CONF-9605186--

Proceedings of the

FOURTEENTH SYMPOSIUM ON ENERGY ENGINEERING SCIENCES

Mechanical Sciences; Solids and Fluids

May 15-16, 1996
at
Argonne
National Laboratory
Argonne, Illinois

Reproduced from the best available copy.

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831
Prices available from (423) 576-8401

Available to the public from the
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, VA 22161

CONF-9605186--

Proceedings of the

FOURTEENTH SYMPOSIUM ON ENERGY ENGINEERING SCIENCES
Mechanical Sciences; Solids and Fluids

May 15-16, 1996

at

ARGONNE NATIONAL LABORATORY

Argonne, Illinois

Cosponsored by

Office of Basic Energy Sciences
U.S. DEPARTMENT OF ENERGY

and

Energy Technology Division
ARGONNE NATIONAL LABORATORY

Coordinated by

Argonne National Laboratory
9700 South Cass Avenue
Argonne, Illinois 60439

Conference Number: 9605186

Do Not Do Analytics on these

Conference 9605186

14. symposium on energy engineering sciences: mechanical sciences; solids and fluids, Argonne, Illinois, 15-17 May 1996

Entered: 06/18/96
Changed: 10/04/96

Papers : 2
Volumes: 1

Activities: 0
Sponsors : 2

Energy Meetings:

Conference Paper 1 of 2

Author : Mondy, L. A.
Report Number: SAND--96-1384C
Secondary : CONF-9605186--1
Trac No. SC: : M96011839

Entered : 06/18/96
Changed :
Description:
Note :

Microstructure and momentum transport in concentrated suspensions

Conference Paper 2 of 2

Author : Telschow, K. L.
Report Number: INEL--96/00162
Secondary : CONF-9605186--2
Trac No. SC: : M96014122

Entered : 09/05/96
Changed :
Description:
Note :

Calculation and measurement of the critical state in superconducting tapes

FOURTEENTH SYMPOSIUM ON ENERGY ENGINEERING SCIENCES
Mechanical Sciences; Solids and Fluids

FOREWORD

This Proceedings Volume includes the technical papers that were presented during the Fourteenth Symposium on Energy Engineering Sciences on May 15-16, 1996, at Argonne National Laboratory, Argonne, Illinois. The Symposium was organized into seven technical sessions, which included 26 individual presentations followed by discussion and interaction with the audience. A list of participants is appended to this volume.

This was the fourteenth annual Symposium sponsored by the Engineering Research Program of the Office of Basic Energy Sciences of the U.S. Department of Energy. The technical areas encompassed in this year's Symposium were mostly from solid and fluid dynamics. Each paper dealt with the research effort being sponsored by the Engineering Research Program.

The DOE Office of Basic Energy Sciences, of which Engineering Research is a component program, is responsible for the long-term mission-oriented research in the Department. It has the prime responsibility for establishing the basic scientific foundation upon which the Nation's future energy options will have to be identified, developed, and built. It is committed to the generation of new knowledge necessary for the solution of present and future problems of energy exploration, production, conversion, and utilization, consistent with respect for the environment.

Consistent with the DOE/BES mission, the Engineering Research Program is charged with the identification, initiation, and management of fundamental research on broad, generic topics addressing energy-related engineering problems. Its stated goals are: 1) to improve and extend the body of knowledge underlying current engineering practice so as to create new options for enhancing energy savings and production, for prolonging useful life of energy-related structures and equipment, and for developing advanced manufacturing technologies and materials processing with emphasis on reducing costs with improved industrial production and performance quality; and 2) to expand the store of fundamental concepts for solving anticipated and unforeseen engineering problems in the energy technologies.

In achieving these goals, the Engineering Research Program supports approximately 130 research projects covering a broad spectrum of topics cutting across traditional engineering disciplines with a focus on three areas: 1) mechanical sciences, 2) control systems and instrumentation, and 3) engineering data and analysis. The Fourteenth Symposium involved approximately one-fifth of the research projects currently sponsored by the DOE/BES Engineering Research Program.

The Fourteenth Symposium was held under the joint sponsorship of the DOE Office of Basic Energy Sciences and Argonne National Laboratory. Local arrangements were handled by Ms. Jacquie Habenicht of ANL Conference Services. Ms. Nina Daly of the ANL Office of Technical Communication Services was responsible for assembling these proceedings and attending to their publication.

I am grateful to all who contributed to the success of the program, particularly to the participants for their excellent presentations and their active involvement in discussions. The resulting interactions made this Symposium a most stimulating and enjoyable experience.

Robert Goulard, ER-15
Division of Engineering and Geosciences
Office of Basic Energy Sciences

FOURTEENTH SYMPOSIUM ON ENERGY ENGINEERING SCIENCES

May 15-16, 1996

Argonne National Laboratory

Argonne, IL

TABLE OF CONTENTS

Page

Technical Session 1 - Superconductors

PHOTOTHERMAL MEASUREMENTS OF SUPERCONDUCTORS 1
G.S. Kino, W.R. Studenmund, and I.M. Fishman (*Stanford Univ., Stanford, CA*)

SCANNED PROBE MICROSCOPY FOR THIN FILM SUPERCONDUCTOR DEVELOPMENT 9
J. Moreland (*NIST, Boulder, CO*)

PROTECTION OF HIGH TEMPERATURE SUPERCONDUCTING THIN-FILMS IN A SEMICONDUCTOR PROCESSING ENVIRONMENT 17
Y. Xu, R. Fiske, S.C. Sanders, and J.W. Ekin (*NIST, Boulder, CO*)

CALCULATION AND MEASUREMENT OF THE CRITICAL STATE IN SUPERCONDUCTING TAPES 27
K.L. Telschow and L.S. Koo (*INEL, Idaho Falls, ID*)

Technical Session 2 - Materials I

SURFACE AND DEFECT MORPHOLOGIES IN ANISOTROPIC ELASTIC AND PIEZOELECTRIC SOLIDS 35
H. Gao and D.M. Barnett (*Stanford Univ., Stanford, CA*)

ELASTIC-PLASTIC FRACTURE MECHANICS OF STRENGTH-MISMATCHING 43
D.M. Parks, S. Ganti, and F.A. McClintock (*MIT, Cambridge, MA*)

DAMAGE MECHANICS - FAILURE MODES 52
D. Krajcinovic and M. Vujosevic (*Arizona State Univ., Tempe, AZ*)

Technical Session 3 - Materials II

T^*_ε INTEGRAL ANALYSIS OF FRACTURE SPECIMENS 60
Y. Omori (*Univ. of Washington, Seattle, WA*), K. Perry, Jr., J.S. Epstein (*INEL, Idaho Falls, ID*), L. Ma (*Univ. of Washington, Seattle, WA*), H. Okada, S.N. Atluri (*Georgia Inst. of Tech., Atlanta, GA*), and A.S. Kobayashi (*Univ. of Washington, Seattle, WA*)

MICROSCOPIC INTERFACIAL PHENOMENA DURING FLOW IN POROUS MEDIA 67
M.J. Miksis and M.P. Ida (*Northwestern Univ., Evanston, IL*)

DETERMINATION OF NEAR-SURFACE MATERIAL PROPERTIES BY LINE-FOCUS ACOUSTIC MICROSCOPY 75
J.D. Achenbach and W. Li (*Northwestern Univ., Evanston, IL*)

CRYOTRIBOLOGY OF DIAMOND AND GRAPHITE 83
Y. Iwasa, A. Ashaboglu, and E.R. Rabinowicz (*MIT, Cambridge, MA*)

Technical Session 4 - Controls

THE TEMPORAL NATURE OF FORCES ACTING ON METAL DROPS IN GAS METAL ARC WELDING 91
L.A. Jones, T.W. Eagar, and J.H. Lang (*MIT, Cambridge, MA*)

PARTIAL CONTROL OF COMPLEX SYSTEMS WITH APPLICATION TO THE FLUIDIZED CATALYTIC CRACKER 98
I.H. Rinard and R. Shinnar (*City College of CUNY, New York, NY*)

NON-LINEAR SYSTEM IDENTIFICATION IN FLOW-INDUCED VIBRATION 106
P.D. Spanos, B.A. Zeldin (*Rice Univ., Houston, TX*), and R. Lu (*Hudson Engineering Corporation, Houston, TX*)

NONLINEAR DIFFUSION AND SUPERCONDUCTING HYSTERESIS 114
I.D. Mayergoyz (*Univ. of Maryland, College Park, MD*)

Technical Session 5 - Fluid Mechanics I

MICROSTRUCTURE AND MOMENTUM TRANSPORT IN CONCENTRATED SUSPENSIONS 124
L.A. Mondy (*Sandia National Laboratories, Albuquerque, NM*), A.L. Graham (*Los Alamos National Laboratory, Los Alamos, NM*), and H. Brenner (*MIT, Cambridge, MA*)

SHEAR-INDUCED PARTICLE DIFFUSION AND ITS EFFECT ON THE FLOW OF CONCENTRATED SUSPENSIONS 136
A. Acrivos (*City College of CUNY, New York, NY*)

NEW DEVELOPMENTS IN THE THEORY OF FLAME PROPAGATION 142
G.I. Sivashinsky (*City College of CUNY, New York, NY*)

ONE-DIMENSIONAL TURBULENCE 150
A.R. Kerstein (*Sandia National Laboratories, Livermore, CA*)

Technical Session 6 - Fluid Mechanics II

COUPLED PARTICLE DISPERSION BY THREE-DIMENSIONAL VORTEX STRUCTURES 158
T.R. Troutt, J.N. Chung, and C.T. Crowe (*Washington State Univ., Pullman, WA*)

EXPERIMENTAL STUDIES OF REYNOLDS NUMBER DEPENDENCE OF TURBULENT MIXING AND TRANSPORT 166
Z. Warhaft (*Cornell Univ., Ithaca, NY*)

CONDUCTION IN NON-UNIFORM COMPOSITES 174
A. Prosperetti and M. Marchioro (*The Johns Hopkins Univ., Baltimore, MD*)

Technical Session 7 - Thin Films

FILM FLOWS AND SELF-ORGANIZED PATTERNS OF 2-D LOCALIZED STRUCTURES 182
A.L. Frenkel (*Univ. of Alabama, Tuscaloosa, AL*)

2-D TRAVELING-WAVE PATTERNS IN BINARY FLUID CONVECTION 191
C.M. Surko and A. La Porta (*Univ. of California, San Diego, La Jolla, CA*)

IMAGE PROCESSING TECHNIQUES FOR MEASURING NONUNIFORM FILM THICKNESS PROFILES 199
S.V. Nitta, A-H. Liu, J.L. Plawsky, and P.C. Wayner, Jr. (*Rensselaer Polytechnic Inst., Troy, NY)*

KINETIC ANALYSIS OF COMPLEX METABOLIC NETWORKS 207
G. Stephanopoulos (*MIT, Cambridge, MA*)

PHOTOTHERMAL MEASUREMENTS OF SUPERCONDUCTORS

G. S. Kino, W. R. Studenmund and I. M. Fishman

Ginzton Laboratory, Stanford University
Stanford, California 94305-4085

ABSTRACT

A photothermal technique has been used to measure diffusion and critical temperature in high temperature superconductors. The technique is particularly suitable for determining material quality and inhomogeneity.

INTRODUCTION

We have been developing a new photothermal microscope technique for studying high-temperature superconductors. The microscope measures differential optical reflectivity (DOR) and thermal diffusivity in thin films and single crystals of cuprate superconductors. Both normal and superconducting phases are characterized. The spatial resolution of the DOR technique is comparable to an optical microscope and allows us to measure small samples several microns in extent. The photothermal technique is extremely sensitive to manifestations of optical and thermal parameters affected by phase transitions, and in particular, to the superconductive transition in high-T_c materials.

Modern technology is able to produce very-high-quality, high-temperature superconducting samples. However, the best available high-T_c samples have a single domain size (~100 μm). Traditional methods of sample characterization (DC and microwave techniques) used for measurements of critical temperature, critical current, and material uniformity have a spatial resolution of ~1 mm.[1,2,3] With our technique, one can measure optical reflectivity and critical temperature with the spatial resolution of an optical microscope (1-2 μm). We can determine the variation of the critical temperature from point to point, and measure the quality of the material, by determining its diffusivity and differential optical reflectivity as a function of temperature. We can also determine the anisotropy of the material in small regions, and hence determine whether it is a single crystal or twinned. Thus, we believe that we have arrived at an excellent method for testing the quality of high-temperature superconducting materials.

For the first time, we used a polarized light probe to observe the anisotropy of single crystals. With it we observe an extremely sharp peak in the probe DOR signal near the critical temperature when the light is polarized along the a axis, and a very different characteristic output as a function of temperature when the light is polarized along the b axis.

THE EXPERIMENT

The basic microscope configuration is shown in Fig. 1. A He-Ne laser (632 nm), acousto-optically modulated by a Bragg cell, is focused on the sample and used as a heating beam; this beam periodically modulates the temperature of the sample at the focus and excites a thermal wave with a diffusion length of the order of 50 μm. Within a few microns from the focus of the heating beam, an infrared semiconductor laser (780 nm) is focused on the sample to probe the reflectivity change due to the temperature variation caused by the thermal wave. We have chosen a semiconductor laser for its minimum noise

and select, with a dichroic (lowpass) filter, the reflected probe laser light which impinges on a PIN photodiode. The phase and amplitude of the modulated signal are measured with a lock-in amplifier.[4]

The photothermal microscope designed for measuring high-temperature superconductors consists of three main components: (1) the optical system of the microscope, which excites and detects the thermal wave, (2) the temperature-controlled cryostat, which houses the sample and carries out the temperature sweep necessary for the measurement, (3) an auto-focusing feedback system, which ensures that the sample always stays at the focal plane of the microscope objective during a temperature sweep, and (4) computer control.[4]

We have combined a photothermal microscope with a high-vacuum cryostat to measure the thermal diffusivity of high-T_C superconductors as a function of temperature. The micron resolution obtained by using such a system allows us to measure single-crystal samples which are too small or irregular to be measured by standard bulk techniques. Therefore, we have the ability to measure the full anisotropic conductivity and measure diffusion within a single crystal or across a grain boundary.

For highest-resolution imaging, we make the two laser beams coincident and measure the amplitude of the reflectivity change. In this manner we can obtain resolutions limited only by the spot size of the microscope objective used for the focusing. For quantitative measurements of the thermal diffusivity in a specific direction, we separate the heating and probe beams by a known distance and measure the thermal phase lag between the two. For distances greater than a few spot widths, the phase lag between two points a distance x apart is $\phi_x = x\sqrt{2\omega / D_x}$, where ω is the heating frequency and D_x is the thermal diffusivity in the x direction.[5] Measurement of thermal phase has the great advantage that it is insensitive to nonthermal effects, such as laser intensities and the variation of refractive index with temperature.

The vacuum inside the cryostat is controlled to be under 2×10^{-7} torr to limit ice on the sample surface to less than 1 μm during the two hours needed for the measurement. A vac ion pump is used with a

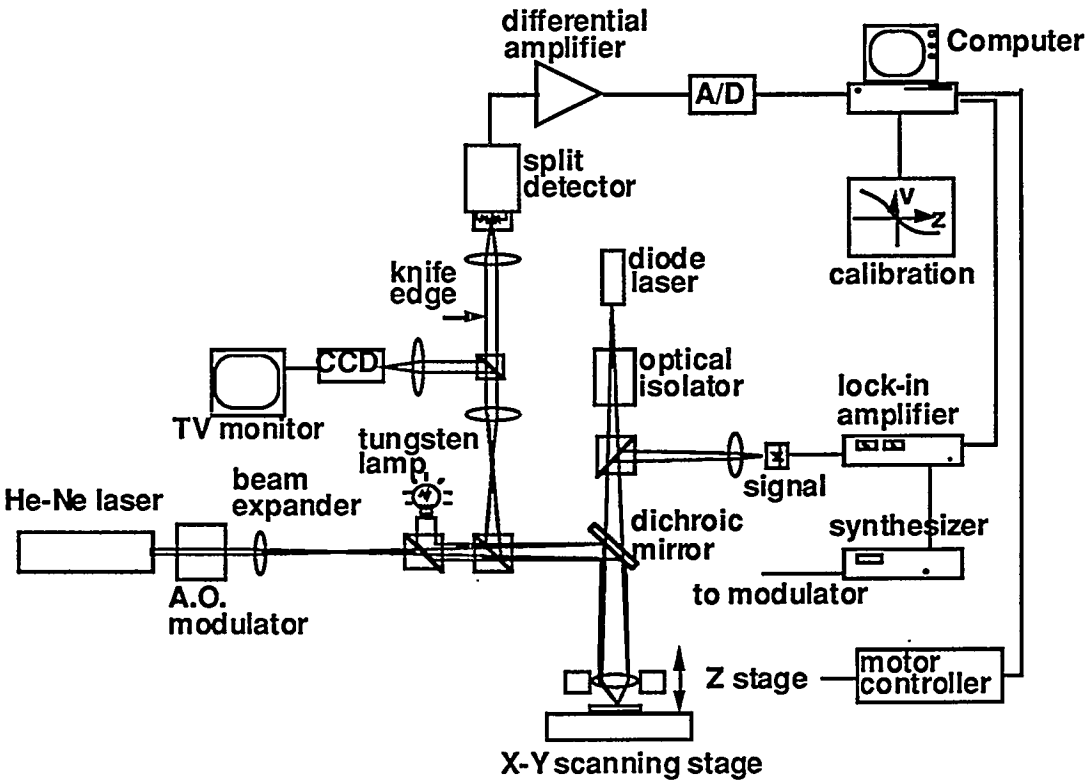


Fig. 1. System for photothermal measurements of superconductors.

gettering pump because of its small size, low weight, low ultimate vacuum, and oil-free operation. This ensures a system free of vibration, and makes it possible to mount the cryostat on a micropositioning system. The focused light beams are incident on the sample passing through a long working distance (7 mm) 0.3 NA objective and a 100 μm thick silica window mounted on the vacuum housing which is made thin to reduce aberration of the focused light.

Special precautions have been taken to ensure an accurate temperature reading of the sample. We limit the power of both heating and probe lasers to only a few tens of microwatts. A three-dimensional theoretical simulation indicates that the dc rise in the surface temperature of the sample caused by laser heating by the pump is less than 2 K, and that by the probe is less than 0.5 K. The temperature reading of a silicon sensor attached to the sample mount is verified with an in-situ ac susceptibility measurement. Underneath the sample, two coils are embedded in a sapphire sample mount, which provides electrical insulation from, but a good thermal conducting path to, the copper block of the cryostat. The larger coil is driven with an ac current; the voltage induced in the smaller coil is measured with a lock-in amplifier. A sharp change in the induced voltage indicates the superconducting transition and calibrates the reading of the silicon sensor.

During the cooling or heating of the cold finger, the tip of the stage tends to move as a result of mechanical contraction or expansion. Large vertical movement of the sample can result in a loss of focus and cause an error in the photothermal signal. We therefore incorporate an autofocusing mechanism to adjust the z-position of the objective and to ensure that the sample surface is always in the focal plane of the objective. A focusing accuracy of 1 μm is easily achieved in our measurements; the error is well within the depth of focus (5 μm) of the 0.33 numerical aperture objective.

The lateral movement of the sample, too, can cause noise in the signal output because of the nonuniformity of the sample reflectance. Therefore, the x- and y-directional movement of the sample is monitored with the CCD camera and is compensated for manually with a motorized stage. The accurate lateral definition (±3 μm) of the experiment allows us to achieve measurements of the thermal diffusivity within one domain of single-crystal high-T_C superconductors.

Phase Measurements of Diffusion: Measurements of the diffusion as a function of temperature in YBCO are shown in Fig. 2. A major advantage of measuring within a small volume is demonstrated by these results, where the diffusivity measurements were conducted for a 10 μm spacing between the excitation and the probe areas in a single YBCO domain, and across the twin boundary. For comparison, the diffusivity measurements for thin YBCO films using another optical technique, the transient grating method, in a region 200 μm across, are also shown. It will be seen that, above the transition temperature, the thermal diffusivity decreases slowly with temperature. The diffusivity measured over large-area thin film samples (200 μm across), and across a grain boundary, agree with bulk measurements, but the diffusivity measured within a grain is approximately twice as large as the measured value across a grain boundary. This is an interesting result, for it indicates that there must be a large change in temperature across the grain boundary, and tends to indicate that phonons (presumably optical phonons) and normal electrons (which contribute to diffusion) are reflected at the grain boundary.[6]

Below T_C, the data for different samples differs dramatically. For the single domain for $T < T_C$, diffusivity increases sharply over two orders of magnitude. This diffusivity enhancement can be explained only by removal of free carriers from the conducting state into the superconducting condensate. For samples of lower quality, the enhancement is not that sharp, especially for the thin film sample where the superconducting transition has almost no effect upon the diffusivity. The measurement through the twin boundary is intermediate between the other two curves. The twin boundary, though thin compared to the beam spacing, creates significant additional resistance for heat transfer. The diffusivity enhancement is evidence of the phonon mechanism of heat transfer inside a single domain. This conclusion does not differ from the traditional viewpoint on heat transfer in high-T_C materials. In thin films, additional mechanisms of phonon scattering, such as scattering by point defects independent of the free carrier density, probably dominate. The experimental results obtained show very conclusively that the scattering rate of phonons is proportional to the number density of normal electrons at temperatures below T_C.[1,6,7,8]

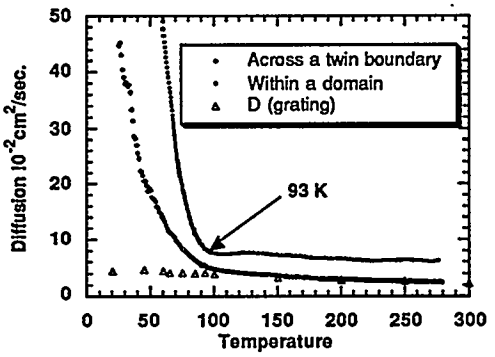


Fig. 2. Measured diffusion in YBCO.

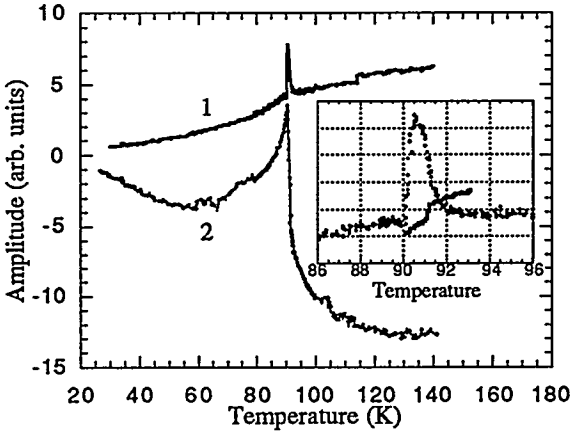


Fig. 3. Temperature dependence of differential reflectivity for the probe beam polarization along the *a* axis (1), and along the *b* axis (2), of a single YBCO domain. Inset: curve 1 in the vicinity of T_C together with the inductive coil response indicating the transition temperature.

Furthermore, the number density of normal electrons varies with temperature much like the predictions of BCS theory, although there is a considerable difference between the experiment and theory near T_C (see Fig. 4).

Differential Optical Reflectivity (DOR) Measurements: In the visible range, the optical constants of high-T_C superconducting materials do not show any noticeable temperature dependence.[9] However, in the vicinity of the critical temperature T_C, the perturbation caused by the transition, though small, is rather sharp and may be detected by a differential technique. Here, we will describe photothermal measurements of differential optical reflectivity (DOR) of single $YBa_2Cu_3O_{7-x}$ crystals[4] and the temperature variation of the thermal wave amplitude response. It is worth mentioning that, though the amplitude and the phase of the thermal wave carry different information, they are obtained in the same experiment.

One set of YBCO samples used in this study was twinned, with critical temperatures in the range of 60-93 K. The modulated signal did not show any anisotropy and polarization dependence. Another set of

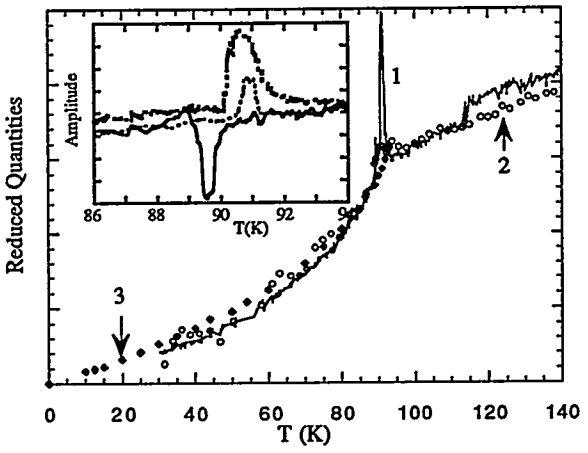


Fig. 4. Comparison of temperature-dependent DOR response in the "a" polarization (curve 1, same as curve 1 in Fig. 1), data for thermal resistance (curve 2, open circles) calculated from thermal wave phase,[6] and for the normal component density (curve 3, solid squares) calculated from microwave measurements.[14] Inset: DOR peaks in the vicinity of T_C for the first sample (top curve, same as curve 1 in this figure), and for two domains of the second sample.

measurements were made with far higher quality single crystals with domain sizes in the 200-500 μm range. In this case we used polarized light for the probe and were able to rotate the angle of polarization of the incident light. Two samples investigated in this study were single YBCO crystals with a critical temperature of 91.2 K. Both samples had comparatively large untwinned areas of approximately 50×200 μm. These areas were imaged under a polarizing microscope as dark and light domains. All measurements were conducted in the *a-b* plane. Standard reflectivity measurements with polarized light on both samples show distinct anisotropy of the reflectivity R in the *a-b* plane, described by:

$$R = A + B\cos 2\theta \qquad (1)$$

with a B/A ratio of 2.56, which is in good agreement with the results obtained on similar samples by Cooper et al[9] [for an optical energy of ~1.5 eV]. This behavior is expected for the reflection from an *a-b* plane with enhanced polarizability along the Cu-O chains oriented along the *b* direction.

Like the results obtained previously on heavily-twinned YBCO samples, the single-domain DOR data for a given probe polarization in Fig. 3 consists of a continuous smooth "background" in the entire temperature range 20-300 K, and a sharp peak in the vicinity of T_C. In earlier measurements we were able to show that the smooth background appears to be associated with the normal component.[6] For a peak-to-peak temperature modulation of approximately 0.2 K, the maximum observed signals are about 10^{-4} of the average reflected intensity of the probe beam. Integration of the DOR response with respect to temperature shows that R changes by less than 1% in the temperature range 20-120 K.

Over most of the temperature range, as might be expected for an anisotropic material with zero off-diagonal terms in the dielectric or reflectivity matrices associated with optical rotation, the DOR data shows angular anisotropy of the same form as in standard reflectivity measurements,[9] but the coefficients A and B vary rapidly with temperature near the critical temperature. Above the critical temperature the DOR response is a weak function of temperature. As we have already discussed, the angular dependence of the DOR signal is of the same form as in the standard reflectivity measurements. However, unlike the results obtained with our earlier measurements of highly-twinned samples, when the polarization vector rotates from *a* to *b*, the DOR signal changes sign,[10] as may be seen in curves 1 and 2 of Fig. 3.

To describe this sign change and some other features of the DOR signal, we use a two-component Drude model[11,12] with a dielectric constant ε of the form:

$$\varepsilon = 1 - \frac{(1-f)\omega_p^2}{\omega^2 + i\Gamma_1\omega} - \frac{f\omega_p^2}{\omega^2 + i\Gamma_2\omega} \tag{2}$$

where ω_p, Γ_1 and Γ_2 are the plasma frequency and electron collision rates for the normal and superconducting phases, respectively, and f is the fraction of superconducting electrons ($f = 0$ for $T > T_c$). The reflection coefficient is

$$R = \left|(\sqrt{\varepsilon} - 1)/(\sqrt{\varepsilon} + 1)\right|^2$$

For small $\Gamma \ll \omega$, $\Gamma \ll \omega_p$, the DOR response takes the form:

$$\frac{dR}{dT} = \frac{2R_0\Gamma_1}{\omega^2 - \omega_p^2}\left[(1-f)\frac{d(\delta\Gamma)}{dT} - \frac{df}{dT}\delta\Gamma\right] \tag{3}$$

where $\delta\Gamma = (\Gamma_1 - \Gamma_2) \ll \Gamma_1$ or Γ_2, $|\omega - \omega_p| \gg \Gamma_1$ or Γ_2, R_0 is the reflection coefficient for an unmodulated probe beam, and it is assumed that $|d\Gamma_2/dT| \ll |d(\delta\Gamma)/dT|$. The DOR signal sign depends on the relation between the incident frequency ω and the plasma frequency ω_p. From comparison with the standard reflectivity measurements in single domains, we expect that the probe frequency ($\lambda = 0.78$ μm) is positioned between the plasma frequencies for the a and b directions.

Beyond the sign reversal, Eq. (3) also predicts that the DOR signal consists of two components proportional to the rates of change of $\delta\Gamma$ and the fraction of superconducting phase f with temperature, respectively. In agreement with this prediction, the DOR response consists of a continuous smooth "background" in the entire temperature range 20-300 K, and a sharp peak in the vicinity of T_c. The background is associated with the normal component response. Above T_c ($f = 0$); only this component is present in the DOR signal. Below T_c the background amplitude decreases and at $T \sim 20K$ practically disappears. Equation (3) predicts that below T_c the first term should decrease with temperature as (1 - f), or the density of the normal component. This prediction allows for comparison with other experiments where the density of the normal component was measured.

In Fig. 4, the densities of the normal component derived from microwave penetration depth (curve 2)[12] and thermal diffusivity (curve 3)[6] measurements are compared with the DOR signal for the "a" polarization (curve 1). The comparison shows similarity of the temperature dependencies for all three measurements, which confirms that the Drude model may be applied, at least for qualitative analysis of what appears to be the normal component of the DOR data. Some noticeable differences may be explained by the temperature dependence of $d(\delta\Gamma)/dT$ below T_c.

We can therefore associate the "smooth" parts of the curves with normal electrons, but the peaks with the superconductive carriers. At the present time, it is difficult to give a good explanation of the behavior near the peaks. Fluctuation theory seems to be inadequate for the purpose, as is BCS theory, and we do not yet have a good explanation of why the peak can change in sign, as shown in the inset of Fig. 4.

Measurement of the Periodic Temperature Variation: Differential optical reflectivity data taken on single-domain samples shows very sharp functions of temperature. Since the resolution of the method is dependent on the variation of temperature caused by the pump heating, there has been some question as to whether our assumption of a small temperature variation is adequate. Our theoretical estimates predicted a 0.2 K variation at the probe. However, there was no direct experimental evidence that our evaluations are correct. We, therefore, developed a method for determining AC temperature variation, based on comparison of the amplitudes of the fundamental (ω) and second harmonic (2ω) components in the reflected signal. The reflection coefficient R may be written as a Taylor expansion in terms of the temperature variation ΔT:

$$R = R_0 + \frac{dR}{dT}\Delta T + \frac{d^2R}{dT^2}\frac{(\Delta T)^2}{2} + \ldots \tag{4}$$

where the periodic variation of temperature ΔT is $\Delta T = \theta \cos \omega t$. It then follows from Eq. (4) that

$$R \approx R_0 + \frac{dR}{dT}\theta \cos \omega t + \frac{d^2R}{dT^2}\frac{\theta^2}{4}\cos 2\omega t \tag{5}$$

The last relation provides a convenient method of measuring the amplitude of the periodic temperature variation. The ω component is proportional to $\theta\, dR/dT$; the 2ω component is proportional to $(\theta^2/4)d^2R/dT^2$. Integrating the 2ω component, we can obtain the value of $(\theta^2/4)dR/dT$. Comparison with the coefficient of the ω component provides a direct measurement of θ. In Fig. 5 the results of these measurements are shown; curve (a) is the first harmonic signal and curve (b) shows the integral of the 2nd harmonic signal over temperature. We measured the small amount of 2nd harmonic in the input signal due to the chopper not being entirely symmetric, and arrived at a value of 0.01 of the 1st harmonic. By taking account of this leakage and the larg2er rate of attenuation of the 2nd harmonic signal, we corrected curve (b) by subtraction of 0.005 of the 1st harmonic signal from it before integration. This process yielded curve (c). It will be seen that two curves representing the differential reflectivity at ω and the integral of the data at 2ω have similar temperature dependences, which agrees with the results of Eqs. (2) and (3).

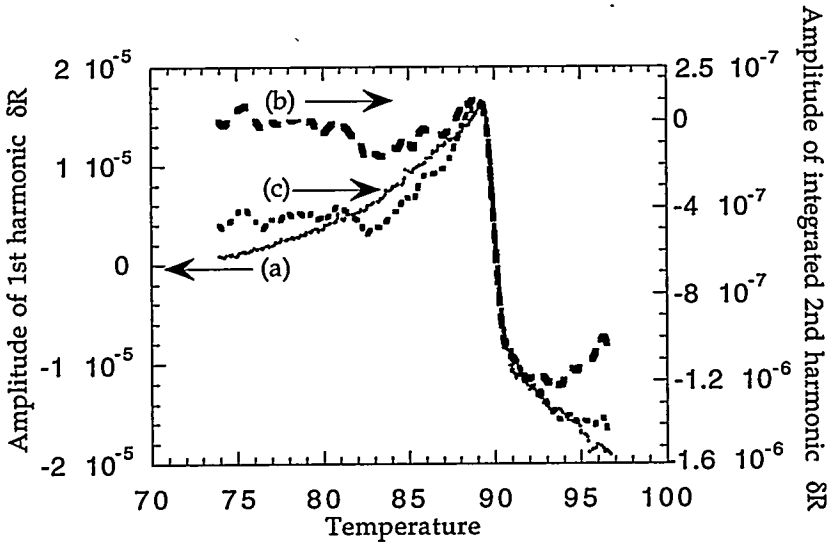


Fig. 5. Comparison of (a) 1st harmonic, (b) integrated 2nd harmonic signal, and (c) integrated 2nd harmonic signal with 0.005 leakage of 1 - f signal with polarization along the *b* axis.

CONCLUSION

The photothermal technique provides information on inhomogeneity, the value of the critical temperature in small regions, and measurements of diffusion (and hence the normal carrier density below the critical temperature). New information has been obtained on the behavior of highly anisotropic single crystals near the critical temperature, and on the temperature variation caused by the measurement process itself. The apparatus is being redesigned to make it more stable and easy to use so that materials researchers can make routine use of it.

ACKNOWLEDGMENT

This work was performed under the auspices of the U.S. Department of Energy. The authors would like to thank J. Giapintzakis of D. Ginsberg's laboratory for the supply of high-quality samples of YBCO, and W. Hardy, and S. Doniach for many useful discussions.

REFERENCES

1. C. Uher, J. Supercond., 3, 337 (1990).

2. R. C. YU, M. B. SALAMON, JIAN PING LU, AND W. C. LEE, Phys. Rev. Lett. 69, 1431 (1992).

3. J. L. COHN, E. F. SKELTON, S. A. WOLF, J. Z. LIU, AND R. N. SHELTON, Phys. Rev. B45, 13144 (1992).

4. X. D. WU, J. T. FANTON, G. S. KINO, AND A. KAPITULNIK, "A Photothermal Microscope for High-T_c Superconductors and Charge Density Waves," Rev. of Scientific Instruments 64, 3321-3327 (1993).

5. J. G. FANTON, D. B. MITZI, A. KAPITULNIK, B. T. KHURI-YAKUB, G. S. KINO, D. GAZIT, AND R. S. FEIGELSON, "Photothermal Measurements of High-T_c Superconductors," Appl. Phys. Lett. **55** (6), 598-599 (7 August 1989)

6. I. M. FISHMAN, G. S. KINO, AND X. D. WU, "Density of Normal Carriers Below T_C and Thermal Resistance of Twin Boundaries in $YBa_2Cu_3O_{7-x}$ Single Crystals," Phys. Rev. B (Condensed Matter) **50**, 7192-4 (1994).

7. L. TEWORDT AND TH. WOLKHAUSEN, Solid State Comm. 70, 839 (1989).

8. L. TEWORDT AND TH. WOLKHAUSEN, Solid State Comm. 75, 515 (1990).

9. S . L. COOPER, D. REZNIK, A. KOTZ, M. A. KARLOW, R. LIU, M. V. KLEIN, W. C. LEE, J. GIAPINTZAKIS, D. M. GINSBERG, B. W. VEAL AND A. P. PAULIKAS, Phys. Rev. B 47, 8233 (1993).

10. I. M. FISHMAN, G. S. KINO, X. D. WU, AND W. R. STUDENMUND, A. A. ABRIKOSOV, A. A. VARLAMOV, AND A. I. BUZDIN, Physica C 249, 304 (1995).

11. T. TIMUSK AND D. B. TANNER, in *Physical Properties of High Temperature Superconductors* I, Ed: D. M. Ginsberg, p.339 (World Scientific, Singapore, 1989).

12. W. N. HARDY, D. A. BONN, D. C. MORGAN, RUIXING LIANG, AND KUAN ZHANG, Phys. Rev. Lett. 70, 3999 (1993).

SCANNED PROBE MICROSCOPY FOR THIN FILM SUPERCONDUCTOR DEVELOPMENT *

John Moreland

Electromagnetic Technology Division
National Institute of Standards and Technology
Boulder, Colorado 80303, USA

ABSTRACT

Scanned probe microscopy is a general term encompassing the science of imaging based on piezoelectric driven probes for measuring local changes in nanoscale properties of materials and devices. Techniques like scanning tunneling microscopy, atomic force microscopy, and scanning potentiometry are becoming common tools in the production and development labs in the semiconductor industry. We present several examples of applications specific to the development of high temperature superconducting thin films and thin-film devices.

INTRODUCTION

Rapid advances in high temperature superconductor (HTS) technology have lead researchers to consider replacing normal metals in several electronic applications including transmission lines, filters, and on-chip as well as multi-chip module interconnects. HTS filters and interconnects have the advantages of low-losses and frequency independent dispersion. In particular, Ajimine, *et al.*[1] have demonstrated the usefulness of a yttria stabilized zirconia (YSZ) layer as a gate insulator between $YBa_2Cu_3O_x$ (YBCO) films and Si substrates. Harvey, *et al.*[2] have made low resistivity contact structures between YSZ/YBCO/Ag trilayers and Si using ion-milled, Au-filled vias with contact resistance's of 10^{-6} Ω-cm^2. Ekin, *et al.*[3] have made significant progress towards optimizing YBCO contact processes that would be compatible with Si integrated circuit processing.

It is clear that HTS thin-film properties are affected by nanometer-scale morphology and electronic processes which directly or indirectly influence the performance of various HTS interconnect devices. Measurements of roughness, critical dimensions, field patterns, and local electronic processes provide information about the

fundamental operation and ultimate performance limitations and are thus critical in the development process. Scanned probe microscopies[4] (SPM) such as scanning tunneling microscopy (STM),[4] atomic force microscopy (AFM),[5] and scanning potentiometry[6] are examples of imaging techniques that are uniquely qualified for these applications because of the nanometer-scale dimensions of the various scanning probes. The purpose of this paper is to demonstrate the usefulness of SPM technology in the development of HTS interconnect devices by presenting several example applications.

MORPHOLOGY AND SURFACE DENSITY OF STATES OF YBCO FILMS

The growth habit of the YBCO films depends on the deposition technique and substrate material.[7,8,9] For example, Fig. 1a shows screw dislocation growth on the surface of a sputtered film on MgO as imaged by STM. Substrate dislocations seed YBCO spiral growth with an average spacing of 1 μm. Many spirals appear to merge into each other within the surface of a single grain of several micrometers in diameter.

The STM growth step height for all of the films measured at room temperature are anomalously large compared to the theoretical unit cell in the c-axis direction. We believe that variations in the surface barrier, density of states, or similar electronic effects cause these anomalies and that they do not represent the true topography of the samples. Hawley, *et al.* have seen similar anomalies in STM data.[10] They have shown, however, by performing AFM, which is relatively unaffected by the surface density of sates, that the true step height is about 1 nm as expected.

The STM step height is temperature dependent. As the temperature decreases from 300 K the growth step height decreases, approaching the expected unit cell value of 1.2 nm at 76. The origin of this striking temperature effect is uncertain. Perhaps there is a semiconducting layer consisting of de-oxygenated YBCO on the films that undergoes carrier freeze-out upon cooling. Carrier freeze out would have the effect of changing the density of states at the surface of the film. In the extreme case of complete freeze-out, the shape of the tunneling barrier between the film and the STM tip may be affected as the layer becomes insulating. Alternatively, at lower temperatures as the Cu-O planes decouple and conduction becomes more 2D in character, the rate for tunneling perpendicular to the Cu-O planes may be affected. Generally, measurements of the resistivity anisotropy for single crystals of YBCO support this possibility. Regardless of the cause, surface density of states of effects will dominate the interface transport properties of YBCO films which are critical to many prospective interconnect technologies.

Comparison of STM and AFM images of YBCO thin-film samples has lead to another interesting observation regarding HTS film growth. AFM images reveal a high density of small distinct nanoparticles, 10 to 50 nm across and 5-20 nm high, which do not appear in the STM images of the same samples as shown in Fig 1b.[11] During STM scanning of the sample, the STM tip breaks off the particles and moves them to the edge of the scanned area. The presence of the particles, which are thought to be yttria,[12] seems to be generic for many types of YBCO thin-films regardless of film thickness or substrate. One intriguing possibility is that the insulating nanoparticles, affectionately referred to as "sand" (as opposed to the much larger "boulders" often observed on laser ablated YBCO thin films), might be the natural pinning sights responsible for the high critical currents of these films. Thus far the microscopic (or nanoscopic, as the case may be) nature of the pining sites has eluded scientists. In any case, the knowledge of the presence of the sand is important for the development of YBCO thin films in multilayer structures for electronic applications of HTS films.

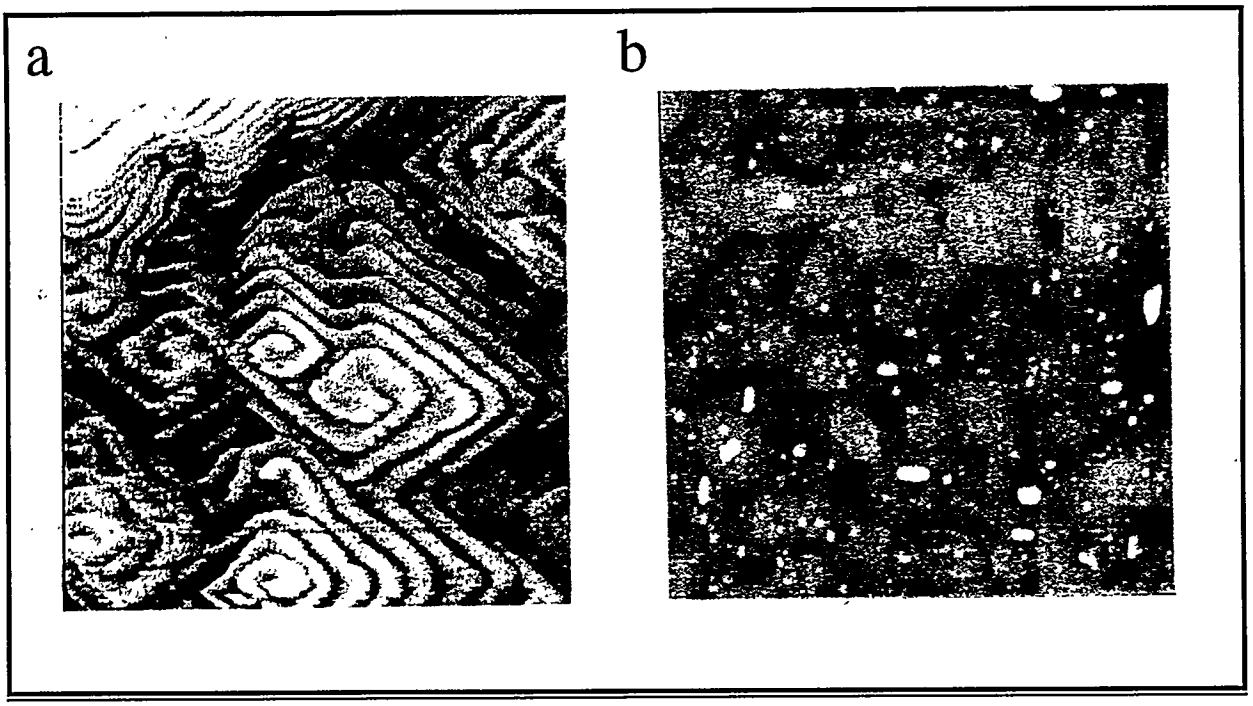


Fig. 1. a) 1 μm STM image of surface density of states "topography" of a YBCO film deposited on a MgO substrate (from ref.9). b) 2.5 μm AFM image of yttria sand on a YBCO film deposited on a $LaAlO_3$ substrate (from ref.11).

YBCO FILM GROWTH ON Si SUBSTRATES

We have studied YBCO thin films deposited on Si substrates.[13] The deposition method was developed by Fork *et al.*[14] and relies on a thin film of YSZ as a chemical buffer as well as a strain layer to reduce lattice-mismatch strain between the YBCO and the Si. R-T measurements of unpatterned films show T_c ranging from 85 to 87 K with a resistivity ratio R_{300K}/R_{100K} of about 2.5. Typically, J_c is 9 x 10^5 A/cm^2 at 4 K. X-ray diffraction shows that the YSZ and the YBCO are c-axis oriented with the YSZ layer having good epitaxy on the Si and the YBCO layer having both a-axis and b-axis oriented grains within the plane of the film.

One of the purposes of a YSZ layer is to prevent chemical reaction between the Si substrate and the YBCO film. Ideally, the YSZ layer should be continuous. However, even if there are pinholes in the YSZ, it is generally thought that they are small in diameter, few in number, and that the reaction product between the YBCO and Si is itself an insulator. This means that the pinholes cannot be detected electrically by measuring conduction leakage between the Si and YBCO. Figure 2a is an AFM image of a YSZ film that was removed from the vacuum system before depositing YBCO. Pinholes penetrating the film are readily visible. The pinhole density is 50/μm^2. The pinholes vary in size with the majority being about 15 nm in diameter. The largest pinhole observed was 44 nm and the smallest was 5 nm. This is direct proof of the existence of pinholes in YSZ buffer layers. Other significant aspects of Fig. 2a are that there is no evidence of cracking in the YSZ film and that, aside from the pinholes, the film is very flat with less than 1 nm roughness indicating good epitaxy on the Si.

Pinholes may be one of several factorsalong with YBCO film strain and cracking that may cause a reduction in T_c and J_c from the highest values observed for films deposited on $LaAlO_3$ of T_c = 91.5 K and

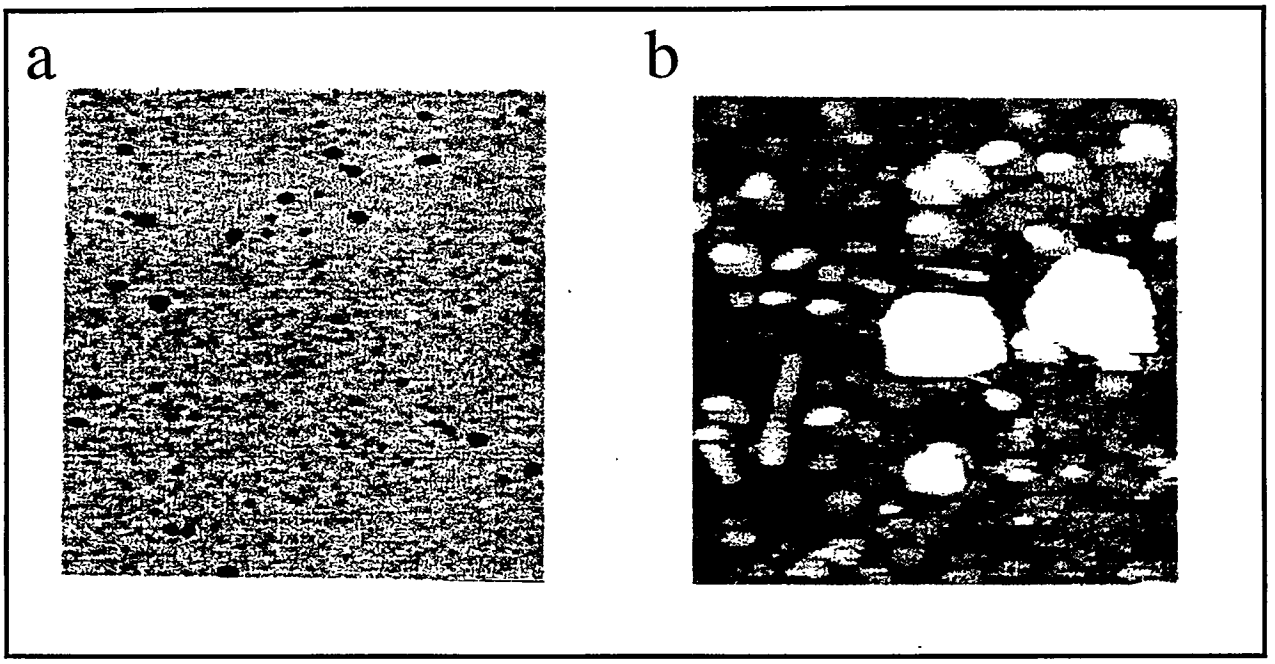


Fig. 2. a) 1 μm AFM image of a YSZ buffer layer on Si showing pinholes (from ref. 13). b) 1 μm AFM image of completed YBCO/YSZ buffer layer.

J_c(4 K) = 10^7 A/cm^2. An AFM image of the YBCO/YSZ bilayer is shown in Fig.2b. Large ablation boulders dominate the image. Between the boulders the film appears flat (roughness less than 10 nm) indicating predominate c-axis growth. We estimate the c-axis grain size to be 90 to 100 nm based on the average distance between roughness peaks. A few long, narrow a-axis grains are also present along with some round protrusions, 50 to 60 nm high and 150 nm in diameter. The density of the round protrusions is about 30/μm^2 which is close to the YSZ pinhole density derived from Fig. 1 of 50/μm^2. We submit that there may be a correlation between the round protrusions and the YSZ pinholes and that they may well be evidence of reaction between the YBCO and the Si.

SCANNING POTENTIOMETRY OF YBCO STEP EDGE JUNCTIONS

The AFM has been adapted for scanning potentiometry.[6] The main features of the modifications to the AFM include a laser beam bounce technique for measuring the deflection of the cantilever which is excited near its resonance frequency by ac capacitive forces between the tip and the sample. The resonant force is minimized by adjusting the dc voltage between the tip and the sample to be the same. This voltage adjustment is plotted as a function of x-y position giving a potential map of the surface. This technique is useful for measuring surface potential due to changes in work function, trapped charges, or current flowing through a device such as a contact via or along an interconnect with 50 nm spatial resolution and millivolt voltage resolution.

We have performed some tests at room temperature on step edge structures for superconductor-normal metal-superconductor (SNS) proximity effect devices. Here, YBCO is deposited at an angle over a previously ion milled step forming a break in the film. Au is then deposited from the other side completing

the junction. Most of the gold is ion milled away leaving a small amount between the banks of the YBCO films on either side of the step forming the junction.

Figure 3a shows the surface potential image of a step-edge junction energized with 10 mA. The corresponding topography is shown in Fig. 3b. Line scans as a function of position across the junction are shown in Fig.s 3b and 3c. The data shows a linear dependence of voltage along the line as expected. Also, a dip in the surface potential at the junction as well as peaks in the surface potential near laser ablation boulders can be seen in the potential images. As mentioned above the technique is sensitive to surface charging which may be the explanation for the dip near the junction. Charge accumulation at the bare interface may have an effect on the operation of these HTS devices.

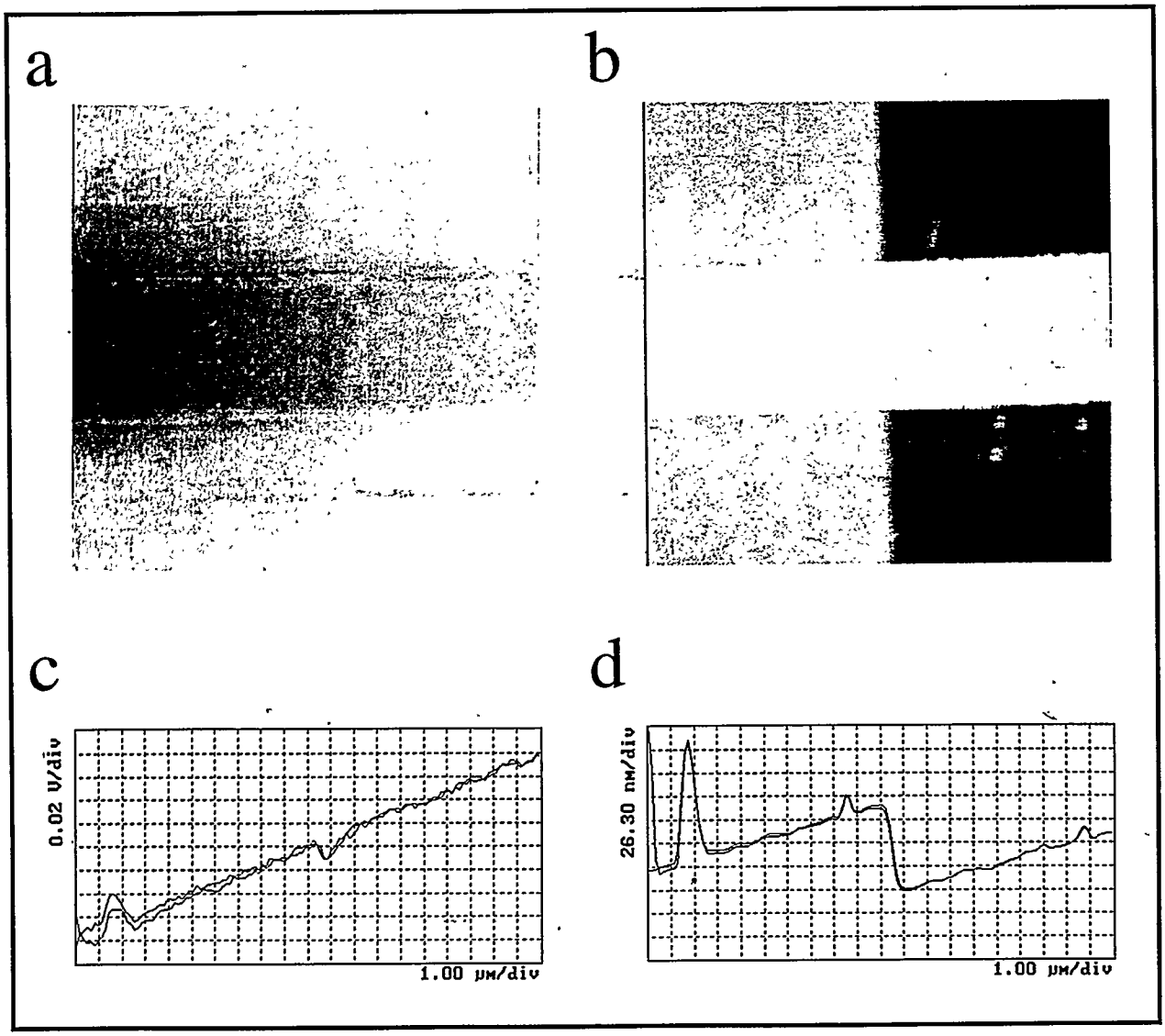


Fig. 3 a) Surface potential image of a YBCO step-edge junction at room temperature. b) Corresponding AFM topography for Fig. 3a. c) Line trace of surface potential vs distance across YBCO step-edge. d) Line trace of the z-height vs distance across the YBCO step-edge.

SUMMARY

In summary, we have shown that STM can be used to image differences in the grain morphology of as-grown, *in situ* YBCO films deposited on various substrates by sputtering or laser ablation. The newly developed long scanning range capabilities combined with the exquisite vertical sensitivity make the STM unique for this type of application. The images are useful for determining the growth habit on individual grains and may be helpful in determining grain boundary structures. We have also looked at STM images of sputtered YBCO films at various temperatures. The images show anomalously large step corrugation at room temperature compared to the expected unit cell step of 1.2 nm. The STM step height is temperature dependent, however, decreasing to close to the expected value at 76 K. Preliminary speculations about the cause of this phenomenon include the possible presence of a semiconducting layer at the surface of the film that undergoes carrier freeze-out or the inception of 2D localization of electrons within CuO planes at lower temperatures. In any case, the tunneling characteristics that effect the STM image are important for understanding the electrical properties of the surfaces of YBCO films. Such information is needed for development of Josephson tunneling barriers and electrical interconnect structures that would be useful for scaleable integrated circuit applications.

The nature and chemical composition of the sand particles are at this time unknown. They may stem from slight variations in YBCO stoichiometry. They are always present in high quality YBCO films made under a variety of deposition conditions. YBCO sand may have significant consequences for multilayer structures being considered in interconnect technologies. If these particles are distributed evenly throughout YBCO films they may occur at high enough volume densities to account for high J_c pinning in YBCO films.

AFM images of YSZ/YBCO bilayers on Si indicate the presence of pinholes in the YSZ layer bridging the Si and the YBCO. This leads to the subsequent formation of nonconducting phases near the pinholes since YBCO and Si react during high temperature processing steps. AFM images show scattered mounds on the completed YSZ/YBCO bilayers having about the same surface density as that of the pinholes. In the future we hope to apply the information gained from STM and AFM to optimize the transport properties of narrow superconducting YBCO lines on Si. Preliminary results show that it is possible make 0.25 μm lines which appear to be largely unharmed by e-beam lithography and subsequent ion milling.

AFM potentiometry based on the force-null detection scheme described above should be useful for studies of microelectronic devices offering better than 100 nm lateral resolution and better than 1 mV voltage resolution. Force-detection scanning potentiometers are sensitive to charging effects. We are developing a simple image subtraction process to minimize the effects of charging.. In addition, we have observed various charging affects including transfer of charge between the AFM tip and the sample. We have also observed switching of the charge state of the AFM cantilever between quasi-stable levels. Oxides on or near the surface of the tip may be responsible. Also, the tip may be picking up charged particles from the sample surface. Several labs have been experimenting with high bandwidth scanning potentiometry of semiconducting devices at frequencies as high as 100 GHz.[15,16,17] Ultimately our goal is to extend their work to low temperature studies of active HTS devices .

NEW DIRECTIONS: MAGNETIC RESONANCE IMAGING AT THE NANOMETER SCALE

Magnetic resonance imaging (MRI) which has the advantages of being non-destructive, chemically specific, and three-dimensional but it is unfortunately limited to the 10 μm spatial resolution level.[18] This barely qualifies MRI as a microscopic technique. Conventional MRI resolution is limited by relatively weak field gradients of order 10^{-1} T/m and the signal-to-noise ratio of the standard pick-up coil circuits

used to detect selected spins. Scanned probe microscopies (SPM), on the other hand, have high resolution, but only probe on or above surface structures. An example is magnetic force microscopy (MFM) which is a well developed SPM technique currently in use by the magnetic data storage industry. MFM routinely gives 50 nm spatial resolution images of the average fields above the surfaces of disk drive components. A new type of SPM related to MFM and referred to as magnetic resonance force microscopy (MRFM) has the main advantages of MRI of being chemically specific, three dimensional, and non-destructive in addition to the high resolution of SPM. MRFM was first proposed by Sidles.[19] It has the potential to greatly improve the current imaging technology for investigations of magnetic resonance phenomena down to the atomic scale.

In MRFM, a force microscope cantilever is used to detect the magnetic force exerted when spins are brought into resonance in response to an applied rf field. In theory, it has the sensitivity required to detect a single nuclear spin by exploiting the very large (~10^8 T/m) field gradient near a sharp magnetized tip similar to those used in MFM. In contrast to MRI, the large field gradients that produce the high spatial resolution also increase the intensity of the magnetic coupling signal between the cantilever and each individual spin residing in the selected slice. This combination is key to overcoming the resolution limits of conventional MRI. Additionally, the scanned probe used in MRFM simplifies the image reconstruction process normally required for conventional MRI.

MRFM differs from other types of SPM in its ability to probe sub-surface structure. Resonant coupling occurs when the Larmor condition is satisfied a points which lie on a surface *remote* from the tip of the probe where the magnetic field is at the appropriate value. This gives MRFM sub-surface sensitivity and a slice selectivity like that in conventional MRI experiments. Additionall, as in conventional magnetic resonance experiments, the spin precession frequency varies for different nuclei or compounds, thus giving MRFM its chemical specificity.

The general operating principles of MRFM are as follows: The spins in the sample are polarized by the magnetic field from the MFM tip, which, as discussed above, also produces the field gradients needed to generate a measurable magnetic force on the cantilever and provide spatial resolution for imaging. The spins in the sample are then resonantly excited by an applied rf field that is modulated at the resonant frequency of the cantilever which producing vibrational motion of the cantilever. Magnetic force maps are produced by measuring the vibrations of the cantilever attached to the sample while scanning the tip with respect to the sample. Real-space imaging information can be recovered from the force maps by using the usual deconvolution image reconstruction methods. The principles of MRFM have been demonstrated by Rugar, *et.al.*[20]. Their prototype microscope has achieved 1 μm resolution, which is 10 to 100 times better than conventional MRI instruments tailored for microscopy of small samples. While 1 μm resolution is far from the 10 pm resolution predicted by theory, the apparatus is relatively crude and not near the current limits of existing technology.

ACKNOWLEDGMENT

The author acknowledges the contributions of R. E. Thomson, R. H. Ono, A. Roshko, S. E. Russek, T. E. Harvey, J.W. Ekin D. A. Rudman, A. J. Nelson, B. Jeanneret, S. C. Sanders, P. Rice and C. Prater. JM was supported by DoE BES contract DE-AIO5-89ER14044

REFERENCES

1. E. M. AJIMINE, F. E. PAGADUAN, M. M. RAHMAN, C. Y. YANG, H. INOKAWA, AND D. E. FORK, AND T. E. GEBALLE, "Electrical Characterization Of Metal-Insulator-Semiconductor Diodes Fabricated From Laser Ablated $YBa_2Cu_3O_x$ /Yttria Stabilized Zirconia Films On Si Substrates," *Appl. Phys. Lett.* 59 , 2889 (1991).
2. T. E. HARVEY, J. MORELAND, B. JEANNERET, R. H. ONO, AND D. A. RUDMAN, "$YBa_2Cu_3O_x$ To Si Interconnection For Semiconductor/Superconductor Integration," *Appl. Phys. Lett.* 61, 225 (1992).
3. see this proceedings
4. H. J. GUNTHERODT AND R. WIESENDANGER, *Scanning Tunneling Microscopy,* Springer Verlag, New York, Springer Series on Surface Science vol.s 20, 28, and 29 (1994).
5. DROR SARID, *Scanning Force Microscopy*, Oxford, New York (1994).
6. Y. MARTIN, D. W. ABRAHAM, H. K. WICKRAMSINGHE, "High Resolution Capacitance Measurement And Potententiometrie By Force Microscopy," *Appl. Phys. Lett.* 52, 1103 (1988).
7. D. P. NORTON, D. H. LOWNDES, X. -Y. ZHENG, S ZHU, AND R. J. WARMACK, *Phys. Rev. B* 44, 9760 (1991).
8. I. D. RAISTRICK, M, HAWLEY, J. G. BEERY, F. H. GARZON, AND R. J. HOULTON, *Appl. Phys. Lett.* 59, 3177 (1991).
9. JOHN MORELAND, PAUL, RICE, S. E. RUSSEK, B. JEANNERET, A. ROSHKO, R. H. ONO, AND D. A. RUDMAN, "Scanning Tunneling Microscopy. Of The Surface Morphology Of $YBa_2Cu_3O_x$ Thin-Films Between 300 K And 76 K," *Appl. Phys. Lett.* 59, 3039 (1991).
10. M. HAWLEY, I. D. RAISTRICK, J. G. BERRY, AND R. J. HOULTON, "Growth Mechnism Of Sputtered Films Of YBCO Studied By Scanning Tunneling Microscopy," *Science* 251 , 1587 (1991).
11. R. E. THOMSON, J. MORELAND, N. MISSERT, D. A. RUDMAN, AND S. C. SANDERS, "Insulating Nanoparticles On $YBa_2Cu_3O_x$ Thin Films Revealed By Comparison Of Atomic Force And Scanning Tunneling Microscopy," *Appl. Phys. Lett.* 63, 614 (1993).
12. private communication J. Mannhart, University of Basel, Switzerland.
13. J. MORELAND, T. E. HARVEY, R. H. ONO, AND A. ROSHKO, "Scanned Probe Microscopy Of $YBa_2Cu_3O_x$ Thin-Film Device Structures On Si Substrates", *IEEE Trans. on Appl. Supercon.* 3, 1586 (1993).
14. D. K. FORK, D. B. FENNER, R. W. BARTON J. M. PHILLIPS, G.A.N. CONELL, J. B. BOYCE, AND T. H. GEBALLE, "High Critical Currents In Strained Epitaxial YBCO On Si", *Appl. Phys. Lett.* 57, 1161 (1990)
15. H. YOKOYAMA, T. INOUE, AND J. ITOH, "Nonresonant Detection Of Electric Force Gradients By Dynamic Force Microscopy, *Appl. Phys. Lett.* 65, 3143, (1994).
16. A. S. HOU, F. HO, AND D. M. BLOOM, "Picosecond Electrical Sampling Using A Scanning Force Microscope," *Electronics Lett.* 28, 2302 (1992).
17. C. BOHM, C. ROTHS, AND E. KUBALEK, "Scanning Force Microscope Test System For Internal Test With High Spatial And Temporal Resolution", *Microelectronic Engineering*, 24 91 (1994).
18. P. T. CALLAGHAN, *Principles Of Nuclear Magnetic Resonance Microscopy*, Oxford, New York, (1991).
19. J. A. SIDLES, "Noninductive Detection Of Single Proton Resonance", *Appl. Phys. Lett.* 58, 2854 (1991).
20. D. RUGAR, C. S. YANNONI, AND J. A. SIDLES, "Mechanical Detection Of Magnetic Resonance", *Nature* 360, 563 (1992).

PROTECTION OF HIGH TEMPERATURE SUPERCONDUCTING THIN-FILMS IN A SEMICONDUCTOR PROCESSING ENVIRONMENT

Yizi Xu, Richard Fiske, S.C. Sanders, and J.W.Ekin

Electromagnetic Technology Division
National Institute of Standards and Technology
Boulder, CO 80303, U.S.A.

ABSTRACT

Annealing studies have been carried out for high temperature superconductor $YBaCuO_{7-\delta}$ in a reducing ambient, in order to identify insulator layer(s) that will effectively protect the superconducting film in the hostile environment. While a layer of magnesium oxide (MgO) sputter deposited directly on $YBaCuO_{7-\delta}$ film provides some degree of protection, we found that a composite structure of YBCO/$SrTiO_3$/MgO, where the $SrTiO_3$ was grown by laser ablation immediately following YBCO deposition (in-situ process), was much more effective. We also address the need for a buffer layer between YBCO and aluminum (Al) during annealing. Al is most commenly used for semiconductor metalization, but is known to react readily with YBCO at elevated temperatures. We found that the most effective buffer layers are platinum (Pt) and gold/platinum (Au/Pt).

INTRODUCTION

Recently there have been several demonstrations of potential applications of superconductor-semiconductor integrated systems. One example is microwave filter subsystem made with high-temperature-superconductor (HTS) yttrium-barium-copper-oxide (YBCO) that can be used in cellular base stations[1]. Combined with cryogenic low-noise amplifiers (made from semiconductors), these compact HTS receiver filter subsystems provide much improved noise/interference rejection and hence higher service quality, more usable channels, and an extended coverage area. Another example of significant potential benefits of superconductor-semiconductor integration technology is the sensor subsystem used in magnetic resonance imaging (MRI) systems[2], currently the biggest commercial market for superconducting technology. The growing interest in low-cost, low-field MRI has created a challenge in MRI sensor technology that can only be met with the development of HTS-based low-noise receiver coils and their successful integration with low-noise semiconductor amplifiers operated at cryogenic temperatures.

The superconductor-semiconductor integration may be achieved in different levels, dependent upon particular applications[3]. At the system and subsystem level, semiconductor wafers are combined with chips carrying superconductor circuits, with each fabricated and processed separately under conditions that are optimized for each of them. This is the approach being followed now in the microwave filter applications. The next step of hybridization involves growing semiconductor and superconductor devices and circuits on the same wafer. More fundamental compatibility issues arise in this step, and it is with some of these issues we will be concerned in this work.

As an example consider the metalization and contact formation process. Aluminum is the most commonly used for metalization and interconnection in semiconductor integrated circuits. A process called contact sintering (or contact alloying) is used to form a reliable and intimate Al/Si

contact. This involves annealing the contact at temperatures close to 400 °C and in a reducing ambient, provided by the "forming gas", which is a mixture of 5-10% at. hydrogen and 95-90% at. nitrogen [4]. This typical semiconductor process will, however, seriously degrade the quality of YBCO film. The properties of HTS YBCO, such as T_c and the critical current density J_c, depends critically on its oxygen stoichiometry, which is readily depleted in such a process. Moreover, aluminum is known to react with YBCO at elevated temperatures by removing oxygen from the latter, thus compounding the problem of oxygen loss. Proper procedures must be taken to protect the YBCO circuits if the YBCO/Si hybrid at the chip level is to fit into the existing, well established semiconductor processing technology.

We have taken a two-step approach to identify these procedures. First we investigate noble metals that can serve as a buffer layer between YBCO and Al during annealing in oxygen atmosphere. The reason for using oxygen instead of forming gas, for example, is to differentiate the effects due to aluminum diffusion into and interact with YBCO, from that due to pure oxygen loss from the YBCO. Second we investigate insulating/dielectric materials that can be used to cover YBCO during argon gas annealing, to prevent oxygen out-diffusion due to the reducing ambient.

EXPERIMENTAL

Buffer Layer For Aluminum Diffusion During Oxygen Annealing

The HTS YBCO films used for aluminum diffusion buffer study were grown on (100) oriented MgO substrates using pulsed laser deposition. Briefly, the MgO substrate was heated to 730 °C, and a laser beam (wavelength=248 nm) with fluence of 2 mJ/cm^2 was focused at a target which was about 9 cm away from the substrate. The pulse rate was 10 pulses/sec and a total of 10,000 pulses gave a nominal thickness of 200 nm. The oxygen background pressure was 27 Pa (200 mTorr) during the deposition, and was brought to 500 Torr during cooling down after deposition. The YBCO films were predominantly c-axis oriented with a high degree of in-plane epitaxy, as revealed by x-ray diffraction. Critical temperatures were in the range of 85–90 K, and the J_c's were typically in excess of 10^6 A/cm^2 at 77 K.

Noble metal buffer layers were deposited on top of YBCO *in situ* (i.e., without breaking the vacuum) by either magnetron sputtering or thermal evaporation, immediately followed by deposition of a top layer of Al, using thermal evaporation. Unless otherwise stated, the Al, as well as each metal layer composing the buffer structure, was 50 nm thick.

The resistance-versus-temperature (R vs. T) measurements were taken in a standard four-probe configuration. The T_c was determined as the temperature at which the sample resistance dropped to 10% of the value at the onset of the transition.

The upper panel of Fig.1 shows the results of 300°C annealing in oxygen for different buffer layers. Buffer layers consisting of Au, Au/Cr, and Au/Ag resulted in the loss of superconductivity down to 76K. The Ag buffer layer was somewhat more effective, but showed a steady and significant decrease of T_c with annealing time. The buffer layers consisting of Pt or Au/Pt were the most effective as barriers for aluminum diffusion. In both cases, there was an increase in T_c after the first 30 minutes of annealing, possibly due to the ordering of initially disordered oxygen atoms on the Cu-O chain sites at this temperature. We also note that there was very little change in T_c beyond the initial 30 minutes, for times up to 2.5 hours.

Encouraged by these results, we raised oxygen annealing temperature to 400°C for samples having Pt and Au/Pt buffers. The results are shown in the lower panel of Fig.1. They are very similar to those displayed in the upper panel. However, 400°C is a typical temperature used in contact alloying in semiconductor process, therefore the significance of this result is that we have identified buffer layers to protect YBCO from aluminum diffusion that are effective at semiconductor processing temperatures.

We now turn to the second step, i.e., to replace the oxygenating ambient with a reducing ambient, e.g., argon, which is actually used in semiconductor contact alloying.

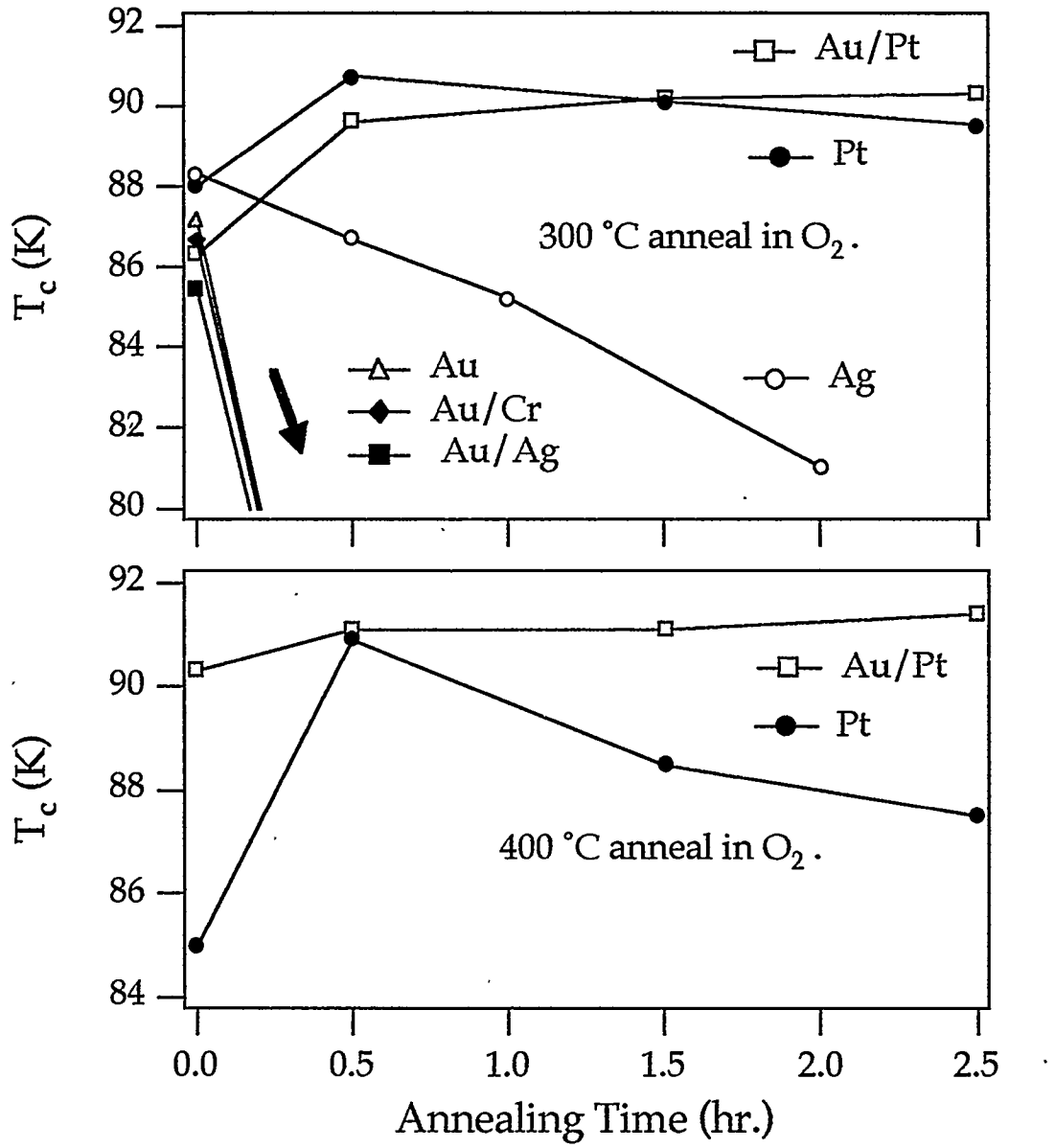


Fig. 1 T_c vs. annealing time of YBCO/buffer/aluminum structure for a number of buffer layers annealed in O_2. The buffer layers made of Au/Pt and Pt were the most effective.

Insulating Layer For Protection of YBCO During Argon Annealing

Here we focus on insulating materials, instead of noble metals, that will protect YBCO during annealing in argon gas. The candidate material we chose was magnesium oxide (MgO) grown by magnetron sputtering.

We have used two types of films in this study. The first type is a bare YBCO film, of 200 nm nominal thickness, on $LaAlO_3$ substrates. The second type, has an overlayer of 30 nm $SrTiO_3$ (STO) deposited *in situ* on top of the 200 nm YBCO film. The purpose of this overlayer is to passivate the YBCO surface thereby preventing oxygen loss during subsequent processing and handling. Both types of films were coated with an MgO layer with a nominal thickness of 75 nm prior to annealing, as depicted in figures 2(a) and 2(b).

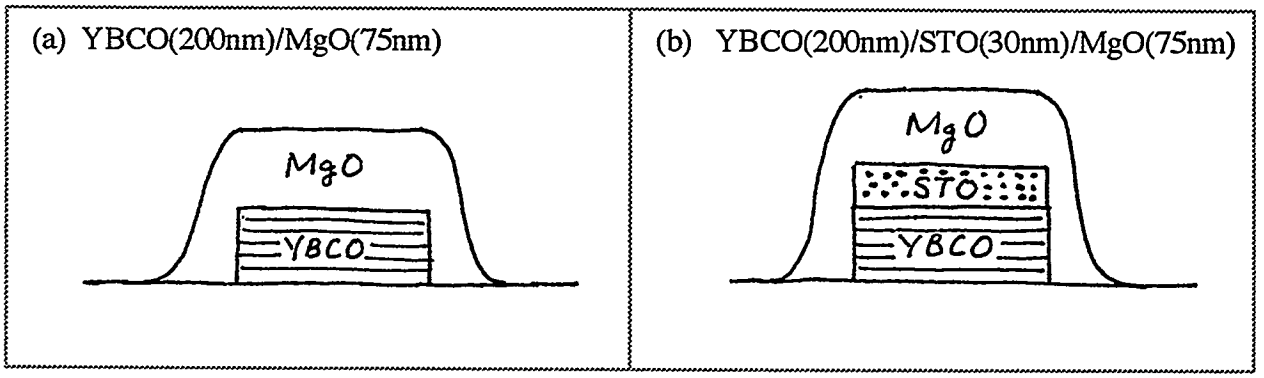


Fig. 2(a) Cross-section of structure A: YBCO film coated with sputter deposited MgO layer.

Fig. 2(b) Cross-section of structure B: YBCO film with a (*in-situ*) STO overlayer, coated with an MgO layer.

Figures 3(a) and 3(b) compare the effects of annealing on these two types of films. The annealing was carried out in a bench top annealing furnace. The gas used was a mixture of 5 % at. hydrogen and 95% at. argon, simulating the "forming gas", and the duration was 20 minutes at 400°C. The T_c measurement was done using an inductive, non-contact technique.

Fig. 3(a) is for a film of bare YBCO coated with 75 nm of MgO. The T_c of the YBCO film was 85.5 K before annealing. It dropped to 47 K after annealing, presumably due to extensive oxygen loss. This is a clear indication that MgO layer alone is not adequate.

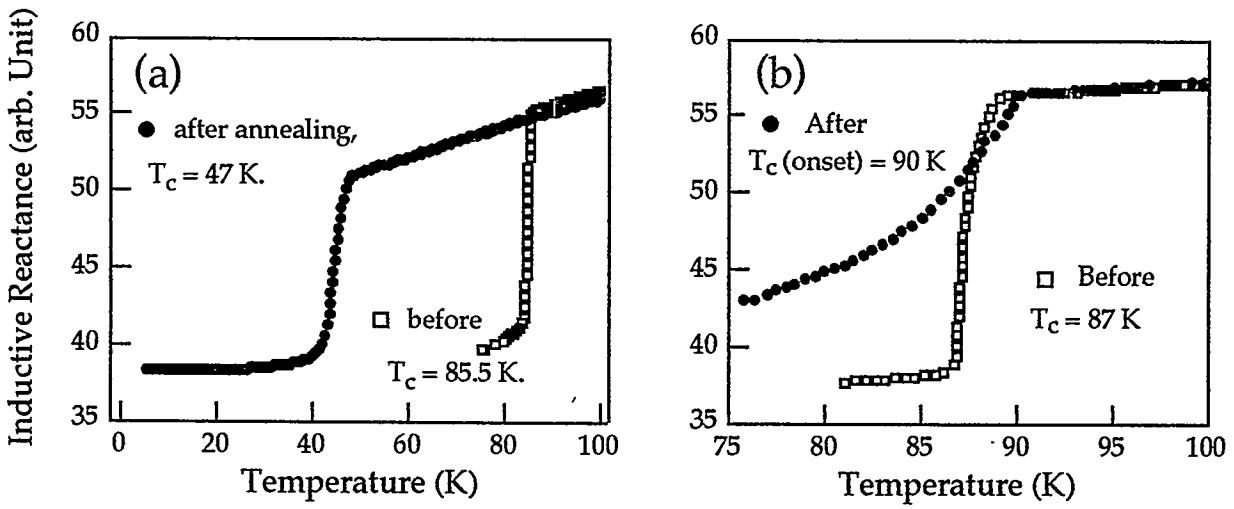


Fig.3 Inductive T_c measurement of unpatterned YBCO films. (a) YBCO(200nm) /MgO(75nm); (b) YBCO(200nm)/STO(30nm)/MgO(75nm). Anneal conditions: 20 mins. @ 400 °C in Ar_2 (95%) + H_2 (5%).

Fig. 3(b) is for a YBCO film with 30 nm STO overlayer. In contrast to the film with MgO layer alone, the onset of the superconducting transition of this film after annealing, close to 90 K, was actually higher than its pre-annealing value. We note that, unlike transport measurement, in the case of inductive measurement the onset of the transition actually signals that a large volume fraction of the sample had become superconducting. This result was the first evidence that a combined structure of STO/MgO could be effectively used to protect YBCO film during forming-gas annealing. However, we also note that was a long tail to the transition, and its depth did not reach the pre-annealing magnitude even at 76 K, indicating a non-uniform distribution of superconductive transitions within the film.

We also patterned several films for transport measurements, in order to compare with the inductive measurements. The pattern consists of a 20 μm wide line with voltage taps 2 mm apart. Four-point resistance was measured using an ac technique with a lock-in amplifier. These films were annealed in a tube furnace in flowing Ar_2, with no hydrogen.

Fig. 4 shows the resistivity versus temperature behavior for a patterned film of the type YBCO(200nm)/ STO(30nm). Before annealing, the resistivity just above the transition is only about 70 μΩ•cm. At 77 K the critical current density J_c was 3.5×10^6 A/cm^2. The film was annealed in a tube furnace at 400 °C in flowing Ar_2 for 20 minutes. We note that after annealing the resistivity just above the transition was nearly doubled, indicating degraded normal state transport properties. The change in the superconducting state was more interesting. The bulk of the transition was 91 K, up by 4 K from the pre-anneal value. However, the film was not fully superconducting even at 77K. The resistance remained finite, at about 10% of its value at the onset of the transition, and showed little temperature dependence thereafter.

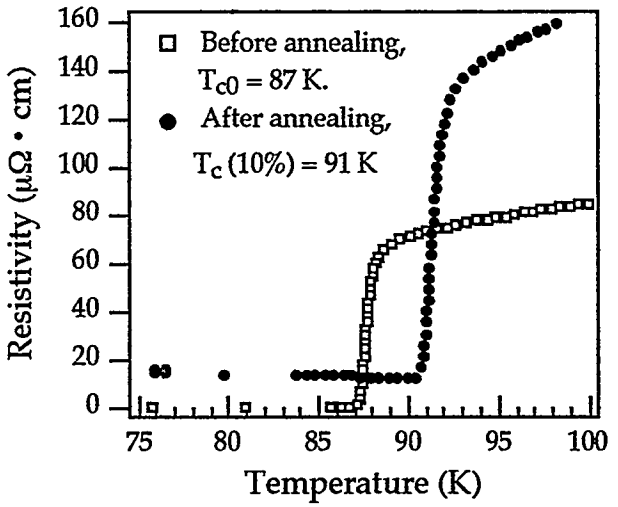


Fig. 4 Resistivity vs. temperature of a patterned YBCO/STO film coated with 75 nm of MgO, before and after Ar_2 anneal.

This result is consistent with that of the inductive measurement (cf. Fig.3(b)). The annealing in argon gas environment increased the transition temperature in the bulk of the YBCO film. It had also created a "foot" in the superconducting transition, which may be associated with localized regions having severe oxygen deficiency. The formation of these regions may conceivably be attributed to oxygen loss through the STO/MgO structure, possibly through pinholes in the MgO, which provided passageways for oxygen out-diffusion during annealing.

To proceed further we modified the MgO deposition procedure. Instead of finishing it in a single vacuum run, we did two sequential runs. We removed the chip from the chamber after the first run, leaving it in the air overnight and then loaded it back to the vacuum chamber for the second run. The total time was kept the same so that the total MgO thickness was the same. However, due to the air exposure, the locations of pinholes in the second MgO layer may not align exactly with those formed in the first MgO layer. This method may therefore enable us to get around the difficulty of pinholes in MgO.

Fig. 5 shows our best result so far. The film was YBCO(200nm)/STO(30nm) type, and was patterned for transport measurement. Before annealing, the room temperature resistivity was about 340 μΩ•cm, and J_c (77K) ~ 3.0×10^6 A/cm^2. After these initial measurements the film was coated with 75 nm of MgO using the two-step sequential deposition procedure, followed by annealing in

flowing argon gas for 20 minutes at 400 °C. We note that the film normal state resistivity increased only slightly after annealing. Moreover, the superconducting transition temperature, T_{c0}, increased from 87.5 K to 88.5 K. In addition the critical current density remained quite high: J_c (77K) ~ 2.0×10^6 A/cm^2. This result demonstrates that YBCO films, protected by an *in-situ* overlayer of STO and sequentially deposited MgO double layers, is capable of withstanding annealing in forming-gas, for temperatures and durations typically found in semiconductor processing. T_{c0} tends to increase slightly, and J_c degradation was small.

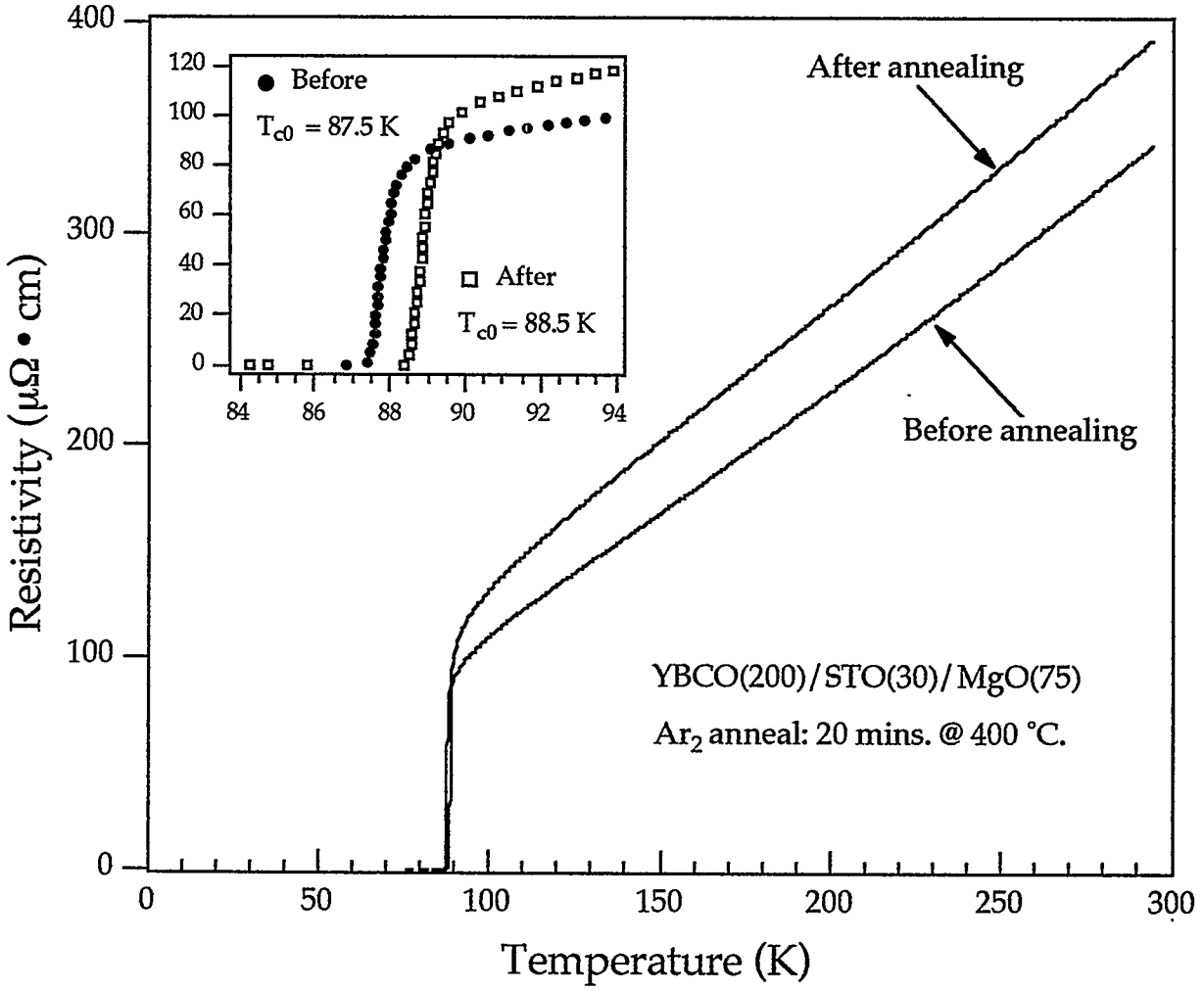


Fig. 5 Resistivity vs. temperature of a patterned YBCO/STO film before and after Ar_2 anneal. The MgO protection layer was grown using a modified 2-step sequential sputter deposition process, see text for details.

