine face should be applicable. Each location should be at ample depth to avoid surface perturbations, some should be several kilometers down. At the same time, water pressure should be monitored in the holes which will be drilled. The present uncertainty about the average shear stress level is a great detriment to a clear analysis of the cause and effect of earthquakes. Moderate accuracy would be sufficient to distinguish between 20 bars and 1000 bars. Changes in stress level with time are much easier to determine and should be correlated with observed relative motion along the fault.

-53-

Similarly, the dilatancy question should be explored by deter-
mining V_p and V_s in a traverse of lengths parallel to and
neighboring the San Andreas fault in addition to transverse to
it. If strike-slip is localized while the associated thrust
fault regions needed to match strike-slip on parallel fault
systems are diffuse, the region adjacent to the San Andreas fault
should show cycles of dilatancy over lengths parallel to the
fault, cycles whose effect would be lost on the average if measure-
ments are made transverse to the fault. However, it is possible
that the dilatancy near strike-slip faults is sufficient to re-
duce the pore water pressure appreciably and strengthen the
region temporarily, yet not be large enough to cause vaporization.
If so, the wave speeds would not be affected but pore water
pressure measurements would show a large variation with time.

Laboratory experiments on both rock healing and the
equivalent of stress corrosion due to water under conditions of
pressure and temperature corresponding to 5 km depth would be
extremely valuable but would require considerable time. They
would be especially worthwhile if accompanied by creep experi-
ments in rocks at temperatures and pressures corresponding to
points in the lower and middle regions of the crust as well as
the upper portion. Such experiments would permit needed improve-
ment and quantification of the deformation pattern of Fig. 1.

Analytical studies of a variety of topics are needed.
Appropriate constitutive relations should be developed for a

-54-

dilatant rock mass in a field of variable stress, temperature, and water pressure. Calculations then could be made of the stability of the rock mass and the localization of failure under increasing compressive or shear stress. Permeability as a function of stress and time at temperature would be intimately related to or form part of the constitutive relations. Fundamental assessment then would be possible of the observation and hypothesis that a drop in V_p/V_s due to dilatancy provides quantitative information that a thrust fault earthquake will occur in a strike-slip dominated region. The calculation of the effect of viscous drag of the mantle on the crust and the viscoelastic response of the lower and middle portions of the crust would provide a beginning for earthquake prediction from material properties. Analytic description, in fracture mechanics terms, of the properties of the surface of slip and the dislocations or cracks moving along this surface would come next. In particular, it is necessary to clarify the picture of stress drop, to distinguish among very different possibilities which lead to very different conclusions about prediction and modification of earthquakes. At one extreme there is the possible actual drop in stress of 30 bars or so over large areas of sliding, many kilometers in extent. The other extreme, which might correspond to melting of the rock on the surface of sliding, is to think in terms of a large (kilobar) drop in stress everywhere sliding is taking place but to suppose only a small fraction

of the total region of slip to be sliding at any one time. Intermediate situations of portions of the total region sliding with rather high friction but appreciable stress drop are equally consistent with the usual analysis of seismic signals. Earth-quake modification procedures are likely to be very sensitive to actual stress drop and the extent of the areas of sliding. A more detailed analysis of existing seismic records, both near and far field, may provide much more information than has yet been brought out.

Acknowledgement

References

1a. Frank D. Stacey, *Physics of the Earth*, John Wiley & Sons (1969).

1b. George D. Garland, *Introduction to Geophysics, Mantle, Core and Crust*, W. B. Saunders Co. (1971).

2a. James N. Brune, "Tectonic Stress and the Spectra of Seismic Shear Waves from Earthquakes," J. Geophys. Res. 75, (1970) pp. 4997-5009.

2b. K. Aki, "Earthquake Mechanism," in A. R. Ritsema (Ed.) "The Upper Mantle," Tectonophysics 13 (1-4), (1972), pp. 423-446.

3a. W. F. Brace, "Laboratory Studies of Stick-Slip and Their Application to Earthquakes," in E. F. Savarensky and T. Rikitake (Eds.) "Forerunners of Strong Earthquakes," Tectonophysics 14, (1972), pp. 189-200.

3b. W. F. Brace and D. K. Riley," Static Uniaxial Deformation of 15 Rocks to 30 kb," Int. J. Rock Mech. Min. Sci. 9, (1972), pp. 271-288.

4. C. B. Raleigh, J. H. Healy, and J. D. Bredehoeft, "Faulting and Crustal Stress at Rangely, Colorado," in Geophysical Monograph Series Vol. 16, Flow and Fracture of Rocks (eds. H. C. Heard et al.), Amer. Geophys. Union, Washington, 1972, pp. 275-284.

5a. Y. P. Aggarwal, L. R. Sykes, J. Armbuster, and M. L. Sbar, "Premonitory Changes in Seismic Velocities and Prediction of Earthquakes," Nature 241, Jan. 1973, pp. 101-104.

5b. J. H. Whitcomb, J. D. Garmany and Don L. Anderson, "Earthquake Prediction Variation of Seismic Velocities before the San Fernando Earthquake," Science 180, (1973), pp. 632-635.

5c. C. H. Scholz, L. R. Sykes, and Y. P. Aggarwal, "The Physical Basis for Earthquake Prediction," Lamont-Doherty Geological Observatory Contribution No. 1939.

5d. Don L. Anderson and James H. Whitcomb, "Time Dependent Seismology," Contribution No. 2363, Division of Geological and Planetary Sciences, Cal. Inst. Tech. 1973.

6a. Amos Nur and Gene Simmons, "The Effect of Saturation on Velocity in Low Porosity Rocks," Earth and Planetary Science Letters 7, (1969) pp. 183-193.

6b.   Amos Nur, "Dilatancy, Pore Fluids and Premonitory Variations of t_s/t_p Travel Times," Bull. Seism. Soc. Am. <u>62</u>, (1972), pp. 1217-1222.

7a.   F. C. Frank, "On Dilatancy in Relation to Seismic Sources," Reviews of Geophysics <u>3</u>, (1965), pp. 485-503.

7b.   W. F. Brace and R. J. Martin, III, "A Test of the Law of Effective Stress for Crystalline Rocks of Low Porosity," Int. J. Rock Mech. Min. Sci. <u>5</u>, 1968, pp. 415-426.

8.   W. F. Brace, B. W. Paulding, Jr., and C. Scholz, "Dilatancy in the Fracture of Crystalline Rocks," J. Geophys. Res. <u>71</u>, (1966), pp. 3939-3953.

9.   C. B. Raleigh, J. H. Healy, J. D. Bredehoeft and J. P. Bohn, "Earthquake Control at Rangely, Colorado," Trans. Am. Geophys. Union <u>52</u>, (1971), pp. 344-

10.   Amos Nur, "Tectonophysics: The Study of Relations Between Deformation and Forces in the Earth," Review Article written in 1973, to appear.

11.   A. Nur and J. R. Booker, "Aftershocks Caused by Pore Fluid Flow," Science <u>175</u>, Feb. 1972, pp. 885-887.

12a. David Griggs, "Hydrolytic Weakening of Quartz and Other Silicates," Geophys. J. R. astr. Soc. <u>14</u>, (1967), pp. 19-31.

12b. C. H. Scholz, "Static Fatigue of Quartz," J. Geophys. Res. <u>77</u> (11), 1972, pp. 2104-2114.

SURFACE

UPPER CRUST

MIDDLE CRUST

LOWER CRUST

MANTLE

Elastic shear strain in upper crust due to overall tectonic shear stress is of order 10X greater than the incremental shear strain

2m

1km

a
b
c
d

e (surface after earthquake)

e

d (surface)
c
b

a
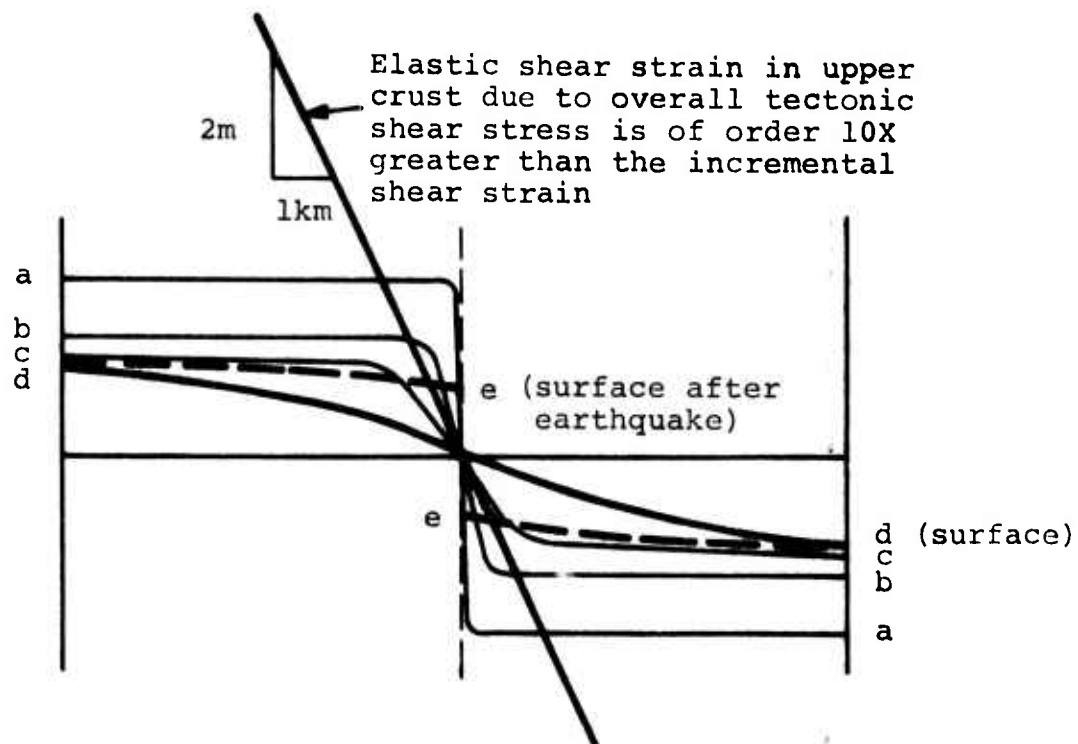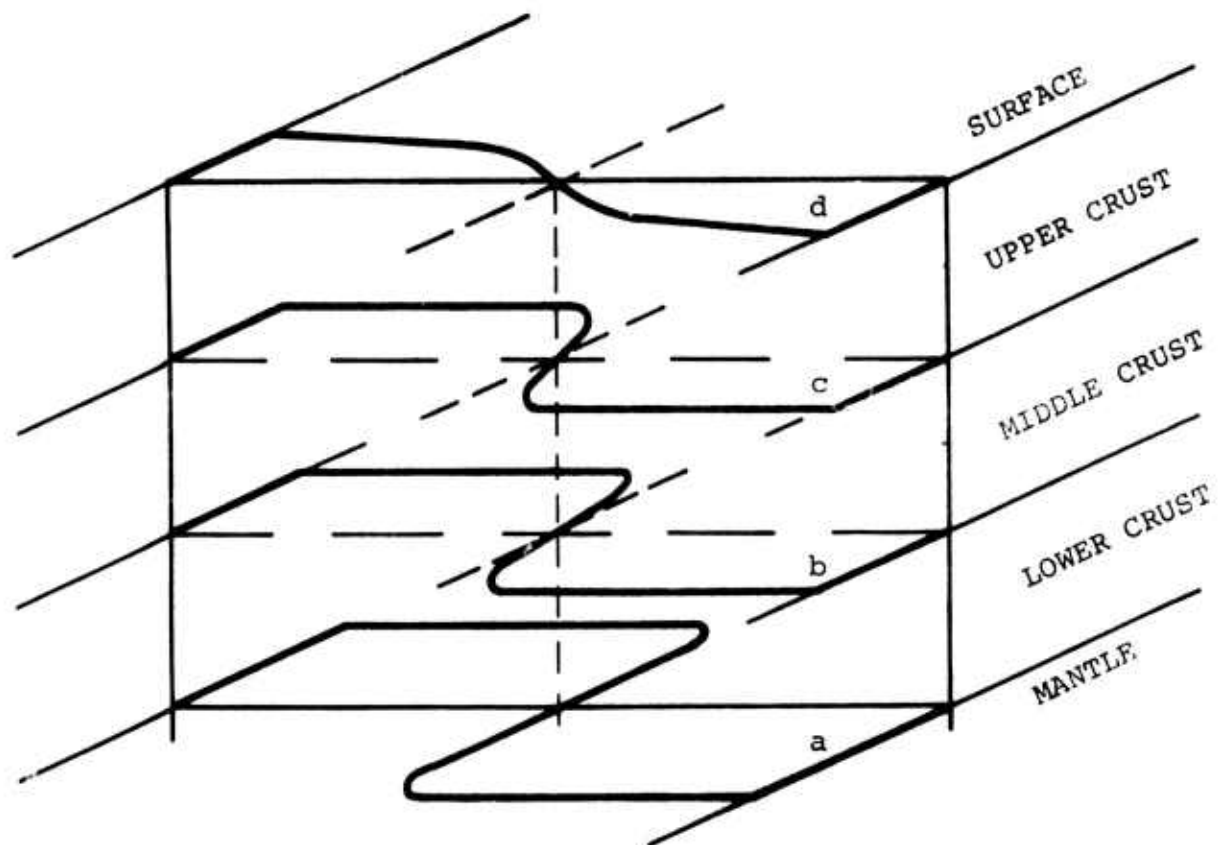
Figure 1.   Incremental displacement picture for strike-slip in any one region from one earthquake to the next.
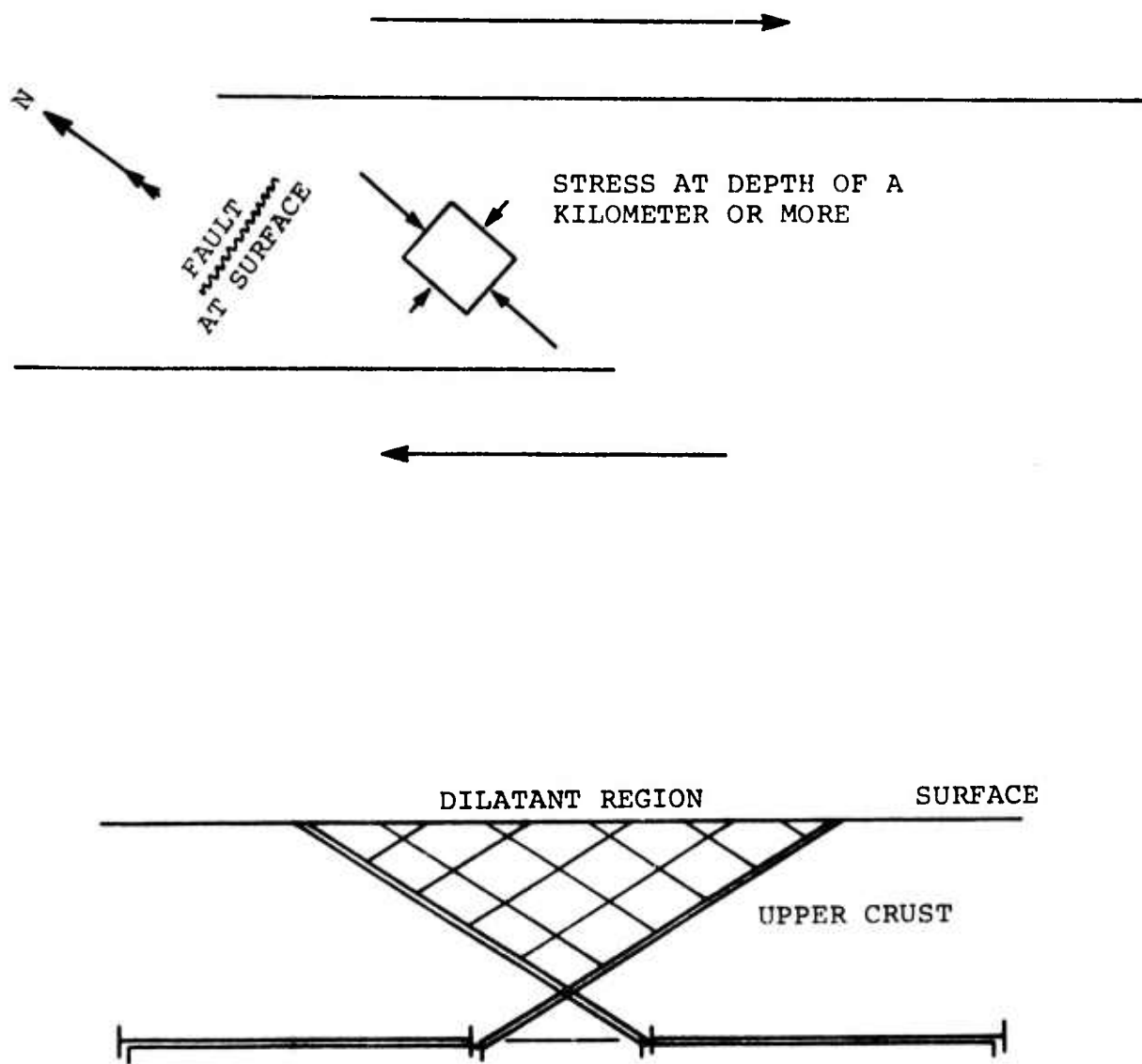
Figure 2.   An associated thrust fault in a strike-slip
            dominated region may have a large dilatant
            region prior to seismic slip.

-60-

THEORY OF SOLUTION STRENGTHENING
OF ALKALI HALIDE CRYSTALS

J. J. Gilman

Abstract

The solution-strengthening of sodium and potassium halide crystals by dissolved alkaline earth atoms is accounted for quantitatively. Strengthening is associated with the change in electrostatic energy that occurs when a divalent ion-vacancy complex is sheared by a dislocation that passes through it. The model yields a hardening coefficient of 11×10^9 d/cm^2; compared with the experimental value of 9.2×10^9 d/cm^2.

THEORY OF SOLUTION STRENGTHENING
OF ALKALI HALIDE CRYSTALS

J. J. Gilman

I.   Introduction

Recent experiments by Chin et al.[1] have shown that crystals of NaCl, Br and KCl, Br are strongly hardened by small additions of divalent ions (Ca^{2+}, Sr^{2+}, and Ba^{2+}). In the solution treated condition the yield strengths, σ_y of their crystals depended only on the concentrations of the additions and not upon the divalent ion species, or upon the host species (see Figure 1). They also found that the strengthening is proportional $C^{\frac{1}{2}}$ where C is the atomic concentration as demonstrated by Figure 1 and expressed in the equation:

$$\sigma_y (10^8 \ d/cm^2) = 0.15 + 92 \ C^{\frac{1}{2}} \tag{1}$$

This result appears to be consistent with the results of previous work as reviewed by Chin et al.[1] Therefore, the problem of a theory is to account for the coefficient of the concentration in terms that are independent of the addition and host species.

The model of Fleischer[2] that is currently used to account for divalent ion hardening gives the following expression for the yield stress:

-62-

$$\tau_y = 1/3 \ G \ \Delta\varepsilon \ C^{\frac{1}{2}} \tag{2}$$

where G is the appropriate shear modulus of the medium, C is the
atomic concentration of divalent ions, and $\Delta\varepsilon = \varepsilon_1 - \varepsilon_2$ where
ε_1 and ε_2 are the principal strains of a tetragonal distortion.

In the present context, this theory has certain diffi-
culties which can be seen by comparing the data listed in Tables
I and II with the experimental observations of Chin et al.[1] In
Table I some properties of the host salts are listed plus the
average value and the range of variation divided by the average.
The variation of the lattice parameter, d_o and of the dielectric
constant, K are small consistent with the lack of variation of
the solid solution hardening. However, the variation of the
elastic shear stiffness, C_{44} is large which should lead to large
differences in hardening for a given species of divalent ion
addition according to Equation (2). Also, in Table II both the
ionic radii and the ionic polarizabilities show large variations
among the three species, and therefore should produce different
values of $\Delta\varepsilon$ for a given host. Yet these species are all ob-
served to have approximately the same effect on the hardnesses
of the four host crystals.

Another deficiency of the Fleischer theory is that it is
not absolute because $\Delta\varepsilon$ is treated as a parameter to be obtained
from experiment or from some other theory.

The invariant feature of the hardening is simply the ionic
charge difference between the host and addition ions. The present

model is based on this factor and the calculation is carried out without introducing a disposable parameter.

II.   Hardening by Divalent Complexes

As a result of the work of Dryden, Morimoto and Cook[3], it is known that for the solutions under discussion, the divalent cations are associated with cation vacancies as shown schematically in Figure 2. For the glide planes that pass through the structure, the interaction between the divalent cation and its associated cation-vacancy presents a strong local impediment to dislocation motion. Since it is localized, the force required to shear the complex cannot be spread out by the dislocation core structure, so the required shearing stress approximately equals the cohesive shear strength of the complex.

The form of Equation (1) is:

$$\sigma_y = A + B \; C^{\frac{1}{2}} \tag{3}$$

where $A \ll B$, so when C becomes unity[4], $\sigma_y = B$. Hence, B is simply twice the cohesive shear strength of the complex (since the resolved shear stress, $\tau = \sigma_y/2$). This can be calculated from electrostatic theory as follows.

III.   Shear Strength of Complex

Before a dislocation with Burgers displacement, b passes through the complex, the distance between the two charge centers is: $d = b = a_o/\sqrt{2}$); and afterwards it becomes $\sqrt{2} \; b = a_o$. Therefore, if K is the static dielectric constant and e = electron

-64-

charge, the energy difference between the initial and final states is (since each charge center has unit excess charge):

$$\Delta U = \frac{e^2}{Ka_o} (\sqrt{2} - 1) \tag{4}$$

The force on the complex caused by a shear stress, τ is τb per unit length, or $\tau b(a_o/2)$ on the segment that shears the complex. Thus the work done, W is:

$$W = 1/2 \ \tau b^2 a_o = \frac{1}{4} \ \tau a_o^3 \tag{5}$$

Equating the work done and the energy increase yields:

$$\tau = \frac{1.66 \ e^2}{K \ a_o^4} \tag{6}$$

But two glide planes pass through the complex so the effective concentration is twice the divalent ion concentration. This introduces a factor of $\sqrt{2}$, so the effective shear strength is:

$$\tau_{eff} = \frac{2.34 \ e^2}{K \ a_o^4} \tag{7}$$

which yields the following hardening coefficient:

$$B = \frac{4.7 \ e^2}{K \ a_o^4} \tag{8}$$

From Table I, the average values of K and a_o are 5.29 and 6.13 $\overset{o}{A}$, respectively which yield a numerical value:

$$B \simeq 11 \times 10^9 \ d/cm^2 \tag{9}$$

compared with the experimental value:

$$B_{exp} = 9.2 \times 10^9 \ d/cm^2 \tag{10}$$

IV.   <u>Summary</u>

A simple theory is presented that accounts quantitatively for the solid-solution hardening of monovalent salt crystals by divalent impurities. It is based on the change in electrostatic energy that occurs when a divalent ion-cation vacancy complex is sheared by a dislocation passing through it.

The solution hardening coefficient that is given by the theory differs from the observed value by only 16%. Since no disposable parameters are used, this good agreement gives credence to the model.

<u>Acknowledgement</u>

References

1. G. Y. Chin, L. G. VanUitert, M. L. Green, G. J. Zydzik, and T. Y. Kometani, J. Amer. Cer. Soc. __56__, 369 (1973).

2. R. L. Fleischer, Acta Met. __10__, 835 (1962); J. Appl. Phys. __33__, 12 (1962).

3. J. S. Dryden, S. Morimoto and J. S. Cook, Phil. Mag. __10__, 379 (1965).

4. Unit concentration yields pure complex which is an imaginary substance, so this is a hypothetical extrapolation.

TABLE I - SOME PROPERTIES OF THE HOST SALTS

COMPOUND	$C_{44} (10^{10}$ $^d/cm^2)$	d_o(A)	K_{static}	IONIC REFRACTION
NaCl	13	5.64	5.68	9.2
NaBr	11	5.97	5.99	12.7
KCl	6.5	6.29	4.70	10.9
KBr	5.2	6.60	4.78	14.2
AVERAGE	8.93	6.13	5.29	11.8
$\Delta x/<x>$	0.87	0.16	0.24	0.43

TABLE II - SOME PROPERTIES OF THE DIVALENT ADDITION IONS

ION	RADIUS (PAULING-A°)	IONIC REFRACTION (POLARIZABILITY)
Ca^{2+}	0.99	1.35
Sr^{2+}	1.13	2.30
Ba^{2+}	1.35	4.30
AVERAGE	1.16	2.65
$\Delta x/<x>$	0.31	1.11
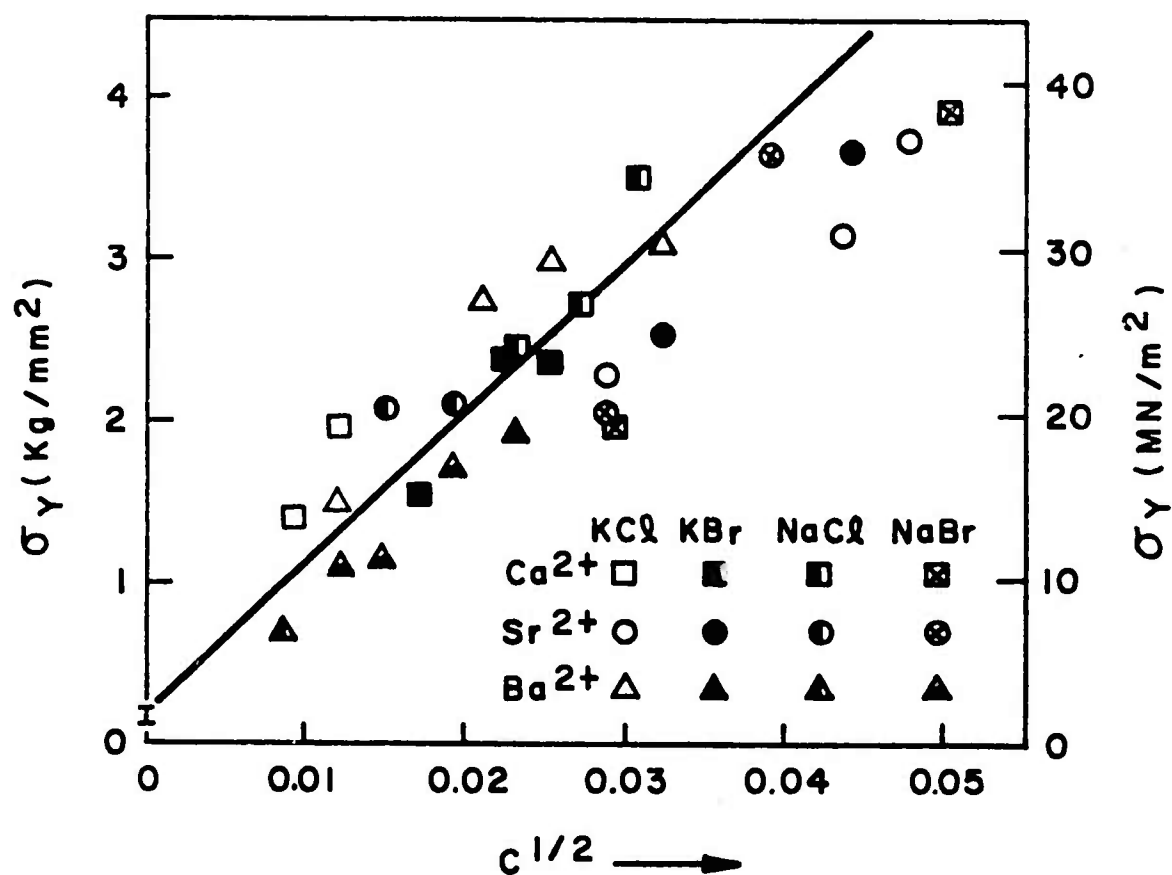
FIGURE 1 — DEPENDENCE OF YIELD STRESS ON
SQUARE—ROOT OF DIVALENT ION—
CONCENTRATION FOR VARIOUS
ALKALI HALIDE CRYSTALS.

GLIDE PLANES
PASSING THROUGH
COMPLEX.
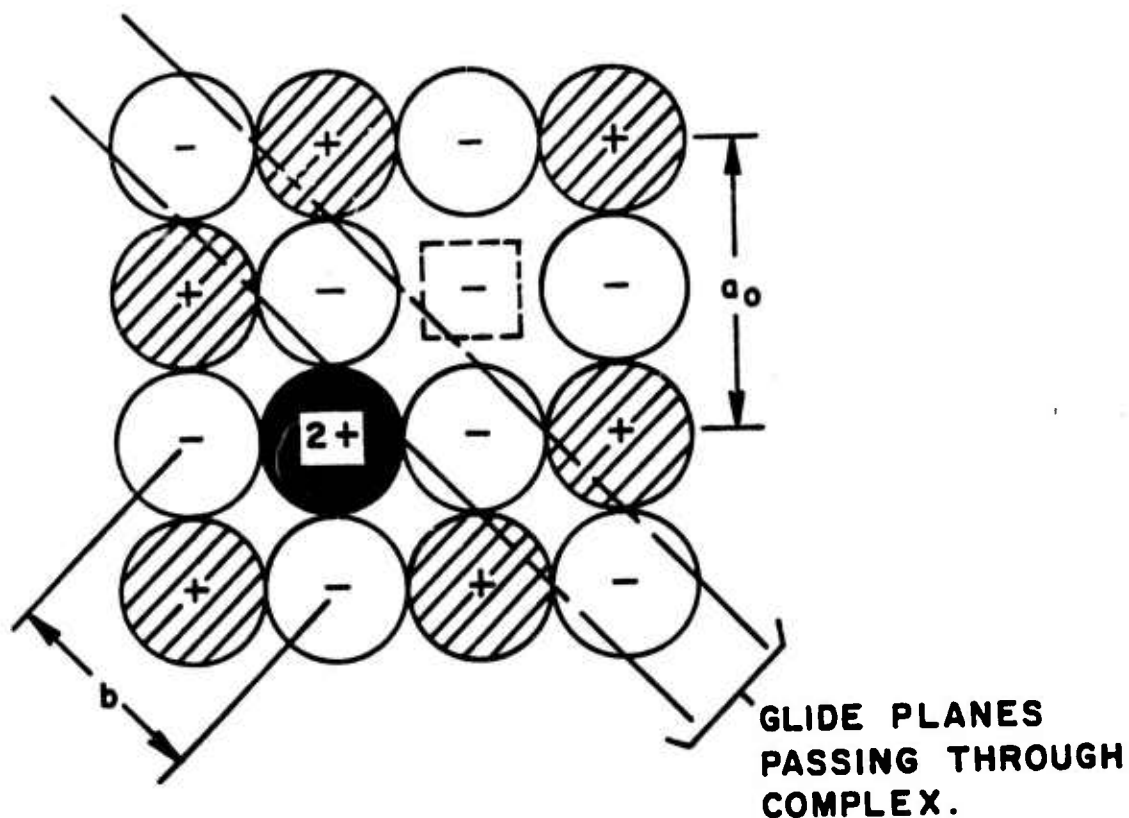
FIGURE 2 — DIVALENT ION — CATION VACANCY COMPLEX
SHOWN SCHEMATICALLY ON THE [100]
PLANE OF THE ROCK—SALT STRUCTURE

MICROWAVE CONDUCTIVITY MEASUREMENTS ON
ANISOTROPIC ORGANIC CRYSTALS

P. L. Richards

Abstract

A microwave technique is described for measuring the
electrical conductivity of anisotropic organic materials.
This technique is useful in the high conductivity (skin effect)
limit which has been encountered in TTF-TCNQ. The sample is
placed in the center conductor of a coaxial resonator so that
a large microwave current flows through it even when the skin
depth is small. Methods for cavity calibration and data
analysis are described.

MICROWAVE CONDUCTIVITY MEASUREMENTS ON
ANISOTROPIC ORGANIC CRYSTALS

P. L. Richards

Coleman et al.[1] have reported electrical conductivities
which are comparable to that of room temperature copper in
several samples of the organic salt TTF-TCNQ. The large con-
ductivity is observed only in a narrow temperature range and in
only one direction in these highly anisotropic crystals. The
possibility that this high conductivity arises from the onset
of superconducting fluctuations has aroused considerable inter-
est. The argument in favor of superconductivity is based on
the very large conductivity compared with that expected from
the carrier density and relaxation frequency.

The conductivity measurements made by Coleman et al.
used a dc four terminal technique which is subject to serious
errors in strongly anisotropic materials. Microwave measure-
ments of conductivity and dielectric constant can in principle
be made without electrical contact to the sample. Such measure-
ments on low conductivity TCNQ's are described by Shchegolov[2]
and by Bloch et al.[3] The technique used for these experiments
is to measure the perturbation of the resonant frequency and
the Q of a cavity resonator when a sample is introduced which
is short compared with the microwave wavelength λ and thin

compared with the microwave skin depth δ. Attempts to use
this technique on the high conductivity samples reported by
Coleman et al. failed because the skin depth at microwave
frequencies $\delta \leq 10^{-4}$ cm becomes so small that a sample thin
compared with δ does not have enough volume to perturb the
resonance by a measurable amount.

A number of microwave techniques have been developed
to measure the absorption in metals in the skin depth limit.
One of these appears to be very convenient for measuring
organic crystals which grow in the form of thin rods oriented
along the high conductivity direction. This is the co-axial
resonator technique used by Pippard to study the anomalous
skin effect in normal metals and the penetration depth in super-
conductors.[4,5] The resonator, which is shown in Fig. 1, is a
section of a TEM mode coaxial transmission line with an open
circuit at each end. The outer conductor extends somewhat be-
yond the ends of the inner conductor and is closed to prevent
radiation loss. The wavelength of the fundamental resonance
is slightly longer than twice the length of the inner conductor
because the spreading of the E-field at the ends introduces
extra capacitance.

The ratio of the power absorbed in the center conductor
(of radius r_1 and surface resistance R_1) to that absorbed in the
outer conductor with the same σ (r_2 and R_2) is approximately
$r_2 R_1 / r_1 R_2$, since the same total current flows in each conductor

in the TEM mode. For this reason the center conductor can dominate the cavity Q even if $R_1 \ll R_2$.

If a sample of high conductivity TTF-TCNQ is used as the center conductor in the above cavity, then the microwave conductivity can be obtained directly from the uncoupled cavity Q. Although this simple procedure can be used under favorable conditions, it has two disadvantages. One is that samples will not usually be available in the proper length to match the chosen microwave frequency range. The second is that the microwave current is a maximum half way along the center conductor and falls to zero at the ends. Consequently, there is some current flow transverse to the sample axis, and the perpendicular conductivity $\sigma_\perp$ will contribute to the power dissipation. We can estimate this effect with a simple calculation. Assume that the microwave surface current is uniform to a depth δ and that it flows parallel to the sample axis over the central half of the sample. The electric field perpendicular to the sample surface near each end implies a surface charge in those regions. If we assume that the microwave current is moving toward (or away from) the surface over the remainder of the sample, the ratio of the power dissipated by the radial currents near the ends to the power dissipated by the axial currents in the central half of the sample is

$$\frac{\delta}{\sigma_\perp \ 2\pi r_1 (\lambda/4)} \ \frac{(\lambda/4)}{\sigma_\parallel \ 2\pi r_1 \delta} = \frac{16 \ \sigma_\parallel \ \delta^2}{\sigma_\perp \ \lambda^2} \quad .$$

-74-

Although the fractional loss due to $\sigma_\perp$ is negligibly small under most circumstances, it is best to avoid this effect as far as possible.

Both problems can be aleviated by extending the sample with metal antennas as is shown in Fig. 1b. The length of the antennas should be cut to obtain the desired resonant frequency. They should be placed on both ends so that the sample is in a region of nearly uniform current. The antennas need not make electrical contact to the sample since a small extra capacitance will have little effect on the measurement. The ends of the sample and the ends of the antennas in contact with the sample should, in so far as possible, be flat, parallel, and equal in diameter so as to minimize fringing fields which would introduce losses related to σ_1. As is shown in Fig. 1b, the three parts of the center conductor can be held in place by inserting them into a capillary of fused silica. The capillary should have thin walls so as to minimize dielectric loss in the regions of strong E-field near the ends of the inner conductor. Capillaries of the desired size can easily be drawn down from larger tubing.

Several factors interact in the choice of the radius r_2 of the outer conductor. For any given sample size and surface resistance, there is a range of sizes which correspond to convenient values of the microwave Q. This Q can be estimated from standard transmission line theory to be

$$Q = \frac{L}{R} \simeq \frac{4\pi\omega \; r_1 r_2}{c^2 \; r_2 R_1 + r_1 R_2} \; \ell n \; (\frac{r_2}{r_1}) \quad .$$

A copper X-band cavity with $r_1 = 0.05$ mm and $r_2 = 5$ mm should have $Q \simeq 700$. In this case 99 percent of the loss occurs in the center conductor.

The coaxial resonator technique discussed here is only useful for high conductivity samples which are in the skin depth limit. For lossy materials for which $r_1 R_2$ can be neglected compared with $r_2 R_1 = r_2 / \delta_1 \sigma_1$ we have

$$Q = \frac{2r_1}{\delta_1} \ln\left(\frac{r_2}{\eta}\right) \ .$$

If δ_1 becomes comparable with r_1 then the Q is of order unity. Although modes other than the TEM could in principle be used to measure low conductivity or even dielectric samples, the analysis would be complicated and there would be no advantage over the conventional small sample perturbation technique.[2] This conventional technique should therefore be used for survey purposes, and the coaxial resonator employed only if very high conductivity is indicated.

The radius of the outer conductor should be chosen small enough to avoid any propagating waveguide modes close to its frequency. This condition is satisfied if $\lambda > 2.62 \ r_2$ for the TM_{01}. Allowing for perturbations arising from the center conductor, $\lambda \gtrsim 5 \ r_2$ should be safe. The ends of the can should be far enough from the sample to avoid excessive spreading resistance, and to avoid dependence of the resonant frequency on the axial position of the sample.

-76-

Depending on the frequency employed, the coupling to the cavity can be accomplished by H-field loops or by E-field stubs as are shown in Fig. 1. Since the coupling is a function of the axial position of the sample relative to the coupling structures, it is convenient to suspend the sample in such a way that this position can be varied during the measurement as is shown in Fig. 1b. Although coupling losses can be calibrated by using a sample of known surface resistance, it is difficult to position different samples precisely enough that the coupling is the same for each. It is preferable to measure the uncoupled cavity Q_o each time a sample is inserted. This can be done by plotting the coupled Q as a function of the cavity transmission coefficient t and then extrapolating to t = 0 to find Q_o. For symmetrical structures this is a simple linear extrapolation.[4] The dielectric loss as well as the loss in the antennas and the outer cavity walls must be subtracted from the uncoupled cavity loss before the sample surface resistance R_S can be obtained from Q_o. This can be done by observing that $\ell_S R_S / r_S$ is a linear function of $1/Q_o$, where ℓ_S and r_S are the length and radius of the sample. The two constants in the linear function can be elvaluated by measuring Q_o for copper wires with two different known values of $\ell_S R_S / r_S$. The surface resistance of copper near room temperature can be computed from the classical skin effect theory to be $R = 1/\sigma\delta = \sqrt{2\pi\omega/\sigma c^2}$. The calibration wire (s) should be carefully electropolished to insure that the surface conductivity is characteristic of the bulk metal. Low temperatures

-77-

can be used to obtain two values of R_S from a single calibration sample if $\sigma(T)$ is first measured. If the computed low temperature skin depth becomes comparable with the mean free path in the copper, then the anomalous skin effect must be used to compute R_S from σ. Such effects become important by 77K for OFHC copper at typical microwave frequencies.[6] A particularly accurate calibration of the residual cavity and antenna losses can be obtained by using a calibration wire which becomes superconducting and by measuring Q_O at room temperature and at a temperature $T < 0.5\ T_c$ where the wire is essentially lossless. Since the loss in a sample with length ℓ_S is expected to vary as $\ell_S R_S/r_S$, it should be possible to scale a single calibration to samples of slightly different lengths. It is well to keep in mind that losses in the antennas and the remainder of the resonator can have significant temperature dependence if they are made from pure metals. Alloys such as brass might be useful for these parts if adequate Q can be maintained.

The classical skin effect theory with $\omega\tau \ll 1$ will usually be valid for high conductivity organic crystals. The conductivity $\sigma_{||} = 2\pi\omega/c^2 R_S^2$ can then be easily obtained from the measured sample surface resistance R_S. A test of the validity of this procedure can be made if changes in resonant frequency as well as Q are measured as a function of temperature.

The effective capacitance C_O of the resonator is essentially independent of the skin depth δ because the depth of the surface charge layer associated with the E-fields normal

to the conducting surfaces is governed by the much smaller electrostatic screening length. The tangential magnetic fields, however, penetrate a distance δ, so changes in the resonant frequency $\omega_0 = (L_0 C_0)^{-\frac{1}{2}}$ arise through the effective inductance L_0.

If we define the surface impedance in the usual way in terms of the tangential electric and magnetic fields at the surface $Z = R+iX = 4\pi E/cH$, then a simple theorem relates the change in resonant frequency $\Delta\omega_0$ associated with a change ΔX_S in the surface reactance of the sample, to the change in the width of the resonance at the half power points $\Delta(\omega_1-\omega_2)$ associated with a change ΔR_S in the surface resistance of the sample:[5]

$$\frac{\Delta X_S}{\Delta R_S} = \frac{2\Delta\omega_0}{\Delta(\omega_1-\omega_2)} \quad .$$

If care is taken to eliminate other sources of frequency shift (or changes in Q), this theorem can be used to obtain an experimental value of $\Delta X_S/\Delta R$. Alternatively, the relationship between $\Delta\omega_0$ and ΔX_S can be obtained like that between $\omega_1-\omega_2$ and R_S from the cavity calibration procedure discussed above. In the classical skin effect theory $R = X = 2\pi\omega\delta$. A measured value of $\Delta X_S \neq \Delta R_S$ is a sign that the classical theory is not valid.

Such an effect could arise from the onset of super-conducting diamagnetism. In this case the real conductivity becomes small for $T < T_c$ and $\omega_0 \ll 2\Delta$, but there is a large imaginary component which is proportional to $1/\omega^2$ and thus leads to a frequency independent reactive penetration depth. Pippard[6]

-79-

used the frequency shift method to measure this penetration depth in type I superconductors. Although this method can be used to detect the onset of superconductivity, dc SQUID magnetometers should be more sensitive.[7]

Another source of $\Delta X \neq \Delta R$ which could prove important might arise from a strong paraelectric or ferroelectric contribution to the imaginary conductivity (real dielectric constant). Although the coaxial resonator is not particularly useful in the large ε limit, it might be desirable to analyze the frequency shifts which mark the onset of such a state.

Two other sources of $\Delta X \neq \Delta R$ are much less likely to occur in organic conductors. If the relaxation frequency $1/\tau$ should become comparable with the microwave frequency or if the mean free path for electron motion perpendicular to the sample surface should become comparable to the skin depth, then the classical skin effect theory will break down. The theory of these relaxation and anomalous skin effects is well-known,[7] but the one-dimensional nature of the high conductivity in materials like TTF-TCNQ is likely to prevent the occurrence of either.

Acknowledgement

References

1. L. B. Coleman, M. J. Cohen, D. J. Sandman, F. G. Yamagishi,
    A. F. Garito and A. J. Heeger, Solid State Communications
    (submitted).

2. I. F. Shchegolev, Phys. Stat. Sol. (a)$\underline{12}$, 9 (1972).

3. A. N. Bloch, J. P. Ferraris, D. O. Cowan, and T. O. Poehler,
    Solid State Communications (submitted).

4. A. B. Pippard, Proc. Roy. Soc. (London) $\underline{A191}$, 370, 385, and
    399 (1947).

5. A. B. Pippard, Proc. Roy. Soc. (London) $\underline{A203}$, 98 (1950).

6. A. B. Pippard, Advanced in Electronics and Electron Physics
    (Academic Press, Inc., New York, 1954).

7. J. P. Gollub, M. R. Beasley, R. Callarotti, and M. Tinkham,
    Phys. Rev. $\underline{B7}$, 3039 (1973).

E-field
Current
X, ⊙ H-field

(a)

(b)
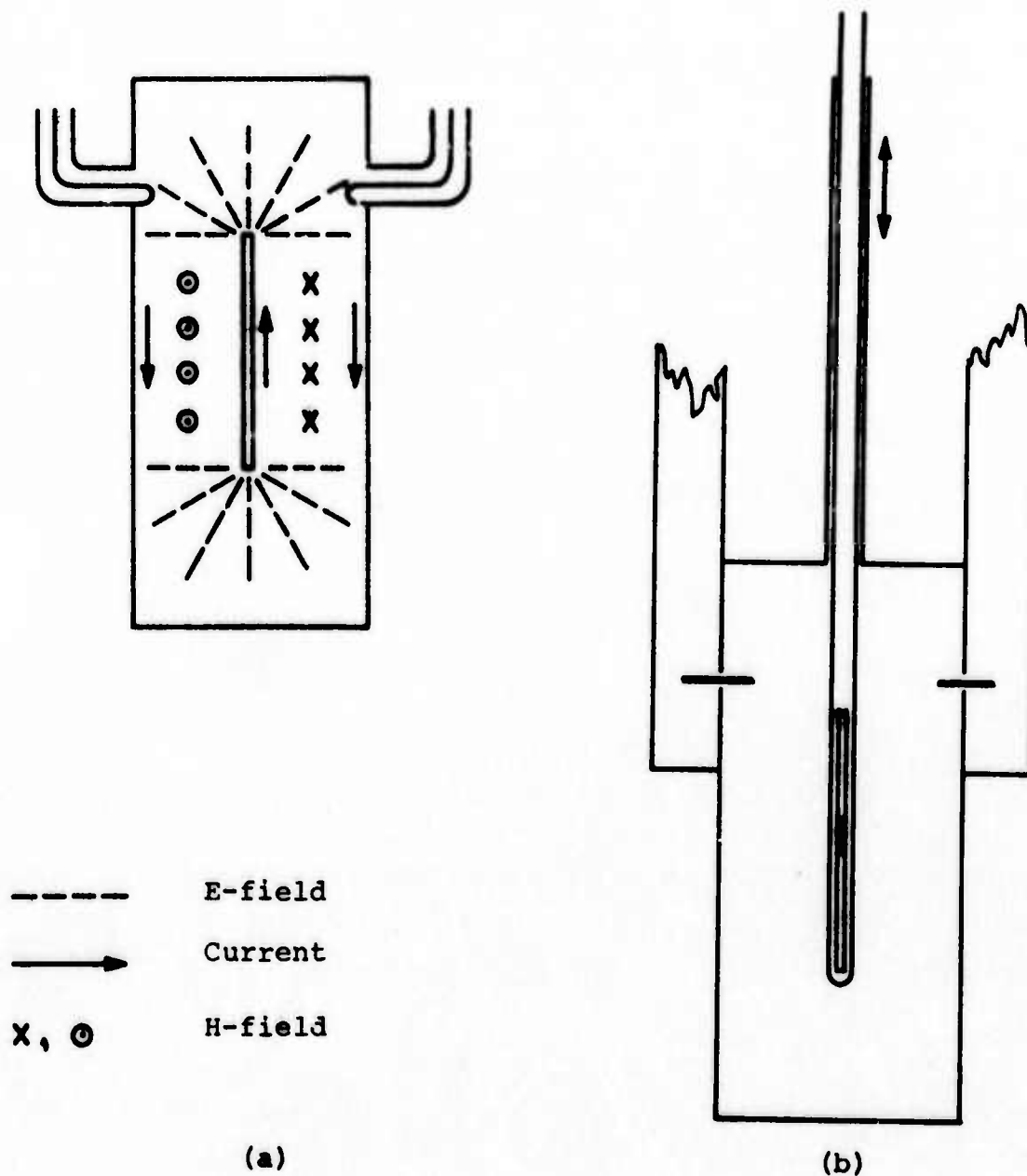
Figure 1(a). Cross section of TEM coaxial resonator showing
    mode pattern and placement of H-Field coupling loops fed
    with coaxial lines.

    (b). Cross section of TEM coaxial resonator showing
    E-Field coupling stubs fed with waveguides, and sample
    (with antennas) suspended in a fused silical tube.

DYNAMICS OF MAGNETIC SUSPENSIONS
FOR HIGH SPEED GROUND TRANSPORTATION

M. Tinkham and P. L. Richards

Abstract

This paper discusses the relation between the power
spectrum of track roughness and the subjective ride quality of
a vehicle traveling over the track. First, we discuss the
suspension dynamics with various combinations of springs and
dashpots to simulate classes of suspensions. A rather detailed
analysis is then given of the effect of a normal conducting
screening plate between superconducting suspension magnets and
the track. Such a plate both screens AC currents from the mag-
nets to reduce AC losses and damps the vertical motion. It is
estimated that a damping time of under one second can be obtained
by this passive means, and the high-frequency properties have
the favorable characteristics of series dashpot damping. A
quantitative discussion is given of the smoothing effect of the
finite length of the magnetic "pads", which average out the short
wavelength variations in track elevation. A somewhat similar
effect arises from the fall-off in subjective sensitivity to
accelerations at frequencies above the natural resonance frequen-
cies of the human body. After taking account of all these
smoothing effects and integrating over all frequencies, it appears

that there will be no serious problem in obtaining satisfactory subjective ride quality with a high-clearance superconducting magnetic suspension system. An analysis of low-clearance ferro-magnetic attractive suspension with feedback stabilization and secondary suspension is also given. This suggests that this system could also give satisfactory performance if the weight of the "magnetic bogie" can be kept small enough compared with the weight of the car.

DYNAMICS OF MAGNETIC SUSPENSIONS
FOR HIGH SPEED GROUND TRANSPORTATION

M. Tinkham and P. L. Richards

I.   Analysis of Ride Quality

A complete analysis fo ride quality in high speed ground
transportation system would entail consideration of all degrees
of freedom of the vehicle's motion under all applied forces, in-
cluding wind buffeting. For simplicity, we consider here only
the vertical motion (heave) of the train due to track roughness.
Similar analytical methods could be used for the other types of
motion and other random forces.

The purpose of this paper is tc work out the relation
between the power spectrum of track roughness $P_z(q)$ and ride
quality, which is defined in terms of the power spectrum of
vertical accelerations of the train $P_{\ddot{z}}(\omega)$. It is found that
finite pad effects resulting from the spatial extent of magnetic
or air cushion levitation can greatly reduce the train acceler-
ations. A subjective ride quality criterion is proposed and
used to quantify the comparison of various configurations.

Vehicle Dynamics

We first work out the relation between $P_{\ddot{z}}(\omega)$ and $P_z(q)$.
Note that the track height power spectrum $P_z(q)$ is referred to
spatial frequency q rather than to the _temporal_ frequency ω;

the two are related by $\omega = qv$, where v is the train velocity. We will consider several different suspension systems in order to illustrate the sensitivity of the predicted ride quality to the suspension design. These include the conventional parallel spring and dashpot illustrated in Fig. la, the series spring and dashpot of lb, and the series-parallel combination of lc. The results are expressed in terms of the height z of the center of mass of the train, a suitably averaged track height $\bar{z}$ and a samping factor $1/Q$ which is the fraction of the stored energy which is dissipated per radian at resonance. The expression for Q in terms of the resonant frequency and dashpot constant is different in each case. Critical damping occurs for $Q = 1/2$.

The suspension response is easily computed in each case by writing down the equation(s) of motion and assuming an $e^{i\omega t}$ variation for both $\bar{z}$ and z. For the parallel combination of Fig. la, we obtain

$$\frac{|z(\omega)|^2}{|\bar{z}(\omega)|^2} = \frac{\omega_0^{\,4} + \omega^2\alpha^2/M^2}{(\omega_0^{\,2} - \omega^2)^2 + \omega^2\alpha^2/M^2} \tag{1a}$$

where for this system $Q = M\omega_0/\alpha$. At the resonant frequency $\omega = \omega_0$ this ratio is $1 + Q^2$ so that a highly damped ($Q \lesssim 1$) suspension is desirable to minimize this resonant amplification of track roughness. For $\omega \gg \omega_0$ the displacement ratio is $\omega_0^{\,2}/Q^2\omega^2$, which is increased by decreasing the damping factor $1/Q$. Thus, increasing the damping effectively decreases the vibration at ω_0, but increases it at higher frequencies. Since

the sensitivity to high frequency vibration is reduced by finite

pad effects (discussed below), it is generally desirable to have

high damping $(Q \leq 1)$.

Studies[1,2] of suspension dynamics which seek to minimize

a linear combination of mean square suspension displacement and

mean square vehicle acceleration conclude that the optimum sus-

pension consists of the series dashpot and spring of Fig. 1b,

rather than the parallel dashpot and spring. Although this opti-

mum is deduced without consideration of finite pad effects, it

is a useful alternative system to consider for transportation

systems which employ active feedback damping. The displacement

response function in this case is

$$\frac{|z(\omega)|^2}{|\bar{Z}(\omega)|^2} = \frac{\omega_0^{\,4}}{(\omega_0^{\,2}-\omega^2)^2+\omega_0^{\,4}\omega^2 M^2/\alpha^2} \quad , \quad \omega \neq 0 \qquad (1b)$$

where $Q = \alpha/M\omega_0$. Equation (1b) is similar to (1a) at low fre-

quencies, but approached zero more rapidly for $\omega \gg \omega_0$. It is

this lack of response at high frequencies that favors the choice

of series dashpot and spring suspensions. As we will see, the

difference between the two types of suspensions is much less

pronounced when finite pad effects are included.

The series spring-dashpot suspension (1b) is not by it-

self satisfactory because it cannot support a static load. From

the equations of motion, the displacement response can be in-

finite at $\omega = 0$. Some additional component, such as the parallel

spring in Fig. 1c, is required to correct the low frequency

-87-

response.   In this case the response ratio can be written

$$\frac{|z(\omega)|^2}{|\bar{z}(\omega)|^2} = \frac{(\omega_1^2+\omega_0^2)^2+\omega_1^4\omega_0^4 M^2/\omega^2\alpha^2}{(\omega_0^2-\omega^2)^2+2(\omega_0^2-\omega^2)\omega_1^4+\omega_1^4+(\omega_1^2-\omega^2)^2 M^2\omega_0^4/\omega^2\alpha^2} \quad (1c)$$

If the parallel spring is chosen to be considerably softer than the series spring, then (1c) resembles (1b) except that the response is increased by a nearly constant factor.

Finite Pad Effects

The effective track height $\bar{z}$ in the above equation is an average over the length L of the vehicle.

$$\bar{z}(x) = \int_{-L/2}^{L/2} \bar{z}(x+x')g(x')dx' \quad , \quad (2)$$

where $g(x)$ is a suitable non-negative weighting function normalized so that $\int_{-L/2}^{L/2} g(x)dx = 1$.   For a magnetic suspension system, $g(x)$ is proportional to $B^2_{surface}(x)$, which gives the force per unit area.   Taking a spatial Fourier transform of (2), and relabeling the dummy variable $(x+x')$ as x, we have

$$\bar{z}(q) \equiv \int_{-\infty}^{\infty} \bar{z}(x)e^{-iqx}dx = \int_{-\infty}^{\infty} z(x)e^{-iqx}dx \int_{-L/2}^{L/2} g(x')e^{iqx'}dx'$$

$$\equiv Z(q)g(-q). \quad (3)$$

Since $g(x)$ is normalized, $g(q) = 1$ for $q = 0$ and $g \leq 1$ for all other q.   The factor $g(q)$ represents the effective smoothing of the track by a spatially extended suspension.

Let us consider several illustrative examples. If $g(x) = 1$, so that the average involves the whole length of the train equally, we have

$$g(q) = \frac{\sin(qL/2)}{(qL/2)} \quad , \tag{4a}$$

the familiar single-slit diffraction pattern, with an envelope falling off as $1/q$ for $qL \gg 1$. In the other limit of hard wheels at each end of the car, $g(x) = \frac{1}{2}[\delta(x-L/2) + \delta(x+L/2)]$, and

$$g(q) = \cos(qL/2) \quad . \tag{4b}$$

This "two-slit diffraction pattern" does not fall off with increasing q.

If the completely localized "wheels" are replaced by extended supports such as magnets at each end, we have $g(x) = \frac{1}{2}[g_0(x-L/2) + g_0(x+L/2)$ where $g_0(x)$ is a normalized weighting function for each separate magnet. This has

$$g(q) = \cos(qL/2)g_0(q) \tag{4c}$$

For example, if the magnet length is $L' = 2x_0$, we might approximate $g_0(x)$ by $e^{-(x/x_0)^2/2}$, in which case $g_0(q) = e^{-x_0^2 q^2/2}$ Thus, the envelope of $g(q)$ will fall off for $q \geq 1/x_0 \simeq 2/L'$.

Finally, if we consider the single Fourier component model of a train with periodic magnets along the entire length of the car, we have $g(x) = 1 + \cos 2kx$, where $k = 2\pi/\ell_x$, ℓ being the magnet period. In this case

-89-

$$g(q) = \frac{\sin(qL/2)}{qL/2} + \frac{1}{2}\frac{\sin(q/2+k)L}{(q/2+k)L} + \frac{1}{2}\frac{\sin(q/2-k)L}{(q/2-k)L} \tag{4d}$$

These various results are plotted in Fig. 2a.

Calculation of the Power Spectrum for Heave Acceleration

Combining the above results, the power spectrum of $\ddot{z}$ is

$$P_{\ddot{z}}(\omega) = \omega^4 P_z(\omega) = \omega^4 \left|\frac{z^2(\omega)}{\bar{z}^2(\omega)}\right| \frac{\bar{z}^2(\omega)}{z^2(\omega)} P_Z(\omega) \tag{5}$$

where $z^2/\bar{z}^2$ is given by (1a-1c), $\bar{z}^2/z^2 = |g(q=\omega/v)|^2$ and $P_Z(\omega) = P_Z(q=\omega/v)/v$. This can be evaluated for any power spectrum $p_Z(q)$. It is known empirically[3], however, that the roughness of roadways approximately follows the law

$$P_Z(q) \simeq A/q^2 \quad , \tag{6}$$

where $A \simeq 5\times10^{-6}$ ft. for an airport runway, while $A \simeq 8\times10^{-5}$ ft. for a highway of reasonable quality. Assuming the form (6), $P_Z(\omega) = Av/\omega^2$. Also, inspection of the examples (4a-4d) shows that for q in the most important range, g(q) can be written as a function $\tilde{g}$ of the dimensionless variable $qL = \omega L/v$. We can then combine factors and write (5) more explicitly as

$$P_{\ddot{z}}(\omega) = Av\omega^2 \left|\frac{z(\omega)}{\bar{z}(\omega)}\right|^2 |\tilde{g}(\omega L/v)|^2 \quad . \tag{7}$$

In order to display the behavior of this function we have plotted the factor $\frac{\omega^2}{\omega_o^2}\left|\frac{z(\omega)}{\bar{z}(\omega)}\right|^2$ in Fig. 3 for three different suspensions and have compared this factor with $|\tilde{g}(\omega L/v)|^2$ in

Fig. 2. The power spectrum of the heave acceleration is given
by the product of a curve from Fig. 2a for the appropriate pad
with a curve from Fig. 2b for the appropriate vehicle suspension.
The correspondence between the horizontal axes in the two plots
has been obtained from the relation $q = \omega/v$ and the train
parameters in Table 1. For lower velocities the scale in Fig. 2a
is contracted relative to that in 2b.

Mean Squared Acceleration

The mean squared heave acceleration $(\ddot{z}_{rms})^2$ is obtained
by integrating the power spectrum (7).

$$(\ddot{z}_{rms})^2 = \int_0^\infty P_{\ddot{z}}(\omega) d\omega \tag{8}$$

Although numerical integration is generally necessary for exact
results, estimates of the magnitude of $\ddot{z}_{rms}$ can be obtained
from various rough approximations. For a well damped parallel
suspension we may take

$$\frac{\omega^2}{\omega_0^2} \left|\frac{z(\omega)}{\bar{z}(\omega)}\right|^2 \simeq 1 \quad ,$$

except for the unimportant frequency range $\omega < \omega_0$. For this
case we obtain

$$(\ddot{z}_{rms})^2 \simeq Av\omega_0^2 \int_0^\infty |g(\omega L/v)|^2 d\omega \tag{9}$$

which diverges for a wheeled vehicle or if finite pad effects
are neglected.[1,2] This infinity is not physical since the

-91-

point mass approximation fails when ω reaches frequencies high enough for train dimensions to be comparable with a sonic wavelength. In magnetic or air cushion vehicles the finite pad function will cut the integral off at $\omega \simeq v/L$ before such frequencies are reached, and we need not pursue this question here.

In the case (4a) of a suspension which involves the whole length L of the train equally, (9) becomes

$$\ddot{z}_{rms} \simeq \omega_0 v (\pi A/L)^{\frac{1}{2}} \simeq 0.05g \qquad , \qquad (10)$$

using values from Table 1. This result is increased by a factor 3/2 for the case (4d) of a single Fourier component suspension of length L. For case (4b) of finite pads of length L' at the ends of a car

$$\ddot{z}_{rms} \simeq \omega_0 v (A/L')^{\frac{1}{2}} \qquad (11)$$

where we have assumed L' << L and set the rapidly oscillating $\cos^2(\omega L/2v)$ factor in the integrand of (9) equal to its average value 1/2. Comparing (11) with (10) we see that $\ddot{z}_{rms}$ is essentially inversely proportional to the square root of the length of the pad per car.

The mean square acceleration for a series suspension at 300 mph can be estimated very roughly by neglecting finite pad effects. The required integral is computed in Ref. 1.

$$\ddot{z}_{rms} = \left[\frac{\pi A v \omega_0^3}{2\sqrt{2}} \right]^{\frac{1}{2}} \qquad (12)$$

This result is smaller than those for the parallel suspension by a factor $\simeq [L\omega_0/v]^{\frac{1}{2}}$ which is the order of unity for a long pad at 300 mph. This result is somewhat inaccurate since the approximations used in obtaining (10) and (12) are poor for the case of a long pad. Numerical integration gives a factor 2 smaller rms acceleration for the series suspension with $Q = \frac{1}{2}$ than for the parallel suspension with $Q = 1$. This corresponds to a factor 4 difference in area under the curves in Fig. 2b out to the cut off frequency of the single slit diffraction pattern.

We can conclude that if a vehicle is in contact with its track over most of its length, the rms heave acceleration is insensitive to the type of suspension employed as long as the suspension is well damped and has a resonant frequency $\omega_0/2\pi \simeq 1$ Hz.

Subjective Roughness of Ride

It is known empirically that the subjective sensitivity of the human body to vertical accelerations depends on frequency. It is a maximum in the frequency range 2-20 Hz, which corresponds roughly to the natural frequencies of organs in the body. In this frequency range, the threshold level for perception is $\ddot{z} \simeq 0.003g$, whereas the intolerable level is reached at $\sim 0.3g$. Thus, a reasonable design goal for human comfort might be $a_{max} \equiv \ddot{z}_{max} = 0.03g \simeq 1$ ft./sec^2. This is the value quoted in studies[4,5] made for the TACV. For frequencies above and below this range, the comfort limit increases roughly as ω and ω^{-1},

-93-

respectively, suggesting an analytic approximation

$$\ddot{z}_{max}(\omega) = \frac{1}{2}\left(\frac{\omega_b}{\omega} + \frac{\omega}{\omega_b}\right) a_{max} \qquad , \qquad (13)$$

where $\omega_b/2\pi \simeq 6\text{-}10$ Hz represents the characteristic body fre-
quency. This function is plotted in Fig. 4, and compared with
the function quoted in the TACV reports.

Now, to the extent that the bodily response is linear[6],
it is plausible to compute a "subjective" root mean square
acceleration $(\ddot{z}_{rms})_{subj}$ by weighting each frequency of the power
spectrum with a factor normalizing it to the frequency ω_b at
which the body is most sensitive. Then our criterion for a satis-
factory ride quality can be stated as

$$(\ddot{z}_{rms})_{subj} \leq a_{max} \simeq 1 \text{ ft/sec}^2 \qquad , \qquad (14)$$

where

$$(\ddot{z}_{rms})_{subj} = \left| a_{max}^2 \int\limits_0^\infty \frac{P_{\ddot{z}}(\omega)}{|\ddot{z}_{max}(\omega)|^2} \, d\omega \right|^2 \qquad .$$

Note that if a single frequency is applied, $P_{\ddot{z}}(\omega)$ is a δ-function,
and our criterion reduced properly to $\ddot{z}(\omega) \leq \ddot{z}_{max}(\omega)$. It can
also be satisfied by a specific limiting power spectrum, such as
that specified in the Urban TACV Ride Quality Specification[7],
which was presumably drawn up on the basis of single-frequency
data with such a criterion in mind. However, such a specification
of a maximum for $P_{\ddot{z}}(\omega)$ at all ω can not be met if any discrete

frequency vibration is present (since it appears as a δ-function).
Such discrete frequencies will arise if periodic structured tracks
are used, as have been given considerable attention. Hence, as
integral criterion such as (14) seems a more generally useful form
of performance specification than either the discrete frequency
or power spectrum approaches.

Assuming the analytic approximation (13),

$$(\ddot{z}_{rms})^2{}_{subj} = 4\int_0^\infty P_{\ddot{z}}(\omega) \; \frac{(\omega/\omega_b)^2}{(1+\omega^2/\omega_b)^2} \; d\omega \; . \tag{15}$$

Comparing with $(\ddot{z}_{rms})^2$ as given by (8), we see that they differ
only by a factor $4(\omega/\omega_b)^2(1+\omega^2/\omega^2{}_b)^{-2}$ in the integrand. The
primary effect of this is to cut off the integral at $\omega \simeq \omega_b$.

In order to illustrate this relation we consider the
case of a parallel spring and dashpot suspension with various
pad configurations. Assuming constant suspension response as
in (9)

$$(\ddot{z}_{rms})^2{}_{subj} = Av\omega_0^2 \int_0^\infty |g(\omega L/v)|^2 \; \frac{(\omega/\omega_b)^2}{(1+\omega^2/\omega_b^2)^2} \, d\omega \quad , \tag{16}$$

where expressions for $g(q)$ are given in (4a-d).

For the case of point supports at the ends of the train

$$(\ddot{z}_{rms})_{subj} = \left[\frac{\pi Av\omega_0^2\omega_b}{2}\right]^{\frac{1}{2}} \simeq 4 \text{ ft/sec}^2 \quad , \tag{17}$$

using values from Table 1. This result is about 4 times our
design limit of 1 ft/sec^2.

For the case of uniform support along the whole train

$$(\ddot{z}_{rms})_{subj} = \left[\frac{2\pi A v^3 \omega_o^2}{L^2 \omega_b}\right]^{\frac{1}{2}} \approx 0.5 \text{ ft/sec}^2 \qquad (18)$$

where we have replaced a rapidly oscillating $(\sin)^2$ factor in the integrand by its average value $1/2$. The finite pad effect has thus yielded a useful 8-fold reduction in subjective acceleration.

Finally, for pads of length L' at both ends of the car

$$(\ddot{z}_{rms})_{subj} \approx \left[\frac{A v \omega_o^2 \omega_b}{1+\omega_b L'/v}\right]^{\frac{1}{2}} \qquad (19)$$

Thus, the criterion for significant reduction of subjective acceleration at high speeds (where the problem is most severe is

$$L' \geq v/\omega_b \approx 10 \text{ ft.} \qquad (20)$$

Even for $L' = 10$ ft., we estimate $(\ddot{z}_{rms})_{subj} \approx 1.5$ ft/sec^2, a bit above the design limit. The reason for the relative ineffectiveness of the finite pad here is that it essentially duplicates the effects of decreased bodily sensitivity at high frequencies.

From our analysis of subjective ride quality and finite pad effects we have reached the general conclusion that the detailed properties of the suspension are not critical. The formal divergence of the rms acceleration of a vehicle with a

-96-

parallel spring and dashpot suspension on a track with a rough-
ness power spectrum $P_z = A/q^2$ is not as important as is implied
in references 1 and 2. The integral (8) can be cut off by finite
pad effects in magnetic or air cushion vehicles. The effects of
high frequencies on subjective ride quality are further reduced
by the insensitivity of the body to frequencies above ω_b. Since
different types of suspensions differ in subjective ride quality
only by factors of order 2, the designer is free to use the most
convenient and economic suspension.

II. Importance of Finite Pads for Ferromagnetic Systems

As an example of the way in which finite pads affect vehicle design, consider the case of the attractive ferromagnetic high-speed transport systems, such as those being developed in Germany. For operation in the 300 mph range, it may be desirable to separate the functions of levitation and of damping of track irregularities so that the latter can be accomplished by conventional means without the complication of relative velocity along the direction of motion. This separation can be obtained in the attractive ferromagnetic systems by means of a low mass primary suspension or "magnetic bogie" which is constrained to follow the track closely and by a soft secondary suspension to isolate the passenger compartment. The primary suspension is feedback controlled for a high resonant frequency ω_{ob} of the bogie. A secondary suspension of the car with resonant frequency $\omega_{oc}/2\pi \simeq 1$ Hz is then required to obtain adequate ride quality.

The maximum allowable bogie acceleration is limited by strength considerations and also by suspension nonlinear effects[2] to a value considerably less than gM_c/M_b at which the dynamic load equals the static load. The inertial loading of the bogie by the car can be neglected in this discussion since the bogie acceleration from this source is approximately the car acceleration increased by the car - bogie mass ratio $\ddot{z}_{rms} M_c/M_b$. This is safely less than the upper limit of $g M_c/M_b$ for the total bogie acceleration, since we have seen that for acceptable ride quality we require $\ddot{z}_{rms} \simeq 1$ ft/sec^2.

The resonant frequency ω_{ob} of the bogie should be chosen
to minimize the suspension excursion subject to the above con-
straint on dynamic load, so that the magnet strength required for
safe operation is minimized. It is desirable to make the bogie
mass M_b small enough that ω_{ob} will be well above the frequency at
which the finite pad effects begin to reduce the track roughness,

$$\omega_{ob}/2\pi > v/L \simeq 5 \text{ Hz}.$$

The bogie will follow the (averaged) track up to this frequency.

The quantitative analysis of suspension excursion in
Ref. 2, Section 2, is approximately valid for this problem des-
pite their neglect of finite pad effects. The series suspension
assumed there is no longer truly optimum when finite pads are
considered but, from the arguments presented earlier, if the
vehicle is in contact with the track over most of its length,
the performance of other types of suspensions will not differ
markedly from that of the series suspension. Finite pad effects
are, however, still important in primary suspension design.
Without them an active suspension with series characteristics
extending to very high frequency is required to avoid the di-
vergence in bogie acceleration. The need for high frequency gain
in the feedback system is reduced by finite pads.

The power spectrum which drives the secondary suspension
is cut off at high frequencies by finite pad effects as well as
by a small amount of filtering from the primary suspension. We
have seen that any well damped suspension with a resonant fre-

quency in the range $\omega_{oc}/2\pi \simeq 1$ Hz will provide adequate ride quality for such a truncated power spectrum. Finite pad effects may thus make it possible to use a conventional passive secondary suspension consisting of a parallel spring and dash-pot.

In order to achieve these advantages, the levitation electromagnets must be distributed over most of the car length. If this is done in such a way that the levitation field at the track is kept essentially uniform, a further benefit is obtained. Eddy current repulsion of the train from a conducting track can be thereby minimized.

III. Dynamics of Superconducting Suspensions

In our discussion above, we have simply characterized the dynamics of a magnetic suspension by a stiffness parameter ω_0 and a damping parameter Q, while concentrating on the spatial characteristics of the suspension (which determine finite pad effects) and exterior quantities (such as track roughness and subjective sensitivity to accelerations). In this part, we shall consider in some detail the effect of the track variations on the superconducting magnets.

The analysis falls into three major parts.

(A) We consider the implications of the constant-flux (as opposed to constant-current) aspect of the superconducting magnets, as they interact with a track which is not completely uniform.

(B) We give a thorough analysis of the use of a conducting sheet between magnets and track to screen out AC currents in the superconducting magnets due to variations in the track as well as to provide a substantial damping of the vertical motion.

(C) We give specific attention to the AC losses in the super-conducting magnets, to determine the conditions of operation required to keep them to an acceptably low level. Our conclusions are summarized in (D).

A. Constant Flux Magnets

The basic ideas can be illustrated by a lumped constant circuit model. In this model, we simulate the time dependence

-101-

due to motion of the train by use of an alternating current at frequency $\omega_v = kv = 2\pi v/\ell_x$, where v is the velocity of the train and ℓ_x is the (full) period of the magnet structure. Let one of the superconducting magnets be replaced by an inductance L_1 carrying an AC current, and the track be replaced by a second inductance L_2, and let the mutual inductance be M. Then the fluxes ϕ_1 and ϕ_2 threading the two coils obey

$$d\phi_1/dt = d(L_1 I_1 + MI_2)/dt \tag{21a}$$

$$d\phi_2/dt = d(L_2 I_2 + MI_1)/dt = -I_2 R_2 \tag{21b}$$

Taking an $e^{i\omega t}$ time dependence and solving for I_1, we have

$$I_1 = \frac{\phi_1/L_1}{1 - \dfrac{k_{12}{}^2}{1-i/\omega\tau}} \tag{22}$$

where $k_{12} = (M^2/L_1 L_2)^{\frac{1}{2}} \leq 1$ is the coefficient of coupling and $\tau = L_2/R_2$ is the time constant of the coil simulating the track. When the train is at rest, $\omega = \omega_v \equiv kv = 0$, and $I_1 = \phi_1/L_1$. When the train is moving at any reasonably high speed (v >> v_0 $\sim$4 mph), $\omega\tau$ >> 1, and $I_1 = (\phi_1/L_1)/(1-k_{12}{}^2)$. Thus, the current in the superconducting magnet increases by a factor $(1-k_{12}{}^2)^{-1}$ when the train goes from rest to speed, since superconducting magnets in persistent current mode are characterized by constant flux. This factor is of significant size, since $k_{12}{}^2$ must be

-102-

reasonably large to give substantial lift.

To make this point more quantitative, we note that

$$I_2 = \frac{-(\Phi_1/M)k_{12}{}^2}{1-k_{12}{}^2-i/\omega\tau} \tag{23}$$

which goes to zero properly if either k_{12} or ω go to zero.
The force is then given by

$$F = -I_1 I \nabla M = \frac{(\Phi_1{}^2/L_1)k_{12}{}^2(\nabla M/M)}{(1-k_{12}{}^2)^2} \frac{(1-i/\omega\tau)}{(1-\frac{i/\omega\tau}{1-k_{12}{}^2})^2} \tag{24}$$

In the limit $\omega\tau \gg 1$, this reduces to

$$F = -L_1 I_1{}^2 k_{12}{}^2(\nabla M/M) \tag{24a}$$

$$= -\frac{(\Phi_1{}^2/L_1)k_{12}{}^2(\nabla M/M)}{(1-k_{12}{}^2)^2} \tag{24b}$$

In these expressions, ∇M refers to the rate of change of M
with separation of the two inductors. Accordingly, $\nabla M/M$
$\approx 1/h$, where h gives the height of the suspension above a
suitable reference plane. Thus, increasing the coupling k_{12}
not only increases the current because of the constant flux
condition $\Phi_1 = $ const, as reflected by the factor $(1-k_{12}{}^2)^{-2}$
in (24b).

Actually, a given magnet will be characterized by a maximum permissible current I_1. Thus, (24a) 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 (22) 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.

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_0 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 ($\sim M_0{}^2/L_1 L_2$) with

-104-

associated intolerable levels of dissipation in the supercon-
ducting 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 $\sim n_R^{-p}$, where the exponent p $\sim$2-3 will depend on such
parameters as the length/width ratio of the train magnet coils
and their separation, which determine the waveform of the flux
density at the track. For example, if the waveform of $B_z(x)$
were a square wave, the fractional modulation would be reduced
by a factor $1/n_R$, 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 $1/n_R$. In fact, by suitable choice of winding con-
figuration 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[8] for a
specific track configuration display the qualitative features
described here. The fractional modulation of the lift is found
to be reduced by a factor of $\sim 10^2$ for $n_R = 10$; and $n_R = 6$ is

-105-

found to give anomalously low modulation ($\sim 3 \times 10^{-3}$), perhaps due to a near cancellation of the 6th Fourier component.

This reduction in the modulation of the coupling by large values of n_R not only reduced the ride roughness, but 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 proporational to $\dot{I}_1 \sim \omega \delta I_1 \sim n_R \delta I_1$ for fixed train magnet length.

B. Use of Conducting Sheet for Screening and Damping

A further reduction in AC losses 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. 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. In our lumped constant model, this sheet would be represented as a third LR circuit couplied to both L_1 and L_2. Rather than follow this schematic approach, we consider a more specific idealization, which allows explicit results to be obtained.

Our model is a variation on the Fourier analysis approach reported earlier.[9] The magnet structure is represented by a transverse sheet current density $\mathcal{J}_{oy}(x) = \sum_m \mathcal{J}_{om} e^{ik_m x}$, where $k_m = 2\pi m/\ell_x$, ℓ_x being the full period of the magnet structure (including one up and one down magnet loop). With reasonable

symmetry, only odd values of m appear. This current sheet is at height h above a ground plane which is taken to be super-conducting for simplicity. (See Fig. 5). The screening sheet is a distance s below the magnet plane, has conductance σd per square, and carries a current $\mathcal{J}_{1y}(x) = \sum_m \mathcal{J}_{1m} e^{ik_m x}$, where the $\mathcal{J}_{1m}$ are to be determined. The solution is found using the method of images to build in the boundary condition $B_z = 0$ at the ground plane. Thus, image currents $-\mathcal{J}_{om}$ and $-\mathcal{J}_{1m}$ are inserted at distances h and (h-s) below the ground plane, fields are calculated everywhere, and self-consistent values determined for the currents.

For a current sheet at z_o, the Fourier coefficients of B_z and B_x are given by

$$B_{zm}(z) = \frac{2\pi i}{c} \mathcal{J}_m e^{-k_m |z-z_o|} \tag{25a}$$

and

$$B_{xm}(z) = \pm \frac{2\pi}{c} \mathcal{J}_m e^{-k_m |z-z_o|} \tag{25b}$$

where $\pm = (z-z_o)/|z-z_o| = \text{Sgn}(z-z_o)$. When the total field B_z due to all currents is evaluated at the magnet plane, one finds

$$B_{zm}(h) = \left(\frac{2\pi i}{c}\right)\left\{\mathcal{J}_{om}\left[1-e^{-2k_m h}\right] + \mathcal{J}_{im}\left[e^{-k_m s}-e^{-k_m(2h-s)}\right]\right\} \tag{26}$$

The condition of constant flux in the superconducting magnets is satisfied if each Fourier component of $B_z(h)$ is constant. Although this may be an overly strong condition, it should be

-107-

a reasonable approximation if a single component is dominant. If we apply this condition in the DC case, where $\mathcal{G}_{1m} = 0$, (26) shows that the currents increase by the factors

$$\frac{\mathcal{G}_{om}(h)}{\mathcal{G}_{om}(\infty)} = \frac{1}{1-e^{-2k_m h}} \tag{27}$$

when the ground plane is brought in from infinity (or equivalently, for normal metal ground plane, when the train reaches high speed). Comparing with (22), for $\omega\tau \gg 1$, we see that in this case

$$k_{12}{}^2(m) = e^{-2k_m h} = e^{-m(4\pi h/\ell_x)} \tag{28}$$

Thus, the higher Fourier components are coupled much more weakly to the track. Accordingly, for simplicity in the following discussion, we shall drop all except the fundamental Fourier component (m=1) and suppress the subscript m.

Now we consider the situation in which the height h is oscillating with amplitude $\delta h\, e^{i\omega t}$ about its equilibrium value, so that AC currents are induced in the normal metal sheet. From (26), $\delta B = 0$ implies that

$$\mathcal{G}_o 2k\delta h e^{-2kh} + \delta\mathcal{G}_o[1-e^{-2kh}] + \delta\mathcal{G}_1[e^{-ks}-e^{-k(2h-s)}] = 0 \tag{29}$$

where $\mathcal{G}_o$ is the DC value, whereas $\delta\mathcal{G}_o$ and $\delta\mathcal{G}_1$ vary as $e^{i\omega t}$. The other basic equation is that

-108-

$$\mathcal{G}_1 = \sigma d\, E_y = -\frac{\sigma d\omega}{kc}\, B_z \qquad (30)$$

where B_z is evaluated at the plane of the sheet using (25b).
The result is

$$\frac{-c^2 k}{2\pi i \omega \sigma d}\, \delta\mathcal{G}_1 = \mathcal{G}_0 2k\delta h e^{-k(2h-s)} + \delta\mathcal{G}_0[e^{-ks}-e^{-k(2h-s)}]$$

$$+ \delta\mathcal{G}_1[1-e^{-2k(h-s)}] \qquad (31)$$

By combining (29) and (31), we can solve for $\delta\mathcal{G}_0$ and $\delta\mathcal{G}_1$ in
terms of δh. The results simplify significantly if we neglect
e^{-4kh} relative to terms in e^{-2kh}. (This approximation is at
the same level as neglecting the difference between constant
flux and constant current.) In that case

$$\delta\mathcal{G}_0 = -\frac{\mathcal{G}_0 2k e^{-2kh}\delta h}{1+i\omega\tau} \qquad (32a)$$

$$\delta\mathcal{G}_1 = -\mathcal{G}_0 2k e^{-k(2h-s)}\delta h \frac{i\omega\tau}{1+i\omega\tau} = i\omega\tau\, e^{ks}\, \delta\mathcal{G}_0 \qquad (32b)$$

where

$$\tau = \frac{2\pi\sigma d(1-e^{-2ks})}{kc^2} = \frac{1-e^{-2ks}}{kv_0} \qquad (33)$$

Here $v_0 = c^2/2\pi\sigma d$ is the characteristic velocity introduced in
our earlier paper. Its value is $\sim$4 mph for 1" aluminum sheet.
Thus, τ will typically be not much greater than $\sim$0.1 sec. un-
less cryogenically cooled conductors are used to increase σ.

The screening effectiveness of the conducting sheet is

given directly by (32a). The magnitude of the AC current changes induced in the superconducting magnets is reduced by a factor $(1+\omega^2\tau^2)^{-\frac{1}{2}}$ for a given amplitude δh. This will not be important for vertical oscillations at the natural frequency of the suspension, $\omega_0/2\pi \simeq 1$ Hz, since $\omega_0\tau \lesssim 1$. However, it will reduce the size of the much higher frequency current changes induced by motion over a structured track. In dealing with a structured track, the higher Fourier components of the magnet structure become important when they match the track period. Since both ω and $\tau_m^{-1} \simeq m\tau^{-1}$, the damping coefficient for the mth component, increase in proportion to n_R, this reduction factor is approximately independent of n_R, having the value $(\omega_v\tau)^{-1} \simeq v_0/v$ for train speed v. This reduction factor is 1/75 for v = 300 mph and the typical value v_0 = 4 mph, so such a shield should be very effective in reducing AC losses in the magnets.

Effect of Conducting Sheet on Equation of Motion

If we assume that both the shielding sheet and the superconducting magnets are rigidly attached to the mass of the train, the total upward force on them gives the lift force. Thus we compute

$$F_z = - \frac{1}{c} [\mathcal{J}_0 B_x(h) + \mathcal{J}_1 B_x(h-s)] \tag{34}$$

where B_x is evaluated using (25b), omitting the sheet under consideration. The result is

$$F_z = \frac{2\pi}{c^2} [\mathcal{J}_0^2 e^{-2kh} + 2\mathcal{J}_0\mathcal{J}_1 e^{-k(2h-s)} + \mathcal{J}_1^2 e^{-2k(h-s)}] \tag{35}$$

-110-

Since only q_0 has a DC component, we write

$$q_0 = \overline{q_0} + \delta q_0$$

$$q_1 = \delta q_1$$

and drop quadratic terms in δq. This yields

$$\overline{F_z} = \frac{2\pi}{c^2} q_0{}^2 e^{-2kh} \tag{36}$$

and

$$\delta F_z = \frac{4\pi}{c^2} \overline{q_0} e^{-2kh}[\delta q_0 + \delta q_1 e^{ks}]$$

or

$$\delta F_z = -\frac{4\pi}{c^2} \overline{q_0}{}^2 e^{-4kh} 2k\delta h \frac{1+i\omega\tau e^{2ks}}{1+i\omega\tau} \tag{37}$$

using (32). Replacing $\overline{F_z}$ by Mg, we can then write (37) as

$$\delta F_z = -4Mgke^{-2kh}[1 + \frac{(e^{2ks}-1)i\omega\tau}{1+i\omega\tau}]\delta h \tag{38}$$

Introducing the definitions

$$\omega_{oo}{}^2 = 4gke^{-2kh} \tag{39a}$$

$$\gamma = e^{2ks}-1 \tag{39b}$$

we can write (38) as

-111-

$$\delta F_z = -M\omega_{oo}^2 \left[1 + \frac{\gamma}{1 - i/\omega\tau} \right] \delta h$$

$$= -M\omega_{oo}^2 \left[\left(1 + \frac{\gamma\omega^2\tau^2}{1+\omega^2\tau^2} \right) + \frac{i\gamma\omega\tau}{1+\omega^2\tau^2} \right] \delta h \tag{40}$$

$$\equiv -M\omega_c^2 \delta h - \alpha i\omega\delta h$$

where

$$\omega_o^2 = \omega_{oo}^2 \left(1 + \frac{\gamma\omega^2\tau^2}{1+\omega^2\tau^2} \right) \tag{41a}$$

and

$$\alpha = \frac{M\omega_{oo}^2\gamma\tau}{1+\omega^2\tau^2} \tag{41b}$$

define effective frequency-dependent force constant and damping parameters. We see that the stiffness of the suspension ($\sim\omega_o^2$) increases by a factor of $(1+\gamma) = e^{2ks}$ in going from $\omega\tau \ll 1$ to $\omega\tau \gg 1$. This factor arises since above $\omega = 1/\tau$, the AC force is essentially on the screen, not the magnet, so the effective height is h-s. Similarly, α falls from a low frequency limit of $M\omega_{oo}^2\gamma\tau$ as $(1+\omega^2\tau^2)^{-1}$.

Damping

A dimensionless measure of the damping is $Q = M\omega_o/\alpha$. Evaluating this at the effective resonance frequency ω_o, we find

$$Q = (\omega_o/\gamma\tau\omega_{oo}^2)(1+\omega_o^2\tau^2) \tag{42}$$

This is a minimum if τ is adjusted to be ω_o^{-1}. In that case

-112-

$$Q_{min} = \frac{2}{\gamma} \frac{\omega_0^2}{\omega_{00}^2} = 1 + \frac{2}{\gamma} = \text{ctnh } ks \approx \frac{1}{ks} \qquad (43)$$

The physical reason for $Q \sim 1/s$, for small s, is that the resistive screen is effectively "shorted out" by the superconducting coil behind it unless there is space between the two. For the typical case in which $s = h/2$, so that the shield sheet is halfway between magnet and ground plane, $\gamma = e^{2ks}-1 = e^{kh}-1$. Since the lift force falls as e^{-2kh}, it is desirable to have $kh \lesssim 1/2$. Thus, typically $\gamma < (e^{1/2}-1) = 0.65$, and is hard to lower the Q below ~ 4. This is somewhat above the optimum value, in view of the resonant amplification of the amplitude by a factor of $(1+Q^2)^{1/2}$ at ω_0. For random noise, the decreased bandwidth with higher Q partially cancels this factor in integrating over the whole noise spectrum. In any case, this damping would prevent any catastrophic buildup of oscillations, and it has the advantage of being completely passive and foolproof.

A better physical feel for the result may be given by considering the die-away time $\tau_D = Q/\omega_0$. For $Q = 4$ and $\omega_0/2\pi = 1$ Hz, this gives $\tau_D = 2/\pi \approx 0.6$ seconds, a quite reasonable value.

By comparison, the damping due to the currents in the normal track is much weaker, τ_D being of the order of 20 seconds as full speed. This long damping time might seem surprising, since the die-away time of the track currents τ is only ~ 0.1 sec. The ineffectiveness of the damping arises since the fields seen from the track have frequency $\omega_v = kv$, amplitude modulated at

$\sim \omega_0$ due to the heaving motion, so $Q \simeq \omega_v \tau >> \omega_0 \tau$. This can be modeled using (24), with $\omega = \omega_v$, and with k_{12}^2 varying with the heaving motion. For simplicity, take $k_{12}^2 << 1$, in which case (24) reduced to

$$\delta F = \frac{(\Phi_1^2/L_1)(\nabla M/M)}{1 - i/\omega_v \tau} \quad \frac{\partial k_{12}^2}{h} \delta h$$

$$\equiv \frac{-M\omega_0^2 \delta h}{1 - i/\omega_v \tau} \simeq -M\omega_0^2 \delta h - \frac{iM\omega_0^2}{\omega_v \tau} \delta h \qquad (44)$$

Noting that in the heaving motion $\dot{h} = i\omega_0 \delta h$, we see that the effective viscous damping force for vertical motion is $-\alpha \dot{h}$, where $\alpha = M\omega_0/\omega_v \tau$. The corresponding damping time is

$$\tau_D = M/\alpha = \omega_v \tau/\omega_0 \qquad (45)$$

and the Q is

$$Q = \omega_0 \tau_D = \omega_v \tau \simeq F_L/F_D \qquad (46)$$

In setting $\omega_v \tau \simeq F_L/F_D$, we draw on the results of the analysis in our earlier work. Thus, any suspension giving a favorable lift/drag ratio will give little damping of the vertical motion. By contrast, a screening plate, fixed relative to the magnets, sees a field varying at the low frequency ω_0, and gives much more effective damping, with $Q \sim \omega_0 \tau$ rather than $\omega_v \tau$.

If the passive damping is deemed insufficient for ride quality, a relatively small amount of active feedback control in addition should be adequate.

Response Function

Although it is illuminating to define the frequency-dependent parameters ω_0 and α, as in (41), for actual computation of system response over a wide frequency range, it is more convenient to work with the constant parameters ω_{00} and γ, defined in (39). Also, we replace δh in (40) by our earlier notation $(z-\bar{z})$, where z and $\bar{z}$ are absolute heights of train and track measured from their respective equilibrium positions. Then, equating δF_z to $M\ddot{z} = -\omega^2 Mz$, (40) becomes

$$\frac{z(\omega)}{\bar{z}(\omega)} = \frac{1+\gamma-i/\omega\tau}{(1+\gamma-\omega^2/\omega_{00}{}^2)-i(1-\omega^2/\omega_{00}{}^2)/\omega\tau} \tag{47}$$

$$\frac{z(\omega)|^2}{|\bar{z}(\omega)|^2} = \frac{\omega_{00}{}^4[(1+\gamma)^2+(1/\omega\tau)^2]}{[(1+\gamma)\omega_{00}{}^2-\omega^2]^2+(\omega_{00}{}^2-\omega^2)^2(1/\omega\tau)^2} \tag{48}$$

This expression reduced to unity at low frequency. At high frequency, it reduces to $(1+\gamma)^2(\omega_{00}/\omega)^4$. For comparison, the high frequency limits for the parallel and series dashpot examples are $\omega_0{}^4/Q^2\omega^2$ and $\omega_0{}^4/\omega^4$, respectively. Thus, the damping plate resembles the series dashpot in that $(z/\bar{z})^2$ falls off as ω^{-4} at high frequencies but it retains a dependence on $Q(\sim 1/\gamma)$ reminiscent of the parallel dashpot case. However, since $Q \sim 1$, the latter factor is relatively unimportant.

C. Concluding Summary

Since superconducting magnets in persistent current mode maintain constant flux, rather than constant current, the current increases by a factor of as much as $(1-k_{12}^2)^{-1}$ when the magnet induces currents in neighboring conductors. Assuming a reasonable value of 0.35 for k_{12}^2, the coupling coefficient between magnets and track, this implies a current increase by a factor of ~ 1.5 in going from rest to speed. Thus, in "charging" the magnets with current, the current must be held to less than $(1-k_{12}^2)$ $\sim 65\%$ of the critical current to allow for this increase. A further margin of safety should be provided to allow for increases in current with excessive weight loads or when the train encounters a "bump" in the track which momentarily increases the coupling. If a structured track is used, these AC induced currents could be excessive; however, by using track loops which are only a fraction (e.g., 1/6) of the length of the magnet loops, the AC component can be greatly reduced.

Further reduction in the AC currents can be obtained by use of a conducting sheet screen between magnets and track. Although this is not very effective at the low natural frequency of the suspension, it should reduce the high-frequency AC currents due to a structured track by a factor of the order of v_o/v, where v_o is a characteristic velocity (4 mph for 1" thick aluminum), proportional to the resistance per square of the screen.

Another important benefit from the conducting sheet is that it provides considerable damping of the vertical motion of

-116-

the train. It appears that a die-away time τ_D of less than 1 second can be obtained using this simple, passive system. Moreover, the high-frequency response of the suspension has the desirable rapid cutoff characteristic of series-dashpot damping, rather than the less desirable properties of the parallel dashpot. By comparison, the damping due to currents in the normal track is much weaker, the die-away time being proportional to the lift/drag ratio, and being of the order of 20 seconds at full train speed.

In a related paper[10], it is shown that cryogenic heating due to AC loss phenomena will be excessive even for the best twisted multifilamentary composite superconducting wire presently available, unless $\dot{B} \leq 10^3$ gauss/second. To satisfy this criterion, the amplitude of heaving motions at the natural frequency of the suspension must be held to under 1% of the height (i.e., to ±0.1"), corresponding to an rms vertical acceleration of about 0.01g; this figure is about an order of magnitude smaller than the limit set by ride quality considerations. Similar numerical results are found at the higher frequencies associated with a structured track, even with a 1" thick room temperature aluminum screening plate. This constraint could be made less severe by cooling the normal metal screen to liquid N_2 temperatures to improve its conductivity and hence its screening. However, this improvement would be bought at the expense of the effectiveness of the damping of vertical motion. Damping is most effective when $\omega\tau \sim 1$, whereas screening goes as $(1+\omega^2\tau^2)^{-\frac{1}{2}}$. Thus there

is little screening when there is good damping. Since τ is ~ 0.1 sec for 1" aluminum plate at room temperature, this gives approximately optimum damping at the natural frequency of the suspension $\omega_0/2\pi \sim 1$ Hz. If the screen were cooled to give useful screening at 1 Hz, the damping would be lost.

Thus, unless somewhat better composite superconductors can be fabricated or unless somewhat greater heat loads can be accepted, it appears that energy dissipation in the superconducting magnets may form a somewhat more severe constraint on tolerable vertical accelerations than does subjective passenger comfort.

Acknowledgement

TABLE 1

Values Chosen for Numerical Estimates

$\omega_0/2\pi = 1$ Hz

$\omega_b/2\pi = 10$ Hz

$L \quad = 100$ ft.

$v \quad = 440$ ft/sec (300 mph)

$A \quad = 10^{-5}$ ft.

$\lambda \quad = 1.5$ ft.

References

1. D. A. Hullender, D. N. Wormley, and H. H. Richardson, "A Preliminary Study of Actively Controlled Air Cushion Vehicle Suspensions," MIT Engineering Projects Laboratory Report EPL-70-76110-11, D.O.T. Contract C-85-65 (1970).

2. S. G. Meisenholder and T. C. Wang, "Dynamic Analysis of an Electromagnetic Suspension System for a Suspended Vehicle System," T.R.W. Systems Group Report 06818-6052-R0.00, D.O.T. Contract DOFFR-C-353-66 (NEG), (1972).

3. Evidence summarized in References 1 and 2.

4. "Technical Extracts from TACV System Study and Analysis Report," T.R.W. Report 06818-6008-R000, for DOT.

5. "Tracked Air Cushion Research Vehicle/Guideway Dynamic Analysis." Grumman Report PMT-B4-R71-07, D.O.T. Contract DOT-FR-00005.

6. The linearity of subjective body response could be tested by seeing whether there is a strong sensitivity to a beat frequency near ω_b between two frequencies both of which are outside the range of high sensitivity.

7. U.S. Dept. of Transportation, May 10, 1971.

8. T. Yamada and M. Iwamoto, Trans. IEEJ, Vol. 92-B, No. 1, p. 19, 1971.

9. P. L. Richards and M. Tinkham, J. Appl. Phys. __43__, 2680 (1972).

10. M. Tinkham, "AC Losses in Superconducting Magnet Suspensions for High-Speed Transportation", 1972 Materials Research Council Report and J. Appl. Phys. __44__, 2385 (1973).
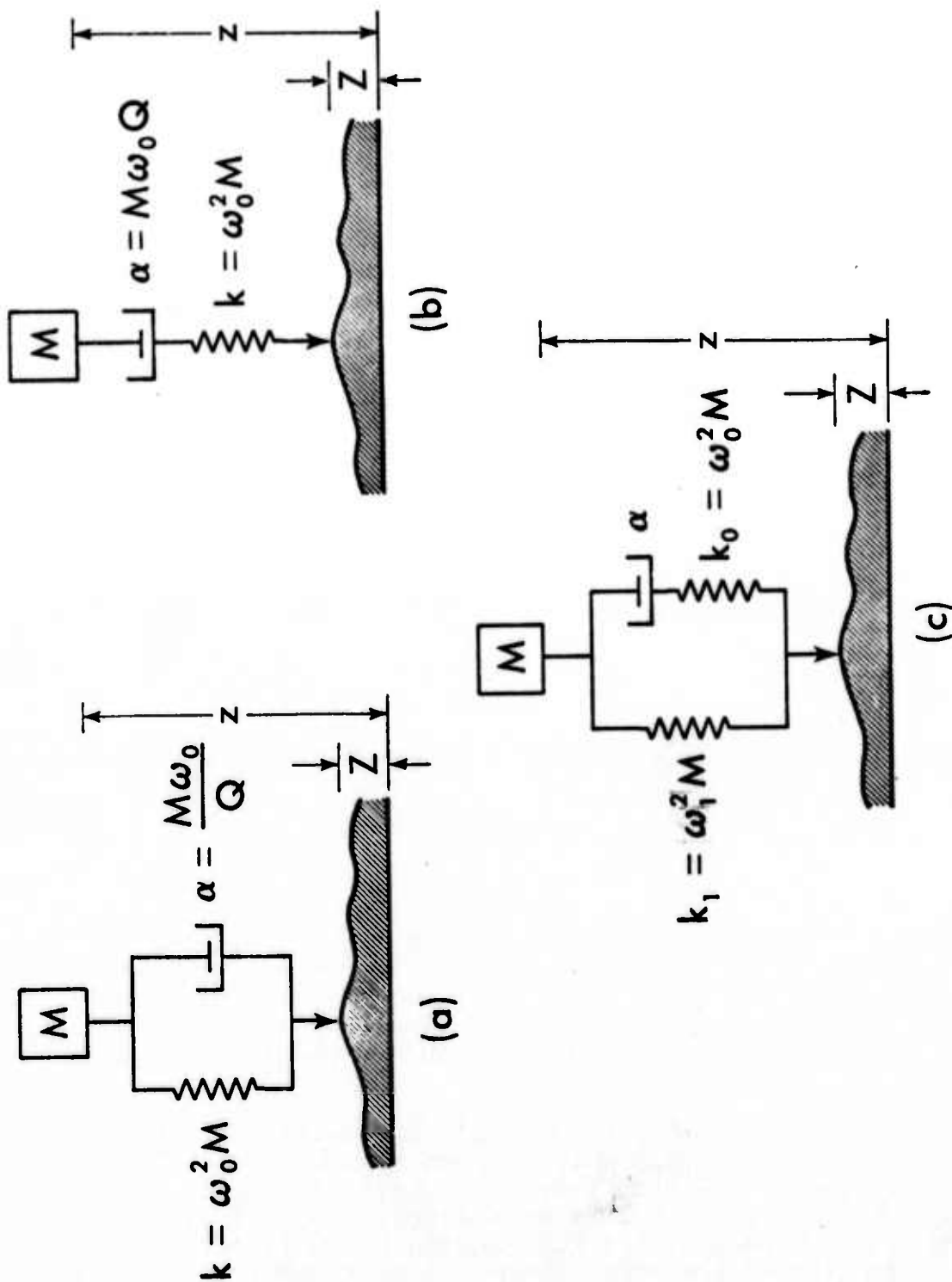
Figure 1. Three suspensions considered in this report. (a) parallel, (b) series, and (c) combined parallel and series.
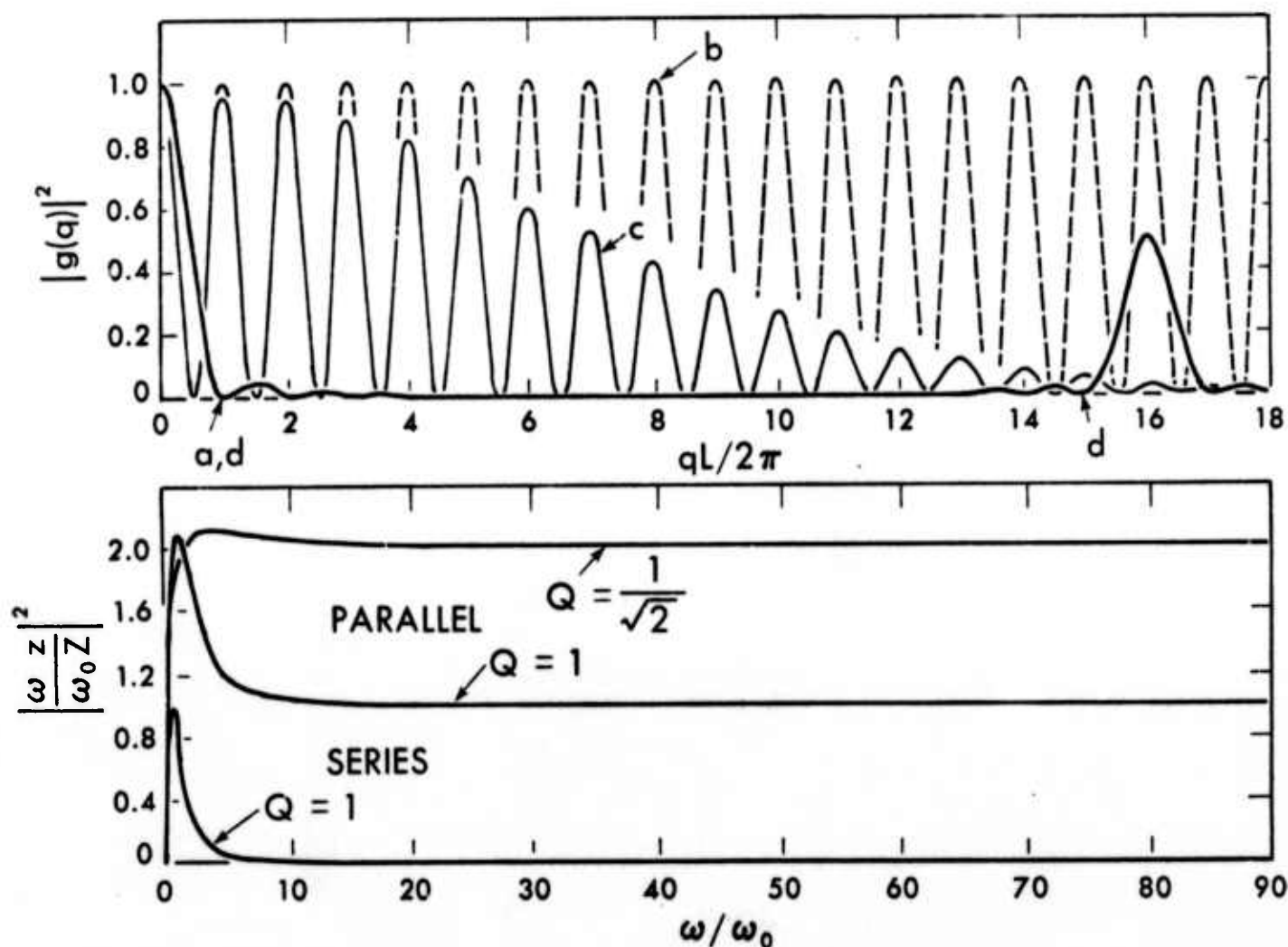
Figure 2 (a). Plots of the finite pad function $g^2(q)$ from
Eqs. (4a-4d). (a) uniform contact of length L, (b) point
contacts separated by L, (c) finite magnets of length $\sim$L/16
separated by L, (d) single Fourier component of length L
which corresponds to 16 uniformly spaced superconducting
magnets.

