KINETIC THEORY AND LONG RANGE CORRELATIONS IN MODERATELY DENSE GASES

T. Petrosky and I. Prigogine

Center for Studies in Statistical Mechanics and Complex Systems,
The University of Texas at Austin, Austin, TX 78712 USA.

ABSTRACT

The complex spectral representation of the Liouville operator is applied to moderately dense gases interacting through hard-core potentials in arbitrary d-dimensional spaces. It is shown that Markovian kinetic equations exist for all d. This provides an answer to the long standing question do kinetic equations exist in two dimensional systems. The non-Markovian effects, such as the long-time tails for arbitrary n-mode coupling, are estimated by superposition of the Markovian evolutions in each subspace as introduced in our spectral decomposition. The long-time tail effects invalidate the traditional kinetic theory based on a truncation of BBGKY hierarchy for $d < 4$, as well as the Green–Kubo formalism, as there appear contributions of order t^{-1}, $t^{-1/2}$, $\cdots$ coming from multiple mode-mode couplings even for $d = 3$.

INTRODUCTION

In spite of much effort the kinetic theory of moderately dense gases faces still a number of difficulties. Does there exist kinetic Markovian equations for two dimensional (2d) systems, such as for example, hard disks? What is the effect of the long time tails on the linear response theory, the so-called Green-Kubo formalism (For an excellent and still up-to-date review of these problems, see ref. [1], as well as the book by Résibois and de Leener [2]. This book will be quoted as *RL*. See also refs. [3-5]). It is well known that reduced distribution functions satisfy master equations which include memory effects (they are "non-Markovian") [6]. Moreover, these equations contain still the full complexity of N-body problem and are mostly formal. We want to show that our recent work on the spectral decomposition of the Liouville operator [7] [†] provides an answer to the questions we mentioned. Using our spectral decomposition the non-Markovian evolution can be split rigorously into independent Markovian evolutions. In line with earlier work (see *RL Ch.*X) we shall consider the case of "hard spheres" in an arbitrary number d of dimensions. We shall show that kinetic Markovian equations exist for all d. Moreover, we can estimate the long-time tails for arbitrary n-mode coupling and show that they lead to divergences of the Green–Kubo integrals for $d \leq 4$.

[†] This will be quoted as *I*. For quantum mechanics see ref. [8].

COMPLEX SPECTRAL REPRESENTATIONS

We start with a short summary of our method (see I). The Liouville equation for the N-particle distribution function (d.f.) $\rho(\mathbf{r}^N, \mathbf{v}^N, t)$ is $i\partial\rho/\partial t = L\rho$. The Liouvillian L consists of a free Liouvillian L_0 and an interaction part δL, i.e., $L = L_0 + \delta L = \sum_{a=1}^{N} L_0^{(a)} + \sum_{b>a=1}^{N} \delta L^{(ab)}$ where $L_0^{(a)} = -i\mathbf{v}_a \cdot (\partial/\partial\mathbf{r}_a)$. Our method deals with the class of ensembles corresponding to the thermodynamic limit (i.e., the number of particle $N \to \infty$, and the volume $V \to \infty$, with the number density $n = N/V$ finite). This class includes the canonical distribution. It describes "persistent" interactions. The eigenfunctions of L_0 are plane waves ($\sim \exp[i\sum_j \mathbf{k}_j \cdot \mathbf{r}_j]$). The distribution function ρ therefore can be decomposed into contributions from different degrees of correlations, $\rho = P^{(0)}\rho + P^{(1)}\rho + P^{(2)}\rho + \cdots$, where $P^{(\nu)}$ is the projection operator which retains the ν-th degree of correlations (see I). For example, $P^{(0)}$ retains the "vacuum component" $\rho_0(\mathbf{v}^N)$, which is the velocity d.f. with vanishing wavevectors for all particles $\mathbf{k}^N = 0$, $P^{(1)}\rho$ the "inhomogeneity components" $\rho_{\mathbf{k}_a}(\mathbf{v}^N)$ with $\mathbf{k}_a \neq 0$ and $\mathbf{k}^{N-1} = 0$ for particles $a = 1$ to N, $P^{(2)}\rho$ the "binary correlation components" $\rho_{\mathbf{k}_a, \mathbf{k}_b}$ with $\mathbf{k}_a \neq 0$ and $\mathbf{k}_b \neq 0$, and so on. We have $L_0 P^{(\nu)} = P^{(\nu)} L_0$, $P^{(\nu)} P^{(\mu)} = P^{(\nu)} \delta_{\nu, \mu}$ and $\sum_\nu P^{(\nu)} = 1$.

As we have shown (see $(I.45)$), the Fourier coefficients with a smaller number of nonvanishing components in wavevectors $\mathbf{k}^N$ have a higher order dependence in the volume factor V in the large volume limit, for example $\rho_0/\rho_{\mathbf{k}_a} \sim V$. The appearance of these delta-function singularities leads to the usual cluster expansion of the reduced distribution functions. Also, because of these singularities, the d.f. ρ *does not* belong to the Hilbert space, as its Hilbert norm vanishes in the thermodynamic limit, $\langle\rho|\rho\rangle \equiv \int d^N x \rho^*(x^N)\rho(x^N) \to 0$, where $x^N = (\mathbf{r}^N, \mathbf{v}^N)$. As a result, the hermitian operator L acquires *complex* eigenvalues $Z_j^{(\nu)}$ breaking time-symmetry. The time evolution of the system then splits into two semigroups. For $\text{Im}\, Z_j^{(\nu)} \leq 0$ equilibrium is approached in our future that is for $t \to +\infty$, while for $\text{Im}\, Z_j^{(\nu)} \geq 0$ equilibrium had been reached in our past. The domain of the two semigroups do not overlap (see I and ref. [8]). To be self-consistent we choose the semigroup oriented towards our future. Then we have the complex spectral representation,

$$e^{-iLt}|\rho(0)\rangle = \sum_\nu \sum_j |F_j^{(\nu)}\rangle e^{-iZ_j^{(\nu)}t}\langle\tilde{F}_j^{(\nu)}|\rho(0)\rangle \qquad (1)$$

where $F_j^{(\nu)}$ and $\tilde{F}_j^{(\nu)}$ are bi-orthonormal sets of right- and left-eigenstates of the Liouvillian with the complex eigenvalue $Z_j^{(\nu)}$, respectively. This spectral decomposition has to be used in conjunction with observables depending on a finite number of variables (see $(I.67)$).

The eigenstates of L are given by $|F_j^{(\nu)}\rangle = n_{\nu,j}[P^{(\nu)} + Q^{(\nu)}\mathcal{C}^{(\nu)}(Z_j^{(\nu)})]|u_j^{(\nu)}\rangle$ and $\langle\tilde{F}_j^{(\nu)}| = n_{\nu,j}\langle\tilde{v}_j^{(\nu)}|[P^{(\nu)} + \mathcal{D}^{(\nu)}(Z_j^{(\nu)})Q^{(\nu)}]$ (see $(I.98)$), where $Q^{(\nu)} = 1 - P^{(\nu)}$ and $n_{\nu,j}$ is a normalization constant. The "creation operator" $\mathcal{C}^{(\nu)}(z)$ and the "destruction operator" $\mathcal{D}^{(\nu)}(z)$ are defined by $\mathcal{C}^{(\nu)}(z) \equiv G_Q^\nu(z)LP^{(\nu)}$ and $\mathcal{D}^{(\nu)}(z) \equiv P^{(\nu)}LG_Q^\nu(z)$ with the propagator $G_Q^\nu(z) = Q^{(\nu)}[z - Q^{(\nu)}LQ^{(\nu)}]^{-1}$.

The states $|u_j^{(\nu)}\rangle$ and $\langle\tilde{v}_j^{(\nu)}|$ are right- and left-eigenstates of the "collision operators" $\psi^{(\nu)}$, i.e., $\psi^{(\nu)}(Z_j^{(\nu)})|u_j^{(\nu)}\rangle = Z_j^{(\nu)}|u_j^{(\nu)}\rangle$ and $\langle\tilde{v}_j^{(\nu)}|\psi^{(\nu)}(Z_j^{(\nu)}) = \langle\tilde{v}_j^{(\nu)}|Z_j^{(\nu)}$ where $\psi^{(\nu)}$ is given by

$$\psi^{(\nu)}(z) = P^{(\nu)}LP^{(\nu)} + P^{(\nu)}LC^{(\nu)}(z)P^{(\nu)} \qquad (2)$$

The collision operators are dissipative operators, and they are central objects of non-equilibrium statistical mechanics (see I). The operators $\mathcal{C}^{(\nu)}$, $\mathcal{D}^{(\nu)}$, and $\psi^{(\nu)}$ are associated to graphs such as represented for $\nu = 0$ in Fig. 1.
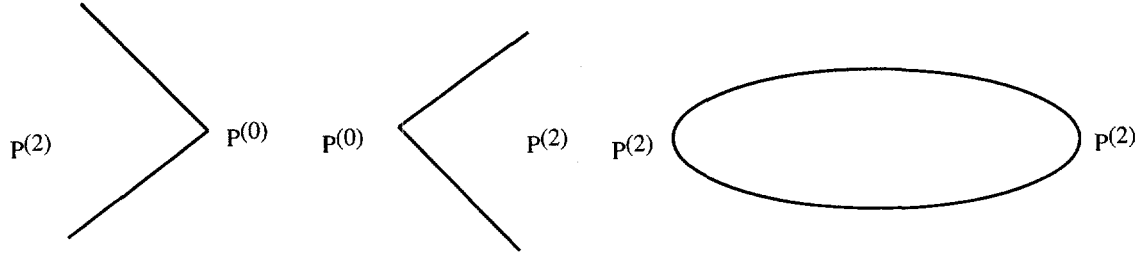
Figure 1: Diagram (a) corresponds to the lowest order contribution of the creation operator $\mathcal{C}^{(0)}$, (b) to the destruction operator $\mathcal{D}^{(0)}$, and (c) to the collision operator $\psi^{(0)}$.

Note that the eigenvalue problem associated to the collision operators $\psi^{(\nu)}(Z_j^{(\nu)})$ is nonlinear as the eigenvalue $Z_j^{(\nu)}$ appears inside the operators. This will play an important role. Also the above relations show that the Liouville operator shares the same eigenvalues with the collision operators. Assuming bi-completeness in the space $P^{(\nu)}$, we may always construct a set of states $\{\langle \tilde{u}_j^{(\nu)}|\}$ bi-othogonal to $\{|u_j^{(\nu)}\rangle\}$, i.e., $\langle \tilde{u}_j^{(\nu)}|u_\beta^{(\mu)}\rangle = \delta_{\nu,\mu}\delta_{\alpha,\beta}$ and $\sum_j |u_j^{(\nu)}\rangle\langle \tilde{u}_j^{(\nu)}| = P^{(\nu)}$. In general, $\langle \tilde{v}_j^{(\nu)}| \neq \langle \tilde{u}_j^{(\nu)}|$ (see I).

In our earlier work, we have repeatedly introduced the concept of "subdynamics." [9,10] The relation of subdynamics to the complex spectral representation can be seen through the projection operators $\Pi^{(\nu)}$ defined by $\Pi^{(\nu)} \equiv \sum_j |F_j^{(\nu)}\rangle\langle \tilde{F}_j^{(\nu)}|$. They satisfy the orthogonality and completeness relations, $L\Pi^{(\nu)} = \Pi^{(\nu)}L$, $\Pi^{(\nu)}\Pi^{(\mu)} = \Pi^{(\nu)}\delta_{\nu,\mu}$, and $\sum_\nu \Pi^{(\nu)} = 1$ (see (I.139)). $\Pi^{(\nu)}$ is an extension of $P^{(\nu)}$ to the total Liouvillian L. Because these projection operators commute with the Liouvillian, each component $|\rho^{(\nu)}\rangle \equiv \Pi^{(\nu)}|\rho\rangle$ individually satisfies the Liouville equation. For this reason, the projection operators $\Pi^{(\nu)}$ are associated with *subdynamics*. Then, (1) leads to a *Markovian equation* which is a closed equation for the "privileged component" $P^{(\nu)}\rho^{(\nu)}(t)$ of each subdynamics $\rho^{(\nu)}(t)$,

$$i\frac{\partial}{\partial t}P^{(\nu)}|\rho^{(\nu)}(t)\rangle = \theta^{(\nu)}P^{(\nu)}|\rho^{(\nu)}(t)\rangle \tag{3}$$

where we have introduced the "global" collision operator

$$\theta^{(\nu)} \equiv \sum_j \psi^{(\nu)}\left(Z_j^{(\nu)}\right)|u_j^{(\nu)}\rangle\langle \tilde{u}_j^{(\nu)}| = \sum_j |u_j^{(\nu)}\rangle Z_j^{(\nu)}\langle \tilde{u}_j^{(\nu)}| \tag{4}$$

This result is important, since as mentioned any component $P^{(\nu)}$ of the total $\rho(t)$ obeys a non-Markovian equation with memory effects (see *RL*). These effects can now be describe by the superposition of Markov processes.