The importance of the STO *in-situ* overlayer can be readily appreciated by comparing Fig. 5 with Fig. 6, which displays the annealing results of another YBCO film. This film was without a STO overlayer, but went through the same procedure as the sample in Fig. 5, i.e., a two-step sequential coating of MgO, followed by annealing in argon at 400 °C for 20 minutes.

We note that both the normal state behavior, as well as the superconducting properties changed markedly. T_{c0} decreased from ~ 88 K to 79 K, just above the liquid nitrogen boiling temperature. Moreover, J_c (77K) decreased form ~2.5×10^6 A/cm^2 to a mere 3.5×10^3 A/cm^2, a nearly ten-fold reduction. This result clearly demonstrates the importance of an overlayer with high crystalline quality.

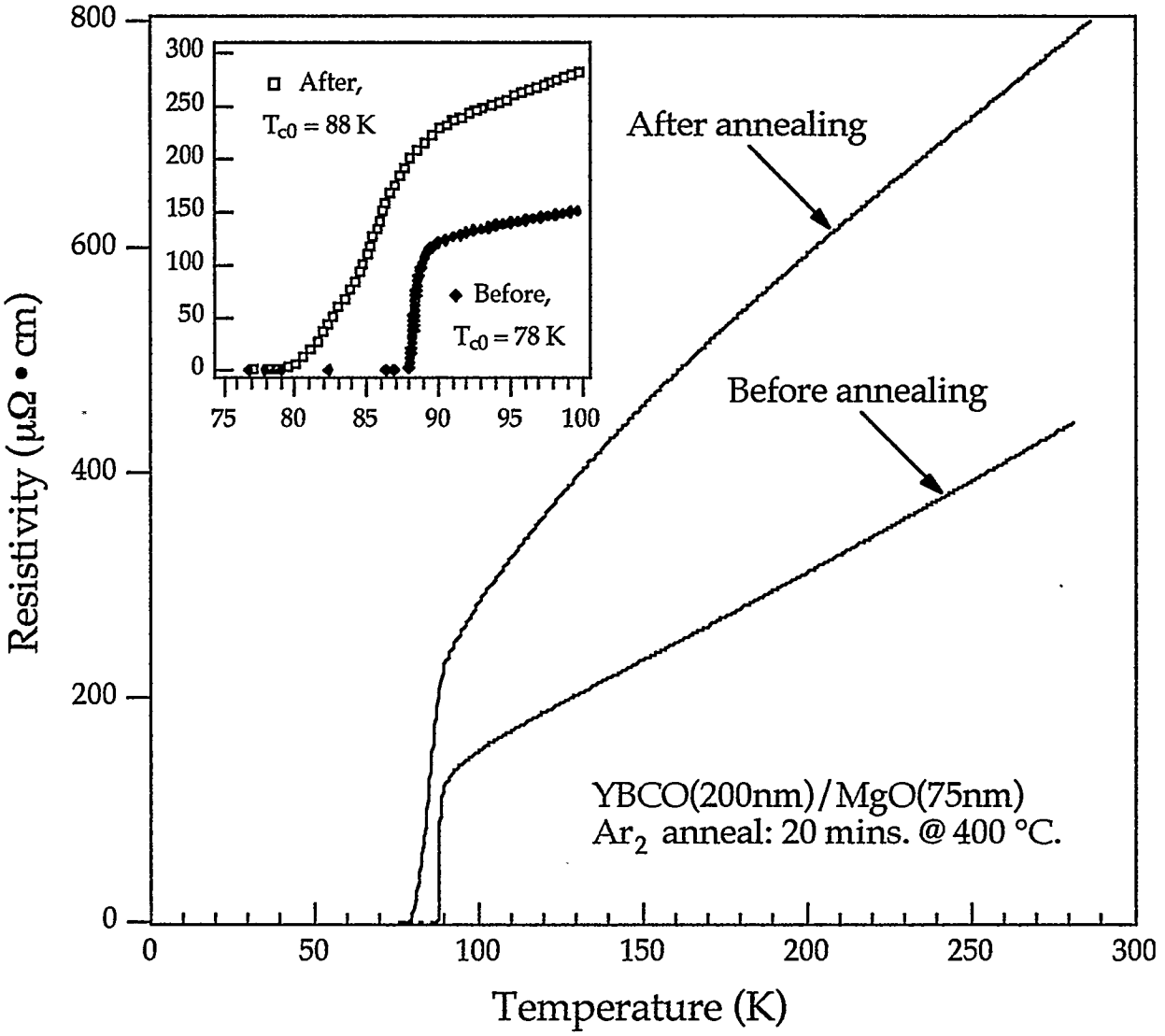


Fig. 6 Resistivity vs. temperature of a patterned YBCO film before and after Ar_2 anneal. The MgO protection layer was grown using a modified 2-step sequential sputter deposition process, see text for details.

The Effect Of Annealing Temperature

We also studied the influence of the annealing temperature. Two more chips, one is identical to that of Fig. 5 (identical in the sense that they were diced from same wafer), and the other identical to that in Fig. 6, were annealed in argon gas at 450 °C for 20 mins. The results are shown in figures 7 and 8. Neither was superconducting at 77 K after annealing. However, less damage was done to the film with STO overlayer: its T_{c0} was 60 K, about 10 degrees higher than the one without STO overlayer, and the increase in the normal state resistivity was several times smaller.

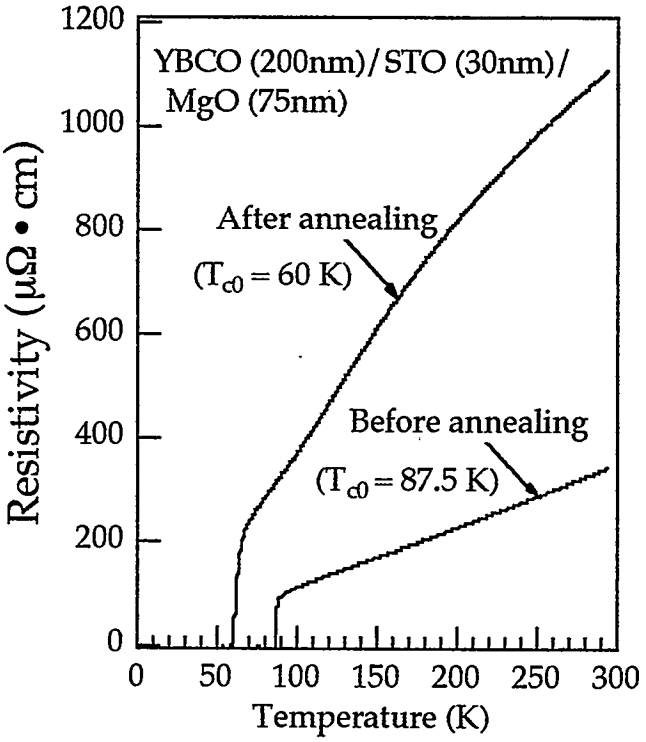


Fig. 7 R vs. T before and after annealing of a patterned film of YBCO/STO. It was covered with a 75 nm MgO using a 2-step sequential deposition process. Ar_2 anneal was at 450 °C for 20 minutes.

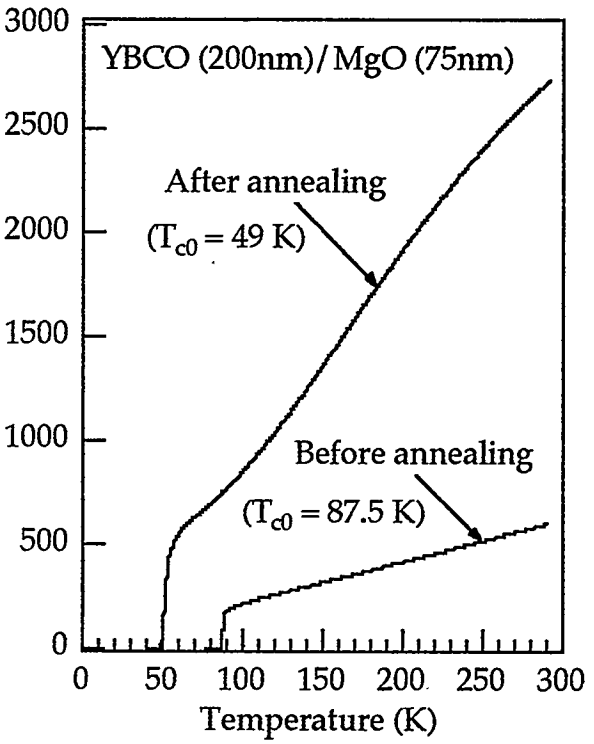


Fig. 8 Similar to Fig. 7, except that the YBCO film did not have STO overlayer.

DISCUSSION

We summarize here systematic trends observed in our Ar_2 gas annealing studies.

First, Ar_2 annealing general degraded the superconducting and normal state properties of YBCO films, the extent depends strongly on the annealing temperature, the structure of the protection layer, and the procedures with which these structures were made. The cause is almost certainly the oxygen out-diffusion, which should have a strong temperature dependence. This is supported by the results shown in figures 5, 6 and figures 7, 8. A mere 50 °C increase in annealing temperature the had resulted in much more extensive loss of oxygen.

Second, it is known that oxygen diffusion is highly anisotropic in YBCO crystals, with the rate in ab-plain being much higher than that along c-axis direction[5]. In thin films the anisotropy may not be as great, due to defects assisted hopping along the c-axis direction. We believe that in our films diffusion in both directions are taking place and are perhaps equally important. For a patterned film such as our J_c test structure, oxygen out diffusion could proceed either along the c-axis direction, which is perpendicular to the film, or in the ab-plain to exit through the edges of the patterned line. However, as long as the annealing temperature is not too high (e.g., less than 400 °C in our experiments), the diffusion rate in neither direction is high, the oxygen loss through the edge diffusion is limited to the periphery of the line. Consequently the annealing results is determined essentially by the perpendicular diffusion, since the films are much thinner than they are wide (0.2 μm vs. 20 μm). This may explain the beneficial effects of the STO overlayer, which is primarily effective in hindering oxygen out-diffusion in the perpendicular direction.

Third, it is also well documented that oxygens on the CuO chain sites in YBCO can be easily displaced under influences of moderate thermal, electrical (e.g., electromigration current[6]) and structural (e.g., strain at high-angle grain boundaries) disturbances. However, the thermally driven oxygen displacement, as in the case of electrically driven oxygen defects migration (electromigration), can have rather different effects on YBCO properties depending on the magnitude of the driving force. A small driving will usually have some beneficial effects due to oxygen reordering. Examples of this are the noticeable increase in the onset of the superconducting transition in Fig.4, and the increase in T_{c0} in Fig. 5. These are not likely due to the increase in oxygen content in the film during annealing, since it was carried out in reducing ambient. Rather, they are due to the improved oxygen vacancy ordering in the interior of the line. Under a larger driving force, rapid and long-range oxygen diffusion occurs, leading to extensive loss of oxygen (mostly through the edge of the patterned lines), and drastic deterioration in film quality. In electromigration experiments on YBCO thin film microbridges it was observed[6] that there was a threshold in the migration current density beyond which the effects turn from beneficial to deterious. Our annealing experiments seem to suggest that there is an analogous threshold, in temperature, between 400 °C and 450 °C.

Forth, the in-situ deposited STO overlayer appears essential. Its effectiveness in preventing oxygen loss from YBCO may be attributed to its high degree of crystallinity and epitaxy, a consequence of *in-situ* high temperature deposition.

Fifth, MgO was chosen for its excellent electrical insulation property. Such MgO has been used in our study of YBCO/noble metal interface resistivity study to provide isolation between the YBCO and the noble metal overlayer, and proved to be sufficient and reliable. For example, isolation test structures formed by sandwiching 75 nm of MgO between two metal cross strips with 0.1×0.1 mm^2 overlap area showed practically infinite resistance. Moreover, their integrity was not affected by annealing in argon gas flow for temperatures up to about 450°C. These qualities make MgO a good choice. In spite of its effectiveness in providing electrical isolation, our MgO layers appear to be less effective in proventing atomic diffusion of oxygen. Since they are deposited at room temperature, they are amorpuous and may well be porous, and may possibly have randomly located pinholes over the entire chip area (about 10×10 mm^2). Therefore there exist passageways for oxygen out-diffusion, which is accelerated by the elevated temperature during annealing. A modified method, i.e., the two-step sequential deposition of MgO appeared to work in the way that was anticipated, i.e., to misalign and hence seal off the pinholes formed in each of the two steps. However a firm conclusion can not be drawn based on just this one example.

ACKNOWLEDGEMENTS

This work was performed under the auspices of the U.S. Department of Energy. The YBCO films with $SrTiO_3$ overlayer used in annealing study were provided by the HTS group at TRW.

REFERENCES

1. J. BROWNE, "HTS Suppliers Take Aim At Commercial Markets," Microwaves & RF, 33, February, 1996.
2. W. A. PHILLIPS, "HTS Subsystem Analysis," Journal of Superconductivity, 8, 735 (1995).
3. T. VAN DUZER, "Superconductor-semiconductor hybrid devices, circuits and systems," Cryogenics, 28, 527 (1988).
4. Simon S. Cohen and Gennady Sh. Gildenblat, "Metal-Semiconductor Contacts and Devices, VLSI Electronics Microstructure Science, Volume 13," Academic Press, 1986.
5. S. J. Rothman, J. R. Routbort, and J. E. Baker, "Tracer diffusion of oxygen in $YBa_2Cu_3O_{7-\delta}$, " Phys. Rev. B 40, 8852 (1989).
6. B. H. Moeckly, D. K. Lathrop, and R. A. Buhrman, Phys. Rev. B 47, 400 (1993).

CALCULATION AND MEASUREMENT OF THE CRITICAL STATE IN SUPERCONDUCTING TAPES

K. L. Telschow and L. S. Koo

Idaho National Engineering Laboratory
Idaho Falls, ID 83415-2209

ABSTRACT

An integral equation approach to solving for the flux front profile in the critical state model is described. Both nonuniform external fields and demagnetizing geometries can be accommodated as long as cylindrical symmetry is preserved. Results for a sphere in a uniform external field and a tape in the field of an external coil parallel to the tape surface are presented.

INTRODUCTION

Quantitative determination of the local low field critical current density, $J_c(H=0)$, in a noncontacting manner is necessary for spatial uniformity inspection during fabrication of high T_c superconducting tapes. This can be accomplished by measuring the magnetic response of the sample with small source/pickup probe coils that spatially scan over the tape surface [1-3] and then using the Bean critical state model [4] to determine J_c from the measured magnetic hysteresis. In this model, the induced currents in the sample are either at the critical value or zero, forming a critical state region bounded by a "flux front profile". The response is calculated by summing the fields produced by the currents within this flux front. J_c can be inferred from the local magnetization measurements. However, demagnetization effects of the tape/coil geometry are not readily accounted for within the normal extension of the Bean critical state

model. This paper describes a method for predicting the flux front profile as a function of the external field that takes into account the demagnetizing effects of the sample surfaces. An integral equation technique is presented that is applicable to sample/probe geometries exhibiting cylindrical symmetry. The net magnetization signal detected by pickup coils is calculated and compared with experimental measurements.

THEORETICAL APPROACH

Whenever a superconducting material is placed in an applied magnetic field, a region of shielding current is induced. According to the Bean critical state model, the induced current begins at the material boundary next to the applied magnetic field and extends inward, satisfying $\nabla x\vec{B} = \mu_0\vec{J}_c$, where $\vec{B}$ is the magnetic induction vector and μ_0 is the free-space permeability. This leads to the net magnetic flux density in the material decreasing to zero at the flux front boundary. The region inside this boundary is free of magnetic field, shielded completely by the induced screening currents. This flux front surface inside a superconducting material can be defined as the surface on which the total magnetic field is zero; a vector equation defines the boundary [5]. However, as is proven in [6], the flux front can also be defined as a surface of zero vector potential, $\vec{A}$, where $\nabla x\vec{A} = \vec{B}$. The use of the vector potential simplifies the calculations since only one component is needed for problems of cylindrical symmetry. For problems with this symmetry, the induced currents inside the material can be modeled as coaxial loops, each carrying a current of constant value, J_c. Hence, the vector potential due to all of the induced currents is a volumetric integral sum of the single loops. The vector potential of a single current loop is well known. The unknown quantities in this integral are the flux front boundary, Ψ, which forms part of the integration limit, and the current density, J_c. In addition, the flux front boundary, Ψ, is a function of both space and the applied magnetic field. To simplify the calculations, the external applied field is normalized by J_c and a characteristic length. With this normalizing scheme, the total vector potential, A_{tot}, becomes a dimensionless implicit function of $\Psi(R,\beta)$, a normalized spatial variable, R, and the normalized applied field, β. Then the total vector potential is given by $A_{tot} = A_\beta - A_{J_c}$, where A_β and A_{J_c} are the vector potentials due to the external field, β, and the induced critical currents, J_c. The minus sign results from the shielding effect. For a given β, the position of the flux front surface is determined by finding spatial points where the total vector potential is zero. This is a difficult problem in general, but can be simplified by reduction to a single integral equation as follows. In general, β is a function of time. This technique deals only with the quasi-stationary states of the critical state. The time scale for changes in the external field is typically very much longer than that exhibited by flux line motion, so the model always assumes a sequence of stationary states uniquely defined by the history and present value of the external field. Changes in the external field then produce a corresponding change in the flux line profile position, but at all times the total vector potential

on this profile is zero. Therefore, a requirement for determining the flux penetration profile is that

$$\frac{\partial A_{tot}(\Psi(R,\beta),R)}{\partial \beta} = F\left(\Psi(R,\beta), \frac{\partial \Psi(R,\beta)}{\partial \beta}\right) = 0. \tag{1}$$

Equation (1) is often a nonlinear integral equation. As suggested by the critical state model, when the external field, β, is initially turned on, flux enters the superconducting material from the surface. Thus the initial profile, $\Psi(R, \beta=0)$, is the material surface. Substitution of the known surface profile into (1) yields a linear integral equation of the first kind for the unknown derivative of Ψ. The algorithm analyzed by Gold [7] is used to resolve (1). As the external field is increased to $\beta = \Delta\beta$, the new flux front profile is approximated by

$$\Psi(R,\beta = \Delta\beta) = \Psi(R,\beta = 0) + \frac{\partial \Psi(R,\beta = 0)}{\partial \beta} \cdot \Delta\beta. \tag{2}$$

This approximation is acceptable as long as the increment $\Delta\beta$ is sufficiently small. With this approximation the flux front profile at $\beta = \Delta\beta$ is known but its derivative is yet to be determined. This is the same situation as at the beginning when $\beta = 0$. The above procedure is repeated and the new β value determined. This methodology results in a progressive incremental numerical scheme in β and an iterative procedure for resolving (1) for each β. Once the flux front profile's dependence on the external field for the zero-field cooled (ZFC) case is known, then the response of the sample to a complete cycle of changes in the external field can be readily calculated, knowing $M_{ZFC}(\beta)$, as follows:

$$M_{\pm}(\beta) = \pm M_{ZFC}(\beta_{max}) \mp 2M_{ZFC}\left(\frac{\beta_{max} \mp \beta}{2}\right) \tag{3}$$

where (+) means decreasing $\beta_{max} > \beta > -\beta_{max}$ and (-) means increasing $-\beta_{max} \rightarrow \beta \rightarrow +\beta_{max}$ in the external field. Examples of this calculation for a spherical sample in a uniform external field and an infinite plate in the field of an external single loop source coil are described below.

SPHERE IN A UNIFORM EXTERNAL FIELD

In this example of a cylindrically symmetric application, let the radius of the sphere be r_0 and the flux front surface be described by $\rho(z,\beta)$ where the external field dependence is included in the normalization parameter $\beta = H_z^{ext}/(r_0 J_c)$. The configuration is depicted in Fig. 1. Also shown in

this figure are two flux fronts trapped inside the sphere as a result of extending the external field to a maximum value below the full penetration value. The total normalized vector potential at (r,z) is given below [6], where all lengths are normalized to the sphere radius $R = r/r_0$; $Z = z/r_0$; $\Psi = \rho/r_0$; etc.; and $\mu_0 a_\phi(R,Z;R',Z')$ is the vector potential at (R,Z) due to a single current loop at (R′,Z′)

$$A_\phi(R,Z,\beta) = \frac{\beta R}{2} - \int_{-1}^{1} dZ' \int_{\Psi(Z',\beta)}^{\sqrt{1-Z'^2}} dR'\, a_\phi(R,Z;R',Z') \tag{4}$$

A single integral equation describing the flux front surface for a sphere results

$$\int_{-1}^{1} dZ'\, a_\phi(\Psi(Z,\beta),Z;\Psi(Z',\beta),Z') \frac{-\partial\Psi(Z',\beta)}{\partial\beta} = \frac{\Psi(Z,\beta)}{2}. \tag{5}$$

The initial profile of the flux front is $\Psi(Z, \beta=0) = \sqrt{1-Z^2}$. All flux fronts for the ZFC case are obtained by following the aforementioned incremental iteration procedure. The full field penetration value obtained yields a normalized external field $\beta^*=0.789$. A value of $\pi/4$ is found analytically for a completely filled sphere. The numerical and analytical values agree to within 0.5% [6] which is consistent with the expected accuracy of the calculation procedure used. The subsequent magnetization hysteresis curve, $M(\beta)$, for the sphere due to a complete cycle of change in the external field can be obtained through equation (3). The results are given in Fig. 2, normalized by the full penetration magnetization $M^* = 3\pi r_0 J_c/32$.

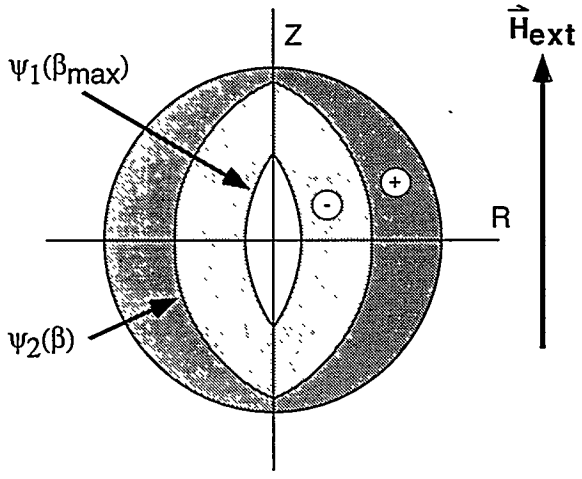


Fig. 1. Flux front profiles for a sphere.

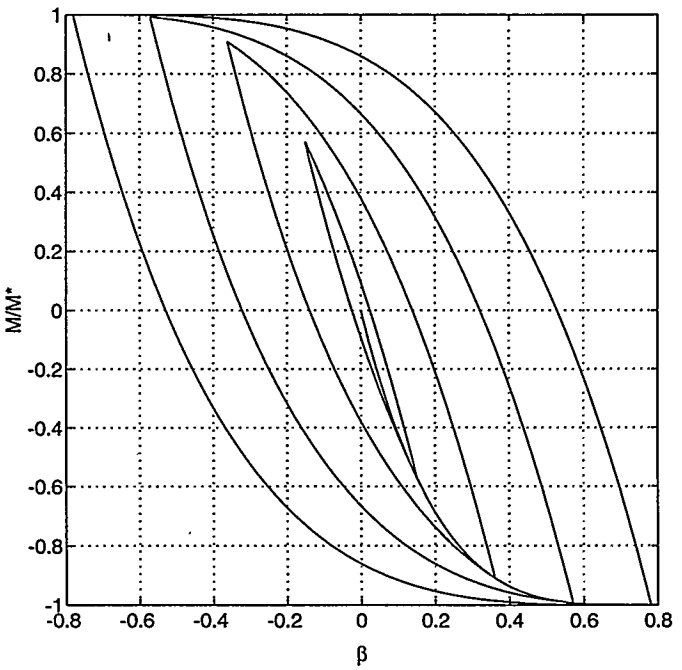


Fig. 2. Magnetization hysteresis of the sphere for various limiting fields.

TAPE GEOMETRY

A single circular coil above a superconducting infinite plate of thickness D is shown in Fig. 3. All geometrical lengths are normalized by the coil radius, r_c. The driving coil is placed at $(1,Z_c)$. The normalized external field is $\beta = I/J_c r_c^2$, where I is the current in the driving coil. The flux front profiles for the ZFC case are also shown in Fig. 3. At $\beta = 0$, $\Psi = 0$ coincides with the plate top surface. As β is increased to the next value, Ψ shows a profile that penetrates deepest directly under the driving coil. This deepest value of Ψ reaches the plate bottom edge at $\beta^* = 0.365$, for $D = -0.2$. Beyond this value, the flux front is broken up into two sections: an inner section, Ψ_1, which approaches R=0, and an outer section, Ψ_2, which approaches R=∞ as β increases. The intersections between Ψ_1, Ψ_2 and the bottom surface, $Z = D$, are denoted as R_1 and R_2, respectively. Before the bifurcation of the flux front surface, the total vector potential amplitude (normalized by $\mu_0 J_c r_c^2$) is

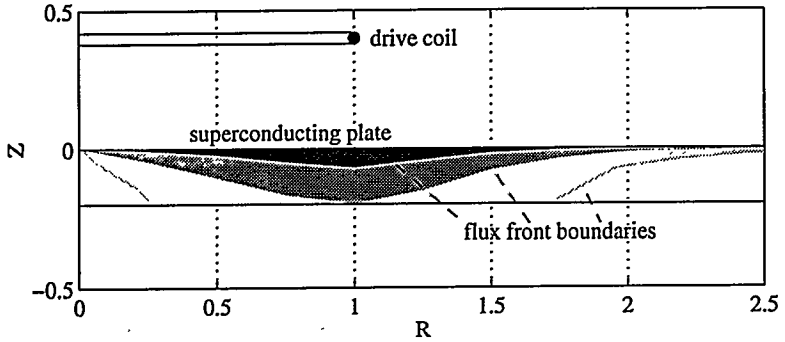


Fig. 3. Geometry for flux penetration in a plate for ZFC case with $\beta = 0.1$, 0.365 and 0.865.

$$A_{tot}(R,Z) = \beta a_\phi(R,Z;1,Z_c) - \int_0^\infty dR' \int_{\Psi(R',\beta)}^0 dZ'\, a_\phi(R,Z;R',Z') \tag{6}$$

where $\mu_0 a_\phi(R,Z;R',Z')$ is the vector potential at (R,Z) due to a unit current loop at (R′,Z′). A single integral equation results for the flux front surface

$$\int_0^\infty dR'\, a_\phi(R,\Psi(R,\beta);R',\Psi(R',\beta)) \frac{-\partial\Psi(R',\beta)}{\partial\beta} = a_\phi(R,\Psi(R,\beta);1,Z_c). \tag{7}$$

The initial profile of the flux front is $\Psi(R,\beta=0) = 0$. As β is increased beyond β^*, the flux front surface Ψ will become two separate surfaces, Ψ_1 and Ψ_2. The total vector potential becomes

$$\begin{aligned} A_{tot}(R,Z) = {} & \beta a_\phi(R,Z;1,Z_c) - \int_0^{R_1(\beta)} dR' \int_{\Psi_1(R',\beta)}^0 dZ'\, a_\phi(R,Z;R',Z') \\ & - \int_{R_1(\beta)}^{R_2(\beta)} dR' \int_D^0 dZ'\, a_\phi(R,Z;R',Z') - \int_{R_2(\beta)}^\infty dR' \int_{\Psi_2(R',\beta)}^0 dZ'\, a_\phi(R,Z;R',Z') \end{aligned} \tag{8}$$

With Ψ_1' and Ψ_2' denoted as $\Psi_1(R',\beta)$ and $\Psi_2(R',\beta)$, two coupled equations result [8]

$$\int_0^{R_1(\beta)} dR'\, a_\phi(R,\Psi_1;R',\Psi_1')\cdot\frac{-\partial\Psi_1'}{\partial\beta}+\int_{R_2(\beta)}^{\infty} dR'\, a_\phi(R,\Psi_1;R',\Psi_2')\cdot\frac{-\partial\Psi_2'}{\partial\beta}=a_\phi(R,\Psi_1;1,Z_c) \quad (9a)$$

and

$$\int_0^{R_1(\beta)} dR'\, a_\phi(R,\Psi_2;R',\Psi_1')\cdot\frac{-\partial\Psi_1'}{\partial\beta}+\int_{R_2(\beta)}^{\infty} dR'\, a_\phi(R,\Psi_2;R',\Psi_2')\cdot\frac{-\partial\Psi_2'}{\partial\beta}=a_\phi(R,\Psi_2;1,Z_c). \quad (9b)$$

These two equations couple both flux fronts together. In the first equation, the observation points are on Ψ_1, while in the second equation the observation points are on Ψ_2. These two equations, expressed in matrix format, are

$$\begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}\begin{pmatrix} f_1 \\ f_2 \end{pmatrix}=\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \text{where} \quad K_{12}f_2=\int_{R_2(\beta)}^{\infty} dR'\, a_\phi(R,\Psi_1;R',\Psi_2')\cdot\frac{-\partial\Psi_2'}{\partial\beta}, \text{ etc.} \quad (10)$$

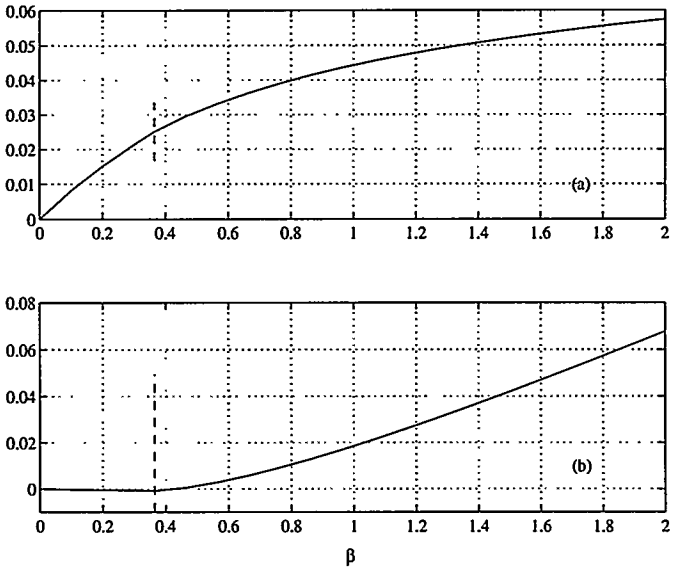


Fig. 4. A_{J_c} for a coil above the plate (a) and A_{tot} for a coil beneath the plate (b).

The incremental and iterative procedures were employed to resolve the two separate flux fronts simultaneously. Two simulated measurement examples are given in Figs. 4a and 4b. In these examples, $r_c = 0.5$ mm, $Z_c = 0.4$, and $D = -0.2$. In Fig. 4a, the normalized screening current vector potential amplitude, A_{J_c}, at the driving coil position (1,0.4) is given, simulating the measurement results found above the sample with a top balanced coil. In Fig. 4b, the normalized total vector potential amplitude, A_{tot}, is given at (1,-0.6), simulating the measurement results found beneath the sample with a single unbalanced coil. The vertical dashed line signifies the value β^* where the flux front fully penetrates to the plate bottom surface.

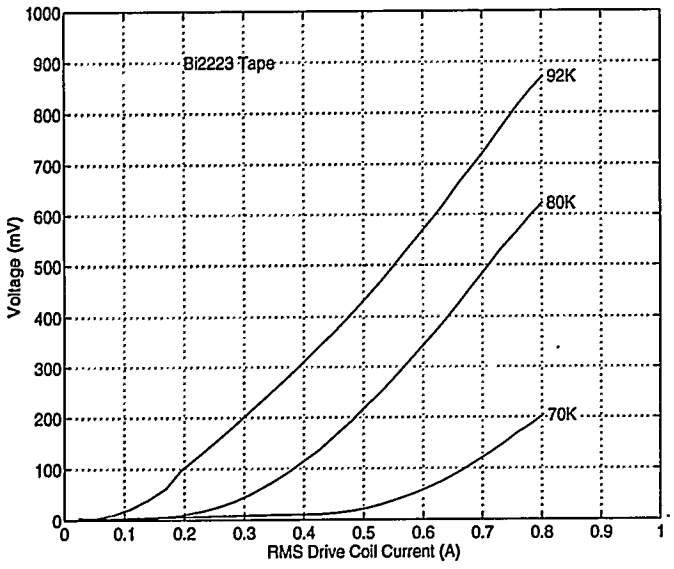


Fig. 5. Measured bottom coil signal amplitudes for different temperatures.

The experimental geometry was that of a single layer superconductor plate deposited between two outside silver layers. Tape samples of Pb-BiSrCaCuO (2223 phase) produced by the powder-in-tube method [3,9] were measured with a small probe source coil (13 turns of #36 copper wire, 1 mm inside diameter). Balanced opposing pickup coils (5 turns each of #36) were wound over the source coil, producing a small source/pickup probe that could be scanned over the sample surface in a liquid nitrogen bath. An additional pickup coil (5 turns #36 copper wire, 1 mm diameter) was positioned below the tape sample. Lift-off distances for both top and bottom coils were about 0.7 mm. Fig. 5 shows the measured results from the bottom coil at one position on the tape for three different temperatures. These results, which are qualitatively similar to the theoretical results of Fig. 4b, represent the magnetization due only to the induced screening currents within the flux front. The shielding effect is clearly visible and, by extrapolation, a unique point corresponding to full plate penetration can be estimated. The local critical current density can then be determined from the corresponding value of β.

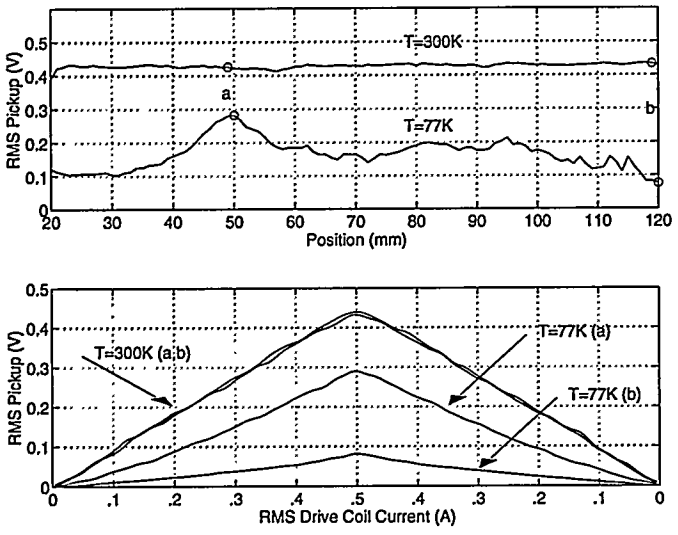


Fig. 6. Maximum signal along a tape (top) and corresponding RMS signal amplitudes (bottom) as a function of the source coil current.

Tapes produced by the "powder in tube" method exhibit thickness variations as well as inhomogeneity in the superconducting layer. Nevertheless, the above approach can be used to provide a measure of uniformity and local critical current density. Fig. 6 shows typical results for a scan of another small probe along a tape, produced by the powder in tube method [9]. Significant variations are observed due to local microstructural inhomogeneities. Detailed information like that shown in Fig. 6 is helpful in

locating problems associated with intergrain connections and thickness or porosity defects.

CONCLUSION

A method has been outlined for calculating the flux front profile for a superconducting sample in either a uniform or nonuniform applied magnetic field possessing cylindrical symmetry. This technique extends the Bean critical state model by fully accounting for demagnetizing effects. The method relies upon finding the flux front penetration profile, which is done by resolving a linear integral equation of the first kind. Measurement-induced voltages and the entire hysteresis loop response can be found by extension of the ZFC magnetization response with a changing external field. Other measured quantities relating to the critical state can be calculated directly from the hysteresis loop if the time dependence of the external field is known.

ACKNOWLEDGMENT

The authors thank M. T. Lanagan of Argonne National laboratory for the tape samples used and K. K. Haulenbeek and F. Bruneel for help in obtaining the measurement results. This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under DOE Idaho Operations Office Contract DE-AC07-94ID13223.

REFERENCES

1. A. M. Campbell and J. E. Evetts, *Critical Currents in Superconductors*, Taylor & Francis Ltd., London, (1972).
2. K. L. Telschow and T. K. O'Brien, *Appl. Phys. Lett.* 59, 730 (1991).
3. K. L. Telschow, T. K. O'Brien, M. T. Lanagan and D. Y. Kaufman, *IEEE Trans. Magn.* 3 (1), part III, 1643 (1993).
4. C. P. Bean, *Phys. Rev. Lett.* 8, 250 (1962) and *Rev. Mod. Phys.* 36, 31 (1964).
5. R. Navarro and L. J. Campbell, *Phys. Rev. B* 44, 10146 (1991).
6. K. L. Telschow and L. S. Koo, *Phys. Rev. B* 50 (10), 6923 (1994).
7. R. Gold, Report ANL-6984, Argonne National Laboratory, December (1964).
8. L. S. Koo and K. L. Telschow, *Phys. Rev. B* 53, 8743 (1996).
9. Samples were provided by M. T. Lanagan at Argonne National Laboratory.

SURFACE AND DEFECT MORPHOLOGIES IN ANISOTROPIC ELASTIC AND PIEZOELECTRIC SOLIDS

Huajian Gao* and David M. Barnett**
Division of Applied Mechanics* and Department of Materials Science & Engineering**
Stanford University
Stanford, CA 94305

ABSTRACT

We investigate issues related to the equilibrium and stability of surface and line defect morphologies in both piezoelectric and anisotropic elastic solids. Following our previous efforts which established that mechanical stresses in purely elastic solids can promote instability of an initially flat surface with respect to surface roughening, we show that the (initially flat) interface between two dissimilar piezoelectric solids can be unstable when subjected to coupled electromechanical loading. Quite recent cross-sectional observations of electrodeposited thin films by Japanese and British researchers provide experimental confirmation of these predictions. We also investigate the occurrence of equilibrium arrangements (zero Peach-Koehler force arrangements) of line defects (dislocations) in anisotropic elastic crystals in the absence of externally applied stresses. Contrary to prevailing opinion, equilibrium arrangements of dislocations under no externally applied stresses appear to be the rule rather than the exception. The existence of such "zero stress arrangements" is fundamental to developing micromechanical models of plastically deforming solids.

INTRODUCTION

In the present work we focus our attention on (i) surface instabilities induced by electromechanical loading in piezoelectric bi-crystals and upon (ii) equilibrium arrangements of line defects (dislocations) in crystals in the absence of externally applied stresses. The former problem (i) represents an extension of our previous study of a purely elastic solid [1] and of a piezoelectric half-space [2], which has established that mechanical stress can induce surface roughening of an initially flat surface, while electrical loading may act to either stabilize or destabilize a flat surface. These results may be of extreme importance for the understanding of crack formation in defect free materials as well for uncovering possible mechanisms for misfit dislocation generation in thin films

deposited on substrates. The present extension allows one to address the broad issue of controlling flat surface/interface instability in materials which may be isotropic or anisotropic, purely dielectric or purely elastic, or piezoelectric, which, quite obviously, has important implications for the fabrication of dielectric and piezoelectric thin films. The motivation for investigating the latter problem (ii) is perhaps best summarized in a very short presentation by Head, et al [3], who noted that although there is total agreement about the role of dislocations in explaining strain hardening in metals, such a vast range of dislocation behavior and arrangements has been observed experimentally that there is little agreement as to which observations are the more important. It seems as if the issue may have to be decided on theoretical grounds by mathematical modeling of dislocation plasticity, and we begin by examining the simplest of such problems, namely, the determination of self-equilibrated dislocation arrangements in solids which are in the unloaded state (zero externally applied stress) following inhomogeneous plastic straining.

SURFACE INSTABILITIES INDUCED BY ELECTROMECHANICAL LOADING

Stress-induced surface instabilities in purely elastic media have been discussed by Gao [1] in connection with diffusive crack formation and with misfit dislocation generation in strained heteroepitaxial thin films. The instability (roughening) of an initially flat surface is driven by the release of stored strain energy due to flaw formation and is opposed by the formation of new surface and the associated increase in surface energy of the system (Asaro and Tiller [4]). Chien and Gao [2] have studied the surface roughening of a piezoelectric half-space when the half-space boundary is either insulating or conducting and have shown that it is indeed possible for electrical loading to help stabilize the initial surface against roughening. The present work treats two dissimilar piezocrystals (half-spaces) joined by perfect electroelastic bonding and studies the stability of the initially flat interface between the two crystals with respect to small interfacial undulations. An analysis of this problem represents an important extension in two respects, namely, (i) it allows us to study roughening of an internal interface as opposed to an external surface, and (ii) depending on the choice of physical properties chosen for the two crystals, either crystal may be considered as isotropic or anisotropic, purely dielectric or purely elastic, or piezoelectric. The former, (i), is important because internal interfaces afford sites for defect initiation in real configurations of technological importance, while the latter, (ii), allows us to study bi-material configurations of importance in modern capacitor, transducer, or integrated circuit fabrication. The ability to control interfacial stability through appropriately coupled electromechanical loading is of paramount significance in these technologies.

Editorial space limitations preclude a complete presentation of the instability analysis, but we shall endeavor to present a brief description of our procedure. The complete treatment is given in Chien, et al [5]. The constitutive relations for a linear piezoelectric medium are [6,7]

$$T_{ij} = C_{ijkl}S_{kl} - e_{kij}E_k \tag{1}$$

$$D_i = e_{ikl}S_{kl} + \varepsilon_{ij}E_j \tag{2}$$

where C_{ijkl} are the stiffnesses measured at constant strain, e_{ijk} the piezoelectric stress constants, and ε_{ij} the permittivities measured at constant strain. T_{ij} , S_{kl}, E_k, and D_i represent components of stress, strain, electric field, and electrical induction, respectively, with

$$S_{kl} = \frac{1}{2}\left(u_{k,l} + u_{l,k}\right) \text{ and } E_k = -\phi_{,k} \tag{3}$$

where $\mathbf{u}$ and ϕ represent displacement and electric potential, respectively. With no body forces or free charges, the governing equations of linear piezoelectricity (quasistatic approximation) are

$$T_{ij,i} = 0 \text{ and } D_{i,i} = 0. \tag{4}$$

The present work considers the stability of an initially flat piezoelectric interface between two semi-infinite dissimilar piezoelectric materials, I (the upper medium) and II (the lower medium), subjected to arbitrary electromechanical loading conditions (Figure 1). In particular we consider electromechanical states which are plane in the sense that all fields are independent of x_3.

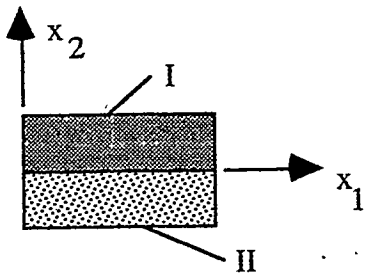


Figure 1. An Initially Flat Interface Between Two Piezoelectric Solids

The composite piezoelectric problem may be solved using the Stroh formalism [8], originally developed for anisotropic elasticity, as extended to the piezoelectric case by Lothe and Barnett [7,9,10]. The important feature to note is that the piezoelectric problem for the two bonded crystals is solved subject to the interfacial boundary conditions that mechanical displacement, normal component of electrical induction, mechanical tractions, and electrostatic potential must be continuous across the interface $x_2 = 0$. To examine stability of the flat interface with respect to small perturbations, we imagine treating the same problem when the interface is now described by a cosinusoidal form, say

$$x_2 = A(x_1) = A\cos kx_1. \quad k = 2\pi/\lambda. \tag{5}$$

The solution for the perturbed interface may be written in terms of two four-vector potentials $\psi^{(i)}$, i = I, II, associated with the respective media, namely,

$$\psi^{(i)} = \psi^{(i0)} + \Lambda^{(i)} \tag{6}$$

where $\Lambda^{(i)}$ represents the effect of the perturbation. Interfacial continuity conditions require that

$$\left[\left|\mathrm{Re}\,\psi^0\right|\right] = 0 \text{ and } \left[\left|\mathrm{Im}\left(\mathbf{Y}\psi^0\right)\right|\right] = 0 \tag{7}$$

where $\mathbf{Y}^{(i)}$ is the 4 x 4 hermitian Stroh admittance matrix for the half-space (i) and $\left[\left|\ \right|\right]$ means the "jump" across the interface $x_2 = 0$. Applying first order perturbation theory allows us to deduce that the perturbed potentials satisfy

$$\left[\left|\mathrm{Re}\,\Lambda\right|\right] = -A\left[\left|\mathrm{Re}\left(\psi^0_{,2}\right)\right|\right] \text{ and } \left[\left|\mathrm{Im}(\mathbf{Y}\Lambda)\right|\right] = -A\left[\left|\mathrm{Im}\,\mathbf{Y}\psi^0_{,2}\right|\right]. \tag{8}$$

Using (8) it is easy to solve for the perturbed potentials in terms of the unperturbed potentials and to compute the change in electrical enthalpy (over one wavelength) induced by the perturbation. Integration of the electrical enthalpy density

$$w = \frac{1}{2}\left(T_{ij}S_{ij} - D_iE_i\right) \tag{9}$$

over the infinite slab between $x_1 = 0$ and $x_1 = \lambda$ shows that the enthalpy change ΔW_λ due to the perturbation is a quadratic form in the two 4-vectors

$$\mathbf{t}^0 = -\mathrm{Re}\left[\left|\psi^0_{,2}\right|\right] \text{ and } \mathbf{e}^0 = \mathrm{Im}\left[\left|\mathbf{Y}\psi^0_{,2}\right|\right], \text{ namely} \tag{10}$$

$$\Delta W_\lambda = \frac{\pi}{2}A^2\left\{\left(\mathbf{t}^0 \bullet \mathbf{a} \bullet \mathbf{t}^0\right) + \left(\mathbf{e}^0 \bullet \mathbf{b} \circ \mathbf{e}^0\right) + 2\left(\mathbf{t}^0 \bullet \mathbf{c} \bullet \mathbf{e}^0\right)\right\} \tag{11}$$

where the 4 x 4 matrices **a**, **b**, and **c** depend on $\mathbf{Y}^{\mathrm{I}}$ and $\mathbf{Y}^{\mathrm{II}}$. One notes that ΔW_λ is independent of λ. The total energy change over one wavelength is the sum of the electrical enthalpy and the surface energy change [5] and is given by

$$\Delta E_{tot} = \Delta W_\lambda + \frac{\gamma A^2 \pi^2}{\lambda}, \tag{12}$$

where γ is the specific interfacial energy. The question of stability of the initially flat interface is studied by examining the sign of the total energy change, i.e., does the undulation raise or lower the energy of the system? If $\Delta E_{tot} > 0$, the interface is stable against perturbation, while if the total energy is negative for at least one perturbation wavelength, a wavy interface is preferred energetically. Clearly, at short wavelengths the surface energy contribution dominates, while the enthalpy should dominate at long wavelengths.

When the electrical enthalpy change tends to destabilize a flat surface (when $\Delta W_\lambda < 0$), a critical wavelength λ_{cr} exists such that

$$\Delta E_{tot}(\lambda) \geq 0 \text{ when } \lambda \leq \lambda_{cr}. \tag{13}$$

Quite apart from being a criterion for stability, the critical wavelength has important implications for the kinetics of any destabilizing mechanism. If diffusion is a possible mechanism for interface roughening, the larger the critical wavelength, the farther matter must diffuse, and thus longer times are required for destabilization. For large critical wavelengths, the interface may be considered stable for all practical purposes. In [5] we have examined a variety of electromechanical loading conditions and material possibilities for media I and II, e.g., vacuum/piezocrystal, dielectric fluid/piezoelectric solid, etc. In particular in [5] it is shown that for a (thin) insulating film deposited on a capacitor plate under constant imposed voltage (a configuration seen in the fabrication of thin film capacitors by anodic oxidation [11]), both mechanical and electrical loading tend to destabilize an interface; for other configurations it is shown that tangential discontinuities in **D** can promote interface stability in the presence of mechanical loading.

Almost simultaneously with the appearance of our paper [5] in the March, 1996 issue of **Proceedings of the Royal Society**, a paper by Shimizu, et al [12], entitled "The scope for studies of thin surface films on metals and alloys by transmission electron microscopy of ultramicrotomed sections" appeared in **Philosophical Transactions of the Royal Society of London.** As the title indicates, these authors used high resolution analytical electron microscopy to examine microtomed sections of films deposited on substrates. They observed that inert markers originally delineating a flat surface in barrier oxide films later delineated a roughened interface following new oxide growth. In addition, their electron micrographs show unmistakable evidence of undulating interfaces between barrier oxide films formed on low voltage type capacitor foil anodized under industrial conditions. Interfaces in electrodeposited Ni-P alloys show undulations, compositional differences, and different layers of differing crystallinity. We believe these experimental results provide striking and convincing evidence for the validity of the theoretical predictions we have presented in [5].

SELF-EQUILIBRATED DISLOCATION ARRAYS

If a solid is subjected to inhomogeneous plastic deformation followed by complete unloading, it is well known that the medium will be in a state of residual stress. On a microscopic scale there is ample experimental evidence that, in the unloaded state, the solid is filled with stationary dislocations, and that the residual stress state is the stress field due to the locked-in dislocation configuration, with the dislocation density typically varying between 10^5 to 10^8 dislocations per square centimeter. These locked-in dislocation distributions are, presumably, in equilibrium in the sense that they arrange themselves so as to minimize the potential energy of the unloaded solid. Alternatively, we may regard the Peach-Koehler force on each dislocation (the negative of the variation of the total energy of the solid with respect to an infinitesimal virtual displacement of each dislocation) as vanishing. The question arises as to whether it is truly possible to have self-equilibrated dislocation configurations in a crystal in the absence of applied stresses. This issue is more than academic, since most direct observations of dislocations in crystals are made on unloaded thin film specimens using transmission electron microscopy. An interpretation of such observations depends on an understanding of either the possibility of the existence of self-equilibrated dislocation configurations or the constraints (for example, rendering some dislocations immobile by "pinning") necessary to ensure their existence. As pointed out by Head, et al [3], there exist trivial examples of such self-equilibrated arrays, namely those distributions which produce only crystal rotations, but no elastic strains; these redundant distributions (called "impotent distributions" by Mura [13]) have no residual stress fields associated with them.

Tighe [14] has studied analytically issues associated with self-equilibrated arrays, but until Lubarda, et al, [15] numerically studied self-equilibrated arrays in elastically isotropic solids, no one had dealt with the issue of self-equilibrated arrays of large numbers of dislocations. Lubarda, et al, indeed showed that self-equilibrated dislocation structures can exist, sometimes with pinning constraints and sometimes without. Our present work has sought to extend the methodology in [15] by allowing for the inclusion of elastic anisotropy, and by adopting a more inclusive and efficient method for treating doubly periodic arrays so as to account for quite large numbers of dislocations. The latter extension is enabled by our ability to sum in closed form the interaction energy between a single dislocation and an infinite dislocation wall, which is significant in view of the long range nature of individual dislocation stress fields.

As the ability to determine equilibrated dislocation configurations involves minimizing the interaction energy between all pairs of dislocations, we have shown that in an infinite solid of arbitrary anisotropy the elastic interaction energy between two straight dislocations at (x_1, y_1) and (x_2, y_2), respectively, (Burgers vectors $\mathbf{b}_1$ and $\mathbf{b}_2$) is

$$E_{int} = \frac{1}{2\pi i}\sum_{\alpha=1}^{6} \pm L_{i\alpha} L_{s\alpha} b_{1i} b_{2s} \ln \frac{(x_1 - x_2) + p_\alpha (y_1 - y_2)}{R} \tag{14}$$

where p_α and $\mathbf{L}_\alpha$ are the anisotropic elastic Stroh [8] eigenvalues and Stroh auxiliary eigenvectors, respectively, and the $\pm$ sign is the sign of Im (p_α). R is an arbitrary constant with dimensions of length; it is of no importance as regards minimization of the total interaction energy. As with Lubarda, et al, we note that the most natural way of studying equilibrium arrays of large numbers of dislocations is to study doubly periodic arrangements of straight dislocations and to focus our attentions on a single repeating "cell". Figure 2 depicts nine such cells of a doubly periodic arrangement with three dislocations per cell, merely for purposes of illustration. In practice we compute with hundreds of dislocations per cell.

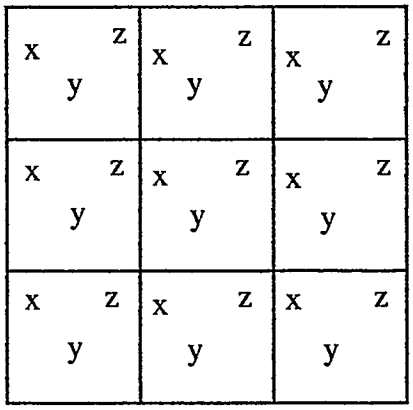


Figure 2. A portion of a doubly periodic arrangement of dislocations.

If we wish to compute the interaction energy for the doubly periodic arrangement shown in Figure 2, it suffices to focus our attention on one unit cell and to compute the interaction of the dislocation labeled "x" with the infinite vertical walls of "x", "y", and "z" dislocations throughout all space. However, and this point seems to have gone unnoticed, the interactions of dislocation x with its own wall or with all other "x" walls is independent of where dislocation "x" sits in the cell, i.e., the variation of this interaction with a virtual displacement of dislocation "x" vanishes; obviously a similar result holds for "y" and "z" dislocations. Thus we need only focus our attentions on the interaction between dislocation x in one cell with all parallel "y" and "z" walls, and on the interaction between dislocation y in the same cell with all parallel "x" and "z" walls, etc. (being careful to count the interactions only once). The second computational savings results from our ability to show that the interaction elastic energy between a single dislocation at (X,Y) in the cell under consideration with a vertical wall containing a dislocation at (X',Y') in the same cell can be expressed exactly as

$$E_{int} = \frac{1}{2\pi i} b_i^X b_s^{X'} \sum_{\alpha=1}^{6} \pm L_{i\alpha} L_{s\alpha} \ln\left[\sin \frac{\pi\left((X - X') + p_\alpha (Y - Y')\right)}{p_\alpha h}\right] \qquad (15)$$

where h is the periodic vertical spacing. The position-dependent part of the total interaction energy per cell is computed by summing terms such as (15) with X' replaced by X' + kH (k= $\pm 1, \pm 2, \ldots \pm \infty$; H is the horizontal periodicity). Although these infinite sums formally diverge, we have shown that the position-dependent part is well-defined and stems from a series whose terms decay exponentially as $|k| \rightarrow \infty$. Thus, we have available the ingredients for an energy minimization scheme which rigorously handles **all** wall dislocations in the vertical direction and whose decay with separation along the horizontal is sufficiently rapid that only a limited number of walls outside the cell under consideration need be considered for good convergence to be obtained.

The most significant results obtained using (15), which are presented with the lecture accompanying this work, involved filling a unit cell of a doubly periodic structure with between 100 and 400 dislocations, distributed on equally spaced parallel horizontal slip planes, with the number of dislocations on a given slip plane assigned by a random number generator, and with the sign (positive or negative) of the dislocations on any slip plane assigned at random. With an equal number of positive and negative dislocations, under zero applied stress an equilibrium distribution **was always found** using a conjugate gradient algorithm. The equilibrium distributions obtained are apparently different for different assigned starting configurations, which illustrates the non-uniqueness of the equilibria obtained. Once the equilibria were found, we examined the Peach-Koehler force on each dislocation in the repeating cell at equilibrium. For typical dislocation densities (which set the length scale for the problem) all Peach-Koehler forces were well below accepted values of lattice friction stresses, indicating that self-equilibrated dislocation distributions under no external stress are the rule rather than the exception. With a 50-50 mix of positive and negative dislocations, no constraints such as "pinning" selected dislocations are necessary to produce equilibrium. The non-uniqueness of these distributions is perhaps at the root of the non-uniqueness of experimental observations of locked-in dislocation configurations in the unloaded state. Like Lubarda, et al, we found that when the dislocation mix is far from a 50-50 mix of positive and negative edge dislocations, the equilibria obtained involve separation of the initial distribution into dislocation walls.

With significantly large numbers of dislocations present, it has yet to be shown that self-equilibrated dislocation distributions occupying only a **finite** portion of either a bounded or an unbounded solid can be found without constraining certain dislocations by pinning. It is possible that by dealing with partial dislocations separated by stacking faults such distributions could be found without invoking pinning constraints. Finally, examining these same issues within the context of the theory of continuously distributed dislocations is indeed worthy of attention, and will be pursued in the future.

ACKNOWLEDGMENTS

This work was performed under a grant to Stanford University from the United States Department of Energy.

REFERENCES

1. H. GAO, "Morphological Instabilities Along Surfaces of Anisotropic Solids," in *Modern Theory of Anisotropic Elasticity and Applications*, pp. 139 - 150, eds. J. J. Wu, T. C. T. Ting, and D. M. Barnett, SIAM, Philadelphia (1991).

2. N. CHIEN and H. GAO, "Stress-Induced Roughening Along Surfaces of Piezoelectric Materials," *Proceedings 11th Symposium on Energy Engineering Sciences*, pp. 16 - 23, Argonne National Laboratory (May, 1993).

3. A. K. HEAD, S. D. HOWISON, J. R. OCKENDON and S. P. TIGHE, "Mathematical Modeling of Dislocation Plasticity," *Physica Scripta*, **T44**, pp. 135 -136 (1992).

4. R. J. ASARO and W. A. TILLER, "Interface Morphology Development During Stress Corrosion Cracking: Part 1. Via Surface Diffusion," *Metallurgical Transactions*, **3**, pp. 1789 - 1796 (1972).

5. N. CHIEN, H. GAO, G. HERRMANN, and D. M. BARNETT, "Diffusive Instabilities Induced by Electromechanical Loading," *Proceedings of the Royal Society, London*, **A**, **452**, pp. 527 - 541 (1996).

6. H. F. TIERSTEN, *Linear Piezoelectric Plate Vibrations*, Plenum Press, N. Y. (1969)

7. D. M. BARNETT and J. LOTHE, "Dislocations and Line Charges in Piezoelectric Insulators," *Physica Status Solidi (b)*,**67**, pp. 105 - 111 (1975).

8. A. N. STROH, 'Steady State Problems in Anisotropic Elasticity," *Journal of Mathematics and Physics*, **41**, pp. 77 - 103 (1962).

9. J. LOTHE and D. M. BARNETT, "Integral Formalism for Surface Waves in Piezoelectric Crystals: Existence Conditions", *Journal of Applied Physics*, **47**, pp. 1799 - 1807 (1976).

10. J. LOTHE and D. M. BARNETT, "Further Development of the Theory for Surface Waves in Piezoelectric Crystals," *Physica Norvegica*, **8**, pp. 239 - 254 (1977).

11. L. I. MAISSEL and R. GLANG, *Handbook of Thin Film Technology*. McGraw - Hill, N. Y. (1970)

12. K. SHIMIZU, K. KOBAYASHI, G. E. THOMPSON, G. C. WOOD, and P. K. SKELDON, "The Scope for Studies of Thin Films on Metals and Alloys by Transmission Electron Microscopy of Ultramicrotomed Sections," *Philosophical Transactions of the Royal Society, London*, **A**, **354**, pp. 213 - 235 (1996).

13. T. MURA, *Micromechanics of Defects in Solids, 2nd Edition*, Martinus Nijhoff Publishers, Dordecht, Netherlands (1987).

14. S. P. TIGHE, "Continuum Modeling of Dislocations", *Thesis for transferal to D. Phil. status in Mathematics*, Michaelmas Term, Oxford University (1990).

15. V. A. LUBARDA, J. A. BLUME, and A. NEEDLEMAN, 'An Analysis of Equilibrium Dislocation Distributions," *Acta Metallurgica et Materialia*, **41**, pp. 625 - 642 (1993).

ELASTIC-PLASTIC FRACTURE MECHANICS OF STRENGTH-MISMATCHING

D. M. Parks, S. Ganti and F. A. McClintock

Massachusetts Institute of Technology
Cambridge, MA 02139, USA.

ABSTRACT

Approximate solutions to stress-fields are provided for a strength-mismatched interface crack in small-scale yielding (SSY) for non-hardening and low hardening materials. Variations of local deformation intensities, characterized by a $J-$type contour integral, are proposed. The softer material experiences a higher deformation intensity level, J_S, while the harder material sees a much lower deformation intensity level, J_H, compared to that obtained from the applied J near the respective homogeneous crack-tips. For a low hardening material, the stress fields are obtained by scaling from an elastic/perfectly-plastic problem, based on an effective mismatch, M_{eff}, which is a function of mismatch, M, and the hardening exponent, n. Triaxial stress build-up is discussed quantitatively in terms of M. The influence of strength-mismatch on cleavage fracture is discussed using Weibull statistics.

INTRODUCTION

Conventional fracture mechanics deals with predicting fracture in large structures by testing small specimens. It relies on the notion that a few parameters (eg: $K_I - T$, $J_I - Q$) characterize the crack-tip stress and deformation fields. Crack-initiation and growth are assumed to occur when these parameters reach critical values (which depend on the geometry and loading) (eg: K_{Ic}, J_{Ic}). In such cases, material resistance is characterized by these few parameters. While this technology, developed for homogeneous materials, is being used for welded joints based more on empiricism than on fundamentals, effective use of this technology is inhibited by the geometrical and material inhomogenities in a weld. The growing concern for quantifying structural resistance in welded structures can be seen from the formation of numerous European and Japanese groups (EPI, FCWRP etc.) to study weld-related fracture problems.

A schematic of a weld is shown in Fig. 1. Complications in the analysis of welded joints include the strength mismatch, location and orientation of the crack with respect to the fusion line, and geometry of the weld. The mismatch in strength between a weld metal and its base plate is termed under-matching and over-matching, respectively,

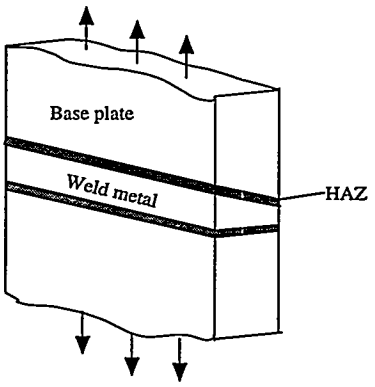


Figure 1: Schematic of a weld.

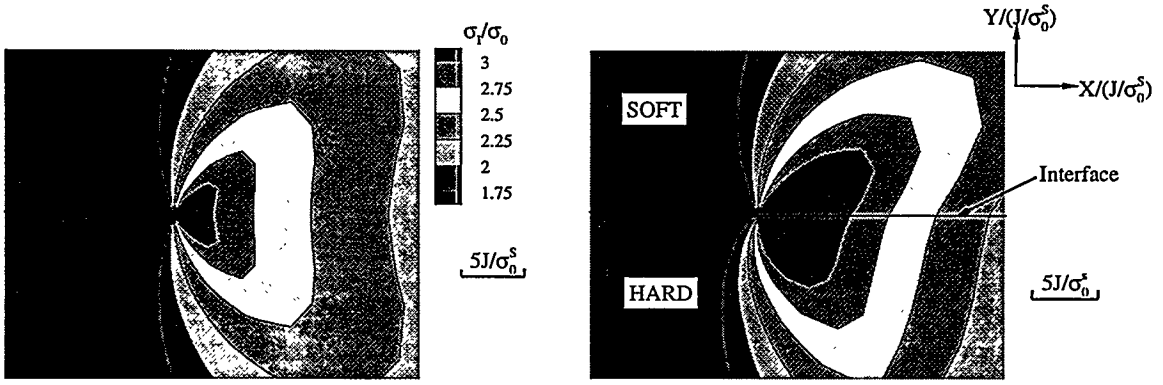


Figure 2: Contours of maximum principal stress for homogeneous and mismatched specimens. (a) Homogeneous (b) Mismatched, $M = 1.5$.

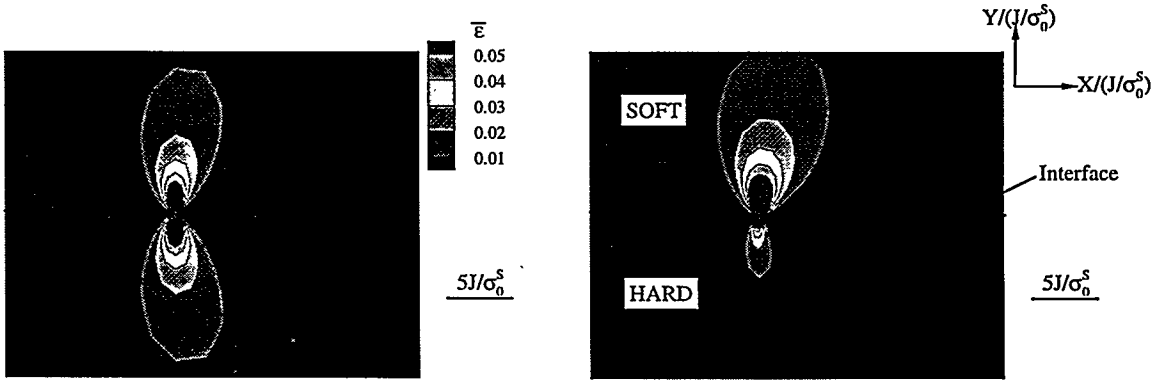


Figure 3: Contours of equivalent plastic strain for homogeneous and mismatched specimens. (a) Homogeneous (b) Mismatched, $M = 1.5$.

if the weld metal yield strength is lower or higher than the base material. Unwanted mis-matching can occur when either the strength of the base or the weld metals deviate from the minimum specified. The analysis is further complicated for a crack located near the fusion line due to the strength and microstructural gradients across the heat affected zone (HAZ).

The variation in properties across the weld influences the development of plastic deformation and stress, and hence the relationship between the resistance and applied loading. At a more fundamental level, it is not clear how measures like the J-integral and $CTOD$ characterize the crack-tip fields in mismatched specimens, even in small scale yielding (SSY). Studies by Koçak et al. [1] and Toyoda et al. [2, 3, 4] show that toughness of the Heat Affected Zone (HAZ) as characterized by $CTOD_c$ decreases with increasing weld metal strength. Due to asymmetric plastic zone development, the slope of J vs Δa decreased for an interface crack, compared to the corresponding base metal [1]. As stated by Koçak et al., the "apparent toughness" of the HAZ depends not only on the microstructure of hardened zones, but also on the mechanical properties of the adjacent material. Figure 2 shows the contours of principal stress and strain for homogeneous and mismatched ($M = 1.5$) specimens at the same far-field J for an interface crack located along the fusion line[1]. Higher triaxial stress builds up in softer material, and over larger regions, compared with 'homogeneous' crack-tip conditions. Deformation is preferentially focussed in a lobe within the soft material. Hence it is not clear what parameters characterize the crack-tip stress and deformation fields. Further, to understand how the crack-tip stress and deformation fields drive the microstructural fracture processes (eg: cleavage, void growth, interface decohesion), it is necessary to construct models of crack-tip fields which are informative, accurately descriptive and tractable.

Here, stress fields near a longitudinal interface crack subjected to Mode-I loading under well-contained yielding are analyzed for non-hardening and low hardening materials. Local measures of stress and deformation intensity are constructed, and their effects on cleavage fracture are studied.

SLIP-LINE MODEL FOR MISMATCHED SPECIMENS NON-HARDENING CASE

For a deeply-cracked homogeneous structure with elastic/perfectly-plastic behavior, subjected to Mode-I loading, Du and Hancock [5] showed that the modified Prandtl field shown in Fig. 4a provides a good approximation for the stress-field ahead of the crack-tip. They showed that geometry-related effects characterized by the T-stress can be modelled as a change in the fan angle, δ, from that of the Prandtl field. The peak normal stress occurs ahead of the crack-tip and is given by $\frac{\sigma_{\theta\theta}}{\sigma_0}\,|_{\theta=0} = (2+\pi-2\delta)/\sqrt{3}$, while the maximum triaxiality there is given by $\frac{\sigma_{kk}}{3\sigma_0}\,|_{\theta=0} = (1+\pi-2\delta)/\sqrt{3}$. The classical Prandtl field can be recovered by setting $\delta = 0$.

A variation of the homogeneous slip-line field solution proposed for small mismatches by Ganti et al. [6] is shown in Fig. 4b. Due to strength mismatch, $M \equiv \frac{\sigma_0^H}{\sigma_0^S}$, which is defined as the ratio of yield strengths between the hard (H) and soft (S) materials, the normalized $T-$stresses are related by $T = \tau^S \sigma_0^S = \tau^H \sigma_0^H$. The stress fields are

[1]Here, the HAZ is neglected in the model. This is valid for characterizing the stress and strain fields as long as the thickness of HAZ is much smaller than the plastic zone size.

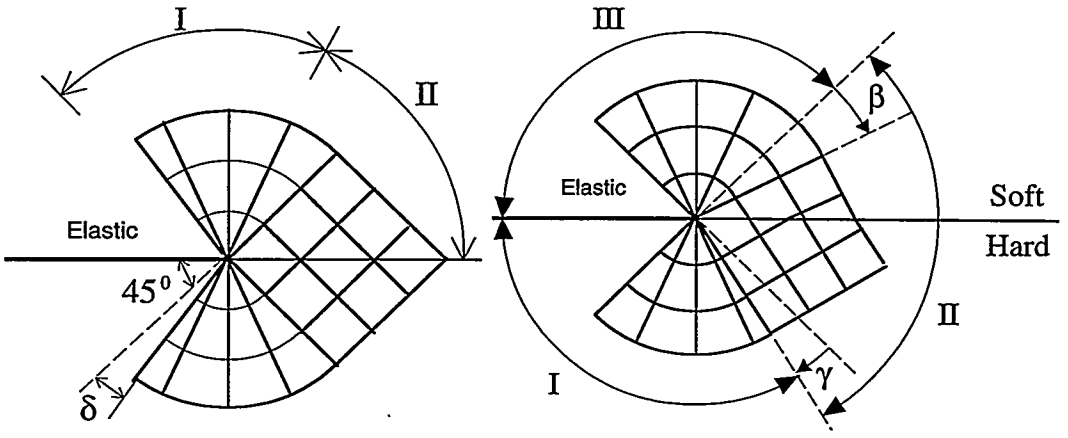


Figure 4: Modified Prandtl field for a (a) homogeneous specimen, (b) mismatched specimen.

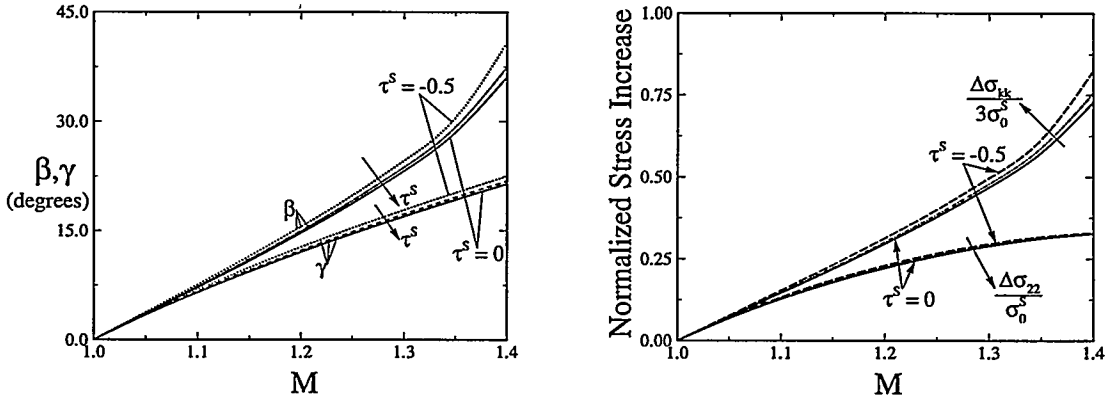


Figure 5: (a)Variation of mismatch field angles, β and γ, with mismatch, M, for various normalized T-stresses. (b)Increase in mean and normal stresses in the soft material near the interface with mismatch, M. $\Delta\sigma_{ij} = \sigma_{ij}^{mis} - \sigma_{ij}^{hom}$

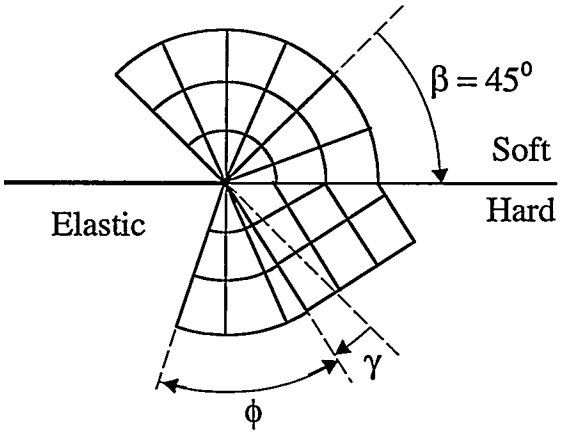


Figure 6: Modified Prandtl field for highly mismatched specimens ($M > 1.421$).

characterized by the angles β and γ which are functions of mismatch, M (Fig. 5a). The ratio of σ_{12}/σ_{22} at the interface increases with increasing mismatch and saturates at a value of $1/(1+\frac{3\pi}{2}-2\delta)$. The increases in triaxial and interface-normal stresses in the soft material near the interface are plotted in Fig. 5b. The effect of $T-$stress on the mismatch angles is of second order, as can be seen from Fig. 5a, though there is considerable reduction in the normal stress ahead of the crack tip. The fields described by Fig. 4b exist when $M \leq 1.421$. For $M > 1.421$, the fan in the soft material is completely developed, with $\beta = 45°$ and a slip-line field shown in Fig. 6 exists.

The non-hardening angular stress distributions, which can be derived from the proposed slip-line solutions, show three distinct regions for small mismatches ($M \leq 1.421$) as shown in Figures 7(a) and 7(b). In regions I and III, the stress-field is same as that of the corresponding homogeneous material, while in region II, mismatch fields differ from the homogeneous fields. Figures 7a and 7b compare the mismatch fields (14–15) with the finite element solutions. Details of the finite element solutions are given by Ganti et al [6]. The slip-line solutions are in good agreement with the corresponding finite element solutions. There is an increase in peak stress triaxiality, and more deformation is forced into the lobe of the softer material with increasing mismatch.

LOW HARDENING MATERIALS

The J-integral gives a combined measure of deformation and stress intensities in homogenoeus hardening materials. As noted in Figures 2–3, soft (hard) material experiences higher (lower) deformation than the corresponding homogeneous material, suggesting that different intensity parameters govern the local fields in each material. Equivalent J-type contour integrals, J_S and J_H, are developed which govern the stress intensity in the soft and hard materials, respectively. A schematic of the contour integrals is shown in Figure 8a. Since the interface is parallel to the crack plane, J_S and J_H are related to J by

$$2J = J_H + J_S. \tag{1}$$

In the case of homogeneous materials, $J_S = J_H = J$. Figure 8b shows the normalized radial variation of J_S with mismatch. The materials used have similar elastic properties and hardening exponent, n, but differing yield strengths. While J showed only small variation with r, the radial variation of J_H and J_S extends to tens of $CTOD$s, giving credibility to the existence of the new fields. Characterization of local deformation intensity parameters quantitatively gives the amount of shielding in the harder material and, conversely, the increased propensity for fracture in the softer material. The degree of deformation redistribution increases with increasing M.

Approximate mismatch fields for low hardening materials are developed based on an effective mismatch and defining an equivalent elastic-plastic problem for which the angular distributions are already discussed. The differing levels of deformation on either side of the interface due to the local J's (and the mismatch itself) change the effective mismatch experienced by the crack-tip. If the SSY fields characterized by J_H and J_S are scaled to satisfy approximately interface traction continuity, we can define effective mismatch as follows:

$$M_{eff} = \frac{\sigma_{\theta\theta}^{SSY(H)}\left(r/(J_H/\sigma_0^H), \theta = 0°\right)}{\sigma_{\theta\theta}^{SSY(S)}\left(r/(J_S/\sigma_0^S), \theta = 0°\right)}, \tag{2}$$

each stress being evaluated in the respective homogeneous material. For non-hardening materials, $M_{eff} = M$. Following an approach similar to that of the non-hardening

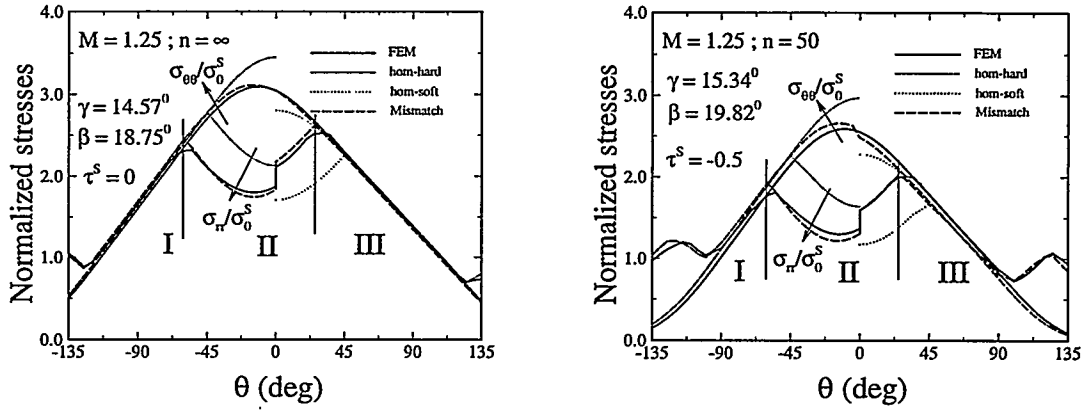


Figure 7: Circumferential variation of radial and hoop stress for (a) $\tau = 0$ and $M = 1.25$ at $r = 2J/\sigma_0^S$ (b) $\tau^S = -0.5$ and $M = 1.25$ at $r = 2J/\sigma_0^S$.

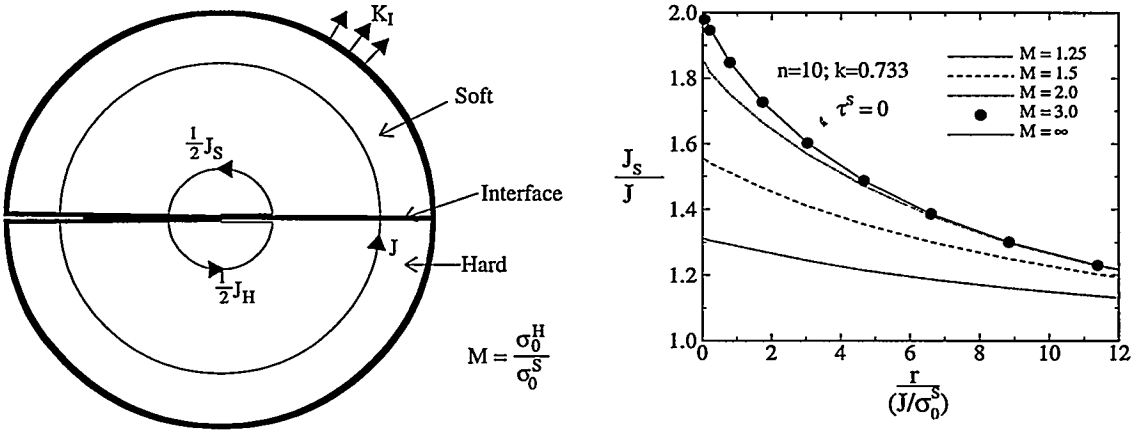


Figure 8: (a)Definition of contour integrals for mismatched specimens. (b)Normalized radial variation of J_S for $\tau = 0$.

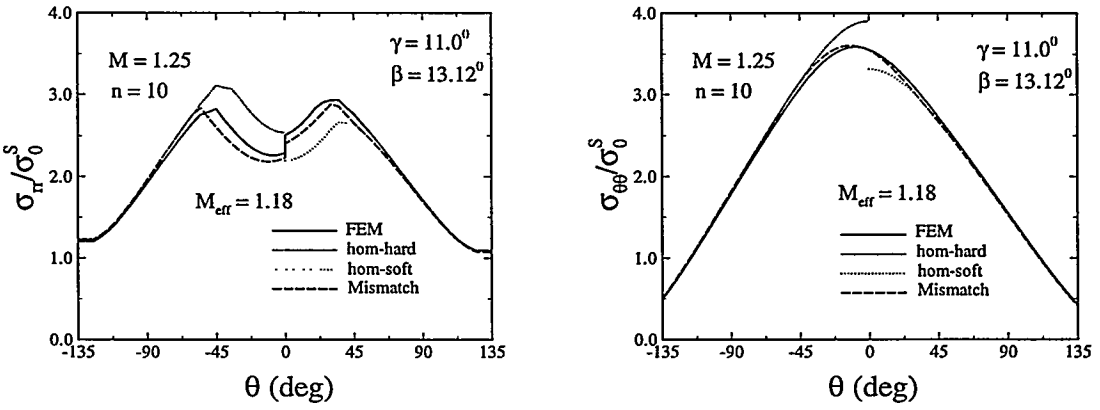


Figure 9: Circumferential variation of normalized radial and hoop stresses at $r = 2J/\sigma_0^S$ for $M = 1.25$. (a) normalized radial stress (b) normalized hoop stress.

materials, the approximate stress distributions in the soft and hard materials are developed. In region I, the solution is the homogeneous SSY one scaled by J_H, and in region III, the local homogeneous solution is scaled by J_S. In region II, the deformation intensities J_H and J_S combine to scale the fields, while the angular distribution is governed by mismatch angles, β and γ, calculated as if the materials are elastic/perfectly-plastic with an effective mismatch of M_{eff}. Figures 8–9 show the simulated and finite element results of the stress fields for different mismatches at the normalized distance $r/(J/\sigma_0^S) = 2$. The agreement between the finite element solutions and the approximate fields is excellent.

CLEAVAGE STUDIES

In the brittle to ductile transition regime for steels, the material resistance is governed by the competition between fracture by cleavage and ductile tearing. Higher stressed volumes as in mismatched specimens (Fig. 2) can bias fracture towards catastrophic failure. Here, neglecting HAZ, we study the effects of mismatch on the cleavage probability of the base metal using weakest link statistics. Assuming a three-parameter Weibull distribution for the strength distribution of the carbide particles [7], the relation between the critical J for cleavage of base metal with an overmatch, J_{Ic}^M, and that of homogeneous base metal, J_{Ic}^H, in contained yielding is given as

$$\frac{J_{Ic}^M}{J_{Ic}^H} = \frac{\left(\int^{A_H} \left(\frac{\sigma-\sigma_u}{\sigma_0}\right)^m dA\right)^{0.5}}{\left(\int^{A_M} \left(\frac{\sigma-\sigma_u}{\sigma_0}\right)^m dA\right)^{0.5}}, \tag{3}$$

where σ_u, σ_0 and m are the parameters for Weibull distribution, A_H is the cross-sectional area of the plastic zone in the homogeneous material, and A_M is the area on the soft side of the mismatched specimen. Equation 3 shows that the critical J is a function of the high-stressed volume, the stress distribution and the material parameters. Figure 11 shows the contours of the stressed areas for the homogeneous and mismatched materials. Note that for homogeneous case, only one-half the area is shown. Figure 12 shows the normalized J_{Ic}^M for mismatched materials for different threshold and Weibull exponents. The cleavage stress decreases with mismatch for large threshold stresses, while there is a moderate increase in J_{Ic} for small threshold stress and mismatch.

CONCLUSIONS

The applicability of homogeneous fracture mechanics to strength-mismatched interface cracks is addressed. New crack-tip stress fields and relevant deformation intensity measures are developed for interface cracks in strength-mismatched specimens. The stress and deformation intensities locally experienced by materials are different from those experienced in homogeneous specimens of either material, when mismatched and homogeneous specimens are loaded to a given applied J. The crack-tip fields in the soft material near the interface show higher normal and mean stresses than the corresponding homogeneous material. Our analysis shows that triaxial constraint increases almost linearly with mismatch, reaching a saturation at $M = 1.421$ in the non-hardening materials.

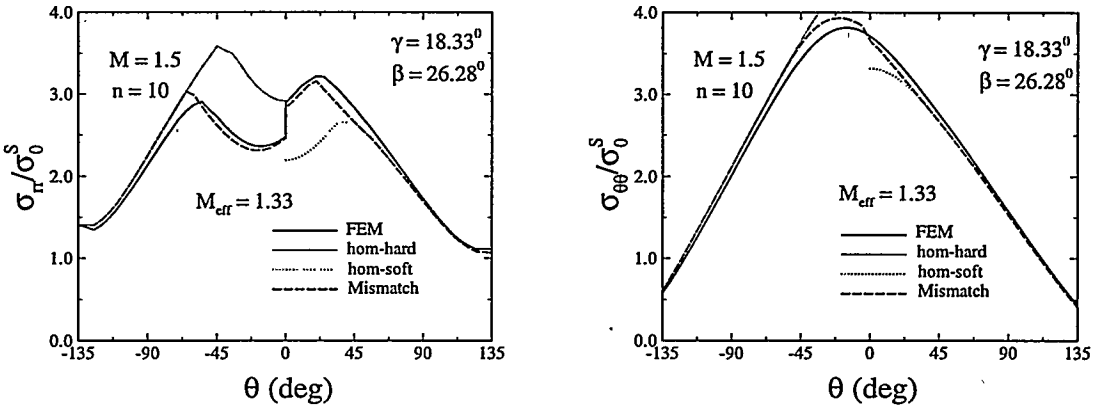


Figure 10: Circumferential variation of normalized radial and hoop stresses at $r = 2J/\sigma_0^S$ for $M = 1.5$. (a) normalized radial stress (b) normalized hoop stress.

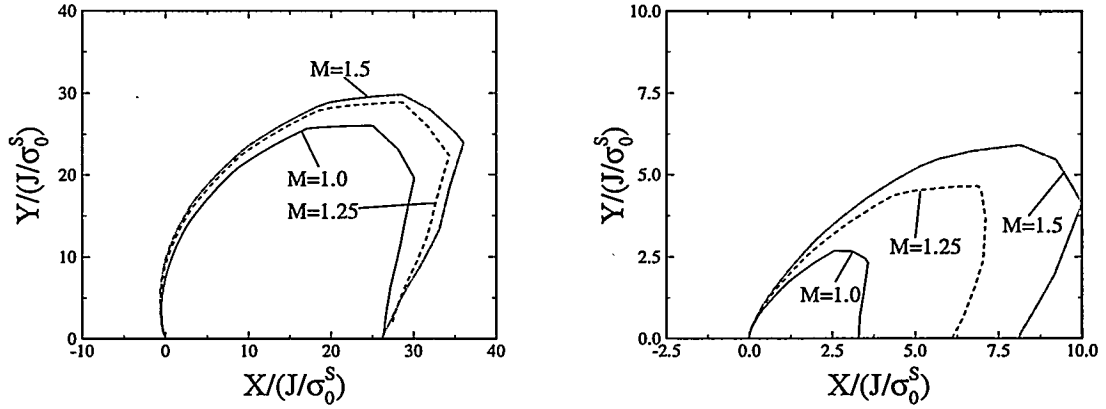


Figure 11: Contours of maximum principal stress for $M = 1, 1.25, 1.5$. (a) $\sigma_I/\sigma_0^S = 2.0$ (b) $\sigma_I/\sigma_0^S = 3.0$.

