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.

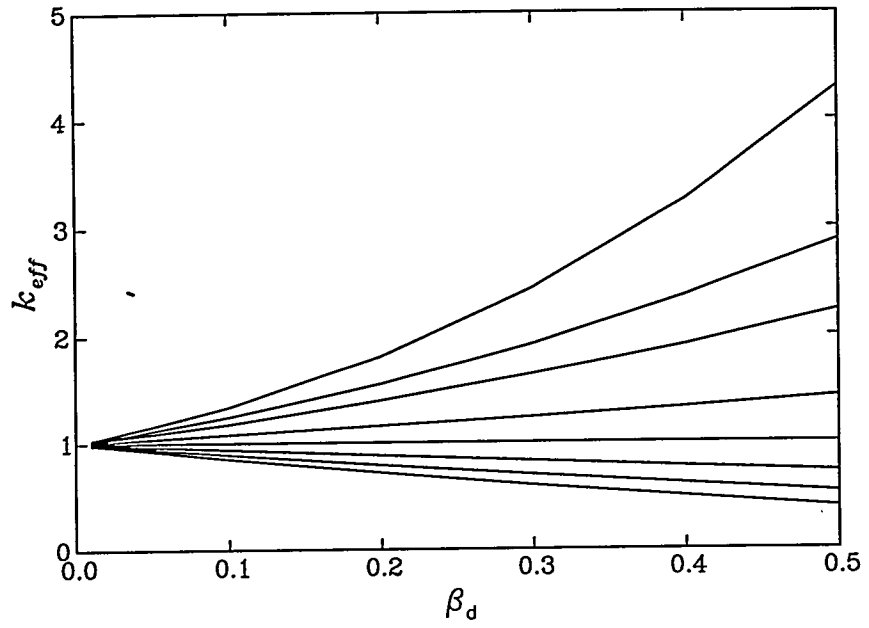


Figure 1: Effective normalized thermal conductivity of a uniform composite with spherical inclusions as a function of the volume fraction of the disperse phase β_D. The lines correspond, from bottom to top, to conductivity ratios k_D/k_C = 0, 0.2, 0.5, 1, 2, 5, 10, 1000.

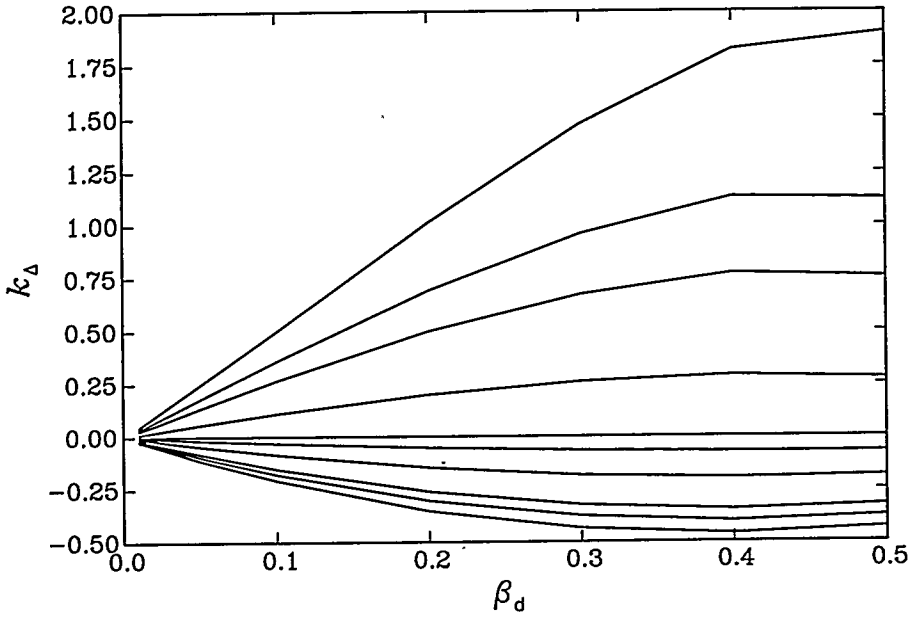


Figure 2: Effective normalized second thermal conductivity composite with spherical inclusions as a function of the volume fraction of the disperse phase β_D. The lines correspond, from bottom to top, to conductivity ratios k_D/k_C = 0, 0.1, 0.2, 0.5, 0.8, 1, 2, 5, 10, 1000.