APPLICATION TO MODERATELY DENSE GASES

After this brief summary we now consider the application of our approach to moderately dense systems which consist of N hard spheres ($d = 3$), or hard disks ($d = 2$) with the same mass m. We use the "pseudo-Liouvillian" formalism to take into account the singular nature of the hard sphere interaction (*RL Ch*.X, see also ref. [11]). We also consider situations near equilibrium, then the *linearized* regime of the kinetic equations is applicable (see *RL Ch*.V). Reducing the kinetic equation (3) to one-particle d.f.s, we obtain Markovian equations for the one particle velocity d.f. $f_0^{(0)}(\mathbf{v}_1)$ (which corresponds to the vacuum of correlation $\nu = 0$) as well

as for the inhomogeneous component $f_{\mathbf{q}}^{(1)}(\mathbf{v}_1)$ (for which $\nu = 1$):

$$\frac{\partial}{\partial t} f_{\mathbf{q}}^{(\nu)}(\mathbf{v}_1, t) = L_{\mathbf{q}}^{(\nu)} f_{\mathbf{q}}^{(\nu)}(\mathbf{v}_1, t) \tag{5}$$

where $L_{\mathbf{q}}^{(\nu)}$ are the reduced one particle collision operators corresponding to (4). More explicitly

$$L_{\mathbf{q}}^{(\nu)} = \sum_j \left[-i\mathbf{q} \cdot \mathbf{v}_1 + nK_1 + \delta\mathcal{L}_{\mathbf{q}}^{\nu}(\xi_j^{\mathbf{q}}) \right] |\chi_j^{\mathbf{q}}\rangle\!\rangle \langle\!\langle \tilde{\chi}_j^{\mathbf{q}}| = \sum_j \xi_j^{\mathbf{q}} |\chi_j^{\mathbf{q}}\rangle\!\rangle \langle\!\langle \tilde{\chi}_j^{\mathbf{q}}| \tag{6}$$

where $\xi = -iz$ and the states $|\chi_j^{\mathbf{q}}\rangle\!\rangle$ and $\langle\!\langle \tilde{\chi}_j^{\mathbf{q}}|$ in (6) are the right- and left-eigenstates of the reduced collision operator $L_{\mathbf{q}}^{(\nu)}$ which shares the same complex eigenvalue $Z_j = i\xi_j$ with the Liouvillian. There are three processes included in (6), the first coming from L_0, the second being the linearlized Boltzmann operator nK_1, and the third coming from mode–mode coupling. To distinguish the reduced one-particle states $|\chi_j^{\mathbf{q}}\rangle\!\rangle$ from the N-particle states $|u_j^{(\nu)}\rangle$, we have used the double-ket notation for the reduced states. The inner product of the reduced states is defined as $\langle\!\langle g|f\rangle\!\rangle \equiv \int d\mathbf{v}_1 \varphi^{eq}(v_1)^{-1} g^*(\mathbf{v}_1) f(\mathbf{v}_1)$, where φ^{eq} is the normalized Maxwellian $\varphi^{eq}(v) = (\beta m/2\pi)^{d/2} \exp(-\beta m v^2/2)$.

The operator K_a is the *linearized Boltzmann collision operator/* defined by (see *RL Chs.* V and X)[‡]

$$K_a \Phi(\mathbf{v}_a) = \int d\mathbf{v}_b \langle 0|t^{(ab)}|0\rangle \left(1 + \hat{P}_{ab}\right) \varphi^{eq}(v_b) \Phi(\mathbf{v}_a) \tag{7}$$

where $|0\rangle \equiv |\mathbf{k}^N = 0\rangle$, and $\hat{P}_{ab}$ is an exchange operator of the velocities $\mathbf{v}_a$ and $\mathbf{v}_b$. The "binary collision operator" $t^{(ab)}$ is defined in the pseudo-Liouvillian formalism by

$$t^{(ab)} = -iV\delta L^{(ab)} \equiv a_0^{d-1} \int_{\hat{\mathbf{s}}\cdot\mathbf{v}_{ab}>0} d\hat{\mathbf{s}} \, (\hat{\mathbf{s}} \cdot \mathbf{v}_{ab}) \left[\delta\left(\mathbf{r}_{ab} - a_0\hat{\mathbf{s}}\right) \hat{b}_{\mathbf{v}}^{(ab)} - \delta\left(\mathbf{r}_{ab} + a_0\hat{\mathbf{s}}\right) \right] \tag{8}$$

Here, a_0 is a diameter of the hard spheres, $\hat{\mathbf{s}}$ is the unit vector, the operator $\hat{b}_{\mathbf{v}}^{(ab)}$ replaces $\mathbf{v}_a$ and $\mathbf{v}_b$ by their pre-collisional values $\mathbf{v}_a'$ and $\mathbf{v}_b'$, and $\mathbf{v}_{ab} \equiv \mathbf{v}_a - \mathbf{v}_b$.

Retaining first only two-mode coupling, the operator $\delta\mathcal{L}_{\mathbf{q}}^{\nu}$ in (6) becomes

$$\delta\mathcal{L}_{\mathbf{q}}^{\nu}(\xi) = \frac{in}{(2\pi)^d} \int d\mathbf{k} \int d\mathbf{v}_b \langle a_{\mathbf{q}}|t^{(ab)}|a_{\mathbf{q-k}}, b_{\mathbf{k}}\rangle \langle a_{\mathbf{q-k}}, b_{\mathbf{k}}|g_2(\xi) t^{(ab)}|a_{\mathbf{q}}\rangle \left(1 + \hat{P}_{ab}\right) \varphi^{eq}(v_b) \tag{9}$$

where $|a_{\mathbf{k}}\rangle \equiv |\mathbf{k}_a = \mathbf{k}, \mathbf{k}^{N-1} = 0\rangle$, $|a_{\mathbf{k}}, b_{\mathbf{k}'}\rangle \equiv |\mathbf{k}_a = \mathbf{k}, \mathbf{k}_b = \mathbf{k}', \mathbf{k}^{N-2} = 0\rangle$, and the reduced two-particle propagator is defined by $g_2(\xi) \equiv [\xi + iL_0^{(ab)} - nK_a - nK_b + t^{(ab)}]^{-1}$ with $L_0^{(ab)} \equiv L_0^{(a)} + L_0^{(b)}$. The propagator has to be evaluated as an analytic continued function from the right-half plane of ξ. To obtain (9) we have approximated the N-particle propagator G_Q^ν in (2) by retaining only the binary correlation subspace, i.e., $G_Q^\nu \approx G_2 \equiv P^{(2)}[z - P^{(2)}LP^{(2)}]^{-1}$.

One can expand the propagator g_2 in (9) in a geometrical series of $t^{(ab)}$ using the propagator $g_2'(\xi) \equiv [\xi + iL_0^{(ab)} - nK_a - nK_b]^{-1}$. Retaining only the lowest order term in the geometrical series (i.e., replacing g_2 by g_2' in (9)) we obtain as an approximation of $\delta\mathcal{L}_{\mathbf{q}}^{\nu}(\xi)$, the operator $\delta\mathcal{L}_{\mathbf{q}}'^{\nu}(\xi)$. This new operator $\delta\mathcal{L}_{\mathbf{q}}'^{\nu}(\xi)$ is well known (see *RL Ch.* X). For $\xi = \epsilon \to 0+$ it corresponds to the so-called "ring operator" whose graphical representation is given in Fig. 2.

[‡] In *RL* it denoted by $C_a^{(1)}$.

135

Figure 2: Graphical representation of the simplest "ring process," where the dots represent the linearized Boltzmann operator.

In Fig. 2 the propagator are renormalized by the linearized Boltzmann operators (c.f. RL(X.113)). From the ring diagrams with $\xi = +\epsilon$, one then obtain the kinetic equation for the one particle velocity d.f. $\varphi(\mathbf{v}_1)$: (c.f. RL (X.73))

$$\frac{\partial}{\partial t}\varphi(\mathbf{v}_1, t) = \left[nK_1 + \delta\mathcal{L}_0'^{\,0}(+\epsilon) \right]\varphi(\mathbf{v}_1, t) \tag{10}$$

In contrast, in our approach the ring operator $\delta\mathcal{L}'^0(+\epsilon)$ is replaced by $\delta\mathcal{L}_0'^{\,0}(\xi_j^0)$. Here, ξ_j^0 is of order γ, instead of $\xi = +\epsilon$, where $\gamma^{-1} = t_r = (na^{d-1}v_0)^{-1}$ is the mean free time for a moderate dense system, and $v_0 = (m\beta)^{-1/2}$ is the "thermal velocity." This difference is essential, because our operator is finite while the ring operator $\delta\mathcal{L}_0'^{\,0}(+\epsilon)$ diverges for $d = 2$. This can be shown by separating the contribution of (9) into two parts, the "hydrodynamic part" and the "nonhydrodynamic part," $\delta\mathcal{L}_0'^{\,0} = \delta\mathcal{L}_h'^{\,0} + \delta\mathcal{L}_{nh}'^0$. Here, $\delta\mathcal{L}_h'^{\,0}$ is defined as the contribution of the integration over $\mathbf{k}$ restricted to the domain $|\mathbf{k}| < k_0$, while $\delta\mathcal{L}_{nh}'^0$ deals with the domain $|\mathbf{k}| > k_0$, where k_0^{-1} is a constant wave number of the order of the mean free path.

For the hydrodynamic part, the real part of the eigenvalues $\lambda_\alpha^{\mathbf{k}}$ of the linearlized Boltzmann collision operator nK_1 are given as $\mathrm{Re}\,\lambda_\alpha^{\mathbf{k}} = -k^2 D_\alpha$, where D_α are "diffusion coefficients" associated to summable invariants $\alpha \in (1, 2, \cdots, d+2)$ (see RL). The integration over the hydrodynamic modes then leads to a contribution of the order

$$\delta\mathcal{L}'^0_h(\xi_\alpha^0) \sim n\left(a_0^{d-1}v_0\right)^2 \int_0^{k_0} dk \frac{k^{d-1}}{k^2 D + \gamma} \tag{11}$$

where we have approximated $\xi_\alpha^0 \sim \gamma$, $D_\alpha \sim D \equiv v_0(a_0^{d-1}n)^{-1}$, and $t^{(12)} \sim a_0^{d-1}v_0$. This integral exists for all dimensions. For example, for $d = 2$ we have $\delta\mathcal{L}_h'^{\,0} \sim (\tau_0/t_r)\ln[1 + (k_0 l_m)^2]$, where we have introduced the notations $l_m \equiv \sqrt{D/\gamma} = (a_0^{d-1}n)^{-1} \sim k_0^{-1}$ for the mean free path, and $\tau_0 \equiv t_c/t_r = a_0^d n$ with $t_c = a_0/v_0$. It is easy to show that the geometrical series of the propagator in g_2 converges for moderately dense systems [12]. Furthermore it is well-known that the non-hydrodynamic part of the ring diagram gives a finite correction to the Boltzmann collision operator (see refs. [13,14] and also RL). Therefore, Markovian kinetic equations exist for all dimensions. This is in contrast to (10), as (11) is the *diverging* integral for $d = 2$ if γ is replaced by $\epsilon \to 0+$.