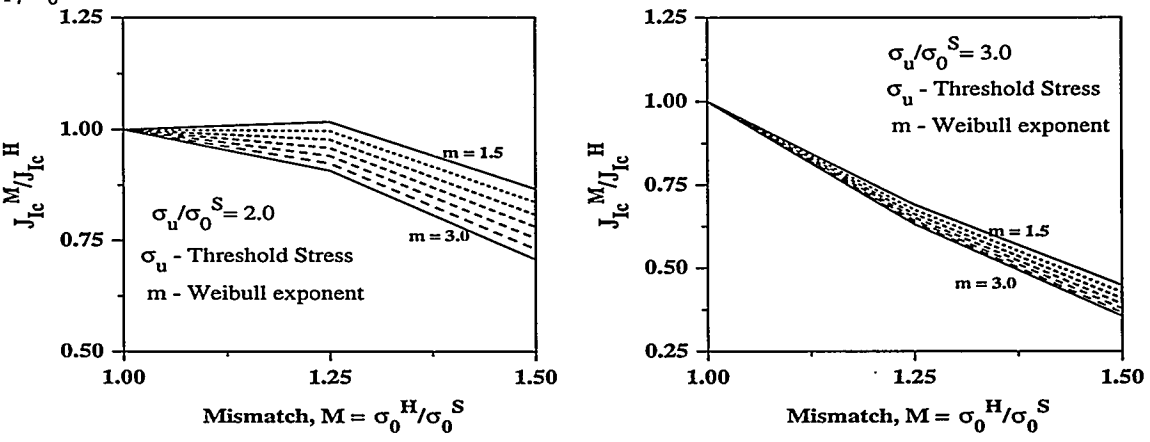


Figure 12: Normalized variation of cleavage J for different threshold stresses. (a) $\sigma_u/\sigma_0 = 2.0$ (b) $\sigma_u/\sigma_0 = 3.0$.

Cleavage studies conducted on the interface cracks in contained yielding show considerable effect of the toughness of the joint on the mismatch, M. In particular, it decreases with increasing M, eventhough, the total area of the plastic zone in the soft material is lower than that in the homogeneous material. This effect is prominent for larger threshold stresses and higher mismatches. Further results in characterization of crack-tip mismatch fields, and in modeling their effects on fracture, are also underway and will be reported elasewhere.

ACKNOWLEDGMENTS

This work was supported by the Office of Basic Energy Sciences, Department of Energy, under Grant $\#DE-FG02-85ER13331$. Computations were performed on Hewlett Packard workstations obtained under the DOE grant and under NSF Grant $\#DDM-8914161$. The ABAQUS finite element program was made available under academic license from Hibbitt, Karlsson, and Sorensen, Inc., Pawtucket, RI. We would like to thank Dr. Simona Socrate and Ms. Hong Dai for the use of their computer programs.

REFERENCES

[1] KOÇAK, M., CHEN, L., TERLINDE, G., GNIRSS, G. and SCHWALBE, K. H., "CTOD testing of HAZ and analysis of pop-in behavior," *Proc. of the 7th Int. Conf. on Offshore Mechanics and Arctic Engineering*, ASME, 1988, pp. 297–304.

[2] TOYODA, M. and SATOH, K., "Fracture toughness of welds: The incidence of mechanical heterogeneity in the correlation of toughness parameters," *Welding in the world*, Vol. 22, No. 3/4, 1984, pp. 76–86.

[3] TOYODA, M., MINAMI, F., RUGGIERI, C., THAULOW, C. and HAUGE, M., "Fracture Property of HAZ-Notched weld joint with mechanical mis-matching-Part I," in *Mismatching of welds, ESIS 17*, Eds. Schwalbe, K. H., Koçak, M., Mechanical Engineering Publications, London, 1994, pp. 399–415.

[4] THAULOW, C., PAAUW, A. J., HAUGE, M., TOYODA, M. and MINAMI, F., "Fracture Property of HAZ-Notched weld joint with mechanical mis-matching-Part II," in *Mismatching of welds, ESIS 17*, Eds. Schwalbe, K. H., Koçak, M., Mechanical Engineering Publications, London, 1994, pp. 417–432.

[5] DU, Z. Z., and HANCOCK, J. W., "The Effect of Non-Singular Stresses on Crack-tip Constraint," *Journal of the Mechanics and Physics of Solids*, Vol. 39, 1991, pp. 555–567.

[6] GANTI, S., PARKS, D. M. and McCLINTOCK, F. A "Analysis of Strength-Mismatched Interface Cracks in SSY," to be published in *Mismatch'96*, Eds. Schwalbe, K. H., Koçak, M.

[7] WANG, Y. Y., "A two parameter characterization of elastic-plastic crack-tip fields and applications to cleavage fracture," *Ph. D.* Thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology, 1991.

DAMAGE MECHANICS – FAILURE MODES

Dusan Krajcinovic and Milena Vujosevic

Arizona State University, Tempe, AZ 85287-6106, USA

Abstract

The present study summarizes the results of the DOE sponsored research program focused on the brittle failure of solids with disordered microstructure. The failure is related to the stochastic processes on the microstructural scale; namely, the nucleation and growth of microcracks. The intrinsic failure modes, such as the percolation, localization and creep rupture, are studied by emphasizing the effect of the micro-structural disorder. A rich spectrum of physical phenomena and new concepts that emerges from this research demonstrates the reasons behind the limitations of traditional, deterministic, and local continuum models.

INTRODUCTION

All failure modes during the inelastic deformation of materials are classified into two classes: intrinsic or material failures and extrinsic or structural failures. A failure will be referred to as being intrinsic if its threshold can be defined in terms of the effective material properties. The thresholds of extrinsic failure modes depend also on the specimen size and shape. The threshold and type of failure depends on many causes. Temperature, strain rate, embrittleing effect of chemical reactions are some of the most prominently studied agents which affect the brittle to ductile transition. This study is concerned only with the brittle and quasi-brittle failures which occur in solids with a relatively modest cohesive strength as a result of damage evolution.

Damage evolution is caused by the nucleation of new microcracks and growth of the already existing microcracks. The pattern and type of damage evolution is controlled by the stress concentrations at microstructural heterogeneities (hot spots) and/or by existence of internal surfaces of inferior cohesive strength (weak links). In micro-heterogeneous (damage tolerant) solids made of materials with strongly dissimilar fracture strength (fibrous and particulate composites) the microcrack growth can be impeded by the strong phase. The microcrack growth is a basic damage evolution mode in damage sensitive materials, which are characterized by statistically homogeneous microstructures.

A typical engineering material has a disordered microstructure. The spatial distribution of hot spots and weak links is not deterministic. Consequently, the distribution of defects is not deterministic either. As long as the concentration of microcracks is dilute the microstructure is often statistically homogeneous on a rather small scale and the specimen response does not depend on the exact locations of defects. At substantial microcrack concentrations the local fluctuations of stress field depend strongly on the microcrack interaction and, therefore, on the distribution of the distances separating closely spaced microcracks. Additionally, the failure can often depend on the

largest defect or on the smallest distance between interacting defects. It seems, therefore, logical that a purely deterministic (continuum) model of failure may in many cases lead to a poor estimate of the failure threshold. The disorder is annealed in the case when the external stimuli (such as temperature, earthquake and wind loads, etc.) fluctuate randomly. In summary, the disorder of the microstructure of brittle materials is ubiquitous and its effect on the failure mode and the threshold is as a rule significant.

CHARACTERISTIC LENGTHS

To quantify the difference between particular failure modes it is necessary to introduce several characteristic lengths. Some of these lengths are easily singled out. For example, the specimen size L is a characteristic length which plays an important role in brittle fractures. More than one characteristic specimen length is needed when the failure threshold depends on the specimen shape.

The linear size L_{rve} of the smallest volume, referred to as the representative volume element [1], within which the material is statistically homogeneous, is also a characteristic length. Another characteristic length is the size ξ of the largest cluster of interacting microcracks. The cluster is defined [2]: (a) either as a string of concatenated microcracks or (b) an ensemble of microcracks which are close enough to affect their growth pattern and rate by direct interaction. A macrocrack and a shear band are two most commonly encountered examples of two respective types of clusters. The acoustic emission test [3] is a useful diagnostic procedure for the determination of the failure mode. The distance L_{ae} between the two consecutive acoustic signals can be used to determine the damage evolution mode and the failure type.

For computational efficiency a system is often subdivided into many sub-systems (finite elements) within which the considered fields are determined using an ad hoc selected interpolation rule. However, the discretization is not only the matter of efficiency since the defects in a sub-volume see the defects in the adjacent sub-volumes only through the effective moduli. Hence, the defects belonging to different sub-volumes are not correlated. Local fluctuations of the stress, strain and damage fields in a sub-volume will affect the response in the rest of the specimen only if the material within the considered sub-volume is statistically homogeneous, i.e. if the conditions

$$\xi \leq L_{rve} \leq L \quad \text{and} \quad L_{rve} \geq L_{tr} \tag{1}$$

are satisfied. In (1) L_{tr} is the wave length of the tractions applied to the external surfaces of the sub-volume. First of two inequalities (1) renders the material statistically homogeneous while the second condition secures homogeneous response. If the inequalities (1) are satisfied the effective continuum, obtained by mapping representative volume elements on material points, is local.

The distribution of distances between the consecutive acoustic signals is uniform, $p(L_{ae}) = const.$, in damage evolution processes controlled by microcrack nucleation, i.e. in materials with a random distribution of weak spots. A propagation of a single crack is characterized by the fact that the majority of signals are clustered at the crack tip. The distances between pairs of consecutive acoustic signals are equal to the resolution length, $L_{ae} \approx l$. The deformation process is dominated by the interaction induced microcrack growth (cooperative effect) if the distribution of lengths L_{ae}: (a) favors small distances in the direction of the normal to a rather thin elongated region, and (b) is perfectly random in the exterior of this region. The material within the localized region is non-local and non-homogeneous while the material in the exterior volume remains local and homogeneous. The damage evolution and the deformation process are dominated by the kinetics of the largest cluster of interacting microcracks (shear band), i.e. by the rate at which the correlation length ξ increases.

On the microscopic scale the damage evolution depends on the outcome of the competition between the microcrack nucleation and microcrack growth. The intrinsic brittle to quasi-brittle transition, defined as a cross-over from the processes controlled by a single crack to processes dominated by many microcracks, depends on: (a) sign of the long range stresses and (b) the bandwidth of the distribution of fracture strengths of the constituent microstructural phases [4],[2]. The traditional, deterministic definition of the brittle to quasi-brittle transition [5], based on the macroscopic phenomenology, is strongly dependent on the specimen size and shape and is, therefore, not a material parameter.

EXTRINSIC FAILURE MODES

Failure of a micro-homogeneous (rock, concrete, glass or ceramic) specimen subjected to a uniaxial tensile stress is a quintessential example of a perfectly brittle failure. The failure occurs when the largest and preferentially oriented (perpendicular to the tensile axis) existing crack starts propagating. The energy barriers in these materials are insufficient to trap the crack rendering the propagation of the growing crack unstable. If the specimen is subjected to a homogeneous stress field it will fail in an almost instantaneous cleavage mode. The force-displacement relation remains linear until the failure is reached indicating that the accumulation of distributed damage is minimal. The failure threshold is a variate dependent on the specimen size since the probability of finding a large pre-existing crack is proportional to the volume. The statistics of extremes represents an appropriate framework for the determination of the failure threshold as a function of stress and specimen size [6],[7].

The failure of an unconfined specimen subjected to uniaxial compression in a longitudinal splitting mode also belongs to the class of extrinsic failures. The growth of cracks, which are roughly parallel to the compression axis, is initially marginally stable [8]. As soon as the largest of these cracks reaches its critical length the specimen splits into two or more parts. A modest concentration of accumulated damage does not affect the specimen failure. The failure threshold can be predicted from Griffith's. The specimen failure depends on the specimen shape (ratio between its length measured along the compression axis and its width in the lateral dimension). Specimen will split only when this aspect ratio is larger than one (slender cylinders). At small aspect ratios the part near the lateral surface will slab off while the rest of the specimen will resist further load increments.

INTRINSIC FAILURE MODES

The threshold of an intrinsic failure mode depends only on the average stresses, average strains and effective material properties. For a material to fail in an intrinsic mode it must be statistically homogeneous (invariant with respect to translations) on a scale of the sub-system, i.e. the conditions (1) must be satisfied. Conditions (1) are satisfied in the mean-field regime, characterized by a dilute concentration of microcracks which renders the effect of the direct interaction of microcracks on the macroscopic response insignificant.

Tensile Failure of a Damage Tolerant Solid

The damage accumulated in damage tolerant materials subjected to a long range tensile field can be substantial. In a stress controlled test a specimen will fail when the (1111) component of the effective tangential stiffness (or elastic) tensor vanishes. In a elongation controlled test the specimen response may crossover to the softening regime signaling the loss of homogeneity reflected as the loss of ellipticity.

Numerical simulations on two-dimensional, damage tolerant, frozen lattices [9] demonstrate the existence of the size effect. The response in the hardening regime can be very accurately estimated by the mean field model [10] adjusted by the size effect term derived in [9]. The response in the softening regime is controlled by large defect clusters and the stress distribution is multifractal [2]. Large scatter of results for different physical realizations of same statistics is indicative of the dominance of extreme statistics. Consequently, the mean field theories are useless for the modeling of the softening regime.

Elastic Percolation

In the absence of a characteristic length the failure threshold also depends on the effective properties since the material is self-similar (scale invariant). The self-similarity takes place when the correlation length tends to infinity during a random crack nucleation process [2]. The failure occurs when the largest cluster, formed by correlated defects, splits the specimen into two parts. This class of problems is studied using the methods of the percolation theory [11].

A typical elastic percolation test [12] consists of a repetitive sequence of two steps: (1) a defect (perforation or a slit) is randomly placed into an unloaded specimen and (b) the specimen is subjected to a very small traction to determine the effective stiffness of the plate without causing plastic deformation. These two steps are repeated until the specimen stiffness is reduced to zero, i.e. when the cluster of correlated defects splits the specimen into two large fragments. The measured stiffness is equal to the effective secant stiffness in the corresponding, deformation controlled deformation process during which the defects nucleate as a result of the monotonically increasing displacements imparted to the specimen [2].

At the percolation threshold ($\xi \to L^-$) the specimen size is the only remaining characteristic length. Due to the randomness of the defect pattern the material on the scale smaller than the correlation length is self-similar. As a consequence of self-similarity the geometry of the largest cluster is fractal and the scaling law for the correlation length is defined in the form $\xi \propto (f - f_c)^\nu$ as $f \to f_c$, where f and f_c are the density and critical density of randomly inserted defects while the exponent ν is the fractal dimension. The rate at which the order parameter or a macro parameter (the vanishing component of the effective stiffness tensor) approaches zero is controlled only by the correlation length. Hence, the scaling law for the vanishing component of the effective stiffness tensor is also fractal $C \propto (f - f_c)^g$ as $f \to f_c^-$. The exponents ν and g, are universal in being independent of the details of local fluctuations of damage, stress and strain fields.

The elastic percolation threshold f_c and universal exponents, which determine the behavior of the system near the percolation threshold, have to be determined for all elastic materials belonging to the same universal class only once. Some of these parameters are already available. The percolation thresholds and scaling laws for the effective stiffness in the case of random distribution of circular [13], [14] and elliptic [15] holes were determined for all homogeneous, isotropic two-dimensional elastic solids. The percolation thresholds for two-dimensional elastic materials weakened by rectilinear slits were determined by simulations [16] and [17]. The corresponding scaling law for the effective stiffness was derived by Krajcinovic [18]. Analytical expressions for the percolation thresholds in the case of holes [19] and slits [20] fit the simulation data extremely well.

The elastic percolation theory can be, firstly, used to compare different mean field models. The percolation model provides the limit to which the mean field estimates of the effective stiffness must tend as $f \to f_c$. Secondly, the random percolation is a correct model for: (a) the response of a rock or concrete specimen subjected to the hydrostatic compression and (b) the cracks induced during the curing of resins and solidification of metals [21].

Localization

From a continuum viewpoint, localization (or faulting) can be, defined as a bifurcation from a state of homogeneous strain to a state of inhomogeneous strain caused by the emergence of a specimen spanning shear band. The strains within the band are much larger than the strains in the exterior of the band which may even relax after the formation of the shear band. From a micromechanical viewpoint, the deformation regime preceding the localization is controlled by the microcrack nucleation while the microcrack growth is prevented by the lateral confinement and the attendant absence of the singular stress fields at the tips of nucleated microcracks. As the nucleation proceeds the probability that the direct interaction of closely spaced (correlated) cracks may lead to the appearance of singular stress fields sufficient to cause microcrack growth increases. The self-organization of microcracks into clusters is a synergistic process since the largest cluster grow faster than smaller ones.

As the clusters of interacting microcracks grow the degree to which the largest cluster controls the deformation increases. Hence, the macroscopic response is in the neighborhood of the localization a function of the correlation length ξ. The amplification of stress intensity factors also depends on the angle subtended by the cluster and compression axes [22]. The angle which optimizes the cooperative effect, varying between 20 and 30 degrees, is very sensitive to the microstructural imperfections.

The general continuum theory of localization, based on the Hadamard's concepts of stability, was formulated by Rudnicki and Rice [23] and Rice [24]. The elegance of these two papers spawned a host of analytical and computational studies. In the process of the development of several sophisticated models the sage advice that the localization conditions "depend critically on subtle features of (constitutive) descriptions" and that the "role of deformation field non-uniformities or imperfections... (are)... of great importance for the initiation and spreading of localized deformation zones" [24] was summarily dismissed. The onset of the localization was defined in terms of the acoustic tensor, i.e. by $det(\mathbf{n} \cdot \mathbf{C}^i \cdot \mathbf{n}) = 0$ where $\mathbf{C}^i$ is the effective tangential tensor at the point of bifurcation and $\mathbf{n}$ the normal to the localized band. The bifurcation is not possible in the von Mises material and the plasticity models must be embellished either by a non-associative flow rule or by allowing the formation of vertices in the yield surface.

The added problem related to the application of local continuum models is related to the absence of a characteristic length. As a result the localized zone has a zero thickness and the density of the energy dissipation is infinite. This fact triggered development of several non–local and gradient localization models.

None of these models treats the material microstructure as being random and the process as being dynamic. In consideration of correlation-induced (Mott) and disorder-induced (Anderson) localizations it was concluded that the delocalization-localization is impossible in the absence of some disorder. It is, therefore, not surprising that the deterministic models were unable to capture the essential non-deterministic nature of the phenomenon which looms behind the tests (Lockner, et al [3]). The localization belongs to the class of the short to long-range transitions. As the damage concentration increases the correlation range ξ grows from zero to the specimen size L. The correlation length ξ controls the specimen macroscopic response only in the phase preceding the onset of localization and the post-localization regime. However, well before that point the influence of ξ on the macroscopic response cannot be neglected. The onset of the localization and exact thickness of the localized band are largely in the eyes of the beholder. In a material such as granite and limestone the band "thickness" varies along its length. The material in the exterior of the band seems to be statistically homogeneous while the characterization of the interior of the non-local material within band is not as simple.

The localization process can be modeled using the molecular dynamics [25]. The disorder is quenched into the lattice. To replicate the Lockner, et al. [3] tests the selected two-dimensional, irregular, elastic triangular lattice is subjected first to biaxial compression $\sigma_{lat} = \sigma_{axial} > 0$. Subsequently, the lattice length in the direction of the longer axis is slowly decreased in order to be able to consider both hardening and softening regimes.

In the early stages of the simulation the damage evolution is controlled by weak spots (rupture of links of inferior strength). As the defect concentration increases the self-organization of defects into the largest cluster becomes evident. The increase of the correlation length ξ is initially slow and crosses over to a power law as $\xi \to L^{-}$. The true nature of the phenomenon is reflected in the transfer of the energy through the disordered lattice. The dynamic and stochastic nature of the phenomenon can be illustrated using the "ant in the labyrinth" model [11]. An ant (random walker), located within the localized band, is allowed to move randomly from a site to any of the other nearest neighbor sites providing that the link between two sites exists. The ant must stay in its site each time it makes an attempt to move along a ruptured link. In an undamaged lattice the mean square distance traversed by the ant is linearly proportional to the walking time. In a localized shear band the time the ant needs to reach the less damaged material increases with the density of ruptured links. However, as the time of walking approaches infinity all ants, starting their walk in different sites within the band, will eventually find their way out from the labyrinth (shear band) assuming that the lattice is connected. Thus, the localized band "width" depends on the walking time allotted to the ant.

As a mechanical analogy of this ant a link within the band can be subjected to a periodic change of its length. The fraction of the imparted energy recorded at a point in the band exterior depends on the damage within the band and the frequency of the imparted vibrations. At very small frequencies (statical loads) large part of the energy percolates through the band. As the frequency of the imparted vibrations increases the fraction of energy recorded in the exterior of the band decreases.

In summary, the question of what is the "width" of the localized band is not well posed. In fact, the width is a function of the correlation length and the frequency. In a purely static approximation the band width is equal to zero.

Annealed Disorder

The creep rupture of resins is an example of failure modes attributed to the annealed disorder. The time to creep rupture depends on the magnitude and sign of the applied tractions, temperature and material microstructure. Most phenomenological and micromechanical [26] models are deterministic in form. Random fluctuations of temperature and stress are neglected. A different viewpoint was taken by Regel', et al. [27] and Vujosevic and Krajcinovic [28] who treat the temperature as the stimulus which enhances the mobility of atoms and increases the probability of the dislocation motion and rupture of atomic bonds.

The microstructure of epoxy resins [21] emphasizes dense nodules interconnected by lower density material (a relatively vacuous network of cross–linked molecular chains). A triangular central force lattice with nodules as sites and molecular chains as links is a realistic model of the described microstructure. The quenched disorder is neglected and the lattice is in the pristine state both geometrically and topologically regular, and all links have identical stiffness and strength.

The damage evolution, i.e. bond rupture sequence, is treated as a random process which is activated by spatially and temporally random thermal fluctuations. The probability that a ith link will rupture during the time interval Δt , was based on the absolute reaction–rate theory [29]. The sequence of the link ruptures is determined by the Monte Carlo lottery. The simulations demonstrate that the temperature driven damage evolution leads to the percolation transition. The

specimen stiffness decreases linearly. The loss of stiffness becomes nonlinear and very rapid when the specimen is subjected to tensile loads. In all cases the strain vs time curves exhibit all trends characteristic of creep. The exponential dependence of the time to failure on the load parameter derived from simulations is in agreement with the experimentally observed trends [27].

The simulations data in [28] were used to construct approximate analytical model which fits the simulation data surprisingly well. A mean field model was derived to estimate the effect that stress concentrations have on the time to creep rupture. Final, and perhaps the most important, result was that the time to rupture does not depend on the lattice size.

SUMMARY AND CONCLUSIONS

The principal message of this research is that the disorder in macrostructure and the random fluctuations of externally imparted stimuli cannot be easily dismissed when estimating the type of failure and its threshold. Purely phenomenological and deterministic continuum models based on fitting experimental data may rather often lead to wrong conclusions. Direct and careless extensions of mean field models beyond the limits of the dilute concentration of defects is not always a prudent strategy. Abandoning familiar and elegant mean field methods may, indeed, be traumatic. At the same time a search of a more rational model, which acknowledges the loss of homogeneity, may in many cases be the only avenue to a reliable analytical prediction of the failure threshold. The development of non-local theories was in the majority of cases channeled along the familiar deterministic routes. Preservation of mathematical elegance may, indeed, be a virtuous goal. At the same time it cannot be the only guideline leading to the improvement of modeling techniques. It seems to the authors of this study that physics should have a precedent over mathematics.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support in form of a research grant from the U.S. Department of Energy, Division of Engineering and Geosciences to the Arizona State University which made this work possible.

REFERENCES

1. S. NEMAT–NASSER and M. HORI, *Micromechanics: Overall Properties of Heterogeneous Materials*, North–Holland, Amsterdam, The Netherlands (1993).
2. D. KRAJCINOVIC, *Damage Mechanics*, North–Holland, Amsterdam, The Netherlands (1996).
3. D.A. LOCKNER, D.E. MOORE and Z. RECHES, 'Microcrack Interaction Leading to Shear Fracture', in: *Rock Mechanics*, J.R. Tillerson and W.R. Wawersik, Eds., Balkema Publ., Rotterdam, The Netherlands, (1992).
4. S. ROUX and A. HANSEN, 'Early Stages of Rupture of Disordered Materials', *Europhys. Lett.* 11, 37 (1990).
5. M.S. PATERSON, *Experimental Rock Deformation–the Brittle Field*, Springer–Verlag, Berlin, Germany (1978).
6. W.A. CURTIN and H. SCHER, 'Algebraic Scaling of Material Strength', *Phys. Rev. B*, 45, 2620 (1992).
7. P.M. DUXBURY, S.G. KIM and P.L. LEATH, 'Size Effect and Statistics of Fracture in Random Materials', *Mat.Science and Engineering*, 176, 25 (1994).

8. H. HORII and S. NEMAT–NASSER, 'Brittle Failure in Compression: Splitting, Faulting and Brittle–Ductile Transition', *Phil.Trans. Royal Soc. London*, 319, 337 (1986).
9. A. HANSEN, S. ROUX and H.J. HERRMANN, 'Rupture of Central–Force Lattices', *J. of Physics France*,50, 733 (1989).
10. D. KRAJCINOVIC and M. BASISTA, 'Rupture of Central–Force Lattices Revisited', *J. Phys. I*, 1, 241 (1991).
11. D. STAUFFER and A. AHARONY, *Introduction to Percolation Theory*, 2nd edition, Taylor & Francis, Washington, DC (1992).
12. L. BENGUIGUI, 'Experimental Study of the Elastic Properties of a Percolating System', *Phys. Rev. Lett.* 53, 2028 (1984).
13. B.I. HALPERIN, S. FENG, P.N. SEN, 'Difference Between Lattice and Continuum Percolation Transport Exponents', *Phys. Rev. Lett.* 54, 2391 (1985).
14. L. BENGUIGUI, 'Elasticity and Fracture near a Percolation Threshold in two Dimensions', in: *Fragmentation, Form and Flow in Fractured Media*, R. Englman and Z. Jaeger, Eds.,*Annals of Israel Phys. Soc.* 8, 288 (1986).
15. K. MALLICK, D. KRAJCINOVIC, D. SUMARAC and M. VUJOSEVIC, 'Critical State of a Two–Dimensional Elastic Continuum Containing Elliptical Voids', *Engng. Fracture Mech.* 46, 553 (1993).
16. E.G. PIKE and C.H. SEAGER, 'Percolation and Connectivity: A Computer Study I, *Phys. Rev. B*, 10, 1421 (1974).
17. P.C. ROBINSON, 'Connectivity of Fracture Systems–A Percolation Theory Approach, *J. Phys. A: Math. Gen.* 16, 605 (1983).
18. D. KRAJCINOVIC, 'Scaling Law for Blue Cheese Model of Damage', *Phys. Lett. A*, 173, 433 (1993).
19. D. KRAJCINOVIC, K. MALLICK, M. BASISTA and D. SUMARAC, 'Elastic Moduli of Perforated Plates in the Neighborhood of Critical States', *Int. J. Solids Structures*, 29, 1837 (1992).
20. D. KRAJCINOVIC, D. SUMARAC and K. MALLICK, 'Elastic Parameters of Brittle, Elastic Solids Containing Slits, *Int. J. Damage Mech.* 1, 386 (1992a).
21. D. KRAJCINOVIC and K. MALLICK, 'Micromechanics of Process Induced Damage in Thermosets, *J. Mech. Phys. Solids*, 43, 1059 (1995).
22. H. HORII and S. NEMAT–NASSER, 'Compression–induced Microcrack Growth in Brittle Solids: Axial Splitting and Shear Failure', *J. Geophys. Res.* 90, 3105 (1985).
23. J.W. RUDNICKI and J.R. RICE, 'Conditions for Localization of Deformation in Pressure–sensitive Dilatant Materials', *J. Mech. Phys. Solids*, 23, 371 (1975).
24. J.R. RICE, 'The Localization of Plastic Deformation', in: *Proc. of 14th International Congres of Theoretical and Applied Mechanics*, W.T. Koiter, Ed., North Holland, The Netherlands (1976).
25. D. KRAJCINOVIC and M. VUJOSEVIC, 'Localization: dynamic transition of the correlation range', (in preparation)
26. A.C.F. COCKS and F.A. LECKIE, 'Creep Continuum Equations for Damaged Materials', in: *Advances in Applied Mechanics*, J.W. Hutchinson and T.J. Wu, Eds., 25, 239, Academic Press, New York, NY (1987)
27. V.R. REGEL', A.I. SLUTKER and E.E. TOMASHEVSKII, *Kinetic Theory of Strength of Solids*, Nauka, Moskow, USSR (in Russian) (1974).
28. M. VUJOSEVIC and D. KRAJCINOVIC, 'Creep Rupture of Polymers: a Statistical Model', *Int. J. Solids Structures* (in print)
29. H. EYRING, 'Viscosity, Plasticity and Diffusion as Examples of Absolute Reaction Rates', *J. Chem. Phys.* 4, 283 (1936).

T^*_ε INTEGRAL ANALYSIS OF FRACTURE SPECIMENS

Y. Omori [1], K. Perry, Jr. [2], J.S. Epstein[2], L. Ma[1], H Okada[3], S. N. Atluri*[3] and A. S. Kobayashi[1]

[1]University of Washington Department of Mechanical Engineering, Seattle, WA 98195-2600
[2]Idaho National Engineering Laboratory, Material Science Division, Idaho Falls, ID 83415
[3]Georgia Institute of Technology, Computational Mechanics Center, Atlanta, GA 30332-0356

ABSTRACT

T^*_ε integral values associated with stable crack growth in thin 2024-T3 aluminum compact (CT) specimens and A606 HSLA steel single edge notched (SEN) specimens were determined directly from the crack tip displacement field obtained by moiré interferometry. Stable crack growth in the SEN specimen was also simulated by an elastic-plastic finite element (FE) model which was driven by the experimentally determined boundary conditions. T^*_ε obtained experimentally and by FE were in reasonable agreements with each other. Unlike the vanishing J integrals with crack extension, T^*_ε reached steady state values with stable crack growth. Thus, for a given integration contour, Γ_ε, near the crack tip, T^*_ε can be used as a stable crack growth as well as a ductile fracture criteria.

INTRODUCTION

In a series of papers, May and Kobayashi [1] showed that the ASTM J resistance curve does not represent the crack tip state of stress since J is not a path independent integral under stable crack growth. Moreover, the near field J integral vanished after reaching a maximum value at the initial phase of stable crack growth. Brust et al [2,3], on the otherhand, showed through finite element (FE) simulation, that the near field T^* integral [4] remained a steady state value under loading, unloading, reloading and under stable crack growth.

Unlike the J integral, the T^* integral is based on the incremental theory of plasticity, T^*_ε, and is defined by Stonesifer and Atluri [4] as

$$\Delta T_{\varepsilon}^{*} = \int_{\Gamma_{\varepsilon}} \left[\Delta W \cdot n_1 - (t_i + \Delta t_i) \cdot \Delta u_{i,1} - \Delta t_i \cdot u_{i,1} \right] ds$$

where W is the work density and t_i and u_i are the traction and displacement, respectively. Here ΔT^*_{ε} denotes the ΔT^* value associated with a contour, Γ_{ε}, in the very vicinity of the crack. The proximity of Γ_{ε} is restricted to the region where plane stress condition prevails and is outside of the region of three dimensional state of stress at the crack tip. The latter is generated by the finite thickness of the plate specimen. For a flat crack without a shear lip, Narasimhan and Rosakis [5] have shown that the state of plane stress prevails outside of one half of the plate thickness.

In theory, the total T^*_{ε} is obtained by summing the incremental ΔT^*_{ε} through the plastic deformation process. Fortunately Pyo et al [6] have found that the total T^*_{ε} computed directly by using the stresses and strains based on the incremental theory of plasticity was for all practical purpose equal to the summed ΔT^*_{ε}. Thus T^*_{ε} can be computed directly without the cumbersome incremental summation procedure provided the stresses and strains, on which T^*_{ε} is based, are obtained through the use of the incremental theory of plasticity. Like the J integral, T^*_{ε} is path dependent in the presence of large scale plasticity and plastic unloading. Thus, if T^*_{ε} is to be used as a fracture criterion, it is imperative that T^*_{ε} be evaluated along a contour, Γ_{ε}, very close to the crack tip, such that it can be considered a crack tip parameter.

In this paper, we present the T^*_{ε} integral values which were determined **directly** from the displacement field surrounding a stably growing crack in thin 2024-T3 aluminum compact (CT) specimens and A606 HSLA steel single edge notched (SEN) specimens.

METHOD OF APPROACH

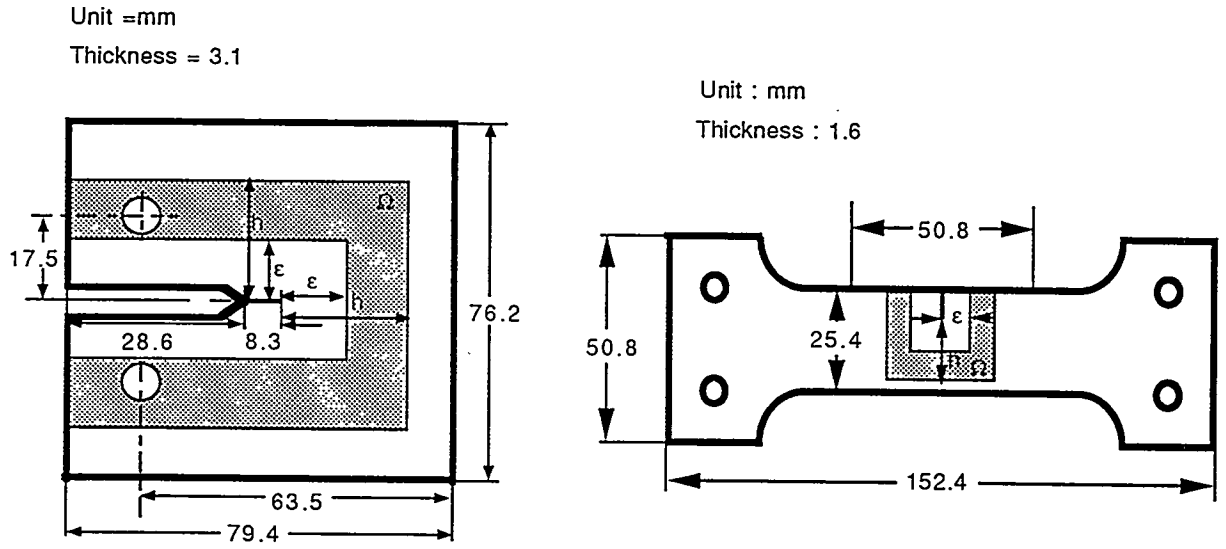
Experimental Procedure

The experimental procedure consisted of measuring the two orthogonal displacement fields surrounding a stably growing crack in the CT and SEN specimens using Moiré interferometry. Figure 1 shows the 2024-T3 aluminum CT specimen and the A606 HSLA steel SEN specimen, respectively. A coarse cross diffraction gratings of 40 lines/mm was used to determine the large plastic strains surrounding the extending crack. A special Moiré interferometry procedure [7], which combines the advantages of geometric Moiré and the traditional Moiré interferometry and uses a low frequency Moiré grating for measuring large strains in the vicinity of the crack tip, was developed for this ductile fracture study. A special four-beam Moiré interferometry bench was also constructed for use with this low frequency Moiré diffraction grating.

The strains along the line integration contour for the T^*_{ε} integral computation were determined directly from the orthogonal displacement field obtained by Moiré interferometry. The stresses corresponding to the total strains were then computed using the equivalent stress-strain and the measured uniaxial stress-strain data of the 2024-T3 aluminum and the A606 HSLA steel sheets. This use of the deformation theory of plasticity to compute stresses does not account for the unloading process which occurs in the trailing wake of the extending crack. However, by restricting the integration contour very close to and along the extending crack, Okada and Atluri

[8] has shown that the contour integration trailing the crack tip can be neglected by virtue of the closeness of the integration path, Γ_ε, to the traction free crack surface. This approximation not only simplified the integration process but also eliminated the undesirable effect of the deformation theory of plasticity which is used to compute the stresses from the measured strains.

The crack tip opening angle (CTOA) was also computed by the angle subtended by the measured crack opening displacement (COD) at a distance 1 mm from the crack tip.



(a) 2024-T3 CT specimen.

(b) A606 HSLA SEN Specimen.

Figure 1. Specimens.

Numerical Procedure

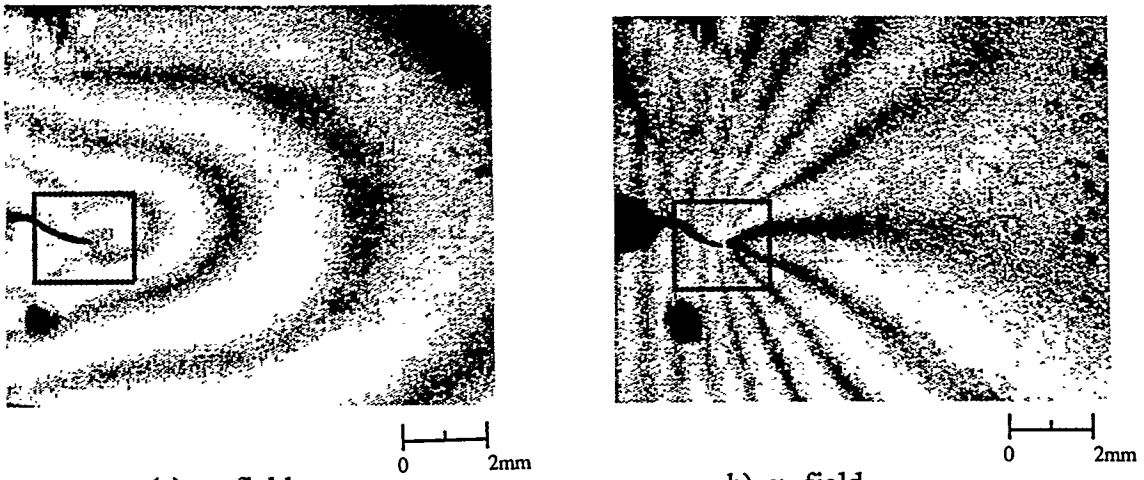
A plane stress finite element (FE) model of portion of the A606 HSLA steel SEN specimen was used for numerical analsysis. The FE model of the SEN specimen was truncated at a 20.8 mm distance from the crack in order to conserve compute time. The measured displacements along the truncated width of the SEN specimen together with the measured instantaneous crack length and the measured equivalent stress-strain relation of A606 HSLA steel were used to drive the FE model in its generation mode based on the incremental theory of plasticity.

The T^*_ε integrals along an elongated contour surrounding the stably growing crack of the SEN specimen was then computed. Unlike the stresses used in the experimental procedure for T^*_ε evaluation, the FE analysis provided stresses which accounted for the unloading effect in the trailing wake of the extending crack tip. Therefore, the entire contour was used for T^*_ε evaluation. Numerical errors in the FE data in the vicinity of the crack tip were masked by replacing the contour integral by the equivalent domain integral of Nikishkov and Atluri [7]. To recapitulate, T^*_ε evaluation procedures for the Moiré and FE studies differ in that the former

involved only the frontal segment of a near-field contour, Γ_ε, while the latter involved an equivalent domain integral over the entire crack length.

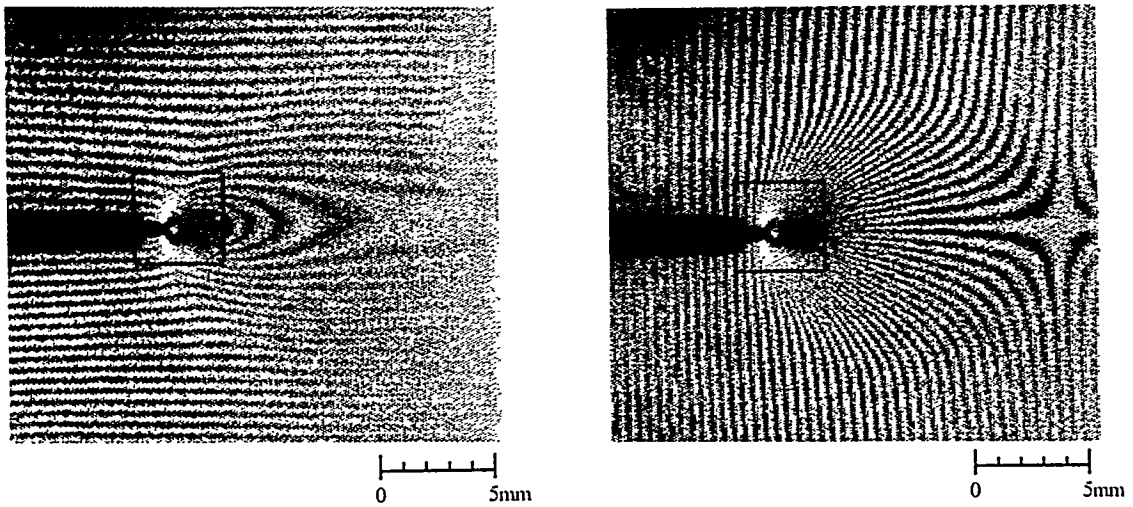
RESULTS

A total of four 2024-T3 aluminum CT specimens and four A606 HSLA SEN specimens were analyzed. Figures 2 and 3 show typical Moiré interferometry patterns of the u- and v-



(a) u- field. b) v- field.

Figure 2. Moiré fringe patterns of 2024-T-3 CT specimen.



(a) u-field. (b) v-field.

Figure 3 Moiré fringe patterns of A606 HSLA SEN specimen.

displacement fields of an 2024-T3 aluminum CT specimen and an A606 HSLA steel SEN specimen, respectively. Also shown are the elongated Γ_ε contour of $\varepsilon = 2.0$ mm along the extended crack in the 2024-T3 aluminum CT specimen and $\varepsilon = 2.0$ mm for the A606 HSLA steel SEN specimen. These ε's are over a one half of a plate thickness away from the crack tip where the state of plane stress is thought to prevail even in the presence of 100 percent shear lips in the crack.

Figure 4 shows the experimentally and FE determined T^*_ε for the 2024-T3 aluminum CT and the A606 HSLA SEN specimens. The maximum crack extension, Δa, for the 2024-T3 aluminum and A606 HSLA steel specimens were 5.5 and 8.0 mm, respectively. Also shown for comparison is the experimentally and numerically determined T^*_ε for the same Γ_ε in a thinner 2024-T3 aluminum SEN specimen [10].

Figure 5 shows the experimentally determined crack tip opening displacements (CTOA). The CTOA for the aluminum SEN specimen is in excellent agreements with that of the 2024-T3 aluminum SEN specimen [10] as well as that of Dawicke et al [11].

Unlike the increasing J resistance curve determined by the ASTM Standard testing procedure, or the vanishing J for the near field solution, the T^*_ε for a given Γ_ε shown in Figure 4 and the CTOA shown in Figure 5 both reached physically more realistic steady state values with stable crack growth.

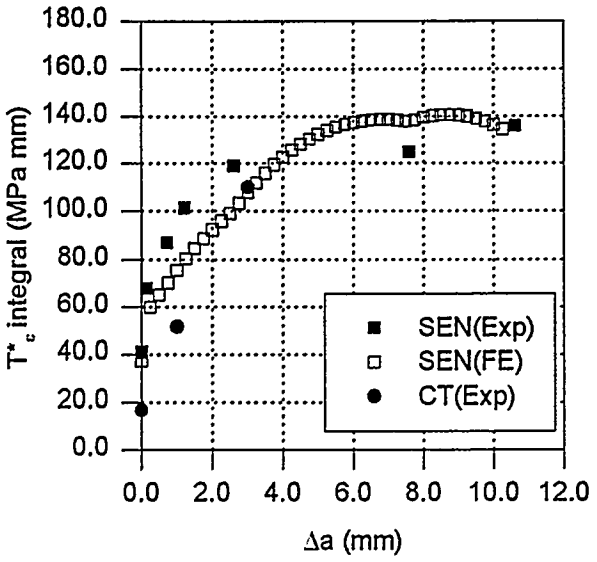
CONCLUSIONS

1. T^*_ε computed from the displacement fields, which were obtained experimentally and from finite element analysis, of the A606 HSLA steel SEN specimens were in good agreement with each other.

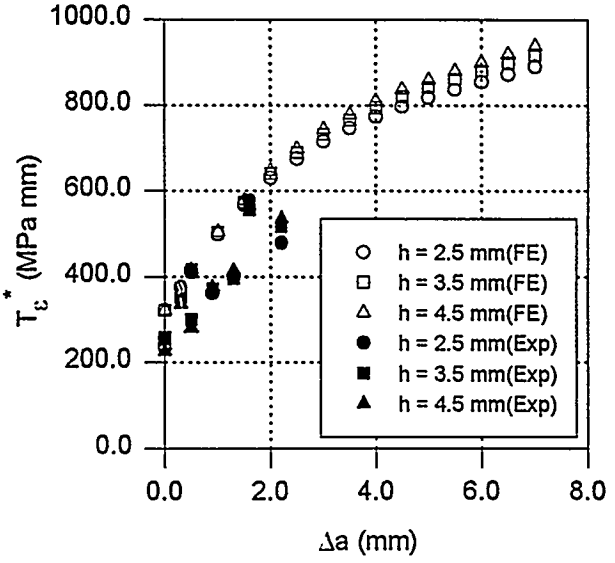
2. Computed T^*_ε values are domain-size independent and tend to converge to a stationary value for a smaller inner contour, Γ_ε of $\varepsilon = 2.0$ mm for the 2024-T3 aluminum CT specimen and $\varepsilon = 2.0$ mm for th A606 HSLA specimen.

3. T^*_ε did not reach a steady state value for $\Delta a = 8.0$ mm during stable crack growth in a thin A606 HSLA SEN specimen.

4. Computed and measured CTOA reached a steady state value of 5^o and 15^o, respectively, after a stable growth of about $\Delta a = 2.0$ mm in the 2024-T3 aluminum CT specimen and A606 HSLA steel SEN specimen.

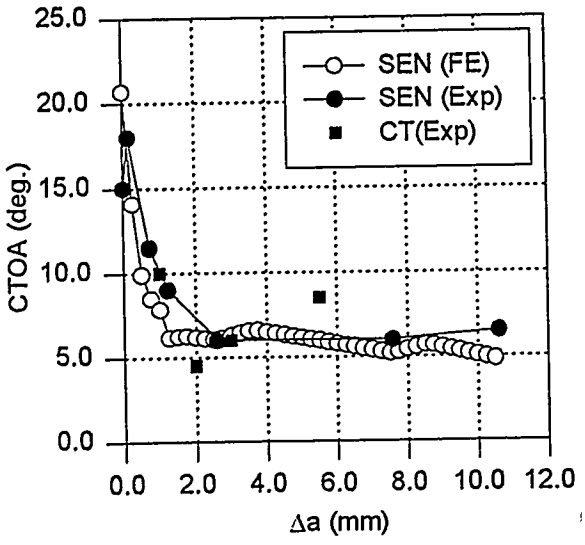


(a) 2024-T3 CT specimen

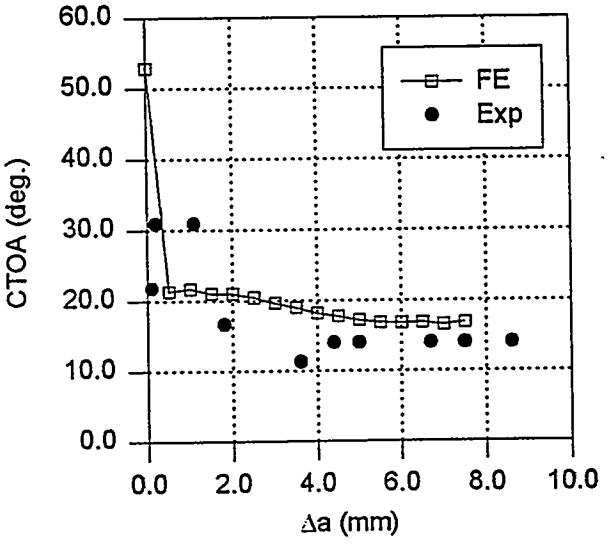


(b) A606 HSLA SEN specimen

Figure 4. Experimentally determined and FE T^*_{ε} integrals.



(a) 2024-T3 SEN and CT specimens



(b) A606 HSLA SEN specimen

Figure 5. CTOA.

ACKNOWLEDGMENT

The results reported here were obtained through DOE Grants DE-FG06-94ER14490 and DE-FG0694ER 14491.

REFERENCES

1. G.B. May. and A.S. Kobayashi, "Plane stress stable crack growth and *J*- Integral/HRR field," *Int. J. Solid and Structures,* 37/617, 857-881 (1995).

2. F.W. Brust, T. Nishioka, S.N. Atluri, and M. Nakagaki, "Further studies on elastic-plastic stable fracture utilizing the *T** integral," *Engineering Fracture Mechanics*, 22, 1079-1103 (1985).

3. F.W. Brust, J.J. McGowan, and S.N. Atluri,, "A combined numerical/experimental study of ductile crack growth after a large unloading, using *T**, *J* and CTOA criteria," *Engineering Fracture Mechanics*, 23, 537-550 (1986).

4. R.C. Stonesifer and S.N. Atluri, "On a study of the (ΔT) and C* integrals for fracture analysis under non-steady creep," *Engineering Fracture Mechanics,* 16, 769-782 (1982).

5. R. Narasimhan and A.J. Rosakis, "Three-Dimensional Effects Near a Crack Tip in a Ductile Three-Point Bend Specimen: Part I--A Numerical Investigation," *ASME Journal of Applied Mechanics,* 57, 607-617, 1990.

6. C.-R. Pyo, H. Okada, and S.N. Atluri, "An Elastic-plastic finite element alternating method for analyzing wide spread fatigue damage in aircraft structures," *Computational Mechanics*, 16, 62-68 (1995).

7. F.X. Wang, G.B. May and A.S. Kobayashi, "Low-spatial-frequency steep grating for use in Moiré interferometry," *Optical Engineering*, 33, 1125-1131 (1994).

8 H. Okada, and S.N. Atluri, "T^*_ε integral evaluation from experimental displacement field for a plate with stably propagating crack: development of calculation procedure and implication of T^*_ε", (1996), to be submitted.

9. G.P. Nikishkov and S.N. Atluri, "An equivalent domain integral method for computing crack-tip integral parameters in non-elastic, thermo-mechanical fracture," *Engineering Fracture Mechanics,* 26, 851-867 (1987).

10. Y. Omori, H. Okada, L. Ma, S.N. Atluri, A.S. Kobayashi, and P. Tan, "Further Studies on T^*_ε Integral Under Plane Stress Crack Growth," to be published in *Localized Damage, '96,* Wessex Institute of Technology.

11. D.S. Dawicke, M. Sutton, J.C. Newman, and C.A. Bigelow, "Measurement and analysis of critical CTOA for thin-sheet aluminum alloy materials," *Fracture Mechanics 25th Symposium,* ASTM STP 1220, eds. F. Erdogan and R. J. Hartranft, 358-379 (1995).

MICROSCOPIC INTERFACIAL PHENOMENA DURING FLOW IN POROUS MEDIA

Michael J. Miksis and Michael P. Ida

Department of Engineering Sciences
and Applied Mathematics
Northwestern University
Evanston, Illinois 60208

Abstract

A fundamental process during any multiphase flow in porous media is the breaking apart of one of the phases into smaller components. Here we investigate this breaking process as applied to a thin liquid film. We study the breaking of both a two dimensional planar film and a cylindrical thread of liquid using both analytical and numerical methods.

INTRODUCTION

Multiphase flows in porous media occur in many situations of practical interest. An example with application to enhanced oil recovery is foam flow in porous media. Here a gas and surfactant solution is injected into the ground. This then generates a foam which is used to drive the oil out. The foam is composed of alternating regions of liquid and gas with the liquid primarily in the form of a thin liquid film coating the solid portions of the pores or in the form of thin liquid lamellae which separate the gas regions and move with the gas down the pore channel. As the foam moves within the porous material, these liquid films can break or rupture forming larger gas bubbles within the material. In addition, the formation of the foam itself is associated with a film rupture process called snap-off where a large gas bubble is driven through a constriction in the pore channel, becomes unstable and splits into two parts (Tsai and Miksis [1].) Our aim here is to study this rupture process. We will consider a model for the dynamics of a thin viscous film which is valid for long wave (relative to the film thickness) disturbances. The model accounts for the effects of surface tension, inertia and van der Waals forces. We begin by studying a two dimensional planar film. Numerical and analytical results close to rupture will be presented. We will also study a thin three dimensional axisymmetric liquid thread.

There have been a number of works recently concerned with the breaking of either thin 2-D liquid films or axisymmetric threads of liquid. For example, Dupont et al. [2] analyzed the interfaces between two liquids under the influence of surface tension (but without van der Waals forces). Rupture in a finite time is observed numerically and the region near rupture is scrutinized for a similarity solution. Considering the finite-time rupture of thin fluid layers in

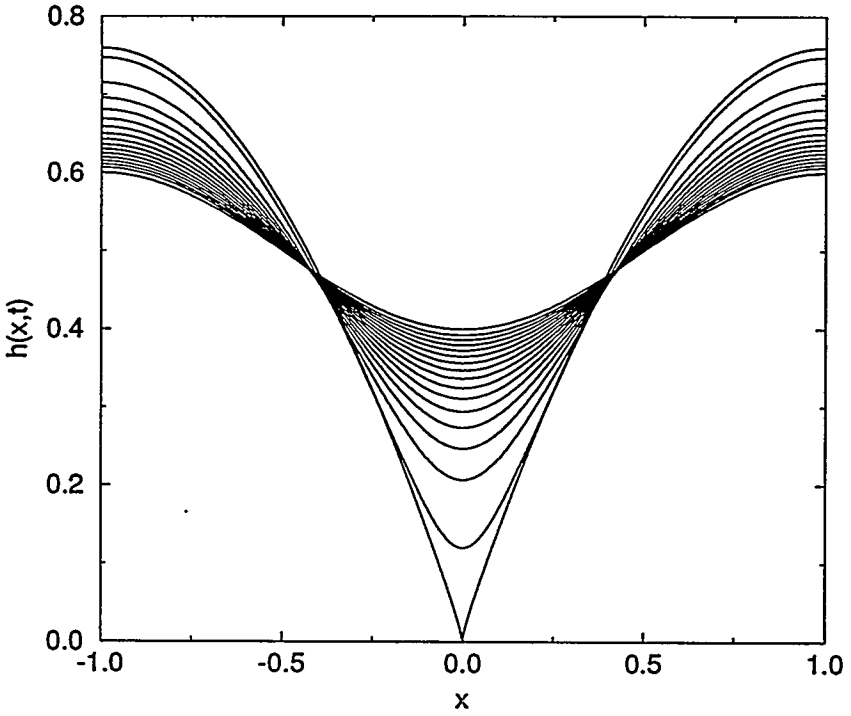


Figure 1: Time series evolution of an unstable thin film for $A = 1$ and $S = A/\pi^2 = S_c/2$. $h^{(0)}(x,t)$ is shown for $t = 0, 0.1, 0.2, \ldots, 1.4, 1.5, 1.52129982$.

Hele-Shaw flows, Goldstein et al. [3] studied a lubrication model for the dynamics of the fluid interfaces. Based upon their numerical studies, they concluded that the Bond number (van der Waals forces were not included in their model) plays a decisive role in the character of the solutions near rupture. Additionally, they investigated the region near rupture numerically and conjectured a power-law form for the interface ($h \propto (t - t_r)^\beta$ where β was numerically found to be equal to 1.45 ± 0.05). Claiming the existence of "universal exponents," Eggers [4] considered the rupture regions of thin axisymmetric threads of liquid. In the same context, i.e., the thinning of threads of liquid, Eggers and Dupont [5] analyzed the "pinching" singularity observed experimentally. Their model equations included the effects of surface tension with the full curvature term, but neglected the role of van der Waals forces.

TWO DIMENSIONAL THIN FILMS

We will use the two-dimensional free film equations of Erneux and Davis [6]. These equations were derived asymptotically in the long wave limit. For convenience, we summarize those equations here. Consider an incompressible film of density ρ and dynamic viscosity μ. Introduce a two-dimensional (x, z) cartesian coordinate system. Let b be the unit of length in the x direction and let h_0, the undisturbed film thickness, be the unit of length in the z direction. In addition let $\mu/b\rho$ be the unit of velocity and $\rho b^2/\mu$ the unit of time. Specify the symmetric film

interfaces by $z = \pm h(x,t)$, here t is time, and the transverse component of the fluid velocity in the x-direction by $u(x,t)$. In the limit $h_0/b \ll 1$, the leading-order evolution equations for these two quantities are then,

$$\frac{\partial h}{\partial t} = -u\frac{\partial h}{\partial x} - h\frac{\partial u}{\partial x}, \tag{1}$$

$$\frac{\partial u}{\partial t} = \frac{4}{h}\frac{\partial h}{\partial x}\frac{\partial u}{\partial x} + 4\frac{\partial^2 u}{\partial x^2} - u\frac{\partial u}{\partial x} + 3S\frac{\partial^3 h}{\partial x^3} + \frac{3A}{8h^4}\frac{\partial h}{\partial x}, \tag{2}$$

where $S = \gamma_0 \rho h_0/3\mu^2$ is a nondimensional constant which encompasses the dimensional surface tension γ_0, and $A = \rho b^2 \tilde{A}/6\pi h_0^3 \mu^2$ is a dimensionless constant which encompasses the effects of van der Waals forces through the Hamaker constant $\tilde{A}$. We will assume periodicity on the interval $x \in [-1,1]$ so that about the origin $x = 0$, h is even ($h(-x,t) = h(x,t)$) and u is odd ($u(-x,t) = -u(x,t)$). Fig. 1 demonstrates a typical unstable numerical solution to these equations. The parameter values are $A = 1$ and $S = A/\pi^2$. The initial conditions used are $u(x,0) = 0$ and $h(x,0) = 1/2 - (0.1)\cos\pi x$, and solutions are shown for $t = 0, 0.1, 0.2, \ldots, 1.4, 1.5, 1.52129982$. Note that near rupture ($h \to 0$ at $x = 0$), the numerical solution appears to have taken on a similarity form. Our aim is to identify this dynamics. Erneux and Davis [6] have show that a uniform thickness film is linearly stable when $S > 2A/\pi^2$. Hence we will select our parameters to fail this stability criterion in order to observe rupture.

In order to resolve the form of the solution near the point of breakage, we define the following similarity variables,

$$\tau = t_r - t, \tag{3a}$$

$$\xi = x\tau^{-\beta}, \tag{3b}$$

where t_r is the time of rupture. We then assume the following forms for h and u,

$$h = \tau^{\alpha}\mathcal{H}(\xi), \tag{4a}$$

$$u = \tau^{\Gamma}\mathcal{U}(\xi). \tag{4b}$$

The forms of these solutions are suggested by the rapid variations in both space and time near the point of rupture. Now substitute the similarity forms into the evolution equations (1) and (2). Suppose we balance all the terms in (1). We find that $\Gamma = \beta - 1$. Now consider the other equation. Here there are several possible terms which could be balanced, each giving a different answer. We claim that the viscous and van der Waals terms must balance (this will be verified numerically later). Hence this assumption implies that we need to balance the terms of $O(\tau^{\Gamma-2\beta})$ with those of $O(\tau^{-3\alpha-\beta})$. This implies that $\alpha = 1/3$. Hence we have determined that $\Gamma = \beta - 1$ and $\alpha = 1/3$ but we have not determined the specific values of β and Γ from this dominate balance argument.

Now we will examine the evolution of the interface numerically. To obtain extreme spatial accuracy in the region near rupture while maintaining a reasonable execution time, an adaptive regridding finite-difference algorithm based upon that employed by Bernoff and Bertozzi [7], Bertozzi [8] and Dupont et al. [2] has been used to solve equations (1) and (2) (see Ida [9] for details). The results of a systematic examination of the observed exponents are summarized in table 1. In each case, the same initial conditions $((h(x,0), u(x,0)) =$

exponents	$A=1$	$A=5$	$A=10$
α	$0.3438428 \pm 0.002986701$	0.3381307 ± 0.00352386	$0.3365617 \pm 0.003716294$
$\alpha - 2\beta$	$-0.4422666 \pm 0.006927182$	$-0.5384250 \pm 0.012368135$	$-0.58326818 \pm 0.01416819$
$\Gamma - \beta$	$-1.0124770 \pm 0.004877930$	$-1.0181044 \pm 0.004328344$	$-1.0192778 \pm 0.004913314$

Table 1: Results of a systematic study of the exponents in the similarity solution for the free film near breaking. For each value of A, values are averaged over $S = \{0, A/4\pi^2, A/2\pi^2, 3A/4\pi^2, A/\pi^2\}$.

$(1/2 - (0.1)\cos\pi x, (0.1)\sin\pi x))$ are used; and for each value of $A = \{1, 5, 10\}$, values of $S = \{0, A/4\pi^2, A/2\pi^2, 3A/4\pi^2, A/\pi^2\}$ are employed. For each value of A used, the values of the exponents are then averaged, with the results and their standard deviations presented in table 1. From the values of the standard deviations, which never exceed approximately 2%, we see that the values of all exponents appear to be independent of S. We find, however, that while α and $\Gamma - \beta$ appear to be independent of A (each is within an interval of approximately two standard deviations in width, and within approximately four standard deviations of their expected values of 1/3 and -1), $\alpha - 2\beta$, the exponent for $\frac{\partial^2 h}{\partial x^2}(0,t)$ which would result from (4), is not (the interval needed to contain it is on the order of 20 standard deviations). Thus, our assumption concerning the balancing of the terms seems to be confirmed.

We would now seem to have satisfactory proof as to the correct form of the similarity solution. Thus, we find that $\alpha = 1/3$, and that $\Gamma - \beta = -1$ but Γ and β are undetermined. We speculate that the constants β and Γ are to be determined through a matching procedure to a more complex solution of the full evolution equations in the outer region away from the point of rupture.

In conclusion we note that we have identified a similarity form for the film thickness and transverse velocity of a two-dimensional thin free film near rupture. We have found that van der Waals and viscous forces dominate the evolution, still leaving a single degree of freedom (the value of either β or Γ) in the precise specification of the form of the solution. This undetermined constant could in principle be determined through a matching procedure to the outer flow field away from the immediate vicinity of rupture. Similarity solutions of this type have been discussed by Barenblatt and Zel'dovich [10].

THREE DIMENSIONAL LIQUID THREADS

In this section, we investigate a set of evolution equations for the dynamics of a thin, axisymmetric thread of liquid [9]. The equations can be derived by a long wave analysis [9]. The effects of a van der Waals like force will be included in our model, in addition to the effects of surface tension, inertia and viscosity.

We begin by considering an infinite, periodic, axisymmetric thread of liquid of density ρ and kinematic viscosity ν oriented along the $\tilde{z}$ axis of a cylindrical coordinate system. Denote the fluid velocity by $\tilde{\mathbf{v}}(\tilde{r}, \tilde{z}) = \tilde{u}_r(\tilde{r}, \tilde{z})\,\mathbf{r} + \tilde{u}_z(\tilde{r}, \tilde{z})\,\mathbf{z}$ where $\mathbf{r}$ and $\mathbf{z}$ are the unit vectors in the $\tilde{r}$ and $\tilde{z}$ directions respectively. Let $\tilde{r} = \tilde{h}(\tilde{z}, \tilde{t})$ represent the fluid interface, $\mathbf{n}$ and γ the surface tension on it. In order to extract a tractable problem we perform a long wavelength analysis of the the equations of motion. Let $\tilde{b}$ be a typical wavelength of the system and $\tilde{h}_0$ be a typical

thread radius such that $\epsilon = \tilde{h}_0/\tilde{b} \ll 1$. The leading order in ϵ equations of motion are [9]

$$\frac{\partial h}{\partial t} = -u_z \frac{\partial h}{\partial z} - \frac{h}{2}\frac{\partial u_z}{\partial z}. \tag{5}$$

and

$$\begin{aligned}\frac{\partial u_z}{\partial t} = &\frac{6}{h}\frac{\partial h}{\partial z}\frac{\partial u_z}{\partial z} + 3\frac{\partial^2 u_z}{\partial z^2} \\ &-u_z\frac{\partial u_z}{\partial z} + \frac{1}{h^2}\frac{\partial h}{\partial z}\left(S + \frac{3A}{h^2}\right).\end{aligned} \tag{6}$$

Here we have introduced the dimensionless variables $\tilde{u}_r = (\epsilon\nu/\tilde{b})u_r$, $\tilde{u}_z = (\nu/\tilde{b})u_z$, $\tilde{z} = \tilde{b}z$, $\tilde{r} = \tilde{h}_0 r$ and $\tilde{h} = \tilde{h}_0 h$. In addition we have scaled time by $\tilde{b}^2/\nu$ and the pressure difference by $(\rho\nu^2/\tilde{b}^2)$. Similar to the previous section we introduce the dimensionless constants $S = \gamma\tilde{b}/\epsilon\rho\nu^2$ and $\tilde{A} = (48\rho\nu^2\tilde{h}_0^3/\tilde{b}^2)A$, where we have assumed that there is an attractive van der Waals like potential of the form $\tilde{A}/\tilde{h}^3$. Together, equations (5) and (6) constitute a set of evolution equations for the two quantities h and u_z. These equations are identical to the evolution equations obtained by Eggers and Dupont [5] with gravity replaced by van der Waals attractions.

Note that the pressure scale specified here is slightly different from that used in the analysis of thin films. In particular, if p_{film} is the pressure scale used in the thin film scalings of the previous section and p_{thread} is the pressure scale used here, then we have that $p_{film}/p_{thread} = \epsilon$. This larger pressure scale is necessary to balance the radial curvature which is now large because of the thinness of the thread and the hoop stress it induces in the normal stress boundary condition.

Suppose that we use the full curvature in the equations, not the asymptotic correction as done above. Then we find that equation (6) becomes,

$$\begin{aligned}\frac{\partial u_z}{\partial t} = &\frac{6}{h}\frac{\partial h}{\partial z}\frac{\partial u_z}{\partial z} + 3\frac{\partial^2 u_z}{\partial z^2} - u_z\frac{\partial u_z}{\partial z} + \frac{3A}{h^4}\frac{\partial h}{\partial z} \\ &+S\left\{\frac{\dfrac{\partial h}{\partial z}}{\sqrt{1+\left(\dfrac{\partial h}{\partial z}\right)^2}}\left[\frac{1}{h^2} + \frac{\dfrac{\partial^2 h}{\partial z^2}}{h\left[1+\left(\dfrac{\partial h}{\partial z}\right)^2\right]} - \frac{3\left(\dfrac{\partial^2 h}{\partial z^2}\right)^2}{\left[1+\left(\dfrac{\partial h}{\partial z}\right)^2\right]^2}\right]\right. \\ &\left. + \frac{\dfrac{\partial^3 h}{\partial z^3}}{\left[1+\left(\dfrac{\partial h}{\partial z}\right)^2\right]^{3/2}}\right\}.\end{aligned} \tag{7}$$

The form of equation (7) is now quite similar to the transverse velocity evolution equations previously derived for thin films (equations (1) and (2)). In particular, the highest-order derivative multiplying the surface tension parameter, S in (7) is now a third-order spatial derivative as it is in the thin film evolution equations. This is in marked contrast to the evolution equation (6) derived using the correct asymptotic expression for the curvature where the the highest-order derivative multiplying S is only of first-order.

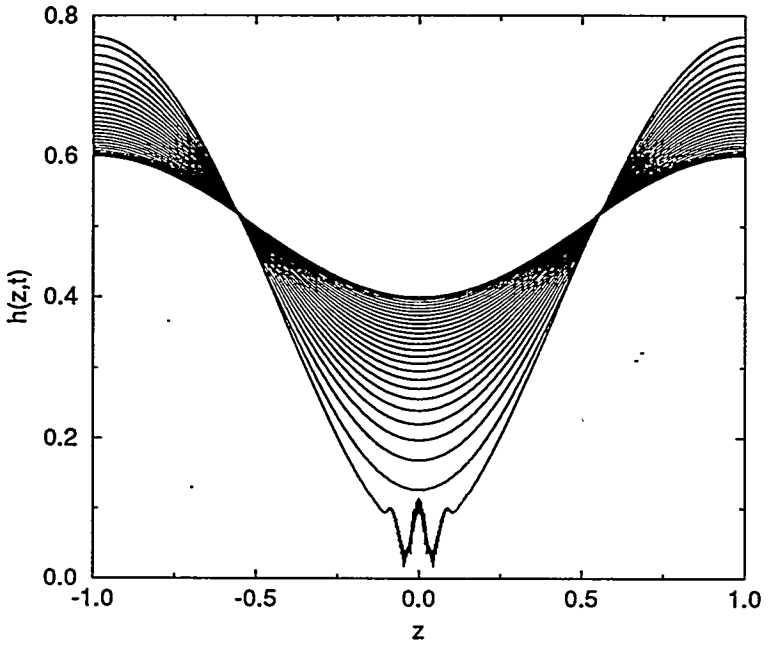


Figure 2: Evolution of an unstable perturbation to the thin thread equations solved using the finite-difference scheme. The parameter values used are $S = A = 1$; and solutions are shown for $t = 0, 0.01, \ldots, 0.25, 0.2572390625$.

To consider the linear stability of the evolution equations (5) and (6) we perturb in normal modes about the steady-state solutions $h = \hat{h} = const.$ and $u_z = 0$,

$$\begin{pmatrix} h \\ u_z \end{pmatrix} = \begin{pmatrix} \hat{h} \\ 0 \end{pmatrix} + \begin{pmatrix} \check{h} \\ \check{u}_z \end{pmatrix} e^{\omega t + ikz}. \tag{8}$$

Linearizing and applying a solvability condition, we obtain the dispersion relation for the growth rate, ω,

$$\omega = \frac{3k^2}{2}\left[-1 \pm \sqrt{1 + \frac{2}{9k^2\hat{h}}\left(S + \frac{3A}{\hat{h}^2}\right)}\right]. \tag{9}$$

Since the larger value of $\omega > 0$ for all values of the parameters involved, we find that the steady-state solution is unconditionally unstable.

If we now use the full expression for the curvature and we analyze the linear stability of equations (5) and (7), we find that the dispersion relation is instead given by,

$$\omega = \frac{3k^2}{2}\left\{-1 \pm \sqrt{1 + \frac{2}{9k^2\hat{h}}\left[S\left(1 - k^2\hat{h}^2\right) + \frac{3A}{\hat{h}^2}\right]}\right\}. \tag{10}$$

Since there are now values of the parameters for which both values of ω may be negative and the steady-state stable, we find that the condition for stability is $k > k_c$ where,

$$k_c = \frac{1}{\hat{h}}\sqrt{1 + \frac{3A}{S\hat{h}^2}}. \tag{11}$$

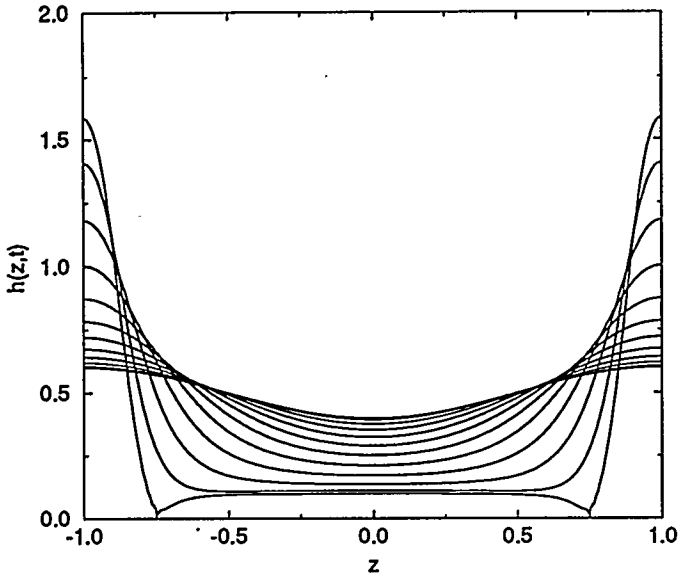


Figure 3: Evolution of an unstable perturbation to the thin thread equations solved using the finite-difference scheme. The parameter values used are $S = 100$ and $A = 1$; and solutions are shown for $t = 0, 0.01, \ldots, 0.09, 0.10, 0.10756$.

In this instance, the largest eigenvalue is positive only over a finite range of k.

Now consider the numerical solution of the evolution equations discussed above assuming periodicity on the interval $z \in [-1, 1]$. The finite-difference approximation used earlier for the two-dimensional thin film (see Ida [9] for details) is used here.

Figs. 2 and 3 are representative of the types of solutions of the coupled nonlinear system (5) and (6). Each figure is started with the same initial conditions, $(h(z,0), u_z(z,0)) = (1/2 - (0.1)\cos\pi z, 0)$, and with the same value of $A = 1$. In fig. 2 we set $S = 1$ and show solutions for $t = 0, 0.01, \ldots, 0.25, 0.2572390625$ and in fig. 3 for which $S = 100$, solutions are shown for $t = 0, 0.01, \ldots, 0.10, 0.10756$. What we find, by resolving the region near rupture more closely, is that a pair of "dimples" forms off of the axis of symmetry ($z = 0$) and leaves an isolated thin packet of fluid remaining. We see that as S increases, the size of this packet of fluid increases with the rupture points moving further away from the axis of symmetry. If we were to focus our attention on the region near rupture, we see the formation of the "dimpled" structure before the onset of numerical instabilities. (We found these instabilities extremely difficult to eliminate, because the growth-rate in equation (9) is monotonically increasing $\forall k > 0$, and the film is unconditionally unstable. Thus, the onset of instabilities exhibited in shorter wavelength modes is not limited by a cutoff wavelength.)

If we solve the equations utilizing the full curvature term, (5) and (7), we obtain solutions where the rupture point is more localized and the dimple seems to disappear. Only order one values of S have been considered and larger values need to be studied to confirm this conclusion.

In conclusion we note that in this section we have studied a set of leading-order evolution equations for the interface shape and transverse velocity of an axisymmetric thin liquid thread

subject to a van der Waals like force. Based upon a linear stability analysis, we have found these equations to be unconditionally unstable. These equations have been solved numerically. We find that as the effects of surface tension are increased, rupture occurs over a wider spatial domain, and a "dimpled" structure is exhibited. This is in marked contrast to the characteristic rupture behavior of thin films which is confined to a small spatial region.

ACKNOWLEDGEMENT

This research was supported in part by Department of Energy grant DE-FG02-88ER13927.

References

[1] T.-M. Tsai and M.J. Miksis, Dynamics of a drop in a constricted capillary tube. *J. Fluid Mech.*, 274, 197-217 (1994).

[2] T. F. Dupont, R. E. Goldstein, L. P. Kadanoff, and S.-M. Zhou, Finite-time singularity formation in Hele-Shaw systems. *Phys. Rev. E*, 47, 4182–96 (1993).

[3] R. E. Goldstein, A. I. Pesci, and M. J. Shelley, Topology transitions and singularities in viscous flows. *Phys. Rev. Lett.*, 70, 2043–2046 (1993).

[4] J. Eggers, Universal pinching of 3d axisymmetric free surface flow. *Phys. Rev. Lett.*, 71, 3458–3460 (1993).

[5] J. Eggers and T. F. Dupont, Drop formation in a one-dimensional approximation of the Navier-Stokes equation. *J. Fluid Mech.*, 262, 205–21 (1994).

[6] T. Erneux and S. H. Davis, Nonlinear rupture of free films. *Phys. Fluids A*, 5, 1117–22 (1993).

[7] A. J. Bernoff and A. L. Bertozzi, Singularities in a modified Kuramoto-Sivashinski equation describing interface motion for phase transition. *Physica D*, 85, 375-404 (1995).

[8] A. L. Bertozzi, Symmetric singularity formation in lubrication-type equations for interface motion. preprint.

[9] M. P. Ida, The Dynamics of Thin Liquid Films, Ph.D. thesis, Northwestern University, Dec. 1995.

[10] G. I. Barenblatt and Ya. B. Zel'dovich, Self-similar solutions as intermediate asymptotics. *Ann. Rev. Fluid Mech.*, 4, 285-312 (1972).

DETERMINATION OF NEAR-SURFACE MATERIAL PROPERTIES BY LINE-FOCUS ACOUSTIC MICROSCOPY

Jan D. Achenbach and Wei Li

Center for Quality Engineering and Failure Prevention
Northwestern University
Evanston, IL 60208, U.S.A.

ABSTRACT

A line-focus acoustic microscope is used in conjunction with a multiple wave-mode method to determine elastic constants from a single $V(z)$ measurement. $V(z)$ curves which include contributions from different wave modes, measured using the line-focus acoustic microscope at 225 MHz, have been compared with theoretical results predicted by a $V(z)$ measurement model. The determination of elastic constants has been achieved numerically by seeking a set of elastic constants that leads to the best fit, in the least square sense, of the theoretical results to the experimental ones. The method has been applied to isotropic materials in bulk, and plate and thin-film configurations. Elastic constants for each of these cases have been determined. The consistency, convergence, sensitivity and accuracy of the procedure have been investigated.

INTRODUCTION

Line-focus acoustic microscopy (LFAM) provides a method to determine the elastic constants of homogeneous specimens and thin-film/substrate configurations. The elastic constants are determined from the velocities of leaky acoustic waves that can be obtained from $V(z)$ measurements. The $V(z)$ curve, which is a record of the transducer output voltage V as a function of the distance z between the lens focus and the specimen surface, is unique to a material and referred to as the material signature. Generally speaking, more than one elastic constant has to be determined, and hence more than one data point is required. Hence, for isotropic materials sufficient data cannot be procured with a single mode. For anisotropic solids the velocity can be measured as a function of the angle defining the propagation direction on the surface to yield a sufficiently large set of data. The technique has been discussed in great detail in a recent review article, which also lists numerous references [1]. For thin-film/substrate configurations measurements at various frequencies or for different film thicknesses may be carried out to obtain sufficient data using standard measurement procedures [1]. There are, however, obvious advantages to work with a single specimen and at a single frequency. As discussed in this paper, this can be done by considering the contributions of more than one leaky wave mode to the $V(z)$ curve.

This paper presents a multiple mode method to determine, with a single-frequency $V(z)$ measurement on a single specimen, the elastic constants of bulk isotropic solids, thin isotropic plates and thin-films on substrates. The $V(z)$ curve which includes contributions from multiple leaky acoustic waves has been measured experimentally using a line-focus acoustic microscope at 225 MHz. $V(z)$ curves have also been simulated numerically using a measurement model with selected elastic constants. Both the experimental and the numer-

ical $V(z)$ curves go through the same $V(z)$ analysis to yield their respective predictions for the leaky acoustic wave velocities. The determination of elastic constants is then achieved through minimization of the differences between the theoretical predictions and the experimental results by a numerical iterative searching procedure known as the simplex method. The accuracy, consistency, convergence and sensitivity of the numerical inversion have also been studied in this paper.

THIN FILMS

Thin film materials are already widely used, and they promise to have significant additional applications in future technology. There are many present and potential applications of configurations consisting of thin films deposited on a substrate.

A first important application is concerned with thin coatings to protect surfaces of components from wear, impact, corrosion, and thermal disturbances. Hard and wear-resistant coatings are an important segment in the US and world economy. The need to extend the wear-life using coatings is not only to save cost, but also reduction of downtime.

Diamond film has a number of remarkable properties. It is the hardest substance known, and it has a higher modulus of elasticity than any other material. When free of impurities, it has one of the highest resistivities. It also combines a very high thermal conductivity with a low thermal expansion coefficient to yield high resistance to thermal shock. Lastly, diamond is very resistant to chemical attack. However, diamond coatings are not good with ferrous alloys because diamond reacts with steel at high temperatures.

Transition-metal nitride films are commonly used as hard, protective coatings for softer surfaces. Superlattice films, including TiN/NbN, TiN/VN, and TiN/VNbN, have, however, been shown to exhibit much higher hardness than homogeneous nitride films. The elastic constants of these films on a substrate are difficult to measure.

LINE-FOCUS ACOUSTIC MICROSCOPY

An acoustic microscope consists of four main components: the acoustic probe, the pulse-mode measurement system for transmitting and receiving electrical signals, the mechanical systems for alignments and movements of the sample and a computer for controlling the system and processing the recorded wave forms.

The propagation of surface acoustic waves along the interface of a solid material and air or water provides a useful means to determine material constants of the solid by measuring propagation velocities. For small specimens, or when local values of the material constants must be determined, the measurements require small wavelengths corresponding to high frequencies. For such cases a single water coupled focused transducer can be used advantageously, and the ultrasonic technique is referred to as quantitative acoustic microscopy.

Particularly useful is a water-coupled line-focus acoustic lens since such a lens allows the measurement of the SAW velocity in specified directions. Hence line-focus acoustic microscopy (LFAM) has been used to determine the elastic constants of anisotropic materials.

The best known technique measures the $V(z)$ curve, which is a record of the magnitude of the transducer's voltage output V as a function of the distance z between the focal line of the lens and the specimen surface. It can be shown that the spacing of peaks or valleys in the oscillatory $V(z)$ curve is directly related to the propagation velocity of the surface wave, see [1]-[2].

$V(z)$ *Curve Measurement*

The experimental results have been obtained with a Honda AMS-5000 ultrasonic measurement system and a line-focus acoustic lens operating at $225MHz$. The detailed description of this system can be found in Reference [1]. Figure 1 shows schematically the configuration of the acoustic probe and the specimen. A ZnO-film transducer generates and detects longitudinal waves at the flat surface of a Z-cut sapphire rod. The acoustic beam is focused by an acoustic lens with a cylindrical concave surface at the other end of the rod. The cylindrical concave surface has a radius of $1.0mm$ and an aperture half-angle of 60°. The operating frequency is around $225MHz$ and the focal length of the lens is $1.15mm$. For efficient transmission of acoustic waves through the lens couplant interface, a chalcogenide glass film with a quarter wavelength thickness is deposited on the cylindrical concave surface. A specimen is placed on a mechanical stage, translated in the vertical direction and rotated around the axis of the rod.

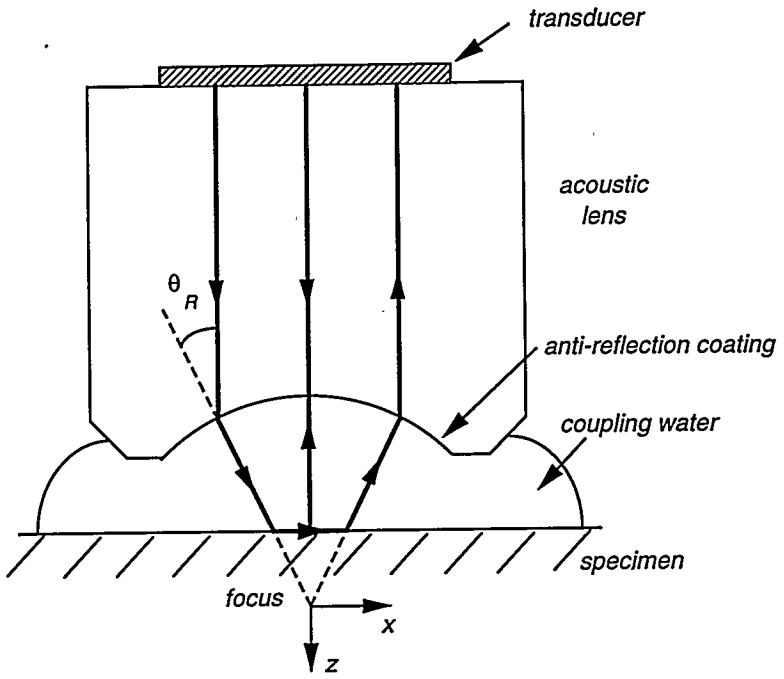


Figure 1: Cross-sectional configuration of the line-focus acoustic probe.

$V(z)$ *Measurement Model*

Basic to the interpretation of $V(z)$ curves is a reliable measurement model. A $V(z)$ measurement model simulates the measurement procedure, including any systemic errors that may occur in the determination of the velocity from experimental $V(z)$ curves. For example, effects due to multiple modes, will be replicated in the numerical model. The material constants obtained from comparisons of results from the measurement model and experiments will, therefore, be free of these systemic errors.

A measurement model for the $V(z)$ curve has been described in detail in Ref. [1]. According to the model, the output voltage of the transducer can be expressed as

$$V(z) = \int_{-\infty}^{+\infty} \exp(2ik_z z) L_1(k_x) L_2(k_x) R(k_x) dk_x \tag{1}$$

where $k_z = (k_w^2 - k_x^2)^{1/2}$ and k_w is the wave number in the coupling fluid. $L_1(k_x)$ is the angular spectrum of the wave field at the focal plane generated by a plane wave in the buffer rod, and $L_2(k_x)$ is the voltage response of the transducer when a plane wave of unit amplitude and wave vector (k_x, k_z) is insonifying the lens. A detailed description and discussion of the angular spectrum function $L_1(k_x)$ and the response function $L_2(k_x)$ can be found in Ref. [1], where integral expressions as well as numerical calculations have been presented. $R(k_x)$, which is the reflectance function of the fluid-loaded specimen, can generally be expressed as:

$$R(k_x) = \frac{A - \beta B}{A + \beta B} \tag{2}$$

where $\beta = i\rho_w \omega^2 / k_z$, and ρ_w is the coupling fluid density. A and B have different definitions for different sample geometries. For a fluid-loaded bare-substrate-specimen or a fluid-loaded layered-specimen, A and B can be found in Ref. [1]. For a fluid-loaded plate-specimen, a numerical approach proposed by Chimenti and Nayfeh [3] has been adopted here. A summary of the calculation of $R(k_x)$ for an isotropic plate is given in the Appendix of Ref. [4].

To determine the leaky surface wave velocities a procedure completely analogous to the one for the experimental $V(z)$ curve is applied to the theoretical result given by Eq. (1).

Determination of Elastic Constants

The elastic constants are obtained by seeking a set of elastic constants that yields the best fit in the least square sense of the theoretical predictions to the experimental results. A numerical iterative searching procedure known as the simplex method has been used to find the set of elastic constants that minimizes the deviation between theoretical and experimental results.

A dimensionless deviation function, D, is defined in terms of the measured and the calculated quantities as:

$$D = \sum_{n=1}^{N} \left\{ W_n^{(v)} \left| \frac{v_n^c - v_n^m}{v_n^m} \right| + W_n^{(A)} \left| \frac{A_n^c - A_n^m}{A_n^m} \right| \right\} \tag{3}$$

where v_n^m and A_n^m are the velocity and the peak-amplitude in the Fourier domain of the n-th leaky surface mode obtained from the measured $V(z)$ curve; v_n^c and A_n^c are the same quantities obtained from the calculated $V(z)$ curve. It is noted that different weights, $W_n^{(v)}$ and $W_n^{(A)}$, have been imposed on the differences in velocities and spectrum amplitudes in order to get the best results. Both experimental and numerical errors have been taken into consideration in choosing these weights. For example, more weight has been put on the velocity part than on the amplitude part, because the accuracy of the velocity measurements is usually about an-order-of-magnitude better.

The selection of the starting values for the application of the simplex method is in general a matter of trial and error. Different selections of the starting values should yield the same answer, but due to the existence of local minimums, a good initial guess of the starting values is very important for quick convergence to the true values.

SOME RESULTS

Specimens with three different sample configurations have been tested. Bulk specimens include a modified borosilicate glass AF45, a glass slide and an aluminum sample; the plate specimen is a modified borosilicate thin glass plate D263 of 50 μm thickness; and thin-film

specimens include a 2.2 μm titanium-film deposited on an aluminum substrate and a 0.1 μm gold-film deposited on a glass-slide substrate.

For the bulk specimens, the leaky Rayleigh wave and the LSSCW have been used for the determination of the elastic constants. For the D263 glass plate the zeroth-order Lamb-mode and a higher-order Lamb-mode have been used. For the thin-film/substrate specimens, multiple LSAW modes have been used to determine the thin-film elastic constants. The dispersive leaky Rayleigh wave and the LSSCW have been used for the titanium-film on an aluminum substrate configuration, and two generalized leaky Lamb modes, namely the dispersive leaky Rayleigh wave and the leaky Sezawa wave mode, have been used for the gold-film on a glass-slide substrate configuration.

The determined elastic constants are listed in Table 1. Known values of the densities and thicknesses have been used. For the thin-film/substrate cases, the elastic constants of the substrates were known. It should be noted that accurate information of the substrate properties is very important for the determination of thin-film constants. Elastic constants determined in this paper for the bulk aluminum and the glass-slide have been used for the substrates.

Table 1: Elastic constants determined for materials in various configurations.

Specimen	Thickness (μm)	Density (kg/m^3)	Waves or Modes used	Determined Elastic Constants (GPa)	
				E	G
AF45 Glass (bulk)	N/A	2720	leaky Rayleigh & LSSCW	63.2	25.6
Glass Slide (bulk)	N/A	2459	leaky Rayleigh & LSSCW	71.0	28.5
Aluminum (bulk)	N/A	2700	leaky Rayleigh & LSSCW	70.4	26.3
D263 Glass (plate)	50	2510	0-th Lamb & higher Lamb	73.1	30.6
Titanium film (on Aluminum)	2.2	4508	dispersive leaky Rayleigh & LSSCW	87.7	33.2
Gold film (on Glass slide)	0.1	19281	dispersive leaky Rayleigh & leaky Sezawa	79.5	28.0

DISCUSSION

Comparisons between Theoretical Predictions & Experimental Results

The sets of elastic constants that have been determined can be verified by comparing the corresponding theoretical velocities with the experimental results. Figure 2(a)(b) shows $V(z)$ curve and $V(k)$ curve comparisons, where the dashed lines are the experimental results

and the solid lines are the theoretical predictions using the determined elastic constants.