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

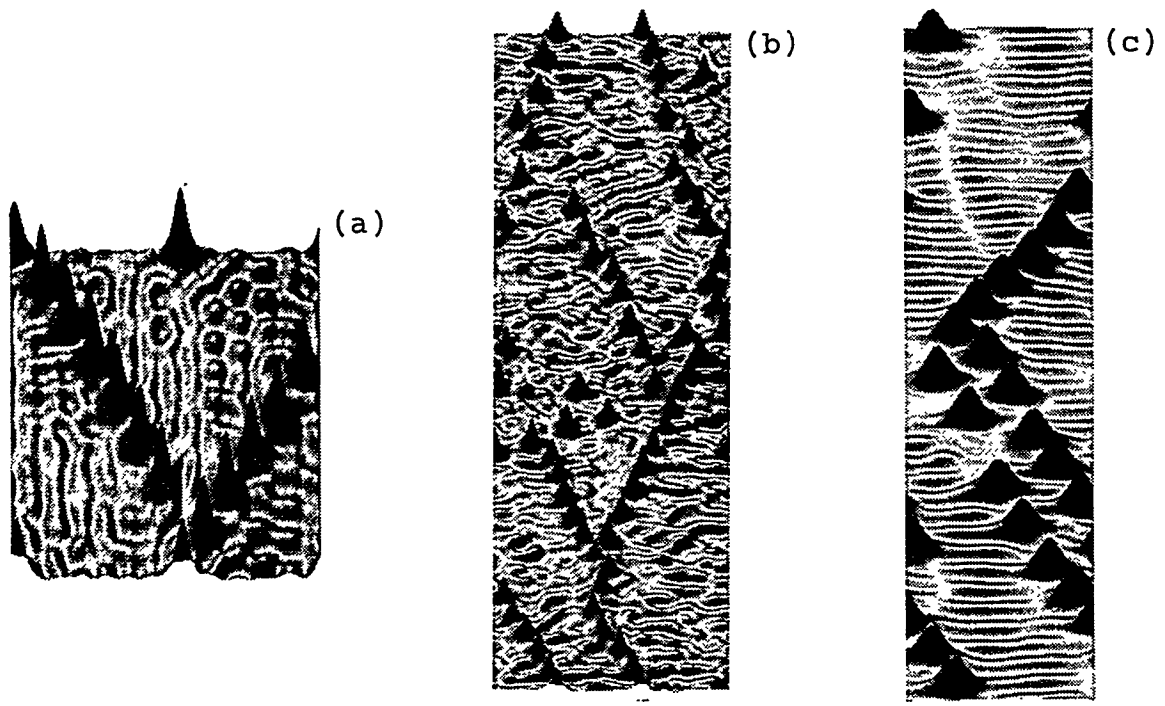


Fig. 1. Snapshots of the time-asymptotic film surface self-organized in simulations of Eq. 5, for three different cases (bulges move down the page here; for convenience of presentation, different axes may have different scales; in reality, all structures have small slopes). (a) $p = q = 16$, $\epsilon^{-1} = 50$, 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 $t = 4.89 \times 10^5$.

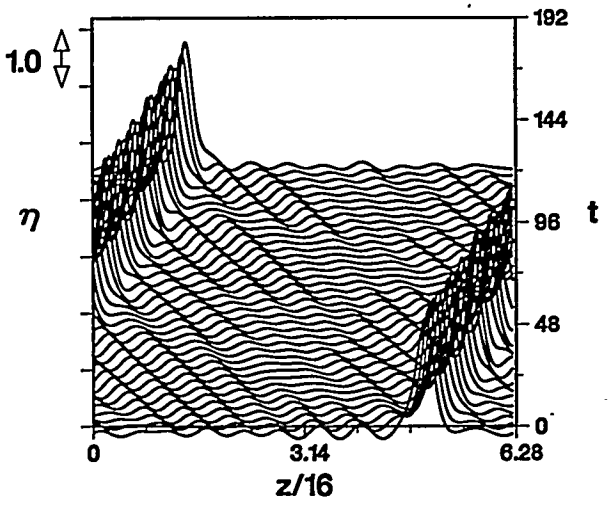


Fig. 2. Time-sequence of instantaneous surface profiles in a fixed cross-section normal to the streamwise direction (for $p = q = 16$, $\epsilon^{-1} = 50$; the time shown as 0 is in fact 1.6×10^5 counting from the start of the run). In particular, it is evident that the (large-amplitude) bulges move in one direction and (small-amplitude) bumps in the opposite direction.