For the inhomogeneous component $\mathcal{L}_{\mathbf{q}}^{(1)}$ the eigenvalue equation for the right-eigenstate is given by

$$\left[-i\mathbf{q}\cdot\mathbf{v}_1 + nK_1 + \delta\mathcal{L}_{\mathbf{q}}^1(\xi_j^{\mathbf{q}}) \right]|\chi_j^{\mathbf{q}}\rangle\!\rangle = \xi_j^{\mathbf{q}}|\chi_j^{\mathbf{q}}\rangle\!\rangle \tag{12}$$

In the hydrodynamic case with $|\mathbf{q}| < k_0$ for $d = 3$, this equation reduces to the same eigenvalue equation as introduced by Ernst and Dorfman to discuss the nonanalytic dispersion relations

in classical fluids. [15] We can repeat this calculation for $d = 2$. We then obtain the following nonanalytic solution for the class of eigenvalues ξ_α^q which vanish in the limit $q \to 0$ for the hydrodynamic modes $\alpha = 1, \cdots, 4$ [12]

$$\xi_\alpha^q = -iqc\sigma_\alpha - q^2 D_\alpha + q^2 \Delta_\alpha \ln\left(\frac{q}{k_0}\right), \qquad \text{for } d = 2 \tag{13}$$

where c is the sound velocity, $\sigma_1 = -\sigma_2 = 1$ for the sound-modes, and $\sigma_3 = \sigma_4 = 0$ for the share-mode and the heat-mode, $D_\alpha > 0$, $\Delta_\alpha > 0$ and $\Delta_\alpha/D_\alpha \sim \tau_0$. In contrast to 3d systems (where the nonanalyticity is given as $\sim q^{5/2}$ [15,16]) the nonanalytic term of order $q^2 \ln q$ in (13) dominates in the real part of the eigenvalues for extremely small values of q satisfying $q/k_0 < \exp(-1/\tau_0)$.

As in the homogeneous case, the original ring diagram $\delta \mathcal{L}_q^1(+\epsilon)$ diverges for the 2d case. Our spectral analysis removes the divergence and leads to a well-defined Markovian kinetic equation (5) in each subspace. Moreover, we have shown that (12) has a well-defined limit for $q \to 0$, that is, the right-hand side of (12) vanishes in this limit [12]. Hence we have a consistent description of the approach to equilibrium.

The qualitative behavior of the eigenvalues of the reduced collision operators for higher-order correlation subspace ($\nu \geq 2$) is the same as in the inhomogenous subspace with $\nu = 1$ in (12). The spectral analysis leads to well-defined Markovian kinetic equations in every subspace. Moreover, we have again a singular contribution proportional to $q^2 \ln q$ for the 2d case in addition to the diffusion process $\sim q^2$ in each subspace [12].

So far, we have considered the time evolution of reduced d.f.s $f_q^{(\nu)}$ that correspond to the privileged components of the distribution functions $P^{(\nu)}\rho^{(\nu)}$ in (3). They obey Markovian kinetic equations. In actual experiments, however, we follow the evolution of the distribution functions obtained by the reduction of $P^{(\nu)}\rho$, such as $\varphi(\mathbf{v}_1)$ in (10), and not of $P^{(\nu)}\rho^{(\nu)}$. In our spectral representation, the long-time tail effects in $\varphi(\mathbf{v}_1)$ is obtained by superposing the Markovian evolution occurring in each subspace. Let us consider the initial condition $|\rho(0)\rangle = P^{(0)}|\rho(0)\rangle$. Then, (1) gives us

$$\varphi(\mathbf{v}_1, t) = L^{Nd/2} \sum_\nu \sum_j \int d\mathbf{v}_2 \cdots \int d\mathbf{v}_N \langle 0| \left(P^{(\nu)} + \mathcal{C}^{(\nu)}\left(Z_j^{(\nu)}\right)\right) |u_j^{(\nu)}\rangle$$
$$\times n_{\nu,j}^2 e^{-iZ_j^{(\nu)}t} \langle \tilde{v}_j^{(\nu)}| \left(P^{(\nu)} + \mathcal{D}^{(\nu)}\left(Z_j^{(\nu)}\right)\right) |0\rangle\langle 0|\rho(0)\rangle \tag{14}$$

Let us analyze more closely the creation operator $\langle 0|\mathcal{C}^{(\nu)}(z)$ which leads from ν to 0. It corresponds to various dynamical processes (involving vertices such as $P^{(2)}\delta L P^{(3)}, \cdots$ and N-particle propagators $G'_\mu(z) \equiv P^{(\mu)}[z - P^{(\mu)}L'P^{(\mu)}]^{-1}$, where the prime on L' denotes that the interactions $t^{(ab)}$ in the corresponding reduced propagators $g_\mu(\xi)$ defined in (9) is excluded, i.e., $G'_\mu(z)$ corresponds to $g'_\mu(\xi)$ introduced just after (9)). For example, for $\nu = 3$ we have

$$\langle 0|\mathcal{C}^{(3)} = \langle 0|G'_0 \delta L G'_2 \left[1 + \delta L G'_2 + \delta L G'_2 \delta L G'_2 + \delta L G'_2 \delta L G'_3 + \delta L G'_3 \delta L G'_3 \right.$$
$$\left. + \delta L G'_3 \delta L G'_4 + \delta L G'_3 \delta L G'_2 + \delta L G'_0 \delta L G'_2 + \cdots \right] \delta L P^{(3)} \tag{15}$$

and a similar expansion for $\mathcal{D}^{(3)}(z)$. As a result, formula (14) corresponds to a variety of dynamical processes, some of which are represented in Fig. 3.

Note that the same diagram may contribute to different subdynamics. For example, diagram (a) gives the contribution $P^{(0)}\exp(-iZ_j^{(0)}t)P^{(0)}$ to the $\Pi^{(0)}$ subspace, as well as $G'_0 \delta L \exp(-iZ_j^{(2)}t)$ $\delta L G'_0$ to $\Pi^{(2)}$.
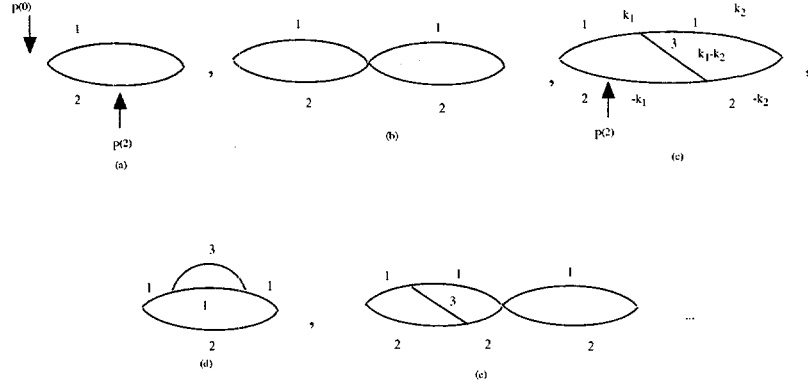
Figure 3: Graphical representation of each term in (14) with the expansion (15). The lines are corresponding to the propagators G'_μ with $\mu \neq 0$ and the vertices to δL. The label of each line corresponds to the index of the velocity, e.g., the line labeled by 1 corresponds to $\mathbf{v}_1$. In diagram (a), $\Pi^{(\nu)}$ indicates the location where the time dependence $\exp(-iZ_j^{(\nu)}t)$ is evaluated. In diagram (c), the wavevectors in the propagators are explicitely written.

Let us sketch the estimation of the long-time tail effect of each contribution in (14) with an example of the $\Pi^{(2)}$ subspace contribution associated to the intermediate wavevector $\mathbf{k}_1$ and $-\mathbf{k}_1$ in diagram (c) (see ref. [12] for detailed calculations). This contribution corresponds to the term which consists of $G'_0 \delta L \exp(-iZ_j^{(2)}t)\delta LG'_3\delta LG'_2\delta LG'_0$. By integration, the N-particle propagators G'_μ reduce to g'_μ; for example, $G'_3(Z_j^{(2)}) \Rightarrow g'_3(\xi_j^{(2)}) = [\xi_j^{(2)} + i\mathbf{k}_2 \cdot \mathbf{v}_1 - i\mathbf{k}_2 \cdot \mathbf{v}_2 + i(\mathbf{k}_1 - \mathbf{k}_2) \cdot \mathbf{v}_3 - nK_1 - nK_2 - nK_3]^{-1}$, and so on. The long-time tails come from the domain of integration over *small* wavevectors $k_i < k_0$ corresponding to the hydrodynamic modes. In this domain the wave-number dependence of the eigenvalues $Z_j^{(2)}$ are given by the same type of k_i dependence as in (13) for the 2d case (and a similar dependence for 3d case, except that there the singular part is replaced by $k_i^{5/2}$). For a large time scale (with the restriction $t/t_r \ll \exp(1/\tau_0)$ for 2d case), we can show that the singular part of the eigenvalues give negligible correction to the diffusive part $k_j^2 D_\alpha$ [12]. Then, substituting the eigenvalue into $\xi_j^{(2)}$, as well as replacing $i(\mathbf{k}_1 - \mathbf{k}_2) \cdot \mathbf{v}_3 - nK_3$ by its hydrodynamical eigenvalues $-|\mathbf{k}_1 - \mathbf{k}_2|^2 D_\beta$, we obtain $g'_3 \sim (ak_1^2 + b\mathbf{k}_1 \cdot \mathbf{k}_2 + ck_2^2)^{-1}$, where a, b and c are constants of the order of D given in (11), and we have estimated only the share-mode and the heat-mode, since they give the slowest decaying process. The propagator g'_3 has a singularity at $k_1 = k_2 = 0$. We have a similar estimation of the wave-number dependence for g'_μ for $\mu \neq 0$. However, we have a different expression for g'_0, since $g'_0 \sim (k_1^2 D_\alpha + \gamma)^{-1} \sim \gamma^{-1}$ for $k_1 < k_0$. Hence, g'_0 has not a singularity at $k_1 = 0$. The reduction of the vertex δL leads to $t^{(ab)}$, and in the hydrodynamic domain, one can neglect the wave-number dependence (see RL (X.100)), then $t^{(ab)} \sim a_0^{d-1} v_0$ as in (11). Combining these order of magnitude estimations, the time dependence of diagram (c) in Fig. 3 is

$$\int_0^{k_0} dk_1 k_1^{d-1} \int d\hat{\mathbf{k}}_1 \int_0^{k_0} dk_2 k_2^{d-1} \int d\hat{\mathbf{k}}_2 \frac{1}{\gamma} e^{-k_1^2 Dt} \frac{1}{ak_1^2 + b\mathbf{k}_1 \cdot \mathbf{k}_2 + ck_2^2} \frac{1}{dk_1^2 + ek_2^2} \frac{1}{\gamma} \sim t^{2-d} \quad (16)$$

where $\hat{\mathbf{k}}_i$ are unit vectors, d and e are also constants of the order of D, and we have changed the dummy variables as $x_i = k_i\sqrt{Dt}$, and replaced the upper limit of the integrals by infinity by neglecting exponentially small terms.

One can easily apply this estimation to an arbitrary term in (14). The result is summarized as follows: We first draw the corresponding graph for each term of (14) with the expansion of

type (15). Then we count the number of lines in each intermediate state. Let us denote the maximum number of lines by η. For fixed η, the slowest decaying process is given by the process in which the number μ of lines first increase (starting with $\mu = 0$) and reaching the maximum value $\mu = \eta$, then decrease reaching $\mu = 0$. For example, diagrams (c) to (e) corresponds to the case $\eta = 3$. Among them diagrams (c) and (d) gives the slowest damping, while (e) corresponds to a more rapid decay. There is therefore a close connection between the topology of the diagrams and the long time contributions.

Denoting the contribution of the slowest process by $\varphi^\eta(\mathbf{v}_1, t)$ for given η, we have obtained

$$\varphi^\eta(\mathbf{v}_1, t) \sim \tau_0^{(d-1)(\eta-1)} (t/t_r)^{(4-d)(\eta-1)/2 - 2} \tag{17}$$

The classical results for mode-mode coupling are of course recovered as $\varphi^2_{d=3} \sim t^{-3/2}$ and $\varphi^2_{d=2} \sim t^{-1}$, but three mode coupling leads to slower process as $\varphi^3_{d=3} \sim t^{-1}$.

The difference between our results and the results for the velocity autocorrelation functions obtained by Dorfman and Cohen [17], as well as by Résibois and Pomeau [18,19], comes from the fact that we avoid the use of any recipe based on the so-called the "most diverging contributions" in a density expansion (see the comment at p271 in RL).[*]

CONCLUDING REMARKS

Let us discuss consequences of our result (17).

1) For $d < 4$, higher order correlations lead to diverging contribution of (17) for large time, so that there is a radius of convergence t_0. For times $t < t_0$, one can use the expansion (15). The value of t_0 can be found by taking the ratio of (17) between η and $\eta + 1$. Then we have $t_0/t_r = \tau_0^{2(d-1)/(d-4)}$, which gives us $t_0/t_r = \tau_0^{-1}$ for $d = 2$ and $t_0/t_r = \tau_0^{-4}$ for $d = 3$. For $d = 2$, this time scale is much less than $t_r \exp(1/\tau_0)$ for moderate dense systems, which is consistent with the approximation we introduced when we have neglected the singular part $k_i^2 \ln k_i$ of the eigenvalue of the Liouvillian in the hydrodynamic modes. For extremely long time scale $t/t_r > \exp(1/\tau_0)$ when we cannot neglect the logarithmic term in the eigenvalue, t exceeds the radius of convergence t_0. Hence, for this time scale the expansion in terms of the ring processes is not applicable to 2d systems.