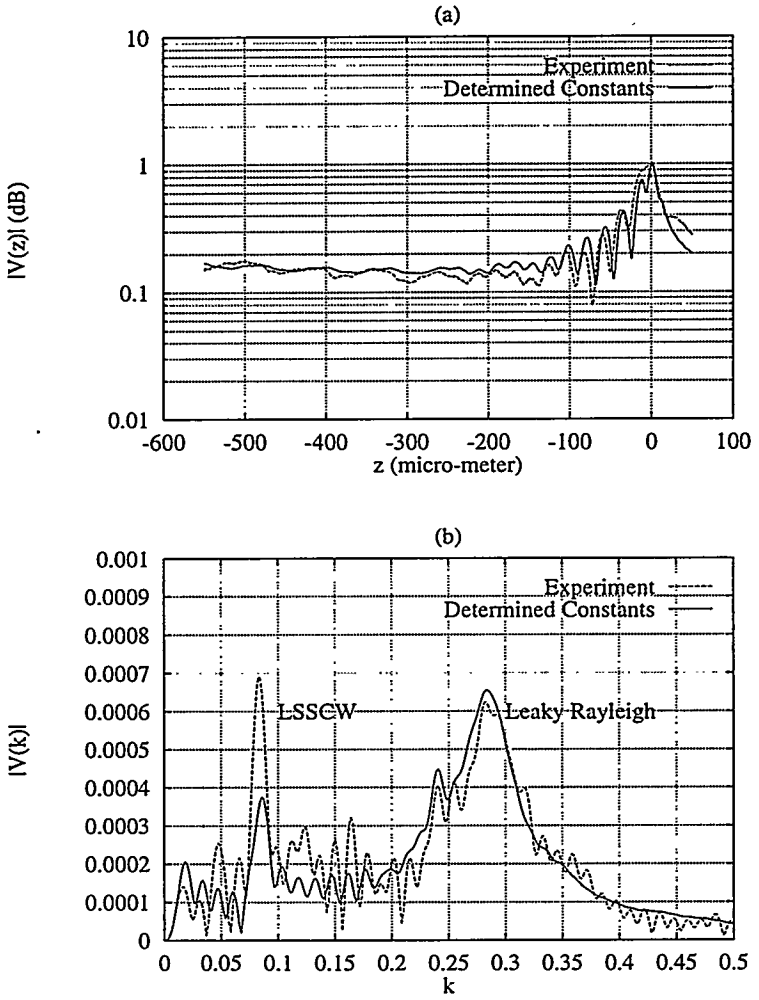


Figure 2: (a) $V(z)$ curve comparisons for AF45 bulk glass. (b) $V(k)$ curve comparisons for AF45 bulk glass.

Consistency of Determined Elastic Constants

The elastic constants that have been determined from one set of experimental data can be checked by using them to calculate $V(z)$ curves for another independent experimental setup. For example, the elastic constants of D263 glass plate that have been determined from the water/plate/air configuration can be verified using the water/plate/water configuration.

Convergence due to Use of Multiple Leaky Acoustic Waves

As discussed in section 2.4, the numerical procedure includes information from different leaky acoustic waves. The use of multiple leaky wave information reduces the region in the plane of the material constants E and G for which the deviation D, defined by Eqn. (3), has

a prescribed small value. This improves the convergence of the procedure. For the AF45 bulk glass sample, Figure 3 shows the reduction of the region for $D < 0.024$ due to the use of multiple leaky acoustic waves. Figure 3 displays the changes of the deviation field D as the Young's modulus and the shear modulus vary around the determined values for the five cases: (a) if only leaky Rayleigh wave velocity information is used for the minimization of D; (b) if only LSSCW velocity information is used; (c) if only leaky Rayleigh wave amplitude information is used; (d) if only LSSCW amplitude information is used; (e) if all the above (a)-(d) have been used and combined in Eqn. (3).

It is interesting to note, from Figures 3(a)-(d), that despite the obvious convergence problems implied by the extent of the dark areas, the areas have different slopes. It is the different directions of the slopes in Figure 3(a) and Figure 3(b)(c)(d) that lead to the convergence of our numerical inversion when Eqn. (3) is used, as indicated by the finite extent of the dark domain in Figure 3(e).

CONCLUSIONS

$V(z)$ measurements of multiple leaky acoustic waves by line focus acoustic microscopy have been used to determine elastic constants. The method presented in this paper has the advantage that it requires only a single V(z)-curve measurement by the LFAM. It serves as a very effective approach to determine elastic constants of a single isotropic specimen since only one single-frequency $V(z)$ measurement is required. Elastic constants for isotropic materials in bulk, and for plate and thin-film/substrate configurations have been determined to satisfactory accuracy. It has been shown that the use of multiple leaky wave information reduces the region in the plane of the material constants E and G for which the deviation D has a prescribed small value. This improves the convergence of the method. Consistency of the results has been shown by verifying the elastic constants determined from one set of experimental data with results from another experimental configuration. The sensitivity of the deviation D to variations of the elastic constants has also been investigated. Generally D is more sensitive to changes of Young's modulus.

ACKNOWLEDGMENT

This work was supported by the Department of Energy, Division of Engineering and Geosciences, Office of Basic Energy Sciences, under Grant DE-FG02-86ER13484.

REFERENCES

1. J. D. Achenbach, J. O. Kim and Y. C. Lee, "Measuring Thin-Film Elastic Constants by Line-Focus Acoustic Microscopy," in *Advances in Acoustic Microscopy*, Vol. 1, edited by Andrew Briggs, 153-208, Plenum Press, New York (1995).

2. J. Kushibiki and N. Chubachi, "Material Characterization by Line-Focus Beam Acoustic Microscope," *IEEE Trans. Sonics Ultrason.*, SU-32, 189-212 (1985).

3. D. E. Chimenti and A. H. Nayfeh, "Ultrasonic Reflection and Guided Waves in Fluid-coupled Composite Laminates," *J. Nondestr. Eval.*, 9, 51-69 (1990).

4. W. Li and J. D. Achenbach, "$V(z)$ Measurement of Multiple Leaky Wave Velocities for Elastic Constant Determination", *J. Acoust. Soc. Am.*, in press.

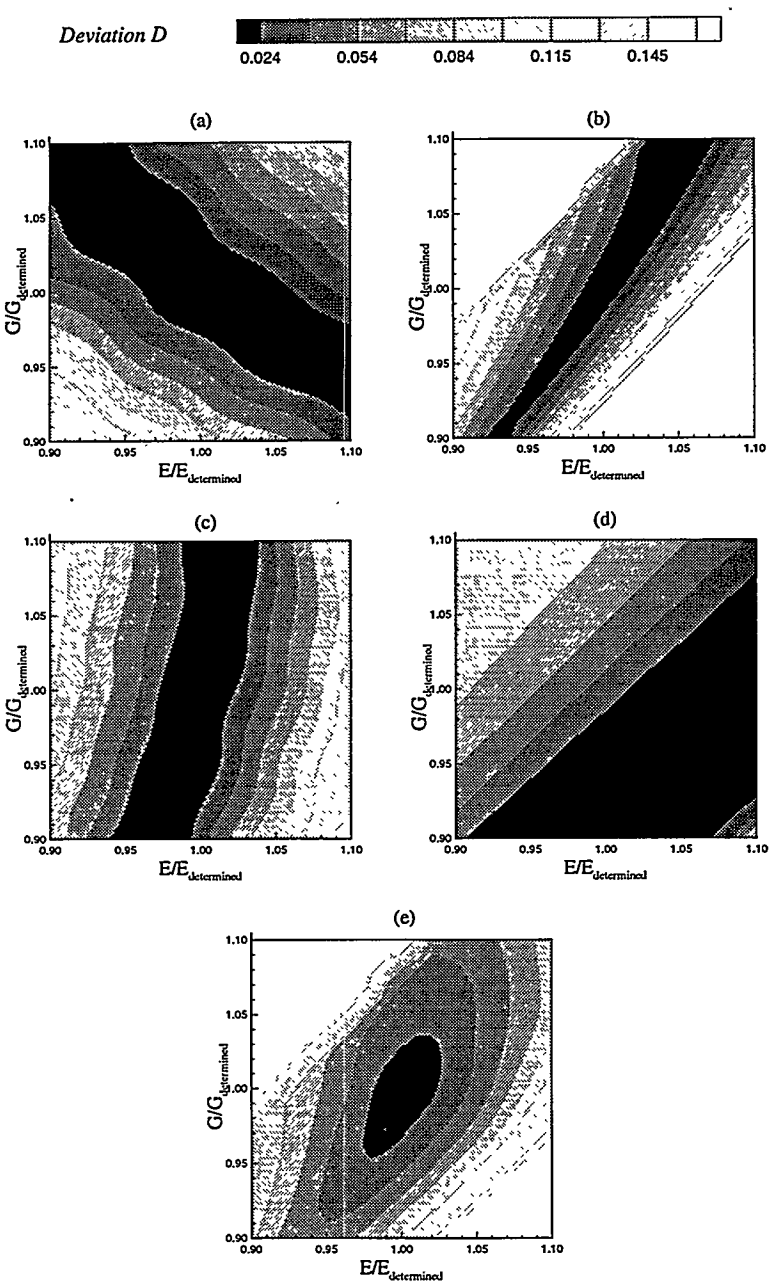


Figure 3: Reduction of region for $D < 0.024$ when multiple leaky acoustic waves/wave parameters are used for AF45 bulk glass. (a) Deviation field when only leaky Rayleigh wave velocity is used; (b) Deviation field when only LSSCW velocity is used; (c) Deviation field when only leaky Rayleigh wave amplitude is used; (d) Deviation field when only LSSCW amplitude is used; (e) all of the above have been used.

CRYOTRIBOLOGY OF DIAMOND AND GRAPHITE

Yukikazu Iwasa, Ahmet F. Ashaboglu, and Ernest R. Rabinowicz

Francis Bitter Magnet Laboratory
and
Department of Mechanical Engineering
Cambridge MA 02139 USA

ABSTRACT

An experimental study was carried out on the tribological behavior of materials of interest in cryogenic applications, focusing on diamond and graphite. Both natural diamond (referred in the text as diamond) and chemical-vapor-deposition (CVD) diamond (CVD-diamond) were used. The experiment was carried out using a pin-on-disk tribometer capable of operating at cryogenic temperatures, from 4.2 to 293 K. Two basic scenarios of testing were used: 1) frictional coefficient (μ) *vs* velocity (v) characteristics at constant temperatures; 2) μ *vs* temperature (T) behavior at fixed sliding speeds. For diamond/CVD-diamond, graphite/CVD-diamond, stainless steel/CVD-diamond pairs, μ's are virtually velocity independent. For each of diamond/graphite, alumina/graphite, and graphite/graphite pairs, the $\partial\mu/\partial v$ characteristic is favorable, *i.e.*, positive. For diamond/CVD-diamond and graphite/CVD-diamond pairs, μ's are nearly temperature independent between in the range 77 – 293 K. Each μ *vs* T plot for pin materials sliding on graphite disks has a peak at a temperature in the range 100 – 200 K.

INTRODUCTION

The principal objectives of this research were to advance the theoretical understanding of low-temperature sliding behavior and to expand the cryogenic tribology data base, particularly for very hard materials. Hard materials are of interest because of their generally high load bearing capacities, low friction coefficients, and wear resistance. The emphasis here is on the frictional behavior of materials of interest to cryogenic applications sliding against CVD-diamond film and graphite.

Early in the next century, high-temperature superconducting (HTS) technology and cryogenics are expected to play key roles in competitive global markets encompassing energy, information, medicine, transportation, space, and basic science [1]. We have been studying

the cryotribology of materials relevant to applications in these fields. A notable recent development in "new materials" is the remarkable progress achieved in materials fabrication techniques, *e.g.*, chemical-vapor-deposition (CVD), for synthesizing diamond films. The CVD-diamond films possess properties very similar to those of bulk diamond, making them promising for new applications such as in HTS magnet technology. The CVD-diamond may be usable in HTS magnets as an interface material as a good electrical insulator as well as a good thermal conductor.

TRIBOLOGY OF DIAMOND AND GRAPHITE

The friction of diamond and graphite has long been known to be low [2]. Gardos and Soriano [3], Samuels and Wilks [4], Kohzaki, *et al.* [5], Jia, *et al.* [6], Bowden and Hanwell [7] have reported on the frictional behavior of diamond, either natural or CVD-processed film, and found both to have low values of friction coefficient, typically 0.05-0.1, with materials exposed in the atmosphere. Mody, *et al.* [8], Petlyuk, *et al.* [9], Khopin [10], Nishiyama, *et al.* [11], Cameron [12] have reported frictional data for graphite and graphite-filled composites. Their results indicate values in the range of 0.05 – 0.15.

EXPERIMENTAL PROCEDURE AND MATERIALS

Apparatus The tests were performed with a rotational pin-on-disk tribometer built to keep specimens at cryogenic temperatures in the range 4.2 – 293 K [13]. A rotating specimen disk slides against three hemispherical specimen pins, symmetrically spaced on a 51-mm bolt circle diameter. The specimens are immersed in a cryogen to achieve and maintain the test temperature: nitrogen (77 K) and helium (4.2 K). The tribometer operates at a constant nominal sliding velocity, ranging from 10^{-7} to 0.1 m/s. Velocity is regulated through a computer-controlled DC servo motor and a set of precision gear reducers. Normal loads (7.5 – 22.4 N) are applied by placing weights on the load ring which transfers the force through a pulley system. A constantan-chromel thermocouple imbedded in the stationary disk sample monitors the specimen temperature.

The friction force is measured by strain arms which prevent the rotation of the specimen pins. The friction force is monitored continuously during test with a computer-based data acquisition system. Although the use of spherical pins results in highly localized contact at the sliding surfaces, the friction coefficients obtained under these conditions are valid because the frictional force is generally independent of the apparent contact area [14].

Materials Table 1 presents a listing of the materials tested, both in disk and pin form. The mechanical properties of the materials are well known at room temperature and to a lesser extent at cryogenic temperatures [15 – 17]. It is not the intention of this project to examine all possible combinations of materials, but rather to test certain combinations that were of interest.

Table 1: Pin and Disk Materials

Pin Materials	*Disk Materials*
Natural diamond	CVD-diamond film
Graphite	Graphite
Copper	Copper
Stainless steel*	Stainless steel†
Alumina	

* AISI440C (pins) and AISI316 (disks).

Both diamond and graphite are crystalline forms of carbon [18]. Diamond is a covalently bonded solid; its bonding arrangement comprises four equivalent covalent bonds directed towards the four corners of a tetrahedron. By comparison, graphite is a layer-lattice material consisting of three equivalent 120° bonds in the basal plane with a fourth, hybrid resonant orbital directed perpendicular to the plane [19]. Graphite's solid lubricant behavior is chiefly due to the easy shear of these interplanar hybrid orbitals. The microwave plasma CVD method was used to synthesize the diamond-coated films for our cryotribological study. In the microwave method, a substrate is placed in a quartz tube and a stream of hydrocarbon gas, *e.g.*, methane diluted by hydrogen, is converted into a film of crystalline diamond by microwave heating. Although silicon is the standard material for substrates, several other substrate materials are possible, including ceramics (silicon carbide, tungsten carbide, silicon nitride, alumina) and metals (tantalum, molybdenum, and tungsten). The diamond is deposited on the substrate at a rate of $0.3 - 0.5\,\mu$m/h. The substrate material used in the experiments was silicon.

Pin and Disk Preparation Graphite pins were obtained as 6.35-mm (1/4-in) diameter rods and turned to their final shape with a radius form tool. Alumina and AISI440C stainless steel pins were obtained in the form of 6.35-mm diameter balls. Three natural diamond pieces, each approximately 1/3 carat and of irregular shape were mounted to allow contact with the specimen disks. The graphite disks were machined from plate stock and faced with a single point cutting tool to remove surface irregularities. The disks were then randomly abraded against 320-grit silicon carbide paper under running water. This cleaned the disks and provided a consistent surface finish between tests. Since water does not wet a surface contaminated with organic contaminants, a disk was deemed clean when a water layer wetted the surface.

Surface Cleaning Before a test, three of the desired pins were mounted in a brass specimen holder, wiped clean with a swab soaked in methanol, allowed to air dry, and placed into the apparatus. Each disk was first rinsed with methanol to displace the water layer and air dried on clean tissue paper. Later, it was placed in an oven at approximately 100°C overnight, to bake away any residual contaminants on the surface.

Testing Two types of measurement were performed: 1) friction (μ) *vs* velocity (v) at constant temperatures (4.2, 77, and 293 K); 2) μ *vs* temperature (T) at a constant sliding velocity of 10^{-5} m/s. For each speed setting, a distance of at least 3 mm was covered to allow the measured friction to achieve a steady value [20]. A velocity of 10^{-5} m/s used in μ *vs* T measurements allowed a sufficient number of data points at a given, though not absolutely constant, temperature and yet limited the total sliding distance of the run to ~2 m. A slowly changing temperature environment could be created in the cooldown process during cryogen transfer into the test cryostat or in the subsequent warm-up process after the cryogen transfer. The friction data could be monitored during either the cooldown or the warm-up process. Both processes showed essentially the same μ *vs* T behaviors. However, because of the lower magnitudes of $\partial T/\partial t$, typically of less than 20 K/h in the warm-up processes compared with 60 – 80 K/h in the cool-down processes, the warm-up processes were used. Because a quiescent helium boil-off rate from the test cryostat was in the range 0.10 – 0.20 l/h, an initial pool of liquid helium provided ~30 hours for each μ *vs* T measurement sequence.

RESULTS AND DISCUSSIONS

Tribological data are often unreproducible and sometimes inexplicable. Friction by its nature is very complex. The factors that contribute to a tribological process may be divided into three classifications: those we know, those we don't, and those we can control. The first is finite, the second may not be, and the third is very finite indeed. The conditions of tribological experiments are particularly crucial. Although utmost care for consistency was taken for each experimental sequence, often ambient conditions, *e.g.*, room temperature, air humidity, cleanliness of the air, have influenced the outcome. When natural (expected) variation of the structures and composition of the materials tested are added to the equation, large variability in results was sometimes encountered. The data gathered will be presented according to two classifications: μ *vs* v; and μ *vs* T. Because of space limitations, only selected sets of data are presented here.

Friction vs Velocity Figures 1 – 6 show μ *vs* v graphs, each with three plots corresponding to temperatures of 4.2, 77, and 293 K. Each μ *vs* v plot is the arithmetic mean of two sets of measurement. Data for the two sets are within 10% at all points for all material pairs. For a given velocity, each point represents the average value of at least 100 data points. The sliding speeds tested ranged from 10^{-6} m/s to 0.1 m/s. For each speed setting, a distance of at least 3 mm was covered to allow the measured friction to achieve a steady value. The applied load was 14.3 N.

Figure 1 – 3 shows, respectively, data diamond/CVD-diamond, stainless steel/CVD-diamond, and graphite/CVD-diamond pairs at three temperatures. Data clearly show that μ at each temperature is virtually velocity independent, confirming a general friction law on the independence of friction on velocity. Data for diamond/graphite, alumina/graphite, and graphite/graphite pairs are shown, respectively, in Figs. 4 – 6. Here for each pair,

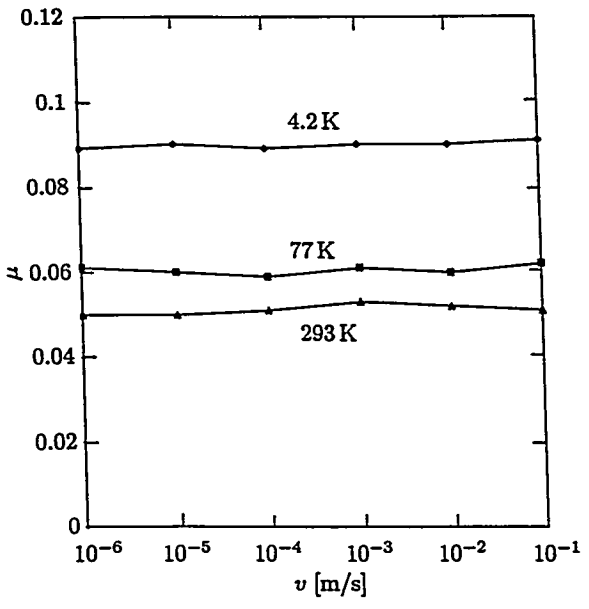


Fig. 1 μ *vs* v for diamond/CVD-diamond pairs.

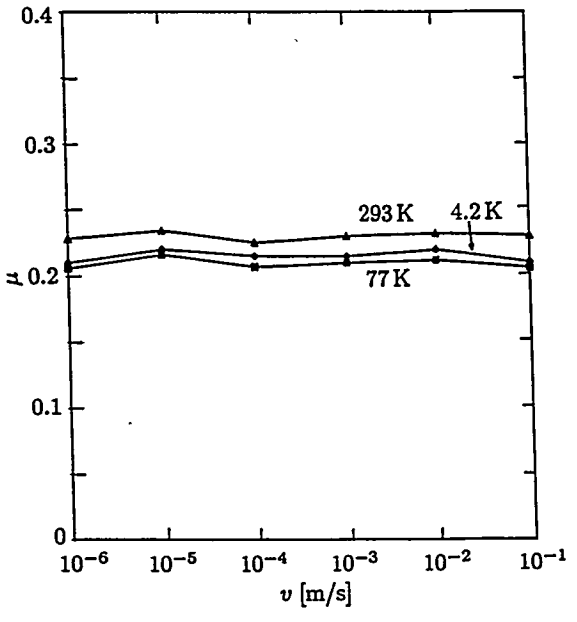


Fig. 2 μ *vs* v for stainless steel/CVD-diamond pairs.

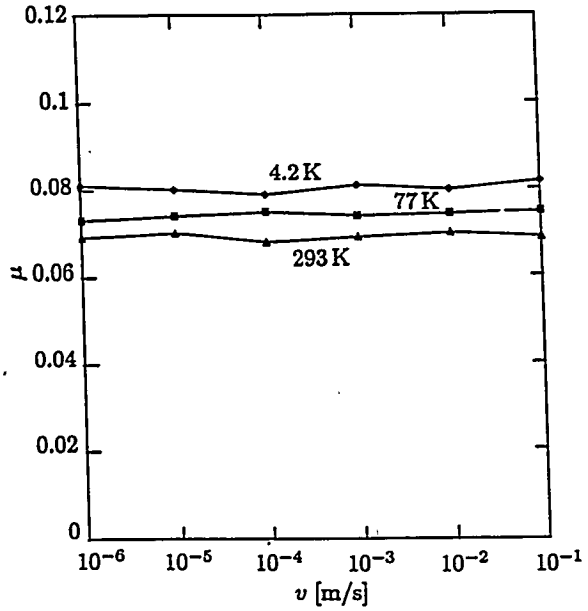


Fig. 3 μ *vs* v for graphite/CVD-diamond pairs.

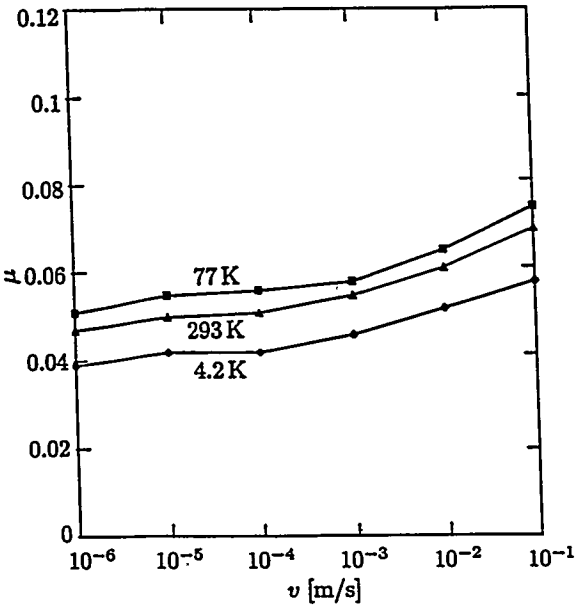


Fig. 4 μ *vs* v for diamond/graphite pairs.

μ increases with v at all three temperatures. The increase can be attributed to the continuous creation of wear particles that are later entrapped between the sliding surfaces. Since graphite is a softer material in comparison with materials like diamond and alumina, the hard pins plow the surface of graphite, wearing graphite and generating wear particles during sliding. When the sliding speed is increased, the rate at which these wear particles are generated is also increased. This results in more wear particles to be entrapped between the sliding surfaces, increasing μ. When graphite pins slide on graphite, again wear particles are generated, this time from both surfaces. Here, because pins and disk are of the same material, adhesion also plays a major role.

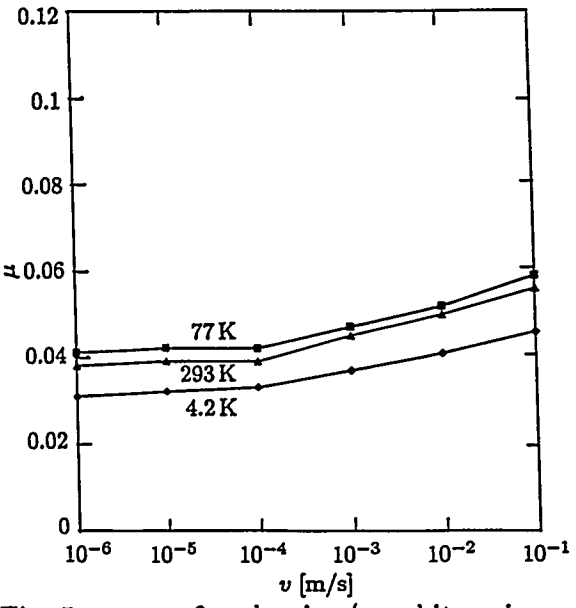


Fig. 5 μ *vs* v for alumina/graphite pairs.

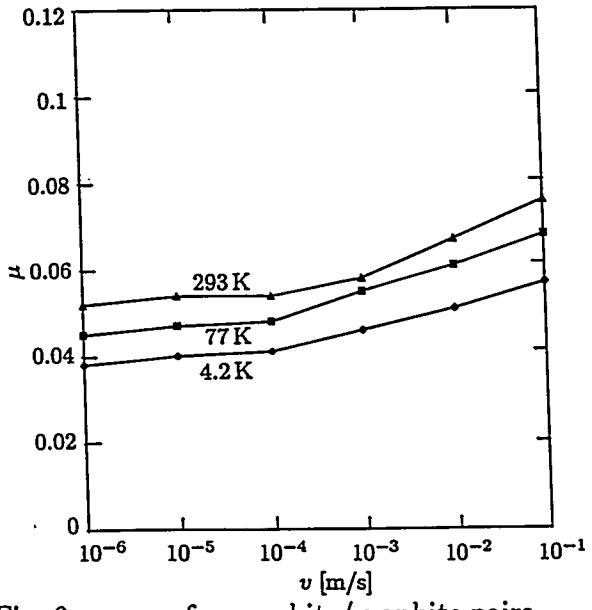


Fig. 6 μ *vs* v for graphite/graphite pairs.

Friction vs Temperature Figures 7 – 10 show μ *vs* T traces, each taken at a constant velocity of 10^{-5} m/s. For each temperature span, the total sliding distance was ~2 m.

Data in Figs. 7 and 8 are, respectively, diamond/CVD-diamond and graphite/CVD-diamond pairs. For the diamond/CVD-diamond pair, μ is 0.09 at 4.2 K and decreases with temperature to 0.05 at 293 K, remaining constant at 0.06 in the middle range 75 – 225 K. According to the adhesion theory, this pair, because both materials are identical, should have high values of μ [21]. Our results appear to be quite the opposite: it has a low value of 0.05 for most of the temperature range tested.

Data for the CVD-diamond/graphite pair (Fig. 8) show a fairly constant μ (0.07 – 0.08) with temperature. Although small, μ varies somewhat in the temperature range 50 – 250 K. This variation may be attributable to the wear particles of graphite, the softer of the material pair. As sliding proceeds wear particles are generated and entrapped between the sliding surfaces, increasing the friction. As the sliding continues, these particles are either plastically deformed or removed from the sliding path, hence reducing the friction. This process of wear particles being generated and later eliminated from the sliding surface could account for the variation in the coefficient of friction. As might be surmised from the three data points at $v = 10^{-5}$ ms/s of Fig. 2, the μ *vs* T plot of stainless steel/CVD-diamond pairs (not shown here) indicates that μ increases gradually above 77 K, reflecting the fact that μ is inversely proportional to stainless steel's hardness.

Data in Figs. 9 and 10 are, respectively, for pins of natural diamond and graphite *vs* graphite disks. For the diamond/graphite pair, μ is roughly temperature independent and equal to ~0.04, from 4.2 K to ~70 K. μ increases with temperature in the range ~70 – ~160 K. This increase is attributed to macroscopic stick-slip, typically observed in this temperature range [22 – 24]. Above ~160 K, μ decreases, reaching at room temperature a value roughly the same as that at 4.2 K. The graphite/graphite data (Fig. 10) are similar to the diamond/graphite data, except here the peak value is ~0.06.

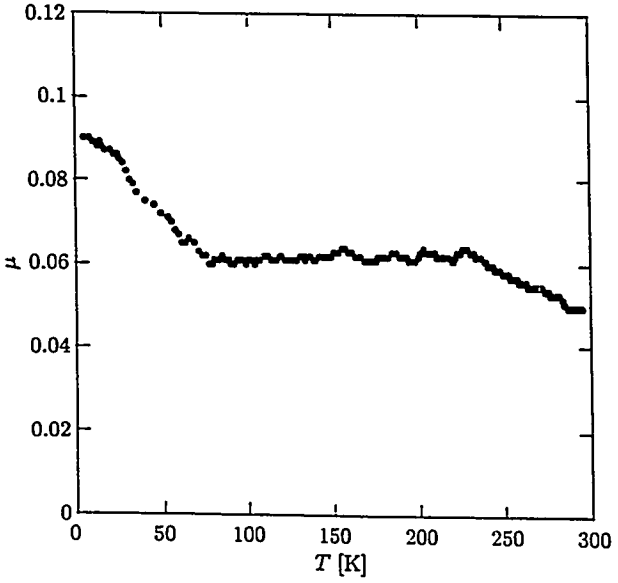


Fig. 7 μ *vs* T at $v = 10^{-5}$ m/s for diamond/CVD-diamond pairs.

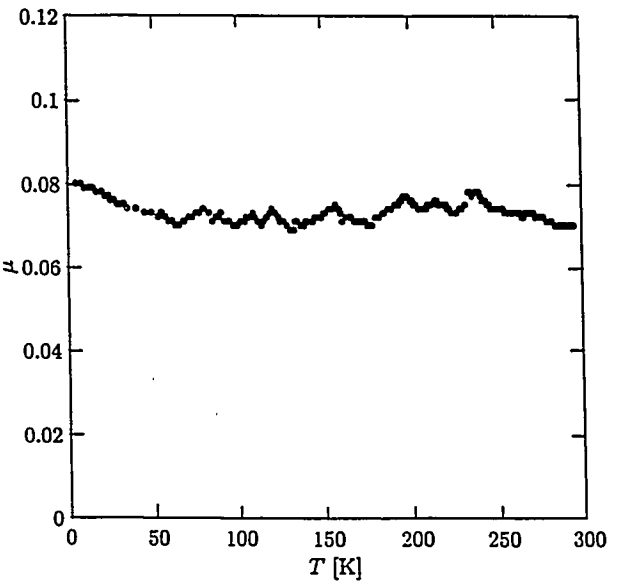


Fig. 8 μ *vs* T at $v = 10^{-5}$ m/s for graphite/CVD-diamond pairs.

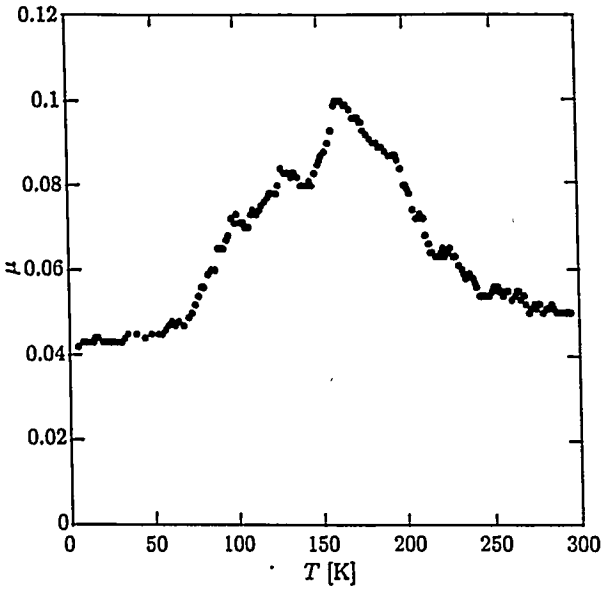


Fig. 9 μ *vs* T at $v = 10^{-5}$ m/s for diamond/graphite pairs.

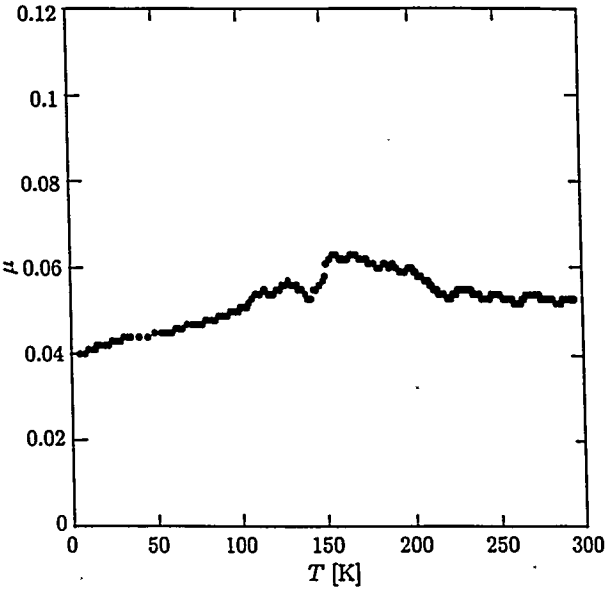


Fig. 10 μ *vs* T at $v = 10^{-5}$ m/s for graphite/graphite pairs.

CONCLUSIONS

We have presented a selection of cryotribology data for diamond (natural and CVD films) and graphite, all in terms of μ *vs* v plots and μ *vs* T plots. Based on our data, both presented here and not presented owing to lack of space, we may make the following conclusions. 1) For natural diamond, graphite, and stainless steel pins *vs* CVD-diamond coated disks, μ's are essentially velocity-independent. For diamond pins sliding on stainless steel disks, μ is roughly velocity-independent at 4.2 and 77 K, but increases with v at 293 K; apparently at temperatures above 77 K and certainly at 293 K, stainless steel is a "soft" material in comparison with diamond. 2.) Although for natural diamond, graphite, and stainless steel pins *vs* CVD-diamond coated disks, μ's are fairly temperature-independent, there are subtle variations. In the diamond/CVD-diamond pairs, μ is at a maximum at 4.2 K and decreases with temperature. In the graphite/CVD-diamond pairs there is a variation in μ with T in the mid temperature range, due perhaps to wear particles generated by the graphite pins. 3.) The materials sliding against graphite disks show a peak frictional coefficient value within the temperature span 100 – 200 K. As is the case with diamond/stainless steel pairs, there is a definite correlation between friction and material hardness: μ is inversely proportional to material hardness. Thus μ is higher at higher temperatures, because hardness generally decreases with temperature.

ACKNOWLEDGEMENT

The work presented here was performed under the auspices of the U.S. Department of Energy, Office of Basic Sciences and is based in part on a thesis submitted by AFA in partial fulfillment of the requirements for the degree of Master of Science in Mechanical Engineering at MIT, February 1996. The authors thank Drs. Koji Kobashi and Takeshi Tachibana of Kobe Steel Ltd. for supply of CVD-diamond coated disks.

REFERENCES

1. Yukikazu Iwasa, *Case Studies in Superconducting Magnets*, (Plenum Press, New York, 1994).
2. F.P. Bowden, and D. Tabor, *Friction and Lubrication*, (Latimer Trend & Co., 1967).
3. M.N. Gardos, and B.L. Soriano, "The effect of environment on the tribological properties of polycrystalline diamond films," *J. Mater. Res.* **5**, No. 11 (1990).
4. B. Samuels, and J. Wilks, "The friction of diamond sliding on diamond," *J. Mater. Science* **23**, 2846 (1988).
5. M. Kohzaki, K. Higuchi, S. Noda, and K. Uchida, "Tribological characteristics of polycrystalline diamond films produced by chemical vapor deposition," *J. Mater. Res.* **7**, (1992).
6. K. Jia, Y.Q. Li, T.E. Fischer, and B. Gallois, "Tribology of diamond-like carbon sliding against itself, silicon nitride, and steel," *J. Mater. Res.* **10**, No. 6 (1995).
7. F.P. Bowden, and A.E. Hanwell, *Proc. Roy. Soc.* **A 295**, 233 (1966).
8. P.B. Mody, T. Chou, and K. Friedrich, "Effect of testing conditions and microstructure on the sliding wear of graphite fiber/PEEK matrix composites," *J. Mater. Science* **23**, 4319 (1988).
9. Y. Yamada, S. Ueda, and K. Tanaka, "Effects of temperature on the friction and wear of poly(amide-imide)-based composites," *J. Japan Soc. Lub. Eng.* **33**, 458-466 (1988).
10. A.M. Petlyuk, L.N. Sentyurikhina, O.V. Lazovskaya, and T.P. Yukhno, "Temperature dependence of wear resistance of solid lubricant coatings containing MoS_2 and graphite," *Sov. J. Fric. and Wear*, **8**, 132 (1987).
11. Katsuhiro Nishiyama, Keisuke Taguchi, Tsuneo Ohdaira, and Abe, "Graphite solid-state lubricant generating low coefficient of friction at high temperatures," *J. Jap. Soc. Powder and Powder Metallurgy* **41**, 613 (1994).
12. A. Cameron, *The Principles of Lubrication*, (Longmans Green & Co., 1966).
13. Y. Iwasa, R. Kensley, and J.E.C. Williams, "Frictional properties of metal-insulator interfaces," *IEEE Trans. Mag.* **MAG-15**, 36 (1979).
14. E. Rabinowicz, *Friction and Wear of Materials*, (John Wiley, New York, 1965).
15. R.B. Scott, *Cryogenic Engineering* (Reprinted by Met-Chem Research, Boulder, CO.)
16. D.A. Wigley, *Materials for Low-temperature Use*, (Oxford University Press, 1978).
17. *Cryogenic Materials Data Handbook*, (Air Force Materials Laboratory, Wright-Patterson Air Force Base, OH, 1970).
18. A.R. Von Hippel, *Molecular Science and Molecular Engineering*, (MIT Press and John Wiley & Sons, New York, 1959).
19. *Physical Properties of Diamond*, Ed. R. Berman, (Clarendon Press, Oxford, 1965).
20. E. Rabinowicz, "Autocorrelation analysis of the sliding process," *J. Appl. Phys.* **27**, 131 (1956).
21. E. Rabinowicz, "Determination of compatibility of metals through static friction test," *ASLE Transactions* **14**, 198 (1971).
22. F.P. Bowden and G.W. Rowe, "The friction and mechanical properties of solid krypton," *Proc. Roy. Soc.* **A.228**, 1 (1955).
23. M.B. Peterson, and S. Ramalingam, "Coatings for tribological applications," *Fundamentals of Friction and Wear of Materials* (ASM, Metals Park, OH) 119 (1981).

THE TEMPORAL NATURE OF FORCES ACTING ON METAL DROPS IN GAS METAL ARC WELDING

Lawrence A. Jones Thomas W. Eagar Jeffrey H. Lang

Massachusetts Institute of Technology
Cambridge, MA, 02139 USA

ABSTRACT

At moderate and high welding currents, the most important forces in gas metal arc welding acting on the molten electrode are magnetic forces arising from the interaction between the welding current and its own magnetic field. These forces drive the dynamic evolution of the drop and also depend on the instantaneous shape of the drop. In this paper, experimentally observed manifestations of magnetic forces are shown, and a technique for approximating the temporal evolution of these forces from experimentally measured drops shapes is reported. The technique provides quantitative data illustrating the large increase in the magnetic forces as a drop detaches from the electrode.

The temporal evolution of the metal drop geometry during drop detachment in gas metal arc welding (GMAW) has an important effect on the process because the forces acting on the drop depend on its shape and they change dramatically over the course of drop detachment. If the geometric evolution of the drop is ignored, it is impossible to quantitatively explain such phenomena as the substantial initial velocity a drop has at the moment it detaches from the electrode. It is also difficult to quantitatively explain the effectiveness of current pulsing. As another example, if the geometric evolution of the drop is ignored, it is difficult to quantitatively explain why some welding conditions result in the stable, axisymmetric detachment of drops and other conditions do not.

Over a wide range of conditions, the effects of magnetic forces may be seen acting on drops detaching from a GMAW electrode. A drop detaching at 260 A and 29 V is shown in Figure 1.* For a 1/16-inch diameter ER70S-3 electrode in Ar–2%O_2, this current is in the upper end of the globular transfer region. The distinctly flattened (oblate) shape of the drop is a result of the magnetic forces acting on the drop. It will be seen below that when a drop of approximately the same size detaches under very low-current conditions, the drop is largely spherical and is not flattened (Figure 5). Also, although it is not apparent

*The images in Figure 1 and subsequent figures were obtained using the optical technique described in [1]. An arc is present in all of the images, but it is rendered virtually invisible by the optical technique.

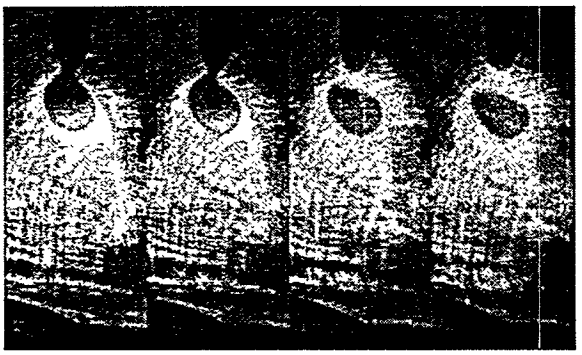


Figure 1: Drop detachment with constant current 260 A and 29 V electrode positive. The electrode is 1/16-inch diameter ER70S-3 in Ar–2%O_2. Note the vertically flattened (oblate) shape of the drop during and after detachment.

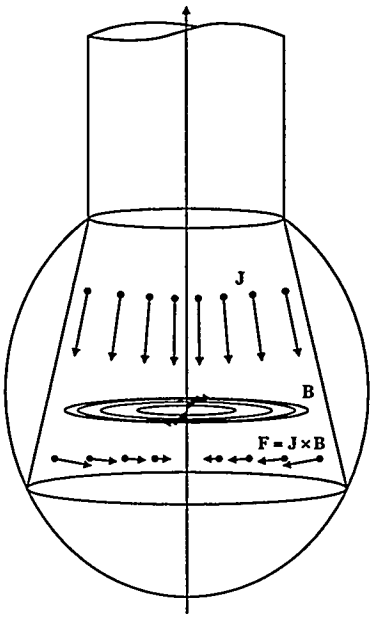


Figure 2: Magnetic forces in a welding drop arise from the interaction between the welding current and its own magnetic field.

from the images in Figure 1, the drops accelerate off the end of the electrode at a rate substantially greater than the acceleration of gravity. This excess acceleration is due to the sharp increase of magnetic forces acting while the drop is detaching and due to the force of plasma flow while the detached drop is in flight.

The magnetic forces arise from the interaction of the welding current with its own magnetic field as illustrated in Figure 2. If the current diverges in the drop (as shown in Figure 2), then downward forces act on the fluid in the drop, and if the current converges in the drop, then upward forces act on the fluid in the drop. Unlike gravity, which acts uniformly in the vertical direction on the fluid (assuming the density of the fluid is spatially uniform) and is an irrotational force, the magnetic forces do not act uniformly and there is a rotational component of force acting on the fluid.

At higher currents (~400–470 A), the heat of the arc causes molten metal to stream off of the electrode forming a column of liquid which then breaks up into drops. The magnetic forces due to the current flowing through this liquid aid the breakup of this column into drops. Even without magnetic forces acting, the breakup of a liquid column will occur due to mechanical disturbances in the fluid flow which cause disturbances in the curvature of the surface of the column. If current is flowing in the column, the magnetic forces aid the breakup of the column. At very high currents, the magnetic forces are apparent in the appearance of rotating streams of metal as shown in Figure 3 for 480 A and 35 V. At such high currents the magnetic forces are significant compared to the inertial forces in the column and slight asymmetries in the column cause asymmetric radial magnetic forces which move the column away from its straight line of flow. It can be shown that such a bend in the current path also results in azimuthal components of the magnetic forces. The combination of asymmetric radial forces and azimuthal forces results in the spiraling mo-

tion of the column as seen in Figure 3. This motion is known as a kink instability [2].

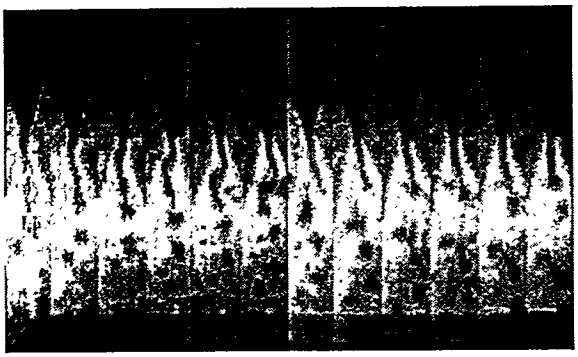


Figure 3: Drop detachment with constant current 480 A, 35 V electrode positive. The electrode is 1/16-inch diameter ER70S-3 in Ar–2%O_2. Note the spiraling detachment of drops from a column of liquid metal.

By pulsing the welding current, magnetic forces may be used to detach drops from the electrode. The shape of the pulse, its magnitude, duration, repetition rate (frequency), and the current level between pulses are all parameters in this process. An incorrect choice of parameters may result in an unstable arc, a reduced metal deposition rate, incorrect base-plate heating, and/or spatter. Figure 4 shows the results of applying a pulse having too great an amplitude. The magnetic forces overwhelm the surface tension forces and the detachment process proceeds asymmetrically. A thin, filamentous neck is violently snapped to the side resulting in fine spatter. Due to the violence of the detachment, the main drop sometimes breaks apart resulting in coarse spatter. If the current is not reduced immediately after the detachment of a drop, the heat of the arc will generate a subsequent small drop (a droplet) and the magnetic forces will detach it asymmetrically which also generates coarse spatter. Such a droplet is seen forming in the last image in Figure 4. In this particular case,

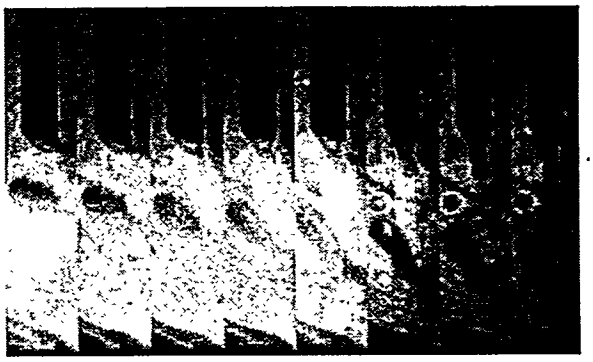


Figure 4: Drop detachment with pulsed current 550/40 A, 6 Hz, 2% duty cycle, and 18 V electrode positive. The electrode is 1/16-inch diameter ER70S-3 in Ar–2%O_2. Note the whipping of the drop neck, the severe distortion of the drop, and the presence of a small droplet on the electrode at the end of the sequence which often detaches and flies to the side as spatter.

the current is reduced before the droplet detaches from the electrode and the droplet is pulled back onto the electrode by surface tension.

Under certain pulsing conditions, drops will detach from the electrode with minimal distortion and no droplets. Such a case is shown in Figure 5. In this case, the current pulse imparts enough momentum to the drop to cause detachment, but ends well before the drop detaches, thereby avoiding the formation of droplets. At the time of detachment, the welding current is 40 A and the drop is almost spherical. The lack of distortion of the drop is in marked contrast to the distortion of the drop seen in Figure 1 at 260 A constant current. When the current is very low at drop detachment, the magnetic forces do not act violently on the neck of the drop as seen in Figure 4, and since the drop is largely undistorted as it detaches, the probability is low that it will break apart.

If a current pulse lower than 290 A

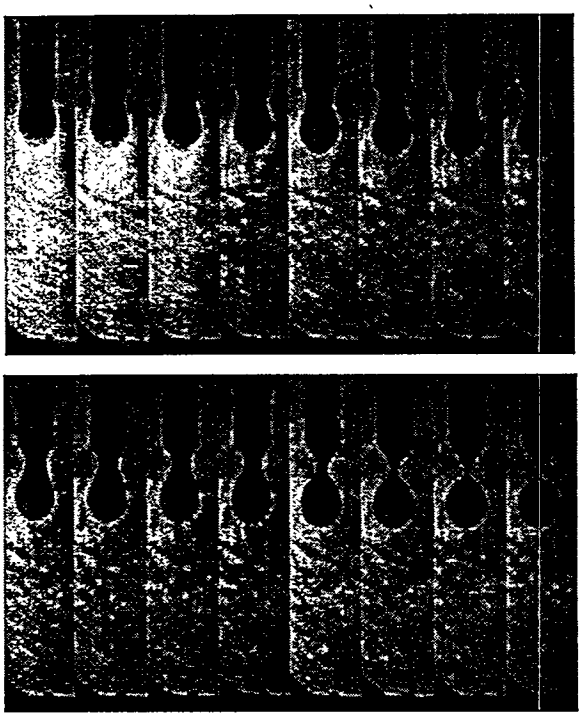


Figure 5: Drop detachment with pulsed current 290/40 A, 5 Hz, 2% duty cycle, and 18 V electrode positive. The electrode is 1/16-inch diameter ER70S-3 in Ar–2%O_2. Note the symmetric detachment of a drop and the lack of drop distortion.

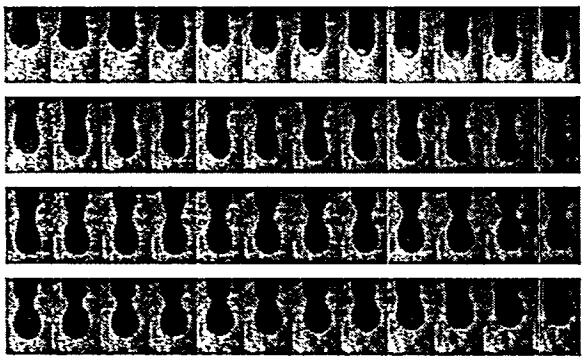


Figure 6: Drop response with pulsed current 280/40 A, 5 Hz, 2% duty cycle, and 18 V electrode positive. The electrode is 1/16-inch diameter ER70S-3 in Ar–2%O_2. The current pulse is insufficient to detach the drop. Subsequent current pulses will also fail to detach the growing drop.

is used, the momentum impulse supplied by the magnetic forces will not be great enough to overcome surface tension and the drop will not detach, as shown in Figure 6. Without the assistance of magnetic forces, the surface tension forces, which act to restore equilibrium, dominate and overcome the inertial force of the drop. During the next several current pulses, the mass of the drop will be greater and the magnetic forces will be unable to impart enough momentum to the drop to make it detach. The drop will continue to grow until gravity, aided by the disturbances caused by current pulses, overcomes the surface tension holding the drop on the electrode.

The magnetic diffusion time and the magnetic Reynolds number in a drop of molten steel on the end of a GMAW electrode are both very small. The very short magnetic diffusion time indicates that on the time-scale of drop motion, the diffusion of the magnetic field throughout the drop is essentially instantaneous. The very small magnetic Reynolds number indicates that the magnetic diffusion process is much faster than fluid convection in the drop. Therefore, in a gas metal arc welding electrode the magnetic field is unaffected by the fluid velocity in the drop. Rather, the distribution of the magnetic field is determined by the instantaneous geometry of the current path in the drop and, assuming uniform fluid conductivity, the current path is determined by factors other than the fluid velocity.

In [3], the magnetic stress tensor is used to calculate the total vertically-directed magnetic force acting on a generalized, axisymmetric drop shape. The motion of the fluid interacts with the magnetic field only in that it affects the resulting shape of the drop and hence the shape of the current path. Therefore, if the shape of the current

path is known at each instant, the magnetic stress tensor may be used to calculate the total vertically-directed magnetic force on the drop (the sum of the irrotational and rotational parts).

As a drop attempts to detach from the solid electrode, a neck forms. The current density in the drop neck increases and the divergence of the current increases. Both phenomena cause the magnetic forces acting on the drop to increase and result in a measurable acceleration of the drop upon detachment. The time during which a neck initiates and collapses is short compared to the total drop growth time but it is during this brief time that the magnetic forces are most important.

During drop necking, the lower part of the drop is well modeled by a truncated ellipsoid and the upper part is well modeled by a volume formed by rotating a third-order polynomial about the vertical axis, as illustrated in Figure 7. Third-order polynomial volumes were chosen to model the neck because these shapes are completely defined by the geometrical boundary conditions at the top of the truncated ellipsoid and the bottom of the electrode.

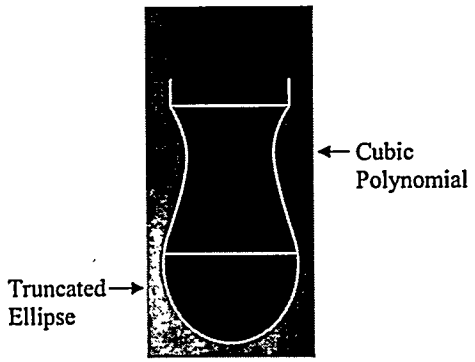


Figure 7: Experimentally-observed necking drop shape modeled with a truncated ellipsoid and a polynomial volume.

A truncated ellipsoid is shown in Figure 8 along with the portion of a polynomial volume below its waist (its narrowest point). In [3], the total vertically-directed magnetic force on a generalized shape is calculated to be

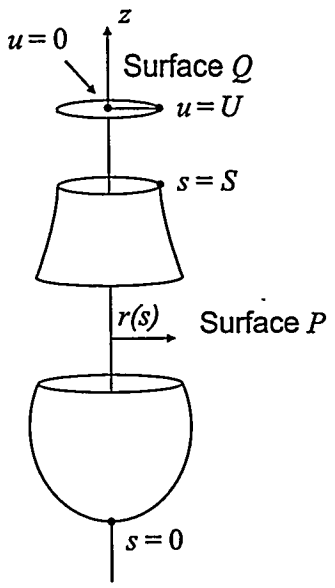


Figure 8: Stress tensor surfaces.

$$f_z = -\frac{\mu_o I^2}{16\pi} + \mu_o \pi \int_0^S \left[\int_0^s j(s')r(s')ds'\right]^2 \frac{dr}{ds'}\frac{ds'}{r(s')}, \quad (1)$$

where I is the portion of the welding current emerging from the drop and $j(s)$ is the density of this current along the surface s of the drop. If there is no drop neck, if the truncated ellipsoid is a truncated sphere, and if the current density $j(s)$ is taken to be constant,* then both integrals in Eq. 1

*No experimental measurements are available in the literature about the distribution of the current density on a GMAW electrode due to the difficulty of making such measurements in the harsh environment of the arc next to the free surface of a drop. However, observations of drops detaching from the electrode empirically suggest that a constant $j(s)$ may be a reasonable approximation [4].

may be solved in closed form. The resulting expression is commonly used to compute the total vertically-directed magnetic force acting on a welding drop:

$$f_z = \frac{\mu_o I^2}{4\pi}\left[\frac{1}{4} - \ln\left(\frac{a\sin\Phi}{r_e}\right) + \frac{1}{1-\cos\Phi} + \frac{2}{(1-\cos\Phi)^2}\ln\left(\frac{2}{1+\cos\Phi}\right)\right], \quad (2)$$

where a is the radius of the sphere, r_e is the radius of the electrode, and Φ is the angle describing the portion of the drop over which current is emerging ($\Phi = 0 \Rightarrow$ no current emerges from the drop, and $\Phi = \pi/2 \Rightarrow$ current emerges from the lower hemisphere of the drop).

While a drop is necking—a time when the magnetic forces increase dramatically—truncated-ellipsoid and polynomial-volume combinations are required to model the shape of the drop, as shown in Figure 7. The integrals in Eq. 1 cannot be solved analytically even for these simple approximations of the fluid surface of a necking drop. Over the surface of the truncated ellipsoid, the inner, squared integral may be solved analytically, but not the outer integral. Over the surface of the polynomial volume, the inner, squared integral cannot be solved analytically (and therefore the outer integral also cannot be solved analytically).

Truncated ellipsoids and polynomial volumes were fit to drop profiles observed during current pulsing similar to those shown in Figure 5. The instantaneous vertically-directed magnetic forces acting on these fitted shapes were then computed numerically using Eq. 1. Constant current density $j(s)$ was assumed to cover the entire surface of the shapes up to the waist of the polynomial volume. Experimentally-measured current pulses and the corresponding computed vertically-directed magnetic forces are shown in Figure 9.

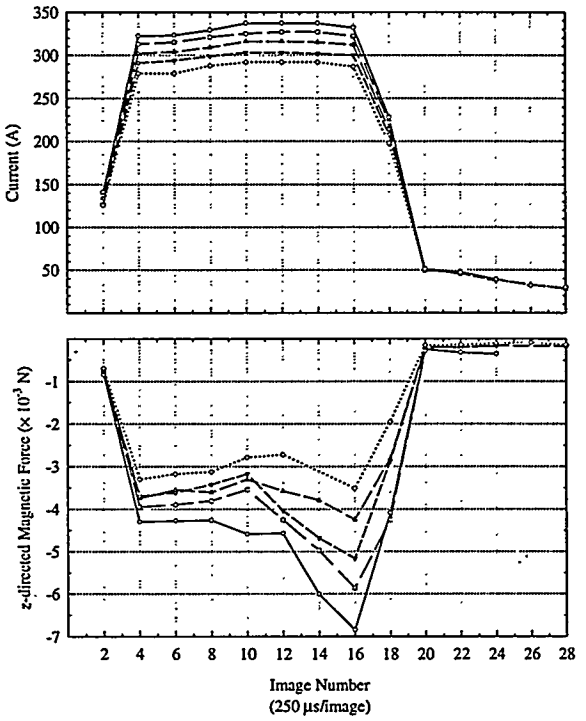


Figure 9: Total vertically-directed magnetic force during current pulsing with the Ar–2%O_2 current emission model.

The forces in Figure 9b show that initially as a drop elongates in response to the downward (negative) magnetic force, the magnitude of the force decreases slightly. This decrease is because in an elongated (prolate) drop the current diverges less. Once a neck begins to form, however, the magnitude of the magnetic force increases rapidly. This increase is because the narrowing of the neck results in greater subsequent current divergence. This observation is not surprising or new since it has long been known that magnetic forces act to "pinch" drops off the end of an electrode. The results shown in Figure 9, however, represent the first time the temporal evolution of the magnetic force has been computed using close approximations of experimentally measured drop shapes.

By using shape approximations as in Figure 7, a simple, dynamic model of drop detachment may be developed, as illustrated in Figure 10. Such a model is developed in [4] and is compared with extensive measurements of drop detachment obtained from high-speed images of gas metal arc welding.

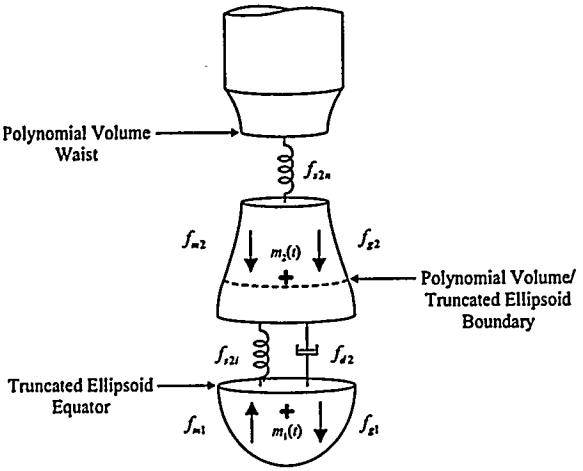


Figure 10: Illustration of the dynamic model developed in [4] using truncated-ellipsoid/polynomial-volume shapes.

Acknowledgement

This work was supported by the United States Department of Energy, Office of Basic Energy Sciences.

References

[1] C. D. ALLEMAND et al., "A Method of Filming Metal Transfer in Welding Arcs," *Welding J.*, **64**(1), 45–47 (1985).

[2] A. DATTNER, "Current-Induced Instabilities of a Mercury Jet," *Arkiv för Fysik*, **21**(7), 71–80 (1962).

[3] J. C. AMSON, "Lorentz Force in the Molten Tip of an Arc Electrode," *British J. of Appl. Phys.*, **16**, 1169–1179 (1965).

[4] L. A. JONES, *Dynamic Electrode Forces in Gas Metal Arc Welding*, PhD thesis, Massachusetts Institute of Technology, 1996.

PARTIAL CONTROL OF COMPLEX SYSTEMS WITH APPLICATION TO THE FLUIDIZED CATALYTIC CRACKER

Irven H. Rinard and Reuel Shinnar

Department of Chemical Engineering
City College of CUNY

ABSTRACT

The research deals with the control of complex nonlinear system with a limited number of manipulated variables. In many chemical processes the number of variables that make up the specifications and constraints exceeds the number of manipulated variables available. Furthermore, model information is limited. The goal of our work is to study the design of the control system and the conditions required to achieve adequate control for such cases. A Fluid Catalytic Cracker was chosen to illustrate and test our approach. This paper presents a short overview and summary of our approach and results.

INTRODUCTION

The fluidized catalytic cracker (FCC) is one of the most important processing systems in oil refineries. It is a complex system to operate and control, one which is not fully understood, and one on which the petroleum industry has spent considerable time and research effort over the past sixty years. The heart of the problem is that there are many more process variables that one needs to control than there are manipulatable variables with which to do so. This is characteristic of most complex systems such as oil refineries and chemical plants as well as social systems, even the economy. The other salient feature of such systems is that their behavior is not well understood. This is true of the FCC; it is certainly true of the economy.

The goal of the present work is to develop a practical theory of partial control that can be applied to complex systems. However, to insure that it is a relevant theory, its development has been closely tied to a significant real-world example, namely, the FCC. Further, this work has been done in close collaboration with Mobil, one of the major petroleum companies. Our research in this area has reached the state where its results can be implemented and utilized by industries in many areas. It

already has led to improvements in the control of FCC's and these have been adopted by industry. It is the goal of this paper to give an overall picture of the theory and its application to control of FCC's. For details the reader is referred to the papers that have already been published under this grant [1-4].

CONCEPTS OF PARTIAL CONTROL

It is not obvious that one can control a complex nonlinear system by only a limited number of manipulated variables, particularly when the behavior of that system is not completely understood. Based on our theoretical results and the study of actual complex processes, we have developed an approach and a theoretical framework as how to choose the manipulated and measured variables for a control system in order to get the best results. The control ideas underlying this research were developed and published by Professor Shinnar in two papers [5,6] and were based on his work in process control in both academia and industry.

In general, in the design of a control system, one first identifies the set of variables that need to be controlled, namely, those that are related to product specifications and process operating constraints. Denote these by Y_{pj} and their number by N_c. Next, one identifies a set of suitable manipulated variables with which to effect this control. Denote these by U_m and their number by N_m. If $N_m = N_c$, then all of the controlled variables can be controlled, at least in principle, to exact set points. We call this exact control. It is the type of control to which the vast majority of the control literature and textbooks are devoted. A different situation arises quite often in the control of large, complex systems, one where $N_m < N_c$. Now, not all the Y_{pj} can be controlled at exact set points. Instead, we must settle for something less strict, namely, that these variables are kept within an acceptable operating space, i.e., $Y_{pjmin} \le Y_{pj} \le Y_{pjmax}$. We call this partial control.

The basic idea of partial control is that we choose a subset of the controlled variables, Y_{cd}, and the manipulated variables, U_{cd}, which results in a square control matrix. This primary control matrix will be used for exact control of the chosen subset of controlled variables. Which and how many variables are chosen depends upon a number of factors that are discussed below. For many systems an appropriate choice of the primary matrix will allow us to maintain all the Y_{pj} within the acceptable operating space by adjusting the set points of the variables in the control matrix (Fig. 1). Choice of this primary control matrix is critical to the success of the overall control. If one has a very good model, one would feed the measurements of Y_{pj} directly into a model based on direct computer control to adjust all U_j. This does not allow integral control and requires a much better model than we normally have.

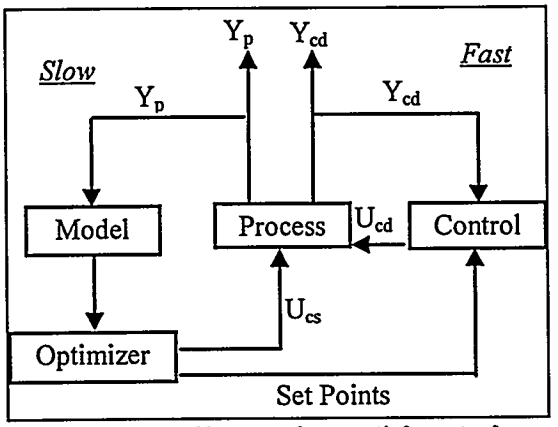


Figure 1: Block diagram for partial control

There are several considerations in the choice of the primary control structure. We would like the primary matrix to be as small as possible. To achieve this, our choice Y_{cd} and U_{cd} must satisfy the following criteria:

A. Dominance

In partial control one chooses a set of measured variables Y_{cd} that are kept by exact control at chosen set points. If one or more of the output variables of concern Y_{pj} moves out of the desired range,

one adjusts the set points of Y_{cd} to keep the operation of the process acceptable, both economically and in terms of product specifications.

It is very important how one chooses this set of measured variables Y_{cd}, some of which may be of no direct interest by themselves. The most important concern is dominance which means that changing the set point of a measured variable has a strong impact on many or all the process variables of interest. To do so requires that we have available manipulated variables which have a strong impact on Y_{pj} and Y_{cd}. A good example is chemical reactor temperature which affects all the kinetic processes just as climatic temperature affects all biological processes. The difference is that in a reactor we can control it. Change the temperature and we change almost everything else: the production rate, the yield or product distribution, even the stability of the operating point. Similarly, interest rates are a dominant variable in the economy. Raise interest rates and many, many other things change. There are many other examples.

B. Sufficiency

It is overly optimistic to believe that a single dominant control loop will lead to acceptable control of the many important variables that arise in complex systems. This leads to the concept of sufficiency. How many and what dominant variables must be controlled so that the system is stable and that all the other variables are maintained within specified limits? What constitutes a sufficient set of such variables?

It is important to ensure that a process has enough manipulated variables with a wide enough range of variability to insure that the process can deliver product that meets its quality specs. The control must simultaneously ensure that process conditions do not violate operating constraints, such as metallurgical limits, flooding or entrainment flow velocities, etc., which we include in Y_p. The control should also achieve this goal in economically acceptable ways. Since the availability of the manipulated variables is determined during the design phase, one needs to understand how the design impacts control. Another important consideration is that of the product specifications. These are determined externally by market requirements but set many of the constraints on Y_{pj}. This will have a major impact on the design of both the unit itself as well as the control system. One has to understand how the design interacts with the specifications in order to come up with an economically acceptable compromise. Better partial controllability is achieved at the expense of adding more manipulated variables but these must be chosen to provide the widest possible range of operating conditions at acceptable cost.

Our research deals with this important problem and we have developed these criteria for the FCC. These results are important in developing strategies for better design procedures, and are a good example demonstrating how basic research in control can lead to better design of chemical processes.

C. Modellability

The choice of dominant variables is itself constrained by the number of manipulated variables available for control. These must satisfy several criteria, the primary ones being that they have a significant gain with respect to the dominant variables to be controlled and that they affect these variables on a suitable time scale.

This leads to the very important question of modellability. To indirectly control a complex system where anywhere from a dozen to a hundred or more variables are of interest and to do this by direct control of a small set, say two to six, dominant variables requires a model. This model must satisfy two criteria:

- How well can the process variables Y_{pj} be related to the dominant variables, particularly in steady-state? This is the major determinant of how well we can do with respect to constraint management.
- How well can the dominant variables be related to the manipulated variables? This is a major

determinant of how well we can stabilize and control the primary control structure. This has to be addressed from system to system; examples will be discussed later. In general full scale reactors have more complex and less uniform flows and hopefully the variables sensitive to these variations are not dominant for Y_p. This is essential not only for control but also for safe scale up.

D. Time Scale of Response

The primary control structure must be chosen so that it responds on a time scale commensurate with the time scale on which the process variables Y_{pj} respond to disturbances to the system. This time scales are determined by the design. The designer faces some compromises for complex processes. In adiabatic processes such as the FCC one introduces intentionally sufficient thermal inertia to allow the operator to intervene in order to prevent catastrophic instabilities and correct mistakes. In the FCC the holdup in the regenerator provides this thermal inertia. The reactor has a much faster response which is beneficial. Fast response in the process industries is often less important than in aerospace.

E. Nonlinearities and Stability

In complex, nonlinear systems assuring stability of the control system is a major concern. Such systems often exhibit multiple steady-states, some of which are open-loop stable, others which are not. Input multiplicities and regions of zero open-loop gain further complicate the task of choosing a control structure. The pertinent nonlinear features of the FCC were analyzed in [2] as part of this program and have critical impact on the control strategies.

APPLICATION TO THE FCC

The fluidized catalytic cracker (FCC) was chosen as an example of the kinds of complex systems encountered in the petroleum and chemical industries which are operated subject to partial control. As shown in Fig. 2, the FCC consists of two principal units. In the first, the reactor, hot catalyst is contacted with oil in a fluidized bed. It supplies heat whereby the oil is first vaporized, then cracked to more desirable products such as gasoline and heating oil. The catalyst is cooled down and simultaneously deactivated by the formation of coke which is a byproduct of the cracking reactions. The catalyst is reactivated and reheated by circulating it to the second unit, the regenerator, where it is contacted with air to combust the coke. The energy released in the regenerator is recycled back to the reactor by the circulating catalyst. One important condition for the operation of the FCC is that the heat generated by combusting all the coke generated equals the heat absorbed in heating and cracking the oil feed.

There are many variables in the FCC that need to be controlled to a greater or lesser extent. These include the fractional conversion of the oil feed to products, the gasoline yield and its composition, the N_{ox} and CO levels of the flue gas leaving the regenerator, and

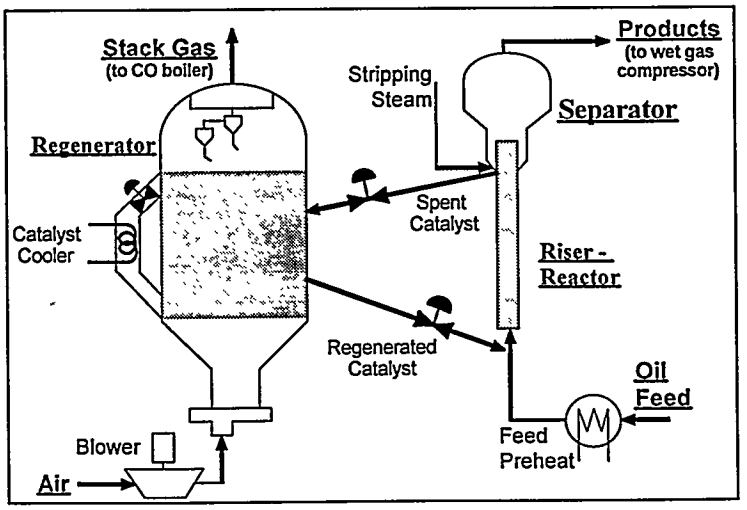


Figure 2: Schematic Diagram of FCC

olefin content of the product gas which is used as a feed for alkylation, to name just a few. There are but a few variables available with which to exert this control on the appropriate time scale, primarily the feed rate of air to the regenerator, F_{air}, the catalyst circulation rate between the two units, F_{cat}, and the temperature of the feed, T_{feed}

Some units have a catalyst cooler to allow processing of heavy feedstocks, for coke make in the FCC is tied to conversion. For highly coking feeds, excess heat has to be removed. Such a cooler decouples the reactor from the regenerator and changes the control.

In addition to U_{cd}, modern FCCs have a range of slowly manipulated variables available, such as catalyst activity and other properties which are controlled by adding and withdrawing catalysts as well as additives that impact on Y_{pj}. [7] reviews how design changes in FCC developed and how they impact on FCC operation.

There is another aspect of design that changes the control of the FCC. Historically, most units operated at regenerator temperatures of 1180 to 1250°F with a CO_2/CO ratio between 1.2 to 1.5. This reduces the amount of heat evolved, allowing higher conversion. The CO is combusted in a CO boiler, but old units often simply emitted it to the atmosphere. Environmental laws forced operations of such units to change the operation to complete CO combustion and combustion promoters allowed to practically eliminate CO emission. Today many units operate in full CO combustion. This changes the control as can be seen from Fig. 3 where we plot T_{rgn} (regenerator bed temperature) versus air rate at constant T_{ris} (riser top temperature). For a given air rate and T_{ris} there are two steady states, one in partial and the other in complete CO combustion.

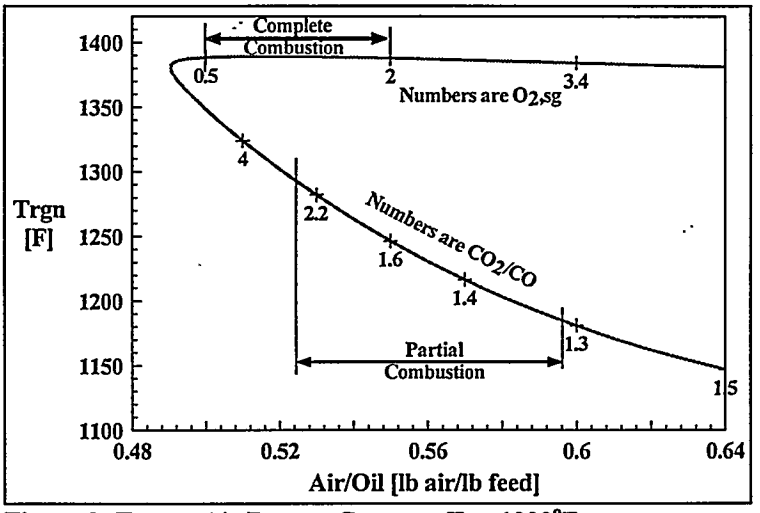


Figure 3: T_{rgn} vs. Air Rate at Constant T_{ris}=1000°F

Fig. 4 gives some options for control matrices and loops that have been used in the partial CO combustion regime. For complete combustion the only option is to use [F_{air} , $O_{2,sg}$] which is essential to keep the unit in complete CO combustion. Excess oxygen varies between 0.5 and 2.0% but aside from keeping the unit stable has no impact on either T_{rgn} or Y_{pj}.

If we look at Fig. 4, the measured variables entering the matrices in partial CO combustion are T_{ris}, T_{rgn}, T_{sg} (stack gas temperature in the regenerator) and ΔT (= T_{sg} - T_{rgn}) which is proportional to the excess oxygen entering the cyclone. While ΔT and T_{sg} have been used in the past and have been preferentially promoted in the control literature [8], they have two disadvantages over T_{rgn} and T_{ris}. They are not dominant and very model

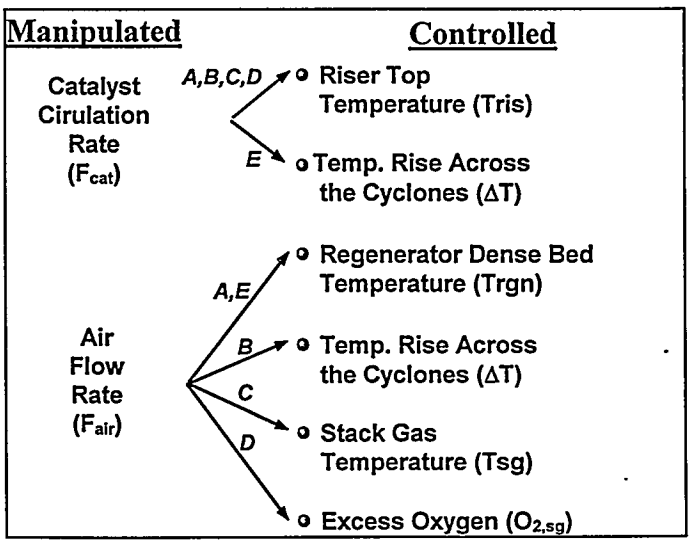


Figure 4: Possible Control Structures for the FCC

sensitive, and therefore not predictable from laboratory experiments. In the FCC the nature of the reactions are such that residence time has a small impact. However, reactor temperature impacts strongly on all reactions. Therefore, reactor outlet temperature T_{ris} as well as inlet temperature T_{mix} strongly impact on Y_{pj}. T_{mix} is a strong function of T_{rgn}. In the regenerator in partial combustion T_{rgn} determines the ratio of CO_2/CO as well as the rate of coke combustion and therefore C_{rgc} (coke on regenerated catalyst). T_{rgn} is therefore dominant. In full CO combustion C_{rgc} is practically zero and CO_2/CO close to infinity. Therefore T_{rgn} is not dominant as long as it is high enough.

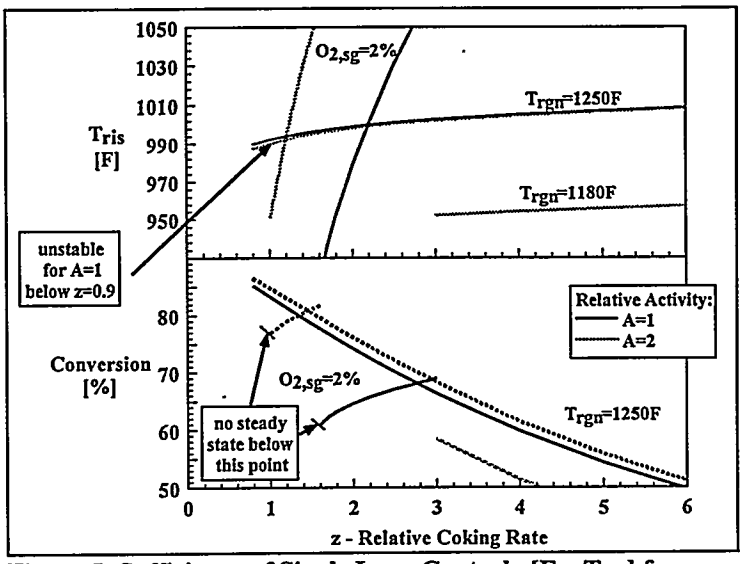


Figure 5: Sufficiency of Single Loop Control. $[F_{air},T_{rgn}]$ for partial combustion; $[F_{air},O_{2,sg}]$ for complete combustion.

The dominant variables are therefore T_{ris} and T_{rgn} in the reactor and T_{rgn} in the regenerator. They are also the only ones of the set that are modellable.

While a discussion of the variables requires far more space than we have, we can use this to illustrate some of the ideas. For example, in partial combustion it is perfectly possible to maintain a stable operation over a wide range of catalyst properties by one single control loop $[F_{air}, T_{rgn}]$. This can be seen from Fig. 5 where we plot T_{ris} vs. relative coking rate (z), as T_{ris} has narrow limits. It has to be maintained between 900 and 1010°F. Below 850°F part of the feed will not vaporize or crack and at about 1030°F thermal cracking becomes the predominant reaction. We note that the impact of F_{air} on T_{rgn} is small, so varying T_{rgn} allows one to keep T_{ris} within the permissible range. At low relative coking rate, z, the unit becomes unstable, but a single PI controller can stabilize it. At very low z it loses the steady state and the only option is a more active catalyst or one that produces more coke.

In full combustion a single control loop $[F_{air}, O_{2,sg}]$ does not allow stable operation. We note that small changes in z will change T_{ris} such that it falls outside the permissible range. The only option is to change catalyst activity. Controlling T_{ris} is here essential. The two loops are equivalent to a single loop in partial combustion.

While a single loop is sufficient to maintain the unit stable in partial combustion (or a 2x2 matrix in complete CO combustion), it is very limiting for controlling Y_{pj}. This is illustrated in Fig. 6 where we show some elements of Y_{pj} for the scheme $[(F_{air}, T_{rgn},), (F_{cat}, T_{ris})]$ for a specific operating point in partial combustion. We note that instead of a line an area of Y_{pj} becomes accessible. Adding slow manipulated variables (U_{csj}) will change the accessible space to a multi-dimensional volume. Similar plots can be made for complete combustion, but here we need three

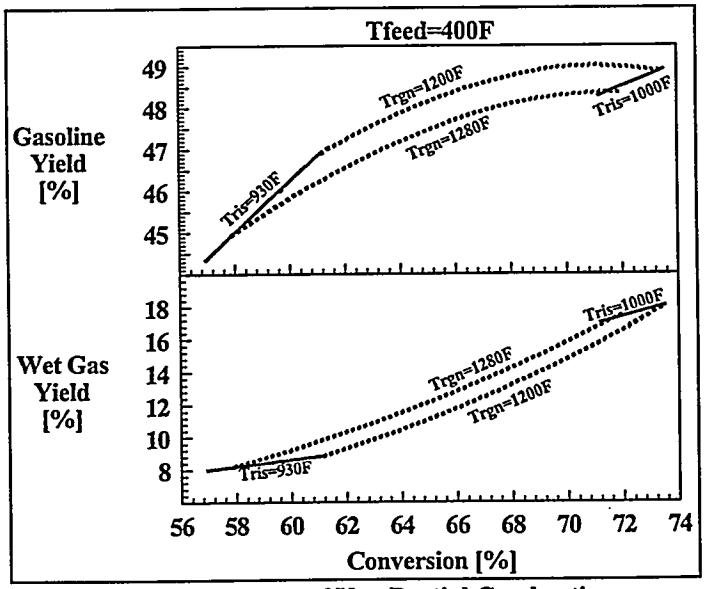


Figure 6: Accessible Area of Yp. Partial Combustion . Control: $[(F_{air},T_{rgn}),(F_{cat}, T_{ris})]$

manipulated variables to increase the accessible space of Y_{pj} from a line to an area.

In complete combustion there is an additional problem. Even three loops do not allow adequate control of T_{rgn} for higher coking rates. The only variable that allows control of T_{rgn} over a wide range of z is catalyst activity. It has a large impact on control in complete combustion whereas its effect is limited in partial combustion. These problems can be completely overcome by the addition of a catalyst cooler which allows T_{rgn} to be adjusted independently both in partial and complete combustion.

Our results show that in partial combustion, in addition to lack of modellability and dominance, ΔT (and therefore T_{sg}) has another problem. If ΔT and T_{ris} are kept constant we can show that increasing z could cause the unit to crash as at constant T_{ris} and ΔT increasing z will decrease T_{rgn} (Fig. 7). If T_{rgn} will drop below 1100°F the unit could lose its uppermost steady state. The same is true for the scheme [(F_{air}, T_{rgn}), (F_{cat}, ΔT)] proposed by [9]. In that case an increase in z will cause T_{ris} to decrease (Fig. 7). If T_{ris} goes below 850°F, the unit also loses its steady state. The exact limits strongly depend on the unit and the catalyst. Thus our results explain the phenomena of crashing due to coking feeds which plagued the industry until the control scheme [(F_{air}, T_{rgn}), (F_{cat}, T_{ris})] was introduced. Crashing occurred when a heavy coking feed was suddenly introduced and did not occur in units with well controlled constant feed composition.

This illustrates an important aspect of our results. Choice of the proper control matrix is in most cases far more important than the multi-variable algorithm itself. Further, while in the FCC linear algorithms are enough for taking care of wide changes in operating conditions and feed perturbations, one must chose the matrix based on nonlinear considerations. The understanding of the nonlinear feature of the system is essential.

One can not even design the linear algorithm itself based solely on the linear model for the matrix. One needs to understand the impact of the specifications. Thus, the matrix [(F_{air}, T_{rgn}), (F_{cat}, T_{ris})] has a negative relative gain array (RGA). In FCC's however one normally designs along the negative diagonal contrary to textbook recommendations. This gives a fast response to T_{ris} and allows tight control of it. As T_{ris} dominates Y_{pj} this advantage outweighs the penalty of operating along the negative diagonal which causes the system to become unstable when one loop opens. Another important aspect of our results that have significance to other systems is that the controllability of a system can totally change if changes occur in the system or operating conditions. This can often occur in reactors due to changes in catalyst properties or changes in Y_{pj} due to market requirements.

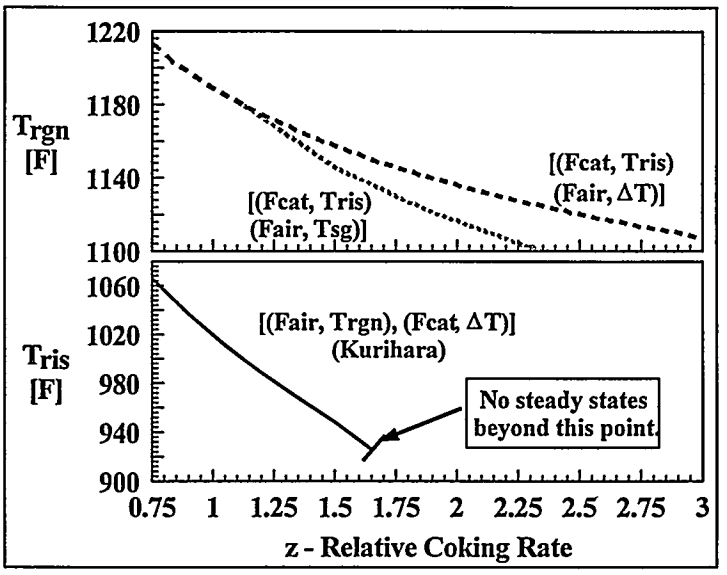


Figure 7: Impact of Coking Rate on Dominant Variables for Different Control Schemes.

Thus our results on the FCC give some important insights as to the design of partial control systems for complex nonlinear systems and raise important challenges for future research. They should also impact on the way we teach process control. The current work focuses on the impact of the design on the control and on the minimum information required to design and operate a successful partial control system. Future work will try to generalize the approach by dealing with other complex systems.

The results are published in a way that indicates that they are useful for a wide range of processes and should have a much wider impact than just FCC control. A paper on how the results impact on economic control was recently submitted for publication. For more details, consult references 1-4.

ACKNOWLEDGMENTS

The authors wish to acknowledge with gratitude the support of the Office of Basic Energy Sciences of the Department of Energy. This research was supported by grant number DE-FG02-91-ER14221.

REFERENCES

1. Arbel, A., Z. Huang, I.H. Rinard, and R. Shinnar, (1995) ,"Dynamics and Control of Fluidized Catalytic Crackers. 1. Modeling of the Current Generation of FCC's," *I&EC Research*, **34**,(4), 1228- 1243.
2. Arbel, A., I.H. Rinard, and R. Shinnar, (1995) , "Dynamics and Control of Fluidized Catalytic Crackers. 2. Multiple Steady States and Instabilities," *I&EC Research*, **34**(9), 3014- 3026.
3. Arbel, A., I.H. Rinard, and R. Shinnar, (1996) , "Dynamics and Control of Fluidized Catalytic Crackers. 3. Designing the Control System: Choice of Manipulated and Measured Variables for Partial Control," *I&EC Research* (accepted for publication).
4. Arbel, A., I.H. Rinard, and R. Shinnar, (1996) , "Dynamics and Control of Fluidized Catalytic Crackers. 4. The Impact of Design on Partial Control," *I&EC Research* (submitted for publication).
5. Shinnar, R. (1981), "Chemical Reactor Modeling for the Purposes of Controller Design," *Chem. Eng. Commun.*, **9**, 73-99.
6. Shinnar, R. (1986), "Impact of Model Uncertainties and Nonlinearities on Modern Controller Design. Present Status and Future Goals," In *Chemical Process Control-CPCIII, Proceedings of the 3rd International Conference on Chemical Process Control, M.* Morari and T.J. McAvoy, Eds., Elsevier, New York, 53-93.
7. Avidan, A. A.; Shinnar, R. Development of Catalytic Cracking Technology. A Lesson in Chemical Reactor Design. *Ind. Eng. Chem. Res.*, **1990**, *29*, 931-942.
8. Hovd, M.; Skogestad, S. Controllability Analysis for the Fluid Catalytic Cracking Process. *Ind. Eng. Chem. Res.*, **1992**, *31*, 1589.
9. Kurihara, H. Optimal Control of Fluid Catalytic Cracking Process. Ph.D. Dissertation, MIT, **1967**

NON-LINEAR SYSTEM IDENTIFICATION IN FLOW-INDUCED VIBRATION

P.D. Spanos and B.A. Zeldin

Department of Mechanical Engineering
Rice University, Houston, Texas 77251

R. Lu

Hudson Engineering Corporation
Houston, Texas 77218

ABSTRACT

The paper introduces a method of identification of non-linear systems encountered in marine engineering applications. The non-linearity is accounted for by a combination of linear subsystems and known zero-memory non-linear transformations; an equivalent linear multi-input-single-output (MISO) system is developed for the identification problem. The unknown transfer functions of the MISO system are identified by assembling a system of linear equations in the frequency domain. This system is solved by performing the Cholesky decomposition of a related matrix. It is shown that the proposed identification method can be interpreted as a "Gram-Schmidt" type of orthogonal decomposition of the input-output quantities of the equivalent MISO system. A numerical example involving the identification of unknown parameters of flow (ocean wave) induced forces on offshore structures elucidates the applicability of the proposed method

INTRODUCTION

Mathematical modeling of wave-induced forces on offshore structures is a complex technical problem. Conventionally, the wave forces on structural elements of size that is small compared to the characteristic wavelength are calculated by Morison's equation [1]. Selecting parameters for the Morison's model which are appropriate for the conditions of a specific site is a delicate procedure. In this context, it is noted that it is rather difficult to account for the effects of waterline, body motion, velocity head, and second-order potential. Thus, the validity of the Morison's force model and its applicability for describing wave forces on flexible offshore structures have been questioned in the literature. In this regard, this paper addresses the issue of determining reliable models of non-linear forces acting on offshore structures by using system identification techniques in conjunction with possibly available experimental data.