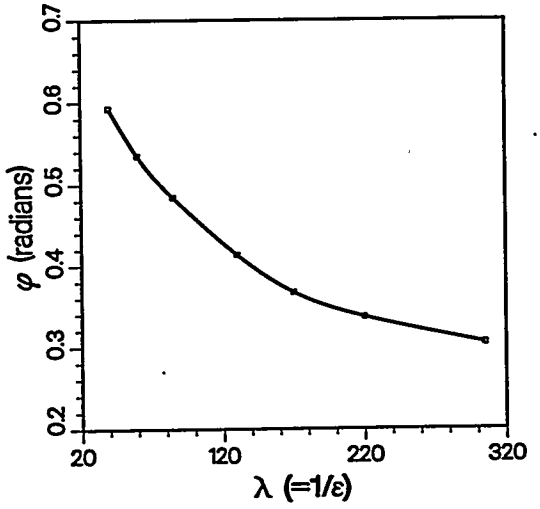


Fig. 3. Angle φ between (every) line of bulges and the streamwise direction varies with ϵ ($p = q = 16$).

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

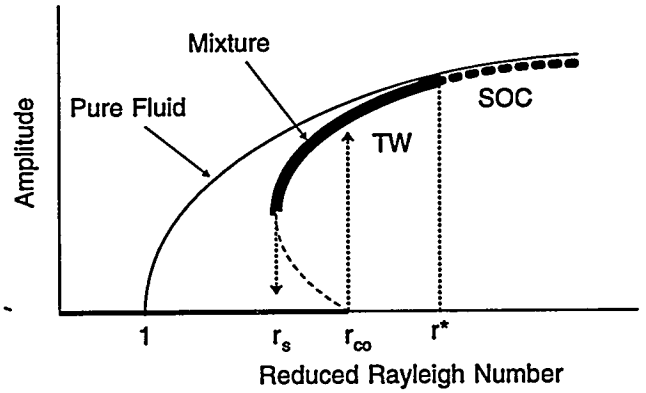


Figure 1: Schematic bifurcation diagram comparing convection in a pure fluid and in a binary mixture with negative separation ratio. In the mixture, the heavy solid line indicates traveling-wave (TW) convection, and the heavy dashed line indicates stationary overturning convection (SOC).