(b). Acceleration response $\omega^2/\omega_0^2 |z(\omega)/Z(\omega)|^2$ of
series and parallel suspension driven by unfiltered track
irregularities. The area under the parallel suspension
curves are infinite, leading to a divergence in $\ddot{z}_{rms}$ in the
absence of finite pad effects. Reasonable Q values have
been selected and the scale chosen to correspond to Fig. 2a
for a typical train at 300 mph.
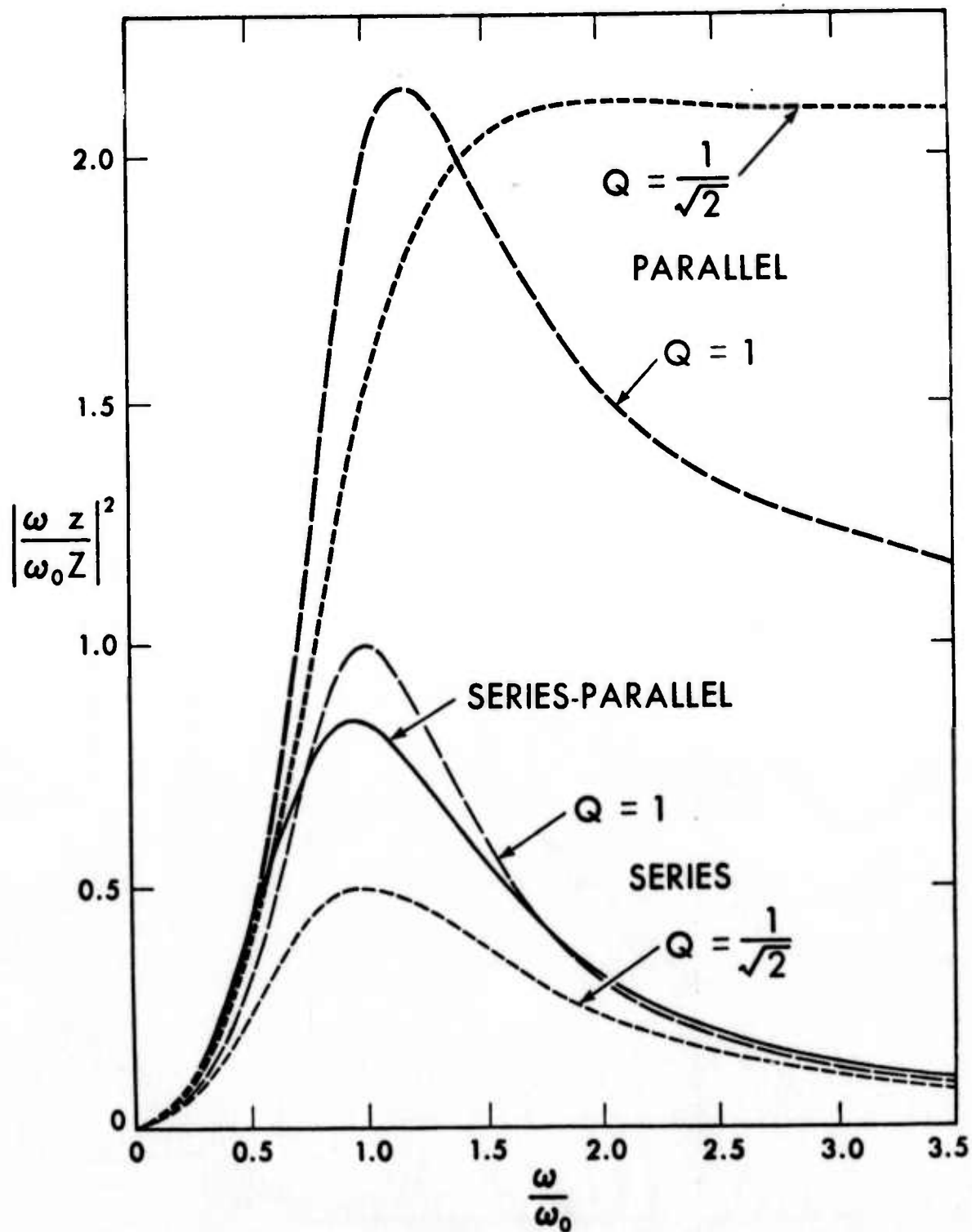
Figure 3. Acceleration response $\omega^2/\omega_0^2|z(\omega)/\dot{z}(\omega)|^2$ of series, parallel and combined suspensions with various Q values driven by unfiltered track irregularities. The dashpot strength and spring constant k_0 for the combined suspension were chosen to be the same as for the series suspension with $Q = 1/\sqrt{2}$. The spring constant k, is 8 times weaker than k_0.
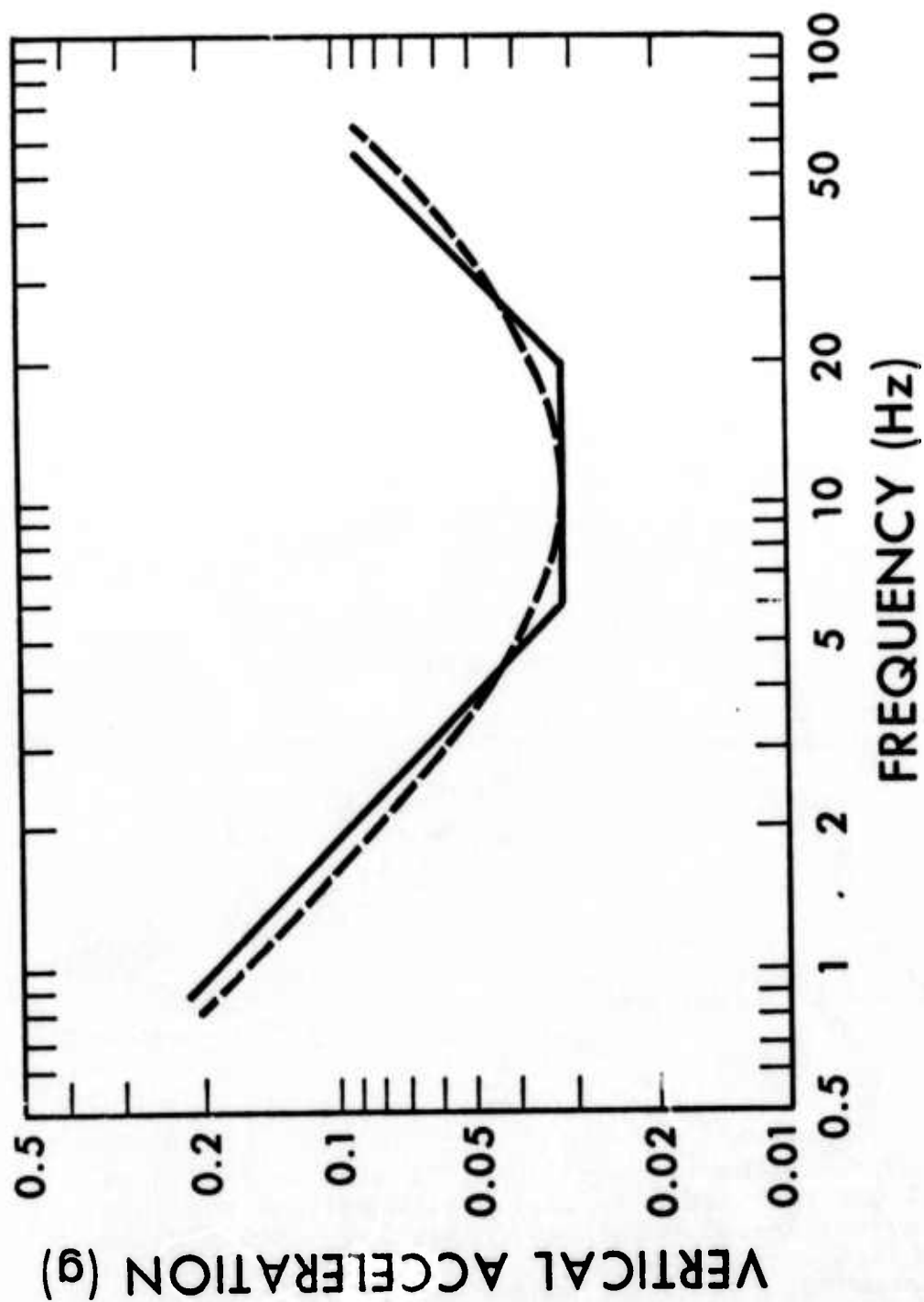
Figure 4. Ride quality standard for vertical acceleration. Solid curve is from Ref. 4. Dashed curve is analytic approximation, Eq. (13), with $\omega_b/2\pi$ = 10 Hz and a_{max} = .03g.

$\mathcal{I}_0$ MAGNET CURRENT

$\mathcal{I}_1$ SCREENING CURRENT

GROUND PLANE

$\left.\begin{array}{c} -\mathcal{I}_1 \\ -\mathcal{I}_0 \end{array}\right\}$ IMAGE CURRENTS
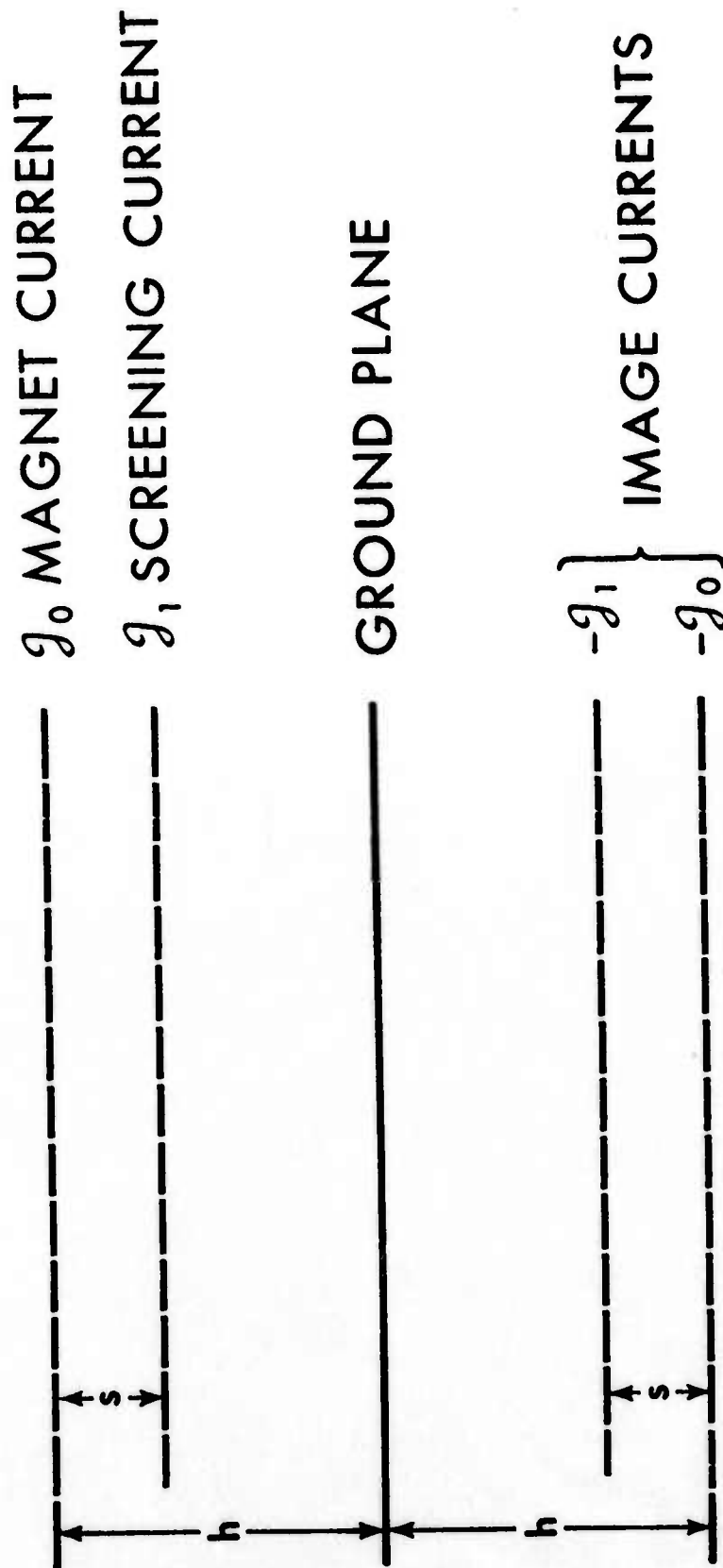
Figure 5. Schematic diagram showing location of current sheets in model.

ADDENDUM TO
AC LOSSES IN SUPERCONDUCTING MAGNET SUSPENSIONS
FOR HIGH-SPEED TRANSPORTATION

M. Tinkham

Abstract

An earlier analysis of AC losses in superconducting
magnets is extended to include conductors with thick as well
as thin filaments. It is shown that thicker filaments may be
superior for applications involving small AC fields, while the
smallest filaments are best for large AC amplitudes.

ADDENDUM TO
AC LOSSES IN SUPERCONDUCTING MAGNET SUSPENSIONS
FOR HIGH-SPEED TRANSPORTATION

M. Tinkham

In a previous paper[1], an analysis was given of the
AC losses to be expected in superconducting magnets used for
levitation of high-speed trains. One of the conclusions
reached was that twisted multi-filamentary superconductors
with filament diameter $d \lesssim 10\mu$ would be needed to keep these
losses sufficiently low. On the contrary, recent experimental
results by the Ford group[2] showed lower losses for coils wound
with single core wire ($d \simeq 200\mu$). As noted by the Ford group,
this result can be understood by recalling that small amplitude
AC fields which do not fully penetrate the filament give a
loss proportional to $(\Delta B)^3$ rather than ΔB, the result for the
case of very fine filaments, used in ref. 1. Because of the
practical importance of this low-loss behavior, it seems worth-
while to extend our previous treatment to include this low
amplitude regime, in order to correct our overly pessimistic
earlier estimate.

For simplicity we consider only a flat slab geometry with
thickness d, with the same field applied to both surfaces. Then
using the usual Bean model[3] calculation, the hysteresis loss per

cycle and per unit volume for field excursion of $\pm\Delta B$ about some operating point is

$$Q = \frac{J_c}{c} \, d(\Delta B) - \frac{4\pi}{3} \frac{J_c^2}{c^2} \, d^2 \qquad \Delta B > H_s \equiv 2\pi d \, J_c/c \qquad (1a)$$

and

$$Q = \frac{(\Delta B)^3}{12\pi^2 dJ_c/c} \qquad \Delta B < H_s \qquad (1b)$$

Equation (6a) of ref. 1, and the estimates in that paper, are based on the linear leading term of (1a). The correction term in (1a) reduces this leading term by a factor of 3 when ΔB is reduced to $H_s = 2\pi dJ_c/c$, the lowest value of AC field for which complete penetration occurs. At this point, a continuous transition is made to (1b), and the loss thereafter decreases as $(\Delta B)^3$. Thus, our estimate based on the leading term of (1a) can seriously overestimate the true loss if $\Delta B \stackrel{<}{\sim} H_s$.

Going further, (1) indicates that $Q \rightarrow 0$ for $d \rightarrow 0$ or $d \rightarrow \infty$. Dissipation is greatest for intermediate values. This can be made more clear by introducing a dimensionless thickness variable

$$x = \frac{2\pi J_c d}{(\Delta B)c} \equiv \frac{d}{d_s} = \frac{H_s}{\Delta B} \qquad (2)$$

which is the ratio of d to the slab thickness penetrated by a field change ΔB. In terms of x, (1) becomes

$$Q = \frac{(\Delta B)^2}{6\pi} \, (3x-2) \qquad x < 1 \qquad (3a)$$

$$Q = \frac{(\Delta B)^2}{6\pi} \frac{1}{x} \qquad x > 1 \qquad (3b)$$

These expressions have a maximum at $x = 3/4$, where $Q = 3(\Delta B)^2/16\pi$, and they fall to zero as $x \to 0$ or ∞. The entire dependence is plotted in Fig. 1. To minimize Q for given ΔB, there are two strategies; make x either small or large, but not near unity.

Considering numerical values, we see that it is hard to get $x \ll 1$ for small values of ΔB. For example, if $J_c \simeq 1.5 \times 10^5$ amp/cm^2 and $d \simeq 10\mu$, $H_s \simeq 100$ Gauss, and $x \simeq 1$ for the typical value $\Delta B = 100$ Gauss. It is hard to fabricate wire with d much less than 10μ; moreover, unless the wire is twisted with a very short pitch, the effective d is larger than the nominal value because of the coupling effects discussed in ref. 1. This suggests that lower loss can be obtained with the other choice, namely by making d (and hence x) as _large_ as possible. But the requirement for stability against thermal fluctuations sets an upper limit on d. For adiabatic stability, the limit on d is

$$d^2 \lesssim \frac{3}{\pi} \frac{c^2 C}{J_c(-dJ_c/dT)} \tag{4}$$

where C is the specific heat per unit volume. For typical parameter values this limits d to values $\sim 200\mu$ or below, for which $H_s \sim 2000$ Gauss.

In Fig. 2, we plot the dissipation per cycle per unit volume for $H_s = 10^2$ and 10^3 Gauss, which correspond roughly to the smallest and largest relevant filament diameters ($d \simeq 10\mu$ and 100μ). Clearly the larger filaments give lower losses for $\Delta B \lesssim 500$ Gauss, whereas the smaller filaments are superior for larger AC amplitudes. Thus, it is not surprising that the

smallest possible filaments are desirable for applications such as pulsed magnets where $\Delta B \simeq B_{max} \simeq H_{c2}$ >> 500 Gauss. On the other hand, the thickest thermally stable filaments should be best for applications such as levitation or rotating machinery, in which the field is basically DC with AC ripple less than a few hundred Gauss.

We conclude with a word of caution: This analysis is highly oversimplified in that it is based on a model of isolated superconducting lamina, exposed to a changing magnetic field which is equal on both sides. The true situations are more complex and varied, since large transport currents are present and the field configurations are quite diverse. These complications can have important effects, as is suggested by the identification of loss components varying with frequency ω as $\omega^{\frac{1}{2}}$ and ω^2, and with $(\Delta B)^2$ in experiments at Westinghouse[4]. For comparison, simple hysteresis losses of the sort discussed here vary as ω and as ΔB or $(\Delta B)^3$. In view of these realities, our point is a modest one; different applications will have different optimum conductors. In particular, it is not even true that it is always a good thing to reduce the superconducting filament diameter.

Acknowledgement

References

1. M. Tinkham, J. Appl. Phys. $\underline{44}$, 2385 (1973).

2. T. K. Hunt, private communication; also J. R. Reitz, R. H. Borcherts, L. C. David, T. K. Hunt, D. F. Wilkie, Ford Motor Report No. FRA-RT-73-27 to Dept. of Transportation (1973).

3. C. P. Bean, Phys. Rev. Lett. $\underline{8}$, 250 (1962).

4. M. S. Walker, Y. W. Chang. J. H. Murphy, and H. E. Haller III, Cryogenic Engineering Conference, Atlanta, August 1973.
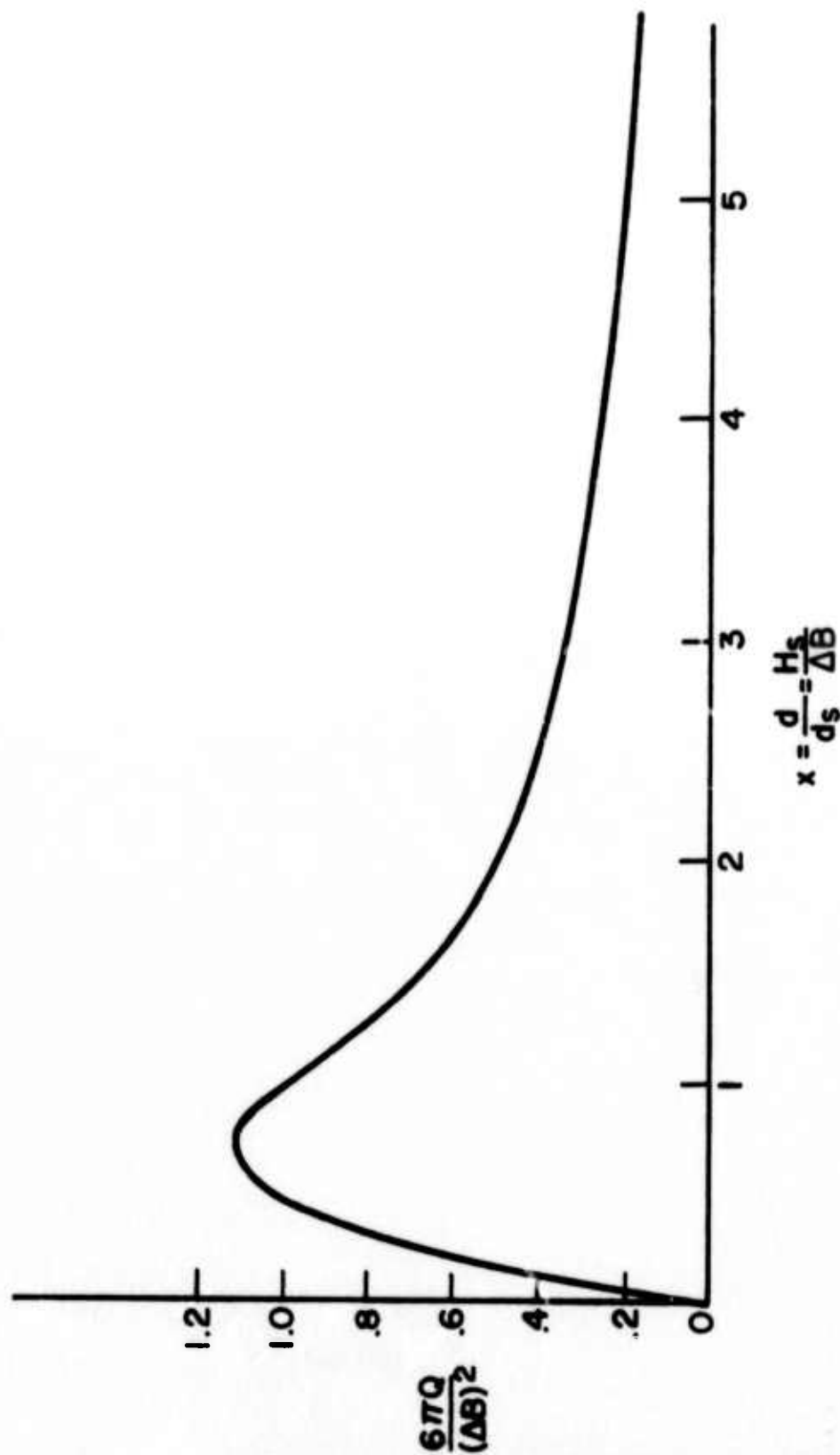
Figure 1. Normalized dissipation per cycle as a function of normalized filament diameter d.
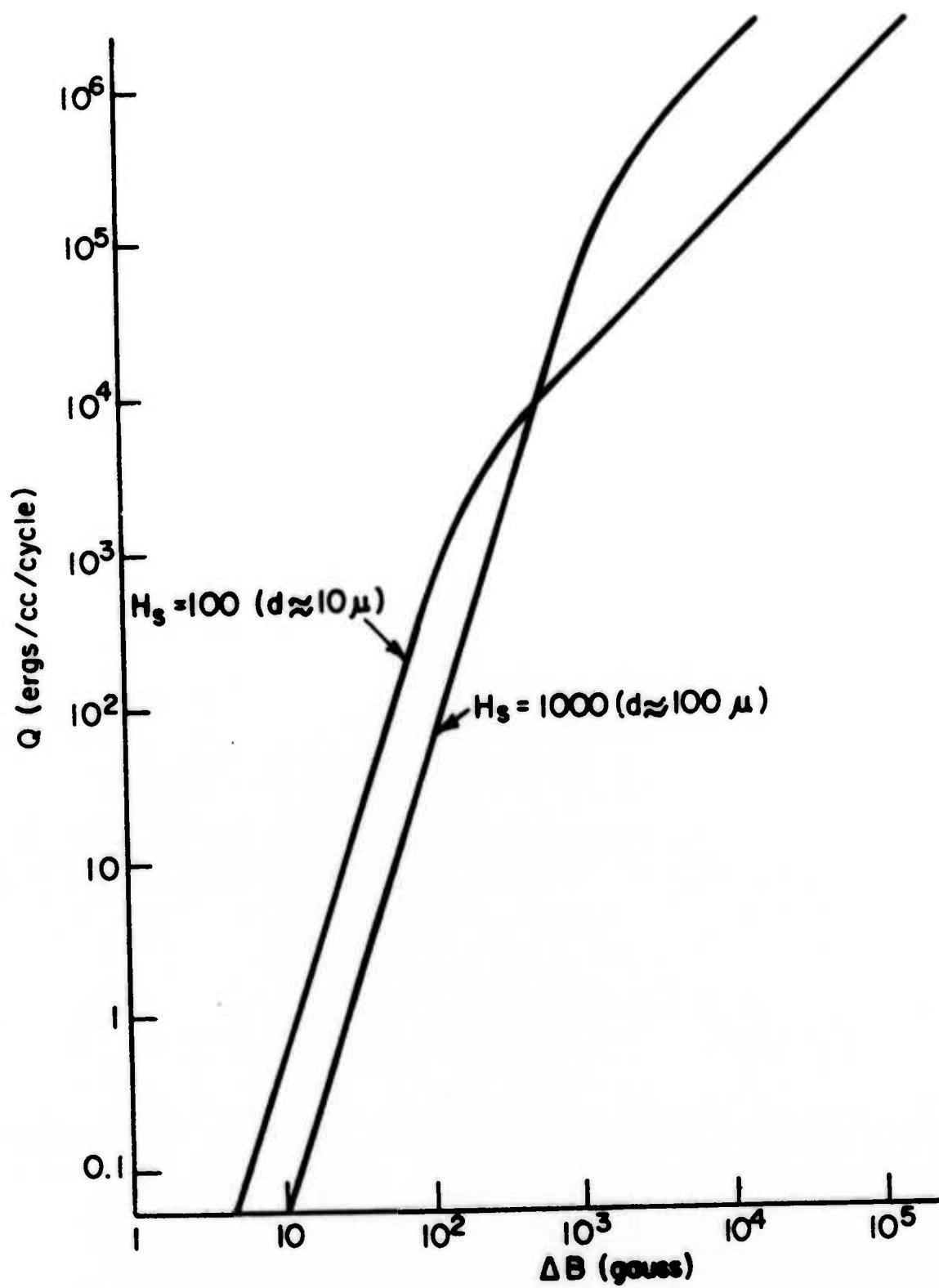
Figure 2. Dissipation per cycle for two different filament diameters.

-134-

COMMENTS ON CERTAIN REPORTS OF
CARBONACEOUS LINEAR POLYACETYLENES

M. F. Hawthorne

Abstract

During the past ten years Kasatochkin and coworkers have studied the carbonaceous products obtained by oxidative-polymerization of acetylene.

$$n \ HC{\equiv}CH \ \xrightarrow[\text{[O]}]{Cu^{+2}} \ H\text{-}(\text{-}C{\equiv}C\text{-})_n\text{-}H$$
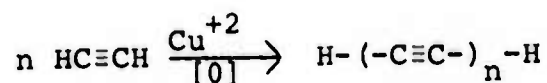
Examination of the original literature discloses that no firm basis exists for formulation of the reaction products as a linear, polymeric acetylene ("Carbyne" or "Carbine"). Additional work should be carried out to carefully characterize the reaction products.

COMMENTS ON CERTAIN REPORTS OF
CARBONACEOUS LINEAR POLYACETYLENES

M. F. Hawthorne

During the past ten years Kasatochkin[1-5] and coworkers have reported the synthesis of carbon-like materials from the oxidative-polymerization of acetylene. Acetylene and Cu^{+2} salts were reacted followed by oxidation of the resulting copper acetylides to produce carbonaceous materials. This synthetic procedure is not novel in that other chemists have employed this method to couple terminal acetylenes during the past twenty years. Perhaps the most striking examples of previous work have been provided by Sandheimer in his classical acetylene cyclization reactions.

$$3HC\equiv C-CH_2CH_2-C\equiv CH \xrightarrow[\text{[O]}]{Cu^{+2}}$$
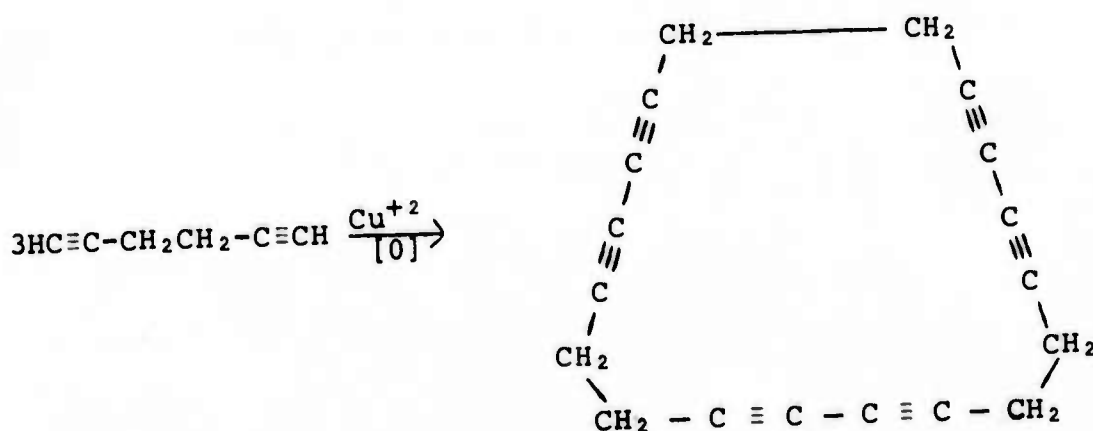
The Kasatochkin reactions were intended to proceed as

follows:

$$n \ HC \equiv CH \xrightarrow[\text{[O]}]{\text{Cu}^{+2}} H-(-C\equiv C-)_n-H$$
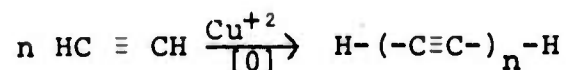
Crosslinking of carbon chains, aromatization and isomerization to cumulene (=C=C=C=) structures were thought to be processes which accompanied polyacetylene ("carbyne" or "carbine") formation. Consequently, attempts were made to identify microscopic regions in the largely amorphous product which contained such structural characteristics.

Examination of the data available[1-4] to the writer strongly suggests that the carbonaceous material prepared from acetylene has not been properly characterized.

Polyacetylenes are highly reactive species which are unstable with respect to elemental carbon and are consequently found to be explosive. The Kasatochkin products were not explosive and appeared to have no unusual chemical properties. The only evidence cited which supports a polyacetylene or cumulene microstructure is found in the infrared spectrum of these materials which contains a strong absorption band between 1900 and 2100 cm^{-1}. This region of the spectrum is characteristic of simple, unsymmetrically substituted acetylenes and cumulenes. Other physical measurements which were made include X-ray powder photographs, electron microscopy, density determinations, etc. None of these methods, except X-ray powder photographs are pertinent to microstructure and the latter show

some evidence of crystalline regions mixed with amorphous materials.

Chemically, the products of the acetylene polymerization process were exposed to extremely severe treatment before physical characterization; namely boiling with concentrated hydrochloric acid, treatment with aqueous ammonia followed by a water wash. Since these manipulations were probably carried out in the air, considerable oxidative degradation of the acetylenic or cumulene structures could have occurred. Finally, the products were heated in the absence of air to 1000°C. The writer contends that the initial product of the acetylene polymerization reactions may have been structurally modified to such an extent that artifacts were observed using physical characterization methods.

Chemical characterization of the final products was attempted using ozonization for multiple bond cleavage.

$$-C\equiv C- + O_3 \xrightarrow{H_2O} \quad \overset{\displaystyle -C = C-}{\underset{\displaystyle \underset{O}{\overset{O \quad O}{\diagdown \diagup}}}{|\quad\quad|}} \longrightarrow \begin{array}{c} COOH \\ COOH \end{array}$$

$$=C=C=C= \quad \xrightarrow[H_2O]{O_3} \quad CO_2 + H_2O_2$$
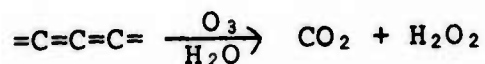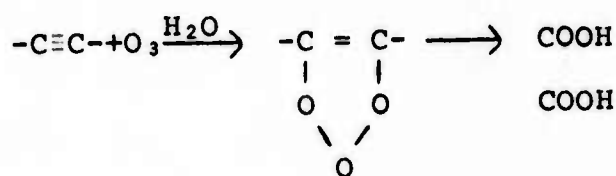
No analysis or yields of products were published except to say that oxalic acid and CO_2 were reaction products. This is quite inconclusive with regard to the true structure of the polymerization product.

-138-

Although other criticisms could be leveled against the reported work, the recommendation may be made that this research should be repeated under scrupulously mild conditions and that chemical work should be performed in a quantitative fashion. Lastly, the now available tool, core electron spectroscopy (ESCA), could be used to determine the number of carbon atom types and their relative populations present in the product. The latter suggestion is based upon the assumption that sp^3, sp^2 and sp (in acetylenic and cumulene bonds) all give discrete ESCA energies.

Acknowledgement

References

1. V. I. Kasatochkin, O. I. Egorova and Y. G. Aseev, Dok. Akad. Nauk, SSSR, $\underline{151}$, 125 (1963).

2. V. I. Kasatochkin, A. M. Sladkov, Y. P. Kudryavtsev, N. M. Popov and V. V. Korshak, ibid, $\underline{177}$, 358 (1967).

3. A. M. Sladov, V. I. Kasatochkin, Y. P. Kudryavtsev and V. V. Karshak, Izvest. Akad. Nauk SSSR, Ser. Khim., $\underline{12}$, 2697 (1968).

4. V. I. Kasatochkin, T. M. Babchinitser, Y. P. Kudryavtsev, A. M. Sladkov and V. V. Korshak, Dok. Akad. Nauk, SSSR, $\underline{184}$, 1969.

5. V. I. Kasatochkin, V. V. Korshak, Y. P. Kudryavtsev, A. M. Sladkov and I. E. Sterenberg, Carbon, $\underline{2}$, 70 (1973).

WORLD WAR II GERMANY AS A MODEL
FUEL-POOR ECONOMY. SYNTHETIC FUELS

M. F. Hawthorne

Abstract

It is suggested that World War II Germany faced an
energy crises in fuels not unlike the crisis which faces the
U.S. at the present time. The German chemical industry un-
doubtedly provided large quantities of synthetic fuel prepared
from limestone, trees, coke, etc., through acetylene as an
intermediate. It is proposed that old reports relating to this
subject be reexamined in light of today's available chemical
technology and fuel requirements.

WORLD WAR II GERMANY AS A MODEL
FUEL--POOR ECONOMY.   SYNTHETIC FUELS

M. F. Hawthorne

It has occurred to the writer that the present energy shortage which faces the United States may in some ways resemble the energy crisis which faced Germany during World War II.   One obvious difference between these two situations is the fact that the U.S. is now pollution conscious while the 1939-1945 Germany was a country at war and, as such, placed no restraints upon the pollution of the environment.   However, the tenuous supply of oil available to WW II Germany from the Balkans is not unlike our present dependence upon the Middle Eastern countries for this same commodity.

The writer has the impression that a fair percentage of Germany's fuel requirements were met by their chemical industry in the form of synthetic fuels.   Ironically, some of these fuels would appear to be "cleaner" than the hydrocarbon fuels cherished by the U.S. and Western Europe.   As an example, methanol (wood alcohol) was probably used as a vehicular fuel in internal combustion engines.   Methanol was probably produced from CO and H_2 which, in turn, were products of the well-known water--gas reaction.   Wood or coal

$$H_2O + C \xrightarrow{\text{heat}} H_2 + CO$$

$$2CO + 3H_2 \xrightarrow{\text{catalyst}} 2CH_3OH$$
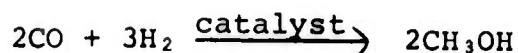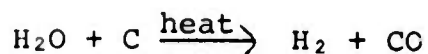
(after coking) would serve as the source of carbon. Along the same line, the Germans developed an entire chemical industry based upon acetylene technology. Acetylene was available through carbon, limestone and heat via calcium

$$CaCO_3\text{(limestone)} + 3C \xrightarrow{\text{heat}} CaC_2 + CO + CO_2$$

carbide (CaC_2).

$$CaC_2 + 2H_2O \longrightarrow Ca(OH)_2 + HC \equiv CH$$
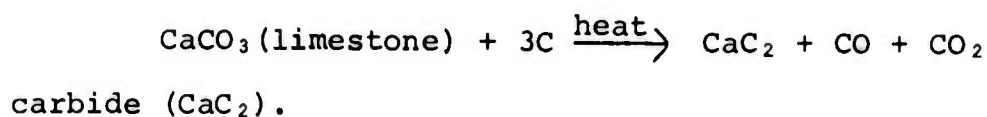
Calcium hydroxide could be recycled via calcium oxide.

$$Ca(OH)_2 \xrightarrow{\text{heat}} H_2O + CaO$$
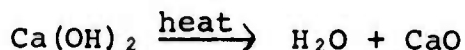
Among the fuel products available from acetylene were ethanol (C_2H_5OH), unsaturated hydrocarbons (which could be hydrogenated to alkanes) and aromatic hydrocarbons.

Following the war, studies were made of the German chemical industry and compiled in a series of U.S. Government reports. It is the writer's thought that those reports might very well yield information, which when updated in terms of present technology, would give us some guidance as to the feasibility of clean synthetic fuel production in the United States. Regardless of the application of esoteric energy resources such as solar energy, clean organic fuels will always

be required.   The question is merely one of price and extent
of use.

Acknowledgement

HYDROGEN TRANSFER BY DIFFUSION THROUGH
A HYDROGENATION-DEHYDROGENATION SYSTEM

M. F. Hawthorne and Howard Reiss

Abstract

Calculations were carried out to explore the feasibility
of hydrogen transfer by the following means. Hydrogenation of
an alkene, diffusion of the resulting alkane through a pipeline
to the delivery point and dehydrogenation of the alkene to pro-
duce hydrogen gas. Alkene produced at the delivery point then
diffuses through the same pipeline to the hydrogen pickup point.
The conclusion is reached that the concept is probably not
feasible for long distances (100 miles), but could be used in
much shorter lines or by using a two-line pumped flow system to
transport alkene to the hydrogenator and alkane to the dehydro-
genator delivery point.

HYDROGEN TRANSFER BY DIFFUSION THROUGH
A HYDROGENATION-DEHYDROGENATION SYSTEM

M. F. Hawthorne and Howard Reiss

The transportation and storage of hydrogen fuel for commercial and home heating as well as its use in the thermal generation of electrical energy poses many problems. However, several options are available which range from finding a solution to the problem of metal embrittlement by H_2 in high pressure (600 psi) pipelines to the transportation of a liquid or solid hydrogen precursor which would liberate H_2 upon mild thermolysis, etc. In this paper we explore another alternative to H_2 transport in which diffusion of a hydrogen precursor and a hydrogen absorber, which is formed by loss of H_2 from the precursor, diffuse through a filled pipe which serves as a reaction vessel. In this way H_2 gas is present within the pipe in very low concentration and undesireable side effects of H_2-metal reactions are mitigated. One could also think of such a system as a H_2 gas "relay" in which little or no H_2 gas appears within the closed system.

As a model chemical system we have chosen the simple hydrogenation-dehydrogenation reaction of aliphatic hydrocarbons in the presence of

$$\begin{array}{c} \text{H} \\ | \\ -\text{C}-\text{C}- \\ | \\ \text{H} \end{array} \quad \overset{\text{Catalyst}}{\underset{K=10^{-25}\ \frac{\text{moles}}{\text{liter}}}{\rightleftarrows}} \quad \text{>C=C<} + \text{H}_2$$

Alkane   Alkene
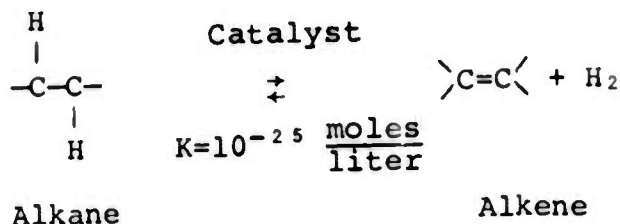
an appropriate noble metal catalyst. Chemical equilibrium is maintained throughout the length of the reaction vessel by means of catalyst present in the form of wire, surface coating on the interior of the pipe or by the use of regularly spaced catalytic baffles. Since the dehydrogenation of a simple alkane to produce H_2 and alkene is endothermic by about 30 Kcal/mole, thermal energy must be supplied to the H_2 entry site and could be put to use.
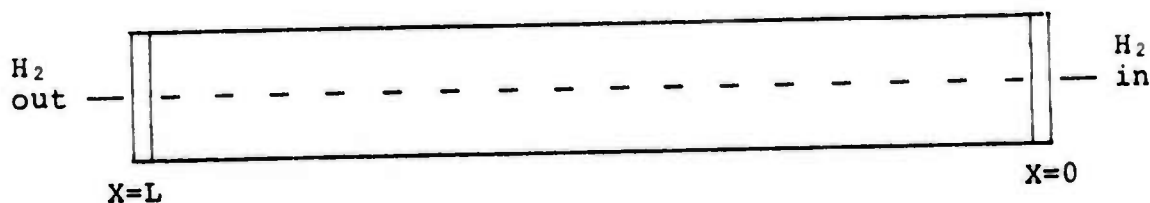
Suppose we could separate H_2 through the semipermeable barrier at an arbitrarily fast rate, ω. Also, suppose there is instantaneous chemical equilibrium everywhere throughout the pipe, so that

$$\frac{C_1 C_3}{C_2} = K \tag{1}$$

where $C_1 = [H_2]$, $C_2 = [\text{alkane}]$

$C_3 = [\text{alkene}]$

-147-

Mechanical equilibrium requires

$$C_1 + C_2 + C_3 = \alpha = \text{constant} \tag{2}$$

We suppose that K is sufficiently small so that

$$C_1 \ll C_2 \tag{3}$$
$$C_1 \ll C_3$$

The flux of alkane is

$$-D_2 \frac{\partial C_2}{\partial x} + v(x) C_2 = J_2 \tag{4}$$

where $v(x)$ is a common convective velocity. Now, since only H_2 passes through the ends of the pipe it follows that

$$J_2 + J_3 = 0 \tag{5}$$

where J_3 is the flux of alkene

$$J_3 = -D_3 \frac{\partial C_3}{\partial} + v(x) C_3 \tag{4a}$$

Furthermore, since the bulk of the H_2 is transported as part of the alkane (because of Eq. (3)) we have, in the steady state

$$J_2 \simeq \omega \tag{6}$$

Now, because of Eqs. (2) and (3)

$$C_2 + C_3 \simeq \alpha \tag{7}$$

Substitution of this equation in Eq. (1) gives

$$\frac{C_1(\alpha - C_2)}{C_2} = K, \quad \frac{C_1 C_3}{\alpha - C_3} = K, \tag{8}$$

from which we derive

$$C_2 = \frac{\alpha C_1}{K + C_1}, \quad C_3 = \frac{\alpha K}{K + C_1} \tag{9}$$

Differentiating Eq. (9) we get

$$\frac{dC_2}{dx} = \frac{\alpha K}{(K + C_1)^2} \frac{dC_1}{dx} \quad \frac{dC_3}{dx} = -\frac{\alpha K}{(K + C_1)^2} \frac{dC_1}{dx} \tag{10}$$

Substituting the first of equations (9) and (10) in Eq. (4) and the latter into Eq. (6)

$$\omega = -D_2 \frac{\alpha K}{(K + C_1)^2} \frac{dC_1}{dx} + v(x) \frac{\alpha C_1}{K + C_1} \tag{11}$$

Substituting Eqs. (9) and (10) into Eq. (4) and Eq. (4a), and the results into Eq. (5), gives

$$-D_2 \frac{\alpha K}{(K + C_1)^2} \frac{dC_1}{dx} + v(x) \frac{\alpha C_1}{K + C_1}$$

$$+D_3 \frac{\alpha K}{(K + C_1)^2} \frac{dC_1}{dx} + v(x) \frac{\alpha K}{K + C_1} = 0 \tag{12}$$

or

$$(D_2 - D_3) \left[\frac{K}{(K + C_1)^2}\right] \frac{dC_1}{dx} = v(x) \tag{13}$$

Substituting this equation into Eq. (11) gives

$$\omega = -D_2 \frac{\alpha K}{(K + C_1)^2} \frac{dC_1}{dx} + \frac{\alpha C_1}{K + C_1} \left\{ (D_2 - D_3) \left[\frac{K}{(K + C_1)^2}\right] \frac{dC_1}{dx} \right\} \tag{14}$$

$$= \frac{\alpha K}{(K + C_1)^2} \left\{ -D_2 + (D_2 - D_3) \frac{1}{K + C_1} \right\} \frac{dC_1}{dx}$$

This can be rearranged to

$$\{\frac{D_2 K + D_3 C_1}{(K+C_1)^3}\} \, dC_1 = - \frac{\omega}{\alpha K} \, dx \tag{15}$$

This can be integrated to yield

$$\frac{(D_2 - D_3)K}{2(K+C_1)^2} + \frac{D_3}{K+C_1} = \frac{\omega}{\alpha K} \, x + A \tag{16}$$

where

$$A = \frac{(D_2 - D_3)K}{2(K+C_1^0)^2} + - \frac{D_3}{K+C_1^0} \tag{17}$$

where C_1^0 is the concentration of C_1 (i.e., $[H_2]$) <u>inside the barrier</u> at <u>x = 0</u>. Eq. (16) can be solved for C_1

$$C_1 = \frac{D_3}{(\frac{\omega x}{\alpha K} + A)} + \{\frac{D_3^2}{4(\frac{\omega x}{\alpha K} + A)} + \frac{(D_2 - D_3)K}{2(\frac{\omega x}{\alpha K} + A)}\}^{\frac{1}{2}} \tag{18}$$

For the special case, $D_2 = D_3 = D$, i.e., no convection, we get

$$C_1 = \frac{D}{\frac{\omega x}{\alpha K} + A} \tag{19}$$

where now

$$A = \frac{D}{K+C_1^0} \tag{20}$$

Thus, in this case,

$$C_1 = \frac{D}{\frac{\omega x}{\alpha K} + \frac{D}{K+C_1^0}} \tag{21}$$

Assume

$K = 10^{-25}$ moles/liter

$\alpha = 5$ moles/liter ~ 100 atm

$$D \simeq 10^{-1} \ cm^2/sec$$

$$x = L \ (pipe \ length) = 10^7 \ cm \simeq 100 \ miles$$

Assume that we wish to pull H_2 at a rate

$$\omega = 1 \ mole/cm^2 \ sec.$$

and that we wish to maintain C_1^0 no higher than 0.01 atm. $\sim 5 \times 10^{-4}$ moles/liter. Then we would require, according to Eq. (21), $C_1(L)$, at the delivery end, to be

$$C_1(L) = \frac{10^{-1}}{\dfrac{10^7}{10^{-25} \times 5} + \dfrac{10^{-1}}{10^{-25} + (5 \times 10^{-4})}}$$

$$\simeq 5 \times 10^{-33} \ moles/liter \tag{22}$$

This is a fraction of a molecule per liter - an impossible vacuum even if one could draw H_2 through the barrier as fast as he chose to.

All may not, however, be lost. Impellar blades might mix the components so as to give an effective diffusion co-efficient 10^{10} times as large as the natural coefficient. This would allow a $C_1(L) \sim 10^{-23}$ moles/liter. This is still small, but could be selecting a system with a larger K. Clearly the major problem is $K \sim 10^{-25}$. If one could arrive at a chemical system with $K \sim 10^{-10}$ and all the other problems could be solved, a viable transport system might result.

An alternative to the diffusion transport scheme would make use of two pipes; one containing alkane feed to the dehydrogenator unit and the other containing alkene feed to the hydrogenation unit. Heat evolved at the hydrogenator could be put to use and the heat required at the dehydrogenator could be supplied by burning 10 to 20% of the clean alkane feed. It seems probable that hydrogen could be transported by this means with a net energy gain if fluid-flow energy losses were minimized to the greatest possible extent.

Acknowledgement

PHASE TRANSITIONS AND RESISTANCE IN
THREE-DIMENSIONAL ARRAYS OF SUPERCONDUCTING
LINEAR CHAINS

M. Tinkham

Abstract

We consider a model solid made up of superconducting unit cells coupled by Josephson coupling energies J_x, J_y, and J_z in the three orthogonal directions. A collection of one-dimensional chains is simulated by making $J_z \gg J_\perp \simeq J_x$, J_y. Our model is a generalization of that of Anderson, Lee, and Saitoh, which considers only the intrachain coupling J_z. Short range order gradually develops along the chains as T is lowered below $T_c^* = 2J_z/k$, the approximate mean-field critical temperature, but a true phase transition into a state with long-range order in 3-D does not occur until a much lower temperature $T_c \simeq T_c^* (J_\perp/J_z)^{\frac{1}{2}}$. For TTF-TCNQ, Anderson et al have estimated $T_c^* = 500°K$; if we take $J_\perp/J_z = 10^{-3\pm1}$, as suggested by the conductivity anisotropy, this gives $T_c = 5-50°K$, a temperature range below the transition to an insulating state which occurs at $\sim60°K$.

The conductivity of a single chain should rise as $(T/T_c^*)e^{T_c^*/T}$ as T decreases. To reach a value 10^6 times that of room temperature copper from that observed in TTF-TCNQ would

require cooling to ∿20°K if the 3-D superconducting transition
to a state of zero resistance did not occur first. In fact, the
metal-insulator transition intervenes. It is also shown that
the difference in conductivity between the ordinary and special
samples of TTF-TCNQ could be roughly accounted for if 1% of the
junctions in the ordinary samples were much weaker than the others.

PHASE TRANSITIONS AND RESISTANCE IN THREE-DIMENSIONAL ARRAYS OF SUPERCONDUCTING LINEAR CHAINS

M. Tinkham

Introduction

It is a well-known theorem of statistical physics that a one-dimensional system can not have a phase transition into a phase with long-range order, such as a classic superconducting phase. It is also well-known, however, that <u>very</u> low resistance <u>can</u> be obtained in superconducting filaments thin enough to be formally one-dimensional [i.e., with diameter $d \gg \xi(T)$]. Roughly speaking, instead of being mathematically sharp, the transition has a fractional width $\Delta T/T_c^* \simeq a^2/A$, where a is the inter-electronic spacing, A is the cross-sectional area, and T_c^* is the mean field T_c. Thus, for a filament of monomolecular cross section, the transition extends down to $T \ll T_c^*$, and unless T_c^* is an order of magnitude higher than the operating temperature, usefully high conductivity will not be obtained. These points, anticipated by Little[1], were elaborated in more modern form in an earlier memorandum.[2]

Now let us shift our attention from isolated filaments to a solid material containing many parallel filaments. If we <u>assume</u> that the high and apparently divergent conductivity of

TTF-TCNQ is a manifestation of superconductivity, we may take it as an example. Evidently there is some coupling between the molecular chains because there is non-zero transverse electrical conductivity. It appears that at room temperature $\sigma_\perp$ is typically 1 (ohm-cm)$^{-1}$, 100-1000 times smaller than $\sigma_\parallel$, but not zero. The intrinsic value of this conductivity could be obscured by sample imperfections and limitations in the measurement techniques, but the values seem reasonably consistent from sample to sample. Even if it were due to localized "shorts" rather than uniform coupling throughout the volume, its effect would be macroscopically similar so long as the density of shorts were high enough to average out over small volumes. Given any distributed interchain coupling, there will be a real 3-D phase transition, but at a much lower temperature T_c than the mean field T_c^* mentioned above. The purpose of this paper is to explore a simple model which illustrates the implications of this weak interchain coupling.

The Model

We consider a model of a solid made up of superconducting unit cells of dimension a_x, a_y, a_z, coupled together by Josephson coupling energies J_x, J_y, and J_z in the three orthogonal directions. To simulate a collection of pseudo-one-dimensional chains, we make $J_z \gg J_x, J_y$. Our model is a generalization of that of Anderson, Lee and Saitoh[3], which considers only the intrachain coupling J_z. Following Anderson et al, we further assume that each element is strongly superconducting in the temperature range

of interest. In this case, we can describe the superconducting order parameter by simply giving the phase factor $e^{i\phi_i}$ in each unit, taking the amplitude to have a limiting constant value. The system interaction energy can then be written as

$$E = \sum_{ij} J_{ij} \cos(\phi_i - \phi_j) \tag{1}$$

where the sum extends over nearest neighbor pairs, and $J_{ij} = J_x$, J_y, or J_z depending on the bond direction.

Clearly, the lowest energy state of (1) has all ϕ_i the same, i.e., long-range phase coherence. But at finite temperatures, there will be phase fluctuations which tend to disrupt this long range order. In a one-dimensional system it turns out that these fluctuations cause $(\Delta\phi)^2 = \langle(\phi_i - \phi_j)^2\rangle$ to diverge with the distance between i and j; whereas in 3-D, long range phase order is possible below a critical temperature T_c. Let us review these matters to allow an estimate of T_c for the case of weakly coupled filaments.

Estimate of T_c

Because of the dominance of the one-dimensional coupling J_z, let us first consider the fluctuations in a single chain. For the low-energy, long-wavelength fluctuations which dominate near T_c, we can expand the cosine in (1), and for a chain of $N(\gg 1)$ units write it as

$$E \simeq -NJ_z + \tfrac{1}{2}J_z \sum_i (\phi_i - \phi_{j+1})^2 \tag{2}$$

-157-

For fluctuations with $\phi = k_z z = (\Delta\phi/Na_z)z$, the energy increase is

$$\Delta E_{1-D} = N\ \frac{J_z}{2}\left(\frac{\Delta\phi}{N}\right)^2 = \tfrac{1}{2}J_z\ \frac{(\Delta\phi)^2}{N} \tag{3}$$

If we set $\Delta E = kT$, this leads to the conclusion that

$$(\Delta\phi)^2 = 2NkT/J_z \tag{4}$$

i.e., that $\Delta\phi$, the fluctuation in relative phases at the ends of the chains diverges as the square root of the length of the chain. Taking $\Delta\phi = \tfrac{1}{2}\pi$ to indicate phase randomization, the phase coherence extends for roughly

$$N_{\pi/2} = \frac{\pi^2}{8}\ \frac{J_z}{kT} \simeq \frac{J_z}{kT} \tag{5}$$

units along the chain. A more rigorous calculation, including an arbitrary collection of long-wave thermal fluctuations in an infinite chain, reproduces this result within a small numerical factor. Thus, we may conclude that there is a gradual growth of the coherence range for short range order as T is reduced, but no long range order until T = 0.

If the same argument were applied to a 3-D system, an N×N×N array of cells with isotropic coupling J, the corresponding energy increase would then be

$$kT = \Delta E_{3-D} = N^3\ \frac{J}{2}\left(\frac{\Delta\phi}{N}\right)^2 = \tfrac{1}{2}NJ(\Delta\phi)^2 \tag{6}$$

so that the fluctuations get smaller rather than larger as

$N \rightarrow \infty$. Thus, long range order is readily achieved. The two-dimensional case is intermediate, with ΔE independent of N for this model; the existence of a rigorous phase transition is model dependent, but one occurs for models which contain any feature to give finite energy to long wave fluctuations.

Returning to the anisotropically coupled system, the physical requirement for the 3-D long range ordering to set in is that the interaction between a coherent region in one chain with similar regions in adjacent chains be strong enough to over-come the thermal energy kT. That is, the phase transition sets in at a T_c given roughly by

$$kT_c \simeq 4N_{\pi/2} \; J_\perp \simeq \frac{4J_z}{kT_c} \; J_\perp$$

where 4 represents the number of neighbor chains. Thus

$$kT_c \simeq 2(J_z J_\perp)^{\frac{1}{2}} \tag{7}$$

For comparison, note that the crude mean field approximation would (incorrectly) give a transition to a long-range ordered state even in a single one-dimensional chain at

$$kT_c^* = 2J_z \tag{8}$$

This value would also be essentially the mean field prediction for the array of chains, since inclusion of the interchain coupling would simply add $2J_x + 2J_y$ to the right hand side, and we assume $J_z \gg J_x + J_y$. Thus the true T_c is less than the

mean field value T_c^* by a ratio

$$T_c/T_c^* \simeq (J_\perp/J_z)^{\frac{1}{2}} \tag{9}$$

Attempting to apply this result to TTF-TCNQ, if we estimate $J_\perp/J_z \simeq \sigma_\perp/\sigma_\shortparallel \simeq 10^{-3\pm1}$, this T_c would be some 10-100 times lower than T_c^*. With $T_c^* \simeq 500°K$ as estimated by Anderson et al from the data, this would give $T_c \simeq 5-50°K$, well below the temperature ($\sim58°K$) of the transition to an insulating state. This result would be consistent with the failure of Coleman et al[4] to observe any real superconductivity. Evidently the numerical values used here are poorly known and are for illustrative purposes only.

Appearance of Resistance

Once the 3-D order has set in at T_c, the resistance (for low currents and fields) should vanish entirely. But what behavior should be expected in the range from T_c to T_c^* and beyond? One might think that phase fluctuations over a length like $N_{\pi/2} a_z$ would give resistance, because of the loss of phase equality between the ends. In fact, this idea was once suggested by Parks and coworkers for the case of superconducting microbridges, but experimental data on 1-D tin whiskers have shown this view to be wrong. Rather, the whisker data confirm the concept proposed by Little[1] and elaborated by Langer and Ambegaokar and by McCumber and Halperin (LAMH)[5]. Namely, resistance occurs when at some point the superconducting phase coherence is ruptured (typically

by having $|\psi| \to 0$) and reforms with a possible irreversible "phase slip" of $\pm 2\pi$. In our present model, this can only occur at one of the Josephson couplings; the energy barrier for such a phase slip (at low currents) is $2J_z$, since $(\phi_i - \phi_{i+1})$ must go through π while moving from ~ 0 to $\sim 2\pi$. With an attempt frequency Ω at each junction, the phase decoupling rate per strand of length L will be $\sim \Omega(L/a_z)e^{-2J_z/kT}$. Following the arguments of LAMH[5], this leads to a resistance for the filament of

$$R = \Omega \frac{L}{a_z} \frac{\pi \hbar^2}{2e^2 kT} e^{-2J_z/kT} \qquad (kT << 2J_z) \qquad (10)$$

Thus, the effective conductivity is

$$\sigma_z = \frac{a_z}{a_x a_y} \frac{1}{\Omega} \frac{2e^2 kT}{\pi \hbar^2} e^{2J_z/kT} \qquad (kT << 2J_z) \qquad (11)$$

Because of the special localized fluctuations required to produce resistance, this formula corresponds to a carrier mean free time which rises _exponentially_ with J_z/kT. This temperature dependence should be contrasted with the only linear rise with J_z/kT of the phase coherence range $N_{\pi/2} a_z$ against long-wave, low-energy fluctuations. The latter retard the onset of 3-D order and the perfect conductivity associated with it, but they do not directly produce resistance. It should further be noted that for $T > T_c$, the interchain couplings are so weak that this one-dimensional result should not be seriously in error.

To the extent that the Josephson junctions resemble classical tunnel junctions, one can relate J_z to the normal state

conductance, as follows: The Josephson coupling energy $J_z = (\hbar/2e)$ $I_0 = (\pi\hbar/4e^2)\Delta/R_n$, where I_0 is the maximum Josephson current at low temperatures, Δ is the low-temperature energy gap, and R_n is the normal resistance of the junction. If we take the normal resistance to be governed by the junctions, neglecting any contribution from within each unit, then $R_n = a_z/\sigma_n a_x a_y$. Combining these relations

$$\sigma_{zn} = \frac{a_z}{a_x a_y} \frac{4e^2}{\pi\hbar\Delta} J_z \tag{12}$$

Thus

$$\frac{\sigma_z}{\sigma_{zn}} = \frac{kT\Delta}{\hbar\Omega 2J_z} e^{2J_z/kT} \tag{13}$$

We presume that $\Delta \simeq kT_c^* \simeq 2J_z$. Moreover, for $T < T_c^*$, we follow Anderson et al and take $\hbar\Omega \simeq kT_c^*$, so that

$$\frac{\sigma_z}{\sigma_{zn}} = \frac{T}{T_c^*} e^{T_c^*/T} \qquad\qquad T < T_c^* \tag{14}$$

which is essentially the result of Anderson et al. On the other hand, for $T > T_c^*$, we presume that $\hbar\Omega \simeq kT$, so that $\sigma_z/\sigma_{zn} \to 1$ as T becomes much greater than T_c^*.

Effect of Chain Defects

We consider two classes of chain defects: (1) those which simply correspond to a reduced coupling constant $J_z' < J_z$ between two successive units, and (2) those in which the coupling is essentially zero, forcing current to flow through the low $\sigma_\perp$ to other chains to bridge around the break.

In case (1), if one assumes that the dominant effect is through the exponential activation factor, one can see from (10) that each weak link contributes $e^{2(J_z-J_z')/kT}$ times as much resistance as a normal one. If one link out of n_ℓ is such a weak one, the observed resistance will be increased by a factor of

$$\frac{R}{R_{ideal}} = 1 + \frac{e^{2(J_z-J_z')/kT}-1}{n_\ell} \tag{15}$$

This will reach a factor of 2 when

$$n_\ell + 1 = e^{2(J_z-J_z')kT} \tag{16}$$

The $\sigma_{\shortparallel}$ data of Coleman et al on "typical" samples are found to fall a factor of 2 below the "special" ones at $T \simeq 90°K$. This could be fitted by (16) if $n_\ell \simeq 100$ and $T_c^{*\prime} = 2J_z' = 100°K$, while $T_c^* = 2J_z = 500°K$. This parameter choice would also account for the general trend of the difference between the two different temperature dependences, but the agreement degenerates for the very high conductivities near 60°K.

In case (2), when there is a complete break in one filament, what is the bridging resistance through $\sigma_\perp$ to other filaments and back on the other side of the break? A rough estimate shows that the optimum degree of current spreading is to $\sim 2.7 r_o$ $[r_o \simeq \frac{1}{2}(a_x a_y)^{\frac{1}{2}}$ ... the filament radius] over a length $2(\sigma_{\shortparallel}/\sigma_\perp)^{\frac{1}{2}} r_o$ on either side of the break, giving an extra resistance equivalent to a length $\ell \simeq \pi r_o (\sigma_{\shortparallel}/\sigma_\perp)^{\frac{1}{2}}$ of filament. Note that $\ell/a_z \simeq 30-100$ at room temperature, but ~ 1000 near 50° in the special crystals.

The corresponding numbers for a weak link ($T_c^*{}' \simeq 100°K$, as treated above) are ~ 4 at room temperature and 800 at 60°. Thus, at room temperature a completely decoupled filament is much more resistive than a weakly coupled one, whereas at low temperatures, they are rather similar.

Concluding Discussion

We have seen that the conductivity of an isolated super-conducting filament of monomolecular cross-section will rise roughly as $(T/T_c^*)e^{T_c^*/T}$ when cooled below its mean-field T_c^*. For example, in TTF-TCNQ, the high-temperature conductivity is of order 1000 $(ohm-cm)^{-1}$, and T_c^* appears to be 500°K. Thus to get really useful "superconductivity" with $\sigma \gtrsim 10^{12}$ $(ohm-cm)^{-1}$ (a rather arbitrary choice $\sim 10^6$ better than room temperature copper), one would need to go to $(T/T_c^*) \sim 0.04$, or $T \lesssim 20°K$. This is far below the insulator phase transition at $\sim 58°K$; but even if this phase transition could be averted, the required temperature would be comparable to the T_c of Nb_3Sn.

The hope is that interchain coupling, whether by tunnelling, cross-linking, or whatever, will lead to a real 3-D superconducting phase transition to a state of zero resistance at a usefully higher temperature. The ("Heisenberg-like") model which we consider most appropriate leads us to estimate a 3-D transition at $T_c/T_c^* \simeq (J_\perp/J_z)^{\frac{1}{2}}$, where the J's are coupling energies between molecular units along a chain and between chains. From the observed con-ductivity ratio in TTF-TCNQ, we estimate that $J_\perp/J_z \simeq 10^2-10^4$, so

-164-

that $T_c/T_c^* \approx 0.1-0.01$. For these parameters, excellent conductivity in single chains would be reached by the time 3-D ordering would be expected. As shown in the Appendix, an Ising-type model leads to a significantly higher, though still small, T_c given by $T_c/T_c^* \approx [\ell n(J_z/J_\perp)]^{-1}$, but given the continuous degree of freedom of the phase in a classical superconductor, this model seems unlikely to be appropriate.

In summary, we conclude that conductivity a million times better than room temperature copper in isolated molecular superconducting chains requires $T/T_c^* \lesssim 0.04$. To get a 3-D phase transition and real superconductivity at a higher temperature than that through interchain coupling appears to require a coupling ratio $J_\perp/J_z \gtrsim (T/T_c^*)^2 \approx 1/600$.

Acknowledgement

References

1. W. A. Little, Phys. Rev. $\underline{156}$, 396 (1967).

2. M. Tinkham, this conference.

3. P. W. Anderson, P. A. Lee, and M. Saitoh, preprint.

4. L. B. Coleman. M. J. Cohen, D. J. Sandman, F. G. Yamagishi,
    A. F. Garito, and A. J. Heeger, Solid State Comm. $\underline{12}$, 1125(1973).

5. J. S. Langer and V. Ambegaokar, Phys. Rev. $\underline{164}$, 498 (1967);
    D. E. McCumber and B. I. Halperin, Phys. Rev. $\underline{B1}$, 1054 (1970).

APPENDIX

Ising Model Estimate of T_c

Our estimate (7) of T_c, namely,

$$kT_c \simeq 2(J_z J_\perp)^{\frac{1}{2}} \simeq kT_c^*(J_\perp/J_z)^{\frac{1}{2}} \tag{A1}$$

was reached by a method taking account of the degradation of order by long-wave low-energy phase fluctuations. This method seems the most appropriate for the model under consideration. None the less, we record here the corresponding estimate of T_c that results from converting the model to an Ising model, in which the phase factors $e^{i\phi_i}$ [and also the $\cos(\phi_i-\phi_j)$] have only two values, ± 1. There are several reasons why this may be useful: (1) The Ising model is exactly soluble in 1 and 2 dimensions, reducing the chance of important errors from mathematical approximations; (2) the Ising model includes the type of fluctuations giving resistance, while excluding the long-wave ones; since we use only heuristic arguments, it is conceivable that this selection is also relevant in determining a T_c; (3) since the nature of the transition in TTF-TCNQ is really unknown, it is possible that the Ising model is more appropriate than it appears on the surface.

In the Ising model (1) is replaced by

$$E = -\sum_{i,j} J_{ij}\sigma_i\sigma_j \tag{A2}$$

where $\sigma_i = \pm 1$ and the J_{ij} are as in (1). In the mean field approximation

$$kT_c^* \equiv kT_c\big|_{\text{mean-field}} = 2(J_x + J_y + J_z) \simeq 2J_z \tag{A3}$$

where the approximate equality holds since we assume $J_z \gg (J_x + J_y)$. This mean field approximation is very bad since it predicts finite T_c even if $J_x = J_y = 0$, whereas in reality $T_c = 0$ for any 1-D system. A better approximation is that of Bethe and Peierls. For the case of γ nearest neighbors, all with the same J, it predicts $kT_c\big|_{B-P} = 2J/\ln(\gamma/\gamma-2)$. This model at least gives correctly $T_c = 0$ if $\gamma = 2$. If we try to introduce the very weak couplings J_x and J_y by the artifice of fractional nearest neighbors, we might write $\gamma \simeq 2 + 2(J_x + J_y)/J_z$ with $J = J_z$ in the above formula. This leads to

$$kT_c\big|_{B-P} = \frac{2J_z}{\ln\left(\frac{J_x + J_y + J_z}{J_x + J_y}\right)} \simeq \frac{2J_z}{\ln\left(\frac{J_z}{J_x + J_y}\right)} \tag{A4}$$

which suggests that $T_c \rightarrow 0$ logarithmically with $(J_x + J_y)$.

Finally, we consider what should be our best approximation, based on the exact solution for T_c in the rectangular Ising net. This models a 2-D Ising system with $J_x \neq J_z$, and has a T_c determined by

$$\sinh\frac{2J_x}{kT_c} \sinh\frac{2J_z}{kT_c} = 1$$

We again try to approximate our 3-D system by replacing J_x by J_x+J_y to account for the weak coupling in two directions from the z-directed chain. Thus we take

$$\sinh \frac{2(J_x+J_y)}{kT_c} \sinh \frac{2J_z}{kT_c} = 1 \tag{A5}$$

Since we assume $(J_x+J_y) << kT_c << J_z$, we can approximate (A5) by

$$\frac{J_x+J_y}{kT_c} e^{2J_z/kT_c} = 1 \tag{A6}$$

or

$$kT_c\Big|_{\substack{\text{Ising rect.} \\ \text{net}}} \simeq \frac{2J_z}{\ln\left(\dfrac{kT_c}{J_x+J_y}\right)} \tag{A7}$$

Presumably this is our most accurate approximation, but it is reassuring that it is so similar to our other approximation (A4), at least for the highly anisotropic systems of interest here.

We also note that (A6) is consistent with, and hence gives some support to, our heuristic argument that the 3-D ordering sets in when the interaction energy between ordered segments of adjacent chains is $\sim kT_c$. The logarithmic dependence on $J_\perp$ in (A7) [as contrasted with the square root dependence in (A1)] results from the exponential growth $\sim e^{2J_z/kT}$ of ordered chain length in the Ising model, which has a finite minimum excitation energy $2J_z$, whereas the "Heisenberg" type model has zero energy excitations of infinite wavelength so that the ordered length is only $\sim(J_z/kT)$, as noted in (5).

Comparing the T_c predicted by (A7) with that of (A1), we see that (A7) gives a higher value. For example, if $\frac{1}{2}(J_x + J_y) \equiv J_\perp = J_z/1000$, $T_c/T_c^* = 0.19$ from (A7), whereas it is only 0.031 from (A1). Thus, if $T_c^* \simeq 500°K$, $T_{c, \text{Ising}} \simeq 95°$ rather than $\sim 16°$. Considering the crudeness of our parameter estimates, it may be significant that the Ising T_c falls within less than a factor of 2 of the phase transition temperature $\sim 58°$.

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, and assuming that the cross-link network is unretarded, it is found that the creep compliance of the entanglement network can be calculated from observations of sample dimensions as a function of time during approach to equilibrium. Further experiments designed to clarify the phenomenon are outlined.

POLYMERIC ENTANGLEMENT NETWORKS
CROSS-LINKED IN STATES OF STRAIN

J. D. Ferry and S. S. Sternstein

I. Introduction

A recent report[1] analyzed the possibilities of the following experiment. An uncrosslinked amorphous polymer of high molecular weight is subjected to a comparatively large deformation and chemical cross-links are introduced within an experimental time scale such that no significant entanglement slippage occurs. Upon release of the sample, its dimensions change to a state of ease in which, qualitatively, the original entanglement network and the cross-link network are pulling in opposite directions. From the dimensions in the state of ease and the elastic modulus in this state, the densities of entanglements and cross-links can be calculated. Equations and numerical evaluations were given, applicable to deformation in simple extension, based on Flory's thorough treatment[2] which confirms the composite network model[3] for sets of permanent cross-links introduced in different states of strain. The potential of additional measurements such as the kinetics of approach to equilibrium and swelling was also mentioned.

Preliminary experiments at the University of Wisconsin, in which 1,2-polybutadiene was cross-linked by γ irradiation in vacuo at 0°C., have demonstrated the utility of the method.[4]

We now summarize the conclusions thus far obtained from these experiments, and explore the kinetics of approach to the state of ease following release of stress after completion of the cross-linking.

II.   Conclusions from Preliminary Experiments

1.   Confirmation of Composite Network Theory. For a given irradiation time, the dependence of λ_s on λ_o fits the theory quite well with an appropriate value of ν_x/ν_N. Here λ_o and λ_s are the stretch ratios during cross-linking and in the state of ease respectively; ν_N and ν_x are the effective moles of network strands per cc. of the entanglement network and the cross-link network respectively. Moreover, ν_x/ν_N is nearly proportional to irradiation time, implying proportionality of ν_x and constancy of ν_N. By combining this information with modulus measurements, ν_N was found to be 0.67×10^{-4} moles cm^{-3}, corresponding to an average molecular weight between entanglement loci of 13,000.

2.   Comparison with Entanglement Density in Original Uncross-linked Polymer. From rheological measurements on the uncross-linked polymer in small deformations,[5] the entanglement strand density was deduced to be 2.5×10^{-4} moles cm^{-3}, corresponding to a molecular weight between entanglements of 3500. It has been suggested that the discrepancy may be due to the entrapment of the entanglements being incomplete, since untrapped entanglements would not contribute to the stress of the N network

-173-

at equilibrium; or to effects of the C_2 term in the Mooney-Rivlin equation, which contributes proportionately more to the stress at low elongations (in the measurements on uncross-linked polymer) than at high elongations (in the composite network experiments). However, these explanations would imply a dependence of ν_N on λ_0 or λ_x, respectively, which has not yet been observed. Possibly with increasing cross-linking more entanglements are trapped but this effect is offset by some chain scission to leave the density of trapped entanglements nearly constant.

3. _Absence of More than One State of Ease_. If some entanglements are untripped, it has been expected that they would contribute at first to retractive force leading to a first state of ease λ_s and then slip to allow adjustment to a second state of ease λ_{sr} ($\lambda_s < \lambda_{sr} < \lambda_0$). But only one is observed, with monotonic change from λ_0 to λ_s following release of stress. Actually, from earlier calculations,[1] a distinction between λ_s and λ_{sr} should be seen only at the lowest level of cross-linking in the present experiments. Its absence indicates that entanglement slippage on dangling branched structures is rapid compared with some other slow configurational changes.

4. _Slow Approach to State of Ease_. From viscoelastic properties of the uncross-linked polymer,[5] it can be estimated that portions of chains between entanglement loci should rearrange their configurations completely at 0° within about 20 sec. However, it requires several hours for the state of ease

to be attained, and in practice the temperature is raised to 25°
to accelerate the process. Thus, some longer-range configur-
ational changes must be taking place. In a formal way, this can
be ascribed to presence of retarded elastic compliance in the N
network, the x network, or both, in a time scale much longer than
that appropriate for rearrangements of single network strands.
If it can be interpreted, it will aid in understanding the
general phenomenon of slow relaxation processes in lightly cross-
linked rubbers.

III.   Kinetics of Approach to State of Ease

Because of the importance of the slow approach to the
state of ease, its relation to the viscoelastic properties of
the two networks is now examined. For simplicity, linear visco-
elasticity is first assumed although this is restricted to smaller
deformations than have been used in the present experiment. For
this case, the tensile strain is represented by $\epsilon = \lambda - 1$ and
the relation between λ_o, λ_s, ν_x, and ν_N given earlier[1] reduces
to

$$\nu_x/\nu_N = \epsilon_s/(\epsilon_o - \epsilon_s) = E_{xe}/E_{Ne} = R \qquad (1)$$

where the E's are the equilibrium Young's Moduli of the re-
spective networks.

The stress which constrains the sample at ϵ_o during
cross-linking has the value $\sigma_o = E_{Ne}\epsilon_o$ just before the sample
is released, where E_{Ne} is the equilibrium modulus of the

entanglement network (N-network)(which may have been changing
during cross-linking because of some chain scission or other
effects). Upon release, σ falls to zero and remains zero, but
the stress σ_N on the entanglement network subsequently rises
and approaches an equilibrium value $E_{Ne}\varepsilon_s$, while the stress σ_x
on the cross-link network becomes increasingly negative
approaching an equilibrium value $-E_{Ne}\varepsilon_s$; throughout the re-
traction process, $\sigma_N = -\sigma_x$.

The strain $\varepsilon(t)$ can be expressed from the Boltzmann
superposition principle as

$$\varepsilon(t) = \sigma_0 [D_\infty - D(t)] + \int_0^t D(t-u)\dot{\sigma}_N(u) \, du \qquad (2)$$

where $D(t)$ is the tensile creep compliance function of the
entanglement network; $D_\infty = 1/E_{Ne}$. The first term expresses the
creep recovery that would occur in the absence of the additional
cross-link network, and the second is the result of the gradually
increasing stress imposed by the latter.

If the cross-link network is unretarded, i.e., retardation
times associated with its configurational changes are small com-
pared with the experimental time scale, σ_x (a negative quantity)
is given by

$$\sigma_x = [\varepsilon(t)-\varepsilon_0]E_{xe} = [\varepsilon(t)-\varepsilon_0]/D_x \qquad (3)$$

where D_x is the equilibrium compliance of the cross-link network.
The time derivative $\dot{\sigma}_N$ is then

-176-

$$\dot{\sigma}_N = -\dot{\sigma}_x = -\dot{\epsilon}/D_x \qquad (4)$$

Substitution into (2) gives

$$\epsilon(t) = \epsilon_0[1-D(t)/D_\infty] - \int_0^t [D(t-u)/D_x]\dot{\epsilon}(u)\,du \qquad (5)$$

The Laplace transform of $\epsilon(t)$ is

$$\bar{\epsilon}(s) = \epsilon_0[1/s-\mathcal{D}(s)/D_\infty]-(\mathcal{D}(s)/D_x)[s\bar{\epsilon}(s)-\epsilon_0] \qquad (6)$$

which can be rearranged to give

$$\mathcal{D}(s)/D_\infty = \frac{\epsilon_0/s-\bar{\epsilon}(s)}{\epsilon_0\{1-(D_\infty/D_x)[\epsilon_0-s\bar{\epsilon}(s)]/\epsilon_0} \qquad (7)$$

Introducing a new variable $y(t) = 1-\epsilon(t)/\epsilon_0$, we have $\dot{y} = -\bar{\epsilon}/\epsilon_0$ and $\bar{y}(s) = 1/s-\bar{\epsilon}(s)/\epsilon_0$;

$$\mathcal{L}\{\dot{y}\} = 1-s\epsilon(s)/\epsilon_0 \equiv g(s) \qquad (8)$$

Substitution into (7), recalling that $D_\infty/D_x = R$, gives

$$\mathcal{D}(s)/D_\infty = \bar{y}(s)/[1-R\,g(s)] \qquad (9)$$

This may be expanded in series:

$$\mathcal{D}(s)/D_\infty = \bar{y}(s) + R\bar{y}(s)\,g(s) + R^2\bar{y}(s)\,g^2(s)\cdots \qquad (10)$$

The inverse transform provides

$$D(t)/D_\infty = y(t) + [Ry*\dot{y}] + [R^2y*\dot{y}*\dot{y}] + \cdots \qquad (11)$$

where

$$y(t) = 1-\epsilon(t)/\epsilon_0$$

$$y * \dot{y} = \int_0^t y(t-\tau) \; \dot{y}(t) \; d\tau$$

$$y * \dot{y} * \dot{y} = \int_0^t \int_0^{t-\tau} y(\tau) \; \dot{y}(t-\tau-u) \; \dot{y}(u) \; du \; d\tau$$

From experimental measurements of the retraction process, viz., $\varepsilon(t)$, the creep compliance function $D(t)$ could in principle be obtained from (11) by numerical evaluation (or possible analytically in terms of an empirically chosen function). However, the results may be strongly weighted by $\dot{\varepsilon}(t)$ at the beginning of the retraction, which is technically difficult to measure.

IV.   Future Experiments

Further experiments which are planned include (1) measurements of the kinetics of approach to state of ease and application of the theory above, (2) measurements of swelling of the composite networks with check of possible anisotropy and also elastic properties in the swollen state, (3) attempts to gauge the extent of chain scission during cross-linking, (4) study of transient changes produced by large deformations, possibly attributable to extensive slippage of trapped entanglements, (5) study of another polymer besides 1,2-polybutadiene.

Acknowledgement

References

1. J. D. Ferry, States of Ease of Polymeric Entanglement Networks Cross-Linked in States of Strain, Preliminary Reports, Memoranda and Technical Notes of the ARPA Materials Summer Conference, Vol. I, July 1971.

2. P. J. Flory, Trans. Faraday Soc., $\underline{56}$, 722 (1960).

3. R. D. Andrews, A. V. Tobolsky, and E. E. Hanson, J. Appl. Phys., $\underline{17}$, 352 (1946).

4. O. Kramer, '. Ty, and J. D. Ferry, Proc. Nat. Acad. Sci. U.S., $\underline{69}$, 2216 (1972).

5. J. F. Sanders, J. D. Ferry, and R. H. Valentine, J. Polymer Sci., A-2, $\underline{6}$, 967 (1968).

HYDROGEN STORAGE IN TRANSITION METALS

H. Ehrenreich

Abstract

Many of the transition metal hydrides contain hydrogen in concentrations comparable to those found in liquid or even solid hydrogen. This fact is of possible technological interest in connection with energy storage: the hydrogen is stored in atomic form and the molecular recombination energy becomes available in a controlled fashion upon moderate heating. This note qualitatively investigates the reasons for the large binding of hydrogen in transition metals in terms of our present understanding of the electronic structure as well as the cohesive energy of these materials. A quantification of these notions may be useful in the search for optimal materials as defined in terms of cost, operating conditions, and storage capacity.

HYDROGEN STORAGE IN TRANSITION METALS

H. Ehrenreich

Introduction

The fact that large densities of hydrogen are readily accommodated in transition metals is of scientific and possibly of technological interest. The density of H atoms (in units of $10^{22}/cm^3$) of liquid and solid H_2 is respectively 4.2 and 5.3. By contrast, TiH_2, ZrH_2, $VH_{0.8}$, and UH_3 accommodate 9.2, 7.3, 5.1, and 8.2, some of which are considerably higher.[1] The hydrogen dissolved in a metal can sometimes be visualized as an interstitial solution. The hydrogen pressure-hydrogen content isotherms of PdH_x are very much like those of a liquid-gas mixture. The low pressure and concentration α phase corresponds to the gas, and the high pressure and concentration β phase corresponds to the liquid. At intermediate concentrations there is an α and β coexistence region. For PdH_x at STP $x \leq 0.6$. On the other hand, it is possible to form a stochiometric form, PdH, at high pressures which is an ordered alloy having the NaCl structure.

For the metals of interest, hydrogen absorption is an exothermic reaction, frequently involving enough energy to dissociate the H_2 molecules. The hydrogen may be released slowly by heating the hydride: in the case of MgH_2, a dissociation pressure of 1 atmosphere corresponds to $\sim 300°C$; in the case of

-182-

TiH_2 the corresponding temperature is about 700°C. These considerations lead to the attractive possibility that atomic H may be safely stored at high densities. Modest heating makes it available as a fuel. The association energy of ~100 kcal/mole involved in the reaction $H + H \rightarrow H_2$ represents the most significant contribution to the energy produced.

Wiswall and Reilly[2] have recently proposed the use of metal hydrides as an essentially pollution free automobile fuel. Specifically, they have suggested its use in a Wankel engine which has been established to operate smoothly on a hydrogen-air mixture without mechanical changes. (Lubrication was provided by injecting an oil spray into the hydrogen feed.) A comparison of the calories per cm^3 available from hydrides formed using a MgNi alloy and V with that of n-octane gasoline is reasonably favorable, the amounts being 4270, 3520 as contrasted to 8000 cal/cm^3 for gasoline. No data concerning refueling speeds is supplied, but certain TiCo alloys are claimed to react, even in large pieces, with hydrogen almost as fast as it can be supplied. While there are obvious questions concerning weight, metal cost, refueling safety, etc., this is one of several applications that has been proposed for energy storage in metallic hydrides. Others are discussed in an accompany paper by Hawthorne.[3]

The principal purpose of this note is to examine why H has a positive heat of solution in many metals and, more specifically, how the electronic energy levels of the metal are altered to produce increased cohesion. Various proposals have been ad-

vanced, such as the protonic and the screened proton model,[4] but none of them has yet been shown to provide a quantitative explanation of all experimental facts. We propose to examine the problem from the solid state physics standpoint of band theory. Our discussion is based on recent results for the energy bands of metal hydrides by Switendick[5] as well as preliminary calculations by Gelatt.[6] The line of argument (although not the quantitative aspects) are followed most readily by reference to the figure.

Transition Metal Band Structures

We begin with a free transition or noble metal atom M, and we shall visualize forming the metal by preparing the atoms in a state most closely resembling that in which they enter the solid. This is accomplished by constructing a Wigner-Seitz cell of the appropriate radius r_{WS} around each metal atom and compressing the charge into it by altering the wave function normalication constant: hence the term "renormalized atom".[7] For the 3d noble and transition atoms about 1/3 to 1/2 of the s charge and about 5% of the d charge lie outside the Widner-Seitz cell. The result of the renormalization process is to raise the atomic d level by about 0.5 Ry because the extra s charge, which is located in the outer regions of the unit cell raises the potential in the region where most of the d charge resides. The behavior of the s level is of less interest. Of primary concern is the resultant position of the conduction band minimum $\varepsilon_{cond}(k=0)$. These are shown schematically in the figure. The positions may

be estimated by calculating

$$\varepsilon_d^{ren} = <\phi_d|-\nabla^2 + v_d^{ren}|\phi_d>$$

$$\varepsilon_{cond}(k=0) = <k=0, OPW|-\nabla^2 + v_d^{ren}|k=0, OPW>.$$

Here v_d^{ren} is the renormalized atom potential, which is obtained from the renormalized atom charge density; $<r|k=0, OPW>$ is a k=0 orthogonalized plane wave.

The upper and lower edges of the d bands are determined by finding the energy of d-electrons in the renormalized atom potential $-\nabla^2 + v_d^{ren}$ and imposing respectively anitbonding $(\phi_d(r_{WS}) = 0)$ and bonding $(d\phi_d(r_{WS})/dr=0)$ boundary conditions.

It is to be emphasized that the foregoing quantities were all obtained from consideration of a renormalized atom. They are, however, quantitatively in agreement with what would be found in the olid, i.e., by bringing together the renormalized atoms, turning on s and d hopping between adjacent sites, as well as s-d hybridization. This point has been verified by performing APW calculations using the same renormalized atom potential and comparing the results: agreement of $\sim.03$ Ry was typical for a variety of transition metals.

Cohesive Energy of Transition Metals

This viewpoint is convenient also in the calculation of the cohesive energy of transition metals since the steps in forming the solid (free atom → renormalized atom → solid) are most simply

defined. By the same token, this approach will be useful also in the case of the transition metal hydrides, MH_x.

One point that may seem surprising when viewed in the context of cohesive energy considerations, is the large upward shift (0.5 Ry/electron) of the d bands upon renormalization. This shift, by itself, does not necessarily affect the cohesive energy deleteriously, since the energies that must be summed to obtain the total are the Hartree-Fock parameters

$$\epsilon_i^{(HF)} = \epsilon_i - \frac{1}{2} \sum_j [U_{ij} = J_{ij} \delta_{\sigma_i \sigma_j}] \quad ,$$

where U_{ij} and J_{ij} are respectively two-electron Coulomb and exchange terms. Both ϵ_i and U_{ij} increase upon renormalization in such a way that $\epsilon_i^{(HF)}$ remains practically unchanged.

This point has been checked in estimates of the cohesive energy of 3d and 4d metals.[8] The principal ingredients contributing to the cohesive energy were found to be: (1) preparation of the free atom in the most nearly self-consistent configuration with which it enters the solid; (2) renormalization of the atom; (3) broadening of atomic d levels into bands; (4) s-d hybridization effects. The contribution of (1) is always negative, (2) may have either sign, and (3), (4) are always positive. Contribution (3), which vanishes for filled d bands, is positive for partly filled bands because a majority of the electrons find themselves in bonding rather than antibonding states. Hybridization effects, (4), have the same size because they lower the energies of some

-186-

s-d levels, at the same time pushing others above the Fermi
energy.

The Hydrides

We now wish to extend the previous considerations to the
transition metal hydrides MH_x. The present results are only semi-
quantitative since the cohesive energy estimates made so far have
been very crude. We adopt the viewpoint that _neutral_ hydrogen
atoms (or perhaps completely screened protons) are inserted into
octahedral sites of the fcc transition metals for $x \leq 1$. For the
stochiometric compound MH the resulting structure has the NaCl
form. In the case of the non-stochiometric solutions HM_x the
situation is more complicated: while the atoms M are situated on
a periodic fcc lattice, the octahedral sites form a disordered
H-vacancy alloy. Furthermore, the H atoms are only situated ran-
domly in the low concentration limit (α-phase). At high H con-
centrations, the solution may be viewed as an NaCl structure with
H vacancies, whereas at intermediate concentrations the α and β
phases coexist. It is in this regime that present alloy theory,
which has been concerned mostly with random substitutional binary
alloys $A_x B_{1-x}$, must be extended.

Let us therefore consider the ordered alloy MH. Referring
to the figure, we ask how the energy levels of the metal M are
modified when neutral H is inserted into the octahedral sites.
The extra potential is attractive to M electrons that extend in-
side it. Among these are the $k=0$ conduction electrons, the

-187-

bonding d electrons, but to a much lesser degree, the antibonding

d electrons. The result is a lowering energy of $\varepsilon_{cond}(k=0)$ and

of the lower d band edge, while the upper d band edge remains

relatively undisturbed. The one-electron energies are thus lowered

and so is presumably the total energy, since the Coulomb integrals

(which have not yet been investigated), would not be expected to

be significantly affected by insertion of the H atoms, provided

there is no charge transfer between the M and H constituents. We

shall return to this point.

One other effect revealed by the band calculations for

many of the metal hydrides and dihydrides is the formation of what

is loosely termed in the figure as a "bonding s band". It results,

in band theoretic language from the fact that the perturbing

periodic H potential produces a sufficient increase in the split-

ting between levels in the d and conduction bands, that they can

never merge with the d bands along any direction of k-space. The

effect is that the part of the band structure in the energy range

around $\varepsilon_{cond}(k=0)$ is separated from the d bands by a finite gap.

From a chemical point of view one may think of these as covalent

bonding levels having s, p and d hybridized M atom and s H atom

wave function components. This bonding effect serves to increase

the cohesive energy further.

Preliminary estimates indicate the cohesive energy in-

crease to amount to $\sim$100 kcal/mole of H_2, for Pd, Ti, and Cu,

which is just enough to dissociate the molecule. The disturbing

feature of this result is its insensitivity to the constituent M.

While it is quite reasonable for Pd, one would expect it to be considerably larger for Ti (which forms a dihydride), and smaller, if not of the opposite sign for Cu. In the latter case, CuH, in a form similar to NiH, has not yet been shown to exist, although this may be the result of the preparation technique rather than an intrinsic instability.[9] Leaving this aside, the difference between PdH or NiH (monohydrides) on the one hand and the ability of metals like Ti to form dihydrides on the other, may well be the result of increasing ionic character (which has been neglected here), or charge transfer, as one moves across the periodic table to the left. One piece of evidence is that LiH is an almost purely ionic insulator. Thus charge transfer effects between Ti and H would lead to an additional stabilizing Madelung energy. By contrast, in NiH and CuH the bond character is probably largely covalent and the constituents are essentially neutral, as has been supposed in the calculations on which the present discussion is based.

Atomic size effects may also play a role in terms of the capability of the octahedral hole to accommodate an H atom without excessive distortion of its electronic cloud. From the pressure -H/M concentration phase diagram and lattice constant measurements, it is clear that the lattice expands by several percent in the β phase and that increased solubility of H is obtained only at the expense of large H pressures.[9] Size effects thus limit the solubility and undoubtedly also affect the cohesive energy.

We finally comment briefly on one qualitative approach to the non-stochiometric solution problem associated with MH_x (x not

an integer). In, what chemists term the "puff" model, one visualizes the neutral H potential to be spread uniformly over an appropriate region of space, in the present case over the part of the Wigner-Seitz cell that has a flat potential in muffin tin schemes for calculating the band structure. The potential shift will be proportional to the concentration for all concentrations. The effects on the band structure obtained from this sort of calculation are in some respects similar to those described above: $\varepsilon_{cond}(k=0)$ and the bonding d states are lowered appreciably; while antibonding d states are largely unaffected. There is one notable qualitative difference, however: the "bonding s states" are not formed in this case, presumably because the directional character of the covalent M-H bond is lost when the effects of H atoms are replaced by uniform shifts of the crystal potential in the muffin-tin portion of the unit cell.

Acknowledgements

TRANSITION METAL HYDRIDES

Free atom M — Renormalized atom M — Solid M — MH
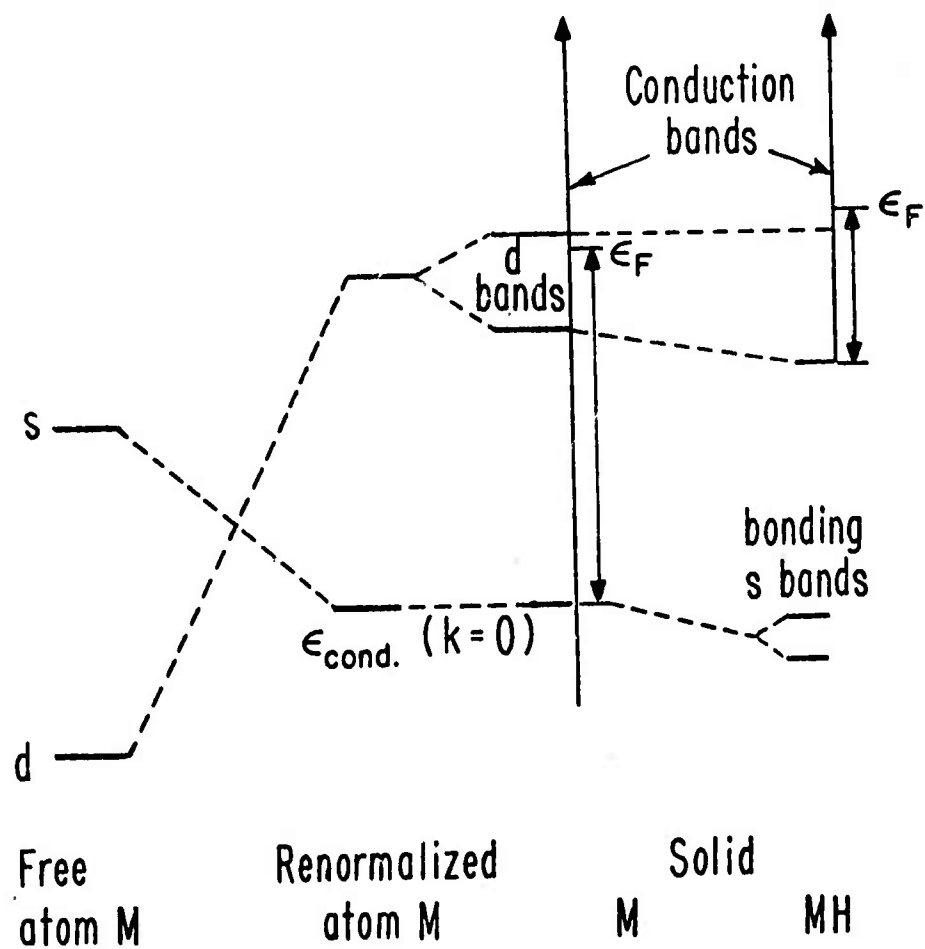
References

1. G. G. Libowitz, *The Solid State Chemistry of Binary Metal Hydrides*, Benjamin, New York (1965).

2. R. H. Wiswall and J. J. Reilly, *7th Intersociety Energy Conversion Engineering Conference*, 1972 (American Chemical Society, Washington, D.C.), p. 1342.

3. F. Hawthorne (These proceedings).

4. Y. Ebisuzaki and M. O'Keefe, *Progress in Solid State Chemistry* (H. Reiss, Ed.) $\underline{4}$, 187 (1967).

5. A. C. Switendick, Per. Bunseges. Phys. Chem. $\underline{76}$, 535 (1972).

6. C. D. Gelatt, Ph.D. thesis, Harvard University, 1973 (unpublished).

7. L. Hodges, R. E. Watson, and H. Ehrenreich, Phys. Rev. $\underline{B5}$, 3953 (1972).

8. R. E. Watson and H. Ehrenreich, Comments in Solid State Physics $\underline{III}$, 109 (1970).

9. F. Hawthorne (private communication).

DYNAMICS OF ELECTRONS IN GRADED LATTICES

H. Reiss

Abstract

The Wannier functions for the lowest lying band of a one dimensional crystal with delta function potential wells are derived analytically for the tight binding limit. A crystal is then considered whose lattice is nonuniform to the extent that the strength of its potential wells varies linearly with position. Using, as basis set, the Wannier functions corresponding to a uniform lattice whose well strength corresponds to a particular point in this crystal, the differential-difference equations which determine the motion of the envelope vector (wave packet) in the graded lattice are formulated. These equations are then solved analytically for the tight binding case.

These solutions may be of value in examining the roles of "effective" Hamiltonians which "drift" (in the sense that the basis set drifts) with an electron in the lattice. Such "local" Hamiltonians are thought to be of value in understanding the behavior of solid state devices predicted on graded lattices.

DYNAMICS OF ELECTRONS IN GRADED LATTICES

H. Reiss

I.   Introduction

Recently, there has been much interest in semiconductor
devices based on crystals in which, aside from defects and im-
purities, the properties of the lattice change with position.
Such devices include those utilizing heterojunctions as well as
less dramatic nonuniformities such as gently graded lattices.[1-5]
The dynamics of electrons in such crystals is imperfectly under-
stood, although there has been some work on the subject.[5-8]

Previous research in this area has either (1) been con-
cerned with the classical limit, or (2) been aimed at the deri-
vation of an effective "local" Hamiltonian which could be used
to define the quantal equations of motion. Although such effec-
tive Hamiltonians can be derived,[6-8] and indeed justified in a
fairly fundamental way, it is not clear how they are to be used
to determine the motion of the electron. They do, however, pro-
vide assurance that a local band structure exists, corresponding
to the situation which would prevail if the crystal were every-
where uniform with the characteristics (lattice parameter, com-
position, etc.) of the local region under consideration. In
mixed crystals, for example III-V compounds, where the nonuni-
formity stems from a continuous variation of composition, it has

been customary to account for disorder by means of the "virtual crystal approximation".[9] Although there is some uncertainty concerning the validity of this approximation beyond first order perturbation theory, we will not be concerned with this problem in the present paper. We will, however, investigate certain aspects of the dynamics of electrons in graded lattices. In fact, we have in mind nonuniform lattices such as might be produced by a continuously varying composition in III-V or II-VI compounds.

To set the matter in better perspective, consider the difficulties which arise out of an attempt to apply the "effective mass formalism"[10] to a nonuniform lattice. Because the lattice properties change with position, so does the effective mass. In the classical limit, we are, in effect, looking for the trajectory of a particle whose mass is constantly changing with position. As a familiar example, consider the trajectory of a rocket whose mass changes because it is constantly ejecting hot gases to provide the driving force. The rocket problem can be solved because we know the ultimate velocity of the ejected matter, and it is possible to apply the principle of conservation of momentum. If the ultimate velocity of the ejected matter is not known, however, we cannot solve the problem.

From another point of view, the instantaneous "conventional" Hamiltonian (or for that matter the local Hamiltonian) for the rocket, regarded as a system (or particle), in itself is of no help in specifying the motion. Since the mass depends on position, the Hamiltonian does not satisfy the canonical equations of motion,

and it is not clear what Hamiltonian would serve this purpose, or it it is even possible to find such a Hamiltonian. Unless we have the appropriate Hamiltonian, we cannot specify dynamical variables which are canonically conjugate, and so it is not clear how one would develop commutation rules which could be used to properly quantize the system.

The local effective Hamiltonians, which can be derived[6-8] within the framework of the effective mass formalism, are subject to all of these difficulties. In substance, we need to derive the proper conservation laws which include the electron and the lattice, and which determine the motion of the envelope function in the effective mass framework. In the usual case, the local Hamiltonian is designed to operate on an "envelope function" for a basis set consisting of Wannier functions[11] corresponding to local lattice conditions. In other words, the basis set drifts with the electron. As the electron moves, there is a reapportionment of "current" between the envelope and the basis functions themselves. The changing effective mass corresponds, of course, to the fact that the envelope function is changing with the basis set. If a wave packet (envelope function) is subject to no external field, it will nevertheless be set in motion because of the nonuniform lattice properties. Given a particular nonuniform lattice, the motion of the wave packet can be specified without ambiguity if the basis set is chosen to correspond to some fixed location within the crystal, and is not allowed to drift with the electron. The changing lattice properties then appear as potentials giving

rise to effective fields.

Because of the subtle nature of the problem, it would be worthwhile to have in hand a example in which the motion of the wave packet in a nonuniform lattice can be specified analytically. This example could then be used to test any general theory that might be proposed. We address ourselves to this task in the present paper. The system for which such an analytical solution can be found consists of a variant of the Kronig-Penney[12] model for a one-dimensional lattice. This consists of a linear array of delta function potentials which, in the usual case, are spaced with perfect periodicity. Three variations are easily considered. In the first example, the lattice parameter does not change, but the strength of the delta function changes linearly with distance. A second example maintains the strength of the delta function but permits a linear change of lattice parameter with distance. The third example would combine these two effects simultaneously. Although it is not necessary for these changes to be so gradual (in order to arrive at an analytical result) that they can be treated as first-order perturbations, it does seem necessary to consider the lattice in the "tight binding" limit[13] where only overlaps between Wannier functions on adjacent sites are of importance. Consequently, in what follows, we limit ourselves to this case. The system treated, therefore, becomes somewhat artificial, but does have the merit of allowing an analytical solution. Furthermore, we treat only the first of the examples discussed above in which the lattice parameter is held constant, but

the strength of the delta functions is allowed to change with distance in a linear fashion. Again, it is emphasized that the problem is solved in a basis set which <u>does</u> <u>not</u> drift with the electron.

II.   The Conventional Kronig-Penney Model in the Tight Binding Limit

A good treatment of the conventional delta function lattice can be found in Smith.[14] In this section, we outline, in essence, the method followed by Smith, making appropriate changes and extentions when necessary in order to prepare the way for the ultimate treatment of a nonuniform lattice.

First, consider the motion of an electron in the vicinity of a one-dimensional potential specified by

$$V(x) = -V_o \quad , \quad -a/2 < x < a/2$$
$$= o \quad , \quad |x| > a/2 \quad , \tag{1}$$

in which V represents the potential energy, x is the coordinate of position, and V_o is a positive quantity,

$$V_o > 0 \quad . \tag{2}$$

The wave equation for this system is

$$d^2\Psi/dx^2 + (2m/\hbar^2)(e-V[x])\Psi = 0 \quad , \tag{3}$$

where Ψ is the wave function for the electron, m is its mass, E is the energy eigenvalue, and $\hbar$ is Planck's constant divided

by 2π. Within the well, we may write this equation in the form

$$d^2\Psi/dx^2 + \beta^2\Psi = 0 \quad , \tag{4}$$

where

$$\beta^2 = \frac{2m}{\hbar^2} (E+V_o) \quad . \tag{5}$$

This quantity is positive even when E is negative, since energy levels within the well always lie above its bottom.

The well is converted into a delta function by allowing

$$a \to 0 \quad , \quad V_o \to \infty \tag{6}$$

in such a manner that the product aV_o remains finite. In this limit, the value of Ψ within the well is effectively constant and may be set equal to $\Psi(o)$. If Eq. (4) is integrated over the well, we obtain

$$(\frac{d\Psi}{dx})_{x=a/2} - (\frac{d\Psi}{dx})_{x=-a/2} = a\beta^2\Psi(o) \quad . \tag{7}$$

Outside the well, Eq. (3) may be written as

$$\frac{d^2\Psi}{dx^2} - \alpha^2\Psi = 0 \quad , \tag{8}$$

where

$$\alpha^2 = -2mE/\hbar^2 > 0 \quad . \tag{9}$$

The well-behaved solutions of Eq. (8) on each side of the well, respectively, are

$$\Psi_+ = A_+ e^{-\alpha x} \tag{10}$$

for positive values of x, and

$$\Psi_- = A_- e^{\alpha x} \tag{11}$$

for negative values of x. A_+ and A_- are constants. Since Ψ is essentially constant over the breadth of the well, we must set

$$A_+ = A_- = A \tag{12}$$

in order to assure continuity. Substitution of Eq. (12) into Eqs. (10) and (11), and the results into Eq. (7), yields

$$2\alpha = a\beta^2 \quad . \tag{13}$$

Since E remains finite as V_o goes to infinity, as long as aV_o remains finite, Eq. (13) determines E when Eqs. (5) and (9) are substituted in it. We obtain

$$E = E_o = -m(aV_o)^2/2\hbar^2 \quad , \tag{14}$$

where E_o is defined by this equation. The energy level specified by Eq. (14) is the only level within the well; there is only one bound state. We may use Eq. (9) to define α_o, the value of α corresponding to E_o. Thus, we have

$$\alpha_o^2 = -2mE_o/\hbar^2 = (maV_o/\hbar^2)^2$$

or

$$\alpha_o = maV_o/\hbar^2 \quad . \tag{15}$$

Now we consider a lattice of delta function potentials,

each separated by the lattice distance λ. A key relation, facili-
tating the easy treatment of this periodic system, is Eq. (7)
which may still be assumed to hold at each potential well, with
the two derivatives taken, respectively, on opposite sides of the
well and $\Psi(o)$ again referring to the constant value of the wave
function within the well. Another key relation derives from the
Bloch theorem.[15] Thus, if Ψ_n denotes the wave function in the
interval, $[n\lambda < x < (n+1)\lambda]$, this theorem requires

$$\Psi_n(x) = e^{ink\lambda}\{Ae^{\alpha(x-n\lambda)} + Be^{-\alpha(x-n\lambda)}\} \quad ,$$

$$n\lambda < x < (n+1)\lambda \quad , \tag{16}$$

where $i = \sqrt{-1}$. In this equation, n is an integer, α is given by
Eq. (9), k is the wave vector, limited to a single Brillouin
zone[16] for a single energy band, and A and B are constants. If
we wish to apply Eq. (7) to the potential well located at $x=n\lambda$,
the derivative on the left of the well refers to Ψ_{n-1}, while that
that on the right involves Ψ_n. Substitution of Eq. (16) into (7)
yields one relation. Another relation comes from the continuity
requirement

$$\Psi_{n-1}(n\lambda) = \Psi_n(n\lambda) \quad ; \tag{17}$$

and these two conditions, when applied simultaneously, lead, after
some algebraic manipulations in which the constants A and B are
eliminated, to the single condition

$$\alpha \cosh \alpha\lambda - \alpha_o \sinh \alpha\lambda = \alpha \cos k\lambda \quad . \tag{18}$$

This is, in fact, the dispersion relation which determines E (since α depends upon E) as a function of wave vector k.

In the tight binding approximation, we allow λ to become very large. Under this circumstance, it can be shown that α lies very close to α_o, and that to the first order in μ, where

$$\mu = e^{-\alpha_o \lambda} \quad ; \tag{19}$$

we can express α as

$$\alpha = \alpha_o + 2\alpha_o e^{-\alpha_o \lambda} \cos k\lambda \quad . \tag{20}$$

Squaring both sides of this equation, keeping terms to the first order in μ, and using Eq. (9), yields

$$E = E_o + 4E_o e^{-\alpha_o \lambda} \cos k\lambda \quad . \tag{21}$$

This equation describes the energy band corresponding to the bound state. The band edge occurs at k=o, and is given by

$$E_b = E_o + 4E_o e^{-\alpha_o \lambda} \quad . \tag{22}$$

If we set the zero of energy at the band edge, and distinguish the energy levels referred to this zero by appending a prime symbol, we have, by combining Eqs. (21) and (22),

$$E' = 4E_o e^{-\alpha_o \lambda} (\cos k\lambda - 1) \quad . \tag{23}$$

Near the band edge, we may expand the cosine in Eq. (23) and obtain

$$E' - 2E_o e^{-\alpha_o \lambda} k^2 \lambda^2 \qquad . \tag{24}$$

The effective mass m_* is defined by the relation

$$E' = \hbar^2 k^2 / 2m_* \qquad . \tag{25}$$

Comparing Eqs. (24) and (25), we have

$$m_* = m e^{\alpha_o \lambda} / 2\alpha_o^2 \lambda^2 \qquad . \tag{26}$$

This relation is, of course, applicable only at small wave numbers near the band edge.

We complete this section by normalizing the wave function specified by Eq. (16). A relationship between the constants A and B in Eq. (16) is obtained by substitution of Eq. (16) into Eqs. (7) and (17). In the case of Eq. (7), the derivative to the left of the well at $x=n\lambda$ involves the function Ψ_{n-1}, while that to the right involves Ψ_n. $\Psi(o)$ is given either by $\Psi_{n-1}(n\lambda)$ or $\Psi_n(n\lambda)$, both of which are equal according to Eq. (17). This same procedure, involving Eqs. (7) and (17), was invoked in arriving at the dispersion relation, Eq. (18). In fact, Eq. (18) was obtained by this very procedure, eliminating the constants A and B by determining the relationship between them. In any event, ex-pressing B in terms of A, we are able to write $\Psi_o(x)$ in the form

$$\Psi_o(x) = A \left\{ \frac{e^{(x-\lambda)} - e^{-\alpha x} e^{ik\lambda} - e^{-\alpha(x-\lambda)} + e^{-\alpha x} e^{ik\lambda}}{e^{-\alpha\lambda} - e^{ik\lambda}} \right\} \quad , \tag{27}$$

where, of course, the dependence of α on k is given by Eq. (20).

According to the Bloch theorem, $\Psi_n(x)$ may be expressed as

$$\Psi_n(x) = e^{ikx} u_k(x) = e^{ink\lambda} \Psi_0(x-n) \quad , \tag{28}$$

where $u_k(x)$ is a function having the periodicity of the lattice, and is, therefore, independent of n. The probability density is given by

$$\Psi_n^* \Psi_n = u_k^* u_k = \Psi_0^*(x-n\lambda) \Psi_0(x-n\lambda) \quad , \tag{29}$$

as is, therefore, as expected, also independent of n. If there are N lattice sites in the crystal, then the integral of the probability density over a single lattice parameter will be 1/N. Since the probability density is independent of n, we might just as well choose the integral lying between x=o and x=λ. Thus, we have

$$\int_0^\lambda u_k^* u_k \, dx = \int_0^\lambda \Psi_0^*(x) \Psi_0(x) \, dx = 1/N \quad . \tag{30}$$

Substitution of Eq. (27) into (30) yields the result

$$|A|^2 = \frac{1}{N} \left\{ \frac{1 + e^{-2\alpha\lambda} - 2e^{-\alpha\lambda} \cos k\lambda}{4\lambda[\cos k\lambda \cosh \alpha\lambda - 1] + \frac{2}{\alpha} \sinh 2\alpha\lambda - \frac{4}{\alpha} \cos k\lambda \sinh \alpha\lambda} \right\}. \tag{31}$$

Notice that α, and therefore A, depends upon k. Careful inspection of Eq. (31) shows that, in the tight binding limit, $|A|^2$ is given by

$$|A|^2 = \alpha e^{-2\alpha\lambda}/N \quad . \tag{32}$$

This completes our discussion of the band structure and the Bloch functions (to the first order in μ) in the uniform lattice. In the next section, we derive the Wannier functions for this case.