Although the identification of linear systems is a well developed subject and is extensively used in engineering practice [2], identification of non-linear structural systems and forces is a less understood scientific area. Several alternative approaches have been proposed in the literature in

conjunction with structural dynamics applications. Non-linear systems can be identified by using the Volterra/Wiener representation theory in which the kernels of the associated integrals are determined from the governing differential equations [3]. In Refs. [4] and [5] the non-linear restoring force was expanded by using a polynomial basis and the coefficients of this expansion were found by minimizing the discrepancy between the observed and simulated responses. By representing the non-linear restoring forces as a set of parallel linear subsystems, the identification of a non-linear system can be developed by adopting a multi-input-single-output (MISO) linear system analysis procedure [6,7]. Using this approach, Vugts and Bouquet [8] verified the accuracy of Morison's equation based on realistic measurement data. Utilizing the representation of a non-linear structural system as a collection of linear subsystems, Spanos and Lu [9] addressed the non-linearity induced by the structure-environment interaction in marine applications; pertinent examples demonstrated the validity of the identification method. Recently, Zeldin and Spanos [10] proposed to use a parametric ARMAX procedure to identify the introduced MISO linear systems. The latter procedure can be especially useful for identification of transfer functions exhibiting sharp peaks; this feature is quite common for structural systems in marine applications.

In this paper, a new procedure is developed for the identification of non-linear models of structural systems for offshore engineering applications. Specifically, the non-linear restoring force is represented as a set of known zero-memory non-linear transformations that are combined with linear subsystems; it reduces a nonlinear system to a linear MISO system [6,7,9]. This method is particularly efficient if preliminary studies of the non-linear system may be used to speculate with regards to the mathematical form of the non-linear force. The identification of the equivalent MISO system is performed by assembling a system of linear equations in the frequency domain with respect to the unknown linear transfer functions of the MISO system. This system is solved by performing the Cholesky decomposition of the associated matrix. The developed identification method can be viewed as a "Gram-Schmidt" type of orthogonal decomposition of the input-output quantities of the equivalent MISO system examined in Refs. [6,7,9]. Pertinent numerical calculations elucidate the implementation of the proposed identification method for marine engineering applications.

MODELING OF NON-LINEAR FORCES

Consider the differential equation

$$m\ddot{q} + c\dot{q} + kq = f(u, \ldots, q, \ldots), \tag{1}$$

where $q(t)$ is the scalar variable representing the structural motion, $u(t)$ is the variable describing the wave kinematics, and f is the random force function which represents the wave induced non-linear force. Further, m, c, and k are the structural mass, damping, and stiffness of the structure, respectively; the dot signifies differentiation with respect to time. To find the unknown structural parameters of Eq. (1), the non-linear force is expanded in terms of base functions of the wave and structural kinematic parameters. In this regard, it is assumed that the system and the excitation nonlinearities can be expressed as a superposition of zero-memory non-linear transformations and linear subsystems. Specifically,

$$f(u, \ldots, q, \ldots) = \sum_{k=1}^{M} H_k\left(\frac{d}{dt}\right) f_k(u, \ldots, q, \ldots), \tag{2}$$

where $H_k(s)$ are polynomials, $f_k(u, \ldots, q, \ldots)$ are non-linear zero-memory transformations, and M is the total number of base functions used in the representation of the non-linear excitation force.

Some preliminary studies may be used to speculate with regards to the mathematical form of the non-linear transformations of Eq. (2). For example, for a Duffing oscillator under random excitation $u(t)$, the non-linear excitation force can be expressed as

$$f(u, \ldots, q, \ldots) = -\varepsilon q^3 + u \tag{3}$$

and the unknown non-linear parameter ε can be determined from measurements by using the proposed identification procedure. For an offshore structure with small structural elements, the non-linear excitation force induced by waves may be represented by Morison's equation. Specifically,

$$f(u, \ldots, q, \ldots) = \alpha_1 |u - \dot{q}|(u - \dot{q}) + \alpha_2 m_a(\dot{u} - \ddot{q}) + m_a \ddot{q}, \tag{4}$$

where α_1 and α_2 are unknown drag and inertia coefficients, respectively. Alternatively, a polynomial expansion of the drag force can be used instead of Morison's equation. That is, the non-linear excitation force can be expressed as

$$f(u, \ldots, q, \ldots) = \sum_{j=0}^{l} \sum_{i=0}^{l-j} \alpha_{ij} u^i \dot{q}^j + \alpha_2 m_a(\dot{u} - \ddot{q}) + m_a \ddot{q}. \tag{5}$$

In this case, the non-linear force is described by $M = (l+1)(l+2)/2 + 2$ parameters.

EQUIVALENT MISO SYSTEM

By substituting Eq. (2) into Eq. (1), the non-linear equation of motion of the structural system can be expressed as

$$H_0\left(\frac{d}{dt}\right) q = \sum_{k=1}^{m} H_k\left(\frac{d}{dt}\right) f_k(u, \ldots, q, \ldots), \tag{6}$$

where $H_0(s) = ms^2 + cs + k$. The proposed identification method rearranges the terms of Eq. (6) by treating them as the input/output quantities of an equivalent MISO linear system. Specifically, the terms q and $f_k(u, \ldots, q, \ldots)$ of Eq. (6) that are associated with the unknown structural parameters are interpreted as the inputs x_k of the MISO system. The remaining terms of Eq. (6) are combined into the system output $y(t)$. Note that this study uses the terms "input" and "output" (x and y) in the context of system analysis terminology, whereas "excitation" and "response" (u and q) denote physically motivated structural analysis terms. The composed MISO system is described by the equation

$$\sum_{k=1}^{n} H_k\left(\frac{d}{dt}\right) x_k(t) = y(t), \tag{7}$$

where n denotes the total number of the input variables used in the equivalent MISO representation of the non-linear structural system.

The time domain description of the MISO system of Eq. (7) leads a frequency domain representation, as well. Specifically, using Fourier transform in Eq. (7) yields

$$\sum_{k=1}^{n} H_k(i\omega) X_k(\omega) + N(\omega) = Y(\omega), \tag{8}$$

where the capital letters denote the Fourier transforms

$$X_k(\omega) = F\{x_k(t)\}, \qquad Y(\omega) = F\{y(t)\}, \tag{9}$$

and the symbol $N(\omega)$ is introduced to represent the extraneous noise. The MISO system with the governing equation (8) is schematically shown in Figure 1.

IDENTIFICATION OF THE TRANSFER FUNCTIONS OF THE MISO SYSTEM

It is assumed that the excitation and the response of the non-linear system are stationary stochastic processes. Correspondingly, the input-output variates of the equivalent MISO linear system are also stationary processes; their correlation is described by the cross-correlation function

$$R_{x_i x_j}(\tau) = E[x_i(t) x_j(t+\tau)] \tag{10}$$

or by the cross-spectral density function

$$S_{x_i x_j}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_{x_i x_j}(\tau) e^{-i\omega\tau} d\tau. \tag{11}$$

In Eq. (10) $E[\circ]$ denotes the operator of mathematical expectation. Alternatively, the cross-spectral density function for stationary processes $x_i(t)$ and $x_j(t)$ can be defined as

$$S_{x_i x_j}(\omega) = \frac{2\pi}{T} E[X_i^*(\omega) X_j(\omega)], \tag{12}$$

where T denotes the length of observation of the processes $x_i(t)$, and $(\ )^*$ is used to denote complex conjugation. It can be shown that the definitions of Eqs. (11) and (12) are identical in the limit as T tends to infinity. In this context, they can be considered as equivalent for an adequately large, but finite, value of T.

In conjunction with the preceding comments, Eq. (8) governing the input-output relationship of the MISO system can be rewritten in the equivalent form

$$\{\underline{X}^t(\omega), Y(\omega)\}\begin{Bmatrix} \underline{H}(i\omega) \\ -1 \end{Bmatrix} = -N(\omega), \tag{13}$$

where $\underline{X} = \{X_1, \ldots X_n\}^t$, $\underline{H} = \{H_1, \ldots H_n\}^t$, and $(\ )^t$ denotes the transposition of the corresponding vectors. Premultiplying Eq. (13) by $2\pi(\{\underline{X}^t(\omega), Y(\omega)\}^*)^t/T$ and taking mathematical expectation yield

$$\begin{bmatrix} S_{\underline{x}\underline{x}} & S_{\underline{x}y} \\ S_{y\underline{x}} & S_{yy} \end{bmatrix}\begin{Bmatrix} \underline{H} \\ -1 \end{Bmatrix} = \begin{Bmatrix} 0 \\ -S_{yN} \end{Bmatrix}, \tag{14}$$

where it is assumed that the noise process $N(\omega)$ is uncorrelated from the inputs of the MISO system. Further, it follows from Eq. (14) that if the observation noise is zero and all non-linear effects are properly captured by the MISO system, then the vector $\{\underline{H}^t, -1\}^t$ is the eigenvector of the spectral matrix of the vector-process $\{\underline{X}^t(\omega), Y(\omega)\}^t$ corresponding to the zero eigenvalue.

The system of equations (14) can be readily used to determine the unknown transfer functions of the MISO system. Indeed, the first n equations of the system can be used to find the vector of the transfer functions $\underline{H}(i\omega)$. The last equation can be utilized to determine the value of S_{yN} which represents the influence of the measurement noise and the non-linear terms which were not included into the MISO system. In this regard it is pointed out that the value S_{yN} is equal to the spectral density of the noise. Indeed, by squaring Eq. (13) one can derive

$$S_{NN}(\omega) = \frac{2\pi}{T}\{(\underline{H}^t)^*, -1\}E\left[\begin{Bmatrix} \underline{X}^* \\ Y^* \end{Bmatrix}\{\underline{X}^t, Y\}\right]\begin{Bmatrix} \underline{H} \\ -1 \end{Bmatrix} = \{(\underline{H}^t)^*, -1\}\begin{bmatrix} S_{\underline{x}\underline{x}} & S_{\underline{x}y} \\ S_{y\underline{x}} & S_{yy} \end{bmatrix}\begin{Bmatrix} \underline{H} \\ -1 \end{Bmatrix} =$$

$$\{(\underline{H}^t)^*, -1\}\begin{Bmatrix} 0 \\ -S_{yN} \end{Bmatrix} = S_{yN}. \tag{15}$$

Upon evaluating the spectral density function S_{yN}, the cumulative coherence function [6,9] can be evaluated from the equation

$$\gamma_{y:\underline{x}}(\omega) = 1 - \frac{S_{NN}(\omega)}{S_{yy}(\omega)} = \frac{S_{yy}(\omega) - S_{yN}}{S_{yy}(\omega)}. \tag{16}$$

The coherence function $\gamma_{y:\underline{x}}(\omega)$ expresses the cumulative linear contribution of the random vector $\underline{x}(t)$ to the MISO system output $y(t)$. This coherence function reflects, in essence, the "goodness" of the selected model in describing the physical system.

In context with the preceding development, the Gaussian elimination procedure applied to the system of equations (14) provides a useful interpretation of the proposed method for numerical evaluation of $\underline{H}(i\omega)$. Note that the Gaussian elimination procedure transforms the matrix of Eq. (14) to an upper triangular matrix. At the step j of this procedure the components of the matrix located on the j-th column below the main diagonal are eliminated. In this regard, every step of the

Gaussian elimination is equivalent to a premultiplication of Eq. (14) by a lower triangular matrix with non-zero components located on the main diagonal and on the j-th column. Thus, during the first step of the Gaussian elimination, Eq. (14) is premultiplied by the matrix

$$A_1 = \begin{bmatrix} 1 & 0 & \dots & 0 \\ -S_{x_2x_1}/S_{x_1x_1} & 1 & & 0 \\ \dots & & & \dots \\ -S_{yx_1}/S_{x_1x_1} & 0 & \dots & 1 \end{bmatrix}. \tag{17}$$

This yields the system of equations

$$\begin{bmatrix} S_{x_1x_1} & S_{x_1x_2} & \dots & S_{x_1y} \\ 0 & S_{x_2x_2;1} & & S_{x_2y;1} \\ \dots & & & \dots \\ 0 & S_{yx_2;1} & \dots & S_{yy;1} \end{bmatrix} \begin{Bmatrix} H_1 \\ H_2 \\ \dots \\ -1 \end{Bmatrix} = \begin{Bmatrix} 0 \\ -S_{yN} \end{Bmatrix}, \tag{18}$$

where $S_{x_kx_j;1} = S_{x_kx_j} - S_{x_kx_1}S_{x_1x_j}/S_{x_1x_1}$. In this manner, during the j-th step of the Gaussian elimination, the system of equation obtained at the end of the previous $(j-1)$ step is premultiplied by the matrix

$$A_j = \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & & \dots \\ 0 & 0 & & 1 & & 0 \\ 0 & 0 & \dots & -S_{x_{j+1}x_j;j-1}/S_{x_jx_j;j-1} & \dots & 0 \\ \dots & \dots & & \dots & & \dots \\ 0 & 0 & \dots & -S_{yx_j;j-1}/S_{x_jx_j;j-1} & \dots & 1 \end{bmatrix}. \tag{19}$$

Then, the Gaussian elimination procedure can be formalized as a premultiplication of Eq. (14) by the matrix

$$A = A_nA_{n-1}\dots A_1. \tag{20}$$

Since all matrices A_j in Eq. (20) are lower triangular, the assembled matrix A is also lower triangular. Also, note that all elements of the main diagonal of the matrix A are equal to 1. The system of equation obtained after completion of the Gaussian elimination can be written as

$$\tilde{S}\begin{Bmatrix} \underline{H} \\ -1 \end{Bmatrix} = \begin{Bmatrix} 0 \\ -S_{yN} \end{Bmatrix}, \tag{21}$$

where $\tilde{S}$ is the upper triangular matrix.

In this regard it is pointed out that Eq. (21) can be obtained directly from Eq. (13) by pre multiplying it by the vector $2\pi(\{\underline{\tilde{X}}^t, \tilde{Y}\}^t)^*/T = 2\pi(A^*\{\underline{X}^t, Y\}^t)^*/T$ and taking mathematical expectation. Note that the matrix S of Eq. (21) can be viewed as the cross-spectral density matrix of the vector-process $\{\underline{\tilde{x}}^t, \tilde{y}\}^t$ and the vector-process $\{\underline{x}^t, y\}^t$. Therefore, it can be concluded from the triangular form of S that the process $\tilde{x}_l$ is uncorrelated from any of the processes x_j for $j < l$, and the process $\tilde{y}$ is uncorrelated from the vector-process $\underline{x}$. Further, the spectral density matrix of the vector-process $\{\underline{\tilde{x}}^t, \tilde{y}\}^t$ can be found from the equation

$$S_{\{\underline{\tilde{x}}^t, \tilde{y}\}^t} = \frac{2\pi}{T}E\left[\begin{Bmatrix} \underline{\tilde{X}} \\ \tilde{Y} \end{Bmatrix}^* \{\underline{\tilde{X}}^t, \tilde{Y}\}\right] = A\begin{bmatrix} S_{\underline{x}\underline{x}} & S_{\underline{x}y} \\ S_{y\underline{x}} & S_{yy} \end{bmatrix}(A^t)^* = \tilde{S}(A^t)^*. \tag{22}$$

Since both matrices $\tilde{S}$ and $(A^t)^*$ are upper triangular, the matrix $S_{\{\tilde{x}^t, \tilde{y}\}^t}$ is also upper triangular. Moreover, the matrix $S_{\{\tilde{x}^t, \tilde{y}\}^t}$ is a symmetric spectral density matrix of a random vector; it is a diagonal matrix. That is, the random vector-process $\{\tilde{x}^t, \tilde{y}\}^t$ is a vector-process with uncorrelated components. Also, note that since the elements of the main diagonal of the matrix $(A^t)^*$ are all equal to 1, the main diagonal of the matrix $S_{\{\tilde{x}^t, \tilde{y}\}^t}$ is identical to the main diagonal of the matrix $\tilde{S}$. That is, upon implementing Gaussian elimination on Eq. (14), the elements of the main diagonal of the obtained matrix S represent the spectral density functions of the corresponding orthogonalized processes.

In conjunction with the preceding developments, it is pointed out that the Gaussian elimination procedure of the system of equations (14) is equivalent to a "Gram-Schmidt" type of orthogonal decomposition of the random vector-process $\{\underline{x}^t, y\}$; this approach was previously pursued in Refs. [6,7,8,9] for identification of the MISO system transfer functions $H_j(i\omega)$.

Finally, it is noted that the matrix A can be determined in a more efficient manner, from a computational standpoint, by using the Cholesky decomposition of the matrix of Eq. (14). Indeed, Eq. (22) yields

$$\begin{bmatrix} S_{\underline{x}\underline{x}} & S_{\underline{x}y} \\ S_{y\underline{x}} & S_{yy} \end{bmatrix} = \left(A^{-1}\sqrt{S_{\{\tilde{x}^t, \tilde{y}\}^t}}\right)\left(\sqrt{S_{\{\tilde{x}^t, \tilde{y}\}^t}}(A^t)^{-*}\right) = LL^{t*}, \tag{23}$$

where L is the lower triangular matrix of the Cholesky decomposition. Since the Cholesky decomposition is unique, one finds

$$L = A^{-1}\sqrt{S_{\{\tilde{x}^t, \tilde{y}\}^t}}. \tag{24}$$

Note that the elements of the main diagonal of the matrix L represent the square root of the spectral density functions of the corresponding components of the orthogonalized vector-process $\{\tilde{x}^t, \tilde{y}\}^t$.

Upon evaluating the transfer functions of the equivalent MISO system by solving equation (14) with the use of the Cholesky decomposition procedure, the unknown structural parameters of the non-linear model can be obtained by using a curve-fitting procedure over the frequency range of interest. Several computational issues of the developed identification method warrant additional remarks. First, the introduced equivalent MISO and the ordering of its components must be selected based on a physical model which makes use of the available knowledge regarding the non-linear structural behavior. A physically motivated model can provide mathematical forms for the base functions of Eq. (2), or indicate the most expeditious expansion of the non-linear excitation force into a series of functions in a statistically equivalent sense. Note that the components that better describe the system physics correspond to the higher value of the coherence function and should be assigned to the first elements of the vector $\underline{x}$. Second, the proposed identification method can be sensitive to the error of estimating the power spectral density functions. It is a common practice in linear spectral analysis to use an appropriate window to reduce the effect of leakage in spectral estimation. Further, overlapping and zooming can be applied to the estimation process to recover the information lost during windowing [11]. Note that a biased estimation of the corresponding spectral density functions and inadequate MISO system representation may yield a matrix $S_{\underline{x}\underline{x}}(\omega)$ which is not positive definite for certain frequencies. Finally, it should be noted that noise may be induced during the measurement. The proposed method relies on the assumption, which is quite reasonable in most circumstances, that the noise is statistically independent from the structural excitation and response quantities; filtering techniques must be applied to reduce this noise as much as possible.

NUMERICAL EXAMPLE

To illustrate the application of the proposed method, a single-degree-of-freedom system excited by a wave force is considered. The wave force is modeled by Morison's equation. Thus, the

motion of the system is governed by the equation

$$m\ddot{q} + c\dot{q} + kq = \alpha_1|u-\dot{q}|(u-\dot{q}) + \alpha_2 m_a(\dot{u}-\ddot{q}) + m_a\ddot{q}. \quad (25)$$

The numerical values $m = 100$, $c = 1$, $k = 0.25$, $m_a = 10$, $\alpha_1 = 0.9$, and $\alpha_2 = 1.25$ are used. This structure has system natural frequency $\omega_n = \sqrt{k/m} = 0.05$, and damping ratio $\zeta = 0.1$. The Gaussian surface wave elevation is generated consistently with the Pierson-Moscowitz spectrum [9] corresponding to a wind speed of 25.74 m/s, using a superposition of sinusoids at 1000 frequency points uniformly spaced in the interval [0, 1.5] rad/sec. The horizontal wave velocity and acceleration at a depth of 10m in a "deepwater" location are computed using linear wave theory. A steady current of 0.5 m/s is superimposed on the velocity at the same location. Note that the assumption of an infinite water depth and of linear wave theory are adopted for simplicity and the proposed identification algorithm does not hinge upon these assumptions. Figure 2 shows the time histories, one-sided power spectra, and probability density functions of the water particle velocity and the displacement response. The low frequency response and the non-Gaussian distribution of the response are clearly discernible.

For the considered example the values of the parameters α_1 and α_2 are assumed to be unknown and are identified based on the simulated data. The base functions are selected as $|u-\dot{q}|(u-\dot{q})$ and $\dot{u}-\ddot{q}$. The following decomposed system is derived: $x_1 = |u-\dot{q}|(u-\dot{q})$, $x_2 = \dot{u}-\ddot{q}$, $y = (m-m_a)\ddot{q} + c\dot{q} + kq$, $H_1(i\omega) = \alpha_1$, and $H_2(i\omega) = \alpha_2 m_a$. Results of identification are shown in Figure 3. The transfer functions $H_1(i\omega) = \alpha_1$ and $H_2(i\omega) = \alpha_2 m_a$ are found to be constant; the parameters are obtained as $\hat{\alpha}_1 = Real(H_1(\omega))$ and $\hat{\alpha}_2 = Real(H_2(\omega))/m_a$, accordingly. The broken lines represent the true parameters, α_1 and α_2. The averaged parameters are $\hat{\alpha}_1 = 0.896$ and $\hat{\alpha}_2 = 1.231$. Note that beyond the cutoff frequency ($\omega = 1.5$ rad/sec) the identified values are of no interest. Shown in Figure 4 are the cumulative coherence functions associated with the SISO system having the input x_1 and with the MISO system with the inputs x_1 and x_2. Full frequency range is displayed, although only the range of [0, 1.5] rad/sec is to be observed. The coherence functions in Figure 4 provide a measure of the validity of the decomposed system and indicates the confidence with respect to the identified transfer functions. Indeed, a quantitative error analysis can be pursued to relate the confidence limit of an identified parameter to the coherence of its associated transfer functions. Also, it was found that the developed identification method is quite robust, with respect to adding noises to the observed excitation and response processes, within the frequency range containing the frequencies which are significant for describing waves and structural behavior.

CONCLUDING REMARKS

Method of identification of non-linear structural systems has been developed. It is based on the decomposition of system nonlinearities in terms of base functions, and on subsequent treatment of these base functions as input-output variables of an equivalent MISO linear system. The transfer functions of the MISO are determined by solving a system of frequency domain equations with the use of the Cholesky decomposition. The numerical example involving identification of the parameters of a non-linear wave force on the offshore structure demonstrates the applicability and usefulness of the proposed identification method.

ACKNOWLEDGEMENT

The financial support of this work by the grant DE-FG03-95ER14500 from the Department of Energy is gratefully acknowledged.

REFERENCES

1. J.R. MORISON, M.P. O'BRIEN, and J.W. JOHNSON, "The Force Exerted by Surface Waves on Piles," *AIME Petroleum Transactions*, 189, 149 (1950).
2. F. KOZIN and H.G. NATKE, "System Identification Techniques", *Structural Safety*, 3, 269, (1986).
3. S.J. GIFFORD and G.R. TOMLINSON, "Recent Advances in the Application of Functional Series to Nonlinear Structures", *Journal of Sound and Vibration*, 135, 289, (1989).
4. S.F. MASRI, H. SASSI, and T.K. CAUGHEY, "Nonparametric Identification of Nearly Arbitrary Nonlinear Systems", *ASME Journal of Applied Mechanics*, 49, 619, (1982).
5. K.S. MOHAMMAD, K. WORDEN, and G.R. TOMLINSON,. "Direct Parameter Estimation for Linear and Nonlinear Structures", *Journal of sound and Vibration*, 138, 471, (1992).
6. J.S. BENDAT, *Random Data: Analysis and Measurement Procedures*, Wiley-Interscience, NY. (1990)
7. J.S. BENDAT, P.A. PALO, and R.N.COPPOLINA, "A General Identification Technique for Nonlinear Differential Equations of Motion", *Probabilistic Engineering Mechanics*, 7, 43, (1992).
8. J.H. VUGTS and A.G. BOUQUET, "A Nonlinear Frequency Domain Description of Wave Forces on an Element of a Vertical Pile in Random Seas," *BOSS*, London, U.K., 693, (1985).
9. P.D. SPANOS and R. LU, "Nonlinear System Identification in Offshore Structure Reliability", *Journal of Offshore Mechanics and Arctic Engineering*, 117, 171, (1995).
10. B.A. ZELDIN, and P.D. SPANOS, "Identification of Nonlinear Systems under Random Excitation," *Proceedings of the ASCE Engineering Mechanics Conference*, Fort Launderdale, May 19-22, (1996).
11. A.V. OPPENHEIM and R.W. SHAFER, *Digital Signal Processing*, Prentice-Hall, NJ, (1975).

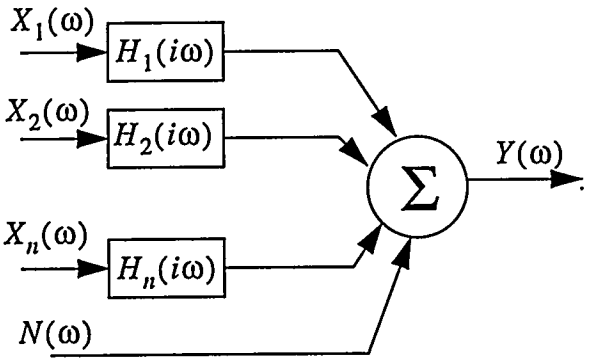


Figure 1. Equivalent MISO system

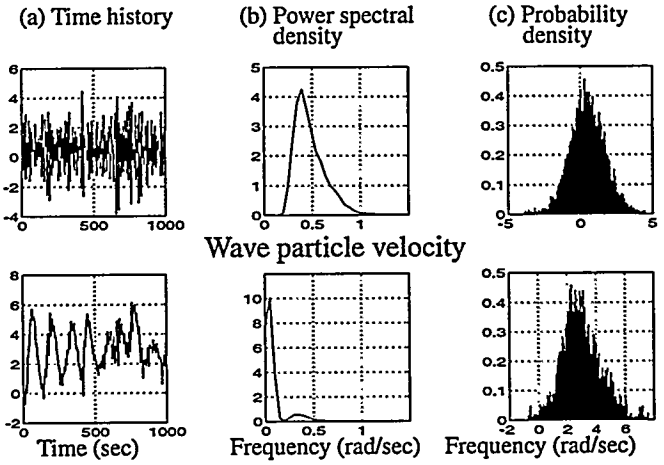


Figure 2. Simulated water particle velocity and structural displacement

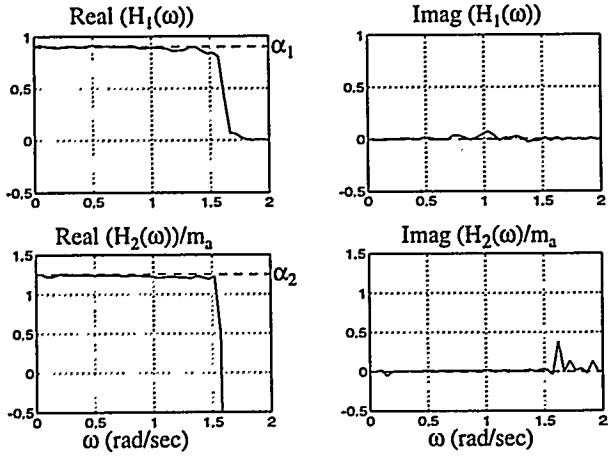


Figure 3. Identified linear transfer functions

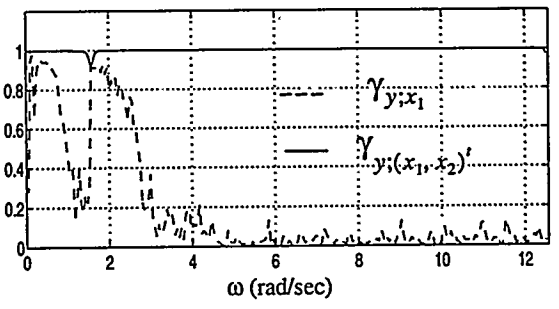


Figure 4. Cumulative coherence functions

NONLINEAR DIFFUSION AND SUPERCONDUCTING HYSTERESIS

I.D. Mayergoyz

Electrical Engineering Department
University of Maryland, College Park, MD 20742

ABSTRACT

Nonlinear diffusion of electromagnetic fields in superconductors with ideal and gradual resistive transitions is studied. Analytical results obtained for linear and nonlinear polarizations of electromagnetic fields are reported. These results lead to various extensions of the critical state model for superconducting hysteresis.

I. INTRODUCTION

It is well-known that type-II superconductors exhibit magnetic hysteresis which is caused by the pinning of the motion of flux filaments. The critical state (Bean) model has been proposed to treat the magnetic hysteresis of hard superconductors [1]-[2]. This model is based on the assumption of ideal (sharp) resistive transition which is described by step-wise E vs. J relation. This assumption leads to the nonlinear diffusion equation which admits simple analytical solutions for 1-D flux configurations. It has been gradually realized that the critical state model has some intrinsic limitations. First, this model leads to the explicit analytical results only for 1-D flux distributions and linear polarizations of external magnetic fields. Second, the critical state model does not account for gradual resistive transitions exhibited by actual superconductors. There are many publications in which extensive efforts are made to further generalize the critical state model. The intent of this paper is to summarize the recent contributions [3]-[7] of the author to this area.

II. NONLINEAR DIFFUSION IN SUPERCONDUCTORS WITH IDEAL RESISTIVE TRANSITIONS

Ideal resistive transitions are described by nonlinear Ohm's law illustrated by Figure 1. According to this law, any electric field, however, small, will induce full (critical) current density j_c to flow.

Consider a superconducting cylinder of arbitrary cross-section (Figure 2a) subject to uniform field $\vec{B}_e(t)$ whose direction does not change with time. As the time varying flux enters the superconductor, it induces screening (shielding) currents of density $\pm j_c$. The distribution of these superconducting screening (shielding) currents is such that they create the magnetic field which at any instant of time completely compensates for the change in the external field $\delta B_e(t)$ in the region interior to superconducting currents.

It is clear that $\delta B_e(t)$ changes its sign as $B_e(t)$ goes through its extremum values. This results in a reversal in the direction (polarity) of superconducting screening currents values. By using this fact, the essence of the critical state model can be summarized as follows.

Each reversal in the external magnetic field results in the formation of a surface layer of superconducting screening currents. This layer extends inward with time until another reversal (extremum) value of the external field is reached. At this point, the inward progress of the previous superconducting current layer is terminated and a new inward extending current layer is formed. The previously induced layers of persisting superconducting currents stay still and they represent past history of the temporal variations of the external field. This past history leaves its mark upon future distributions of superconducting currents.

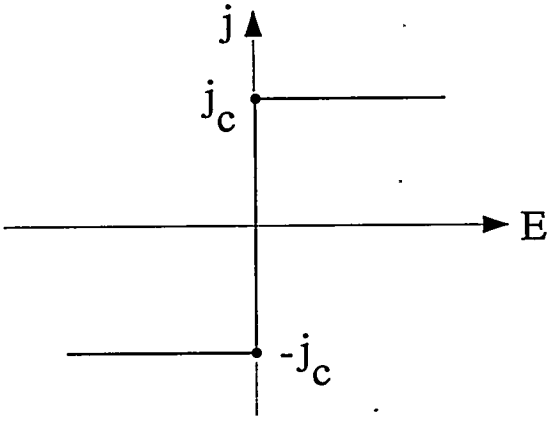


Figure 1. Nonlinear Ohm's law.

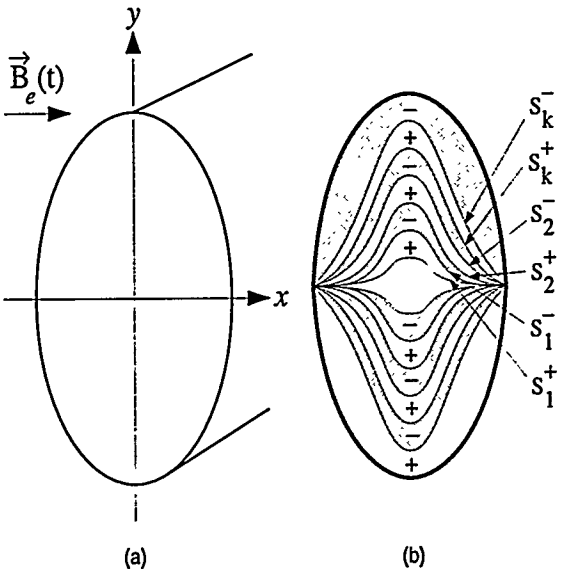


Figure 2. Superconducting cylinder subject to external field.

Thus, it can be concluded that at any instant of time there exist several (many) layers of persisting superconducting currents (see Figure 2b). These persisting currents have opposite polarities (directions) in adjacent layers. The interior boundaries S_k^+ and S_k^- of all the layers (except the last one) remain still and they are uniquely determined by the past extremum values of $B_e(t)$. The last induced current layer extends inward as the external field changes in time monotonically.

The magnetic moment, $\vec{M}$, of the superconductor is related to the distribution of the superconducting screening currents as follows:

$$\vec{M}(t) = \int\int_S (\vec{r} \times \vec{j}(t)) ds, \tag{1}$$

where the integration is performed over the superconductor cross-section.

In general, this magnetic moment has x- and y-components. According to (1), these components are given by the expressions:

$$M_x(t) = \int\int_S yj(t)ds, \quad M_y(t) = -\int\int_S xj(t)ds. \tag{2}$$

It is apparent from the previous discussion that the instantaneous values of $M_x(t)$ and $M_y(t)$ depend not only on the current instantaneous value of the external field $B_e(t)$ but on the past extremum values of $B_e(t)$ as well. This is because the overall distribution of persisting superconducting currents, j, depends on the past extrema of $B_e(t)$. Thus, it can be concluded that the relationships $M_x(t)$ vs. $B_e(t)$ (and $M_y(t)$ vs. $B_e(t)$) exhibit **discrete** memories which are characteristic and intrinsic to rate independent hysteresis.

It is clear from the above description of the critical state model that a newly induced and inward extending layer of superconducting currents will wipe out (replace) some already existing layers of persisting superconducting currents if they correspond to the previous extremum values of $B_e(t)$ which are exceeded by a new extremum value. In this way, the effect of those previous extremum values of $B_e(t)$ on the overall future current distributions will be completely eliminated. According to (2), the effect of those past extremum values of the external magnetic field on the magnetic moment will be eliminated as well. This is the "wiping out" property of the superconducting hysteresis as described by the critical state model.

Next, we proceed to the discussion of the "congruency" property. Consider two distinct variations of the external field, $B_e^{(1)}(t)$ and $B_e^{(2)}(t)$. Suppose that these two external fields have different past histories and, consequently, different sequences of local past extrema. However, starting from some instant of time they vary back-and-forth between the same reversal values. It is apparent from the description of the critical state model and expressions (2) that these two identical back-and-forth variations of the external field will result in the formation of two minor loops for hysteretic relation $M_x(t)$ vs. $B_e(t)$ (or $M_y(t)$ vs. $B_e(t)$). It is also apparent from the same description of the critical state model that these two back-and-forth variations of the external field will affect in the **identical** way the **same** surface layers of a superconductor. Unaffected layers of the persistent superconducting currents will be different for $B_e^{(1)}(t)$ and $B_e^{(2)}(t)$ because of their different past histories. However, according to (2), these unaffected layers of persistent currents result in constant in time ("background") components of the magnetic moment. Consequently, it can be concluded that the same **incremental** variations of $B_e^{(1)}(t)$ and $B_e^{(2)}(t)$ will result in equal **increments** of M_x (and M_y). This is tantamount to the congruency of the corresponding minor loops.

The experimental testing of the "congruency" and "wiping out" properties has been recently undertaken [8] and it has been found that these properties are in good compliance with experimental data for tested superconductors.

It has been established [9],[10] that the "wiping out" property and the "congruency" property constitute the necessary and sufficient conditions for the representation of actual hysteresis nonlinearities by the Preisach model. Thus, the description of the superconducting hysteresis by the critical state model is equivalent to the description of the same hysteresis by the Preisach model, which is defined as follows:

$$-M(t) = \iint\limits_{\alpha \geq \beta} \mu(\alpha, \beta)\hat{\gamma}_{\alpha\beta}B_e(t)d\alpha d\beta, \tag{3}$$

where: $M(t)$ can be x- or y-component of the magnetic moment, negative sign in (3) accounts for diamagnetic nature of superconductor, $\hat{\gamma}_{\alpha\beta}$ are elementary hysteresis operators which are represented

by rectangular loops with α and β as the "up" and "down" switching values, respectively; the function $\mu(\alpha,\beta)$ is not specified in advance and should be determined from matching first-order transition curves [10].

The question can be asked "What is to be gained from the above result?" The answer is as follows. There is no readily available analytical machinery for the calculation of the interior boundaries of superconducting current layers for specimens of arbitrary shapes. For this reason, the critical state model does not lead to the mathematically explicit results. The application of the Preisach model allows one to circumvent these difficulties by using some experimental data. Namely, for any superconducting specimen, the "first-order transition" curves can be measured and used for the identification of the Preisach model [10] for the given specimen. By using these curves, complete prediction of hysteretic behavior of the specimen can be given at least at the same level of accuracy and physical legitimacy as in the case of the critical state model. In particular, cyclic and "ramp" losses can be explicitly expressed in terms of the first-order transition curves [10].

III. NONLINEAR DIFFUSION IN SUPERCONDUTORS WITH GRADUAL RESISTIVE TRANSITIONS (LINEAR POLARIZATION)

Actual resistive transitions of superconductors are gradual and they are usually described by the "power law"

$$E = (j/k)^n, \qquad (n > 1), \tag{4}$$

where E is an electric field, j is a current density, and k is the parameter which coordinates the dimensions of both sides of the last equation.

The exponent "n" is a measure of the sharpness of the resistive transition and it may vary in the range 7-1000. At first, the power law was regarded only as a reasonable empirical description of the resistive transition. However, recently there has been a considerable research effort focused on the theoretical justification of power law (4). In this paper, power law (4) is used as a constitutive equation for superconductors.

It is easy to show that this constitutive relation leads to the following nonlinear diffusion equation for the current density:

$$\frac{\partial^2 J^n}{\partial z^2} = \mu_0 k^n \frac{\partial J}{\partial t}. \tag{5}$$

We shall first consider the solution of this equation for the following boundary and initial conditions:

$$J(0,t) = ct^p, \qquad (t \geq 0, p > 0), \tag{6}$$

$$J(z,0) = 0 \qquad (z > 0). \tag{7}$$

By using the dimensionality analysis, the self-similar solution to the initial-boundary value problem (5)-(7) can be found. For $n \geq 7$, this self-similar solution can be written (with sufficiently high accuracy) as follows [3]:

$$J(z,t) = \begin{cases} ct^p \left(1 - \frac{z}{dt^m}\right)^{1/(n-1)}, & \text{if } z \leq dt^m, \\ 0, & \text{if } z \geq dt^m, \end{cases} \tag{8}$$

where

$$d = \sqrt{(nc^{n-1})/[\mu_o k^n m(n-1)]}. \tag{9}$$

The close examination of self-similar solution (8) leads to the following conclusion: in spite of the wide range of variation of boundary conditions (6) (see Figure 3), the profile of electric current density $J(z,t)$ remains approximately the same. For typical values of n, this profile is very close to a rectangular one. This suggests that the actual profile of electric current density will be close to a rectangular one for any monotonically increasing boundary conditions $J_0(t) = J(0,t)$. Thus, we arrive at the following generalization of the critical state model.

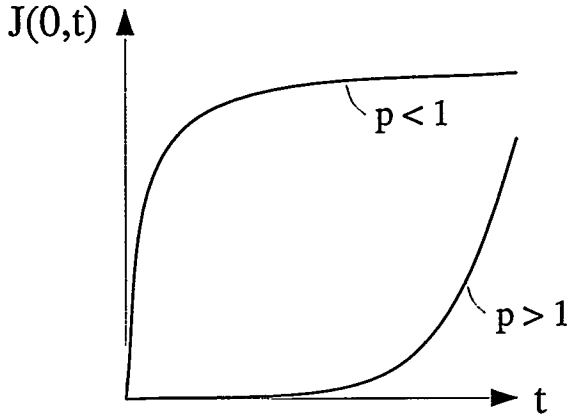


Figure 3. Boundary conditions (6) for various p.

Current density $J(z,t)$ has a rectangular profile with the height equal to the instantaneous value $J_0(t)$ of electric current density on the boundary of superconductor (see Figure 4). Magnetic field $H(z,t)$ has a linear profile with a slope determined by instantaneous value of $J_0(t)$.

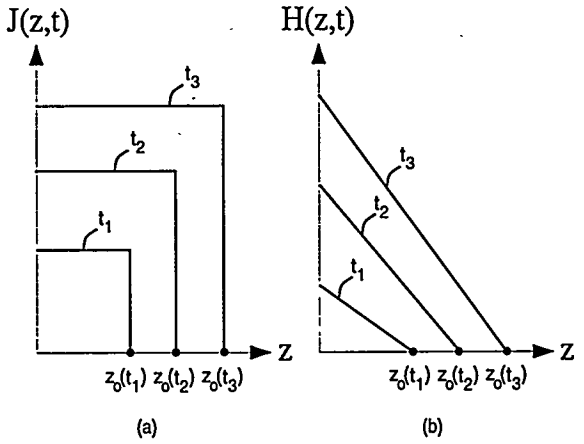


Figure 4. Generalization of the critical state model.

To better appreciate the above generalization, we remind that in the critical state model the current has a rectangular profile of constant (in time) height, while the magnetic field has a linear profile with constant (in time) slope.

For zero front $z_0(t)$ of the current (or magnetic field) profile, we have:

$$z_0(t) = \frac{H_0(t)}{J_0(t)}. \tag{10}$$

To find $J_0(t)$ in terms of $H_0(t)$, we multiply (2) by z and integrate from 0 to $z_0(t)$ with respect to z and from 0 to t with respect to t. After simple transformations, we arrive at the following expression:

$$\mu_0 k^n \int_0^{z_0(t)} zJ(z,t)dz = \int_0^t J_0^n(\tau)d\tau. \tag{11}$$

By using the rectangular profile approximation for $J(z,t)$, we obtain

$$\frac{\mu_0 k^n}{2} J_0(t) z_0^2(t) = \int_0^t J_0^n(\tau)d\tau. \tag{12}$$

By substituting expression (10) into (12), we find

$$\frac{\mu_0 k^n}{2}\frac{d}{dt}\left[\frac{H_0^2(t)}{J_0(t)}\right] = J_0^n(t). \tag{13}$$

By integrating equation (13), we can find the expression for $J_0(t)$ in terms of $H_0(t)$, which, after substitution into (10), leads to the following formula for $z_0(t)$ in terms of the boundary values of the magnetic field $H_0(t)$:

$$z_0(t) = \frac{1}{H_0(t)}\left[\frac{2(n+1)}{\mu_0 k^n}\int_0^t H_0^{2n}(\tau)d\tau\right]^{1/(n+1)}. \tag{14}$$

It is instructive to point out that nonlinear diffusion in superconductors with gradual resistive transitions may exhibit a peculiar (anomalous) mode which does not exist in superconductors with ideal resistive transitions. This is a "standing" mode. In the case of this mode, the electromagnetic field on a superconductor boundary increases with time, while the region occupied by the electromagnetic field does not expand.

The "standing" mode is the exact solution of the nonlinear diffusion equation (5) which corresponds to the following initial and boundary conditions:

$$J(z,0) = \begin{cases} \left[\frac{(n-1)\mu_0 k^n (z_0-z)^2}{2(n+1)nt_0}\right]^{\frac{1}{n-1}}, & \text{if } 0 \le z \le z_0, \\ 0, & \text{if } z \ge z_0, \end{cases} \tag{15}$$

$$J_0(t) = J(0,t) = \left[\frac{(n-1)\mu_0 k^n z_0^2}{2(n+1)n(t_0-t)}\right]^{\frac{1}{n-1}}, t \ge t_0. \tag{16}$$

The standing mode solution itself is given by the expression [6]:

$$J(z,t) = \begin{cases} \left[\frac{(n-1)\mu_0 k^n (z_0-z)^2}{2(n+1)n(t_0-t)}\right]^{\frac{1}{n-1}}, & \text{if } 0 \le z \le z_0, \\ 0, & \text{if } z \ge z_0, \end{cases} \tag{17}$$

. This solution is illustrated by Figure 5.

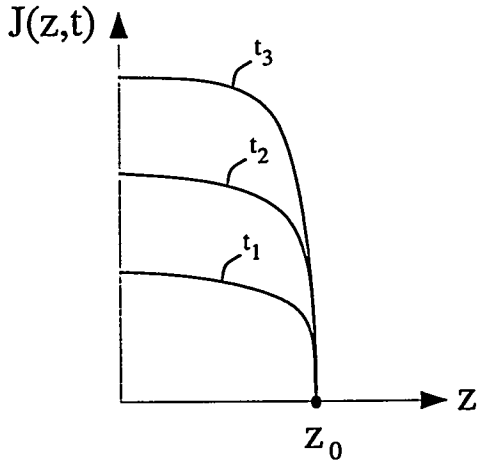


Figure 5. "Standing" mode.

The origin of the "standing" mode can be elucidated on physical grounds as follows. Under the boundary condition (16), the electromagnetic energy entering the superconducting material at any instant of time is just enough to affect the almost uniform increase in electric current density in the region $(0 \le z \le z_0)$ already occupied by the field, but insufficient to affect the further diffusion of the field into

the material. Mathematically, it is clear that the origin of the "standing" mode is closely related to the fact that in the case of nonlinear diffusion equation (5) one deals with "finite support" solutions, that is with the solutions which have finite zero front. In the case of linear diffusion, the zero front is always infinite.

IV. NONLINEAR DIFFUSION IN SUPERCONDUCTORS WITH GRADUAL RESISTIVE TRANSITIONS (NONLINEAR POLARIZATIONS)

Most of the literature on the critical state model is concerned with scalar superconducting hysteresis. This is because the study of vector hysteresis requires the investigation of nonlinear diffusion of electromagnetic fields into superconductors for the case when these fields are not linearly polarized. This is a very difficult analytical problem that requires the solution of coupled nonlinear partial differential (diffusion) equations. It turns out that this problem can be attacked as follows. First, we consider the case of circular polarization of electromagnetic fields. In this case, the exact analytical solutions of coupled nonlinear diffusion equations can be found due to the high degree of symmetry associated with this polarization. Then, the case of arbitrary polarization is treated as perturbation of the circular polarization.

To start the discussion, consider a plane circularly polarized electromagnetic wave penetrating superconducting half-space $z > 0$. This implies the following boundary conditions.

$$E_x(0,t) = E_m \cos(\omega t + \gamma),$$
$$E_y(0,t) = E_m \sin(\omega t + \gamma), \tag{18}$$

$$E_x(\infty,t) = E_y(\infty,t) = 0. \tag{19}$$

The distribution of electric field E in half-space $z > 0$ is governed by the coupled nonlinear diffusion equations:

$$\frac{\partial^2 E_x}{\partial z^2} = \mu_0 \frac{\partial}{\partial t}[\sigma(\sqrt{E_x^2 + E_y^2})E_x], \tag{20}$$

$$\frac{\partial^2 E_y}{\partial z^2} = \mu_0 \frac{\partial}{\partial t}[\sigma(\sqrt{E_x^2 + E_y^2})E_y], \tag{21}$$

where according to (4)

$$\sigma\left(\sqrt{E_x^2 + E_y^2}\right) = \sigma(E) = kE^{1/n-1}. \tag{22}$$

It turns out that exact (periodic in time) analytical solution to the boundary value problem (18)-(22) can be found [4]. This solution has the following form:

$$E_x(z,t) = E_m\left(1 - \frac{z}{z_0}\right)^{\alpha'} \cos[\omega t + \theta(z) + \gamma], \tag{23}$$

$$E_y(z,t) = E_m\left(1 - \frac{z}{z_0}\right)^{\alpha'} \sin[\omega t + \theta(z) + \gamma], \tag{24}$$

where $z < z_0$ and

$$z_0 = \sqrt[4]{\frac{2n(n+1)(3n+1)^2}{\sqrt{\omega}\mu_0\sigma_m(n-1)}},$$
$$\sigma_m = kE_m^{1/n-1}, \tag{25}$$

$$\theta(z) = \alpha'' ln(1 - z/z_0), \tag{26}$$

$$\alpha' = \frac{2n}{n-1}, \quad \alpha'' = \sqrt{\frac{2n(n+1)}{N-1}}. \tag{27}$$

For $z \geq z_0$, the electric field is equal to zero.

The remarkable property of the above solution is the fact that the circular polarization is preserved everywhere within the superconducting medium. As a result, there are no higher order time harmonics of the electromagnetic field anywhere within the medium despite its nonlinear properties. This fact can be easily understood on the physical grounds if we take into account the rotational symmetry of the boundary value problem (18)-(22).

Next, we consider the nonlinear diffusion of the electromagnetic field for the case when the electric field on the boundary is specified as follows:

$$\begin{aligned} E_x(0,t) &= E_m[\cos(\omega t + \gamma) + \epsilon f_x(t)], \\ E_y(0,t) &= E_m[\sin(\omega t + \gamma) + \epsilon f_y(t)], \end{aligned} \tag{28}$$

where ϵ is some small parameter, while $f_x(t)$ and $f_y(t)$ are given periodic functions of time. It is apparent that this case can be construed as a perturbation of the circular polarization. We shall look for the periodic solution in the form:

$$\begin{aligned} E_x(z,t) &= E_x^0(z,t) + \epsilon e_x(z,t), \\ E_y(z,t) &= E_y^0(z,t) + \epsilon e_y(z,t). \end{aligned} \tag{29}$$

By using the machinery of perturbation technique, it can be shown [5] that E_x^0 and E_y^0 coincide with the "circular polarization" solution (23)-(27), while e_x and e_y satisfy the following equations:

$$\begin{aligned} \frac{\partial^2 e_x(z,t)}{\partial z^2} = \mu_0 \sigma_m \Big(1 - \frac{z}{z_0}\Big)^2 \frac{\partial}{\partial t}\Big[\Big(\frac{1+n}{2n} \\ + \frac{1-n}{2n}\cos[2\omega t + 2\theta(z)]\Big) e_x(z,t) \\ + \frac{1-n}{2n}\sin[2\omega t + 2\theta(z)] e_y(z,t)\Big], \end{aligned} \tag{30}$$

$$\begin{aligned} &\frac{\partial^2 e_y(z,t)}{\partial z^2} = \\ &\mu_0 \Big(1 - \frac{z}{z_0}\Big)^2 \frac{\partial}{\partial t}\Big[\frac{1-n}{2n}\sin[2\omega t + 2\theta(z)] e_x(z,t) \\ &+ \Big(\frac{1+n}{2n} - \frac{1-n}{2n}\cos[2\omega t + 2\theta(z)]\Big) e_y(z,t)\Big]. \end{aligned} \tag{31}$$

Equations (30) and (31) are coupled linear partial differential equations of parabolic type with a variable in time and space coefficients. To find the periodic solutions of these equations, we introduce new complex valued state variables:

$$\begin{aligned} \varphi(z,t) &= e_x(z,t) + i e_y(z,t), \\ \psi(z,t) &= e_x(z,t) - i e_y(z,t), \end{aligned} \tag{32}$$

and the following Fourier series expansions for them:

$$\varphi(z,t) = \sum_{k=-\infty}^{\infty} \varphi_{2k+1}(z) e^{i(2k+1)\omega t}, \tag{33}$$

$$\psi(z,t) = \sum_{k=-\infty}^{\infty} \psi_{2k+1}(z) e^{i(2k+1)\omega t}. \tag{34}$$

Here, it is tacitly assumed that $f_x(t)$ and $f_y(t)$ (and with them $\varphi(z,t)$ and $\psi(z,t)$) are functions of half-wave symmetry.

By using (32), (33), and (34) in (30) and (31), the following equations for φ_{2k+1} and ψ_{2k+1} can be derived [5]:

$$\begin{aligned}&\left(1-\frac{z}{z_0}\right)^2\frac{d^2\varphi_{2k+1}}{dz^2}\\&=i\chi_{2k+1}\left[a\varphi_{2k+1}-\left(1-\frac{z}{z_0}\right)^{i2\alpha''}\psi_{2k-1}\right],\end{aligned}\tag{35}$$

$$\begin{aligned}&\left(1-\frac{z}{z_0}\right)^2\frac{d^2\psi_{2k-1}}{dz^2}=\\&=i\chi_{2k-1}\left[a\psi_{2k-1}+\left(1-\frac{z}{z_0}\right)^{-i2\alpha''}\varphi_{2k+1}\right]\\&(k=0,\pm1,\pm2,\ldots),\end{aligned}\tag{36}$$

where

$$a=\frac{1+n}{1-n},\qquad \chi_{2k+1}=(2k+1)\omega\mu_0\sigma_m\frac{1-n}{2n}.\tag{37}$$

Thus, the problem of integration of partial differential equations (30)-(31) is reduced to the solution of infinite set of ordinary differential equations with respect to Fourier coefficients φ_{2k+1} and ψ_{2k-1}. The remarkable property of these simultaneous equations is that they are only coupled by pairs. It allows one to solve each pair of these coupled equations separately. After φ_{2k+1} and ψ_{2k-1} are found, we can compute $\varphi(z,t)$ and $\psi(z,t)$, and then $e_x(z,t)$ and $e_y(z,t)$.

A solution of coupled equations (35)-(36) can be found in the form:

$$\begin{aligned}\varphi_{2k+1}(z)&=A_{2k+1}\left(1-\frac{z}{z_0}\right)^{\beta},\\\psi_{2k-1}&=B_{2k-1}\left(1-\frac{z}{z_0}\right)^{\beta-i2\alpha''},\end{aligned}\tag{38}$$

and details of calculations of β, A_{2k+1} and B_{2k-1} are given in [5].

It is easy to show [5] that in the important case of elliptical polarization, when

$$f_x(t)=\cos\omega t,\qquad f_y(t)=\sin\omega t,\tag{39}$$

only the first ($k=1$) pair of equations (35)-(36) must be solved. This means that only first and third harmonics are not equal to zero.

ACKNOWLEDGEMENT

The reported research is supported by the U.S. Department of Energy, Engineering Research Program.

REFERENCES

1. C.P. BEAN, Phys. Rev. Lett. 8, 250 (1962).
2. H. LONDON, Phys. Lett. 6, 162 (1963).
3. I.D. MAYERGOYZ, J. Appl. Phys. 76, 7130 (1994).
4. I.D. MAYERGOYZ, J. Appl. Phys. 75, 6963 (1994).
5. I.D. MAYERGOYZ, J. Appl. Phys. 75, 6956 (1994).

6. I.D. MAYERGOYZ, J. Appl. Phys. 79, 6473 (1996).

7. I.D. MAYERGOYZ and M. NEELY, J. Appl. Phys. 79, 6602 (1996).

8. G. FRIEDMAN, L. LIN, and J.S. KOUVEL, J. Appl. Phys. 75, (5685) (1994).

9. I.D. MAYERGOYZ and T. KEIM, J. Appl. Phys. 67, 5466 (1990).

10. I.D. MAYERGOYZ, Mathematical Models of Hysteresis, Springer-Verlag, 1-204 (1994).

MICROSTRUCTURE AND MOMENTUM TRANSPORT IN CONCENTRATED SUSPENSIONS

Lisa A. Mondy

Sandia National Laboratories, Dept. 9112, Albuquerque, New Mexico 87185-0834

Alan L. Graham

Los Alamos National Laboratory, ESA-EPE, Los Alamos, New Mexico 87545

Howard Brenner

Dept. of Chem. Eng., Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

ABSTRACT

This paper reviews several coupled theoretical and experimental investigations of the effect of microstructure on momentum transport in concentrated suspensions. An expression to predict the apparent suspension viscosity of mixtures of rods and spheres is developed and verified with falling-ball viscometry experiments. The effects of suspension-scale slip (relative to the bulk continuum) are studied with a sensitive spinning-ball rheometer, and the results are explained with a novel theoretical method. The first noninvasive, nuclear magnetic resonance imaging measurements of the evolution of velocity and concentration profiles in pressure-driven entrance flows of initially well mixed suspensions in a circular conduit are described, as well as more complex two-dimensional flows with recirculation, *e.g.* flow in a journal bearing. These data in nonhomogeneous flows and complementary three-dimensional video imaging of individual tracer particles in homogeneous flows are providing much needed information on the effects of flow on particle interactions and effective rheological properties at the macroscale.

INTRODUCTION

Many industrial processes include the transport of suspensions of solid particles in liquids, such as coal and other solid feedstock slurries. Oil, gas, and geothermal energy production rely on the transport of suspensions such as muds, cements, proppant, and gravel slurries in the drilling and completion of a well. Suspensions are also found in high-energy-consumption industrial processes such as found in pulp and paper manufacturing. The complex rheological response of suspensions often limit the efficiency of the design of such processes, causing loss of productivity, increased cost, and increased energy consumption. Because of

the importance of particulate two-phase flows in the applications described above, the study of suspension rheology remains an important technical research topic for the Department of Energy.

This overview of our recent research supported by the Department of Energy, Office of Basic Energy Sciences, will focus on flow of suspensions of relatively large particles in which colloidal and inertial effects are negligibly small. There is growing evidence that even in this restricted range of flows, the rheology of a suspension with a nondilute particle concentration cannot be characterized by a material function. Instead, the microstructure of the suspension determines the overall macroscopic properties, and the flow history of the suspension determines aspects of the microstructure. Advances in the ability to predict the rheological response of concentrated suspensions depend on answering three broad questions: 1) How does the microstructure of a suspension affect the rheological properties? 2) How do boundaries, such as walls, affect the microstructure and properties? 3) How does the macroscopically imposed flow field affect the microstructure of a suspension? Aspects of these questions are being addressed in our work.

In the following section we will explore the first question by discussing the use of falling-ball rheometry as a means to circumvent the shear-induced changes in microstructure that can be encountered when using conventional rotational devices to measure suspension viscosity. We will discuss falling-ball rheometry used to determine the apparent viscosity of a suspension of particles of two shapes. In the third section we will discuss experimental and theoretical aspects of spinning-ball rheometry in otherwise quiescent suspensions and show that this can provide a sensitive measure of slip at the surface of a particle.

The fourth section of this paper focuses on efforts to develop capability to predict the evolution of concentration and velocity profiles of an initially well mixed suspension as it demixes when subjected to nonhomogeneous shear flows. If the local concentration is known, one can then use the falling-ball information to determine the *local* viscosity in a flow field. Global behavior can then be determined by incorporating a spatially varying viscosity field into the usual balance equations. We will illustrate the existence of flow-induced microstructural changes with data on the time evolution of concentration and velocity profiles in suspensions undergoing flow in pipes and between counter-rotating eccentric cylinders (journal bearings). When the suspended particles are small in comparison to the characteristic dimensions of the flow apparatus, steady-state concentration and velocity profiles are in good agreement with predictioins of the shear-induced migration model [1,2]. However, another avenue to modeling particle migration is to use a kinetic theory approach, which has been applied successfully in granular flows [3,4]. In this theory the intensity of the velocity fluctuations, caused by particle interactions, is characterized by a 'granular temperature' analogous to the temperature in classical kinetic theories and governed by a balance of fluctuation energy. Under some situations this approach leads to the same balance equations as with the first model, but with a hydrodynamic diffusion that can be determined in homogeneous flow fields. We will describe experiments where we use particle tracking techniques, originally developed in falling-ball studies, to determine the granular temperature of various suspensions undergoing homogeneous flow between parallel moving belts.

FALLING-BALL RHEOMETRY IN COMPLEX SUSPENSIONS

In previous work, we have shown that falling-ball rheometry is an excellent tool to probe the rheological properties of a suspension without significantly changing the properties through the very act of measuring them. Unlike conventional viscometers, which employ flow fields that tend to influence the microstructure of the suspension, falling-ball rheometry can be used to determine the macroscopic viscosity of a suspension with little effect on the microstructure [5]. This is especially useful for suspensions of particles with aspect ratio greater than one, whose alignment is especially sensitive to the flow field. We have recently begun to use falling-ball rheometry to study suspensions of particles with a mix of shapes.

Most investigations on the rheology of concentrated suspensions have focused on monodisperse suspensions of either spherical or rodlike particles. In practice, most suspensions contain particles that are polydisperse both in size and shape. Only a limited number of studies have been devoted to the problem of size polydispersity in suspensions of spherical particles, and even less is known about the behavior of suspensions composed of particles of different shapes.

Farris [6] develop a model for the viscosity of suspensions of spheres with multimodal diameter distributions. In his model, for each fraction of a given particle size, the smaller particles in suspension have the same effect as a homogeneous fluid with Newtonian viscosity similar to the viscosity of a suspension made up only of the fraction of smaller spheres. In other words, the smaller suspended particles do not interact with the larger particles and are 'sensed' by the larger particles as part of the continuous suspending fluid. We will apply these concepts to develop an equation for the relative viscosity of a suspension composed of a mixture of rodlike and spherical particles. If the rods are large enough relative to the spheres, we may consider the spherical particles as part of the homogeneous suspending continuum. Let us define an apparent sphere volume fraction $\phi^*_s = V_s/(V_0 + V_s) = \phi_s/(1-\phi_r)$, where V is volume and the subscripts s, 0, and r stand for the spheres, fluid, and rods, respectively. If we assume the viscosity of a suspension composed of spheres and rods is the same as the viscosity of a suspension of rods suspended in a Newtonian homogeneous fluid of viscosity identical to the viscosity of an equivalent suspension of spheres with a volume fraction ϕ^*_s, we may write (after Farris):

$$\mu_{rel}(\phi_T) = \mu_{rel,\,spheres}(\phi^*_s)\mu_{rel,\,rods}(\phi_r) \quad , \tag{1}$$

where the relative viscosity of a suspension is the viscosity of the suspension normalized by the viscosity of the suspending continuum. Several expressions are available for the relative viscosity of suspensions of spheres (e.g. those listed by Graham et al. [7]). Here we adopt the following empirical relation [8], which has agreed well with previous falling-ball measurements:

$$\mu_{rel,\,spheres}(\phi) = 1 + 2.5(\phi) + 10.05(\phi)^2 + 0.00273e^{(16.6\phi)} \tag{2}$$

For the viscosity of a suspension of randomly oriented rods, we have the following empirical relation for rods with aspect ratio of 20 [5]:

$$\mu_{rel,\,rods}(\phi) = 1 + 28.50\phi \qquad \phi < 0.125 \tag{3a}$$

$$= 1 + 2.040\phi^3 \qquad \phi > 0.125 \tag{3b}$$

The relative viscosity of a mixed suspension may now be calculated for any combination of rods (of aspect ratio 20) and spheres by the set of equations (1)-(3).

The falling-ball experimental apparatus, materials, and methodology have been described in detail previously [5,9]. The suspensions were composed of mixtures of poly(methyl methacrylate) particles in a Newtonian liquid. The particles were mixtures of spheres, with diameters of 3.175 mm, and aspect-ratio-20 rods, with length of 31.65 mm. The rod-sphere mixtures were suspended in a liquid solution with three primary ingredients (50wt% alkylaryl polyether alcohol, 35wt% polyalkylene glycol, and 15wt% tetrabromoethane). The weight fractions of the ingredients were adjusted so that the density and the refractive index of the fluid would match those of the particles. Three different suspensions were prepared with total solids volume fraction ϕ_T of 0.35 (volume fraction of rods ϕ_r=0.05 and of spheres ϕ_s=0.30), 0.40 (ϕ_r=0.10 and ϕ_s=0.30), and 0.45 (ϕ_r=0.05 and ϕ_s=0.40) respectively. The falling balls were either chrome-plated steel, monel, or tungsten-carbide ball bearings with diameters between 6.35 mm and 15.88 mm.

The trajectories of the falling balls were recorded on a high-speed digitizing video system. An average velocity for an individual experiment was determined by measuring the elapsed time for a ball to settle a known distance on the screen. The results of up to 40 individual experiments with a falling ball of one nominal size (not necessarily of one material) were averaged to determine a reproducible effective viscosity of the suspension. Two or three sizes of falling balls were used for each suspension and the results showed no significant effect of the relative sizes of the falling ball and the suspended particles over the size ranges listed above. The average apparent relative viscosity for each suspension was obtained by averaging the entire

set of experimental results (up to 120 individual experiments).

These results are shown in Figure 1 along with the values predicted by equations (1)-(3). The solid and dotted lines represent the relative viscosity for a suspension of randomly oriented rods of aspect ratio a_r=20 [eq. 3a and 3b] and for a suspension of spheres [eq. 2], respectively. The broken lines represent the calculated viscosity for mixed rod-sphere suspensions with the indicated fraction of rods (5%, 10%, and 15%) based on the predictions of eq. 1 combined with eqs. 2 and 3. The agreement between the falling-ball experimental points and the calculated lines is very good.

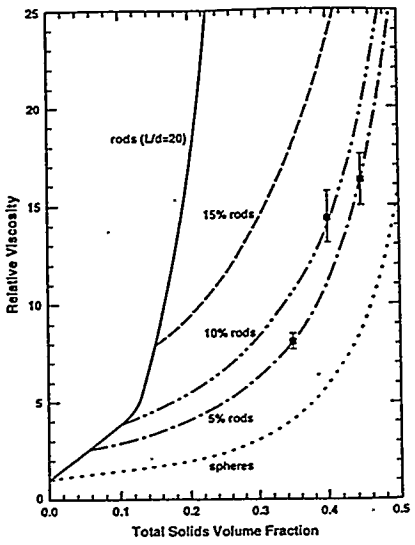


Figure 1. Falling-ball viscosities of mixtures of suspended rods and spheres.

SLIP STUDIES

Spinning-Ball Experiments

In the falling-ball experiments described in the section above, the drag on the ball appeared to be that found in a Newtonian liquid with *no slip* at the ball boundaries. Instead of measuring the mean velocity of a falling ball, we could instead measure the mean torque on a spinning ball. This geometry is more sensitive to slip at the ball boundary. Whereas the force F on a ball moving slowly through an unbounded Newtonian liquid without slip can be described as $F=6\pi a\mu v$ (where μ is the viscosity of the liquid and a and v are the radius and velocity of the ball, respectively), the force with perfect slip is $4\pi a\mu v$. In contrast, the torque T on a ball spinning slowly in a Newtonian liquid is described by Kirchoff's law, $8\pi a^3\mu\Omega$. (where Ω is the angular velocity of the ball); however, the torque on a ball with perfect slip at the boundaries is zero [10].

Kunesh and coworkers studied the torque on balls spinning in single-phase Newtonian liquids, verified the formula above, and quantified the effects of the free surface [11]. We have completed similar experiments to measure the torque on balls spinning in otherwise quiescent suspensions. We measured the torque on three sizes of balls (0.32, 1.27, and 2.54 cm in diameter) spinning in various suspensions. Suspension with solids volume fractions of 0.25, 0.40, and 0.50 were studied. Three sizes of suspended spheres (0.07, 0.32, and 0.64 cm in diameter) were used in the suspending oil described earlier. The suspensions were well mixed prior to the start of an experiment.

Typical traces of the torque on a 1.27 cm-diameter ball in terms of the relative viscosity (the measured spinning-ball viscosity normalized by capillary viscosity measurements of the suspending liquid) in the suspending liquid and in a suspension with ϕ=0.5 is shown in Figure 2. The suspending liquid measurements agree well with capillary measurements and show no variation with time (number of revolutions). On the

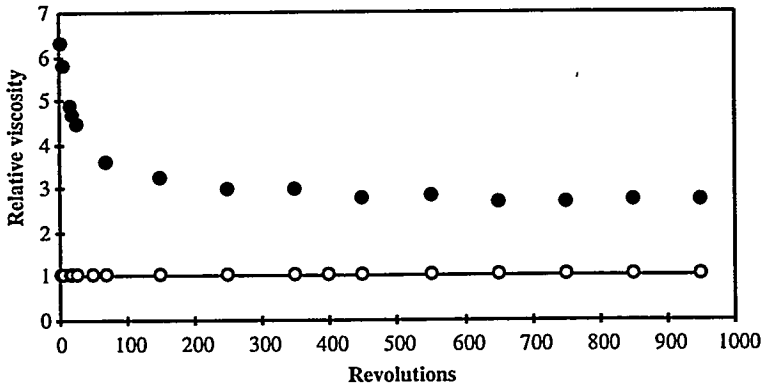


Figure 2. Spinning-ball viscosities (relative to capillary measurement of the suspending liquid viscosity) for the suspending liquid (o) and for a suspension with ϕ=0.50 (●).

other hand, the suspension measurements show a distinct fall off in the measured viscosity with increasing numbers of revolution. This is expected, as this nonhomogeneous flow induces particle migration (which will be discussed in the following section). The short-time behavior (an average torque for the first four revolutions) is taken as an indication of any apparent slip at the ball's surface in the still homogeneous suspension. The effects of the relative sizes of the spinning ball and the suspended spheres are shown in Fig. 3. Here, all data are taken in suspensions with ϕ=0.5; however, the suspended-sphere size varies as well as the spinning-ball size. As the suspended spheres become small compared to the spinning ball, the spinning-ball viscosity increases. The torque experienced (initially) on a 2.54-cm-diameter ball spinning in a suspension of 0.07-cm-diameter particles is correspondent to the viscosity measured with falling-ball rheometry. Conversely, when the spinning ball and the suspended spheres are comparable in diameter, the presence of significant 'Kirchoff-law slip' is observed.

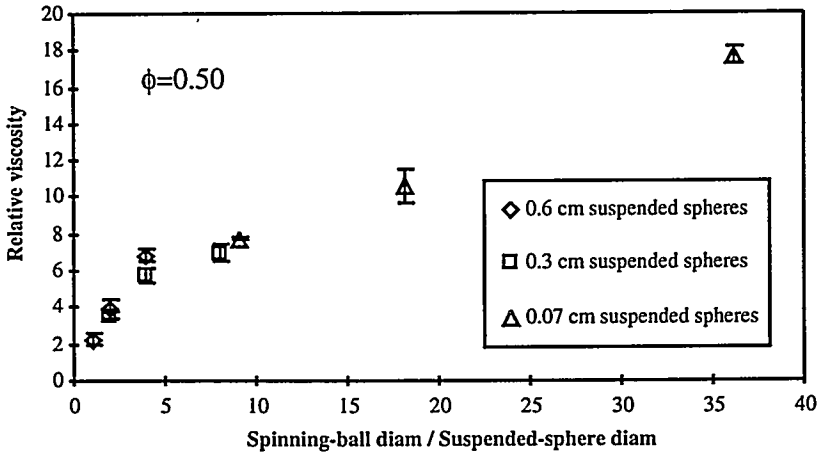


Figure 3. The effect of the relative sizes of the spinning ball and the suspended spheres on the initial apparent viscosity measured assuming Kirchoff's law. The data point to the far right is close to the value predicted by falling-ball studies [23] and conventional rheometry [8].

Theoretical Developments

Einstein's [12] classical analysis of the rheology of a dilute suspension related the increased viscosity of the suspension to the additional dissipation occurring within a 'suspension cell' owing to the perturbing

presence of a freely suspended sphere in an otherwise uniform shear field. However, scalar dissipation arguments are viable only in cases where the suspension behaves macroscopically as a homogeneous isotropic fluid. In particular, these methods are inapplicable in circumstances where the suspension-scale stress/strain-rate relationship is anisotropic. Batchelor [13] and Brenner [14] developed a general theory from which the stress/strain-rate relation may be obtained. Their methods are based on averaging over a 'suspension cell' the interstitial-scale stress and velocity gradient tensors. Higher-order terms in the relative viscosity/suspended-particle concentration expansion have been obtained by Batchelor & Green [15,16], based on an 'ensemble average' approach. Each of the above methods is essentially local in nature; that is, effects of bounding walls as well as spatial nonhomogeneities in the ambient velocity gradient are neglected. When the ensemble-average approach is applied, and the existence of walls ignored, nonconvergent integrals arise, which require *ad hoc* renormalization methods to overcome.

Recently, a new method has been developed for rheologically homogenizing a dilute suspension composed of freely suspended spherical particle dispersed in a Newtonian fluid [17]. The method is global in nature; that is, wall effects and spatial dependence of both the ambient flow and the particle number density are encountered, enabling known classical results for the suspension viscosity to be obtained without the need for renormalization.

When the ambient flow is singular (as for example in the case of a small sedimenting or rotating ball comparable in diameter to the suspended spheres) it is possible to use this technique to estimate the velocity at points far from the singularity. In a recent paper, it is shown that even far from the singularity (relative to the freely suspended sphere radius), the suspension does *not* behave like a homogeneous medium [17]. Specifically, due to interparticle hydrodynamic interactions, the average extra-torque exerted on a ball rotating at a given angular velocity (and, conversely, the average reduction in angular velocity experienced by a sphere on which a given torque is exerted), are not related by the Kirchoff's law linear factor $8\pi d^3\mu$, instead a suspension-scale 'slip' occurs at the surface of the spinning ball in agreement with the experimental work discussed above. Furthermore, the extra-torque felt by a ball held at constant angular velocity in a suspension and the reduction in angular velocity felt by the same ball held at constant torque do not correspond directly. In fact, when the ratio of the spinning-ball diameter to the suspended-sphere diameter is roughly one, the dimensionless extra-torque is almost 25 per cent larger than the comparable reduction in angular velocity. This phenomenon cannot occur in a homogeneous medium for which the constitutive stress/rate-of-strain relationship is an intrinsic material property of the system.

In the case of a sedimenting ball, the 'apparent viscosity' obtained experimentally by the supposed applicability of Stokes law agrees with the viscosity of the suspension measured by standard viscometric methods if the ball is the same size or larger than the suspended spheres [9]. However, if the ball is somewhat smaller, the reduction in sedimentation velocity is *less,* apparently because of a 'slip' at the surface of the sedimenting ball [18]. Recent theoretical results show this as well for dilute suspensions [19]. It is interesting to note that the appearance of 'slip' occurs over a smaller range of the ratio of the tracer (in this case, falling) ball to the suspended particles than in the rotating ball case. Furthermore, the theory predicts that if the falling ball is yet smaller relative to the suspended spheres than those studied experimentally, the reduction in sedimentation velocity then becomes significantly *higher.* This contrasting behavior arises from the difference in the respective probability density functions for the cases of sedimenting and rotating spheres. For the case of a rotating sphere, the probability density function $P(\mathbf{x}_1/s_0)$ is constant (where $\mathbf{x}_1$ is the location of the center of a suspended sphere and s_0 corresponds to the domain inside a sedimenting or rotating sphere) and independent of the relative sizes of the rotating and suspended spheres. In contrast, in the case of a sedimenting sphere, it exhibits large gradients near s_0 for very small sedimenting spheres, rendering the near-field contribution dominant. Since the settling velocity decreases significantly when the settling and freely suspended spheres nearly touch, the reduction in sedimentation velocity increases proportionally. In other words, whereas the rotating sphere has only one 'interaction' mechanism (namely that the overall effect of hydrodynamic interactions decays monotonically with decreasing ratio of the rotating-to-suspended sphere diameter since the domain in which the effect of the suspended sphere is sensible shrinks), the settling sphere has two competing mechanisms. The first is similar to that of the rotating

sphere. The second is the increase of probability density function in the vicinity of the singularity with decreasing ratio of the rotating-to-suspended sphere diameter, which makes the near-field contribution dominant in spite of the fact that the domain in which this effect is significant shrinks.

EFFECTS OF FLOW ON THE MICROSTRUCTURE OF SUSPENSIONS

Pressure-Driven Pipe Flow

Flow-induced migration of suspended particles is thought to occur whenever particle interactions are more frequent in one part of a flow field than in another, as could occur in the presence of spatially varying shear rate, concentration, or viscosity fields. The spatial distribution of suspended particles present in concentrated suspensions is difficult to measure because most suspensions are opaque even at relatively low particle concentrations. However, under the auspices of the Department of Energy, Office of Basic Energy Sciences, noninvasive techniques based on nuclear magnetic resonance (NMR) imaging have been developed by Fukushima and coworkers to study both concentration and velocity profiles in multiphase flows [20,21]. We have employed these NMR imaging techniques to study the flow-induced migration of particles in suspension when subjected to a variety of flow fields.

One of our more recent studies involved low-particle-Reynolds-number pressure-driven flow in a circular conduit of suspensions ranging in solids volume fraction ϕ of 0.1 to 0.45. Measurements were made using 3.175-mm-diameter particles in a 50.4-mm-diameter tube ($a/R = 0.0625$) and 675-mm-diameter particles in a 25.4-mm-diameter tube ($a/R = 0.0266$). The primary data obtained from these experiments were NMR images of the concentration (ϕ) and velocity (v) fields at various locations downstream of an in-line mixer.

During flow development, significant migration to the axis of the tube and 'plug-like' velocity profiles were observed at all solids volume fractions. Full flow development occurred sooner than predicted by existing scaling arguments. Evidence suggests that, at higher concentrations (>30%), evolution of the ϕ and v profiles occur on different length scalses. Two flow rates were tested (9.89 mm/s and 197.7 mm/s). The development of the ϕ and v profiles were independent of flow rate.

Example steady-state profiles are shown in Figure 4. At the lower ratio of a/R the ϕ profile achieves a cusp at the center of the flow. The higher a/R (0.0625) is significantly above the ratio suggested by Seshardi & Sutera [22] and Mondy et al. [23] to be the upper limit of continuum behavior. Particle size effects manifest themselves as somewhat more blunted concentration profiles at larger a/R. The depletion in ϕ apparent near the wall would result in a 'layer' of lower viscosity and a reduction in the pressure drop required to flow the suspension, which is also consistent with the findings of Mondy et al. [23]

Piston-Driven Pipe Flow

In contrast to pressure-driven pipe flow, piston-driven pipe flow is not unidirectional. At the surface of the moving piston the velocity profile is necessarily uniform, yet downstream the velocity profile becomes parabolic for a Newtonian liquid. In order for this to occur, continuity requires that liquid near the pipe walls be swept into the center of the pipe. We find that this complex flow leads to particle migration in both the radial and axial directions.

Recently, we have studied the flow of a concentrated suspension in a 38 cm long by 5 cm diameter pipe equipped with a driving piston at one end and a freely moving piston at the other. This geometry results in a closed system ideal for an NMR imaging study of a two-dimensional flow. Two suspensions were imaged, one of 0.07-cm-diameter spheres and one of 0.32-cm-diameter spheres, both at an overall solids volume fraction of 0.50. Figure 5A shows an image of the suspension of smaller spheres in the region near the driving piston after the piston has traveled approximately 5 pipe diameters. A region of high liquid content has formed near the pipe walls and has been swept to the pipe axis along the piston face. Figure 5B shows the radially averaged solids volume fraction, in the suspension of larger spheres, along the axis of the pipe from

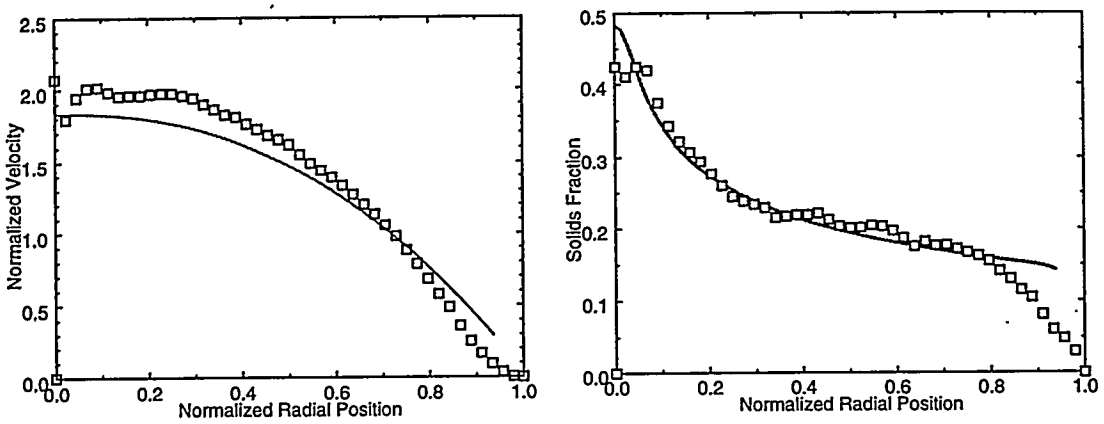


Figure 4. Steady-state ν and ϕ profiles for a/R = 0.0625 at $\phi = 0.20$ as measured with NMR (symbols) and predicted values (lines) from the improved shear-induced migration model.

the driving piston to the freely moving one. The particles 'lead' the fluid and concentrate at the far end of the flow (farthest from the driving piston).

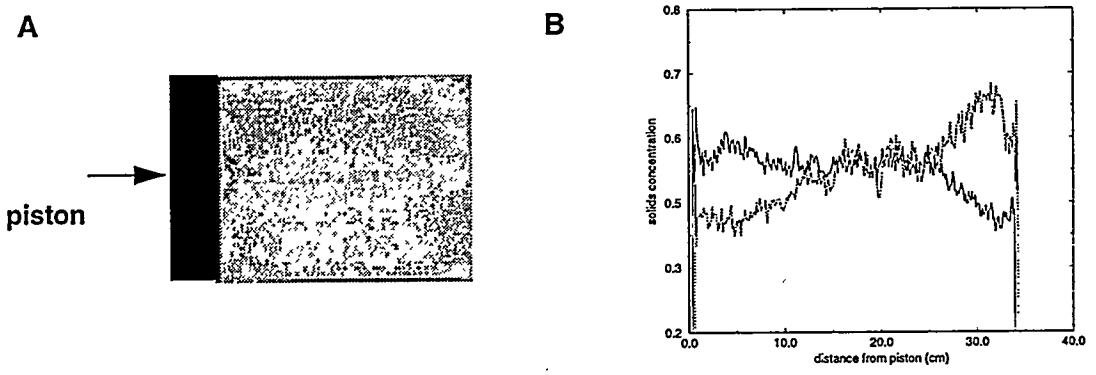


Figure 5. A. NMR image near the driving piston. Dark areas represent regions of higher liquid fraction. B. Radially averaged solids volume fraction along the axis of the pipe from the driving piston to the freely moving one.

Numerical Modeling and Constitutive Equation Development

The NMR data discussed above has been compared with the results of a constitutive model developed by Phillips et al. [2] after Leighton & Acrivos [1]. This constitutive model consists of both a Newtonian constitutive equation, in which the viscosity depends on the local particle volume fraction, and a diffusive equation that accounts for shear-induced particle migration. Two adjustable parameters arise in the diffusive equation, which describe the relative strength of the mechanisms for particle migration. These two rate parameters can be evaluated empirically with experimental measurements of velocity and concentration profiles in a wide-gap Couette apparatus. We have recently determined these parameters as functions of ϕ.

One criticism of this model was its prediction of a cusp-shaped concentration profile in pressure-driven pipe flow. However, the recent NMR experiments have shown that such concentration profiles can occur, as shown in Figure 4, which is of one of the more blunted obtained. The original Phillips formulation [2] has been modified so that the shear rate is averaged over the size of a particle. This results in different fully

developed profiles for different particle sizes, in agreement with the data. Examples of the predictions and the close agreement with data can be seen in Figure 4. However, the original formulation and the new model both overpredict the entrance length in comparison to the data.

The second criticism was the inherently one-dimensional treatment of the particle migration (it is dependent on gradients in a scalar shear-rate). Although this could still be an impediment to generalizing the model to complex flows, we have found that this simple model can often predict the very complex behavior of suspensions. The constitutive expression previously described by Phillips et al. [2] has been expanded to two-dimensional flows by describing the flow in terms of the strain rate tensor $\mathbf{D}$ and the migration in terms of gradients in the generalized shear rate $\dot{\gamma} = (2 \text{ tr } \mathbf{D}^2)^{1/2}$. The equation set was then solved numerically and the predictions compared to NMR imaging data. NMR imaging has also been used to study the flow of concentrated suspensions in the gap between a rotating inner cylinder placed eccentrically within an outer fixed cylinder (a journal bearing). We reported earlier [24] that this model, when coupled with a finite volume solver, failed to capture the qualitative nature of this two-dimensional flow. Specifically, at certain values of the eccentricity a very slow recirculation occurs and concentrated suspensions evolve a concentration profile with the maximum concentration of solids occurring, not at the outer wall, but inside the gap. The earlier numerical results always predicted a monotonic increase in solids volume fraction from the rotating inner cylinder to the outer wall, and no recirculation zone was predicted. However, recently we have used a finite element technique with more success. Figure 6 shows the development of a spatially varying solids volume fraction as the number of turns of the inner cylinder increases. The predicted profiles are remarkably similar to the NMR images. In addition, the calculations do indeed predict a recirculation zone.

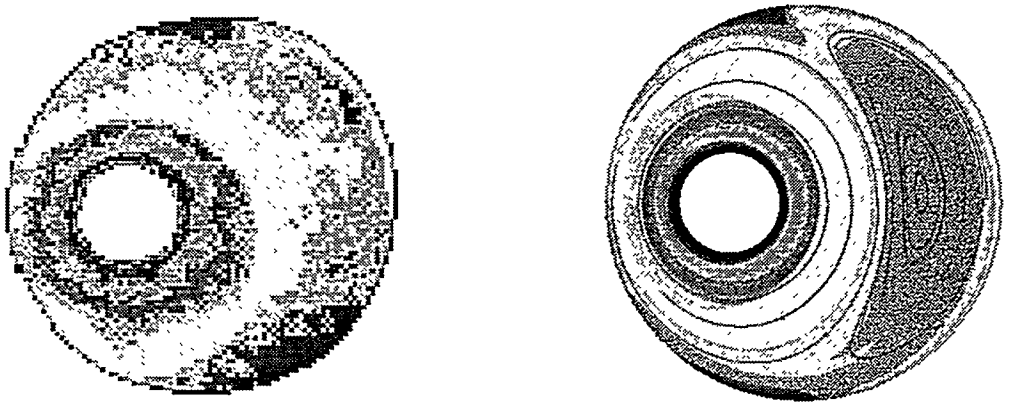


Figure 6. NMR image of liquid volume fraction contours (left) and predicted contours of liquid volume fraction and streamlines (right) for a suspension undergoing flow in a cylindrical journal bearing after 1000 turns.

The finite element model was also used to predict the behavior of concentrated suspensions undergoing piston-driven flow. Figure 7 shows the predicted spatially varying concentration obtained under the same conditions as the experiments described in the previous subsection. The contour plot is remarkably similar to the NMR results shown in Figure 5.

Microrheological Observations

Despite some successes with the above model, we feel that other avenues should continue to be explored to ensure that the particle migration phenomena is adequately understood and appropriately generalized to multiple dimensions. One such avenue recently suggested is to use a kinetic theory approach, which has

been applied successfully in granular flows [3,4]. In this theory the intensity of the velocity fluctuations, caused by particle interactions, is characterized by a "granular temperature" analogous to the temperature in classical kinetic theories and governed by a balance of fluctuation energy. This approach emphasizes the importance of measuring not only average behavior of suspensions but the details of the fluctuations about those averages. Under some situations this approach leads to the same balance equations as with the first model, but with a hydrodynamic diffusion that can be determined in homogeneous flow fields.