2) There is a critical dimension at $d = 4$. Higher order correlations lead to slower decay process for $d < 4$, while they lead to quicker decay process for $d > 4$.

3) For $d \to \infty$, the long-time tail effects vanish, and the evolution of the one-particle velocity d.f. $\varphi(\mathbf{v}_1, t)$ is governed by a Markovian kinetic equation in the $\Pi^{(0)}$ subspace.

Let us conclude with some general remarks. The traditional approach of the kinetic theory based on the BBGKY hierarchy relays upon a *truncation* of the hierarchy at a certain order of correlations. Because the higher order correlations become more important for asymptotic times and for $d < 4$, this truncation is incorrect. There are also difficulties due to truncation in the non-hydrodynamic contributions. With the renormalization of the ring processes binary correlations leads n^m contribution in the density expansion, but so lead also higher order correlations up to $(m + 1)$-th order of correlations.

The long-time tail effects described by (17) invalidate the Green–Kubo formalism for $d < 4$. This was well known for $d = 2$, but it is also true for $d = 3$ as there appear contributions of t^{-1}, $t^{-1/2}, \cdots$ coming from multiple mode-mode couplings. Still the linear response formalism remains a valuable tool when used for times where the Markovian approximation to transport

---

[*] We have obtained a different estimation of their propagator $X_k(t)$ from the one postulated in refs. [18-20], i.e., we have obtained $\|X_k(v_1; t)\| \geq t^{1/2} \exp(-bk^2 t)$ for higher mode processes [12], which in contrast with their assumption of the existence of an upper bound of the propagator.

theory is valid [21]. Also it is rigorous for $d \to \infty$. There appears an amusing analogy with the mean field approach in equilibrium statistical mechanics.

In conclusion our approach based on the spectral decomposition of the Liouville operator which avoids carefully non-dynamical assumptions appears to be of special interest when going beyond the limit of dilute gases.

ACKNOWLEDGMENTS

References

1. Y. POMEAU and P. RÉSIBOIS, *Phys. Report*, 19C, 63 (1975).

2. P. RÉSIBOIS and M. de LEENER, *Classical Kinetic Theory of Fluids*, John Wiley and Sons, New York (1977).

3. J. R. DORFMAN and T. KIRKPATRICK, in *Lecture Notes in Physics: Systems Far From Equilibrium –Proceedings*, Sitges, 263, Springer-Verlag, Berlin, Heidelberg, New York (1980).

4. J. A. McLENNAN, *Introduction to Non-equilibrium Statistical Mechanics* Prentice Hall, New Jersey (1989).

5. M. MARESCHAL, *Adv. Chem. Phys.* 100, 31, John Wiley and Sons, Inc. (1997).

6. I. PRIGOGINE, *Non-Equilibrium Statistical Mechanics* , Wiley Interscience, New York, (1962).

7. T. PETROSKY and I. PRIGOGINE, *Chaos, Solitons & Flactals*, 7, 441 (1996).

8. T. PETROSKY and I. PRIGOGINE, *Adv. Chem. Phys.* 99, 1 John Wiley and Sons, Inc. (1997).

9. I. PRIGOGINE, C. GEORGE, and F. HENIN, *Physica*, 45, 418 (1969).

10. T. PETROSKY and H. HASEGAWA, *Physica*, 160A, 351 (1989).

11. M. H. ERNST, J. R. DORFMAN, W. R. HOEGY and J. VAN LEEUWEN, *Physica*, 45, 127 (1969).

12. T. PETROSKY, unpublised.

13. J. VAN LEEUWEN and A. WEYLAND, Physica 36, 457 (1967).

14. Y. KAN and J. R. DORFMAN: *Phys. Rev. A*, 16, 2447 (1977).

15. M. H. ERNST and J. R. DORFMAN, *Physica*, 61, 157 (1972).

16. M. H. ERNST and J. R. DORFMAN, *J. of Stat. Phys.*, 12, 311 (1975).

17. J. R. DORFMAN and E. G. D. COHEN, *Phys. Rev. A*, 6, 776 (1972); 12, 292 (1975).

18. P. RÉSIBOIS, *Physica*, 70, 431 (1973).

19. P. RÉSIBOIS and Y. POMEAU, *Physica*, 72, 493 (1974).

20. M. THEODROSOPULU and P. RÉSIBOIS, *Physica*, 82A, 47 (1976).

21. P. RÉSIBOIS, *J. of Chem. Phys.*, 41, 2979 (1964).

Double Phase Slips and Bound Defect Pairs in Parametrically Driven Waves

Hermann Riecke and Glen D. Granzow
Department of Engineering Sciences and Applied Mathematics
Northwestern University, Evanston, IL 60208, USA

Spatio-temporal chaos in parametrically driven waves is investigated in one and two dimensions using numerical simulations of Ginzburg-Landau equations. A regime is identified in which in one dimension the dynamics are due to double phase slips. In very small systems they are found to arise through a Hopf bifurcation off a mixed mode. In large systems they can lead to a state of localized spatio-temporal chaos, which can be understood within the framework of phase dynamics. In two dimensions the double phase slips are replaced by bound defect pairs. Our simulations indicate the possibility of an unbinding transition of these pairs, which is associated with a transition from ordered to disordered defect chaos.

I. INTRODUCTION

While low-dimensional chaotic dynamics are quite well understood this is not the case for chaotic dynamics of high dimension as, for instance, spatio-temporal chaos. Spatio-temporal chaos arises in systems in which spatial degrees of freedom play an important role and the structures are not only chaotic in time but also in space. These dynamics arise often in pattern-forming systems when all ordered patterns become unstable. Of particular interest are systems in which the chaos is extensive, i.e. systems in which quantities like the attractor dimension and the number of positive Lyapunov exponents grow linearly with the system size, so that one can think of them as a large number of coupled, chaotic entities.

Spatio-temporal chaos is found in many experimental systems. It has been extensively studied in convection where a number of different types have been observed. In the presence of rotation, a classic result is the occurrence of domain chaos. It is due to the Küppers-Lortz instability which renders steady convection rolls unstable to rolls with a different orientation [1]. Since the new rolls are susceptible to the same instability a persistent switching of patches of rolls of different orientations is observerd [2]. Very recently spatio-temporal chaos has also been found (without rotation) in regimes in which the convection rolls are in fact linearly stable. Sufficiently large perturbations, however, lead to a state of spiral-defect chaos in which spirals and other types of defects dominate [3].

Another type of spatio-temporal chaos arises in electroconvection in nematic liquic crystals in a traveling wave regime [4]. Due to the axial anisotropy of this system the waves travel only in certain directions relative to the axis of anisotropy. In the regime in question they travel obliquely to that direction and because of the reflection symmetries of the system the dynamics is governed by the competition of waves traveling in 4 directions. The chaotic dynamics arise immediately at the onset of convection and is characterized by defects in the various wave components. Associated with each defect is a suppresion of the corresponding wave amplitude leading to domains in which one or two of the wave components dominate. Some understanding of these dynamics and their possible origin has been obtained within a set of coupled Ginzburg-Landau equations [5]. Spatio-temporal chaos of traveling waves in an isotropic system arise in convection in binary mixtures [6].

A very rich class of pattern-forming systems are parametrically driven waves, the standard realization being surface waves on a fluid that are excited by a vertical shaking of the container at twice the frequency of the waves. Depending on the fluid parameters and the driving frequency spatially periodic and quasi-periodic patterns of various kinds have been found as well as transitions to spatio-temporal chaos [7]. Strikingly, very similar phenomena can be found even if the fluid is replaced by a granular medium like sand [8].

In the present communication we present theoretical results for the dynamics of parametrically driven waves in one and two dimensions within the framework of Ginzburg-Landau equations. In the one-dimensional analysis we address the important question how to characterize the behavior of a spatially and

temporally chaotic state on length scales that are much larger than the typical wavelength. The response of regular periodic patterns to long-wave perturbations is well understood; in some sense it is a dissipative analogue to sound waves in crystals. For steady patterns the response is typically diffusive whereas for oscillatory patterns it is propagative. Both can be described by phase equations (or equations for the local wavenumber). Here we describe a chaotic state for which the same type of description is possible, i.e. although the dynamics on short scales is chaotic in space and time, the large-scale behavior of that state is diffusive. This striking behavior is due to the fact that the chaotic state is driven by *double phase slips* as described below. We show that the homogeneously chaotic state can become diffusively unstable on large scales and separates into arrays of chaotic and non-chaotic domains very similar to phase separation in mixtures. This provides a mechanism that can lead to the *localization* of the chaotic dynamics in space.

Experimentally, localized spatio-temporal chaos has been observed in Taylor vortex flow [9], Rayleigh-Bénard convection [10], and parametrically excited surface waves [11]. So far, the localization mechanism in these systems is, however, only poorly understood.

In the second part we present ongoing work on the dynamics that arise in two dimensions in the regime in which double-phase slips occur in one dimension. As discussed below it leads to 'fluctuating bound defect pairs'. We present evidence that indicates a transition from an ordered state of defect chaos to a disordered one. This transition appears to be associated with an 'unbinding' of the defect pairs.

II. GINZBURG-LANDAU EQUATIONS FOR PARAMETRICALLY DRIVEN WAVES

To obtain a tractable model for parametrically driven waves we consider systems that exhibit a supercritical Hopf bifurcation to traveling waves, i.e. when the control parameter is increased beyond a certain threshold value the basic state becomes unstable to small-amplitude traveling waves. This is, for instance, the case in electroconvection of nematic liquid crystals [4][1]. Just below the Hopf bifurcation the traveling-wave modes are only weakly damped. Therefore a small driving is sufficient to excite standing waves of small amplitude. Consequently, a weakly nonlinear description is possible by expanding about the unforced basic state and treating the forcing and the damping as small perturbations. This leads to two coupled equations for the complex amplitudes of the traveling-wave components, i.e. physical quantities like the vertical fluid velocity u in the midplane of a convection system are given by

$$u(\mathbf{r}, t) = \epsilon A(X, Y, T) e^{i(\mathbf{q_c} \cdot \mathbf{r} - \frac{\omega_c}{2} t)} + \epsilon B(X, Y, T) e^{i(\mathbf{q_c} \cdot \mathbf{r} + \frac{\omega_c}{2} t)} + c.c. + o(\epsilon). \tag{1}$$

The complex amplitudes A and B vary on the slow time and space scales, $T = \epsilon^2 t$, $X = \epsilon x$, and $Y = \epsilon y$ with $\mathbf{r} = (x, y)$, and $\epsilon \ll 1$.

Using standard symmetry arguments one obtains for the amplitudes A and B in one dimension the Ginzburg-Landau equations [13]

$$\partial_T A + s \partial_X A = d \partial_X^2 A + a A + b B + c A(|A|^2 + |B|^2) + g A |B|^2, \tag{2}$$

$$\partial_T B - s \partial_X B = d^* \partial_X^2 B + a^* B + b A + c^* B(|A|^2 + |B|^2) + g^* B |A|^2. \tag{3}$$

The coefficients in (2,3) are complex except for s and b, which are real. The real part a_r of the coefficient of the linear term a gives the linear damping of the traveling waves in the absence of the periodic forcing and is proportional to the distance from the Hopf bifurcation. The coefficient of the linear coupling term b gives the amplitude of the periodic forcing as can be seen from the fact that it breaks the continuous time-translation

---

[1]Convection in binary mixtures also exhibits a Hopf bifurcation to traveling waves. It is, however, subcritical [6] and the waves appear right away with finite amplitude. An analysis of the parametric forcing of that system is therefore more complicated [12].

symmetry $t \rightarrow t + \Delta t$ that implies the transformation $A \rightarrow A e^{i\Delta t \omega_e/2}$ $B \rightarrow B e^{-i\Delta t \omega_e/2}$. The imaginary part a_i of the coefficient of the linear term a gives the difference between the frequency of the unforced waves and half the forcing frequency ω_e.