III.   Wannier Functions for the Delta Function Lattice

In this section, we employ the symbol n to denote the lattice site at which a given Wannier function is centered. We only consider Wannier functions associated with the lowest lying band, treated in the last section. The argument of the Wannier function centered at $x=n\lambda$ is $x-n\lambda$, and the Wannier function itself will be written as

$$a(x-n\lambda) = (1/\sqrt{N}) \sum_k e^{-ikn\lambda}\Psi(x,k) \quad , \tag{34}$$

where $\Psi(x,k)$ is the Bloch function, corresponding to wave vector k, discussed in the last section. We now replace n in Eq. (28) with r, also an integer, so as not to conflict with n as we now use it for the site at which a Wannier function is centered. Combining Eq. (33) with (28), we have, for the interval, $r\lambda < x < (r+1)\lambda$,

$$\Psi_r(x,k) = 2(\frac{\alpha}{N})^{\frac{1}{2}}e^{-\alpha\lambda}e^{ikr\lambda}\{\sinh \alpha[x-(r+1)\lambda]-e^{ik\lambda}\sinh \alpha[x-r\lambda]\}. \tag{35}$$

Since $\Psi(x,k)$ has a piecewise description in terms of $\Psi_r(x,k)$, the same will be true of the Wannier function expressed by Eq. (34). We define $a_r(x-n\lambda)$ as a piece of the Wannier function, centered at site n, lying in the interval between $x=r\lambda$ and

$x = (r+1)\lambda$, and write

$$a_r(x-n\lambda) = (1/\sqrt{N}) \sum_k e^{-ikn\lambda} \psi_r(x,k)$$

$$= (2/N) \sum_k \alpha^{\frac{1}{2}} e^{-\alpha\lambda} e^{ik(r-n)\lambda} \{\sinh \alpha[x-(r+1)]\lambda - e^{ik\lambda}\sinh \alpha[x-r\lambda]\} . \quad (36)$$

In the sum of Eq. (36), k extends over a single Brillouin zone and, again, it should be noted that α depends on k, in accordance with Eq. (20). If we assume N to be even (without loss of generality), then

$$k = 2\pi s/N\lambda, \quad s = -N/2, \quad -(\frac{N}{2} - 1), \cdots, \quad \frac{N}{2} - 1 . \quad (37)$$

The sum over wave vector in Eq. (36) can be carried out in closed form as long as we are interested in orders of μ, no greater than the first, i.e., in the tight binding case. To accomplish this, it is necessary to express $\alpha^{\frac{1}{2}}$, using Eq. (20), as

$$\alpha^{\frac{1}{2}} = \alpha_o^{\frac{1}{2}}[1+e^{-\alpha_o\lambda}(\frac{e^{ik\lambda}+e^{-ik\lambda}}{2})] \quad , \quad (38)$$

and to express all exponentials, having α as a coefficient in the exponent, as

$$e^{\alpha\theta} = e^{\alpha_o\theta}[1+2\alpha_o e^{-\alpha_o\lambda}(\frac{e^{ik\lambda}+e^{-ik\lambda}}{2})\theta] \quad . \quad (39)$$

where we have used θ to represent any one of the variables which α might multiply in the exponent. Of course, Eqs. (38) and (39) are only correct to the first order in μ. The summation procedure is somewhat tedious, but it is greatly facilitated by the relation[16]

-206-

$$\sum_k e^{ikj\lambda} = N\delta_{jo} \quad , \tag{40}$$

where j is an integer, δ_{jo} is the Kronecke δ, while the sum goes over the k values specified by Eq. (37), i.e., over a Brillouin zone.

After a quite lengthy calculation, we find that the Wannier function, centered on the site located at $x=n\lambda$, survives to the first order in μ, for only four values of r, namely, $r=n-2$, $n-1$, n, and $n+1$. The four surviving parts are

$$a_{n-2}(x-n\lambda) = -\alpha_o^{3/2}[x-(n-1)\lambda+(1/2)\alpha_o]e^{\alpha_o(x-n\lambda]} \quad , \tag{41}$$

$$a_{n-1}(x-n\lambda) = \alpha_o^{1/2}e^{-\alpha_o[x-(n-2)\lambda]}$$

$$+ \alpha_o^{3/2}[x-(n-1)\lambda-(1/2)\alpha_o]e^{-\alpha_o[x-(n-2)\lambda]}$$

$$- \alpha_o^{1/2}e^{\alpha_o(x-n\lambda]} \quad , \tag{42}$$

$$a_n(x-n\lambda) = \alpha_o^{1/2}e^{\alpha_o[x-(n+2)\lambda]}$$

$$- \alpha_o^{3/2}[x-(n+1)\lambda+(1/2)\alpha_o]e^{\alpha_o(x-(n+2)\lambda]}$$

$$- \alpha_o^{1/2}e^{-\alpha_o[x-n\lambda]} \quad , \tag{43}$$

$$a_{n+1}(x-n\lambda) = \alpha_o^{3/2}[x-(n+1)\lambda-(1/2)\alpha_o]e^{-\alpha_o[x-n\lambda]} \quad . \tag{44}$$

This completes the derivation of the Wannier functions,

appropriate to the first order in μ, for the delta function lattice. In the next section, we consider a nonuniform lattice.

IV. Equations of Motion for the Envelope Function in a One-Dimensional Lattice with Linear Variation of Well Strength

An arbitrary time-dependent electronic wave function, $\Psi(x,t)$, may be expressed as a linear combination of Wannier functions.[17] Thus, we may write

$$\Psi(x,t) = \sum_n f(n,t) a(x-n\lambda) \quad , \tag{45}$$

where the sum is over all lattice sites, n. The time-dependent coefficients $f(n,t)$ constitute the "envelope vector", or if n is regarded as continuous, the envelope function, in the basis set comprised of the Wannier functions. Strictly speaking, the sum in Eq. (45) should be so general as to encompass the set of Wannier functions from every band. However, in what follows, perturbing potentials will be so slowly varying that interband transitions can be ignored. Such transitions are further reduced in importance because we work in the tight-binding limit. Thus, in Eq. (45), we only include the Wannier functions corresponding to the lowest-lying band associated with the isolated atom bound state wave function.

Since the Wannier functions have already been derived in the previous section, the general solution for $\Psi(x,t)$ requires that we formulate and solve the equations of motion for the envelope vector.

-208-

If an external field is present with potential $U(x)$, then it can be shown[18] that the envelope vector is determined by the following set of differential-difference equations:

$$\sum_n \{E_{n-n'} + U_{n,n'}\} f(n,t) = i\hbar \frac{\partial f(n',t)}{\partial t} \qquad , \qquad (46)$$

one for each value of n'. In Eq. (46),

$$E_{n-n'} = 1/N \sum_k E(k) e^{-ik(n-n')\lambda} \qquad , \qquad (47)$$

where E is given by Eq. (21), and where the sum over wave vector corresponds, again, to a single Brillouin zone. In addition, in Eq. (46),

$$U_{n,n'} = \int a^*(x-n'\lambda) U(x) a(x-n\lambda) dx \qquad , \qquad (48)$$

where the integral goes over the entire crystal.

These equations can be adapted so they are useful in treating the nonuniform lattice without an external field. To see how this is done, consider a uniform lattice of delta function wells of the type introduced in Section 2. We may refer to aV_0 as the well strength. We now add to the potential, defined by this lattice of wells, the function Δ defined as follows:

$$\Delta = -aV_0 \sum_n n\delta(x-n\lambda) \qquad , \qquad (49)$$

where δ is the Dirac delta function and aV_0 is a "unit" of added well strength. Clearly, Δ adds the strength naV_0 to the original well located at the nth lattice site. Furthermore, this added

strength depends linearly, through the multiplying factor n, upon lattice distance. In this way, we have produced a nonuniform lattice in which the well strength changes linearly with position. However, from the formal point of view, we may regard Δ as an external field (although somewhat complex in nature) and observe that it should play the same role as U(x) defined above. With this convention, we are choosing to solve the problem in a Wannier representation corresponding to a <u>uniform</u> lattice whose well depth is the same as the depth at the n=o site in the nonuniform lattice. Thus, our basis set will <u>not</u> drift with the electron. On the other hand, once having solved the problem in a representation corresponding to a fixed point, a simple change of parameters provides the solution for a representation referred to any other point in the lattice.

Substituting Δ for U(x) in Eq. (48) and using r in place of n in Eq. (49), we obtain

$$U_{n,n'} = -av_o \sum_r r \int a^*(x-n'\lambda)\delta(x-r\lambda)a(x-n\lambda)dx \quad . \quad (50)$$

Both $U_{n,n'}$ and $E_{n-n'}$ can now be evaluated through use of Eqs. (41) through (44) and Eq. (21), respectively. We first evaluate $E_{n-n'}$. Substitution of Eq. (21) into (47), and making use of Eqs. (15) and (40), yields

$$E_{n-n'} = -(\hbar^2\alpha_o^2)/2m\{\delta(n-n')+2e^{-\alpha_o\lambda}[\delta(n-[n'+1]+\delta(n-[n'-1])]\}. \quad (51)$$

The sum involving $E_{n-n'}$ in Eq. (46) then becomes

-210-

$$\sum_n E_{n-n'} f(n,t) = -(\hbar^2 \alpha_o^2)/2m \{ f(n',t) + 2e^{-\alpha_o \lambda} [f(n'+1,t) + f(n'-1,t)] \} \; .$$

(51)

For the calculation of $U_{n,n'}$, we make use of Eq. (50), together with Eqs. (41) through (44). Since the Wannier functions specified by this last set of equations are real, it is clear that

$$U_{n,n'} = U_{n',n} \qquad .$$

(53)

Furthermore, since in the rth term in Eq. (50) the delta function is located at the rth site, the Wannier functions which enter the integrand are restricted to those which have terms of order μ or lower at $x=r\lambda$. In fact, a careful analysis shows that the only matrix elements which are of order μ or lower are $U_{n,n}$, $U_{n,n-1}$, and $U_{n,n+1}$. We outline the analysis for these surviving matrix elements. The analysis for those which do not survive is performed in a similar manner.

We represent, in semidiagrammatic fashion, the Wannier functions whose four parts appear, respectively, as Eqs. (41), (42), (43), and (44) as follows:

$$a(x-n) = \overset{\vdash}{a}_{n-2} + \overset{\bowtie}{a}_{n-1} + \overset{-}{a}_n + \overset{\bowtie}{a}_{n+1} \qquad .$$

(54)

The four terms on the right of this equation correspond, respectively, to the functions in Eq. (41), (42), (43), and (44). The Wannier function is continuous (to the first order in μ) at the junction between the various integrals. For example, the junction between the first two terms on the right of Eq. (54) occurs at $x=(n-1)\lambda$. While the second term on the right is nonzero in the

-211-

interval between $x=(n-1)\lambda$ and $x=n\lambda$. At the junction between these two intervals located at $x=(n-1)\lambda$, the functions belonging to the respective intervals are equal, out to the first order in μ. Therefore, it is redundant to assign respective "closed" intervals to both terms.

In Eq. (54), only the term which is nonzero in the interval between $x=(n-1)\lambda$, and $x=n\lambda$ is defined on a closed interval. This is indicated by the two vertical bars at the ends of the horizontal bar over the term. In contrast, the first term in Eq. (54) is only defined on an interval closed at one end, i.e., at $x=(n-2)\lambda$. It is not necessary to define to the first term at $x=(n-1)\lambda$, because the function is appropriately defined there by the second term in Eq. (54). The same considerations apply to the third and fourth terms and their respective intervals in Eq. (54). It will be noted that a vertical bar is associated with every junction between intervals as well as at the ends of the outermost intervals, so that the function is defined at the end of each interval which appears in Eq. (54). There are five such distinct ends.

Our preoccupation with the _ends_ of the intervals stems from Eq. (50), where the delta function in the integrand makes it necessary to know the values of the Wannier functions only at the junctions between intervals, or, what is the same thing, at the ends of intervals.

Suppose we begin by evaluating $U_{n,n}$. Then in Eq. (50) applied to this case, the first end which needs to be considered lies at $x=(n-2)\lambda$ or at $r=n+2$. If we substitute Eq. (54) into (50),

written for $U_{n,n}$, then terms in the integrand, such as

$$-av_o(n-1)[\overset{\llcorner}{a}_{n-2}(x=[n-1]\lambda)][\overset{\llcorner}{a}_{n-1}(x=[n-1]\lambda)] \tag{55}$$

must be set equal to zero because the first factor in curly brackets, by reference to the open overhead bar, is zero at $x=(n-1)\lambda$. (note that the asterisk denoting the complex conjugate has been dropped in Eq. (55) since the Wannier functions are real.) Using this kind of analysis, $U_{n,n}$ may be represented as

$$U_{n,n} = -av_o\{(n-2)\overset{\llcorner}{a}_{n-2}\overset{\llcorner}{a}_{n-2}+(n-1)\overset{\lrcorner}{a}_{n-1}\overset{\lrcorner}{a}_{n-1}+n\overset{\lrcorner}{a}_{n-1}\overset{\lrcorner}{a}_{n-1}$$

$$+ (n+1)\overset{\lrcorner}{a}_{n+1}\overset{\lrcorner}{a}_{n+1} + (n+2)\overset{\lrcorner}{a}_{n+1}\overset{\lrcorner}{a}_{n+1}\} \tag{56}$$

in which the various "a" terms are to be evaluated at the site corresponding to the prefactor of each product. Making these evaluations with the aid of Eqs. (41) through (44), we arrive at the result

$$U_{n,n} = -av_o n\alpha_o \quad , \tag{57}$$

in which we have neglected all terms of the order μ^2 or greater. The evaluation of $U_{n,n-1}$ proceeds in the same manner. However, we require the Wannier function centered at $x=(n-1)\lambda$. This may be obtained directly from Eqs. (41) through (44) by simply replacing n in these equations by n-1. The first integral of importance will now lie between $x=(n-3)\lambda$ and $x=(n-2)\lambda$. To avoid confusion, we denote the nonzero parts of this Wannier function by b instead of by a as in Eq. (54), written for the function

-213-

centered at $x=n\lambda$. In place of Eq. (54), we now have

$$a(x-[n-1]\lambda) = \overleftarrow{b}_{n-3} + \overleftrightarrow{b}_{n-2} + b_{n-1} + \overrightarrow{b}_n \quad , \tag{58}$$

and substituting <u>both</u> Eqs. (54) and (58) into (50) we get

$$U_{n,n-1} = -av_o\{(n-2)\overleftrightarrow{b}_{n-2}\overleftarrow{a}_{n-2}+(n-1)\overleftrightarrow{b}_{n-2}\overleftrightarrow{a}_{n-1}+n\overleftrightarrow{b}_n\overleftrightarrow{a}_{n-1}+(n+1)\overleftrightarrow{b}_n\overleftrightarrow{a}_{n+1}\} \quad . \tag{59}$$

References to Eqs. (41) through (44) and to the shifted version for $a(x-[n-1]\lambda)$ shows that this has the value,

$$U_{n,n-1} = U_{n-1,n} = av_o(2n-1)\frac{\alpha_o}{2}e^{-\alpha_o\lambda} \quad , \tag{60}$$

in which terms of order μ^2 and higher have been neglected. By symmetry, we obtain

$$U_{n,n+1} = U_{n+1,n} = av_o(2n+1)\frac{\alpha_o}{2}e^{-\alpha_o\lambda} \quad , \tag{61}$$

The sum involving $U_{n,n'}$ in Eq. (46) may now be computed using Eqs. (55), (60), and (61). The result is,

$$\sum_n U_{n,n'}f(n,t) = -\alpha_oav_o\{n'f(n',t) + \frac{\mu}{2}(2n'-1)f(n'-1,t)$$
$$+ \frac{\mu}{2}(2n'+1)f(n'+1,t)\} \tag{62}$$

where we have used Eq. (19).

We may now write Eq. (46) in its entirety, combining in it Eqs. (52) and (62). In so doing, we drop the prime on the n. We get

-214-

$$-[(\hbar^2\alpha_o^2)/2m]\{f(n,t)+2\mu[f(n+1,t)+f(n-1,t)]\}$$

$$-\alpha_o av_o\{nf(n,t)+(\mu/2)[(2n-1)f(n-1,t)+(2n+1)f(n+1,t)]\}$$

$$= i\hbar(\partial f[n,t]/\partial t) \qquad . \qquad (63)$$

This equation represents the set of differential-difference equations which determine the envelope vector. These equations are correct to the first order in μ. Furthermore, they are linear so that we have to solve them only for the initial condition

$$f_k(n,o) = A_k\delta_{nk} \qquad , \qquad (64)$$

where A_k is a constant and δ_{nk} is the Kronecke δ. The most general initial condition can then be expressed in the form

$$f(n,o) = \sum_k A_k\delta_{nk} \qquad . \qquad (65)$$

And since the equations are linear, the general time dependent solution will be given by

$$f(n,t) = \sum_k f_k(n,t) \qquad , \qquad (66)$$

in which $f_k(n,t)$ is the solution corresponding to the initial condition, Eq. (64). In the next section, we obtain analytical expressions for $f(n,t)$, correct out to the first order in μ.

V.   Solution of the Set of Differential-Difference Equations

In order to solve the set of equations (63), we introduce a generating function $G(r,t)$ defined as follows:

-215-

$$G(r,t) = \sum_{n=-\infty}^{n=\infty} f(n,t)r^n \quad , \tag{67}$$

in which r is not to be confused with any previous use, in this paper, of that symbol. From the definition of $G(r,t)$, it follows that

$$\partial G/\partial r = \sum_{n=-\infty}^{n=\infty} nf(n,t)r^{n-1} \quad , \tag{68}$$

$$r^2 \, \partial G/\partial r + rG = \sum_{n=-\infty}^{n=\infty} nf(n-1,t)r^n \quad , \tag{69}$$

$$\partial G/\partial r - G/r = \sum_{n=-\infty}^{n=\infty} nf(n+1,t)r^n \quad . \tag{70}$$

Equation (63) is now multiplied by r^n and summed over n. Equations (68) through (70) can be substituted into this result to yield a partial differential equation for the generating function. This equation, Eq. (71), follows, simplified by the use of the quantities ω and τ which are themselves defined in Eqs. (72 and (73).

$$G + \omega r(\partial G/\partial r) - (1/i)(\partial G/\partial \tau) + \mu\{2(r+[1/r])G$$
$$+ (\omega/2)(r-[1/r])G + \omega(r+[1/r])r(\partial G/\partial r)\} = 0 \quad , \tag{71}$$

$$\omega = 2mav_o/\hbar^2\alpha_o \quad , \tag{72}$$

$$\tau = (\hbar\alpha_o^2/2m)t \quad . \tag{73}$$

It is possible to provide an analytical solution of Eq. (71) to the first order in μ. We accomplish this by first writing G as a power series in μ, neglecting terms beyond the first order:

-216-

$$G = G_o(r,\tau) + \mu G_1(r,\tau) \qquad . \tag{74}$$

If we make use of the initial condition, Eq. (64), in Eq. (67), we obtain

$$G(r,o) = A_k r^k \qquad . \tag{75}$$

The use of this condition implies that the coefficients in G are $f_k(n,t)$.

Since μ depends on λ, and A_k is arbitrary, not depending on λ, it follows, by comparison of Eqs. (74) and (75),

$$G_o(r,o) = A_k r^k \qquad , \tag{76}$$

$$G_1(r,o) = 0 \qquad . \tag{77}$$

Substitution of Eq. (74) into (71), and setting the coefficient of each power of μ equal to zero, leads to the following two equations:

$$G_o + \omega r (\partial G_o/\partial r) = (1/i)(\partial G_o/\partial \tau) \qquad , \tag{78}$$

$$G_1 + \omega r(\partial G_1/\partial r) - (1/i)(\partial G_1/\partial \tau) + 2(r+[1/r])G_o$$
$$+ (\omega/2)(r-[1/r])G_o + \omega(r+[1/r])r(\partial G_o/\partial r) = 0 \quad . \tag{79}$$

The solution to Eq. (78), satisfying the initial condition, Eq. (76), is immediate:

$$G_o(r,\tau) = A_k r^k e^{i(1+k\omega)\tau} \qquad . \tag{80}$$

-217-

That this is the proper solution can be made evident by direct substitution of the result into Eq. (78). The solution of Eq. (79) is not difficult, but somewhat more involved. Matters are facilitated by defining the quantity ϕ by the following equation:

$$G_1(r,\tau) = \phi(r,\tau)e^{i(1+k\omega)\tau} \qquad ; \tag{81}$$

and things are made still easier by defining

$$\xi = \ln r \qquad . \tag{82}$$

Substitution of Eqs. (80), (81), and (82) into (79) leaves us with

$$(\partial\phi/\partial\xi)-(1/i\omega)(\partial\phi/\partial\tau)-k\phi+A_k e^{k\xi}\{2(2+[\omega k])\cosh\xi+\omega\sinh\xi\} = 0 \quad . \tag{83}$$

If we further define y through the relation

$$\phi = y-A_k e^{k\xi}\{\cosh\xi+2([2/\omega]+k)\sinh\xi\} \qquad , \tag{84}$$

and substitute this into Eq. (83), we are left with

$$\partial y/\partial\xi = (1/i\omega)(\partial y/\partial\tau) \qquad . \tag{85}$$

Equation (85) has the solution

$$y = y(z) \qquad , \qquad z = \xi + i\omega\tau \qquad . \tag{86}$$

The initial condition on ϕ is obtained by substituting Eq. (81) into (77) with the result,

$$\phi(r,o) = 0 \qquad . \tag{87}$$

-218-

Substituting Eq. (86) into (84) and the result into (87) gives,

$$y = A_k e^{k(\xi+i\omega\tau)} \{\cosh(\xi+i\omega\tau)+2([2/\omega]+k)\sinh(\xi+i\omega\tau)\} \qquad . \quad (88)$$

Substituting Eq. (88) into (84) and the latter into (81), at the same time using Eq. (82) to convert ξ into r, gives

$$G_1 = A_k e^{i(\omega+1)\tau}\{(1/2)e^{i(k+1)\omega\tau}+([2/\omega]+k)(e^{i\omega\tau}-1)-1/2\}r^{k+1}$$

$$+A_k e^{i(\omega+1)\tau}\{(1/2)e^{i(k-1)\omega\tau}-([2/\omega]+k)(e^{-i\omega\tau}-1)-1/2\}r^{k-1}. \quad (89)$$

Substituting Eqs. (80) and (89) into (74) gives, finally, for $G(r,\tau)$,

$$G(r,\tau) = A_k e^{i(1+k\omega)\tau}r^k$$

$$+\mu A_k e^{i(1+\omega)\tau}\{(1/2)e^{i(k-1)\omega\tau}-([2/\omega]+k)(e^{-i\omega\tau}-1)-1/2\}r^{k-1}$$

$$+\mu A_k e^{i(1+\omega)\tau}\{(1/2)e^{i(k+1)\omega\tau}+([2/\omega]+k)(e^{i\omega\tau}-1)-1/2\}r^{k+1}.$$
$$(90)$$

Equation (90) shows that, to the first order in μ, an electron confined initially to a single site located at $x=k\lambda$, makes transitions only to the adjacent sites located at $x=(k-1)\lambda$ and $k=(k+1)\lambda$. This follows from the fact that only the $k-1$, k, and $k+1$ terms appear in the relevant generating function specified by Eq. (89), and that the coefficients of powers of r are the corresponding components of the envelope function. In order to observe transitions to more distant sites, it is necessary to solve the theory out to a higher order in μ, i.e., it is necessary

examine a less tightly bound case.

As a result of this situation, Eq. (66) can be simplified in the case of tight binding, to

$$f(n,t) = f_{n-1}(n,t) + f_n(n,t) + f_{n+1}(n,t) \quad . \tag{91}$$

The first term on the right of Eq. (91) is obtained as the coefficient of r^{k+1} in Eq. (90), with $k=n-1$. The second term on the right of Eq. (91) is the coefficient of r^k in Eq. (90), with $k=n$. Similarly, the last term on the right of Eq. (91) is the coefficient of r^{k-1} in Eq. (90), with $k=n+1$. As a result, we have

$$f(n,\tau) = A_n e^{i(1+n\omega)\tau}$$
$$+ \mu A_{n+1} e^{i(1+\omega)\tau} \{(1/2) e^{in\omega\tau} - ([2/\omega]+n+1)(e^{-i\omega\tau}-1)-1/2\}$$
$$+ \mu A_{n-1} e^{i(1+\omega)\tau} \{(1/2) e^{in\omega\tau} + ([2/\omega]+n-1)(e^{i\omega\tau}-1)-1/2\}. \tag{92}$$

This is the general solution, to the first order in μ, of the differential-difference equation represented by Eq. (64). The arbitrary initial condition is determined by the set of constants A_n appearing in Eq. (92). The probability density corresponding to the envelope packet itself is obtained from

$$f^*(n,\tau)f(n,\tau) = A_n^2 + \mu A_n A_{n+1}\{\cos\omega\tau - 2([2/\omega]+n+1)\cos n\omega\tau$$
$$+ 2([2/\omega]+n+1/2)\cos(n-1)\omega\tau\}$$
$$+ \mu A_n A_{n-1}\{\cos\omega\tau + 2([2/\omega]+n-1)\cos(n-2)\omega\tau$$

$$- 2([2/\omega]+n-1/2)\cos(n-1)\omega\tau\} \qquad . \qquad (93)$$

This too, is carried only as far as the first order in μ.

VI.   The Long Wavelength Limit

The usual effective mass formalism deals with long wavelength packets, and, to the extent that off-diagonal elements $u_{n,n'}$ can be neglected (as well as interband transitions), n can be regarded as continuous, and $f(n,t)$ or $f(n\lambda,t)$ then satisfies a normal Schrödinger wave equation with the effective mass m_*, Eq. (26), replacing the actual mass. $u_{n,n}=u(n\lambda)$ replaces the conventional potential. Unfortunately, in the present case, as Eqs. (60) and (61) clearly show, off-diagonal terms persist in the matrix for the potential out to order $\mu=e^{-\alpha_o\lambda}$. Consequently, a simple Schrödinger-type wave equation is not recovered. Since the off-diagonal terms connect adjacent sites in the tight binding case, this is not surprising.

On the other hand, if the properties of the lattice change slowly enough compared to the component wavelengths of the wave packet, we can still generate a Schrödinger-like equation with an effective potential proportional (but not equal) to the change in lattice potential. To prove this, we substitute Eqs. (72) and (73) into (63) and obtain

$$f(n,\tau) + 2\mu[f(n+1,\tau)+f(n-1,\tau)]$$

$$+ \omega\{nf(n,\tau)+(\mu/2)[(2n-1)f(n-1,\tau)+(2n+1)f(n+1,\tau)]$$

$$= -i(\partial f[n,\tau]/\partial\tau) \qquad . \qquad (94)$$

Expanding $f(n-1,\tau)$ and $f(n+1,\tau)$ about n, and retaining terms
through the second order, we get, treating n as continuous,

$$f + \omega nf + i(\partial f/\partial \tau)$$

$$+ \mu\{4f+2(\partial^2 f/\partial n^2)+2\omega nf+2\omega(\partial f/\partial n)+\omega(\partial^2 f/\partial n^2)-\omega(\partial f/\partial n)\} = 0. \quad (95)$$

Now assume that the "unit" of changing well strength, av_0, is as
small as μ or smaller, and that we consider the motion of the
wave packet close enough to $n=0$, so that the perturbing potential
is always of order μ or less. Then Eq. (72) shows that ω is of
order μ. Terms in Eq. (95) multiplied by $\mu\omega$ are of second order
and may therefore be neglected. The remaining terms give

$$2\mu(\partial^2 f/\partial n^2) + [1+4\mu+\omega n]f = -i(\partial f/\partial \tau) \qquad . \qquad (96)$$

The term $(1+4\mu)f$ in this equation appears only because we have
not chosen the band edge, E_b, of Eq. (22) as the energy zero.
As such, it only introduces the phase factor $e^{i(1+4\mu)\tau}$ into f.
We can remove both $(1+4\mu)f$ and the phase factor by defining f'
through

$$f = f'(n,\tau)e^{i(1+4\mu)\tau} \qquad (97)$$

and substituting in Eq. (96). The result is

$$2\mu(\partial^2 f'/\partial n^2) + \omega nf' = -i(\partial f'/\partial \tau) \qquad . \qquad (98)$$

Clearly, f' is the wave function, with E_b chosen as zero of
energy. Introducing the continuous coordinate of position

$$x = n\lambda \tag{99}$$

into Eq. (98) gives

$$2\mu\lambda^2(\partial^2 f'/\partial x^2) + (\omega/\lambda)xf' = i(\partial f'/\partial\tau) \quad . \tag{100}$$

Finally, using Eqs. (72) and (73) to transform back to t, and to eliminate ω, gives

$$-(\hbar^2/2m_*)(\partial^2 f'/\partial x^2) - (\alpha_o av_o/\lambda)xf' = i\hbar(\partial f'/\partial t) \quad , \tag{101}$$

where m_* is given by Eq. (26). This is the Schrödinger equation in which the real mass is replaced by the effective, and where the potential is related to the changing well strength. In fact, for the potential we have, from Eq. (101),

$$v(x) = -(\alpha_o av_o x/\lambda) = -(ma^2 V_o v_o x/\lambda\hbar^2) \quad , \tag{102}$$

where we have used Eq. (15). This potential varies linearly with lattice position, and a solution to Eq. (101) is easily obtained by using the WKB method.[19]

However, Eq. (101) is limited to very long wavelength envelope functions, and to extremely gradual changes in lattice properties. In contrast, Eq. (63) and its general solution, Eq. (92), is not limited in this manner. Both equations are restricted, however, to the tight binding case.

Acknowledgement

-223-

References

1. H. Kroemer, Proc. IRE $\underline{45}$, 1535 (1957).

2. R. L. Anderson, IBM J. Res. $\underline{4}$, 283 (1960).

3. R. L. Anderson, Solid State Electronics $\underline{5}$, 341 (1962).

4. S. S. Perlman and D. L. Feucht, Solid State Electronics $\underline{7}$, 911 (1964).

5. A. R. Riben and D. L. Feucht, Int. J. Electronics $\underline{20}$, 583 (1966).

6. T. Gora and F. Williams, International Conference on II-VI Semiconducting Compounds (Benjamin, New York, 1967), ed. by D. G. Thomas.

7. T. Gora and F. Williams, Phys. Rev. $\underline{177}$, 1179 (1969).

8. W. Kohn, Elsewhere in these Proceedings.

9. R. H. Parmenter, Phys. Rev. $\underline{97}$, 587 (1955).

10. J. M. Ziman, Principles of the Theory of Solids, Chapter 6 (Cambridge University Press, 1964).

11. Ibid., p. 148.

12. R. Kronig and W. G. Penney, Proc. Roy. Soc. $\underline{A130}$, 499 (1931).

13. R. A. Smith, Wave Mechanics of Crystalline Solids (Wiley and Sons, New York, 1961), p. 149.

14. Ibid., Chapter 4.

15. J. M. Ziman, Principles of the Theory of Solids, p. 15.

16. Ibid., p. 149.

17. Ibid., p. 151.

18. Ibid., p. 152

19. D. Rapp, Quantum Mechanics, Chapter 9 (Holt, Rinehart and Winston, New York, 1971).

DUCTILE VS. BRITTLE BEHAVIOR OF CRYSTALS

J. R. Rice and R. M. Thomson

Abstract

A necessary criterion for brittle fracture in crystals is established in terms of the spontaneous emission of dislocations from an atomically sharp cleavage crack. We have calculated the stability of a sharp crack against emission of a blunting dislocation for a number of crystals and crystal types in two dimensions and the energy to form a stable loop of dislocation from the crack tip in three dimensions. We find that contrary to previous expectations, an atomically sharp cleavage crack is stable in a wide range of crystal types, but that in the face centered cubic investigated, blunting reactions occur spontaneously. Of the body centered metals investigated, iron is an intermediate case between the brittle and ductile cases, and the ionic and covalent crystals investigated are all stable against dislocation emission. Qualitatively, when the dimensionless parameter $\mu b/\gamma$ is larger than about 7.5 - 10, we expect the crystal to support a sharp cleavage crack, though a more precise criterion must include core size of the dislocation and geometrical factors.

DUCTILE VS. BRITTLE BEHAVIOR OF CRYSTALS

J. R. Rice and R. M. Thomson

I.   Introduction

This paper is concerned with the problem of the ductile
vs. brittle response of crystals. There is in the literature
of this subject a conventional understanding that a solid will
either be ductile or truly brittle depending upon the ratio of
theoretical shear strength to theoretical tensile strength.
See Kelly (1966). We believe that a correct description of this
competition should include actual dislocation processes at cracks,
since the ductile response of the solid must produce dislocations
in order to yield.

A truly ductile material like pure copper apparently can-
not sustain a cleavage crack, but may fail by plastic instability
and necking on a gross scale. Stronger materials apparently also
exhibit the same essentially plastic necking phenomenon on a more
microscopic scale through the process of hole growth, although
the macroscopic appearance of the failure is crack-like. On the
opposite end of the scale, some materials like diamond and mica
apparently can undergo pure brittle cleavage with no discoverable
plasticity associated with the process. In between these two
extremes, there apparently exists a class of materials where a
cleavage crack in the true atomically sharp sense exists, but is

surrounded and associated with an atmosphere of dislocations. (Burns and Webb, 1970; Burns, 1970). This intermediate case exhibits many complexities, such as high effective surface energies, plastic zones surrounding the crack tip, etc., but there is no reason to suppose that, provided hole growth is not occurring, the crack tip is not sharp on the atomic level. Of course, experimental proof of this statement in any given case will be indirect at best, but the theoretical descriptions of cleavage and of hole growth are sufficiently different as to make one wish to distinguish as clearly as possible between the separate physical cases.

Kelly, Tyson and Cottrell (1967) were the first to pose this problem of brittle vs. ductile fracture in an essentially proper way when they attempted to test the self-consistency of the proposition that a cleavage crack can exist in a particular type of crystal. They, in effect, asked: "If a cleavage crack were created by some process in a crystal, would the tip spontaneously blunt as the result of shear by the atoms of the tip region?" They then postulated that such would be the case if the highest shear stress in the vicinity of the crack exceeded the theoretical shear strength of the material. However, this criterion cannot be sufficient for the crack to blunt, because the shear stress near a crack is not everywhere constant on the shear plane as it would have to be to cause the atoms to shear past one another uniformly. Instead, the stress is highly localized in the vicinity of the crack tip. By geometrical necessity, localized shear on a plane intersecting the crack tip caused by

-227-

the high shear stresses there, matched to a non-sheared region at greater distances on the same plane, where the stress is below the theoretical strength simply defines a dislocation. Hence, a blunting reaction at the crack tip requires the production (or annihilation) of dislocations.

Kelly, Tyson, and Cottrell indeed discussed one kind of dislocation reaction for NaCl. They calculated the approximate energy to form a full metastable loop of dislocation near the crack in NaCl, and found the energy to be prohibitively high. Others have also discussed dislocation formation near crack tip. (Armstrong, 1966; Kitajima, 1966) Armstrong has, in particular, estimated the formation energy of a dislocation dipole loop completely surrounding a circular crack in a crystal. However, this type of dislocation interaction does not correspond to a blunting reaction and does not directly address the question we pose.

In this paper, we shall propose models for the production of a dislocation from the tip of the crack in such a way that after the dislocation expands under the external stress field as concentrated by the presence of the crack, an atomically sharp crack will have been blunted by one atomic plane, Figure 1. This blunted crack will then be trapped at the original lattice position until the external stress is increased substantially (when probably further dislocation blunting may be possible). Crystals for which there exists a large energy barrier for this emission can be expected to be good candidates for brittle cleavage (but perhaps where the crack has associated with it clouds of disloca-

tions which are formed or captured through other processes in the nearby lattice). In order for a dislocation to blunt a crack, it is necessary for the Burgers vector to have a component normal to the crack plane, and for the slip plane to intersect the crack line (or crack front) along its whole length, i.e., the crack line must be contained within the slip plane.

It is, of course, possible to conceive of a process by which the crack may be blunted by dislocations which are formed from nearby sources and which are emitted on precisely the right plane to blunt the crack tip. In view of the fact that the stress field in a region surrounding a cleavage crack will be above the macroscopic plastic flow stress for the material, one might suppose this could be an effective blunting mechanism. However, macroscopic yield is a property associated only with regions of the size of many microns, even for fairly ductile materials. Hence, for a random position of the crack tip, the probability of finding a source on the correct plane at a distance from the crack for which the source can operate is small.

The plan of the paper is as follows. In the next section, we shall consider the various forces operating between a crack and a dislocation in two dimensions. These forces are 1) the force on a dislocation due to the stress field surrounding the crack, 2) the surface tension force caused by creating more surface at the blunted crack, and 3) the image force of the dislocation in the free surface of the crack. The first term repels the dislocation, and the latter two attract it toward the crack

-229-

tip, giving rise to the possibility of a position of unstable equilibrium. In Section III we estimate the activation energy for formation of a dislocation half loop out of the crack under the action of these forces when an energy barrier exists. Finally, in Section IV we discuss the physical consequences of our calculations.

II.   Forces on a Dislocation Near a Crack Tip

Let us suppose that the loads on the body considered act symmetrically about the plane of a straight crack, so that before emission of the dislocation, only the tensile opening mode of relative crack surface displacement is present. If K_I is the "elastic stress intensity factor" (see, for example, Rice, 1968) due to the loads, then in two dimensions the in-plane shear stress acting at distance ρ on the slip plane of Figure 2 is

$$\sigma_{\rho\phi} = K_I (8\pi\rho)^{-\frac{1}{2}} \sin \phi \cos \phi/2 \tag{1}$$

For this equation to be valid, ρ must be a small fraction of overall crack length. The anti-plane shear stress component (i.e., in the direction of b_s) is zero. The release of potential energy of the body and load system per unit of new crack area is

$$G = \frac{1-\nu^2}{E} K_I^2 \tag{2}$$

(E = Young's modulus, ν = Poisson ratio.) At the fracture load predicted on the Griffith theory, $G = 2\gamma$ where γ is the true surface energy of the crack plane. Thus, if the applied load on the

body is chosen as that which would cause fracture if no dislocations were emitted, then the force (shear stress times Burgers vector) on the dislocation segment shown in Figure 2 due to the applied load is

$$f_\sigma = \sigma_{\rho\phi} \, b_e = \left[\frac{E\gamma b}{4\pi(1-\nu^2)\xi} \right]^{\frac{1}{2}} \sin\phi \, \cos\phi/2 \, \cos\psi \qquad (3)$$

Here $\xi = \rho/b$ and $b_e = b \cos\psi$ is the edge component. The screw component b_s does not appear since the applied load induces no shear stress in this direction.

There will be an "image" force which tends to pull the dislocation back into the crack. This may be inferred from direct solution of the corresponding elasticity problem, as presented in most general form by Atkinson (1966) for anisotropic materials. A rather different derivation is presented in the Appendix, for a straight dislocation parallel to the crack tip in an isotropic materials. (The procedure of the Appendix is based on energy considerations and properties of point functions, rather than on direct solution of the elastic field equations for a dislocation near a crack. The approach is readily generalized to other elastic interaction problems, and may be of some interest in itself.) The resulting image force (Equation A-14) is

$$f_i = - \frac{E b_e^2}{8\pi(1-\nu^2)\rho} - \frac{E b_s^2}{8\pi(1+\nu)\rho} \qquad (4)$$

This is a remarkable result, because precisely the same expression for the image force is obtained in the case of a dislocation in a half space with its core lying parallel to, and at perpendicular

-231-

distance ρ from, the free surface (see, for example, Hirth and Lothe, 1968). Both of the above forces are calculated on the assumptions of "infinitesimal" elasticity, and actual geometry changes at the crack tip due to emission of the dislocation have been neglected. The image force term may be rewritten from Figure 2 as

$$f_i = - \frac{Eb(1-\nu \sin^2 \psi)}{8\pi(1-\nu^2)\xi} \tag{5}$$

$$\xi = \rho/b$$

In Figure 1, as the dislocation is formed, a ledge is left behind, and as the core comes through the surface, forces due to the formation of the ledge must be included. As a function of the position, $\xi = \rho/b$, of the dislocation, the ledge energy is given by

$$V_\ell = \frac{2}{\pi} \gamma b \cos\psi \sin\phi \tan^{-1} \frac{2\xi}{e^{3/2}\xi_0} \tag{6}$$

In deriving (6), we have used the misfit function for a Peierls model of the dislocation with a width or core cut off ξ_0, which is consistent with the quantity r_0/b where r_0 is defined by Hirth and Lothe (1968), p. 212. The $\sin\phi$ dependence is taken to approximately represent the modification of ledge energy from γb_e due to slip plane inclination. The force is consequently

$$f_\ell = - \frac{2}{\pi} \frac{\gamma\alpha \cos\psi \sin\phi}{\xi^2 + \alpha^2} \quad \text{where} \quad = e^{3/2} \xi_0/2$$

-232-

Note that the ξ^{-1} and ξ^{-2} attraction back toward crack, due to the image and ledge forces, outweights the $\xi^{-\frac{1}{2}}$ force resulting from the applied load when ξ is small, whereas just the opposite happens when ξ is large. Hence, the equilibrium position of the dislocation is unstable, and the dislocation will be driven away indefinitely, until it reaches some obstacle, if it ever attains a distance from the crack tip greater than the equilibrium distance.

The critical distance, ξ_c at which a straight dislocation is in unstable equilibrium under these three forces is, from (3), (5), and (6), given by the solution of

$$f_{tot.} = \mu b \left[- \frac{1}{4\pi\xi} \frac{1-\nu}{1-\nu} \sin^2\psi - \frac{2}{\pi\eta^2\beta'} \frac{\alpha}{\xi^2+\alpha^2} + \frac{1}{\eta\beta}\left(\frac{1}{2\pi(1-\nu)\xi}\right)^{\frac{1}{2}} \right] = 0 \tag{7}$$

We have used the following abbreviations,

$$\frac{1}{\beta} = \cos\psi \, \sin\phi \, \cos\phi/2,$$

$$\frac{1}{\beta'} = \cos\psi \, \sin\phi, \tag{8}$$

$$\eta^2 = \mu b/\gamma,$$

where μ is the shear modulus. If the value of ξ_c is less than the core cut off, we presumably have a case where the method does not apply, and spontaneous generation is a good possibility. Equations (3) and (5) diverge for $\xi \to 0$, but, of course, the forces they represent must actually in toto approach zero as $\xi \to 0$ because of nonlinear core effects.

-233-

We display in Table I the relevant physical data with the values we have chosen. The values of γ contain the greatest degree of uncertainty, and we discuss the problems associated with surface energy for our calculation in Section IV. Table II displays the results of the solution of Eq. (7) for the various solids, together with suggested values from Hirth and Lothe (1968) for the elastic cut-off for comparison. We also list some cruder estimates of the critical distance, ξ_c' and ξ_c''. In ξ_c', we neglect the effect of the ledge in Eq. (7), and then the condition is given by

$$\xi_c' = \frac{(1-\nu \sin^2\psi)^2}{8\pi(1-\nu)} \beta^2 \frac{\mu b}{\gamma} \tag{9}$$

In the second estimate, we average over the geometrical and crystallographic effects still present in (9), giving an even cruder estimate,

$$\xi_c'' \simeq \frac{\mu b}{10\gamma} \tag{10}$$

It is interesting to enquire how splitting the dislocation will affect the calculations, since certainly in the face centered metals, splitting does occur. In this case, the Burgers vector will be lowered in magnitude, and the angle, ψ, will be changed. So far as the size of the Burgers vector is concerned, Eq. (7) is dimensionless, and is not affected. The size of the cut-off is sensitive to b, however, and in fact the value we have chosen is just that appropriate to the split dislocation in the face centered metals. Eq. (7) does contain the crystallography of the Burgers

-234-

vector through the angle, ψ, however, in a rather complicated manner. For the whole dislocation, $\psi = 30°$, and there are two possibilities for the two partials, $\psi = 0°$ and $\psi = 60°$. For the first value, the ratio of the various terms in Eq. (7) is only changed about 5% from their values for the whole dislocation, thus leaving the results of Table II unchanged. For the second case, where $\psi = 60°$, the first term in Eq. (7) becomes about 50% larger than for the whole dislocation, thereby increasing the value of ξ_c. We show a few values for this partial in Table II.

III. Energy Considerations

If the equilibrium point, ξ_c, is larger than the core cut-off, then there is an energy hump for the dislocation to jump in order to be emitted from the crack. Clearly, in this case, it will be impossible for a uniform straight line to be emitted from the crack, because for an infinite length of dislocation, the energy diverges. Instead, a local fluctuation in the form an irregular loop will be formed which, beyond the saddle point configuration, will expand under the external stress. The factors determining the saddle point configuration are the same as before. Image and ledge forces will predominate for small loop sizes, while for larger ones the external stress is dominant.

In the elastic three dimensional problem, even in the isotropic regime, the problem of determining the saddle point configuration and consequently its energy is intractable in part because the force fields are variable, and in part because rigorous

-235-

treatments of three dimensional dislocation problems are not possible when the shape is complicated. In fact, although we know the image force for a straight dislocation parallel to the crack, image terms are known to be complex when the shape is more complicated, even without the additional complexity of the crack half surface. We shall thus proceed by developing some qualitative insights into the problem, and then make some simplifying assumptions about the saddle point configuration which allow us to make meaningful calculations.

First of all, the stress field due to the applied load is straightforward, and yields a normal force, f_σ, on a dislocation element, $d\underline{\ell}$, given by the local value of the stress, $\underline{\underline{\sigma}}$. This force is

$$f_\sigma = (\underline{\underline{\sigma}} \cdot \underline{b}) \times d\underline{\ell} \tag{11}$$

The ledge force is always a minor term, except possibly very near the crack surface. It will have the primary effect of drawing together the two ends of the loop where they touch the crack surface. It will act just like a pure surface tension on these ends because of the surface energy required to expose the ledge as the loop expands.

The image force, as mentioned above is more complex. In problems of dislocations interacting with normal surfaces, it is a useful first approximation to simply replace the image terms by the action of an image dislocation reflected in the surface. This replacement is not normally rigorous, but yields a fair

approximation. Since in our case of the crack, the straight dis-
location yields the same result as for a normal surface, we shall
adopt this simple stratagem for the crack. This means that in
order to calculate the energy of a dislocation half loop configur-
ation which ends at the crack surface, we need do no more than
calculate the energy of a <u>full</u> loop including the reflected image,
and take half of the result. This energy will then include the
image terms.

One further very important result follows from the presence
of the image term. The image term requires that when a dislocation
approaches an open surface, it must cut the surface at normal in-
cidence. The predominance of the image term in the immediate
vicinity of the crack surface will require the same boundary con-
dition on the dislocation at the crack surface. This condition,
in conjunction with the magnitude of the stress in the region be-
yond ξ_c are the crucial determinants of the total activation
energy.

We note one final qualitative characteristic of the two
dimensional force field which is very suggestive for our three-
dimensional treatment. Except for the ledge term in Eq. (7),
which is important only very near the crack tip, the net force
is the difference between $1/\xi$ and $1/\sqrt{\xi}$, a very broad function
which we have plotted in Fig. 3. The point at which this function
reaches one half its maximum value is about 1.5 ξ_c, and it does
not fall below this value again till ξ becomes approximately 50 ξ_c.
Thus, throughout this very board range, the balance between these

two major forces is nearly constant. Under a constant normal

force, of course the equilibrium shape of the dislocation has

constant curvature, and is circular. For values of ξ less than

ξ_c, the force becomes negative, and including the effect of the

ledge, the curvature will also reverse. Hence the general shape

of the saddle point configuration will be as sketched in Fig. 4.

In view of these qualitative background comments, we feel

justified in proceeding with the assumption that the approximate

equilibrium shape is the simple half circle depicted in Fig. 4a.

In view of the broad maximum in the effective force field, we

believe the energy of the activated state thus computed is a

reasonable estimate of the true energy, and our use of the circu-

lar shape probably introduces no major errors in the calculation.

We thus calculate the total energy of the activated state

of the system, which consists of three parts. 1) The self energy

of the dislocation half loop. As mentioned before, this energy

will automatically include the image term contribution. 2) The

energy of the ledge. 3) The energy gained by the dislocation

loop as it expands under the influence of the stress surrounding

the crack. In order to determine the size of the loop of the

activated state, we locate the maximum energy of the loop as a

function of the loop radius.

The self energy of a dislocation half loop as given by

Hirth and Lothe (1968) is

$$U_{self} = \mu b^3 r \frac{2-\nu}{8(1-\nu)} \ln \frac{8r}{e^2 \xi_o} \qquad (12)$$

-238-

In this equation, r and ξ_0 are respectively the radii of the half circle and radii of the core cut-off in units of the Burgers vector, b. ξ_0 is the same quantity as used for the core cut-off in Section II.

To the self energy must be added the energy of the ledge formed as the loop expands. We write this in the form

$$U_{ledge} = 2\gamma b^2 \cos\psi \sin\phi (r-\xi_0) \tag{13}$$

Eq. (13) simplifies the expression for a ledge as we used it in Section II. Here we assume that when the radius r is greater than the cut-off radius, the ledge is fully formed, and that the energy is linear in the radius of the loop.

Finally, we compute the energy gained by the half loop in the stress field of the crack tip. In terms of Eq. (3) and Fig. 5, the energy to expand the loop from the initial radius, ξ_0, to r is

$$U_\sigma = - \left[\frac{E\gamma b}{4\pi(1-\nu^2)} \right]^{\frac{1}{2}} b^2 \cos\psi \sin\phi \cos\phi/2$$

$$\int_{\xi_0}^{r} dr \int_{0}^{\pi} d\theta \frac{r}{\sqrt{r \sin\theta}} \tag{14}$$

Noting that

$$\int_{0}^{\pi} \frac{d\theta}{\sqrt{\sin\theta}} = \pi^{\frac{1}{2}} \frac{\Gamma\left(\frac{1}{4}\right)}{\Gamma\left(\frac{3}{4}\right)} \tag{15}$$

we then have

$$U_\sigma = -0.9862 \sqrt{\frac{2}{1-\nu}} \, \mu b^3 \, \sqrt{\frac{\gamma}{\mu b}} \, \sin\phi \, \cos\psi \, \cos\phi/2$$

$$\left(r^{\frac{3}{2}} - \xi_0^{\frac{3}{2}}\right) \tag{16}$$

The total energy change for a crack which has emitted a dislocation

$$U_{act} = \mu b^3 \left[r U_0 \, \ell n \, \frac{r}{\xi_0} + U_\ell (r-\xi_0) - \frac{2}{3} U_s \left(\frac{3}{r^2} - \xi_0^{\frac{3}{2}} \right) \right]$$

$$U_0 = \frac{2-\nu}{8(1-\nu)} \tag{17}$$

$$U_\ell = 2 \, \frac{\gamma}{\mu b} \, \cos\psi \, \sin\phi$$

$$U_s = \frac{2.092}{\sqrt{1-\nu}} \sqrt{\frac{\gamma}{\mu b}} \, \sin\phi \, \cos\psi \, \cos\phi/2$$

$(8/e^2 \simeq 1)$. From (17), the condition for the activated state is

$$\frac{dU_{act}}{dr} = 0 = U_0 \, \ell n \, er/\xi_0 + U_\ell + U_s \, r^{\frac{1}{2}} \tag{18}$$

Schematically, the three terms in (17) are sketched in Fig. 6. The stress term eventually always dominates the other two for large r, but depending upon the parameters, the self energy plus the ledge term may be dominant for values near ξ_0. That is to say, depending upon the parameters, there may be a

-240-

spontaneous emission of a dislocation loop with no activation energy, or there may be a finite activation energy to form the loop whose radius is determined by Eq. (18).

We have solved Eq. (17) for the critical loop size, and calculated the activation energy for the list of materials given in Table I. The results are listed in Table III. Further, the functional variation of the activation energy with $\gamma/\mu b$, ξ_0, and orientation parameters is shown in Fig. 7. In this plot, the activation energy [Eq. (17)] is plotted after γ is eliminated through Eq. (18). Here it is convenient to plot the dimensionless energy

$$u_{act} = \frac{U_{act}}{\frac{(2-\nu)^2}{8(1-\nu)} \frac{\beta^2}{\beta'} \mu b^3} \tag{19}$$

in terms of

$$S = \frac{16(1-\nu)}{5\beta'(2-\nu)} \frac{\gamma}{\mu b} \quad \text{and} \quad R_0 = \frac{16\beta'}{5(2-\nu)\beta^2} \xi_0 \quad . \tag{20}$$

Indeed, these forms have been chosen because, for typical values of ν and the orientation parameters, they reduce to

$$u_{act} \simeq \frac{U_{act}}{\mu b^3} , \quad S \simeq \frac{\gamma}{\mu b} , \quad R_0 \simeq \xi_0 \quad . \tag{21}$$

It is seen that the behavior divides mainly into two kinds: If the orientation-dependent core parameter, R_0, is small, there is a substantial energy barrier to dislocation nucleation for all

-241-

values of $\gamma/\mu b$ (i.e., $U_{act} \gtrsim 0.1 \mu b^3$). On the other hand, if R_0 is large there will be an energy barrier only if the surface energy parameter, S, is smaller than that at which the curves cross the S axis; otherwise there is spontaneous nucleation. The curves for the larger values of R_0 rise so steeply that there is essentially a critical value of S below which the energy barrier is substantial, and above which there is spontaneous nucleation, although finer examination shows that there is a narrow range of R_0 and S values for which U_{act} is low enough that thermal activation could blunt the crack for sufficiently long time scales of load application.

IV.   Physical Results and Interpretation

a) <u>Surface Energy</u>. Among the experimentally determined values listed in Table I, the values of γ are the ones for which the only important uncertainties exist.

For tungsten, Cordell and Hull (1969) find that the surface energy varies from a low value of 1700 cgs at low temperature to 6000 cgs at higher temperature. The higher value apparently is due to plasticity induced in the vicinity of the moving crack. Ohter values have been measured by other workers intermediate to these extremes, but we believe the low value listed represents the bare crack.

NaCl is claimed by Class (1964) in his thesis to have a surface energy of 115 cgs, and this value is adopted by Kelly (1966). On the other hand, we believe the value 250 is more

-242-

likely. It is in the range of the experiments of Gilman (1960),
Wiederhorn (1970), and Benson, et al (1955, 1956, and 1959), and
is in reasonable agreement with the theoretical calculations,
which for this crystal should not have a large error, see Tosi
(1964) and MacMillan and Kelly (1972). We list both values, and
results for both, however.

For Al_2O_3, we have a range of values in the literature all
the way from 1000 cgs as given by Kingery (1954) for the basal
plane to 6000 cgs as determined in cleavage experiments on the
rhombohedral planes by Wiederhorn (1969). Wiederhorn also reports
that cleavage is not possible on the basal planes, and in a pri-
vate communication has noted that the cleavage surface energy for
these planes must in consequence be in excess of 40,000 cgs! One
wonders if the value reported by Kingery (not measured in cleavage)
is not actually either that for a restructured surface or a com-
posite value for a dimpled surface whose average orientation is
[0001], but not the true basal plane. Consequently, we use γ =
6000 cgs in our work.

In Zn, Maitland and Chadwick (1969) have measured a range
of values from 100 cgs (confirmed by other previous authors), to
575 cgs, depending upon the technique used for making the measure-
ment. Because of weaknesses in the analysis as applied to the
experiments yielding the lower value, they believe the value of
575 cgs is to be preferred.

But experimental difficulties do not exhaust our problems
with the surface energy. One must also be certain that the surface

energy as we use it are synonymous. In our work, the correct γ is derived as the energy necessary to break bonds at a crack tip. Even in a pure cleavage event where no dislocations are produced, this process may not correspond to the thermodynamic surface energy because of relaxation effects which are possible on some crystal surfaces, such as Si, where the surface is entirely restructured. For this reason, the value chosen for Si is derived in a cleavage experiment, and for diamond is calculated from the value of the carbon-carbon bond. Unfortunately, no cleavage surface energies are available for Ge. We also note a further point or rigor. In very hard crystals where lattice trapping is important, the γ measured to grow a crack is different from the γ measured to heal a crack, Hsieh and Thomson (1973).

b) <u>Two Dimensional Results</u>. The results in Table III suggest that the face-centered crystals, with the possible exception of Ni which has a borderline ξ_c, are unstable to dislocation formation, since the critical distance, ξ_c, is less than the core radius, ξ_0. Since all the elastic forces become impossible to define inside the core radius, and since the repulsive forces are dominant for all distances larger than the core, we believe a crack in these crystals cannot sustain the large shear forces at its tip without forming the dislocation spontaneously. Na, of the body-centered crystals, probably also spontaneously emits the dislocation because of the small size of ξ_c, and also because we believe that in this case ξ_0 is not so small as 2/3, as is assumed for the other body-centered cubics. In all other cases, ξ_c is sensibly

-244-

larger than the core size, and we believe this gives rise to an energy hump the dislocation must negotiate as it is formed.

We have listed in Table II two other cruder estimates of ξ_c. In the first, we neglect the ledge term and in the second average over the geometry. Both are pretty fair approximations to the more accurate values for the brittle materials in the lower half of the table. The ledge term is seen to be important for the softer crystals, as one would expect.

The dimensionless ratio, $\mu b/\gamma$, related to our ξ_c'' has already entered into Armstrong's (1966) discussion of the brittle vs. ductile competition, and we see that it is indeed explicit in our own work. However, in our theory, as one progresses down the list of crystals in Table II, it is the opposite tendency of the growing value of $\mu b/\gamma$ and the decreasing value of the core cut-off which makes the difference between the ductile face-centered cubics and the more brittle body-centered cubics.

We note in Table II the positions of iron and nickel between the obviously ductile and obviously brittle solids. We give two calculations for iron. In one, the crack line runs along a cube direction, [100], and in the second, the crack line runs along (110]. The second is actually that observed by Tetelman and Robertson (1963) in silicon iron, but it is the former which has the interesting low value of ξ_c.

We have listed the results for a split dislocation for a noble metal and for nickel. (Aluminum has small or not splitting.)

-245-

The results shown are for only one of the partials. In the case shown, the partial has unfavorable geometrical factors, since it is not oriented to cause the crack to open effectively. In this unfavorable case, the critical radii are slightly increased, but not enough to change our general results.

c) <u>Three Dimensional Results</u>. The three dimensional results for the estimated activation energy as given in Table III confirm the two dimentional findings. We find a negative activation energy for the face-centered cubics, which again means that since all energies are cut off at the core radius, by the time any dislocation is well formed, it is under the primary influence of the replusive forces. Again, Na would seem to have an essentially spontaneous emission, not only because of the small value of the calculated activation energy, but also because we have probably underestimated the true dislocation core size of this very soft crystal.

The main surprise when comparing the two- with the three-dimensional results is the very large values obtained for the activation energy of the brittle crystals, even though ξ_c is usually a small number. One might suppose that it would take only a small energy to push a local segment of dislocation past ξ_c, say in NaCl, over a small front of the crack tip, and that the repulsive force would then be able to dominate the picture sufficiently to expand the fluctuation indefinitely. However, the reason this is not so is that the dislocation must meet the crack surface at normal incidence. Otherwise, the forces there will collapse the fluctuation no matter what happens to that portion lying over the hump

beyond ζ_c. In order for the dislocation to meet the surface per-
pendicularly, a full half loop of dislocation must be formed in
the region of repulsive forces, which is costly in dislocation
line energy.

d. _Conclusions._ We find a very strong tendency for crystals
to be either completely ductile, or completely brittle, so far as
dislocation emission is concerned. Iron and nickel are the only
interesting cases where our calculations suggest that the acti-
vation energy may be sufficiently low that thermal fluctuations
could play a role. With the various uncertainties in our calcu-
lations, we are unfortunately not able to pin down the values in
these cases sufficiently to make a definite prediction.

Except for iron, sodium, and the face-centered cubics, we
feel confident in predicting that no spontaneous or thermally
assisted blunting of a crack tip can occur in the other crystals
investigated. Even with the uncertainties in the γ and the other
approximations made, crystals like LiF and NaCl (to say nothing
of the hexagonal metals) seem immune enough to this process. This
conclusion means that if blunting does occur it must be because
of dislocations produced _outside_ the tip region, and which are
then attracted toward, and collide with it.

Our conclusions are at variance with the often quoted
claim that cracks cannot remain sharp (and the crystals thus be
brittle) if the shear stress at the crack tip is larger than the
theoretical shear stress of the homogeneous crystal. As we stated
in the introduction, a shear stress at the value of the theoretical

theoretical shear strength will cause the crystal to break down only if the stress is homogeneously applied across an entire plane running through the crystal. If the stress is only applied locally on that plane, then atoms where the stress has dropped below the maximum value will not be displaced from their lattice positions, and atoms where the force is maximum will not necessarily be displaced even though the stress is above the theoretical shear strength because of the resistance offered by atoms in the undeformed region of the plane.

Our conclusions suggest that so far as spontaneous emission is concerned, atomically sharp cleavage cracks are by no means a minority occurrence in materials. When the condition $\mu b/\gamma > 7.5-10$ is approximately satisfied, we have reason to expect the crack to remain sharp, though geometrical factors and variations in core size make this condition only approximate. This condition, in the cases investigated, is satisfied for all but the face-centered cubic metals, and certain borderline body-centered metals.

Our conclusions leave the precise morphology of cleavage cracks ultimately undetermined, however, because they do not address blunting reactions due to dislocations which might be attracted to the crack tip from the surrounding crystal. Processes which lower the cleavability by increasing the effective cleavage surface energy through the action of dislocation atmospheres trapped by the stress field of the crack are also not addressed here. On the other hand, our work does allow an evaluation of environmental effects which can be understood in terms of a

lowering of γ. By increasing the ratio $\mu b/\gamma$, these could presumably make possible atomically sharp cracking, even in solids such as fcc crystals which normally show ductile response. Then the 'cleavage' is not necessarily fast-running, but can proceed only as fast as reactions take place to bring about the requisite lowering of γ at the tip. It is possible that some environmentally assisted crack growth can be understood in this way.

Acknowledgement

TABLE I

Physical Parameters of Crystals

Crystal	Cleavage System	Slip System	γ cgs	$\mu \times 10^{11}$	ν	b Angstroms	$n=\sqrt{\dfrac{\mu b}{\gamma}}$
Face Centered Metals							
Pb	(001)	($\bar{1}\bar{1}1$)	440[1]	0.727	0.387	3.49	2.401
Au	[110]	[10$\bar{1}$]	1418[2]	2.37	0.412	2.87	2.190
Cu			1688[2]	4.05	0.324	2.55	2.473
Ag			1135[2]	2.56	0.354	2.88	2.549
Al			840[1]	2.51	0.347	2.85	2.918
Ni			1725[3]	7.48	0.276	2.49	3.286
Body Centered Metals							
Na	(001)	(0$\bar{1}$1)	237[1]	0.243	0.201	3.66	1.937
Fe	[100]	[111]	1975[4]	6.92	0.291	2.49	2.954
W			1700[5]	16.0	0.278	2.74	5.078
Fe*	(001)/[1$\bar{1}$0]	(110)/[$\bar{1}$11]	1975[4]	6.92	0.291	2.49	2.954
Ionic							
LiF	(001)	(0$\bar{1}$1)	480[6]	4.40	0.187	2.848	5.109
NaCl	[100]	[011]	250[7]	1.63	0.248	3.988	5.099
			(115)[8]	"	"	"	(7.518)
MgO			1200[9]	11.57	0.173	2.974	5.355
Al$_2$O$_3$	($\bar{1}$012)/[1$\bar{2}$10]	(0001)/[$\bar{1}\bar{1}$20]	6000[10]	23.3	0.200	4.75	4.295
Diamond Cubic							
Si	(111)	($\bar{1}$11)	1200[4]	6.05	0.215	3.83	4.394
Ge	[011]	[$\bar{1}\bar{1}$0]	600[1]	4.92	0.200	4.00	5.727
C			5400[12]	50.9	0.068	2.52	4.874
Hexagonal Metals							
Be	(0001)	(ijk0)	1620[1]	15.5	0.333	3.58	5.849
Zn	[0001]	[0001]	575[9]	3.83	0.333	4.94	5.736

Surface Energy References

1. J. W. Taylor, Metallurgia 50, 161 (1954).

2. Structure and Properties of Solid Surfaces (1952) Ed.
   R. Gomer and C. Smith, U. of Chicago Press, Chicago, Ill.
   Grain Boundaries in Metals (1957), D. McLean, Oxford U.
   Press, London, p. 76.

3. M. Inman and H. Tipler, Met. Reviews 8, 105 (1963)

4. A. Price, M. Hall and A. Greenough, Acta Met 12, 49 (1964).

5. J. Cordwell and D. Hull, Phil. Mag. 19, 951 (1969).

6. S. Burns and W. Webb, J. Appl. Phys. 41, 2078, 2086 (1970).

7. M. Tosi, Solid State Physics, Vol. 16, Ed. F. Seitz and
   Turnbull, Academic Press, New York (1964). See also,
   N. MacMillan and A. Kelly, Proc. Roy. Soc. 330, 291 (1972).

8. W. Class, Ph.D. thesis, Columbia University (1964).

9. A. Maitland, G. Chadwick, Phil. Mag. 19, 645 (1969).

10. S. Wiederhorn, J. Amer. Cer. Soc. 52, 485 (1969).

11. W. Kingery, J. Amer. Cer. Soc. 37, 42 (1954).

12. W. Hawkins, J. Chem. Phys. 10, 269 (1941).

TABLE II

Two Dimensional Results

Crystal	Core Cut-Off ξ_0	Critical Distance ξ_c	Approximate Values	
			ξ_c' w/o ledge	$\xi_c'' = \mu b/10\gamma$
Pb	2	1.1	0.88	0.58
Au	2	0.85	0.65	0.48
Cu	2	1.00	0.77	0.61
Ag	2	1.09	0.85	0.65
Al	2	1.4	1.1	0.85
Ni	2	1.7	1.3	1.08
Na	2/3	1.2	0.54	0.375
Fe	2/3	1.9	1.3	0.87
W	2/3	4.0	3.9	2.6
Fe*	2/3	2.7	1.9	0.87
LiF	0.25	3.2	2.9	2.6
NaCl	{ 0.25 / (0.25) }	{ 3.4 / (7.0) }	{ 3.2 / (7.0) }	{ 2.6 / (5.6) }
MgO	0.25	3.4	3.2	2.9
Al_2O_3	0.25	2.3	2.1	1.8
Si	0.25	2.2	2.0	1.9
Ge	0.25	3.7	3.3	3.3
C	0.25	2.4	2.2	2.4
Be	2/3	4.5	4.1	3.4
Zn	2/3	4.3	3.9	3.3

Partial dislocations $b = \dfrac{1}{\sqrt{12}} [2\bar{1}\bar{1}]$, $\psi = 60°$

Cu	2	1.4	0.91	0.35
Ni	2	2.3	1.6	0.62

TABLE III

Three Dimensional Results

Crystal	Activation Energy (Electron Volts)	Radius of Activated Loop (in units of b)
Pb		
Au		
Cu		
Ag	Spontaneous Emission - No activated state for $r > \xi_0$.	
Al		
Ni		
Na	0.02	1.5
Fe	2.2	5.1
W	329	50.7
Fe*	19	17
LiF	58	32
NaCl	$\left\{ \begin{array}{c} 62 \\ (240) \end{array} \right\}$	$\left\{ \begin{array}{c} 33 \\ (95) \end{array} \right\}$
MgO	205	37
Al_2O_3	852	20
Si	111	20
Ge	260	42
C	351	27
Be	180	23
Zn	107	21.2

References

R. Armstrong, (1966) Mat. Sci. & Eng. $\underline{1}$, 251

C. Atkinson, (1966) Int. J. Fract. Mech. $\underline{2}$, 567

G. C. Benson and G. W. Benson, (1955) Can. J. Chem. $\underline{33}$, 232

G. Benson, H. Schreiber and F. Van Zeggeren, (1956) Can. J. Chem. $\underline{34}$, 1553

G. Benson and P. Balk, (1959) J. Phys. Chem. $\underline{63}$, 1009

S. Burns and W. Webb, (1970) J. Appl. Phys. $\underline{41}$, 2078, 2086

S. Burns, (1970) Acta Met. $\underline{18}$, 969

W. Class, (1964) Ph.D. thesis, Columbia University

J. Cordwell and D. Hull, (1969) Phil. Mag. $\underline{19}$, 951

J. Gilman, (1960) J. Appl. Phys. $\underline{31}$, 2208

J. Hirth and J. Lothe, (1968) Theory of Dislocations, McGraw Hill, New York

C. Hsieh and R. Thomson, J. Appl. Phys. to be published

A. Kelly, (1966) Strong Solids, Oxford University Press, London

A. Kelly, W. Tyson and A. Cottrell, (1967) Phil. Mag. $\underline{15}$, 567

W. Kingery, (1954) J. Amer. Cer. Soc. $\underline{37}$, 42

K. Kitajima, (1966) Proc. 1st Int. Conf. Fracture, Vol. 2, Sendai

N. MacMillan and A. Kelly, (1972) Proc. Roy. Soc. $\underline{330}$, 291

A. Maitland and G. Chadwick, (1969) Phil. Mag. $\underline{19}$, 645

J. Rice, (1968) Fracture, ed. H. Liebowitz, Vol. 2, Academic Press, New York, pp. 191-311

J. Rice, (1972) Some Remarks on Elastic Crack Tip Stress Fields, International Journal of Solids and Structures, Vol. 8, pp. 751-758

A. Tetelman and W. Robertson, (1963) Acta Met. $\underline{11}$, 415

M. Tosi, (1964), Solid State Physics, Ed. F. Seitz and D. Turnbull, Vol. 16, Academic Press, New York

S. Wiederhorn, (1969) J. Amer. Cer. Soc. $\underline{52}$, 485

S. Wiederhorn, R. Voses and B. Bean, (1970) J. Amer. Cer. Soc. $\underline{53}$, 18

APPENDIX

Image Force on Dislocation Near a Crack Tip

To compute the image force, consider Fig. 8 in which a straight dislocation line lies parallel to a crack front, with the slip plane intersecting the fracture plane at distance a ahead of the tip. We are ultimately interested in the case $a=0$, but by considering a as variable we can use energy methods to compute the force.

Suppose that the body containing the crack is subjected to three generalized boundary forces Q_I, Q_{II}, and Q_{III}, each of which if exerted singly upon a dislocation free body would cause only a Mode I, II, or III crack tip singularity, respectively. Let q_I, q_{II}, and q_{III} be the associated generalized displacements. These are defined so that

$$\sum_J Q_J \; dq_J$$

is the work of boundary loadings per unit thickness into the plane of the figure. The sum on J extends from I to III. Hence, if we let U be the strain energy of the body per the same unit thickness,

$$dU = \sum_J Q_J \; dq_J + G \; da - f \; d\rho \quad , \tag{A1}$$

where the respective contributions to energy changes come from load-point displacements dq_J, crack advance $-da$, and dislocation glide $d\rho$; G is the crack extension force (or energy release rate) and f is the force on the dislocation.

The last equation may be rewritten as

$$d(U - \sum_J Q_J q_J) = -\sum_J q_J dQ_J + G\, da - f\, d\rho \quad .$$

Further, since the terms on the right constitute an exact differential, we may write the Maxwell relation

$$\left(\frac{\partial G}{\partial \rho}\right)_{\ell,Q} = -\left(\frac{\partial f}{\partial a}\right)_{\rho,Q} \tag{A2}$$

This is a fundamental relation for it allows us to compute the force on the dislocation solely from a knowledge of elastic fields induced by each of the forces Q_J. To see this, recall that the crack energy release rate is given by (e.g., Rice, 1968)

$$G = \frac{1-\nu^2}{E}(K_I^2 + K_{II}^2) + \frac{1+\nu}{E}K_{III}^2 \quad , \tag{A3}$$

where the K's are stress intensity factors for the three modes. These are due both to the boundary forces Q_J and to the interaction of the dislocation stress field with the crack, and take the forms

$$K_J = k_J(a)Q_J + L_J(a,\rho) \quad , \qquad J = I, II, III \tag{A4}$$

Here each k_J is the Mode J stress intensity factor induced per unit of the corresponding boundary force. Each is assumed to be known for all crack lengths and hence can be considered as given function of a. The unknown functions L_J are the stress intensity factors induced by the dislocation. The force on the dislocation is given by

$$f = b_e[t_I(a,\rho)Q_I + t_{II}(a,\rho)Q_{II}] + b_s\, t_{III}(a,\rho)Q_{III} + g(a,\rho) \tag{A5}$$

Here b_e is the edge and b_s the screw Burgers vector component; t_I and t_{II} are the shear stresses induced on the glide plane, at the dislocation position, in the edge direction per unit boundary force Q_I and Q_{II}, respectively; t_{III} is the shear stress in the screw direction induced per unit force Q_{III}. These shear stresses are to be considered as known functions of crack length and dislocation position, and hence of a and ρ. The unknown force term g is that due to the dislocation itself, i.e., the image force, and our object is to compute it.

By substituting for G and f as above in the Maxwell relation (A2), we obtain

$$\frac{2(1-\nu^2)}{E}[(k_I Q_I + L_I)\frac{\partial L_I}{\partial \rho} + (k_{II}Q_{II} + L_{II})\frac{\partial L_{II}}{\partial \rho} + \frac{1}{1-\nu}(k_{III}Q_{III} + L_{III})\frac{\partial L_{III}}{\partial \rho}]$$

$$= -b_e \frac{\partial t_J}{\partial a} Q_I - b_e \frac{\partial t_{II}}{\partial a} Q_{II} - b_s \frac{\partial t_{III}}{\partial a} Q_{III} - \frac{\partial g}{\partial a}$$

This must hold for all values of the Q's. Hence by equating coefficients of each Q_J,

$$\frac{\partial L_J}{\partial \rho} = \frac{-Eb_e}{2(1-\nu^2)k_J}\frac{\partial t_J}{\partial a}, \text{ for } J=I,II; \quad \frac{\partial L_{III}}{\partial \rho} = \frac{-Eb_s}{2(1+\nu)k_{III}}\frac{\partial t_{III}}{\partial a}.$$

$$\tag{A6}$$

And by equating sides when each $Q_J = 0$,

$$\frac{\partial g}{\partial a} = -\frac{1-\nu^2}{E}\frac{\partial}{\partial \rho}(L_I^2 + L_{II}^2 + \frac{1}{1-\nu}L_{III}^2) \tag{A7}$$

The calculation of the image force now involves simple integrations: (i) First, we note that the right sides in (A6) involve known functions of a and ρ; each L_J is determined by integrating these at fixed a from the ρ of interest to a large value of ρ, say to ∞ or to the specimen boundary, at which the L's may be taken as zero. (ii) Thus the right side of (A7) is determined, and g is obtained by integration from the a value of interest (zero in our case) to a large value of a at which g may be taken as zero.

Now, since we are interested in a dislocation rather near to the crack tip, it will suffice to consider a semi-infinite crack in an infinite body. Also, the solution for the L's and g cannot depend on the particular nature of the load systems denoted by the Q's. Indeed, the right sides of (A6) are universal functions, the same for all load systems inducing a given crack tip mode (see Rice, 1972). Hence, for simplicity, we can choose load systems which act so far from the crack tip that the characteristic inverse-square-root elastic stress distribution gives the entire stress field, and that the k's are virtually constant, for all crack distances a of interest. Let $(\sigma_{ij})_J$ denote the stress field for a given loading mode, J; this will be of the form

$$(\sigma_{ij})_J = \frac{k_J Q_J}{\sqrt{r}} \, [F_{ij}(\theta)]_J \quad , \qquad i,j = x,y,z \quad . \tag{A8}$$

Here the F's are functions of θ, particular to each mode, that are tabulated e.g., by Rice, 1968, pp. 216-217. From these the

shear stresses on the dislocation, per unit boundary loads, are identified from

$$Q_J t_J = \frac{(\sigma_{yy})_J - (\sigma_{xx})_J}{2} \sin 2\phi + (\sigma_{xy})_J \cos 2\phi \quad ,$$

for $J = I$, II, and

$$Q_{III} t_{III} = (\sigma_{yz})_{III} \cos\phi - (\sigma_{xz})_{III} \sin\phi \quad .$$

Upon substitution from (A8), these give

$$t_I = k_I (8\pi)^{-\frac{1}{2}} \sin\phi \ \mathrm{Re}[e^{i\phi}(\xi^{-\frac{1}{2}} - a\xi^{-3/2})]$$

$$t_{II} = k_{II}(8\pi)^{-\frac{1}{2}} \mathrm{Re}[2e^{2i\phi}\xi^{-\frac{1}{2}}] + \sin\phi \ \mathrm{Im}[e^{i\phi}(\xi^{-\frac{1}{2}} - a\xi^{-3/2})] \quad \text{(A9)}$$

$$t_{III} = k_{III}(2\pi)^{-\frac{1}{2}} \mathrm{Re}[e^{i\phi}\xi^{-\frac{1}{2}}] \quad ,$$

where i is the unit imaginary number, Re and Im denote real and imaginary parts, and where

$$\xi = re^{i\theta} = a + \rho e^{i\phi}$$

Hence the first of (A6) becomes

$$\frac{\partial L_I}{\partial \rho} = - \frac{Eb_e}{2(1-\nu^2)k_I} \left\{ k_I(8\pi)^{-\frac{1}{2}} \sin\phi \ \frac{\partial}{\partial a} \mathrm{Re}[e^{i\phi}(\xi^{-\frac{1}{2}} - a\xi^{-3/2})] \right.$$

$$= \frac{3 \ Eb_e \sin \phi}{8(2\pi)^{\frac{1}{2}}(1-\nu^2)} \mathrm{Re}[e^{i\phi}(\xi^{-3/2} - a\xi^{-5/2})]$$

-260-

Multiplying by $d\rho$ and recognizing that $e^{i\phi} d\rho = d\xi$, so that integration on ρ becomes integration on ξ inside the brackets, we have

$$L_I = \frac{3\ Eb_e\ \sin\phi}{8(2\pi)^{\frac{1}{2}}(1-\nu^2)}\ \text{Re}\left[-2\xi^{-\frac{1}{2}} + \frac{2}{3}\ a\xi^{-3/2}\right]$$

(A10)

$$= \frac{-Eb_e}{4(1-\nu^2)(2\pi r)^{\frac{1}{2}}}\left[3\ \sin\phi\ \cos\frac{\theta}{2} - \sin(\phi-\theta)\ \cos\frac{3\theta}{2}\right]\ ,$$

where the constant of integration is chosen so that L_I vanishes at $\rho = \infty$. In a similar way we find

$$L_{II} = \frac{-Eb_e}{4(1-\nu^2)(2\pi r)^{\frac{1}{2}}}\left[2\ \cos\phi\ \cos\frac{\theta}{2} - \sin\phi\ \sin\frac{\theta}{2} + \sin(\phi-\theta)\ \sin\frac{3\theta}{2}\right]$$

$$L_{III} = \frac{-Eb_s}{2(1+\nu)(2\pi r)^{\frac{1}{2}}}\ \cos\frac{\theta}{2}$$

(A11)

Now that the L's are known, we have determined the effect of the dislocation on the stress intensity factors K_J (A4) and, further, we can substitute into (A7) to determine the image force. Since each of the L's is proportional to $r^{-\frac{1}{2}}$, (A7) is of the form

$$\frac{\partial g}{\partial a} = -\frac{\partial}{\partial \rho}\left[\frac{Q(\theta,\phi)}{r}\right]\qquad,$$

where

$$Q(\theta,\phi) = \frac{1-\nu^2}{E}\ r\left(L_I^2 + L_{II}^2 + \frac{1}{1-\nu}\ L_{III}^2\right)\qquad.$$

(A12)

But by using the relations

$$\frac{\partial}{\partial a} = \cos\theta \frac{\partial}{\partial r} - \frac{\sin\theta}{r} \frac{\partial}{\partial\theta} \quad ,$$

$$\frac{\partial}{\partial\rho} = \cos(\phi-\theta) \frac{\partial}{\partial r} - \frac{\sin(\phi-\theta)}{r} \frac{\partial}{\partial\theta}$$

between partial derivatives, one may readily derive the identity

$$- \frac{\partial}{\partial\rho} \left[\frac{Q(\theta,\phi)}{r} \right] = \frac{\partial}{\partial a} \left[\frac{\sin(\phi-\theta)}{\sin\theta} \frac{Q(\theta,\phi)}{r} \right]$$

This lets us integrate (A12) immediately. We must, however, append a 'constant' of integration, which may depend on ρ ($= r \sin\theta/\sin\phi$), and which assures that $g \to 0$ as $a \to \infty$. This means that $g \to 0$ as $r \to \infty$, $\theta \to 0$ in such a way that $r \sin\theta$ remains finite. Integration of (A12) subject to this condition gives the image force

$$g = \frac{\sin(\phi-\theta)Q(\theta,\phi) - \sin\phi \, Q(0,\phi)}{r \sin\theta} \quad , \tag{A13}$$

where Q is defined from (A10,11) via (A12).

In the text of the paper we were concerned with the case $a = 0$, so that $\theta = \phi$ and $r = \rho$. In that case

$$g = - \frac{Q(0,\phi)}{\rho} = - \frac{Eb_e^2}{8\pi(1-\nu^2)\rho} - \frac{Eb_s^2}{8\pi(1+\nu)\rho} \quad , \tag{A14}$$

where the latter form comes from identifying $Q(0,\phi)$, which is found to be independent of ϕ. Remarkably, this is the same image force as for a dislocation line at distance ρ along its slip plane from the boundary of a half-space. In fact, the result for this latter case is also obtainable from (A13) by letting the crack tip

pass far beyond the site of the dislocation, $a \to -\infty$. This means that we let $\theta \to \pi$, $r \to \infty$ while keeping $r \sin\theta$ finite (it equals $\rho \sin\phi$). By direct evaluation one finds that $Q(\pi, \phi) = 0$, so that the image force from the half-space boundary is also given by (A14).
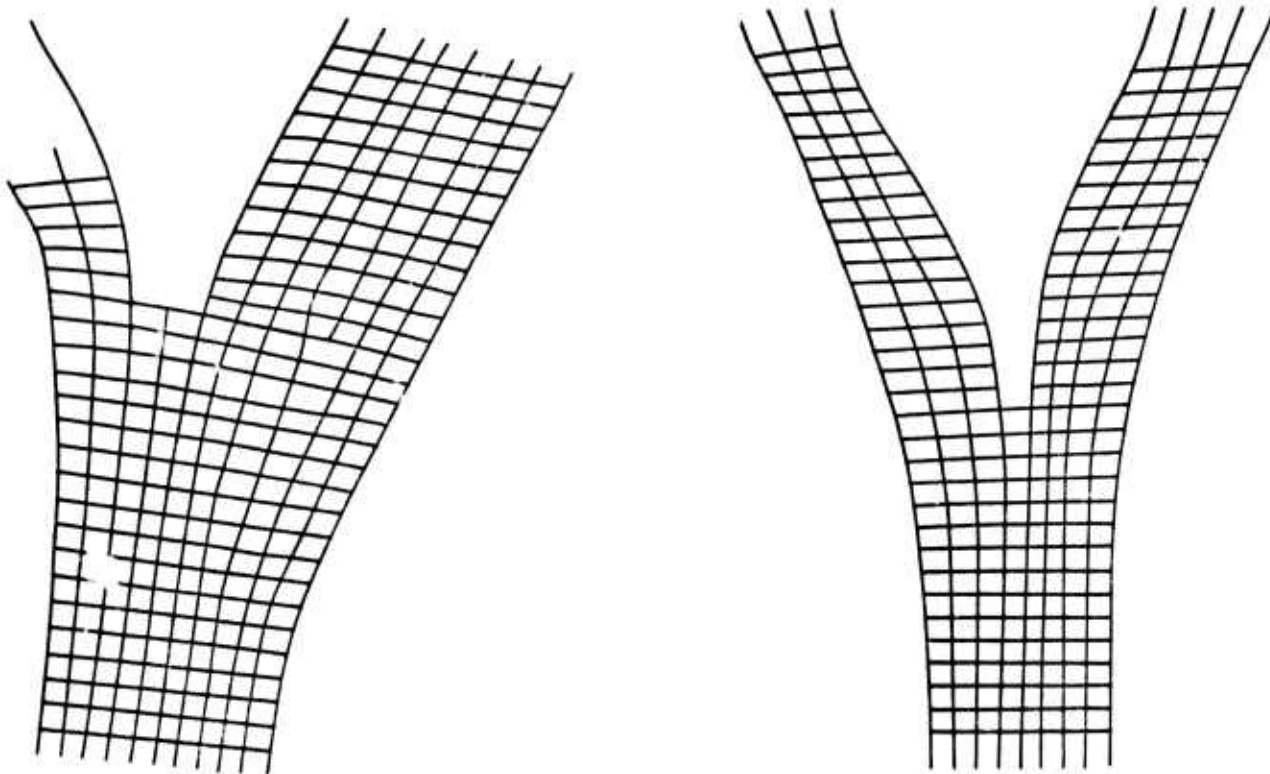
Figure 1.   An atomically sharp crack is blunted when a dis-
            location is emitted from the tip when the Burgers
            vector has a normal component to the fracture plane.
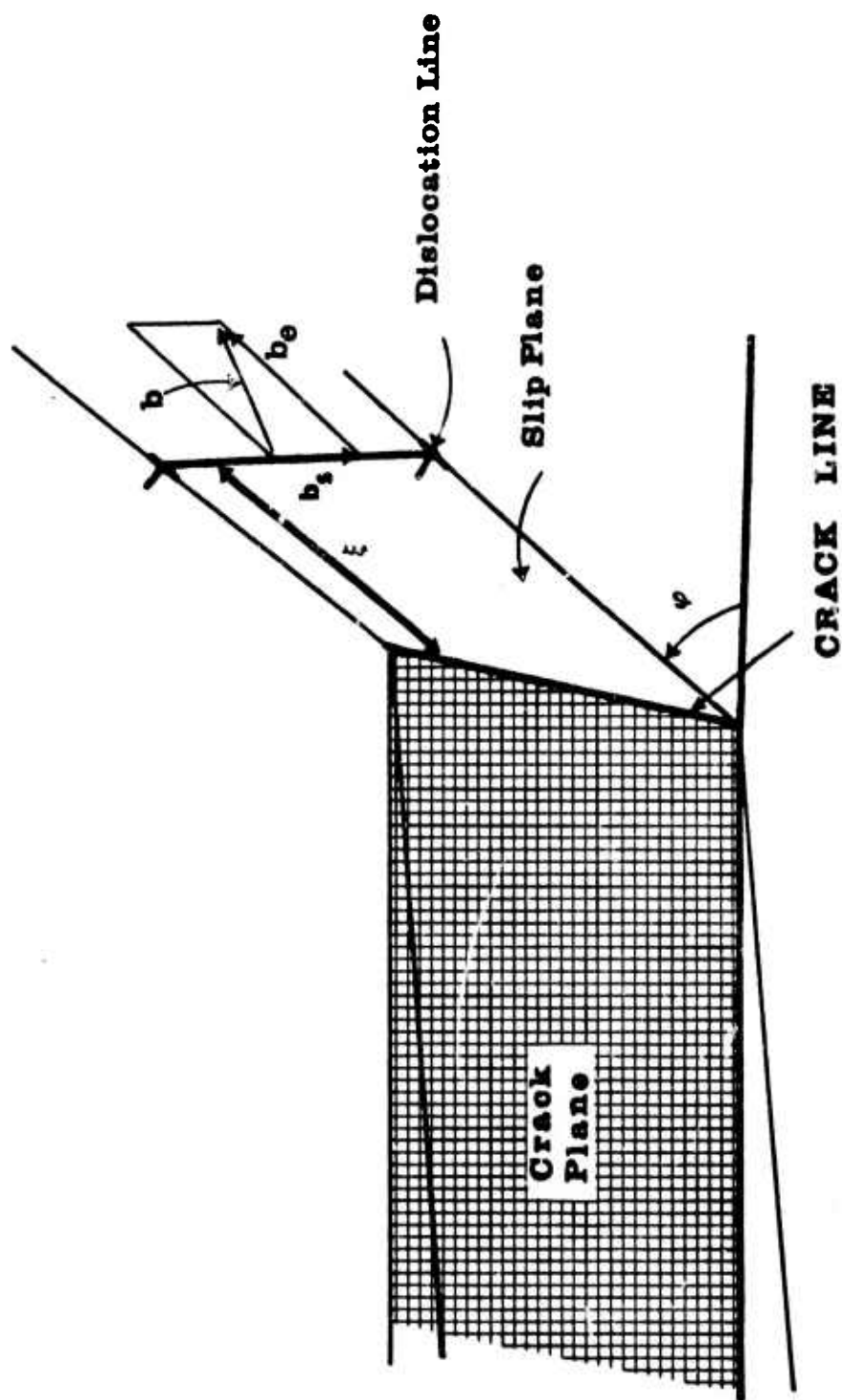
Figure 2. Geometry of the dislocation, crack configuration in two dimensions. b_e and b_s are perpendicular and parallel components of the Burgers vector relative to the crack.
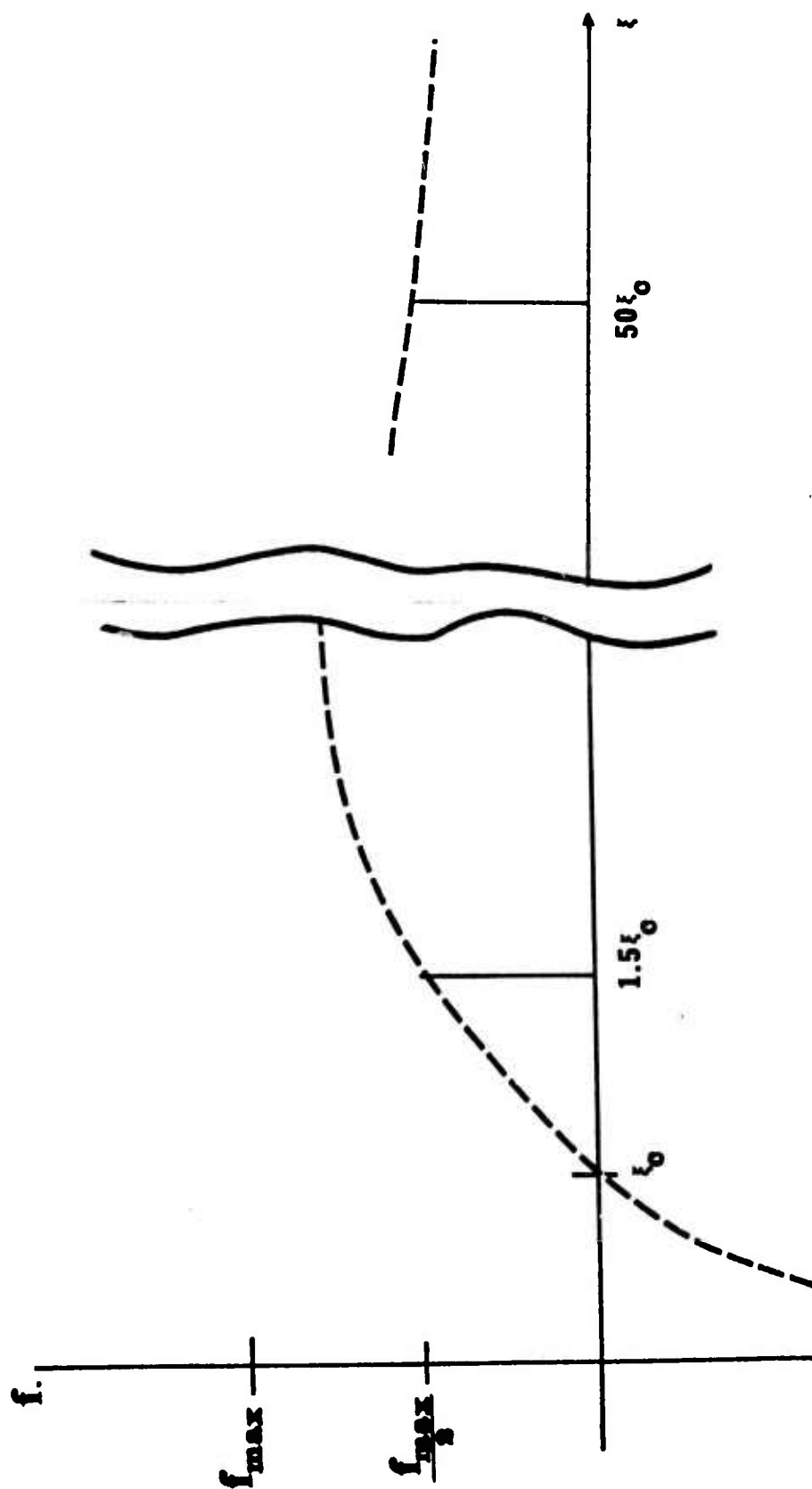
Figure 3. The function $f = -A/\xi + B/\sqrt{\xi}$ plotted to show its broad maximum.
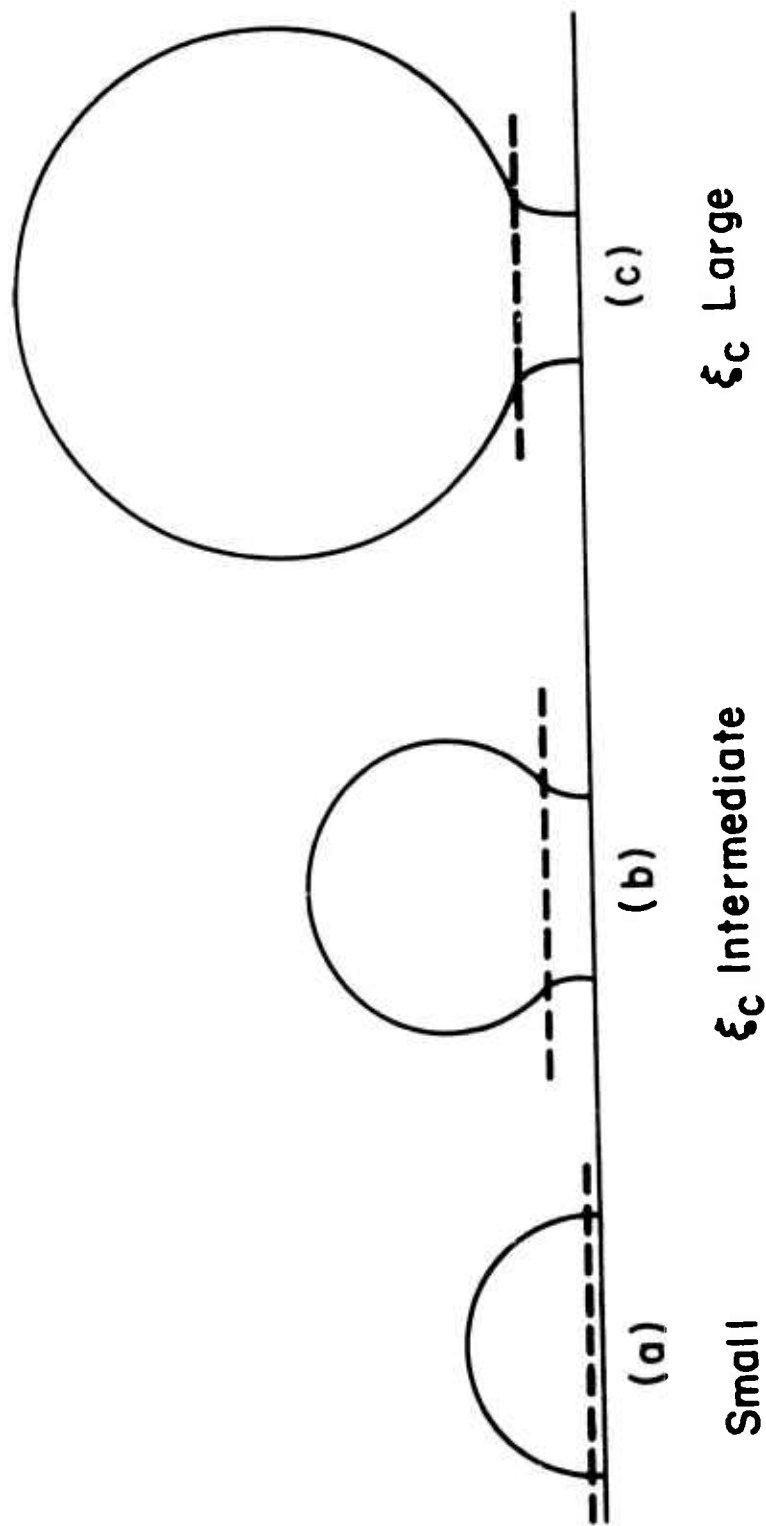
Figure 4. Schematic representations of saddle point dislocation configurations. At distances greater from the crack than ξ_c, the curvature is positive, while at less than ξ_c, the curvature is negative. The dislocation always has normal incidence at the surface because of the preponderance of the image term near the surface. a) ξ_c is small and the region of negative curvature is negligible. b) Intermediate ξ_c. c) Large ξ_c.

-267-
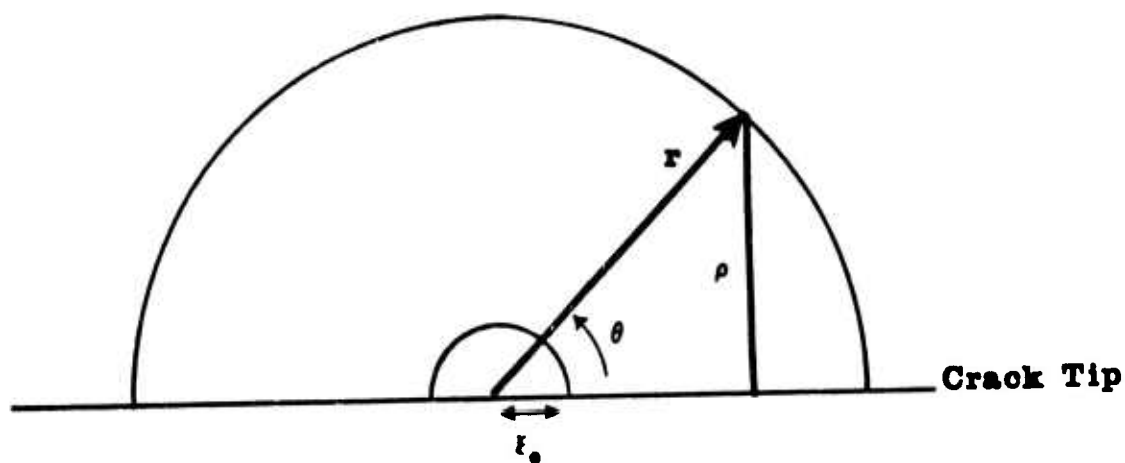
Figure 5. Confiruration for calculating the work done under the crack stress by the expanding dislocation. The work integration has the lower limit shown when $r = \xi_0$.
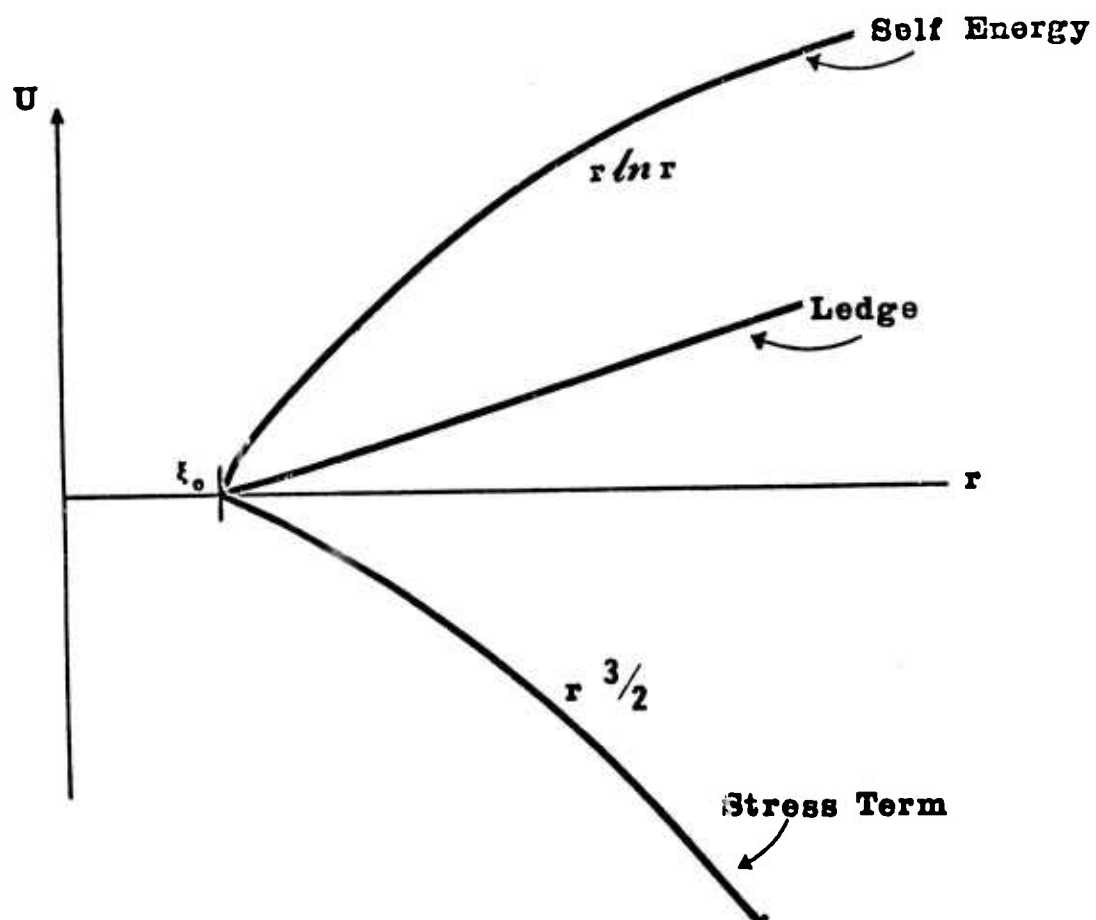
Figure 6. Schematic variation of the three terms in Eq. (17) as a function of the radius.
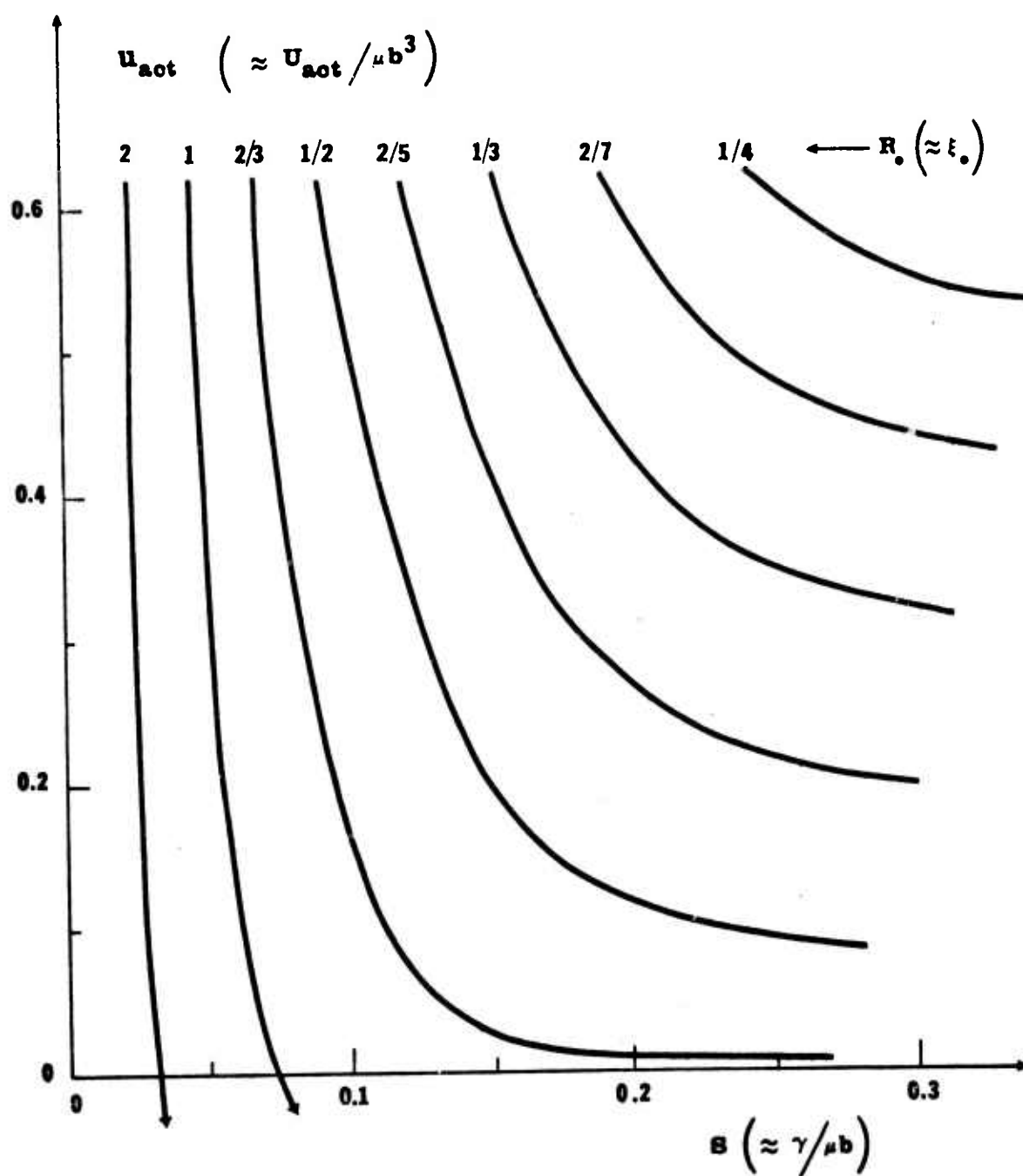
Figure 7. Dependence of activation energy for dislocation nucleation on $S(\approx\gamma/\mu b)$ and $R_0(\approx\xi_0)$; exact definitions of S and R_0, and normalization of energy, are given in text.
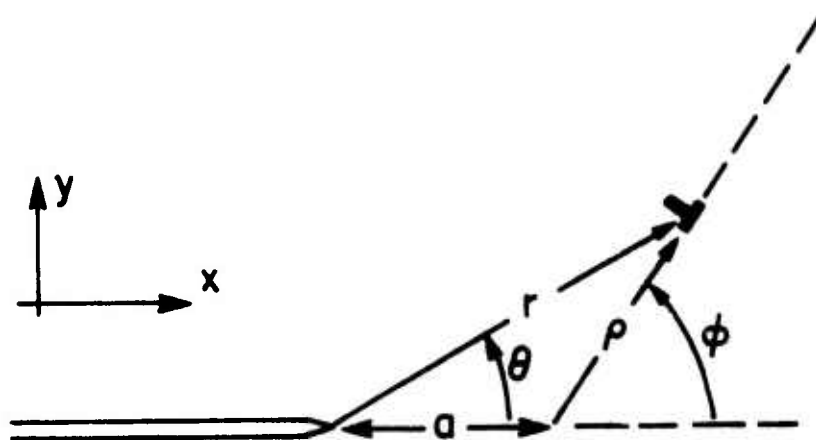
Figure 8. Configuration of crack and dislocation for computation of the image force.