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-

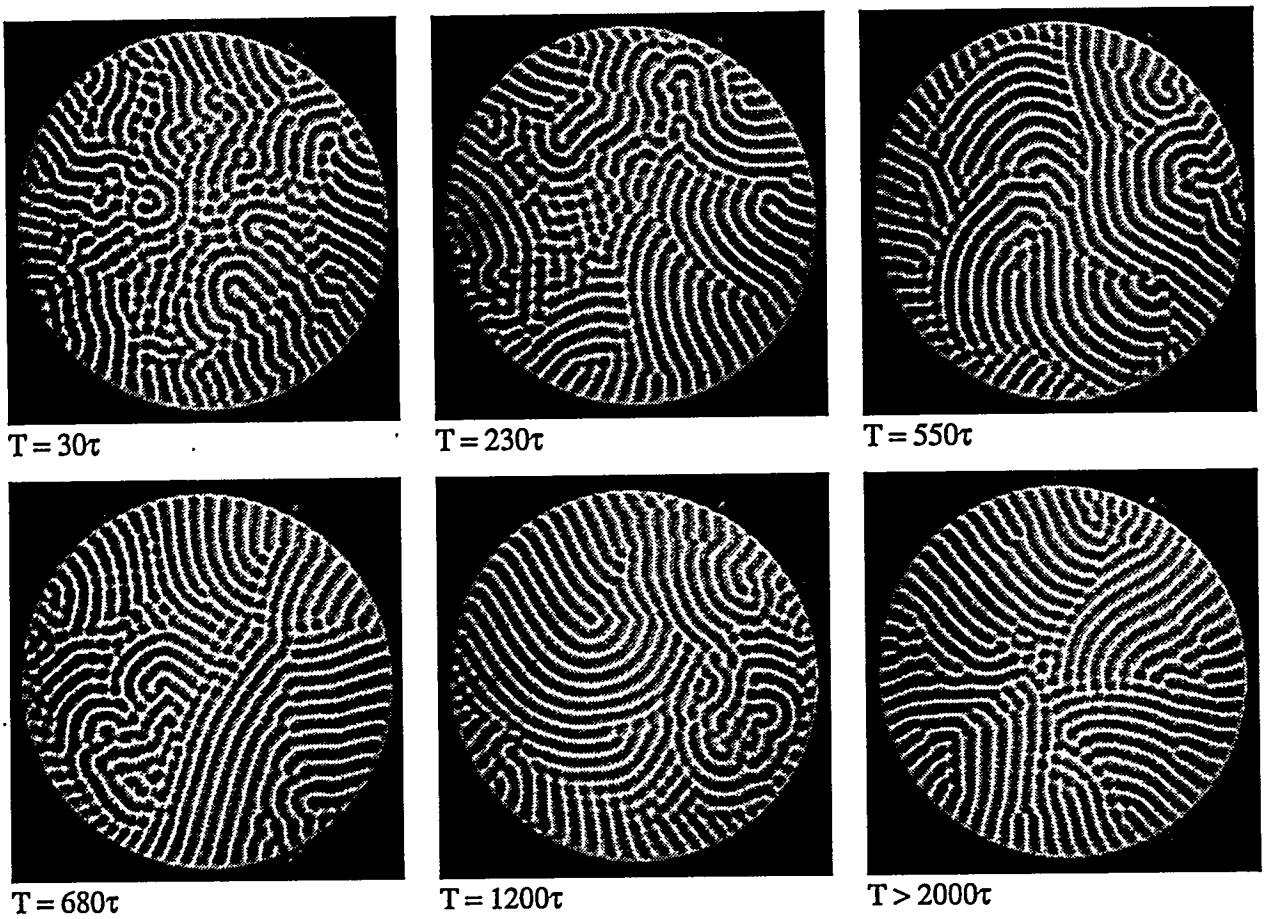


Figure 2: The long-term evolution of a traveling-wave pattern at r=1.28. Time is expressed in terms of the vertical thermal diffusion time, $\tau = 124$ sec.