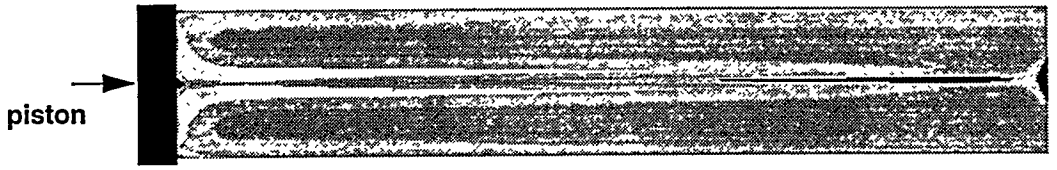


Figure 7. Predicted concentration contours in piston-driven flow using a finite element code.

Currently, we are using particle tracking techniques, originally developed in falling-ball studies, to determine the granular temperature of various suspensions undergoing homogeneous flow between parallel moving belts. The suspensions consist of poly(methyl methacrylate) 0.635-cm-diameter spheres neutrally buoyant in the oil mixture described earlier. The suspended spheres are primarily transparent, with the same index of refraction as the suspending liquid. A few opaque tracer spheres, otherwise identical to the others, are added to the suspension.

To date, 100 detailed three-dimensional trajectories of the tracer spheres in a suspension with ϕ=0.20 undergoing flow at two different shear rates have been recorded. Figure 8A shows one such trajectory (in only two of the three directions), where the y-direction is parallel to the belt and in the direction of motion and the x-direction is in the direction of the overall velocity gradient. The origin is at the center of the device. Figure 8B shows the velocity fluctuations in the x-direction for an ensemble of 50 particles at a shear rate of 3.34 sec^{-1}. From our preliminary measurements, it appears that the velocity fluctuations are more or less isotropic. The granular temperature in each direction is defined as $T = \langle u' \cdot u' \rangle_p$, where u' is the velocity fluctuation of a particle about its local mean velocity and the angled brackets denote ensemble averaging over all the particles. Preliminary results for T_x (in the x-direction) at two shear rates are shown in Figure 8C.

Currently we are instrumenting the homogeneous flow apparatus with piezoelectric pressure sensors that will allow the measurement of the frequency of particle-wall interactions, as well as the additional pressure due to the presence of the particles. These are pieces of information critical to the evaluation of granular-flow based suspension rheology models.

CONCLUSIONS

We have performed a variety of experimental, theoretical, and numerical studies to elucidate the linkage between the microstructure and the macroscopically observed responses of suspensions of particles in liquids. NMR imaging studies and visual observations have confirmed that a suspension's microstructure can change dramatically during flow. Falling-ball viscometers, on the other hand, can be used (under certain circumstances) to determine an apparent viscosity of a *homogeneous* suspension, without significantly affecting the microstructure during the measurement. Quiescent suspensions can also be used to examine effects of boundaries. We have described one such measurement: the torque on a rotating ball in otherwise quiescent suspensions. Recent theoretical results have also shed light on experimental results indicating Stokes law and Kirchoff's law could be presumed to hold only under limited circumstances.

Information obtained with quiescent suspensions can be combined with information about the evolving microstructure in a flow to predict the spatial variations in viscosity and the global behavior. We have had successes in modeling multidimensional flows with an approach that describes shear-induced particle migration with a diffusive equation. However, further studies of the details of particle interactions are needed

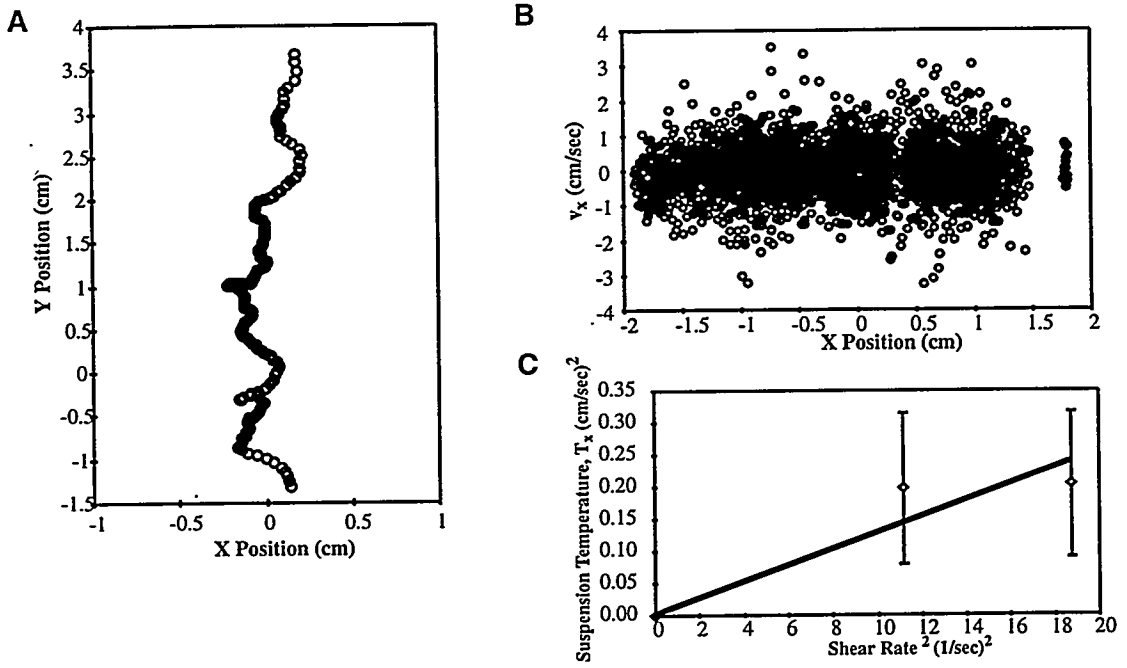


Figure 8. A. The trajectory of a tracer sphere in a suspension with ϕ=0.20 undergoing uniform shear. B. Velocity fluctuations of 50 particles in this flow. C. Resultant granular temperatures at two shear rates.

before definitive predictive capabilities can be developed. Measurement of the detailed fluctuations of the velocity of particles in suspension undergoing flow is an example of one such study.

ACKNOWLEDGMENTS

This work was sponsored by the U. S. Department of Energy, at Los Alamos National Laboratory under Contract W-7405-ENG-36 with the University of California and at Sandia National Laboratories under Contract DE-AC04-76DP00789. The authors would like to acknowledge gratefully support for this work by the U. S. Department of Energy, Division of Engineering and Geosciences, Office of Basic Energy Sciences. Partial funding for the work done at Los Alamos and Sandia was also provided by SDIO/IST and managed by the Office of Naval Research.

Much of this work was done in collaboration with other scientists. The falling-ball studies on mixtures of rods and spheres were done in collaboration with Prof. Moshe Gottlieb, Ben Gurion University. The spinning-ball theoretical work was done in collaboration with Prof. Yaniv Almog, Massachusetts Institute of Technology. The NMR imaging studies were made possible with the expertise and hard work of Drs. Eiichi Fukushima and Steve Altobelli, The Lovelace Institutes. Theory development and numerical modeling of particle migration has been successful through collaborative efforts with Dr. Andrea Mammoli, Los Alamos National Laboratory, and Dr. Sam Subia and Prof. Marc Ingber, University of New Mexico.

REFERENCES

1. Leighton, D. and Acrivos, A., *J. Fluid Mech.* 275, 155-199 (1987).

2. R. J. Phillips, R. C. Armstrong, R. A. Brown, A. L. Graham, J. R. Abbott, *Phys Fluids A* 4, 30-40 (1992).

3. D.F. McTigue and J. T. Jenkins, "Channel flow of a concentrated suspension," In *Advances in Micromechanics of Granular Materials*, H. H. Shen et al., Editors, Elsevier Science Publishers, New York, 1992.

4. P. R. Nott and J. F. Brady, *J. Fluid Mech.*, 275, 157-199 (1994).

5. W. J. Milliken, L. A. Mondy, M. Gottlieb, A. L. Graham, and R. L. Powell, *J. Fluid Mech.* 202, 217 (1989).

6. R. J. Farris, *Trans. Soc. Rheol.* 12, 281 (1968).

7. A. L. Graham, R. D. Steele, and R. B. Bird, *Ind. Eng. Chem. Fundam.* 23, 420 (1984).

8. D. G. Thomas, *J. Colloid Sci.* 20, 267 (1965).

9. L. A. Mondy, A. L. Graham, and J. Jensen, *J. Rheol.* 30, 1031(1986).

10. H. Lamb, *Hydrodynamics*, Dover Publications, New York, 6th edition, 1945.

11. J. G. Kunesh, H. Brenner, M. E. O'Neill, and A. Falade, *J. Fluid Mech.* 154, 29-42 (1985).

12. A. Einstein, *Ann. Physik* 19, 289-306 (1906). Errata, *Ann. Physik* 34, 591-592 (1911).

13. G. K. Batchelor, *J. Fluid Mech.* 41, 545-570 (1970).

14. H. Brenner, *Chem. Engng. Sci.* 27, 1069-1107 (1972).

15. G. K. Batchelor, and J. T. Green, *J. Fluid Mech.* 56, 375-400 (1972).

16. G. K. Batchelor, and J. T. Green, *J. Fluid Mech.* 56, 401-427 (1972).

17. Y. Almog and H. Brenner, "Renormalization-free homogenization of a dilute suspension of spheres," preprint (1996).

18. W. J. Milliken, L. A. Mondy, M. Gottlieb, A. L. Graham, and R. L. Powell, *Phys. Chem. Hydro.* 11, 341 (1989).

19. Y. Almog and H. Brenner, "Non-continuum anomalies in the apparent viscosity experienced by a test sphere moving through an otherwise quiescent suspension," preprint (1996).

20. A. Caprihan and E. Fukushima, *Physics Reports* 198, 195-235 (1990).

21. P. D. Majors, R. C. Givler, and E. Fukushima, *J. Magn. Reson.* 85, 235-243 (1989).

22. V. Seshadri and S. P. Sutera, *Trans. Soc. Rheol*, 14, 351 (1970).

23. L. A. Mondy, A. L. Graham, and M. Gottlieb, *Proceedings of the Xth International Congress on Rheology*, Sydney, Australia 2, 137 (1988).

24. N. Phan-Thien, A. Graham, S. A. Altobelli, J. R. Abbott, and L. A. Mondy, *Ind. Eng. Chem Res.* 34, 3187-3194 (1995).

SHEAR-INDUCED PARTICLE DIFFUSION AND ITS EFFECT ON THE FLOW OF CONCENTRATED SUSPENSIONS

Andreas Acrivos

Levich Institute
Steinman Hall, #1M
City College of CUNY
140th Street & Convent Avenue
New York, New York 10031

ABSTRACT

The mechanism underlying shear-induced particle diffusion in concentrated suspensions is clarified. Examples are then presented where this diffusion process plays a crucial role in determining the manner by which such suspensions flow under laminar conditions.

INTRODUCTION

During the past ten years, it has been widely recognized that concentrated suspensions of non-colloidal particles undergoing shear give rise to a number of curious and important phenomena which play a vital role in governing the rheology of such systems. Examples include the resuspension of a settled bed of heavy particles even under conditions of varnishingly small inertial forces, and the migration of neutrally buoyant particles across streamlines from regions of high shear to low thereby creating a non-uniform concentration profile with a highly concentrated center plug in a suspension flowing in a tube. All these effects are due to the existence of shear-induced particle diffusion at vanishingly small particle Reynolds numbers [1].

At first glance, it might appear that such a diffusive process which leads to particle migrations across streamlines would run counter to the well-known reversibility property of the creeping flow equations. Thus, one might argue that if, for example, migration were to occur in a given direction where the flow is, say, from left to right, reversing the flow direction should also reverse the sense of this migration. This argument, however, applies only to systems consisting of a finite (relatively small) number of particles in the absence of non-

hydrodynamic effects. In contrast, a space filling suspension having an effectively infinite number of particles behaves in a fundamentally different way.

The reason is as follows: When dealing with a finite number of particles, their velocities $\mathbf{U}_i$ can, in principle, be determined exactly via the solution of the creeping flow equations, given their positions $\mathbf{r}_i$ as well as the imposed flow field. Thus, the evolution of their configuration is governed by the Smoluchovski equation, shown in the figure, which is entirely deterministic in the sense that, given the position of all the spheres at some instant of time, the configuration at later, or even earlier, times can be computed exactly via the solution of this equation.

When the number of particles is infinite, however, the straightforward determination of $\mathbf{U}_i$ fails because the sum of the weak contributions to $\mathbf{U}_i$ from the infinitely many distant particles at "infinity" diverges owing to the slow algebraic decay of any velocity disturbance in Stokes flow. Thus, one replaces these distant particles by an effective medium containing a random distribution of particles of given concentration whose hydrodynamic influence is felt only via their contribution to the effective viscosity of this effective medium. In other words, in calculating the velocity of a given particle, say the black particle in the figure, one supposes that this particle is surrounded, as before, by a finite number of (white) particles but that, in addition, the whole set is immersed into an effective fluid extending to infinity. The evolution of the configuration then proceeds in a deterministic fashion via the solution of the Smolucovski equation with the particle velocities $\mathbf{U}_i$ calculated in the manner discussed above, but only for small times, i.e. only as long as the marked particle is surrounded by its original neighbors. For later times, however, the marked particles is surrounded by new neighbors which were originally located within the effective medium and whose original position was unspecified. Thus, the motion of the marked particle acquires a random component which means that its location can no longer be determined exactly but can only be represented via a probability density function $\mathbf{p(r,\tau)}$ given by the solution of the Fokker-Planck equation also shown in the figure. In addition, when the bulk properties of the suspension, such as the bulk shear rate or the particle concentration, vary over distances which are much larger than the particle radius a, one can obtain the evolution equation for the particle concentration ϕ, simply by replacing p by ϕ in the Fokker-Planck equation referred to above.

The quantity $\mathbf{D(r)}$ is the particle tracer diffusivity, while $\mathbf{V(r)}$ is the mean velocity of a representative particle. The latter consists of two parts: the bulk velocity $\mathbf{U(r)}$ of the suspension viewed as an effective medium, and a "drift" particle velocity $\mathbf{V^*(r)}$ relative to $\mathbf{U(r)}$. This drift velocity vanishes of course for neutrally buoyant particles in a simple shear flow when the concentration ϕ and the shear rate γ are constant because, under these conditions, there exists no preferential direction for particles to migrate across the streamlines of the bulk flow. In the presence of a macroscopic concentration gradient, however, a given particle will be displaced by its neighbors more frequently on one side than on the other and, hence, will tend to "drift" towards the region of lower concentration. A similar drift will occur in the presence of a gradient in the shear rate or in the shear stress.

As will be seen presently, the existence of such a drift velocity is primarily responsible for the various phenomena referred to above which arise when concentrated suspensions are subjected to shear.

EXPERIMENTAL MEASUREMENTS

In principle, the particle tracer diffusivity **D(r)**, which, from scaling arguments, is proportional to the product of the local shear rate times the square of the particle radius, should be easy to measure by following the motion of an individual tagged particle in a suspension of uniform concentration and shear and then computing its mean square displacement. This was the technique originally used [2] which has since been modified and rendered more accurate [3,4]. The particle tracer diffusivity can also be computed via the so-called "Stokesian dynamics" calculations [5]. Finally, analytic expressions for the transverse components of **D(r)** for smooth equi-sized spheres along and normal to the plane of shear in a simple shear flow were recently derived [6] using the computed trajectories of triads of interacting particles.

In contrast to **D(r)**, which can be measured directly, the drift velocity $\mathbf{V}^{*}(\mathbf{r})$ has, to-date, been determined only indirectly, specifically by matching experimental results to model equations. For example, in the presence of a macroscopic concentration gradient, say $d\phi/dy$, in a simple shear flow along x, the component of $\mathbf{V}^{*}$ the y-direction is proportional to $d\phi/dy$, with the constant of proportionality being minus the so-called gradient diffusivity, which also scales as the produce of the local shear rate and the square of the particle radius. The component of this gradient diffusivity normal to the plane of shear was then determined from the observed long-term decrease in the effective suspension viscosity as measured in a Couette device, by fitting the data using the solution of a one-dimensional unsteady-state diffusion equation with the gradient diffusivity being an adjustable parameter [7]. Similarly, the component of the gradient diffusivity along the plane of shear, as well as the diffusivity which enters into the expression relating $\mathbf{V}^{*}$ to the gradient in the shear rate, were determined from the observed short-term increase in the effective suspension viscosity when the Couette device was activated after loading the sample [7].

Thus, reliable values for all of these diffusivities, as functions ϕ, currently exist for monodisperse suspensions of solid spheres which can be used for modeling purposes.

COMPARISONS OF MODELING CALCULATIONS WITH EXPERIMENTS

We summarize below a few of the many cases in which model calculations using no adjustable parameters have been successfully compared quantitatively with experimental observations.

1. Thirty years ago, Karnis, Goldsmith and Mason [8] measured, at the end of a long circular tube, the particle velocities in flowing suspensions of monodisperse neutrally buoyant solid

spheres and reported that the profile was blunted at the center rather than parabolic as in pure viscous fluids. Moreover, for fixed tube dimensions, the bluntness was found to increase with an increase in ϕ and with the particle radius a. This is consistent with the premise that particles diffused from regions of high shear, i.e. the wall, to the centerline and the fact that the diffusivity is a monotonically increasing function of ϕ and is proportional to a^2 [7]. Indeed, model calculations [9] were found to be in very good agreement with the experimental results referred to above [8].

2. When a settled suspension of heavy particles with a clear fluid layer above it is placed in a Couette device which is then turned on, the bed expands and the suspension is observed to flow [10, 11]. The height to which the suspension rises can then be calculated via the solution of model equations and excellent agreement is found between the results, as predicted from the model calculations, and the experimental measurements [11]. A similar agreement exists under transient conditions [8, 12].

3. When heavy particles in a suspension sediment under gravity in a settler having inclined walls, the dense concentrated sediment that overlays the upward facing wall is able to flow freely only because the upward shear-induced particle flux due to gradients in the particle concentration and in the shear stress balance the downward gravitational flux. The theoretically determined particle velocity profile within this sediment layer as well as the local sediment layer thickness were found to be in excellent agreement with experimental measurements [13].

4. The model equations developed to-date apply only to monodisperse suspensions under conditions where the shear flow is laminar and either uni- or quasi-unidirectional. An attempt was recently made to extend the applicability of these equations to fully three-dimensional flows by simply replacing the local shear rate γ by the second invariant of the deformation tensor [14]. Although this is, admittedly, a very crude step, the resulting theoretical predictions [14] for the concentration and particle velocity profiles in a tube were quantitatively consistent with experimental results reported earlier [15].

The above are but a few examples of suspension flows where shear-induced particle diffusion manifests itself in a crucial way.

ACKNOWLEDGEMENT

My work on suspension rheology was supported by the U.S. Department of Energy and is currently being performed in collaboration with B. Khusid, R. Mauri, A. Tripathi and Y. Wang.

REFERENCES

1. A. ACRIVOS, "Bingham Award Lecture - 1994; Shear-Induced Particle Diffusion in Concentrated Suspensions of Noncolloidal Particles," *J. Rheol.* 39, 813 (1994).

2. E.C. ECKSTEIN, D.G. BAILEY and A.H. SHAPIRO, "Self-Diffusion of Particles in Shear Flow of a Suspension," *J. Fluid Mech.* 79, 191-208 (1977).

3. D. LEIGHTON and A. ACRIVOS, "Measurement of the Shear-Induced Coefficient of Self-Diffusion," *J. Fluid Mech.* 177, 109-131 (1987a).

4. S.E. PHAN and D. LEIGHTON, "Measurement of the Shear-Induced Tracer Diffusivity in Concentrated Suspensions," *J. Fluid Mech.* (submitted, 1993).

5. G. BOSSIS and J.F. BRADY, "Self-Diffusion of Brownian Particles in Concentrated Suspensions Under Shear," *J. Chem. Phys.* 87, 5437 (1987).

6. Y. WANG, R. MAURI and A. ACRIVOS, "Transverse Shear-Induced Liquid and Particle Tracer Diffusivities in a Dilute Suspension of Spheres Undergoing a Simple Shear Flow," *J. Fluid Mech.* (to appear, 1996).

7. D. LEIGHTON and A. ACRIVOS, "The Shear-Induced Migration of Particles in Concentrated Suspensions," *J. Fluid Mech.* 181, 415-439 (1987b).

8. A. KARNIS, H.L. GOLDSMITH and S.G. MASON, "The Kinetics of Flowing Dispersions. I. Concentrated Suspensions of Rigid Particles," *J. Colloid Interface Sci.* 22, 531-553 (1966).

9. R.J. PHILLIPS, R.C. ARMSTRONG, R.A. BROWN, A.L. GRAHAM and J.R. ABBOTT, "A Constitutive Equation for Concentrated Suspensions that Accounts for Shear-Induced Particle Migration," *Phys. Fluids A* 4, 30-40 (1991).

10. D. LEIGHTON and A. ACRIVOS, "Viscous Resuspension," *Chem. Eng. Sci.* 41, 1377-1384 (1986).

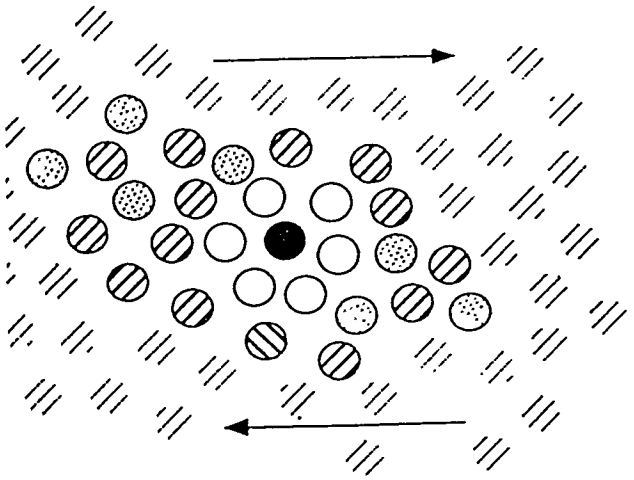
11. A. ACRIVOS, R. MAURI and X. FAN, "Shear-Induced Resuspension in a Couette Device," *Int. J. Multiphase Flow* 19, 797-802 (1993).

12. A. ACRIVOS, X. FAN and R. MAURI, "On the Measurement of the Relative Viscosity of Suspensions," *J. Rheol.* 38, 1285-1296 (1994).

13. B. KAPOOR and A. ACRIVOS, "Sedimentation and Sediment Flow in Settling Tanks with Inclined Walls," *J. Fluid Mech.* 290, 39-66 (1995).

14. K. ZHANG and A. ACRIVOS, "Viscous Resuspensions in Fully-Developed Laminar Pipe Flows," *J. Int. Multiphase Flow* 20. 579 (1994).

15. S.A. ALTOBELLI, R.C. GIVLER and E. FUKUSHIMA, "Velocity and Concentration Measurements of Suspensions by Nuclear Magnetic Resonance," *J. Rheol.* 35, 721-734 (1991).

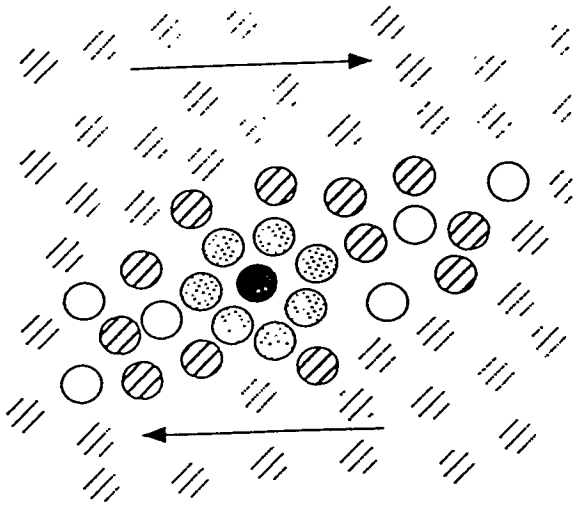


The Smoluchovski equation in the high-Peclet limit

$$\frac{\partial P}{\partial \tau} + \sum_{i=1}^{N} \nabla_i (\mathbf{U}_i P) = 0$$

$$\frac{d\mathbf{r}_i}{d\tau} = \mathbf{U}_i(\mathbf{C}_N, \tau)$$

τ is the microscopic time



The Fokker-Planck equation

$$\frac{\partial p}{\partial t} + \nabla(\mathbf{V}p) = \nabla^2(\mathbf{D}p)$$

$$\mathbf{V} = \lim_{\tau \to \infty} \frac{\langle \Delta \mathbf{R}(\tau) \rangle}{\tau}$$

$$\mathbf{D} = \lim_{\tau \to \infty} \frac{\langle \Delta \mathbf{R}(\tau) \Delta \mathbf{R}(\tau) \rangle}{2\tau}$$

NEW DEVELOPMENTS IN THE THEORY OF FLAME PROPAGATION

Gregory I. Sivashinsky

The Benjamin Levich Institute for
Physico-Chemical Hydrodynamics
The City College of the City University of New York
New York, N.Y. 10031

ABSTRACT

Two topics in combustion fluid mechanics are discussed. The first is a theory of the outward propagating spherical flame in the regime of well-developed hydrodynamic instability. In a qualitative agreement with experimental observations it is shown that the flame assumes a fractal-like wrinkled structure resulting in the overall burning rate acceleration. In contrast to hydrodynamically unstable flames, the expanding flame subject exclusively to the effect of diffusive instability does not indicate any disposition toward acceleration.

The second topic concerns the dynamics of diffusively unstable flames subjected to radiative heat losses. At high enough heat losses the flame breaks up into separate self-propagating cap-like flamelets while a significant portion of the fuel remains unconsumed.

INTRODUCTION

A spherical flame spreading out from an ignition source is one of the most basic configurations of premixed combustion. While such flames are quite feasible in the laboratory, under certain conditions a nominally spherical flame becomes unstable and displays an irregular pattern of wrinkles. As is now well established, there are two principal mechanisms for the intrinsic flame instability: (i) thermal expansion of the burnt gas and (ii) high mobility of the deficient reactant (e.g. [1]). The first, the so-called hydrodynamic or Darrieus-Landau mode of instability, is an invariable feature of any exothermic premixed gas flame. On the other hand, the occurrence of the second, the diffusive mode of instability, clearly depends on the composition of the mixture.

The outward propagating spherical flame in the regime of well-developed

hydrodynamic instability assumes a multiple-scale pebbly structure. To observe such a configuration the aspect ratio of the system should be rather large. For conventional hydrocarbon-air mixtures under normal pressure this would require the flame to be of several meters in diameter. In relatively small-scale systems the hydrodynamically unstable flames are either completely smooth or exhibit a few wide-spaced ridges that are well maintained even under the deformation and extension of the flame. Unlike the former, the diffusive mode of instability manifests itself in the emergence of the small-scale irregularly recombining cellular structure and therefore relatively easily produced under normal laboratory conditions.

In recent years new aspects of hydrodynamic and diffusive instabilities have been revealed. It was observed that in the regime of well-developed hydrodynamic instability, the average radius, $\overline{R}$, of the large-scale wrinkled flame moving into an initially quiescent homogeneous premixture grows as $\overline{R} \sim t^{3/2}$ [2]. This striking effect implies, that the wrinkled sphere surface area grows as $\overline{R}^{7/3}$, i.e. faster than $\overline{R}^2$, that may well be regarded as the self-fractalization of the flame interface with 7/3 being its fractal dimension. The first-principle description of the phenomenon is one of the topics of the present progress report. Another novel effect concerns the diffusive instability of near-limit low-Lewis-number flames. It was observed that in mixtures of very weak reactivity the point ignition leads to an outward propagating cellular flame with rapidly separating cells resulting in the flame self-fragmentation [3]. In some circumstances the fragments close up upon themselves to form stationary spherical structures called the flame-balls. An equilibrium theory of the flame-balls was proposed in [4]. The present report deals with the dynamics of their formation.

SELF-FRACTALIZATION OF HYDRODYNAMICALLY UNSTABLE FLAMES

To describe an outward propagating hydrodynamically unstable flame the following weakly nonlinear evolution equation was employed [5],

$$\frac{\partial R}{\partial t} - \frac{U_b}{2\overline{R}^2}\left(\frac{\partial R}{\partial \theta}\right)^2 = \frac{D_m}{\overline{R}^2}\frac{\partial^2 R}{\partial \theta^2} + \frac{\gamma U_b}{2\overline{R}} I\{R\} + U_b \qquad (1)$$

where $0 < \theta < 2\pi$ and $I\{R\}$, $\overline{R}$ are defined as

$$I\{R\} = \frac{1}{\pi}\sum_{n=1}^{\infty} n \int_0^{2\pi} \cos\left[n(\theta - \theta^*)\right] R(\theta^*, t) d\theta^*, \quad \overline{R} = \frac{1}{2\pi}\int_0^{2\pi} R(\theta, t) d\theta \ .$$

Here $r = R(\theta, t)$ is the interface of the outward propagating flame; U_b – speed of a planar flame relative to the burned gas; D_m – Markstein diffusivity; D_{th} – thermal diffusivity of the mixture; $\alpha = \frac{1}{2}\beta(Le^{-1} - 1)$ – Zeldovich number; Le – Lewis number; γ – thermal expansion coefficient ($\gamma < 1$).

In the limit of weak thermal expansion ($\gamma \ll 1$), Eq. (1) is an exact asymptotics provided R, θ, t are appropriately scaled.

To ensure the well-posedness of the associated initial value problem, Eq. (1) should provide dissipation of the short-wavelength disturbances. Hence, D_m should be positive, that pertains to high-Lewis-number premixtures. The numerical simulation of Eq. (1) yields the following picture of the flame evolution. At the beginning of the process the expanding interface exhibits one or two folds and the flame evolves in a self-similar manner. Somewhat later, almost instantaneously, a considerable portion of the interface acquires a nearly periodic cellular structure. As the flame sphere grows the quasi-periodic corrugations gradually stretch while the interface becomes less regular. The sufficiently enlarged cells eventually also acquire a fine structure and so on. The whole process, thus, assumes the character of a cascade, quite in line with the general concept of the fractal curve (Fig. 1a). The most interesting feature of the system, however, is that the average speed, $\overline{R}_t$, does not stabilize but rather enjoys a noticeable amplification. As one can see from Figure 1b the $\overline{R}_t(t)$ dependence is not incompatible with the experimental $\overline{R}_t \sim t^{1/2}$ power law. Moreover, the fractal analysis of the numerical solution [5] (via the standard box counting procedure) yields the fractal dimension $D_1 \simeq 4/3$ which is quite consistent with $D_2 = D_1 + 1 = 7/3$ suggested by the experimental data [2].

Eq. (1) also transpires to be accessible to analytical explorations. Similar to Burgers' equation it admits an infinite number of exact solutions of the form [6,7],

$$R = \gamma U_b t + 2D_m U_b^{-1} \sum_{n=1}^{2N} \ln\left[\sin\tfrac{1}{2}\big(M\theta - Z_n(t)\big)\right] \tag{2}$$

where M is an integer and time-dependent Z_n's are poles of $R(\theta,t)$ in the complex plane, appearing in conjugate pairs. Their dynamics is governed by the system of ODE's

$$\frac{dZ_n}{dt} = -\frac{\gamma^2 M^2 D_m}{\overline{R}^2} \sum_{n\neq m} ctg\left(\frac{Z_n - Z_m}{2}\right) - \frac{i\gamma^3 M U_b}{2\overline{R}} sign(ImZ_n)\ . \tag{3}$$

For a pole solution the number of flame cusps never exceeds MN. Moreover, $\overline{R}_t \to U_b$ as $t \to \infty$. Yet, direct numerical simulations of the original Eq. (1) show that the flame does accelerate and the number of cusps constantly grows as the flame expands. Comparing the flame dynamics governed by Eq. (1) with the exact pole solutions, the following picture of the flame self-fractalization has been revealed [7]. The flame dynamics evolves through successive instabilities and births of poles where the flame closely follows a $2N$ pole solution before approaching a $2N+2$ pole solution. The process keeps repeating itself as the time increases.

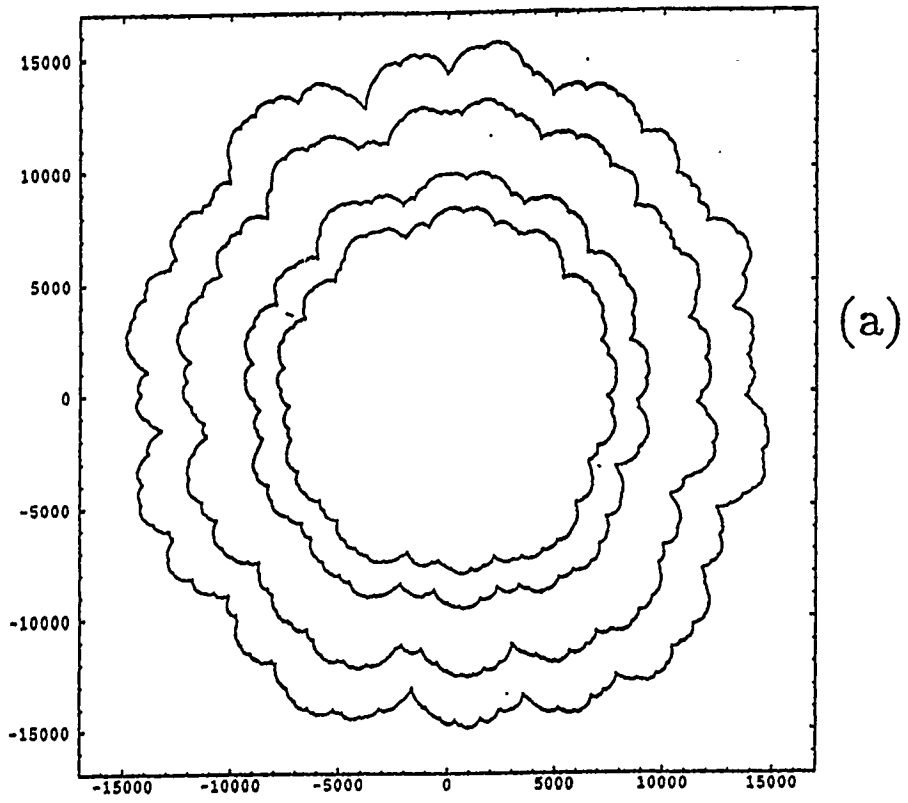
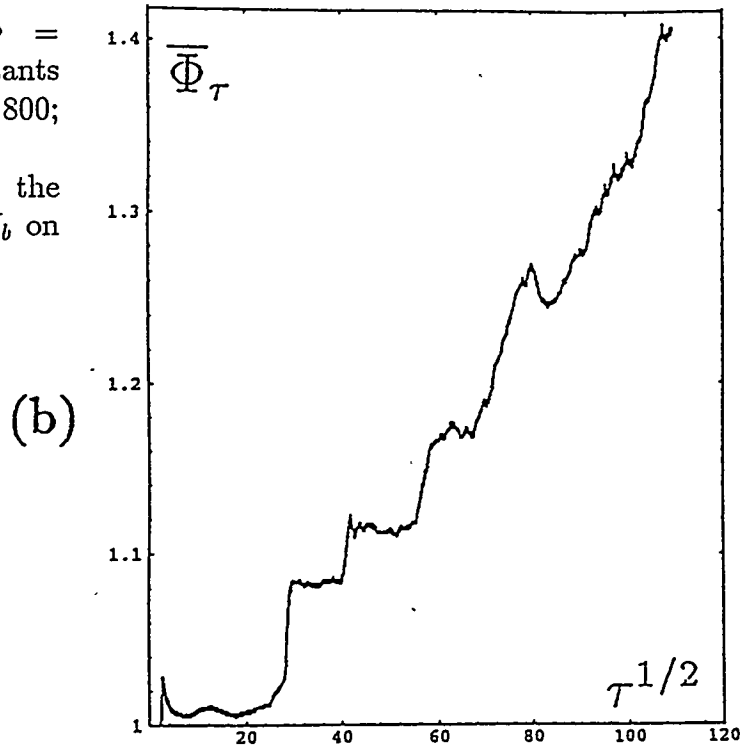


Figure 1. Numerical simulation of Eq. (1) for $\gamma = 0.8$ [5].
(a) – scaled flame interface $\Phi = U_b R/D_m$ at four consecutive instants of time $\tau = \gamma^2 U_b^2 t/D_m = 6,800$; 8,000; 10,000; 12,000.
(b) – temporal dependence of the scaled flame speed $\overline{\Phi}_\tau = \overline{R}_t/\gamma^2 U_b$ on $\tau^{1/2}$.



If Markstein's diffusivity, D_m, is negative, which occurs in low-Lewis-number premixtures, the earlier mentioned diffusive mechanism of instability enters the play. In this case dissipation of small-scale disturbances is provided by the new term $\sim -\overline{R}^{-4} R_{\theta\theta\theta\theta}$ which should be added to the r.h.s. of Eq. (1). In relatively small-scale flames the diffusive instability may dominate over the hydrodynamic one whose impact on the overall flame dynamics may be disregarded. The pertinent evolution equation for the outward propagating wrinkled flame reads,

$$\frac{\partial R}{\partial t} = \frac{U_b}{2\overline{R}^2}\left(\frac{\partial R}{\partial \theta}\right)^2 + \frac{D_m}{\overline{R}^2}\frac{\partial^2 R}{\partial \theta^2} - \frac{4D_{th}\ell_{th}}{\overline{R}^4}\frac{\partial^4 R}{\partial \theta^4} + U_b \ . \tag{4}$$

Here $\ell_{th} = D_{th}/U_b$ is the thermal width of the flame. Similar to the hydrodynamic case the diffusive instability leads to the flame wrinkling which in turn results in the enhancement of the effective flame speed, $\overline{R}_t$. However, in contrast to the situation with hydrodynamic instability, here $\overline{R}_t$ rather rapidly comes to saturation. Thus, the well-developed wrinkled flame does not accelerate. Such an outcome is apparently due to the fact that in the diffusive case the wrinkling occurs not as a multiple-scale cascade but rather as a generation of fixed-size cells, whose width is entirely controlled by the small perturbations maximum growth rate. In sufficiently large-scale systems the diffusive instability will clearly interact with the hydrodynamic one which may result in the flame acceleration similar to that occuring for Eq. (1) with positive D_m.

SELF-FRAGMENTATION OF DIFFUSIVELY UNSTABLE FLAMES

The morphology of the flame fragmentation clearly cannot be described within the weakly non-linear model such as Eq. (1), and requires a more general coordinate-free approach. The systematic derivation of the pertinent reduced equation is a difficult problem still awaiting a solution. Yet, it appears that near the planar flame quenching point certain aspects of such a model may be captured through the following semi-phenomenological reasoning. Near the quenching point the dispersion relations of the diffusive and hydrodynamic instabilities are known to be formally identical. The hydrodynamic problem allows for a rather well founded coordinate-free reduced equation [8]. Hence, by the appropriate redefining of the parameters one may try to apply this equation to the near-limit diffusively unstable flames as well, where it reads [9],

$$\mathbf{n}\cdot\frac{d\mathbf{r}}{dt} = -U_q + D_{th}K + \tfrac{1}{2}\gamma_{eff}U_q\left(1 + \frac{1}{\pi}\int_S \frac{(\mathbf{r}-\mathbf{s})\cdot\mathbf{n}}{|\mathbf{r}-\mathbf{s}|^2}dS\right) \ . \tag{5}$$

Here $\mathbf{r}$ and $\mathbf{s}$ are the points on the flame interface, $\mathbf{n}$ is the normal directed to the burned gas at the point $\mathbf{r}$, $K = -\nabla\cdot\mathbf{n}$ is the flame curvature, $U_q = U_b/\sqrt{e}$ is the flame speed at the quenching point: $\gamma_{eff} = \sqrt{4\alpha/\alpha + 6}$. For the hydrodynamically unstable flame $U_q, \gamma_{eff}, D_{th}$ should be replaced by U_b, γ, D_m, respectively. While the thermal expansion parameter γ never exceeds unity, its counterpart γ_{eff} may come rather close to 2 provided $\alpha = \frac{1}{2}\beta(Le^{-1} - 1)$ is large enough. This is easily achieved for low-Lewis-number premixtures. Numerical simulations of Eq. (4) show that at $\gamma_{eff} < 1$ it produces

a wrinkled flame similar to that of Eq. (1) (Fig. 1a). However, at $\gamma_{eff} > 1$ the flame evolution occurs as a fingering instability that may well be regarded as the incipient stage of the flame self-fragmentation (Fig. 2).

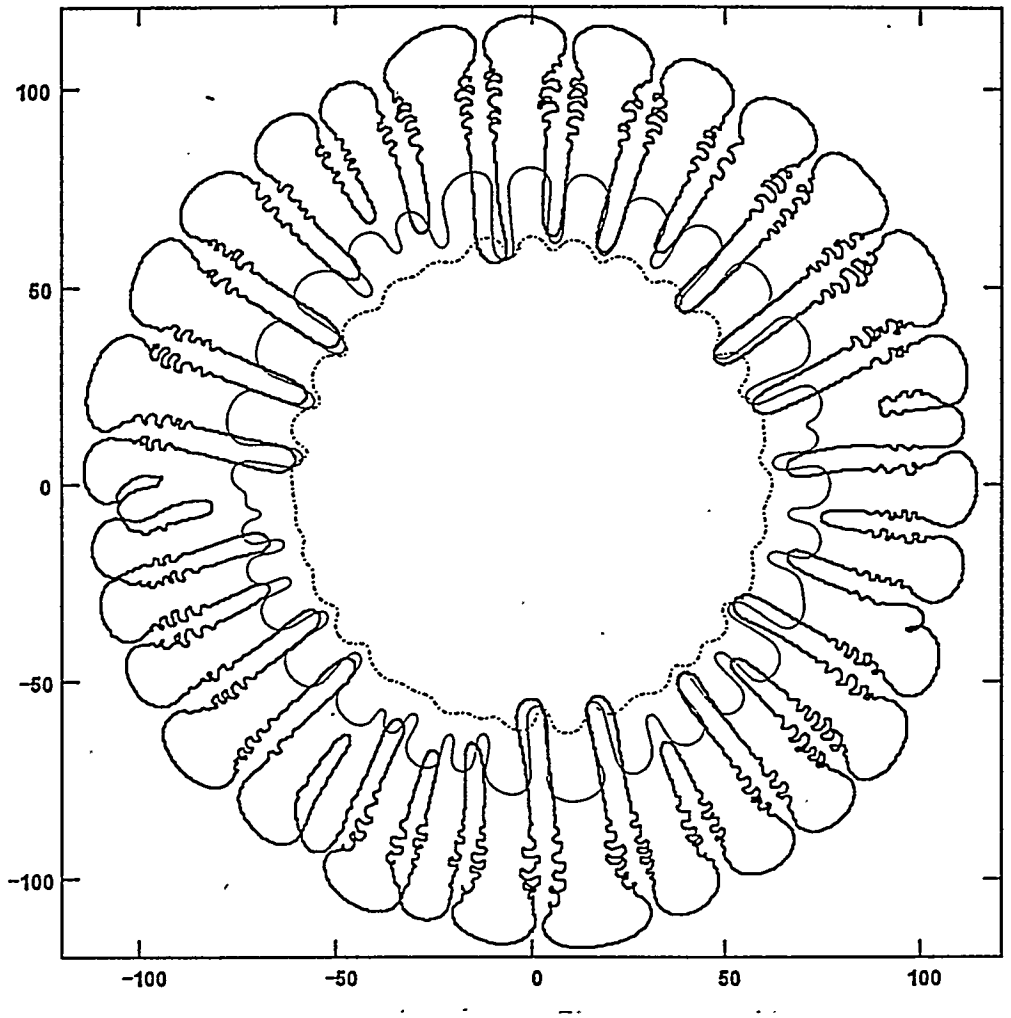


Figure 2. Numerical simulation of Eq. (5) for a near-limit diffusively unstable flame at $\gamma_{eff} = 1.5$. Shown are flame configurations at three consecutive instants of time [9].

Apart from exploring the reduced equation (5) we also undertook a direct numerical simulation of the pertinent reaction-diffusion system based on a finite rate Arrhenius kinetics [10]. Figure 3 shows some of the results obtained. The emerging cap-like flamelets appear as localized solitary waves spreading through the reactive premixture and leaving most of it unconsumed. At strong enough heat losses this combustion mode also becomes unfeasible resulting in a total suppression of the flame.

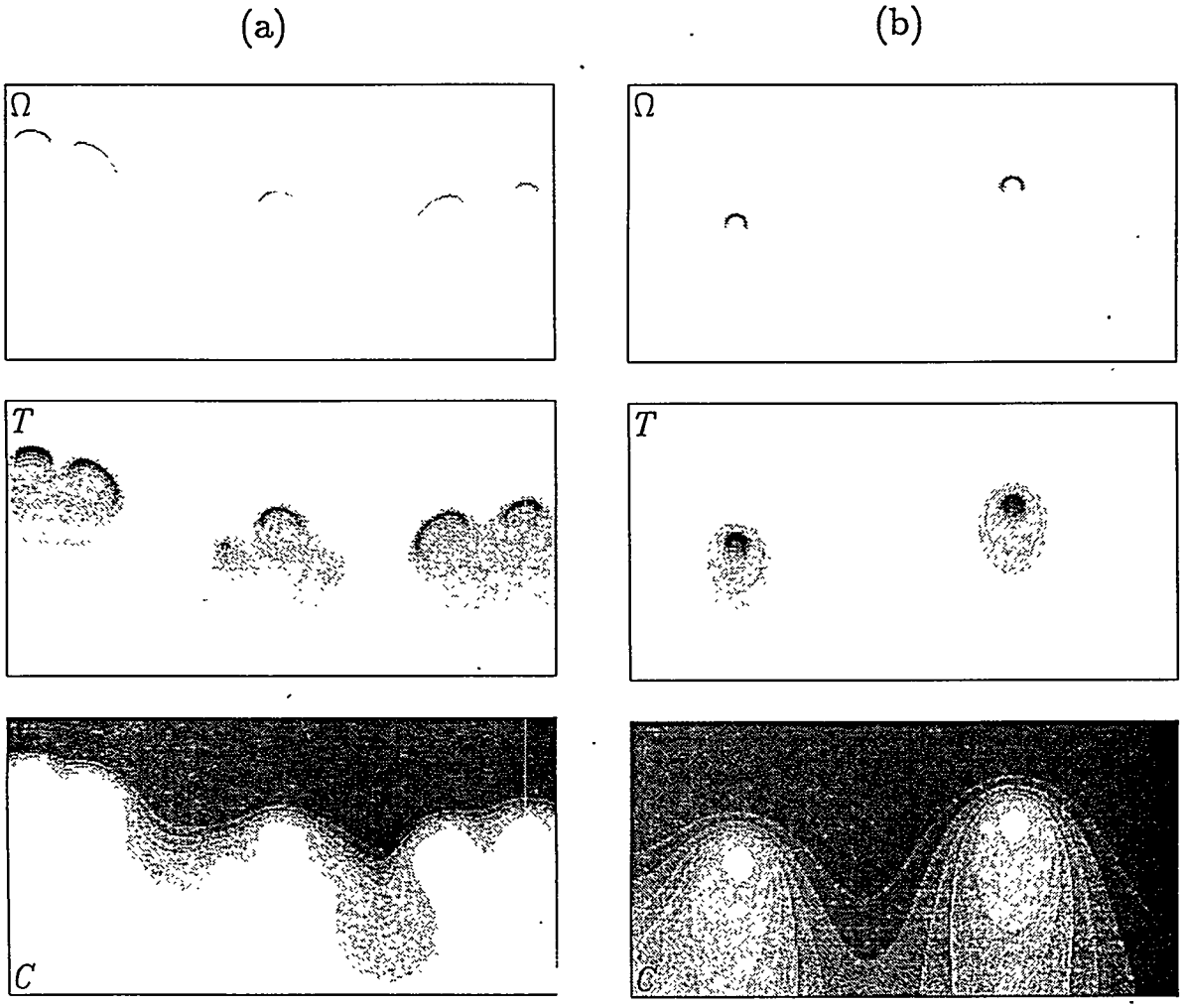


Figure 3. Reaction rate (Ω), temperature (T), and deficient reactant concentration (C) distributions in low-Lewis-number premixed flames. (a) – moderately nonadiabatic case, (b) – strongly nonadiabatic case [10].

ACKNOWLEDGEMENT

These studies were supported by the U.S. Department of Energy under Grant No. DE-FG02-88ER13822.

REFERENCES

1. G.I. SIVASHINSKY, *Phil. Trans. R. Soc. Lond. A* **49**, 332 (1990).

2. Y.A. GOSTINTSEV, A.G. ISTRATOV and Y.V. SHULENIN, *Combus. Expl. Shock Waves* **24**, 70 (1988).

3. P.D. RONNEY, *Combust. Flame* **62**, 121 (1990).

4. J. BUCKMASTER, G. JOULIN and P. RONNEY, *Combust. Flame* **70**, 381 (1990).

5. L. FILYAND, G.I. SIVASHINSKY and M.L. FRANKEL, *Physica D* **72**, 110 (1994).

6. G. JOULIN, *Phys. Rev. E* **50**, 2030 (1994).

7. M. RAHIBE, N. AUBRY, G.I. SIVASHINSKY and R. LIMA, *Phys. Rev. E* **52**, 3674 (1995).

8. M.L. FRANKEL, *Phys. Fluids A* **2**(10), 1879 (1990).

9. M.L. FRANKEL and G.I. SIVASHINSKY, *Phys. Rev. E* **52**, 6154 (1995).

10. L. KAGAN and G.I. SIVASHINSKY, *Combust. Flame* (1996), in press.

ONE-DIMENSIONAL TURBULENCE

Alan R. Kerstein

Sandia National Laboratories
Livermore, CA 94551-0969

ABSTRACT

One-Dimensional Turbulence is a new turbulence modeling strategy involving an unsteady simulation implemented in one spatial dimension. In one dimension, fine scale viscous and molecular-diffusive processes can be resolved affordably in simulations at high turbulence intensity. The mechanistic distinction between advective and molecular processes is thereby preserved, in contrast to turbulence models presently employed. A stochastic process consisting of mapping 'events' applied to a one-dimensional velocity profile represents turbulent advection. The local event rate for given eddy size is proportional to the velocity difference across the eddy. These properties cause an imposed shear to induce an eddy cascade analogous in many respects to the eddy cascade in turbulent flow. Many scaling and fluctuation properties of self-preserving flows, and of passive scalars introduced into these flows, are reproduced.

INTRODUCTION

Many aspects of turbulent flow, and of physical and chemical processes within turbulent flow, can be captured only in a fully resolved, unsteady simulation. To address these aspects by a more economical method than direct numerical simulation (DNS), one possible strategy is to use a model with reduced spatial dimensionality. Several two-dimensional (2D) formulations have been employed for this purpose, such as 2D Navier-Stokes simulations, discrete-vortex methods, and simulations involving synthetic 2D velocity fields. Further reductions in computational cost can be achieved by adopting a 1D formulation.

1D turbulence models are often used to represent turbulent transport, in particular, vertical transport in geophysical flows. These models are neither fully resolved nor, in most cases, unsteady. They typically incorporate the average effect of fine-scale unsteady processes by means of empirical parametrizations.

Here, a new approach [1] involving fully resolved, unsteady simulation on a 1D domain is outlined. The approach, denoted 'One-Dimensional Turbulence' (ODT), is intended to extend the scope of computationally accessible turbulent flow phenomena.

MODEL FORMULATION

The foundation of the model is the recognition that the key mechanisms of flow modification by a turbulent eddy, compressional strain and rotational folding, can be represented in one dimension.

This observation motivated the formulation of the linear-eddy model (LEM) [2], a turbulent mixing model that is the antecedent of ODT.

ODT is formulated as follows. Specializing to boundary-layer type flows for clarity, the computational domain y represents the transverse coordinate. The fluid state is represented by the transverse profile of streamwise velocity $u(y,t)$, the kinematic viscosity ν, any advected scalar profiles $\theta(y,t)$, and their corresponding molecular transport coefficients. The formulation encompasses spatial development parametrized by (y,x), where x is the streamwise coordinate, as well as temporal development parametrized by (y,t). For present purposes, consider the latter.

Viscous evolution is implemented deterministically, governed by equations of conventional form. This implementation is mechanistically literal because the viscous scales are fully resolved in the computation. Advection by a single turbulent eddy is represented by the triplet map, an instantaneous transformation of a segment of the computational domain. The mapping rule is illustrated in Fig. 1. The straining and folding properties of this map mimic the corresponding attributes of turbulent eddies [2]. The affected segment is denoted $[y_0, y_0 + l]$, where y_0 and the segment size l are randomly selected. The space-time-size sequence of mappings is a stochastic process governed by the rate distribution

$$\lambda(l; y_0, t) = \frac{1}{l^2 \tau(l; y_0, t)}. \tag{1}$$

Here, τ is the eddy time scale

$$\tau(l; y_0, t) = \frac{l}{A \Delta u}, \tag{2}$$

where Δu is some measure of the velocity difference across $[y_0, y_0 + l]$. Here, Δu is taken to be the difference of velocities averaged over the intervals $[y_0, y_0 + l/2]$ and $[y_0 + l/2, y_0 + l]$, respectively. Other reasonable definitions are possible. In most cases, computed results are found to be insensitive to the precise definition.

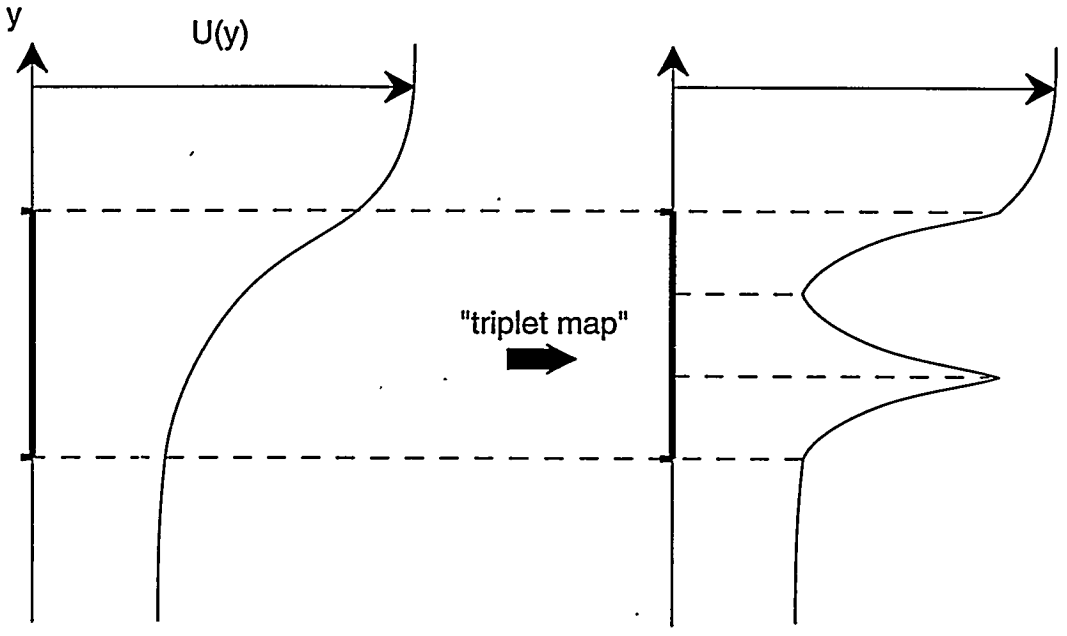


Figure 1. Application of the triplet map to a designated interval of the transverse profile of streamwise velocity. The map replaces the profile in the segment by three copies, each compressed by a factor of three, with the middle copy inverted.

Equations (1) and (2) provide a simple, dimensionally consistent prescription of the ensemble of mappings, without introducing extrinsic quantities. A is the only free parameter in the present formulation. Work in progress involves generalization of Eq. (2) to account for buoyancy effects in density-stratified flows.

Scalar fields $\theta(y,t)$ can be introduced. They are subject to molecular transport, implemented deterministically in a conventional manner, and to advection by mappings.

RELATION TO 3D FLOW

Operationally, ODT is a numerical method for generating realizations of a class of stochastic initial-boundary-value problems on a 1D domain. The initial and boundary conditions can be chosen to represent homogeneous turbulence, free shear flow, or wall-bounded flow. A variety of such flows have been simulated, and results have been compared to measurements.

Representative cases are discussed shortly. First, the interpretation of ODT in the context of 3D flow is considered.

For spatially developing flow, viscous evolution is represented by the usual boundary-layer equations. In the absence of the mapping process, the model reproduces the well known laminar solutions for the planar boundary layer, shear layer, jet, etc. These solutions are deterministic, so all simulated realizations of a given flow are identical, and no fluctuations are predicted.

The stochastic mapping process introduces fluctuations that can be characterized statistically using the methods generally applied to an ensemble of measured or computer-simulated flow realizations. Single-point velocity and scalar statistics of any order can be extracted, and multipoint statistics such as spectra and spatial correlations can be obtained. Lagrangian, fractal, and wavelet analyses are applicable.

In particular, the Reynolds stress component $\langle u'v' \rangle$ can be obtained although motions along the computational domain are governed by a sequence of maps rather than a velocity field $v(y)$. To see this, consider the interpretation of $\langle \theta'v' \rangle$, where θ is any fluid property and θ' is the residual after subtraction of the ensemble mean value at a given y location. Then $\langle \theta'v' \rangle$ is the flux of θ induced by velocity fluctuations v' at that location. In ODT, this flux is determined by monitoring mapping-induced transfers of θ across any y level of interest. This definition is applicable not only to the Reynolds stress ($\theta = u$) but also to fluxes of density, temperature, chemical composition, etc.

An important aspect of the relation between ODT and 3D flow is energy transfer. The mapping process is measure-preserving. In the spatially discrete numerical implementation, the triplet map is a permutation of the cells of the 1D domain. Spatial profiles of fluid properties, including the streamwise velocity u, are rearranged accordingly. Therefore any spatially integrated function of u, in particular the (normalized) kinetic energy u^2 integrated over the 1D domain, is invariant under mappings.

In 3D flow, the velocity component u associated with a fluid element can change by two mechanisms, pressure effects and viscous effects. Viscous effects are represented within ODT, but pressure effects, which transfer kinetic energy among velocity components, are not. Therefore the quantity u^2 in ODT is more closely analogous, from an energetic viewpoint, to the total kinetic energy of 3D flow than to the 3D quantity u^2. Accordingly, the budget of the production, transport and dissipation of u^2 determined from ODT simulations is compared to measured budgets of kinetic energy.

These formal correspondences between ODT and 3D flow do not explain why the former should reproduce behaviors of the latter. The explanation is that ODT is formulated so as to reproduce properties of Navier-Stokes turbulence that are mandated by dimensional relations.

There are two types of dimensionally mandated properties. Microstructural properties are flow independent at high Reynolds number (Re). Salient microstructural properties are the $k^{-5/3}$ wavenumber scaling of the inertial range energy spectrum and the $Re^{3/4}$ scaling of the high-wavenumber viscous cutoff of the inertial range. Macrostructural properties characterize the overall evolution of flow structure subject to initial and boundary conditions and conservation laws. Salient examples are the self-similar decay of homogeneous turbulence, the self-similar growth of free shear flows, and the log-law regime of wall-bounded flows.

Dimensionally mandated scaling is a facet of self-similarity. The evolution of a self-similar process is governed by the length and time scales describing the state of the system at a given instant, rather than extrinsic length and time scales. The governing length and time scales, in conjunction with the boundary conditions and conservation laws, determine a unique set of dimensionally consistent scaling laws governing the evolution of the process. (If the similarity is 'incomplete' [3], initial conditions may also influence the scaling.)

The rate distribution specified by Eqs. (1) and (2) generates eddies of a given size at a rate governed by flow fluctuations on that length scale. The eddies in turn wrinkle the u profile, thereby creating fluctuations that sustain the mapping process. This feedback process may be viewed as a form of turbulence closure. In this instance, the closure involves a postulated two-way coupling between a velocity profile and a stochastic process rather than the usual closure involving a postulated relation among terms in a formal expansion of the Navier-Stokes equation.

The time scale specified by Eq. (2) is the ODT analog of the eddy turnover time. Though mappings are instantaneous, they introduce finite-time effects through their imprints on the velocity field and the influence of these imprints on the time scales governing subsequent events. These finite-time effects have two significant consequences.

First, the subsequent evolution is governed by intrinsic length and time scales, leading to self-similarity and thus to dimensionally mandated scalings. As noted, the mapping process conserves spatially integrated quantities, so the applicable conservation laws are those governing viscous evolution. The viscous equations are of conventional form, conserving momentum (momentum flux) in temporally (spatially) developing flow. Therefore the conservation laws determining the dimensionally consistent scalings are the same as in 3D flows.

Second, the finite-time effects induce an eddy cascade. The triplet map steepens velocity gradients and reduces the length scale of velocity fluctuations (see Fig. 1). Mild gradients over large length scales induce mappings that generate steeper gradients over over shorter length scales. The resulting time-scale shortening increases the frequency of subsequent smaller-scale mappings. This self-acceleration process is the ODT turbulent cascade. Because this cascade is driven by intrinsic length and time scales, and because u^2 is conserved, the dimensionally mandated scalings are the same as in the inertial range of 3D turbulent flow.

These considerations explain the performance of ODT with regard to self-similar evolution. The model can also capture transient behaviors, as illustrated shortly. This is because transients are generated by changes in boundary conditions or external inputs that are represented within the model. The length and time scales of flow response to these changes are governed by the eddy distribution within the flow at the epoch of the change (assuming, for this discussion, a sudden change). Provided that the fluctuation spectrum of the ODT u profile is a faithful rendering of the 3D spectrum, the length and time scales of the ODT transient response will emulate the 3D flow response.

It is thus plausible, though by no means guaranteed, that ODT can capture transient as well as self-similar evolution. The model is limited, of course, to flows whose symmetries (in the ensemble average) admit a low-dimensional characterization. Also, the model cannot capture pressure-dominated effects such as flow separation.

The attributes of ODT are analogous in many respects to those of mixing-length and related models. An important distinction is that dimensionally consistent relations between eddy length and time scales have heretofore been applied to the mean flow, or in more advanced models, to low-order fluctuations. In ODT, the self-consistent closure implicit in Eqs. (1) and (2) is applied at the level of the individual turbulent eddy. This approach yields a wide scope of predictive capability, including high-order fluctuations and multi-point statistics as well as mean properties.

In subsequent sections, applications to a wall-bounded flow and a free shear flow are discussed. For these and other flows involving a turbulent inner flow and a constant-velocity outer flow, the model formulation is supplemented by an 'eddy exclusion' rule. Any mapping that is more than half contained within the outer flow is disallowed. This prevents rare events much larger than the turbulent zone width. These events violate the scaling principle that the largest eddy size should be of the order of the turbulent zone width.

PLANAR BOUNDARY LAYER

The spatially developing boundary layer above a planar wall is simulated by setting the initial u profile equal to a constant value and imposing the no-slip boundary condition $u = 0$ at $y = 0$. The eddy rate, Eq. (1), is identically zero initially. However, the eddy rate distribution immediately becomes nonnull owing to velocity differences induced by the viscous evolution. The eddy time scale is much longer than the time scale for viscous dissipation of mapping-induced perturbations of the velocity profile during an initial transient period. The early development is therefore viscous-dominated. As the profile spreads, equality of these time scales is achieved in some y range owing to viscous growth of the layer, allowing a turbulent cascade to develop. The subsequent coexistence of a viscous-dominated wall layer and a nearly inviscid outer flow is evident in simulated flow realizations.

The qualitative picture is consistent with 3D evolution, though the early development lacks the laminar instability mechanisms that govern the transition to turbulence in the 3D boundary layer. Quantitative comparisons are meaningful only in the fully developed turbulent boundary layer.

The free parameter A in ODT is assigned the value 0.23 to obtain the best fit to the measured dependence of the friction coefficient on streamwise distance. The functional form of this dependence, and all other quantities of interest, are then determined from the simulations with no further empirical input.

Quantitatively accurate predictions of boundary-layer width and the shape factor as a function of streamwise distance are obtained. ODT exhibits log layers that collapse in wall-scaled coordinates, with a Von Kármán constant of 0.25, versus the measured value 0.41. Velocity-defect scaling of the outer flow is obtained. Second and third order fluctuation statistics statistics have been examined, namely the streamwise velocity variance and skewness, the Reynolds stress, and the turbulent kinetic energy budget. Qualitative features are in good overall conformance with measurements, and reasonable quantitative agreement is obtained.

PLANAR JET

The planar jet is simulated by assigning a top-hat initial u profile. The similarity scalings of this flow are reproduced, and transverse profiles of mean and fluctuation properties through third order are in reasonable agreement with measurements. A is assigned the value 0.14 to match the measured spreading rate.

To illustrate transient relaxation within ODT, model predictions are compared to measurements of passive scalar mixing downstream of pairs of coaxial, coplanar ring sources in a turbulent

round air jet [4]. Rings of six different radii were employed in the experiment, the largest radius being 0.71 times the half-width at half-maximum of the mean velocity profile in the plane of the rings. Labelling the rings in order of increasing radius, their radii were chosen in the proportion $1 : \frac{3}{2} : 2 : 3 : \frac{7}{2} : 4$. In successive runs, the rings were heated individually or in pairs, and the data were combined so that the two rings of a pair could be interpreted as sources of two distinct passive scalar species. The concentration covariance of these two species as a function of radial offset r and distance downstream of the source plane was deduced from the measurements. Here, only centerline results are considered.

Using ODT, the planar-jet analog of this mixing process is simulated. (The round jet can also be simulated using ODT, but this is not implemented for reasons discussed elsewhere [1].) It is reasonable to simulate the round-jet mixing process using a planar-jet simulation, despite the differences between the similarity scalings of these flows, provided that the streamwise development is parametrized by the normalized convective time [4] $t/\tau \equiv \int_{x_{\text{ring}}}^{x} \mathrm{d}x \, \langle u'^2 \rangle^{1/2} / (l_h u_m)$. Here, l_h and u_m are the jet half-width and the mean centerline velocity.

Spanwise line sources of a passive scalar ($Pr = 0.7$) are placed at a streamwise location x_{ring} within the self-preserving region of the simulated flow. (In the experiment, the sources were placed upstream of the self-preserving region, potentially complicating the interpretation of the results.) The transverse offsets of the line sources, scaled by the jet width, are set equal to the scaled radial offsets in the experiment.

Figure 2 shows measured and computed profiles of the segregation parameter α, defined as the scalar covariance divided by the product of scalar means. Measurements are plotted with the convective time scale stretched by a factor of two relative to the ODT convective time scale. The plot indicates that functional dependences are well predicted by ODT, but the simulated mixing process evolves roughly half as fast as indicated by the measurements.

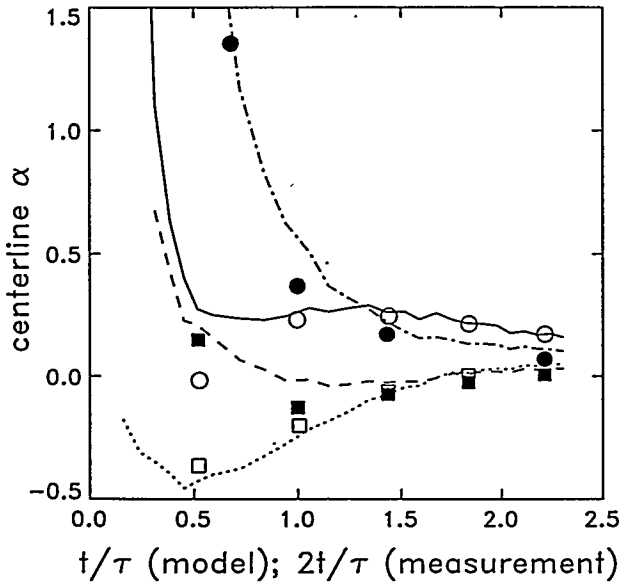


Figure 2. Centerline evolution of the segregation parameter α versus normalized convective time t/τ. Curves: planar-jet computations. Symbols: round-jet measurements [4]. Plotted data is keyed to the scaled radii r/r_{min} of the pair of ring sources, where r is the ring radius and r_{min} is the radius of the smallest ring: ——, open circle, $r/r_{\text{min}} = (1, \frac{3}{2})$; ·····, open square, $r/r_{\text{min}} = (1, 3)$; —·—, filled circle, $r/r_{\text{min}} = (\frac{7}{2}, 4)$; – – – –, filled square, $r/r_{\text{min}} = (2, 4)$.

The implication of this discrepancy is that the predicted relationship between small-scale motions governing mixing within the jet and large-scale motions governing jet entrainment differs from the relationship found in 3D turbulence. Put another way, if A had been assigned so as to match the observed mixing rate, then the computed overall spreading rate of the jet would have been too fast. This may be attributed a deficiency of either the simulated microstructure or the simulated large-scale entrainment. The latter is suspect because it is sensitive to the eddy exclusion rule, which somewhat arbitrarily disallows eddies more than half contained in the outer flow. This rule can be (but has not yet been) generalized to include an adjustable parameter that allows large scale and small scale evolution to be reconciled.

The comparison of measured and computed results yields insight concerning the significance of a set of measurements for which no detailed explanation had previously been offered. One could not *a priori* rule out sensitivity of the measured results to organized large-scale motions, to near-field effects owing to the proximity of the scalar sources to the potential core of the jet, or to other possible influences. The simplicity of ODT, together with the data comparisons, suggests the following explanation of the observations. It is inferred [1] that the observations reflect a transition from near-field radial transport dominance to far-field axial transport dominance, in conjunction with known transient properties of three-stream turbulent mixing.

DISCUSSION

One-dimensional turbulence as formulated here is intended to identify the range of turbulence phenomena that can be captured with a minimal representation of the interplay between advective and viscous processes in turbulent flow. The essential elements of such a representation are identified as a mechanistically literal numerical implementation of viscous evolution (requiring, at a minimum, one spatial coordinate), a stochastic advection process incorporating the essential ingredients of vortical motion (compressive strain and rotational folding), and a simple dimensionally-based prescription of the ensemble of such motions. To incorporate these elements in a 1D formulation, the triplet map (Fig. 1) is adopted as a 1D representation of an individual turbulent eddy and Eqs. (1) and (2) are used to determine the eddy rate distribution as a function of the instantaneous streamwise velocity profile. Diverse phenomena are reproduced by integrating these elements into a numerical simulation in which the initial and boundary conditions corresponding to various turbulent flow configurations can be imposed.

This modeling approach has features in common with various other turbulence models. For example, the analogy between the dimensional principles embodied in ODT and those underpinning mixing-length models has been noted.

In another context, ODT may be viewed as the real-space analog of shell models [5] and other mode-coupling models formulated in Fourier spaces of reduced dimensionality. In ODT, the mapping events are the 'modes' and the streamwise velocity profile is the field that couples them.

The discreteness of the events is reminiscent of the discrete-vortex method, a 2D model that can also be rendered in 1D, though with limited applicability [6]. In ODT, however, the events are discrete in time as well as space. This formulation reflects the ephemeral nature of turbulent eddies, in contrast to vortical structures that are relatively persistent at the largest (energy-containing) and smallest (viscosity-dominated) length scales.

The nonlocality of fluid motion implicit in the mapping process is reminiscient of nonlocal transport models [7]. Those models describe the evolution of mean quantities rather than individual flow realizations. The relation between those models and averages of ODT flow realizations will be investigated in future work.

Mathematically, ODT is somewhat analogous to a deterministic sequence of 1D maps in which the structure of each map is a function of the profile to which it is applied [8]. In ODT, it is the statistical ensemble of maps, rather than map structure, that evolves.

These analogies, though suggestive, do not reflect the key features of the modeling approach. The key features are the representation of eddy creation by shear that is externally imposed and amplified by previous eddies, and the mechanistically literal implementation of molecular (viscous and diffusive) effects by maintaining full spatial resolution. By capturing the interplay of shear, vortical motion, and viscosity, the model emulates the mechanisms underlying a variety of turbulent flow phenomena.

ACKNOWLEDGEMENT

This research was supported by the Division of Engineering and Geosciences, Office of Basic Energy Sciences, U.S. Department of Energy.

REFERENCES

1. A. R. KERSTEIN, "One-Dimensional Turbulence. Part 1. Homogeneous Turbulence and Shear Flows," preprint (1995).

2. A. R. KERSTEIN, "Linear-Eddy Modeling of Turbulent Transport. Part 6. Microstructure of Diffusive Scalar Mixing Fields," *J. Fluid Mech.* 231, 361 (1991).

3. G. I. BARENBLATT, *Similarity, Self-Similarity, and Intermediate Asymptotics*, Consultant Bureau, New York (1979).

4. C. TONG and Z. WARHAFT, "Passive Scalar Dispersion and Mixing in a Turbulent Jet," *J. Fluid Mech.* 292, 1 (1995).

5. L. P. KADANOFF, "A Model of Turbulence," *Phys. Today* 48, 11 (1995).

6. P. CONSTANTIN, P. D. LAX and A. MAJDA, "A Simple One-Dimensional Model for the Three-Dimensional Vorticity Equation," *Commun. Pure Appl. Math* 38, 715 (1985).

7. R. B. STULL, "Review of Non-Local Mixing in Turbulent Atmospheres: Transilient Turbulence Theory," *Bound.-Layer Meteor.* 62, 21 (1993).

8. T. SHINBROT and J. M. OTTINO, "Maps, PDE's and Solitary Waves," *Int. J. Bifurc. Chaos* 5, 955 (1995).

COUPLED PARTICLE DISPERSION BY THREE-DIMENSIONAL VORTEX STRUCTURES

T.R. Troutt, J.N. Chung and C.T. Crowe

School of Mechanical and Materials Engineering
Washington State University

ABSTRACT

The primary objective of this research program is to obtain understanding concerning the role of three-dimensional vortex structures in the dispersion of particles and droplets in free shear flows. This research program builds on previous studies which focused on the nature of particle dispersion in large scale quasi two-dimensional vortex structures. This investigation employs time dependent experimental and numerical techniques to provide information concerning the particulate dispersion produced by three dimensional vortex structures in free shear layers. The free shear flows investigated include modified plane mixing layers, and modified plane wakes. The modifications to these flows involve slight perturbations to the initiation boundary conditions such that three-dimensional vortex structures are rapidly generated by the experimental and numerical flow fields. Recent results support the importance of these vortex structures in the particle dispersion process.

INTRODUCTION

During the past twenty years considerable research efforts have been directed towards investigating the important deterministic flow structures present in free shear flows. Initially this work concentrated on the large scale vortex structures that were primarily dependent on the geometry of the mean flow velocity gradient. For flows in which the mean velocity field is two-dimensional, such as mixing layers, wakes and axisymmetric jets, the large scale structures are typically quasi two-dimensional with instantaneous vorticity approximately aligned, in the time average vorticity direction. Large scale flow structures of this type have been documented extensively in the literature. (Ho and Huerre 1984, Browand and Troutt 1980, 1985). Although

these large scale structures are extremely important in the global development of free shear flows they do not provide a complete description of the mixing processes associated with these flows.

Significant research attention concerning free shear flows has focused recently on the nature of a deterministic three dimensional vortex structure which apparently co-exists with the large scale structures. The presence of this three dimensional structure in plane mixing layers was initially documented through experimental flow visualization studies by Breidenthal (1981), and Bernal and Roshko (1986). Later visualization studies by Jimenez *et al.* (1985) and Lasheras *et al.* (1986, 1988) examined more closely the origin and evolution of the three dimensional vortical structures. The three dimensional structures are typically aligned over much of their length in the streamwise direction. The structures are also typically referred to as braids or ribs or simply streamwise vortices. Extensive hot-wire measurements by Huang and Ho (1990) have confirmed the concentrated vortical nature of these three dimensional structures and their sensitivity to initial conditions. The measurements also demonstrated that the spanwise wavelength of the streamwise structure scales with the spacing of the large scale structures. The ratio of these two scales was found to be approximately 2/3 after initial pairing interactions were completed.

Numerical simulations of three dimensional vortex structures have been performed by Ashurst and Meiburg (1988) for a spatially developing plane mixing layer. A review concerning the fundamentals of three dimensional discrete vortex simulations is available by Leonard (1985). Metcalfe *et al.* (1987) employed spectral techniques to simulate the dynamics of three dimensional perturbations in a temporally developing mixing layer. This work involved a pseudo-spectral type approach where finite difference techniques were applied for computing non-linear terms in the governing equations. Pseudo-spectral simulations of the temporally developing plane mixing layer by Rogers and Moser (1992) investigated in more detail the complex three dimensional interactions between of the streamwise rib vortices. Later pseudo-spectral simulations by Moser and Rogers (1993) followed the three dimensional evolution of the plane mixing layer vortex structures through three pairing interactions and up to the onset of transition to turbulent flow. The role of these three dimensional vortex structures in the two-phase particle dispersion process in free shear flows is a subject that has received little attention in the past, mainly because of it complexity (Crowe, Chung, Troutt 1993).

The primary objective of this research program is to obtain understanding concerning the importance of three dimensional vortex structures in the dispersion of particles and droplets in free shear flows. The free shear flows investigated include modified plane mixing layers, and modified plane wakes. The modifications to these flows involve slight perturbations to the initiation boundary conditions such that three dimensional vortex structures will be rapidly generated by the experimental and numerical flow fields. The particulate dispersion process associated with these structures is then focused upon using advanced experimental and numerical techniques.

RESEARCH PROGRAM

This overall research program is being pursued using both experimental and numerical approaches. Recent experimental work has concentrated on characterizing and quantifying the three dimensional streamwise vortex structures in plane wakes using phase-averaged hot-wire anemometry techniques. The flow experiments have been carried out in a low turbulence shear

flow wind tunnel with test facility crosssection dimensions of 60cm by 45cm. A computer controlled system for probe positioning, active shear flow forcing and data acquisition has been developed expressly for these measurements.

The numerical techniques for simulating the free shear flow and the particle dispersion process have proceeded along two avenues. One approach has employed a time dependent three dimensional spectral technique based on the work of Metcalfe *et al.* (1987). This approach has been employed to generate three dimensional vortex structures in a temporally developing mixing layer. Initial three dimensional simulations have also been performed involving one-way coupled particle dispersion using this technique.

In addition to the one-way coupled three dimensional simulations, extensive numerical studies involving two-way coupling effects in two-dimensional wakes have also been performed. These simulations have involved two dimensional discrete vortex techniques to evaluate mass, momentum and energy transfer effects between the particulate and gas phases. Previous simulation results emphasized momentum coupling effects. The present efforts focus primarily on mass and energy coupling effects.

THREE-DIMENSIONAL NUMERICAL RESULTS

Numerical simulations of a temporally developing three dimensional mixing layer at Re=500 based on initial instability wavelength have been recently computed. The simulation is initiated from a hyperbolic tangent velocity profile with two dimensional fundamental and subharmonic perturbations. In addition an isotropic random phase three dimensional energy spectrum is initially specified. Figure 1 displays the time development of spanwise vortex structures from the three dimensional simulation. The spanwise vortex structure development for the three dimensional simulation is very similar to that observed previously from two-dimensional numerical results.

The development and persistence of three dimensional stream-wise vortex structures is displayed in Figure 2. The general character of these rib like structures is in agreement with previous experimental and numerical findings concerning these flow structures. Three dimensional particle concentration contours produced by the simulated flow are shown in Figure 3 for various Stokes number particles. Considerable order is apparent in the dispersion pattern at St=1.0.

TWO-WAY COUPLING RESULTS

Recent analytical and numerical efforts have focused on the effects of two-way mass and energy coupling produced by vaporizing liquid droplets in a heated gas flow. The analysis begins from the compressible flow continuity equation

$$\vec{\nabla}\cdot\vec{u} = \frac{S_m}{\rho} - \frac{1}{\rho}\frac{D\rho}{Dt}$$

where S_m is a mass source per unit volume produced by droplet vaporization or condensation. With the assumptions of low Mach number and constant droplet temperature the continuity equation can be rewritten as

$$\bar{\nabla}\cdot\bar{u}=\frac{S_m}{\rho}+\frac{1}{T}\frac{DT}{Dt}=\frac{S_m}{\rho}+\frac{\dot{Q}}{\rho c_p T}$$

The right side of the continuity equation represents the coupling effect of the droplet on the gas flow. The first term represents the mass source from the droplet to the gas and the second term the heat transfer source.

The velocity field in the mass and energy two-way coupling model can be decomposed into three parts: the field corresponding to the base potential flow, the flow velocity induced by the vortices in the field, and the flow velocity produced by the effects of mass and energy coupling. The base flow is simulated with a two-dimensional discrete vortex technique.

The effect of droplet mass and energy coupling on the gas flow can best be evaluated through the use of coupling parameters. The mass coupling parameter can be written as

$$\Pi_m=\frac{S_m}{\rho U_o/\ell}$$

where U_o and ℓ are characteristic flow velocity and length scales.
The energy coupling parameter can be written as

$$\Pi_E=\frac{\dot{Q}}{\rho c_p TUo/\ell}$$

Figure 4 shows the results of two-way mass and energy coupling effects for a wake downstream of a bluff body. The results indicate that the development of vortices downstream of the body may be altered somewhat due to the two-way coupling effects.

Three Dimensional Experimental Results

Experimental results concerning the three-dimensional nature of the vortex structures in a (Re≈1200 based on momentum thickness) plane wake flow have recently been acquired and analyzed. Examples of the experimental data from a cross stream section of a perturbed wake flow at a selected downstream position are shown in Figure 5. The vorticity contours illuminate the alternating sign character and the organized nature of the streamwise vortices in this flow. Eventually experimental results similar to these will be directly compared to simulation results.

Summary

Numerical and experimental investigations concerning three dimensional flow and particle dispersion processes are presently underway. Numerical and experimental results have confirmed the presence of the alternating sign patterns and concentrated vorticity levels associated with these flow phenomena. Initial simulation results demonstrate that three dimensional vortices can have a substantial effect on the particle dispersion patterns at intermediate Stokes numbers.

Extensive numerical simulations of two-way mass and energy coupling effects between vaporizing droplets in gas flows indicate that the initial vortex formation and development may be somewhat delayed. However changes in the overall character of the wake flow are not substantial over the range of parameters presently investigated.

Acknowledgments

The authors gratefully acknowledge the support of Department of Energy Grant No. DE-FG06-94ER-14450 under the direction of Dr. Oscar Manley and Dr. Robert Goulard.

References

Ashurst, W.T. and Mieburg, E., 1988, Three-dimensional shear layers via vortex dynamics, *J. Fluid Mech.*, vol. 189, pp. 87-116.

Bernal, L.P. and Roshko, A., 1986, Streamwise vortex structure in plane mixing layers, *J. Fluid Mech.*, **170**, 499-525.

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

Browand, F.K. and Troutt, T.R., 1980, A note on spanwise structure in the two-dimensional mixing layer, *J. of Fluid Mechanics*, **97**, 4, 771-781.

Browand, F.K. and Troutt, T.R., 1985, The turbulent mixing layer. Geometry of large vortices, *J. Fluid Mech.*, **158**, 487.

Crowe, C.T., Chung, J.N. and Troutt, T.R., 1993, Particle dispersion by organized turbulent structures, *Particulate Two-Phase Flow*, (ed. M.C. Roco), Butterworth-Heinemann, 626-669.

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

Huang, L.S. and Ho, C.M., 1990, Small scale transition in a plane mixing layer, *J. Fluid Mech.*, **210**, 475-500.

Jimenez, J., Cogollos, M. and Bernal, L.P., 1985, A perspective view of the plane mixing layer, *J. Fluid Mech.*, **152**, 125.

Lasheras, J.C., Cho, J.S. & Maxworthy, T., 1986, On the origin and evolution of streamwise vortical structures in a plane, free shear layer, *J. Fluid Mech.*, **172**, 231-258.

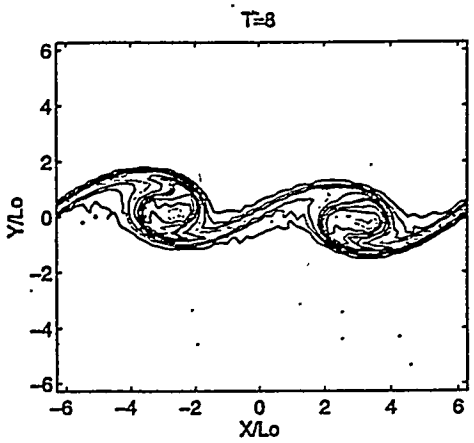
Lasheras, J.C. & Choi, H., 1988, Three dimensional instability of a plane free shear layer: an experimental study of the formation and evolution of streamwise vortices, *J. Fluid Mech.*, **189**, 53-86.

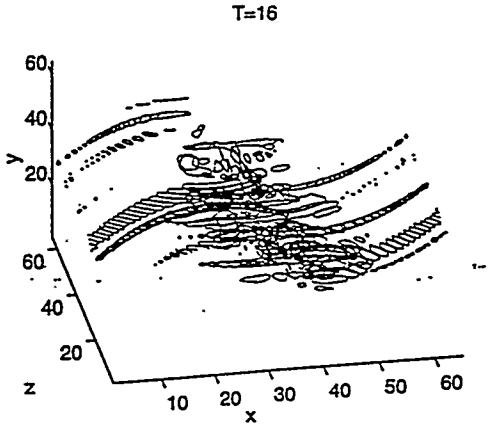
Leonard, A., 1985, Computing three-dimensional incompressible flows with vortex elements, *Ann. Rev. Fluid Mech.*, **17**, 523.

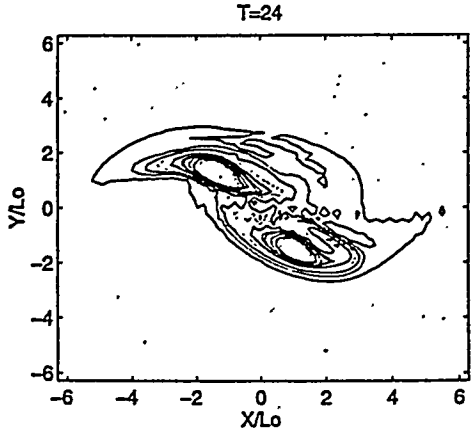
Metcalfe, R.W., Orszag, S.A., Brachet, M.E., Menon, S. & Riley, J.J., 1987, Secondary instability of a temporally growing mixing layer, *J. Fluid Mech.*, **184**, 207-243.

Moser, R.D. and Rogers, M.M., 1993, The three-dimensional evolution of a plane mixing layer: pairing and transition to turbulence, *J. Fluid Mech.*, **247**, 275-320.

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







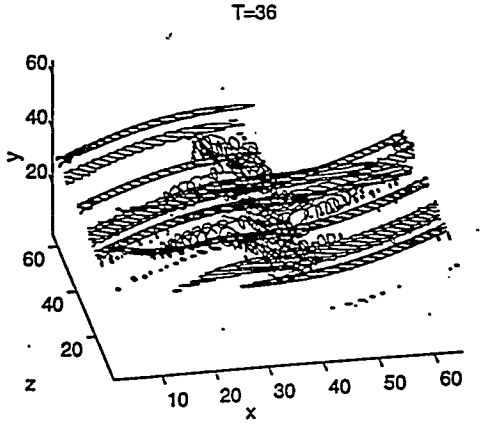


Figure 1. Time development of spanwise vorticity contours in a mixing layer

Figure 2. Time development of three dimensional streamwise vorticity contours in a mixing layer

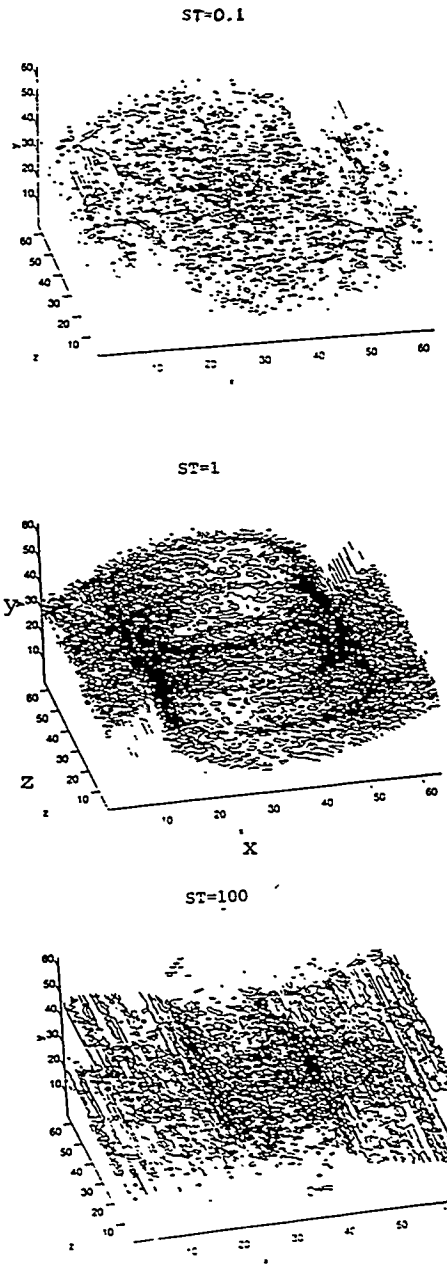


Figure 3. Particle concentration contours for three-dimensional layer at various Stokes numbers.