STATISTICS OF BRITTLE FRACTURE

F. A. McClintock

Abstract

From a simple statistical model of occasional cracked grain boundaries, a statistical distribution of strength is derived which does not fall into any of the three asymptotic forms of extreme value distributions. The size effects for this new extreme-value distribution are similar to those of the third asymptote with an exponent of about $m = 4$, but it is necessary to drop 5 or 6 standard deviations below the median in order to reduce the failure probability to 1 in 10^6, which corresponds more to the first asymptote, $m = \infty$. Stress gradient effects, leading to notch insensitivity, are reviewed for the third asymptotic distribution, and a method for correlating scatter in strength with position of failure in three-point bend specimens is derived and illustrated.

Preceding page blank

STATISTICS OF BRITTLE FRACTURE

F. A. McClintock

1.   INTRODUCTION

In planning to extend the application of ceramics to
structures, in designing for applications where proof testing
is impractical or in considering the production costs of parts
rejected in proof testing, there is a need to use a small sample
to predict the behavior of large parts, or a large number of
parts, with as high a confidence as possible. Such predictions
require knowledge of the underlying distribution function. With-
out it, sampling would have to involve very large numbers to
obtain the desired very low probability of failure. Roughly
speaking, in order to assure a failure probability of the order
of 1 in N, tests must be run on the order of N specimens. Thus
for failure probabilities of the order of 10^{-4} to 10^{-6}, sample
sizes of 10^4 to 10^6 would be required. Even such large samples
would be misleading, since they would consist of a small sample
of batches or suppliers. With it so expensive to obtain a dis-
tribution function empirically, it is desirable to have the
theoretical basis for statistical distribution functions.

The normal distribution function is a good approximation
when the result depends more or less equally on a large number

of events. It is hardly appropriate for fracture, particularly for fracture initiation where the failure depends more on the most serious flaw. Even this is an oversimplification, however, for the very existence of acoustic emission, as well as petrographic evidence from compressed rocks, indicates that it is not the first crack growth, but rather cooperative crack growth over a number of elements, that leads to final fracture. Nonetheless, the extreme value distribution, in which fracture is postulated to occur in specimens under uniform stress reaches the strength of the weakest element, provides a limiting case to observe as a contrast with the normal or log normal distributions. Estimates can be made of the reliability of structures using both of these distribution functions and judgement applied to select a reasonable working stress. This judgement must of course be tempered by the fact that in many cases, failures are due to special accidents in production which in effect give a multi-modal distribution.

2. EXTREME VALUE DISTRIBUTIONS

Following Weibull's (1939) study of size effect, we shall start with the density of flaws per unit surface area that give strengths in the range S to S + dS. Later we shall develop this density from an idealized model of cracks; here we simply denote it by g(S) dS. The probability of no fracture taking place in an infinitesimal area δA below a stress S_1 is

$$1 - \delta\phi = 1 - \delta A \int_0^{S_1} g(S) \, dS \quad . \tag{1}$$

In a total area A there will be $A/\delta A$ infinitesimal areas δA. The probability, $1 - \phi$, that none of them will have strengths below S_1 is the product of their individual probabilities of no fracture, assuming that the strengths of the areas are statistically independent of each other. (This is the most questionable part of the analysis, since defects in production are not random, and cracks at the edge of one infinitesimal area would affect the neighbors.)

$$1 - \phi = (1 - \delta\phi)^{A/\delta A} = \left[1 - \delta A \int_0^{S_1} g(S) \, dS \right]^{A/\delta A} \tag{2}$$

Replacing the term of the form $(1-\epsilon)^n$ by $\exp n \ln(1-\epsilon) \simeq \exp(-n\epsilon)$ gives a limiting form for the probability of fracture below the stress S_1:

$$\phi = 1 - \exp\left(-A \int_0^{S_1} g(S) \, dS \right) \quad . \tag{3}$$

Before developing Eq. 3 further, note its structure. As $S_1 \to 0$, the argument of the exponent goes to zero; the exponent goes to unity; and the probability of failure ϕ goes to zero, as it should. At the upper limit as the stress increases, if the integral increases indefinitely the exponent goes to zero and the probability of failure goes to unity. Likewise, for any given integral, the greater the area, the smaller the exponent

and the greater the probability Φ of prior fracture.

Statistical extreme value theory was developed by Frechét in 1927 and Fisher and Tippett in 1928, and proposed empirically by Weibull (1939). It was presented in detail by Gumbel (1958) and was summarized by Kendall and Stuart (1963). The asymptotic extreme value distributions arise from functions $g(S)$ that give a constant form for the function Φ as the size of the sample is increased. In this context, "constant form" means that the distributions are of similar shape, differing only in origin and scale factor.

For the distributions of the least values S, the functions $g(S)$ are of just three possible kinds:

1) $\quad g(S) = ae^{\alpha S}$

2) $\quad g(S) = a(S_u - S)^{-k}$ for $S < S_u$ $\qquad\qquad$ (4)

3) $\quad g(S) = a(S - S_\ell)^k$ for $S > S_\ell$.

Statistical studies of extreme values have focused on these distributions. There is some reason for this, since the third distribution can be thought of as the limiting case in which there is a lower limit to the strength S, and since in large samples we are concerned only with the weakest elements, we need consider only the leading term in a power series expansion of the function $g(S)$ about that lowest limit. The distribution of the first kind can be thought of as the limiting

-277-

case of the third, with the lower limit to the strength going to $-\infty$ and the exponent increasing at such a rate as to keep the resulting distribution finite. The distribution of the second kind is the leading term in a distribution with a finite upper limit to the distribution. Since the asymptotic forms have been relatively well developed, in the last part of the paper their applications to brittle materials will be reviewed. These limiting distributions are by no means universal, however; as we shall see in the next section, if the distribution of the crack lengths follows an extreme value distribution of the first kind, the resulting distribution of strengths does not follow any of these asymptotic distributions. Therefore, the asymptotic distributions should be considered for the convenient insight they give, but not as being definitive.

3.   A CRACK MODEL FOR THE DISTRIBUTION OF STRENGTH

It is helpful to have a simple, if inexact, model of the strength of a brittle material to suggest an appropriate form for the distribution of strength per unit area, $g(S)$. For simplicity, assume that there are a number of independent units such as grain boundary segments, and that for each there is a finite probability p of being cracked. As in Fig. 1 assume that the material consists cf a linear string of units. Call an element (a crack) any set of continuous cracked units, including the following uncracked unit. This assumption implies neglecting the effects of crack branching or meandering.
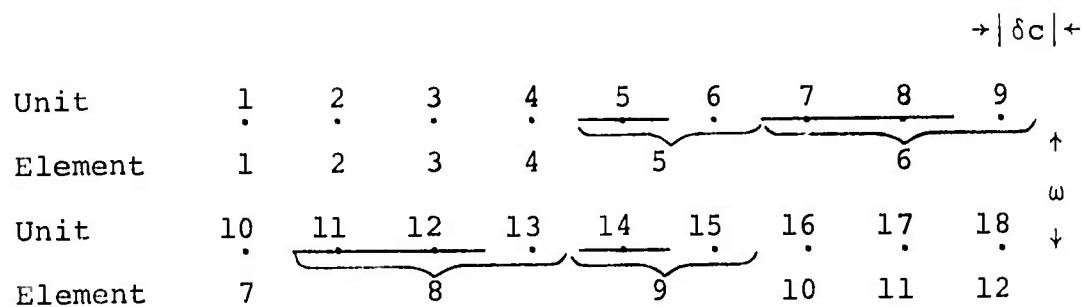
Figure 1. String of elements containing n cracked units each $(n \geq 0)$.

The probability that an element is uncracked is that its first (and only) unit is uncracked:

$$f(0) = 1-p \qquad .$$

(5)

For the element to contain a crack just one grain boundary segment long, the first unit must be cracked and the second uncracked, for which the probability is

$$f(1) = p(1-p)$$

(6)

Similarly, for an element to consist of a crack that is n grain boundary segments long, the probability is

$$f(n) = p^n(1-p) \qquad .$$

(7)

The frequency distribution function given by Eq. 7 is discrete; it can be approximated by a continuous distribution by regarding each discrete value of n as a variable running from $n-\frac{1}{2}$ to $n+\frac{1}{2}$. In Eq. 3 we need the integral of the distribution of strengths

-279-

per unit area <u>up to</u> a given value. With fracture mechanics we can get this from the integral of the distribution of crack lengths per unit area <u>above</u> a given value. Therefore consider N elements, and divide the number of cracks longer than $c/\delta c$ by the total area. If the strings of Fig. 1 are spaced w apart, the area for each unit is $w\delta c$. Then the cumulative frequency distribution of long cracks is

$$\int_{c/\delta c}^{\infty} h(n)\,dn = \frac{N \displaystyle\int_{c/\delta c}^{\infty} p^n(1-p)\,dn}{N \displaystyle\int_{-\frac{1}{2}}^{\infty} p^n(1-p)(n+1)w\delta c\,dn} \quad . \tag{8}$$

The integrals are evaluated with the substitution $p^n = \exp(n\ \ell n\ p)$ and a table of integrals:

$$\int_{c/\delta c}^{\infty} h(n)\,dn = \frac{[\exp(n\ \ell n\ p)]_{c/\delta c-\frac{1}{2}}/\ell n\ p}{w\delta c[\exp(n\ \ell n\ p)((n/\ell n\ p)-(1/\ell n\ p)+(1/\ell np))]_{-\frac{1}{2}}^{\infty}} \tag{9}$$

which, since $\ell n\ p < 0$, becomes

$$\int_{c/\delta c}^{\infty} h(n)\,dn = \frac{\exp((c/\delta c)\ell n\ p)}{w\delta c\ \exp((-\frac{1}{2})\ell n\ p)(\frac{1}{2}-1/\ell n\ p)} \quad , \tag{10}$$

which is of the form

$$\int_{c/\delta c}^{\infty} h(n)\,dn = a\ \exp(-bc) \quad .$$

Note that differentiating this cumulative distribution function shows it to be an asymptotic distribution function of the first kind for the length of the largest crack.

The corresponding strengths can be estimated from linear elastic fracture mechanics, with the assumption that the cracks are far enough apart so that they do not interact. This assumption is not correct, but judging from the results of Ishida (1969), the separations need only be a crack half-length for the interactions to be reduced to the order of 10%. It is further necessary to assume that the cracks are sharp. Sprunt and Brace (1973) have shown that many natural cracks in rocks have a finite tip radius, as observed with the scanning electron microscope, but the radius is small enough for the sharp crack to be a useful approximation. They also find that there are often small bridging elements in the center of a crack; the analysis considered here applies only after the bridging elements have been broken. For sharp cracks the applied stress S necessary to propagate a crack of total length c, according to linear elastic fracture mechanics, is given in terms of a critical stress intensity factor k_{1c} by

$$S = k_{1c}/\sqrt{c/2} \quad \text{or} \quad c = 2k_{1c}^2/S^2 \quad . \tag{11}$$

Substitution into Eq. 10 gives the cumulative distribution function of strengths per unit area, and differentiation gives the frequency distribution per unit area:

$$\int_0^{S_1} g(S)\,dS = \frac{\sqrt{p}}{w\delta c(.5-1/\ln p)} \exp \frac{2k_{1c}^2 \ln p}{\delta c S_1^2} \tag{12}$$

-281-

$$g(S) = - \frac{4\sqrt{p}\ k_{1C}^2\ \ell n\ p}{w\delta c^2 S^3 (.5-1/\ell n\ p)}\ \exp \frac{2k_{1C}^2\ \ell n\ p}{\delta c\ S^2} \quad . \tag{13}$$

Note that Eq. 13 does not correspond to any one of the forms of Eq. 4 giving asymptotic extreme value distributions.

Equation 13 can be simplified by defining the strength of a specimen containing a single isolated grain boundary crack, that is a unit crack, by

$$S_u = k_{1C}/\sqrt{\delta c/2} \quad . \tag{14}$$

With this definition Eq. 12 can be rewritten as

$$\int_0^{S_1} g(S)\,dS = \frac{1}{w\delta c}\ \frac{\sqrt{p}}{(.5-1/\ell n\ p)}\ \exp \frac{\ell n\ p}{(S_1/S_u)^2}$$

$$= \frac{a}{w\delta c}\ \exp \frac{-b}{(S_1/S_u)^2} \tag{15}$$

where a and b are not those of Eq. 10, but

$$a = \sqrt{p}/(.5-1/\ell n\ p) \quad \text{and} \quad b = -\ell n\ p \quad . \tag{16}$$

The cumulative extreme value distribution corresponding to Eq. 3, for strengths below S in specimens of area A, is of the form

$$\phi(S) = 1-\exp(- \frac{Aa}{w\delta c}\ \exp(-b/(S/S_u)^2)) \quad . \tag{17}$$

The corresponding frequency distribution giving strengths between S and S + dS is

$$\phi(S) = \frac{Aa}{w\delta c} \frac{2b}{S_1^3} \exp(-b/(S/S_u)^2) \exp(-\frac{Aa}{w\delta c} \exp(-b/(S/S_u)^2)) \quad . \quad (18)$$

Since the integrations to provide the mean and standard deviation of strength look awkward, find the n-tile strengths at ϕ:

$$-(w\delta c/Aa)\ln(1-\phi) = \exp(-b/(S/S_u)^2) \quad ,$$

$$\frac{S(\phi)}{S_u} = \sqrt{\frac{b}{\ln(\frac{Aa}{w \ c}/(-\ln(1-\phi)))}} \quad . \quad (19)$$

Note that Eq. 19 can only be valid for areas large enough or ϕ levels small enough so that the crack is longer than a unit crack:

$$S(\phi) < S_u \quad \text{or} \quad b < \ln(\frac{Aa}{w\delta c}/-\ln(1-\phi)) \quad \text{or} \quad \phi < 1-\exp(-\frac{Aae}{w\delta c})^b \quad . \quad (20)$$

For small ϕ, $\phi < (A/w\delta c)p^{3/2}/(.5-1/\ln p)$.

The coefficients a and b in Eqs. 15-20 can be evaluated in a number of ways. Physically, they can be evaluated from a knowledge of the defect distribution of the specimen through a comparison of Eqs. 3, 13 and 14. Starting from experimental data, they can be found from values of the mode and the quartiles of the experimental distribution. Alternatively, they could be found by plotting the order statistics on appropriately scaled graph paper.

Physically, the model developed here is only valid when the probability of fracture of any given unit, p, is small compared with unity, so that the cracks are widely spaced. For

example, take p = 0.1, .001, and .0001. Lower values of p would seem to require unrealistic degrees of perfection. As regards the area ratio, consider a grain size of the order of .3mm. The lateral distance, w, between cracks should be roughly proportional to $\sqrt{1/p}$, here 10 to 100 δc, giving an area subtended by a unit crack to be of the order of 1 to 10mm^2. The area of a part, on the other hand, is likely to be of the order of 1 to 100 cm^2, giving $A/w\delta c = 10$ to 10^4. Valid Φ's from Equation 20 are $\Phi < 1$ to 10^{-5}. The probabilities of failure Φ which are measured experimentally are covered by the range .01 to .99. On the other hand, limits of reliability corresponding to failure probabilities of as low as 10^{-6} are of interest in certain components of central station power plants, of aircraft, or of hazardous consumer products, so we shall consider a range of variables extending over this set of values.

Experimentally, the values of a and b can be found from the mean and interquartile values of a sample, assuming that the sample is representative of the population. Define the ratio of the third to the first quartile strength by R. Then from Eq. 19,

$$\frac{S(.75)}{S(.25)} \equiv R = \sqrt{\frac{\ln\left((Aa/w\delta c)/(-\ln(.75))\right)}{\ln\left((Aa/w\delta c)/(-\ln(.25))\right)}}$$

or

$$\left(\frac{Aa}{w\delta c}\right)^{R^2}/(\ln 4)^{R^2} = \frac{Aa}{w\delta c}\ /\ln(4/3)$$

$$\frac{Aa}{w\delta c} = \ln 4\left[(\ln 4)/\ln(4/3)\right]^{1/(R^2-1)} \qquad . \qquad (21)$$

Equation 21 is evaluated in Table 1. Since from Eq. 16 a is
always less than unity, it is clear that extremely large parts
or fine grain sizes are required for small coefficients of vari-
ation in strength ($R \simeq 1$).

TABLE 1.   Evaluating Eq. 21 for Area Ratio Parameter $Aa/w\delta c$
in Terms of the Ratio of Quartile Strengths,
$R = S(.75)/S(.25)$

R	$Aa/w\delta c$	$\ln((Aa/w\delta c)/\ln 2) = b/(S(.5)/S_u)^2$
1.01	1.3145×10^{43}	78.928
1.02	1.1116×10^{17}	39.616
1.04	3.2454×10^{8}	19.964
1.08	1.7633×10^{4}	10.144
1.16	1.3120×10^{2}	5.243
1.32	1.1528×10^{1}	2.811
1.64	3.5161	1.624
1.96	2.4109	1.247

With the a parameter determined, the b parameter can be found
directly from Eq. 19, using the median strength, $S(.5)$. For
this purpose, the denominator under the radical is included as
the last column of Table 1.

For large sample sizes, the parameters a and b can be
estimated more accurately by plotting the entire sample using
coordinate axes that are chosen to give a straight line relation
for an infinitely large sample. Since we are interested in the

minimum strengths, the plotting will be similar to that on
ordinary extreme value paper, rotated 180° to change the sign
of both axes, and with the reciprocal of the square of strength
plotted as the abscissa. The plotting position for the ith
point of a sample of N is at $i/(N+1)$ (Gumbel, 1958, p. 34). An
example of the coordinate paper is shown in Fig. 2, which also
gives the graphical construction from which the parameters in-
volving a and b can be obtained. In practice a different scale
might be used for determining the parameters; this scale is
needed for extrapolation to very safe structures. b is given by
the slope of the curve; from b and S at $\Phi = 1-1/e$, Aa/wδc can
be calculated from Eq. 19:

$$b/(S(.632)/S_u)^2 = \ln Aa/w\delta c \tag{22}$$

A measure of the skewness of the distribution is given
by the difference between the mode and first quartile strengths,
divided by the interquartile strengths. From Eq. 19 this ratio
has the form

$$\frac{S(.5) - S(.25)}{S(.75) - S(.25)} = \frac{\sqrt{1/\ln(Aa/w\delta c\ln 2)} - \sqrt{1/\ln(Aa/w\delta c\ln 4)}}{\sqrt{1/\ln(Aa/w\delta c\ln(4/3))} - \sqrt{1/\ln(Aa/w\delta c\ln 4)}} \tag{23}$$

Clearly the skewness and hence the shape of the distribution
depends on the size of the sample, A/wδc. This is further
evidence that this extreme value distribution function is not
one of the asymptotic types, which are defined as having similar
shapes, with only the mean and scale factor changing with sample
size.
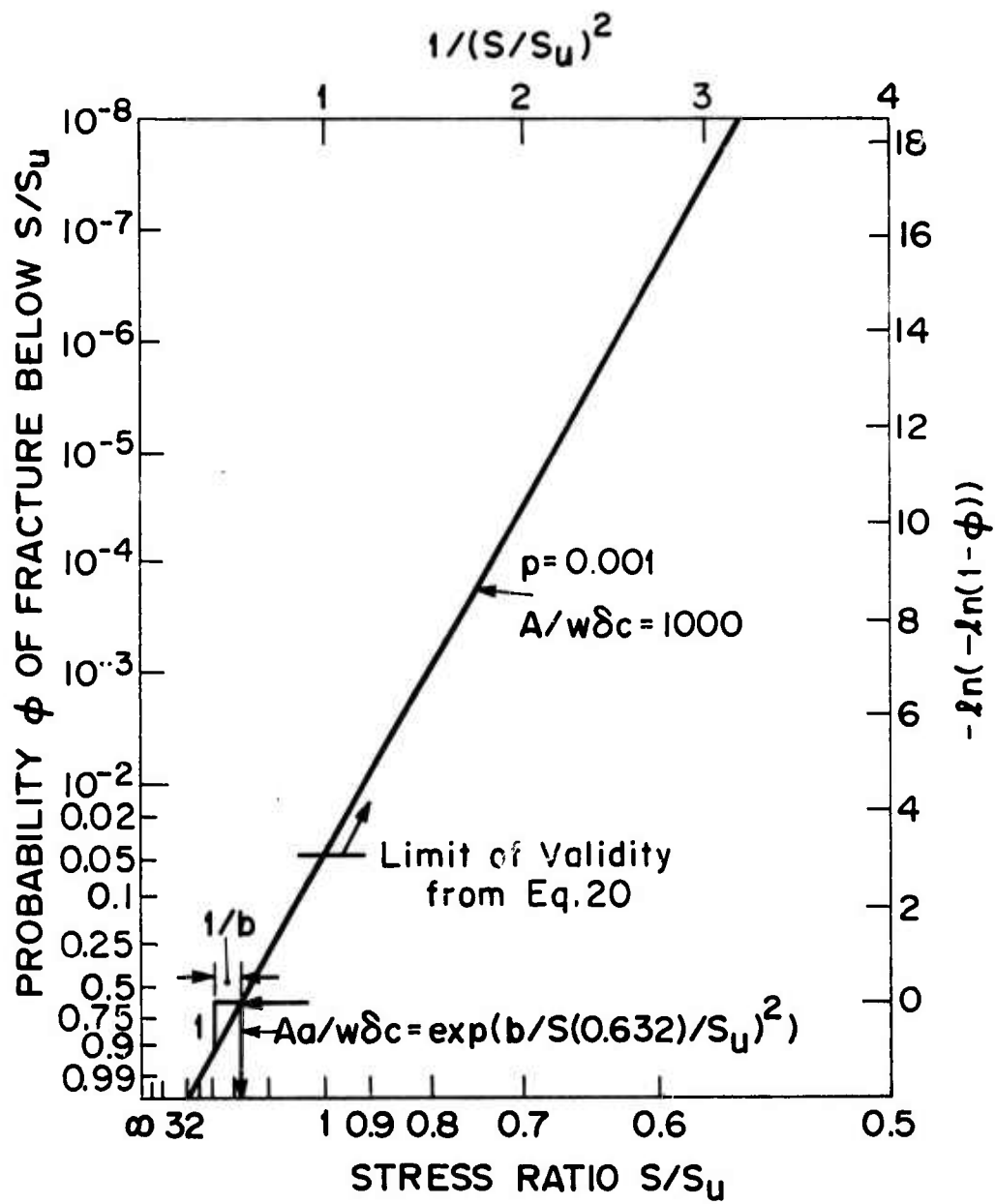
Figure 2. Plotting Crack Statistics

4.   SIZE EFFECT FROM CRACK STRENGTH STATISTICS

With increased specimen size one would expect that the median value of the strength and the scatter in strength, as expressed by the interquartile range, would both decrease. The median and quartile strengths are given in Table 2, along with the interquartile range as a measure of scatter. Only those values of probability of grain failure p and specimen-to-crack area ratio $A/w\delta c$ are listed for which the quartile strength ratios S/S_u are less than unity; higher strengths are fictitious in this model, and arise from approximating the integer number of cracked grain segments by a continuous variable. In reality, the variability in this regime is caused by variations in grain boundaries. This illustrates one of the dangers of extrapolating statistics obtained at one level of probability to a widely different one; different physical mechanisms may be governing. If so, a single distribution function is not appropriate, and the correct one cannot be found from a few observations alone.

Returning to the results given in Table 2, increasing the area ratio by a factor of 100 decreases the median by 2 to $2\frac{1}{2}$ interquartile ranges or about 3 or 4 standard deviations. As we shall see below, this is rather large compared to the values for the extreme value distribution of the third kind, and is nearly equal to that for the distribution of the first kind, which has the possibility of a negatively infinite strength. On the other hand, the standard deviation for the asymptotic ex-

TABLE 2. Quartile and Median Strengths from Crack Strength
Statistics, Eq. 19.

S_u is strength with One Crack of Unit Length; p is probability of a
segment being cracked. Values of $A/w\delta c$ satisfy Eq. 20 for each p.

Area ratio,

$A/w\delta c$	$\dfrac{S(.25)}{S_u}$	$\dfrac{S(.5)}{S_u}$	$\dfrac{S(.75)}{S_u}$	$\dfrac{S(.75)-S(.25)}{S_u}$
p = .01				
10^3	.863	.932	.999	.136
10^4	.737	.778	.816	.079
10^6	.593	.614	.632	.039
10^8	.510	.523	.534	.024
10^{10}	.454	.464	.471	.017
p = .001				
10^5	.842	.883	.919	.077
10^6	.757	.787	.812	.055
10^8	.644	.662	.677	.033
10^{10}	.570	.582	.592	.022
p = .0001				
10^6	.917	.956	.991	.074
10^8	.769	.792	.812	.043
10^{10}	.676	.691	.704	.028
p = .00001				
10^8	.893	.921	.946	.053
10^{10}	.777	.796	.812	.035

treme value distribution of the first kind does not vary with size, whereas for this distribution increasing the size by a factor of 100 drops the interquartile range by a factor of $1\frac{1}{2}$ to 2. Thus the prediction of size effect is substantially different from that for the asymptotic extreme value distributions.

One of the primary practical questions is how many inter-quartile ranges one must drop below the median to get the probability of failure below some acceptable value. This is presented in Table 3. The ratio is dependent on specimen size, as expected, since the distribution is not one of the asymptotic types (of constant shape). For a failure probability of 10^{-6}, one must drop 4 to 6 interquartile ranges (about 6 to 9 standard deviations) below the median. This corresponds to the drop for an asymptotic extreme value distribution of the third kind with $m = 10$ to 20. Expressed another way, for a failure probability of 10^{-6} one must lower the stress to 2/3 to 1/2 of the median value.

5.   SIZE AND SHAPE EFFECTS FOR ASYMPTOTIC DISTRIBUTIONS

For comparison with the size effect arising in the crack strength statistics, the size effect on the median for the third asymptotic distribution is (see for example McClintock 1971)

$$\frac{\bar{S}_1 - \bar{S}_2}{\sigma_1} = \frac{1 - (A_1/A_2)^{1/m}}{\{(2/m)!/[(1/m)!]^2 - 1\}^{\frac{1}{2}}} \tag{24a}$$

where m is related to the exponent k of Eq. 4 by $m = k + 1$. In the limit as m approaches infinity, Eq. 24a becomes

-290-

TABLE 3.   Stress Relative to Median for Failure Probability ϕ

$$\frac{S(\phi)-S(.5)}{S(.75)-S(.25)}$$

$\phi =$	10^{-2}	10^{-4}	10^{-6}	10^{-8}

Area ratio

$A/w\delta c$

Grain boundary crack probability p = .01

	10^{-2}	10^{-4}	10^{-6}	10^{-8}
10^3	-1.74	-2.65	-3.20	-3.58
10^4	-1.94	-3.13	-3.90	-4.46
10^6	-2.17	-3.74	-4.86	-5.72
10^8	-2.29	-4.10	-5.48	-6.58
10^{10}	-2.36	-4.35	-5.92	-7.22

Grain boundary crack probability p = .001

10^5	-2.02	-3.33	-4.21	-4.86
10^6	-2.13	-3.63	-4.68	-5.47
10^8	-2.26	-4.03	-5.36	-6.41
10^{10}	-2.35	-4.30	-5.84	-7.09

Grain boundary crack probability p = .0001

10^6	-2.08	-3.50	-4.47	-5.20
10^8	-2.24	-3.95	-5.22	-6.22
10^{10}	-2.33	-4.24	-5.74	-6.94

Grain boundary crack probability p = .00001

10^8	-2.21	-3.86	-5.07	-6.00
10^{10}	-2.31	-4.18	-5.63	-6.79

$$\frac{\bar{S}_1 - \bar{S}_2}{\sigma_1} = .78 \ell n \; A_1/A_2 \qquad . \qquad\qquad (24b)$$

The ratio of the standard deviations of the two sizes is

$$\sigma_2/\sigma_1 = (A_1/A_2)^{1/m} \qquad . \qquad\qquad (25)$$

These relations are plotted in Figure 3 for an area ratio of 100.

The effect of strain gradients and hence specimen shape on strength can be found by assuming that a three dimensional strain distribution can be given in terms of the maximum stress S_p the stress S_ℓ below which the probability of failure of an asymptotic distribution of the third kind goes to zero (see Eq. 3-4), the coordinates x_1, x_2, and x_3 from the point of maximum stress, and corresponding scale factors ℓ_1, ℓ_2 and ℓ_3 by the following equation which is convenient for statistics, if awkward for stress analysis:

$$S-S_\ell = (S_p - S_\ell) [1-(x_1/\ell_1)^{p_1}][1-(x_2/\ell_2)^{p_2}][1-(x_3/\ell_3)^{p_3}] \qquad (26)$$

For example, exponents p_i of infinity, 2, or 1 give stress distributions that are constant, fall off parabolically, or fall off linearly from the maximum. The substitution of Eq. 26 into Eq. 3 gives asymptotic distributions for the maximum stresses S_p in various specimens, except that the area of Eq. 3 is replaced by a product of the effective lengths ℓ_i, which are (correcting an error in McClintock and Argon, 1966, p. 597)
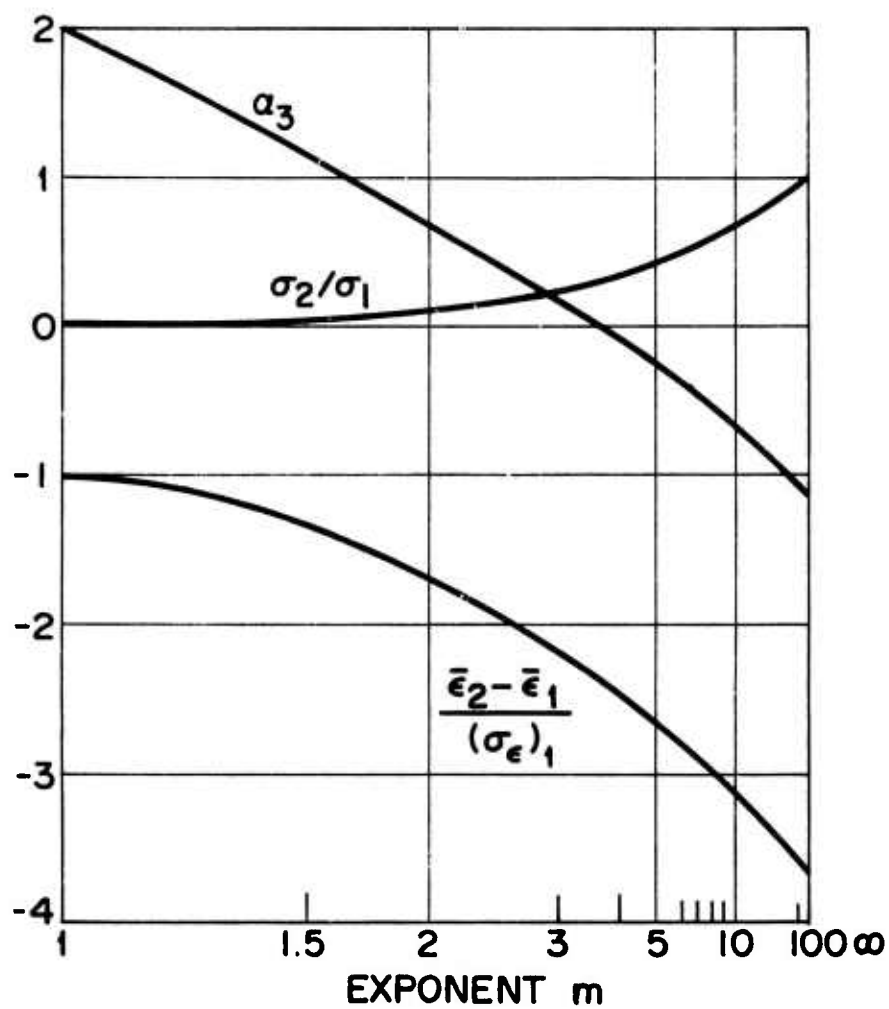
Figure 3.   Skewness α_3, standard deviation in strength σ, and mean strength $\bar{S}$, as for $A_2/A_1 = 100$ as functions of m for the third asymptotic distribution.

$$\ell_i^e = \frac{\ell_i}{p_i} \frac{\Gamma(m+1)\Gamma(1/p_i)}{\Gamma(m+1+1/p_i)} \qquad . \tag{27}$$

For example, for exponents p_i of ∞ or 1, the effective lengths $1_i^e/1_i$ are 1 or $1/(m+1)$, respectively. The size effect equations, Eqs. 24 and 25, now may be used directly for the statistics of the most stressed point, S_p.

6.   ESTIMATING VARIATION IN STRENGTH FROM VARIATION
                    IN POSITION IN FRACTURE

In testing brittle materials, the question often arises as to whether the variability in the test arises from the test technique itself or is variability inherent in the material. For example, only that inherent in the material is important in estimating the size effect. McClintock (1955 a,b) showed that in the testing specimens in which the stress fell off parabolically from the point of maximum stress, according to an equation of the form

$$S = S_p(1-az^2) \qquad , \tag{28}$$

the standard deviation in position of failure about the point of maximum strength could be related to the expected standard deviation in strength by an equation of the form

$$\sigma_S = 2.3 \ a\sigma_z^2 \qquad . \tag{29}$$

The coefficient a is determined by the geometry and method of loading and is illustrated in Fig. 5. The cases of pure bending
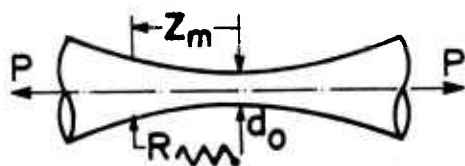
$$Z_m = 0$$

$$a = 2/d_0 R$$
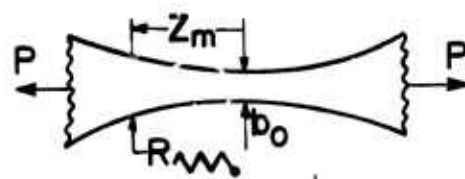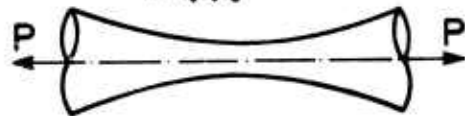
$$Z_m = 0$$

$$a = 1/b_0 R$$
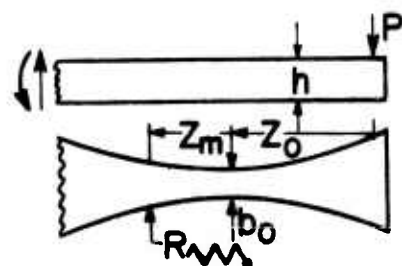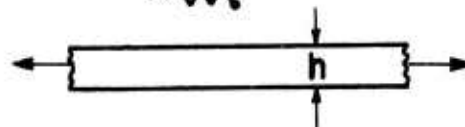
$$Z_m = \frac{b_0 R}{2Z_0} - \frac{b_0^2 R^2}{8Z_0^3} + \cdots$$

$$a = 1/b_0 R - 1/4Z_0^2 + \cdots$$
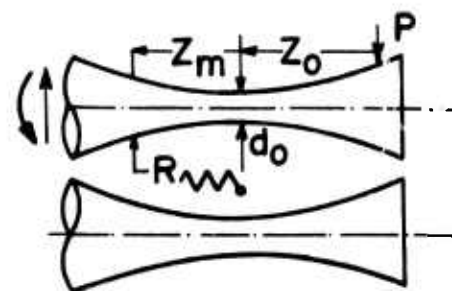
$$Z_m = \frac{d_0 R}{6Z_0} - \frac{5d_0^2 R^2}{216Z_0^3} + \cdots$$

$$a = 3/d_0 R + 1/4Z_0^2 + \cdots$$
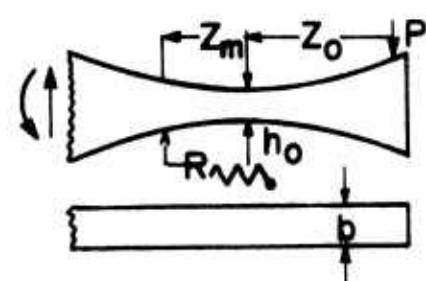
$$Z_m = \frac{h_0 R}{4Z_0} - \frac{3h_0^2 R^2}{64Z_0^3} + \cdots$$

$$a = 2/h_0 R + 1/8Z_0^2 + \cdots$$

Figure 5. Equations for parabolic stress distributions in test specimens.

can be obtained from cantilever bending by letting z_0 approach infinity.

For three-point bending tests with scatter in position of failure large compared to the specimen thickness, the stress varies linearly with position. An approximate relation between the standard deviations in strength and position can be obtained by considering a material that is perfect except for flaws of a single strength which reduce its strength to S_p-R. We can find the ratio of standard deviation in strength to that in position of failure for this bi-modal distribution, and, judging from the fact that the results for standard deviations are generally independent of the shape of the distribution function, we can apply the ideas to experimental data.

Let the density of flaws per unit area be p and the specimen width be b. Then the probability of no flaw within the first $2z/\delta z$ elements and also a flaw occurring at $\pm z$ from the origin is given by

$$(1-pb\delta z)^{2z/\delta z} \; 2pb\delta z \quad . \tag{30}$$

Going to the limit as δz approaches zero, as in deriving Eq. 9, then gives the frequency distribution of failure in space:

$$f(z)dz = e^{-2pbz} \; 2pbdz \quad . \tag{31}$$

The most distant location of fracture is at the point where the applied stress has dropped to the flawed strength:

-296-

$$z_c = (\ell/2)(R/S_p) \tag{32}$$

The standard deviation of position of failure about $z = 0$ is now found by integration

$$\sigma_z = \sqrt{\int f(z)z^2 dz} = \sqrt{\int_0^{z_c} z^2 e^{-2pbz}\, 2pb\, dz}$$

$$\sigma_z = \sqrt{\frac{1}{(2pb)^2}[-\exp(-2pbz_c)[(2pbz_c)^2 + 4pbz_c + 2] + 2]}\ .$$

For convenience define $Q = 2pbz_c = pb\ell R/S_p$:

$$\sigma_z = \frac{\ell R}{2QS_p}\sqrt{2 - \exp(-Q)[Q^2 + 2Q + 2]}\ . \tag{33}$$

The standard in strength is most easily calculated by basing it on the local stress at the site of the fracture, which will either be S_p, for the specimens which fail at the center without a flaw, or will be $S_p - R$ for the remaining specimens which fail at a flaw. The mean strength of this bi-modal distribution is

$$\bar{S}_\ell = S_p - R(1 - \exp(-Q)) \tag{34}$$

and the standard deviation is found from the deviations of the two modes from the mean:

$$\sigma_{S_\ell} = \sqrt{e^{-Q}[R(1 - e^{-Q}]^2 + (1 - e^{-Q})[Re^{-Q}]^2}\ , \tag{35}$$

which reduces to

$$\sigma_{S_\ell} = R \sqrt{e^{-Q} - e^{-2Q}} \quad . \tag{35}$$

The ratio of the standard deviation in strength to the mean strength $\bar{S}$ can now be expressed as a fraction of the standard deviation in position of failure, normalized with respect to specimen half length $\ell/2$. For the limiting case of small flaw probabilities, $Q \rightarrow 0$, this ratio can be shown to approach $\sqrt{3}$ by keeping 3rd order terms in the denominator.

$$\frac{\sigma_{S_\ell}/S_\ell}{\sigma_z/(\ell/2)} = \frac{Q\sqrt{\exp(-Q) - \exp(-2Q)}}{[1 - (R/S_p)(1 - \exp(-Q))]\sqrt{2 - \exp(-Q)[Q^2 + 2Q + 2]}} \tag{36}$$

$$\rightarrow \sqrt{3} \text{ as } Q \rightarrow 0$$

Matthews (1973) compared Eq. 36 with data on a number of specimens of potassium chloride, .875 inches long and .125 inches square cross-section. The results, in Table 4, show satisfactory agreement with the expected ratio, indicating that there were no sources of variability in testing beyond the effects of surface preparation present in the material itself. A more formal comparison of the two sources of variability could be obtained by deriving approximate confidence limits for the ratios of the two standard deviations as done by McClintock (1955b).

7.   DISCUSSION

In his book on extreme value distributions, Gumbel (1958) introduced his chapter on the first asymptote with the quotation,

TABLE 4. Variation in Position of Failure and Strength of
3-Point Bend Potassium Chloride Specimens.
$\ell = .875$ inches

Prep	σ_z, in	σ_S, psi	$\bar{S}$, psi	$\dfrac{\sigma_S/\bar{S}}{\sigma_z/(\ell/2)}$	SAMPLE SIZE
1	.064	960	3140	2.09	10
2	.123	1500	2370	2.32	9
3	.040	710	2840	2.72	10
4	.031	620	2830	3.11	4
5	.070	1030	2730	2.38	5
6	.053	550	2720	1.66	5
7	.036	1470	3630	4.91	4
8	.067	950	2320	2.68	7
9	.034	380	2030	2.41	11
10	.054	500	1960	2.06	10
Theoretical				$\rightarrow \sqrt{3}$	

"'sint ut sunt aut non sint.' (Accept them as they are or deny their existence.)" In the past, I have been accepting them as they are. In light of the crack model discussed above, I now take a more moderate approach. While some distributions may approach a constant shape (be asymptotic), there are others in which the shape will continue to change. Probably we should make estimates from several bases as an aid in judging what to do in a particular situation. The crack distribution presented here is not in contradiction with the existence of asymptotes. It has been overlooked simply because Gumbel's book was published only a year after Williams' and Irwin's pioneering papers of fracture mechanics, so that it is not surprising that Gumbel made the assumption that "the breaking strength is a linear function of the size of the largest flaw alone." What is rather surprising is that apparently so many people, including myself, have failed to note this restriction and to look at the possibility of circumventing it.

The next improvement in the model is to take into account the interactions of cracks with each other, but this effect is likely to be so complicated that it must be handled numerically. An analysis such as that given here will be of help in correlating and interpreting the results of any such numerical calculations, some of which are currently being carried out.

Another limitation to the use of extreme value distributions is the lack of small sample estimates for the parameters

involved. For practical reasons the engineer usually has only small samples available. Even if he were to test hundreds of specimens they would probably come from one lot or one supplier, or at most a small sample of lots and suppliers. The possibility of developing a small sample estimate does not seem very bright except for the rather involved method which Lieblein has developed for the asymptotic extreme value distribution of the first kind. This method could be used here by transforming the stress variable to the square of its reciprocal. Alternatively, it might appear that one could determine the sampling statistics for small sample distribution empirically by programming a computer to carry out the sampling. In practice, very large sample sizes are needed to get high levels of confidence for the desired statistics.

Finally, I must apologize that time has not allowed surveying the literature to find data to which this crack strength distribution might apply. Nor can I do more than mention some of the other work in the field, such as that by Batdorf and Cruse (1973) on multiaxial stress, Dukes (1971) on design with brittle materials, and Oh and Finnie (1970) on location of fracture.

REFERENCES

Batdorf, S. B.   1973   "A Statistical Theory for the
  Crose, J. G.   Fracture of Brittle Structures Sub-
                           jected to Non-uniform Polyaxial
                           Stresses," Aerospace Corp. Report
                           No. TR-0073 (3450-76)-2 for Space
                           and Missile Systems Organization,
                           Air Force Systems Command, Los
                           Angeles Air Force Station, Report
                           SAMSO-TR-73-162, 34 pp.

Dukes, W. H.   1971   Handbook of Brittle Material Design
                           Technology, Advisory Group for Aero-
                           space Research and Development,
                           AGARDograph Report No. AGARD-AG-152-
                           71 (AD-719 712) 160 pp., Paris (France).

Fisher, R. A.   1928   "Limiting Forms of the Frequency Dis-
  Tippett, L. H. C.   tribution of the Largest or Smallest
                           Member of the Sample," Proc. Camb.
                           Phil. Soc. 24, 180-190.

Fréchet, M.   1927   "Sur la loi de probabilité de l'écart
                           maximum," Annales de la Soc. Polonaise
                           de Math 6, 92.

Gumbel, E. J.   1958   Statistics of Extremes, Columbia
                           University Press, New York.

Ishida, M.   1969   "Analysis of Stress Intensity Factors
                           in a Plate with any Given Distribution
                           of Cracks," Nihon Kikai Gakkai Rombunshu,
                           (Trans. Japan Soc. Mech. Eng.) 35, 1815-
                           1822. Transl. in "Preliminary Reports,
                           Memoranda and Technical Notes of the
                           ARPA Materials Summer Conference," 1971
                           1, Defense Documentation Information
                           Center.

Kendall, M. G.   1963   The Advanced Theory of Statistics,
   Stuart, A.   2nd ed., Vol. 1, Charles Griffin,
                                   London.

Matthews, J. R.   1973   Personal communication.

McClintock, F. A.   1955a   "The Statistical Theory of Size and
                                  Shape Effects in Fatigue," J. Appl.
                                  Mech. 22, 421-426.

McClintock, F. A.   1955b   "A Criterion for Minimum Scatter in
                                  Fatigue Testing," J. Appl. Mech. 22,
                                  421-431.

McClintock, F. A.   1971   "Plasticity Aspects of Fracture,"
                                  Fracture, H. Liebowitz, ed., Vol. 3,
                                  Academic Press, New York, 47-225.

McClintock, F. A.   1966   Mechanical Behavior of Materials,
   Argon, A. S., eds.   Addison-Wesley, Reading, Mass.

Oh, H. L.   1970   "On the Location of Fracture in
   Finnie, I.   Brittle Solids - I. Due to Static
                                  Loading," Int. J. Fracture Mech. 6,
                                  287-300. See also "-II. Due to Wave
                                  Propagation in a Slender Rod," 333-
                                  339.

Sprunt, E. S.   1973   Personal communication.
   Brace, W. F.

Weibull, W.   1939   "A Statistical Theory of Strength of
                                  Materials," Ing. Vet. Ak. Handl. 151,
                                  Royal Inst. of Tech., Stockholm.
                                  See also, "The Phenomenon of Rupture
                                  in Solids," 153.

RESISTANCE OF TSUEI'S DUCTILE COMPOSITE
SUPERCONDUCTORS

M. Tinkham

Abstract

It is shown that a volume fraction f_S of superconducting
filaments of length L and diameter d embedded in a normal metal
matrix yields a composite with an effective resistivity $\bar{\rho}_{eff} \simeq$
$(\rho_{matrix}/f_S)(d^2/L^2)$. Since $\bar{\rho}_{eff}/\rho_{matrix}$ may be as small as 10^{-7},
ultrasensitive techniques (such as use of a superconducting volt-
meter or decay of persistent currents) may be needed to detect
such small resistance in short sample tests. Moreover, this
resistance may be low enough for practical applications, even if
the matrix is fully normally resistive. The current density in
the matrix is less than that in the filaments by a factor of
$\sim d/2L$. If this factor is small enough, even the weak "proximity
effect" superconductivity between filaments may be able to carry
it as a supercurrent, either without resistance or with only
flux-flow resistance. Evidently great advantages follow from
obtaining the largest possible ratio of L/d (as well as the
largest possible superconducting fraction and the lowest possible
matrix resistivity) by appropriate development of metallurgical
and fabrication techniques.

Preceding page blank

RESISTANCE OF TSUEI'S DUCTILE COMPOSITE
SUPERCONDUCTORS

M. Tinkham

Tsuei[1] has reported making conductors containing many
parallel filaments of Nb_3Sn, with breaks at random positions,
all embedded in a copper matrix. The question addressed here
is to what extent the absence of continuous Nb_3Sn filaments
extending throughout will degrade the performance. If the
spacing between filaments is small enough, or if there are
particles of Nb_3Sn dispersed throughout the Cu, there may be
superconductivity throughout by the proximity effect. However,
such superconductivity is always "weak"; i.e., it has a small
critical current density and it extends only a distance $\sim\xi_N$
~ 1000 Å into the matrix from a strong superconductor. Thus,
even if the specimen is truly superconducting at small current
densities, it may be slightly resistive at useful current den-
sities and fields.

An obvious limiting case is that in which the matrix has
simply the normal resistance of Cu, while the superconducting
filaments are all operating without resistance below their local
J_c. Let us take the filaments to be of length L, diameter d,
and typical spacing s between centers. The current per filament,

I, will typically enter and leave over a length L/2 at either end of the segment, giving a radial current density $J_\perp = I/\pi rL$. Assuming 6 neighbors, the associated voltage drop in a length L is

$$V \simeq \rho \, \frac{2I}{\pi L} \, \left(1 + \frac{1}{6}\right) \int_{d/2}^{s/2} \frac{dr}{r} = \frac{7}{3\pi} \, \frac{\rho I}{L} \, \ell n \, \frac{s}{d} \tag{1}$$

This is the same as if the filament had an effective resistivity

$$\rho_{eff} = \frac{V}{I} \, \frac{\pi d^2}{4L} = \rho \, \frac{7}{12} \, \frac{d^2}{L^2} \, \ell n \, \frac{s}{d} \tag{2}$$

The logarithmic factor is about 1.5 if the filaments form 5% of the volume, so that we may simplify (2) to

$$\rho_{eff} \simeq \rho \, (d^2/L^2) \tag{2a}$$

But if the filaments result from drawing down Nb grains of roughly spherical shape, d^2/L^2 will be approximately $1/R^3$, where R is the factor by which the cross-sectional area has been reduced in drawing. This was $R = 600$ in the best materials quoted by Tsuei, leading to

$$\rho_{eff} \simeq \rho/2{\times}10^8 \simeq 5{\times}10^{-16} \text{ ohm-cm} \tag{3}$$

if $\rho_{Cu} \simeq 10^{-7}$ ohm-cm as he quoted for this material. Referred to the total cross section (including the copper), the effective average resistivity is given by

$$\bar{\rho}_{eff} = (\overline{\sigma_{eff}})^{-1} \simeq \rho_{Cu}(1+f_S \, L^2/d^2)^{-1} \simeq \rho_{Cu}(1+f_S R^3)^{-1} \tag{4}$$

where f_S is the volume fraction of superconducting filaments. For the Tsuei material, this is approximately

$$\bar{\rho}_{eff} \simeq 5 \times 10^{-16} \, f_S^{-1} \text{ ohm-cm} \qquad (5)$$

or $\sim 10^{-14}$ ohm-cm if $f_S \sim 5\%$. Since this is 4 orders of magnitude below the quoted sensitivity of the measurements, it appears that even fully resistive copper could be taken to be super-conducting. Note, however, our idealized assumption that the filaments have not broken in the drawing out by a factor of 600, so that the length L should be of the order of 1 cm while $d \simeq 1$ micron. If the actual continuous filament length were less, the resistance would rise as $1/L^2$. But even if the filaments were only 100 microns long, the resistance would still be at the limit of detectability. If $\bar{\rho}_{eff}$ were 10^{-14}, at J_c of 5×10^4 amps/cm^2, the dissipation would be 2.5×10^{-5} W/cm^3, which is marginally satisfactory for large scale applications. Measurements with higher sensitivity, either using a superconducting voltmeter (sensitive to 10^{-15} V, typically) or decay of persistent currents, would be needed to make a decisive check of the resistance at the level required for practical applications.

From the above, it appears that sufficiently long filaments will give useful performance even if the copper is fully normal. Nonetheless, it would clearly be better if the proximity effect kept the copper superconducting. From the model discussed above, we see that the perpendicular current density

in the copper is less than that in the filament by a ratio

$$\frac{J_\perp}{J_{fil}} = \frac{d}{2L} \simeq 3 \times 10^{-5} \tag{6}$$

Even rather weak proximity effect superconductivity could carry such a small current density. It is also worth noting that since this current flows perpendicular to the filament in all directions to other filaments, some of the current can flow more or less parallel to any large transverse field due to other windings. The low Lorentz force in this orientation should make the superconductivity less sensitive to the magnetic field. Nonetheless, pinning forces should be so weak in the copper that flux flow resistance will probably occur, but this will never give resistance higher than that of the normal copper, estimated in the previous paragraph.

Acknowledgement

Reference

1. C. C. Tsuei, Science 180, 57 (1973); also report at this conference.

SHIELDING OF THE MAGNETIC FIELD
ONBOARD SUPERCONDUCTING TRAINS

P. L. Richards

Abstract

The problem of shielding passengers from the static magnetic field of a train levitated by superconducting magnets is discussed. A general calculation is presented which shows that iron shielding would be prohibitively heavy. Two dimensional potential calculations are used to estimate the effectiveness of superconducting sheet shields. It is concluded that a combination of iron and superconducting shields is required to achieve satisfactory results. Model experiments are suggested as a method for optimizing specific designs.

SHIELDING OF THE MAGNETIC FIELD
ONBOARD SUPERCONDUCTING TRAINS

P. L. Richards

The use of superconducting magnets onboard a high
speed transportation vehicle to provide levitation and guidance
forces is currently under study.[1-3] The presence of easily
detectable magnetic fields in the passenger compartment would
impair public acceptance of the transportation system even if it
can be shown to be harmless. Relatively little is known about
the biological effects of small magnetic fields.[4] It seems
likely however that the tolerable level will be set by the effect
on items carried by the passengers. Fields of the order of 100
Gauss are sufficient to stop watches, damage magnetic recordings,
and cause substantial forces on ferromagnetic objects. Any
acceptable design for a magnetically levitated vehicle must
therefore include shielding to reduce the field level in the
passenger areas to a value substantially less than 100 Gauss.

Ferromagnetic Shields

One useful form of magnetic shield is a sheet of soft
ferromagnetic material such as iron placed approximately parallel
to the flux lines. As long as the material remains unsaturated,
the magnetic reluctance is minimized for flux lines which pass

through the sheet. This procedure is very effective in reducing already small magnetic fields and making them negligible. A 1/16 in. thick iron sheet (with saturation induction B_{sat} = 2×10^4 Gauss) surrounding a region with transverse dimensions $\simeq 10$ ft. can concentrate a field of 20 Gauss and effectively remove it from the interior. The difficulty with magnetic shielding is not associated with such small values of field, but with the large field values which exist close to the levitation magnets. A much larger amount of iron is required for the primary shielding of these high fields.

As an example, consider a large superconducting magnet in a horizontal plane at a distance h below the floor of a train car. The magnet is also assumed to lie a distance h above a diamagnetic track. If the floor of the train is covered with a soft iron sheet sufficiently thick to avoid saturation, essentially all of the field passes through this "flux return" and is shielded out of the passenger compartment. The efficiency of the levitation magnets will be increased by the reduced reluctance of the magnetic circuit. We can estimate the weight of iron required if a car of total weight W_T is supported on an rms magnetic field B over an area A of track.

$$W_T = \frac{B^2 A}{8\pi} \tag{1}$$

The thickness of iron of area A required to avoid reaching the saturation induction B_{sat} is given approximately by

$$d = \frac{\lambda B}{B_{sat}} \tag{2}$$

Here λ is $\sim\frac{1}{2}$ of the length or the width of the magnet, whichever is smaller. The weight of the shield of density ρ is thus given by

$$W_S = Ad\rho g = \frac{8\pi W_T d\rho g}{B^2} \tag{3}$$

so that for a magnet dimension $\lambda = 30$ cm, the fractional weight of an iron shield is

$$\frac{W_S}{W_T} = \frac{8\pi\rho g\lambda}{B\ B_{sat}} \simeq \frac{300}{B}\ (cgs)^{-1} \qquad , \tag{4}$$

where B is in Gauss. For a 45,000 Kg car with magnets over an area of 3×30 m^2, $B \simeq 10^3$ Gauss so that the shield must be 1.5 cm thick and comprise 1/3 of the total train weight. This very severe condition can be eased somewhat by localizing the magnets at each end of the car. If the area covered by magnetic field is reduced by a factor α, then the shield weight required is reduced by $\alpha^{\frac{1}{2}}$. Perhaps a more important advantage of local-izing the magnets at the ends of the car is the opportunity to locate relatively field insensitive cryogenic equipment rather than passengers over the magnets. The field area should not be made too small, however, or averaging of track irregularities, which is important for ride quality, will be lost.

Superconducting Shields

We have seen that a ferromagnetic flux return can be used to shield the passenger compartment, but that in many cases the

weight of material required is excessive. An alternative
approach to the primary shielding problem is to use a super-
conducting sheet located above the levitation magnet as a dia-
magnetic shield. This has the disadvantages of complicating
the cryogenic system and, since the shielding currents tend to
cancel the field at the track, lowering the efficiency of the
levitation magnets. In order to investigate these effects, we
have solved several approximate field distribution problems by
digital computer.

The calculation of the magnetic field of a current loop
in the vicinity of a diamagnetic sheet of finite extent is the
fundamental problem of magnetic levitation. This problem can be
solved in the general case only by a very laborious calculation
using a self-consistent vector potential.[5] Fortunately, reasonably
accurate values of lift and drag can be obtained by methods which
neglect edge effects. Such approximations are useless, however,
for evaluating the shielding effectiveness of finite sheets.

In order to approximate the shielding problem we replace
the train magnet loops by two parallel wires which carry current
in opposite directions parallel to the train motion. A two di-
mensional solution is then easily obtained which approximates
the real shielding problem for long narrow train magnets. In
the real problem the fields vanish in planes of antisymmetry
which pass between train loops and are maximum in planes of
symmetry which bisect train loops. The approximate calculation
is valid for these latter planes for distances from the current

-315-

sources which are small compared with the actual coil length. Where it is in error, the approximation overestimates the magnetic field.

In order to determine the range of validity of the calculation we have compared the field caused by a line of current, $B = 2I/cr$, to that in the symmetry plane of a line in which the same current flows in segments of length ℓ, but alternating in direction,

$$B = \frac{2I}{cr_0} \sum_{n=0}^{\alpha} \left\{ \frac{2n}{\sqrt{(2n)^2+(2r/\ell)^2}} - \frac{2(2n+1)}{\sqrt{(2n+1)^2+(2r/\ell)^2}} + \frac{2n+2}{\sqrt{(2n+2)^2+(2r/\ell)^2}} \right\} \tag{5}$$

where the series is grouped for rapid convirgence. The uniform current approximation overestimates the field by a factor 2 when $r = 0.5\ell$, and a factor 10 when $r = 2.5\ell$.

The shielding calculations are best described by reference to Figures 1a and 1b which show a single line of train megnets symmetricly placed between a diamagnetic shield (above) and a diamagnetic track (below). Since the magnetic field has right-left symmetry, only the right half plane is computed. The numbers in Fig. 1a are the solutions to Laplace's equation $\nabla^2 A = 0$ for a source (with $A = 1$ at a single computational grid point) which is marked with a circle and for potential $A = 0$ at the two shield planes which are shown as horizontal lines. The potential vanishes along the left margin which is a line of antisymmetry for A and at infinity, which is taken as a rectang-

ular box slightly larger than the region for which the potential is plotted.

The computation of A was done by a Gauss-Seidel relaxation method using a computational grid space equal to 1/4 of the plotting grid in the vicinity of the source and shields and 1/2 of the plotting grid further away. The calculation was terminated when the change between successive iterations was less than 10^{-5} of the largest value of A. The results proved slightly sensitive to the location of boundaries, so that only the first significant figure is valid. This accuracy is sufficient because of the approximate nature of the two dimensional model.

The potential A can be thought of as a magnetic vector potential oriented perpendicular to the plane. Contours of constant A plotted in Fig. 1a are therefore flux lines. The magnitude $|B|$ of the magnetic field in Fig. 1b is obtained from the square root of the sum of the squares of the derivatives of A in the vertical and horizontal directions. Derivatives are obtained from the differences between values of A on either side of a computational grid point. The finite value of $|B|$ at the source point is an artifact of this calculational method.

The results in Fig. 1b allow a comparison of the magnetic field in the passenger compartment with the field at the surface of the track which gives levitation. These results are valid when the velocity is high enough that the track is effectively diamagnetic. We see that without further shielding the field at the floor of the passenger compartment is at least 10 times less

than the average levitation field and that it is ∿50 times less over most of the area occupied by passengers. Figs. 2a,b and 3a,b give corresponding results for cases in which the shield plate (and track) extends less far beyond the edges of the train magnet. A narrower shield means a smaller cryostat, but larger stray fields. In Fig. 4a,b we give one example of the field distribution for a single shielding plane. This can be interpreted either as a shielded magnet moving slowly enough that track diamagnetism can be neglected, or (inverted) as an unshielded magnet over a diamagnetic track.

Unless the superconducting shields extend rather far beyond the edges of the train magnets, they do not by themselves reduce the fields in the passenger area to an acceptable level. For the train design sketched in Fig. 5a, the best shield would combine both iron and superconducting sheets. The former should be positioned vertically along the sides of the train as shown in Fig. 5a, so as to lie parallel to the fringing flux lines and provide a low reluctance return path. The fields in this region are more than an order of magnitude less than the levitation field so the weight of such a shield should not be prohibitive.

An alternative train design under consideration in Japan consists of two parallel sets of train loops over relatively narrow looped tracks. In order to provide levitation such looped tracks must be reasonably diamagnetic at high speeds so, for purposes of shielding estimates, can be replaced by homogeneous sheets as shown in Fig. 5b. The computations summarized in Figs. 1-4 can

-318-

also be used to estimate shielding requirements for this geometry.

Model Experiments

The optimization of magnetic shielding for a specific train design is complicated by the need for both diamagnetic and ferromagnetic shield materials as well as by the multiple magnets and/or complicated track geometries required for transverse stability. Although computer codes could be developed to investigate this problem, they would be very involved even in the two dimensional approximation.

A much simpler approach is to use model experiments. It is relatively easy to wind coils of copper wire several inches on a side to represent train magnets. If such coils are connected in series and driven from an oscillator, the ac magnetic field can be detected with a pick-up coil and the skin effect can be used to simulate superconducting shields. At 100 kHz, 1/16 in. thick aluminum sheets are essentially diamagnetic. The overall dimensions of the model should be chosen for ease of fabrication of shield or track sections. For example, measurements could be made near the center of a line of 5 train magnets with a total length of 36 in. ($\sim$1/12 scale). A wide variety of track and shield configurations could then easily be cut and bent from standard widths of sheet aluminum.

Ferromagnetic foil with a well-defined saturation magnetization is available commercially.[6] In order to reproduce the effects of both diamagnetic and ferromagnetic shields in static fields, the frequency at which the model train magnet

-319-

coils are driven must be chosen so that very little flux penetrates 1/16 in. aluminum, but penetration of 1 mil iron is nearly complete. The ratio $R = B_\ell/B_u$ of magnetic field on either side of a conducting sheet in the presence of a single Fourier component of ac magnetic field is easily obtained from the well-known result for the lift force F_L in a magnetic train in the low velocity limit.[1]

$$F_L = \frac{(B_u^2 - B_\ell^2)A}{8\pi} = \frac{B_u^2 A}{8\pi}(1-R^2) = F_\infty \frac{v^2/v_0^2}{1+v^2/v_0^2} \qquad (6)$$

so that

$$R = (1+\omega^2\tau^2)^{-\frac{1}{2}} \qquad , \qquad (7)$$

where $\tau = 1/kv_0 = \sigma dZ_0/2ck$. For train coils with a 12 in. period and a frequency of 10^4 Hz, 1/16 in. aluminum has a penetration ratio of $R \simeq 10^{-2}$ and 1 mil iron $R \simeq 0.9$. These values are almost near enough the ideal values of zero and unity to allow accurate conclusions to be obtained from the model. A larger spread of values could be obtained by reducing the thickness of the ferromagnetic foil or increasing the thickness of this aluminum. Aluminum thicker than 1/8 in. would be very much more difficult to fabricate.

Since ferromagnetic shields should not be used in a highly saturated region, it is not necessary to exactly model saturation phenomena in order to gain insight into the use of such shields. It is interesting to observe, however, that if the dimensions and materials are scaled so that $B\ell/B_{sat}d$ is kept

constant then a corresponding amount of saturation can be obtained. If the characteristic dimension ℓ of a current loop is reduced by 12, and the shield thickness decreased by 500, then for a factor 3 smaller B_{sat},[6] the field must be reduced by a factor ~ 100. Since the magnetic field from a current loop B varies as I/ℓ, the required current is $\sim 10^3$ less than in a full scale system, or $\sim 10^2$ Ampere turns. A set of such magnets of the suggested dimensions would require copper coils with 1/16 in.^{2} cross-sectional area and a $\sim 25W$ source. Commercial ferromagnetic shielding foils are not ideal for this application since the materials are selected for maximum B_{sat}. Thinner foil or special materials with lower B_{sat} would ease the magnet current requirements considerably.

Conclusions

We have seen that for many designs of magnetically levitated trains, adequate shielding of the passenger compartment from stray magnetic fields requires a combination of superconducting dia-magnetic and iron ferromagnetic shields. Calculations of shielding efficiency for these combined systems are difficult to do accurately, but small scale models can easily be built which will aid in shield optimization.

Acknowledgement

acknowledge many stimulating discussions with Professor M. Tinkham on this subject. He is greatly indebted to Dr. B. Carnahan for writing the computer programs used.

References

1. P. L. Richards and M. Tinkham, "Magnetic Suspension and Propulsion Systems for High Speed Transportation", J. Appl. Phys. 43, 2680 (1972).

2. H. T. Coffey, F. Chilton, and L. O. Hoppie, "The Feasibility of Magnetically Levitating High Speed Ground Vehicles" S.R.I. Report FRA-RT-72-39, D.O.T. Contract-FR-10001 (1972).

3. J. R. Reitz, R. H. Borcherts, L. C. David, and D. F. Wilkie, "Technical Feasibility of Magnetic Levitation as a Suspension System for High-Speed Ground Transportation Vehicles, Vol. I", Ford Motor Co. Report, D.O.T. Contract DOT-FR-10026 (1972).

4. M. F. Barnothy, ed., "Biological Effects of Magnetic Fields," (Plenum Press, New York, 1964 and 1969) Vol. 1 and 2.

5. L. Hannakam, Elektrotech. Z. (ETZ-A) 86, 427 (1965).

6. Co-Netic AA from the Perfection Mica Co., 1322 N. Elston Avenue, Chicago, Illinois.

```
0.0    4.0E-05 8.1E-05 1.2E-04 1.6E-04 1.9E-04 2.0E-04 2.0E-04 1.8E-04 1.5E-04 1.1E-04

0.0    8.5E-05 1.7E-04 2.6E-04 3.4E-04 4.0E-04 4.3E-04 4.3E-04 3.9E-04 3.2E-04 2.3E-04

0.0    1.4E-04 2.8E-04 4.2E-04 5.5E-04 6.5E-04 7.0E-04 6.9E-04 6.4E-04 5.3E-04 3.8E-04

0.0    2.1E-04 4.2E-04 6.3E-04 8.2E-04 9.6E-04 1.0E-03 1.0E-03 9.5E-04 7.8E-04 5.6E-04

0.0    3.0E-04 6.0E-04 8.9E-04 1.2E-03 1.4E-03 1.5E-03 1.5E-03 1.3E-03 1.1E-03 7.9E-04

0.0    4.1E-04 8.2E-04 1.2E-03 1.6E-03 1.9E-03 2.1E-03 2.1E-03 1.9E-03 1.6E-03 1.1E-03

0.0    5.4E-04 1.1E-03 1.7E-03 2.2E-03 2.6E-03 2.9E-03 2.8E-03 2.6E-03 2.1E-03 1.5E-03

0.0    6.9E-04 1.4E-03 2.2E-03 3.0E-03 3.6E-03 3.9E-03 3.9E-03 3.5E-03 2.9E-03 2.0E-03

0.0    8.3E-04 1.8E-03 2.8E-03 3.9E-03 4.9E-03 5.4E-03 5.4E-03 4.8E-03 3.9E-03 2.7E-03

0.0    9.1E-04 2.0E-03 3.4E-03 5.1E-03 6.7E-03 7.5E-03 7.4E-03 6.5E-03 5.1E-03 3.5E-03

0.0    8.4E-04 2.0E-03 3.7E-03 6.5E-03 9.4E-03 1.1E-02 1.0E-02 8.7E-03 6.7E-03 4.5E-03

0.0    5.3E-04 1.3E-03 3.0E-03 7.3E-03 1.4E-02 1.6E-02 1.4E-02 1.1E-02 8.4E-03 5.5E-03

0.0    0.0     0.0     0.0     0.0     2.6E-02 2.4E-02 1.9E-02 1.4E-02 1.0E-02 6.5E-03

0.0    1.2E-01 2.2E-01 1.5E-01 8.0E-02 5.0E-02 3.4E-02 2.4E-02 1.7E-02 1.1E-02 7.2E-03

0.0    2.1E-01 1.0E 00 2.5E-01 1.1E-01 6.2E-02 3.8E-02 2.6E-02 1.8E-02 1.2E-02 7.5E-03

0.0    1.2E-01 2.2E-01 1.5E-01 8.0E-02 5.0E-02 3.4E-02 2.4E-02 1.7E-02 1.1E-02 7.2E-03

0.0    0.0     0.0     0.0     0.0     2.6E-02 2.4E-02 1.9E-02 1.4E-02 1.0E-02 6.5E-03

0.0    5.3E-04 1.3E-03 3.0E-03 7.3E-03 1.4E-02 1.6E-02 1.4E-02 1.1E-02 8.4E-03 5.5E-03
```

Figure 1a. Calculated values of vector potential A for a pair
of line current sources between two diamagnetic
sheets. The left margin is a plane of anitsymmetry
for A. Contours of constant A are flux lines.

2.0E-05 5.8E-05 9.4E-05 1.3E-04 1.7E-04 2.0E-04 2.1E-04 2.1E-04 1.9E-04 1.6E-04 1.2E-04

4.2E-05 9.9E-05 1.3E-04 1.7F-04 2.1E-04 2.3E-04 2.4E-04 2.4E-04 2.3E-04 2.0E-04 1.7E-04

7.0E-05 1.5E-04 1.9E-04 2.3E-04 2.7E-04 2.9E-04 3.0E-04 3.0E-04 2.9E-04 2.6E-04 2.3E-04

1.0E-04 2.2E-04 2.6E-04 3.1E-04 3.5E-04 3.8E-04 3.9E-04 3.9E-04 3.7E-04 3.5E-04 3.2E-04

1.5E-04 3.2E-04 3.6E-04 4.2E-04 4.7E-04 5.0E-04 5.1E-04 5.1E-04 5.0E-04 4.7E-04 4.5E-04

2.0E-04 4.3E-04 4.9E-04 5.6E-04 6.3E-04 6.7E-04 6.9E-04 6.9E-04 6.7E-04 6.4E-04 6.1E-04

2.7E-04 5.7E-04 6.4E-04 7.4E-04 8.3E-04 9.0E-04 9.3E-04 9.3E-04 9.0E-04 8.6E-04 8.2E-04

3.4E-04 7.2E-04 8.3E-04 9.8E-04 1.1E-03 1.2E-03 1.3E-03 1.3E-03 1.2E-03 1.2E-03 1.1E-03

4.1E-04 8.7E-04 1.0E-03 1.3E-03 1.5E-03 1.7E-03 1.8E-03 1.7E-03 1.7E-03 1.6E-03 1.5E-03

4.4E-04 9.7E-04 1.2E-03 1.7E-03 2.2E-03 2.5E-03 2.6E-03 2.5E-03 2.3E-03 2.1E-03 1.9E-03

4.0E-04 9.5E-04 1.4E-03 2.2E-03 3.4E-03 4.2E-03 4.0E-03 3.5E-03 3.1E-03 2.7E-03 2.5E-03

2.5E-04 7.5E-04 1.5E-03 3.1E-03 7.0E-03 8.3E-03 6.5E-03 5.1E-03 4.1E-03 3.4E-03 3.0E-03

0.0   0.0   0.0   0.0   0.0   1.9E-02 1.0E-02 7.1E-03 5.3E-03 4.2E-03 3.6E-03

1.1E-01 1.3E-01 2.9E-01 1.7E-01 7.5E-02 3.1E-02 1.5E-02 9.0E-03 6.3E-03 4.8E-03 4.0E-03

1.8E-01 2.8E-01 1.0E 75 2.3E-01 7.7E-02 3.4E-02 1.7E-02 9.9E-03 6.7E-03 5.0E-03 4.1E-03

1.1E-01 1.8E-01 2.9E-01 1.7E-01 7.5E-02 3.1E-02 1.5E-02 9.0E-03 6.3E-03 4.8E-03 4.0E-03

0.0   0.0   0.0   0.0   0.0   1.9E-02 1.0E-02 7.1E-03 5.3E-03 4.2E-03 3.6E-03

2.5E-04 7.5E-04 1.5E-03 3.1E-03 7.0E-03 8.3E-03 6.5E-03 5.1E-03 4.1E-03 3.4E-03 3.0E-03

Figure 1b.   Calculated values of the magnitude of the magnetic
field |B| for the geometry of Figure 1a.

0.0   1.5E-04 2.8E-04 4.1E-04 5.1E-04 5.7E-04 5.9E-04 5.7E-04 5.1E-04 4.2E-04 2.9E-04

0.0   3.0E-04 5.9E-04 8.5E-04 1.1E-03 1.2E-03 1.2E-03 1.2E-03 1.1E-03 8.7E-04 6.1E-04

0.0   4.9E-04 9.5E-04 1.4E-03 1.7E-03 1.9E-03 2.0E-03 1.9E-03 1.7E-03 1.4E-03 9.8E-04

0.0   7.1E-04 1.4E-03 2.0E-03 2.5E-03 2.8E-03 2.9E-03 2.8E-03 2.5E-03 2.0E-03 1.4E-03

0.0   9.9E-04 1.9E-03 2.8E-03 3.4E-03 3.9E-03 4.0E-03 3.8E-03 3.4E-03 2.8E-03 1.9E-03

0.0   1.3E-03 2.6E-03 3.8E-03 4.7E-03 5.3E-03 5.4E-03 5.2E-03 4.6E-03 3.7E-03 2.6E-03

0.0   1.8E-03 3.5E-03 5.1E-03 6.3E-03 7.1E-03 7.3E-03 7.0E-03 6.2E-03 4.9E-03 3.4E-03

0.0   2.3E-03 4.6E-03 6.7E-03 8.4E-03 9.5E-03 9.8E-03 9.3E-03 8.1E-03 6.5E-03 4.5E-03

0.0   2.9E-03 5.8E-03 8.8E-03 1.1E-02 1.3E-02 1.3E-02 1.2E-02 1.1E-02 8.3E-03 5.7E-03

0.0   3.3E-03 7.1E-03 1.1E-02 1.5E-02 1.7E-02 1.8E-02 1.6E-02 1.4E-02 1.1E-02 7.1E-03

0.0   3.3E-03 7.8E-03 1.4E-02 2.1E-02 2.4E-02 2.4E-02 2.1E-02 1.7E-02 1.3E-02 8.7E-03

0.0   2.3E-03 6.3E-03 1.6E-02 3.1E-02 3.5E-02 3.3E-02 2.7E-02 2.1E-02 1.6E-02 1.0E-02

0.0   0.0   0.0   0.0   5.5E-02 5.3E-02 4.3E-02 3.4E-02 2.6E-02 1.8E-02 1.2E-02

0.0   1.2E-01 2.2E-01 1.6E-01 1.1E-01 7.4E-02 5.4E-02 3.9E-02 2.9E-02 2.0E-02 1.3E-02

0.0   2.1E-01 1.00E 00 2.6E-01 1.4E-01 8.5E-02 5.8E-02 4.1E-02 3.0E-02 2.1E-02 1.3E-02

0.0   1.2E-01 2.2E-01 1.6E-01 1.1E-01 7.4E-02 5.4E-02 3.9E-02 2.9E-02 2.0E-02 1.3E-02

0.0   0.0   0.0   0.0   5.5E-02 5.3E-02 4.3E-02 3.4E-02 2.6E-02 1.8E-02 1.2E-02

0.0   2.3E-03 6.3E-03 1.6E-02 3.1E-02 3.5E-02 3.3E-02 2.7E-02 2.1E-02 1.6E-02 1.0E-02

Figure 2a.   Similar to Fig. 1a except for the decreased shield
              length.

7.3E-05 2.1E-04 3.2E-04 4.4E-04 5.3E-04 5.9E-04 6.1E-04 5.9E-04 5.3E-04 4.4E-04 3.3E-04

1.5E-04 3.4E-04 4.3E-04 5.3E-04 6.1E-04 6.6E-04 6.8E-04 6.6E-04 6.1E-04 5.3E-04 4.4E-04

2.4E-04 5.2E-04 5.9E-04 6.8E-04 7.5E-04 7.9E-04 8.1E-04 7.9E-04 7.4E-04 6.7E-04 5.9E-04

3.6E-04 7.4E-04 8.1E-04 8.8E-04 9.5E-04 9.9E-04 1.0E-03 9.8E-04 9.3E-04 8.7E-04 8.0E-04

5.0E-04 1.0E-03 1.1E-03 1.2E-03 1.2E-03 1.3E-03 1.3E-03 1.2E-03 1.2E-03 1.1E-03 1.1E-03

6.7E-04 1.4E-03 1.5E-03 1.5E-03 1.6E-03 1.6E-03 1.6E-03 1.6E-03 1.6E-03 1.5E-03 1.4E-03

8.9E-04 1.8E-03 1.9E-03 2.0E-03 2.1E-03 2.2E-03 2.2E-03 2.1E-03 2.0E-03 1.9E-03 1.8E-03

1.1E-03 2.4E-03 2.5E-03 2.7E-03 2.9E-03 2.9E-03 2.9E-03 2.8E-03 2.6E-03 2.5E-03 2.4E-03

1.4E-03 3.0E-03 3.3E-03 3.6E-03 3.9E-03 4.0E-03 3.9E-03 3.7E-03 3.4E-03 3.2E-03 3.0E-03

1.6E-03 3.5E-03 4.2E-03 5.1E-03 5.7E-03 5.7E-03 5.4E-03 4.9E-03 4.4E-03 4.1E-03 3.8E-03

1.6E-03 3.7E-03 5.3E-03 7.7E-03 9.2E-03 8.7E-03 7.6E-03 6.5E-03 5.7E-03 5.1E-03 4.7E-03

1.1E-03 3.3E-03 6.9E-03 1.5E-02 1.8E-02 1.4E-02 1.1E-02 8.6E-03 7.1E-03 6.2E-03 5.5E-03

0.0 0.0 0.0 0.0 4.0E-02 2.2E-02 1.5E-02 1.1E-02 8.6E-03 7.2E-03 6.3E-03

1.2E-01 1.8E-01 2.9E-01 1.6E-01 6.6E-02 3.2E-02 1.9E-02 1.3E-02 9.8E-03 7.9E-03 6.9E-03

1.8E-01 2.9E-01 1.0E 75 2.1E-01 7.8E-02 3.6E-02 2.1E-02 1.4E-02 1.0E-02 8.2E-03 7.1E-03

1.2E-01 1.8E-01 2.9E-01 1.6E-01 6.6E-02 3.2E-02 1.9E-02 1.3E-02 9.8E-03 7.9E-03 6.9E-03

0.0 0.0 0.0 0.0 4.0E-02 2.2E-02 1.5E-02 1.1E-02 8.6E-03 7.2E-036.3E-03

1.1E-03 3.3E-03 6.9E-03 1.5E-02 1.8E-02 1.4E-02 1.1E-02 8.6E-03 7.1E-03 6.2E-03 5.5E-03

Figure 2b.   Similar to Fig. 1b except for the decreased shield
length.

```
0.0    3.6E-04 7.0E-04 9.8E-04 1.2E-03 1.3E-03 1.4E-03 1.3E-03 1.2E-03 9.3E-04 6.5E-04

0.0    7.5E-04 1.5E-03 2.0E-03 2.5E-03 2.7E-03 2.8E-03 2.7E-03 2.4E-03 1.9E-03 1.3E-03

0.0    1.2E-03 2.3E-03 3.3E-03 4.0E-03 4.4E-03 4.5E-03 4.3E-03 3.8E-03 3.0E-03 2.1E-03

0.0    1.8E-03 3.4E-03 4.8E-03 5.8E-03 6.3E-03 6.5E-03 6.1E-03 5.4E-03 4.4E-03 3.0E-03

0.0    2.5E-03 4.7E-03 6.6E-03 8.0E-03 8.8E-03 8.9E-03 8.4E-03 7.4E-03 5.9E-03 4.1E-03

0.0    3.4E-03 6.5E-03 9.1E-03 1.1E-02 1.2E-02 1.2E-02 1.1E-02 9.8E-03 7.8E-03 5.4E-03

0.0    4.5E-03 8.8E-03 1.2E-02 1.5E-02 1.6E-02 1.6E-02 1.5E-02 1.3E-02 1.0E-02 7.0E-03

0.0    6.1E-03 1.2E-02 1.7E-02 2.0E-02 2.1E-02 2.1E-02 1.9E-02 1.7E-02 1.3E-02 8.9E-03

0.0    8.0E-03 1.6E-02 2.2E-02 2.7E-02 2.9E-02 2.8E-02 2.5E-02 2.1E-02 1.6E-02 1.1E-02

0.0    1.0E-02 2.1E-02 3.0E-02 3.7E-02 3.8E-02 3.6E-02 3.2E-02 2.6E-02 2.0E-02 1.4E-02

0.0    1.2E-02 2.7E-02 4.2E-02 5.1E-02 5.2E-02 4.7E-02 4.0E-02 3.3E-02 2.4E-02 1.6E-02

0.0    1.0E-02 3.1E-02 6.3E-02 7.4E-02 7.0E-02 6.1E-02 5.0E-02 3.9E-02 2.8E-02 1.9E-02

0.0    0.0     0.0     1.1E-01 1.1E-01 9.4E-02 7.5E-02 5.9E-02 4.5E-02 3.2E-02 2.1E-02

0.0    1.4E-01 2.6E-01 2.2E-01 1.6E-01 1.2E-01 8.8E-02 6.6E-02 4.9E-02 3.5E-02 2.2E-02

0.0    2.3E-01 1.0E 00 3.2E-01 1.9E-01 1.3E-01 9.3E-02 6.9E-02 5.0E-02 3.6E-02 2.3E-02

0.0    1.4E-01 2.6E-01 2.2E-01 1.6E-01 1.2E-01 8.8E-02 6.6E-02 4.9E-02 3.5E-02 2.2E-02

0.0    0.0     0.0     1.1E-01 1.1E-01 9.4E-02 7.5E-02 5.9E-02 4.5E-02 3.2E-02 2.1E-02

0.0    1.0E-02 3.1E-02 6.3E-02 7.4E-02 7.0E-02 6.1E-02 5.0E-02 3.9E-02 2.8E-02 1.9E-02
```

Figure 3a. Similar to Fig. 2a except for decreased shield length.

```
1.8E-04 5.1E-04 7.8E-04 1.0E-03 1.2E-03 1.4E-03 1.4E-03 1.3E-03 1.2E-03 9.9E-04 7.3E-04

3.8E-04 8.5E-04 1.0E-03 1.2E-03 1.4E-03 1.5E-03 1.5E-03 1.5E-03 1.4E-03 1.2E-03 9.6E-04

6.1E-04 1.3E-03 1.4E-03 1.6E-03 1.7E-03 1.8E-03 1.8E-03 1.7E-03 1.6E-03 1.5E-03 1.3E-03

8.9E-04 1.8E-03 1.9E-03 2.1E-03 2.2E-03 2.2E-03 2.2E-03 2.1E-03 2.0E-03 1.9E-03 1.7E-03

1.2E-03 2.5E-03 2.6E-03 2.7E-03 2.8E-03 2.8E-03 2.8E-03 2.7E-03 2.5E-03 2.4E-03 2.3E-03

1.7E-03 3.4E-03 3.5E-03 3.6E-03 3.6E-03 3.6E-03 3.5E-03 3.4E-03 3.2E-03 3.1E-03 2.9E-03

2.3E-03 4.6E-03 4.7E-03 4.8E-03 4.8E-03 4.7E-03 4.5E-03 4.3E-03 4.1E-03 3.9E-03 3.7E-03

3.1E-03 6.2E-03 6.3E-03 6.4E-03 6.4E-03 6.2E-03 5.9E-03 5.6E-03 5.2E-03 5.0E-03 4.7E-03

4.0E-03 8.2E-03 8.5E-03 8.9E-03 8.8E-03 8.4E-03 7.8E-03 7.2E-03 6.6E-03 6.2E-03 5.9E-03

5.1E-03 1.1E-02 1.2E-02 1.3E-02 1.2E-02 1.1E-02 1.0E-02 9.2E-03 8.3E-03 7.6E-03 7.2E-03

5.7E-03 1.3E-02 1.7E-02 2.0E-02 1.9E-02 1.6E-02 1.4E-02 1.2E-02 1.0E-02 9.2E-03 8.6E-03

4.5E-03 1.4E-02 3.2E-02 3.7E-02 2.9E-02 2.3E-02 1.8E-02 1.5E-02 1.2E-02 1.1E-02 9.9E-03

0.0      0.0      0.0      8.2E-02 4.6E-02 3.1E-02 2.3E-02 1.8E-02 1.4E-02 1.2E-02 1.1E-02

1.3E-01 2.0E-01 3.1E-01 1.5E-01 7.0E-02 4.0E-02 2.7E-02 2.0E-02 1.6E-02 1.3E-02 1.2E-02

2.0E-01 3.1E-01 1.0E 75 2.5E-01 8.5E-02 4.5E-02 2.9E-02 2.1E-02 1.6E-02 1.4E-02 1.2E-02

1.3E-01 2.0E-01 3.1E-01 1.5E-01 7.0E-02 4.0E-02 2.7E-02 2.0E-02 1.6E-02 1.3E-02 1.2E-02

0.0      0.0      0.0      8.2E-02 4.6E-02 3.1E-02 2.3E-02 1.8E-02 1.4E-02 1.2E-02 1.1E-02

4.5E-03 1.4E-02 3.2E-02 3.7E-02 2.9E-02 2.3E-02 1.8E-02 1.5E-02 1.2E-02 1.1E-02 9.9E-03
```

Figure 3b.   Similar to Fig. 2b except for decreased shield length.

```
0,0        2.2E-04 4.3E-04 6.1E-04 7.5E-04 8.4E-04 8.7E-04 8.5E-04 7.6E-04 6.2E-04 4.4E-04

0 0        4.6E-04 8.9E-04 1.3E-03 1.6E-03 1.7E-03 1.8E-03 1.8E-03 1.6E-03 1.3E-03 9.0E-04

0 0        7.3E-04 1.4E-03 2.0E-03 2.5E-03 2.8E-03 2.9E-03 2.8E-03 2.5E-03 2.0E-03 1.4E-03

0 0        1.1E-03 2.0E-03 2.9E-03 3.6E-03 4.0E-03 4.2E-03 4.0E-03 3.6E-03 2.9E-03 2.1E-03

0 0        1.4E-03 2.8E-03 4.0E-03 4.9E-03 5.5E-03 5.7E-03 5.5E-03 4.9E-03 4.0E-03 2.8E-03

0 0        1.9E-03 3.7E-03 5.4E-03 6.6E-03 7.5E-03 7.7E-03 7.5E-03 6.6E-03 5.4E-03 3.8E-03

0 0        2.5E-03 4.9E-03 7.1E-03 8.8E-03 1.0E-02 1.0E-02 9.9E-03 8.8E-03 7.1E-03 5.0E-03

0 0        3.2E-03 6.3E-03 9.2E-03 1.2E-02 1.3E-02 1.4E-02 1.3E-02 1.2E-02 9.3E-03 6.4E-03

0 0        3.9E-03 7.9E-03 1.2E-02 1.5E-02 1.8E-02 1.8E-02 1.7E-02 1.5E-02 1.2E-02 8.2E-03

0 0        4.4E-03 9.4E-03 1.5E-02 2.0E-02 2.4E-02 2.4E-02 2.3E-02 1.9E-02 1.5E-02 1.0E-02

0 0        4.4E-03 1.0E-02 1.9E-02 2.8E-02 3.3E-02 3.3E-02 3.0E-02 2.5E-02 1.9E-02 1.3E-02

0 0        3.0E-03 8.1E-03 2.0E-02 4.0E-02 4.7E-02 4.5E-02 3.9E-02 3.1E-02 2.3E-02 1.5E-02

0 0   0.0    0.0    0.0    6.9E-02 7.0E-02 6.0E-02 4.9E-02 3.8E-02 2.8E-02 1.8E-02

0 0        1.3E-01 2.3E-01 1.8E-01 1.3E-01 9.9E-02 7.6E-02 5.8E-02 4.4E-02 3.1E-02 2.0E-02

0 0        2.3E-01 1.0E 00 3.0E-01 1.8E-01 1.2E-01 8.8E-02 6.5E-02 4.8E-02 3.4E-02 2.2E-02

0 0        1.7E-01 3.0E-01 2.5E-01 1.7E-01 1.2E-01 8.9E-02 6.6E-02 4.9E-02 3.4E-02 2.2E-02

0 0        9.6E-02 1.6E-01 1.6E-01 1.4E-01 1.1E-01 8.3E-02 6.3E-02 4.7E-02 3.3E-02 2.15E-02

0 0        5.6E-02 9.5E-02 1.1E-01 1.0E-01 8.7E-02 7.1E-02 5.6E-02 4.2E-02 3.1E-02 2.0E-02

0 0        3.4E-02 6.0E-02 7.3E-02 7.4E-02 6.7E-02 5.8E-02 4.7E-02 3.7E-02 2.7E-02 1.8E-02

0 0        2.2E-02 4.0E-02 5.0E-02 5.3E-02 5.1E-02 4.6E-02 3.9E-02 3.1E-02 2.3E-02 1.5E-02

0 0        1.5E-02 2.7E-02 3.5E-02 3.9E-02 3.9E-02 3.6E-02 3.1E-02 2.5E-02 1.9E-02 1.3E-02

0 0        1.0E-02 1.9E-02 2.5E-02 2.8E-02 2.9E-02 2.7E-02 2.4E-02 2.0E-02 1.5E-02 1.0E-02

0 0        6.9E-03 1.3E-02 1.8E-02 2.0E-02 2.1E-02 2.1E-02 1.9E-02 1.6E-02 1.2E-02 8.3E-03

0 0        4.8E-03 9.2E-03 1.3E-02 1.5E-02 1.6E-02 1.5E-02 1.4E-02 1.2E-02 9.4E-03 6.5E-03
```

Figure 4a.   Similar to Fig. 2a except that one shiled has been
             omitted. Comparison with Fig. 2a shows that the
             addition of a diamagnetic shield reduces the field
             at the track slightly.

```
1,1E-04 3.2E-04 4.8E-04 6.5E-04 7.8E-04 8.7E-04 9.0E-04 8.7E-04 7.9E-04 6.5E-04 4.9E-04

2,3E-04 5.2E-04 6.4E-04 7.7E-04 8.9E-04 9.7E-04 9.9E-04 9.7E-04 8.9E-04 7.8E-04 6.4E-04

3,7E-04 7.7E-04 8.6E-04 9.7E-04 1.1E-03 1.1E-03 1.2E-03 1.1E-03 1.1E-03 9.7E-04 8.7E-04

5,3E-04 1.1E-03 1.2E-03 1.2E-03 1.3E-03 1.4E-03 1.4E-03 1.4E-03 1.3E-03 1.3E-03 1.2E-03

7,3E-04 1.5E-03 1.5E-03 1.6E-03 1.7E-03 1.8E-03 1.8E-03 1.8E-03 1.7E-03 1.6E-03 1.5E-03

9,7E-04 2.0E-03 2.0E-03 2.1E-03 2.2E-03 2.3E-03 2.3E-03 2.2E-03 2.2E-03 2.1E-03 2.0E-03

1,3E-03 2.5E-03 2.6E-03 2.8E-03 2.9E-03 3.0E-03 3.0E-03 2.9E-03 2.8E-03 2.7E-03 2.6E-03

1,6E-03 3.2E-03 3.4E-03 3.6E-03 3.8E-03 3.9E-03 3.9E-03 3.8E-03 3.7E-03 3.5E-03 3.4E-03

1,9E-03 4.0E-03 4.4E-03 4.8E-03 5.2E-03 5.4E-03 5.3E-03 5.0E-03 4.8E-03 4.6E-03 4.4E-03

2,2E-03 4.7E-03 5.5E-03 6.6E-03 7.5E-03 7.6E-03 7.2E-03 6.7E-03 6.2E-03 5.8E-03 5.5E-03

2,1E-03 4.9E-03 7.0E-03 1.0E-02 1.2E-02 1.1E-02 1.0E-02 9.0E-03 8.0E-03 7.3E-03 6.8E-03

1,4E-05 4.3E-03 8.9E-03 2.0E-02 2.2E-02 1.8E-02 1.4E-02 1.2E-02 1.0E-02 9.0E-03 8.2E-03

0,0     0.0     0.0     0.0         4.9E-02 2.8E-02 2.0E-02 1.5E-02 1.2E-02 1.1E-02 9.6E-03

1,2E-01 1.9E-01 3.0E-01 1.7E-01 7.6E-02 3.9E-02 2.5E-02 1.8E-02 1.4E-02 1.2E-02 1.1E-02

2,0E-01 3.0E-01 1.0E 75 2.0E-01 8.4E-02 4.4E-02 2.8E-02 2.0E-02 1.6E-02 1.3E-02 1.2E-02

1,7E-01 1.9E-01 2.3E-01 1.2E-01 6.7E-02 4.1E-02 2.7E-02 2.0E-02 1.6E-02 1.3E-02 1.2E-02

1,0E-01 1.0E-01 9.4E-02 7.1E-02 4.8E-02 3.3E-02 2.4E-02 1.9E-02 1.5E-02 1.3E-02 1.1E-02

5,9E-02 5.8E-02 5.3E-02 4.4E-02 3.4E-02 2.6E-02 2.0E-02 1.6E-02 1.4E-02 1.2E-02 1.1E-02

3,6E-02 3.5E-02 3.3E-02 2.9E-02 2.4E-02 2.0E-02 1.6E-02 1.4E-02 1.2E-02 1.0E-02 9.4E-03

2,3E-02 2.2E-02 2.1E-02 1.9E-02 1.7E-02 1.5E-02 1.3E-02 1.1E-02 9.8E-03 8.8E-03 8.1E-03

1,5E-02 1.5E-02 1.4E-02 1.4E-02 1.3E-02 1.1E-02 1.0E-02 9.0E-6Gddww-03 7.3E-03 6.8E-03

1,0E-02 1.0E-02 1.0E-02 9.7E-03 9.2E-03 8.5E-03 7.8E-03 7.1E-03 6.5E-03 6.0E-03 5.6E-03

7,1E-03 7.1E-03 7.1E-03 7.0E-03 6.8E-03 6.5E-03 6.1E-03 5.6E-03 5.2E-03 4.8E-03 4.5E-03

4,9E-03 5.0E-03 5.1E-03 5.1E-03 5.1E-03 4.9E-03 4.7E-03 4.4E-03 4.1E-03 3.8E-03 3.5E-03
```

Figure 4b.   Similar to Fig. 2b except that one shield has been omitted.   Comparison with Fig. 2b shows that the addition of a diamagnetic shield reduces the field at the track slightly.

FERROMAGNETIC SHIELD

SUPERCONDUCTING SHIELD

TRACK

TRAIN MAGNET
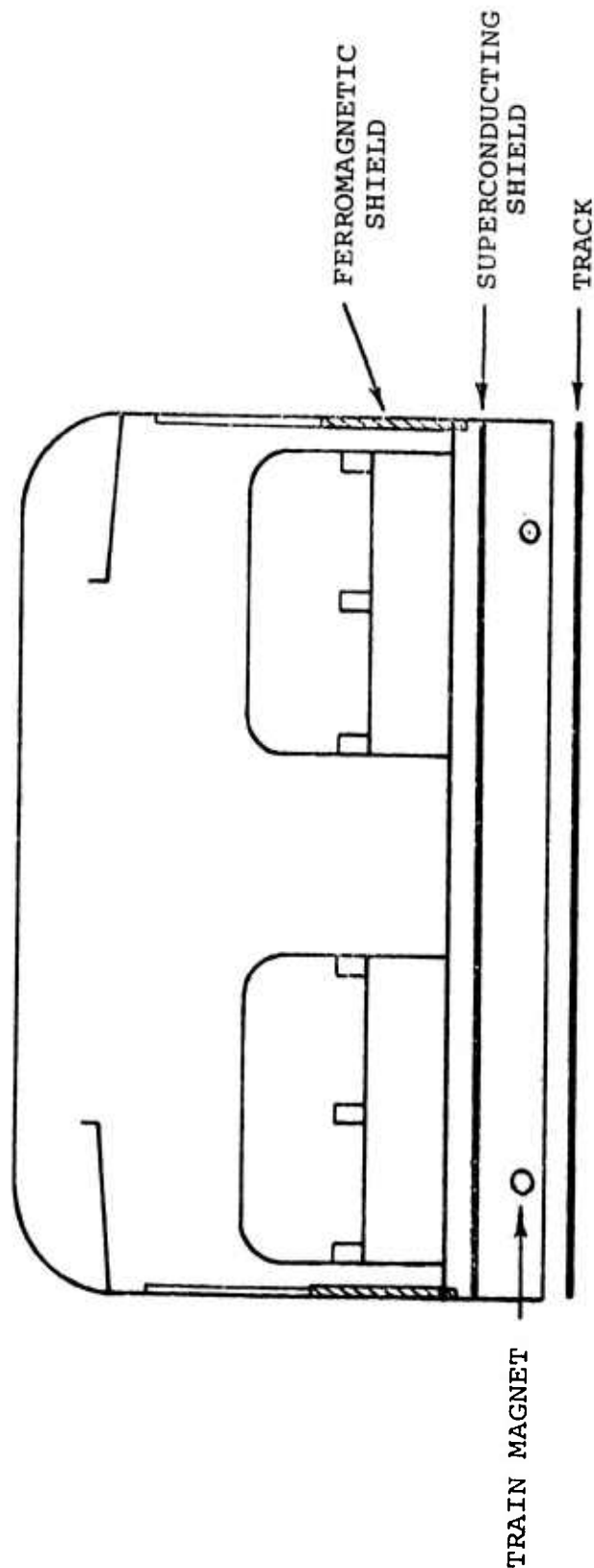
(a)

Figure 5a, b.   Cross sections of two possible magnetic levitation configurations showing location of ferromagnetic and superconducting shields.

(b)

FERROMAGNETIC
SHIELD

SUPERCONDUCTING
SHIELD

TRACK

TRAIN MAGNET

-332-
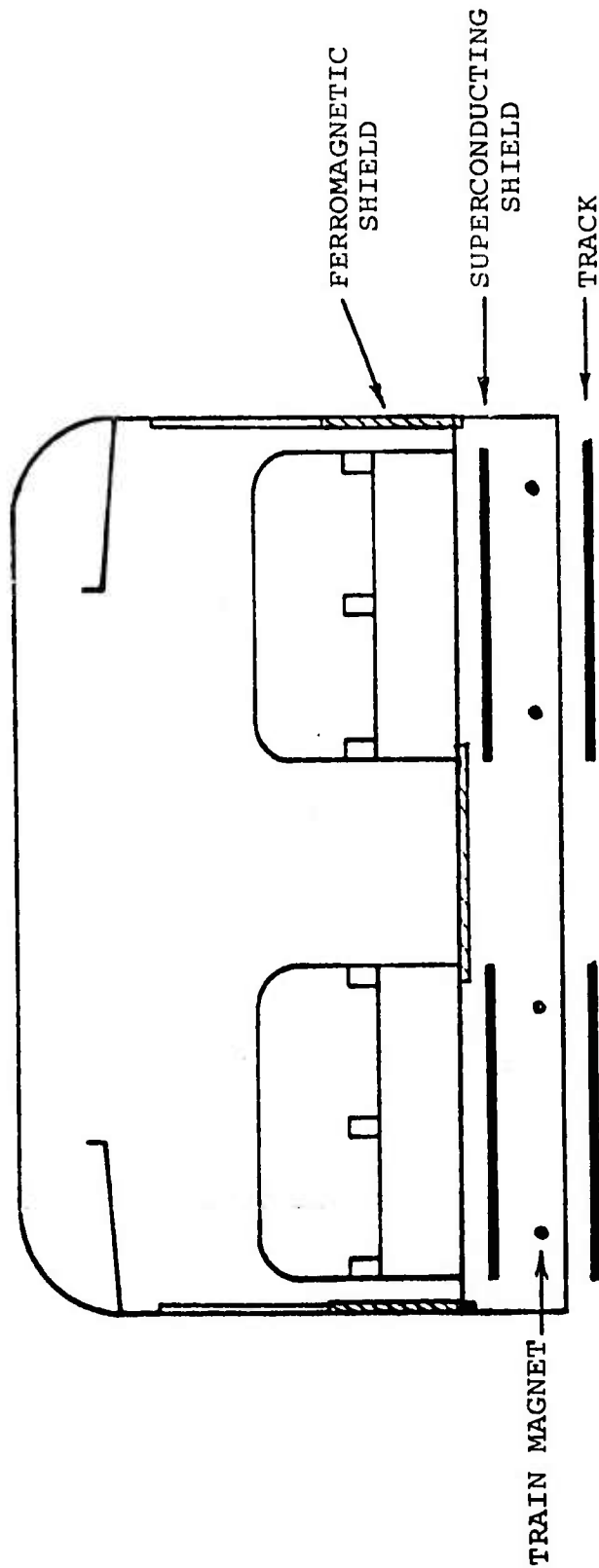
STRUCTURAL CHARACTERIZATION
OF AMORPHOUS SEMICONDUCTORS*

A. Bienenstock

Abstract

Structural characterization studies of a-Ge and a-Si
are reviewed, with particular emphasis on density and void
characterization. Systematic studies show that an ideal material
is approached through high substrate temperatures, low base
pressures, low deposition rates and long substrate to sample dis-
tances. The "ideal" films have densities and refractive indices
close to crystalline values. It is suggested that a large portion
of the apparent discrepancies in existing density, void shape and
refractive index data can be rationalized if it is assumed that
sputtering or high evaporation rates lead to large numbers of
small voids, whereas low evaporation rates lead to smaller numbers
of larger voids. In addition, it is assumed that annealing leads
to void coalescence and that an important role of oxygen is to
decrease the surface mobility of adatoms. Structural studies of
amorphous Ge-Chalcogen films are also reviewed, with the tentative
conclusion that a structure with fourfold coordinated Ge and two-
fold coordinated chalcogens is most appropriate for the germanium
monochalcogenides.

STRUCTURAL CHARACTERIZATION OF
AMORPHOUS SEMICONDUCTORS

A. Bienenstock

I. General Introduction

The structural characterization of amorphous materials
is difficult, as is evidenced by the contrast with the character-
ization of crystalline materials. In the latter, one first
characterizes the ideal periodic structure with a determination
of the short range coordination and, through a knowledge of the
translational symmetry, the long range order. Further character-
ization then involves a determination of the concentration of
various types of defects and impurities as deviations from the
ideal structure.

In the case of amorphous materials, there are no ideal
structures which can be accepted with complete confidence and
which can be used as starting points for the listing of defects
and impurities. Experimental techniques now exist for the de-
termination of short-range order with varying degrees of ambig-
uity - little for elemental materials - large for alloys.
Similarly, experimental techniques exist for the observation of
gross defects like large voids. Yet, when two amorphous samples
of the same composition have different densities, it is only
rarely that we can determine the source of that difference. Some

of it may be due to observable voids, some to unobservable voids, and the remainder to small variations in bond angles and distances which add up to appreciable density differences.

The approach to characterization must, it seems, consist of three parts. The first is the determination of short and intermediate range average atomic arrangements with decreasing ambiguity. The second is the determination of the concentrations of observable defects and impurities. The third is to proceed towards the characterization of an ideal system and the changes of physical properties with known deviations from ideality. In this sense, the ideal structure is conceived of as the state of minimum free energy within the constrained phase space made up of microscopic states which are noncrystalline. The approach towards the ideal system is through preparation techniques which are likely to bring the system into a metastable equilibrium at some temperature with a minimum of defects and impurities.

This review of our attempts at such characterization deals with two model systems. The simplest is that of amorphous Si and Ge. Here, as described previously (Moss, 1973), there is almost universal agreement that the tetrahedral coordination of the crystals is maintained. As he has discussed, the classical dispute between the proponents of the random network and microcrystalline models is active. In this paper we deal with attempts to characterize the grosser defects and to relate them to some physical properties.

-335-

The second system contains the Ge-chalcogen alloys, in which the short-range coordination remains ambiguous, but where results constituting considerable advances may be anticipated at this meeting.

II.   Amorphous Si and Ge

A.   Introduction

Prior to late 1970, the characterization of the physical properties of amorphous Ge and Si was chaotic, with markedly different values having been published for the long wavelength index of refraction, the density, the position and shape of the absorption edge and the conductivity.

In late 1970 and early 1971, the situation changed with the publication of systematic studies of the variation of physical properties with deposition conditions and annealing. Three of special significance, but somewhat conflicting results, are discussed briefly below.

Two quite different films of a-Ge were obtained (Donovan et al. 1970a) by varying the substrate temperature in the deposition. With a base pressure, P_b, of 10^{-8} torr and deposition rates, R_d, in the range 10-50 Å/sec, they found that densities, ρ, of 4.73 ±0.05 and 5.40 ±0.15 g/cm^3 and band edges of 0.6 and 0.7-0.8 eV were obtained with the substrate temperature, T_s, in the range 20 to approximately 250°C and the range 250 to 300°C, respectively. The infrared indices of refraction of the two types of films appeared identical, to within 2.5% to the crystal-

-336-

line value of n = 4.05.

At approximately the same time, systematic changes of various a-Ge physical properties with deposition (Theye 1970) and annealing (Theye 1971) conditions were noted. Starting with films deposited with R_d = 200 Å/sec and P_b = 10^{-7} torr, she found that annealing at successively higher temperatures to 400°C yielded an increase in the absorption edge energy from 0.87 to 1 eV, a decrease in n from 4.37 to 4.24 and a marked increase in the resistivity. All of her films, however, appeared to have $\rho = \rho_0$, the crystalline density (5.35 g/cm^3). Similar increases in the resistivity and edge energy, as well as decreases in n were obtained by increasing the substrate temperature from 25 to 300°C. A decrease in n was also noted when R_d was decreased to 2 Å/sec.

Thus, although these authors did not agree on values of measured parameters, their collective works brought forth the following hypotheses:

1. The physical properties of a-Ge films are strongly dependent on the deposition conditions. Hence, the large differences in previously reported physical properties are due, in large part, to differences in such conditions.

2. Films in a state close to some ideal might be achieved through systematic variation of the deposition conditions. One characerisitic of this ideal is a density equal to or greater than that of crystalline Ge.

3. Since the differences in the films depend on differences in deposition conditions, they are due, in turn, to differing

-337-

structural-chemical states of the films which might be charater-izable. Hence, it might be possible to relate the differences in physical properties to different structural-chemical states.

These three hypotheses, plus others to be introduced below, have governed much of the experimental work in the past three years. Since much of the work depends on experimental attempts to approach ideality, let me summarize the rules governing those approaches.

B. Deposition and Annealing Conditions

To approach "ideality", the substrate temperatures must be sufficiently high so that the atoms have sufficient mobility to rearrange after their initial attachment to the film. Unfortu-nately, T_s is upper bounded by the crystallization temperature.

The deposition rate must be very low so that atoms can take advantage of high surface mobilities to rearrange before they are covered by other atoms. In this case, however, there is a trade-off because impurity atoms yielding the base pressure of the deposition system compete for attachment to the film.