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

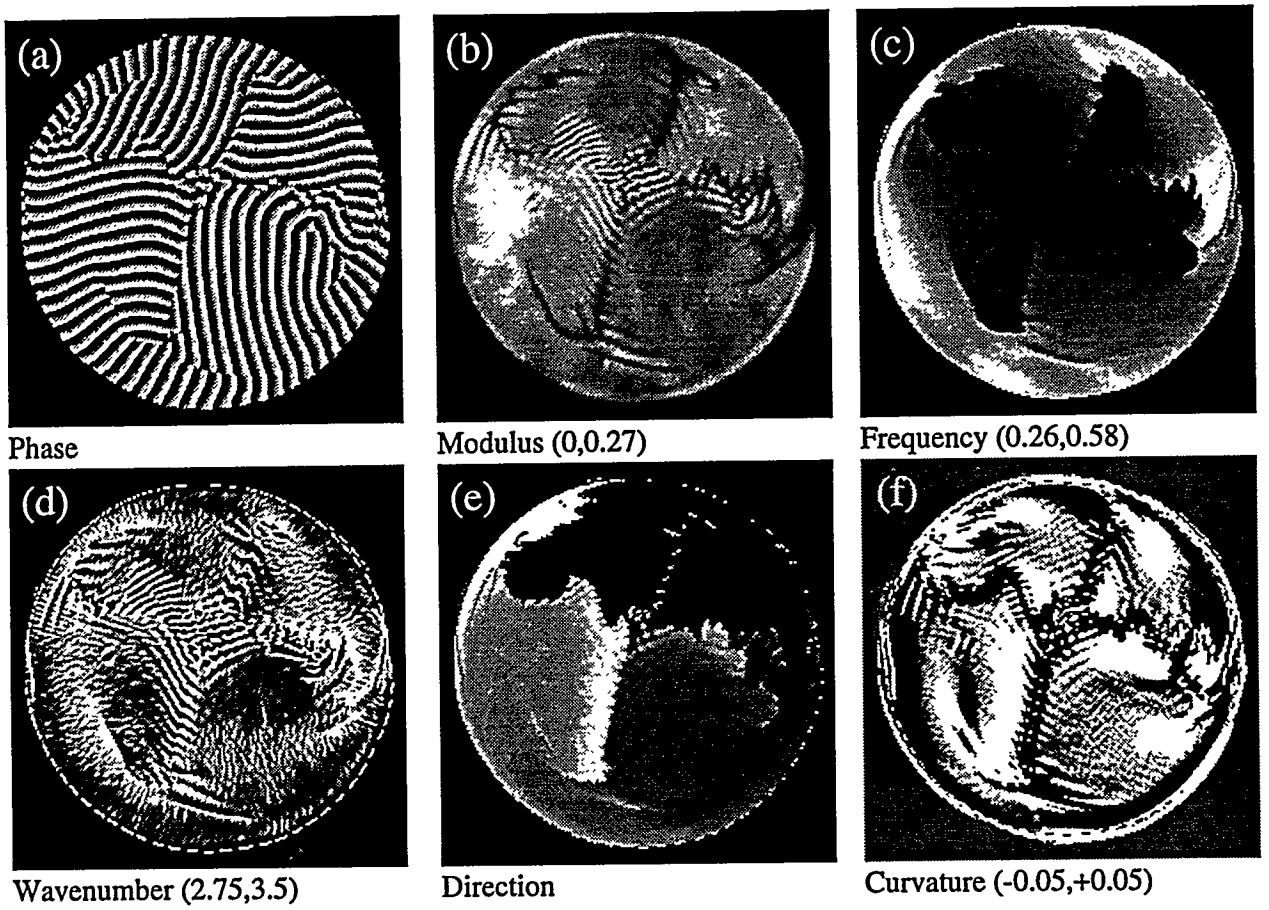
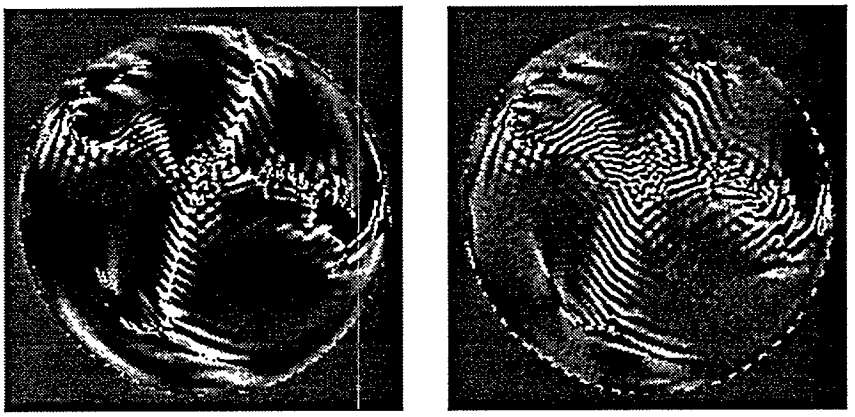


Figure 3: Analysis of a traveling-wave convection pattern showing (a) the phase, ϕ; (b) the modulus, $\|A\|$; (c) the frequency, ω; (d) the wave number $\nabla\phi$; (e) the direction, $tan^{-1}(k_y/k_x)$; and (f) the curvature, $\nabla \cdot \hat{n}$.

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

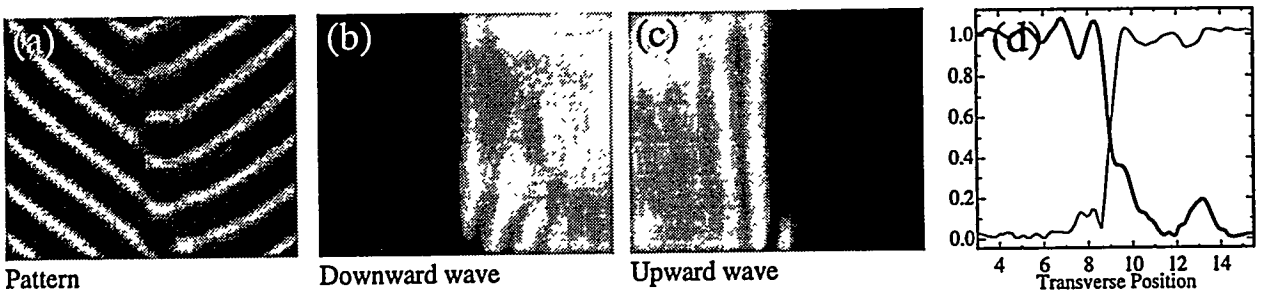


Figure 5: (a) A shadowgraph image of a zipper boundary recorded at $r = 1.37$, (b) the modulus of the wave component measured on the right side of the boundary, (c) the modulus of the wave component measured on the left side of the boundary, (d) the moduli of the two amplitudes along a line transverse 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.