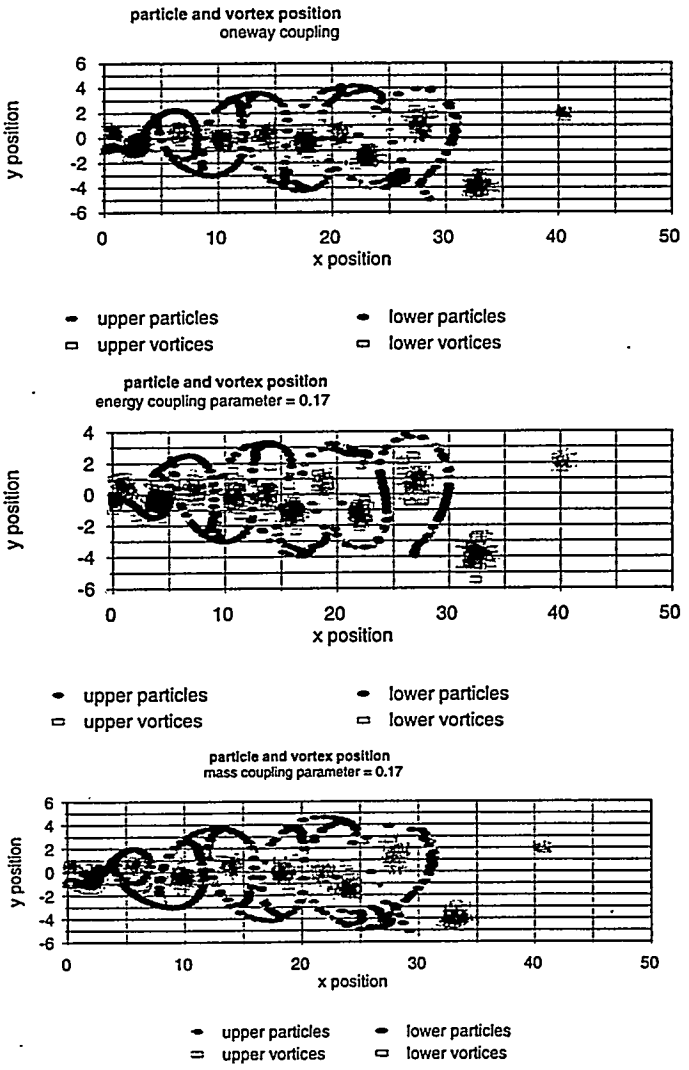


Figure 4. Two-way mass and energy coupling effects in a Plane Wake.

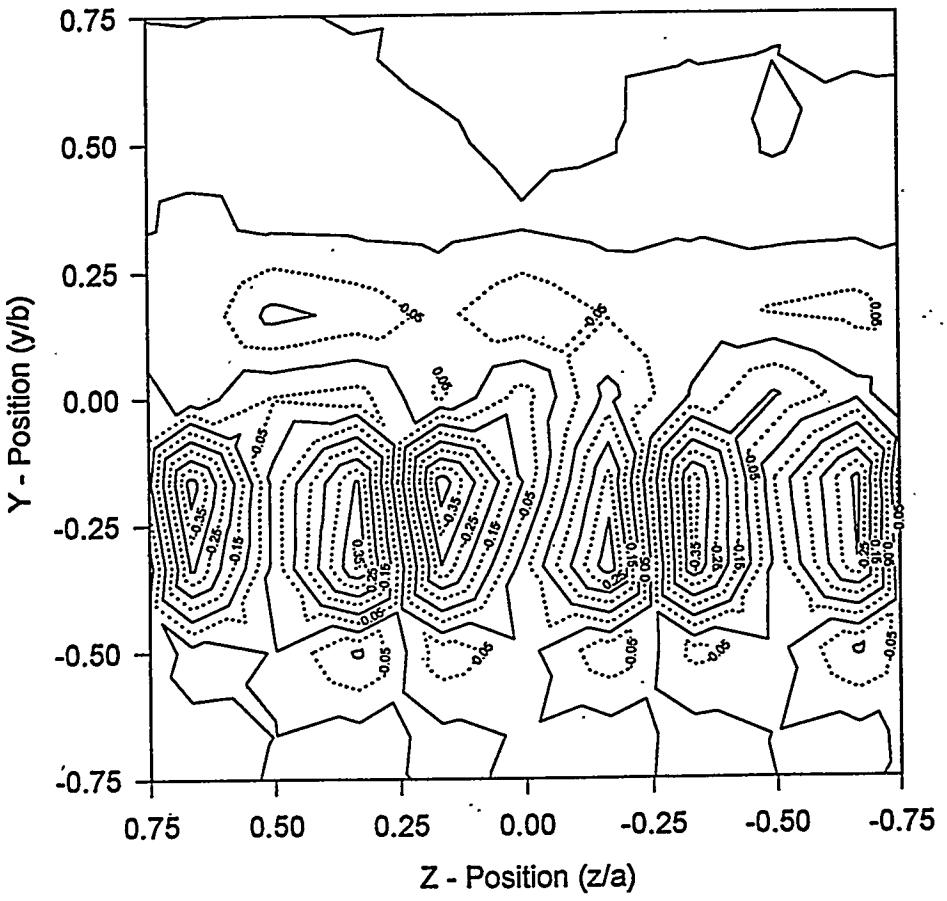


Figure 5. Phase averaged experimental streamwise vorticity contours in a plane wake.

EXPERIMENTAL STUDIES OF REYNOLDS NUMBER DEPENDENCE OF TURBULENT MIXING & TRANSPORT

Z. Warhaft
Sibley School of Mechanical and
Aerospace Engineering
Cornell University, Ithaca NY 14853

ABSTRACT

An overview of our recent experiments, in which we generate high Reynolds number homogeneous grid turbulence, is provided. We show that in a small wind tunnel we are able to achieve Reynolds numbers that are sufficiently high ($R_\lambda \sim 800, R_\ell \sim 36,000$) such that many of the aspects of turbulence that hitherto have only been observed in large scale anisotropic shear flows, are obtained. In particular we study the evolution of the spectrum with Reynolds number, the Kolmogorov constant and the internal intermittency, showing the way they tend to their high Reynolds number asymptotes. Thus we link previous low Reynolds number laboratory experiments with large scale environmental measurements.

INTRODUCTION

A detailed understanding of turbulence is required in order to predict mixing and combustion rates, drag, pollutant transport and many other aspects of fluid motion that are of concern in industry and the environment. For over fifty years the Kolmogorov (1941) similarity theory has provided the basis for the scaling of experimental data as well as for the modelling and simulation of turbulent flows. No other phenomenology has such generality or simplicity. Although there are delicate issues concerning some of its predictions, it does have good experimental confirmation for flows at high Reynolds numbers, suggesting that the scaling is approximately correct, at least for variance quantities such as the turbulence energy and the dissipation rate.

The Kolmogorov spectrum is used widely in engineering simulations and models, both in homogeneous and inhomogeneous flows. For example, Large Eddy Simulation (LES) is based on the ability to model the small scales *à la* Kolmogorov [1]. Yet engineering flows often occur at moderate Reynolds numbers where the Kolmogorov assumptions may not hold. Here the form of the spectrum is different. Remarkably, there has been no systematic study of

how the turbulence spectrum evolves as a function of Reynolds number. Indeed the question of what is high Reynolds number turbulence has not been properly addressed although it is known that by about $R_\lambda \sim 400$ the spectrum shows an appreciable inertial range, close to the $-5/3$ slope predicted by Kolmogorov [2]. Here, R_λ is the Taylor Reynolds number defined in the usual way, as $(\langle u^2 \rangle)^{1/2}\lambda/\nu$ where u is the fluctuating longitudinal velocity, λ is the Taylor microscale ($\lambda \equiv (\langle u^2 \rangle / \langle (\partial u/\partial x)^2 \rangle)^{1/2}$) and ν is the kinematic viscosity.

The Kolmogorov inertial range scaling assumes that the turbulence is locally isotropic. It has been most studied in flows that are anisotropic at the large scales due to mean shear or buoyancy, or some other effects. The recent boundary layer experiments of Saddoughi and Veeravalli [3] show that local isotropy does indeed occur as the shear stress co-spectrum (affecting the large scales) begins to diminish. However, for lower Reynolds number shear flows (for R_λ less than 300 or 400), the co-spectrum (due to the shear) extends to the dissipation range, providing almost no locally isotropic turbulence. (e.g., figure 21 of reference 3). Thus while the experiments with shear indicate at what R_λ there becomes an appreciable region of local isotropy in the wave number spectrum, they do not address whether lower R_λ experiments would be consistent with Kolmogorov scaling if the anisotropic effects of the large scales were absent, since the effects of shear dominate all of the spectrum for the lower R_λ. This issue can only be addressed if the turbulence is isotropic at all scales, including those in the energy containing range.

Unfortunately, grid turbulence, the only type of turbulence that is free of large scale anisotropy that can be generated in the laboratory, has been limited to low Reynolds numbers ($R_\lambda < 150$) since the grid generates very low turbulence intensities (less than 3%). However, over the past three years following Makita [4], we have developed and built active turbulence generator grids that produce high intensity turbulence with large integral scales. We recently have achieved an R_λ of nearly 800, around 3 times the highest R_λ that can be obtained using direct numerical simulations. The active grid turbulence has the essential ingredients of very high R_λ atmospheric flows but without their complicating effects of large scale spatial and temporal inhomogeneity.

Briefly, Makita's grid works as follows: Triangular shaped wings are placed on each mesh of the grid (figure 1). A pulse generator rotates the grid bars (and hence the wings) and at the same time the motor is fed by a random pulse which reverses the rotation of the grid bar. Thus the wing, always in rotational motion, reverses its direction randomly, providing a flapping motion. Each bar is separately controlled, providing random flapping between bars. For our experiment the mesh length was 5 cm and the mean speed was varied from 3 to 14 m/s. The rotation speed of the bars was around $2Hz$ and was slightly varied with wind speed. The velocity variance decay was similar to that of conventional grids. The experiments were carried out in our $40 \times 40\,\mathrm{cm}^2 \times 5\mathrm{m}$ wind tunnel [5]. Velocity and temperature fluctuations were measured using conventional hot wire anemometry.

Here we will provide an overview of our recent experiments. Further results can be found in Mydlarski and Warhaft[6,7].

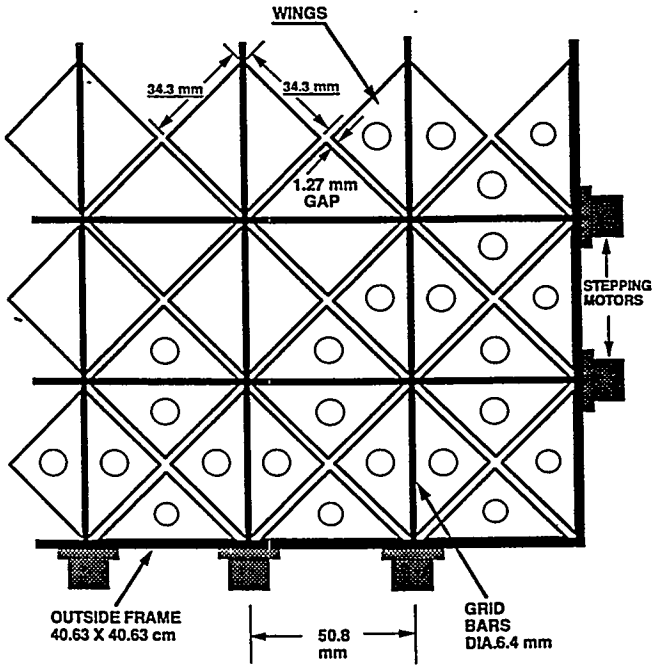


Figure 1: A sketch of the active grid.

THE SPECTRA

Figure 2(a) shows u and v spectra for $50 \leq R_\lambda \leq 473$ obtained with the active grid.[1] Our results show a significant scaling range close to $-5/3$. Figure 2(b) shows a compilation of results from Saddoughi and Veeravalli (1994) taken in a number of different flows, both in the laboratory and in the atmosphere. Although their results are mainly for shear flows, they show a very strong resemblence to our homogeneous measurements. Both the sets of results (figures 2(a) and (b)) have provided strong support for Kolmogorov scalings.

A more stringent test for Kolmogorov scaling is to display the data in compensated form. Figure 3(a) shows the F_{11} spectra from our experiments multiplied by $\varepsilon^{-2/3}k_1^{5/3}$. The inertial range should be horizontal on such a plot. The 4 representative compensated spectra are for $R_\lambda = 99, 199, 373$ and 448. While there is a trend towards becoming horizontal with R_λ, there is still a significant slope at the highest Reynolds number. The dashed lines show the accepted value of the Kolmogorov constant. It falls approximately midway between the low and high ends of the scaling range.

The spectra of figure 3(a) show, then, that while there clearly is a scaling region, its slope is not $-5/3$ even for the highest R_λ. It appears that the spectra are not yet self similar. In order to describe them, we use a modified similarity form:

$$F_{11}(k_1) = C_{1*}\varepsilon^{2/3}k_1^{-5/3}(k_1\eta)^{5/3-n_1} \tag{1}$$

$$F_{22}(k_1) = C_{2*}\varepsilon^{2/3}k_1^{-5/3}(k_1\eta)^{5/3-n_2} \tag{2}$$

[1]Very recently we have achieved an R_λ of 780 using a larger active grid. The trends in the new data are consistent with the lower R_λ described here.

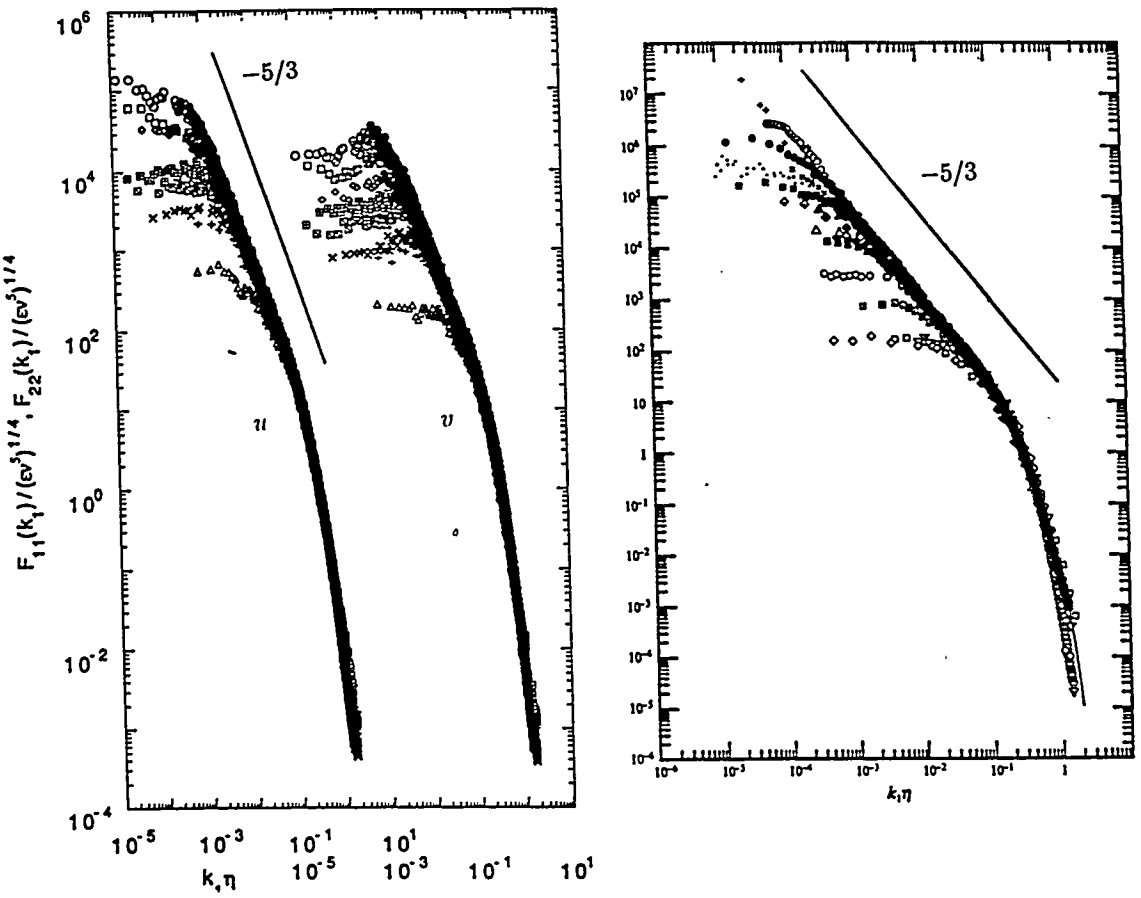


Figure 2: (a) The longitudinal (u) and transverse (v) power spectra for R_λ varying from 50 (open triangles) to 473 (open circles) in the present experiment. (b) A compilation of shear flow u spectra [3]. Here the R_λ variation is from 23 to 3,200.

where n_1 and n_2 are the slopes of the scaling region for the u and v spectra respectively and C_{1*} and C_{2*} are now Kolmogorov variables: both C_* and n are functions of R_λ (and as $R_\lambda \to \infty, n_1, n_2 \to 5/3$ and $C_* \to C$ (the traditional Kolmogorov constant)). In figure 3(b) we have plotted $C_{1*} = F_{11}(k)\varepsilon^{-2/3}k_1^{n_1}\eta^{n_1-5/3}$ *vs.* $k_1\eta$ for the four spectra of figure 3(a). The value of n_1, which varied from $1.40(R_\lambda = 99)$ to $1.58(R_\lambda = 448)$ was determined by trial and error such that the scaling region would be horizontal. Note that C_{1*} decreases as R_λ increases but even for the high R_λ case its value is approximately 0.7, well above the accepted high Reynolds number estimate of approximately 0.5.

Figure 4 is a summary of the best fit scaling exponent, n_1, for all the u spectra we measured over the range $50 < R_\lambda < 473$. These results show that below $R_\lambda \sim 100$ the spectra have a scaling region in the range 1.3 to 1.4. The relatively large uncertainty is due to the small width of the scaling region. There is then a relatively well defined transition region extending from $R_\lambda \sim 100$ to $R_\lambda \sim 200$ where the scaling exponent steepens to a value of about 1.52. Beyond $R_\lambda \sim 200$ the slope tends to increase very slightly. Our maximum slope was 1.58 at $R_\lambda \sim 473$. We emphasize the high degree of reproductability of the high Reynolds number experiments ($R_\lambda > 250$). The experimental scatter was ± 0.01.

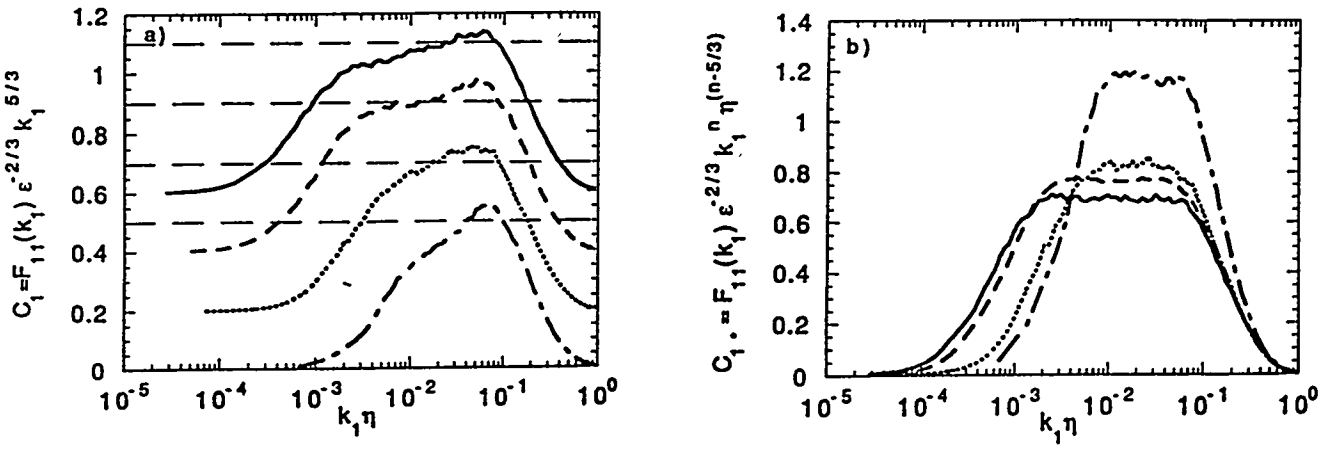


Figure 3: Compensated u spectra for $R_\lambda = 99, 199, 373$ and 448. (a) $F_{11}(k_1)\varepsilon^{-2/3}k_1^{5/3} = C_1$ (b) $F_{11}(k_1)\varepsilon^{-2/3}k_1^{n_1}\eta^{n_1-5/3} = C_{1*}$. The horizontal dashed, line in (a) is 0.5, the accepted value for C_1 (equation 1) for high Reynolds number turbulence. Each of the curves in (a) has been successively shifted by 0.2 with respect to the lower one. The Reynolds number increases from the bottom curve upwards in (a) and from the upper curve downwards in (b).

We have fitted a $-2/3$ power law curve to figure 4. Defining $p_1 = 5/3 - n$, we find $p_1 = 5.25R_\lambda^{-2/3}$. This suggests that a 5/3 scaling region will not occur until $R_\lambda \sim 10^4$, a very high Reynolds number indeed. Figure 5 shows a plot of C_{1*} as a function of p. (The values of C_* were determined from all of the measured spectra in the same manner as for the four spectra in figure 2(b).) The best fit line to C_{1*} is

$$C_{1*} = 0.51 + 2.39p_1. \tag{3}$$

Thus, when $p_1 = 0 (n_1 = 5/3), C_{1*} = C_1 = 0.51$. The generally accepted value of the three-dimensional Kolmogorov constant C is 1.5 [4] and the one- and three-dimensional constants are related by $C_1 = 18C/55$. Thus $C_1 \approx 0.49$. Our extrapolated value of 0.51 is remarkably consistent with this value. We emphasize that equation 3 is a best fit.

In a similar way, we determined C_2, the Kolmogorov variable for the transverse velocity, v. The ratio of the Kolmogorov constants C_2/C_1 must be 4/3 if $n_1 = n_2 = 5/3$. In order to determine the best fit line for C_{2*} we have used the value $C_2 = 4/3 \times 0.51$ for $p_2 = 0$. The best line fit is

$$C_{2*} = 0.68 + 3.07p_2. \tag{4}$$

INTERNAL INTERMITTENCY

While approximately Gaussian at the large scales, turbulence at the small scales is strongly intermittent and non-Gaussian. If the Reynolds number is high enough, the intermittent structure should be observable not only in the dissipation range, but also in the

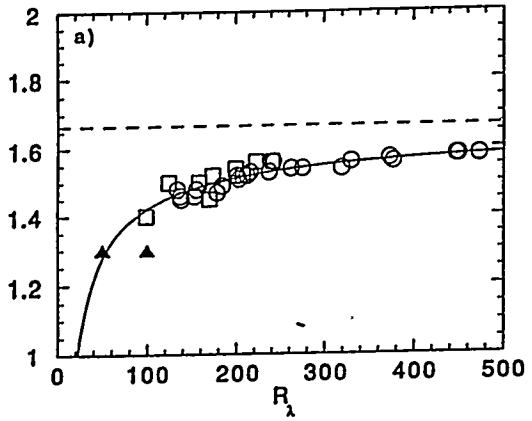


Figure 4: The slope of the u spectrum as a function of R_λ for the various active grid experiments.

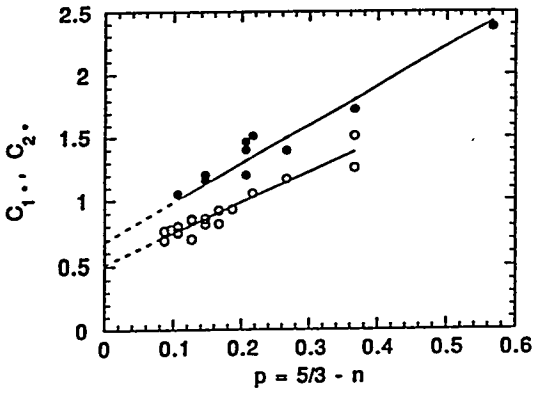


Figure 5: C_{1*} and C_{2*} (equations 1 and 2) plotted as a function of $p = 5/3 - n$ where n is the slope of the respective spectrum. Open circles, C_{1*}; closed circles, C_{2*}. For C_{1*}, the line is that of best fit. For C_{2*}, the best fit line has been forced through $C_{1*}(p = 0) \times 4/3$ at $p = 0$.

inertial subrange. Kolmogorov [2] hypothesised that for high R_λ, the energy dissipation rate averaged over a radius r, ε_r, (where r is within the inertial subrange) is related to $\Delta u(r)$ by the relation

$$\Delta u(r) = V(r\varepsilon_r)^{1/3} \tag{5}$$

where V is a stochastic variable independent of r and ε_r. Thus the quantities $\Delta u(r)$ and ε_r must be statistically dependent. In figure 6(a) and (b) we have plotted $(r_a\varepsilon^{11}_{r_a})^{1/3}$ and $(r_a\varepsilon^{21}_{r_a})^{1/3}$ conditioned on $\Delta u(r_a)$. Here $\varepsilon^{11}_{r_a}(\equiv 15\nu U^{-2}\langle(\partial u/\partial t)^2\rangle)$ and $\varepsilon^{21}_{r_a}(\equiv (7.5\nu U^{-2}\langle(\partial v/\partial t)^2\rangle)$ are (one-dimensional surrogates for the total dissipation) determined over a record of length r_a, from which we also obtained $\Delta u(r_a)$ from velocity difference between the start and end of the record. For high R_λ, figures 6(a) and (b) show both $\varepsilon^{11}_{r_a}$ and $\varepsilon^{21}_{r_a}$ are statistically dependent on $\Delta u(r_a)$: the curves have a pronounced V shape indicating that higher dissipation (averaged over r_a) is associated with larger velocity differences. For low R_λ, the conditional statistics are considerably flatter, showing only a weak dependence on $\Delta u(r_a)$. This is particularly so for the $\varepsilon^{21}_{r_a}$ case (figure 26(b)). Notice the asymmetry in those curves for low R_λ.

Recently it was pointed out [8,9] that a statistical dependence between $\Delta u(r)$ and $r\varepsilon^{11}_r$ must occur, even if the Kolmogorov revised similarity hypothesis, KRHS, does not hold. A correlation will occur between ε^{11}_r and $\Delta u(r)$ on purely kinematic grounds. For a given $\Delta u(r)$, there exists a minimum possible value of the dissipation ε^{11}_r which corresponds to a linear variation in u over the distance r $(\varepsilon^{11}_r|_{\text{min}} = 15\nu(\Delta u(r)/r)^2)$. On the other hand, the existence of a statistical dependence between ε^{21}_r and $\Delta u(r)$ suggests a dynamical contribution. For our high R_λ experiments (figure 6(a) and (b)) both ε^{11}_r and ε^{21}_r conditioned on $\Delta u(r)$ show statistical dependence on $\Delta u(r)$, giving strong support for KRHS. Thus the combination of our high and low Reynolds number cases and the conditional statistics of both ε^{11}_r and ε^{21}_r on $\Delta u(r)$ enable us to separate the kinematical from the dynamical contributions to the statistical dependence: at low R_λ the dependence appears to be kinematical while at high

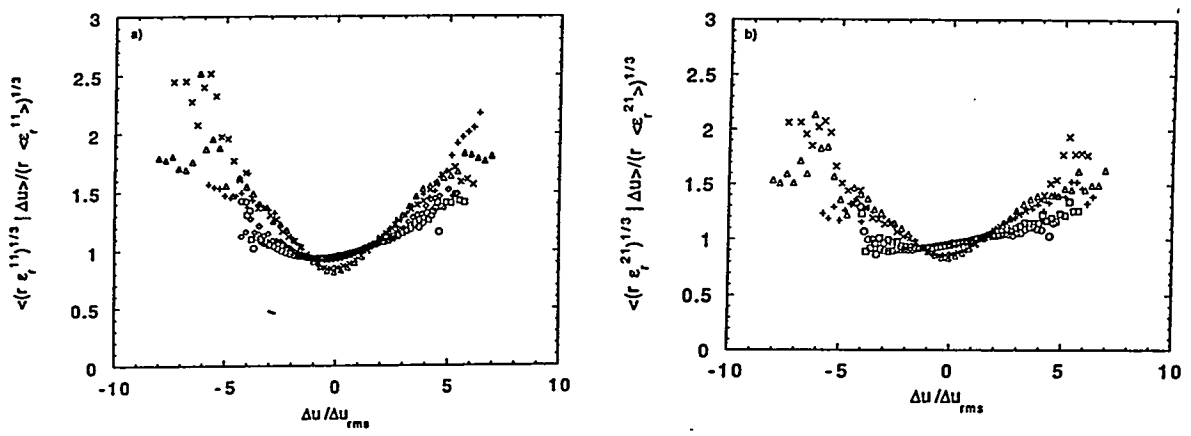


Figure 6: The expectation of $\varepsilon_{r_a}^{11}$ conditioned on $\Delta u(r_a)$ and $\varepsilon_{r_a}^{21}$ conditioned on $\Delta u(r_a)$. (a) $[\langle (r_a \varepsilon_{r_a}^{11})^{1/3} | \Delta u(r_a) \rangle]/(r_a \langle \varepsilon_{r_a}^{11} \rangle)^{1/3}$ (b) $[\langle (r_a \varepsilon_{r_a}^{21})^{1/3} | \Delta u(r_a) \rangle]/(r_a \langle \varepsilon_{r_a}^{21} \rangle)^{1/3}$. For both (a) and (b) the symbols are : $\triangle, R_\lambda = 473R; +, R_\lambda = 275R; \times, R_\lambda = 207S; \square R_\lambda = 99S; \diamond, R_\lambda = 100$ (conventional grid, $M = 10.2$ cm), $\bigcirc, R_\lambda = 50$ (conventional grid, $M = 2.54$cm).

R_λ a significant dynamical effect is observed.

Finally, we have determined the intermittency exponent (K62) from the autocorrelation of ε [2]:

$$\rho_{\varepsilon\varepsilon}(r) = \frac{\langle \varepsilon(x)\varepsilon(x+r) \rangle}{\langle \varepsilon^2 \rangle} \propto r^{-\mu}. \tag{6}$$

Our estimates of μ are plotted as a function of R_λ in figure 7. Below $R_\lambda \sim 100, \mu \sim 0$. There is then a steep rise to a value of around 0.11 at $R_\lambda \approx 450$. Measurements in very high Reynolds number flows show μ is approximately constant, with a value of approximately 0.2 [2]. Evidently, we are not yet at a sufficiently high R_λ to attain this value. This is consistent with the spectra (figure 3) which are still evolving.

CONCLUDING COMMENTS

We have implemented a powerful new, cheap, experimental method of studying homogeneous turbulence at Reynolds numbers that are sufficiently high so that many of the issues raised in contemporary turbulence theory may be studied. We have found that the R_λ must be well above 200 before fully developed internal intermittency first appears. We note that Direct Numerical Simulations tend to be in the range $R_\lambda \sim 200$ and it is unlikely that they will achieve R_λ in the 500 to 800 range in the near future. We are presently using the active grid to study scalar mixing and dispersion, as well as further elucidating the issues raised here.

ACKNOWLEDGEMENTS

I am grateful to my students Chenning Tong and Laurent Mydlarski and to Edward Jordan for technical assistance. This work was funded by the Department of Energy, Basic Energy Sciences under grant DE-FG02-88ER13929.

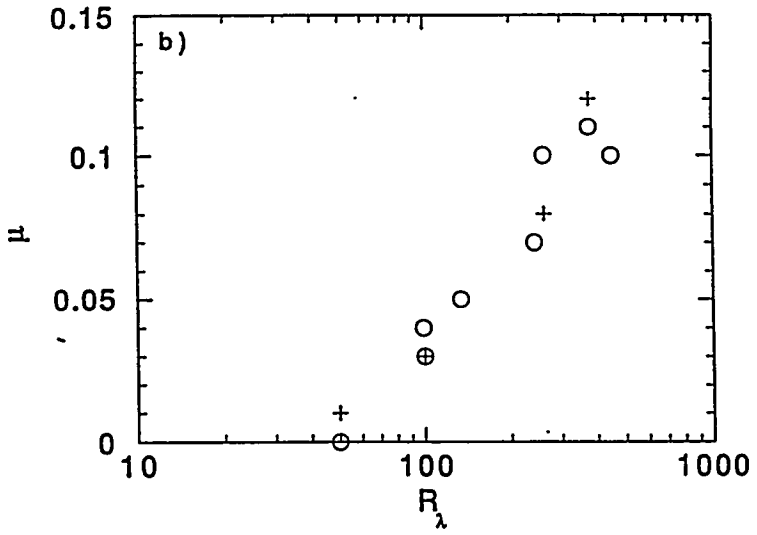


Figure 7: The estimate of μ using equation 6.

REFERENCES

1. W.C. Reynolds, 1990. "In Whither Turbulence". *Turbulence at the Crossroads*, ed. J.L. Lumley, 313–343, Berlin: Springer-Verlag, 535.

2. M. Nelkin, 1994. Universality and scaling in fully developed turbulence. *Adv. in Physics* **43**, 143–181.

3. S.G. Saddoughi, & Veeravalli, S.V. 1994. Local isotropy in turbulent boundary layers at high Reynolds number. *J. Fluid Mech.* **268**, 333–372.

4. H. Makita, 1991. Realization of a large-scale turbulence field in a small wind tunnel. *Fluid Dynamics Research* **8**, 53–64.

5. Jayesh, C. Tong, and Z. Warhaft, 1994. On temperature spectra in grid turbulence. *Phys. Fluids* **6**, 306–312.

6. L. Mydlarski and Z. Warhaft, 1996. On the onset of high Reynolds number grid generated wind tunnel turbulence. *J. Fluid Mech.* (In press).

7. L. Mydlarski and Z. Warhaft, 1996. Temperature spectra and related statistics in grid turbulence up to $R_\lambda \sim 500$. (In preparation).

8. S. Chen, S. D. Doolen, R. H. Kraichnan and Z-S. She, 1993. On statistical correlations between velocity increments and locally averaged dissipation in homogeneous turbulence. *Phys. Fluids* A **5**, 458–463.

9. S.T. Thoroddsen, 1995. Reevaluation of the experimental support for the Kolmogorov refined similarity hypothesis. *Phys. Fluids* **7**, 691–693.

CONDUCTION IN NON-UNIFORM COMPOSITES

A. Prosperetti and M. Marchioro
Department of Mechanical Engineering, The Johns Hopkins University
Baltimore MD 21218

ABSTRACT

A method for the numerical simulation of non-uniform mixtures is described and applied to the heat conduction problem. It is found that, when the inclusions are not uniformly distributed in space, the standard single-phase Fourier law of conduction, with an effective conductivity multiplying the gradient of the average temperature, is not satisfied.

INTRODUCTION

In several recent papers we have developed an approach to the derivation of averaged equations for disperse multiphase flows that appears promising in that the closure problem is phrased in terms of computable quantities (Zhang and Prosperetti 1994a, 1994b, 1996). The theory is developed in terms of ensemble averages that are notoriously difficult to calculate numerically. However, the complexity of the task decreases by many orders of magnitude in the case of homogeneous suspensions for which the average quantities become spatially uniform, as in this situation the ensemble average can be reduced to a volume average over many (as opposed to a large number of) realizations of the flow. In Zhang and Prosperetti (1994a) we have given an explicit example of the procedure for one simple case of this type. We started out with a representation of the quantity of interest in terms of the unconditionally averaged fields multiplied by unknown coefficients, and determined the coefficients from the simulations. We believe that this is a promising approach that can ultimately lead to a useful closed set of averaged equations provided adequate simulations can be carried out. Unfortunately, the method cannot evidently deal with non-uniform suspensions, a situation that it is imperative to consider in order to develop a complete theory. Indeed, it is widely recognized that there is little hope of developing realistic models of multiphase flows without second-order spatial derivatives (see e.g. Batchelor 1988).

These considerations motivate the present study in which we propose a method for the evaluation of gradient terms that only requires the calculation of averages for uniform suspensions. In order to develop and test the method, we have deemed it desirable to work with a system simpler than two phase flow, namely heat conduction in a composite with spherical inclusions. In this case the quantity to be determined is the mean heat flux in the mixture $\mathbf{q}_m$. It can readily be shown that this quantity can be calculated from the expression (Batchelor 1974):

$$\mathbf{q}_m = -k_C \nabla T_m - (k_D - k_C) < \nabla T_D >, \tag{1}$$

where the subscripts C and D refer to the continuous and disperse phase respectively, the angle brackets denote the phase-ensemble average, k is the thermal conductivity, T_D the disperse-phase temperature and

$$T_m = \beta_C < T_C > + \beta_D < T_D > \tag{2}$$

is the mean temperature defined in terms of the individual phase average temperatures and volume fractions $\beta_{C,D}$. Note that $\beta_C + \beta_D = 1$.

THE NON-UNIFORM ENSEMBLE

To explain our procedure, consider for example the standard Fourier law of single-phase heat conduction relating the heat flux $\mathbf{q}$ to the temperature gradient ∇T, $\mathbf{q} = -k\nabla T$. If this relation is true, the thermal conductivity k can be calculated, or measured, no matter how small or large ∇T is. Indeed, in the classical Chapman-Enskog expansion of the Kinetic Theory of Gases, this relation is derived and k calculated on the assumption of a small temperature gradient.

We base our approach on the assumption that the closure relations that we are seeking establish a functional relation among the average quantities included in the theory endowed with a similar "intrinsic" nature. This remark suggests that we proceed perturbatively, setting up a "nearly uniform" suspension and expanding in the degree of non-uniformity in such a way that all the actual numerical calculations are conducted on a uniform composite. The idea is similar to the familiar asymptotic method of domain perturbation, where the problem in the perturbed domain is approximated as a series of problems on the simpler unperturbed one.

We consider composites occupying the entire space and consisting of an infinite number of copies of a fundamental cubic cell L much greater than the particle radius a. We can therefore simply deal with such a fundamental cell replacing the rest of the composite by periodic boundary conditions. By standard techniques we can generate numerically an ensemble of such fundamental cells, each one with N particles randomly and uniformly distributed in the mean. Let $P_0(N) \equiv P(\mathbf{y}^1, \mathbf{y}^2, \ldots \mathbf{y}^N)$ in which $\mathbf{y}^\alpha$, $\alpha = 1, 2, \ldots, N$ are the positions of the particle centers, be the probability density for this ensemble. In each realization, subject now each particle to the small displacement

$$(\mathbf{y}^\alpha)' = \mathbf{y}^\alpha + \epsilon \mathbf{f}(\mathbf{y}^\alpha), \tag{3}$$

where $\mathbf{f}$ is a given, deterministic function equal for all particles and ϵ a small parameter. The new probability density then becomes

$$P(N) = P_0(N)\,[1 + \epsilon\Phi(N)], \qquad \Phi(N) = \sum_{\alpha=1}^{N} \nabla_{\mathbf{y}^\alpha} \cdot \mathbf{f}(\mathbf{y}^\alpha). \tag{4}$$

For the purposes of the present paper we take

$$\nabla \cdot \mathbf{f}(\mathbf{y}) = \sin \mathbf{k} \cdot \mathbf{y}, \tag{5}$$

where $\mathbf{k}$ equals $2\pi/L$ times a unit vector in one of the coordinate directions. This may be considered as a single term of a Fourier expansion. A more general analysis will be presented elsewhere.

The volume fraction of the disperse phase β_D is defined by

$$\beta_D(\mathbf{x},t) = \frac{1}{N!}\int d\mathcal{C}^N P(N;t)\,\chi_D(\mathbf{x};N)\,, \tag{6}$$

where χ_D is the characteristic, or indicator, function of the disperse phase. For a suspension of equal spherical particles of radius a an explicit representation is (Lundgren 1972)

$$\chi_D(\mathbf{x};N) = 1-\chi_C(\mathbf{x},N) = \sum_{\alpha=1}^{N} H\left(a-|\mathbf{x}-\mathbf{y}^\alpha|\right), \tag{7}$$

where H is the Heaviside distribution. Upon substitution of (4) into this definition of β_D we have

$$\beta_D = \beta_D^0 + \epsilon\beta_D^1 \sin\mathbf{k}\cdot\mathbf{x}, \tag{8}$$

where

$$\beta_D^0 = \frac{1}{N!}\int \chi_D(\mathbf{x};N)\,P_0(N;t)\,d\mathcal{C}^N = \frac{Nv}{L^3}, \tag{9}$$

where $v=\frac{4}{3}\pi a^3$ is the particle volume, and

$$\begin{aligned}\beta_D^1 &= \frac{2}{L^3}\frac{1}{N!}\int d\mathcal{C}^N\,P_0(N;t)\,\Phi(N)\sum_{\alpha=1}^{N}\sin\mathbf{k}\cdot\mathbf{y}^\alpha\int_{|\mathbf{z}|\le a} d^3z\cos\mathbf{k}\cdot\mathbf{z}\\ &\simeq \frac{2}{L^3}\frac{v}{N!}\int d\mathcal{C}^N\,P_0(N;t)\,\Phi^2(N)+O\left(\frac{a^2}{L^2}\right).\end{aligned} \tag{10}$$

The last form is obtained upon a Taylor series expansion of the cosine in the inner integral of the previous line, and upon recognizing that the remaining sum of sines, if f is given by (5), is just (4).

TEMPERATURE FIELDS

For the determination of the microscopic, exact temperature field for each configuration we use the multipole expansion method closely following the approach of Sangani and Yao (1988). We thus set

$$T_C(\mathbf{x};N) = \mathbf{G}\cdot\mathbf{x}+\sum_{\alpha=1}^{N}\sum_{m} A_m^\alpha\Delta_m S_1(\mathbf{x}-\mathbf{y}^\alpha), \tag{11}$$

where we have used an abbreviated notation for the inner summation that, written out in detail, is

$$\sum_m \Delta_m A_m^\alpha S_1(\mathbf{x}-\mathbf{y}^\alpha) = \sum_{n=1}^{\infty}\sum_{m=0}^{n}\left[A_{nm}^\alpha\partial_1^{n-m}\Delta_m+\tilde{A}_{nm}^\alpha\partial_1^{n-m}\tilde{\Delta}_m\right]S_1(\mathbf{x}-\mathbf{y}^\alpha). \tag{12}$$

Here ∂_1, Δ_m, and $\tilde{\Delta}_m$ are differential operators with respect to the components of the field point $\mathbf{x}$ and S_1 is Hasimoto's function (for details see Sangani and Yao 1988). The vector $\mathbf{G}$ is a constant that can be interpreted as the "overall" temperature gradient in a large piece of the composite.

The temperature field inside the generic particle is represented as a spherical harmonic expansion:

$$T_D^\alpha = \sum_{n=1}^{\infty}\sum_{m=-n}^{n}\left(C_n^m+D_n^m a^{-2n-1}\right)r^n Y_n^m(\theta,\phi) \tag{13}$$

Imposing the continuity of temperature and heat fluxes at the particle surfaces gives relations between the coefficients A, C, and D that we do not need to write down in detail. Suffice it to say that, as expected, all these constants are found to depend linearly on $\mathbf{G}$.

EXPRESSIONS FOR THE AVERAGES

We are going to seek representations of the average fields in terms of Fourier series truncated at a low order in correspondence with the ansatz (5) for f. For the continuous-phase average temperature we write

$$\beta_C\left(<T_C>-\mathbf{G}\cdot\mathbf{x}\right) = \tau_0^C + \epsilon\tau_s^C \sin\mathbf{k}\cdot\mathbf{x} + \epsilon\tau_c^C \cos\mathbf{k}\cdot\mathbf{x}. \tag{14}$$

According to Fourier's theorem, the coefficients τ^C are found by taking projections. For example

$$\begin{aligned}\tau_0^C &= \frac{1}{N!L^3}\int dC^N\, P_0(N;t)\sum_{\alpha=1}^{N}\int_{V_C} d^3x \sum_m A_m^\alpha \Delta_m S_1(\mathbf{x}-\mathbf{y}^\alpha)\\ &= -\beta_D^0\,\frac{1}{N!}\int dC^N\, P_0(N;t)\frac{1}{N}\sum_{\alpha=1}^{N}\left[T_C^r(\mathbf{y}^\alpha)-\mathbf{G}\cdot\mathbf{y}^\alpha\right],\end{aligned} \tag{15}$$

where the expression in the second line is found upon substituting the representation (11) for T_C into the one in the first line. Similarly one finds that the other coefficients may be conveniently written as

$$\tau_s^C = -\beta_D^0\left[(\tau_s^C)_1 + 3\,\mathbf{m}\cdot\tau_s^C\right], \qquad \tau_c^C = 3\beta_D^0\,\mathbf{m}\cdot\tau_c^C, \tag{16}$$

where $\mathbf{m} = \mathbf{k}/k$ and

$$(\tau_s^C)_1 = \frac{1}{N!}\int dC^N\, P_0(N;t)\Phi(N)\frac{2}{N}\sum_{\alpha=1}^{N}\left[T_C^r(\mathbf{y}^\alpha)-\mathbf{G}\cdot\mathbf{y}^\alpha\right]\sin\mathbf{k}\cdot\mathbf{y}^\alpha \tag{17}$$

$$\tau_s^C = \left[\frac{1}{N!}\int dC^N\, P_0(N;t)\Phi(N)\frac{2}{Na^3}\sum_{\alpha=1}^{N}\mathbf{A}_2^\alpha \sin\mathbf{k}\cdot\mathbf{y}^\alpha\right]\cdot\mathbf{m} \tag{18}$$

$$\tau_c^C = \frac{1}{N!}\int dC^N\, P_0(N;t)\Phi(N)\frac{2}{N}\left[\frac{1}{ka^3}\sum_{\alpha=1}^{N}\mathbf{A}_1^\alpha\right]\sin\mathbf{k}\cdot\mathbf{y}^\alpha. \tag{19}$$

Here $\mathbf{A}_1$ and $\mathbf{A}_2$ are a vector and a symmetric second-order tensor having components related to the 3 and 6 scalar coefficients corresponding to $n = 1$ and 2 in the expansion (11).

One can readily deduce some information about the necessary structure of these complicated expressions as follows. Let us start from τ_0^C. This quantity must be a scalar linearly dependent on $\mathbf{G}$. This is only possible if the vector nature of $\mathbf{G}$ can be neutralized by taking a scalar product with another vector, but no other vector is available here as τ_0^C is calculated on the basis of a uniform mixture. Hence we expect $\tau_0^C = 0$, a result that is confirmed by the numerical calculations.

For $(\tau_s^C)_1$, again we expect a linear dependence on $\mathbf{G}$ but, since this quantity "knows" about the non-uniformity of the mixture, also on $\mathbf{m}$. Then we are led to the form

$$(\tau_s^C)_1 = \frac{t}{k}\,\mathbf{m}\cdot\mathbf{G}, \tag{20}$$

with t a dimensionless quantity to be calculated numerically. For τ_s^C, τ_c^C we need a vector linearly dependent on $\mathbf{G}$ and possibly, linearly or nonlinearly, upon $\mathbf{m}$. We are thus led to

$$\tau_{c,s}^C = \frac{u_{c,s}^1}{k}\mathbf{G} + \frac{u_{c,s}^2}{k}\,(\mathbf{m}\cdot\mathbf{G})\,\mathbf{m}. \tag{21}$$

For the disperse-phase temperature field we write

$$\beta_D\,(< T_D > -\mathbf{G}\cdot\mathbf{x}) = \tau_0^D + \epsilon\tau_s^D \sin\mathbf{k}\cdot\mathbf{x} + \epsilon\tau_c^D \cos\mathbf{k}\cdot\mathbf{x}, \tag{22}$$

and, from the representation (13), we find

$$\tau_0^D = -\tau_0^C, \qquad \tau_s^D = (\tau_s^C)_1 = \frac{t}{k}\mathbf{m}\cdot\mathbf{G}, \tag{23}$$

$$\tau_c^D = \frac{3}{5}\frac{a^2}{1-\kappa}\beta_D^0\,(u_c^1 + u_c^2)\mathbf{k}\cdot\mathbf{G}. \tag{24}$$

The last quantity to be evaluated is the mean temperature gradient in the disperse phase. Upon setting, as before,

$$(1-\kappa)\,\beta_D < \nabla T_D >= \beta_D^0\,[\mathbf{d}_0 + \epsilon\mathbf{d}_s \sin\mathbf{k}\cdot\mathbf{x} + \epsilon\mathbf{d}_c \cos\mathbf{k}\cdot\mathbf{x}]\,, \tag{25}$$

with $\kappa = k_D/k_C$ the ratio we find

$$\mathbf{d}_0 = (1-\kappa)\,\frac{1}{N!}\int dC^N\,P_0(N;t)\frac{1}{N}\sum_{\alpha=1}^{N}\nabla T_D(\mathbf{y}^\alpha) = -3u_0\mathbf{G}, \tag{26}$$

$$\mathbf{d}_s = (1-\kappa)\,\frac{2}{N!}\int dC^N\,P_0(N;t)\,\Phi(N)\frac{1}{N}\sum_{\alpha=1}^{N}\nabla T_D(\mathbf{y}^\alpha)\,\sin\mathbf{k}\cdot\mathbf{y}^\alpha = -3\left[u_c^1\mathbf{G} + u_c^2\,(\mathbf{m}\cdot\mathbf{G})\,\mathbf{m}\right], \tag{27}$$

$$\begin{aligned}\mathbf{d}_c &= -(1-\kappa)\,\frac{a^2}{5}\left[\frac{2}{N!}\int dC^N\,P_0(N;t)\,\Phi(N)\frac{1}{N}\sum_{\alpha=1}^{N}\nabla\nabla T_D(\mathbf{y}^\alpha)\,\sin\mathbf{k}\cdot\mathbf{y}^\alpha\right]\cdot\mathbf{k} \\ &= -3\left[u_s^1\mathbf{G} + u_s^2\,(\mathbf{m}\cdot\mathbf{G})\,\mathbf{m}\right].\end{aligned} \tag{28}$$

THE AVERAGE HEAT FLUX

With the results of the previous section, we can now write down an expression for the mean flux $\mathbf{q}_m$. It turns out that, numerically, we find $t = 0$, $u_s^1 = 0$, $u_s^2 = 0$. Hence we drop the corresponding terms and have

$$-\frac{1}{k_C}\,\mathbf{q}_m = \left(1 + 3u_0\beta_D^0 + 3\epsilon\beta_D^0\,u_c^1 \sin\mathbf{k}\cdot\mathbf{x} - 3\epsilon\beta_D^0\,u_c^1\,\mathbf{m}\,\mathbf{m}\sin\mathbf{k}\cdot\mathbf{x}\right)\cdot\mathbf{G}. \tag{29}$$

This form of the result is dependent on the specific way in which the problem has been set up, rather than reflecting an intrinsic relation among the fundamental quantities of the theory, namely $< T_{C,D} >$ and β_D. As explained before, we proceed on the postulate that such an intrinsic form does exist and therefore we try to express $\mathbf{G}$ and $\mathbf{m}\,(\mathbf{m}\cdot\mathbf{G})\,\sin\mathbf{k}\cdot\mathbf{x}$ in terms of such fundamental quantities, or some convenient combination of them. We shall make use of ∇T_m and $\nabla\,(< T_D > - < T_C >)$.

Upon substitution of the previous results into the definition (2) of the mean temperature and differentiation, we find

$$\nabla T_m = \mathbf{G} - 3\epsilon\beta_D^0\,\left(u_c^1 + u_c^2\right)(\mathbf{m}\cdot\mathbf{G})\,\mathbf{m}\sin\mathbf{k}\cdot\mathbf{x}, \tag{30}$$

$$\nabla(<T_D> - <T_C>) = 3\epsilon \frac{\beta_D^0}{1-\beta_D^0} \left(u_c^1 + u_c^2\right) (\mathbf{m}\cdot\mathbf{G})\, \mathbf{m} \sin \mathbf{k}\cdot\mathbf{x}. \tag{31}$$

It is obvious from this expression that, for a uniform mixture in which $\epsilon = 0$, $\nabla < T_D > = \nabla < T_C >$, as is well known. For this relation to hold also in the present case, one would need $u_c^1 + u_c^2 = 0$, which is not supported by the simulations.

These two relations are now solved for $\mathbf{G}$ and $\mathbf{m}\,(\mathbf{m}\cdot\mathbf{G})\,\sin(\mathbf{k}\cdot\mathbf{x})$ and the result inserted into the expression (29) for $\mathbf{q}_m$. The result is

$$\begin{aligned} -\frac{1}{k_C}\mathbf{q}_m \;&= -\left(1 + 3\beta_D^0 u_0 + 3\beta_D^0 u_c^1 \epsilon \sin \mathbf{k}\cdot\mathbf{x}\right) \nabla T_m \\ &\quad + \left(1-\beta_D^0\right)\left(3\beta_D^0 u_0 - \frac{u_c^2}{u_c^1 + u_c^2}\right) \nabla(<T_D> - <T_C>)\,. \end{aligned} \tag{32}$$

In writing this expression we have retained the sine in the first term for the following reason. Let $\kappa_{eff} = \kappa_{eff}(\beta_D)$ be the effective conductivity of a uniform composite normalized by k_C. If we use the expression (8) for β_D, we have

$$\kappa_{eff}\left(\beta_D^0 + \epsilon\beta_D^1 \sin \mathbf{k}\cdot\mathbf{x}\right) \simeq \kappa_{eff}\left(\beta_D^0\right) + \epsilon\beta_D^1 \sin \mathbf{k}\cdot\mathbf{x}\, \frac{d\kappa_{eff}}{d\beta_D}. \tag{33}$$

The result (32) thus suggests that

$$\kappa_{eff} = 1 + 3\beta_D u_0\,, \qquad \frac{d\kappa_{eff}}{d\beta_D} = 3\beta_D \frac{u_c^1}{\beta_D^1}\,, \tag{34}$$

which are both results that we have verified numerically. If we also define a second effective thermal conductivity by

$$\kappa_\Delta = (1-\beta_D)\left(3\beta_D u_0 - \frac{u_c^2}{u_c^1 + u_c^2}\right), \tag{35}$$

we have the final result in the form

$$-\frac{1}{k_C}\mathbf{q}_m = \kappa_{eff}\,\nabla T_m + \kappa_\Delta\,\nabla(<T_D> - <T_C>)\,. \tag{36}$$

As noted before, for a uniform composite the second term is absent because the mean gradient is the same in the two phases.

RESULTS AND DISCUSSION

We generate an infinite composite by placing N spheres at random in a fundamental cell, and by filling up the whole space with copies of this cell. Experience shows that the results are not strongly affected by the artificial periodicity introduced in this way. The ensemble averages are calculated by the Monte Carlo method using 500 configurations, 32 particles per configuration, and truncating the summation in (12) to $n = 3$. The numerical results for κ_{eff} and κ_Δ obtained in this way are shown in Figs. 1 and 2 as functions of β_D for several values of $\kappa = k_D/k_C$. The former quantity, κ_{eff}, is the same as the result obtained by several authors for a uniform composite. The slight numerical differences between the present results and those published by Sangani and Yao are to be imputed to our use of 3, rather than 7, singularities. As noted before, we have also calculated β_D^1 and verified numerically the second relation in (34).

In the steady problem, of course, any temperature gradient is linearly dependent on either $\mathbf{G}$ or ∇T_m, and to that extent the relation (36) is less determinate than the result (29). However, (29) is not a general result, but only represents the solution to a specific problem. The connection between $\nabla(< T_D > - < T_C >)$ and $\mathbf{G}$ in any other situation must be worked out anew. Furthermore, in a time-dependent problem, the temperature gradients in the two phases are independent in general, and in this case (36) may be expected to be applicable, while no relation of the type (29) is likely. In this sense, it may be stated that a composite material of the type considered here satisfies Fourier's law of conduction with an effective conductivity only when it is uniform. Spatial non-uniformities in the distribution of the disperse phase give rise to a qualitatively new effect. Evidently, the actual prediction of the mean heat flux requires information on the spatial structure of both $< T_C >$ and $< T_D >$. Two separate energy equations for the two phases are therefore required for the full solution of the problem.

The only other study of this situation that we are aware of is a very recent paper by Buyevich and Ustinov (1995) who carried out a perturbation expansion on the assumption of a small gradient of β_D, rather than of a small non-uniform part as in (8). Their result is therefore different from the present one. In particular, they find an effect of order a^2/L^2, which is smaller than ours.

ACKNOWLEDGMENT

This work was supported by DOE under grant DE-FG02-89ER14043

REFERENCES

G.K. Batchelor (1974). Transport properties of two-phase materials with random structure. *Ann. Rev. Fluid Mech.*, 6:227–255.

G.K. Batchelor (1988). A new theory of the instability of a uniform fluidized bed. *J. Fluid Mech.*, 193:75–110.

Yu. A. Buyevich and V.A. Ustinov (1995). Effective thermal conductivity of a microscopically inhomogeneous dispersion. *Int. J. Heat Mass Transfer*, 38:381–389.

T.S. Lundgren (1972). Slow flow through stationary random beds and suspensions of spheres. *J. Fluid Mech.*, 51:273–299.

A.S. Sangani and C. Yao (1988). Bulk conductivity of composites with spherical inclusions. *J. Appl. Phys.*, 63:1334–1341.

D.Z. Zhang and A. Prosperetti (1996). Energy and momentum equations for disperse two-phase flows and their closure for dilute suspensions. *Int. J. Multiphase Flow*, submitted.

D.Z. Zhang and A. Prosperetti (1994a). Averaged equations for inviscid disperse two-phase flow. *J. Fluid Mech.*, 267:185–219.

D.Z. Zhang and A. Prosperetti (1994b). Ensemble phase-averaged equations for bubbly flows. *Phys. Fluids*, 6:2956–2970.

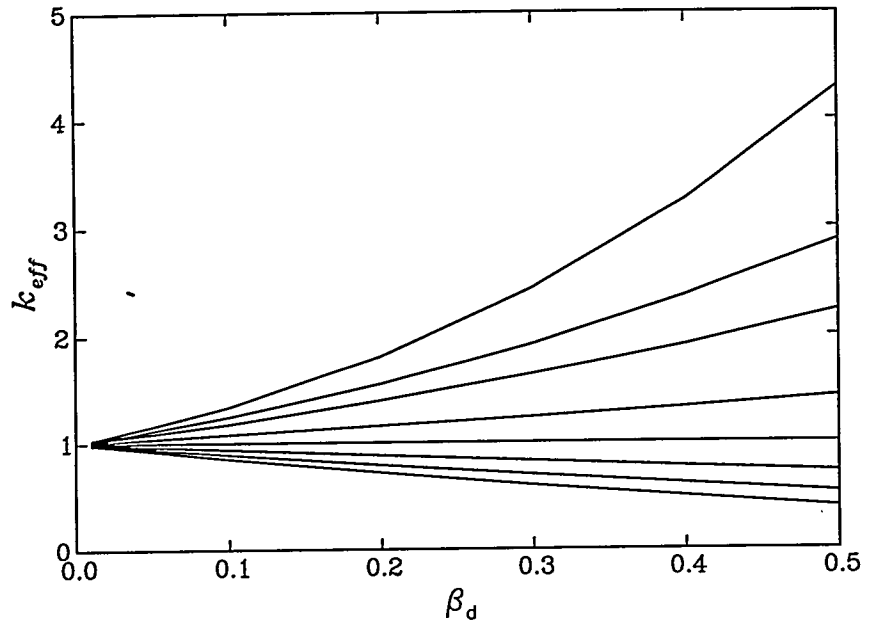


Figure 1: Effective normalized thermal conductivity of a uniform composite with spherical inclusions as a function of the volume fraction of the disperse phase β_D. The lines correspond, from bottom to top, to conductivity ratios k_D/k_C = 0, 0.2, 0.5, 1, 2, 5, 10, 1000.

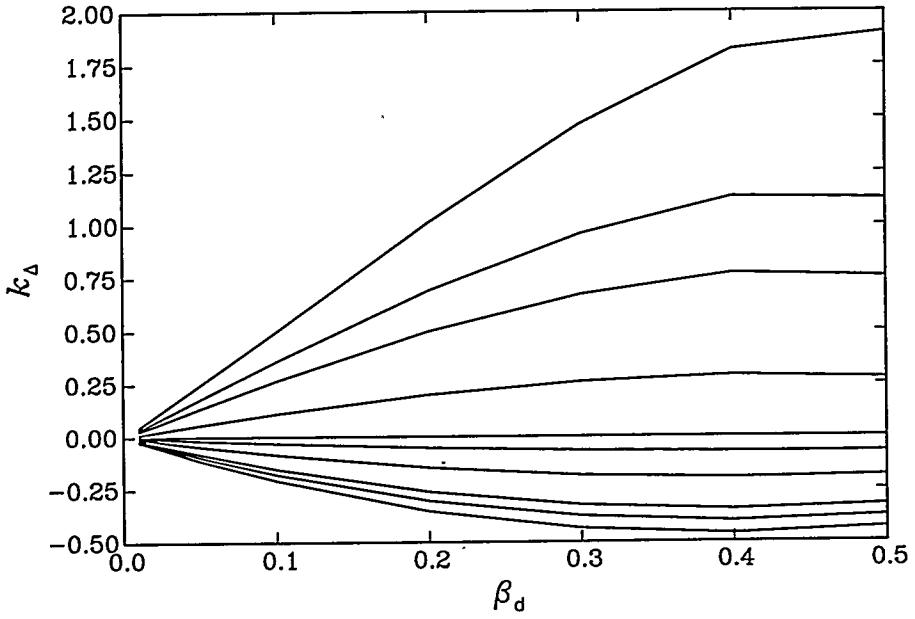


Figure 2: Effective normalized second thermal conductivity composite with spherical inclusions as a function of the volume fraction of the disperse phase β_D. The lines correspond, from bottom to top, to conductivity ratios k_D/k_C = 0, 0.1, 0.2, 0.5, 0.8, 1, 2, 5, 10, 1000.

FILM FLOWS AND SELF-ORGANIZED PATTERNS OF 2D-LOCALIZED STRUCTURES

Alexander L. Frenkel

Department of Mathematics
University of Alabama
Tuscaloosa, Alabama 35487-0350

ABSTRACT

Films flowing down an inclined plane are considered. An unconventional perturbation approach is discussed. It yields the most general evolution equation for film thickness and the least restrictive conditions for its validity. Results of numerical simulations of the dissipative-dispersive evolution equation indicate that novel, more complex type of spatiotemporal patterns can exist for strange attractors of nonequilibrium systems. It is suggested that real-life experiments satisfying the validity conditions of this theory are possible.

INTRODUCTION

Thin liquid layers ("films") flowing along solid surfaces occur in both natural and man-made environments (and the wavy film flows can captivate the occasional observer—this author, for one). Industrial applications of film flows started as long ago as the 1800s, and have been growing in their scope and importance ever since (see e.g. [1]).

Naturally, the studies of such film flows (the "Kapitza problem") have a considerable history. However, the nonlinear dynamics of wavy films is still far from being fully understood (see e.g. [2]; [3] and [4] are some recent reviews of film flow studies). The Navier-Stokes (NS) system of partial differential equations couples together the velocity components which are functions of four independent variables (time and the three spatial coordinates). In addition, there are boundary conditions (BCs), including those at the moving interface, whose position itself is determined by a partial differential equation (PDE) which involves the unknown velocity values. Such a three-dimensional (3D) problem is prohibitively difficult to simulate even with the most powerful presently available computers.

Even the simpler 2D computations were undertaken only recently, and only under the additional simplifying restrictions of short computational intervals and/or time-independence (see e.g. [5]). However, the 2D flows are frequently unstable to 3D disturbances, and therefore three-dimensionality can be important in many film flows (see e.g. recent experiments [2]).

Fortunately, there are certain domains in the parameter space for which the complicated NS-dynamical evolution of 3D film waves can be captured by much simpler *approximate* descriptions. In the most favorable cases, such a theory hinges on a *single* PDE which governs the evolution of film thickness, a function of at most *two* spatial coordinates. [The theory also leads to explicit expressions for the 3D velocity and pressure fields in terms of the (2D) film thickness.] Recently, we [3] obtained the most general of such evolution equations (EEs) for a film flowing down an inclined plane [which, in a certain sense (explained below), includes any other such equation that can be valid for all time]. Numerical simulations of that equation

revealed the spontaneous formation of ordered patterns consisting of self-organized coherent structures. These unusual patterns are the subject of the present communication.

The phenomenon of pattern formation in nonequilibrium, driven dissipative systems is currently a topic of active experimental and theoretical research (see e.g. [6] for a recent progress review). However, the self-organization of patterns was mainly studied in closed-flow situations, such as the Rayleigh-Bénard convection. The planforms studied up to now in fluid-dynamical experiments—as well as in solid state physics, nonlinear optics, chemistry, and biology—can be divided into two classes. In patterns of the first kind, the elementary unit is "one-dimensional", in the sense that one of its dimensions is much larger than the others (like a rope or a thread). The convection rolls are an example, as are the "spiral" and "target" chemical waves. The second kind of patterns are *two-dimensional* arrays of 2D structures—for example, hexagonal cells in some large-aspect-ratio convection experiments.

Our studies reported here show that patterns of coherent structures can spontaneously form in *film flows* (which, of course, are *open-flow* systems). Remarkably, these patterns are of a type different from both the above classes (and typically of a more complex character). In the rest of the text, we discuss the derivation of the most general EE (and of the conditions of its validity); some results of its numerical simulations; and some theoretical explanations and possible experiments regarding the novel patterns.

PERTURBATION THEORY

Consider a layer of an incompressible Newtonian liquid flowing down an inclined plane under the action of gravity. We introduce the coordinates as follows: the x axis is normal to the plane and directed into the film; the y axis is in the spanwise direction; and the z axis is directed streamwise. The corresponding components of velocity are u, v, and w. We denote the pressure field in the film by p; the pressure of the ambient passive gas is neglected for simplicity.

The system is determined by the following independent parameters: the average thickness of the film $\overline{h}_0$ (the overbar here and below indicates a *dimensional* quantity); the liquid density $\overline{\rho}$, viscosity $\overline{\mu}$, and surface tension $\overline{\sigma}$; gravity acceleration $\overline{g}$; and the angle of the plane with the horizontal θ.

There is a well-known, time-independent, constant-thickness solution of the NS problem called Nusselt's flow. The only nonzero component of velocity is the streamwise one. It only changes across the film, starting from the zero value at the solid plane. The free-surface value $\bar{U}$ of the Nusselt velocity is $\bar{U} = \bar{g}\bar{h}_0^2 \sin\theta/(2\bar{\nu})$ (where $\bar{\nu} = \overline{\mu}/\overline{\rho}$ is the kinematic viscosity). We nondimensionalize all quantities with units based on $\overline{\rho}$, $\overline{h}_0$, and $\bar{U}$ (e.g. $\overline{\rho}\bar{U}^2\overline{h}_0$ is the unit of measurement for the surface tension). Exactly three independent *basic parameters* (BPs) appear in the *dimensionless* equations and boundary conditions; one can choose e.g. the inclination angle θ, the Reynolds number $R \equiv \bar{h}_0\bar{U}/\bar{\nu}$ $[= \bar{g}\bar{h}_0^3 \sin\theta/(2\bar{\nu}^2)]$, and the Weber number $W \equiv \sigma R/2$ $[= \overline{\sigma}/(\overline{\rho}\bar{g}\bar{h}_0^2 \sin\theta)]$, as such BPs.

In [3] we discussed different perturbation approaches to obtaining single-EE approximations of the film dynamics. Here we briefly reiterate some points. The conventional approach uses formal series in powers of a single (small) long-wave parameter, say ε. In particular, the three basic parameters must each be prescribed certain orders of magnitude in terms of powers of ε—such as $W \sim \varepsilon^{-2}$. Thus, artificial inter-dependences are forced on those BPs. The three degrees of freedom in the parameter space (corresponding to the three BPs) collapse into just *one* degree of freedom, ε. Because of this, the validity conditions(VCs) for the derivation are unnecessarily restrictive [7] in this single-parameter approach (SPA). Also, if the exponents of powers prescribed to the BPs are changed, one can obtain a different EE. To determine if this is the case, one has to repeat the EE derivation for each new choice of the set of exponents—and there is an infinite number of such choices.

These drawbacks of the SPA can be remedied (see [3]). It is essential to notice here that each derivation of this type essentially amounts to neglecting certain terms in the NS equations for velocities, so that they become *ordinary* differential equations (ODEs) in x. These ODEs are linear and have constant coefficients. They can be easily solved in terms of the film thickness h and its derivatives. Substituting these expressions for velocities into the kinematic BC, one arrives at the single closed PDE for the thickness $h(y, z, t)$. So, it is possible to determine the *minimal* simplifications of the original NS equations, the terms which simply have to be discarded—and therefore require that they be small—if the goal is to arrive at the *solvable* ODEs (we call this requirement the *principle of derivability* of a single EE). It is clear that in this way, the most

universal single EE approximating the exact NS evolution is arrived at; every other valid EE (including those obtained by any single-parameter derivation) can be obtained by simply discarding some terms of the universal EE. This approach also yields the conditions of validity for the universal equation as well as the (stronger) VCs for each of its simplified versions. The validity conditions have the form of inequalities requiring that (a finite number of) certain dimensionless parameters must be *independently* small. Thus, this is a version of the multi-parameter perturbation approach (MPA) suggested and developed in our earlier papers over a number of years (see [3] and references therein).

As was discussed in those publications, one kind of condition is that of the *local* validity of the theory. These local-validity conditions (VCs) involve, along with the BPs, the *local* (in time) parameters, such as the characteristic lengthscale, timescale, and wave amplitude. The film flow is a dissipative system which evolves to an attractor and forgets the initial conditions. Accordingly, the characteristic scales gradually change from their initially prescribed values to the attractor-appropriate ones. Thus, these local parameters (LPs) can depend on time (before the system approaches the attractor), so that it may happen that the validity conditions cease to be satisfied after a finite time. If this is the case, the EE is *not* valid for all time (in other words, it is non-uniform in time). On the other hand, for the appropriate values of BPs (since, clearly, the attractor values of the LPs are determined by the BPs only), the local validity conditions can remain satisfied even when we substitute for the LPs their *attractor* expressions in terms of BPs; in this case the evolution equation is clearly valid for all time. In this way, one arrives at the *global*-validity conditions, which involve *only* the "global"parameters, the BPs. The result, then, is that the evolution equation is uniformly valid in time, provided that certain three groups of the three original BPs (we call those groups the modified BPs) are (independently) small.

It is clear that globally valid description of evolution by a single equation can only be possible in certain *restricted* domains of the space of BPs. We argued earlier (see [3] and references therein) that such a single-EE description cannot exist globally for those parametric regimes of inclined-film flow which lead to the amplitude of surface waves being "large"—comparable to the average film thickness. In the present communication, we are interested in the large-time behavior, when the system is already close to the attractor, and we want a single-EE description of the wavy film dynamics. Therefore, for the film thickness deviation, $\eta \equiv h - h_0$, we assume from the outset that its amplitude $A(t) \equiv \max|\eta|$ is small (for all time). Our derivation (a refinement of the one presented in [3]) will be described only briefly here; details will be given elsewhere [8]. It is essentially an iteration procedure. We write the fields in the form of sums containing the known Nusselt parts, e.g. $w = w_N + \tilde{w}_0$, where $w_N(x) = 2x - x^2$ is Nusselt's streamwise velocity; clearly, the $\tilde{w}_0$ is the unknown error in the approximation of w by w_N. The z-NS equation is rewritten in the form $\tilde{w}_{0xx} = \cdots$ [here and everywhere below, a subscript x, y, z, or (time) t, indicates differentiation with respect to that variable.], and all the terms (and *only* those terms) of the r.h.s. which contain the unknown error functions are discarded. The solution of the thus simplified equation with appropriate BCs (the—similarly simplified—tangential-stress balance at the free surface the no-slip at the solid plane) is found ($w_0 = 2\eta x$). Next, the approximation to the normal velocity is found from the incompressibility (no-divergence) equation, $\tilde{u}_{0x} = \cdots$; to the pressure from the x-NS equation, $\tilde{p}_{0x} = \cdots$(and the normal-stress BC); and to the spanwise velocity from the y-NS equation, $\tilde{v}_{0xx} = \cdots$. The procedure can be repeated, leading to increasingly refined approximations; e.g. $w \simeq w_N + w_0$, $w \simeq w_N + w_0 + w_1$, etc., where $w_{i+1} \ll w_i$ $(i = 0, 1, 2, \cdots)$. Similarly, $u = u_0 + u_1 + u_2 + u_3$ (note $u_N = v_N = 0$), etc.

As a result (by substituting the velocity expressions in terms of η into the kinematic BC), one obtains the evolution equation in the form

$$\eta_t + 4\eta\eta_z + \frac{2}{3}\delta\eta_{zz} - \frac{2}{3}\cot\theta\eta_{yy} + \frac{2}{3}W\nabla^4\eta + 2\nabla^2\eta_z = 0 \qquad (1)$$

where $\nabla^2 = \partial^2/\partial z^2 + \partial^2/\partial y^2$ and $\delta \equiv (4R/5 - \cot\theta)$. (We will always assume $\delta > 0$, because otherwise the infinitesimal disturbances would not grow and the interesting finite-amplitude waves would never appear.) Equation (1) holds in a reference frame moving (with respect to the solid plane) with the velocity $V = 2$ in the streamwise direction. [This choice of the reference frame removes the trivial fast-time oscillations of film thickness (at a fixed station) arising because of the uniform translation of waves past the station with their phase velocity (cf. [9]).]

The local-validity conditions yielded by the theory are $\max\left[A,\ 1/L^2,\ R/L,\ W/L^3\right] \ll 1$. Here L is the characteristic z-lengthscale (such that $\partial/\partial z \sim 1/L$; $\eta_{zz} \sim L^{-2}$, etc.; and we have assumed for the sake

of simplicity that the characteristic y-lengthscale is $\geq L$—which has been the case in all experiments we know). Note that the requirement $L^2 \gg 1$ appears as a necessary consequence of the derivability principle. Thus, no single-EE theory can avoid the small-slope requirement.

The analysis of derivation of the above EE shows that its third term originates from the inertia members of the NS equations; this term is destabilizing, as is readily seen from the linear stability theory. The (stabilizing) fourth and fifth terms are due to hydrostatic and capillary (i.e. surface-tension) parts of the pressure, respectively. Finally, the last, odd-derivative term is due to the *viscous* part of the pressure. This term is purely *dispersive:* it does not lead to either growth or decay of the amplitude of surface deviation η—in contrast to the dissipative, third through fifth terms (we regard the production of energy by the destabilizing terms as negative dissipation). Such a term also appeared in the EE obtained by Topper and Kawahara [10] for the case of an almost vertical plane: they used the small angle of the plane with the *vertical* as their (single) perturbation parameter. Also, under their assumptions all the terms of the EE have to be of the same order of magnitude. Our derivation shows that the Topper-Kawahara (TK) equation of the form (1) holds under much less restrictive conditions than those stipulated in [10]. Different limiting cases of Eq. (1) are identified as simpler well-known equations (see [3]).

When this (dissipative) system has evolved sufficiently close to the attractor, the average wave amplitude ceases to change in time. So, the destabilizing and stabilizing terms of the EE must be of the same order of magnitude. Estimating the derivatives in terms of the lengthscale as was mentioned above yields $L_a \sim \sqrt{W/\delta}$ for the value L_a of the lengthscale that is characteristic of the attractor. The amplitude A_a on the attractor is determined from the balance of the nonlinear term of EE (1) with the dispersive or dissipative (linear) terms, whichever is greater: $A_a \sim \max(L_a^{-2}, W/L_a^3)$. By substituting these expressions into the local validity conditions above, the global validity conditions are obtained. Namely, the following modified basic parameters—which we denote as α and β—are required to be small: $\alpha \equiv 1/L_a^2 (= \delta/W) \ll 1$ and (noting that $\delta < R$ and $W/L_a^3 \sim \delta/L_a, < R/L_a$) $\beta \equiv R/L_a (= R\sqrt{(\delta/W)}) \ll 1$.

In certain domains of the space of basic parameters, the dispersive term is small; then it can be neglected. [This was the case when we numerically simulated our EE (1) with the parameter values pertaining to the recent experimental studies [2] of an inclined-film flow. Although their global-VC parameter β is ~ 1 rather than $\ll 1$, we had a qualitative agreement with the transient phenomena observed in those experiments (see [3] for details of those results.)]

However, if the *dissipative* terms are small, they still play an important role, on a slower time scale. Namely, the 2D KdV equation obtained by neglecting the dissipative terms, similarly to the usual case of 1D KdV equation, has a whole *family* of soliton solutions (these axially-symmetric solutions were found numerically in Ref. [11]). The wider soliton is shorter and moves slower. If disturbed, it relaxes on a faster time scale to the solution corresponding to the new value of the family parameter. The additional small dissipative terms of the TK equation gradually change the parameter—say, the lengthscale—of the soliton, until the equilibrium value L_a, determined by the balance between the stabilizing and destabilizing dissipative terms, is reached. (For the 1D case, this phenomenon was first described in Ref. [12].) Therefore, the derivation must correctly determine the small dissipative terms (to their leading order). This is not guaranteed by the derivation of Eq. (1) in which the (cross-stream) velocity approximation was truncated at u_1. Therefore, we [8] have analyzed higher-order iterations.

We find that after taking into account, e.g., the corrections u_2 and u_3, the EE reads

$$\eta_t + 4\eta\eta_z + (2/3)\delta\eta_{zz} - (2/3)\cot\theta\eta_{yy} + [2\nabla^2\eta_z + (40/63)R\delta\eta_{zzz} - (40/63)R\cot\theta\eta_{zyy}]$$

$$+[(2/3)W\nabla^4\eta - (6/5)\cot\theta\nabla^4\eta + (157/56)R\nabla^2\eta_{zz} + (8/45)R\cot^2\theta\nabla^4\eta$$

$$+(1213952/2027025)R^3\eta_{zzzz} - (138904/155925)R^2\cot\theta\nabla^2\eta_{zz}] + (8/5)R(\eta\eta_z)_z = 0. \qquad (2)$$

[The 1D ($\partial/\partial y = 0$) version of this equation, with the same numerical coefficients, was obtained before [13], but our 2D version is new. The same numerical coefficients, in a linearized 1D context, are found in an even earlier paper [14].] There are additional dissipative terms, e.g. $(157/56)R\nabla^2\eta_{zz}$, which in certain ranges of BPs can be larger than the original stabilizing term, $(2/3)W\nabla^4\eta$ (which is clearly the case for $W \ll R$). Can they provide saturation by balancing the destabilizing term? Our answer is negative: equation (2) can only be valid locally. It turns out that the last (nonlinear) term in the Eq. (2) is destabilizing and blows up the solution. Physically, all the important dissipative terms (which, by analyzing the iterations through all

orders, we have shown to be of the form—for simplicity, in the 1D case—$R^{2n-1}\partial^{2n}\eta/\partial z^{2n}$), are traced back to *inertia*, the same factor which is responsible for the destabilizing (linear, second-derivative) term of the EE (2). It is unlikely that the same physical factor will provide both the destabilization and a balancing stabilizing term. We conclude that the long-time behavior in such cases cannot be described by a single-EE theory (although a conclusive mathematical demonstration would require massive calculations to determine the sign of the higher-order dissipative terms, which we have not undertaken). To exclude such parametric regime, one requires $\max(R, R^3) \ll W$.

Thus, EE (1) is the universal (most general) all-time valid evolution equation of the inclined-plane film flow. The global validity conditions ($\alpha \ll 1$, $\beta \ll 1$, and $\max(R, R^3) \ll W$) are all satisfied if we take

$$\alpha_R \equiv R/W \ll 1 \tag{3}$$

$$\beta_R \equiv R\sqrt{(R/W)} \ll 1 \tag{4}$$

(recall that $\delta < R$).

A possible simplification of the universal EE is obtained by omitting its dispersive term; this results in a KS-type equation $\eta_t + 4\eta\eta_z + (2/3)[\delta\eta_{zz} - \cot\theta\eta_{yy} + W\nabla^4\eta] = 0$. [It can be seen that the additional (global) VC expressing the smallness of dispersion is $(W\delta)^{-1/2} \ll 1$.] The long-time behavior of the KS equation is chaotic. In the present paper, we are interested in the opposite case—we show that ordered patterns arise under the parametric conditions of the *large* dispersive term, $\sqrt{(W\delta)} \ll 1$. (As we noted earlier, the dissipative terms, albeit small, must be retained in the EE, to their leading order.)

In addition to large dispersiveness (in other words, small dissipativeness), we limit ourselves in the present paper to the case of a *vertical* plane, $\cot\theta = 0$ (note that $\delta = 4R/5$ in this case). We can transform Eq. (1) to a "canonical" form, which contains only one "tunable" constant—by rescaling $\eta = N\widetilde{\eta}$, $z = L_u\widetilde{z}$, $y = L_u\widetilde{y}$, and $t = T\widetilde{t}$, where $N = 2R/(5W)$, $L_u = \sqrt{5W/(4R)}$, and $T = (5^{3/2}/16)(W/R)^{3/2}$. Dropping the tildes in the notations of variables, the resulting canonical form of the EE is

$$\eta_t + \eta\eta_z + \nabla^2\eta_z + \epsilon\left(\eta_{zz} + \nabla^4\eta\right) = 0. \tag{5}$$

The single control parameter in this equation is

$$\epsilon = (2/3)\sqrt{WR/5}, \tag{6}$$

and we will be mostly concerned with the parametric domains [in the (W, R)-space] for which $\epsilon \ll 1$.

NUMERICAL SIMULATIONS AND SOME RESULTS

We have carried out numerical simulations of Eq. (5) with periodic boundary conditions. To exhibit interesting spatial behavior, a system should be sufficiently "large". For the periodicity domain $0 < y < 2\pi p$ and $0 < z < 2\pi q$ in our simulations of Eq. (5), we chose $5 \le p \le 16$ and $16 \le p \le 80$. We used spatial grids of up to 256×256 nodes, with the Fourier pseudospectral method for spatial derivatives and with appropriate dealiasing. The initial values of η were chosen—independently at each spatial node—from the interval $[-0,05, 0.05]$ with a uniform probability distribution. Time marching was done (in the Fourier space) by using Adams- Bashforth and/or Runge-Kutta methods. We checked the results by refining the space grids and time steps; by verifying the volume conservation, $\int \eta dydz = 0$; etc. A typical simulation run took $\sim 10^5 - 10^6$ time-steps.

The main focus of the present communication is the presence of highly nontrivial patterns in time-asymptotic states for the *strongly dispersive* cases, $\epsilon \ll 1$. Figure 1 shows snapshots of the film surface at large times for three different sets of parameter values. We will speak of such numerically identified time-asymptotic states as *attractors*, although one needs to be cautious here: It is known that such extended systems may sometimes exhibit long transients.

There are two subpatterns in Fig. 1a: The V-shaped formation consisting of 13 large-amplitude bulges aligned into two straight lines moves as a whole downstream with a certain velocity, and the small-amplitude

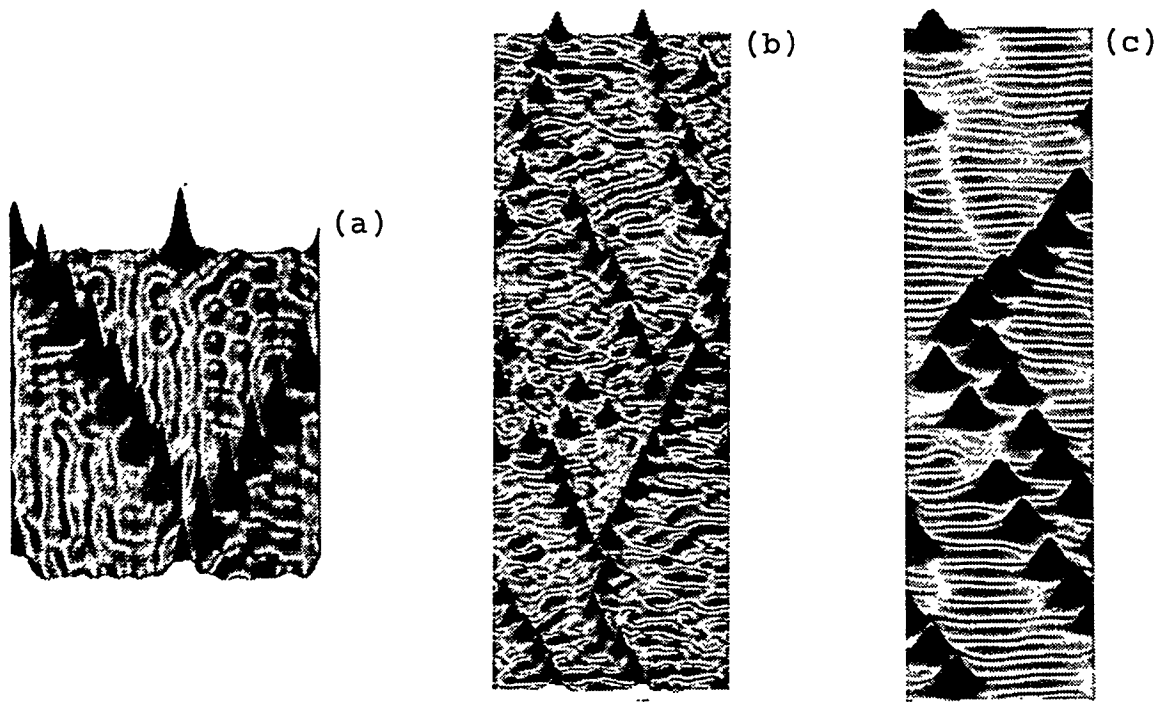


Fig. 1. Snapshots of the time-asymptotic film surface self-organized in simulations of Eq. 5, for three different cases (bulges move down the page here; for convenience of presentation, different axes may have different scales; in reality, all structures have small slopes). (a) $p = q = 16$, $\epsilon^{-1} = 50$, and $t = 1.6 \times 10^5$; (b) $(p, q) = (16, 80)$, $\epsilon^{-1} = 30$, and $t = 5.98 \times 10^4$; (c)$(p, q) = (5, 60)$, $\epsilon^{-1} = 25$, and $t = 4.89 \times 10^5$.

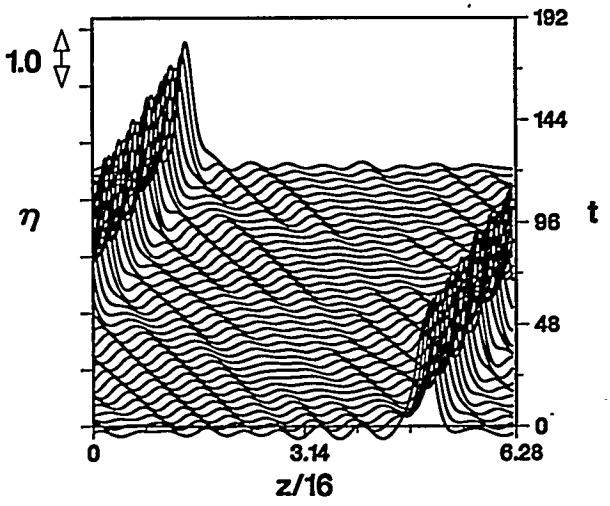


Fig. 2. Time-sequence of instantaneous surface profiles in a fixed cross-section normal to the streamwise direction (for $p = q = 16$, $\epsilon^{-1} = 50$; the time shown as 0 is in fact 1.6×10^5 counting from the start of the run). In particular, it is evident that the (large-amplitude) bulges move in one direction and (small-amplitude) bumps in the opposite direction.

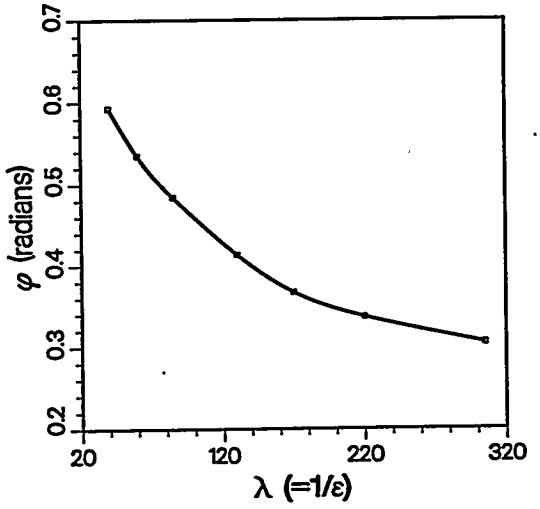


Fig. 3. Angle φ between (every) line of bulges and the streamwise direction varies with ϵ ($p = q = 16$).

background moves uniformly as well, but in the opposite (if the reference frame is appropriate) direction . (Similar segregation of coherent structures into two subpatterns of different amplitudes is also seen for the non-square, large-aspect-ratio domains, Figs. 1b and 1c.) This movement of the two subpatterns in opposite directions is evident in the time-delay plots shown in Fig. 2. The background subpattern in Fig. 1a consists of small-amplitude "bumps" which form a lattice-like structure. Even though the bumps change their shape and interact with each other in an irregular manner, they seem to maintain their identity. In particular, they do not seem to coalesce or break up (see Fig. 2). As one sees in Fig. 2, the bumps weakly interact with one another. Also, the height of a bulge *chaotically fluctuates* (see more in [3], where these fluctuations are also reflected in the energy plot, Fig. 2 there).