In order to have high purity films, therefore, the base pressure of the deposition system must be low and the sources must be very pure.

In addition, the substrate to sample distance must be large in order that most of the atoms reach the film along a perpendicular approach path (see, e.g., Orlowski and Spicer 1972). Otherwise, hills in the film shadow valleys.

Less explored, but believed by some to be importance, is the nature of the vapor produced in the vaporization process.

-338-

Some believe that surface rearrangement is less possible with a molecular vapor than with an atomic vapor. Although no systematic study of this has been, to my knowledge, published, the high resistivity of films produced by glow discharge of silane (Le Comber et al. 1973) that molecules containing up to six or seven atoms form a significant portion of the vapor.

Finally, post-deposition annealing might be employed. There is little evidence, however, that such annealing has an appreciable effect on films produced at high T_s, low P_b and low R_d.

C. The Void Hypothesis

A significant structural characterization variable was introduced through the observation (Moss and Graczyk 1969) of intense electron small-angle scattering (SAS) from an a-Si film with $\rho = 0.85\rho_o$, as determined (Moss and Graczyk 1970) by radial distribution function (rdf) techniques. These authors attributed the SAS to voids whose diameter, I estimate from the lowest angle of measurement, must be in the size range of 7 to 20 $\overset{o}{A}$ in diameter. From these measurements, the following conclusions might be drawn:

1. Voids in this size range account for an appreciable fraction of the amorphous density deficit, since the density determined by excluding the SAS from the rdf is ρ_o.

2. The density of void-free a-Si is approximately ρ_o.

3. Annealing the film at 400°C yielded an increase of 7% in ρ and a corresponding decrease in the SAS without crystallization.

-339-

This result is in apparent contradiction to a-Ge annealing studies (Theye 1971).

4. Previously observed electron spin resonance signals (Brodsky and Title 1969) could be attributed to unpaired spins associated with atoms on the void surfaces. As a result, the spin resonance signal became a more meaningful characterization tool.

D. The Densities of "Ideal" a-Si and a-Ge

With this general introduction to the hypotheses which arose in late 1970 and early 1971, let me provide a constricted review of papers published since and questions to be discussed at this meeting, concentrating first on the conceptually simplest property, the density.

The densities of sputtered a-Ge films have been shown (Paul et al. 1973) to vary very slowly from 4.85 to 4.90 g/cm³ with T_S in the range 25 to 150°C. They then increase sharply to 5.14 at $T_S = 300°C$ and increase only to 5.18 at $T_S = 350°C$. This result corroborates and extends that of Donovan et al. (1970a), although the upper limit for ρ/ρ_0 is 0.97 in the more recent work and 1.01 in the earlier. Nevertheless, Paul et al. have shown that at least one of their high density films contains 1.8 at.% Ar, so that the figure 0.97 is a lower limit to the actual density of the a-Ge, which is probably closer to $0.99\rho_0$. The unusual temperature dependence observed by both groups is consistent with a model (Paul et al. 1973) which involves the probability that an adatom configuration can change to a configuration of lower local

-340-

free energy in the time for the deposition of a monolayer. These authors also measure n and the room temperature resistivity and show that their variations with T_s are also consistent with the model.

Corroboration of the high ρ/ρ_0 for the ideal amorphous films is also obtained from plots of a-Si density, measured by alpha-particle back scattering, versus spin concentration (Brodsky et al. 1972). The measured ρ/ρ_0 range is from 0.89 to 0.97, with the highest values obtained through high T_s and low R_d. Their values of ρ_0/ρ for thin films vary linearly with spin concentration to an extrapolated value of $\rho/\rho_0 = 1.01$ at zero spin concentration, which might be taken as a measure of ideality.

These values of ρ/ρ_0 are significantly higher than the 0.97 predicted by the Polk (Polk 1971) model. As explanation of this has not, to my knowledge, been provided. Nevertheless, they now seem reasonably well established.

E. Characterization of the Voids and their Contributions
    to the Density

Let me now turn to two interrelated questions which have been the subject of much study:

1. Can observable voids account for the density deficits in low density films?

2. What are the sizes and shapes of the voids?

The observation (Moss and Graczyk 1969) that SAS scattering from voids accounts for the density deficits in their a-Si films is not corroborated by later a-Ge studies.

-341-

SAS measurements and bulk density measurements have been performed (Shevchik and Paul 1972) on evaporated, sputtered and electrolytic films with densities 10, 11 and 6% less than ρ_0. While the SAS contribution to the rdf determined density can account for approximately 1/2 of the evaporated film density deficit, that contribution is much too small (an an order of magnitude) to account for the deficit in the sputtered and electrolytic films.

A detailed study of the variation of SAS intensity with sample angle and scattering angle (Cargill 1972) of an evaporated film (with ρ/ρ_0 probably equal to 0.9) indicates that the voids have pipelike shapes with axes perpendicular to the surface of the film and that the void volume can only account for about 10% of the excess volume. He also notes that the observed SAS intensity is of the same order of magnitude as that observed by Moss and Graczyk (Moss and Graczyk 1969) and suggests that they overestimated the void volume as a result of their assumption that the voids are spherical. His conclusion about the density deficit rests strongly on his determination of the void volume which, in turn, rests strongly on his assumption that a principal axis of the ellipsoids used to describe the voids is perpendicular to the film surface. This assumption may possibly be in error, as is discussed below.

Direct comparison (Temkin et al. 1973) of the SAS intensity from low and high density sputtered a-Ge films indicates that voids in the radius range from 7 to 250 Å can account for only 1/10 of

-342-

the density difference. An important result of this work is that bulk and rdf density measurements agree. Hence, there is good reason to believe that the structural defect causing the density deficit is contained within the X-ray scattering data, but not within the SAS intensity associated with voids in that size range. They also show that larger voids cannot be observed by SEM with 250 $\overset{\circ}{A}$ resolution and present strong evidence from rdf first neighbor peak areas that the deficit is contained in voids with radii smaller than 7 $\overset{\circ}{A}$.

On the other hand, optical measurements (Galeener 1971a-b, Bauer et al. 1972) on evaporated low density a-Ge films (Donovan et al. 1970b) indicate the presence of a connected network of microcracks about 6 $\overset{\circ}{A}$ in width and about 120 $\overset{\circ}{A}$ in length separating much larger islands of void-free material. Transmission electron microscopy studies (Donovan and Heinemann 1971) appear to confirm the existence of such cracks in low density films and their absence in high density films.

The validity of this conclusion is called into question, however, by the observation (Hauser and Staudinger 1973) of similar electron microscopy contrast in sputtered a-Ge films. As a result of stereoscopic studies, they attribute the contrast to a network of pipelike voids of the shape determined from SAS studies (Cargill 1972) which run between the two film surfaces, but which are not normal to the surfaces (possibly negating one of Cargill's basic assumptions). They also find that the contrast is present in high T_s films, but do not believe that their results contradict

those of Donovan and Heinemann because "These differences can again be ascribed to the different deposition processes, sputtering versus evaporation: it is a well-known fact that the atomic mobility is much lower in a sputtered film than in an evaporated film, principally because of the presence of argon during sputtering."

F. Partial Resolution of the Apparent Contradictions

Faced with apparent contradictions in apparently high quality work, it seems reasonable to assume that:

1. Difference deposition procedures are yielding different size and shape voids with different concentrations.
2. Oxygen and/or other impurities are playing significant roles in some measurements.

The direct influence of oxygen on the optical properties of their sputtered a-Ge films has been shown (Connell et al. 1973) to be too small in samples whose oxygen content has been well characterized (Paul et al. 1973). This result may be taken to imply that the influence is also small on the properties of other (Theye 1970, 1971 and Donovan et al. 1970a, 1970b) films. On the other hand, it seems extremely likely that the presence of even a small oxygen concentration can retard the surface rearrangement and alter the void characteristics and density. Evidence for this is contained in the lower density of a-Ge films obtained through increased base pressure (Donovan et al. 1970a).

With this in mind, let me return to the first hypothesis. First, there is direct evidence from studies in which the same experiments are employed by the same workers on different samples that different deposition procedures are yielding different sizes

-344-

and shapes. This is contained in the observation (Shevchik and Paul 1972) that SAS determined void volume could account for 1/2 the density deficit in an evaporated film, but not in a sputtered or electrolytic film.

It is also implicitly contained in the studies of ρ_o/ρ versus spin concentration (Brodsky et al. 1972). The linear relation, observed for thin films, applies only if the spin density (void surface, presumably) to void volume remains relatively constant with different densities. They find significant deviations for thick films prepared at low T_s and high R_d, which may be interpreted as a much higher spin density than the density deficit would lead one to expect. Such a spin density could result from a higher concentration of small voids in the thick films.

If we assume that, for a given T_s, sputtering or high R_d lead to smaller voids than does evaporation at low R_d, some semblance of order can be placed on the measurements described above.

First, low R_d evaporated a-Ge films of high and low density (Donovan et al. 1970a) yield n approximately equal to n_o, the crystalline value, whereas high R_d samples yield (Theye 1970, 1971) higher values of n which decrease with annealing or increasing T_s. The high values of n have been confirmed on sputtered a-Ge films (Connell et al. 1973) of low density. Since they are not likely to result from the void volume, it seems wise to assume (Connell and Paul 1972) that they result from the unpaired electrons leading also to the spin resonance signals. If the Donovan et al. low density samples consist of islands of high density a-Ge surrounded

-345-

by cracks, as is indicated by the electron microscopy studies (Donovan and Heinemann 1971), one would expect their films to yield n approximately equal to n_o, since it is the islands which are likely to contribute to the fringes. The achievement of such an atomic arrangement might result from the unique combination of high vacuum, and consequent low oxygen concentration, with the low R_d employed by them. The Theye samples, on the other hand, are more likely to have small voids distributed homogeneously, leading to a high average n, as a result of the high R_d.

If this is the case, one must also explain the constant high density observed (Theye 1971) in the annealing experiments on high R_d a-Ge samples evaporated on room temperature substrates. From her evaporation techniques, one would expect ρ to be considerably lower than ρ_o. Yet, the densities of her films, as observed by total X-ray reflection fringes, are all equal to that of a crystallized sample so measured. This apparent contradiction might be explained if one assumed that the effect of annealing was to coalesce the voids, but not eliminate them. The coalescense is likely (Chaudhari 1973 and Turnbull 1973) to take place at surface oxide layers. This would lead to a decrease in n without an increase in ρ. Crystallization might be accompanied by a massive increase in the average void size, leading to the observation of n_o, but with the density maintained. This absence of a significant density change with annealing has also been observed (Brodsky and Stiles 1970) in a-Si films whose spin density did decrease as a result of such annealing, leading to the same conclusion.

-346-

The hypothesis that annealing leads to coalescence of voids is also consistent with the apparent densification through annealing at 400°C films as observed by SAS (Moss and Graczyk 1969). They were not able to observe SAS from voids larger than about 20 Å in diameter with their experimental technique, as discussed above. Hence, it is possible that such voids existed but were unobservable so that the density change upon annealing was unobservable.

This hypothesis that sputtering or high R_d evaporation leads to smaller voids than does low R_d evaporation can be used to rationalize a great deal of the apparent discrepanices in void characterization discussed above. It would seem best, however, to check the hypothesis through direct experiments, such as SAS or spin density-mass density correlations.

III.   Structural Studies of Ge-Chalcogen Alloys

In the preceding sections dealing with amorphous silicon and germanium, no discussion has been presented of the local atomic arrangement for two reasons. The first is that such discussions will have been present elsewhere (Moss 1973). The second is that the rdf's presented thus far lead to the relatively unambiguous conclusion that almost every germanium or silicon is fourfold coordinated. This lack of ambiguity disappears as soon as one deals with alloy amorphous films and is shown most strikingly by the work performed thus far on the germanium-chalcogen alloys discussed below.

The basic problem here is that X-ray or electron diffraction is primarily sensitive to the electron pair correlation function or, in other words, the probability of finding two electrons separated by a distance r. Straightforward Fourier transformation of the diffraction data would yield this pair correlation function directly. Usually, however, one manipulates the data so that the electrons associated with each atom are made to appear concentrated at the nuclear site. As a result, one might think of the radial distribution as the probability of finding nuclei separated by a distance r, with the probability weighted by the product of the atomic numbers. In an A-B alloy or compound, it is often impossible to distinguish A-A, A-B and B-B pairs in the pair correlation function. As a result ambiguities arise.

In many simple compounds, like SiO_2, the ambiguity is negligible. This results from the fact that the apparent short range coordination as revealed by the radial distribution function is very similar to that shown in crystalline compounds. It is, therefore, relatively simple to identify peaks in the radial distribution function with different types of interatomic pairs.

Such an identification is not possible, however, in rdf's of amorphous alloys near the composition of, or of compounds with the composition of GeS (Rowland et al. 1972), GeSe (Fawcett et al. 1972, Bienenstock et al. 1973) and GeTe (Betts et al. 1970, Dove et al. 1970). The nearest neighbor distances in these materials are 0.2-0.3 Å less than, and the average coordination numbers

appear to be approximately one-half of, those in the corresponding crystals. Hence, the short range coordinations of the amorphous and crystalline materials are markedly different, so that the crystal structures provide little guidance in the determination of the amorphous structures.

In this case, one is truly faced with the ambiguity discussed above. For example, crystalline Ge-Ge and Te-Te near neighbor distances are 2.45 and 2.86 Å, respectively. The sum of the Pauling single bond covalent radii of Ge and Te is 2.59 Å. All such distances could have contributed to the broad nearest neighbor radial distribution peaks observed for a-GeTe films.

Two possible coordination pictures have been put forth (Betts et al. 1970) to explain the GeTe rdf. In the first, each Ge and Te is threefold coordinated by atoms of the opposite atomic species. Such a coordination might be associated with crystalline GeS and GeSe and would be analogous to the fourfold coordination in III-IV and II-VI compounds.

In the second model, each Ge is fourfold and each Te is twofold coordinated in keeping with the 8-N rule for covalent bonding of the individual atoms. In the calculation of rdf first neighbor peak areas, no chemical ordering was assumed. It was, however, indicated that the calculated area is only weakly sensitive to such ordering.

These two models have competed for acceptance despite efforts to distinguish between them. At first, it appeared as if the nearest neighbor distances were quite appropriate for the

fourfold-twofold model, but too short for the threefold coordinated model. Short distances in an amorphous threefold bonding scheme have, however, been rationalized (Bienenstock 1973).

ESCA results (Betts et al. 1972a) appeared to indicate that Ge core shifts in crystalline GeTe are significantly larger than in amorphous, indicating the applicability of the fourfold-twofold, rather than threefold, coordination. These results have been shown to be incorrect and to have been caused by the surface oxidation of the crystalline GeTe (Shevchik et al. 1973, Fisher et al. 1973).

The fourfold-twofold model appeared, initially, to have the virtue of simply explaining (Betts et al. 1970) the first neighbor peak areas of a number of Ge_xTe_{1-x} alloys. It was subsequently shown (Betts et al. 1972b) that these areas could be equally well fit, for $x > 1/3$, by a phase-separated model in which one component is a threefold coordinated GeTe. Such phase separation would be expected (Bienenstock 1973) if the GeTe were so coordinated. Similar results obtain for the Ge_xSe_{1-x} system. It is also shown in that work, however, that the difference in calculated areas for the compound GeS is quite large. Since this is the case, let me review some of the structural work on the Ge_xS_{1-x} alloys.

Radial distribution functions were obtained (Rowland et al. (1972) on samples with $x = 0.33$ and 0.42. With these two radial distributions, an artificial radial distribution was constructed for GeS by assuming that the $x = 0.42$ sample was phase-separated

-350-

into GeS_2 and GeS. The resulting first neighbor peak area (with the sample normalized so that its composition is $Ge_{.5}S_{.5}$) is approximately 2000 electrons^2. This is completely inconsistent with threefold coordination, for which the area would be 1536 electrons^2, but is consistent with fourfold coordination of Ge by S, for which the area would be 2048 electrons^2. If we assume, however, that the coordination in amorphous GeS is similar to that in GeSe and GeTe, then such fourfold coordination is extremely unlikely since it is inconsistent with the areas observed for these materials. The area is also consistent, though, with models based on fourfold Ge and twofold S coordination. Hence, this work provides some support for the fourfold-twofold coordination model and seems inconsistent with the threefold coordinated model.

Similar tentative support for the fourfold-twofold co-ordinated model is provided by a preliminary rdf (Verhelle and Bienenstock 1973) on a sputtered film of GeS. The first neighbor peak area is essentially the same as that determined in the earlier work. It is my understanding from the abstract that similar results have been obtained elsewhere (Cervinka and Hruby 1973).

As indicated previously, one would anticipate phase separa-ration if the threefold coordinated model is appropriate. Tenta-tive electron microscopy observations (Feltz et al. 1972) of such separation in a Ge-Se alloy were put forth, but then withdrawn (Feltz et al. 1973). The latter work suggests that replica studies are likely to yield spurious effects which may be misinterpreted as phase separation. Unfortunately, it has also been shown

-351-

Bienenstock (1973) that little contrast is likely to be observed from such separation if it exists in Ge_xTe_{1-x} or Ge_xSe_{1-x} alloys, so that such studies are not too promising.

At this meeting we are likely to see this problem resolved through X-ray absorption edge fine structure (Sayres et al. 1973), photoemission spectroscopy (Fisher et al. 1973, Shevchik et al. 1973), as well as infrared and Raman spectroscopy (Fisher and Tauc 1973, Lucovsky 1973) studies. It is my understanding that the fourfold-twofold coordinated model is appropriate and that the Ge is either fourfold coordinated by chalcogens or by Ge.

IV.   Acknowledgements

References

Barna, A., Barna, P. B., Bodo, Z., Pocza, J. F., Pozgai, I., and Randoczi, G., 1973, Proceedings of The Fifth International Conference on Amorphous and Liquid Semiconductors, Garmisch-Partenkirchen, Germany, September, 1973.

Bauer, R. S., Galeener, F. L., and Spicer, W. E., 1972, Journal of Non-Crystalline Solids, 8-10, 196.

Betts, F., Bienenstock, A., and Ovshinsky, S. R., 1970, Journal of Non-Crystalline Solids, 4, 554.

Betts, F., Bienenstock, A., and Bates, C. W., Jr., 1972a, Journal of Non-Crystalline Solids, 8-10, 364.

Betts, F., Bienenstock, A., Keating, D., and deNeufville, J. P., 1972b, Journal of Non-Crystalline Solids, 7, 417.

Bienenstock, A., 1973, Journal of Non-Crystalline Solids, 11, 447.

Bienenstock, A., Mortyn, F., Narashimhan, S., and Rowland, S. C., 1973, unpublished work.

Brodsky, M. H., Kaplan, D., and Ziegler, J. F., 1972, Appl. Phys. Letters, 21, 305.

Brodsky, M. H., and Stiles, P. J., 1970, Phys. Rev. Letters, 25, 798.

Brodsky, M. H., and Title, R. S., 1969, Phys. Rev. Letters, 23, 581.

Cargill, G. S. III, 1972, Phys. Rev. Letters, 28, 1372.

Cervinka, L., and Hruby, A., 1973, Proceedings of the Fifth International Conference on Amorphous and Liquid Semiconductors, Garmisch-Partenkirchen, Germany, September 1973.

Chaudhari, P., 1973, private communication.

Connell, G. A. N., and Paul, W., 1972, Journal of Non-Crystalline Solids, 8-10, 215.

Connell, G. A. N., Temkin, R. J., and Paul, W., 1973, Proceedings of the Fifth International Conference on Amorphous and Liquid Semiconductors, Garmisch-Partenkirchen, Germany, September 1973, and unpublished work.

Donovan, T. M., Ashley, E. J., and Spicer, W. E., 1970a, Phys. Letters, 32A, 85.

Donovan, T. M., Spicer, W. E., Bennett, J. M., and Ashley, E. J., 1970b, Phys. Rev. B, 2, 397.

Donovan, T. M., and Heinemann, K., 1971, Phys. Rev. Letters, 27, 1794.

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

Fawcett, R. W., Wagner, C. N. J., and Cargill, G. S. III, 1972, J. Non-Crystalline Solids, 8-10, 369.

Feltz, A., Buttner, J. H., Lippmann, F. J., and Maul, W., 1972, J. Non-Crystalline Solids, 8-10, 64.

Feltz, A., Lippmann, F. J., and Zacharias, A., 1973, J. Non-Crystalline Solids, 12, 129.

Fisher, G. B., Lindau, I., Spicer, W. E., Verhelle, A. and Weaver, H. E., 1973, Proceedings of the Fifth International Conference on Amorphous and Liquid Semiconductors, Garmisch-Partenkirchen, Germany, September 1973.

Fisher, G. B., and Tauc, J., 1973, ibid.

Galeener, F. L., 1971a, Phys. Rev. Letters, 27, 421.

Galeener, F. L., 1971b, Phys. Rev. Letters, 27, 1716.

Hauser, J. J., and Staudinger, A., 1973, Phys. Rev. B, 8, 607.

LeComber, P. G., Loveland, R. J., Spear, W. E., and Vaughan, R. A., 1973, Proceedings of the Fifth International Conference on Amorphous and Liquid Semiconductors, Garmisch-Partenkirchen, Germany, September 1973.

Moss, S. C., 1973, ibid.

Moss, S. C., and Graczyk, J. F., 1969, Phys. Rev. Letters 23, 1167.

Moss, S. C., and Graczyk, J. F., 1970, Proceedings of the Tenth International Conference on the Physics of Semiconductors, edited by Keller, S. P., Hensel, J. S., and Stern, F. (U.S. Atomic Energy Commission, Division of Technical Information), p. 658.

Orlowski, B. A., and Spicer, W. E., 1972, Mat. Res. Bull. 7, 793.

Paul, W., Connell, G. A. N, and Temkin, R. J., 1973, Proceedings of the Fifth International Conference on Amorphous and Liquid Semiconductors, Garmisch-Partenkirchen, Germany, September 1973, and unpublished work.

Polk, D. E., 1971, J. Non-Crystalline Solids, $\underline{5}$, 347.

Rowland, S. C., Narasimhan, S., and Bienenstock, A., 1972, J. Appl. Phys., $\underline{43}$, 2741.

Sayers, D. E., Lytle, F. W., and Stern, E. A., Proceedings of the Fifth International Conference on Amorphous and Liquid Semiconductors, Garmisch-Partenkirchen, Germany, September 1973.

Shevchik, N. J., and Paul, W., 1972, J. Non-Crystalline Solids, $\underline{8-10}$, 381.

Shevchik, N. J., Tejada, J., and Cardona, M., 1973, Proceedings of the Fifth International Conference on Amorphous and Liquid Semiconductors, Garmisch-Partenkirchen, Germany, September 1973.

Temkin, R. J., Paul, W., and Connell, G. A. N., 1973, ibid and unpublished work.

Theye, M. L., 1970, Optic Communications $\underline{2}$, 329.

Theye, M. L., 1971, Mat. Res. Bull., $\underline{6}$, 103.

Turnbull, D., 1973, private communication.

Verhelle, A., and Bienenstock, A., 1973, unpublished work.

DECISION ANALYSIS APPLIED TO PENETRANT
INSPECTION OF CERAMIC TURBINE BLADES

F. A. McClintock, R. L. Coble & R. M. Thomson

Abstract

Decision analysis forces the analyst to express his hunches
and insight in a quantitative way. As regards inspection of
brittle materials, the very strong dependence of crack growth
rate on crack length means that only a very narrow range of
cracks need be detected in an inspection scheme. Longer cracks
must be removed or prevented in manufacture. Shorter cracks are
of no concern. When no hazard to human life is anticipated,
even the high cost of failure in terms of reputation and further
sales was not enough to justify penetrant inspection of gas
turbine blades. This conclusion seemed to hold in spite of a
wide variability in expert opinion about the frequency of internal
versus external cracks or about the distribution of crack lengths
themselves.

DECISION ANALYSIS APPLIED TO PENETRANT
INSPECTION OF CERAMIC TURBINE BLADES

F. A. McClintock, R. L. Coble & R. M. Thomson

I.   Introduction

The decision to inspect parts as a means of preventing
failure is often undertaken on an intuitive basis, just as is
the choice of the working stress or factor of safety. Decision
analysis is a mathematical tool for bringing to a common focus
the various factors and hypotheses which enter into such a
decision. This tool has been developed for management decisions,
but it should also be useful in connection with assuring high
reliability in mechanical parts. Here its main advantage is in
providing a means for describing and using estimates of the very
low probabilities of failure that are desired in systems of high
reliability.

II.   Introduction to Decision Analysis

The discussion given here follows the concise introduction
given by Howard (1968). Decision analysis is of primary value
for problems that involve many variables, each with a degree of
uncertainty, and which are changing with time. The example given
here is simplified in that it contains relatively few variables
and is rather static in time, although the degree of uncertainty
is very important.

The essential steps in a decision analysis are illustrated in Fig. 1. The first step is to define the structure of the problem by listing the decision variables (those under the manager's control), the state variables (either deterministic or probabilistic and not under the manager's control), and the relations between the decision and state variables. The structure takes on the form of a tree which develops the various possible outcomes.

The second step of Fig. 1 is assigning probabilities to the various paths. These probabilities are assigned in the subjective sense as the manager's estimates, rather than in the objective sense as the result of repeated experimentation in the actual world.

The third step of Fig. 1 is estimating the value of reducing the uncertainty of the estimates by obtaining further information (e.g., by experimentation or other research).

Not shown in Fig. 1 is the coding of the risk-preference, which gives a measure of whether the decision should be based on the average value or cost of an action or whether, for example, because of limited resources, it is important to reduce the risk of large losses even more than their average cost would indicate. There may also be a corresponding time preference for money beyond that dictated by normal interest rates. Once these steps have been carried out, the problem is most conveniently solved with the aid of a computer program which keeps track of the various possible series of events and the values of each.

III.   The Decision Structure for Dye Penetrant Inspection of

Ceramic Turbine Blades

As an illustrative example, consider the question as to
whether or how often to use dye penetrant inspection to examine
the blades of a 5000 kilowatt turbine for a fleet of 100 patrol
craft. The fleet is to be put in service in three years and will
acquire 10,000 hours of operating experience in the following
three year period. Preliminary calculations indicated that cracks
would not grow to critical size if they had been undetectable on
first inspection. Therefore, the problem taken was whether or
not to apply dye penetrant crack detection at the time of manu-
facture.

The variability in the detectable length of cracks was
estimated to be much smaller than the variability in lengths of
the prior existing cracks, so it was assumed that the crack de-
tection methods were perfect once the cracks reached a given size.
The dye penetrant method only allows detection of surface cracks.
From any given surface indication there may be a variety of crack
depths, denoted by the ratio of depth d to length ℓ of surface
indication. From these two dimensions it was assumed that an
effective crack length could be calculated and the time to failure
estimated exactly. If failure occurred, there would be loss of
the engine, loss of availability of the boat for some time, and
a loss of manufacturer's future business. These considerations
led to the decision tree in Fig. 2, which will now be discussed
in detail.

In accordance with the needs of the computer program, the nodes or branch points of the three are given names. The first node is where the decision is made whether or not to use dye penetrant. It is given the name DYE. This decision leads to either of two chance nodes YD (yes-dye) and ND (no-dye). Consider first the YD branch. There are two possibilities. The worst crack in the turbine many be an internal crack, IC, in which case it cannot be detected by dye penetrant, or it may be an external crack, EC, in which case it may be detected. Note that subsequent nodes carry the letters of all preceding nodes (except the first) to provide a unique name for each node.

At the next level of branching, some of the internal cracks are long enough so that in the 10,000 hour life they will suffer field failures, FF; others will have cracks short enough to survive in the field, FS. Turning to the external cracks, there will be a distribution of lengths of indications, L_i. As a compromise between precision and convenience, consider just five possible crack lengths. At each of these nodes there is the chance matter of the depth to length ratio of the crack, d/ℓ, leading to five more nodes, D_j, which are terminal nodes (T).

A similar structure of the possibilities of cracks and field failures occurs if no dye penetrant inspection is carried out; the only difference is that external as well as internal cracks will lead to field failures if they are long enough.

This completes the structure phase of the decision analysis indicated by the first block in Fig. 1. Note that

simplifying approximations have been made where possible. One
that has not been mentioned is that it has been assumed that the
reliability is high enough so that the change of having a second
field failure, in a replacement turbine, is negligible within the
design lifetime. For a preliminary study, the structure should
be simple; the subsequent results will show which areas need
better modeling or further study, and which are not important in
reaching a decision.

IV.   Assignment of Probabilities

Probabilities at the chance nodes were assigned by inter-
viewing seventeen workers in the field who were representatives
of turbine manufacturers, the government, and universities. All
had been selected for their interest in the reliability of brittle
parts for ceramic gas turbines. The participants were first asked
to estimate the probability that the worst crack would be internal
in a turbine with the proposed polycrystalline silicon nitride
blades. Time did not allow the recommended procedure of individual
interviews, but we did make a preliminary survey, report the re-
sults, ask those with differing opinions to state their views and
the background information from which they made their decisions,
and then repeat the survey. The raw results are given in Table 1.
First consider the first pair of columns, giving the expected
probabilities of worst flaws being internal and therefore not
detectable by dye penetrant techniques. The range of estimated
probabilities was very broad. It also changed very little from

before to after the discussion, indicating at least that this was a group of independent-minded people! The results left us little alternative but to run the problem with two assumptions: the probability that the worst crack was internal was chosen to be .1 in one case and .7 in the other. This possibility of conveniently making runs for alternative hypotheses is of course one of the strengths of using a computer program.

The next set of nodes for which expert opinion was solicited was that associated with the length of the worst indication arising in the penetrant inspection. Participants were asked to estimate the total length in microns of the longest surface crack length which would be encountered in 10%, 1% or .1% of the turbines. The results are given in the last three columns of Table 1 and again indicate wide disagreement. For convenience, the authors gathered the probabilities into three groups, added a fourth, and extrapolated them to crack lengths typical of the worst to be encountered in 90% of the turbines and to those extremely rare ones which would be encountered in only 10^{-4} of the turbines. The probabilities were chosen to sum to unity, giving a discrete approximation to what is in reality a continuous distribution curve. The resulting distributions or probabilities, which were used in the analysis, are listed in Table 2. The distributions of depth ratios assumed by the authors are also given in Table 2.

The possibilities of field failures arising from internal cracks were chosen by calculating the critical crack length, esti-

-363-

mating the crack growth during the 10,000 hour service period and selecting probabilities of failure that would correspond to those for external cracks, taking into account the somewhat greater seriousness of external cracks of the same depth. Thus in the four different ranges of probability which were to be considered, the probabilities of field failures were taken to be 10^{-2}, 3×10^{-3}, 3×10^{-4}, or 3×10^{-5}.

The rewards (or values, which are the negative of costs) were assigned to the various nodes and branches by the authors. First, the cost of inspection of 200 blades from 100 turbines, before assembly during manufacture, was taken to be $1.00 each or $20,000. Next, the cost of a field failure was taken to be the value of the turbine at 5,000 kilowatt $\times$ $50.00 per kilowatt, plus a loss of availability of the craft for one month at a charge of 20% per annum of the $5 million cost of the craft. In addition, there was two week's loss of crew time at $1,000 per man-month. Finally, it was postulated that five failures would cost the profit on three similar orders of 100 turbines each. These orders would be assumed lost if the company incurred a bad reputation from five field failures in these turbines. Charging one fifth of this cost to each field failure gives a cost of $1.5 million, which is by far the dominant cost. In the actual problem the total cost of field failure was then about $2 million. Finally, the cost before assembly of discarding a blade with a bad penetrant indication was taken to be $20.

Although not part of the usual decision analysis calcu-
lations, it turns out to be convenient to carry along the
variables, here the crack length and crack depth ratio, which
affect the rewards or costs at subsequent points in the tree.
The nodes and associated probabilities, rewards and variables
are all summarized in Table 2. There remains the task of de-
ciding which crack lengths would lead to failure in the service
time, and as such which should be removed from service if en-
countered by dye penetrant inspection.

V.   Supplementary Variables: the Fracture Analysis

From the data of Fig. 3 obtained by Wiederhorn and Evans
(1973), it is reasonable to assume that cracks grow with time
according to

$$\frac{dc}{dt} = \dot{c}_0 \left(\frac{K}{K_0}\right)^n \tag{1}$$

where $\dot{c}_0$ and K are constants read from Fig. 3. Except very
close to fracture, the exponent n has a value of approximately
10. Integrating Eq. 1 to a final half-length c_f at a lifetime
t_f and assuming that the stress intensity factor K is related to
the crack half-length c, the applied stress σ_a, and a geometrical
factor Y by

$$K = \sigma_a \sqrt{c} \, Y \text{ where } Y = \sqrt{\pi} \tag{2}$$

gives

$$\frac{1}{-n/2+1}\left(c_f^{-n/2+1} - c^{-n/2+1}\right) = \dot{c}_0 \left(\frac{\sigma_a Y}{K_0}\right)^n t_f \quad . \tag{3}$$

The negative exponential of the final crack length is usually negligible compared to that of the initial crack length, so Eq. 3 can be solved for the initial crack length in the form

$$t_f = \frac{1}{n/2-1}\left(\frac{K_0}{\sigma_a \sqrt{c}\, Y}\right)^n \frac{c}{\dot{c}_0} \quad . \tag{4}$$

One important result can already be obtained form Eq. 4. With an exponent of n=10, Eq. 4 indicates that the time to fracture varies as the inverse fourth power of the crack length. This means, for instance, that if cracks of length c would grow to fracture in the 10,000 hour period, then if it were possible only to detect cracks twice as long as this, even ideally perfect inspections would have to be carried out 16 times during the 10,000 hour period to prevent failure. Conversely, if it were possible to detect cracks of half the critical length for failure within the given period, then the life of the part could be assured for 160,000 hours. With this narrow range of crack length in which to operate, it seems best to concentrate all the inspection at the beginning of service rather than to count on repeated inspections to prevent fracture. This analysis of course overlooks the possibility of stress corrosion cracks which would accelerate the crack growth rates of Fig. 3 in ways which should be accounted for in some more complete decision analysis.

In applying Eq. 4, two different effects should be con-
sidered: short time thermal shock and long time steady stress.
In the first case, consider an applied stress due to thermal
shock reaching a level of 140 MNm^{-2} at 1250°C. From Fig. 3 the
constants $\dot{c}_0$ and K_0 are respectively, $\dot{c}_0 = 10^{-7}$ m/sec at a stress
intensity factor of $K_0 = 5$ MNm$^{-3/2}$. Assume that only 1% of the
10,000 hour service period is to be spent under the thermal shock
condition. In this case, the critical initial crack half-length
is

$$c = \left[\left(\frac{n}{2} - 1\right)\dot{c}_0\left(\frac{O_a Y}{K}\right)^n t_f\right]^{-1/(n/2-1)} = \left[(4)(10^{-7})\left(\frac{140\sqrt{\pi}}{5}\right)^{10}(.36\times10^5)\right]^{-\frac{1}{4}}$$

$$= .094 \text{ mm} \simeq .100 \text{ mm} \qquad . \tag{5}$$

Alternatively, at 1400°C and a stress of 35 MNm^{-2}, the critical
initial crack length is

$$c = \left[(4)(.5\times10^{-4})\left(\frac{35\sqrt{\pi}}{5}\right)^{10}(36\times10^6)\right]^{-\frac{1}{4}} = .2 \text{ mm} \qquad . \tag{6}$$

For illustration we take the shorter initial crack, that which
grows to fracture under the thermal shock conditions, and neglect
the growth due to steady stress. A crack half-length of greater
than 100 microns will then cause fracture from internal cracks.
This is close to the lower limit possible for penetrant inspection,
which is usually thought to be a total length of the order of
.005 in. or 125 microns.

For determining crack growth, the length ℓ of the penetrant indication on the surface should be modified by the unknown depth d of the crack. Because cracks of given length on the surface would be unlikely to swell out underneath, and because for shallow cracks the effective crack length is nearly the depth, one might take the effective half-length of a plane crack to be

$$c_e = (\ell/2)(2d/\ell)/\sqrt{1 + 4d^2/\ell^2} \quad .$$

Actually, we used a simpler expession,

$$c_e = (\ell)(d/\ell)/2 \quad , \tag{7}$$

but the results, discussed below, would not be essentially different.

After inspection, a blade should be rejected if the surface indication is longer than some critical value. Actually, for simplicity, since the cost of rejection was small we based rejection on half the effective half-length of the crack. Again the results would not be essentially different.

These premises led to the terminal set of rewards listed at the end of Table 2. In the present form of the program it appeared necessary to hand-calculate those rewards that were dependent on the variables. The extension of the program to provide such calculations would not seem to be difficult, except in so far as any modification of someone else's program is likely to be difficult. Perhaps a supplementary program to provide batch input would be a simpler approach.

-368-

VI.   The Computer Program

The computer program used to carry out this analysis is the program TREE by Rosseau and North (1971). It is available in time sharing mode through several computer communication services. Details of the program execution are not of concern here. Suffice it to say that learning to use the system and the program and to run the first problem required about 20 man-hours plus 7½ console hours for a skilled programmer and an assistant. The computational cost was mostly in console time charges. Later problems of similar size would probably require only about 3 man-hours plus 1 console hour.

VII.   Results

Results are presented in Table 3. We had speculated about whether we could predict the results, but none of us did. While the results may at first seem surprising, they were checked in a number of instances and found to be consistent with the premises. In no case did it appear that the cost of carrying out the penetrant inspection would be warranted, in spite of the very high penalty for the field failure of a turbine. Changes in any premises might reverse the decision in some cases, but any such changes would have to be justified by a more accurate study of the costs involved.

The principal conclusions of this study are
1. The use of decision analysis forces one to express his hunches and insight in a quantitative way that can be discussed with

others and left as part of the permanent design record.

2. As regards inspection of brittle materials, the very strong dependence of crack growth rate on crack length means that only a very critical range of cracks need be detected in an inspection scheme. Longer cracks must be removed or prevented in manufacture and shorter cracks are of no concern.

3. For this application, where no hazard to human life was anticipated, even the high cost of failure in terms of reputation and further sales was not enough to make penetrant inspection worthwhile.

4. The wide variability in expert opinion about the frequency of internal versus external cracks or about the distribution of crack lengths was not enough to upset these conclusions.

Acknowledgement

References

Decision analysis has been treated in a number of books and articles. Of the books, Raiffa (1968) and Schlaifer (1969) are two of the best known. Schlaifer (1971) describes a series of computer programs useful with his text. For the beginner, however, the concise article by Howard (1968) is suggested, along with examples such as by North (1968).

Howard, Ronald A.	1968	"The Foundations of Decision Analysis", IEEE Trans., Systems Sci. and Cybernetics, Vol. SCC-4, pp. 211-219.
North, D. Warner	1968	"A Tutorial Introduction to Decision Theory," IEEE Trnas. Systems Sci. and Cybernetics, Vol. SCC-4, pp. 200-210.
Raiffa, Howard	1968	Decision Analysis, Addison-Wesley, Reading, Massachusetts.
Rousseau, W. F. North, D. W.	1971	A User's Manual for TREE, A Program for the Analysis of Decision and Probability Trees, SRI Project 8104-6, Stanford Research Institute, Menlo Park, California.
Schlaifer, Robert	1969	Analysis of Decisions Under Uncertainty, McGraw-Hill Book Co., New York, New York.
Schlaifer, Robert	1971	Computer Programs for Elementary Decision Analysis, Division of Research, Graduate School of Business Administration, Harvard University, Boston, Massachusetts.
Wiederhorn, S. Evans, A. G.	1973	Unpublished research.

TABLE 1. Participants' estimates of internal cracking and of worst surface indications for Si_3N_4 turbine blades.

Probability that the worst crack in a turbine is internal		Length in microns of the longest indication in any blade in the turbine if the crack is external, for a given level of probability.		
Before Discussion	After Discussion	$p = .1$	.01	.001
.3	.3	5	10	20
.5	.5	10	50	100
.8	.8	1	3	8
.4	.5	10	50	140
.8	.7	1	10	100
0	0	10	20	100
.25	.3	25	250	2500
.5	.5	50	100	300
10^{-3}	10^{-2}	10	50	100
.1	.1	15	150	300
.3	.3	150	200	400
.5	.4	300	750	2000
.1	.1	50	100	500
.2	.4	10	20	100
.7	.5	10	100	1000
.4	.4	5	20	200
.1	.2	1000	2000	4000

TABLE 2. Probabilities, rewards, and variables

Node (Type)	Probability	Reward	Variable
YD	----	-$20,000	
ND	----	0	
-DIC	$P_{IC} = .1$ or $.7$	0	
-DEC	$1-P_{IC}$	0	
-DICFF	$P_{ICFF} = 10^{-2}$, 3×10^{-3}, 3×10^{-4}, or 3×10^{-5}	-2×10^6	
-DICFS	$1-P_{ICFF}$	0	assumptions corresponding to alternatives of -DICFF $\ell,\mu m$

			(a)	(b)	(c)	(d)
DECLi	.9	0	20,	10,	2,	or 2
	$.9\times10^{-1}$	0	50,	20,	5	or 6
	$.9\times10^{-2}$	0	300,	100,	20,	or 20
	$.9\times10^{-3}$	0	2000,	500,	100,	or 60
	10^{-4}	0	10,000,	2000,	500,	or 200
			d/ℓ			

Node (Type)	Probability	Reward	Variable
-DECLiDj	.2	0	.1
	.3	0	.3
	.4	0	1
	.09	0	3
	.01	0	10
T-YD	If $2c_e > 100\mu$, $R = -$20		
T-ND	If $2c_e > 200\mu$, $R = -$2\times10^6$		

TABLE 3. Average rewards with and without dye penetrant inspection

Probability worst crack internal, P_{IC}	Probability of field failure from internal crack, P_{ICFF}	Set of indicated lengths (Table 2)	Average Rewards	
			with dye inspection YD	with no dye inspection ND
.1	.01	(a)	-22,000.42	-13,196.00
.7	.01	(a)	-34,000.14	-17,732.00
.1	.003	(b)	-20,600.05	- 3,174.00
.7	.003	(b)	-24,200.02	- 5,058.00
.1	.0003	(c)	-20,060.00	- 312.00
.7	.0003	(c)	-20,420.00	- 504.00
.1	.00003	(d)	-20,006.00	- 40.20
.7	.00003	(d)	-20,042.00	- 53.40
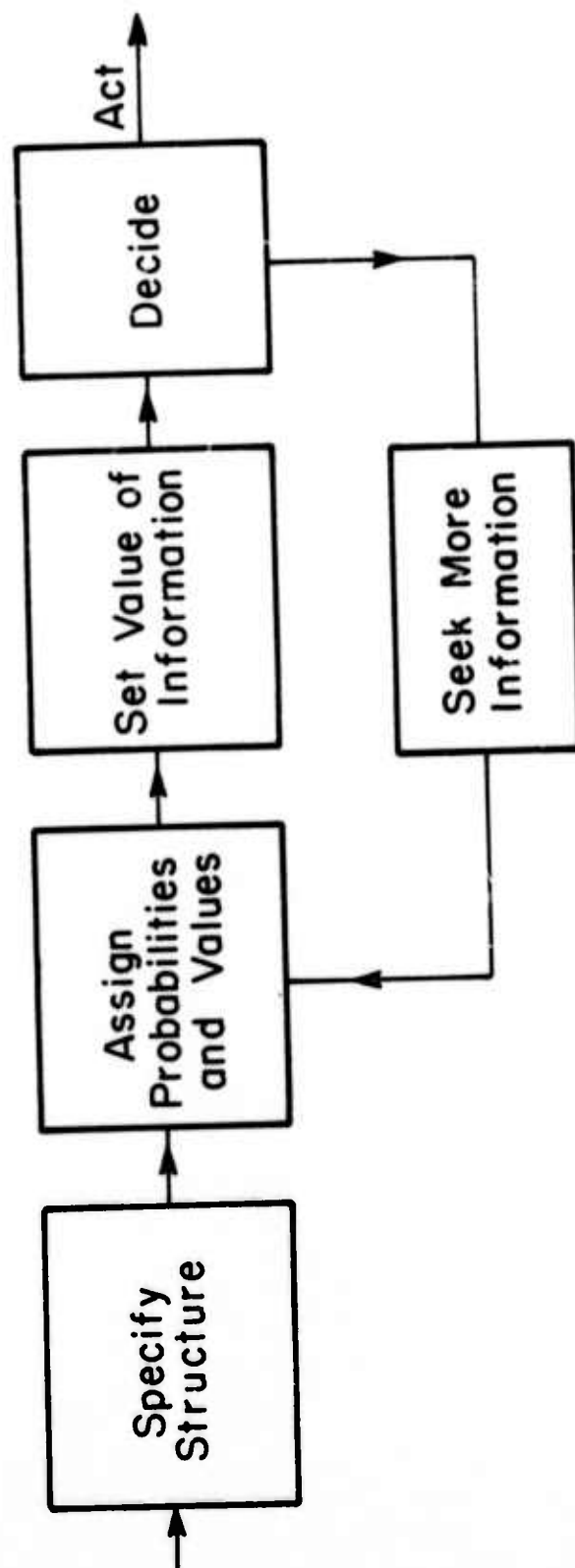
Figure 1. Steps in decision analysis.

-375-
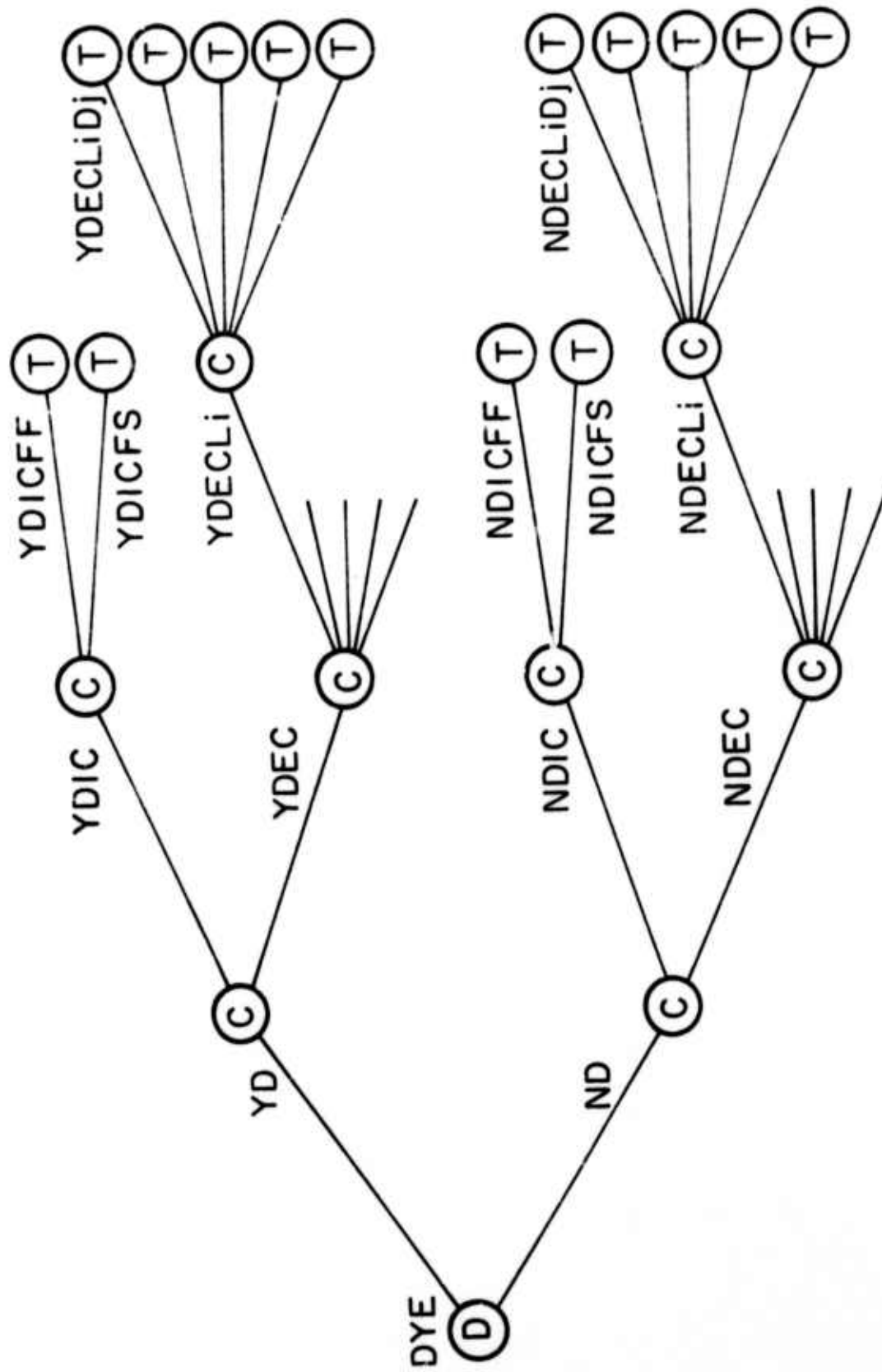
Figure 2. Decision tree for dye penetrant inspection.
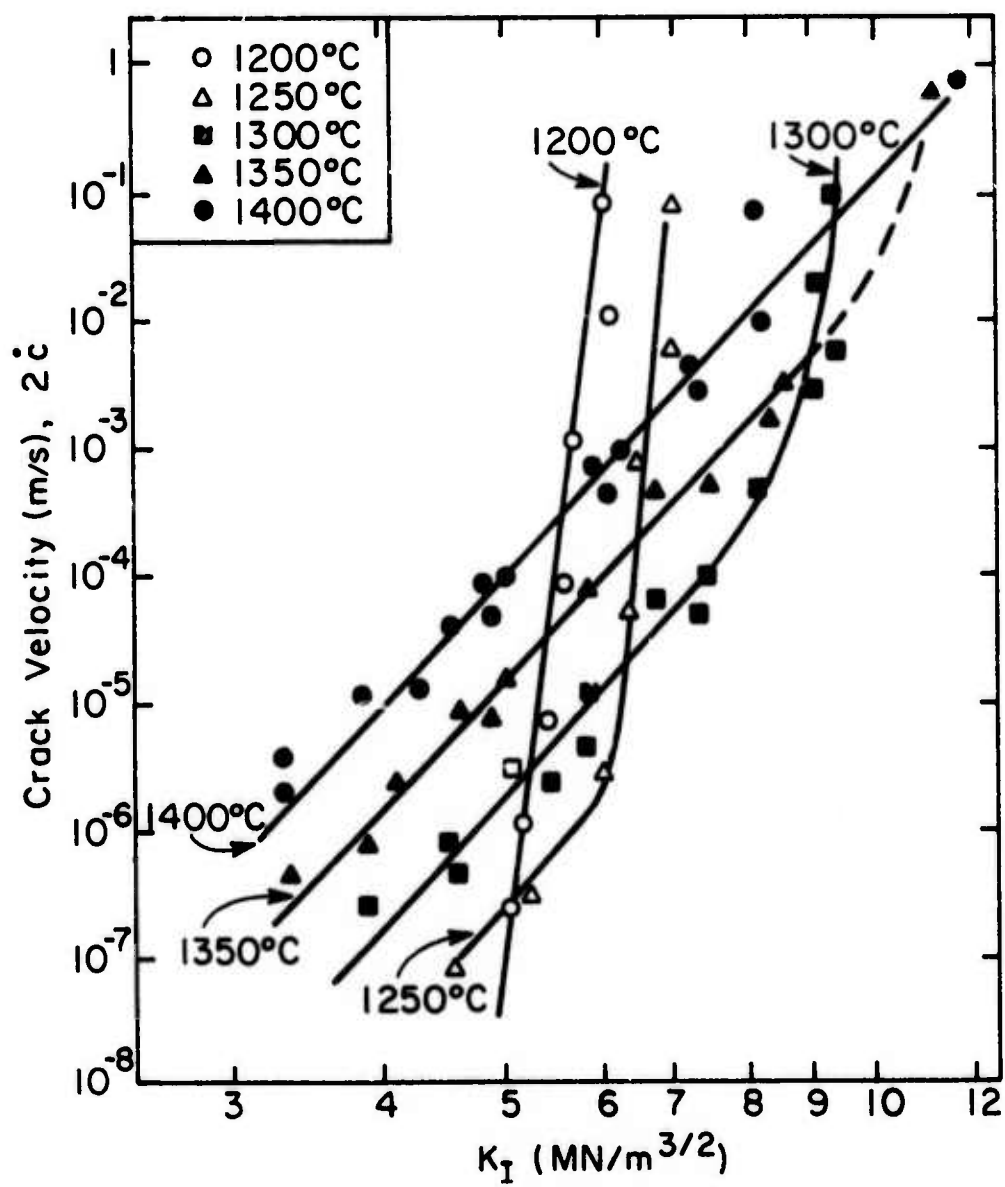
Figure 3. Crack growth rate as a function of stress intensity.

INTERPRETIVE COMMENTS ON
"REMARKS ON GIANT CONDUCTIVITY IN TTF-TCNQ"

M. Tinkham

Abstract

In a recent preprint, Anderson et al. derived the temperature dependence $\sim T\, e^{T_c/T}$ for the fluctuation-limited conductivity below T_c in a novel model for a superconducting molecular chain. Experimental data on TTF-TCNQ were fitted to this form with $T_c = 500°K$. It is shown here that essentially this same form of temperature dependence is predicted by the conventional theory of thin metallic superconductors, if the diameter is scaled down to the molecular level.

INTERPRETIVE COMMENTS ON
"REMARKS ON GIANT CONDUCTIVITY IN TTF-TCNQ"

M. Tinkham

The thrust of the remarks of Anderson et al.[1] is that
a one-dimensional (1-D) system of very small cross-section will
have a broad superconducting transition, and that the resistance
will stay high until fluctuations are gradually frozen out in
some temperature range far below T_c, the "mean field" critical
temperature. Although Anderson et al. model TTF-TCNQ as a string
of discrete weak links between molecular units, it is useful
first to review the predictions of the theory of ordinary con-
tinuous uniform 1-D systems, such as the tin whiskers with diameter
$\sim 0.5\mu$, whose resistive transitions have been studied experimentally
by Lukens, Warburton, and Webb[2] and by Newbower, Beasley, and
Tinkham[3].

The theory is well developed in two limits: $T \ll T_c$ and
$T \gg T_c$. In the first limit, well below T_c, long range supercon-
ducting phase coherence is established, but broken occasionally
by fluctuation-induced phase slips. The theory was worked out
by Langer and Ambegaokar[4] and by McCumber and Halperin[5], and is
usually called the LAMH Theory. In the other limit, T well
above T_c, the system is basically normal, i.e., no long range

phase coherence - but fluctuations create momentary islands of weak superconductivity, which reduce the overall normal resistance. The theory for this limit was first worked out by Aslamazov and Larkin[6] (AL). In between these two limits is the so-called "critical region" where neither of these limiting approximations is adequate. The width of this critical region increases as the cross-section of conductor decreases and may be very wide in a conductor of molecular dimensions. Now let us review the predictions of these two limiting case theories.

The AL prediction for the <u>extra</u> conductivity above T_c due to superconducting fluctuations is

$$\sigma' = \sigma - \sigma_n = \frac{e^2 \xi(0)}{16 \hbar A} \left(\frac{T}{T - T_c}\right)^{3/2} \tag{1}$$

where A is the cross-sectional area, $\xi(0)$ is the zero-temperature coherence length, and T_c is the mean-field transition temperature. Note that the extra conductance of a filament, $\sigma'A$, which is the directly measureable quantity, is independent of A. Physically speaking, the reason for this result is that the fluctuation effects get larger and cancel the smaller area. It is normally assumed that this formula is only justified for $\sigma' \lesssim \sigma_n$, i.e., for modest enhancements of the conductivity. Thus, its use by Coleman et al.[7] to fit a 500-fold conductivity increase seems inappropriate and may be misleading.

To account for such large decreases in resistance, one must assume that long-range phase coherence persists for some significant time between the phase slip processes giving resistance.

In other words, the LAMH model for $T < T_c$ should be a better starting point than the AL theory. In the LAMH theory the conductivity is limited only by superconducting phase slips, so the normal resistance plays no direct role. The predicted resistance is

$$R \simeq \Omega \, \frac{\pi \hbar^2}{2e^2 kT} \, e^{-\Delta F_o/kT} \tag{2}$$

where $\Omega \simeq (L/\xi)(1/\tau_{GL})$ is a temperature-dependent attempt frequency, typically of order 10^{12} sec^{-1}. The activation energy ΔF_o is approximately the superconducting condensation energy in a length $\sim \xi$ of whisker, since this minimum length must go nearly normal to decouple the phase from one end to the other. More precisely, it turns out that

$$\Delta F_o = \frac{8\sqrt{2}}{3} \, A\xi \, \frac{H_c^2}{8\pi} \propto (1-t)^{3/2} \qquad (t \simeq 1) \tag{3}$$

where $t \equiv T/T_c$, if we insert the temperature dependence $H_c \simeq H_c(0)(1-t^2)$ and $\xi \simeq \xi(0)(1-t)^{-\frac{1}{2}}$. Putting in the numbers for a 1/2 micron tin whisker, the exponent is approximately

$$\left. \frac{\Delta F_o}{kT} \right|_{\text{tin whisker}} \simeq 6 \times 10^6 \, \frac{(1-t)^{3/2}}{t} \tag{4}$$

which gives astronomically small resistance unless $t \simeq 1$. For example, the resistance will be unobservably small if $\Delta F_o/kt \simeq 50$, so that $e^{-\Delta F_o/kT} \simeq 10^{-20}$. This occurs when $(1-t) \simeq 4 \times 10^{-4}$, or $T_c - T \simeq 1.6$ mK, as is confirmed experimentally.

Now let us rather blindly scale this calculation to the resistance of a superconducting fiber of molecular dimensions.

In the absence of numerical parameters for H_c and ξ, it is useful to use standard relations (even though they may not be directly applicable) to rewrite the exponent $\Delta F_o/kT$. Specifically, following BCS theory, we write the condensation energy at $T=0$ as $H_c^2(0)/8\pi = \frac{1}{2} N(0)\Delta^2(0)$, where $\Delta(0) = 1.76 \ kT_c$, and then use the approximation $H_c(t) = H_c(0)(1-t^2)$; we also approximate the temperature dependence of $\xi(t)$ by the "two-fluid" formula $\xi(t) \simeq \xi(0)[(1-t^2)]^{\frac{1}{2}}$, and take the clean limit result $\xi(0) = 0.74\xi_o$, where $\xi_o = \hbar v_F/\pi\Delta(0)$. If we also use the free-electron relation $N(0)\hbar v_F = k_F^2/2\pi^2 \simeq \frac{1}{2}(N/V)^{2/3}$, where (N/V) is the number of conduction electrons per unit volume, the exponent reduces to

$$\frac{\Delta F_o}{kT} \simeq 0.4(N/V)^{2/3}A\frac{(1-t^2)(1-t^4)^{\frac{1}{2}}}{t} \xrightarrow[(t\simeq 1)]{} 1.6(N/V)^{2/3}A\frac{(1-t)^{3/2}}{t} \qquad (5)$$

For one electron per atom, $(N/V)^{2/3}A$ is roughly the number of atoms in a cross-sectional plane of the filament. With lower carrier densities, the energy is less in proportion to $(N/V)^{2/3}$.

In applying this formula to a stacked column of planar molecules, as in TCNQ complexes, it is useful to write the molecular volume as Ad, where d is the interplanar distance. Then, if there is one conduction electron per molecule, $(N/V)^{2/3}A = (A/d^2)^{1/3}$. In TCNQ complexes, $A \simeq 6\text{Å} \times 9\text{Å}$, whereas $d \simeq 3\text{Å}$. Thus $(A/d^2)^{1/3} \simeq 1.8$ and $\Delta F_o/kT \simeq 0.7(1-t^2)(1-t^4)^{\frac{1}{2}}/t$, which reduce to $\sim 0.7 T_c/T$ for $t \ll 1$. Finally, if the prefactor in (2)

is computed taking $\hbar/\tau_{GL} = 8kT_c/\pi$ (which is the result of extra-
polating TDGL theory to $T \ll T_c$), one finds a conductivity $\sigma = L/AR$
of

$$\sigma_{esu} = 30c \; \sigma_{practical} = \frac{c}{4}(\frac{e^2}{\hbar c}) \; \frac{\xi}{A} \; \frac{T}{T_c} \; e^{\Delta F_o/kT} \qquad (T \ll T_c) \qquad (6)$$

If we take $\xi = 10\overset{\circ}{A}$ and $A = 5 \times 10^{-15} \; cm^2$, this becomes, for the
ΔF_o found above

$$\sigma = 1200 \, (T/T_c) \; e^{0.7Tc/T} \qquad (ohm\text{-}cm)^{-1} \qquad (T \ll T_c) \qquad (6a)$$

Apart from minor numerical factors, this is equivalent to
equation (7) of Anderson et al. Thus, the form of the predicted
conductivity is not sensitive to the distinction between a dis-
crete "Josephson-coupled" model and a continuous model. On the
other hand, our expression (5) shows clearly the importance of
the small cross-section.

As noted by Anderson et al., the observed conductivity
is well fitted by an exponential form such as (6a), with $T_c \sim$
500°K (or $\sim$700°K with our factor of 0.7 in the exponent). The
prefactor is also of the correct order of magnitude. Despite
the high "T_c", which implies that local pairing commences at
these high temperatures, the conductivity does not start to rise
strongly until a much lower temperature. Near T_c, the long
range phase coherence would be interrupted too frequently by
thermal fluctuations. Obviously the specific numbers should
not be taken seriously, but this example illustrates how the
mean field T_c can be much larger than the temperature at which

the resistance starts to get very small.

Remarks:

(1) To get this very broad transition, one must assume no coupling exists between neighboring chains. If this coupling is not <u>zero</u>, it should lead to 3D behavior and strong super-conductivity at some lower "effective" transition temperature. Examples of this sort are familiar in magnetism: short-range order develops strongly along chains well above the temperature at which long-range order develops by the weak coupling between chains.

(2) Apparently some sort of Peierls transition to an insulating state occurs to account for the collapse of con-ductivity before a strongly superconducting state is set up. It is not obvious that one could not have the superconducting state set up first.

(3) Of course, the whole interpretation of this enhanced conductivity as reflecting incipient superconductivity remains speculative. It is entirely possible that the explanation is of totally different nature.

Acknowledgement

References

1. P. W. Anderson, P. A. Lee, and M. Saitoh, preprint.

2. J. E. Lukens, R. J. Warburton, and W. W. Webb, Phys. Rev. Letters $\underline{25}$, 1180 (1970).

3. R. S. Newbower, M. R. Beasley, and M. Tinkham, Phys. Rev. $\underline{B5}$, 864 (1972).

4. J. S. Langer and V. Ambegaokar, Phys. Rev. $\underline{164}$, 498 (1967).

5. D. E. McCumber and B. I. Halperin, Phys. Rev. $\underline{B1}$, 1054 (1970).

6. L. G. Aslamazov and A. I. Larkin, Phys. Letters $\underline{26A}$, 238 (1968).

7. L. B. Coleman, M. J. Cohen, D. J. Sandman, F. G. Yamagishi, A. F. Garito, and A. J. Heeger, Solid State Comm. $\underline{12}$, 1125 (1973).

PROPAGATION OF TRANSIENT ELASTIC WAVES
IN PERIODIC COMPOSITES

J. A. Krumhansl
E. H. Lee

Abstract

The application of Fourier-Floquet methods to transients
in periodic composite media is presented. The head of the wave
solutions in the far field are dominated by the maximum group
velocity contributions from each frequency band.

In one dimension, the results of the wave front tracing
approach for an initial impulse have been recovered from Fourier-
Floquet analysis in the case of layers with commensurate transit
times.

Motion of a periodic layered medium under applied forces
is discussed, including both a formal representation of the
response Green's function and the treatment of short pulse
pressure loading on a surface.

We conclude that important features of transient elastic
wave propagation can be obtained using Fourier-Floquet methods.

PROPAGATION OF TRANSIENT ELASTIC WAVES
IN PERIODIC COMPOSITES

J. A. Krumhansl
E. H. Lee

I.   Introduction

In previous work[1], we have given principal consideration
to steady state Floquet solutions for waves in periodic composites.
As is standard in linear problems, such solutions are of interest
because they form a basis for Fourier development of transient
solutions. The technique for Fourier-Floquet expansions is a
conceptually straight-forward, though sometimes analytically
complex, procedure. The purpose of this paper is to outline the
Fourier-Floquet description of transients. Application is made
to layered composites, with propagation normal to the layer
planes, x = constant.

Before going on to details, we first note what is of
principal experimental interest.

(i) Given initial velocity and displacement variations,
say localized around $x \simeq 0$, $t = 0$, what is the asymptotic be-
havior at large x and t? As has been found many times before
in such problems, this asymptotic behavior is dominated by phase
coincidence of various Fourier components. Therefore, the
principal features are determined by the dispersion relations,
ω vs. q, particularly the group velocity ($d\omega/dq$), where ω is the
frequency and q the wave number.

(ii) Another problem is that arising when a prescribed force distribution $F(x,t)$ is applied to the composite and the resulting displacement is desired. The relation between the two quantities is given by a response function, which is an appropriately defined Green's function.

We proceed as follows. First, we develop the formal description of transients in Fourier-Floquet form. Second, we examine the asymptotic behavior and show that there are general features common to all models, resulting from dispersion. We use the method of stationary phase and the criterion of minimum attenuation to obtain these results. Third, we show the connection with "geometric optics" for the particular case of propagation normal to the planes of a periodically layered composite. Fourth, we comment on applied stress-displacement relations, i.e., response functions. The present work complements that of Peck and Gurtman[2], who considered propagation parallel to layers of a composite.

II. Fourier-Floquet Representation of Solutions

Everything which is done here for one dimension can be formally extended to three dimensions. We present the discussion in one dimension for pedagogical reasons.

Given the elastic equation in a periodic medium[1], with elastic constant $\eta(x)$ and density $\rho(x)$,

$$\frac{\partial}{\partial x}\left(\eta\frac{\partial u}{\partial x}\right) - \rho\frac{\partial^2 u}{\partial t^2} = -F(x,t) \tag{1}$$

for the case $F(x,t) = 0$, $u = u_\nu \exp[+i\omega t]$, and $\eta(x)$ and $\rho(x)$ varying with period $\Delta x = a$, the solutions of (1) in the form

$$\frac{\partial}{\partial x} \left(\eta \frac{\partial u_\nu}{\partial x} \right) + \rho \omega^2 u_\nu = 0 \quad , \tag{2}$$

yield a set of displacements and dispersion relations of the form

$$u_\nu(x;q) = e^{iqx} v_\nu(x;q) \tag{3}$$

$$\omega = \omega_\nu(q) \tag{4}$$

These lie in ν denumerable regions of ω, and $-(\pi/a) \leq q \leq (\pi/a)$ for each; $v_\nu(x,q)$ being periodic with period $\Delta x = a$.

These solutions, being solutions of a second order self-adjoint equation and boundary conditions, may be used to construct an orthonormal set (see Appendix). Thus,

$$\int_{-\infty}^{\infty} dx \, \rho(x) u_\nu^*(x,q) u_{\nu'}(x,q') = \delta_{\nu\nu'} \, \delta(q-q') \tag{5}$$

which implies the normalization

$$(2\pi/a) \int_{-a/2}^{a/2} dx \, \rho \, u_\nu^*(x,q) u_\nu(x,q) = 1 \tag{6}$$

where the asterisk denotes complex conjugate. Thus the band orthogonality is given by Kronecker delta, and the q space by Dirac delta. Then we may construct solutions of the homogeneous equation by the expansion

$$u(x,t) = \sum_{\nu} \int_{-\pi/a}^{\pi/a} dq \, [c_\nu(+,q) e^{i\omega_\nu t} + c_\nu(-,q) e^{i\omega_\nu t}] u_\nu(x;q) \qquad (7)$$

$$u(x,t) = \sum_{\nu} \int_{-\pi/a}^{\pi/a} dq \, [c_\nu(+,q) e^{i\omega_\nu t} - c_\nu(-,q) e^{-i\omega_\nu t}] (i\omega_\nu) u_\nu(x;q) \qquad (8)$$

The coefficients are given by the usual inverse relation

$$\int_{-\infty}^{\infty} dx \, \rho(x) u_\nu^*(x;q) u(x,0) = c_\nu(+,q) + c_\nu(-,q) \qquad (9)$$

$$\int_{-\infty}^{\infty} dx \, \rho(x) u_\nu^*(x;q) u(x,0) = i\omega_\nu [c_\nu(+,q) - c_\nu(-,q) \qquad (10)$$

These general formulae specify the homogeneous solution com-
pletely. When extending to three dimensions, the displacements
are vector fields and orthonormalization is of the scalar
product

$$\iiint_{-\infty}^{\infty} dxdydz \, \rho \, \underline{u}_\nu^*(\underline{r},\underline{q}) \cdot \underline{u}_\nu{}'(\underline{r},\underline{q}) = \delta_{\nu\nu'} \delta(\underline{q}-\underline{q}') \qquad (11)$$

and three components of initial velocity and displacement
specified as functions of position are necessary to specify
the problem.

III. Propagation of a Pulse

(i) Consider an infinite, layered, medium at rest until a momentum impulse of strength s is given to the element $-\frac{\varepsilon}{2} < x < \frac{\varepsilon}{2}$ centered at $x = 0$ (in application it would be convenient to arrange $x = 0$ at a symmetry plane of the periodic lattice). The details of the stress and particle velocity will be discussed further in Section V. In the limit of $\varepsilon \to 0$ we have the initial conditions:

$$u(x,t = 0+) = 0 \qquad \text{for all } x$$
$$\dot{u}(x,t = 0+) = s\delta (x-0) \tag{12}$$

where $\delta(x-0)$ is the Dirac delta function. For these initial conditions, we work out the complete Fourier-Floquet solutions from (9) and (10):

$$c_\nu(+,q) = -c_\nu(-,q) \tag{13}$$

$$\rho(0)su_\nu^*(0;q) = i\omega_\nu 2c_\nu(+,q) \tag{14}$$

Whence

$$\dot{u}(x,t) = \sum_\nu \int_{-\pi/a}^{\pi/a} dq\rho(0)su_\nu^*(0;q)u_\nu(x;q)\, \cos(\omega_\nu t) \tag{15}$$

The solutions $u_\nu(x;q)$ have the property that $u_\nu^*(x'q)=[u_\nu(x;-q)]$, and with a little algebra one finds

$$\dot{u}(x,t) = \sum_\nu \int_0^{\pi/a} dqA_\nu(q,x)\,[\cos(\omega_\nu(q)t-\theta(q,x))+\cos(\omega_\nu(q)t+\theta(q,x)] \tag{16}$$

where A_ν and θ are defined by

$$s\rho(0) \; u_\nu^*(0;q) u_\nu(x;q) = A_\nu(q,x) e^{i\theta(q,x)} \qquad (17)$$

In this case of an infinite medium, the pulse splits into two equal components, one propagating to the right, and the other to the left.

(ii) Now consider the asymptotic behavior for large positive x. While equation (16) is the general result, we can take advantage of the special quasi-periodic property of Floquet solutions, and examine the velocity at $x = 0 + na$, for integer n. Then $u_\nu(na;q) = e^{iqna} u_\nu(0;q)$; also, $A_\nu(q,na) = s\rho(0) u_\nu^*(0;q) u_\nu(0;q)$, and $\theta(q,x) = qna$. We will consider the solution many layers from the origin, i.e., $n \gg 1$, and (16) becomes:

$$\dot{u}(na,t) = \sum_\nu \int_0^{\pi/a} dq A_\nu \; \cos[\omega_\nu(q) t - qna] \qquad (18)$$

We will examine the asymptotic behavior of this quantity by an adaptation of the method of stationary phase[3,4] to the integral over q in equation (18). In the usual case of wave packet motion $A_\nu(q)$ is a function which is peaked around some particular value of q, say q_0. Then it is argued that for large t and na the phase of the cosine function $\phi_\nu(q) = [\omega_\nu(q) t - qna]$ is a rapidly varying function of q unless in

$$\phi_\nu(q) = \phi_\nu(q_o) + \left(\frac{d\phi_\nu}{dq}\right)_{q_o} (q-q_o) + \cdots \qquad (19)$$

we choose t for given na to make $(\frac{d\phi_\nu}{dq})_{q_o} = 0$. The higher order terms in the Taylor series lead to attenuation of the wave packet and will be discussed shortly. The "stationary phase" condition then leads to a condition on t, for given na, such that

$$[(\frac{d\omega_\nu(q)}{dq})_{q_o} t - na] = 0 \tag{20}$$

where

$$(\frac{d\omega_\nu}{dq})_{q_o} = v_g(q_o)$$

is the group velocity for waves of mean wave number q_o.

The physical problem we are considering is slightly different. The choice of an initial momentum impulse leads to an $A_\nu(q)$ which is not peaked in q space, but rather is distributed more or less uniformly; thus we do not have a wave "packet" centered at a given q_o. However, by examining the more complete Taylor series we now show that something like a packet can develop asymptotically, serving to define the arrival and shape of the modified impulse at large t and na.

The first four terms of the expansion of $\phi(q)$ about $q = q_o$ are

$$\phi_\nu(q) = (\omega_\nu(q_o)t - q_o na) + [(\frac{d\omega_\nu}{dq})_{q_o} t - na](q-q_o)$$

$$+ (\frac{d^2\omega_\nu}{dq^2})_{q_o} \frac{(q-q_o)^2}{2!}t + (\frac{d^3\omega_\nu}{dq^3})_{q_o} \frac{(q-q_o)^3}{3!}t + \cdots \tag{21}$$

It will be noted that beyond the first order the higher co-efficients do not depend on na.

Now ask the question: given that the spectral density $A_\nu(q)$ is more or less uniform in q, which components of $A_\nu(q)$ dominate at large distance?

For given na, and prescribed $q \simeq q_0$, stationary phase arguments determine arrival time, as just discussed. But going to next order the contribution of components near q_0 to the integral is proportional to

$$\int_{-\infty}^{\infty} dq \; Re\left[\exp\left(\frac{i(q-q_0)^2}{2} \frac{d^2\omega}{dq^2}\Big|_{q_0} t\right)\right] = \frac{1}{2}\sqrt{\frac{\pi}{t \frac{d^2\omega}{dq^2}\Big|_{q_0}}}$$

and it is seen that for large t (i.e., arrival at large na) the amplitude attenuates as

$$\frac{d^2\omega}{dq^2}\Big|_{q_0}^{-\frac{1}{2}} t^{-\frac{1}{2}}$$

This leads to a further consideration at large distances; the behavior will be dominated by those components of A(q) for which $\frac{d^2\omega}{dq^2}\Big|_{q_0}$ is small. Indeed, it is possible in the present instance to find q such that

$$\left(\frac{d^2\omega}{dq^2}\right)_{q_0} = 0$$

This, incidentally, happens to be the condition for maximum group velocity $\frac{d}{dq} (\frac{d\omega}{dq}) = 0$, whence we adopt the terminology $q_0 = q_{\nu m}$ to signify that q in the νth band for which the group velocity is a maximum.

Thus, at large na we may compute the dominant behavior by using two conditions:

(a) minimum attenuation selects $q \simeq q_m$ components of $A_\nu(q)$.

(b) constant phase conditions with $q \simeq q_m$ determine times.

However, the approximation just given actually fails at $\frac{d^2\omega}{dq^2}\Big|_{q_0} = 0$ and it is necessary to go to the next terms in the Taylor series, which we now proceed to examine.

Consider the Taylor expansion about $q = q_{\nu m}$:

$$\phi_\nu(q) = (\omega_\nu(q_{\nu m})t - q_{\nu m}na) + [(\frac{d\omega_\nu}{dq})_{\nu m}t - na](q - q_{\nu m})$$
$$+ 0 + \frac{1}{3!}(\frac{d^3\omega_\nu}{dq^3})_{\nu m}(q - q_{\nu m})^3 t + \cdots \qquad (23)$$

If terms beyond the cubic in $(q - q_{\nu m})$ are dropped, the result is given in closed form by a set of Airy functions.[5] The final result is

$$\dot{u}(x,t) = \sum_\nu A_\nu(q_{\nu m})(1/2\omega'''_{\nu m}t)^{-1/3}2\pi \cos(\omega_{\nu m}t - q_{\nu m}na)$$
$$\times Ai[-(1/2\omega'''_{\nu m}t)^{-1/3}(na - (\frac{d_{\omega\nu}}{dq})_{\nu m}t)] \qquad (24)$$

where Ai is the Airy function of the first kind. The appearance
of Airy functions is common in analogous problems, for example[2].
The Airy function of the first kind may be expressed in terms of
Hankel functions of order one-third; for negative arguments $Ai(z)$
vanishes exponentially, while for positive arguments it is an
algebraically damped oscillatory function. The function $Ai(z)$ is
plotted in Fig. (1). Physically, for $-[na-(d\omega_\nu/dq)_{\nu m}t]$ less than
zero, the envelope of the νth term in (24) decreases exponentially
toward decreasing times; when this argument is near zero the
envelope rises rapidly, so that the main content arrives with
group velocity $(d\omega_\nu/dq)_{\nu m}$ for the νth band; thereafter, the en-
velope oscillates and damps. Throughout this behavior, the en-
velope modulates the oscillatory term $\cos(\omega_{\nu m}t-q_{\nu m}na)$ for each
band.

While the specific features may vary from one band to
another, as determined by the particular case, the following
general features are found for the asymptotic behavior of a
pulse initially uniform over a plane:

(1) The main arrival of each band's contribution is
determined by the maximum group velocity for that band. Thus,
except for scale, the dispersion relation $\omega_\nu(q)$ determines the
response.

(2) Because of the factor $(1/2\ \omega_{\nu m}'''t)^{-1/3}$ in (24) and
$t \simeq na(d\omega_\nu/dq)_{\nu m}^{-1} \simeq x(d\omega_\nu/dq)_{\nu m}^{-1}$, we find that (a) the
asymptotic peak amplitude decreases as $x^{-1/3}$, (b) the pulse
spreads out and becomes less steep with rise time proportional

to $x^{+1/3}$, (c) peak stress is complicated to calculate, but also falls off at large x.

Thus, the general effects expected of dispersive propagation are found; sharp pulse fronts are dispersed and peak stresses are attenuated with distance. In addition, we draw attention to the fact that these general results are almost model independent, being the same whether one is propagating along layer planes (Peck and Gurtman)[2] or normal as here, and can be anticipated for real fiber reinforced geometries which will differ from layered composites only in details of the dispersion relations. The specific asymptotic variation above refers to initially planar pulses; for initially cylindrical or spherical pulses the method of constant phase might be used in two or three dimensions, leading to appropriately different asymptotic variation of the head of the wave structure.

This completes the general discussion. Application to a particular case requires a knowledge of the dispersion relations. For illustrative purposes here, we continue with a case which can be treated exactly, the periodic layered composite.

IV.   Pulse Propagation in a Layered Medium and the Geometric
       Optics Description

For alternating layers of internally homogeneous elastic materials, propagation normal to the layers, and either purely longitudinal of purely transverse displacements, it is straightforward to obtain exact Floquet solutions[1]. To discuss the asymptotic properties, we need only have the dispersion relation:

$$\cos qa = [1 + \frac{(1-p)^2}{4p}] \cos(\alpha+\beta) - \frac{(1-p)^2}{4p} \cos(\alpha-\beta) \quad (25a)$$

or equivalently

$$\cos qa = \frac{1}{1-R^2} \cos(\alpha+\beta) - \frac{R^2}{1-R^2} \cos(\alpha-\beta) \quad (25b)$$

with the "matrix" and "filler" materials denoted by subscripts "m" and "f", and of thickness (a-b) and b, respectively, then the quantities appearing in (25) are as follows:

$$\alpha = \omega\tau_f = \omega(b/c_f)$$

$$\beta = \omega\tau_m = \omega[(a-b)/c_m]$$

$$p = (\eta_f\rho_f/\eta_m\rho_m)^{1/2}$$

$$R = \frac{1-p}{1+p} , \text{ the reflection coefficient}$$

That is, τ_m and τ_f are transit times in the respective layers. For arbitrary τ_f and τ_m, the quantities α and β are incommensurate and (25) is a general doubly periodic function of α and β. The roots of (25) may be obtained by the graphical construction shown in Fig. (2). It is to be seen directly that the maximum possible absolute value of the right-hand side of Eq. (25) is $[1 + \frac{2R^2}{1-R^2}]$, as shown in Fig. (2). The case $\omega = 0$ is a special point, since both $\cos(\alpha+\beta)$ and $\cos(\alpha-\beta)$ are always unity. The possible values of frequency for which qa is real (i.e., propagating solutions) are those for which the function lies between +1 and -1. Thus, the allowed frequencies for

-399-

propagation lie within bands generally separated by gaps as indicated in Fig. 2. As drawn, the graph corresponds to p near unity, or a small reflection coefficient R; for greatly different media R is large and the maximum excursion of the right-hand side of (25) becomes large and the pass bands narrow.

Equations (25) provide an implicit relation between ω and q. It is not likely that the general case can be discussed further exactly. We can, however, expect that the maximum group velocity will occur near the middle of each band. The group velocity may be computed by differentiating either of Eqs. (25), and eliminating trigonometric terms in α and β in favor of q. The calculation leads to an implicit equation in q, which in general is cumbersome. For two special cases we can obtain simply interpretable results:

Case A: if R is small we neglect the term in $\cos(\alpha-\beta)$ and find, for all bands

$$\frac{d\omega}{dq}\Big|_{max} \simeq \left(\frac{a}{\tau_m+\tau_f}\right) \frac{1}{1+R^2} \tag{26a}$$

which occurs very near qa = odd multiples of $\pi/2$ after the first band. Thus the precursor wave speed is reduced by the factor $(1+R^2)^{-1}$.

Case B: if the magnitude of the reflection coefficient is nearly unity it is not in general possible to obtain a simple expression, but if we take $\tau_m = \tau_f$ (not unreasonable for a stiff filler material in a small amount of soft matrix) then

-400-

$$\frac{d\omega}{dq}\bigg|_{max} \simeq \frac{a}{\tau_m + \tau_f} \left(\frac{1-R^2}{1+R^2}\right)^{1/2} \tag{26b}$$

When the reflection coefficient approaches unity, the group velocity is drastically reduced from the wave speed. Again this occurs for qa = odd multiples of $\pi/2$ after the first band.

These considerations suggest that we should compare Fourier description with that of wave fronts propagated with successive reflection at interfaces. We would expect that at x = na (again for a $\delta(x-0)$ initial velocity pulse) there would be a first pulse (precursor) arriving at $t = n(\tau_m + \tau_f)$; thereafter, there would be echoes at times delayed by integer multiples of $2\tau_m$, $2\tau_f$, and combinations thereof.

If τ_m and τ_f are incommensurate, then for large n there would be a dense distribution of arriving pulses, modulated by the Airy function envelope discussed in the previous section. It would be interesting if a behavior like this could be extracted from the Fourier representation. We have not been able to do so generally, because of the complicated time distribution of arrivals in the incommensurate case.

However, if we make the special assumption that $\tau_m = k\tau_f$, where k is an integer, we find very special properties for the dispersion relation, which now becomes a singly periodic function of ω, with period $(2\pi/\tau_f)$. This follows from Eq. (25) since $\cos(\alpha+\beta)$ and $\cos(\alpha-\beta)$ are then both periodic in $(2\pi/\tau_f)$. Now

-401-

$$\text{for } \omega = (\pi/\tau_f) \qquad (\alpha+\beta) = (k+1)\pi$$

$$(\alpha-\beta) = (k-1)\pi$$

$$\text{for } \omega = (2\pi/\tau_f) \qquad (\alpha+\beta) = 2(k+1)\pi$$

$$(\alpha-\beta) = 2(k-1)\pi$$

These relations lead to very special features in the dispersion relations and make possible a nearly complete mathematical description. Again, the graphical description is useful; Fig. (3) shows what happens for the case $k = 2$.

Two notable features have appeared. First, there is no band gap at $(\alpha+\beta) = 3\pi$, 6π, etc., i.e., bands coalesce. Second, it is apparent that all properties of the dispersion relation are periodic in $(\pi/\tau_f) = (2\pi/2\tau_f)$. Then the general form of the velocity given by Eq. (24) will be dominated by the coherent superposition of all those terms in the series having precisely the same group velocity, which will also have the same envelope function. In fact, their parameters will be the same as those for ω and q approaching zero. Thus, we find

$$\dot{u}(na,t) \to F_0(t;na-c_0t) \sum_{\ell=-\infty}^{\infty} 2\pi \cos[2\pi\ell(\frac{t}{2\tau_f} - n)] \qquad (27)$$

where

$$F_0 = |1/2\ \omega_0'''t|^{-1/3} Ai[|1/2\ \omega_0'''t|^{-1/3}(na-c_0t)] \qquad (28)$$

and in which ω_0''' and $c_0 = (d\omega/dq)_0$ are evaluated for $\omega,q \to 0$. But

$$\int_{-\infty}^{\infty} \cos [2\pi\ell(\frac{t}{2\tau_f})] = 4\pi \sum_m \delta(t-2\tau_f m) \qquad (29)$$

a series of uniformly spaced δ-functions, with spacing $2\tau_f$, which is the delay time for an echo in the thinner material. Thus, the dominant asymptotic signal is, as expected, a series of pulses, modulated by an envelope function. The behavior is that which could be found by one dimensional "ray tracing", and is sketched in Fig. (4). The only physical feature not found in this asymptotic development of the Fourier-Floquet representation is that in the ray treatment there is an earliest arrival time, prior to which all quantities must vanish identically. In the present discussion, the envelope function decreases exponentially before the arrival of the body of the pulse, but never vanishes identically.

V. Response Functions and Force-Displacement Relations

(i) We next outline how one may develop the response functions for periodic media. The approach follows quite closely that of the so-called "double-time" Green's functions used in physics, particularly in discussing lattice vibrations.[7]

The problem is that of the forced vibrations, and again we illustrate with a one-dimensional case. We have

$$- \frac{\partial}{\partial x} [\eta(x) \frac{\partial u}{\partial x}] + \rho(x) \frac{\partial^2 u}{\partial t^2} = F(x.t) \tag{30}$$

For a linear system, the general solution may be developed in terms of the displacement Green's function $G(x,x';t-t')$ which is defined by

$$- \frac{\partial}{\partial x} [\eta(x) \frac{\partial G}{\partial x}] + \rho(x) \frac{\partial^2 G}{\partial t^2} = \delta(x-x')\delta(t-t') \tag{31}$$

-403-

For an infinite medium, we can develop G in a time-space Fourier-Floquet representation

$$G(x,x';t-t') = \sum_\nu \int_{-\pi/a}^{\pi/a} dq \int_{-\infty}^{\infty} d\omega e^{i\omega t} g_\nu(q,\omega) u_\nu(x;q) \qquad (32)$$

Substituting (32) in (31) and using the orthonormality properties of the Floquet solutions, we easily find that

$$g_\nu(q,\omega) = \frac{1}{[\omega_\nu^2(q)-\omega^2]} \cdot \frac{e^{-i\omega t'}}{2\pi} u_\nu^*(x';q) \qquad (33)$$

so

$$G(x,x';t-t') = \int_{-\infty}^{\infty} d\omega \frac{e^{i\omega(t-t')}}{2\pi} \overline{G}(x,x';\omega) \qquad (34)$$

where

$$\overline{G}(x,x';\omega) = \sum_\nu \int_{-\pi/a}^{\pi/a} dq \frac{u_\nu(x;q)u_\nu^*(x';q)}{[\omega^2(q)-\omega^2]} \qquad (35)$$

From these quantities one can construct the displacement and velocities for a general force, e.g.,

$$u(x,t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx'dt' \; G(x,x';t-t')F(x',t') \qquad (36)$$

or the Fourier transform

$$\overline{u}(x,) = \int_{-\infty}^{\infty} dx' \; G(x,x';\omega)F(x',\omega) \qquad (37a)$$

$$u(x,t) = \int_{-\infty}^{\infty} e^{i\omega t} \, \overline{u}(x,\omega) \, d\omega \qquad (37b)$$

-404-

In all of the time Fourier transforms, a small imaginary part
of ω may be introduced to ensure causalty.

While these equations constitute formal solutions, it
should be recognized that their application in detail requires
not only the dispersion relations, but also a complete knowledge
of the Floquet displacement functions.

(ii) Since transient loading of a composite commonly
occurs on an external surface, we can exploit the fact that waves
generated by such a time dependent short duration pressure dis-
tribution on a surface can be analyzed within the framework of
the infinitely extended body theory when the surface is a plane
of symmetry of the lattice. This is achieved by first considering
plane wave propagation with reflection and transmission at the
matrix-reinforcement interfaces for the half-space, say, $x \geq 0$.
This is easy to evaluate for the short duration of load application,
but would be cumbersome for long times because of the multiplicity
of reflections and transmissions which continue to occur. If a
tensile variation of surface traction, the negative of the
pressure pulse, is applied to the half-space $x \leq 0$, the surface
velocity at any time will be the same for both half-spaces by
symmetry and the linearity of the problem. Thus the two half-
space surfaces will move with the same velocity and displacement.
They can thus be considered to be in contact and welded together
without modifying the stress wave distribution in either, so
that the displacement and velocity at any instant after pressure

cessation can be used as initial values for the full space, in order to evaluate the subsequent motion. Since surface forces on the half-spaces are then no longer acting, no body forces at the weld arise in the combined problem, so that the theory[1] can be applied.

Because these "initial" distributions of stress and velocity are finite in magnitude and spread over one or more cells of the composite, the Fourier-Floquet integral is likely to exhibit practical convergence, the higher modes not being too highly stimulated. This form of the solution will permit study of the far field by asymptotic methods, with the determination of "head of the wave" contributions. Although these effects would eventually appear in the reflection-transmission procedure mentioned above, they are likely to lag far behind the precursor and thus involve so many interactions that the Fourier-Floquet approach is likely to be advantageous in determining far field response.

Wave Transmission and Reflection Solution

Consider a pressure variation $p(t)$, $t \geq 0$, on the surface of a semi-infinite block, $x \geq 0$, of layered composite, as shown in Fig. 5. The matrix and reinforcement are considered each to be homogeneous linear elastic materials, and hence waves transmitted through them satisfy the wave equation:

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{c_i^2} \frac{\partial^2 u}{\partial t^2} = 0 \tag{38}$$

where c_i, the wave velocity, is equal to c_m in the matrix and c_f in the reinforcement, and is related to the elastic modulus for waves of one-dimensional strain $\eta_i = \lambda_i + 2\mu_i$ and density ρ_i, by the relation:

$$\rho_i c_i^2 = \eta_i \tag{39}$$

for i equal to m or f in the two materials, respectively. The λ_i and μ_i are the Lamé elastic constants. Perhaps the most convenient way of evaluating the waves generated by the applied pressure p(t) in Fig. 5 is to utilize the theory of characteristics (see Ref. 8, p. 119, for example) which determines the relations

$$\rho_i c_i v \pm \sigma = \text{constant} \tag{40}$$

on

$$\frac{dx}{dt} = \pm c_i \tag{41}$$

where v is the particle velocity $\partial u/\partial t$ and σ is the stress. Thus for the initially undisturbed composite block, Fig. 5, with a layer of matrix material of thickness (a-b)/2 at the surface (since the surface is considered to be on the centerline of a matrix slab, which is of width [a-b]) the stress field for $t < (a-b)/2c_m$ is given by

$$\sigma(x,t) = -p(t-x/c_m), \quad x < c_m t \tag{42}$$

$$= 0 \qquad , \quad x > c_m t$$

since tensile stress is considered positive, and the velocity by

$$v(x,t) = -\sigma(x,t)/\rho_m c_m \qquad (43)$$

These results follow from (40), since the constant for the lower sign is zero because the body is initially undisturbed before the pressure is applied, and thus both v and σ are zero throughout the half-space at $t = 0$.

For the half-space $x \leq 0$ and tensile surface traction of magnitude $p(t)$, the dependent variables are related to the above by evenness as a function of x for velocity v and displacement u, and oddness for the stress.

Suppose, for example, that $p(t)$ has the constant value p_o, applied at $t = 0$ and removed at $t = t_i < (a-b)/2c_m$. Then for $t = t_i$, the stress displacement and velocity in the full space are as illustrated in Fig. 6, and these could then be considered as "initial" conditions for the whole space, which will determine the motion of the half-space $x \geq 0$ with the pressure applied to its surface.

If the pressure application continues for $t > (a-b)/2c_m$, the stress wave reaches the first interface at $x = (a-b)/2$ while the pressure is still acting. Reflection and transmission at this interface must therefore be included in the analysis of wave propagation which must be extended through the time until the surface pressure ceases. We must therefore add to the characteristic relations (40) and (41) the requirement that the

-408-

displacement and traction must be continuous across each inter-
face, thus

$$u[(a/2 \pm b/2 \pm na)^+, t] = u[(a/2 \pm b/2 \pm na)^-, t] \qquad (44)$$

$$\sigma[(a/2 \pm b/2 \pm na)^+, t] = \sigma[(a/2 \pm b/2 + na)^-, t] \qquad (45)$$

where

$$\sigma(x,t) = \eta_i \epsilon(x,t) = \eta_i \partial u/\partial x \qquad (46)$$

and n is an integer. Since the arguments in (44) and (45) are
interface positions (see Fig. 5), η_i takes on the values η_f
and η_m on either the plus or minus sides of the interface. In
(46), ϵ denotes the strain.

Combination of (40), (41), (44) and (46) permits the
solution to be extended indefinitely, but the multiple re-
flections and transmissions at the interfaces associated with
repeated application of (44) and (45) with n = ±1, ±2, etc.,
can become cumbersome as t increases. For short time loading,
no difficulty arises. A computer code based on such a procedure
is mentioned by Lundergan and Drumheller[9].

As already described, as soon as the applied pressure
has ceased, this solution can be used for initial values for
displacement and velocity to express the subsequent motion in
Floquet form[1].

There is no need to limit the characteristic solution
presented in this section to prescribed surface pressure. The
solution for prescribed velocity can be obtained from (40) and

(41) in a similar manner, but after some prescribed time the surface must be permitted to move freely without force applied, so that the problem for the two combined half-spaces is free from body force for the Floquet representation part of the analysis. Moreover, impact due to plate slap can also be handled in this way. The impact surface, $x = 0$, would become an additional interface if the impacting plate were not of matrix material, otherwise it would simply be a surface incapable of transmitting tensile stress. The characteristic solution with free rear surface of the impacting plate will lead to rebound and separation of the plate from the surface of the half-space. This time could be used to provide initial conditions for the two half-space problem, the solution of which can be expressed in Floquet form for the subsequent motion. Thus this approach can be applied to a variety of loading situations, to which can thus be applied stationary phase methods for analysis of response in the far field.

Acknowledgement

Bibliography

1. W. Kohn, J. A. Krumhansl and E. H. Lee, J. Appl. Mech 39, 327-36, 1972.

2. J. C. Pack and G. A. Gurtman, J. Appl. Mech. 36, 479-84, 1969.

3. A. Papoulis, "Fourier Integral and Applications," McGraw Hill, New York, 1962.

4. K. Gottfried, "Quantum Mechanics," I.W.A., Benjamin, New York, pp. 100-101, 1966.

5. M. Abramowitz and I. A. Stegan, "Handbook of Mathematical Functions," U.S. Government Printing Office.

6. H. Margenau and G. M. Murphy, "The Mathematics of Physics and Chemistry," Van Nostrand, New York, p. 253, 1956.

7. R. J. Elliot, "Phonons," Scottish University Summer School Lectures 1965, ed. R. S. Stevenson, Oliver and Boyd, 1966.

8. E. H. Lee, "The Continuum Mechanics Aspect of Material Properties Determination," Energetics, Vol. III, pp. 85-122, 1967.

9. C. D. Lundergan and D. S. Drumheller, J. Appl. Phys. 42, 669-675, 1971.

APPENDIX

Orthogonality of Floquet Solutions

The differential equation for steady oscillatory solutions for displacement, $u(x) e^{\pm i\omega t}$, is:

$$\frac{d}{dx} \left(\eta(x) \frac{du}{dx}\right) + \rho(x)\omega^2 u = 0 \qquad (A.1)$$

and the Floquet quasi-periodic boundary conditions are[1]

$$u(a/2) = u(-a/2) e^{iqa} \qquad (A.2)$$

$$u'(a/2) = u'(-a/2) e^{iqa} \qquad (A.3)$$

Following the usual procedure for establishing orthogonality of modes of vibration, $u_\nu(x;q)$, where ν is the mode number, we consider the integral:

$$\int_{-a/2}^{a/2} u_{\nu'}^* \left[\frac{d}{dx}\left(\eta\frac{du_\nu}{dx}\right) + \rho\, \omega_\nu^2\, u_\nu\right] dx = 0 \qquad (A.4)$$

Integrate by parts, interchange ν and ν' and take the complex conjugate. Subtracting this from the original integral gives:

$$(\omega_\nu^2 - \omega_{\nu'}^2) \int_{-a/2}^{a/2} \rho\, u_{\nu'}^*\, u_\nu dx = 0 \qquad (A.5)$$

and hence orthogonality over a single cell $-a/2 \leq x \leq a/2$ for a particular wave number q is established if $\omega_\nu \neq \omega_{\nu'}$.

To consider transients over $-\infty < x < \infty$, we need to

integrate over the complete range of x, and utilize quasi-periodicity, (3), of $u_\nu(x;q)$ in the form

$$u_\nu(x+a;q) = u_\nu(x;q)\ e^{iqa} \qquad (A.6)$$

Then for integers n:

$$\int_{-(a/2)+na}^{(a/2)+na} \rho(x)\ u_\nu(x;q)\ u_\nu^*{}'(x;q')\ dx = e^{i(q-q')na} \int_{-a/2}^{a/2} "\,dx \qquad (A.7)$$

Thus

$$\int_{-(a/2)-Na}^{(a/2)+Na} \rho(x)\ u_\nu(x;q)\ u_\nu^*{}'(x;q')\ dx = \sum_{-N}^{N} e^{i(q-q')na} \int_{-a/2}^{a/2} "\,dx \qquad (A.8)$$

Where N is an integer. But

$$\sum_{-N}^{N} e^{i(q-q')na} = \frac{\sin[(2N+1)(q-q')a/2]}{\sin[(q-q')a/2]} \qquad (A.9)$$

and by an extension of the result cited in reference[6]

$$\lim_{N \to \infty} \left\{ \frac{\sin[(2N+1)(q-q')a/2]}{\sin[(q-q')a/2]} \right\} = \frac{2\pi}{a} \sum_{n=-\infty}^{\infty} \delta(q-q'+2n\pi/a) \qquad (A.10)$$

for n ranging over the integers. Thus from (A.8 and for $q \neq q'$:

$$\int_{-\infty}^{\infty} \rho(x)\ u_\nu(x;q) u_\nu^*{}'(x;q')\ dx = 0 \qquad (A.11)$$

with appropriate consideration of convergence of the improper

-413-

integral. Note that because of the periodicity of the composite structure, only the range $0 \leq q \leq \pi/a$ of wave numbers needs to be considered, so that only one of the delta functions in (A.10) makes a contribution. Combination of (A.5) and (A.11) gives (5) and yields the possibility of selecting normalizing factors for $u_\nu(x;q)$ to give (6).
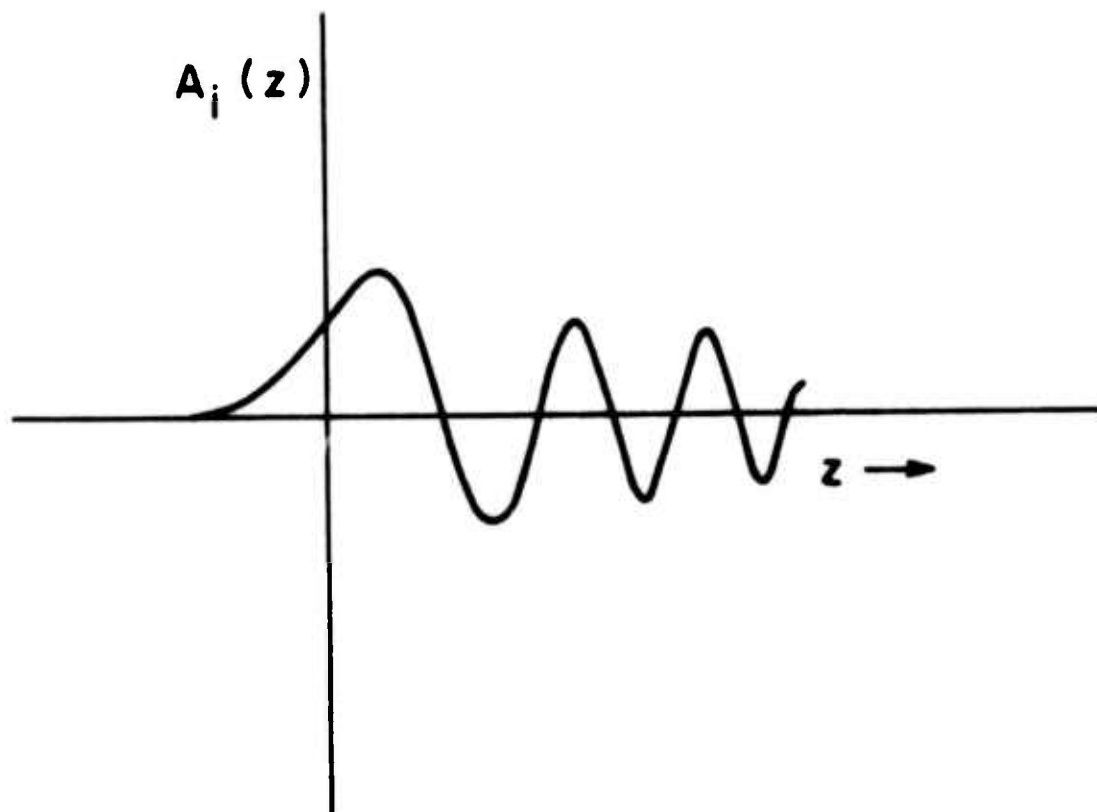
Figure 1. Airy Function Ai(z) vs. z. In the example of layered composites, $z = -(1/2\omega'''t)^{-1/3}(na - \frac{d\omega}{dq}|_{\nu m}t)$ so that Ai(z) vanishes for early times, then rises at $na - \frac{d\omega}{dq}|_{\nu m}t$.
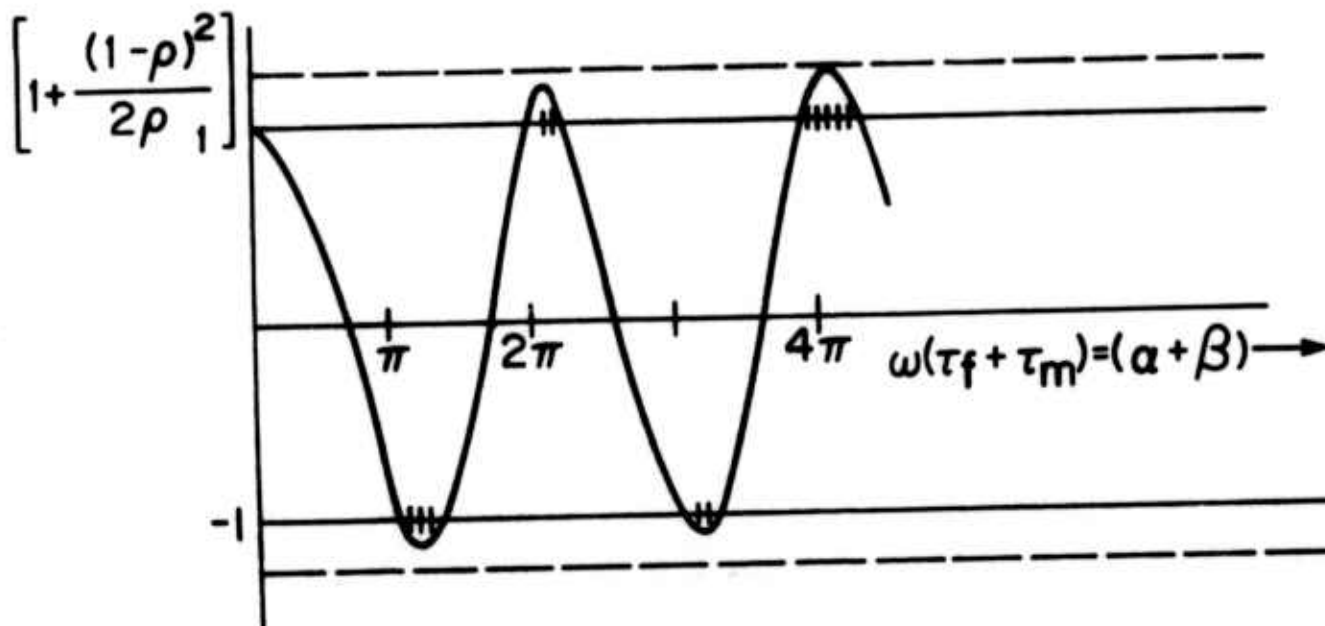
Figure 2. A plot of $[\frac{1}{1-R^2} \cos(\alpha+\beta) - \frac{R^2}{1-R^2} (\alpha-\beta)]$ vs.