The same Ginzburg-Landau equations are obtained also for systems that do not exhibit a Hopf bifurcation, if they have weakly damped traveling wave modes. This is the case for surface waves on fluids with small viscosity. Then again only a small driving is necessary to excite the surface waves and one can perform the same kind of expansions. In that case $a_r < 0$ represents the damping of the waves and the group velocity parameter s is in general complex, indicating that the dissipation depends already linearly on the wavenumber. However, all dissipative terms are small.

In addition to the trivial solution $A = B = 0$ (2,3) possess three types of simple solutions: $|A| = |B| = const$, $|A| \neq |B|$ (both constant), and $|A| = |B|$ (both time-periodic). We are in particular interested in the first type, which corresponds in the physical system to standing waves which are phase-locked to the parametric forcing, i.e. they are excited by the forcing. With increasing a_r they become unstable to solutions of the second type, which correspond to traveling waves as they exist also in the absence of the periodic forcing. Solutions of the third type correspond to standing waves that are not phase-locked to the forcing. For $g_r < 0$ they are unstable to the traveling waves.

The response of the phase-locked standing waves to long-wave perturbations can be described using a phase equation [14],

$$\partial_T \varphi = D(q) \partial_X^2 \phi \qquad \text{with } q = \partial_X \phi. \tag{4}$$

which due to the spatial reflection symmetry of the waves is a (nonlinear) diffusion equation. The diffusion coefficient is not necessarily positive and its sign-change indicates an instability of the waves, the Eckhaus instability. For the one-dimensional case the diffusion coefficient was given in [14]. The resulting stability limits as they are relevant for the first part of the paper are shown in fig.1a. The neutral curve, above which the basic state is unstable to standing-wave perturbations, is given by the dashed line. The solid line gives the Eckhaus instability of the phase-locked standing waves. In the second part of the paper we will consider a case in which traveling waves appear in the absence of forcing, i.e. $a_r > 0$. The corresponding stability regions are shown in fig.1b. The neutral curve is given by the dashed line, neutral curve for the appearance of traveling waves by a dotted line. Their Eckhaus instability is denoted by a solid line. In addition to the Eckhaus instability also a parity-breaking instability arises in which the standing waves become unstable to traveling waves. It is indicated by a dashed-dotted line. Over some range of parameters the parity-breaking instability is preempted by a mode that arises first at finite modulation wavenumber (open squares). It emerges from the parity-breaking instability. The standing waves are stable only inside the region marked by the solid lines and the squares.
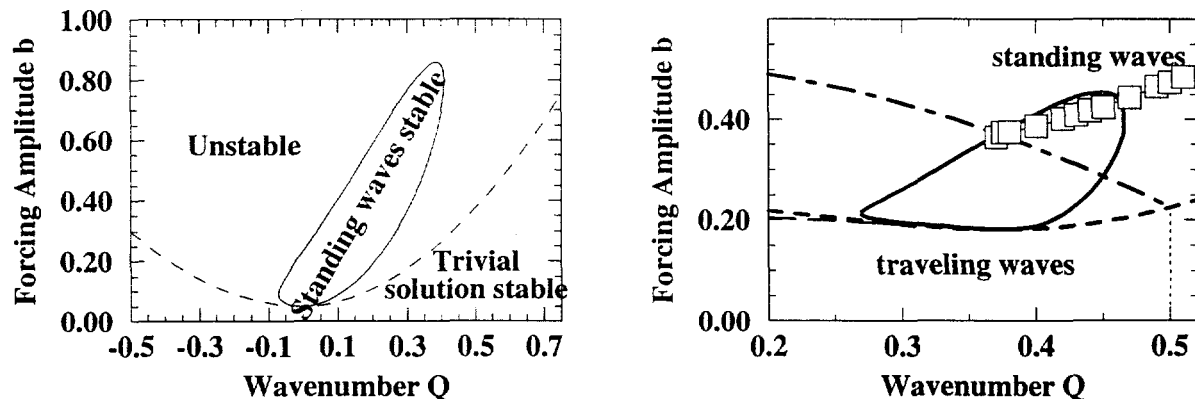

FIG. 1. Linear stability diagram for $c = -1 + 4i$, $d = 1 + 0.5i$, $s = 0.2$, $g = -1 - 12i$. a) $a = -0.05$, b) $a_r = 0.25$.

143

III. DYNAMICS IN A SMALL SYSTEM

The central new feature of the standing waves to be discussed in this paper is the appearance of 'double phase slips' [15,16]. Usually the Eckhaus instability leads to a single phase slip which changes the total phase in the system and through which the wavenumber of the pattern jumps from outside the stable band to inside the band. Such an event is shown in fig.2a. While this occurs also in this system when the Eckhaus instability is crossed for weak forcing, it is not the case for larger forcing: for $b \approx 0.4$ and above the same perturbation leads to a double phase slip, which consists of two consecutive phase slips that undo each other as shown in fig.2b. After the double phase slip the pattern has the same wave number as before and can undergo the same instability again leading to persistent dynamics. It can be periodic or irregular as discussed in sec.IV.

Single phase slip   Double phase slip

Time

$Real((A(X,T)+B(X,T))e^{iqx})$
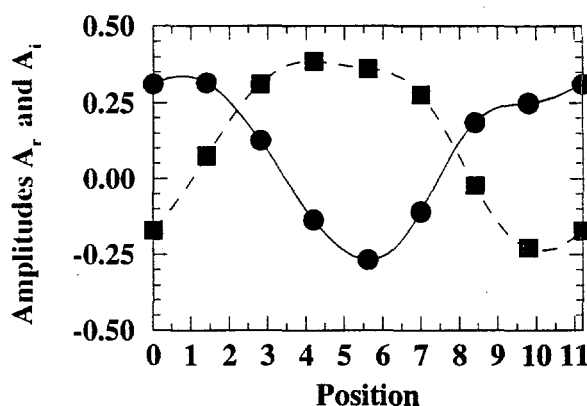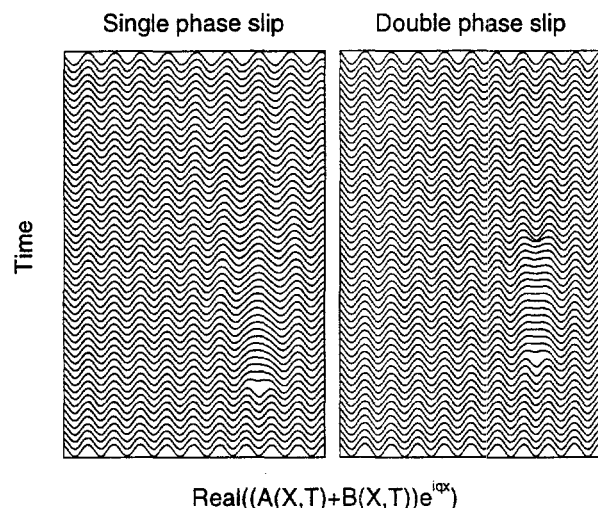
Amplitudes A_r and A_i

Position

FIG. 2. Space-time diagrams showing a) single phase slip for small values of the forcing amplitude (e.g. $b = 0.1$), and b) double phase slip observed for larger values of the forcing amplitude (e.g. $b = 0.6$). Other parameters as in fig.1a.

FIG. 3. Mixed-mode solution for $L = 11.2$

To study the origin of the double phase slip we consider a minimal system in which only the Fourier modes 0 and 1 are important and simulate (2,3) numerically with a pseudospectral code keeping only the Fourier modes -4 to +4. Changing the length of the system at fixed $b = 1.0$ we find the phase portraits shown in fig.4a and fig.4b. In each run the same pattern very close to the solution with one wavelength in the system is chosen as initial condition (solid diamond). For small length (large wavenumber) this initial condition rapidly evolves to the homogeneous solution (short dashes and open triangle).

When the length is increased the homogeneous solution becomes unstable and a new stable fixed point corresponding to a mixed mode involving Fourier modes 0 as well as 1 arises (open circle). Consequently the solution converges to that mixed mode. Fig.3 shows the real and imaginary parts of the amplitude A of the mixed mode for $L = 11.2$. With increasing L the mixed-mode fixed point moves to the left and the trajectory turns around. For $L = 11.2$ the mixed-mode fixed point, which was a stable node for smaller L, becomes a stable spiral point (open square). When L is increased to $L = 11.3$ the spiral point becomes unstable and generates a stable limit cycle as is shown in fig.4b. Closer inspection of that phase portrait shows that the trajectory follows a three-dimensional path: while it starts *outside* the limit cycle it intersects itself a number of times and eventually approaches the limit cycle from *inside*. This indicates that a model that is to capture the mixed mode, its limit cycle, and the two fixed points corresponding to the homogeneous solution and that with one wavelength has to be at least three-dimensional. For yet larger L, a second mixed-mode

fixed point appears, which is associated with the Eckhaus instability of the solution with one wavelength. It becomes important when the limit cycle grows and eventually becomes homoclinic to this additional fixed point around $L = 12$.
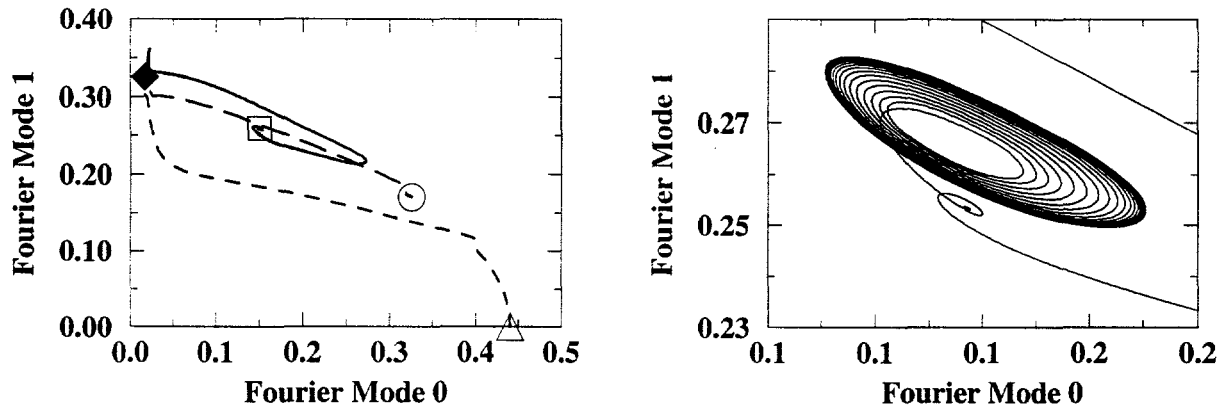

FIG. 4. Phase space projected onto the magnitude of the Fourier modes 0 and 1 for increasing values of the system length L. a) $L = 8$ (short dashes), $L = 10$ (long dashes), and $L = 11.2$ (solid line). Initial condition is marked by a solid diamond. the final fixed point by open symbols. b) $L = 11.3$.

IV. DYNAMICS IN A LARGE SYSTEM

While in the small system discussed in sec.III only simple dynamics was found, complex dynamics arises if the double phase slips can occur at more than one location. Here we describe results for large systems where a surprisingly simple description of the large-scale behavior is possible. Our search for such a description was motivated by states of *localized* spatio-temporal chaos [15,16]. An example is shown in fig.5 where each of the double phase slips is marked as a dot. An initial perturbation triggers a double phase slip. In its vicinity more double phase slips arise and the chaotic activity starts to spread. However, by $t = 50,000$ the width of the chaotic domain stops growing and a stable state is reached, in which the chaotic activity is confined to part of the homogeneous system. At first sight, this result is very surprising; one might have expected that the chaotic activity would always spread through the whole system.
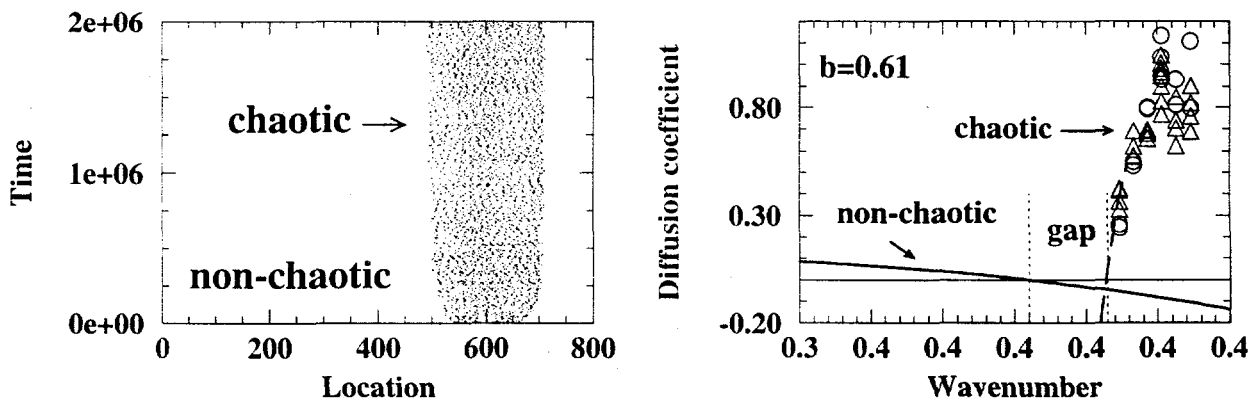

FIG. 5. Localized spatio-temporal chaos for $b = 0.6$ (other parameters as in fig.1a); the averaged effective wavenumber is $< q >= 0.377$).