We investigated how the pattern changes as the value of ϵ for the above simulations (with $p = q = 16$) is varied from $\epsilon \gg 1$ to a small value, $\epsilon = 1/305$. When $\epsilon \gg 1$, the 1D version of Eq. (5) is essentially the KS equation, which on extended spatial intervals yields *chaotic* attractors. Accordingly, we observe chaotic (although 2D) waveforms in our simulations of Eq. (5) with large ϵ. For smaller ϵ (larger dispersiveness), however, the ordering effect of dispersiveness becomes increasingly evident: The amplitude separation into bulges and smaller-amplitude background structures becomes noticeable for $\epsilon \sim 1/5$ and continues to grow as ϵ is decreased. At smaller value of ϵ ($\sim 1/25$), the bulges start lining up, even though each straight-line segment (which is inclined by the characteristic angle φ to the streamwise direction) consists of just two or three bulges. It appears that for the longer lines (the V-shaped pattern) to form, the value of ϵ has to be sufficiently small, $\epsilon < 1/30$.

It is natural to inquire as to how the various quantities of the pattern scale with ϵ. We varied ϵ^{-1} between 25 and 305 for $p = q = 16$. In one set of simulations, ϵ^{-1} was gradually decreased from 50 in relatively small steps of 5 (to allow the system to "adiabatically" adjust to the new parameter value), up to $\epsilon^{-1} = 25$, at which point the line formations of bulges break down. In another set of simulations, ϵ^{-1} was increased from 30 in steps of 10 or 15 up to $\epsilon^{-1} = 305$. In all cases, we find that the characteristic width of the bulge as well as that of the bump is ~ 1, independent of ϵ. The amplitude of bulges is also constant, ~ 1, as is the velocity of bulges and that of bumps (of course, the *signs of* these two velocities are different). Only the bump amplitude varies: it scales like $\sim \epsilon$.

The V-shaped formation of bulges retains its form when ϵ is changed from 1/30 to 1/305. However, the (absolute value of the) angle of each bulge-filled line with the streamwise axis decreases with ϵ, probably approaching some asymptotic value in the limit $\epsilon \to 0$ (see Fig. 3. Since there is no parameters remaining in this limit, the asymptotic angle should be just 0.) This decrease of the angle seems to be determined mainly by the increase of the *streamwise* separation between the neighboring bulges, while the spanwise one stays approximately constant

When $\epsilon \ll 1$, the dissipative terms in Eq. (5) can be treated as perturbations $\sim \epsilon$ of the 2D KdV equation $\eta_t + \eta\eta_z + \nabla^2\eta_z = 0$. This equation does not seem to have any analytical solutions. However, by transforming to a reference frame moving with a velocity $c > 0$ [replace η_t with $(-c\eta_z)$ in the equation], Petviashvili and Yankov [11] numerically obtained a stationary axially-symmetric solitary-wave solution. By balancing the first term with the nonlinear term, $c\eta_z \sim \eta\eta_z$, and the latter term with the dispersive term, the characteristic amplitude and velocity of these solutions are found to be $\eta \sim c$ and $c \sim 1/L_s^2$ where L_s is the characteristic lengthscale, which is not uniquely determined by the KdV equation. However, as was discussed above, there is also the balance between the dissipation terms in Eq. (5), a necessary outcome of slow-time evolution. It selects uniquely the soliton of $L_s \sim 1$, which results in $c \sim 1$ and $\eta \sim 1$ as well, independent of ϵ. These estimates are consistent with the numerical results reported above.

Motivated by the discovery of the second, small-amplitude subpattern, we examined the possibility of a corresponding second travelling-wave solution. If we transfer to the frame moving with a *negative* velocity $c = -a^2$, where a is a constant, there are such solutions—with the nonlinear term being as small as the dissipative ones. Indeed, the leading-order equation then is $\nabla^2\eta_z + a^2\eta_z = 0$, which is the well-known Helmholtz equation for η_z. There are solutions $\propto \sin Jy \sin Kz$ ($J^2 + K^2 = a^2$). The balance between the (small) dissipative terms again determines $K \sim c \sim 1$, and the balance of the dissipative terms with the nonlinear term yields $\eta \sim \epsilon$. We see that these lengthscale, amplitude and velocity (including its sign) agree with those observed for the bumps in the numerical experiments as described above.

Note that our assumption of the *negative* velocity was essential: with a positive velocity, one arrives at the *modified* Helmholtz equation, which does not have any oscillating solutions. There are only *exponential* solutions, which are physically unsuitable here. [We note that the Helmholtz equation has axially-symmetric

solutions as well, $\propto J_0(ar)$ where J_0 is the Bessel function (r is the radial coordinate). This solution is only weakly localized: it decays at spatial infinity like a power rather than exponentially.]

Similar to Eq. (5), we have derived an equation for a film flowing down a vertical *cylinder* (see Ref. [3]). In particular, one can see that if the (dimensionless) radius b of the cylinder is not too small ($b \gg \beta^{-1}$), the flow is well approximated by the planar-film equation1 (with periodic BC in the azimuthal direction; we note that this also justifies our use of *spanwise*-periodic BCs in the numerical simulations discussed above. As to the streamwise BC, we believe the solution is insensitive to the BC type in the limit of large aspect ratio q/p.). One finds that with $\overline{h_0} \sim 1$ mm, the cylinder radius $\overline{b} \sim 1$ cm, and under parametric conditions $\alpha_R \ll 1$, $\beta_R \ll 1$, and $\epsilon \ll 1$, for the waves (evolving as they propagate from the entrance end of not-too-long a cylinder to the exit end) to have enough time to approach the attractor stage, the liquid should be several hundred times as viscous as water. For example, it could be glycerin with an admixture of water. [It is interesting to note that one can see a straight row of bulges in the photograph of a film flowing down a cylinder, Fig. 2 of Ref. [15]; however, the β-condition of validity was not strictly satisfied for the parameter values of those experiments.]

SUMMARY AND CONCLUDING REMARKS

Numerical 2D simulations of a realistic evolution PDE indicate that nonequilibrium dissipative systems can spontaneously form spatial patterns which are significantly more complex than those known before. Namely, whereas the conventional patterns (such as thermal-convection rolls, "target" and "spiral" chemical reaction-diffusion waves, etc.) are essentially almost periodic and stationary (at least locally), and are either arrays of 1D structures or 2D arrays of 2D structures—the novel patterns exhibit soliton-like, 2D spatially-localized excitations which can spontaneously line up into 1D arrays. They make *non*periodic, but nevertheless highly ordered arrangements. Moreover, these patterns typically consist of subpatterns—each of a different amplitude and each moving as a whole with its own velocity—"percolating " through one another. Thus, the overall complex spatiotemporal pattern is *non-stationary* in any reference frame, even if consideration is restricted to small domains containing only a few soliton-like structures. This is in contrast to any other pattern we could find in the literature. (For an example of the conventional pattern, if one considers a *small* piece of a chemical spiral wave, much smaller than the spiral radius, that local part of the spiral pattern will appear almost stationary—and almost periodic as well—in an appropriate reference frame.) It is interesting to note that the observed complex order appears "on the edge of chaos" (which overtakes at smaller dispersiveness), in accordance with some ideas of the recent "science of complexity".

The particular dissipative-dispersive evolution PDE under consideration here has been consistently derived from the full Navier-Stokes problem. It is the most general single EE to provide a controllably close approximation to the evolution of a liquid film flowing down an inclined plane. The unconventional perturbation approach used in this derivation has the advantage of yielding clear and comparatively non-restrictive parametric conditions of the validity of the theory. To satisfy those validity conditions for a possible (terrestrial) experiment designed to observe patterns of the novel type on a film flowing down a vertical cylinder, the film liquid should be much more viscous than water (e.g. a glycerin-water solution).

To construct a theory of *interactions* of coherent structures which could explain the observed patterns remains a fascinating challenge.

ACKNOWLEDGMENTS

The research reported above was partially supported by the U.S. DOE and carried out in collaboration with Dr. K. Indireshkumar.

References

[1] G. D. Fulford. The flow of liquids in thin films. In *Advances in Chemical Engineering*, vol. 5 (T. B. Drew, J. W. Hoopes, and T. Vermeulen, eds.; Academic Press, New York, 1964) pp. 151–236.

[2] J. Liu, J. B. Schneider, and J. P. Gollub. Three dimensional instabilities of flowing films. *Phys. Fluids.*, **7**, 55–67 (1995).

[3] A. L. Frenkel and K. Indireshkumar. Derivations and simulations of evolution equations of wavy film flows. In *Mathematical modeling and simulation in hydrodynamic stability* (Daniel N. Riahi, ed.; World Scientific, Singapore, 1996) pp. 35–81.

[4] H.-C. Chang. Wave evolution on a falling film. *Annu. Rev. Fluid Mech.*, **26**, 103–136 (1994).

[5] T. R. Salamon, R. C. Armstrong, and R. A. Brown. Travelling waves on vertical films: Numerical analysis using the finite element method. *Phys. Fluids*, **6**, 2202–2220 (1994).

[6] M. C. Cross and P. C. Hohenberg. Pattern formation outside of equilibrium. *Rev. Mod. Phys*, **65**, 851–1112 (1993).

[7] A. L. Frenkel. On asymptotic multiparameter method: Nonlinear film rupture. In *Proc. Ninth Symp. on Energy Engineering Sciences, CONF-9105116 (Argonne Natl. Lab., Argonne, Illinois, 1991)*, pp. 185–192.

[8] A. L. Frenkel and K. Indireshkumar. Wavy film flows down an inclined plane. part I: Perturbation theory and the universal evolution equation for the film thickness. To be published.

[9] A. L. Frenkel. On evolution equations for thin films flowing down solid surfaces. *Phys. Fluids A*, **5**, 2342–2347 (1993).

[10] J. Topper and T. Kawahara. Approximate equations for long nonlinear waves on a viscous film. *J. Phys. Soc. Japan*, **44**, 663–666 (1978).

[11] V. I. Petviashvili and V. V. Yan'kov. Two-layered vortices in a rotating stratified liquid. *Dokl. Akad. Nauk. SSSR*, **267**, 825–828 (1982).

[12] T. Kawahara. Formation of saturated solitons in a nonlinear dispersive system with instability and dissipation. *Phys. Rev. Lett.*, **51**, 381–383 (1983).

[13] C. Nakaya. Long waves on a thin fluid layer flowing down an inclined plane. *Phys. Fluids*, **18**, 1407–1420 (1975).

[14] D. J. Benney. Long waves in liquid films. *J. Math. Phys.*, **45**, 150–155 (1966).

[15] A M. Binnie. Experiments on the onset of wave formation on a film flowing down a vertical plane. *J. Fluid Mech.*, **2**, 551–553 (1957).

2-D Traveling-Wave Patterns in Binary Fluid Convection

C. M. Surko and A. La Porta
Department of Physics, University of California, San Diego, La Jolla CA 92093

Abstract

An overview is presented of recent experiments designed to study two-dimensional traveling-wave convection in binary fluid convection in a large aspect ratio container. Disordered patterns are observed when convection is initiated. As time proceeds, they evolve to more ordered patterns, consisting of several domains of traveling-waves separated by well-defined domain boundaries. The detailed character of the patterns depends sensitively on the Rayleigh number. Numerical techniques are described which were developed to provide a quantitative characterization of the traveling-wave patterns. Applications of complex demodulation techniques are also described, which make a detailed study of the structure and dynamics of the domain boundaries possible.

Introduction

When a spatially extended system is driven far from equilibrium a breaking of translational symmetry is sometimes observed which results in the formation of a pattern. In contrast to equilibrium systems, which are governed by a free-energy minimization principle, patterns in nonequilibrium systems typically exhibit nonrelaxational dynamics. As a result, a much richer variety of phenomena is observed, and it has proven very difficult to understand the general principles of pattern selection in nonequilibrium systems. Despite this, the effort to find relationships between the patterns and the symmetries of the system in which they occur has been very successful[1], especially in systems in which the primary instability is stationary. It is important to extend this work to the broad class of systems in which the primary instability is oscillatory and the patterns consist of traveling waves. Work in this area has potential relevance to many important applications, such as reaction-diffusion systems, large aspect ratio lasers, and oceanographic flows.

Convection in binary mixtures of ethanol and water is an example of a pattern forming system with an oscillatory instability. Binary fluid convection is a double-diffusive system in which two quantities (heat and concentration) diffuse in the fluid and are advected by the velocity field. In the ethanol-water system, there is a coupling between temperature and solute concentration known as the Soret effect. Therefore, if the convection cell is heated from below a concentration gradient forms in the fluid layer, the effect of which is parameterized by the separation ratio,

$$\psi \equiv -c\,(1-c)\,S_t\frac{\beta}{\alpha}, \tag{1}$$

where c is the ethanol concentration, S_t is the Soret coefficient, and α and β are the thermal and concentration expansion coefficients, respectively[1]. For the 8% ethanol mixture studied here, $\psi = -0.24$, and the ethanol concentration gradient tends to stabilize the fluid layer against thermal convection[2]. This stratification of the fluid layer and the strong separation of time scales for

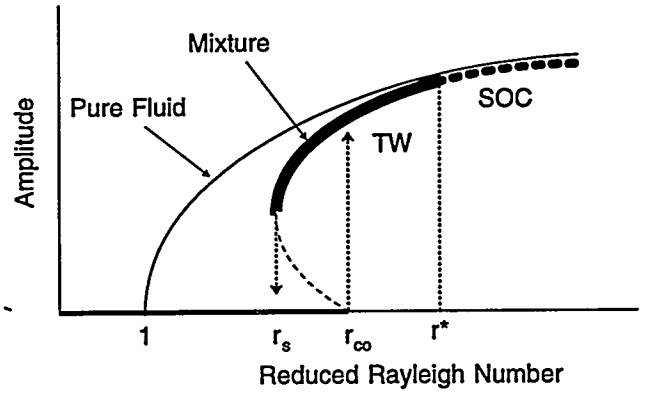


Figure 1: Schematic bifurcation diagram comparing convection in a pure fluid and in a binary mixture with negative separation ratio. In the mixture, the heavy solid line indicates traveling-wave (TW) convection, and the heavy dashed line indicates stationary overturning convection (SOC).

ethanol diffusion and heat diffusion significantly influence the onset of convection in the mixture. In this case, a subcritical Hopf bifurcation to a state of oscillating convection[3] is observed at onset, as indicated in Fig. 1. After convection begins in the mixture, the interaction of the Soret effect with the convective flow can produce complex phenomena such as chaotic growth and collapse of traveling waves[2], and the formation of pulses in one or two dimensions[4, 5]. The Soret effect is responsible for the strongly nonlinear traveling-wave state which is observed for large negative ψ and also influences the texture of stationary convection patterns which are observed above r^*[6], as indicated in Fig. 1.

Convection in fluid mixtures has several advantages as a system in which to study the dynamics of traveling-wave patterns. The underlying physical equations governing the system, the Navier-Stokes equation coupled with the equations for the diffusion of mass and heat, are well known, and the physical parameters of the system can be precisely controlled. Furthermore, the degree of nonlinearity of the traveling-wave state is dependent on ψ, which may be varied over a wide range by changing the concentration of ethanol in the mixture[2].

Apparatus

The experiments were performed in a convection cell consisting of a resistively heated bottom plate and a sapphire top plate which is cooled by a temperature regulated water flow bath. A window in the flow channel provides optical access for visualization of the convection cell through the bath. The cylindrical convection container has an unusually large diameter of 21 cm and a height of 0.4 cm, which corresponds to an aspect ratio (r/h) of 26. The cell was specifically designed to have a large aspect ratio, in order to study traveling-wave patterns which are separated as much as possible from the influence of the physical boundaries of the cell.

The convection cell has several unique features. The bottom plate is a 1.91 cm thick polycrystalline silicon cylinder which is mirror polished to a flatness of one wavelength per inch. Although the thermal conductivity of silicon is a factor of 2.5 smaller than that of copper, it is still adequate. Silicon is very hard and does not interact with water, so that the quality and durability of the mirrored surface is far superior to that typically obtained from plated copper mirrors. The main technical difficulty in this experiment is maintaining a constant and uniform Rayleigh num-

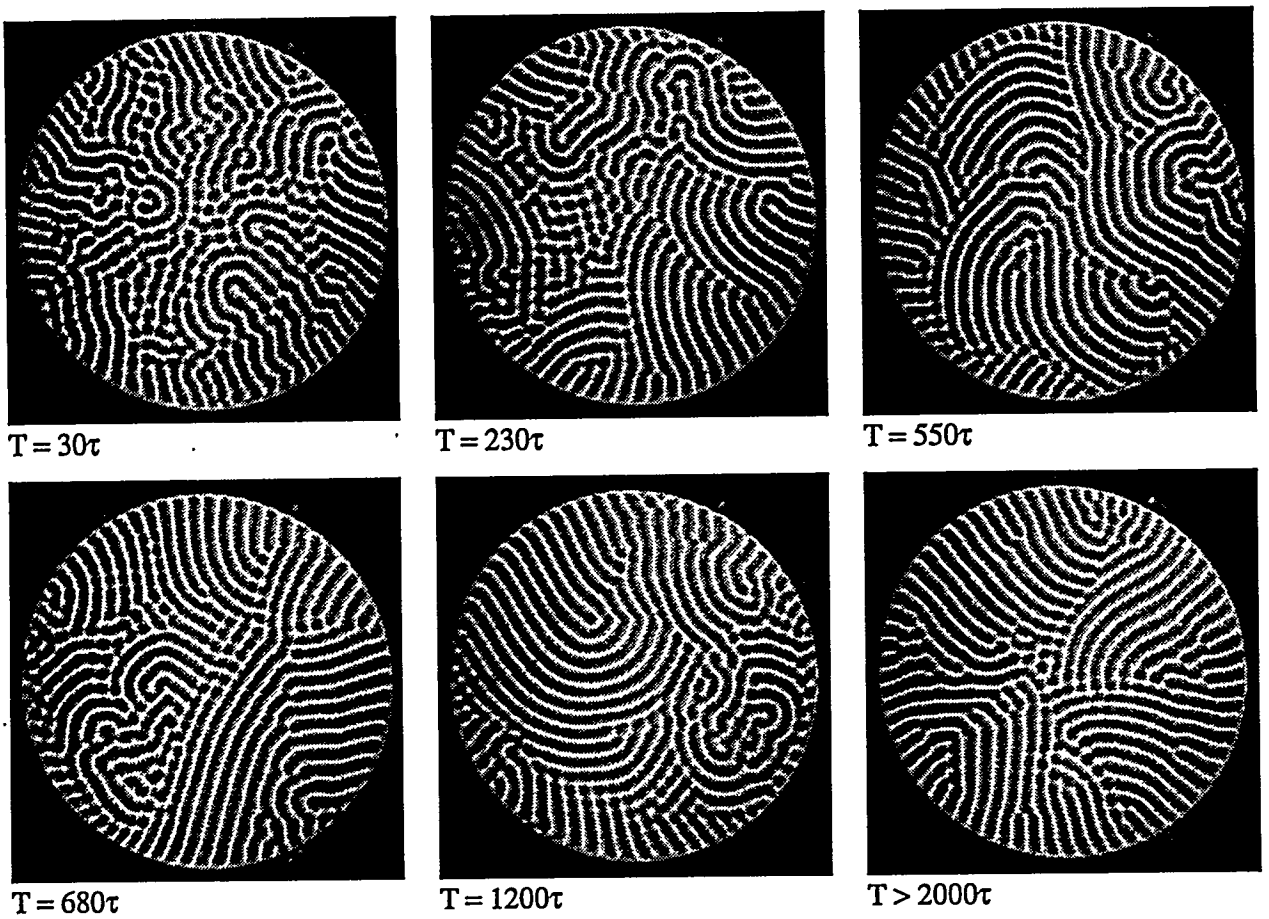


Figure 2: The long-term evolution of a traveling-wave pattern at r=1.28. Time is expressed in terms of the vertical thermal diffusion time, $\tau = 124$ sec.

ber across the large convection cell. To this end, a linear channel flow was employed to cool the top plate. This flow is much simpler in structure than the circularly symmetric flows which are typically employed in cylindrical convection cells. By using a flow with a high Reynolds number, the temperature variation over the top plate is small and predicatable, and can be compensated by creating a matching temperature distribution on the bottom plate. Our measurements indicate that the peak-to-peak variation in the Rayleigh number over the cell is 0.2%. The visualization of such a large convection cell also presents some unique problems. The cell's diameter of 21 cm makes refracting optics impractical. Therefore we have employed a parabolic mirror as the main focusing element of our white-light shadowgraph. Images are acquired using a CCD camera and digitized with a PC frame grabber. A time-lapse VCR is used to monitor the long term evolution of the patterns. A flexible computer process control system has been developed which controls the experiment and acquires data. The apparatus has been described in detail elsewhere[7].

Survey of Traveling-Wave Patterns

The first experiments in traveling-wave convection in binary fluid mixtures observed 2-D patterns in small and medium aspect-ratio convection cells[8, 9]. Early on, investigators found that the patterns and their associated dynamics were very complex, and experiments quickly turned to

1-D patterns, either in narrow channels[10] or in an annular geometry[11]. Our experiment has a significantly larger aspect ratio than previous work on 2-D TW convection[4], and so our first task has been to make a survey of phenomena in the system.

The evolution of a typical traveling-wave pattern in a cylindrical container is shown in Fig. 2. The basic conclusion is that extremely disordered patterns are normally created when traveling-wave convection is initiated, but that a coarsening of the pattern occurs over time scales of the order of thousands of vertical thermal diffusion times. One of the important mechanisms for the coarsening of the pattern is the launching of organized rolls from sources on the boundaries which tend to sweep chaotic fluctuations to the boundary of the pattern. This mechanism makes an interesting contrast to the case of rotating Rayleigh-Bénard convection[12], where the coarsening occurs via the growth of domains in the bulk of the pattern. The coarsening of the TW pattern continues until the pattern has organized itself into a few domains of straight or slightly curved rolls, separated by well-defined domain boundaries. Such a pattern is shown in the final panel of Fig. 2.

We have found that organized, multi-domain patterns are formed over the entire traveling-wave branch of convection, but that the character of the patterns is sensitive to the Rayleigh number[7]. The domain boundaries in Fig. 2 typically separate patches of rolls which are nearly perpendicular to each other, with the domain boundary parallel to one set of rolls. At slightly higher Rayleigh number, near $r = 1.35$, "zipper" boundaries are more commonly observed, in which counter propagating rolls shear past each other, with the domain boundary parallel to the direction of propagation. At higher Rayleigh numbers, near r^*, there are typically no clear domain boundaries. Here the convection patterns are organized around point defects and rotate more or less rigidly. Initially, we have concentrated our effort on organized patterns such as the one that develops in Fig. 2.

Development of 2-D Traveling-Wave Analysis Techniques

As mentioned above, there is a large body of work on 1-D traveling-wave convection patterns. In this restricted geometry, the patterns consist of superpositions of counter-propagating waves. By making use of the fact that the waves have a narrow spectral content, it is possible to demodulate them in time and space to obtain the local amplitude and wave number of the wave components[13]. This technique enables us to make meaningful tests of theoretical predictions and to make precise statements about the dynamics of the patterns. Examples include studies of the Eckhaus instability[14], the chaotic evolution of the convective amplitude[15], and the behavior of pulses of traveling-wave convection[5].

Unfortunately there is no straightforward way to extend this technique to 2-D patterns. Although the modulus of the wave number, $\|\mathbf{k}\|$, is narrowly distributed in the 2-D patterns, the direction of propagation is arbitrary, so that $(\mathbf{k})_x$ and $(\mathbf{k})_y$ are broadly distributed and spatial demodulation cannot be performed. We have developed an algorithm which is similar to complex demodulation in that it extracts a complex amplitude from the pattern, but is flexible enough to represent the complex structure of the 2-D traveling-wave patterns[7].

The basis of the algorithm is the fact that, despite the complex spatial structure of the TW patterns, the time series at a typical point in the pattern is periodic and has a narrow frequency spectrum. Within the large domains of the mature patterns, extremely regular oscillations are observed, and the slowly moving domain boundaries merely cause phase dislocations in the time series in a small fraction of the area of the pattern. The pattern can therefore be represented as an array of oscillators and the description of the pattern then consists of the complex amplitude and

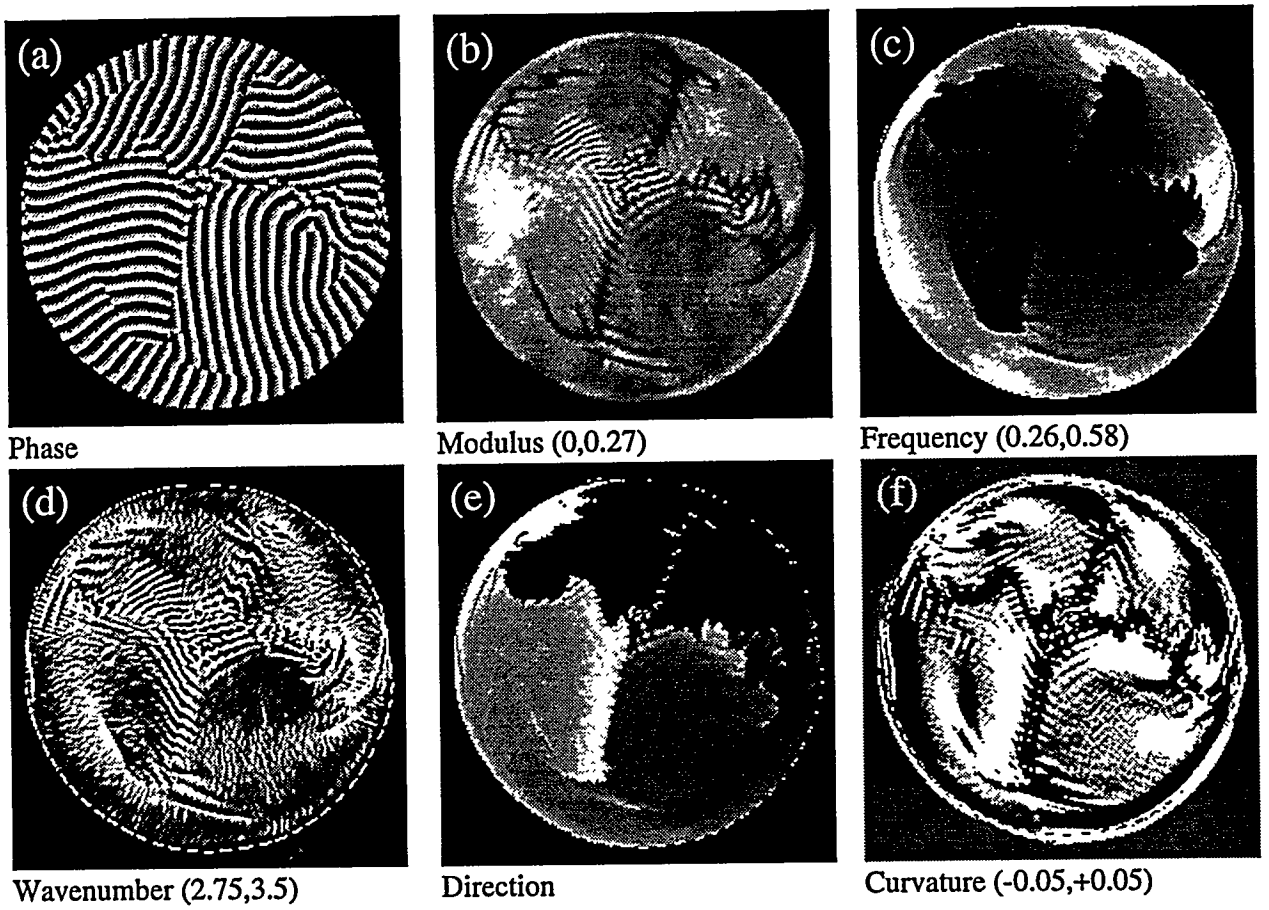
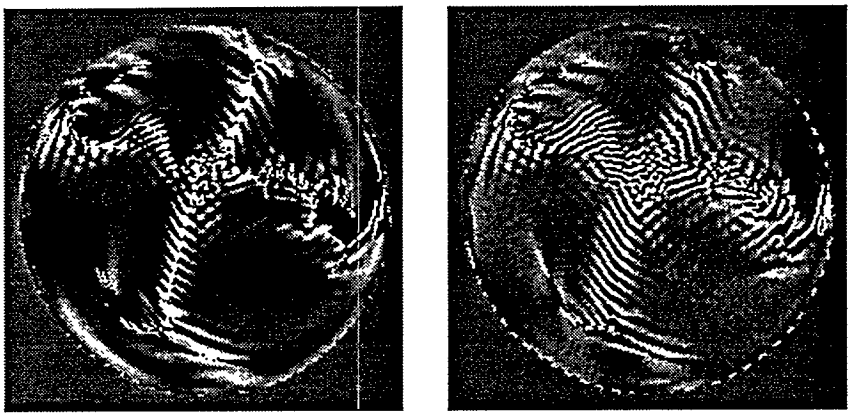


Figure 3: Analysis of a traveling-wave convection pattern showing (a) the phase, ϕ; (b) the modulus, $\|A\|$; (c) the frequency, ω; (d) the wave number $\nabla\phi$; (e) the direction, $tan^{-1}(k_y/k_x)$; and (f) the curvature, $\nabla \cdot \hat{n}$.

the frequency of oscillation of each pixel. In the strongly nonlinear TW patterns, the modulus of the complex amplitude is approximately constant in time and space. Therefore information about the spatial structure of the pattern is contained mainly in the phase, and the instantaneous evolution of the pattern is largely determined by the local frequency of oscillation. The dynamical properties of the pattern can be derived from the complex amplitude and frequency fields as described below.

Quantitative Characterization of Patterns

Fig. 3 shows the dynamical analysis of a traveling-wave convection pattern. Fig. 3(a) is the phase of the complex amplitude. The phase gives a clean representation of the pattern and resolves the ambiguity of the direction of propagation of the rolls. The modulus of the complex amplitude is shown in Fig. 3(b), and it is nearly constant within the large domains, justifying the assumption that the pattern is determined mainly by the phase. In Fig. 3(c), the frequency distribution of the pattern is shown. It is clear that the frequency, and hence the phase velocity, is strongly modified by the proximity of domain boundaries. The remaining panels of Fig. 3 show quantities derived from the phase field. The wave vector is the gradient of the phase field. The modulus of this vector is shown in Fig. 3(d), and is found to have a rather narrow distribution compared with the



Twist (-0.022,+0.022) Stretch (-0.022,+0.022)

Figure 4: (a) The twist, $\frac{1}{2\pi}\nabla\omega \times \hat{\mathbf{n}}$, and (b) the stretch, $\frac{1}{2\pi}\nabla\omega \cdot \hat{\mathbf{n}}$.

frequency. The direction of the wave vector is represented in Fig. 3(e), and the divergence of the normal vector, shown in Fig. 3(f), is a measure of the curvature.

The data in Fig. 3 characterize the spatial pattern and its instantaneous rate of evolution. It is also of interest to measure the deformation of the pattern. This may be done by comparing the phase field with the gradient of the frequency field. Clearly, the pattern can evolve without deformation only if the frequency field is uniform. If the frequency field is different at two points in the pattern, then a different number of rolls will pass these points in a given time interval, and the pattern will be deformed. The nature of the deformation depends on the relationship between the gradient of the frequency, $\nabla\omega$, and the normal vector, $\hat{\mathbf{n}}$. If $\nabla\omega$ and $\hat{\mathbf{n}}$ are parallel, then local stretching of the pattern occurs, and if $\nabla\omega$ and $\hat{\mathbf{n}}$ are perpendicular, a local twisting of the pattern occurs. Therefore, the dot product and the cross product of $\nabla\omega$ and $\hat{\mathbf{n}}$ are measures of the stretching and twisting, respectively, and they are shown in Fig. 4.

The algorithm described above represents an attempt to obtain detailed quantitative information from complex 2-D traveling-wave convection patterns. It allows precise comparisons between experimental data and analytical models. For example, the data of Fig. 3(d) indicate that the wave number is sharply peaked and remains within the Eckhaus stable band, probably because rolls can be created and annihilated freely at the domain boundaries. The data of Fig. 3(c) exhibit a wide range of frequencies, even though the variation of wave number is narrow, indicating that linear dispersion is not responsible for the frequency spread. Fig. 3(f) indicates a pervasive positive curvature of the convection rolls.

The data in Fig. 4 place interesting constraints on a model. The twist is substantial, although there is no measurable stretching or compression of the convection rolls as they propagate. This is equivalent to the statement that the gradient of the frequency is everywhere perpendicular to the direction of propagation. These data and those in Fig. 3(d) indicate great rigidity of wave number in the TW patterns. A successful model for TW convection should reproduce these properties.

Interaction of Traveling Waves at Domain Boundaries

Our study of ordered traveling-wave patterns indicates that the properties of the domain boundaries are important in the pattern selection mechanism. It appears that the multi-domain patterns are stable because the kinetics of the domain boundaries are consistent with the circular cell geometry

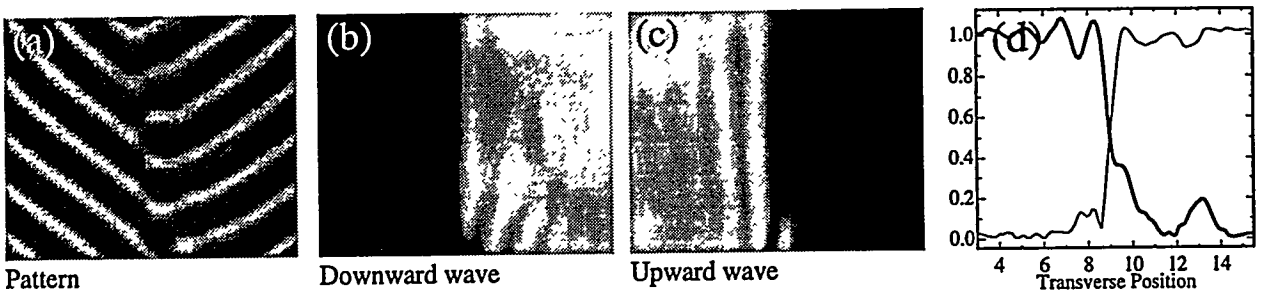


Figure 5: (a) A shadowgraph image of a zipper boundary recorded at $r = 1.37$, (b) the modulus of the wave component measured on the right side of the boundary, (c) the modulus of the wave component measured on the left side of the boundary, (d) the moduli of the two amplitudes along a line transverse to the boundary.

and with the kinetics of the traveling-waves, including the suppression of the phase velocity which occurs near the domain boundaries. In order to study the interaction of traveling waves at domain boundaries, we have employed 3-D complex demodulation of small areas of patterns in which the wave vectors are well defined. Using complex demodulation, it is possible to calculate the complex amplitude of a certain spatio-temporal wave component, defined by k_x, k_y and ω. By calculating this amplitude for the wave components on either side of a domain boundary, it is possible to measure both the penetration of waves through domain boundaries and the movement of the boundaries.

As an example of the use of this technique, Fig. 5 shows the demodulation of a "zipper" boundary. Rolls on the left of the pattern are moving up and toward the boundary, and rolls on the right are moving down and away from the boundary. Figs. 5(b) and (c) are maps of the modulus of the two dominant wave components in the pattern and indicate the general structure of the domain boundary. Fig. 5(d) shows the amplitudes along a line perpendicular to the domain boundary. From these curves it is evident that the region over which the two wave components overlap is approximately one cell height. Another interesting feature of this data is the ripple which is visible on the modulus for waves on the left side of the boundary. This indicates a standing wave pattern which is consistent with 10% reflection of the incoming waves from the domain boundary. No evidence of a standing wave is observed on the other side of the boundary. Here, the waves move away from the boundary and no reflection would be expected.

Conclusion

We have conducted a survey of 2-D traveling-wave convection patterns which occur in ethanol-water mixtures at a large negative value of the separation ratio. We have used a variety of numerical techniques to characterize the dynamics of the complex 2-D patterns, and we have indicated the properties which mathematical models should reproduce. These include the stability of certain configurations of domain boundaries, the suppression of the phase velocity near domain boundaries, and the interpenetration and reflection of traveling-waves at the domain boundaries.

Acknowledgement

This work was supported by the U. S. Department of Energy, Office of Basic Energy Sciences, under Grant No. DE-FG03-90ER14148.

References

[1] M. C. Cross and P. C. Hohenberg. Pattern Formation Outside of Equilibrium. *Reviews of Modern Physics*, **65**, 851–1112 (1993).

[2] P. Kolodner, H. Williams, and C. Moe. Optical Measurement of the Soret Coefficient of Ethanol/Water Solutions. *Journal of Chemical Physics*, **88**, 6512–6524 (1988).

[3] P. Manneville. *Dissipative Structures and Weak Turbulence.* Academic Press, Boston, 1990.

[4] K. Lerman, E. Bodenschatz, D. S. Cannell, and G. Ahlers. Transient Localized States in 2D Binary Liquid Convection. *Physical Review Letters*, **70**, 3572–3575 (1993).

[5] P. Kolodner. Interactions of Nonlinear Pulses in Convection in Binary Fluids. *Physical Review A*, **43**, 4269 (1991).

[6] A. La Porta, K. D. Eaton, and C. M. Surko. Transition from Curved to Angular Texture in Binary-Fluid Convection. *Physical Review E*, **53**, 570 (1996).

[7] A. La Porta and C. M. Surko. Dynamics of Two-Dimensional Traveling-Wave Convection Patterns. *Physical Review E*, **53**, in press (1996).

[8] R. W. Walden, P. Kolodner, A. Passner, and C. M. Surko. Traveling Waves and Chaos in Convection in Binary Fluid Mixtures. *Physical Review Letters*, **55**, 496 (1985).

[9] V. Steinberg, W. Moses, and J. Fineberg. Spatio-Temporal Complixity at the Onset of Convection in a Binary Fluid. *Nucl. Phys. B*, **2**, 109 (1987).

[10] J. J. Niemela, G. Ahlers, and D. S. Cannell. Localized Traveling-Wave States in Binary-Fluid Convection. *Physical Review Letters*, **64**, 1365–1368 (1990).

[11] P. Kolodner, D. Bensimon, and C. M. Surko. Traveling-Wave Convection in an Annulus. *Physical Review Letters*, **60**, 1723–1726 (1988).

[12] L. Ning, Y. Hu, R. Ecke, and G. Ahlers. Spatial and Temporal Averages in Chaotic Patterns. *Physical Review Letters*, **71**, 2216–2219 (1993).

[13] P. Kolodner and H. Williams. Complex Demodulation Techniques for Experiments on Traveling-Wave Convection. In F. H. Busse and L. Kramer, editors, *Nonlinear Evolution of Spatio-Temporal Structures in Dissipative Continuous Systems*, pages 73–91. Plenum Press, New York, 1990.

[14] G. W. Baxter, K. D. Eaton, and C. M. Surko. Eckhaus Instability for Traveling Waves. *Physical Review A*, **46**, R1735–R1738 (1992).

[15] P. Kolodner, J. A. Glazier, and H. Williams. Dispersive Chaos in One-Dimensional Traveling-Wave Convection. *Physical Review Letters*, **65**, 1579–1582 (1990).

IMAGE PROCESSING TECHNIQUES FOR MEASURING NON-UNIFORM FILM THICKNESS PROFILES

Satyanarayana V. Nitta, An-Hong Liu, Joel L. Plawsky, and Peter C. Wayner, Jr.

The Isermann Department of Chemical Engineering
Rensselaer Polytechnic Institute
Troy, New York 12180-3590

ABSTRACT

The long term objective of this research program is to determine the fluid flow and drying characteristics of thin liquid/solid films using image processing techniques such as Image Analyzing Interferometry (IAI) and Image Scanning Ellipsometry (ISE)[1]. The primary purpose of this paper is to present experimental data on the effectiveness of IAI and ISE to measure nonuniform film thickness profiles.

Steady-state, non-isothermal profiles of evaporating films were measured using IAI. Transient thickness profiles of a draining film were measured using ISE. The two techniques are then compared and contrasted. The ISE can be used to measure transient as well as steady-state profiles of films with thickness ranging from 1 nm to > 20 μm, whereas IAI can be used to directly measure steady-state and transient profiles of only films thicker than about 100 nm. An evaluation of the reflected intensity can be used to extend the use of the IAI below 100 nm.

INTRODUCTION

The dynamics governing drying and evaporation phenomena in thin films have been studied extensively. However, our understanding of these phenomena is far from complete. In systems where drying or evaporation takes place, it is crucial to understand the complex effects of the interfacial and intermolecular forces on the intermediate and final film properties. These forces are a function of the film's thickness profile. While theoretical analyses of these film profiles have been available for some time, experimental evidence to ratify these theories has

been lacking only because of the lack of suitable techniques. We believe that ISE, and to a certain extent, IAI are well suited to study such films experimentally.

IAI is a technique based on optical interference, developed to measure non-uniform film thickness profiles. IAI enjoys among other advantages, excellent thickness sensitivity, lateral resolution and the ability to study every point on a surface simultaneously by using image processing techniques. Steady-state, non-isothermal, thickness profiles of Pentane on Quartz[2] were measured using IAI to demonstrate it's effectiveness. One fundamental limitation of IAI is that the measurement of film thicknesses below approximately 100 nm. are less accurate because they depend on the relative intensity of the reflected light. A null ellipsometer and/or an intensity analysis can be used to enhance the accuracy. Since this is time consuming, IAI is not well suited to measuring transient film thickness profiles when the thickness is less than 100 nm. However, transient profiles for thicker films can be easily obtained.

ISE is a technique which can measure both liquid and solid film thicknesses ranging from 1 nm $\rightarrow$ >20 μm. Moreover, it can "handle steady-state and transient processes, measure the entire surface profile, and be non-destructive".[3] ISE can also measure refractive index variations across the surface under observation. From these measurements, if the surface is a mixture of two chemical species, their chemical compositions can be easily extrapolated. In this paper, we demonstrate that the ISE can be used successfully to study the transient draining profiles of a fully wetting and a partially wetting film. The two techniques, ISE and IAI, are then compared and contrasted. We also document the shortfalls of the initial design, and briefly mention the improvements that were made on the second generation ISE. A concise summary of current investigations of the drying of spin-coated sol-gel films, using the improved second generation ISE, will also be given.

EXPERIMENTAL SET-UP

Image Scanning Ellipsometry

The details of the theory, design, calibration and operation of the ISE are documented elsewhere[4], hence no attempt will be made to reproduce them here. However, a schematic overview of the ISE will be presented in the succeeding sentences. The image scanning ellipsometer is based on conventional null ellipsometry, which records the change in phase and amplitude upon reflection of incident polarized light, from a surface. Through appropriate models, the recorded phase and amplitude differences upon reflection are converted into film thickness and refractive index data.

Figure 1 contains a schematic of the image scanning ellipsometer. The ISE has two arms, the polarizing arm, and the analyzing arm. The polarizing arm of the ISE has a light source which provides nonpolarized light, a polarizer which polarizers the light linearly, and a compensator which changes the state of polarization of the light from linear to circular. The analyzing arm has a polarizing analyzer which records the polarization state of the analyzer, and an imaging package. The imaging package consists of a long working distance microscope and a CCD camera. The CCD camera is controlled by a desktop PC through a frame grabber. In a normal null ellipsometer, a photomultiplier or light intensity detector would be used in place of the imaging package.

For a null ellipsometer, the analyzing polarizer is used to null or extinguish the reflected light and the photomultiplier is used to detect the null point. At the null point, depending on the azimuthal angles of the Polarizer, Analyzer and the Compensator, the film thickness and refractive index can be calculated. For the ISE however, if the film from which reflection takes place is non-uniform, a series of bright and dark fringes are produced, as shown in Figure 2.

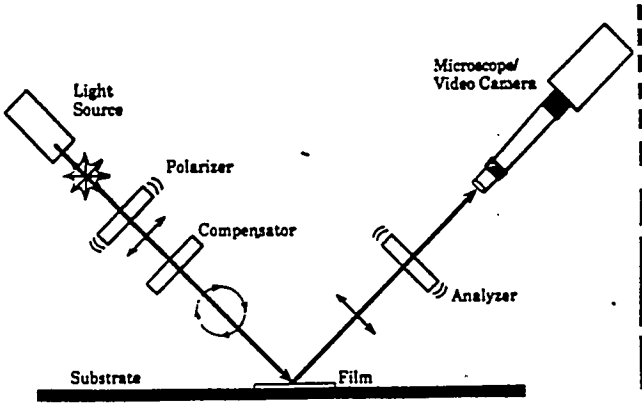


Figure 1. Schematic diagram of the ISE

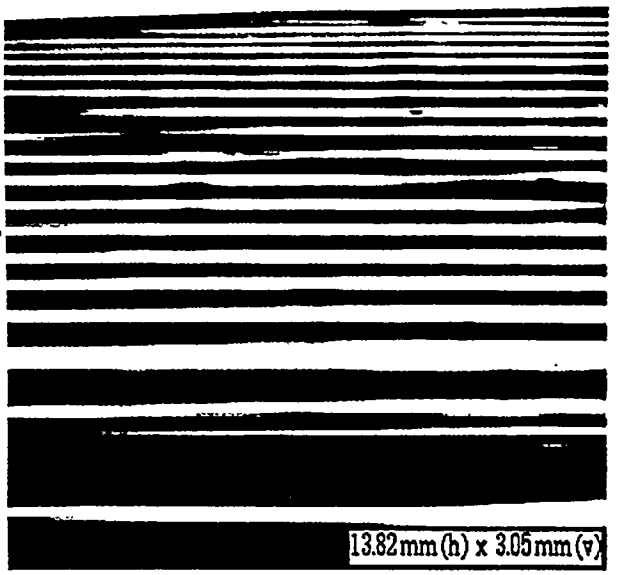


Figure 2. Reflectance fringes from a draining film

These fringes are reflectivity fringes, and the center of every dark fringe corresponds to a null point. The centers of the dark fringes are calculated by processing the acquired image by employing an image analysis software. Thus the film thickness and the refractive index can be determined at the center of these dark fringes. Changing the polarizer angle changes the state of polarization of the incident light. This will in turn change the position of the bright and dark fringes. That is, the fringes are displaced. Hence the film thickness corresponding to the center of the new dark fringes (the minimum intensity point) can be obtained. Therefore, by moving the polarizer while keeping the analyzer and the compensator fixed, the thickness profile over the entire surface can be plotted. One advantage of the ISE is that the technique is reasonably insensitive to the overall intensity of the image, because it depends only on the relative intensity of the image. Another advantage is that the initial null point can be associated with a very thin film (film thickness on the order of 1 nm.)

The film thickness at the null points can be calculated only up to a repeat thickness. This thickness, δ_r, is given by

$$\delta_r = \frac{\lambda\beta}{2\pi\eta_f \cos\theta_f} \tag{1}$$

where, λ is the wavelength of the light, β, is the polarizer angle, η_f is the refractive index of the film, and θ_f is the angle of refraction of light through the film. In this study, the repeat thickness was found to be about 302.5 nm.

Image Analyzing Interferometry

The IAI consists of a CCD camera mounted on top of a normal optical microscope. The CCD camera is in turn controlled by a desktop PC through a framegrabber. Interference images captured by the framegrabber are stored in the PC and later analyzed using an image analysis software. The principle of operation of the IAI is based on the interference of light rays upon reflection from a transparent thin film. Alternating constructive and destructive interference patterns occur when coherent light undergoes reflection from the two interfaces of a thin liquid film and recombine. Constructive interference occurs when the two reflected beams are in

phase, while destructive interference occurs when the two beams are out of phase by π. The equation used to calculate the film thickness is:

$$\delta = \frac{(2L+1)\lambda}{4n_f} \qquad (2)$$

where, n_f is the refractive index of the film, L is the order of the destructive (dark) fringes, and λ is the wavelength of monochromatic light used. The first destructive fringe occurs at a relatively large thickness of approximately 100 nm.

Test Cell

A fused silica cell was used for the draining film experiments. The cell was trapezoidal in shape to ensure that light entering and leaving the cell was perpendicular to the cell walls. The cell was cleaned by a standard RCA cleaning procedure, sealed with the test liquids (a fully wetting FC-70 film in one set of experiments and a partially wetting dodecane film in another) and placed on a hinged sample holder on the ISE. The holder was hinged in such a way that fluid would drain on to the part of the silica surface that was under observation with the ellipsometer.

The experiment[3] was initiated by tilting the test cell and returning it to it's original position. A series of images were recorded at various time intervals from the start of the experiment. These images were digitized and stored in the computer's memory and were later analyzed by an image analysis program to determine the minimum intensity points (null points.)

In addition to this, experiments sponsored by NASA were also carried out to determine the thickness profile using IAI for a Pentane film over Quartz. Details of these experiments are presented elsewhere[2].

RESULTS AND DISCUSSION

ISE Results

Figures 3 and 4 show experimentally determined profiles of the partially wetting and fully wetting draining films respectively, at different time intervals from the start of flow. From the figures, the flow can be divided into four draining regions, interfacial, transition, hydrodynamic and meniscus. The region farthest from the liquid pool is the interfacial region and it consists of the adsorbed film. The transition region connects the interfacial region and the hydrodynamic region. Film thicknesses in the interfacial and transition regions are below 100 nm. The hydrodynamic draining region extends for another 1-2 mm down from the interfacial region. As can be seen from the figure, the extent of this region shrinks with time. The last region is the meniscus region, and it lies near the liquid pool. It is characterized by a rapidly increasing film thickness profile.

IAI Results

Figure 5 shows an experimentally determined profile of a Pentane film on quartz, at steady-state. As can be seen from the picture, the smallest thickness that could be resolved here was about 100 nm. Hence, the interfacial region is below the threshold of the IAI. Therefore, the evaluation of this region (below 100 nm) depends on modeling.

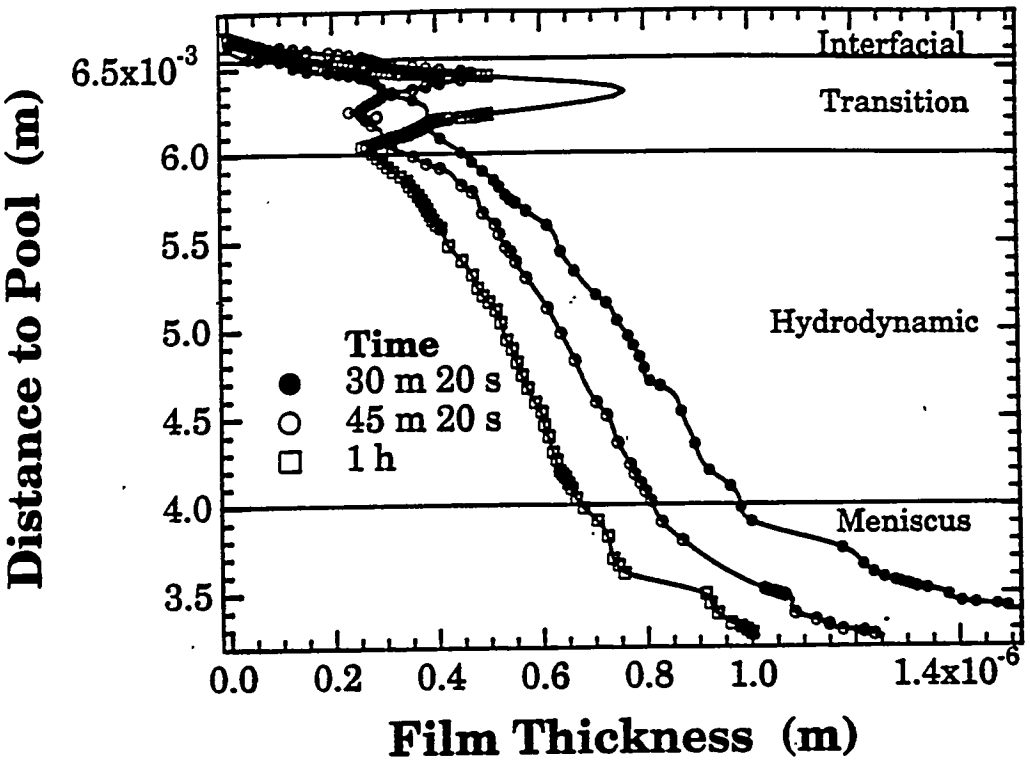


Figure 3. Draining profile of a partially wetting

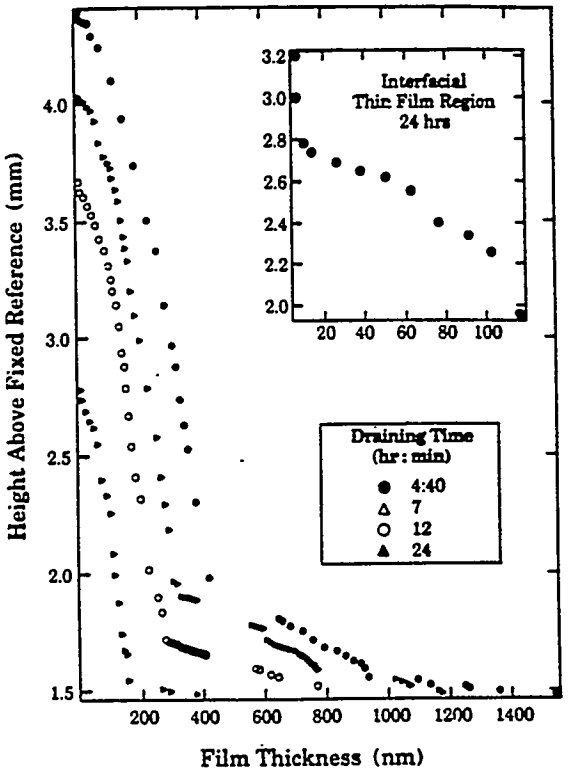


Figure 4. Draining profile of a fully wetting film

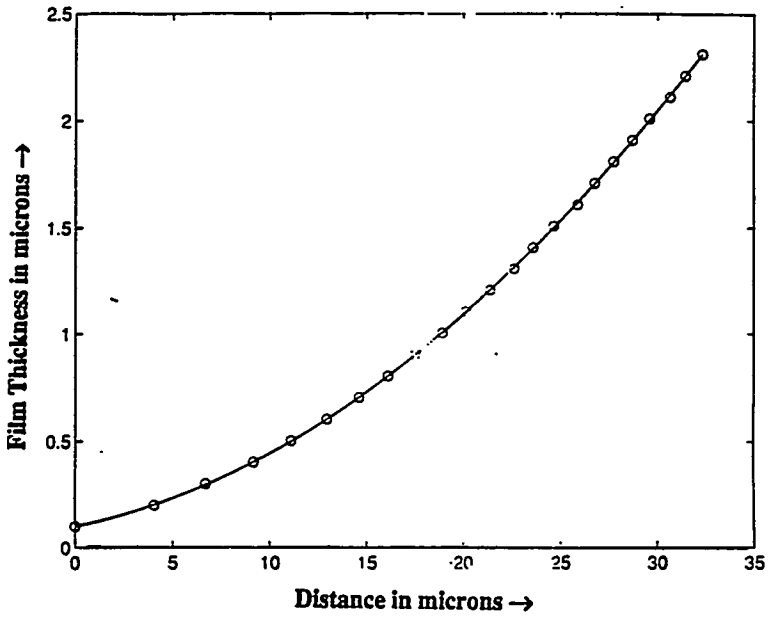


Figure 5. Steady-state film profile of Pentane on a Quartz substrate obtained by IAI

ISE vs. IAI

A comparison of the two techniques reveals some features of interest. For clarity, the comparison will be split up into two categories as shown below:

Advantages of ISE over IAI

- ISE can be used to measure thicknesses from a monolayer upto about 20 μm, whereas IAI can only be easily used from about 100 nm and upwards. It is noted that an evaluation of the reflected intensity can be used to extend the use of the IAI below 100 nm.
- ISE enjoys a larger lateral resolution than IAI. Hence larger variations in thickness profiles can be observed using ISE.
- ISE can be used on films whose R.I. is unknown. At the same time, R.I. variations can also be handled facilely by using ISE. IAI needs R.I values to determine film thickness unless two different wavelengths are used.

Advantages of IAI over ISE

- IAI is more accurate as the accuracy of ISE is, depending on the algorithm used, highly dependent on the initial guesses for the thickness and R.I (Ellipsometry algorithms use models to relate changes in phase and amplitude to thickness and R.I.)
- IAI is easier to use and is less expensive.
- It is conceivable that IAI can be used to measure steeper profiles than ISE because of it's larger magnification. At this point, we are still unsure of ISE's limitation in this regard.

Modifications to the initial design of ISE

Modifications were made to the initial design of the ISE using recent advances made in Imaging technology to overcome some of the limitations with respect to weakly reflecting films. However, some of these modifications created new problems while solving some of the older ones. These modifications are:

The use of a laser (He-Ne green laser at 543.5 nm) as the light source. In the initial design, we used a UV lamp. The biggest asset of laser light is that it is an intense, coherent, collimated light source. Thus, most of the intensity problems we faced with the UV lamp were

eliminated. However, the trade-off is with respect to the speckle that is a characteristic of laser light. While laser speckle has it's uses, it can be a great nuisance in imaging applications. In our case, the speckle makes it difficult to determine the exact center of a dark fringe as it produces variations in the intensity profile. There are two techniques to tackle this problem. The first is to eliminate the speckle at the source itself, by using diffusers[5,6] or fiber optic probes. The second technique involves "cleaning up" the image by certain algorithms that can eliminate the speckle by employing complex statistical formulae. Currently, we are employing both techniques to help remove this obstacle

The use of a higher resolution, lower light level CCD camera. In the previous version of the ISE, we used a standard 512x512 pixel camera. However, in this version, we deciding to replace the old camera with a 1024x1024 pixel camera. This augmented resolution is mildly offset by the slightly slower rate at which frames can be "grabbed" using this camera. With the standard 512x512 pixel camera, we can grab upto 30 frames a second. However, with the new camera, we can only grab upto 10 frames a second. This is mainly due to the extra demands placed on the computers' memory interface by the higher resolution camera. We do not expect this slightly lower transferring power to be a handicap.

The mechanical infrastructure on which the current version of the ISE is mounted offers several advantages over the previous one. The chief among these are the ability to change the angle of incidence (multiple angle of incidence ellipsometry[7] offers a statistically superior alternative to conventional null ellipsometry), and the improvement in the number of degrees of freedom enjoyed by the substrate holder.

CURRENT INVESTIGATIONS USING THE MODIFIED ISE

Currently, we are applying the ISE to study the complex transport effects which occur during the drying of spin-coated sol-gel films. It has recently been found that organically modifying sol-gel systems by partially replacing the surface hydroxyls has an anomalous effect on the final film thickness of dip-coated sol-gel films[8]. These films exhibit a 'springback' during the final stages of drying. This ability to control the degree of shrinkage during drying could have an enormous impact on many potential applications.

To study this effect in spin-coated systems, we are currently carrying out the following investigation: An organically modified sol-gel system is spin-coated over a 0.5 μm. high by 500 μm. wide aluminum line (to provide a meniscus for the ISE). The film is then dried on a hot-plate and the evolution of the drying film thickness and refractive index profiles is observed in real-time using the ISE. The thickness and refractive index profiles will yield information about the porosity of the resulting film. Film porosities are responsible for effects such as electrical and thermal insulation, etc. The ultimate goal is to understand from a fundamental perspective, how the drying related stresses will effect the final porosity of the film.

CONCLUSIONS

1. The ISE and the IAI are efficient and accurate techniques to study the complex interfacial effects that occur in very thin films.

2. Both ISE and IAI can be used to study transient as well as steady-state thickness profiles. IAI cannot be easily used below 100 nm which is where the intermolecular forces are most important.
3. The initial design of the ISE has been modified to take advantage of the increasing data processing available in imaging applications today.

ACKNOWLEDGMENTS

This material is based on work supported by the Department of Energy under Grant # DE-FG02-89ER14045.A000. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the DOE.

REFERENCES

1. A-H. Liu, J. L. Plawsky, S. DasGupta, & P. C. Wayner, Jr., *Proceedings of the Tenth International Heat Transfer Conference, Brighton, UK*, 1994, Vol. 2, pp 267-272
2. Karthikeyan, M., Huang, J., Plawsky, J. L., & Wayner, P. C. Jr., *ASME National Heat Transfer Conference, Houston, TX*, 1996 (submitted for publication)
3. A-H. Liu, P. C. Wayner, Jr., & J. L. Plawsky, *Phys. Fluids* **6** (6), 1994, pp 1963-1971
4. A-H. Liu, P. C. Wayner, Jr., & J. L. Plawsky, *Applied Optics*, Vol. 33, No. 7, 1994, pp 1223-1229
5. S. Lowenthal, & D. Joyeux, *J. Opt. Soc. Am.*, Vol. 61, No. 7, 1971, pp 847-851
6. Ih, C. S., *Applied Optics,* Vol 16, No. 6, 1977, pp 1473-1474
7. Smit, M. K., *Thin Solid Films,* Vol. 189, No. 2, 1990, pp 193-203
8. S. S. Prakash, C. J. Brinker, A. J. Hurd, & S. M. Rao, *Nature,* Vol. 374, 1995, pp 439-443

KINETIC ANALYSIS OF COMPLEX METABOLIC NETWORKS

Gregory Stephanopoulos

MIT, Department of Chemical Engineering
Cambridge, MA 02139, U.S.A.

ABSTRACT

A new methodology is presented for the analysis of complex metabolic networks with the goal of metabolite overproduction. The objective is to locate a small number of reaction steps in a network that have maximum impact on network flux amplification and whose rate can also be increased without functional network derangement. This method extends the concepts of Metabolic Control Analysis to *groups* of reactions and offers the means for calculating group control coefficients as measures of the control exercised by groups of reactions on the overall network fluxes and intracellular metabolite pools. It is further demonstrated that the optimal strategy for the effective increase of network fluxes, while maintaing an uninterrupted supply of intermediate metabolites, is through the *coordinated* amplification of multiple (as opposed to a single) reaction steps. Satisfying this requirement invokes the concept of the concentration control coefficient, which emerges as a critical parameter in the identification of *feasible* enzymatic modifications with maximal impact on the network flux. A case study of aromatic aminoacid production is provided to illustrate these concepts.

INTRODUCTION

Many industrial applications make use of the unique capabilities of microorganisms to convert simple carbohydrates into a variety of products. Microbial processes for the production of chemicals, materials, and pharmaceuticals and specialty chemicals, are presently employed in many parts of the world. The above products are synthesized by complex networks of biochemical reactions catalyzed by specific enzymes. The throughput of these networks is determined by the specific rate of glucose, or other carbohydrate, uptake and the relative activities of the network enzymes participating in the production of these products. Although gains in volumetric productivities and yields have increased the competitiveness of biological

processes, the majority of chemical and materials production presently is carried out by chemical synthesis because of the superior yields and productivities obtained by such processes.

Biological systems, on the other hand, offer some distinct advantages such as enhanced selectivities and environmentally benign operations utilizing renewable resources as raw materials. Furthermore, they allow greater flexibility for process optimization using techniques from genetic engineering. These techniques can be employed to extend the range of substrates that a microorganism can utilize to alter the product profile secreted by a producing cell, or overproduce a product normally secreted by a microorganism.

Although genetic engineering has been instrumental in the construction of strains with enhanced or unique properties, a greater challenge would be to effect multi-fold throughput increases through the metabolic pathways of industrial microorganisms such as yeasts, *Streptomyces*, bacilli, and *Escherichia coli*. Such flux (i.e. throughput) amplifications are needed to significantly increase the specific productivities of biological systems and thus make them competitive with chemical processes. To accomplish this objective the central carbon metabolic pathway must be similarly amplified as it is the main line of carbon processing and, as such, it constitutes the main supply route to all product-forming pathways. Amplification of central carbon metabolism, however, is a very demanding undertaking since it involves many interconnected reactions with sophisticated feedback controls and regulations that have, to date, evaded most attempts at directed manipulation. In this context, it becomes very important to identify the key branchpoints where such controls are exercised and the specific reactions within metabolic pathways that must be specifically amplified to effect a direct change in the overall network flux. Relevant questions are also whether one or more reactions need to be modified and whether activity amplification should take place in a sequential or simultaneous manner.

The objective of this paper is to analyze the kinetic behavior of complex metabolic networks and provide a framework within which answers to the above questions can be systematically sought. By necessity, this is a theoretical paper. The alternative, namely, the experimental evaluation of networks in the absence of a rational framework, would be an unfocused and time-consuming undertaking. We have opted, instead, to employ a rather sophisticated network of biochemical reactions as a surrogate cell to facilitate our investigation into the dynamics of such networks operating in real microorganisms. The purpose here is not to simulate biological reality, but rather to take advantage of a system which exhibits all aspects of regulation, tight control, and feedback mechanisms that are likely to be encountered with real biological systems.

Our presentation utilizes the framework of metabolic control analysis (MCA) [1-4] and its various extensions. In order to facilitate the investigation of networks with a large number of reactions, we present a novel method of reaction *grouping* and extensions of MCA to groups of reactions, a concept that is very valuable in describing the kinetic behavior of complex networks. The magnitude of control coefficients is used as a measure of the kinetic control exercised by single reactions or groups of reactions. As such, much of the focus of the paper is on techniques for the determination of such control coefficients as a means for identifying network limiting steps. This is demonstrated with a model of aromatic aminoacid biosynthesis. We close our presentation with an evaluation of the extent to which such limiting steps can be amplified and the effect such amplifications can have on the stability of the overall metabolic network. It turns out that it is not possible to significantly amplify rate controlling steps without complete derangement of the network. On balance, it may be more desirable to focus on steps that allow greater amplification even though they have lesser impact on the overall kinetics of the network.

In this context, the *concentration control coefficient* emerges as an important parameter that determines the extent to which various steps can be amplified without compromising the stability of the system.

THEORY

Rudiments of Metabolic Control Analysis

A microbial cell is often viewed as a black box typically processing carbon and nitrogen sources to derive the energy and carbon skeletons needed for growth along with the secretion of various metabolic products. Figure 1 shows such a schematic illustrating the utilization of glucose and ammonia by *Saccharomyces cereviciae* with simultaneous excretion of ethanol, glycerol (Gol), polysaccharides (Pol), CO_2 and the aromatic amino acids: tryptophan, phenylalanine, and tyrosine. Lumped parameter models have proven useful in producing macroscopic kinetic expressions for the rates of substrate uptake, microbial growth, and product production for use in the design of fermentation equipment for the propagation of these microorganisms.

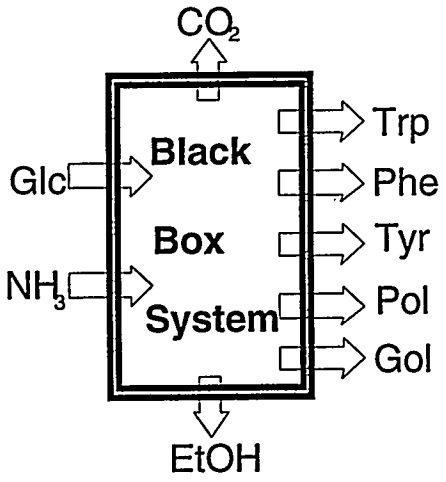


Figure 1

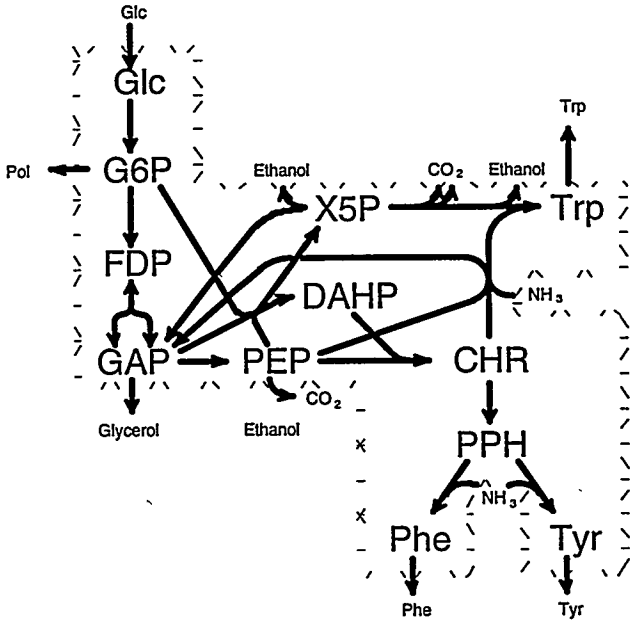


Figure 2

If, on the other hand, one is interested in altering the metabolism in order to increase the overall flux through these organisms, then one needs to be concerned about the exact biochemical network that catalyzes the transformation of the substrates into energy and the secreted products. Figure 2 is a schematic of the network of biochemical reactions operating in *S. cereviciae* and leading to the production of aromatic amino acids. Energy, in the form of ATP, is produced primarily by the ethanol pathway and consumed at various reaction points. In order to bring about significant increases in the overall rate of product production then a number of relevant questions arise: (a) which enzymatic step or series of steps should be targeted for

modification in order to effect such an outcome; (b) what is the optimal type and magnitude of modification; and (c) should the modifications take place in a sequential or simultaneous manner.

The above questions are complex and difficult to answer due to the high non-linearity and interaction among the various reaction steps and intracellular metabolites. They are a part of the general quest to elucidate the *control of flux* in metabolic networks, a central tenet of **metabolic engineering**. For the past 22 years a convenient framework and school of thought have evolved, currently referred to as metabolic control analysis, that can be used to address these questions.

A key parameter in MCA is the flux control coefficient (FCC). FCC's provide a measure of the impact that a change in a single reaction step of a metabolic network can have on an overall flux through the network. They are defined as the ratio of the fractional change of a metabolic flux J to the fractional change in the velocity of enzymatic reaction i:

$$C_i^J = (dJ/J)/(dv_i/v_i) \qquad \text{(Eq. 1)}$$

Another parameter introduced by MCA is the reaction elasticity, defined as the fractional change in the velocity of an enzymatic reaction v_i divided by the fractional change in the concentration of a metabolite M_j:

$$\varepsilon_j^i = (\partial v_i/v_i)/(\partial M_j/M_j) \qquad \text{(Eq. 2)}$$

Clearly, the elasticity with respect to metabolites that have no effect on the reaction velocity is zero. The elasticity with respect to another metabolite can be obtained from the partial derivative of the reaction velocity with respect to the concentration of the metabolite in question. In this regard, elasticity can be considered as a pseudo order of the reaction rate.

Elasticities are local parameters, while FCC's systemic properties of the network at steady state. MCA theorems provide equations relating FCC's with elasticities so that FCC's can be determined if the elasticities are known. This means that if kinetic models are available for the individual reaction steps of the network, they can be used for the exact determination of the flux control coefficients and, through them, the control architecture of the network.

Reaction Grouping, Group Control Coefficients

Accurate *in vivo* reaction kinetics are, in general, not available. Furthermore, typical metabolic pathways involve many reactions that make it infeasible to evaluate the impact of each one of them on the overall network kinetics. A useful concept is that of reaction grouping introduced by the top down metabolic control analysis (TDCA) [5,6]. The main tenet of this approach is to focus on *groups of reactions* instead of individual reaction steps and evaluate the effect of different groups on the kinetics of the overall network. This is an intuitive approach whose success depends on the correct definition of reaction groups, which frequently differ from groupings suggested by the topology of biochemical maps.

We have developed a method for the systematic definition of reaction groups. This method makes use of the steady-state internal metabolite stoichiometry (SIMS) matrix defined as an mxr matrix in which m is the number of explicit steady state metabolites in the network and r is the number of explicit reactions. Each element N_{ji} of the SIMS matrix is the stoichiometric coefficient v_j of metabolite X_j participating in reaction i written as:

$$\sum_{\text{reactants}}(-v_j X_j) \rightarrow \sum_{\text{products}}(v_j X_j) \qquad \text{(Eq. 3)}$$

It is crucial that the direction of each reaction of Eq. 3 be the same as the net flux in the actual network. Following the construction of the SIMS matrix, the reaction groups are identified from the membership of the vectors of the kernel matrix, K, of N defined as: $N \cdot K = 0$.

It should be noted that the reactions comprising the columns of the kernel of the SIMS matrix also define the independent pathways of the network. Independent pathways reflect the smallest set of reactions connecting a single network output with the necessary network inputs in such a way that permits a steady state to be reached by all internal metabolites. For a network consisting of r reactions and m internal metabolites at steady state, the number of independent pathways P is shown to be equal to $P = r - m$. Besides independent pathways, reaction grouping also identifies the *link metabolites,* (also referred to as branch points), as the intervening metabolite at the point of separation between two or more independent pathways.

When the above method is applied to the metabolic network of Figure 2, the pathways and link metabolites indicated in Fig. 3 are identified: ATP; glucose-6-phosphate (G6P); fructose-1, 6-diphosphate (FDP) and glyceraldehyde-3-phosphate (GAP); xylulose-5-phosphate (X5P) and chorismate (CHR); and prephenate (PPH). It should be noted that the FDP/GAP branch point actually consists of an equilibrium between the two species. In the case of the X5P/CHR branch point, *both* species are produced by the common pathway and consumed by alternate routes; consequently, this branch point consists of *dual* link metabolites. It is also critical to realize that the first link metabolite to be identified in this network is ATP, although this fact may not be immediately apparent from the reaction schematic. The glycolytic production of ATP is common to all pathways. Because the independent pathways utilize ATP differently, a pathway separation occurs at the ATP junction, so ATP is indeed found to be a link metabolite [7].

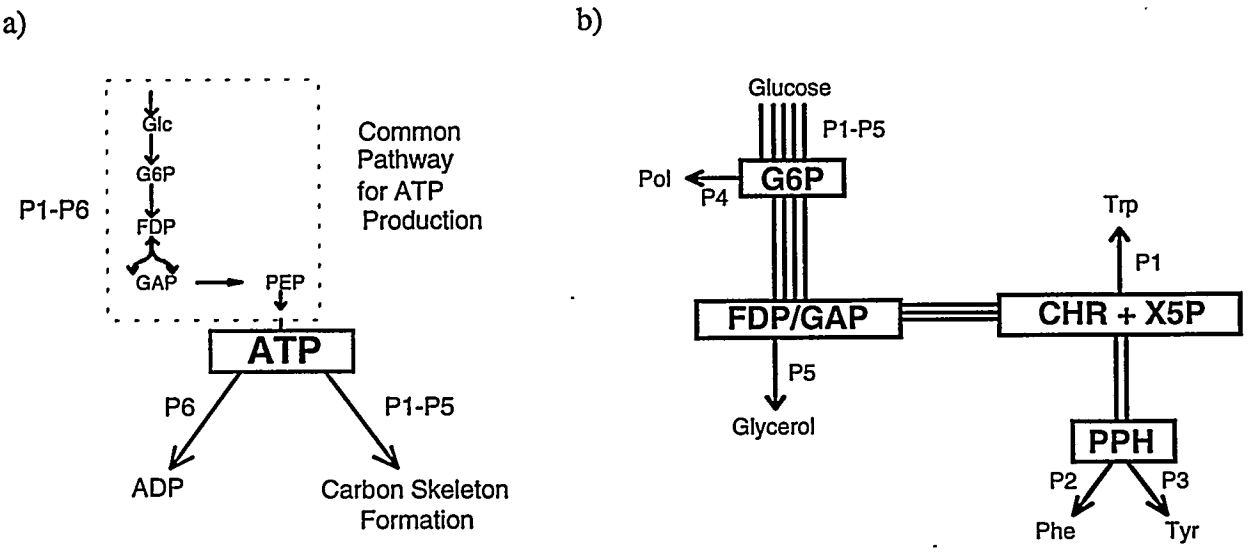


Figure 3

Concurrent with reaction grouping is the concept of the *group flux control coefficient (gFCC).* The latter is defined as the flux control coefficient which would exist, were the entire group of reactions actually a single step. Since it is practically impossible to implement changes of the same magnitude in all reactions in a group, the experimental determination of gFCC's relies on the measurement of carbon fluxes and flux changes following perturbations in one or more reactions in the group. The details of the calculation can be found in [8].

It should be noted that each independent pathway is usually associated with a secreted product. In fact, the accumulation of secreted metabolites can provide a measure of the flux through each independent pathway [9,10]. Furthermore, changes in product accumulation, during different phases of a fermentation process or in response to introduced perturbations, can be used for the quantitative evaluation of the kinetic control exercised by each reaction group, through the determination of the corresponding group control coefficients. Through the recursive analysis of overlapping reaction groups around different link metabolites, the search for the controlling steps of the network can, in fact, be focused within a small group of reactions.

RESULTS

An in-depth case study was carried out for the aminoacid biosynthetic network of Fig. 2. *In-vitro* kinetic expressions (from [11] with minor modifications) were used for each of the indicated reactions. The network reaction kinetic model allows the determination of the steady-state concentrations of the intracellular metabolites and, through them, the calculation of the reaction elasticities and FCC's. MCA theorems are invoked in the latter calculation following usual practice. The magnitude of the FCC for the phosphofructokinase (PFK) reaction, (G6P→FDP in Fig. 2), was found to be significantly greater than any other reaction in the network, indicating that the PFK enzyme exercises a significant fraction of the total control on the network flux.

Although the above approach based on FCC calculation is practically infeasible (due to lack of reliable *in-vivo* kinetic models), it can serve as a guide in the development of experimentally feasible methods, such as one that involves grouping of reactions. It is reminded that the calculation of the *group* control coefficients requires the measurement of the fluxes through the simplified network of Fig. 3b, a task that can be normally accomplished from the measurement of the indicated extracellular metabolites. Through successive reaction groupings around the different link metabolites, the network flux control can be localized to a single intermetabolite linkage, that becomes the focus of further investigation and genetic modification.

The introduction of kinetic perturbations to the reactions of the network of Fig. 2, [8], allowed the calculation of the gFCC of the various reaction groups. It was found that the controlling reaction group (as assessed by the magnitude of the gFCC's), should lie upstream of the FDP/GAP branch point and downstream of G6P. Thus, the step exerting the most control is the reaction of PFK, in agreement with the conclusion reached from the magnitude of the individual FCC's.

The large magnitude of the phosphofructokinase FCC would indicate that this particular step should be the primary amplification target in order to bring about the maximum effect on amino-acid overproduction. However, simulations of proportional increases in the kinetic parameters of the PFK reaction revealed that the structure of the network prevents PFK amplification beyond an 11% increase in activity. At greater amplifications the overall system is unstable, *i.e.*, unable

to converge to a steady state condition. The reason for this particular instability turned out to be a bifurcation, at the above PFK amplification value, into a space where a steady state for the chorismate metabolite does not exist. In other words, there are no acceptable intracellular metabolite concentrations that can balance the rates of chorismate production and depletion. Although it cannot be claimed that this would happen in a similarly-modified strain of *S. cerevisiae*, this result is analogous to typical cellular responses to the introduction of a catastrophic metabolic disturbance. In such cases, secretion of metabolites, induction of degradation pathways, and drastic changes in product profiles are commonly observed.

One way to limit network instabilities following the introduction of a genetic perturbation is to design perturbations that minimize the departure of the altered cell from a normal steady state. This can be implemented by a *coordinated* modification of more than one steps that effect an increase of flux through the network while maintaining intracellular metabolite levels near their original steady state. By allowing modest changes in metabolite levels, it can be shown that significant increases in the overall network flux can be obtained from the modification of *only a small number* of carefully selected enzymatic steps.

Group or individual *concentration control coefficients* (CCC) emerge as key parameters in the optimization process, since they provide a measure of metabolite sensitivity to reaction rate modifications. The problem then is to determine the best single reaction step, or perhaps the best two or three steps, that should be amplified in order to effect the largest possible increase in the flux of the network, subject to the constraint that all intracellular metabolite levels remain within a reasonable range of their original steady-state values. As it happens, the results are rather insensitive to the allowed metabolite range, owing to the fact that, once the bifurcation borders are approached, progression towards network instability occurs rather precipitously.

Omitting details, the optimal reaction step(s), as well as the recommended rate amplification are obtained as the solution of a constrained optimization problem [8]. For the network of Fig. 2, although PFK is clearly a limiting reaction, if a single step is to be amplified, it is most profitable to do so for the reaction producing the desired amino acid. The large FCC for PFK is, in essence, nullified by a much larger CCC that restricts the allowable kinetic amplification of PFK. Reactions 14, (CHR→Trp), and 12, (PPH→Phe), emerge as the optimal single steps for tryptophan and phenylalanine production, respectively, because they offer the best balance between the magnitude of the allowed amplification (measured by the CCC) and the impact on the network flux per unit of reaction rate amplification (measured by the FCC). When two reaction steps are allowed to be changed, the optimization procedure predicts that the flux for tryptophan and phenylalanine overproduction can be doubled through the increase of the reaction pairs (PFK, 14) and (PFK, 12), respectively. In essence, the effect of these pairings is to moderate the level of chorismate, by *pulling* away from chorismate (through reaction 14 or 16) part of the overwhelming carbon flux which is *pushed* into chorismate by the amplification of PFK. Thus, adjustment of one reaction in each pair serves to alleviate the metabolic instabilities otherwise caused by amplification of the second. Thus, *simultaneous* adjustment of multiple steps is favorable to sequential amplification.

Our results suggest that greater flux increases can be achieved through the simultaneous amplification of two reactions rather than one, and with three rather than two. Amplification of four or more steps, however, was found to be marginally better and occasionally problematic, due to unpredictable instabilities resulting from significant changes in such a large number of steps, as well as the unfeasibility of experimental implementation of this many simultaneous

alterations. It is instead suggested that an optimal pair or triplet of reactions be amplified, followed by analysis and further modification of the resulting system. In most any case, significant flux amplification should be achievable through adjustment of a small number of reactions.

ACKNOWLEDGEMENT

The financial support of this work by the Department of Energy Award Number DE-FG02-94ER14487 is gratefully acknowledged.

REFERENCES

1. H. KACSER and J. A. BURNS, "The control of flux," *Symp. Soc. Exp. Biol.* 27, 65-104 (1973).
2. J. R. SMALL and H. KACSER, "Responses of metabolic systems to large changes in enzyme activities and effectors. I. The linear treatment of unbranched chains," *Eur. J. Biochem.* 213, 613-624 (1993a).
3. J. R. SMALL and H. KACSER, "Responses of metabolic systems to large changes in enzyme activities and effectors. II. The linear treatment of branched pathways and metabolite concentrations. Assessment of the general non-linear case," *Eur. J. Biochem.* 213, 625-640 (1993b).
4. D. A. FELL, "Metabolic Control Analysis: a survey of its theoretical and experimental development," *Biochem. J.* 286, 313-330 (1992).
5. G.C. BROWN, R. P. HAFNER and M. D. BRAND, "A 'top-down' approach to the determination of control coefficients by metabolic control theory," *Eur. J. Biochem.* 188, 321-325 (1990).
6. P. A. QUANT, "Experimental application of top-down control analysis to metabolic systems," *Trends Biochem. Sci.* 18, 26-30 (1993).
7. T. W. SIMPSON, G. E. COLÓN and G. STEPHANOPOULOS, "Two paradigms of metabolic engineering applied to amino acid biosynthesis." *Trans. Biochem. Soc.* 23, 381-387 (1995).
8. G. STEPHANOPOULOS and T. W. SIMPSON, "Flux Amplification in Complex Metabolic Networks," *Chem. Eng. Science,* (submitted) (1996).
9. G. STEPHANOPOULOS and J. J. VALLINO, "Network rigidity and metabolic engineering in metabolite overproduction," *Science* 252, 1675-1681 (1991).
10. J. J. VALLINO and G. STEPHANOPOULOS, "Metabolic flux distributions in *Corynebacterium glutamicum* during growth and lysine overproduction," *Biotechnol. & Bioeng.* 41, 633-646 (1993).
11. J. L. GALAZZO and J. E. BAILEY, "Fermentation pathway kinetics and metabolic flux control in suspended and immobilized *Saccharomyces cerevisiae*," *Enzyme Microb. Technol.* 12, 162-172 (1990). See also "Errata," 13, 363 (1991).

Final List of Participants

FOURTEENTH SYMPOSIUM ON ENERGY ENGINEERING SCIENCES

May 15-16, 1996

Argonne National Laboratory
Argonne, Illinois

Jan D. Achenbach
Center for Quality Engineering & Failure Prevention
Northwestern University
2137 North Sheridan Road
Evanston, IL 60208

Andreas Acrivos
Levich Institute
City College of CUNY
T-1M, City College
New York, NY 10031

David M. Barnett
Dept. of Materials Science & Engineering
Stanford University
Building 550, Rm. 550K
Stanford, CA 94305-2205

Thomas W. Eagar
Department of Materials Science & Engineering
Massachusetts Institute of Technology
Building 8-309
77 Massachusetts Avenue
Cambridge, MA 02139

Jack W. Ekin
Electromagnetic Technology Division
NIST
325 Broadway
MS 814.05
Boulder, CO 80303

Alexander Frenkel
Department of Mathematics
University of Alabama
Tuscaloosa, AL 35487-0350

Fred M. Glaser
Department of Fossil Energy
U.S. Department of Energy
Germantown, MD 20874

Richard J. Goldstein
Regents Professor & Head ME Department
University of Minnesota
111 Church Street, SE
Minneapolis, MN 55455

Robert Goulard
Division of Engineering & Geosciences, ER-15
U.S. Department of Energy
Office of Basic Energy Sciences
19901 Germantown Road
Germantown, MD 20874-1290

Yuki Iwasa
Francis Bitter Magnet Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139

Alan R. Kerstein
Diagnostics & Reacting Flow
Sandia National Laboratories
MS 9051
Livermore, CA 94551-0969

Isaak D. Mayergoyz
Department of Electrical Engineering
University of Maryland
College Park, MD 20742

Gordon S. Kino
Ginzton Laboratory
Stanford University
MC 4085
Stanford, CA 94305

Michael J. Miksis
Dept. of Engineering Sci. & Applied Mathematics
Northwestern University
2145 Sheridan Road
Evanston, IL 60208

Albert S. Kobayashi
Department of Mechanical Engineering
University of Washington
Box 352600
Seattle, WA 98195-2600

Lisa A. Mondy
Multiphase and Energetic Processes
Sandia National Laboratories
MS 0834
Albuquerque, NM 87185-0834

Dusan Krajcinovic
Department of Mechanical & Aerospace Engineering
Arizona State University
Tempe, AZ 85287-6106

John Moreland
Electromagnetic Technology Division
NIST-Boulder
MS 814.05
325 Broadway
Boulder, CO 80303

Oscar P. Manley
Division of Engineering & Geosciences, ER-15
U.S. Department of Energy
Office of Basic Energy Sciences
19901 Germantown Road
Germantown, MD 20874-1290

David M. Parks
Department of Mechanical Engineering
Massachusetts Institute of Technology
Room 1-308
77 Massachusetts Avenue
Cambridge, MA 02139

Robert E. Price
Division of Engineering & Geosciences, ER-15
U.S. Department of Energy
Office of Basic Energy Sciences
19901 Germantown Road
Germantown, MD 20874-1290

Gregory Stephanopoulos
Department of Chemical Engineering
Massachusetts Institute of Technology
Room 66-552
25 Ames Street
Cambridge, MA 02139

Andrea Prosperetti
Department of Mechanical Engineering
Johns Hopkins University
122 Latrobe Hall
Baltimore, MD 21218

Clifford M. Surko
Department of Physics
University of California-San Diego
9500 Gilman Drive
LaJolla, CA 92093-0319

Reuel Shinnar
Department of Chemical Engineering
City College of CUNY
New York, NY 10031

Kenneth L. Telschow
Department of Materials Physics
Idaho National Engineering Laboratory
LITCO, P.O. Box 1625
Idaho Falls, ID 83415-2209

Gregory Sivashinsky
Benjamin Levich Inst. for Physico-Chemical Hydrodynamics
City College of CUNY
Convent Avenue at 138 Street
New York, NY 10031

Timothy R. Troutt
Department of Mechanical Engineering
Washington State University
Pullman, WA 99164-2920

Pol D. Spanos
Department of Civil/Mechanical Engineering
Rice University
MS 321
6100 Main Street
Houston, TX 77005-1892

Zellman Warhaft
Dept. of Mechanical & Aerospace Engineering
Cornell University
Upson Hall
Ithaca, NY 14853

Peter C. Wayner, Jr.
Department of Chemical Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180-3590

Yizi Xu
NIST
MS 814.05
325 Broadway
Boulder, CO 80303