$\alpha+\beta - \omega(\tau_f+\tau_m)$.
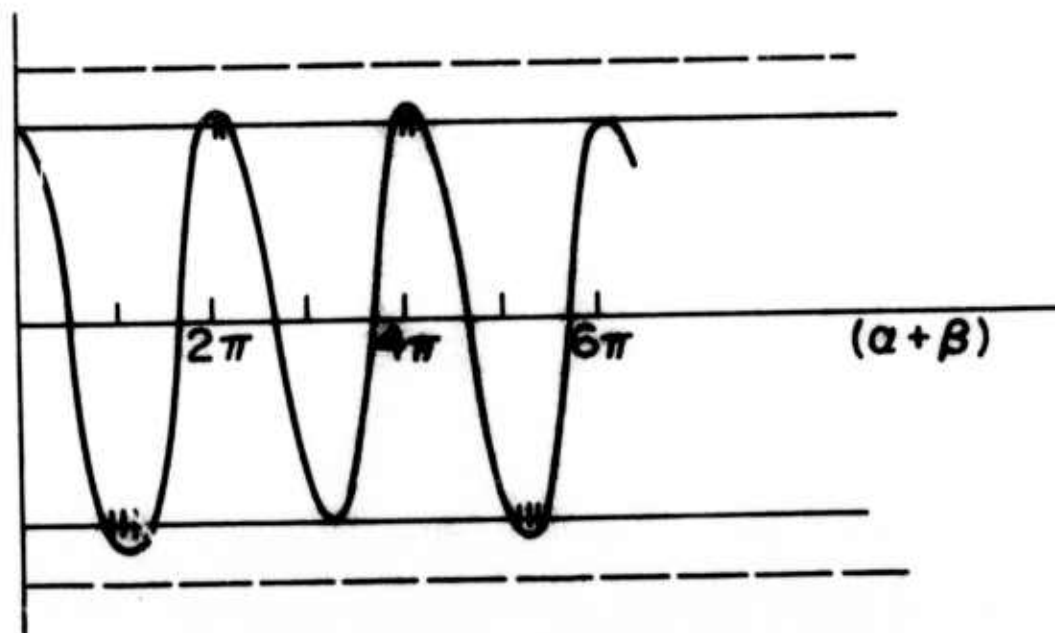
Figure 3.   Graphical determination of dispersion relation for
            commensurate layer transit times.
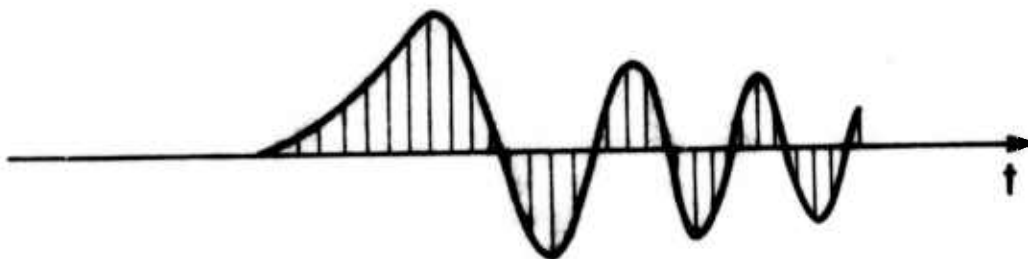
Figure 4. Geometric optics limit for anymptotic propagation
in layered composite with commensurate transit times
$\tau_m = k\tau_f$ in the layers. The lines represent a train
of pulses with envelope given by the Airy function.
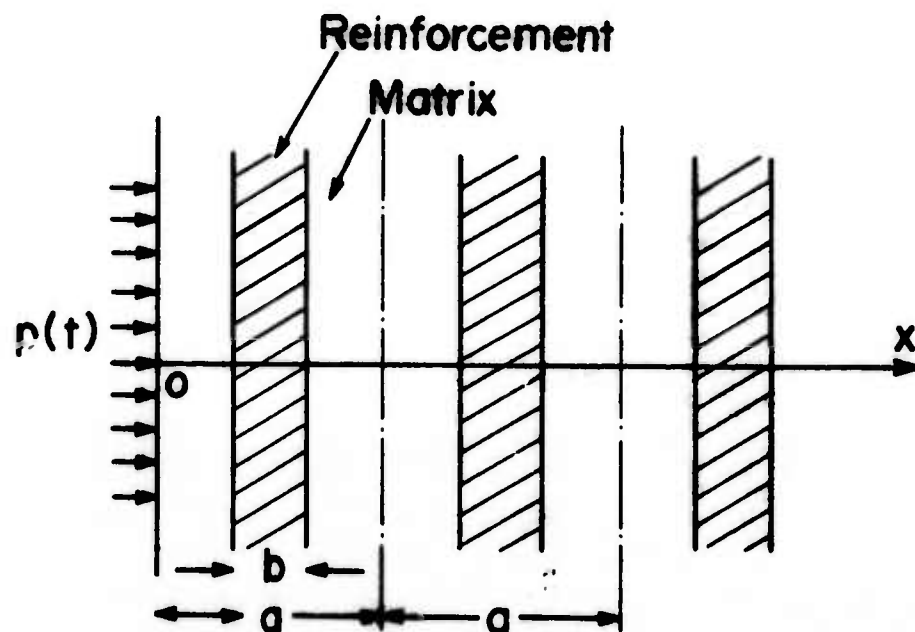
Figure 5.   Composite Half-Space with Applied Surface Pressure
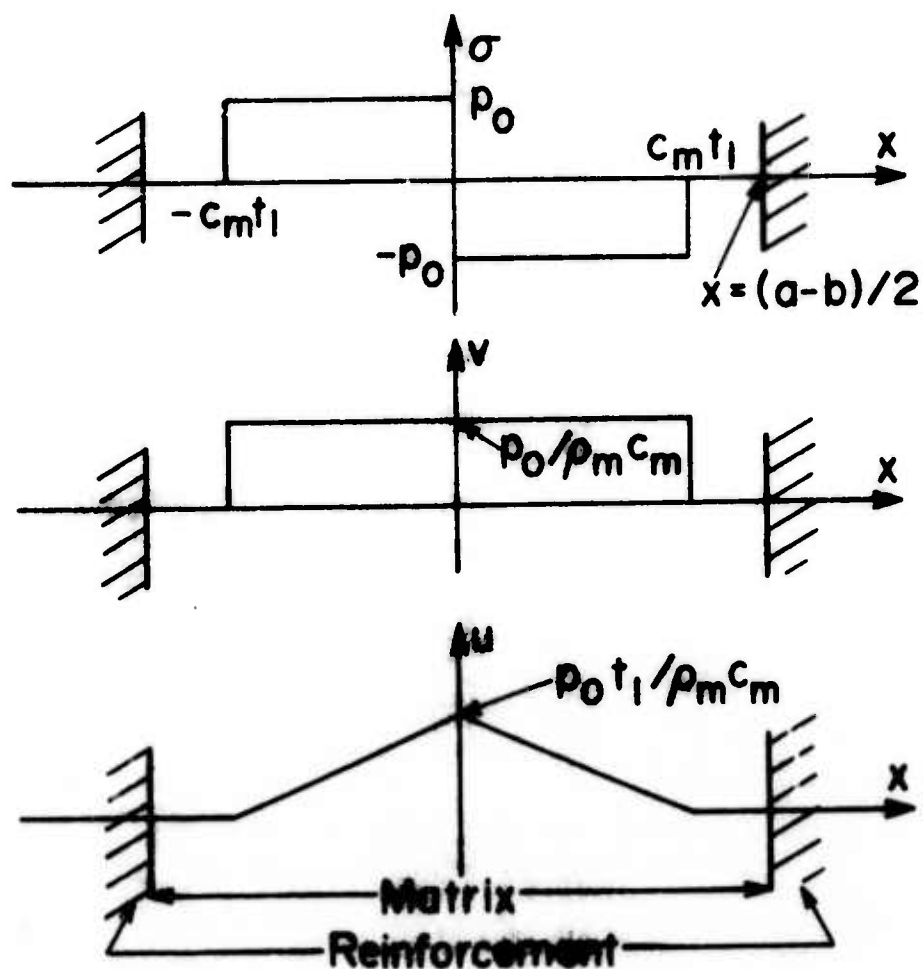
Figure 6. Stress, Velocity & Displacement for Constant Applied
Pressure, p_o, until $t = t_1 < (a-b)/2c_m$.

ON WAVES IN COMPOSITE MATERIALS
WITH PERIODIC STRUCTURE

E. H. Lee
Wei H. Yang*

Abstract

The propagation of waves through non-homogeneous elastic
material, with a periodic structure of elastic constants and
density variation, can be conveniently treated in terms of
Floquet waves. These are quasi-periodic waves, propagating in
one direction, the displacement profiles of which change with
time, but repeat after the wave has traversed a complete cell
of the material. Locally this solution is the superposition of
refracted and reflected components. Basic concepts of the
theory are presented for a laminated composite material and
propagation normal to the laminations. The frequency spectrum
has a banded structure, comprising pass or propagating bands
and stop bands. It is shown that the frequencies at the
boundaries of the bands correspond to wave profiles which are
normal modes of vibration of the individual cells with fixed or
free surfaces. Both types occur at each limiting frequency.
Properties of Floquet waves are interpreted in terms of normal

*Department of Mechanical Engineering, University of Michigan

mode theory. From these considerations the high frequency limit
for Floquet waves is investigated and interpreted in terms of
geometrical optics type analysis.

ON WAVES IN COMPOSITE MATERIALS
WITH PERIODIC STRUCTURE

E. H. Lee
Wei H. Yang

Introduction

In recent years, Floquet or Bloch theory has been applied to the analysis of the propagation of waves through composite media with periodic structures.[1] For layered composites with elastic constant $\eta(x)$ and density $\rho(x)$ varying periodically in x with period a:

$$\eta(x + a) = \eta(x) \tag{1}$$

$$\rho(x + a) = \rho(x) \tag{2}$$

waves propagated in the x direction must satisfy the partial differential equation:

$$\frac{\partial}{\partial x} \left[\eta \frac{\partial U}{\partial x} \right] = \rho \frac{\partial^2 U}{\partial t^2} \tag{3}$$

where U is the displacement in the x direction for longitudinal waves, or perpendicular to this direction for shear waves. We consider steady oscillatory waves of circular frequency ω, and hence write

$$U(x,t) = u(x) \, e^{i\omega t} \tag{4}$$

so that (3) becomes

$$\frac{d}{dx}\left[\eta \frac{du}{dx}\right] + \rho\omega^2 u = 0 \tag{5}$$

By Floquet's theorem there are solutions of (5) of the form

$$u(x) = v(x) \, e^{iqx} \tag{6}$$

where $v(x)$ is periodic of period a and q is the wave number for the wave expression (4). Such solutions for $U(x,t)$, (4), known as Floquet waves, can be evaluated over a single cell $-a/2 \leq x \leq a/2$ on the basis of (5) and the quasi-periodic boundary conditions:

$$u(a/2) = u(-a/2) \, e^{iqa} \tag{7a}$$

$$u'(a/2) = u'(-a/2) \, e^{iqa} \tag{7b}$$

The solution for all x is then given by the quasi-periodic recurrence relation:

$$u(x + a) = u(x) \, e^{iqa} \tag{8}$$

which follows from (6).

Relations (5) and (7) form a Sturm-Liouville system with eigenfunctions $u_\nu(x;q)$ corresponding to the eigen-frequencies, $\omega_\nu(q)$. These satisfy the orthogonality relation[2]:

$$\int_{-a/2}^{a/2} \rho(x) \, u_\nu^*(x;q) u_{\nu'}(x;q) dx = 0, \quad \nu \neq \nu' \tag{9}$$

where superscript * denotes the complex conjugate. Relations

-424-

(5) and (7) yield a denumerably infinite set of eigenfunctions for each value of q. Because of (6), with $v(x)$ periodic of perio a, it is only necessary to consider q limited to the range $-\pi/a \leq q \leq \pi/a$, since for

$$q = q_1 + 2n\pi/a \tag{10}$$

n an integer, the exponential in (6) of the second term of (10) yields a periodic function which can be considered to be in-orporated into the function $v(x)$. Moreover, since in view of (5) and (7),

$$u_\nu(x;-q) = u_\nu^*(x;q) \tag{11}$$

q can be considered limited to $0 \leq q \leq \pi/a$, and a banded frequency spectrum results, comprising bands which transmit Floquet waves, and no-pass bands which do not, as depicted in Fig. 1.

For some purposes it is convenient to work with q covering all positive real values. In the case of a homogeneous medium (3) becomes the simple wave equation with solutions

$$U(x,t) = e^{i[qx + \omega(q)t]} \tag{12}$$

where

$$\omega/q = c, \quad c = \sqrt{\eta/\rho} \tag{13}$$

c being the elastic wave velocity. Relation (12) corresponds to a left going wave, and a right going wave is obtained simply by changing the sign of ω or q. In this case $v(x)$ in (6) is simply unity. It is of course possible to use the reduced zone

scheme for q, $0 \leq q \leq \pi/a$, with a arbitrary, and then v is a complex exponential:

$$u(x) = e^{i2n\pi x/a} e^{i(q-2n\pi/a)x} \tag{14}$$

but this in general is not fruitful. However, it is sometimes useful to utilize the whole positive range of q in the periodic structure case and thus have the extended zone scheme for the wave number. Fig. 2 shows the dispersion curves of Fig. 1 plotted in this manner. By analogy with the homogeneous case, ω/q will then give the phase velocity corresponding to the wave velocity c in the homogeneous case, (12) and (13). This whole question of defining the wave number arises because, for the Floquet wave (6), q is only defined modulo $2\pi/a$, a change in q simply modifying $v(x)$. This is associated with the fact that observation of a wave only at equidistant positions does not uniquely define the wavelength and hence wave number and phase velocity,[3] p. 7.

The determination of the Floquet wave (6) can be conveniently carried out[4] by obtaining two linearly independent solutions of (5), and selecting a linear combination of them to satisfy the boundary conditions (7). For the common situation of a composite with properties symmetric in a single cell:

$$\rho(x) = \rho(-x) \ , \ \eta(x) = \eta(-x) \ , \ -a/2 \leq x \leq a/2 \tag{15}$$

it is particularly convenient to select the basic solutions to be even and odd in the cell. This can be achieved by integrating

-426-

(5) from the origin with zero u´ or zero u there respectively.

If $u_e(x)$ and $u_0(x)$ are the even and odd functions respectively, then

$$u(x) = u_e(x) + C\, u_0(x) \tag{16}$$

satifies the quasi-periodic condition (7a) if

$$C = i\, \frac{u_e(\frac{a}{2})}{u_0(\frac{a}{2})}\, \tan \frac{qa}{2} \tag{17}$$

and (7b) if

$$C = -i\, \frac{u_e´(\frac{a}{2})}{u_0´(\frac{a}{2})}\, \cot \frac{qa}{2} \tag{18}$$

Compatibility of (17) and (18) yields the dispersion relations

$$\omega = \omega_\nu(q) \tag{19}$$

This procedure can be evaluated in terms of trigono-metric functions[4] for the composite consisting of a homogeneous filler material "f" occupying

$$-b/2 < x < b/2 \tag{20}$$

the rest of the cell

$$-a/2 < x < -b/2\ ,\ b/2 < x < a/2 \tag{21}$$

being composed of homogeneous matrix material "m". Suffixes f and m refer to corresponding material constants. The Floquet

wave (16) takes the form:

$$u(x) = \cos\frac{\omega x}{c_f} + i\frac{(\cos\frac{\alpha}{2}\cos\frac{\beta}{2} - p\sin\frac{\alpha}{2}\sin\frac{\beta}{2})}{(\sin\frac{\alpha}{2}\cos\frac{\beta}{2} + p\cos\frac{\alpha}{2}\sin\frac{\beta}{2})}\tan\frac{qa}{2}\sin\frac{\omega x}{c_f}, -b/2 < x < b/2 \quad (22)$$

where $\alpha = \frac{\omega b}{c_f}$, $\beta = \frac{\omega(a-b)}{c_m}$, $p = \frac{\eta_f c_m}{\eta_m c_m}$, c being the elastic wave

speed. The dispersion relation takes the form

$$\cos(qa) = \cos\alpha\cos\beta - \frac{1}{2}(p + \frac{1}{p})\sin\alpha\sin\beta \quad (23)$$

which yields dispersion curves of the type shown in Fig. 1 and 2.
It will be shown that these relations take on particularly simple
forms at the ends of the bands in Fig. 1, that is for qa equal
to 0 or π. Examples of stress wave profiles deduced from (22) as
extended by (8), for other values of qa, are presented in Ref. 4.

Properties at Band Ends

At the ends of the bands in Fig. 1, which correspond to
limiting frequencies between passing bands of frequencies with
real wave numbers, and stopping bands associated with imaginary
wave numbers, qa = 0 or π, and particularly simple motions arise.
For qa = 0, tan(qa/2) = 0, and for qa = π, cot(qa/2) = 0, but
these factors do not, of themselves, determine the combination
(22) since the other factor of $\sin(\omega x/c_f)$ can be singular. The
dispersion relation (23) can be manipulated into either of the
two forms

$$\cos^2\frac{qa}{2} = (\cos\frac{\alpha}{2}\cos\frac{\beta}{2} - p\sin\frac{\alpha}{2}\sin\frac{\beta}{2})(\cos\frac{\alpha}{2}\cos\frac{\beta}{2} - \frac{1}{p}\sin\frac{\alpha}{2}\sin\frac{\beta}{2}) \quad (24)$$

-428-

and

$$\sin^2 \frac{qa}{2} = (\sin\frac{\alpha}{2} \cos\frac{\beta}{2} + p \cos\frac{\alpha}{2} \sin\frac{\beta}{2})(\sin\frac{\alpha}{2} \cos\frac{\beta}{2} + \frac{1}{p} \cos\frac{\alpha}{2} \sin\frac{\beta}{2}) \quad (25)$$

The first root of (23) for $q = 0$, is $\omega = 0$, and

$$u_1(x;0) = \text{constant} \quad (26)$$

using the notation of (9), since (26) clearly satisfies (5) and (7) for q and ω zero.

For $q = 0$, $\omega \neq 0$, the zeros of the two factors on the right hand side of (25) give pairs of values of ω. For the first factor zero, the equation (25) determines the coefficient of $\sin(\omega x/c_f)$ in (22) to be infinity, which determines $\sin(\omega x/c_f)$ to be the eigenfunction. Note that this is equivalent to formally dividing (16) by C, and re-normalizing $u(x)$, to give the eigenfunction $u_0(x)$ when $C \to \infty$. For the second factor of (25) equal to zero, the coefficient of $\sin(\omega x/c_f)$ in (22) is zero, and hence $\cos(\omega x/c_f)$ is the corresponding eigenfunction. Over the whole cell these eigenfunctions are the odd and even functions u_0 and u_e respectively.

Similar results are obtained from (24) for the case $qa = \pi$. Since these eigenfunctions are real or pure imaginary, the Floquet waves given by (4) are simply normal modes of oscillation of the system, all elements moving in phase. There is no propagation, and in conformity with this circumstance the group velocity at the ends of the bands given by (23) is zero.

The eigenfunctions must satisfy both (7a) and (7b) which for $q = 0$ requires

$$u(a/2) = u(-a/2) \qquad (27a)$$

$$u'(a/2) = u'(-a/2) \qquad (27b)$$

and for $qa = \pi$:

$$u(a/2) = -u(-a/2) \qquad (28a)$$

$$u'(a/2) = -u'(-a/2) \qquad (28b)$$

Thus for odd eigenfunctions for $q = 0$, (27a) demands that

$$u(a/2) = u(-a/2) = 0 \qquad (29)$$

and for even eigenfunctions that

$$u'(a/2) = u'(-a/2) = 0 \qquad (30)$$

The latter requires that strain at the boundaries of the cell $-a/2 \leq x \leq a/2$ be zero, hence also the stress by Hooke's law. Thus (29) corresponds to vibration of a rod with fixed ends, and (30) for free ends.

Contrary results apply for $qa = \pi$ according to (28).

We can thus discuss the Floquet waves for the ends of the pass frequency bands in terms of the normal modes for a single cell with fixed and free boundaries. The first few modes are depicted figuratively in Fig. 3. They are shown as sine curves, whereas the actual modes for a composite of the usual

type with stiff filler laminae have sharp discontinuities of gradient[4], but Fig. 3 shows the correct configuration of modes. The modes are labelled according to the notation of (9):

$$u_\nu(x;q)$$

and the reference numerals to the left, correspond to those which label points in Figs. 1 and 2. For the usual case of a composite with more dense and much stiffer filler laminae ($\rho_f > \rho_m$, $\eta_f \gg \eta_m$), the natural frequencies increase with reference integer. Even though the wavelength is the same, mode 2 clearly has lower frequency than mode 3, since for 2 the dense material of the filler is concentrated in the region of maximum displacement thus enhancing the kinetic energy, and the potential energy is reduced since the maximum strain occurs in the softer matrix material. For a homogeneous medium, the modes 2 and 3, and other pairs, can be brought into coincidence by simply translating the boundaries of the cell, there are no frequency gaps, and the curves in Fig. 2 simply reduce to the straight line $\omega = qc$, (13). One could clearly analyze the composite using a cell centered on a matrix lamina rather than a filler lamina, and the same conclusions would be reached with the new p in (22) and (23) being replaced by 1/p and the odd-even property of the eigenfunctions being interchanged.

It is interesting to observe that the orthogonality relation (9) for either $qa = 0$ or π in each case involves modes of both the fixed ends and free ends sequences. For modes within

each sequence, (9) follows from the usual orthogonality property of normal modes, and the integrals (9) involving terms from both sequences reduce to zero because of the evenness-oddness properties deduced already. For example for $q = 0$, only even modes for free ends, and odd modes for fixed ends appear, and this guarantees orthogonality according to (9).

High Frequency Limit

Fig. 3 shows that after the zero frequency mode, the modes for $qa = 0$ and π fall in pairs having the same wave length, the number of waves in the unit cell increasing by one half for each successive pair. Thus modes labelled 2 and 3 contain half a wave, 4 and 5 one complete wave, and so on. This sequence will continue because of the prescription of fixed and free ends, and will apply for more general but still symmetric distribution of properties in the unit cell. Thus as the frequency increases, an additional wave arises by increasing the frequency through two new bands. This corresponds to an increase in q of $2\pi/a$ on the extended zone difinition of q.

For high order modes of a vibrating system, there are asymptotic representations for natural frequencies[5]. These are based on a geometrical optics type result, and effectively show that the mode shape away from the boundaries is independent of the boundary conditions. In the case presently under study, of waves in one dimension for a composite consisting of laminae of two homogeneous media, the wave length in the filler if $2\pi c_f/\omega$, and that in the matrix: $2\pi c_m/\omega$. Thus at high frequency, the

number of waves in a cell is:

$$\frac{b\omega}{2\pi c_f} + \frac{(a-b)\omega}{2\pi c_m} = \frac{\omega}{2\pi}(\tau_f + \tau_m) \tag{31}$$

where τ_f and τ_m are the travel times for waves traversing a filler and matrix lamina respectively. Thus $\Delta\omega$ for an additional wave is:

$$2\pi/(\tau_f + \tau_m) \tag{32}$$

As shown above, this corresponds to $\Delta q = 2\pi/a$. For large ω, this gradient from band to band will dominate the phase velocity, and thus yield a limiting phase velocity of

$$\frac{\Delta\omega}{\Delta q} = \frac{a}{\tau_f + \tau_m} = c_\infty \tag{33}$$

This corresponds to the effective precursor or geometrical optics speed for waves crossing a cell, and gives the shortest transmission time for a local disturbance to influence initially undisturbed points of the composite. The fact that the phase velocity for Floquet waves attains the value c_∞, is to be expected, since Fourier-Floquet integrals[2] must embody the ability to represent the precursor in a transient. The result is consistent with the fact[6] that the propagation of a discontinuity and the high frequency limit for oscillatory waves can both be interpreted by geometrical optics analysis.

Discussion

The dispersion relations for Floquet waves depicted in Figs. 1 and 2, and embodied in (23), give phase velocities which for low frequencies are usually appreciably less than the limiting phase velocity c_∞. Group velocities are generally much smaller, since they are associated with the gradients of the curves in Fig. 2, with the phase velocities being determined by the gradient of the secant from the origin. As the filler material becomes stiffer relative to the matrix material, the stop frequency bands become wider and the pass bands narrower. Hence the group velocities decrease while the phase velocities maintain their magnitude. The limit c_∞ tends to become independent of the elastic constant η_f, since τ_f decreases towards zero as η_f, and hence c_f, increase without limit. In order to assess the changes in phase velocity, it is interesting to compare the limiting phase velocity at zero frequency, c_0, with the high frequency limit, c_∞, (33). The former[1] is given by:

$$\frac{1}{c_0^2} = (\rho_f \frac{b}{a} + \rho_m \frac{a-b}{a})(\frac{b}{a\eta_f} + \frac{a-b}{a\eta_m}) \tag{34}$$

The inequality:

$$c_\infty^2 > c_0^2 \tag{35}$$

can be reduced to

$$(\rho_f c_f - \rho_m c_m)^2 > 0 \tag{36}$$

Thus (35) is guaranteed unless

$$\rho_f c_f = \rho_m c_m \tag{37}$$

which is the condition for no wave reflection at interfaces between the filler and matrix. Thus c_0 decreases relative to c_∞ as the acoustic disparity between the two materials increases. For the condition (37), waves of all frequencies simply travel through the composite without distortion, with speeds c_f through the filler and c_m through the matrix. This would be an unusual condition for a composite, and in general the relative speeds given in the body of the paper determine the dispersive properties of a laminated composite material.

Acknowledgement

References

1. W. Kohn, J. A. Krumhansl and E. H. Lee, J. Appl. Mech. $\underline{39}$, 327-336, 1972.

2. J. A. Krumhansl and E. H. Lee, "Propagation of Transient Elastic Waves in Periodic Composites," to appear in J. Appl. Mech.

3. L. Brioullin, Wave Propagation in Periodic Structures, Dover Publications Inc., 1953.

4. E. H. Lee, "A Survey of Variational Methods for Elastic Wave Propagation Analysis in Composites with Periodic Structures. Dynamics of Composite Materials," ed. E. H. Lee, ASME, 122-138, 1972.

5. R. Courant and D. Hilbert, Methods of Mathematical Physics, I, Interscience Inc., p. 336 ff., 1953.

6. R. D. Luneburg, Mathematical Theory of Optics, University of California Press, 1964.
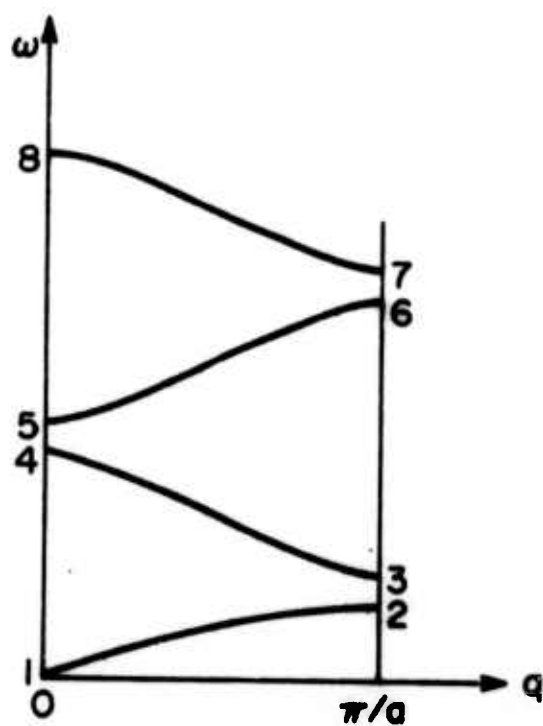
Figure 1. Frequency spectrum using the reduced zone scheme for wave number q.
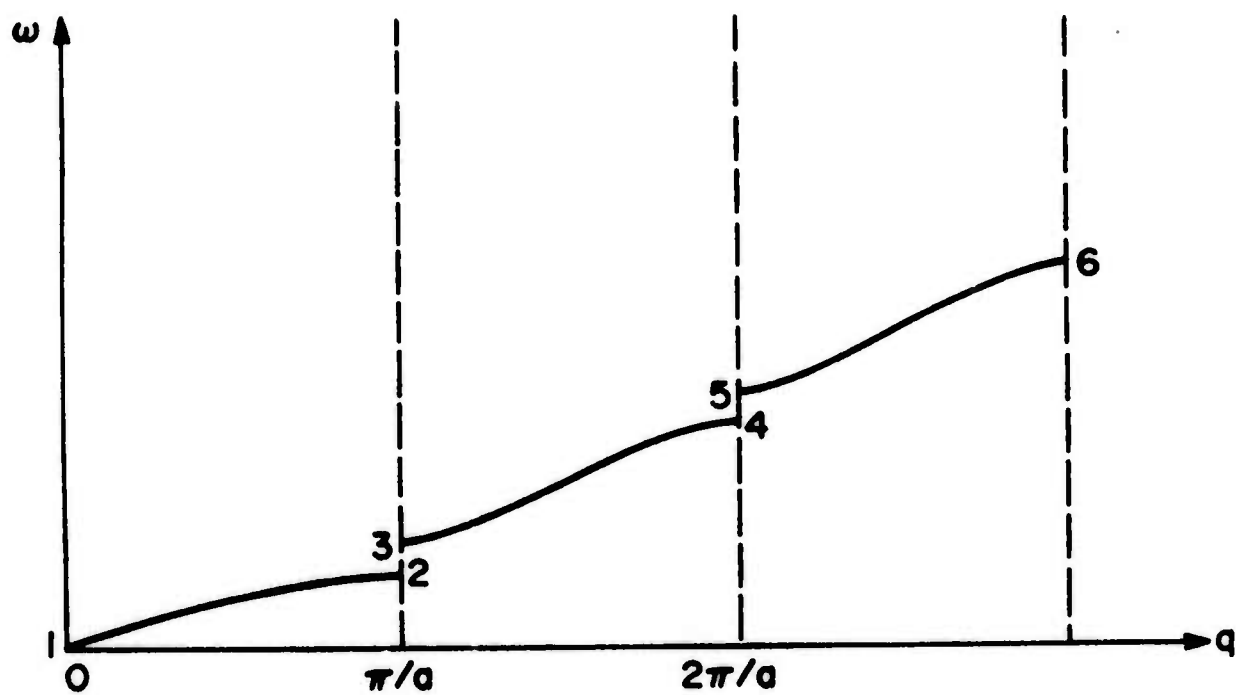
Figure 2.   Frequency spectrum (dispersion curves) plotted on
the extended zone scheme for q.
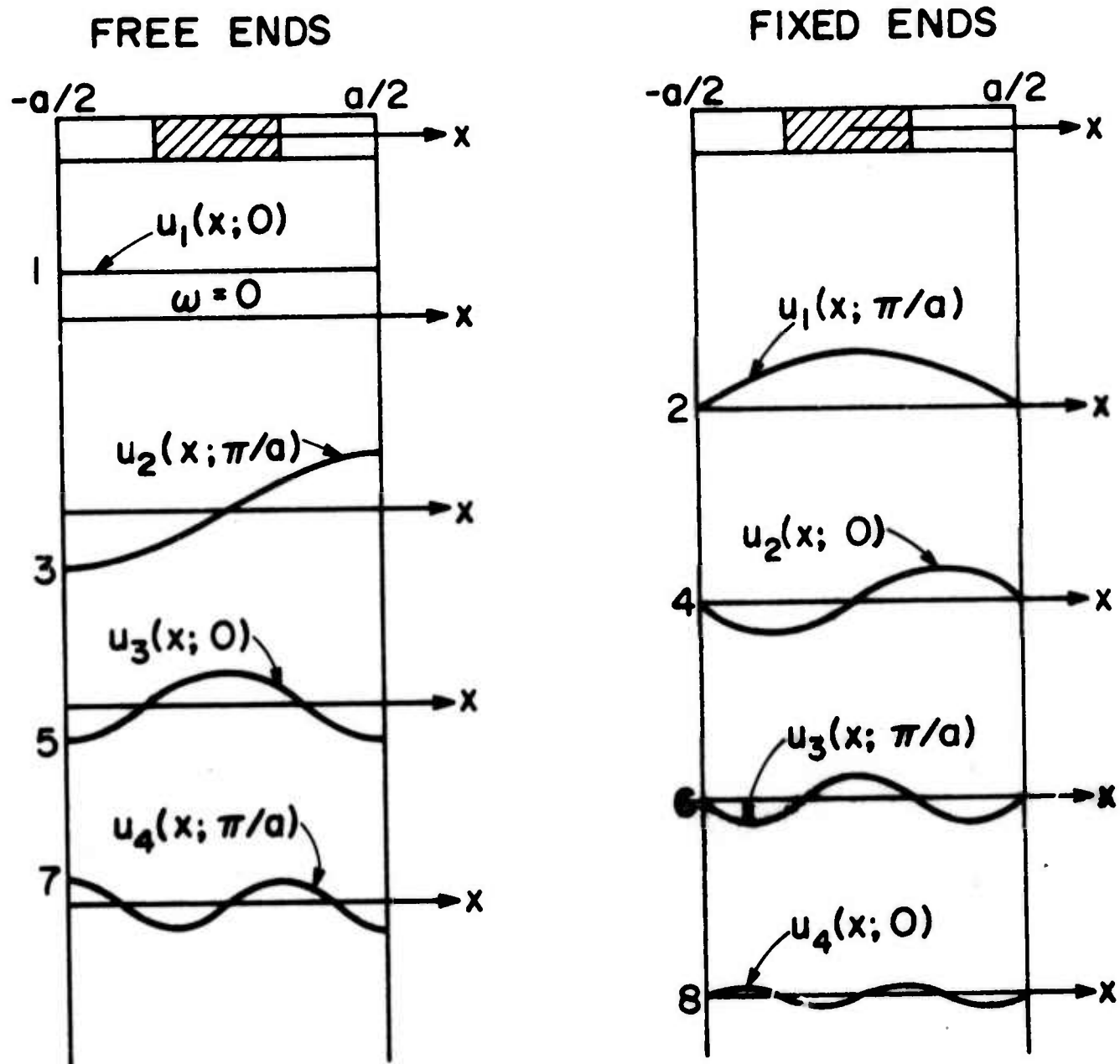
Figure 3. Normal modes for free and fixed ended composite cells.

ELASTIC SOLUTIONS FOR IMPERFECTIONS
IN BRITTLE MATERIALS

F. A. McClintock

Abstract

An annotated bibliography of 42 references is given to help provide research workers in the field of ceramics and other brittle materials with a collection of the information, especially applicable to their field, which is available from the theory of elasticity. The categories of references include inclusions, corners in inhomogeneous or dissimilar materials, elastic constants of polycrystals, three-dimensional solutions, cracks in dissimilar materials, and arrays of cracks. The author would welcome any additions to this brief collection.

ELASTIC SOLUTIONS FOR IMPERFECTIONS
IN BRITTLE MATERIALS

F. A. McClintock

Introduction

The following collection of references was gathered during the ARPA Materials Research Council in 1972 to help provide research workers in the field of ceramics and other brittle materials with a collection of the information available from the theory of elasticity. This collection is necessarily brief, not only because of the shortness of time; but also because of the limited perspective and lapses in memory of the author. Any further suggestions will be welcomed. That the collection is possible at all is due to the efforts of the Research Council staff, in particular Molly Garfin and Michael Gordon.

In reading the papers, one is struck with the need for both the newer and the older work in the field. The newer work is both more concise and general, but it is at the same time much more difficult to interpret physically and, in fact, in many cases I have had to give up. The pioneering work in a field is more likely to give a good introduction to the analysis and to quote the results in concrete terms.

-442-

1. Inclusions

Holes and rigid inclusions in the shape of spheres and circular cylinders were reported by Goodier (1933) and Donnell (1941). Neither of these authors recognized at the time the constancy of stress that holds within a two-dimensional elliptical inclusion and which was shown also for the 3-dimensional ellipsoid, if my recollection is correct, by Eshelby (1957) and (1961). The effects of elastic modulus of the inclusion can be roughly linearly interpolated between the limiting cases of holes, uniform material, and the rigid inclusion. For spheres, stress concentrations are of the order of two.

Sternberg (1958) gave a review of 3-dimensional solutions in elasticity, some of which will be discussed below. Of particular interest here is the solution for two spherical holes close together, showing only a 5% interference effect even when there is only one diameter between the holes (Sternberg and Sadowsky, 1952). Similar two-dimensional solutions show more interference and also show infinite stress concentrations as holes approach a free surface.

2. Corners in homogeneous or dissimilar materials

In the pioneering work on this topic, (Williams, 1952) found that the stress singularity near a corner varied as r^{-n} where $\underline{n}$ varies from 0 for a straight boundary to 1/2 for a sharp crack. Thus a 90° corner is much less serious than a sharp crack in producing a stress some point a microscopic dis-

-443-

tance below the tip of the corner. It was this work which
later led Williams to find the stress and strain distributions
about cracks which, along with the contemporary work of Irwin,
was the basis for linear elastic fracture mechanics. Solutions
for square cutouts were given by Savin (1965) in his book on
stress concentrations around holes, and work in Holland by Koiter
should also be reviewed. Karp and Karal (1962) and Kurshin (1959)
also worked on this problem. Lukas and Erdogan (1966) included
the effects of thermal stress.

For corners or wedges with dissimilar materials, the
first solution was by Knein (1927) for a 90° wedge bonded to an
infinite rigid plane. Recent papers by Bogy (1968), (1970),
(1971) provide general solutions for arbitrary material constants
and wedge angles. The presentation is simplified by Dundurs'
(1969) grouping of the four elastic constants into just two
independent terms. While this grouping makes presentation of
the general case possible, it also obscures the general result
and it may be preferable to refer to papers by Hein and Erdogan
(1971), who presented their results more directly using a
Poisson's ratio of 0.20 for cne of the materials, which is
appropriate for semi-conductors bonded to substrates.

3.   Elastic constants of polycrystals

Although specific references have not been located here,
the problem of estimating the average elastic constants from
those of polycrystals has received much attention recently,

particularly in view of the problem of estimating the modulii of composites. Upper and lower bounds to the estimates can be obtained by assuming either homogeneous strain or stress (the Voigt and Reuss models). Kroner in a German publication showed that the root mean square of two estimates was often very accurate. Estimates taking some account of the configuration of the structure have been given by Hill in the Journal of the Mechanics of Physics of Solids and also by Hashin.

4.   Three-dimensional solutions

Sternberg (1958) reviewed work to that date. One of the classical problems is that of a cylindrical hole breaking through to a surface, which was treated by him and Sadowsky (1949), but he recommends the later work by Alblas (1957). It is my impression that the plane strain solution is valid at depths large compared to disturbed field of the hole (perhaps 10 times the diameter). The displacement distributions are much closer to the plane strain solutions than are the stresses. The surface of the body, even though traction free, is not under plane stress conditions because of the influence of the stress gradients normal to the surface on the equilibrium equations.

With cracks imbedded in a body the original solution for a penny-shaped crack first obtained by Green and Sneddon (1950), when compared with the stress intensity factor determined by Irwin (1962), indicates not much difference between the two and three dimensional problems. When cracks approach a free

surface, the stress concentration can increase markedly, as shown approximately by Kobayashi et al. (1965).

The 3-dimensional problem of a crack emerging to a surface, with its leading edge at 90° to the surface, has been studied numerically by Cruse and Van Buren (1971). Again they found that the plane strain displacement field extended much closer to the free surface than did the stress field. The analytic character of the singularity is not clearly determined from their work, although it seems to me that it should have a conical shape.

The other numerical methods for 3-dimensional problems are those by Argyris (1965, see also his book) and Rizzo and Shippy (1961) who dealt specifically with the inhomogeneous elastic inclusion problem.

5. Cracks in dissimilar materials

In the pioneering paper on this topic, Williams (1959) showed that stresses are of the order $\sigma = r^{-\frac{1}{2}} \sin(\lambda \ell n r)$, indicating an oscillatory behavior. The oscillations should lie within 10^{-3} crack lengths of the tip. There is a possibility of a higher order singularity, of order r^{-1}, which might be excited by nonlinearities or inhomogeneities near the crack tip. This problem was further extended by Rice and Sih (1965), extended to cracks normal to a bi-material interface by Zak and Williams (1963) and to a number of cracks along a boundary by Salganik (1963) and Cherepanov (1962). Both Salganik and

-446-

England (1964) found that overlapping displacements might occur in the region close to the crack tip. Finally Bogy (1971) studied the crack approaching an interface from any angle. It seems that many physical conclusions of these studies are yet to be drawn in the light of possible changes in elastic constants between phases, structural inhomogeneities at the scale of grains or smaller, and appropriate fracture criteria derived from a still finer scale.

6.   Arrays of cracks

Arrays of cracks have been studied beginning with the approximate work of Westergaard (1939) and running through a variety of authors including Smith (1965) and Yokobori et al. (1965, 1967), Ichikawa et al. (1965), Ohashi et al. (1965). A very complete numerical method was prepared by Isihida (1969), who checked his computer results against a number of known analytical solutions. Essentially the result is that cracks must be within a half length of each other before the local stress intensity factors rise by more than 10 or 20%. He did not, however, look at the problem of the T-shaped crack in which one crack cuts across the path of another.

Acknowledgement

REFERENCES

Alblas, J. B.   1957   "Theory of the Three-dimensional state of stress in a plate with a hole (in Dutch), Dissertation, Technische Hogeschool Delft; Amsterdam, H. J. Paris

Argyris, J. H.   1965   "Matrix Analysis of Three-Dimensional Elastic Media, Small and Large Displacements", AIAA J. $\underline{3}$, 45.

Bogy, D. B.   1968   "Edge-Bonded Dissimilar Orthogonal Elastic Wedges Under Normal and Shear Loading," J. Appl. Mech $\underline{35}$, 460-466.

Bogy, D. B.   1970   "On the Problem of Edge-Bonded Elastic Quarter-Planes Loaded at the Boundary," Int. J. Solids Structures $\underline{6}$, 1287-1313.

Bogy, D. B.   1971   "On the Plane Elastostatic Problem of a Loaded Crack Terminating at a Material Interface," J. Appl. Mech. $\underline{38}$, Ser. E, No. 4, 911-918.

Bogy, D. B.   1971   "Two Edge-Bonded Elastic Wedges of Different Materials and Wedge Angles Under Surface Tractions," J. Appl. Mech. $\underline{38}$, 377-386.

Chen, W. T.   1968   "Axisymmetric Stress Field Around Spheroidal Inclusions and Cavities in a Transversely Isotropic Material," J. Appl. Mech. $\underline{35}$, Trans. ASME, Vol. 90, Ser. E, 770-773.

Cherepanov, G. P.   1962   Izv. Akad. Nauk SSR, OTN, No. 1, 131-137.

Cruse, T. A.   1969   "Numerical Solutions in Three-Dimensional Elastostatics," Int. J. Solids Struct. $\underline{5}$, 1259.

Cruse, T. A.   1970   "Lateral Constraint in a Cracked, Three Dimensional Elastic Body," Int. J. Fract. Mech. $\underline{6}$, 326-328.

Cruse, T. A.   1971   "Three-Dimensional Elastic Stress
    VanBuren, W.   Analysis of a Fracture Specimen with
                                an Edge Crack," Int. J. Fract. Mech.
                                7, No. 1, 1-15.

Dundurs, J.   1969   "Edge-Bonded Dissimilar Orthogonal
                                Elastic Wedges Under Normal Shear
                                Loading," J. Appl. Mech. 36, 650-652.

Erdogan, F.   1963   "Stress Distribution in a Nonhomogeneous
                                Elastic Plane with Cracks," J. Appl.
                                Mech. 30, Trans. ASME, vol. 85, Series
                                E, 232-236.

Erdogan, F.   1969   "Stresses in Fiber-Reinforced Composites
    Ozbek, T.   with Imperfect Bonding," J. Appl. Mech.
                                36, 865-869.

Eshelby, J. D.   1957   "The Determination of the Elastic Field
                                of an Ellipsoidal Inclusion and Related
                                Problems," Proc. Roy. Soc. London (A)
                                241, 376-396.

Eshelby, J. D.   1961   "Elastic Inclusions and Inhomogeneities,"
                                Progress in Solid Mechanics II, I. N.
                                Sneddon and R. Hill, Eds., North Holland
                                Publishing Co., Amsterdam, 89-142.

Goodier, J. N.   1933   "Concentration of Stress Around Spher-
                                ical and Cylindrical Inclusions and
                                Flaws," Trans. ASME 55, 7, 39-44.

Green, A. E.   1949   "The Distribution of Stress in the
    Sneddon, I. N.   Neighborhood of a Flat Elliptical
                                Crack," Proc. Camb. Phil. Soc. Pt. 1
                                46, 159.

Hein, V. L.   1971   "Stress Singularities in a Two-Material
    Erdogan, F.   Wedge," Int. J. Fract. Mech. 7, No. 3,
                                317-330.

Ichikawa, M.   1965   "Interaction Between Parallel Cracks
    Ohashi, M.   in an Elastic Solid and Its Effect on
    Yokobori, T.   Fracture," Reports of Res. Inst. for
                                Strength and Fracture of Materials,"
                                Tohoku U., Sendai, Japan, Vol. 1, No. 1,
                                1-14.

Ishida, M.   1969   "Analysis of Stress Intensity Factors in a Plate with any Given Distribution of Cracks," Nihon Kikai Gakkai Rombunshu (Trans. Japan Soc. Mech. Engineers), 35, 1815-1822. Translated in Reports and Notes of the ARPA Materials Summer Conference, 1971, Vol. 1, pp. 483-514, University of Michigan.

Karp, S. N.
Karal, F. C.   1962   "The Elastic-Field Behavior in the Neighborhood of a Crack of Arbitrary Angle," Communications on Pure and Applied Math., Vol. XV, 413-421.

Kirsch, G.   1898   "Die Theori der Elastizitaet und die Bedürfnisse der Festigkeitslehre," Z. VDI 42, 797.

Knein, M.   1927   "Zur Theorie der Druckversuchs," Abhandlungen der Aerödynamische Inst. u.d. Technische Hochschule, Aachen, Germany, Vol. 7, 43-62.

Kobayashi, A. S.
Ziv, M.
Hall, L. R.   1965   "Approximate Stress Intensity Factor for an Embedded Elliptical Crack Near Two Parallel Free Surfaces," Int. J. Fract. Mech. 1, 81-89.

Kurshin, L. M.   1959   "Mixed Plane Boundary Value Problem of the Theory of Elasticity for a Quadrant," Appl. Math. and Mech. (transl. PMM) Vol. 23, No. 5, 1403-08.

Lucas, R. A.
Erdogan, F.   1966   "Quasi-State Transient Thermal Stresses in an Infinite Wedge," Int. J. Solids and Structures 2, 205-222.

Ohashi, M.
Ichikawa, M.
Yokobori, T.   1965   "The Interaction of Two Parallel Elastic Cracks Not Perpendicular to Each Other," Reports of the Res. Inst. for Strength and Fracture of Materials, Tohoku U., Sendai, Japan, Vol. 1, No. 1, 1-14.

Rice, J. R.
Sih, G. C.   1965   "Plane Problems of Cracks in Dissimilar Media," J. Appl. Mech. 32, Trans. ASME, 87, Series E, 418.

Rizzo, F. J.   1968   "A Formulation and Solution Procedure
    Shippy, D. J.   for the General Non-Homogeneous Elastic
                                Inclusion Problem," Int. J. Solids
                                Struct., $\underline{4}$ 1161.

Salganik, R. L.   1963   PPM (translated by ASME), Vol. 27,
                                1468-1478.

Savin, G. N.   1961   Stress Concentration Around Holes,
                                Pergamon Press, London.

Smith, E.   1966   "Fracture at Stress Concentrations,"
                                Proc. 1st International Conf. on
                                Fracture, Vol. 1, Sendai, Japan,
                                133-151.

Sternberg, E.   1958   "Three-Dimensional Stress Concentrations
                                in the Theory of Elasticity," J. Appl.
                                Mech. Reviews $\underline{11}$, No. 1, 1-4.

Sternberg, E
    Sadowsky, M. A. 1949   "Three-Dimensional Solution for the
                                Stress Concentration Around a Circular
                                Hole in a Plate of Arbitrary Thickness,"
                                J. Appl. Mech. $\underline{16}$, 1, 27.

Sternberg, E.
    Sadowsky, M. A. 1952   "On the Axisymmetric Problem of the
                                Theory of Elasticity for an Infinite
                                Region Containing Two Spherical
                                Cavities," J. Appl. Mech. $\underline{19}$, 1, 19.

Westergaard, H. M. 1939   "Bearing Pressures and Cracks," J. Appl.
                                Mech., Trans. ASME $\underline{61}$, 49.

Williams, M. L.   1952   "Stress Singularities Resulting From
                                Various Boundary Conditions in Angular
                                Corners of Plates in Extension," J.
                                Appl. Mech. $\underline{19}$, 526-529, Trans. ASME
                                $\underline{74}$.

Williams, M. L.   1959   "The Stresses Around a Fault or Crack
                                in Dissimilar Media," Bulletin of the
                                Seismological Society of America $\underline{49}$,
                                No. 2, 199-204.

Yokobori, T.   1965   "The Interaction of Two Collinear
    Ohashi, M.   Asymmetrical Elastic Cracks," Reports
    Ichikawa, M.   of the Res. Inst. for Strength and
                                Fracture of Materials, Tohoku U.,
                                Sendai, Japan, Vol. 1, No. 2, 33-40.

Yokobori, T.   1967   "The Interaction of Parallel Elastic
  Ichikawa, M.   Cracks and Parallel Slip Bands Re-
                             spectively Based on the Concept of
                             Continuous Distribution of Dislocations
                             II," Reports of the Res. Inst. for
                             Strength and Fracture of Materials,
                             Tohoku U., Sendai, Japan, $\underline{3}$, 15-37.

Zak, A. R.
  Williams, M. L.   1963   "Crack Point Stress Singularities at
                             a Bi-Material Interface," J. Appl. Mech.
                             $\underline{30}$, 142-143.

MODELS OF SPALL FRACTURE BY HOLE GROWTH

F. A. McClintock

Abstract

After an introduction to the static fracture of metals
by hole growth, the physically possible sources of rate effects
are reviewed: those arising from inertia terms as in wave motion;
from quantum mechanical tunnelling effects; and from thermal
activation. Limitations due to the size of the hole relative to
the mean free path of slip lines are also estimated. It is con-
cluded that dislocation viscosity aided by temperature rise,
thermal diffusivity and mechanical inertia play markedly different
roles in different alloys at different intensities of loading.
The largest uncertainty is in the mobile dislocation density of
hard alloys. In an extensive author's discussion, it is concluded
that inertia effects on hole growth are negligible when the loading
time exceeds 20 to 50 nanoseconds for high and low strength alloys,
respectively. Otherwise the strength under shock loading is of
the order of four times the yield strength.

MODELS OF SPALL FRACTURE BY HOLE GROWTH

F. A. McClintock

I.   Static Fracture by the Growth of Holes

Since scanning electron micrographs show strong similar-
ities between static and dynamic fracture by hole growth, it is
worth reviewing the stages in static fracture, as they are
currently understood.

Nucleation can occur in a perfect lattice at a stress
related to the modulus of elasticity E by the shape of the inter-
atomic force law. Bubble raft studies by McClintock and O'Day
(1965), originally motivated by observations of spall fracture,
gave values of E/10 to E/20. Estimates can also be obtained from
a power-law estimate of the interatomic forces, but Lomer (1949)
felt that the bubble model is more accurate for copper, at least.
Since the atomic frequency is of the order of 10^{13}/sec, similar
values would be expected for dynamic fracture. Such a theoret-
ical strength was apparently obtained by McQueen and Marsh (1962).
The bubble model shows that dislocations and grain boundaries
reduce the strength by a factor of two. Very possible the mosaic
pattern of nucleation observed by Seaman et al. (1971) in tests
on 99.999% pure aluminum at E/25 indicates nucleation at a dis-
location substructure.

More commonly, hole growth nucleates at second phases, either by cracking of brittle phases or by separation of the inclusions from the surrounding matrix. Once nucleated, holes grow by plastic flow. Both the deformation and the triaxial stress applied to the surrounding matrix play a role, as indicated by the following approximate equation giving the change in semi-axis $\underline{b}$ of holes with spacing $\underline{\ell}_b$ in material that strain-hardens according to the law

$$\bar{\sigma} = \bar{\sigma}_1 \bar{\epsilon}^{\bar{n}} \quad , \tag{1}$$

and is subject to transverse principal stress components $\underline{\sigma_a^\infty}$ and $\underline{\sigma_b^\infty}$:

$$\frac{d\ell n(2b/\ell_b)}{d\bar{\epsilon}^\infty} = \frac{\sinh[(1-n)(\sigma_a^\infty + \sigma_b^\infty)/(2\bar{\sigma}/\sqrt{3})]}{(1-n)} \tag{2}$$

Localization of void growth into a sheet which does not deform in its plane, but across which there is relative motion, is the next stage of fracture, as recently pointed out by Berg (1972). This localization should be reduced by the spatial variations in stress accompanying dynamic fracture. More diffuse fractures are possible dynamically, because the straining at one region can be continuing for a moment after a nearby region has begun to unload by hole growth.

The final stage in formation of a crack is the separation process across the localized sheet of holes. As shown recently by Nagpal, et al., (1972), the traction falls off nearly linearly

with displacement, and the work per unit area during this final separation process is of the order of the yield strength Y times the spacing of the holes, ℓ.

Crack growth generally occurs by re-nucleation of holes ahead of the crack, followed by joining up. Statically, this process favors holes at either side of the crack tip, as shown by Carson (1970), and for viscous flow by Cipolla (1972). On a more macroscopic scale, the fracture criteria can be provided by the well-known equations of linear elastic fracture mechanics, which uniquely define the stress in an elastic region distant from the plastic zone but close compared to the total crack length or the distance to the next nearest boundary. This stress distribution is distorted by dynamic effects, but will not be discussed here.

II. Rate Effects

Rate effects in materials can only occur in one of three ways: by inertia terms, as in wave motion; by quantum mechanical tunneling effects; and by the time required for thermal activation, including any temperature effects due to adiatatic temperature rise.

For interpreting experimental results and estimating the magnitudes of effects, it is convenient to have a number of parameters worked out in consistent units such as Table 1. The properties in Table 1 were taken to be those of the pure metal except for the tensile strength, which was taken to be that of the strongest common commercial alloy. The loading parameters

TABLE 1.   Parameters for Spall by Hole Growth

Material: condition:	Aluminum pure, alloy		Copper pure, alloy		Iron pure, alloy	
Young's modulus, E, ksi kbar $\simeq 10^3$ atm	10,300 710		17,500 1210		30,000 2070	
Poisson's ratio, ν	.33		.345		.28	
Tensile strength, T.S., kbar	.47	6.1	2.2	13.8	2.1	20.6
Density, ρ, g/cm^3	2.7		8.9		7.7	
Specific heat, cal/g°C	.22		.093		.11	
Thermal conduct. cal/sec cm°C	.49		.92		.11	
Loading parameters						
Uniaxial strain modulus $E_1 = E(1-\nu)/((1+\nu)(1-\nu))$, kbar	1050		1900		2660	
Uniaxial strain wave velocity, $c_1 = \sqrt{E_1/\rho}$, cm/μsec	.62		.46		.59	
Shear wave velocity, c_2, μsec	.215		.225		.324	
Stress/impact velocity, $\sqrt{\rho E_1}$, kbar/(cm/μsec)	1685		4100		4530	
Pulse duration/flyer thickness, $2/c_1$, μsec/cm	3.2		4.3		3.4	
Thermal parameters						
Temperature rise for $r_2/r_1=10$, $2\ell n10$(T.S.)/ρC, °C	90	1150	300	1850	380	2700
Melting point °K	950		1380		1830	
Thermal diffusivity $\alpha = k/\rho C$, cm^2/sec	.82		1.11		.13	
Quench time for a 10μ hole R^2/α, μsec	.3		.22		1.9	
Viscous parameters						
Viscosity, $.3kT/b^4 \Lambda c_2$, poise	5400		12,900		9800	
Strain rate for viscous stress = T.S. at 10^8 disl./cm^2, μsec^{-1} (end of rate-indep. plast)	.044(.45)		.085(.29)		.15(.40)	
Strain rate at 10^8 disl./cm^2 & .5 $c_2, \dot{\gamma} = b\Lambda.5c_2$, μsec^{-1} (end of linear viscosity)	.45		.29		.40	
Viscous stress at .5 c_2, $\mu\dot{\gamma}$ (end of linear visc.) kbar	2.4		3.7		3.9	

are calculated on the basis of elastic collisions which should
be valid when the stress pulses are only a small fraction of
the modulus. In the expression for pulse duration per unit via
thickness, the factor 2 arises from the total length of the
pulse being set by its travel through the flier and return. This
total length also governs the time of loading, for a square wave
it is the time required for the reflected wave to pass the spall
plane. Further discussion of rate effects due to inertia will
be deferred to the last section.

Quantum mechanical effects would be more important than
effects due to thermal motion only under conditions for which

$$h\nu > kT. \tag{3}$$

The frequency ν of concern here is not $\nu = 10^{13}$/sec, which holds
for individual atoms, but something more like $\nu = 10^{11}$/sec, which
is the vibrational frequency of the dislocations pinned against
an obstacle. (Teutonico, et al. 1964) This implies that thermal
motion will free a dislocation sooner than tunneling provided the
temperature is greater than about 1/100 of the Debye temperature,
T_d:

$$T \geq T_D/100. \tag{4}$$

Since this limit is only a few °K, quantum mechanical tunneling
will be be important in these problems.

Under ordinary conditions, the strain rate effects in

reasonably hard metals are quite unimportant. For instance in expressing the strain rate as a power function of stress, exponents are of the order of 60 to 120 (McGregor and Fisher, 1945; Hoggatt and Recht, 1969). On the other hand, research workers dealing with pure metals have observed effective viscosities corresponding to a linear dependence of strain rate on stress. For instance Lothe (1962) gives a dislocation phonon drag of

$$\tau b = Bv, \text{ where } B = 3kT/10C_2b^2. \tag{5}$$

Eliminating the velocity v in terms of the strain rate $\dot{\gamma}$, by the mobile dislocation density Λ_m and the Burgers vector b

$$\dot{\gamma} = b\Lambda_m v \tag{6}$$

gives the viscosity as

$$\mu = \tau/\dot{\gamma} = .3kT/b^4\Lambda_m c_2 \quad . \tag{7}$$

The evaluation of these equations is illustrated in Fig. 1 for copper. Assuming a mobile dislocation density of $10^8/cm^2$, the mean dislocation velocity can be related to the strain rate by Eq. 7 and the stress calculated from the viscosity. For the annealed material, there is a small range of rate-independent plasticity, where obstacles limit the flow stress. Soon phonon drag becomes limiting. Finally inertial effects limit the dislocation velocity, and, if the mobile dislocation density remains constant by annihilation, the strain rate is also limited, giving a hole growth rate proportional to radius.
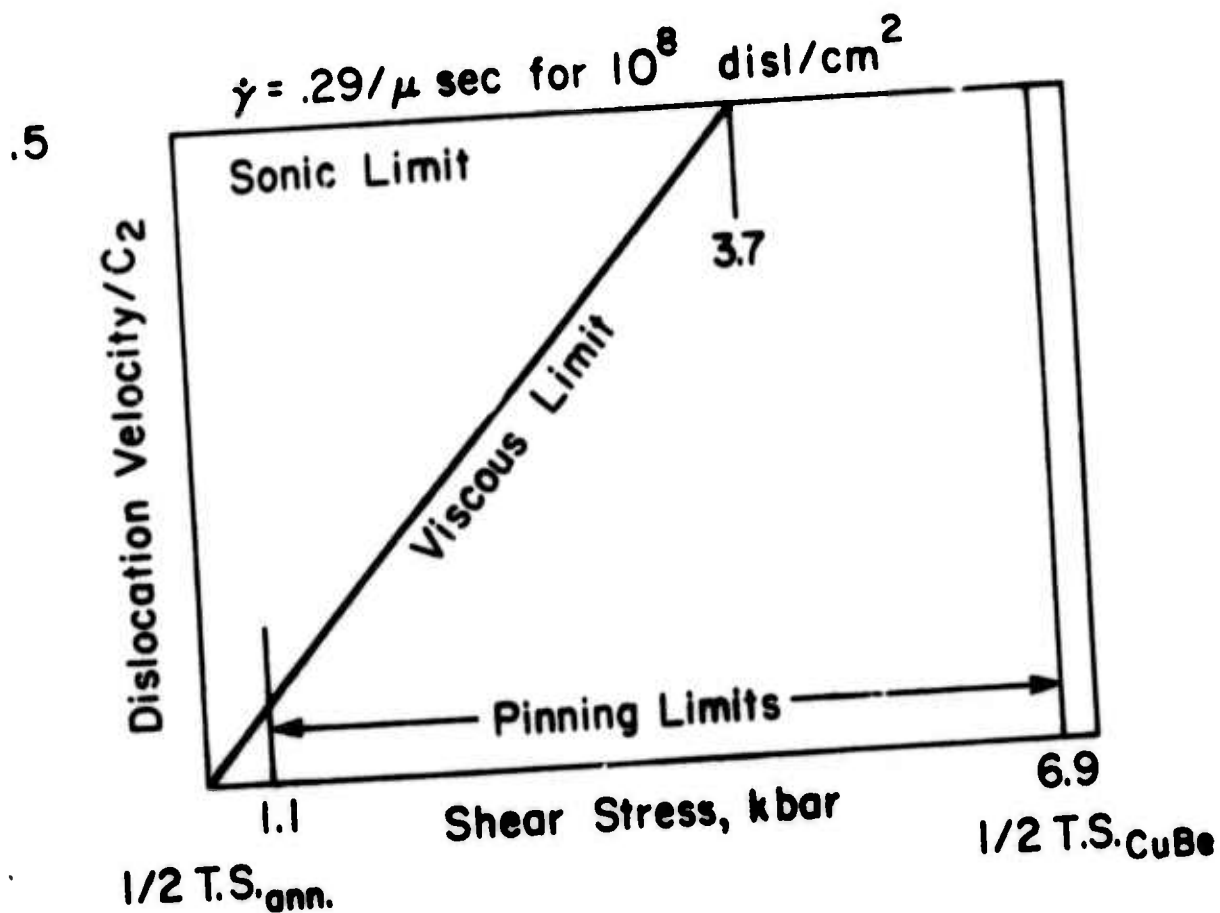
Figure 1.   Regimes of rate effects in annealed copper and hard beryllium copper.

For hard Beryllium copper alloy, however, there is no viscous regime. This illustrates the need to carry out studies in the regime of practical interest. Similar differences between situations are observed for aluminum, iron and their alloys.

These calculations may not leave the rate-independent theory of plasticity a clear field for hard alloys, however, as indicated in the section of Table 1 dealing with thermal behavior. Increasing the radius of a hole by a factor of ten by stretching its surface in each direction causes a local straining of $2\ln(10)$, corresponding to a strain of 4.6. As shown in Table 1, the temperature rises for the alloys are so great that melting would have to occur if the process were adiabatic. Actually these high strains are only encountered in a very thin skin, and if one estimates the quench time on the basis of the hole radius, even in that thin skin the heating only becomes important for microsecond type pulses after the holes reach a radius of the order of 5μ.

III. Size Effects

Another factor limiting classical plasticity is the size of the hole relative to the in free path of slip lines. The distance ℓ between pinning points on dislocation loops sets a lower bound to this dimension and can be estimated from the well-known equation for hardening:

$$\tau = Gb/\ell \ .\tag{8}$$

The correlation of flow in slip bands, however, means that the slip could first be regarded as being homogeneous at a scale several orders of magnitude larger than this, say 100ℓ.

$$100\ell = 30\mu \text{ and } 3\mu, \text{ respectively,} \tag{9}$$

indicating that the use of homogeneous plasticity becomes doubtful for holes smaller than a micron even in hard alloys. With other solutions so difficult, however, one is tempted to push his luck.

IV. Inertia Model of Spall Fracture by Hole Growth

Neglecting the limitations discussed above, we can estimate the inertia effects themselves by using the equation of Hopkins (1960) for the expansion of the hole of radius $\underline{a}$ in a fully plastic regime subjected to an applied stress σ_R at the radius $\underline{R}$ midway between holes. Hopkins' Eq. 7.65 is

$$\sigma_R/\rho = (2Y/\rho)(\ell n\ R/a) + \ddot{a}(a-a^2/R) + \dot{a}^2(3/2+a^4/2R^4-2a/R) . \tag{10}$$

A measure of the relative importance of the acceleration and the velocity terms can be obtained by first setting the radial velocity to zero and solving for the acceleration,

$$\ddot{a} = (\sigma_R - 2Y\ell nR/a)/\rho(a-a^2/R) , \tag{11}$$

and in turn setting the acceleration to zero and solving for a quasi-steady state velocity:

$$\dot{a} = \sqrt{\frac{\sigma_R - 2Y \ell n\,(R/a)}{\rho\,(3/2 + a^4/2R^4 - 2a/R)}}\tag{12}$$

The distance required to accelerate to steady states is of the order of

$$\dot{a}^2/\ddot{a} = \frac{a/R - (a/R)^2}{3/2 + a^4/2R^4 - 2a/R} \simeq 2a/3R \quad . \tag{13}$$

It appears that steady state develops relatively soon, so that the radial velocity is found reasonably closely from Eq. 12. This provides an equation of state for the porosity, which could then be used with a uniaxial code, as was done by Seaman, et al. (1971) with viscous hole growth models. In addition to the effects of hole growth on the porosity and average properties of the material itself, the unloading waves due to hole growth should also be taken into account. The plastic calculations outlined here should be good for high strength materials only slightly above their limit loads.

V. Conclusion

Order of magnitude estimates indicate that the importance of the variables such as dislocation viscosity, adiabatic temperature rise, thermal diffusivity, and mechanical inertia play markedly different roles with different alloys at differest intensities of loading. The weakest assumption in these estimates is that of the mobile location density in hard alloys, about which further comment would be welcome.

VI. Acknowledgement

References

Berg, C. A.   1972   "Ductile Fracture by Development of
                                    Surfaces of Unstable Cavity Expansion,"
                                    Journal of Research of the National
                                    Bureau of Standards, v. 76C, 33-39.

Carson, J. W.   1970   "A Study of Plane Strain Ductile
                                    Fracture," Ph.D. Thesis, Dept. of
                                    Mech. Eng., M.I.T., Cambridge, Mass.

Cipolla, R. C.   1972   "A Study of the Initiation of Ductile
                                    Fracture from Grooves," M.S. Thesis,
                                    Dept. of Mech. Eng., M.I.T. Cambridge,
                                    Mass.

Hoggatt, C. R.   1969   "Dynamic Stress Strain Relationships
  Recht, R. F.   Obtained from an Expanding Ring
                                    Experiment," SESA Spring Meeting,
                                    Paper 1487.

Hopkins, H. G.   1960   "Dynamic Expansion of Spherical
                                    Cavities in Metals," Progress in
                                    Solid Mechanics, v. 1, 86-164.

Lomer, W. M.   1949   "A Dynamical Model of A Crystal
                                    Structure, III," Proc. Roy. Soc. A,
                                    v. 196, 132.

Lomer, W. M.   1949   "The Forces Between Floating Bubbles
                                    and a Quantitative Study of the
                                    Bragg 'Bubble Model' of a Crystal,"
                                    Proc. Camb. Phil. Soc., v. 45, 660.

Lothe, J.   1962   "Theory of Dislocation Mobility in
                                    Pure Slip," J. Appl. Phys., v. 33,
                                    2116-2125.

MacGregor, C. W.   1945   "Tension Tests at Constant Strain
  Fisher, J. C.   Rates," J. Appl. Mech. v. 12, 217-227.

McQueen, R. G.   1962   "Ultimate Yield Strength of Copper,"
  Marsh, S. P.   J. Appl. Phys. v. 33, 654-665.

Nagpal, V.   1973   "Traction-Displacement Boundary
  McClintock, F. A.   Condition for Plastic Fracture by
  Berg, C. A.   Hole Growth," International Symposium
  Subudhi, M.   on the Foundations of Plasticity,
                                    Warsaw, Poland, to be published in
                                    Proceedings.

Seaman, L.
  Barbee, T. W. Jr.
  Curran, D. R.

1971

"Dynamic Fracture Criteria of Homogeneous Materials," Stanford Research Institute, TR. AFWL-TR-71-156. Menlo Park, California

Teutonico, L. M.
  Granato, A. V.
  Lücke, K.

1964

"Theory of the Thermal Breakaway of a Pinned Dislocation Line with Application to Damping Phenomena," *J. Appl. Phys.*, v. 35, 220-234.

See Also:

Granato, A. V.
  Lücke, K
  Schlipt, J.
  Teutonico, L. J.

1964

"Entropy Factors for Thermally Activated Unpinning of Dislocations," *J. Appl. Phys.* v. 35, 2732-2745.

Author's Discussion on
"Models of Spall Fracture by Hole Growth"

F. A. McClintock

Further comment on two points in the paper may be in
order: the question of mobile dislocation density and detailed
estimates of limiting stresses and growth times for voids.

The question of mobile dislocation density may be seen
more clearly by re-plotting Fig. 1 as tensile strength vs. strain
rate, using a logarithmic scale, as shown in Fig. 2. The engineer
is interested in a phenomenal range of strain rates, from nano-
seconds to hundreds of years; only the right half is shown. Con-
sider an annealed metal. At low strain rates, obstacle and dis-
location pinning limit the flow; the rate dependence is very
small, with a slope of only 1-2% on a log-log scale. If the
mobile dislocation density were assumed to remain nearly constant,
the dislocation velocity would rise, eventually reaching a
limiting viscous behavior determined by phonon drag. The stress
now increases linearly with strain rate. Further increases in
stress would accelerate the dislocations until they approached
say half the shear wave velocity, indicated by the vertical line
labelled "sonic limit". Eventually, increases in stress would
cause the ideal strength to be reached, at which shear can occur
with no dislocation motion. Here the plot of flow stress vs.

-467-

strain rate would again become horizontal, until the effects of atomic inertia were reached at frequencies of the order of $10^{13}/$ sec. For high strength alloys, Fig. 2 indicates that a viscous limit would never be reached whatever the mobile dislocation density, since that shifts both the viscous and sonic limits by the same amount. The assumption of a constant mobile dislocation density seems very questionable and the author would appreciate comment from others about data or more reliable estimates for the rate dependence in the high strain rate regime for strong alloys. One would expect that relatively moderate increases in strength, of the order of a factor of two occuring in several decades of strain rate, would allow the mobile dislocation density to approach the total density. This density could be estimated from either dislocation interactions at a distance ℓ or the critical stress for nucleating sources. In terms of shear strength τ, shear modulus G, and Burgers vector b it would be of the order of

$$\Lambda = \frac{1}{\ell^2} = \left(\frac{\tau}{Gb}\right)^2 \qquad . \tag{14}$$

If the rate dependence does indeed flatten out so that the flow stress only rises by a factor of 2 in 4 orders of magnitude of strain rate, then the calculations of dynamic hole growth could reasonably be based on a rate-insensitive calculation, with the mean flow stress simply elevated by an appropriate amount. For low strength materials, if there is an abrupt ramp the viscous equations might still be appropriate.
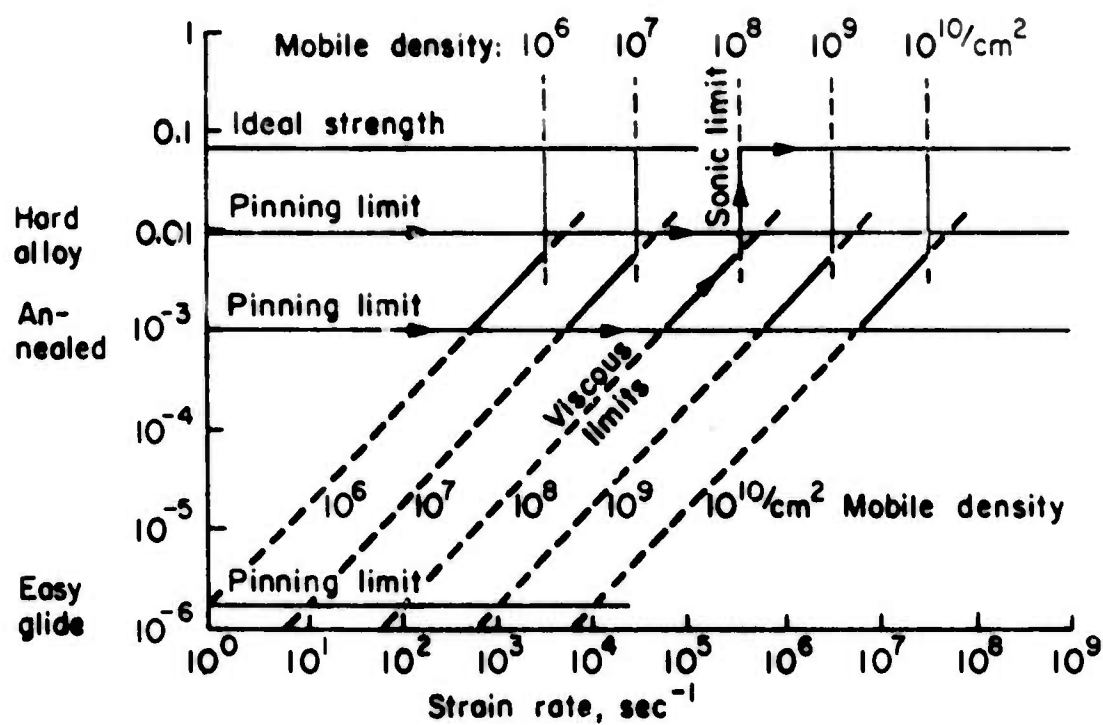
Figure 2.   Hypothetical strain rate dependence for constant
dislocation density.

As a second point of amplification of the paper, upper limits to the time required by inertia effects in spall growth can be obtained by approximate integration of Hopkins' equations for incompressible material.

First for a hole initially at rest on the spall plane, the time required for an elastic wave appearing there to spread out to the radius R is

$$t_{load} = R/c_1 \qquad . \tag{15}$$

For typical hole spacings of $R = 5\mu$ and $c_1 = .6$ cm/μsec, the loading time is only .8 nanoseconds.

Next, for elastic growth of the hole to initial yield, from Hopkins' Eq. 7.55 with $\sigma_r = 0$ at $r = a$ and $\sigma_r = \sigma_R$ at $r = R$,

$$\sigma_R/\rho = (4E/9\rho)(1-a_0^3/a^3)(1-a^3/R^3) + \ddot{a}a(1-a/R)$$
$$+ \dot{a}^4(3/2-2a/R + (1/2)a^4/R^4) \tag{16}$$

Hole growth in the elastic regime is only a fraction of the radius, so estimate the time from the acceleration term alone. First, the hole radius at initial yield, $\sigma_\phi = \sigma_\theta = Y$ for incompressible material is

$$a-a_0 = a_0(Y/E)(1-1/2) \qquad . \tag{17}$$

For a/R small compared to unity, substituting Eq. 16 into the acceleration equation gives

$$t = \sqrt{2(a-a_o)/\ddot{a}} = \left[a_o(Y/E) \frac{a_o}{\sigma_R/\rho - (4E/9\rho)(a^3/a_o^3 - 1)} \right]^{\frac{1}{2}} \quad (18)$$

The second term in the denominator increases from zero to $2Y/3\rho$ as the hole grows. Taking the higher value gives an upper limit to the time of initial yielding:

$$t = a_o / \sqrt{E/\rho} \, (\sigma_R/Y - 2/3) \quad . \quad (19)$$

Comparison with the loading time of .8 nanoseconds from Eq. 15, and taking σ_R/Y twice that for static yield, gives time of the order of a fraction a/R of a nanosecond, quite negligible.

The time for elastic-plastic growth is found similarly. From Hopkins' Eq. 7.65, 7.68, and 7.71 and for $R \gg a$,

$$\mathfrak{I}_R = (2Y/3\rho)(-b^3/R^3 + 1 + \ell n(2E/3Y) + \ell n(1 - a_o^3/a^3))$$

$$+ \ddot{a}a + 3\dot{a}^2/2 \quad (20)$$

when the plastic zone radius $\underline{b}$ is between $\underline{R}$ and $\underline{a}$:

$$1 \geq b^3/a^3 = (2E/3Y)(1 - a_o^3/a^3) \leq R^3/a^3 \quad .$$

When the plastic zone has reached R, the inner radius is

$$a^3/a_o^3 = 1 + (R^3/a_o^3)(3Y/2E) \quad . \quad (21)$$

Except for very soft, dirty metals $((R/a_o)^3 < E/Y)$, the holes will have grown by a number of radii by the time the plastic zone reaches R, so estimate the growth time from the steady state velocity:

$$t = \int \frac{da}{\dot{a}} = \int \frac{\sqrt{3/2} \quad da}{\sqrt{\sigma_R/\rho + (2Y/3\rho)(b^3/R^3 - 1 - \ln(2E/3Y) - \ln(1 - a_0^3/a^3))}} \quad (22)$$

An upper limit to the time is found from a lower limit to the denominator, neglecting b^3/R^3 and a_0/a. For σ_R 10% above that for zero growth rate,

$$t < \frac{a}{\frac{2}{3}\sqrt{.1\frac{Y}{E}\frac{E}{\rho}(1+\ln\frac{2E}{3Y})}} \quad (23)$$

For $Y/E = .001$ to $.01$, and $c = \sqrt{E/\rho} = .5$cm/μsec, $R = 5\mu$, and from Eq. 21

$$a = R^3\sqrt{3Y/2E} \quad , \quad (24)$$

the times are less than 6 to 5 nanoseconds. Lower values of applied stress would require longer times, but are not of much practical interest because only a 10% decrease in applied stress would stop the hole growth completely.

Turning to the fully plastic regime, obtain an estimate from Eq. 12 of the original paper, with Eq. 24 for an initial value of a/R in the numerator, and a/R terms neglected in the denominator of that equation:

$$t < R/a_{min} = \frac{R}{\sqrt{\frac{2}{3}\frac{\sigma_R}{\rho} - \frac{4Y}{9\rho}\ln\frac{2E}{3Y}}} \quad (25)$$

Again for σ_R 10% above the limit for hole growth,

$$t \leq (3/2)R/\sqrt{.1(E/\rho)(Y/E)\ln(2E/3Y)} \qquad (26)$$

For $\sqrt{E/\rho}$ = .5 cm/μsec, Y/E = .001 to .01, and R = 5μ, the times are less than 60 to 23 nanoseconds.

In conclusion, then, inertia effects are negligible (except within 10% of the limit of zero hole growth rate) when the loading times exceed 20 to 50 nanoseconds for high and low strength alloys, respectively. The limiting values of the applied stresses are typically of the order of a 4 times the yield strength, as given by the first term on the right side of Eqs. 10 or 17.

MATERIALS NEEDS FOR MORE EFFICIENT DEGRADATION
OF HIGH-TEMPERATURE ENERGY

J. L. Margrave

Abstract

In the current energy crisis the advantages of higher
operating temperature are multiple--higher efficiencies, gross
conservation of fuel, less pollution, etc.-- but the achieve-
ment of these goals is inevitably hampered by the lack of satis-
factory high temperature materials. Not only are extended
efforts of synthetic chemists and metallurgists required, but
measurement of physical and chemical properties at extreme con-
ditions are basic to the development of feasible engineering
designs. As an example of the discrepancy between practical
developments and the availability of fundamental chemical know-
ledge, consider the Si_3N_4 program. As of July, 1973, neither
the low-temperature heat capacity nor the absolute entropy of
this material are known. Within the last year, new gas-phase
species have been observed over the heated solid. Coordinated
programs in synthesis and property measurement are desirable.
Tabulations of basic thermodynamic data are needed.

Li/CFX BATTERIES

J. L. Margrave

Abstract

Thermodynamic data have been obtained for various compounds in the $CF_{0.7-1.2}$ range, including low temperature heat capacities, heats of formation, and thermal stabilities. Theoretical EMF's calculated for $CF_{1.1}$/Li and for C_4F/Li batteries suggest cell voltages close to 5 volts. The basic patents in this area are held by the U.S. Signal Corps (Breuer on C_4F/Li) and by the Japanese (Watanabe on $CF_{1.0}$/Li). Current problem areas are electrolyte stability, electrode construction and a general lack of fundamental knowledge about the electrochemical process at the CF-electrode-electrolyte interface.

Preceding page blank

LIGHT ELEMENT HYDRIDES AS ENERGY
TRANSPORT MATERIALS

J. L. Margrave
M. F. Hawthorne

Abstract

A summary of potential compounds and systems involving
lithium and boron mixed hydrides, lithium and boron mixed
alkyls, lithium and aluminum hydrides and alkyls will be pre-
pared. Thermodynamic data and preparative chemistry are al-
ready available for many of these materials and their large
scale use as energy transport materials should be practical
from both economic and engineering viewpoints. In simple terms,
hydrogen at one of the pipes is converted to a convenient boron
compound, e.g., B_2H_6; pumped through existing pipelines; and
then the hydrogen and some other boron compound, probably H_3BO_3,
are separated by hydrolysis. The boron would be re-cycled.

CHEMISTRY WITHOUT CONTAINERS

J. L. Margrave

Abstract

A review of methods and devices used for levitation
of conductors--metals, alloys, carbides, nitrides, certain
oxides, etc.--and the application of this approach for the
measurement of thermodynamic properties of liquid metals, for
metallurgical processing, for chemical synthesis, etc.

Preceding page blank

ON THE SCALE OF MORPHOLOGICAL PHENOMENA
IN POLYMERIC ALLOYS

M. B. Bever and M. Shen

Abstract

The morphological phenomena in polymeric alloys result from the spatial arrangement on a supramolecular scale of assemblies of macromolecules. Their complexity increases rapidly in going from amorphous homogeneous polymers to the various types of heterogeneous polymers.

In this memorandum polymeric alloys are classified by their chain constitution. The major experimental methods for obtaining quantitative morphological data are listed. The size ranges of various morphological features are summarized in a graph and the nature of these features is explained briefly.

ON THE SCALE OF MORPHOLOGICAL PHENOMENA
IN POLYMERIC ALLOYS

M. B. Bever and M. Shen

The morphology of polymeric alloys results from the spatial arrangement on a supramolecular scale of assemblies of macromolecules. The complexity of these arrangements increases rapidly in going from amorphous homogeneous polymers to the various types of heterogeneous polymers.

The morphological aspects of polymeric alloys have been discussed in detail elsewhere[1]. In this memorandum, we shall be concerned with those quantitative features for which information is available.

General

Polymers, including polymeric alloys classified by their chain constitution, are shown in Table 1. The size ranges of various morphological features are plotted in Fig. 1; the documentation of the data is given in a forthcoming publication[2], which is a revised and expanded version of Ref. 1.

Quantitative morphological data can be obtained by the use of the following experimental methods:

1. Small angle X-ray scattering

2. Electron and X-ray diffraction

3. Electron microscopy

    a. Transmission

    b. Replica

    c. Scanning

4. Light scattering

5. Optical microscopy

    a. Conventional

    b. Polarized light

    c. Phase contrast

The evaluation of the results obtained by these techniques usually involves methods of quantitative stereology[3,4]. To date only the simpler of these methods have been applied to polymers and the potential of quantitative stereology remains largely unexploited.

Specific Morphological Features

The following comments refer to the morphological features shown in Fig. 1.

The crystalline phases in polymeric alloys have characteristic <u>unit cell sizes</u>. Their size range depends on the complexity of the repeating unit.

<u>Single crystals</u> of polymers differ from typical inorganic single crystals. Polymer single crystals occur as thin plates which have characteristic ranges of thickness and lateral dimensions. Polymeric alloys do not usually occur as single crystals, but single crystals are included here because of their

relation to the crystallites that form spherulites.

Crystalline phases in polymeric alloys can be present as spherulites, which grow from the melt as spheres but ultimately attain polyhedral shapes because of mutual impingement. Each spherulite consists of a large number of crystallites, each of which in some respects resembles the single crystals already mentioned.

Some recent evidence suggests that supramolecular structures called nodules exist in amorphous homogeneous polymeric alloys. This evidence consists of density variations which have been attributed to localized chain folding.

In block and graft copolymers, two amorphous phases may form structures consisting of a matrix and dispersed particles. These particles may be spheres or rods. The characteristic parameters of these particles are their diameters and the mean free distances between them. In addition, rods have characteristic lengths. The size ranges of these parameters are shown in Fig. 1. Under some circumstances, particles dispersed in a matrix are arranged in a regular periodic structure, which is referred to as a macrolattice.

Another two-phase configuration is the lamellar structure, which is characterized by interlamellar spacings.

The two phases present in polyblends are characteristically coarser than the dispersed particles in block and graft copolymers. The particles in polyblends also have less definite shapes.

Conclusion

The interpretation of structure-property relations of polymers has been almost exclusively concerned with their chemical structures. By contrast, most of the morphological features referred to in this memorandum have been investigated only very recently. It may be expected that these features will be studied more extensively in the future and that their effects on the properties of polymeric alloys will be increasingly recognized.

Acknowledgement

References

1. M. B. Bever and M. Shen, On the Morphology of Polymeric Alloys, in Preliminary Reports, Memoranda and Technical Notes of the ARPA Materials Summer Conference, July 1972, pp. 547-573.

2. M. B. Bever and M. Shen, The Morphology of Polymeric Alloys, Mat. Sci. Eng. (accepted for publication).

3. E. E. Underwood, Quantitative Stereology, 1970, Addison-Wesley.

4. E. E. Underwood, Applications of Quantitative Metallography, in Metals Handbook, Vol. 8, 1973, American Society for Metals, pp. 37-47.
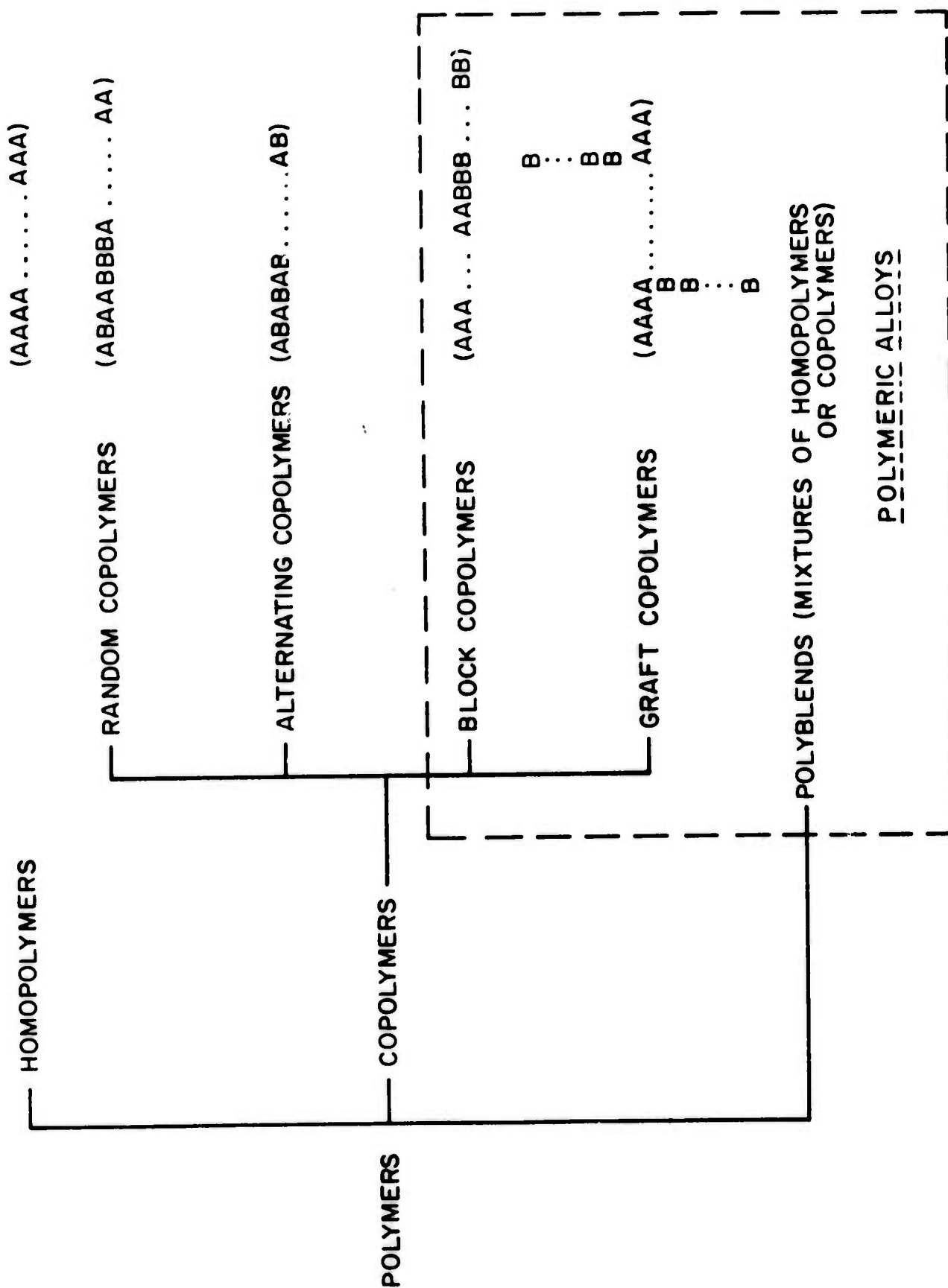
Table I. Classification of Polymers on the Basis of Chain Constitution
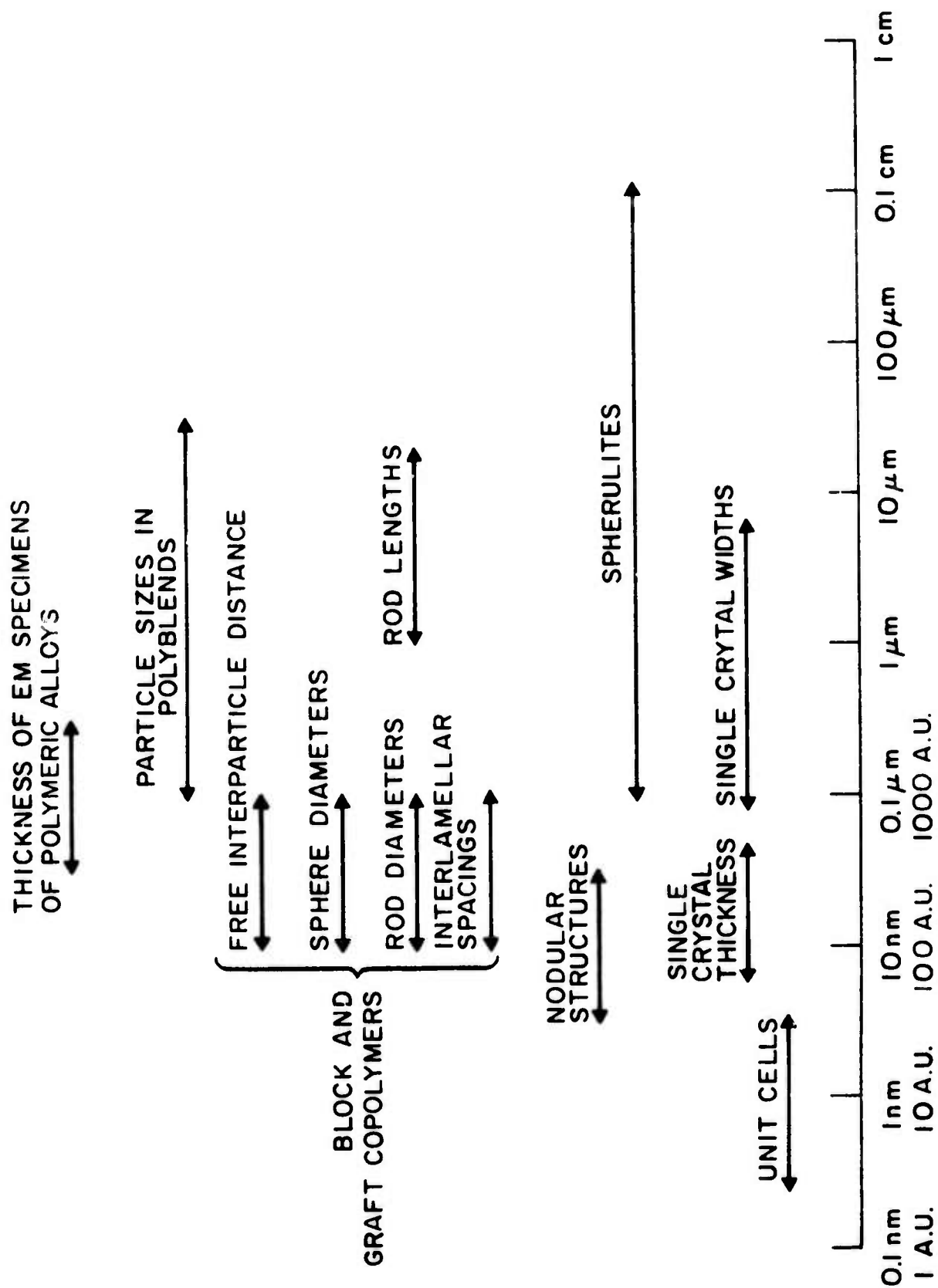
Fig. 1 Size Ranges of Morphological Features of Polymeric Alloys

MATERIALS PROBLEMS IN THE TRANSPORT
AND STORAGE OF HYDROGEN

M. B. Bever

Abstract

This memorandum is concerned with the following
questions: If hydrogen is to be transported and stored in
large volumes in a routine manner, what materials problems
will arise? What information is available regarding these
problems? What additional information is required and how
can it be obtained?

Preceding page blank

MATERIALS PROBLEMS IN THE TRANSPORT
AND STORAGE OF HYDROGEN

M. B. Bever

1.   Introduction

Hydrogen has been suggested as a medium for transmitting and storing energy.[1] The use of hydrogen in this manner raises various materials problems. This memorandum deals briefly with the nature of these problems, the available information concerning them and the need for additional information.

Hydrogen can be transported and stored in the following forms:

1. as the element

    1.1 gaseous

    1.2 liquid

2. as a compound or in solution

    2.1 gaseous

    2.2 liquid

    2.3 solid

Each of these modes of hydrogen transport and storage poses specific materials problems. Modes 1.1, 1.2 and 2.1 are considered in this memorandum; modes involving compounds are also dealt with in concurrent memoranda[2-5].

2.   Transport and Storage of Hydrogen Gas

2.1   General

        Systems of pipelines and storage tanks made of steel
have been proposed for the transport and storage of hydrogen gas.
This proposal is based on the following current applications
which are assumed to be satisfactory:

        (1)   small, high-pressure hydrogen tanks;

        (2)   equipment for process use of hydrogen in the chemical
                and petroleum refining industries;

        (3)   distribution systems for manufactured (or "town") gas,
                which contains up to 50% H_2;

        (4)   natural gas pipelines;

        (5)   two pipelines (one in Texas and one in Germany) re-
                ported to carry hydrogen over appreciable distances.

        How relevant for large-scale transport and storage are
these applications? In particular, what justification is there
for statements which assert apparently without critical assess-
ment that pipelines now used for natural gas can be switched to
hydrogen? The following quotation is an example of this approach[6]:

        "...method of transportation...for hydrogen would probably
be the same as for natural gas now. The two gases are similar
enough in most properties so that it is entirely feasible to use
the same technology (or even the existing systems) of transmission
pipelines, intermediate storage, and urban distribution systems
used for natural gas..."

The following considerations show that such extrapolations cannot be accepted uncritically:

Concerning items (1) and (2) above - small hydrogen tanks and similar equipment have relatively thick walls.

Item (3) - Distribution systems for manufactured gas are operated at low pressure; they have relatively thick walls; the gas probably contains one or several inhibitors (discussed below).

Item (4) - Natural gas does not contain hydrogen; H_2S which may be present is known to cause embrittlement of pipelines.

Item (5) - Detailed information on pipelines used for hydrogen is not readily available.

2.2 Hydrogen Embrittlement

Johnson has called attention to the possibility of hydrogen embrittlement (HE) in hydrogen pipelines and storage tanks.[7] Although the conditions for and the mechanism of HE are not fully understood, its major features are:[8,9]

(1) HE is caused by atomic hydrogen which may originate in diatomic hydrogen gas, acids or other compounds.

(2) Low-strength steels are less susceptible to HE than high-strength steels and apparently steels are immune below a certain strength level.

(3) O_2, CO, CH_4 act as inhibitors of embrittlement.

(4) Some other gases, especially H_2S, promote HE.

(5) Metallurgical variables affect the susceptibility of steels to HE to a relatively minor degree. Microstructure,

alloy content and impurities have effects which are generally similar to their effects on temper brittleness.

(6) Welding can increase the susceptibility to HE by causing hard local regions.[10,11] Segregation of elements such as manganese has the same effect. Steelmaking and processing practices may have adverse effects by altering the distribution and arrangement of inclusions.

References 8-11 review the current state of knowledge of HE.

2.3   Prevention

The following approaches to the prevention of HE of pipelines and storage tanks should be considered:

(1) The use of low-strength steels.

This is imperative and will reduce the risk of HE but may not provide absolute protection. In particular, it does not eliminate problems arising from segregation or bad welding practice. The use of low-strength steels also runs counter to a trend toward higher stress levels in pipeline design.

(2) Elimination of contact between hydrogen and the steel surface.

Hydrogen can be separated from the steel surface by mechanical barriers. Continuous layers on all surfaces exposed to hydrogen or other coatings and liners such as teflon may be considered, but they do not seem to provide a dependable solution. Aside from the problem of adding a coating or liner to an existing pipeline, it is doubtful that the required total protection

-495-

of all steel surfaces could be achieved in practice.

A more dependable means of preventing contact of hydrogen with steel surfaces and thus eliminating the risk of HE exists on the molecular (or atomic) level through the use of inhibitors, specifically, oxygen. Relevant experimental data have been published[8], but information based on operating experience is also needed. It is not known whether a temporary absence of an inhibitor causes irreversible damage. There is no detailed information for CH_4 as an inhibitor of HE.

It may appear reasonable to consider the use of other metals for hydrogen pipelines. Although such use may be technically feasible, however, it would require a great deal of experimental and developmental work; moreover, the cost of other metals seems to rule them out.

2.4 Other Potential Problems

In addition to HE, the following phenomena may cause problems in the transport and storage of hydrogen gas:

(1) Loss by diffusion: this is proportional to $\sqrt{P_i}$.

(2) Loss by leakage: the leakage coefficient of H_2 is 3 to 6 times as large as the leakage coefficient of CH_4. The entire system intended for hydrogen, therefore, must be tighter than a system for natural gas.

(3) Action of hydrogen on special metals present in valves and other specialized equipment.

(4) Action of hydrogen on nonmetallic materials, especially gaskets.

-496-

3.   Transport and Storage of Liquid Hydrogen

Transport of liquid hydrogen by pipeline does not seem
likely in the foreseeable future. Liquid hydrogen will continue
to be shipped and stored in insulated containers. These containers,
however, may be of larger dimensions, especially for water trans-
port.

Hydrogen may be stored as a liquid at terminals. Ul-
timately, it may be used as fuel for vehicles and possibly for
the cooling of superconductors (if superconducting materials with
higher transition temperatures than at present become available).

The experience of AEC and NASA with the handling of liquid
hydrogen provides a base for further developments. The scaling up
and the greater dispersion of liquid hydrogen operations, however,
are likely to present new problems, including materials problems.

Liquid hydrogen appears not to cause HE of steel. This
subject, however, needs further investigation. Also, information
regarding materials used and other relevant practices involving
liquid hydrogen should be collected. In general, it may be assumed
that the hydrogen potential at the temperatures involved is too
low for HE, but the possibility of problems arising from accidental
or local heating, especially in connection with transfer oper-
ations, seems to exist.

The materials problems in the use of liquid hydrogen are
those encountered in cryogenic engineering generally and are
the direct result of the low temperatures. Thermal stresses are
an obvious example. Carbon steels, even low-strength steels, are

inevitably brittle at the temperature of liquid hydrogen and alloying may not be able to change this. The use of face-centered cubic metals in spite of their cost and other drawbacks may be necessary. This subject calls for more information.

As a speculation, it has been suggested that a twin system could use liquid hydrogen as coolant for a superconducting cable (again assuming a suitable superconductor) at the same time as the liquid hydrogen would be transported along the cable.

4.   Transport and Storage of Gaseous Hydrogen Compounds

The transport and storage of gaseous hydrogen compounds is mentioned here for completeness. (Compounds are also discussed in Refs. 2-5.) Methane appears to be the most promising gaseous compound and could indeed be handled by natural gas technology and in natural gas equipment. No new major materials problems are likely to arise from the substitution of pure methane for natural gas.

5.   Conclusions and Recommendations

This memorandum is intended as a preliminary statement regarding materials problems in the transport and storage of hydrogen and especially regarding certain potential hazards that may be involved in these operations. To some degree it represents a reaction to uncritical statements by proponents of a "hydrogen economy" who have assumed that a direct and unqualified extrapolation can be made from natural gas to hydrogen. In contrast to this approach it is argued here that certain risks primarily involving HE by hydrogen gas and to a lesser extent risks involving the handling of liquid hydrogen must still be assessed in detail.

-498-

This will require the collection and analysis of additional
information, some of which may already be available, especially
in the pipeline industry. Also, further experimental data may
still be needed. Only careful evaluation rather than simple ex-
trapolation can provide the basis for decisions involving the
transport and storage of hydrogen.

Acknowledgement

References

1. W. E. Winsche, K. C. Hoffman and F. J. Salzano, "Hydrogen: Its Future Role in the Nation's Economy", Science, Vol. 180 (1973) 1325.

2. M. F. Hawthorne and Howard Reiss, "Hydrogen Transfer by Diffusion Through a Hydrogenation-Dehydrogenation System", Memoranda of Materials Research Conference, 1973.

3. J. L. Margrave, "Light Element Hydrides as Energy Transport Materials", ibid.

4. R. Gomer, M. F. Hawthorne and J. L. Margrave, "Some Comments in Favor of Hydrides Rather than Hydrogen as an Energy Transport Medium," ibid.

5. H. Ehrenreich, "Hydrogen Storage in Transition Metals," ibid.

6. D. P. Gregory, D. Y. C. Ng and G. M. Long, "The Hydrogen Economy," Chapter 8 in Electrochemistry of Cleaner Environments, J. O'M. Bockris, ed., Plenum Press, 1973, p. 251.

7. H. H. Johnson, "Hydrogen Embrittlement," Letter to Editor, Science, Vol. 179 (1973) p. 228.

8. H. H. Johnson, "Hydrogen Brittleness in Hydrogen and Hydrogen-Oxygen Gas Mixtures," Proceedings of the International Conference on Stress-Corrosion Cracking and Hydrogen Embrittlement of Iron-Base Alloys, Unieux, Firminy, France, June 1973.

9. C. S. Carter and M. V. Hyatt, "Review of Stress-Corrosion Cracking in the Low-Alloy and Low-Strength Steels," ibid.

10. R. R. Fessler, T. P. Groeneveld and A. R. Elsea, "Stress-Corrosion and Hydrogen-Stress Cracking in Buried Pipelines," ibid.

11. R. S. Treseder, "Oil Industry Experience with Hydrogen Embrittlement and Stress Corrosion Cracking," ibid.

COMMENT ON A
PRESENTATION BY DR. ROBERT MEHRABIAN
ON CASTING OF SEMI-SOLID METALS

D. C. Drucker and E. H. Lee

As described, the hot billet in its "solid" form is
primarily a viscous metal which surrounds clumps of solid metal.
The thixotropy or drop to a very low flow stress following
shearing was explained by the speaker to be a consequence of
the breaking of contact welds between the solid particles. Al-
though the stresses due to the billet weight alone are low,
there is some difficulty with this explanation, even when the
viscosity is highly nonlinear, if the fluid has the same low
viscosity in the billet that it has when filling the mold after
entering a small opening.

An alternative suggestion is that there may be a more
complex filamentary or sheet-like reaching out or joining of
the solid particles which would drastically constrict any flow
at low stress and so cause the billet to behave as a solid under
its own weight. This structure would be broken down by strong
shearing action and the metal then would behave as a fluid of
low viscosity.

Another possible explanation of the thixotropic softening
due to deformation is that the effect is associated with dif-

-501-

ferent phases of the material in this region of temperature. A
simple example of this type of phenomenon for elastic bodies has
been studied (E. Varley and Alan Day, "Equilibrium Phases of
Elastic Materials at Uniform Temperatures and Pressure" Archive
for Rational Mechanics and Analysis 22, 1966, 253-269) which
illustrates the essentially free deformation which can occur in
such circumstances. The Gibbs free energy is taken to be a
function of deformation variables, pressure and temperature which
differ for different phases. Entropy balance and deformation
compatibility between regions containing different phases, lead
to possible equilibrium geometry changes at effectively zero
shear stress. It seems possible that the structure of this theory
which combines general deformation variables with pressure and
temperature in the analysis of phase transformation may be ex-
tended to deal with the essentially viscous fluid or plastic
characteristics of a metal near its melting point.

Acknowledgement

PROPAGATION OF LOW FREQUENCY ELASTIC DISTURBANCES
IN A THREE-DIMENSIONAL COMPOSITE MATERIAL

W. Kohn

Abstract

This paper is a generalization to three dimensions of a earlier study for one-dimensional composites. We show here that in the limit of low frequencies the displacement vector $u_i(r,t)$ can be written in the form $u_i(r,t) = \delta_{ij} + v_{ij1}(r)\partial/\partial x_1 + \ldots)$ $U_j(r,t)$. Here $U_j(r,t)$ is a slowly varying vector-function of r and t which describes the mean displacement of each cell of the composite. Its components satisfy a set of three coupled partial differential equations with constant coefficients. These coefficients are obtainable from the three-by-three secular equation which yields the low-lying normal mode frequencies, $\omega(k)$. Information about local strains is contained in the function $v_{ij1}(r)$, which is characteristic of static deformations, and is discussed in detail. Among applications of this method is the structure of the head of a pulse propagating in an arbitrary direction.

PROPAGATION OF LOW FREQUENCY ELASTIC DISTURBANCES
IN A THREE-DIMENSIONAL COMPOSITE MATERIAL

W. Kohn

1. Introduction

In a recent paper[1] we have presented a theory of low

frequency* elastic disturbances in a one-dimensional composite

to which the reader is referred as background to the present

paper. The characteristic feature of this theory is that the

problem is clearly divided into its macro- and microscale aspects.

It was shown that the displacement $u(x,t)$ could be written in a

product form

$$u(x,t) = (1 + v_1(x) \frac{\partial}{\partial x} + \cdots) U(x,t) \quad . \tag{1.1}$$

Here $U(x,t)$ describes the average macroscale displacement field

and satisfies a macroscale wave equation of the form

$$\frac{\partial^2}{\partial t^2} - \bar{c}^2 \frac{\partial^2}{\partial x^2} - \beta \frac{\partial^4}{\partial x^4} + \cdots) U(x,t) = 0 \quad , \tag{1.2}$$

where the constants $\bar{c}$, β, etc., are derived from the small wave

number expansion of the lowest band of eignemodes

$$\omega^2(k) = \bar{c}^2 k^2 - \beta k^4 + \cdots) \quad . \tag{1.3}$$

---

*Low frequency means that many unit cells are traversed in one
 period; or that a characteristic wavelength is much greater
 than a cell dimension.

The pre-factor $(1 + v_1(x) \partial/\partial x + \cdots)$ in (1.1) carries the information about strains and stresses on the scale of the microstructure of the composite. $v_1(x)$ is the function which describes the microstrains due to a homogeneous static macro-strain.

By means of this theory we discussed general long wave-length disturbances in one-dimensional composites and also the so-called "head of the wave" for an arbitrary initial pulse.

In the present paper we generalize most of these results to arbitrary three-dimensional composites.

The wave equation for a general periodic elastic medium has the form[2]

$$\rho(\underset{\sim}{r}) \frac{\partial^2 u_i(\underset{\sim}{r},t)}{\partial t^2} = \frac{\partial}{\partial x_j} [c_{ijlm}(\underset{\sim}{r}) \frac{1}{2}(\frac{\partial u_1(r,t)}{\partial x_m} + \frac{\partial u_m(r,t)}{\partial x_1})]. \qquad (1.4)$$

Here $u_i(\underset{\sim}{r},t)$ is the displacement, $\rho(\underset{\sim}{r})$ the density, and the $c_{ijlm}(\underset{\sim}{r})$ the local elastic moduli. The functions $\rho(\underset{\sim}{r})$ and $c_{ihlm}(\underset{\sim}{r})$ have the periodicity property.

$$f(\underset{\sim}{r} + \underset{\sim}{\tau}_\mu) = f(\underset{\sim}{r}) \quad , \qquad (1.5)$$

the $\underset{\sim}{\tau}_\mu$ being the fundamental translation vectors of the medium.

These equations have solutions of the Floquet form

$$u_i(\underset{\sim}{r},t) = u_i(\underset{\sim}{r},\underset{\sim}{k}) e^{-i\omega t} \quad , \qquad (1.6)$$

where

$$u_i(\underset{\sim}{r}, \underset{\sim}{k}) = v_1(\underset{\sim}{r}, \underset{\sim}{k}) e^{i\underset{\sim}{k}\cdot\underset{\sim}{r}} \tag{1.7}$$

and

$$v_i(\underset{\sim}{r} + \underset{\sim}{\tau}_\mu) = v_i(\underset{\sim}{r}) \quad . \tag{1.8}$$

We show in Section 2 that the three acoustic frequency branches $\omega(\underset{\sim}{k})$ are obtained from a solution of the following three coupled equations

$$M_{ij}(\underset{\sim}{k}) U_j = \omega^2 U_i \quad , \tag{1.9}$$

where, for small $\underset{\sim}{k}$,

$$M_{ij}(k) = M^{(2)}_{ij,pq} k_p k_q + M^{(4)}_{ij,pqrs} k_p k_q k_r k_s + \cdots) \quad . \tag{1.10}$$

The coefficients $M^{(2)}_{ij;pq}$ are related to the standard macroscopic elastic moduli $\bar{c}_{ijkl}$ as follows[3]

$$M^{(2)}_{ij;pq} = \bar{c}_{ipjq} \quad . \tag{1.11}$$

The secular equation

$$\text{Det } |M_{ij}(\underset{\sim}{k}) - \omega^2 \delta_{ij}| = 0 \tag{1.12}$$

leads, for any given direction of $\underset{\sim}{k}$, to three solutions of the form

$$\omega^2(k;j) = c^2(\theta,\phi;j)k^2 - \beta_4(\theta,\phi;j)k^4 + \cdots \quad (j=1,2,3) \tag{1.13}$$

Here $c(\theta,\phi;j)$ are the three sound velocities in the direction θ,ϕ. We recall however that, at $\underset{\sim}{k} = 0, \omega^2$ does not have a power series expansion in the components k_p.