FIG. 6. Diffusion coefficient of the non-chaotic and the chaotic pattern for $b = 0.6$ (other parameters as in fig.1a).

The mechanism for the localization is due to the fact that double phase slips conserve the total phase, i.e. the wavelength of the pattern is the same before and after the double phase slip. Thus, a time-averaged pattern has a well-defined wavenumber $\hat{q}$ and a well-defined phase $\hat{\phi}$. Since the phase is conserved it is expected to satisfy a slow evolution equation, which on symmetry grounds has the form of a diffusion equation

$$\partial_T \hat{\phi} = \hat{D}(\hat{q}) \partial_X^2 \hat{\phi}. \tag{5}$$

Through a detailed numerical study of the response of the extended chaotic state to a time-periodic perturbation we have explicitly demonstrated the diffusive behavior of the chaotic state on large scales [15,16]. This allowed us also to determine the effective diffusion coefficient $\hat{D}(\hat{q})$ as a function of $\hat{q}$. The result is shown in fig.6. The solid line shows the analytic result for the usual diffusion coefficient $D(q)$ (cf. (4)) which becomes negative at the Eckhaus stability limit. For larger wavenumbers the pattern is unstable to phase slips and undergoes persistent double phase slips. However, a state in which this chaotic activity is homogeneously distributed over the system is not stable to long-wave perturbations as long as the effective diffusion coefficient $\hat{D}$ is negative. Thus, there is a stability gap in wavenumber for which neither the regular nor the chaotic state are diffusively stable (see fig.6). If the initial wavenumber is chosen in that range the pattern has to break up into domains with large wavenumber, which are chaotic, and domains with smaller wavenumber, which are not chaotic. It cannot go into the chaotic (or the non-chaotic) everywhere in space since the total phase, i.e. the integral over the wavenumber, is conserved in the process. Thus, if the wavenumber increases in some part it has to decrease in another part of the system. This separation into domains is very much the same as the phase separation found in equilibrium (fluid) mixtures when they are quenched into the miscibility gap.

V. DYNAMICS IN TWO DIMENSIONS

The wavelength-changing process that occurs in one dimension *via* a phase slip involves in two dimensional patterns the creation (or annihilation) of a pair of defects (dislocations). It is therefore tempting to speculate that a double phase slip will correspond to the creation and annihilation of a 'bound defect pair', i.e. two defects will be created together and will annihilate each other soon thereafter. This would be in contrast to the dynamics observed usually (e.g. in the single complex Ginzburg-Landau equation) where the defects that are created together are not strongly correlated [17,18]. If such a regime of fluctuating bound defect pairs exists one may expect also a transition in which the defects become unbound. In thermodynamic equilibrium such unbinding transitions have found great interest in the context of two-dimensional melting [19] and of vortex unbinding in thin-film superconductors [20]. We are currently investigating the possibility of such a transition in this non-equilibrium system.

We have obtained preliminary results that indicate that such a transition may exist for parameters corresponding to the stability diagram shown in fig.1b. Since no double phase slips seem to occur in the single Ginzburg-Landau equation describing traveling waves we consider a regime in which there exists a transition from parametrically forced standing waves to traveling waves, i.e. a parity-breaking instability. This is the case for $a_r > 0$ as shown in fig.1b. For large b the standing waves are unstable at all wavenumbers as in the case discussed above and double phase slips occur. With decreasing b the parity-breaking instability is approached and we expect that the double phase slips may become replaced by single phase slips in its vicinity.

Fig.7a,b show space-time diagrams for the y-location of defects for simulations for $b = 2$ and $a_r = -0.05$, and $b = 0.5$ and $a_r = 0.25$. For these simulations (2,3) have been extended to two dimensions by replacing the second derivative by a Laplacian. For large b (fig.7a) the defects behave as expected: after their creation they move apart, turn around and annihilate each other again, forming a loop in the space-time diagram. Even in this regime there are loops containing more than one defect pair. Fig.8 shows the statistics of loops

of different sizes (solid symbols): although larger loops occur, their relative frequency decays very rapidly (exponentially) with size. For smaller b the space-time diagram of the defect location is considerably more complicated and offers no simple picture of the dynamics. The corresponding statistics of loop sizes is shown in fig.8 with open symbols. Still most loops contain only one defect pair; this reflects essentially the fact that even if the two defects that are created together moved in a completely uncorrelated fashion they would still be annihilated most likely by their 'partner' since it is closest initially. However, fig.8 shows that large loops are now considerably more frequent. In fact, a power-law decay is more consistent with the data than an exponential decay (cf. fig.8b). It should be mentioned, that the two-dimensional correlation functions of the patterns themselves show also a drastic difference: while for $b = 2$ the pattern is strongly correlated and consists of quite ordered stripes, the correlation function decays rapidly and almost isotropically for $b = 0.5$ [21]. Thus, the two regimes could also be called ordered and disordered defect chaos.
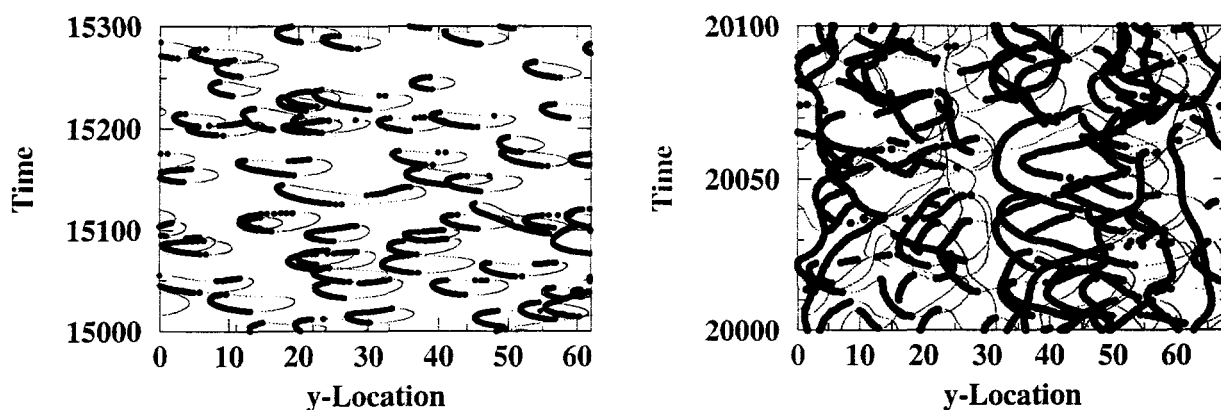

FIG. 7. y-component of the trajectory of defects. a) $b = 2$ (other parameters as in fig.1a); $b = 0.5$ (other parameters as in fig.1b).
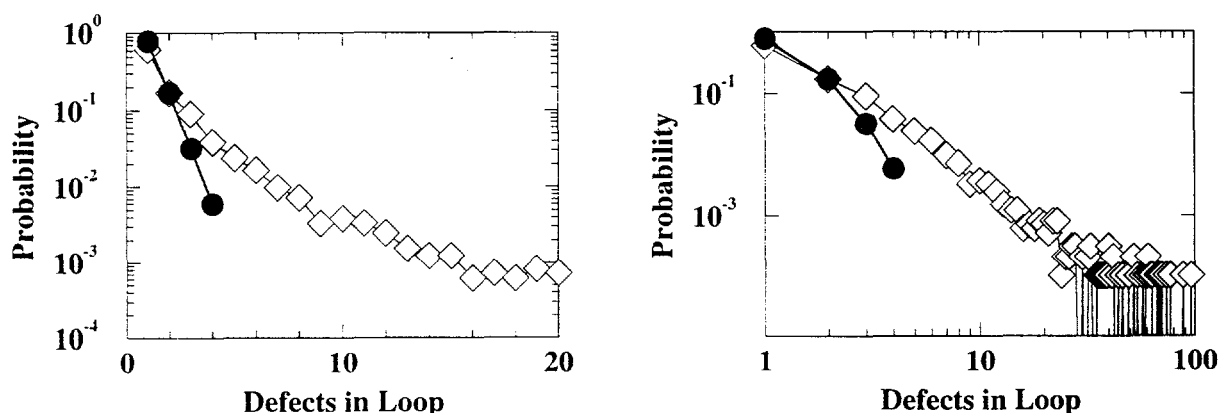

FIG. 8. Relative frequency of defect loops as a function of the number of defect pairs being part of them. open symbols: $b = 2$ (other parameters as in fig.1a); solid symbols: $b = 0.5$ (other parameters as in fig.1b). The double-logarithmic plot indicates that for $b = 0.5$ the distribution is better approximated by a power law than by an exponential.

Current and future work is directed to identify whether the change between the two regimes shown in this paper is smooth or whether it involves a true transition. As diagnostics we are not only using the loop-size distribution, but also the defect life-time, the spatial extent of the loops, the distances travelled by each defect, as well as correlation functions.

---

[1] G. Küppers and D. Lortz. Transition from laminar convection to thermal turbulence in a rotating fluid layer. *J. Fluid Mech.*, 35:609, 1969.

[2] Y. Hu, R.E. Ecke, and G. Ahlers. Time and length scales in rotating Rayleigh-Bénard convection. *Phys. Rev. Lett.*, 74:5040, 1995.

[3] S.W. Morris, E. Bodenschatz, D.S. Cannell, and G. Ahlers. The spatio-temporal structure of spiral-defect chaos. *Physica D*, 97:164, 1996.

[4] M. Dennin, G. Ahlers, and D.S. Cannell. Spatiotemporal chaos in electroconvection. *Science*, 272:388, 1996.

[5] H. Riecke and L. Kramer. On the stability of standing waves and traveling rectangles. *in preparation.*

[6] A. La Porta and C.M. Surko. Phase defects as a measure of disorder in traveling-wave convection. *Phys. Rev. Lett.*, 77:2678, 1996.

[7] A. Kudrolli and J.P. Gollub. Patterns and spatiotemporal chaos in parametrically forced surface waves: a systematic survey at large aspect ratio. *Physica D*, 97:133, 1996.

[8] F. Melo, P.B. Umbanhowar, and H.L. Swinney. Hexagons, kinks, and disorder in oscillated granular layers. *Phys. Rev. Lett.*, 75:3838, 1995.

[9] B.W. Baxter and C.D. Andereck. Formation of dynamical domains in a circular Couette system. *Phys. Rev. Lett.*, 57:3046, 1986.

[10] S. Ciliberto and M.A. Rubio. Local oscillations, traveling waves, and chaos in rayleigh-bénard convection. *Phys. Rev. Lett.*, 58:2652, 1987.

[11] A. Kudrolli and J.P. Gollub. Localized spatiotemporal chaos in surface waves. *Phys. Rev. E*, 54:1052, 1996.

[12] H. Riecke, J.D. Crawford, and E. Knobloch. Temporal modulation of a subcritical bifurcation to travelling waves. In P. Coullet and P. Huerre, editors, *The Geometry of Nonequilibrium*, pages 61–64. Plenum Press, 1991.

[13] H. Riecke, J.D. Crawford, and E. Knobloch. Time-modulated oscillatory convection. *Phys. Rev. Lett.*, 61:1942, 1988.

[14] H. Riecke. Stable wave-number kinks in parametrically excited standing waves. *Europhys. Lett.*, 11:213, 1990.

[15] G.D. Granzow and H. Riecke. Phase diffusion in localized spatio-temporal amplitude chaos. *Phys. Rev. Lett.*, 77:2451, 1996.

[16] G.D. Granzow and H. Riecke. Double phase slips and spatio-temporal chaos in a model for parametrically excited standing waves. *SIAM J. Appl. Math.*, submitted.

[17] L. Gil, J. Lega, and J.L. Meunier. Statistical properties of defect-mediated turbulence. *Phys. Rev. A*, 41:1138–1141, 1990.