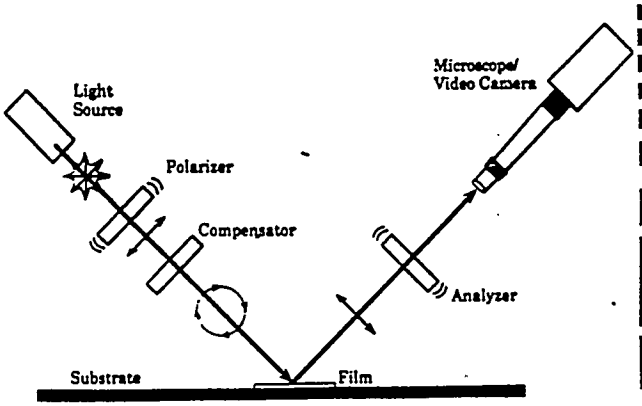


Figure 1. Schematic diagram of the ISE

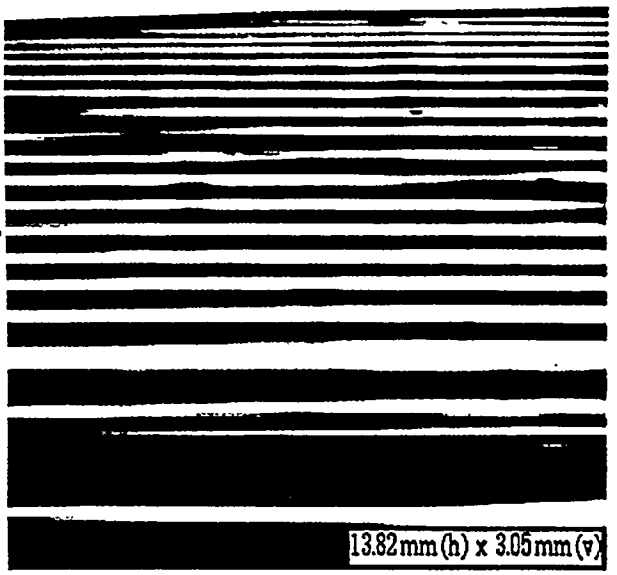


Figure 2. Reflectance fringes from a draining film

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.