We also show that a long wavelength disturbance has
the form

$$u_i(\underset{\sim}{r},t) = (\delta_{ij} + v_{ij1}(\underset{\sim}{r}) \frac{\partial}{\partial x_1} + \cdots) U_j(\underset{\sim}{r},t) \quad . \tag{1.14}$$

Here $U_i(\underset{\sim}{r},t)$ is a slowly varying function of $\underset{\sim}{r}$ and t which
satisfies the wave equation

$$\frac{\partial U_i(\underset{\sim}{r},t)}{\partial t^2} + M_{ij}(\frac{1}{i} \underset{\sim}{\nabla}) \ U_j(\underset{\sim}{r},t) = 0 \quad , \tag{1.15}$$

or

$$\frac{\partial^2 U_i(\underset{\sim}{r},t)}{\partial t^2} + (-M_{ij;pq}^{(2)} \frac{\partial}{\partial x_p} \frac{\partial}{\partial x_q} + M_{ij;pqrs}^{(4)} \frac{\partial}{\partial x_p} \frac{\partial}{\partial x_q} \frac{\partial}{\partial x_r} \frac{\partial}{\partial x_s}$$

$$+ \cdots) \ U_j(\underset{\sim}{r},t) = 0 \quad , \tag{1.16}$$

where the M's are the matrix elements of the secular equation
(1.12). The function $v_{ij1}(\underset{\sim}{r})$ describe the elastic deformation
on a microscale due to a macroscopic, static homogeneous
$\partial U_j(\underset{\sim}{r})/\partial x_1$.

The problem of calculating the propagation of an arbi-
trary low frequency disturbance is thus split into two quite
distinct parts:

I. The Macroscopic Displacement Field

Obtain first the functions $M_{ij}(\underset{\sim}{k})$ for small $\underset{\sim}{k}$. If the
disturbance in question is limited to very low frequency
(low k-) components, the first term in (1.10), quadratic in k,

is sufficient, whose coefficients are determined by the elastic moduli according to (1.11). However, if the smoothing out of a sharp wave front is at issue (head of the pulse solutions[4,5,1]), it is necessary to include the next term in (2.7), of fourth order in k. The coefficients $M_{ij,pqrs}^{(4)}$ can be obtained by fitting the solutions of the secular equation (1.12) to measured or calculated[6,7] dispersion curves of the composite up to order k^4.

After the function $M_{ij}(k)$ have been obtained to sufficient accuracy, equation (1.15) or (1.16) must be solved for the macroscopic displacement field $U_i(r,t)$, subject to the given boundary and initial conditions. The execution of such solutions by means of Green's functions is further discussed in Section 3. In particular we solve there explicitly the problem of the propagation of a plane disturbance through a general three-dimensional composite.

II. The Microscopic Strain and Stress Fields

For many practical applications one needs to know the local stresses and strains set up in the medium as a result of the passage of a dynamic disturbance. The calculation of these quantities requires a knowledge of the static, elastic properties of the medium. Specifically, we require the local elastic moduli $c_{ijlm}(r)$ (see Eq. 1.4) and the local displacements $v_{ijl}(r)$ associated with a microscopic strain $\partial U_j(r)/\partial x_1$ as follows

$$u_i(r) = U_i(r) + v_{ijl}(r) \frac{\partial}{\partial x_1} U_j(r) \quad . \tag{1.17}$$

The quantities $v_{ijl}(r)$ can either be determined experimentally in static strain experiments or calculated by methods developed specifically for treating elastic deformations of inhomogeneous media[8,9]. Finally they can also be calculated by the method of long waves developed by Born[10]. This is carried out in the Appendix.

Once $U_j(r)$ has been determined from the solution of Eq. (1.15) or (1.16), and $c_{ijlm}(r)$ and $v_{ijl}(r)$ are known, the local strains and stresses are given by

$$\frac{\partial u_i(r,t)}{\partial x_m} = [\delta_{ij}\delta_{lm} + \frac{\partial v_{ijl}(r)}{\partial x_m}] \frac{\partial U_j(r,t)}{\partial x_l} \quad , \tag{1.18}$$

$$\sigma_{ij}(r,t) = c_{ijkl}(r) \frac{1}{2}(\frac{\partial u_k(r,t)}{\partial x_l} + \frac{\partial u_l(r,t)}{\partial x_k}) \quad . \tag{1.19}$$

Finally we wish to add some remarks about the class of composites with discontinuous densities, ρ, and elastic moduli, c_{ijlm}, which includes many materials of practical importance. For such systems the required input functions $M_{ij}(k)$ and $v_{ijl}(r)$ may be obtained either by taking $\rho(r)$ and $c_{ijlm}(r)$ as limits of continuous functions, or by special methods appropriate to discontinuous media[6,7]. However the description of low frequency time dependent phenomena in terms of these quantities, which is the principal subject of the present paper, is the same for continuous and discontinuous media.

2.   Periodic Waves

We consider an elastic medium whose density $\rho(\underset{\sim}{r})$ and local elastic moduli $c_{ijlm}(\underset{\sim}{r})$ satisfy the periodicity conditions (1.5). Such a medium has a complete set of elastic waves of the form

$$u_i(\underset{\sim}{r},t) = v_i(\underset{\sim}{r},\underset{\sim}{k})\, e^{i(\underset{\sim}{k}\,\underset{\sim}{r}-\omega t)} \quad , \tag{2.1}$$

where $v_i(\underset{\sim}{r},\underset{\sim}{k})$ are completely periodic functions in the sense of Eq. (1.5). The displacements $v_i(\underset{\sim}{r},\underset{\sim}{k})$ satisfy the following coupled equations

$$-\omega^2(\underset{\sim}{r})v_i(\underset{\sim}{r},\underset{\sim}{k}) = \frac{\partial}{\partial x_j}\,[c_{ijlm}(\underset{\sim}{r})\,\frac{v_1(\underset{\sim}{r};\underset{\sim}{k})}{\partial x_m}]$$

$$+ k_p i[\frac{\partial}{\partial x_j}\,c_{ijlm}(\underset{\sim}{r})\,\delta_{pm}v_1(\underset{\sim}{r},\underset{\sim}{k})$$

$$+ c_{ijlm}(\underset{\sim}{r})(\delta_{pj}\,\frac{v_1(\underset{\sim}{r},\underset{\sim}{k})}{\partial x_m} + \delta_{pm}\,\frac{\partial v_1(\underset{\sim}{r},\underset{\sim}{k})}{\partial x_j})]$$

$$+ k_p k_q(-1)[c_{ijlm}(\underset{\sim}{r})\delta_{pj}\delta_{qm}v_1(\underset{\sim}{r},\underset{\sim}{k})].$$

Here we have used the symmetry property of the elastic moduli

$$c_{ijlm}(\underset{\sim}{r}) = c_{ijml}(\underset{\sim}{r}) \quad . \tag{2.2}$$

For compactness we now write (2.1) in the form

$$\omega^2\rho(\underset{\sim}{r})\,v_i(\underset{\sim}{r},\underset{\sim}{k}) = m_{ij}(\underset{\sim}{k})\,v_j(\underset{\sim}{r},\underset{\sim}{k}) \quad , \tag{2.3}$$

where $m_{ij}(\underset{\sim}{k})$ is a differential operator of the second degree

in k_p. We thus write

$$m_{ij}(\underset{\sim}{k}) = m_{ij}^{(0)} + m_{ij}^{(1)} + m_{ij}^{(2)} \quad , \tag{2.4}$$

with

$$m_{ij}^{(1)} = k_p m_{ij,p}^{(1)}; \quad m_{ij}^{(2)} = k_p k_q m_{ij,pq}^{(2)} \quad . \tag{2.5}$$

The operators $m_{ij}^{(\mu)}$ ($\mu = 0,1,2$) can be directly identified from Eq. (2.1).

We write the solutions for $\underset{\sim}{k} = 0$ as

$$v_i(\underset{\sim}{r},0) = u_i(\underset{\sim}{r},0) = u_i(\underset{\sim}{r}) \quad . \tag{2.6}$$

These vectors satisfy the equations

$$\omega^2 \rho(\underset{\sim}{r}) u_i(\underset{\sim}{r}) = m_{ij}^{(0)} u_j(\underset{\sim}{r}) \tag{2.7}$$

and constitute a complete set of fully periodic vectors. We label the corresponding set of eigenvalues by ω_μ^2 and the associated eigenvectors by $u_{\mu\nu;i}$. Here the additional label ν allows for degeneracies, such as the important three-fold degeneracy associated with $\omega^2 = 0$. Thus, we re-write (2.7) as*

$$\omega_\mu^2 \rho(\underset{\sim}{r}) u_{\mu\nu;i}(\underset{\sim}{r}) = m_{ij}^{(0)} u_{\mu\nu;j}(\underset{\sim}{r}) \quad . \tag{2.8}$$

The orthonormality property is

$$(u_{\mu\nu;j}, u_{\mu'\nu'j}) = \delta_{\mu\mu'} \delta_{\nu\nu'} \quad , \tag{2.9}$$

where the scalar product of two periodic functions is defined

---

*There is obviously no summation on μ.

as

$$(f,g) \equiv \frac{1}{\bar{\rho}\Omega} \int f^*(\underset{\sim}{r}) \, \rho(\underset{\sim}{r}) \, g(\underset{\sim}{r}) \, d\underset{\sim}{r} \quad , \qquad (2.10)$$

Ω being the volume of the unit cell and $\bar{\rho}$ the mean density.

We now follow a method developed by Kohn and Luttinger[11] for electronic wavefunctions in periodic media, to discuss the solutions of Eq. (2.3) for general $\underset{\sim}{k}$ in terms of the solutions $u_{\mu\nu;i}(\underset{\sim}{r})$ for $\underset{\sim}{k} = 0$. Since the operator m_{ij} is a polynomial in the components of the wave-vector k_p one might expect that $v_i(\underset{\sim}{r},\underset{\sim}{k})$ in powers of k_p. This is however not the case in the vicinity of degenerate points, such as the most important point $\omega^2 = 0$, where ω^2 itself is not a power series in k_p. Hence, a more appropriate method is required.

Following Kohn and Luttinger we expand $v_i(\underset{\sim}{r},\underset{\sim}{k})$ in terms of the complete set $u_{\mu\nu;i}(\underset{\sim}{r})$,

$$v_i(\underset{\sim}{r},\underset{\sim}{k}) = a_{\mu\nu}(\underset{\sim}{k}) \, u_{\mu\nu;i}(\underset{\sim}{r}) \quad , \qquad (2.11)$$

where

$$a_{\mu\nu}(\underset{\sim}{k}) = (u_{\mu\nu;i}(\underset{\sim}{r}), \, v_i(\underset{\sim}{r},\underset{\sim}{k})) \quad . \qquad (2.12)$$

Note particularly that for $= 0$ and $\omega^2 = 0$ the three basic vectors may be taken as uniform displacements along the x,y, and z directions, i.e.,

$$u_{o\nu;i}(\underset{\sim}{r}) = \delta_{i\nu} \quad . \qquad (2.13)$$

Substitution of (2.12) into (2.4) gives the set of linear equations

$$\omega^2 a_{\mu\nu} = (\mu\nu|m|\mu'\nu') \, a_{\mu'\nu'} \quad , \tag{2.14}$$

where

$$(\mu\nu|m|\mu'\nu') = \frac{1}{\rho\Omega} \int_{\text{cell}} u_{\mu\nu;i}(\overset{*}{\underset{\sim}{r}}) \, m_{ij} \, u_{\mu\nu;j}(\underset{\sim}{r}) \, d\underset{\sim}{r} \quad . \tag{2.15}$$

For $\underset{\sim}{k} = 0$,

$$(\mu\nu|m|\mu'\nu') = \mu\nu|m^{(0)}|\mu'\nu') = \omega_\mu^2 \, \delta_{\mu\mu'} \delta_{\nu\nu'} \tag{2.16}$$

and the diagonalization of Eq. (2.14) is trivial. However, for $\underset{\sim}{k} = 0$, $(\mu\nu|m|\mu'\nu')$ has also off-diagonal elements, small if $\underset{\sim}{k}$ is small. We shall now, to successive powers of k_p, eliminate all matrix elements connecting $\mu = 0$ with $\mu \neq 0$, by means of a unitary transformation

$$a_{\mu\nu} = (\mu\nu|W|\mu'\nu') \, b_{\mu\nu} \quad , \tag{2.17}$$

where

$$\sum_{\mu\nu} (\mu\nu|W|\mu'\nu')^* \, (\mu\nu|W|\mu''\nu'') = \delta_{\mu'\mu''}\delta_{\nu'\nu''} \quad . \tag{2.18}$$

Such a matrix may be put in the form

$$W = e^{iS} \quad , \tag{2.19}$$

where S is a hermitian matrix,

$$(\mu\nu|S|\mu'\nu')^* = (\mu'\nu'|S|\mu\nu) \quad . \tag{2.20}$$

For $\underset{\sim}{k} = 0$ we have evidently $W = 1$ and $S = 0$. For finite $\underset{\sim}{k}$ we make the power series Ansatz

$$S(k) = k_p S_p^{(1)} + k_p k_q S_{pq}^{(2)} + \cdots = S^{(1)} + S^{(2)} + \cdots, \quad (2.21)$$

where we have dropped the explicit matrix notation. Writing now

$$a = e^{iS} b \quad , \quad (2.22)$$

and substituting in (2.14) gives

$$\omega^2 f = Mb \quad , \quad (2.23)$$

where the matrix M is given by

$$M = e^{-iS} m e^{iS} \quad . \quad (2.24)$$

We wish to expand M in powers of k. We begin by making the commutator expansion

$$M = m + i[S,m] + 1/2[[S,m],S] + \cdots \quad (2.25)$$

Next, we expand both m and S in powers of k and group the resulting terms appropriately. This gives

$$M = m^{(0)} + \{m^{(1)} + i[S^{(1)}, m^{(0)}]\} + \{m^{(2)} + i[S^{(1)}, m^{(1)}]$$
$$+ i[S^{(2)}, m^{(0)}] + 1/2[[S^{(1)}, m^{(0)}], S^{(1)}]]\} \quad . \quad (2.26)$$

We now solve successively for $S^{(1)}$, $S^{(2)}$, etc., to eliminate interband matrix elements of M of order k, k^2, etc., connecting $\mu = 0$ with $\mu \neq 0$. This generates a new matrix M in which the $(0\mu|M|0\nu')$ is decoupled to arbitrary order in k from the rest of the matrix, so that the problem of finding the lowest three

eigenvalues $\omega_0^2(\underset{\sim}{k})$ is reduced to the solution of a 3×3 detrimental equation. It is evident that to each order in k, $S^{(\alpha)}$ may be taken as off-diagonal. Restoring explicit matrix notation and observing that, by construction, $m^{(0)}$ is diagonal in μ and ν we have, from (2.27)

$$(\mu\nu|M|\mu'\nu') = \omega_\mu^2 \, \delta_{\mu\mu'} + \{(\mu\nu|m^{(1)}|\mu'\nu')$$

$$+ i(\mu\nu|S^{(1)}|\mu'\nu')(\omega_{\mu'}^2 - \omega_\mu^2)\} \quad . \qquad (2.27)$$

To eliminate first order off-diagonal matrix elements connecting $\mu = 0$ with $\mu \neq 0$ we choose

$$(0\nu|S^{(1)}|\mu'\nu') = (\mu'\nu'|S^{(1)}|0\nu)^* = \frac{i}{\omega_{\mu'}^2} \, (\mu\nu|m^{(1)}|\mu'\nu') \qquad (2.28)$$

and all other matrix elements zero. We proceed similarly for the higher order terms of S. Recalling that $(0\nu|m^{(0)}|0\nu') = 0$, we thus obtain the following power series for $(0\nu|M(\underset{\sim}{k})|0\nu')$

$$M(\underset{\sim}{k})_{\nu\nu'} \equiv (0\nu|M(\underset{\sim}{k})|0\nu') + M^{(2)}_{\nu\nu',pq} \, k_p k_q$$

$$+ M^{(4)}_{,pqrs} \, k_p k_q k_r k_s + \cdots \qquad (2.29)$$

where, using $S^{(1)}$ only, we can calculate from (2.27),

$$M^{(2)}_{\nu\nu',pq} = m_{\nu\nu'pq} + \sum_{\mu''\nu''} \frac{1}{\omega_{\mu''}^2} \, (0\nu|m_p^{(1)}|\mu''\nu'')(\mu''\nu''|m_q^{(1)}|0\nu') \qquad (2.30)$$

Calculation of the higher order terms can be carried out in a similar manner but becomes increasingly more complex.

The eigenvectors and eigenvalues associated with $\mu = 0$ are now obtained from the three coupled equations

$$M_{\nu\nu'}(\underset{\sim}{k})\ U_\nu = \omega^2 U_{\nu'}\ , \qquad (2.31)$$

where we have written

$$U_\nu \equiv b_{0\nu} \qquad (2.32)$$

or from

$$[M^{(2)}_{\nu\nu';pq}\ k_p k_q + M^{(4)}_{\nu\nu';pqrs}\ k_p k_q k_r k_s + \cdots]\ U_\nu(\underset{\sim}{k}) = \omega^2 U'_\nu(\underset{\sim}{k}). \quad (2.33)$$

The corresponding third degree secular equation is

$$\left| M_{\nu\nu'}(\underset{\sim}{k}) - \omega^2 \delta_{\nu\nu'} \right| = 0. \quad \nu,\nu' = 1,2,3 \qquad (2.34)$$

Apart from different labelling, Eqs. (2.32), (2.33) and (2.34) will be seen to be identical with Eqs. (1.9), (1.10) and (1.12).

The eigenvectors $v_i(r,k)$ of Eq. (2.4) are given by

$$v_i(r,\underset{\sim}{k}) = u_{\mu\nu;i}(\underset{\sim}{r})\,(\mu\nu|e^{iS}|\mu'\nu')\,b_{\mu'\nu'}$$

$$= u_{\mu\nu;i}(\underset{\sim}{r})\,(\mu\nu|1 + iS + \cdots|0\nu')\,U_\nu,$$

$$= \delta_{i\nu'} + u_{\mu\nu;i}(\underset{\sim}{r})ik_p(\mu\nu|S^{(1)}_p|0_{\nu'}) + \cdots)U_\nu,$$

$$= (\delta_{i\nu'} + v_{i\nu'p}(\underset{\sim}{r})ik_p + \cdots)\,U_\nu,\ \qquad (2.35)$$

where the quantity $S^{(1)}_p k_p (\equiv S^{(1)})$ is given in Eq. (2.28) and

$$v_{i\nu'p}(\underset{\sim}{r}) \equiv \sum_{\mu\nu} u_{\mu\nu;i}(\underset{\sim}{r})\,(\mu\nu|S^{(1)}_p|0\nu)\ . \qquad (2.36)$$

Eq. (2.35) expresses the function $v_{i\nu'1}(\underset{\sim}{r})$ as an infinite series in the complete set of periodic displacements $u_{\mu\nu'i}(\underset{\sim}{r})$ of the normal modes associated with $\underset{\sim}{k} = 0$. In the Appendix, using Born's method of long waves[10], we re-derive Eq. (2.31) up to second order in k. At the same time we obtain an alternative definition of $v_{i\nu'1}(\underset{\sim}{r})$ as the solution of differential equation which is probably of greater practical value than the infinite series (2.36).

The full displacement field corresponding to $v_i(\underset{\sim}{r},\underset{\sim}{k})$ is

$$u_i(\underset{\sim}{r},t) = v_i(\underset{\sim}{r},\underset{\sim}{k})\ e^{i(\underset{\sim}{k}\cdot\underset{\sim}{r}-\omega t)}$$

$$= (\delta_{ij} + v_{ij1}(\underset{\sim}{r})\frac{\partial}{\partial x_1} + \cdots)\ U_j\ e^{i(\underset{\sim}{k}\cdot\underset{\sim}{r}-\omega t)}\ , \quad (2.37)$$

where we have relabelled dummy indices to accord with Section 1.

Now define

$$U_j(\underset{\sim}{r},t)\ U_j\ e^{i(\underset{\sim}{k}\cdot\underset{\sim}{r}-\omega t)} \quad\quad\quad (2.38)$$

Then clearly, by (2.31), $v_j(\underset{\sim}{r},t)$ satisfies the equation

$$M_{ij}(\underset{\sim}{k})\ U_j(\underset{\sim}{r},t) = \omega^2 U_i(\underset{\sim}{r},t) \quad\quad (2.39)$$

or

$$M_{ij}(1/i\nabla)\ U_j(\underset{\sim}{r},t) = -\frac{\partial^2}{\partial t^2}\ U_j(\underset{\sim}{r},t)\quad . \quad\quad (2.40)$$

Superposition of Periodic Waves

In Eq. (2.40) the wave-vector k and frequency ω no longer explicitly appear.

Therefore _any_ displacement field which is a linear superposition of modes belonging to the three lowest branches has the form

$$u_i(\underset{\sim}{r},t) = (\delta_{ij} + v_{ij1}(\underset{\sim}{r}) \frac{\partial}{\partial x_1} + \cdots) U_j(\underset{\sim}{r},t) \quad , \qquad (2.41)$$

where $U_i(\underset{\sim}{r},t)$ satisfies Eq. (2.4).

3. Solution of Initial Value Problems

In the preceding section we have reduced the problem of calculating low frequency disturbances in a composite medium to the determination of the static function $v_{ij1}(\underset{\sim}{r})$ and the solution of the differential equation (1.15) for the macroscopic displacement vector $U_i(\underset{\sim}{r},t)$. In the present section we address the problem of solving this equation in an infinite medium, subject to given initial conditions:

$$U_i(\underset{\sim}{r},0) = F_i^{(1)}(\underset{\sim}{r}) \quad , \qquad (3.1a)$$

$$\frac{\partial U_i(\underset{\sim}{r},t)}{\partial t}\bigg|_{t=0} = F_i^{(2)}(\underset{\sim}{r}) \quad , \qquad (3.1b)$$

where $F_i^{(1)}$ and $F_i^{(2)}$ are both slowly varying functions of $\underset{\sim}{r}$.

As usual this solution is effected by the appropriate two Green's functions $G_{ij}^{(\alpha)}(\underset{\sim}{r}-\underset{\sim}{r}',t)(\alpha = 1,2)$. These satisfy the differential equations

$$[\frac{\partial^2}{\partial t^2}\delta_{ij} - M_{ij}(\frac{1}{i}\nabla)] G_{jk}^{(\alpha)}(\underset{\sim}{r}-\underset{\sim}{r}') = 0 \quad \alpha = 1,2; \; t>0 \quad , \qquad (3.2)$$

-518-

and the following initial conditions:

$$G_{ij}^{(1)} (\underset{\sim}{r}-\underset{\sim}{r}',0) = \delta_{ij}\delta(\underset{\sim}{r}-\underset{\sim}{r}') \qquad , \qquad (3.3a)$$

$$\frac{\partial}{\partial t}[G_{ij}^{(1)} (\underset{\sim}{r}-\underset{\sim}{r}',t)]_{t=0} = 0 \qquad , \qquad (3.3b)$$

and

$$G_{ij}^{(2)} (\underset{\sim}{r}-\underset{\sim}{r}',0) = 0 \qquad , \qquad (3.4a)$$

$$\frac{\partial}{\partial t}[G_{ij}^{(1)} (\underset{\sim}{r}-\underset{\sim}{r}')]_{t=0} = \delta_{ij}\delta(\underset{\sim}{r}-\underset{\sim}{r}') \qquad . \qquad (3.4b)$$

The solution of the initial value problem defined by the differential equation (1.15) and the initial condition (3.1) is then

$$U_i(\underset{\sim}{r},t) = \int d\underset{\sim}{r}' [G_{ij}^{(1)} (\underset{\sim}{r}-\underset{\sim}{r},t)F_j^{(1)} (\underset{\sim}{r}') + G_{ij}^{(2)} (\underset{\sim}{r}-\underset{\sim}{r},t)F_j^{(2)} (\underset{\sim}{r}')] \quad , \quad (3.5)$$

as may be immediately verified.

The Green's functions can be formally constructed from the plane wave solutions of the differential equation (1.15). We write these solutions as

$$U_{m,i}^{\pm}(\underset{\sim}{r},t;k) = U_{m,i}(\underset{\sim}{k}) \, e^{i(\underset{\sim}{k}\cdot\underset{\sim}{r}\pm\omega_m(\underset{\sim}{k})t)}, \ m=1,2,3, \qquad (3.6)$$

where m denotes the three branches of the vibration spectrum and $\omega_m(\underset{\sim}{k})$ is the positive square root of one of the solutions of the secular equation (1.12). The vectors $U_{m,i}(\underset{\sim}{k})$ satisfy the orthogonality and completeness relations.

$$U^*_{m,i}(\underset{\sim}{k})U_{m',i}(\underset{\sim}{k}) = \delta_{mm'} \quad , \tag{3.7}$$

$$U_{m,i}(\underset{\sim}{k})U^*_{m,j}(\underset{\sim}{k}) = \delta_{ij} \quad . \tag{3.8}$$

$G^{(1)}_{ij}$ and $G^{(2)}_{ij}$ can now be written as

$$G^{(1)}_{ij}(\underset{\sim}{r},t) = \frac{1}{(2\pi)^3} \int d\underset{\sim}{k} \sum_m U_{m,i}(\underset{\sim}{k})U^*_{m,j}(\underset{\sim}{k})e^{i\underset{\sim}{k}\cdot\underset{\sim}{r}} \cos[\omega_m(\underset{\sim}{k})t] \tag{3.9}$$

and

$$G^{(2)}_{ij}(\underset{\sim}{r},t) = \frac{1}{(2\pi)^3} \int d\underset{\sim}{k} \sum_m U_{m,i}(\underset{\sim}{k})U^*_{m,j}(\underset{\sim}{k})e^{i\underset{\sim}{k}\cdot\underset{\sim}{r}} \frac{\sin[\omega_m(\underset{\sim}{k})t]}{\omega_m(\underset{\sim}{k})} \tag{3.10}$$

as can be directly verified.

One-Dimensional Disturbance

Because of the complicated dependence of ω_m on $\underset{\sim}{k}$ we are not able to evaluate $(G^{\alpha}_{ij}(\underset{\sim}{r})$ explicitly for a general disturbance. However, if $F^{(\alpha)}_i(\underset{\sim}{r})$ depends on only one coordinate, say x, effective one-dimensional Green's functions $G^{(\alpha)}_{ij}(x)$ can be explicitly calculated. In this case we have

$$U_i(x,t) = \int dx' [G^{(1)}_{ij}(x-x',t)F^{(1)}_j(x') + G^{(2)}_{ij}(x-x',t)F^{(2)}_j(x')], \tag{3.11}$$

where

$$G^{(1)}_{ij}(x) = \int G^{(1)}_{ij}(\underset{\sim}{r})dy\,dz$$

$$= \frac{1}{2\pi} \int dk_x \sum_m U_{m,i}(k_x)U^*_{m,j}(k_x)e^{ik_x x} \cos\,_m(k_x)t \tag{3.12}$$

and a corresponding expression for $G^{(2)}_{ij}$. Here

$$U_{m,i}(k_x) \equiv U_{mi}(k_x,0,0) \qquad , \tag{3.13a}$$

$$\omega_m(k_x) \equiv \omega_m(k_x,0,0) \qquad . \tag{3.13b}$$

Now for $\underset{\sim}{k}$ in the x-direction, the displacement vector $U_{m,i}(k_x)$ is a slowly varying function of k_x and hence can be taken out of the integral giving

$$G_{ij}^{(1)}(x) = \sum_m U_{m,i} U_{m,j}^* \, g_m^{(1)}(x) \qquad , \tag{3.14}$$

where

$$U_{m,i} = \lim_{k_x \to 0} U_{m,i}(k_x) \tag{3.15}$$

and

$$g_m^{(1)}(x) = \frac{1}{2\pi} \int e^{ik_2 x} \cos[\omega_m(k_x)t] \qquad . \tag{3.16}$$

The integral over dk_x in (3.16) is identical to that encountered in our treatment of one-dimensional composites[1]. Let us write

$$\omega_m(k_x) = c_m k_x + \frac{\omega_m^{\prime\prime\prime}}{3!} k_x^3 + \cdots \tag{3.17}$$

Then if the cubic term of $\omega_m(k_x)$ in (3.17) can be neglected, which is the case when $F^{(1)}(x)$ is extremely slowly varying, $g_m^{(1)}(x-t)$ takes the simple form, familiar from the theory of homogeneous elastic media

$$g_m^{(1)}(x,t) = \frac{1}{2} [\delta(x-c_m t) + \delta(x+c_m t)] \qquad . \tag{3.18}$$

If the expression of $\omega_m(k_x)$ is carried to third order, $g_m^{(1)}(x,t)$ is given in terms of Airy functions[1], (Eq. 3.14)

$$g_m^{(1)}(x,t) = \frac{1}{2} \left(\frac{1}{2}\omega_m''' t\right)^{-1/3}$$

$$\times \left\{ A_i \left[\left(\frac{1}{2}\omega_m''' t\right)^{-1/3} (x-c_m t) \right] \right.$$

$$\left. + A_i \left[\left(\frac{1}{2}\omega_m''' t\right)^{-1/3} (-x-c_m t) \right] \right\} \quad . \tag{3.19}$$

This form allows for the dispersion of plane waves and gives a satisfactory description of the head of a pulse (see ref. 1, Sec. 3).

Similarly we find

$$G_{ij}^{(2)}(x,t) = \sum_m U_{m,i} U_{m,j}^* \; g_m^{(2)}(x,t) \quad , \tag{3.20}$$

where, when the cubic term of $\omega_m(k_x)$ is neglected,

$$g_m^{(2)}(x,t) = \begin{cases} c_m^{-1} & |x| < c_m t \\ 0 & |x| > c_m t \end{cases} \quad . \tag{3.21}$$

When the cubic term is included one finds

$$g_m^{(2)}(x,t) = \int_0^t \frac{1}{2}\left(\frac{1}{2}\omega_m''' t'\right)^{-1/3}$$

$$\times \left\{ A_i \left[\left(\frac{1}{2}\omega_m''' t'\right)^{-1/3} (x-c_m t') \right] \right.$$

$$\left. + A_i \left[\left(\frac{1}{2}\omega_m''' t'\right)^{1/3} (-x-c_m t') \right] \right\} \quad . \tag{3.22}$$

as

$$U_i(x,t) = \sum_m U_{m,i} \int dx' [g_m^{(1)}(x-x',t) \; f_m^{(1)}(x')$$

$$+ \; g_m^{(2)}(x-x',t) \; f_m^{(2)}(x')] \quad , \tag{3.32}$$

where

$$f_m^{(\alpha)}(x) \equiv \sum_m U_{m,j}^{\star} F_j^{(\alpha)}(x) \qquad . \qquad (3.24)$$

Eq. (3.22) describes the superposition of three disturbances, each with the polarization vector associated with one of the elastic waves propagating in the x-direction.

Simple boundary value problems on a half-space can be immediately solved in the conventional manner. Thus, for $U_i(0,t) = 0$ and given $F^{(\alpha)}(x)$ for $x > 0$, we solve the problem in infinite space with $F^{(\alpha)}(-x) = -F^{(\alpha)}(x)$. The general boundary value problem associated with the differential equation (1.15) is beyond the scope of the present paper.

We wish to close with a conjecture concerning <u>disordered</u> systems. It would appear from the nature of our solution for ordered systems, Eqs. (1.14) and (1.16), that for sufficiently long wave lengths, Eq. (1.14) should remain valid up to the order shown and Eq. (1.16) up to the second order term. In the present case $v_{ij1}(\underset{\sim}{r})$ in Eq. (1.14) again represents local displacements due to a homogeneous mean-strain, but is of course no longer a periodic function of $\underset{\sim}{r}$. Eq. (1.16), now takes the familiar form for an isotropic solid

$$\frac{\partial^2 U_i}{\partial t^2} = \frac{1}{p} \quad \lambda \left\{ \frac{\partial^2 U_k}{\partial x_i \partial x_k} + \mu \left(\frac{\partial^2 U_i}{x_j x_j} + \frac{\partial^2 U_j}{x_i x_j} \right) + \cdots \right\} \qquad (3.25)$$

where λ and μ are the Lamé constants. We expect that the higher-order terms, both in (3.25) and (1.14) will reflect

features specific to disordered systems. These remain to be elucidated.

Acknowledgement

References

1. Kohn, W., "Propagation of Low Frequency Elastic Disturbances in a Composite Material," *Journal of Applied Mechanics*, to be published.

2. Huntington, H. B., "The Elastic Constants of Crystals," in *Solid State Physics*, Vol. 7, F. Seitz and D. Turnbull, eds., Academic Press, New York, 1958.

3. Musgrave, M. J. P., "Crystal Acoustics," Holden-Day, 1970.

4. Peck, J. C. and Gurtman, G. A., "Dispersive Pulse Propagation Parallel to the Interfaces of a Laminated Composite," *Journal of Applied Mechanics*, Vol. 35, 1969, pp. 479-484.

5. Krumhansl, J. A., "Propagation of Transients in Periodic Composites," in Preliminary Reports, Memoranda and Technical Notes of the ARPA Materials Summer Conference, Department of Chemical and Metallurgical Engineering, University of Michigan, 1970, pp. 176-185.

6. Kohn, W., Krumhansl, J. A. and Lee, E. H., "Variational Methods for Dispersion Relations and Elastic Properties of Composite Materials," *Journal of Applied Mechanics*, Vol. 35, 1972, pp. 327-336.

7. Nemat-Nasser, S., "General Variational Methods for Waves in Elastic Composites," *Journal of Elasticity*, Vol. 2, 1972, pp. 73-90.

8. Hashin, Z. and Shtrikman, S., "A Variational Approach to the Theory of the Elastic Behavior of Multiphase Materials," *Journal of the Mechanics and Physics of Solids*, Vol. 11, 1963, pp. 127-140.

9. Sun, C. T., Aschenback, J. D. and Hermann, G., "Continuum Theory for a Laminated Medium," *Journal of Applied Mechanics*, Vol. 35, 1968, pp. 467-475.

10. Born, M. and Huang, K., "Dynamical Theory of Crystal Lattices," Chapter V, Oxford Clarendon Press, 1954.

11. Luttinger, J. M. and Kohn, W., "Motion of Electrons and Holes in Perturbed Periodic Fields," *Physical Review*, Vol. 97, 1954, pp. 869-883.

APPENDIX

In this appendix we present an alternative discussion of the subject matter of Section 2, based on Born's method of long waves[10].

We begin with Eq. (2.3), in which we substitute for $m_{ij}(k)$ from Eqs. (2.4) and (2.5) and write

$$k_p \equiv \epsilon \kappa_p \quad , \tag{A.1}$$

where κ_p is a unit vector, __held fixed for the time being__, and ϵ is a small parameter.

$$\omega^2 \rho(\underset{\sim}{r}) \; v_i(\underset{\sim}{r},\underset{\sim}{k}) = [m_{ij}^{(0)} + \epsilon \kappa_p m_{ij;p}^{(1)} + \epsilon^2 \kappa_p \kappa_q m_{ij;pq}^{(2)}]$$
$$\times \; v_j(\underset{\sim}{r},\epsilon \underset{\sim}{\kappa}). \tag{A.2}$$

Now, for fixed κ_p, we can write

$$\omega = \epsilon \omega_1 + \epsilon^3 \omega_3 + \cdots \tag{A.3}$$

and solve (A.2) by the power series Ansatz

$$v_i(\underset{\sim}{r},\epsilon \underset{\sim}{k}) = v_i^{(0)}(\underset{\sim}{r}) + v_i^{(1)}(\underset{\sim}{r}) + 2v_i^{(2)}(\underset{\sim}{r}) + \cdots \tag{A.4}$$

Substitution in (A.4) and collecting terms of successive powers in ϵ gives

$$0 = m_{ij}^{(0)} v_j^{(0)}(\underset{\sim}{r}) \tag{A.5a}$$

$$0 = m_{ij}^{(0)} v_j^{(1)}(\underset{\sim}{r}) + \kappa_p m_{ij;p}^{(1)} v_j^{(0)}(\underset{\sim}{r}) \tag{A.5b}$$

-526-

$$-\omega_i^2 \rho(\underset{\sim}{r}) v_i^0(\underset{\sim}{r}) = m_{ij}^{(0)} v_j^{(2)}(\underset{\sim}{r}) + \kappa_p m_{ij;p}^{(1)} v_j^{(1)}(\underset{\sim}{r})$$

$$+ \kappa_p \kappa_q m_{ij;pq}^{(2)} v_j^{(0)}(\underset{\sim}{r}) \qquad \text{(A.5c)}$$

etc. Since $m_{ij}^{(0)}$ is a differential operator (see Eq. (2.2), the first of these equations is satisfied by <u>any</u> $\underset{\sim}{r}$-independent vector

$$v_i^0(\underset{\sim}{r}) = U_i \qquad \text{(A.6)}$$

The second equation is a linear, inhomogeneous equation for $v_i^{(1)}(\underset{\sim}{r})$ in terms of $v_i^0(\underset{\sim}{r})$ $(=U_i)$. In order for this equation to have a solution it is necessary that the inhomogeneity be orthogonal to the solutions of the homogeneous equation, i.e.,

$$\int U_i^* \kappa_p m_{ij;p}^{(1)} U_j \, d\underset{\sim}{r} = 0 \qquad . \qquad \text{(A.7)}$$

By Eq. (2.2) this is equivalent to

$$\kappa_p i \, U_i^* \, [\int \frac{\partial}{\partial x_1} c_{iljp}(\underset{\sim}{r}) d\underset{\sim}{r}] \, U_j = 0 \qquad . \qquad \text{(A.8)}$$

This last equation follows immediately from the periodicity of $c_{iljp}(\underset{\sim}{r})$. Thus (A.5b) can in fact be solved and determines $v_i^{(1)}(\underset{\sim}{r})$ in terms of U_i

$$[\frac{\partial}{\partial x_1} c_{iljm}(\underset{\sim}{r}) \frac{\partial}{\partial x_m}] v_j^{(1)}(\underset{\sim}{r}) = -k_p i [\frac{\partial}{\partial x_1} c_{iljp}(\underset{\sim}{r})] U_j \qquad . \quad \text{(A.9)}$$

We write this solution as

$$v_i^{(1)}(\underset{\sim}{r}) = ik_1 \cdot v_{ij1}(\underset{\sim}{r}) U_j \qquad . \qquad \text{(A.10)}$$

Finally, in order that the third equation (A.5c) have a solution it is again necessary that the inhomogeneous part of the equation be orthogonal to all solutions of the homogeneous equation, U_i, i.e., to all space-independent vectors. Hence we integrate the i-component of the inhomogeneous term over a cell and equate the result to zero:

$$-\omega_1^2 \frac{1}{\bar{\rho}\Omega} \int_{cell} d\underset{\sim}{r}\rho(\underset{\sim}{r})U_i = \kappa_p\kappa_q \frac{1}{\bar{\rho}\Omega} \int_{cell} d\underset{\sim}{r}[m_{ij;p}^{(1)}iv_{jlq}(\underset{\sim}{r})U_i$$

$$+ m_{ij;pq}^{(2)}U_j] \qquad (A.11)$$

or

$$\omega_1^2 U_i = M_{ij;pq}^{(2)}\kappa_p\kappa_q U_j \qquad , \qquad (A.12)$$

where

$$M_{ij;pq}^{(2)} = -\frac{1}{\bar{\rho}\Omega} \int_{cell} d\underset{\sim}{r}[m_{il;p}^{(1)}iv_{ljq}(\underset{\sim}{r}) + m_{ij;pq}^{(2)}] \qquad . \qquad (A.13)$$

If we re-introduce k_p and ω by Eqs. (A.1) and (A.3), we find that for small k_p, the eigenfrequencies are determined by the secular equation

$$|M_{ij;pq}^{(2)} k_p k_q - \omega^2\delta_{ij}| = 0 \qquad , \qquad (A.14)$$

and the eigenvectors are given by

$$v_i(\underset{\sim}{r},\underset{\sim}{k}) = (\delta_{ij} + ik_lv_{ijl}(\underset{\sim}{r}) + \cdots)U_j \qquad (A.15)$$

where U_i are solutions of Eq. (A.12). These results are of the same form as those obtained in Section 2.

Comparison of Born's method of long waves, used in this Appendix and the method of Kohn and Luttinger used in Section 2 shows the following differences. The Born method gives the low orders eigenvectors and eigenfrequencies in terms of a function $v_{ijl}(\underset{\sim}{r})$ (A.9, A.10) defined as solution of a differential equation. It does not require, as a preliminary, the construction of the complete set of periodic eigenvectors $u_{\mu\nu;i}(\underset{\sim}{r})$ of Eq. (2.8) nor the summation of an infinite series as in Eqs. (2.30) and (2.35). On the other hand, the method of Section 2 yields the structure of the solution to arbitrary order of perturbation theory, in particular the form (2.34) of the secular equation for arbitrary $\underset{\sim}{k}$ and the corresponding form (2.40) of the three coupled differential equations for the macroscopic displacement vector $U_i(\underset{\sim}{r},t)$.

A NOTE ON THE STORAGE OF ENERGY

M. B. Bever

Abstract

This note briefly considers general features of energy
storage. It then reviews some materials aspects of the storage
of thermal energy and chemical energy.

A NOTE ON THE STORAGE OF ENERGY

M. B. Bever

Energy systems may require the storing of energy for operating and economic reasons. Energy storage assists in meeting peak loads and improves the load factor thus reducing capacity needs. In the utilization of an intermittent energy source such as solar energy, the ability to store energy is usually essential.

Different kinds of energy can be stored by different means. Typical examples are:

Energy	Storage
thermal	reservoirs; regenerators
potential	pumped storage
kinetic	flywheels
elastic	mainsprings
chemical	synthetic fuels; other raw materials
electrical	condensers
electrochemical	batteries
magnetic	special devices

Energy may be converted from one form to another for effective storage. For example, surplus electrical energy may be stored as electrochemical energy (batteries) or as potential energy (pumped storage).

Energy storage facilities may involve materials problems. For example, the ideal material for flywheels has high mechanical strength and low density; apparently, these requirements can best be met by making the wheels out of composite materials.[1] Other examples are the mechanical properties of mainsprings, the thermodynamic properties of synthetic fuels and the electrochemical properties of materials for batteries. In all these cases, other properties, such as corrosion resistance, may also be required.

This note briefly reviews some materials problems involved in storing thermal energy and chemical energy.

Storage of Thermal Energy

Thermal energy can be stored as the enthalpy difference of a material

(1) at two temperatures ("sensible heat"),

(2) in two different states of aggregation (e.g. heat of fusion),

(3) at two concentrations (heat of solution).

The utilization of solar energy provides examples of thermal energy storage.[2] Experimental installations of solar house heating have used mechanisms (1) and (2) listed above. Water and rocks are suitable media for thermal energy storage. Water has the advantages of ready availability, large heat capacity and suitability for direct operation without heat exchangers; its major disadvantage is the limited range of operating temperatures. Rocks allow considerably higher storage temperatures and greater efficiency of the collection system.

-533-

Sodium sulphate decahydrate (Glauber's salt) and sodium phosphate dodecahydrate have been considered for storage mechanism (2). These salts melt and dissolve in their water of crystallization. Their melting points (90° and 96°F), however, are too low for best operation; also, the reversibility of the reaction seems to deteriorate with time because of increasing difficulties of nucleation and gradual phase separation.

Sulphuric acid in concentrated and dilute solution and nickel chloride solutions have been suggested as systems for utilization of the heat of solution. Although these systems have some promise for storing energy, corrosion problems, complexity of operation and costs seem to militate against their use.

The thermal storage systems mentioned in the foregoing have been considered for storing low-temperature thermal energy. For high-temperature storage rocks or molten salts have been suggested.

The storing of thermal energy requires thermal insulation. This introduces further materials requirements.

A detailed analysis of thermal energy storage for solar energy applications is presented in Ref. 2. Ref. 3 gives some numerical data for low-temperature and high-temperature energy storage.

Storage of Chemical Energy

Energy extracted from a source such as solar energy can be utilized to produce a material which is then available for

use in an exothermic reaction. The production of synthetic fuels is an example; in particular hydrogen has been suggested as the basis of a "hydrogen economy".[4] Hydrogen would be a clean fuel - especially if burned with pure oxygen in a twin H_2-O_2 system; it could also be used in an engine or fuel cell.

Metals, for example aluminum, represent stored energy. In the thermite reaction aluminum liberates high-temperature heat which has been used in special applications, but metals in general are not well suited for chemical energy storage.

It has been argued that aluminum represents an "energy bank" because it can be recycled with the expenditure of a small fraction of the energy required for the primary production of aluminum. Within limits, which are mainly due to carrying costs, surplus energy can be used to produce energy-intensive materials for inventory.

Chemical energy is free energy rather than enthalpy. It is stored as such in fuels and raw materials but, as the example of synthetic fuels shows, it may be released as thermal energy.

Acknowledgement

References

1. G. F. Morganthaler and S. P. Bonk, Composite Flywheel Stress Analysis and Materials Study, Advances in Structural Composites, Vol. 12, Society of Aerospace Material and Process Engineers, (1967) D-5.

2. G. O. G. Löf, The Heating and Cooling of Buildings with Solar Energy, Chapter 11, Introduction to the Utilization of Solar Energy, ed. by A. M. Zarem and Duane D. Erway, 1963.

3. W. E. Morrow, Jr., Solar Energy Utilization, Prepared for Federal Power Commission National Power Survey, June 1973.

4. W. E. Winsche, K. C. Hoffman and F. J. Salzano, Hydrogen: Its Future Role in the Nation's Economy, Science Vol. 180 (1973) 1325.

SOME COMMENTS IN FAVOR OF HYDRIDES RATHER THAN
HYDROGEN AS AN ENERGY TRANSPORT MEDIUM

R. Gomer, M. F. Hawthorne and J. L. Margrave

It has become fashionable to cite hydrogen as the ideal energy transport medium in the (near?) future when natural gas, fuel oil and other petrochemicals may not be available. Elemental hydrogen is obtainable from water by electrolysis and from organic matter through bacterial processes. It is light and produces a large increment of energy per gram combusted in oxygen. The product of combustion is water so that there is a minimal pollution problem. On the other hand, hydrogen as an energy source would have to be transported, probably by high pressure pipelines, and utilized in industries, in power plants, and possibly in automobiles and homes. There are problems of leakage and the danger of explosion when leaks occur. Furthermore, the general problem of hydrogen embrittlement of metals appears formidable and it is not currently known whether or not high pressure hydrogen can be transported through available pipelines made of available materials without prohibitive capital costs.

The purpose of this note is to call attention to the basic chemical knowledge which exists in the area of hydride chemistry and to point out the many advantages of hydrides as energy

sources. The hydrides of interest are (1) the simple hydro-
carbons, which currently come directly from the earth or are
refined from petroleum; (2) the boron hydrides, which can be
utilized to "fix" hydrogen in a somewhat more energetic form
than in ordinary hydrocarbons, and for which an extensive pre-
parative chemistry and engineering experience were obtained in
the 1955-61 high energy fuels program; and (3) other light
element hydrides -- $LiBH_4$, $B(CH_3)_3$, etc. -- which have been
extensively studied.

Since the hydrocarbons and the boron hydrides are fairly
similar in many properties, there should be no difficulty in
transporting these compounds through existing pipelines. In the
case of boron compounds, one would probably seek to recover
hydrogen, burning it as the fuel and to recycle the boron, both
because B is a relatively scarce element and because the products
of combustion, boric oxide or boric acid, would be solid pollu-
tants of some toxicity. In the case of lithium, there is also
a very limited amount of the element in the earth's crust, so it
would have to be cycled if it were part of the energy system.
These limitations would have to be considered against the higher
energy content per gram of fuel in choosing a system.

In contrast with the lithium and boron hydrides, the
familiar hydrocarbons can be easily synthesized from common
starting materials by well-known reactions. Ideally, synthesis
from atmospheric CO_2 and H_2O would lead to a pure energy cycle,

in which the combustion products, mixed in the atmosphere, are re-used. The losses (apart from Carnot losses inevitable in any combustion) arise from the fact that high pressure gases must be used to synthesize CH_4 while CO_2 and H_2O will be released at low pressure. Thus, per mole of CH_4 (at 1 atm.) a loss of RT $\ln 1/P_{CO_2 \text{ atm.}} \simeq 6$ Kcal/mole will be incurred. This is not serious, relative to other factors. The materials can be transported readily in the systems we now have available, and the products of combustion are CO_2 and water, with no sulfur dioxide pollution, as long as one uses pure synthetic CH_4 or C_2H_6. Current engines and devices, slightly modified, could burn methane or ethane effectively.

In summary, it may be uneconomical to resort to an energy system based on elemental hydrogen. Hydrocarbons like methane and ethane can be prepared efficiently from cheap starting materials, transported in existing pipelines and burned in existing engines, furnaces and power generators.

Acknowledgement

FIELD OBSERVATIONS ON THE DEFORMATION OF ROCKS

F. A. McClintock

In the consideration of earthquake mechanisms, the
question frequently arises as to the deformability of rocks in
the immediate neighborhood of the fault zone. Can large blocks
slide over each other or do they become pulverized with very
broad zone deformation in which the sheer strain is low? Some
observations were made which may bear on this phenomena while
climbing in Cuaymacho Ranch State Park, forty miles west of San
Diego. Stonewall Peak has an outcropping of what the layman
would call granite but which was characterized in the museum,
as I recall, as granerite. It was mostly white or cream color
in appearance with a few percent of a black phase and a grain
size of the order of a few millimeters. Three observations were
made on the ridge running north from Stonewall Peak which was
reached by bushwhacking across from the first switchback of the
trail, although easier access could be obtained by going over
the peak and down.

The first observation was made about 200 meters down the
ridge and 20 meters off the ridge to the right. Here a block
perhaps half a meter to a meter thick had slid across a joint
through a displacement of the order of one-half meter. The under-
lying rock showed that the black material was extended in streaks

Preceding page blank

-541-

two or three centimeters long, but approximately the original
width and rather uniformly distributed over the base of the rock.
The second observation was made at the crest of the ridge just
above the first observation. Here on the crest itself was a slab
perhaps one-half by three meters on whose upward facing surface
similar markings were found. They were remarkably uniformly
distributed over the surface, but surface weathering indicated
that a few millimeters down the original grain structure existed.
The character of surface faulting suggested that perhaps the
surface layer of the rock was badly cracked and susceptible to
spall. Close observation indicated an appearance of the rock
really being deformed and being hard, although perhaps somewhat
more cracked than normal. There seemed to be no doubt that the
markings were minerals in the rock and not simply stain.

The third observation was made about 30 meters below the
summit of the peak itself on a rock separate from the original
cliff. The location was again on the western side of the north
ridge at a place where two chimneys provide obvious climbing
routes to the ridge. Here there seemed to be a transfer of black
rock which was found in the vicinity, and here also the markings
were more unevenly distributed as one would expect with a slightly
wavy surface.

Before making these conclusions, the author would have
said that smearing of rocks was unlikely as near to the surface
as these rocks must have been. Barring other evidence, he is

convinced that some of the minerals in these rocks were smeared
by sliding. He would welcome more informed comment on these
observations. If they are indeed correct, there appears a real
possibility that earthquake faults can be so confined that a
deformation is limited to a surface only millimeters thick with
attendant heating and smearing of the sliding surfaces.

Acknowledgement

ON STRESS WAVE PROPAGATION DIFFERENCES
BETWEEN GRADED AND GRADIENT MATERIALS*

E. H. Lee, B. Budiansky and D. C. Drucker

Abstract

This work contains a comparison of the resistance to
impact of built-up plates with stepped mechanical properties
(such as density and elastic constant) with plates having a
similar range of properties distributed continuously. Marked
differences arise in the stress generated in such types of
plates under the same impact loading. In particular, the ten-
dency to spall can be quite different in the two cases.

This work continues an initial effort last year for
which the computing was then not completed. We now have the
solution to a much more meaningful problem, for which we can
vary the width of transition zone between homogeneous regions
with different properties, and so assess in detail the change
from continuous variation to discontinuous steps in properties.

This problem arose from a discussion with visitors from
AMMRAC in the sessions on gradient materials arranged by
Professor Bever.

Note: Full paper will be included in a later report.

Preceding page blank

WORKSHOP ON RELIABILITY OF BRITTLE MATERIALS

R. M. Thomson, R. L. Coble
and F. A. McClintock

Abstract

A workshop of approximately 15 people was organized to review the methods for flaw detection for ceramic materials and methods for assuring mechanical reliability. This paper reviews the collected opinion of the workshop regarding the present status of the field and argues that periodic inspection for flaws will no. be as effective as careful quality control during manufacturing because of the extreme sensitivity of crack growth to temperature, intiial crack size, and operating stress.

Preceding page blank

SUMMARY OF DISCUSSIONS ON THE
WORKSHOP ON RELIABILITY OF BRITTLE MATERIALS

R. M. Thomson, R. L. Coble
and F. A. McClintock

I. INTRODUCTION

On July 16 through 18, a small group met in La Jolla,
California, under the auspices of the Materials Research Council
(MRC), to consider the present status of our ability to assure
the reliability of brittle materials in high temperature struc-
tural applications. The workshop was organized with the general
goals in mind of the ARPA program in brittle materials design.
This paper in part reports the results of the discussion of the
workshop, and in part summarizes independent views of the authors
regarding the reliability of brittle materials. It is our belief
that when the materials fabrication problems have finally been
worked through and the final designs have been adopted, one will
still be left with the long-term problem of assuring the reli-
ability of the material in its application, and perhaps in moni-
toring the material for possible deleterious changes in use.
These problems are important for a ceramic for the very reason
that the size of flaws critical to the reliability of a structural
part is an order of magnitude smaller than in the usual conven-
tional metallic materials.

There is a basic trade-off between the insistence on the fabrication of an absolutely perfect material, on the one hand, and on the other hand, allowing some small probability for the appearance of flaws, but coupled with a flaw detection scheme which can find these flaws. In a practical application, the optimum trade-off will rest somewhere in the latter category. The major attention in programs for the application of ceramics to structural applications has so far emphasized the development of the material's properties and the fabrication techniques themselves. The object of our workshop was to survey where we stand in the longer term problem of reliability assurance.

II.   ENGINEERING REQUIREMENTS

A.   Transient and Steady State Requirements

The engineering requirements of the two parts of the ARPA program represent in some respects opposite ends of the applications spectrum for ceramics. For the large land-based gas turbine for electrical power production, material cost is not nearly so important as in the small automobile turbine, and likewise sophisticated inspection procedures can be envisioned as a maintenance procedure through the life of the system. In the small automobile turbine, although quality control in manufacture can be sophisticated, maintenance will have to be highly simplified. In this latter case, a safe reliable life will surely be a requirement. One advantage of the small turbine is that pre-

sumably the total life will perhaps be considerably shorter than in the large electrical powered turbine.

The susceptibility of either design to failure is of course a highly complex subject, since stresses and temperatures vary from place to place and depend strongly on operating conditions. From the most recent progress report[1], it is possible to pinpoint the apparent most critical situations. These are listed in the accompanying Table I. Generally speaking, two threats exist. The first is due to high stresses caused by transients set up by sudden changes in operating conditions, such as an emergency shutdown of the power turbine or a full power acceleration from a cold start in the small turbine. These transients set the condition for instantaneous strength in the design. The second threat is posed by slow crack growth over the lifetime of the material due to a combination of high temperature and moderate stress. From the data available in the most recent progress report, we have again listed in Table I what appear to be the worst cases for slow crack growth in both turbines. The design stresses and temperatures given in Table I than make it possible to estimate the critical sized flaw and the time to fracture due to slow crack growth. To make these estimates, we use the slow crack growth data of Evans and Wiederhorn[2], and assume that the crack velocity law is given by

$$\dot{c} = \alpha K^n e^{-E/RT} \tag{1}$$

where K is the stress intensity factor, and $\dot{c}$ is the velocity of the crack, α is a constant, E is the activation energy for slow crack growth, R is the gas constant, and T is the temperature. From Evans and Wiederhorn, the various constants are given by

$$\alpha = 1.37 \times 10^{-36}$$
$$n = 9$$
$$E = 220 \text{ kcal/mole}$$

With substitution of the relation,

$$K = \sigma\sqrt{\pi c/2} \tag{3}$$

where σ is the external stress, we can then integrate Eq. (1) to give an expression for the time to failure, t.

$$t = \frac{2}{n-2} \left(\frac{2}{\pi}\right)^{n/2} \frac{e^{E/RT}}{\alpha\sigma^n} \frac{1}{c^{(n/2)-1}} \tag{4}$$

The expression for c in Eq. (3) is written for the entire crack length, instead of the usual half length. The reason is that only one tip of the crack is assumed to be growing. Also, Eq. (3) is written for the nominal plane stress case, and for three-dimensional configurations, the effective crack length is modified by geometrical factors. When the crack is on the surface, and its depth, c^*, is small compared to its length on the surface, then c of (1) is related to c^* by $c = 2c^*$. This flaw size is then used as a nominal longest flaw for the entire design.

Using this flaw size as the initial crack size for the slow crack growth conditions, we have derived the time to failure.

Equation (4) can be inverted to give an expression for the initial crack length for a given time to failure.

$$c = \left(\frac{2}{\pi}\right)^{n/(n-2)} \left[\frac{2e^{E/RT}}{(n-2)\alpha t}\right]^{2/(n-2)} \frac{1}{\sigma^{\frac{2}{1-2/n}}} \tag{5}$$

Equations (4) and (5) are derived on the assumption that the final crack length is much larger than the initial crack. We can also recombine these equations to give the result,

$$t = \frac{2}{n-2} \frac{c_o}{v_o} \tag{6}$$

From this equation, the only parameters necessary to find the time to failure are the initial crack length, c_o, and the initial crack velocity, v_o, taken from the plots of Fig. 1.

In computing the critical crack length for the transient cases in Table I, we take K_{Ic} to be given by the values for the crack velocity where $V = 10^{-1}$m/s in the results of Evans and Wiederhorn and use Eq. (3). (We show the results of Evans and Wiederhorn by permission in Fig. 1.)

According to Table I, the most serious transient stress for the small turbine appears in the hub area of the motor during a fast run-up from a cold start, and leads to a critical flaw size of about 100μm. The most critical area for slow crack growth appears in the neck area of the small turbine, and taking

this same sized flaw, we see that it will not grow to fracture under the given conditions for approximately 5000 hours.

The most serious transient stress for the large turbine occurs in the leading edge of the vane with a critical sized flaw in the neighborhood of 240μm. Under slow crack growth conditions in the vane, these same cracks can grow to fracture in about 20,000 hours. (A. Evans and E. Fuller have indicated privately to the authors that the situation may be somewhat worse than this when slow crack growth during the cool-down of the transient is included.)

The important conclusion from these design requirements is that means are presently required for finding flaws of a size in the neighborhood of 100μ.

B. Safe Life vs. Periodic Inspection

An important strategy for reliability assurance is visible in the sensitivity of the crack growth laws to operating and initial conditions. Namely, the time to fracture is an extremely steep function of the three parameters, stress, initial crack size, and temperature. Thus, if these parameters are set to give a safe life which is less than the total operating life, then periodic flaw detection procedures will be a necessary part of the maintenance procedure. However, by slightly decreasing the initial flaw size, operating stress, or operating temperature, the time to fracture can be enormously increased so that the entire expected life of the part in encompassed and the need for periodic testing is eliminated. (We note that halving the crack

-553-

size in Si_3N_4 increases the safe life by a factor of ten, decreasing the operating stress by 30% increases the safe life by a factor of ten, and decreasing the temperature by 50°C at 1000°C increases the safe life by a factor of 34.)

Quantitatively, we can express these results from Eq. (4) for the sensitivity of the time to fracture on the crack size as

$$t = ac^{-p} \qquad (7)$$

If we wish to compute the ratio by which the time to fracture has changed from t_1 to t_2 for a change in the initial crack length from c_1 to c_2, then

$$\frac{t_1}{t_2} = \left(\frac{c_2}{c_1}\right)^p \qquad (8)$$

where the subscripts relate to the two different choices. Since for Si_3N_4, $p = 3.5$, a factor of two change in the initial crack size changes the time to fracture by a factor of more than ten. The decision of whether to conduct ten periodic inspections during the total lifetime of the part or to carry out a more stringent search for flaws at the beginning with no periodic inspections depends upon comparing the cost for searching for flaws of a given size just once, or searching ten times during the life of the part for flaws of twice the size. Generally speaking, one would expect the slightly more stringent initial quality control to be a marginal cost compared to the cost of ten complete inspections.

There are other factors, however, in Eq. (4). For example, if the operating stress is not known to within 15%, in the same way, the time to fracture changes by

$$\frac{t_1}{t_2} = \left(\frac{1}{.85}\right)^9 = 4.3 \tag{9}$$

However, the way to catch this error is not to carry out periodic inspections! A crack which during a given maintenance period will grow to critical size will not have grown enough in previous maintenance periods to be observable. Thus, suppose that a crack in two maintenance periods, $2t_m$, will grow to critical size. The question is, by what ratio does its length change at the close of the first maintenance period, t_m? The answer, from Eq. (5), is

$$c_0 = a(2t_m)^{-1/p}$$
$$\tag{10}$$
$$c_1 = a(t_m)^{-1/p}$$

where c_0 is the length of a crack which will fracture after two maintenance periods, and c_1 is the length of a crack which will fracture after one. Thus,

$$\frac{c_1}{c_0} = 2^{1/3.5} = 1.22 \tag{11}$$

and the crack only changes its length by about 20% from the beginning to the end of the first maintenance period. Similar results are valid for temperature uncertainties, and for un-

certainties in the material parameters of the crack growth cases.

Our conclusion thus is that the gains from periodic inspection are marginal, provided all the cracks are present initially, and the quality control is perfect in catching all flaws larger than the specified size. While the reliability function for flaw detection certainly does depend upon flaw size, an exhaustive treatment of the gains to be realized from repeated inspection because of the imperfect flaw detection scheme is beyond the limits of this discussion. We can dispose readily of the creation of new flaws, say by foreign particle impingement, however.

The problem of new flaws introduced by foreign particles is actually the same as the previous case. Thus, we suppose that all particles are excluded which will create Hertzian cracks which can grow to fracture in one maintenance period, t_m. If $t_m = \frac{1}{10} t_{tot}$, where t_{tot} is the total life, then the ratio of crack size to be caught in one maintenance period, to the crack size which will grow to fracture in the total life is given as before by

$$\frac{c_m}{t_{tot}} = (10)^{\frac{1}{3.5}} \simeq 2 \tag{12}$$

Presumably, the foreign particles are controlled by a mesh screen. Hence, the size of the Hertzian crack is determined by the size of the mesh, and we have a relation between the size of a crack and the size of particle[3],

$$\beta_1 r = \left(\frac{c}{\beta_2 r}\right)^q \tag{13}$$

where β_1, β_2 are constants, r is the particle size, and q is about 1.5. Thus, for two different crack sizes in this equation, we have

$$\left(\frac{r_m}{t_{tot}}\right)^{q+1} = \frac{c_m}{c_{tot}} = 2 \tag{14}$$

$$\frac{r_m}{r_{tot}} = (2)^{\frac{1}{2.5}} = 1.3$$

Thus, the mesh to catch particles which can be allowed in the turbine which will be periodically inspected ten times during its life is only 30% different from that for an uninspected engine. Doubtless, it will be cheaper to make this marginal change in particle mesh than to carry out the inspections.

III. MATERIALS ASPECTS

In addition to the design variables, the special characteristics of ceramics as materials are important for reliability assurance, and in particular the type and distribution of flaws in the starting material is of special interest.

Internal flaws in the material are primarily associated with either voids which are left over from the sintering process, or are formed around precipitate particles having a different thermal expansion from the matrix material. Pores and voids situated on grain boundaries are more serious than when they occur within a grain because of the relative ease of propagating

a crack along the grain boundary relative to propagating the crack through the grain. Thus, a pore on a grain boundary will tend to form a crack running the full length of the grain boundary on either side of the pore to the point where the grain boundary branches as shown in Fig. 2. According to this picture, the materials as grown will then possess numerous cracks associated with aglomerate and voids with the peak in the crack distribution generally associated with the grain size of the material. Thus, uniformity in grain structure is important since out size grains are likely sites for cracks. In the same way, very large impurity aglomerate particles and large voids will be sites for crack formation.

A considerable difference of opinion developed in the workshop relative to the importance of surface versus internal cracks, and relative to the general range of crack size distributions. We take this divergence to be indicative of the fast rate of progress in material development, currently. Present commercially available hot pressed Si_3N_4 has a grain size of about 1μ, and often fails in strength tests from internal flaws associated with large voids and impurity aglomerates. On the other hand, recent research material prepared by the General Electric Research Laboratory (sintered S_iC) has a grain size of about 0.1μ, and exhibits no internal flaws larger than 10μ. In strength testing with this material, surface flaws due to surface preparation and handling are responsible for fracture initiation. These flaws are formed especially in the region of corners and

fillets. One hopes that the G.E. experience with research grade S_iC is typical of the results one can eventually expect of commercially available ceramics. If so, very high strengths of 100,000 psi and over will be a practicality, and the major reliability problem will then shift to controlling damage to the surface during manufacture and operation.

Another issue arose during the workshop relative to the mechanical testing which had been performed. Because of problems with gripping ceramics, the conventional mechanical test is a bend test. However, this emphasizes surface flaws over internal flaws, and the volume of material subjected to the maximum stress is smaller than one typically obtains in a tension test. Katz, at AMMRC, has developed a procedure for tension testing which uses a hydro-static loading technique. General Electric, on the other hand, has been developing statistics on strength by the use of a type of hardness testing and the generation of Hertzian surface cracks. Much work, of course, also remains to be done to gather the relevant statistics for that material which will eventually be used in the turbines. Finally, one is left with the dilema of how and whether one can extrapolate the statistical data in the extreme value regions. At this stage of the game, fully deterministic testing procedures seem to be the only reliable course.

We did not address in the workshop two other important material variables which in their own way contribute to the overall reliability, namely, long-term chemical changes on the surface and at cracks due to the operating environment, and creep.

In conclusion, so far as flaws in the material are concerned, typical strength in commercial material corresponds to a flaw size of about 100μ, and this flaw is at least as likely to be internal as on the surface. Present materials technology is in a fast state of flux, however, and research results indicate that materials containing critical sized internal flaws in the neighborhood of 10μ may be a practical goal. In this case, the flaw detection problem becomes very severe if the maximum strength is to be assured, and surface flaws caused by fabrication or operating conditions gain the center of attention. In any case, a quality assurance strategy should involve some kind of initial test of the materials before fabrication of the entire part to check for internal flaws, and subsequent testing of the surface to catch flaws initiated by fabrication processing of the finished product. Likewise, in service, the major threat would seem to be surface damage.

IV. TECHNIQUES FOR RELIABILITY ASSURANCE AND FLAW DETECTION

The techniques which the workshop considered to have the greatest potential for use in flaw detection and quality control were proof testing, acoustic emission, ultrasonics, penetrants, X-rays, and new techniques. The general opinion was that each of these techniques had important potential applications, but that none of them was capable of doing the entire job. In Table II, we have summarized the general opinions regarding the various techniques as expressed by the workshop. We also took a poll of

the attendees regarding which techniques were of greatest interest
for immediate development as inspection tools for ceramics. The
combination of proof test-acoustic emission and penetrants were
considered to have the highest potential for the ceramic applica-
tion.

A. Proof Testing

The general theory of a proof test, as it might be applied
to a ceramic, is that the part is subjected to a stress higher
than the operating stress. If the part does not break under the
proof stress, then the largest possible flaw in the material is
less than the critical size flaw for the proof stress. One can
therefore assume that under operating conditions, when slow crack
growth can occur, the part will be safe until the largest flaw
present at the time of testing can grow to its critical size.
From the data on slow crack growth, the period of safe operation
can be predicted from Eq. (4). In terms of the ratio of the
proof stress to the normal operating stress, Wiederhorn and Evans[4]
have derived the following expression for the safe operating period
(expressed in our notations).

$$
t = \frac{2(\sigma_0/\sigma)^{n-2} \exp(E/RT)}{(n-2)\alpha\sigma^2 \ K_{IC}^{n-2}} \tag{15}
$$

K_{IC} is the critical stress intensity factor for fracture, and
σ_0/σ is the ratio of proof stress to operating stress. The proof
test theory has been applied to cases where the stress is not a
simple constant, but varies in time by numerically integrating

Eq. (1). The theory does not apply to situations where there is an extra component to the crack velocity which is due to cycling effects, as in fatigue. Ceramics in their brittle range are not expected to show this effect, although it is a major term for ductile metals. In the higher operating regime where creep occurs, the possibility does exist that fatigue effects will become important for ceramics. This point has not so far been checked for silicon nitride in the higher temperature regimes. There are two major problems associated with the practical application of proof test for the turbine application. The first has to do with how to simulate the complex stress distribution in any of the various turbine parts where critical stress levels occur. For example, in the end caps that hold the vanes in the large power turbine, the high operating stresses appear only in a small area of a groove into which the vane fits. Likewise, in the automobile turbine, the major operating stress appears in the small portion of the neck region. It is probably impractical to attempt to simulate these stresses by mechanical means, but some discussion took place at the workshop regarding the possibilities for simulating the stresses by thermal quenching.

Thermal quenching is a very familiar technique in the laboratory for producing nonequilibrium vacancy concentrations in in metal processing, but clearly a good bit of work will be necessary in order to develop an adequate proof test by the use of thermal quenching. One obvious difficulty with quenching is

that unless the process is carefully controlled, additional cracks will be formed which were not originally in the material. Another difficulty is the detection of critical size cracks after the quench, since the normal result of a quench would probably not be the complete fracture of a whole specimen, as in the case of a mechanical proof test. Acoustic emission has been suggested as one way of detection the growth of critical sized flaws during a quench. The second major difficulty with the proof test is that if it were used in a full scale assembly by overspeeding the turbine, the failure of the turbine or a single part of the turbine could prove disastrous to the whole system. Thus, a straightforward overspeed proof test is not a desirable candidate as a maintenance procedure of the turbine. Again, there are possibilities that acoustic emission might be used in conjunction with a proof test to indicate crack growth of critical sized cracks before catastrophe occurs. The problems of detecting the sound emissions of cracks in the turbine in the presence of all of the extraneous noises which a rotating turbine will produce as well as the problem of attaching a transducer to the rotating turbine will produce as well as the problem of attaching a transducer to the rotating turbine will be difficult problems to work out.

Thus, in summary, although the proof test is an attractive idea which is based on very sound principles regarding the behavior of flaws in ceramics, and can be easily used in cases where the part is subjected during operation to a simple stress

distribution or where the stress distribution can be easily simulated, in the case of the ceramic turbine, considerable development will be necessary before a practical application will be possible.

B. Acoustic Emission

Acoustic emission as a flaw detection scheme relies on the fact that observable sonic pulses are emitted by a crack which is growing under stress in a polycrystalline ceramic. Obviously, acoustic emission is not a passive technique, and is really a special case of a proof test, but it has the advantage that it does detect flaws which are actually pathologic. (Passive techniques such as ultrasonics may, for example, be sensitive to open voids which may not be as pathologic as their physical size may indicate.) Figure (3) shows how acoustic emission for the case of porcelains closely follows the slow crack growth law, and a similar result is expected for silicon nitride and silicon carbide. Proof testing and acoustic emission are really part of the same basic technique, since in both cases the material must be under a stress at least partially simulating operating conditions.

Some very interesting work by Schuldies was reported at the workshop in which a laser beam was used to selectively heat and stress a small area so that the flaw responsible for the emission could be localized. This could be an important consideration, for example, in the case of the leading edge of a turbine vane where only a small portion of the total vane must

be critically tested. This technique has considerable promise as a method of "proofing" a part without the danger that the whole engine will explode, and has the advantage that the area causing difficulty is automatically located.

The basic problem with acoustic emission has to do with the difficulty of interpretation. Although such work as that of Evans and Linzer in Fig. (3) shows an empirical relation between emission and crack growth, the theoretical relationship between the type and intensity of emission from various types of sources has not been adequately established. Thus, an efficient use of this technique and full confidence in its interpretation will require a deeper level of understanding.

C. Ultrasonic Detection

Ultrasonics is used to detect flaws in much the same way that sonar is used to detect submarines. It is a highly developed technique which has proved very successful for metals.

Various modifications have been developed which make use of shear waves, surface waves, Lamb waves, etc., and acoustic holography has been developed to improve the imaging of the flaws in the material. Basically, the technique is likely to become an important tool for flaw detection in ceramics, but the increased resolution necessary to detect the small flaws in ceramics and the higher sound velocity in ceramics as compared with metals, both dictate the use of a higher frequency of order 10^8 or perhaps 10^9 Hz, than is normal. Some experimentation along this line is underway, notably by Seidel, at the University of Michigan. The

-565-

basic limit to the frequency which can be used is set by the
grain size and perhaps by dislocation damping. One additional
problem in the ceramic is the fact that the expected pathological
flaws are quite likely to be very tight brittel cracks rather
than the open cracks observed in metals. Theoretical work by
Yee shows that when the opening of the crack is of the order of
the wavelength of the sound in air, the reflectivity of the crack
is reduced.

Thus, although ultrasonics has been intensively developed
to find the relatively open large cracks of metals, considerable
work will be necessary before it can be applied directly to the
ceramic case. At this point, no fundamental reasons exist to
suggest that it cannot be modified to serve an important role in
ceramics provided the necessary detection schemes can be developed,
and the damping is not too severe at the necessary frequency.

Some discussion took place in the workshop regarding the
possible use of ultrasonics to detect large grains. Since the
overall quality of the materials will depend on having a well-
specified grain size, it will be desirable to be able to detect
occasional grains falling well outside the normal distribution.
Ultrasonics may be a candidate provided the necessary frequency
can be obtained.

D. Penetrant Techniques

The basic idea of the use of the penetrant to detect a
flaw is to wash the surface of the material with a penetrating
oil. The oil then seeps into small surface irregularities and

-566-

flaws. Next, the surface is wiped clean of the excess oil, leaving deposits behind which coat the inside surface of minute cracks and dimples. Next a "developer" is painted on the surface which draws the penetrant out of the cracks and into the film of the developer. If the original penetrant contained either a bright dye or a fluorescent material in its composition, then the developing film will contain a brightly visible map of the flaw structure on the surface. The general process described is simple in application, cheap, and can be made sensitive to very small cracks and flaws. (Cracks 125 microns deep and 500 microns in length, have been detected in ceramics by penetrant techniques at United Aircraft.) The major disadvantage of the technique is that it is sensitive only to cracks and voids which intersect the surface. It is also more a qualitative than a quantitative tool in its traditional form, since a measure of the depth of the flaw is not provided. It has not been demonstrated yet whether a penetrant will, in fact, be sensitive to a tight surface crack and thus be able to detect all the pathological flaws occurring on the surface of the ceramic. Suggestions have bee   ie for ways in which penetrants can be made more quantitative   nd one of these methods is to adsorb an isotope of krypton on the surface cracks and measure the radioactivity. This method is likely to be more sensitive than liquid penetrants, in addition to having potential for quantitative work.

E.   Radiography

Both X-ray and neutron radiography have a demonstrated usefulness as detection schemes, however, in order to detect a void or crack, a general rule of thumb is that the X-ray density through the region containing the flaw must differ from the X-ray density through good material by more than about 1%. Thus, only voids of a radius of about 1% of the total thickness of a sample can be detected. On the other hand, heavy particulates, such as tungsten compounds, are detectable to much greater sensitivity, depending upon the difference in atomic number between precipitate and matrix. Also, neutrons can be used in the same way for detecting impurity particles whose neutron scattering cross section differs strongly from the matrix. Since boron is one of the important impurities in sintered silicon nitride, the effectiveness of neutron radiography for the detection of boron particles will be surveyed.

F.   New Techniques

Considerable discussion occurred relative to possible development of new techniques and of improvement and modification of older methods so that they are more suitable for the ceramic application.

Intensive development is currently taking place for the application of holographic techniques, both optical and acoustic, to the flaw detection problem in other materials. Optical holography as applied to flaw detection is a method of measuring differences in strain on a surface as a function of stress, and

it is a very sensitive technique. It should be surveyed for possible application to the tight ceramic cracks. Acoustic holography is, in effect, a process by which ultrasonic scattering and reflection by flaws in the material is made visible. If one can solve the basic problem of getting the resolution necessary by ultrasonics, this technique could then have some interest for imaging the result.

Some very interesting ideas were discussed relative to possibilities for using dielectric breakdown around a crack which has been painted by a conducting layer. This suggestion is discussed separately in another report of the MRC.

Other possibilities exist for taking advantage of the fact that silicon nitride is a good insulator, through dielectric loss measurements. It could be used, for example, with microwave probing as a method for surveying relatively large cracks. Detection of the 100 micron crack would, of course, require sub-millimeter microwaves which is a difficult region. For thin samples, the material is transluscent and a simple optical scanning could be used for flaw detection for very near surface flaws.

V.   CONCLUSIONS

We are still too deeply involved with important materials improvements and even with design concept changes to be able to develop an exact picture of what the eventual details of the reliability requirements will be and what precise techniques will

be most effective in assuring this reliability. However, certain broad generalizations appear to be possible. Briefly, these are:

1. So far as previously existing cracks are concerned, the extreme sensitivity of slow crack growth to temperature, stress, and initial flaw size indicates that it will be relatively easy to operate on a safe life principle. Since decreasing the largest initial flaw size in a part by a factor of two increases the safe life of the part by a factor of ten, no advantage accrues to monitoring slow crack growth during the service life of the ceramic part.

2. Problems of flaw generation during operation have not generally been studied. However, the same principles will be applicable there also. Namely, periodic inspection for flaws created during service does not seem to be feasible, and screens which eliminate particles causing flaws which can grow to critical size during the life of the engine will be necessary.

3. Materials development is moving in the direction that one may be able to eliminate internal flaws as a serious problem after initial quality control check of the material, with surface damage being the primary flaw control problem.

4. Under these conditions, penetrants and proof testing (possibly in combination with acoustic emission) were the techniques which appeared to the working group to have the greatest potential. No one in the group, however, felt that the standard flaw detection techniques which have been developed and optimized for the metallic application would be adequate in their present

guises, and that considerable adjustment and development will be necessary to optimize any of these techniques to the ceramic case. (The resolution of u trasonics will have to be considerably increased, cracks in ceramics are likely to be very tight compared to those in metals, advantage may be taken of the non-conducting property of the ceramic, etc.)

5. A number of new possibilities for flaw detection in ceramics occurred to us such as the use of activated krypton, corona discharge at crack tips, laser microstrain techniques, etc.

6. As in the case of traditional metallic flaw detection, it is believed that for ceramics, a "bag of tricks" approach to flaw detection is desirable so that flexibility is assured for meeting various requirements.

ACKNOWLEDGEMENT

REFERENCES

1. Brittle Materials Design, High Temperature Gas Turbine, AMMRC CTR 73-9, March 1973.

2. S. Wiederhorn and A. Evans, to be published, Int. J. Fract. Mech.

3. A. Evans, to be published, J. Am. Cer. Soc.

4. A. Evans and S. Wiederhorn, to be published, J. Matl. Sci.

5. A. Evans and M. Linzer, to be published, J. Am. Cer. Soc.

TABLE I: ENGINEERING REQUIREMENTS REGARDING CERAMIC STRENGTH

Small Turbine

Comment	Stress Level	Critical Crack Size (Instantaneous Fracture)	Temperature	Time to Fracture
Transient Conditions				
Maximum stress is in neck and hub. Neck has higher temp. Transient lasts ≈ 30 sec.	320 MN/m² (48,000 psi)	120μm	684°C	
Steady State Operating Conditions				
Neck area 1st stage	231 MN/m² (34,000 psi)	360μm	1129°C	4800 hrs. (C=100μm)
Neck area 2nd stage	292 MN/m² (43,000 psi)	180μm	906°C	10⁹ hrs. (C=100μm)

Large Turbine

Comment	Stress Level	Critical Crack Size (Instantaneous Fracture)	Temperature	Time to Fracture
Transient Conditions				
Vane leading edge. Maximum stress lasts ≈ 10 sec.	279 MN/m² (41,000 psi)	240μm	1106°C	
Steady State Operating Conditions				
Vane end caps	143 MN/m² (21,000 psi)	700μm	1073°C(?)	4×10 hrs. (C=240μm)
Vane (gas load)	34 MN/m² (5,000 psi)		1400°C	21,000 hrs. (C=240μm)

TABLE II: RELATIVE COMPARISON OF NDT TECHNIQUES FOR CERAMICS

Technique	Present Status	Potential Utility	Major Limitations
Proof Testing	Lab development. Room temp. crack growth studies by NBS, Westinghouse. Presently useful in simple stress, shape situations.	Has potential for use in more complex stress situations, especially time effects. Also may have a potential for more complex shapes, and thermal distributions.	Presently difficult to apply to complex shapes, combinations of stress and thermal stress. Must simulate service conditions accurately.
Acoustic Emission	Lab development at NBS, Westinghouse, Ford, Penn. State, etc.	Potential for: 1) Flaw detection by itself. 2) High potential in conjunction with proof testing. 3) Emission from rubbing of tight cracks.	Noise discrimination is serious. Correlation with μ structure of material presently poor. Needs development for high temp. applications. Needs technique development to make use of all information available from an event. Application to rotating parts not yet possible.
Ultra-sonics	Gross flaws, survey. Lab development of high frequency.	Potential for practical use for small flaw detection limited to higher frequency development. May have a potential for finding fluctuations in grain size. Has potential for void detection.	Random flaw orientation makes interpretation difficult. Tight cracks difficult to distinguish.

-574-

TABLE II: (continued)

Technique	Present Status	Potential Utility	Major Limitations
Pene-trants	Has been able to detect flaws >500µ long, 125µm deep.	Has been developed for metals as result of empirical development. Has potential for greater quantization. Needs optimization for ceramics, i.e., better wetting of ceramic surface and diffusion into tight cracks. Potential is high if background roughness can be controlled. Kr development looks good.	Limited to surface flaws. Present experience is variable. Surface roughness is a problem. Cannot give depth quantitatively.
Radiog-raphy	Survey tool good for inclusion detection when W is a problem.	Westinghouse will look at B inclusions with neutrons. Can improve sensitivity by µ focus, instrumentation.	Sensitivity is a function of thickness. Tight cracks unobserved. Random orientation is a problem.
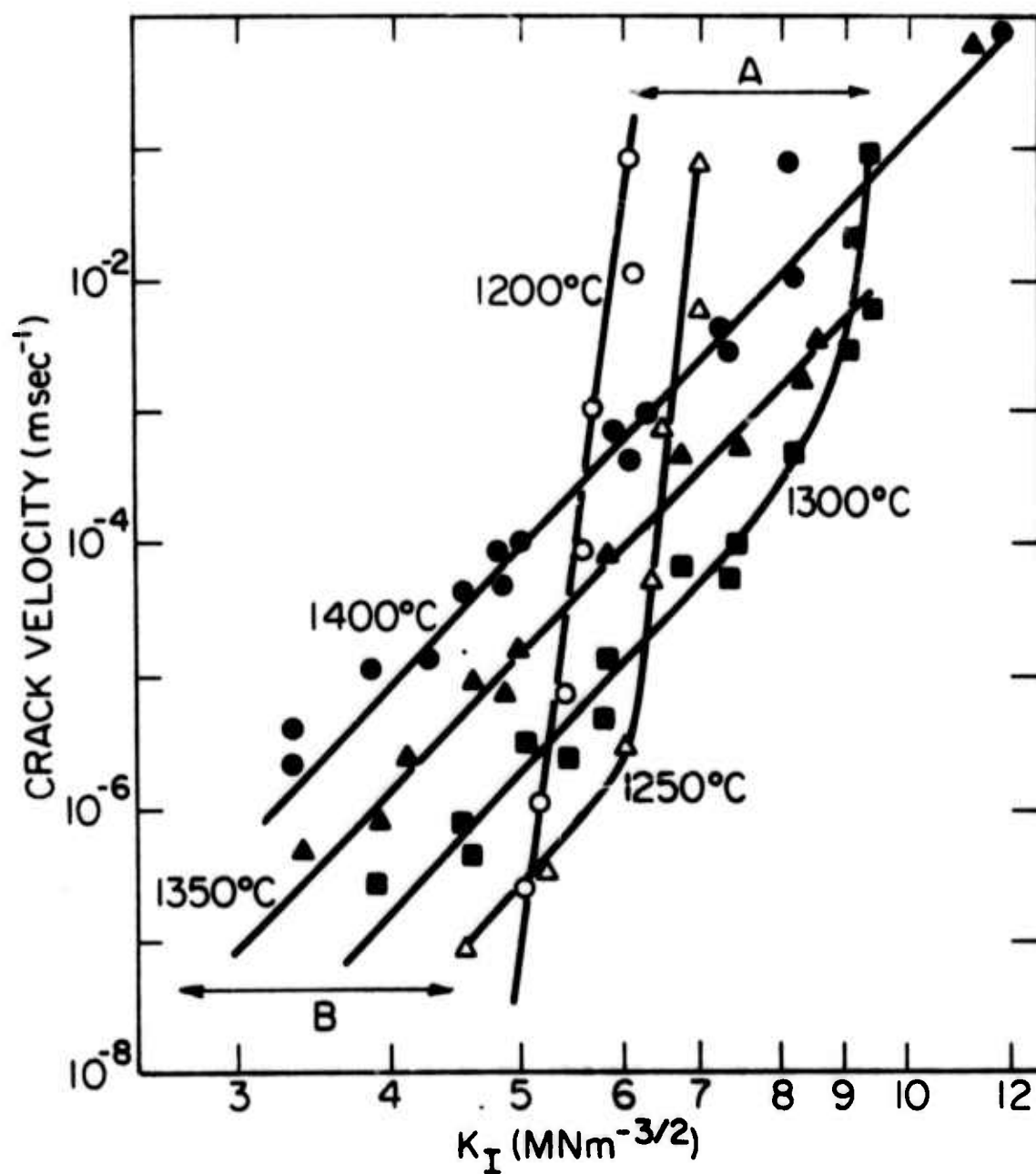
Figure 1. Crack velocity as a function of stress intensity in Si_3N_4 by permission from Evans and Wiederhorn.
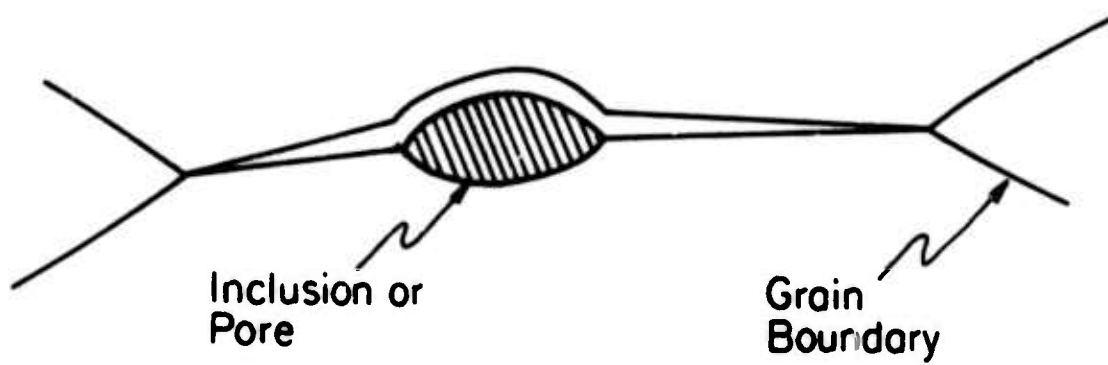
Figure 2. Schematic drawing of a crack caused by a pore on
a grain boundary.
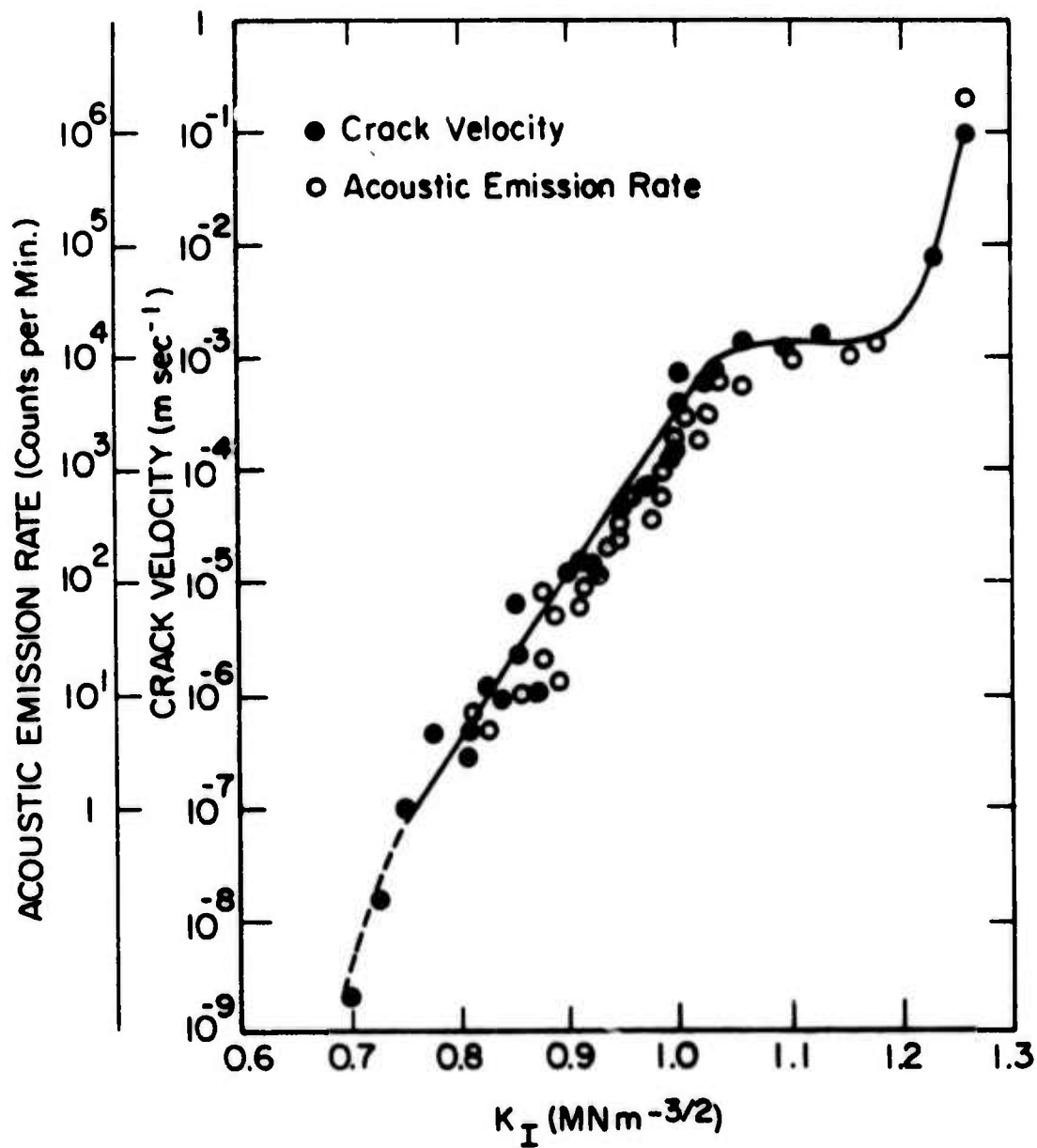
Figure 3. Acoustic emission and crack growth in porcelain as a function of stress intensity factor showing close correlation of emission rate with crack velocity. By permission from Evans and Linzer.

REPORT OF

MEETING ON SCIENTIFIC BARRIERS IN BATTERY SYSTEMS

R. A. Huggins, W. H. Flygare,
E. J. Cairns and C. W. Tobias

I.   Introduction

Battery systems are receiving increasing attention at the present time because of their importance in connection with a number of different types of applications. Among these are electrical vehicle propulsion, large scale power leveling, remote power sources, standby (emergency) power sources, and portable electric devices.

This report will first review some of the basic principles of electrochemical generation and storage of energy by galvanic cells. This will be followed by some brief comments concerning important practical parameters, as well as theoretical limits and realistic expectations for various battery systems. Specific problem areas in different battery systems will then be pointed out, with emphasis upon materials problems and limitations. Lastly, a number of general comments and specific recommendations are presented for basic and applied research into battery materials (electrodes, electrolytes, separators, and containers) that should lead to more rapid technological development in this important area.

II.   General Principles Related to Battery Systems

If we bring in contact two materials of which one is a stronger electron donor than the other (i.e., a stronger re-

ducing agent), the two will react, resulting in the first material giving up electrons (thereby undergoing oxidation), while the second accepts the electrons (thereby undergoing reduction). The energy liberated in association with this spontaneous reaction is, at constant T and P, the heat of reaction, ΔH.

Such reactions may be imagined to occur in two halves, for instance:

overall: $Zn + CuSO_4 = ZnSO_4 + Cu$

halves: $Zn = Zn^{++} + 2e$, $Cu^{++} + 2e = Cu$

overall: $Pb + PbO_2 + 2H_2SO_4 = 2PbSO_4 + 2H_2O$

halves: $Pb + SO_4^{=} = PbSO_4 + 2e$,

$$PbO_2 + H_2SO_4 + 2H^+ + 2e = PbSO_4 + 2H_2O$$

overall: $2H_2 + O_2 = 2H_2O$

halves: $2H_2 = 4H^+ + 4e$, $O_2 + 2H_2O + 4e = 4OH^-$
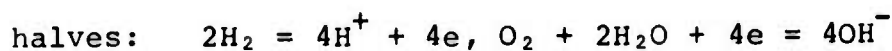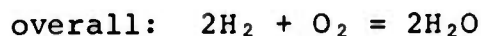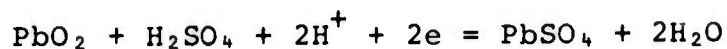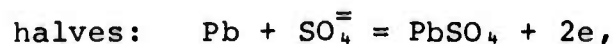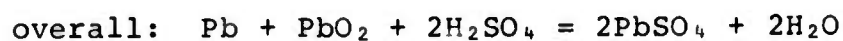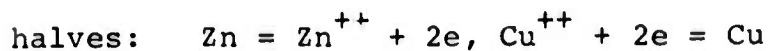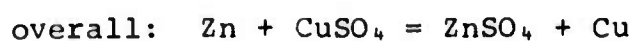
Now if we wish to carry out reactions such as illustrated above so that part of the energy liberated will perform useful electrical work, we have to spatially separate the two half reactions so that the electrons liberated by the donor (e.g., Zn, Pb, H_2, etc.) should not reach the acceptor half (Cu^{++}, O_2, PbO_2, etc.) by direct molecular contact, but rather only through an external load (resistance, motor, etc.). Thus, in principle, the free energy, $\Delta G = \Delta H - T\Delta S$ of the reaction can, as a maximum, be delivered to do external work and the heat

absorbed by the reacting system will only be $T\Delta S$.

To render the above outlined cell scheme feasible, we have to complete the circuit by placing between the two half reactions an ionically conducting medium (the electrolyte). The electrolyte will have to satisfy two key criteria:

(a) It cannot be an electronic conductor, because otherwise the overall reaction would proceed without delivering the useful electrical work in the outer circuit.

(b) It has to be an ionic conductor, which also serves as a suitable source of ions to enable the completion of the two half reactions.

For our examples the electrolytes will be $ZnSO_4$ and $CuSO_4$ solutions (in contact with Zn and Cu, respectively) separated by a porous diaphragm to slow down direct access of $CuSO_4$ to Zn; aqueous H_2SO_4 for the lead acid battery example; and a neutral salt solution (e.g., Na_2SO_4) for the H_2/O_2 cell. The latter can be operated in acid or basic medium as well, with appropriate changes in the half reactions (but not in the overall reaction).

Since the reacting materials or products may themselves be poor electronic conductors for the half reactions to proceed, it is often necessary to provide intimate contact with an "inert" metal which then serves as an electron collector. Thus the electrode may be the reacting material itself (such as in

the case of Cu, Zn, Pb) or an "inert" metal in intimate contact
with the reactants (in case of PbO_2, H_2, O_2). Such a contact
may be best achieved by mixing the metal powder with the
reacting material and pasting it into a suitable "inert" metal
grid. In the case of gaseous reactant, the contact can be
produced by simply bubbling the gas past a metal surface, or
much better, by producing a three phase contact between electro-
lyte, gas and electron collector in the form of a porous electrode.
If properly designed, such a porous electrode forms a gas-
electrolyte interface within the pores, so that neither the gas
nor the electrolyte percolates through.

Ideally in a galvanic cell at rest (open circuit) no
reactions occur, and the energy can be therefore stored for an
indefinite length of time. Upon closing the outer circuit,
current begins to flow, its magnitude being determined by the
thermodynamic driving force of the cell reaction ($\Delta G = -nFE$)
and by the impedance of the outer and inner parts of the circuit.

The internal impedance of the circuit consists of the
ohmic resistance of the electrolyte connecting the electrodes,
R_i, and of overpotential components associated with the rate
processes involved in the electrode reactions. In the ideal
case when these rate processes are very fast, the terminal
voltage E_T is related to the thermodynamic potential (at rest),
E_{Emf} by

$$\frac{E_T}{E_{Emf}} = \frac{R_o}{R_o + R_i}$$

where R_o is the load resistance. Maximum power may be derived from such a device when $R_i = R_o$, in which case half the energy generated is liberated as heat within the cell itself.

The Emf of a cell depends on the strength of the electron donor or acceptor property of the materials involved, characterized by the standard electrode potentials of the half reactions referred to conditions when all reactants and products are at their respective standard states. The dependence of the cell Emf, E, on the activities (a) of the reactants and products for the cell reaction

$$\nu_A A + \nu_B B = \nu_C C + \nu_D D$$

is given by

$$-\Delta G = nFE = RT \ln K - RT \ln \Pi a_i^{\nu_i};$$

$$E = E^\circ - \frac{RT}{nf} \ln \frac{a_C^{\nu_C} a_D^{\nu_D}}{a_A^{\nu_A} a_B^{\nu_B}}$$

where ΔG is the change in free energy for the reaction, n is the number of electrons involved in the oxidation and reduction halves, ν_i is negative for reactants and positive for products. E° is the standard Emf of the cell.

Returning to our cell schemes, it is evident that the Emf of the cell will change if the chemical potentials (concentrations or activities) of the reacting or product species

-583-

change during operation. A stable Emf requires a reaction scheme where only minimal changes in activities will occur upon progress of the cell reaction.

In the selection of half reactions considerations will include high $E°$, compatibility with the other half reaction and with the electrolyte, and stability with respect to the electrolyte itself. Considering that the standard decomposition potential of water is only 1.23 volts at 25°C, the possibilities for designing energetic galvanic cells using aqueous solutions appears in principle quite discouraging. However, because of the large activation energies required for the decomposition of water on some electrodes, in practice it has been possible to operate galvanic cells having much larger Emf's than 1.23 V. The incentive for using half reactions with high E-'s can be easily understood if one considers the role of internal resistance of the cell as it affects power delivered to the outside circuit. If five rather than ten cells need to be connected in series to yield the same terminal voltage at a given load, the internal losses will be half as much with five cells than with ten. On the other hand, high Emf reflects high energy and often high reactivity; spontaneous reaction with the electrolyte is more likely to occur.

The nature of the ionic conductor has a decisive role in determining the limitations of present-day batteries. To begin with, because of the small number of charge carriers and their low mobility, the conductance of electrolytes is low: a moderately concentrated electrolyte has 10^5-10^6 times higher

-584-

resistivity than a typical metal. To minimize ohmic losses it is therefore essential to place electrodes as closely together as possible. However, not only are there mechanical limitations to approaching electrodes, it is also necessary to provide sufficient reacting electrolyte (ions required in the half reactions) if extended use of the cell is contemplated. Alternately, the electrolyte may be continuously replaced by flow between the electrodes; this technique, however, involves a greater mechanical complexity and increases weight and volume.

In electrolytic solutions the current is carried by all ions present, as determined by the relative numbers and mobilities. Thus, the passage of current unavoidably results in causing the formation of concentration gradients. Considering either half reaction, it is evident that the ionic species participating in it (or produced by it) is carrying only part of the current. Since its transference number is less than unity, a deficiency or an excess of the species is created, depending on whether it is consumed or produced by the reaction and on the polarity of the ion relative to that of the electrode. These concentration changes always render the Emf of the reaction lower than at rest and the higher the current density, the larger these concentration differences become. At sufficiently high discharge rates the concentration of a reacting ionic species may reach zero, and the terminal voltage consequently will approach zero as well. Drastic composition changes in batteries

are very undesirable as they may cause a steeply declining voltage upon discharge at constant load (and consequent loss of usable capacity). In general the most favorable cell reactions are those which don't affect the bulk composition of the electrolyte (example, Cd-NiOOH).

Since the direct access of the reactant of one of the half cells to the other half cell results in spontaneous chemical reaction, a loss of energy efficiency and potentially a total loss of the battery can occur if this access is not prevented. Thus battery reactants should be preferably quite insoluble solids (the metals, their oxides, halides), or a porous barrier (membrane) must separate the two electrolyte compartments, which at least can slow down the access of reactants from one compartment to the other.

The employment of solid-solid reactions in battery technology is particularly pronounced in secondary or rechargeable systems. To a large degree the reason for this is that to be able to discharge and charge efficiently and at moderately high rates, it is necessary that mass transfer not be a limiting factor. If in the charge or discharge process the supply of one of the species is limited by its transport from the body of the electrolyte rather than from the immediate vicinity of an electron transfer site on the surface (as in the case of solid-solid reactions), the cell operation will be problematical, and to say the least it will have an unfavorably steep voltage characteristic. The solid, if it is a non-

conductor (such as most oxides, halides) should have such a
state of aggregation on the surface as to allow the passage of
ions to the electron transfer sites. The art of formulating
electrode compositions in which metals, including "inert metals",
are mixed with oxides or halides is highly empirical, one of the
sensitive areas of battery technology. The plate should have
adequate porosity to allow penetration of electrolyte into the
depth of pore channels so that the conversion reaction can pro-
ceed throughout the thickness of the plate. It is, however,
evident that the most favorable reaction site will be on the
external surface of the electrode -- both because it receives
the best supply of ions and also because it is the closest to
the electrode of the opposite polarity. Thus as the charge or
discharge progresses, the reaction penetrates further into the
depth cf the porous electrode. Since this involves a longer
transport path for the reacting ions and a higher ohmic re-
sistance, the output voltage (E_T) of the cell will decrease.
Reacting porous electrodes therefore inherently cause adverse
voltage characteristics. Ideally, an infinitely thin planar
electrode would yield the flattest discharge curve; however,
it would do so at the expense of infinitely small capacity.

The intricate manner in which the reacting masses are
arranged in a porous electrode is largely responsible for our
inability so far to determine the "state of charge" in batteries
which have already been subject to some use or which were simply

allowed to stand and undergo deteriorative processes of some sort or another. The problem is a difficult one because at different load (current density) levels the penetration pattern of current is different and, furthermore, the penetration pattern depends more or less sharply on prior load history.

Thin electrodes, although they have obvious advantages from the point of view of transport, require a disproportionally large fraction per reacting mass of support (electronically conducting inert metal grid). Thus while the energy efficiency and voltage characteristics are improved, the specific energy declines with thinning the electrodes. The capacity remains the same if the same amount of reacting mass is included in the cell (i.e., if we appropriately increase the number of plates). Since for the same load we have decreased the superficial current density, the internal losses (ohmic and overpotential) will be lower than in case of thick plates.

The net reaction describing the gross chemical change associated with the transfer of electrons at the interface gives no indication of the physical nature of the products formed at the surface. As mentioned earlier, in essentially all commonly used batteries solid-solid transformations occur upon charge or discharge. Many of the relevant solid materials are poor electronic conductors. Upon formation of a coherent, non-porous layer the current would cease to flow; or even if such layers are porous, the ohmic drop through the layers may involve severe voltage penalties. The state of aggregation on the micro

scale of reactants and products is then of key importance in the design and operation of batteries.

The mechanism of electrocrystallization, after many years of effort at clarification, is still very much in need of further work. The problem in its most general form may be stated as follows: we need to be able to predict and control the manner in which a new solid phase is formed. The micro-morphology, nucleation and crystal growth of relevant metals, oxides and halides requires a far deeper understanding than we possess today. Many otherwise attractive electrode re-actions had to be discarded as unsuitable for battery appli-cations because of problems related to crystallization, high apparent impedances associated with the formation of dielectric layers, high contact resistances, etc.

Because the rates of charge transfer and transport processes increase with increasing temperature, significant advantages are associated in principle with high-temperature operation of galvanic cells. Developmental work on the Broers-Ketelaar fuel cell, the $Li-Cl_2$ cell, and the Na(Li)-S recharge-able battery indicates that electrode reactions which exhibit high charge transfer overpotential at ambient temperatures can be operated at high current density with very low overpotentials at elevated temperature and that the ohmic losses across the cell are very much smaller than in room temperature operation.

Medium or high temperature cells make it possible to employ solid electrolytes rather than molten single or multi-

component melts. The advantages of solid electrolytes include the following:

(a) Only one of the reacting ionic species carries the current and therefore there are no concentration gradients generated by the flow of charge.

(b) Solid electrolytes have certain structural, configurational advantages; they may serve to separate two liquid reactants (electrodes).

(c) Ionic species which cannot exist in aqueous systems can be the current-carrying species. Of particular interest in this regard are solids in which either $O^=$ ions or alkali metal ions are mobile.

Although there are numerous advantages associated with high temperature galvanic cells, the difficulties involved in developing economically feasible hardware are just as numerous. To begin with, presently known solid electrolytes which promise to have suitable mechanical and chemical properties are small in number and require further intensive R&D effort. The development of metallic and dielectric materials suitable for cell components will require more ingenuity and more time than have the corresponding problems in low temperature cell applications.

During the past ten years or so a class of solid ionic conductors was discovered which demonstrate at room temperature adequate conductance to allow their employment in primary and secondary batteries. Batteries employing alkali metal-silver

iodide (e.g., RbAg$_4$I$_5$) and β-alumina are undergoing increasing development. The inherent advantages of a thin solid ionic conductor attracted much attention, especially since exploratory research has indicated that other solid materials (e.g., those with tunnel-type structures) may also demonstrate favorable conductive properties.

Because of the inherent limitations of aqueous electrolytes, various organic and inorganic solvents have been sought to replace water in battery applications. It has been conclusively demonstrated that a number of common organic and inorganic solvents are suitable for reversible charge-discharge of alkali metal electrodes. Unfortunately, the development of energetic materials (oxidizing agents) compatible with these solvents has been less successful.