[18] B.W. Roberts, E. Bodenschatz, and J.P. Sethna. A bound on the decay of defect–defect correlation functions in two-dimensional complex order parameter equations. *Physica D*, 99:252, 1996.

[19] D.R. Nelson. *Defect-mediated phase transitions*, volume 7 of *Phase Transitions and Critical Phenomena*, chapter 1. Academic, New York, 1983.

[20] P. Minnhagen. The two-dimensional Coulomb gas, vortex unbinding, and superfluid-superconducting films. *Rev. Mod. Phys.*, 59:1001, 1987.

[21] G.D. Granzow and H. Riecke. Ordered and disordered defect chaos. *Physica A*, submitted.

ABSOLUTE INSTABILITY FROM LINEAR CONVERSION OF COUNTER-PROPAGATING POSITIVE AND NEGATIVE ENERGY WAVES

A.N. KAUFMAN, A.J. BRIZARD, J.J. MOREHEAD

Lawrence Berkeley National Laboratory
Berkeley, California 94720

E.R. TRACY

College of William and Mary
Williamsburg, VA 23185

ABSTRACT

The resonant interaction of a negative-energy wave with a positive-energy wave gives rise to a linear instability. Whereas a single crossing of rays in a nonuniform medium leads to a *convectively* saturated instability, we show that a double crossing can yield an *absolute* instability.

A negative-energy wave has the property that the system in the *presence* of the wave has *less* energy than in its *absense*. If a positive-energy wave (a) interacts *resonantly* with a negative-energy wave (b), *both* waves may grow in time, while conserving total wave energy:

$$0 = dW/dt = dW_a/dt + dW_b/dt .$$

For exponential growth at amplitude-rate γ, we have

$$\begin{cases} dW_a/dt = 2\gamma W_a , \\ dW_b/dt = -d|W_b|/dt = -2\gamma|W_b| = 2\gamma W_b ; \end{cases}$$

thus $W_b(t) = -W_a(t)$.

For waves propagating freely in a weakly *nonuniform* medium (variation in x, say), we represent the waves ($j = a, b$) in eikonal form, with $k_x^a \neq k_x^b, k_y^a = k_y^b, k_z^a = k_z^b, \omega^a = \omega^b$:

$$\psi_j(\mathbf{x}, t) = \tilde{\psi}_j \mathrm{exp} i [\int^x k_x^j(x')dx' + k_y y + k_z z - \omega t],$$

valid in the *non-resonant* regions. *Resonance* occurs at regions where $k_x^a \simeq k_x^b$.

In Figure 1, we illustrate a situation that occurs when both waves [represented by their *rays* $k_x^j(x)$] have *caustics* ($dk_x/dx = \infty$). Then there are *two* ray *crossings*, or resonances. At these crossings (where the wave amplitudes $\tilde{\psi}_j$ change rapidly), both waves *increase* their absolute energy as a result of their interaction.

Consider the lower crossing. Let the ratio of transmitted to incident energy flux of wave a be $T > 1$, and of converted flux of b to incident flux be $C < 0$. Then energy conservation states

$$T + C = 1 \,,$$

or

$$|C| = T - 1 \,.$$

In order that amplification of *both* waves occurs, we need $|C| > 1$, or $T > 2$.

The transmission T can be expressed [1] in terms of the coupling η between the two waves, and the Poisson Bracket magnitude B of the two dispersion functions $D_j(x, k_x)$:

$$T = \exp 2\pi |\eta|^2 / B \,,$$

where

$$
\begin{aligned}
B &\equiv |\{D_a, D_b\}| \\
&\equiv |(\partial D_a/\partial x)(\partial D_b/\partial k_x) - (\partial D_a/\partial k_x)(\partial D_b/\partial x)| \,.
\end{aligned}
$$

The condition $T > 2$ thus yields a *threshold* coupling strength η_{th} for the instability:

$$|\eta_{th}|^2 = B(\ln 2)/2\pi \,.$$

To evaluate B, we examine the dispersion functions in more detail. In the neighborhood of a caustic, each ray ($j = a, b$) is a parabolic curve: $x(k_x) = x_j + \beta_j k_x^2$, where $x_j = x(k_x = 0)$ is the caustic location, and $\beta_j \equiv dx/dk_x^2$ is the ray curvature. The two rays thus cross at $k_x = \pm(\Delta x/\Delta\beta)^{1/2} \equiv \pm k_x^c$, where $\Delta x \equiv x_a - x_b$ is the caustic separation, and $\Delta\beta \equiv \beta_b - \beta_a$ is the curvature difference. It follows that the dispersion functions have the form

$$D_j(x, k_x) = (D_x^j)(x - x_j - \beta_j k_x^2) \,,$$

where

$$D_x^j \equiv \partial D_j/\partial x = -(\partial D_j/\partial\omega)(\partial\omega/\partial x)_j = (\partial D_j/\partial\omega)\dot{k}_x^j < 0$$

is proportional to the (constant) ray velocity in the k_x direction. We obtain (after straightforward algebra)

$$|B| = dD_x^a D_x^b(\Delta\beta)k_x^c \equiv 2D_x^a D_x^b(\Delta\beta\Delta x)^{1/2} \,. \tag{1}$$

Thus, for given parameters, the threshold coupling is minimized by *minimizing* the caustic separation Δx.

To determine the latter, we impose the requirement of phase matching: the phase change $\Delta\phi$ of ray a (say), after one circuit, must be an integer multiple N of 2π: $\Delta\phi = 2\pi N$. To find $\Delta\phi$, we use eikonal theory (to lowest order in η): $\Delta\phi = -\oint x(k_x)dk_x - \pi/2$, where the first term is the standard phase integral (the area enclosed by the rays), and the second term is the sum of the two (lowest-order) phase shifts at the conversion points [2]. The evaluation of the phase integral is elementary: $\oint x(k_x)dk_x = (4/3)(\Delta\beta)^{-1/2}(\Delta x)^{3/2}$. The minimum separation $\Delta x_{\min}$ is thus obtained by setting $N = 0$:

$$\Delta x_{\min} = (3\pi/8)^{2/3}(\Delta\beta)^{1/3} \,. \tag{2}$$

Inserting Eq.(2) into Eq.(1), we have $|B|_{\min} = (3\pi)^{1/3} D_x^a D_x^b (\Delta\beta)^{2/3}$, yielding the threshold

$$|\eta_{\mathrm{th}}|^2 = [(\ln 2)/2\pi](3\pi)^{1/3} D_x^a D_x^b (\Delta\beta)^{2/3} . \tag{3}$$

To obtain the positive growth rate γ (when $T > 2$), we track the energy around one circuit, obtaining $\gamma = [\ln(T-1)]/\tau$, where τ is the time interval for a circuit:

$$\tau = \sum_{j=a,b} \int dk_x / |\dot{k}_x^j|$$

$$= (|\dot{k}_x^a|^{-1} + |\dot{k}_x^b|^{-1})(3\pi)^{1/3}(\Delta\beta)^{1/3} ,$$

and Eq.(2) has been used.

In conclusion, the general double-crossing mode-conversion problem supports a discrete set of modes, labelled by N above. As the coupling between the modes increases, more of these modes are (absolutely) unstable. For couplings below the threshold value, Eq.(3), all modes are stable. This double-crossing instability can occur in physically different types of wave phenomena.

Acknowledgements

References

1. A.J. Brizard, J.J. Morehead, A.N. Kaufman, and E.R. Tracy, "Double-Cross Instability: An Absolute Instability Caused by Counter-Propagating Positive- and Negative-Energy Waves ," *Phys. Rev. Lett.* **77**, 1500 (1996).
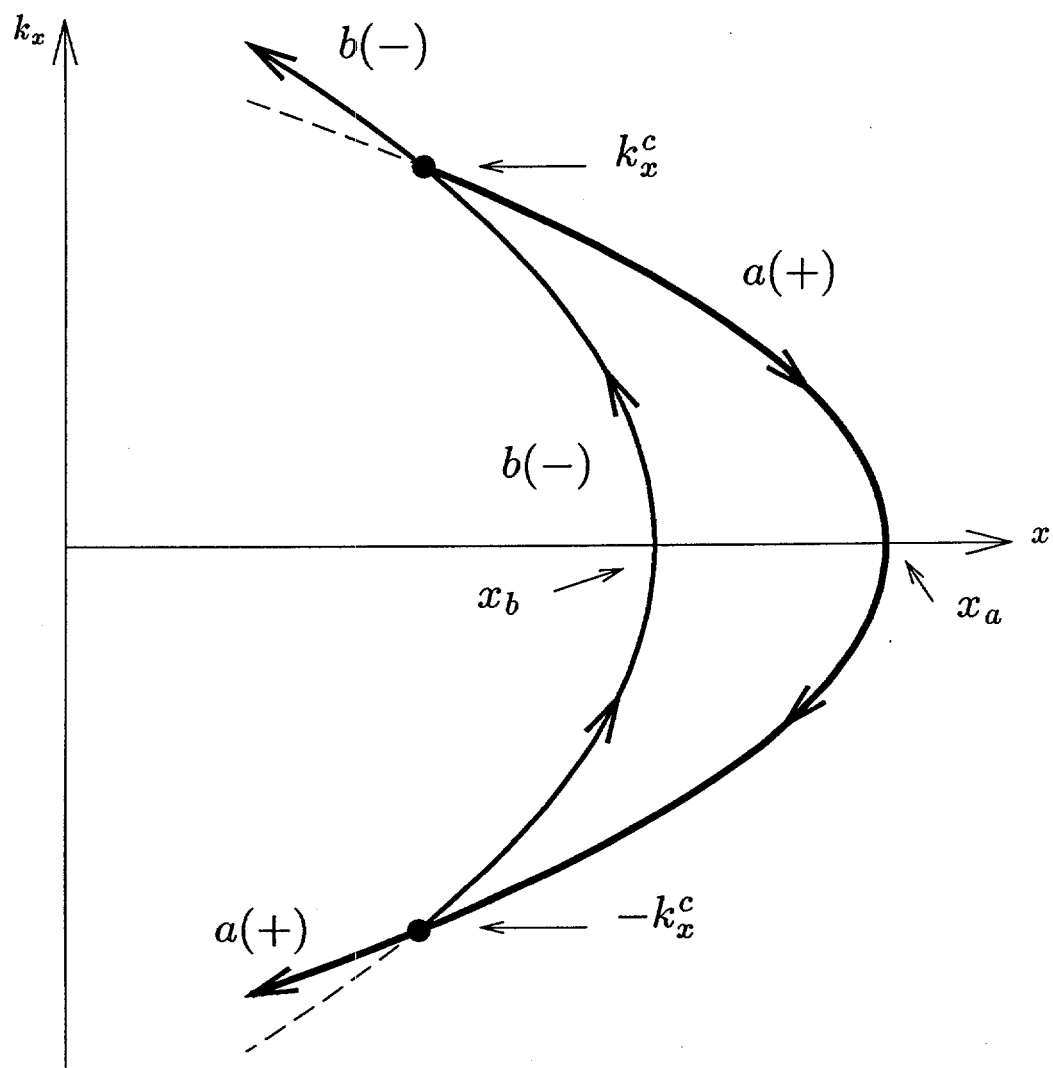2. J.J. Morehead, "Phase Shifts of Mode Conversion," to be published.

Figure 1. Orbits of the positive-energy wave (a) and negative-energy wave (b)

ENZYME STRUCTURE AND ACTIVITY AT LIQUID-LIQUID INTERFACES

C. J. Beverung, M. J. Tupy, C. J. Radke and H. W. Blanch

Department of Chemical Engineering
University of California, Berkeley, CA 94720

ABSTRACT

Understanding the behavior of proteins interaction at oil/water interfaces is crucial to the design of two-phase bioprocesses (aqueous/organic). An examination of the mechanism of protein adsorption at the oil/water interface was undertaken using tensiometry, transmission electron microscopy (TEM) and a novel total internal reflection fluorescence spectrometer (TIRFS), constructed to monitor adsorption dynamics. Dynamic interfacial tension measurements of protein adsorption show three regimes which can be described by diffusion to the interface, adsorption and denaturation of the adsorbed protein. TEM micrographs show a network of proteins in the adsorbed layer at long times. TIRFS data show that this network formation or protein entanglement in the adsorbed state requires a long period of time to occur. A series of two-monomer random polyamino acids used as model proteins demonstrate many of the adsorption characteristics observed for natural proteins.