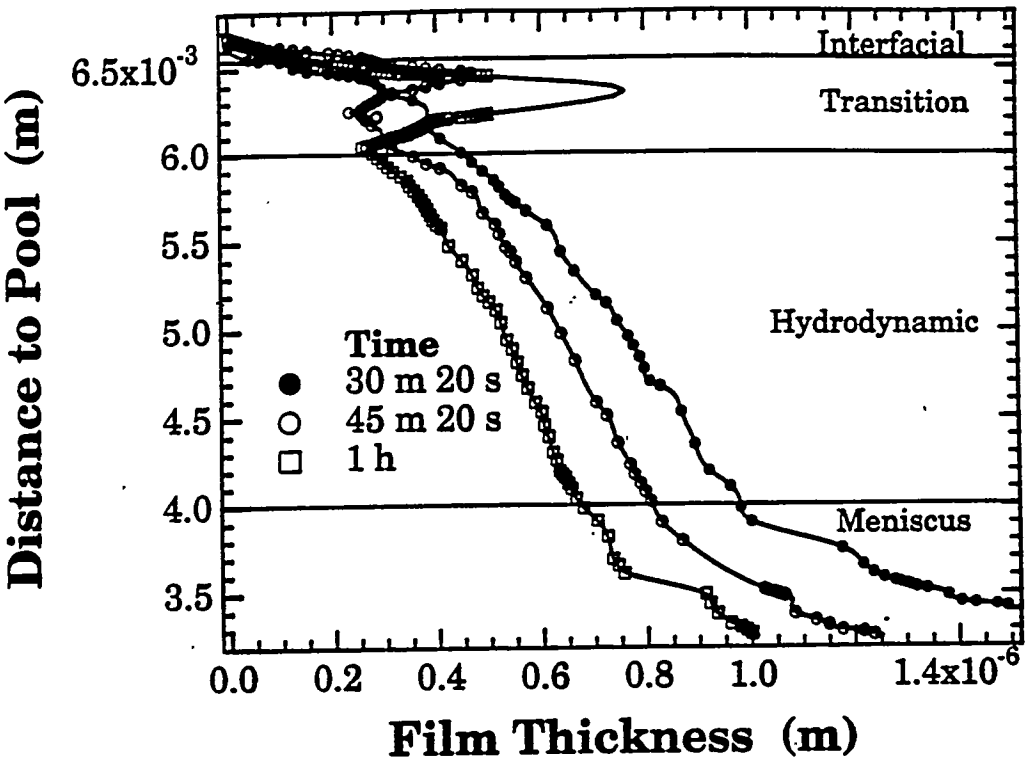


Figure 3. Draining profile of a partially wetting

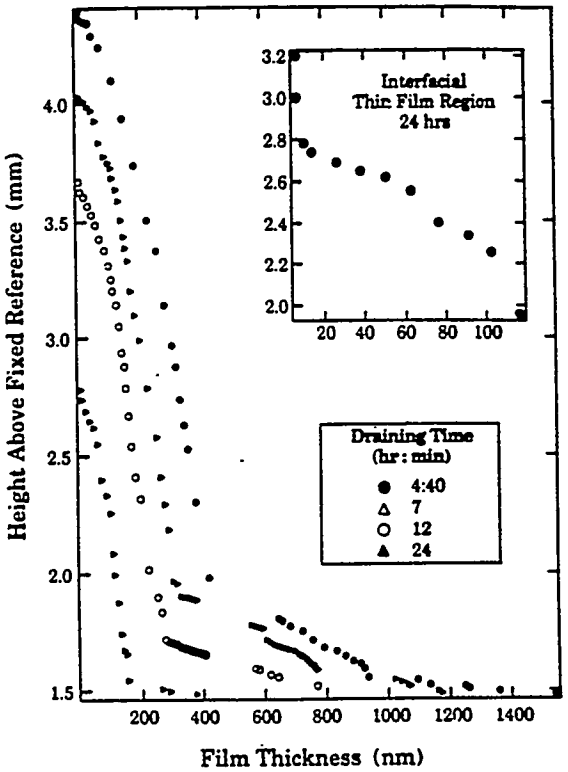


Figure 4. Draining profile of a fully wetting film

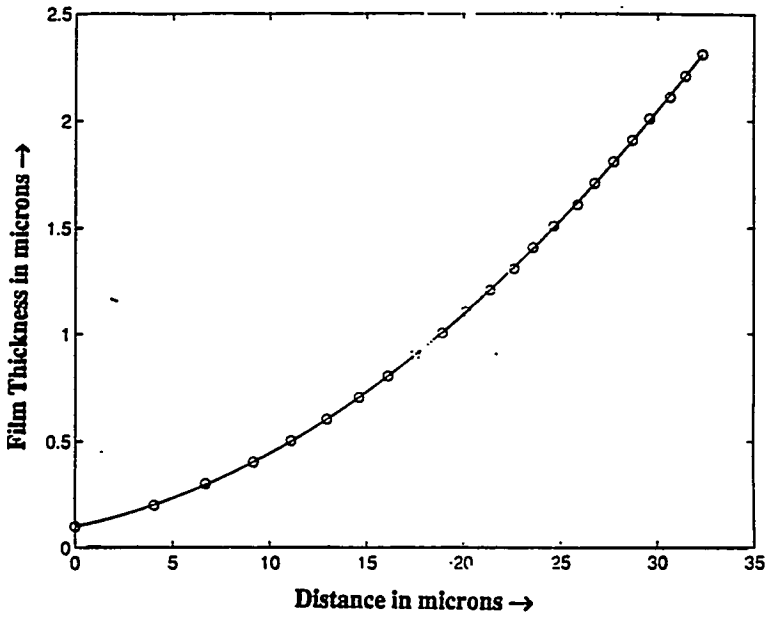


Figure 5. Steady-state film profile of Pentane on a Quartz substrate obtained by IAI

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.