The problems most commonly encountered with non-aqueous solvents include their relatively low solvent power for ionic materials, high viscosity, hence low conductance, and attack by direct oxidation or reduction. It is very likely that practical hardware can be developed which will deliver high (100 wh/lb) energy density, but it is questionable whether high power density can also be achieved.

III.   Important Parameters and Properties

There are a number of important parameters and properties relating to the use of battery systems in practical applications. These can be listed as follows:

1. Basic system parameters

   a. Specific energy (watt hours per kilogram)

   b. Specific power (watts per kilogram)

   c. Energy density (watt hours per cm^3)

   d. Power density (watts per cm^3)

   e. Capacity density (ampere hours per cm^2 electrode

2. Other important parameters are

   a. Rechargeability

   b. Energy efficiency

   c. Voltage efficiency

   d. Operating temperature range

   e. Shelf life (storage life)

   f. Mechanical ruggedness

   g. State-of-charge indication

   h. Price

   i. Safety

   j. Large scale availability

One can draw some general conclusions regarding targets for the values of the basic system parameters as well as the price necessary in order to compete with other types of energy storage for certain types of applications. These are shown in Table I.

The ultimate limits that might be achieved in various different types of electrochemical systems depend upon the

parameter under consideration. It is relatively easy to determine the maximum theoretical values of specific energy that might be obtained from particular electrochemical systems on the basis of thermodynamic data, where available, coupled with information concerning the specific weight of appropriate species. Such calculations have been made for a number of common systems and are listed in Table II. In such calculations it is assumed that all of the weight is assigned to reactants alone. These theoretical maximum values of specific energy are plotted versus the equivalent weight of the reactants in Fig. 1.

Of course, in practical systems additional weight is utilized in other important components of any battery system, such as the electrolyte, the container, current collector and leads, structural members, etc. It is generally found that the practical limit is between 20 percent and 30 percent of the maximum theoretical value in most designs. One can estimate the values which appear to be practically feasible in some of the present battery systems. Projections concerning practical values of specific energy and specific power for various proposed electrochemical systems are plotted in Fig. 2.

IV.   Preliminary Overview of Problem Areas in Battery and Fuel Cell Systems

The following pages contain some comments relating to a number of different materials-related aspects of battery and fuel cell systems. The authors recognize that this discussion

is certainly _not_ complete, and that it may even omit some important items.

In addition, it should be emphasized that it represents only a hastily taken snapshot of the current state of this rapidly moving field at a particular time, seen through an admittedly imperfect lens, quite limited in both resolution and angle of field, represented by the judgment of a limited number of people, some with restricted experience in certain aspects of this field of technology.

Furthermore, it should be pointed out that there has been no effort to put these items in order of priority with respect to either magnitude or importance. Some problems mentioned are obviously more readily solved (or avoided) than others, and there is a wide variation in the degree of scientific sophistication necessary for technological progress.

A. Ambient temperature aqueous acid electrolyte systems.

1. Electrode problems

a. The lead negative electrode of a $Pb-PbO_2$ cell can experience sulfation, which involves the development of a blocking layer of lead sulphate upon the surface. Also, under rapid charging conditions hydrogen evolution occurs.

b. The positive electrode (PbO_2) in a $Pb-PbO_2$ cell experiences loss of the active material from the plate which can result in both the loss of energy storage

-594-

capacity and the possibility of electrically shorting between plates by contact with material that has fallen to the bottom of the cell. Problems here relate to adhesion and cohesion of microstructural constituents during cyclic operation. Dynamic changes in the micro-structure are related to the phenomena involved in the solution and deposition of solid phases. Corrosion of the grid material (typically lead alloys containing antimony, calcium, etc.) is also a problem.

c. The air or oxygen positive electrode in acid aqueous electrolyte systems involves the use of electro-catalytic materials and suffers problems with recharge-ability. This three-phase system evidently becomes "degraded" for the forward (discharge) reaction by phenomena which occur during recharging.

d. The chlorine positive electrode in acid aqueous electrolyte systems suffers difficulty with recharge-ability. In addition, of course, there are problems of handling and storage of the chlorine reactant.

2. Electrolyte problems

Sulfuric acid does not have any specific prob-lems when used in the lead acid battery. However, acid-electrolyte hydrocarbon fuel cells have been shown to operate best when "superacids" are employed. Further investigation into the possible electrolyte uses of

these acids is desireable.

3. Problems with static component materials (containers, seals, current collectors, etc.)

    a. Interesting superacids cause corrosion problems, and wetting problems with PTFE.

    b. Porous PTFE separators which are currently used in acid systems are expensive.

B. Ambient temperature aqueous alkaline electrolyte systems

    1. Electrode problems

    a. Zinc negative electrodes typically have problems with the formation of dendrites as well as changes in macroscopic shape.

    b. Magnesium negative electrodes have problems because of passivation.

    c. Nickel oxide positive electrodes have problems with charge retention under static conditions, and suffer a decrease in rechargeability above room temperature.

    d. The air (oxygen) positive electrode and hydrogen negative electrodes require the use of electrocatalysts. A platinum replacement would be very important here, especially for use in acidic electrolytes. Even with platinum, there is difficulty with cycling of the oxygen electrode due to some type of irreversible changes which occur during cycling.

2. Electrolyte problems

There are no pressing problems with alkaline aqueous electrolytes, aside from those related to carbonate accumulation during operation with air.

3. Static component materials problems

Difficulty is experienced with separators and seals in systems of this type.

C. Ambient temperature non-aqueous electrolyte systems

1. Electrode problems

a. The negative electrode is typically lithium in these systems and problems of passivation and dendrite formation are evident.

b. Positive electrode constituents in non-aqueous electrolyte systems include metal halides and sulfides. Problems are experienced with solubility of these electrode constituents in the electrolyte and in re-chargeability.

c. Inorganic constituents which react in a positive electrode system may include SO_2, $SOCl_2$, $POCl_2$ and others. Insufficient experience is at hand to evaluate problems in such systems.

2. Electrolyte problems

This area is currently quite undeveloped. There are a number of organic-solvent electrolytes which have been investigated, and several important inorganic con-

stituents of the electrolyte-reactant solution have become publicly visible recently (these include SO_2, $SOCl_2$, and $POCl_3$). This appears to be a fertile area for further study. Appropriate materials must have proper values of conductivity, viscosity, liquidus range, dielectric constant, solubility and dissociation parameters, and be compatible with both electrodes and reactants.

3. Static component materials problems

Materials for seals and separators present problems in both organic and inorganic ambient temperature non-aqueous liquid electrolyte battery systems.

D. Elevated temperature solid electrolyte systems

The sodium-sulphur cell using a solid electrolyte of the beta alumina family is the primary present example of this class.

1. Electrode problems

a. The liquid sodium negative electrode presents no problems directly, although there are occasional difficulties with the wetting of the beta alumina interface by sodium. This can cause a substantial reduction in the active electrode area with resultant problems, such as hot spots and reduction in power.

b. The positive (sulphur) electrode presents a number of problems at high current densities and deep discharge conditions because the electrode consists of

-598-

several Na-S phases of varying composition (see the characteristics of the sodium-sulphur phase diagram). This situation is complicated by the necessity of having a current collector (typically graphite felt) present at the interface as well. Very little is known about the actual microstructure at the electrolyte-current collector interface during charge and discharge in this system, but it is suspected that a high sodium content solid phase is present. Voids can also be present in this viscous liquid-solid polyphase system. Much is yet to be learned in order to optimize the behavior near deep discharge and during rapid recharging.

One can categorize the required information into matters which are geometry and rate-insensitive on the one hand, and those which depend upon the local geometry and current density on the other. The first group include thermodynamic relationships between species and phases, kinetic phenomena within phases as well as at phase boundaries, and the properties of the pertinent phases and species that are present. The geometry and rate-sensitive problems include the spatial distribution of constituents as well as the species present within the various phases under dynamic conditions. In addition, the driving forces and transport behavior during both discharge and charge cycles need to be better under-

stood in order to optimize the behavior of this important electrode system.

2. Electrolyte problems

Enough is currently known about materials of the beta alumina family to identify some of the problems that require further understanding in connection with this solid electrolyte material.

a. Cracking occurs on the sodium electrode side during recharging with high current densities. The effect of chemical composition of the electrolyte as well as various microstructural effects and surface morphology are not fully understood.

b. On the sulphur electrode side there is a substantial disparity in results reported by different groups working is this area. Some indicate that they have no problems with the sulphur side of the solid electrolyte, whereas others have experienced substantial amounts of damage (micro fractures and flaking). The reasons for this are not understood at the present time.

c. The influence of composition (both major constituents and impurities) and microstructural variables upon the ionic and electronic conductivities need to be studied in order to understand these important properties and to improve this characteristic for battery use.

d.   There are indications that impurities acquired during operation of cells of this type may have deleterious effects. These include silicon, potassium, iron, etc.

e.   It is quite obvious that there is much to be done to both understand and improve the preparation and processing techniques for the production of satisfactory beta alumina shapes. In order for cells of this type to become a commercial reality, production technology must be developed to a much higher degree than is characteristic of the ceramic industry.

f.   It is important to search for other materials which might compete with beta alumina. Of special interest would be solid electrolytes which are good conductors for lithium ions.

3.   Static component materials

a.   Containment materials for liquid sodium do not present a problem.

b.   On the other hand, containment of liquid sulfur presents a serious corrosion problem if one chooses to use metallic materials.

c.   The seals, particularly those in contact with the sulfur side of the cell, present a serious problem. Also, difficulties have been encountered in joining beta alumina to other ceramics (such as alpha alumina) or metals.

-601-

d.   The current collector in the positive (sulfur) electrode is presently carbon felt. This material is too expensive (about \$1000/lb) at the present time. It also is fragile.

E.   Elevated temperature liquid-electrolyte systems.

These systems presently utilize lithium as the negative electrode reactant and alkali halide electrolytes.

1.   Electrode problems

a.   Retention of liquid metal in the porous current collectors presently used in this type of cell is not sufficiently reliable.

b.   Solid lithium-aluminum alloys have also been used for the negative electrode, but they offer fabrication problems because of their mechanical properties and have a reduced current density capability to a limited diffusion coefficient for the lithium in the alloy. It would be desirable to find other lithium-containing alloys in which the diffusion coefficient is higher so that depletion does not occur at the surface as readily. Shape change can also be a problem in these materials.

c.   The sulfur positive electrode suffers because of the "solubility" of sulfur in the liquid electrolytes. Also, the sulfur electrode has the same difficulties in liquid electrolyte systems as were mentioned above in connection with solid electrolyte cells.

d.   The selenium positive electrode has similar prob-

-602-

lems to those of the sulfur electrode, although they are not quite so severe because selenium has a significant electronic conductivity.

e. If chlorine is used as a positive electrode constituent in molten electrolyte systems it suffers difficulties in rechargeability. The carbon current collector undergoes wetting by the electrolyte, causing flooding of the electrode.

2. Electrolyte problems

a. One of the difficulties with the alkali halide electrolytes is the solubility of the alkali metal in the alkali halide. This is lowest (best) in the case of lithium in lithium halides, whereas sodium and potassium halides exhibit deviations from stoichiometry in the metal-rich direction which lead to electronic conductivity and hence self-discharge problems. It would be desirable to find molten salt electrolytes with considerably lower melting points to alleviate some of these problems.

b. Self-discharge or capacity reduction can also occur in molten alkali halide electrolyte systems due to solubility and transport of chalcogens (sulfur, selenium) or halogens (Cl_2, etc.).

c. These electrolyte materials sometimes flood materials used as current collectors (e.g., carbon) thereby inhibiting the positive electrode reaction by reducing the extent of the reaction zone.

3. Static component materials

a. There is no serious problem in finding containment materials for liquid lithium but some problems are experienced in the shorting of insulating materials by Li.

b. The problem of the containment of sodium in these systems is similar to that mentioned for the solid electrolyte systems above.

c. Separators present a serious problem in this type of cell. What is needed is a highly porous insulator with high strength that is corrosion resistant in the electrolyte. This separator material should be in good contact with (wetted by) the electrolyte.

V. General Comments

1. Lithium is generally chosen as the preferred alkali metal for use in the negative electrode in ambient temperature cells because:

a. It is the easiest alkali metal to handle. Some operations can actually be done in air because of the formation of a passive protective surface layer.

b. Its melting point (180°C) is high enough that the heat of reaction (hydration) in contact with water is not sufficient to cause it to melt under normal conditions. As a result, new surface contact is not continually generated as in the case of sodium and, particularly potassium.

-604-

c. It is the most electropositive element avail-
able in solid form.

d. It has the lowest equivalent weight of electro-
positive solids.

e. It is readily available in a variety of forms,
including sheets already pressed onto mesh substrates.

2. Lithium is typically chosen for molten alkali halide
electrolyte systems because it is relatively insoluble in
its own halide salts. The other alkali metals are more
soluble and the requirements of electroneutrality cause
the presence of excess alkali metal ions to be compensated
by "free" electrons leading to self-discharge due to
electronic conduction.

3. Sodium is the best alkali metal to use with beta
alumina solid electrolytes because it has been found to have
the largest value of ionic conductivity in the beta alumina
structure.

4. Li and Na are the current candidates for negative
electrode reactants of importance in elevated temperature
systems. Other candidates include aluminum, magnesium,
beryllium, calcium, zinc, and potassium.

5. Dendrite formation is not an important problem in
the lead acid battery system, but does become an important
consideration in some other cases, particularly systems
involving alkaline electrolytes and zinc.

6. The cadmium electrode is generally not a problem

from a performance standpoint. However, its cost and limited availability must be recognized.

7. It is now too early to define a complete set of problems in non-aqueous inorganic electrolyte systems.

VI. Recommended Areas of Research into Materials Problems in Batteries

The following is a list of some materials research areas which we feel deserve support in order to increase the rate of technological development of battery systems. This list is not complete, nor is it in order of priority. Much of the discussion was concentrated on high specific energy and high specific power systems. Recommendations for research studies will be emphasized in that direction.

1. Intensive research should take place on the solid state properties of solid materials which show high alkali metal ion conductivity. Currently β-alumina is a good solid electrolyte only for Na$^+$ ions. It is evident that Li/S, Na/S, and perhaps K/S systems have high promise for high energy and power density batteries. More work is needed along the following lines:

a. General study of diffusion and conductivity in solid electrolytes as a function of temperature, microstructure and impurity content is important. In particular, search for an effective solid electrolyte

for Li^+ ion diffusion is very important. New solids
and perhaps glasses should be examined for alkali metal
ion conductivity, and stability toward alkali metals.

b. Studies of the liquid polysulfide and liquid
alkali metal-β-alumina interfaces. These studies should
include the phenomena and characterization of wetting
of the liquid-solid interface as a function of impurity
level in both the liquid and the solid β-alumina.

c. The influence of grain boundaries, microstructure,
and glassy state (amorphous) character on cation con-
ductivity should have more attention.

d. Development of processing techniques for ceramics
(e.g., β-alumina) to produce controlled purity, compo-
sition, and microstructure are likely to yield high
dividends.

e. Replacement materials for the carbon felt
current collector in the positive electrode in Li/S
and Na/S systems would be desirable.

2. High-temperature batteries using molten salt electro-
lytes need work in:

a. Finding new negative electrodes such as solid
Li alloys with a high Li diffusion rate.

b. Finding new electrolytes which have lower
melting points and are inert to alkali metals.

c. Identifying insulating materials stable in the
presence of alkali metals for use in seals, feedthroughs,

-607-

separators, and insulators.

    d.   Identifying (metallic) materials of construction
stable toward alkali metal-sulfur mixtures in the presence
of alkali halides.

3.   In non-aqueous Li battery systems work is needed in:

    a.   Studies of new inorganic electrolytes.

    b.   Examination of the passive layer on Li and
how it can be avoided.

4.   Other battery systems need work in:

    a.   Dendrite and shape-change avoidance

    b.   Charge retention at the nickel oxide electrode in
alkaline electrolyte systems. Basic study of the electro-
chemical phenomena at the nickel oxide electrode is desirable.

    c.   Passivation of the Mg electrode is a problem.
In general, the nature of the passivation of highly
electropositive solids is worthwhile to study.

    d.   Flooding of the positive electrode by the
electrolyte in secondary gas-electrode systems is a problem
that should receive attention. Perhaps some protective
membrane type structure can he used to prevent the flooding.

    e.   The general problem of electrocatalysts (plati-
num presently being the most effective) deserves some
fundamental study in the expectation that understanding
of the unusual behavior of Pt will lead to the develop-
ment of less expensive substitutes.

f. Superacid type organic acids should be examined for activity in fuel cells of the direct hydrocarbon, and other direct types using organic fuels.

4. Other general problems

a. The mechanisms involved in the tendency for flooding of porous gas-electrodes upon aging, and measures for avoiding this phenomenon represent an important problem area requiring attention.

b. The deployment of catalysts (spatial distribution, state of aggregation, etc.), especially of noble metals, in such a manner as to maximize their effect on lowering overpotentials while reducing the amounts of catalyst required deserves further attention. It is by now well understood that the charge transfer processes in gas electrodes occur only at- or very near the electrolyte-gas-electronic conductor interface, and that therefore catalyst materials located elsewhere perform no useful function.

VII. Concluding Remarks

It is very important to recognize that we have undertaken a very precarious task, and that this discussion of problem areas and research objectives is presented only as an introductory working paper or point of departure. There is a high probability that it contains errors and omissions because of both the limited time available for its compilation and the restricted horizons

and biased judgment of those involved.

In addition to this general disclaimer, the great importance of the rate of progress in this field should be emphasized. Some old problems become solved, others are avoided, and new ones become more critical with the passage of time. This is particularly true when major new concepts or approaches arise. As an example, the emergence of the potential use of cationically-conducting solid electrolytes with liquid electrode systems has led to interest in systems in which many of the important problems and limitations related to solid electrodes are no longer even relevant. Instead, a whole new set of pressure points has come into view. Of equal, if not greater, importance than many of the presently visible items mentioned here is the continual need for the exploration of new ideas, concepts, and systems by which energy can be stored and converted electrochemically.

The authors hope that this report will focus attention upon the large number and variety of problems involving materials behavior in the electrochemical energy conversion and storage area. Serious, sustained efforts in some of the nation's leading materials laboratories are necessary to assure good progress in this field.

Acknowledgement

-610-

APPENDICES

1. Meeting Program

2. List of Attendees

PROGRAM OF MEETING ON

SCIENTIFIC BARRIERS IN BATTERY SYSTEMS

General Chairman: W. H. Flygare

Discussion Leader: R. A. Huggins

Monday, July 16

9:00 - 12:00   Dr. C. Martin Stickley, Director for
Materials Sciences, ARPA; General
Questions

Dr. C. Tobias, University of California,
Berkeley; "Introduction to Electro-
Chemical Systems"

1:00 - 3:30   Dr. George Kugler, Electric Storate
Battery, Inc.; "Electric Vehicle Power"

Dr. Elton Cairns, General Motors;
"High Specific Energy Batteries"

Tuesday, July 17

9:00 - 12:00   Dr. Boone Owens, Gould, Inc.;
"Solid Electrolyte Batteries

Dr. C. C. Liang, P. R. Mallory, Inc.;
"Solid State Batteries"

Dr. Morris Eisenberg, Electrochemica;
"High Energy Organic-Lithium Batteries"

1:00 - 3:30   General Discussion

SCIENTIFIC BARRIERS IN BATTERY SYSTEMS

Attendance List

Materials Research Council

Name	Affiliation
H. Ehrenreich	Harvard University
W. H. Flygare	University of Illinois
J. J. Gilman	Allied Chemical Corporation
R. Gomer	University of Chicago
F. Hawthorne	University of California, Los Angeles
E. E. Hucke	University of Michigan
R. A. Huggins	Stanford University
W. Kohn	University of California, San Diego
J. L. Margrave	Rice University
H. Reiss	University of California, Los Angeles
C. M. Stickley	Advanced Research Projects Agency
E. C. van Reuth	Advanced Research Projects Agency

Visitors

E. J. Cairns	General Motors Research Laboratories
M. Eisenberg	Electrochemica Corporation
N. Elsner	Gulf General Atomic
J. R. Huff	USAMERDC
G. Kugler	ESB, Inc.
C. Liang	P. R. Mallory and Co.
R. Minck	Ford Scientific Laboratory
S. Mitoff	General Electric Company
B. Owens	Gould, Inc.
L. J. Rogers	USAECOM
S. Ruby	Advanced Research Projects Agency
S. Schuldiner	Naval Research Laboratory
S. Pzpak	Naval Electronics Laboratory Center
C. Tobias	University of California, Berkeley

TABLE I

Approximate Targets for Several Types of Applications

Application	Specific Energy (w-hrs/lb)	Specific Power (w/lb)	Capacity Density* (amp-hrs/cm²)	Active Life	Price
Large Vehicle Propulsion	100	100	>0.3	3 yr.	<$50/kw-hr
Urban Vehicle Propulsion	50	50	>0.3	3 yr.	<$50/kw-hr
Large Scale Load Leveling	200		>0.3	5 yr.	$15/kw-hr
Heart Pacemaker		<0.01	>0.3	10 yr.	$200/kw $200/battery
Military timers, fuses, etc.	50	variable	various	various	**

* Assuming parallel-plate construction
**Broad range, depending upon application

-614-

TABLE II

Maximum Theoretical Specific Energy for Selected Electrochemical Systems

System	Maximum Theoretical Specific Energy
$Li/CuCl_2$	503 watt-hr/lb
Ag/Zn	208
Pb-acid	78
Ni/Cd	107
Ag/Cd	120
Ag/Zn	208
Zn/air	671
Li/CuF_2	750
Li/AgCl	300
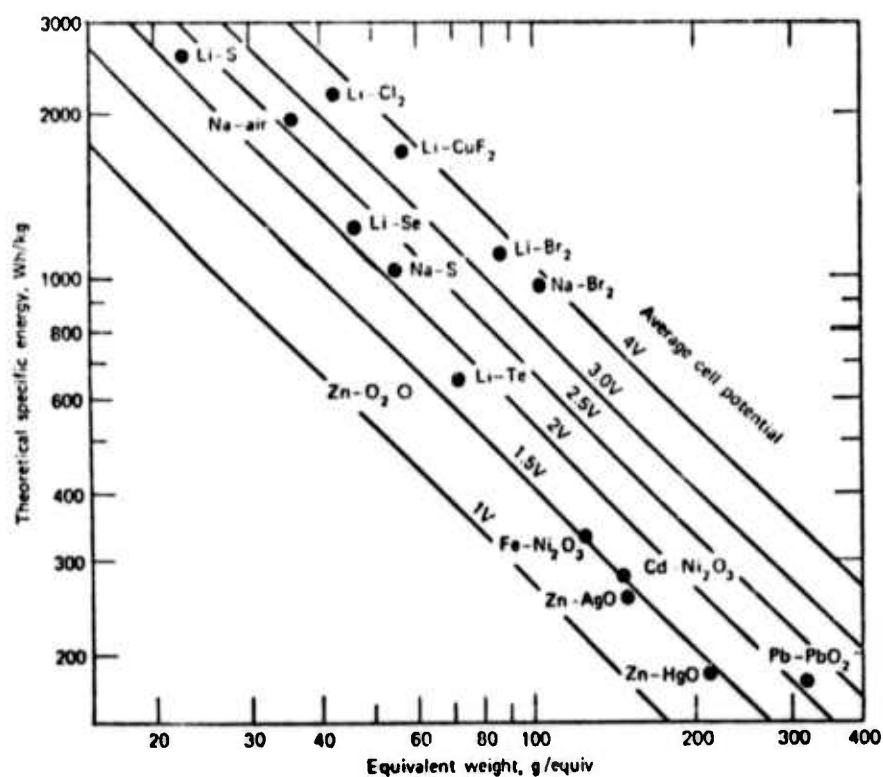
Figure 1. Effect of the difference of electronegativity
between anode and cathode (as reflected by the
cell voltage) and of the equivalent weight (of
anode plus cathode materials) on the maximum
theoretical specific energy of various electro-
chemical cells. (Cairns)
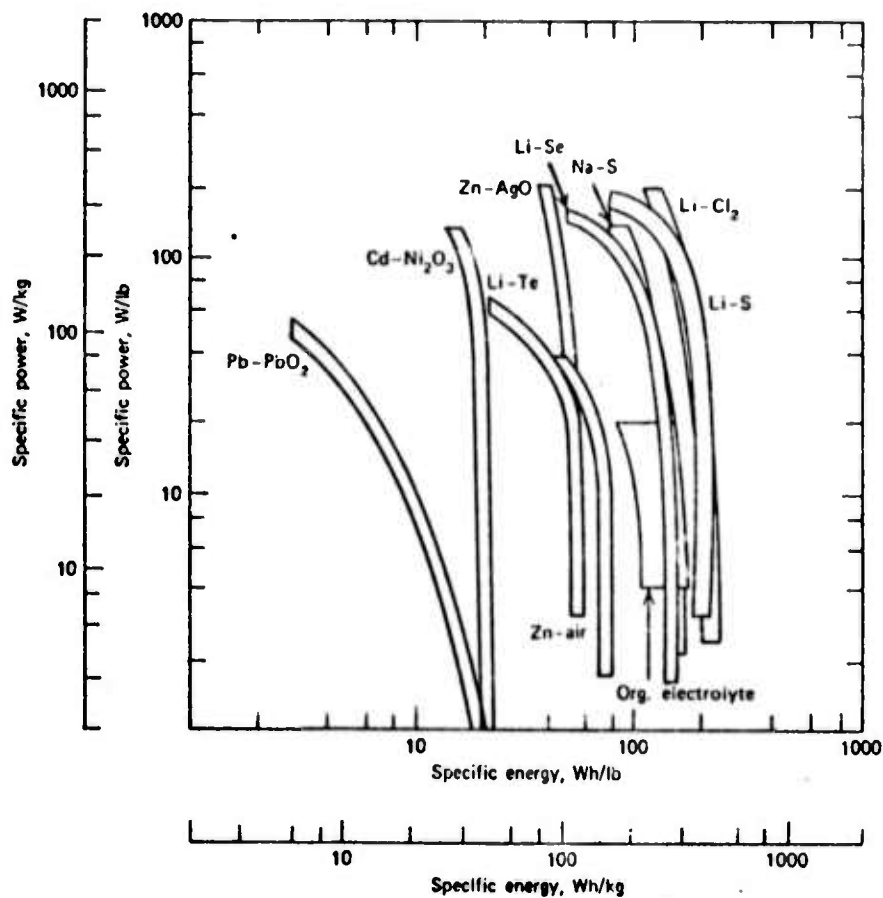
Figure 2.   Estimated specific power–specific energy curves
for high-temperature batteries, compared to those
for conventional batteries.   (Cairns)

MATERIALS PROBLEMS RELATED TO TRANSMISSION
AND REFLECTION OF UV RADIATION

M. Sparks
H. Ehrenreich
W. H. Flygare
C. M. Stickley

Approximately 15 outside users and developers of UV lasers and materials met with several Materials Research Council members to examine the nature of materials limitations as related to high power UV generation and transmission.

The general frequency region of interest was the vacuum ultraviolet. The xenon laser, with wavelengths near 1720 A, has received the greatest attention to date.[1] The currently available power is approximately one millijoule in ~10 nsec pulses. This power level must be viewed in the context that laser fusion requires 10^4 joules and exoatmospheric applications require 10^5 to 10^6 joules.

The problem of obtaining acceptable materials for windows and mirrors for use at high intensities is more difficult in the vacuum ultraviolet region than in the infrared region. The latter has been the subject of numerous investigations in the last three years, while the former has received little attention.

The schedule of formal presentation attempted to cover four general areas:

Preceding page blank

(1) Fundamental Principles of Reflection and Transmission of UV Radiation; H. Ehrenreich, F. C. Brown, J. Stafford and A. Glass.

(2) Current Materials Problems in UV Lasers; P. Hoff.

(3) Materials Selection, Preparation, Polishing and Characterization of Bulk Materials and Surfaces; H. Bennet, W. Plummer, W. Hunter, V. Rehn, E. Catalano.

(4) Methods of Generation of Fundamental Optical Constants in Materials in the UV and Soft X-ray Regions; F. Brown, W. Spicer, H. Ehrenreich.

In addition to the above, R. Andrews discussed the prospects of X-ray lasers but the associated materials problems were not discussed extensively.

The principal points that emerged from the discussion which followed the conference and which was based on information supplied by the speakers, focused on (A) the fundamental limitations of the materials used for UV mirrors and windows and (B) recommendations for future research.

A.   Fundamental Limitations

1. Al appears to be the mirror material that comes closest to having ideal properties. Its reflectivity cut-off lies beyond 15 eV; still, in pure Al above about 1 eV, 8% of the incident photon flux is absorbed. The interband mechanism responsible for this absorption is also present in other polyvalent metals that might be alternate candidates.

-620-

Theoretical estimates indicate that the value of the absorbance A = 8% is intrinsic.[2] Alloys and elements other than aluminum might be considered. It seems unlikely, however, that a material better than aluminum will be found; alloying tends to increase the absorption since it decreases the electron relaxation time, τ. In addition, the simple polyvalent metals, which have a sufficiently high plasma frequency that the reflectivity cut-off occurs at sufficiently small wavelengths, have weak interband absorptions in the infrared.[3] These increase the absorption to unacceptable levels.

The achievement of 92% reflectivity in pure Al to 7 eV is possible only in the absence of any surface oxide contamination. Overcoating with a dielectric material such as MgF_2 or LiF prevents oxidation, but may alter the optical properties in the UV to some extent. This value can be maintained to about 10 eV if the surface is ultrasmooth (roughness $\sim$10 A), since elimination of roughness prevents absorption by surface plasmons (see below).

Other possible methods of protecting aluminum and other metals from oxidation include the following: Thin films of materials, such as Al_2O_3, BeO, and SiO_2, in addition to MgF_2 and LiF, could be tried. Since thin films, say $\sim$100 A thick, should suffice, materials having values of absorption coefficient, α greater than the values required for windows could be considered thus greatly increasing the number of candidate coating materials. For a 100 A-thick film with the large value of $\alpha = 10^3$ cm^{-1}, 0.1 percent of the radiation is absorbed. Sputtering is a promising

method of deposition. Using a Knudsen cell for deposition was also suggested.

Since thick samples of Al_2O_3 readily transmit radiation down to $\sim$1500-1600 A, it is initially surprising that oxidation drastically reduces the reflectivity of aluminum for $\lambda \sim$1000-2000 A, especially since the oxide films are believed to be only a few tens of angstroms thick. For a 20 A-thick film, the intensity is reduced by 1/e for $\alpha = 5 \times 10^6$ cm^{-1}, a value greater than that of aluminum at any wavelength. This dramatic effect may be caused by increased reflectivty in the oxide surface of the radiation which has been reflected from the metal surface thereby trapping the radiation. If smooth single crystal aluminum films can be grown and carefully oxidized, they may be stable and free from the reduction in reflectivity. Passivating surface reactions, similar to a mono-layer of germanium on tungsten, possibly could render aluminum free from the oxidation effect. Possible reactants include: F_2, Si, Ni, Au, Cu, and Ge. Additional comments on surfaces and materials are given by J. J. Gilman as an appendix to this report.

A broad absorption band extending from $\sim$1300 to $\sim$1800 A has been attributed to excitation of surface plasmons. The electromagnetic field will not excite surface plasmons on a perfect surface. It is observed that the strength of the absorption band indeed decreases as the surface roughness σ is is decreased, and becomes unobservable at $\sigma \simeq 10$ A. Studies of

scattered light and photoemission have been useful in studying surface roughness. The surface-plasma wavelength is $\lambda_{sp} = (1+n^2)^{\frac{1}{2}}\lambda_p$, where n is the index of refraction of the overcoating material and the bulk plasma wavelength $\lambda_p = 814$ A for aluminum. For n = 1.4, $\lambda_{sp} = 1400$ A, in fair agreement with peak near 1550 A. The possibility of shifting the surface-plasmon frequency by using films thinner than the surface-plasmon penetration depth could be considered if the residual absorption in smooth films becomes important.

Mirrors other than normal-incidence metallic reflectors may also be considered. The reflectivities of dense materials at near-grazing angles are large. For example, 90% reflectance at an angle of 1° has been obtained even at 67 A, where the reflectivity at normal incidence is near zero. The small loss and beam degradation at each of the many small-angle reflections required to redirect a beam through a large angle, of say 180°, would exclude this method in many applications. Nevertheless, it should be noted that the threshold for irreversible damage is lower than that of a single, normal incidence reflector, since the absorption by each mirror is small.

Internal reflectors, such as corner cubes, and multi-layer structures of dielectrics and/or metals may be effective, but a tolerance of $\lambda/20 \simeq 100$ A is required in some applications. In structures with metal-to-metal interfaces, diffusion that could spoil the properties of the structure should be considered.

Some materials have n < 1 in the ultraviolet region. Such materials conceivably could be used as total external reflectors.

2. Materials for use as high-power windows should have a large bandgap, E_g. The value of E_g must be greater than $\hbar\omega$, where ω is the ultraviolet frequency, in order to avoid direct absorption of the photons across the gap. All insulators have gaps smaller than about 13 eV. This limitation is imposed by the fact that the spacing of adjacent atomic energy levels are usually of the order of, or smaller than, 1 Rydberg. LiF and MgF_2 appear to be prime candidates since their band gaps are larger than 10 eV. The number of materials is limited, the band gap sequence being LiF > BeO > NaF > MgF_2 > Al_2O_3 > SiO_2.

3. The extent of non-linear absorption mechanisms in the UV is only incompletely understood at the present time. Present extimates of the power density limits for both windows (in the two photon absorption region) and mirrors pertaining to Xe lasers are $\sim$1 megawatt/cm^2.

4. The role of imperfections is also incompletely understood. The failure mechanisms in laser glasses considered by Bloembergen are applicable here.[4] In addition, the formation of such imperfections, as well as color centers (in UV windows) with time must be considered.

5. The mechanical properties of windows must be such that they can withstand appreciable pressure differentials and shock waves (30-40 atm). Unstrengthened LiF may be poor for that reason.

-624-

B.   Recommendations

1.   The non-linear optical constants should be measured
from the infrared through the UV region.   These would include:
(a) $\chi^{(2)}(\omega)$ in crystals lacking inversion symmetry; in this case
a 2ω photon is generated and absorbed by a single step mechanism;
(b) $\chi^{(3)}(\omega)$ which involves two photon absorptions, simulated
Raman and Brillonin scattering, as well as the non-linear index
of refraction.   This might be accomplished by using SLAC UV
photons together with a coherent source (say, a ruby laser).

2.   Lower energy UV lasers ($\hbar\omega$ = 3-4 eV) should be seriously
considered since windows can be chosen for which the dissipative
part of χ^{NL} vanishes to third order.   It is not clear whether
the mirror problem would be resolved by this expedient although
a phased array of insulating partial reflectors (like LiF)
followed by an Al-MgF$_2$ mirror could be visualized.

3.   UV mirror technology should be developed along the
following lines: (a) use of sputtering techniques to make curved
self-supported mirrors;   (b) use of combined mechanical polishing
and chemical etching, the former to produce correct curvature,
the latter to remove surface damage;   (c) the development of
other techniques such as ion polishing.

4.   High purity LiF and MgF$_2$ should be prepared and their
optical and mechanical properties, as well as resistance to
radiation damage and tendency to form color centers should be
investigated more thoroughly than heretofore.   Also, the con-
trolled addition of impurities such as Mg to harden LiF (pure

-625-

LiF is very soft and therefore difficult to polish) should be investigated.

5. The use of SLAC for some of the foregoing experiments should be investigated. At $\hbar\omega = 10$ eV, there are $\sim 10^{13}$ photons/Angstrom-mr-sec. For $\Delta\lambda \sim 10^2$ A in 10 mr this would lead to 10^{16} photons/sec and an intensity of 10^5 watts/cm^2 which is only an order of magnitude smaller than the presently achievable limit of 10^6 watt/cm^2 (this is strictly a materials limit). It is estimated that a factor of 10 scale up would be possible with the presently available system, were it not for this limitation.

Acknowledgement

References

1. C. K. Rhodes, "Review of Ultraviolet Lasers," _IEEE Journal of Quantum Electronics_, in press.

2. H. Ehrenreich, H. R. Philipp and B. Segall, Phys. Rev. 132, 1918 (1963).

3. W. A. Harrison, Phys. Rev. 147, 467 (1966).

4. N. Bloembergen, "The Role of Cracks, Pores and Absorbing Inclusions on Lasers Induced Damage Threshold at Surfaces of Transparent Dielectrics," NRC Report, 1972.

APPENDIX A

The following is a list of participants in the ARPA Materials Research Council summer meeting on Materials Problems Related to Reflection and Transmission of High-Power Ultraviolet Radiation held 19-20 July 1973 in La Jolla, California.

Dr. R. A. Andrews	NRL
Dr. H. E. Bennett	NWC, China Lake
Prof. A. Bienenstock	Stanford University, MRC
Dr. F. C. Brown	Xerox, Palo Alto
Dr. E. Catalano	LLL
Prof. R. Coble	MIT, MRC
Prof. H. Ehrenreich	Harvard University, MRC
Prof. W. H. Flygare	University of Illinois, MRC
Dr. J. Gilman	Allied Chemical, MRC
Lt. Col. W. Goldberg	AFML
Prof. R. Gomer	University of Chicago, MRC
Dr. M. Hass	USNRL
Prof. F. Hawthorne	UCLA, MRC
Dr. P. W. Hoff	LLL
Prof. E. E. Hucke	University of Michigan, MRC
Dr. W. R. Hunter	USNRL
Dr. A. Kahan	AFCRL
Prof. P. L. Kelley	MIT/LL
Prof. F. A. McClintock	MIT, MRC
Dr. H. W. Paxton	Carnegie Mellon
Prof. W. Plummer	University of Pennsylvania
Prof. G. M. Pound	Stanford University
Dr. V. Rehn	NWC, China Lake
Prof. P. L. Richards	University of California, Berkeley, MRC
Prof. A. J. Sievers	Cornell University, MRC
Dr. M. S. Sparks	Xonics
Prof. W. E. Spicer	Stanford University
Dr. J. L. Stanford	NWC, China Lake
Dr. C. M. Stickley	ARPA, MRC
Dr. R. Thomson	NBS, MRC
Prof. M. Tinkham	Harvard University, MRC
Dr. R. E. Watson	Brookhaven National Laboratory

APPENDIX B

Materials Problems Related to
Transmission and Reflection of U.V. Radiation

General Chairman: W. H. Flygare
Discussion Leader: Marshall Sparks

Conference Room #1

Thursday, July 19, 1973

9:00 - 12:00   Dr. C. Martin Stickley, Director for Materials
Science, ARPA; General Statement of Meeting Goals.

Dr. Paul Hoff, LRL; High Power U.V. Lasers

Dr. H. Ehrenreich, Harvard; Theory of U.V.
Properties of Solids

Dr. F. C. Brown, Xerox; Studies of U.V. Properties
of Materials

1:00 - 3:30   Dr. H. E. Bennett, China Lake; High Reflectivity
U.V. Mirrors.

Dr. E. W. Plummer, NBS; Submicron Surface Roughness
and Chemical Analysis Studies of U.V. Reflectivity

Dr. J. Stanford, China Lake; Optical Excitations
of Surface Plasmons

Friday, July 20, 1973

8:00 - 12:00   Dr. W. Hunter, NRL; U.V. Mirrors

Dr. V. Rehn, China Lake; U.V. Fluorescence
Spectroscopy

Dr. W. E. Spicer, Stanford; U.V. Photoemission
Spectroscopy

Dr. E. Catalano, LRL; Transparent Materials for
High Power U.V. Windows

1:00 - 2:30   Dr. A. Glass, LRL; Multiphoton Absorption

Dr. R. Andrews, NRL; X-Ray Lasers.

2:30   General Summary

APPENDIX C

COMMENTS ON MATERIALS FOR VACUUM
ULTRA-VIOLET OPTICAL COMPONENTS

J. J. Gilman

I.   The Effect of Oxidation on the Reflectivity of Al

A dramatic reduction in the UV reflectivity of Al occurs when a freshly prepared film of Al is removed from high vacua. The reduction occurs over a wide range of wavelengths (roughly 1000-2000$\overset{\circ}{A}$) and the amount of reduction ranges up to ~50%. At first sight, this is a puzzling effect because it is known that high purity Al_2O_3 readily transmits UV radiation down to 1500-1600$\overset{\circ}{A}$. Possible causes are:

(a) The oxidation process roughens the Al surface on a fine scale that increases surface plasmon scattering.

(b) Nitrogen is incorporated in the oxide as an impurity level within the bandgap.

(c) Since the structure of Al_2O_3 is hexagonal and Al films are polycrystalline, the resulting oxide film has a spacial variation of refractive index that may lead to intense UV scattering; especially if the particle sizes are comparable with the incident wavelengths.

(d) Grain boundaries do not strongly perturb the properties of simple metals because the binding energy is largely

APPENDIX C

COMMENTS ON MATERIALS FOR VACUUM
ULTRA-VIOLET OPTICAL COMPONENTS

J. J. Gilman

I.   The Effect of Oxidation on the Reflectivity of Al

A dramatic reduction in the UV reflectivity of Al occurs when a freshly prepared film of Al is removed from high vacua. The reduction occurs over a wide range of wavelengths (roughly 1000-2000Å) and the amount of reduction ranges up to ∿50%. At first sight, this is a puzzling effect because it is known that high purity Al_2O_3 readily transmits UV radiation down to 1500-1600Å. Possible causes are:

(a) The oxidation process roughens the Al surface on a fine scale that increases surface plasmon scattering.

(b) Nitrogen is incorporated in the oxide as an impurity level within the bandgap.

(c) Since the structure of Al_2O_3 is hexagonal and Al films are polycrystalline, the resulting oxide film has a spacial variation of refractive index that may lead to intense UV scattering; especially if the particle sizes are comparable with the incident wavelengths.

(d) Grain boundaries do not strongly perturb the properties of simple metals because the binding energy is largely

independent of the structure factor. This is not true
for oxides, however, where the boundaries have sub-
stantially higher energies than the crystals. The
boundaries also have lower densities and therefore
higher refractive indices. Thus, grain boundary
scattering may be serious at UV wavelengths.

(e) Because of dispersion, Al_2O_3 may become partially
reflective at wavelengths near but substantially
greater than the absorption edge. In stoichiometric,
pure crystals this is an intrinsic effect. It may be
sensitive, however, to stoichiometry and impurities.

Of these causes, (a) is unlikely to be the important one
because the reduction of reflectivity does not show a peak as
would be expected for a plasmon effect. The second cause (b) is
trivial to correct by means of pre-oxidation with pure oxygen.

Causes (c) and (d) require either; protection to prevent
oxidation without introducing a different problem; or the use of
carefully oxidized monocrystals.

Cause (e) requires elimination of the oxide film.

A. Protection

1. Chemical - Very little is known about the chemistry
of clean surfaces, but it is known that surface compounds exist
that are not found in bulk form. Therefore, it is reasonable
to expect that passivation reactions might be found that would
prevent rapid oxidation of Al. The surface compounds must be

more, or nearly as stable as Al_2O_3 in order to resist reduction. Some possible reactants are: F_2, Si, Ni, Au, Cu, Ge, etc.

    2.   <u>Physical</u> - Although it would be difficult to build the structure, in principle a thin transmitting oxide window could be placed over an Al film, separated from it by a vacuum. Perhaps the freshly deposited Al could create the vacuum. It could be cooled initially to prevent oxidation, then covered, and then warmed to room temperature.

    B.   <u>Monocrystalline Structure</u> - Films that consist of individual crystals have been produced in recent years by means of:

        1.   Liquid-phase-epitaxy

        2.   Controlled sputtering

They might also be produced by:

        1.   Chemical-vapor-deposition in a controlled time varying directional temperature gradient

        2.   Recrystallization of an amorphous or fine-grained film in a temperature gradient

After a monocrystalline Al film had been produced, it would need to be carefully oxidized to produce a thin epitaxial overlay of monocrystalline Al_2O_3.

II.   <u>Polishing Techniques</u>

    Because of the sensitivity of UV reflectivity and refraction to both geometric and structural irregularities at surfaces, special shaping and polishing techniques are needed for making UV optical components. First of all, the structural damage and

irregularities that are induced by conventional abrasive shaping and polishing need to be minimized or eliminated. After this has been done it may also be necessary to consider surface chemical effects.

The art of polishing is perhaps most advanced in the yield of the preparation of substrates for thin-film electronics and data-processing; especially for magnetic films used in magnetic bubble memories. In this case both flatness and freedom from surface damage and irregularities is required. This is accomplished by means of chemo-mechanical polishing.

Chemical polishing alone is not satisfactory because optical flatness is not obtained. Therefore, in chemo-mechanical polishing a surface reaction is used to actually remove material from the surface, followed by mechanically wiping away the reaction products. In this way, optically flat and parallel sided garnet wafers ($\sim$100 microns thick) have been made with surfaces that are free of surface damage. Presumably, closely related methods could be used to make components with non-flat figures.

III. Thin Coatings for UV Optics

For high-power level devices optical coatings must be applied to a damage-free substrate surface, must strongly adhere to the substrate, and must have highly homogeneous optical properties.

The ideal way in which these conditions can be met is by epitaxially growing a mono-crystalline film on a monocrystal

-633-

substrate. In fact, this may be the only means for making

successful components.

At the present time, epitaxial films can be made by:

(a) high temperature evaporation

(b) controlled sputtering

(c) liquid-phase epitaxy

The most successful of these to date (as a commercial process)

is liquid phase epitaxy. Other processes that might be used

are:

(a) epitaxial chemical-vapor deposition in a temperature
    gradient

(b) recrystallization of an amorphous film in a temper-
    ature gradient

Less than ideal, but adequate, behavior might be obtained

from amorphous films lying on either crystalline or amorphous

substrates.

It is apparent that if optimal combinations of films

and substrates are to be found it is desirable to have numerous

candidates to make combinations from. Therefore, knowledge of

optical properties at UV wavelengths needs to be substantially

expanded.

IV.   Compositions for UV Optical Materials

Some suggestions:

(a) LiF:Mg - High purity LiF is very soft and therefore difficult

to shape and polish. It might be possible to harden it by adding

small amounts of Mg without significantly degrading the optical properties. The hardening caused by Mg additions is known to be: $\sim 2\times10^{12}$d/cm^2-unit conc., but the solubility limit and the optical properties of doped crystals are not known to the extent of my knowledge. Since the base hardness is $\sim 2\times10^7$, this can be increased an order of magnitude by adding about 100 ppm of Mg.

Other methods for hardening LiF include:

i. quench-hardening-fast cooling from an annealing temperature induces a high concentration of dislocation loops.

ii. radiation-hardening-F-centers cause substantial hardening and can be bleached out subsequent to the fabrication of a component.

(b) Yttrium Aluminum Garnet ($Y_2O_3\cdot Al_2O_3$ or other clear garnets) - At some sacrifice of band gap, this crystal gives exceptionally low piezo-optic coupling which minimizes themal lensing of an intense beam.

(c) Chrysoberyl ($BeO\cdot Al_2O_3$) - Unlike pure BeO this compound can be melted and therefore readily made into optical components. It may have some advantages over Al_2O_3.

(d) Combinations of BeO and B_2O_3 - I do not know the crystal chemistry but there may be a combination of interest.

(e) Spinels - Especially a lithium containing one, or perhaps even the prototype: $MgO\cdot Al_2O_3$.

(f) Beryl and related compounds.

SUPERCONDUCTING MATERIALS AND
LONG RANGE PROSPECTS

P. E. Duwez

A symposium on Superconducting Materials and Long Range Prospects was held on July 11, 12 and 13, 1973 at La Jolla. The subjects under discussion included

1) the presently available metallic superconductors such as Nb-Ti alloys and the A-15 type phases and the search for new superconducting metallic phases with higher T_c, H_c and J_c

2) the fabrication techniques and their shortcomings

3) the long range possibility of achieving a true superconductivity state in the charge-transfer organic salts.

The participants and their subjects were as follows:

LIST OF PARTICIPANTS

Name	Subject
Professor M. R. Beasley Harvard University	Fluctuations above the Transition Temperature as a Probe for High Temperature Superconductors
Dr. M. G. Bowman Los Alamos Scientific Laboratory	Basic Properties of High Temperature Superconducting Materials Related to Developing Fabrication Techniques
Dr. D. O. Cowan Johns Hopkins University	New Organic Compounds that Behave like Metals
Dr. A. L. Giorgi Los Alamos Scientific Laboratory	New High Temperature Superconductors and the Periodic Table
Dr. D. H. Gurinsky Brookhaven National Laboratory	Nb_3Sn Superconductors for AC Power Transmission

Preceding page blank

Professor A. J. Heeger
University of Pennsylvania

The Charge Transfer Organic Salts:
"A New Class of Materials"

Dr. John K. Hulm
Westinghouse Research
  Laboratories

Superconducting Materials Research
at Westinghouse

Dr. A. C. Lawson
University of California
  at San Diego

Lattice Instabilities in A-15
Type Superconductors

Professor W. A. Little
Stanford University

Organic Superconductivity

Professor Margaret L. A.
  MacVicar
Massachusetts Institute of
  Technology

A-15 Superconducting Materials -
Nb Based Alloys

Professor B. T. Matthias
University of California
  at San Diego

Metallic Superconductivity

Dr. Myron Strongin
Brookhaven National Laboratory

Superconductivity and Surfaces

Dr. M. Suenaga
Brookhaven National Laboratory

Preparation and Properties of
Multifilamentary A-15 Super-
conductors

Dr. C. C. Tsuei
California Institute of
  Technology

Copper-Rich Superconducting Alloys
Obtained by Melting

Dr. E. C. van Reuth
ARPA

Naval Superconducting Propulsion
Systems Requirements

Metallic Superconductors*

Matthias opened the first session of this symposium and
emphasized the importance of raising the transition temperature
of superconductors. The urgent need for superconductors with a
transition temperature of the order of 25°K is due to the un-

---

*Paragraphs between quotes were submitted in writing by their
authors.

certainties in the availability of helium within the next 10 to 20 years, and liquid hydrogen would be the only alternative. Matthias stated that one promising approach is in true ternary systems (in opposition to the pseudo-binary systems containing 3 elements). Lattice instabilities are also favorable to high T_c. Present obstacles are the lack of data on ternary phase diagrams and a poor understanding of the nature of lattice instability and changes. The lack of interest in phase diagram work results in the fact that it is a rather routine and unrewarding field of research, although it requires great skill and imagination. The same remarks apply to the determination of thermodynamic properties of alloys which are required to understand and even predict phase relationships in alloys.

"As more elements are found to become superconducting, it becomes increasingly evident that the normal state for a metal at low temperatures is the superconducting one. However, until this time none of the many reports of high transition temperatures, >25K, has ever been corroborated and are wrong in all probability. The same holds for the theoretical papers in this field. The major obstacle in the path to higher temperatures is the increasing instability the higher the critical temperature, in a given compound."

"There are many kinds of instability, such as lattice, stoichiometry, symmetry and existence. Most binary systems have been considered by now. Improvements in these may eventually lead to 25°K at most. Therefore, superconductivity of true ternary

phases, as compared to 3 element configuration in pseudobinary compounds, will become essential for higher transition temperatures. We have discovered several such phases, for example, $X_{.5}MO_3S_4$, with X = Cu, Ag, Mg, Zn, Cd, In, Sn, Pb at present $T_{max} \simeq 17°$, $Li_{0.3}Ti_{1.2}S_3$, T_c near 18°K. Both these sulfide systems are not even cubic any longer, something never found before in this temperature range. Even binary oxides such as $LiTi_2O_4$ in the spinel structure now gives 14°K. This is only the beginning of ternary compounds. They are a difficult field but will lead eventually to higher critical temperatures and free us from the requirement of being restricted to niobium."

The research program on new high T_c superconductors in Los Alamos was described by Giorgi.

"High transition temperature materials (T_c >15°K) are essential for the many proposed applications of superconductivity. Up to the present only a limited effort is expanded on investigations of new materials, and most of the investigations are concerned with materials containing Nb or V in the NaCl-type or A-15 type structure."

"Recent work by Matthias and coworkers with LiTiSx and $Mo_3M_xS_4$ (M_x = Cu,Pb,Ag, etc.) has shown T_c >15°K possible with non-cubic structures. Work at Los Alamos on $Y_2(Th)C_3$-type materials has given T_c = 17.5°K. So high T_c is possible with materials containing no V or Nb and having a complex cubic structure. Thus T_c >15° should occur in many other systems. The most likely metals to study are the 4-d transition metals Yttrium

through Technetium. More effort should therefore be expended on new preparations involving these metals for possible higher T_c materials. A better understanding of metastable high temperature phases in various systems involving these metals should be investigated."

Commenting on Giorgi's presentation, Hirth stated, "In Mo-Re alloys T_c increases to a maximum at about 35% Re. This is also the point where stacking fault energy and twin boundary energy are at a minimum. The correlation between the two may simply reflect that both depend on interatomic bond characteristic at that composition. Alternatively, it may indicate that stacking faults and twins directly affect T_c."

MacVicar reviewed her work on the structure and superconducting properties of the A-15 type alloys. She firmly believes that the A-15 superconductors are still the most promising ones because of their favorable crystal structure (3 orthogonal chains of atoms) achieving perfect ordering in the structure is an important factor. Experiments based on tunneling provide useful information on the degree of order in these alloys.

Phase transition at low temperature in the A-15 structure was reviewed by Lawson. It is important to study the connection between lattice instability and superconductivity because it will suggest the point at which the ultimate limitations of a given class of materials has been reached. Lattice instability must have an effect on the low temperature mechanical properties of superconductors. In this connection Hirth had the following

comment: "Disorder produced thermally decreases T_c in A-15 alloys. This could result from either lattice strain or from thermally induced antiphase boundaries. It might be of interest to compare these results with T_c measurements on crystals with deformation induced disorder, which should produce a different configuration of anti-phase boundaries (planar)."

The effect of ultra-thin films of oxygen on the surface of Nb was discussed by Strongin. Leed and Auger techniques were used to characterize the films. Ultra-thin films of Al on Si and Al on PbTe were also discussed in connection with a search for the exitonic mechanism.

The subject of multi-filamentary superconductors with A-15 compounds was discussed by Suenaga. The preparation of these materials starts by introducing wires (or rods) of the ductile superconducting metal A in the A_3B structure in a Cu ingot and processing the composite ingot by rolling and drawing. At a certain stage during the processing the element B is introduced by diffusion through the Cu to transform the A filaments (at least partially) into the A-15 structure. An alternative approach is to start with an ingot of the Cu-B alloy instead of pure Cu. This happens to be possible because the B elements such as Sn, Ga, Al, Si in the interesting A_3B superconducting alloys have relatively high solubilities in Cu. So far, A-15 multifilamentary superconductors have been successfully processed in appreciable lengths with Nb_3Sn, V_3Ga and V_3Si. The values of T_c, J_c and H_{c2} are comparable with known bulk values and are even better in some

cases. Work is in progress on Nb_3Ga and Nb_3Al for higher magnetic fields applications. The brittleness of the A-15 filaments is still a problem which may impose some limitations in handling these materials during fabrication. At the present time there is no easy way to join sections of the superconductor which must therefore be fabricated in a single length for each application.

Tsuei presented a new approach to Cu-rich alloys containing a superconducting phase and obtained by melting and subsequent cold working and heat treatment. The program started with a study of Cu-Nb alloys in which there is no intermediate phase. The solubility of Cu in Nb is very small and that of Nb in Cu is about 1 at.% at about 1100°C and decreases to practically zero at room temperature.

In the copper-rich alloys (0 to 10 at.% Nb) superconductivity is due to the proximity effect between the very finely divided Nb particles resulting from solid state precipitation. After cold rolling, the larger Nb primary dendrites are elongated and eventually form a quasi-continuous superconducting path. (C. C. Tsuei and L. R. Newkirk, J. Materials Science).

These alloys are of little practical interest since their superconducting properties are not better than those of pure Nb. The next step was to add Sn as one of the constituents in the melt and promote the formation of Nb_3Sn by suitable heat treatment after rolling. As an example, an alloy of composition $Cu_{88.5}Nb_{10}Sn_{1.5}$ (in atomic percentages) heat treated for 2 days at 650°C after rolling gave $T_c \simeq 15°K$, $J_c \simeq 5 \times 10^4$ A/cm² at 4.2°K

and zero field. It should be pointed out that this value of J_c is calculated by taking the total cross-section of the wire including the Cu (see C. C. Tsuei, Science 180, 57, 1973). No degradation of the superconducting properties were detected after a 10 mil wire was bent and then straightened 100 times around a one inch mandrel.

Additional work is being carried out on

1) the effect of Nb and Sn concentrations on properties,
2) the best cold working and heat treatment conditions for each alloy composition,
3) the determination of J_c as a function of the field,
4) the extension of the melting technique to other alloy systems such as Cu with NbAl, NbAlGe, VGa, NbGa and VSi.

All work so far has been on a laboratory scale (less than 10 g. ingots). Scaling-up the process will undoubtedly bring new problems, among which will be that of obtaining a rather rapid rate of freezing of the melt, which seems to be necessary to achieve a suitable as-cast microstructure. It is probable that consumable electrode melting will be the answer to that problem.

Gurinsky gave a report on recent work on Nb_3Sn + Cu super-conductors for application in BNL superconducting ac transmission project. What is required is a "flexible" coaxial cable which can be drawn into Dewar pipes much as is now done in pipe cable. Nb_3Sn has been selected because of its high T_c, J_c and H_c. In order to improve the economics of cryogenic equipment, the line

-644-

will be operated at 7 to 8°K. The disadvantages of Nb_3Sn materials are brittleness, cost, ac losses and Dewar, dielectric and hysteretic losses. A variety of materials are under investigation, including commerical tapes and composite materials prepared at BNL (see report by Suenaga above). It appears that losses are smaller in the BNL materials than in standard tapes. The smoothness of the surface of the superconductor is important in reducing losses.

Bowman submitted the following comments

1) Opportunities exist for the use of A-15 superconductors in several important applications. However, except for the commercial Nb_3Sn (and V_3Ga) tapes, these materials are not available for use.

2) Available commercial (tapes) are not optimized for use in terms of physical form (i.e., tapes vs. stabilized filaments), and are probably not optimized in composition and certainly not microstructure.

3) It seems evident that new fabrication techniques are needed not only for the A-15 structures now known, but also for adaptation to new, higher T_c structure that may be discovered.

4) New and better fabrication techniques will require imagination, but also better knowledge of basic materials properties such as:

   a) High temperature phase relationships particularly homogeneity ranges versus temperature.

   b) Diffusion coefficients as functions of temperature and of composition.

c) Thermochemical properties such as free energies of formation versus temperature and activities of components versus composition over homogeneity ranges.

5) It seems probable that existing data on A-15 systems are inaccurate or incomplete. This is the result of the fact that complete and accurate studies are difficult, and also because such studies have not been encouraged and supported. However, the promising properties of these materials will not be
  · realized in applications of superconductivity until adequate programs in fabrication development are carried out. More basic studies of the materials involved is a necessary part of such fabrication development.

The design and construction of a superconducting generator were presented by Hulm. Superconductivity will probably become economical for very large power plants in the range of 5000 to 10,000 M watts. The purpose of the Westinghouse project, however, is to demonstrate the feasibility of superconducting generators on a small scale, about 5M watts. The concept includes a rotating superconducting dc magnet with a normal ac stator. The field would be about 50 kG, so that use can be made of the NbTi copper stabilized multifilamentary superconducting wires presently fabricated on a large enough scale. The properties of the superconducting material in the Westinghouse design are summarized in Table I:

TABLE I. Factors of Importance in Superconductor Applications

	Present Material (Nb-Ti)	Future Need
1. Electrical		
T_c	9°K	>20°K
H_c	∿100 kG (4°K)	Adequate
J_c	∿2×10⁵ A/cm² (50kG, 4°K)	Could use more
Protection	Cu shunts	Use less
Stability	Filaments, Twist	Use less
AC Losses	Subdivision	Reduce
Zero-resistance joints	Presently impossible for filaments	Desirable in coils
Long perfect lengths	Possible in miles	Maintain
2. Mechanical		
Yield strength	300,000 psi	As high as possible
Bending radius without damage	Adequate	Maintain
Insensitivity to handling	Adequate	Maintain
3. Cost Per Pound	Too high	Reduce

The properties of the Nb-Ti multifilamentary superconductor listed in the table are somewhat marginal. In particular, the fact that the wires cannot be joined requires the use of some single length for a given application. This condition can be met for a 5M watts generator but for larger machines this would require fabricating many miles of wire from a single ingot. The rather low T_c of NbTi filaments is also marginal, and a summary of the desirable properties for future applications is given in the last column of the table. To meet those requirements NbTi must be replaced by other alloys, probably of the A-15 type, or nitrides or more complex ternary alloys. There is not much hope in achieving ductility in any of these higher T_c superconductors and fabrication techniques will have to be improved to deal with these essentially brittle alloys. Work is now in progress at Westinghouse on the fabrication of Nb_3Sn filamentary superconductors by a technique similar to that developed at BNL, filaments of Nb in Cu followed by diffusion of Sn through the Cu matrix.

The application of superconductivity to naval propulsion systems was summarized by van Reuth, in the absence of W. J. Levedahl who was unable to attend the meeting. The adoption of high speed gas turbines has imposed new design requirements because of the mismatch between high free-turbine speeds and low propeller speeds. In addition, the precise alignment required for mechanical drives leads to locating the turbines deep into the hull and this in turn leads to large duct volumes. Electrical

connections between electrical turbine driven generators and motors offers the possibility of using constant speed turbines located on the deck and motors with speeds compatible with propeller speeds. The main advantages of a superconducting system is the large reduction in space requirements. The Naval Ship Research and Development Center at Annapolis, Maryland, is designing and constructing a small propulsion system with an LM 1000 gas turbine of about 1000 hp, dc superconducting generators, motors and transmission line. At the present time, it is planned to use commercially available materials of the multifilamentary NbTi type, unless other higher performance superconductors become practical within one or two years.

Four presentations were concerned with the possible existence of superconductivity in organic materials. Abstracts of these talks (as submitted by their authors) are given below.

D. O. Cowan

Recent efforts in the synthesis and study of new organic materials have led to the development of the organic salts TTF-TCNQ, ATTF-TCNQ, TMTTF-TCNQ with electrical conductivities comparable with those of common metals. The work from the Hopkins Group that was presented included (a) The first synthesis of these materials, (b) dc and ac conductivity studies as a function of temperature, (c) X-ray crystal structure of TTF-TCNQ, (d) magnetic susceptibility as a function of temperature, (e) ESCA characteristics.

Work in this area has just begun. The flexibility of organic synthesis places at our disposal many of the important parameters for the development of new organic conductors. Such factors, for example, as bandwidth, band overlap, degree of disorder, correlation energies, and local polarizabilities can be varied systematically.

W. A. Little

I would like to stress that the value of the study of the development of organic materials as conductors or possible superconductors lies in the great versatility of organic structures. If the behavior of these materials can be understood and explained in physical terms, the possibility of developing a vast number of new materials of controlled properties could be realized. However, one should recognize the difficulties of preferring such materials; the materials problems of purifying them and of obtaining good crystals. These problems are made doubly difficult because of the need for close collaboration between organic chemist and physicist.

The principal point I wish to make in regard to the search for high temperature organic metallo-organic superconductivity is that in the period 1914-1971 virtually all theoretical arguments for or against such a phenomenon were based on model systems in no way related to any specific materials. Only in the last couple of years have questions been directed at the possibility of achieving high T_c in real materials. It is possible

now to get a much better idea of the practicality of achieving this, from a knowledge of the normal state properties and geometry of possible modified structures of such materials as TTF-TCNQ; $K_2 Pt(CN)_4 Br_{0.3} nH_2O$ and the interrelated layer structure such as TaS_2. This can be expected to be done in the next few years. The long term prospects should then be more clearly resolved.

In commenting on the recent results on the TCNQ-TTF by Coleman and co-workers, one should point out that first there seems to be some real question whether the pin point conductivity measurements may be responsible for yielding spurious results which could appear to give the anomaly. If, however, these experimental results have to be valid then the various theoretical explanations need careful scrutiny.

The original suggestion by Coleman et al that the high conductivity in the TTF-TCNQ could arise from superconducting fluctuation induced by the Kohn anomaly soft mode, I believe is invalid. At temperatures near 60°K the thermal excitation of any low lying phonon modes would lead to a phonon induced repulsion.

Secondly, the Bardeen suggestion that the enhanced conductivity could be due to a moving Peierl distorted lattice state has been criticized on various grounds and it appears that the actual state, in fact, carries no current, though this model ignores any role played by TTF.

As an alternate suggestion we point out that the longitudinal polarization of the TTF chain could provide an attractive

interaction for the electrons on the TCNQ chain and this might lead to an excitonic superconductor. Numerical parameters are currently being calculated to see if this is a real possibility. If this should prove to be the case then the enhanced conductivity should be considered as fluctuations into the superconducting state but with a Ginzburg-Landau T_C of about 530 K. The fluctuations in the molecular chains will suppress the superconducting contributions to the conductivity until one reaches temperature of the order of $T/T_C \approx 0.2$ as shown in a paper by the author in Phys. Rev. 1967.

If this latter model is valid one may obtain a general class of high temperature superconductors by preparing filamentary compounds consisting of two types of chains, A and B. In this the longitudinal polarization of the B chain would provide the excitonic interaction for the A chain.

A. J. Heeger

With regard to the Organic Charge Transfer Salt Materials, I want to emphasize the following points:

1. We are able to synthesize (to some extent design) organic metals. These are narrow band systems in which the el-el Coulomb repulsion is reduced by π-polarizability to the point where the metallic state is possible.

2. Significant progress has been made in understanding these systems and their electronic properties; i.e., band widths, effective masses, etc. Particularly conductivity, IR data,

thermopower, susceptibility, and nuclear relaxation studies demonstrate that (TTF) (TCNQ) is a narrow band, pseudo 1-d metal with negligible Coulomb repulsion, carrier density $\sim 4 \times 10^{21}$, and strong el-phonon coupling. Thus, in a general sense, such systems are of potential interest to the field of superconductivity.

3. The 1-d electronic structure is new and its particular advantages to superconductivity (or disadvantages) remain to be fully clarified. Because of the Peierls instability such 1-d systems are inherently unstable.

4. The giant conductivity observed is exciting and if indeed correct, must be a many-body effect for simple mean free paths of the implied magnitude cannot be understood. There are potential difficulties with such measurements in anisotropic systems which must be ruled out. This is being worked on both with dc and ac techniques. However, the data at present seem to indicate that the intrinsic conductivity is high in all samples but that impurities and defects are limiting.

5. The materials problems are severe. We have taken considerable effort to prepare high quality materials. The techniques used should lead to a new standard in purity. However, we clearly (even ourselves) do not have control of the materials in a detailed way; witness the nonreproducibility of the data and even the observation of different temperatures for σ_{max} in different samples. Even given pure materials the question of crystal perfection is very difficult in these compounds,

and particularly important in view of the l-d electronic properties. Evidently, the different results from different laboratories is the result of the relatively primitive state of the materials in this area.

6. The problems in developing organic materials for possible use (in the context of their electronic and magnetic properties) are clearly enormous. However, the potential associated with the coupling of the flexibility of organic chemistry to solid state phenomena makes this an area worthy of development.

M. R. Beasley

Superconductivity in systems with quasi-reduced-dimensionality--the one dimensional-like A-15 compounds and charge transfer organic salts, and the two-dimensional-like layered compounds--was discussed with regard to the effects of superconducting fluctuations in these systems and other superconducting properties which are sensitive to the reduced dimensionality. In particular, the interpretation of the conductivity maximum observed in TTF-TCNQ by Heeger et al as resulting from superconducting fluctuations was discussed. Following Anderson, Lee and Saitoh, it was shown that, if this conductivity maximum is due to superconductivity, it is more consistent with the known behavior of fluctuations in one-dimensional superconductors below T_c rather than above T_c. In this view, resistance below T_c results from momentary fluctuations from the superconducting state

to the normal state. Such behavior is well established in experiments on superconducting filamentary tin whiskers. In this view the superconducting transition temperature inferred from Heeger's data on TTF-TCNQ is 500K and the transition at 60K into the insulating phase is unrelated to the superconductivity. Many questions about this interpretation still remain, however.

For example, what role does the interchain coupling play? In the 2-D layered compounds, the coupling between the layers produces modified 3-D superconducting behavior above and below T_c, whereas the interpretation presented above on TTF-TCNQ is based on a totally 1-D model of the behavior in these systems. It was also pointed out that, again based on the experience with layered compounds, a substantial diamagnetism or Meissner effect should be present in TTF-TCNQ if it is really superconducting. Finally it was pointed out that superconductors with quasi-reduced-dimensionality may have peculiar and interesting properties below T_c which result uniquely from their reduced dimensionality. Very high critical fields and the possibility of coupled Josephson junction behavior in layered compounds were cited as examples. While the critical fields in layered compounds are known to be high (increasing with decreasing temperature at rates as high as $10^5 Oe/K$), the possibility of useful Josephson behavior is speculative.

Comments from Fred Hawthorne

As a chemist, I worry a great deal about the criteria of purity used in many superconductor preparations and the reproducibility of experiments from laboratory to laboratory. Perhaps a "Round-Robbin" evaluation scheme could be devised among ARPA contractors in which a particularly nice material is examined (independently) by three or more groups before the initial observations are accepted as fact. (This method was successfully employed in the 1950's when applied to rocket propellant performance and propellant stability).

I wish to most strongly support the search for new organic and organometallic superconductors and conductors. This field has fantastic breadth (and depth) which has been virtually untouched. Materials prepared and carefully characterized in such a program could be exceedingly useful if they were merely conductors, etc., which were cheap, physically tractable and had such special secondary properties as strong electronic radiation absorption, neutron absorption and high thermal stability, etc. The TTF/TCNQ system is the first step in this direction and I could care less if this material is superconductor or only a conductor at 60°K. What counts is the fact that phenomena have been observed which demand clarification and extension regardless of the theory which applies to the particular case. I foresee this area as becoming very important in materials research regardless of particular, preset expectations.

Acknowledgement

SUMMARY OF DISCUSSIONS ON
STRUCTURE AND PROPERTIES OF DISORDERED CARBONS

E. E. Hucke

Representatives of the four ARPA contractors investi-
gating various aspects of disordered carbons met July 23 and 24
with interested members of the MRC and guests. Those present
for all or a portion of the meeting were E. E. Hucke, University
of Michigan; P. L. Walker, Pennsylvania State; J. C. Bokros,
Gulf Medical Products Division; G. Tingey, Battelle Northwest;
M. Stickley, ARPA; A. Bienenstock, Stanford; M. Cohen, MIT;
Frank McClintock, MIT; H. Reiss, UCLA; J. Hirth, Ohio State;
F. Hawthorne, UCLA; H. Paxton, Carnegie-Mellon; P. Shewmon,
National Science Foundation and J. J. Gilman, Allied Chemical
Corporation.

The purpose of the meeting was fourfold:

1. To appraise MRC members of progress and gain ideas
for solutions of the outstanding problems of understanding the
evolution of the various structures and the resulting properties
of the disordered carbons ("glassy carbons").

2. To allow review of the major findings of the first
two years work by each of the contractors.

3. To map out the third year's effort for each contractor
and provide a coordination in sample exchange and property deter-
mination.

4. To discuss potential applications to DoD problems of some of the materials produced and characterized during the past two years.

A short introductory presentation was given by E. Hucke in which he discussed the broad range of structures obtainable in the non-crystalline forms of carbon. It was pointed out that solid carbons are always formed from molecular decompositions with loss of mass which in most cases gives considerable void space within the structure of the remaining high carbon polymers and therefore is importantly influenced by the temperature, pressure and time sequence occurring in the pyrolysis. While this behavior allows a very wide range in properties to be produced by varying processing it becomes difficult to either reproduce or characterize different disordered carbons. This behavior is in sharp contrast to the crystalline metals and ceramics and other amorphous materials where structural changes occur at constant mass.

The major findings of the various groups is summarized below:

University of Michigan

1. Thick sections of sound material with controllable porosity can be reproducibly achieved in short processing times.

2. Structural variations induced by processing changes can alter every property measured over wide ranges, often several orders of magnitude.

3. A comparatively simple thermodynamic measurement can differentiate the structural differences in various disordered carbons according to their enthalpy and entropy of disorder and yield information on the occupation of active surface sites.

Gulf Medical Products Division

1. High strength, fatigue and wear resistant carbons can be produced in both batch and steady-state pyrolytic beds.

2. The mechanical properties of the low temperature isotropic carbons can be well-correlated to their microstructures and can be varied over wide ranges with process parameters.

3. Alloy additions of Si and Ti can significantly improve mechanical properties, especially wear resistance.

Pennsylvania State

1. Fine dispersions of large amounts of Fe can be incorporated into carbons by way of organometal additions to organic precursors.

2. High pressure during pyrolysis significantly affects the structure of the carbon residue.

3. Useful molecular sieve carbons can be produced in open pored structures by suspension polymerization, pressing, and pyrolysis.

4. Novel carbon-metal structures can be produced by sputtering.

Battelle-Northwest

1. Some disordered carbons have significantly greater oxidation resistance per unit area than the best graphites.

2. Processing changes can improve oxidation resistance by several orders of magnitude.

3. Strength properties can be improved over commercial glassy carbon by controlling processing variables.

4. Toughness can be improved by several orders of magnitude by proper incorporation of carbon fibers.

It was decided that the third year's effort would concentrate on each laboratory following and extending the major result areas summarized above. Considerably more sample exchange between laboratories was recommended.

It was agreed that a meeting would be set up with Bob Meyers of Aerospace Corporation to discuss details of carbon materials requirements in certain nozzle and heat shield applications.

Martin Stickley will endeavor to appraise the group of any potential ways that materials developed by the ARPA contractors can be of use in the NOL re-entry vehicle program.

Considerable discussion of unique properties of disordered carbons led to the conclusions that:

1. Some unusually difficult seal and bearing problems would be solved with disordered carbons.

2. That carbon materials including those in a low pyrolysis state offer a solution to several problems as insulators, current collectors, and catalyst supports in advanced electrical cell systems for direct conversion, including high energy density batteries, fuel cells, and concentration cells.

3. Disordered carbons now offer the only satisfactory materials for many biomedical devices.

4. The inherent corrosion resistance of carbon for acids molten salts, H_2, Cl_2 offers potential solution to several difficult materials problems in solar energy storage and conversion systems.

Recommendation for Support

of a

HANDBOOK ON STRESS-CORROSION CRACKING

Submitted to:

    Dr. C. Martin Stickley
    Advanced Research Projects
     Agency
    1400 Wilson Boulevard
    Arlington, Virginia 22209

Submitted by:

    SCC Task Group
      B. F. Brown
      J. P. Hirth
      H. H. Johnson
      H. W. Paxton
      R. W. Staehle
      M. Cohen, Chairman

ARPA Materials Research Council
La Jolla, California
23 July 1973

RECOMMENDATION FOR SUPPORT OF A

HANDBOOK ON STRESS-CORROSION CRACKING

Table of Contents

1. **Introduction - Stress-Corrosion Cracking as a National Problem**

Corrosion is one of the most formidable technological problems in the United States; the cost to the nation has been estimated at $15 billion per year, with $1 billion of this directly related to the Department of Defense. The figures encompass all forms of corrosion, including stress-corrosion cracking (SCC), but the latter is particularly dangerous and can be exceedingly costly. In fact, SCC has become a limiting factor in several advanced technologies, such as deep submergence vehicles, rocket-motor casings, aerospace pressure vessels, critical members of the F111 aircraft, and deep-well drilling. Recently, S. H. Bush documented 88 episodes of SCC failures in nuclear power plants. The certification of several new nuclear installations has been delayed because of SCC threats. The near-disastrous SCC failure of titanium-alloy tanks in contact with methanol and N_2O_4 has imposed critical limitations on the Apollo Project. In a tabulation of materials problems encountered by an aerospace industry (compiled by H. Reiss in 1968),over one-third could be classed as stress corrosion.

The term stress-corrosion cracking, as we use it here, covers many important aspects of stress corrosion including corrosion fatigue, hydrogen embrittlement, and liquid-metal embrittlement. It is clear that existing knowledge on SCC is not adequately appreciated by the engineering and design community, even when preventive measures have been reported in the technical literature. As a consequence, structures are frequently designed, fabricated and operated in ignorance

of the stress-corrosion hazards; at least some of the resulting catastrophies could have been avoided if the lessons of experience were more readily accessible. It has been pointed out that SCC "epidemics" appear to follow a cycle of seven years, due to the circumstance that practical information on preventive measures arising from a series of SCC failures tends to be lost when the problems subside and the knowledgable personnel move on to other jobs.

In numerous instances, SCC may not seem to create difficulties becuase, intentionally or not, ultraconservative designs are adopted. However, as the performance requirements become more stringent, the growing danger of SCC will pose new constraints, particularly in view of the potentially high cost per incident. These disconcerting considerations have been recognized by industry and federal agencies. We estimate that, over the past 10-15 years, about $50 million has been spent on stress-corrosion research and development, with some $28 million of this coming from the government. And yet the available body of knowledge on SCC and its control has neither been digested authoritatively nor made available in a tractable form that could help eliminate many costly failures. We are calling attention to this "educational vacuum" as a critical national anomaly which can be removed, and suggest that the allocation for this purpose of some 1% of the aforementioned $50 million R & D expenditure on SCC would do much to secure that investment.

2. Response of the ARPA Materials Research Council

In recognition that SCC has emerged as a limiting factor in many important areas, the ARPA Materials Research Council sponsored an interdisciplinary conference on "Environmental Degradation of Stressed Materials" in July 1971 at Woods Hole, Massachusetts; the program was organized by J. P. Hirth and F. A. McClintock. Although the chief focus was on further research that might be required to advance the understanding of SCC, a consensus developed among the participants that a key need, if not the paramount one, was for an enlightened awareness of the nature and control of SCC on the part of engineers and designers. The idea arose that this objective might be achieved most effectively through a handbook or handbook series on stress corrosion.

This conclusion was reinforced at the July 1972 meeting of the ARPA Materials Research Council in Centerville, Massachusetts, where a small group chaired by Morris Cohen met to examine the feasibility, and to lay out the content, of a Handbook on Stress-Corrosion Cracking. The participants were:

  B. F. Brown, formerly Naval Research Laboratory,
     now The American University

  M. Cohen, Massachusetts Institute of Technology

  J. P. Hirth, The Ohio State University

  H. H. Johnson, Cornell University

  J. Kruger, National Bureau of Standards

F. A. McClintock, Massachusetts Institute of Technology

H. W. Paxton, formerly National Science Foundation
       now Carnegie-Mellon University

E. N. Pugh, University of Illinois

H. H. Uhlig, Massachusetts Institute of Technology

The group recommended that a major effort be undertaken to collect, systematize and evaluate SCC knowledge, and that the resulting publications be widely distributed in the engineering community. It was tentatively suggested that the program proceed in three parts. The first was to be a compilation of case histories of important SCC failures, with accompanying text and illustrations, to describe the salient lessons and guidelines derived from each example. The second was visualized as an organized assembly of qualitative SCC information on major alloy systems operating in various environments, with the information presented in convenient tabular form by both alloy system and environment. The third part was to be a full compilation of quantitative engineering data on SCC for design purposes. In all three instances, the formats selected were to be of maximum usefulness to design, operating, and maintenance engineers.

In the light of these considerations, B. F. Brown and R. W. Staehle were identified as the experts to undertake the handbook program, and H. W. Paxton was chosen to play a central role in the coordination process.

In June 1973, with impetus from the prior Council sessions on SCC, an International Conference on "Stress Corrosion and

Hydrogen Embrittlement in Iron Base Alloys" took place in Firminy, France, and several Council-related individuals participated. A measure of the concern for SCC problems was indicated by the fact that 230 investigators attended from some 10 countries. A novel feature of the proceedings, again stemming from the earlier Council meetings, was the preparation of critical interdisciplinary assessments of the presented subject matter by scientists outside the SCC field.

The International Conference provided still further evidence that a massive SCC educational effort is needed for the general engineering and design community, and that this situation applies to other industrial countries as well as to the United States.

3. Framework of the Proposed SCC Handbook Program

3.1 Plan for Two Volumes

The present SCC Task Troup consisting of B. F. Brown, J. P. Hirth, H. H. Johnson, H. W. Paxton, R. W. Staehle, and Morris Cohen, Chairman, convened at the July 1973 meeting of the ARPA Materials Research Council in La Jolla, California, to prepare a definite handbook proposal for possible support by ARPA.

The earlier three-part handbook concept was finally coalesced into a two-volume plan, the first and second parts to become Vol. I with B. F. Brown as the Principal Investigator, and the third part to become Vol. II with R. W. Staehle as the Principal Investigator. Appropriate, but tentative, titles with estimated number of printed pages are as follows:

Handbook on Stress-Corrosion Cracking

Vol. I - Control of Stress-Corrosion Cracking (200 pp. max.)

Vol. II - Engineering Design Data on Stress-Corrosion
Cracking (~1000 pp.)

We also recommend that an Advisory and Publication
Board be established, with H. W. Paxton as the Principal
Investigator.

3.2 Organization of the Program

The proposed organizational structure for the SCC Hand-
book program, as endorsed by the SCC Task Group, is dis-
played on page 9.

In view of the fact that ARPA funds its R & D programs
through various DoD agencies for contracting and monitoring
purposes, we propose that the Office of Naval Research be
chosen for that function. The ONR Metallurgy and Ceramics
Branch has considerable background in the corrosion field.
Under the suggested arrangement, presumably ONR would con-
tract with The American University, The Ohio State University,
and Carnegie-Mellon University, after receiving the necessary
funds from ARPA and after receiving separate, but related,
proposals from the three universities. The technical substance
of these proposals, as prepared by the respective Principal
Investigators, is summarized in the attachments to this docu-
ment. Detailed budgets will be submitted with the official
proposals to ONR. Our plan for the overall funding is presented
in the next section.   ·671<

Organizational Plan for Proposed Handbook Program

```
┌──────────┐
│ Sponsor  │
│  ARPA    │
└──────────┘
     │
┌──────────┐
│ Monitor  │
│   ONR    │
└──────────┘
```

Vol. I	Advisory and Publication Board	Vol. II
B. F. Brown American University (Contract I)	H. W. Paxton Carnegie-Mellon University (Contract III)	R. W. Staehle Ohio State University (Contract II) Battelle Memorial Institute (Subcontract IIA)
Responsibilities 1. To prepare a handbook type of manuscript that will alert the designer to the extent and seriousness of SCC. 2. To provide concise guidelines for preventing or controlling SCC.	**Responsibilities** 1. To coordinate the preparation of the two SCC Handbook parts. 2. To assure design and engineering emphasis. 3. To administer the publication of the two handbook volumes.	**Responsibilities** 1. To compile, systematize, and evaluate SCC data. 2. To present data in concise format usable for engineering design. 3. To provide design rules for alloy systems, environments, and applications.

672<

3.3 Funding Schedule

Vol. I (under Contract I) can be brought to the final manuscript stage by the end of the 1st contract year, and made ready-for-printing by the middle of the 2nd year. The book should be available for market during the early part of the 3rd year.

Vol. II (under Contract II) constitutes an appreciably larger effort, and involves data-collection by Battelle (under Subcontract IIA) as well as analysis and assessment by The Ohio State University. The final manuscript should be finished within 2 1/2 contract years, and made ready-for-printing by the end of the 3rd year. The book should be on the market during the 4th year.

The proposed funding schedule on page 11 is shown as a 3-year program; it is anticipated that any expenditures incurred during the 4th year in connection with the printing of Vol. II can be covered by a no-cost extension of Contract III. This contract will have the funds, not only for administering the Advisory and Publication Board, but also for the copy-editing, art work, proofreading and, in part, for the manufacturing costs. The latter will be necessary, whatever the publication mechanism turns out to be, in order that the two volumes can be aggressively marketed by the publisher at a modest price, say $5 for Vol. I and $10 for Vol. II. It also seems highly desirable to involve the appropriate technical societies, perhaps via the Federation of Materials Societies, for their endorsement and distribution services. The Government Printing Office and

Proposed Funding Schedule
(in thousands of dollars)

Year	Contract I American Univ.	Contract II Ohio State	(and Battelle)	Contract III Carnegie-Mellon	Book Publi- cation	Totals
1973-74	30	25	(25)	15		95
1974-75	15	65	(60)	15	15	170
1975-76	-	70	(55)	15	35	175
	45	160	140	45	50	440

674<

commercial publishers will likewise be considered in that frame-
work.

In any event, the Advisory and Publication Board will
carry the responsibility for getting the two volumes into print
and onto the market. We estimate that approximately 5000 copies
of each volume can be sold, although in the opinion of one informed
source, the sales of Vol. I might reach 20,000 copies at a $5
price.

4. Closure

The varied studies on SCC that have emanated from activities
of the ARPA Materials Research Council over the past few years
point clearly to the urgent need for a handbook-type of publi-
cation that can be used effectively by design and operating
engineers.

To that end, the present SCC Task Group recommends that
ARPA undertake the sponsorship of a two-volume Handbook on
Stress-Corrosion Cracking, with the approximate subtitles:
Vol. I - Control of Stress-Corrosion Cracking, and Vol. II -
Engineering Design Data on Stress-Corrosion Cracking.

We also recommend an organizational plan and funding
schedule, based on three contracts to be monitored by the Office
of Naval Research at an estimated cost of $440,000 over a three-
year period. The proposed budget includes an allowance for part
of the publication costs so that the two volumes will have the
necessary sales distribution to reach a maximum audience in the
engineering and design community

ATTACHMENT I

DEVELOPMENT OF A STRESS-CORROSION CRACKING

CONTROL PLAN FOR THE ENGINEER

B. F. Brown
Department of Chemistry
The American University
Washington, D. C. 20016

Introduction

The history of the development of advanced structures --
structures for which there has been little opportunity for the
evolutionary processes to guide designers -- has been replete
with examples of premature structural failures. A non-negligible
proportion of these failures have been and continue to be caused
by stress-corrosion cracking (SCC). SCC has caused failures
of large rocket-motor cases of both steel and titanium, aluminum
aircraft fuselages and landing gear components, stainless steel
components of power plants (civilian and military, nuclear and
non-nuclear), and in surface ships. It has also caused important
loss of integrity of underground pipelines, and it places a major
restriction on the selection of materials for drilling deep oil
wells. It led to redesigning of the Deep Submergence Rescue
Vehicle and of the supersonic transport, and it has caused
numerous ugly surprises in a variety of aerospace pressure ves-
sels. The problem is thus widespread throughout both civilian
and military technology.

Although research has produced and is producing impressive
advances in our understanding of the problem, from the complexity
of the phenomenon it is clear that research alone will not
eliminate the problem for a very long time. In the meantime, a

676<

methodology is needed to <u>control</u> the problem. The development of such a methodology and mode of presentation which invites study by the busy engineer and which is comprehensible to him without unduly compromising scientific accuracy is the essence of this proposal.

Stress-Corrosion Cracking Control Plan

Not only is there no general research solution to the problem, but there is no <u>general</u> engineering solution, either. But there are <u>particularized</u> solutions, or partial solutions, which have been developed for certain structures, alloys, and environments. The control plan which is the goal of the proposed work is the structuring of these particularized solutions into a methodology from which the engineer can derive guidance either for similar situations or for novel structures which may match components of the control plan as far as SCC is concerned.

Included in the presentation of the control plan there would be material to

1) alert the target readers to the reality and potential gravity of the SCC threat, by presenting pictorial evidence of failures in various structures, alloys, and environments;

2. explain the basics of metallurgy, mechanics and chemistry as these interact to cause SCC, including the elements of current theories;

3. emphasize that tests are needed to verify SCC behavior in fundamentally new combinations of alloys and environments, indicating how these tests are commonly conducted and the results interpreted; and

4) prepare the reader either for digging more deeply into the literature or for discussing a potential problem with a consultant where unusual complexities or hazards might be involved.

The primary target reader is the design engineer, materials engineer, and maintenance engineer. A control plan is needed for these engineers because available guidelines are either fragmentary, exhortative only, outdated, proprietarily held, or comprehensible only to the specialist.

It should be emphasized that the proposed development in no wise indicates that all the needed research work has been done -- quite the opposite. The problem is simply so complex that the research solution is a long way off, and until the scientifically neater solution becomes available we need better engineering guidelines than are now available.

The format in which the control plan would be presented is outlined below.

STRESS-CORROSION CRACKING CONTROL PLAN

I. Preface, including warning that there is no safe substitute for testing new combinations of alloys and environments in order to be sure of controlling SCC.

II. Methodology of SCC control strategies

(a) according to structure; primarily aimed at the designer. See Table I as an example of such a tabular presentation in preliminary (and incomplete) form.

(b) according to environment

(c) according to alloy

The purpose of this highly condensed presentation is to provide the reader with perspective before he is confronted with slabs of text on an unfamiliar discipline.

III. Theory of SCC (highly condensed), with such background material in metallurgy, electrochemistry, and fracture mechanics as is needed in order to comprehend the control plan essentials and to be able to enter more

TABLE I

Methodology of SCC Control Plans

(First Draft; provisional and incomplete)

Structure/Alloy System	Mandatory	Primary	Alternate	Supplemental
1) Aerospace-hydrospace/high-strength steel	Avoid sustained or repetitive tensile stress, or -	Measure KI_{scc} in pertinent environment & operate below it.		Cd-plate; avoid opportunity for local acidification; hydrotest pressure vessels with inert liquid. Use steel with high KI_{scc}
2) Power plants/austenitic SS	Control purity (Cl⁻) of water & quality of steam. Meticulous cleanliness during assembly.	Minimize stresses due to fit-up; minimum use of cold-worked steel; avoid sensitization.		Minimize design temperature; use more resistant alloys.
3) Aerospace pressure vessels, chemical reaction vessels/titanium		Select resistant alloy; measure KI_{scc} if new environment, operate below it, if feasible.		Clean up after final assembly by anodizing. Avoid contaminating surface with Fe from tools
4) Boilers/ferritic steel (caustic cracking)		Use balanced treatment of water and controlled heat flux to prevent development of high local causticity.	Use steel with some Cr (e.g., Croloy 1 1/4).	
5) Heat exchangers/Cu alloys	Avoid both mercury & its salts and ammonia	Avoid excessively high hardness tempers	Use a resistant alloy (e.g., cupro-nickel)	
6) Aerospace/Al alloys	Avoid contamination with mercury & its salts.	Heat treat & alloy to trade off strength for SCC resistance.	Exploit texturing	Use cathodic protection, coating, cladding.
7) Chemical process vessels/austenitic stainless steels.				

(etc.)

sophisticated literature or consult with experts.

IV.   Expansion of Table I section by section. The sections
       would be illustrated by photographs of failures, micro-
       graphs, and fractographs. Cautions about exceptions and
       uncertainties. Bibliography for each section to get the
       interested reader an entry into the specialist literature.

       This section is the heart of the control plan, for
       which the tables in II above only serve for orientation.

V.   Common test methods, with illustrations of specimens,
       discussion of test procedures, and analysis of test
       results. Bibliography (brief) on test methods, for
       which good recent summaries are available.

VI.   Exhaustive index, including cross-indexing by alloy and
       environment.

It should be emphasized that this format represents a

number of compromises, and that as attempts are made to organize

information into this format, it may become desirable to

modify the format in one or more ways. The final form of the

complete control plan will constitute the final report on the

contract.

Just as any comprehensive control plan for _fatigue_ must give

proper balance between initiation and propagation of a _fatigue_

crack so must the SCC control plan give proper attention to both

the initiation and propagation of a stress-corrosion crack. The

strategies to control initiation do not always help to control

propagation, and in fact may accelerate propagation. There are

essentially no data treating both initiation and propagation in

the same system; therefore a great deal of the control plan will

have to be based upon subjective evaluation of inhomogeneous data.

Such subjective evaluations, even by consensus of the most

experienced specialists, is less satisfying by far than the usual

requisite objective data; but in the absence of such data, the

alternative to the subjective judgments of seasoned specialists

is the judgment of nonspecialists. Thus, as in so many engineering situations, much of the control plan will have to rest on informed estimates.

Operating Mode

The modus operandi of the project will be for the Principal Investigator to proceed with trial organization, based upon his experience with various aerospace, hydrospace, ordnance, power plant, and other hardware projects of the Army, Navy, Air Force, AEC, and NASA, supplemented by literature and especially by discussions with specialists and advisory bodies. Care will be taken to maintain close liaison with other SCC and related ARPA projects. Repeated recycling of components of the control plan, and of the entire plan, is proposed as a necessary procedure for getting the necessary critical in-put. Much of this recycling can be done effectively only by face-to-face discussions, at least in the early stage, with geographically dispersed experts; this recycling process will account for the entirety of the travel budget, which will be disproportionately large compared to a typical laboratory project of the same total budget.

Personnel

The Principal Investigator will be Dr. B. F. Brown, who will devote half time to the project on a first-priority basis. Dr. Brown was formerly Head of the Physical Metallurgy Branch of the Naval Research Laboratory (GS 16) and was for one year Superintendent of the Metallurgy Division (PL 313). He is

presently Senior Research Scientist at The American University.
As a consultant to the project, Dr. R. T. Foley will be available
without cost to the project budget. Dr. Foley is an electro-
chemist with extensive basic and applied experience in corrosion
and with the additional background as supervisor of a large indus-
trial metallurgy group. He is Editor of the Corrosion Monograph
Series of The Electrochemical Society, and is Professor of
Chemistry at The American University. The professional quali-
fications of both men are well known at the Office of Naval
Research, which is understood to be under consideration for the
agency to execute and monitor the contract if the proposal is
accepted by ARPA. Full professional biographies are available
if needed.

It is estimated that the final report will be written not
later than one year from the contract starting date, with a
one-year budget of $30,000. The second year, with a budget of
$15,000 will be occupied with preparation of the book manuscript
for publication.

ATTACHMENT II

Proposal for the Development

of a

Design Handbook on

Stress-Corrosion Cracking
(Volume II)

R. W. Staehle
Professor
Department of Metallurgical Engineering
The Ohio State University
Columbus, Ohio 43201

1.0   Introduction

There is presently a major need to make the available data from stress-corrosion cracking specifically useful in the design of engineering structures.   This is a proposal for a joint effort by The Ohio State University and Battelle Memorial Institute to produce a handbook volume containing such information.   Purposes of the book are the following:

1.   Compile and evaluate existing information.

2.   Present data in a concise and usable form.

3.   Provide design guidelines on alloy systems, environ-
     ments, and applications.

Specific reasons for such a handbook are the following:

1.   A large number of very expensive failures continue to
     occur which are caused by stress-corrosion cracking
     (SCC).

2.   There does not exist in any form a systematic, focused,
     and coherent approach to incorporating the broad range
     of SCC data in design.

3.   The available data have been obtained by a large number
     of investigators using widely different test conditions;
     and the data obtained depend very sensitively upon the
     method of test.   Thus, it is difficult for the user to
     interpret the significance of the data.

4. The influence of a given environmental species depends greatly upon the specific molecular form. For example, the hydrogen-affected growth velocity of a crack in high-strength steels in gaseous environments varies about 10^7 depending upon whether the molecule is H_2S or NH_3. The phase of the environment is also crucial; for example, wetting and drying is probably 10^3 times more aggressive in terms of time-to-failure than a single phase gaseous environment. Finally, very small changes in the alloy-composition can profoundly influence the tendency to crack. Decreasing the nitrogen from 0.04 w/o to 0.004 w/o in austenitic stainless steels stops cracking in some environments.

Despite these complexities, we need to provide simple, accessible, and reliable guides to designers in the form of useful parameters. These parameters must be in terms which are readily understood and which can be translated into standard specifications.

Further, we need to provide the designer with options for specifying alloy composition and heat treatment, environmental composition, surface alterations, acceptable growth rates for cracks, cyclic stressing conditions, and interaction with existing flaws.

2.0 Scope

The scope of the handbook (Vol. II) includes the following:

2.1 Materials

The present project includes only metallic materials, although we recognize the eventual need to consider glasses, ceramics, and polymers.

The alloy categories, for example, are the following:

1. High-strength quenched and tempered steels.

2. Ferritic, duplex, and austenitic steels including higher-nickel alloys.

3. Carbon steels.

4. Aluminum-base alloys.

5. Titanium-base alloys.

6. Copper-base alloys, including brasses, monels, cupro-nickels.

2.2 Phenomena and Environments

All likely environments are to be considered; these include gases, aqueous liquids, and liquid metals. Classes of phenomena include hydrogen effects, corrosion fatigue, and liquid-metal embrittlement.

2.3 Sources of Information

The handbook will utilize the entire available data-base, including domestic and foreign literature. At least the following categories of literature will be utilized:

1. Open published literature.

2. Government reports.

3. Public company reports.

4. Unpublished information which can be verified.

The existing data-base from MCIC at Battelle will be expecially useful.

2.4 Variables

The design variables of interest depend greatly upon specific applications as well as the nature of the information.

Key design parameters will involve crack velocity vs. stress intensity, or time to failure vs. applied stress. The stress (stress intensity) dependencies will then be modified by the parameters of environmental species and their concentrations,

alloy additions, temperature, alloy structure, phase of environ-
ment (e.g., wetting and drying), galvanic coupling, and surface
alterations (e.g., shot peening, decarburization).

2.5 Interpretive Commentary and Correlations

Where appropriate, suggested interpretations and in-
junctions will accompany the data. In some cases caution will
be advised; in others, useful approaches to utilizing the data
will be suggested.

In addition, the possible applicability of data from one
alloy will be suggested, where appropriate, to apply to another.

Ranges of validity of the data and possibly legitimate
extrapolability will be defined.

2.6 Cross Reference by Environment

The essential presentation will be organized by alloy
systems; a separate page will be utilized for each alloy in the
section. However, many designers approach their designs by
asking for the optimum alloy for a given environmental circum-
stance. There will be a cross-referencing system wherein alloy
performance in a given environment will be rated, and indexing
information will be given to the alloy data.

2.7 Design Procedure

Incorporating SCC information often is a new venture
as well as a departure from conventional procedures. Thus,
specifically suggested or examplary procedures for utilizing the
information are required. The following specific areas might be
considered in this regard: design for non-standard as well as

standard conditions; procedures for incorporating residual stresses in design in view of the large number of SCC failures caused thereby; the "multiple-fix" approach of changing the stress, environment, and alloy, or surface treatment in some relative combination; procedures for estimating extent of damage during service operation; and procedures to adopt when data are not available.

2.8 Examples of Case Histories

Selected examples of failures will be included and will be chosen to illustrate schemes for specific design procedures and how a correct approach could have prevented the failure. These cases will be coordinated with Vol. I.

3.0 Structure of Presentation

3.1 General

The specific organization of the handbook volume is described in this section. In the final publication, full use of curves, tables, schematic presentations, and explanatory notes will be utilized.

3.2 Introduction

Components of the introductory section include:

1. General effect of SCC on mechanical response of metals.

2. Description of component phenomena which comprise SCC.

3. Explanation of various types of data presentation and directions for their general use.

4. Procedures for incorporating SCC data into design.

5. Examples of major equipment failures and demonstration of specific design procedures for avoiding such failures.

6. Procedures for incorporating residual-stress considerations.

7. Annotated bibliography.

8. Multiple approach to ameliorating SCC involving simultaneous changes in several influences such as stress, environment, alloy and surface treatment.

9. Procedures for total system design incorporating SCC: includes stress-time and environment-time profile together with crack growth response in each regime.

10. Initiation vs. propagation considerations.

11. Significance of smooth surface and pre-cracked specimens.

3.3 Alloy Class

For each alloy class there will be an introductory section including the following specific items:

1. Composition of alloys in the class.

2. Identification of major existing and possible areas of concern for SCC.

3. Systematic dependence of SCC upon alloy composition within the class.

4. Metallography and fractography of SCC within the class and discussion of patterns and possible meaning to designer.

5. General physical metallurgy, chemistry, and mechanical response (without SCC) of the class.

3.4 Alloy Module

Each alloy module will contain the following:

1. Primary information will be on the effect of stress arranged according to one of the following modes:

a. Crack velocity vs. stress intensity.

b. Initial stress intensity vs. failure time.

c. Initial stress (smooth specimen) vs. failure time.

d. Stress vs. number of cycles to failure (smooth specimen).

2. The variables of alloy composition, alloy structure, environmental conditions, surface treatments will be summarized. Attempts to correlate work from many investigations will be incorporated in envelope plots and correlation equations.

3. The modules will be subdivided according to major environmental classes such as aqueous chloride environments, aqueous caustic environments, and liquid-metal embrittlement, if the particular alloy-environment system indicates appropriate patterns.

3.5 Environmental Cross Reference

Important industrial environments will be divided into appropriate categories. Within these categories the alloys or alloy classes will be organized according to an approximate figure of merit. Examples of such environments are:

1. Sea water.

2. Liquid metals (further subdivided by Hg, Na, etc.).

3. Hydrogen gas.

4. Anhydrous ammonia.

5. Hydrogen sulfide-containing environments.

6. Caustic solutions.

Careful attention will be given to organizing environmental classes in order to group similar ones together as well as to distinguish subtle differences.

3.6 Analyses of Failures

Specific procedures and precautions will be outlined

for analyzing the evidence from failures and drawing appropriate conclusions. For example, in the observation of fatigue failures an approximate measurement of da/dN can be obtained from photographs of the fracture surface; and this, in turn, can be used to determine ΔK. Appropriate chemical analyses can distinguish many important causative agents such as those promoting hydrogen embrittlement or chloride cracking.

4.0 Approach and Tasks

1. Begin immediately to collect and catalog data. Total data collecting will require 1.5 years. Process will continue as material is circulated.

2. Establish prototype mode for presenting alloy module, and complete in nine months. This will require major agreements of Advisory Board and consultants.

3. Analyses of data will start in nine months, and continue until completion of handbook.

4. As sections are completed, they will be sent to recognized authorities for critiques.

5.0 Personnel and Facilities

Principal Investigator is R. W. Staehle. He will coordinate activities at OSU with those of Battelle Memorial Institute. He is Professor of Metallurgical Engineering at The Ohio State University.

The work on this handbook will be a collaboration between the Metals and Ceramics Information Center of Battelle Memorial Institute in Columbus, Ohio and the Department of Metallurgical Engineering at The Ohio State University. The files of information at BMI will be readily available for this purpose. Of particular importance to the proposed program is a data-collecting program by BMI-MCIC on fracture mechanics parameters now being sponsored

by the Air Force Materials Laboratory. This project has pro-
duced a "Fracture Control Handbook" after the first year and will
contain two additional supplements, concentrating on high-strength
steels, stainless steels, and alloys of aluminum, titanium, and
nickel.

Important aspects of the Principal Investigator's background
relative to the proposed project are the following:

1. Chief Editor of Corrosion Magazine.

2. Co-Editor (With M. G. Fontana) of Advances in
   Corrosion Science and Technology by Plenum Press,
   Publisher (3 volumes).

3. Senior Editor and Organizer of NACE International
   Conference Series:

   a. Fundamental Aspects of Stress Corrosion Cracking
      Co-Editors: Forty and Van Rooyen, published 1969.

   b. Corrosion Fatigue: Chemistry, Mechanics, Micro-
      structure, Co-Editors: McEvily and Devereux,
      published 1973.

   c. Localized Corrosion, Co-Editors: Brown, Kruger,
      and Agarwal, to be published 1973.

   d. High Temperature High Pressure Electrochemistry
      Aqueous Solutions, Co-Editors: Jones and
      Slater, to be published 1974.

   e. Stress Corrosion Cracking and Hydrogen Embrittle-
      ment of Iron Base Alloys, Co-Editors: Hochmann
      and Slater, to be published 1974.

6.0 Budget

The total budget for the projected handbook is as
follows:

	OSU	BMI	Total
First Year	$25,000	$25,000	$ 50,000
Second Year	65,000	60,000	125,000
Third Year	70,000	55,000	125,000
			$300,000

692<

ATTACHMENT III

ADVISORY AND PUBLICATION BOARD

H. W. Paxton
Professor
Department of Metallurgy and Materials Science
Carnegie-Mellon University
Pittsburgh, Pa. 15213

The production of the two handbook volumes under discussion needs input from a variety of sources. The SCC Task Group recommends that a committee be operated which will stay in very close contact with both writing projects to provide advice and oversee publication. This committee (the Advisory and Publication Board) will be made up of a suitable mixture of people knowledgable in stress-corrosion cracking and representatives of the potential user community, e.g. designers, operators, and maintenance engineers. The goal of maximum utility to the user will be kept clearly in mind. Scientific sophistication will be maintained only at a level appropriate to induce credibility. The Board will also concern itself explicitly with coordination between the two handbook volumes.

The operational plan visualizes two-day meetings four times each year for the first two years. At these meetings, the Board will initially assist in laying out a detailed working outline for the programs. At later meetings, reviews on the significance and value of the work to-date will be held based on material submitted in advance to the Board.

It seems quite likely that differences of opinion will arise on presentation of text material. It is a primary responsibility of the Advisory Board to resolve any such differences. The Board does not see itself as directing the two

Principal Investigators and this is explicitly shown on the organization diagram.

During production of the texts for the volumes, the Board will be concerning itself with the creation of an attractive format and an energetic marketing and distribution mechanism. In keeping with the spirit of encouraging widespread use, the Board will examine a variety of possible methods, not necessarily mutually exclusive. We will also explore ways of working with the technical societies, or perhaps with their recently-formed vehicle, the Federation of Materials Societies. It seems reasonable to expect that books given the blessing of technical societies will be more likely to receive convinced usage.

We also intend to examine possibilities with commercial book publishers and the Government Printing Office. The lowest possible cost consistent with a production of good quality and maximum distribution is a prime desideratum for extensive usage.

Some probable members for the Board are:

H. W. Paxton	Carnegie-Mellon University (Chairman)
H. H. Johnson	Cornell University
S. H. Bush	Battelle Northwest
J. H. Westbrook	General Electric Company

We expect to draw in an informal way on the advice of M. Cohen, Chairman of the Task Group on Stress Corrosion, as well as other members of the ARPA Materials Research Council.

Selection of "user" members of the Advisory and Publication Board cannot be made at this time. There are many sectors of industry and government involved and expertise in all cannot be represented. The relevant industries include:

chemical
nuclear
petroleum
construction
pressure vessels

transportation, including
aerospace
electrical
communications

It may be that we can obtain representation from some of these industries and government groups by correspondence, questionnaires, and editorial critiques. This approach will also be explored to obtain input especially from Western Europe and Japan.

A tentative budget for three years of operation is attached. It includes, during the second and third years, book-preparation items such as professional editorial assistance, art work, and some manufacturing costs in order to make the two volumes marketable at an attractive selling price.

BUDGET FOR CONTRACT III
Carnegie-Mellon University, Contractor

Year 1

H. W. Paxton (5% time)	$ 2000
Secretarial (15% time)	1000
Overhead Fringe Benefits	2100
Telephone	500
Supplies and Services	500
Travel of Board Members (50 man-days)	5000
Consulting Fees (50 man-days at $100)	5000
	$16100

695<

Year 2

Same as for Year 1	$16100
Plus Vol. I Preparation Costs	15000
	$31100

Year 3

H. W. Paxton (5% time)	$ 2200
Secretarial (15% time)	1100
Overhead	2300
Fringe Benefits	
Telephone	500
Supplies and Services	500
Travel of Board Members (20 man-days)	2000
Consulting Fees (20 man-days)	2000
Vol. II Preparation Costs	35000
	$45600

696<