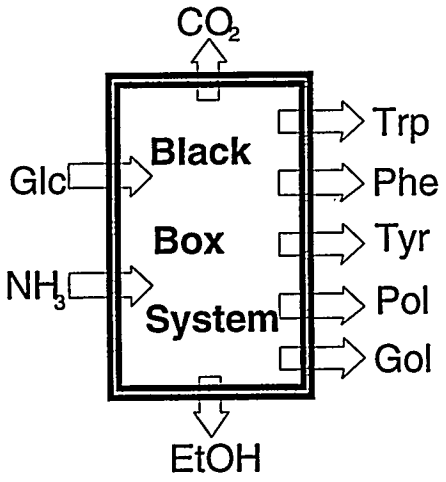


Figure 1

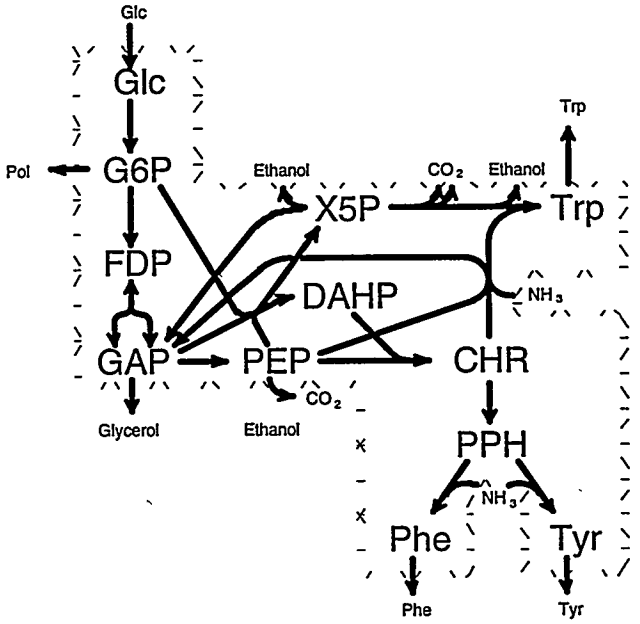


Figure 2

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].

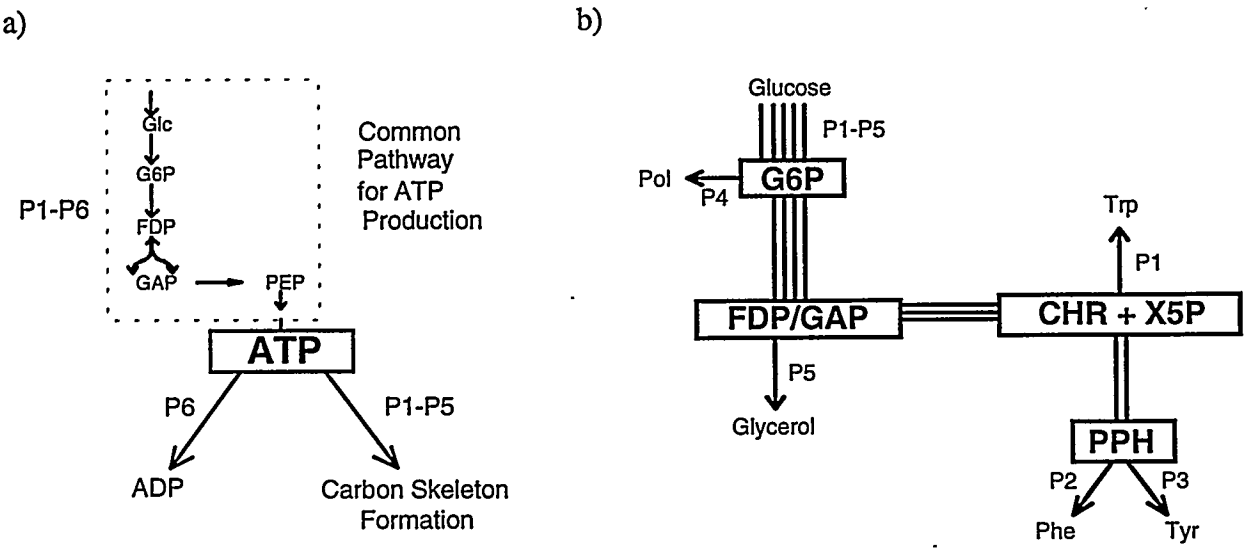


Figure 3

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