INTRODUCTION

Biological approaches for the treatment of hazardous or mixed wastes and for the processing of hydrocarbon mixtures have demonstrated utility [1,2]. Many of these applications require interaction of the reacting phase with a non-aqueous liquid phase. However, many biological catalysts requires an aqueous milieu for activity. In order to exploit

the advantages of these biological mechanisms an efficient means to contact the reacting aqueous medium with the non-aqueous feed is required. Two-phase liquid-emulsion bioreaction systems have been considered for these applications. Although two-phase bioreactors have been developed for various processes, the lack of fundamental understanding presents challenges for design and scaleup.

One aspect of understanding emulsion bioreactor performance is characterizing the oil/water (O/W) interface. The nature of the O/W interface determines emulsion stability, interfacial mass transport and enzyme reactivity. Adsorption of proteins from the aqueous phase at the O/W interface affects all these properties. We have employed interfacial tensiometry, transmission electron microscopy (TEM) and total internal reflection fluorescence spectroscopy (TIRFS) to understand the time scales of adsorption and the affect of the O/W interface on protein conformation.

Understanding protein adsorption is made complex by the nature of the proteins. To facilitate understanding of the adsorption process we have employed poly-amino acid copolymers as model proteins. The dynamic interfacial tension of a series of glutamic acid copolymers with varying hydrophobicities of the second amino acid has been studied.

PROTEIN ADSORPTION

Pendant Drop Tensiometry.

We have applied pendant drop tensiometry to study the dynamics of protein adsorption at a heptane/water interface. This technique involves fitting the Young-Laplace equation, using the interfacial tension as the fitting parameter, to the coordinates of the drop edge as measured by image analysis [3,4]. The pendant drop technique is well-suited to study dynamics because the system is not perturbed by deformation of the interface in order to perform the measurement. This technique is more convenient to apply to O/W systems than DuNouy ring or Wilhelmy plate techniques.

Our recent work has examined the dynamics of adsorption of ovalbumin, β-casein, lysozyme and bovine serum albumin (BSA) at the heptane/buffer interface. Figure 1 shows the concentration dependence of ovalbumin tension at pH 7.1. Low bulk concentrations show a lag period where diffusional effects are significant. As concentration is increased, this lag time is eliminated, and only the long-term effects are seen. At longer times the interface becomes saturated with protein molecules and a steep decrease in the tension is observed. At very long times, on the order of 24 hours, a continuous, steady decrease in interfacial tension is observed. Continuous reduction in tension may be attributed to slow conformational change and aggregation of this globular protein at the O/W interface [5].

Figure 2 shows the dynamic interfacial tension of four proteins studied at a single, low concentration. The same dynamic characteristics discussed above for ovalbumin at low concentrations are observed for lysozyme and BSA. The adsorption of the random coil protein, β-casein, however shows markedly different dynamics. Low concentrations promote short-time diffusional effects. However, at extended periods β-casein approaches a constant
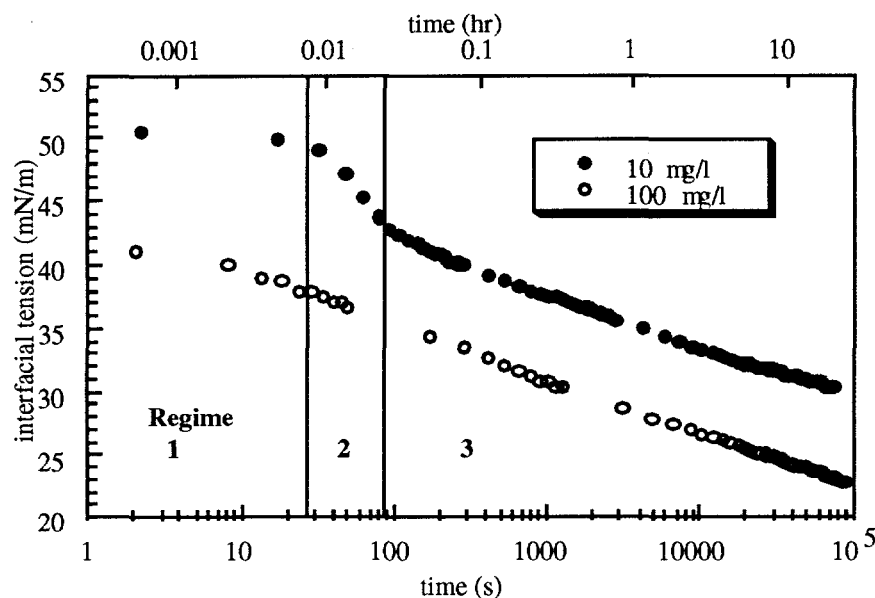
Figure 1. Dynamic Interfacial Tension of Ovalbumin. Tensions of the heptane/water interface were measured at room temperature. The water is a 0.1M sodium phosphate buffer solution at pH 7.1.
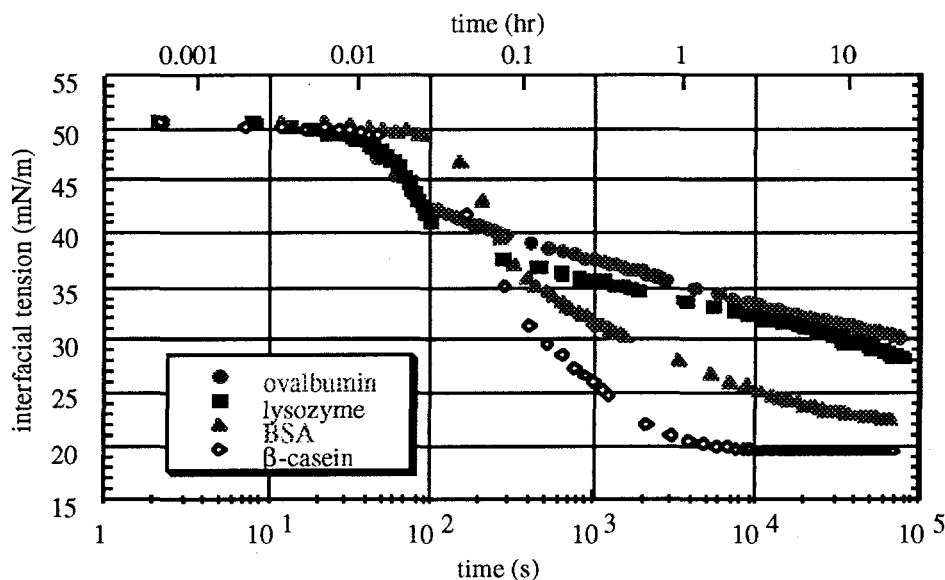

Figure 2. Dynamic Interfacial Tension of Protein Solutions. Tensions of the heptane/water interface were measured at room temperature. The water is a 0.1M sodium phosphate buffer solution at pH 7.1. Protein concentration was 10mg/l for all of the proteins.

interfacial tension of 19 mN/m at a concentration of 10 mg/l. This apparent equilibrium suggests the random-coil nature of this protein enables a more rapid approach to a final adsorption state.

Qualitative information on the adsorbed protein layer has been obtained from a perturbation of the pendant drop after the protein has interacted with the O/W interface for long times (> 18 hours). Withdrawal of the aqueous solution from the drop interior results in

compression of the adsorbed protein layer. Compression of the protein layer leads to collapse of the drop shape and a wrinkling of the gel-like protein film. This phenomenon is macroscopically visible, and suggests a significant interaction between protein molecules present at the interface in order to form an agglomerated network or an entangled polymeric matrix. The strength of this protein film is sufficient to permit it to be harvested in a manner similar to Langmuir-Blodgett techniques.Transmission Electron Microscopy.

Compression of protein layers adsorbed at the O/W interface leads to collapse of the drop shape and a wrinkling of the gel-like protein film. This phenomenon is macroscopically visible, and suggests a significant quantity of interacting protein molecules are present at the interface. TEM of protein layers collected from a planar interface was performed to validate this observation and obtain general structural information of gel layer

For TEM experiments, a glass slide with copper electron microscopy grids is covered with a microporous polymeric formvar film. This film provides structural support for the protein layer, yet allows direct observation through the 0.5 micron holes. The microporous formvar support can be seen in Figure 3 through the protein film and on the right edge of the photograph. The slide was immersed in a protein solution, which was then covered with a layer of heptane. The protein was allowed to adsorb to the O/W interface overnight.

Figure 3 shows a layer of ovalbumin collected in this manner from the heptane/buffer interface. The presence of a protein skin is seen, with significant protein-protein interactions observable in the form of large spherical aggregates. Higher magnification shows an apparent network formation on two length scales. One consists of a homogeneous layer, with strands on the order of the size of individual protein molecules (50 Å). The second is comprised of much larger protein aggregates, with strand thicknesses on the order of hundreds of angstroms and much larger pores. The latter case of aggregations appears to give the protein film its strong mechanical properties and gel-like elastic attributes. Similar results were observed for a film of β-casein adsorbed at the heptane/buffer interface.

The structures seen in the TEM experiments are similar to those seen upon heat-denatured of proteins. The oil phase is an effective medium to induce the order/disorder transition necessary for the denaturation and aggregation of proteins molecules. This effect appears to propagate over length scales beyond single monolayer coverage.

Total Internal Reflection Fluorescence Spectroscopy.

We have employed total internal reflection spectroscopy to investigate the protein conformational changes that were observed from interfacial tension data and TEM images. The interfacial tension at constant temperature is a function of adsorption and the chemical potential of the adsorbed species. Use of extrinsic fluorescent probes and TIRF enable us to follow the adsorption of labeled proteins. By combining the adsorption information and interfacial tension data we gain insight into the energy state of the adsorbed species.

Although more commonly used to investigate the solid/water interface systems [6,7], we have designed and constructed a novel apparatus to measure the fluorescence emitted by adsorbed species at the O/W interface by means of total internal reflection fluorescence. Using light focused at the interface, the evanescent wave produced at the point of total internal
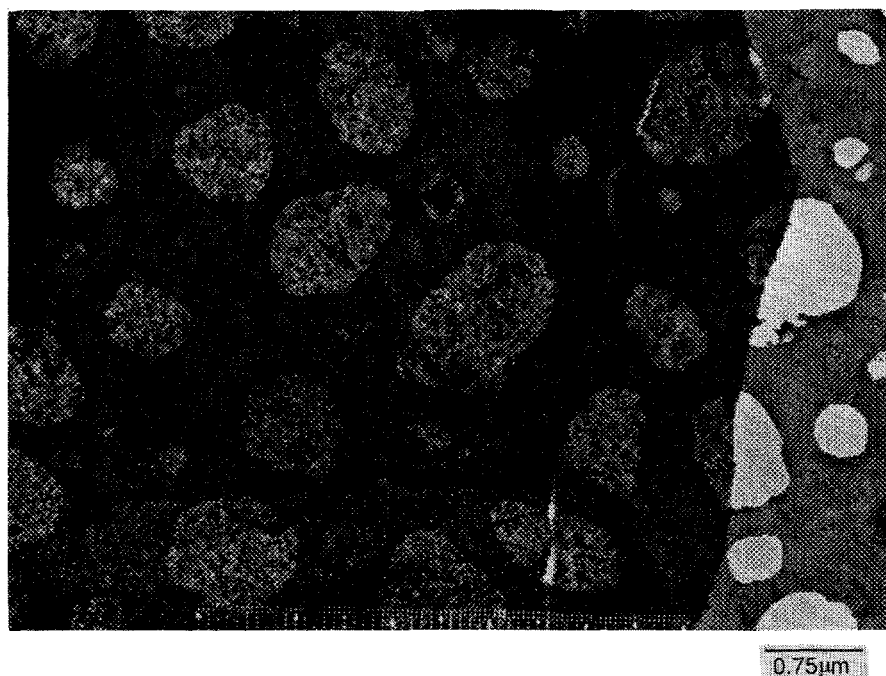
Figure 3. TEM Micrograph of Adsorbed Ovalbumin Film. Film collected from the heptane/buffer interface after 18 hours. The buffer is a pH 7.1, 0.1M sodium phosphate buffer with 100mg/l ovalbumin.

reflection excites fluorophores present in the adsorbed layer. Therefore the fluorescence generated and detected is dominated by those species adsorbed at the interface.

The unique design of our apparatus enables us to follow the dynamic TIRF at O/W interfaces. The centerpiece of the apparatus is the TIRFS cell shown schematically in Figure 4. The cell is designed to ensure diffusive transport of a fluorescent, surface active species from an aqueous bulk solution to the O/W interface. Diffusion of surfactant to the interface occurs in a thin stagnant water layer that is separated by a porous polymeric membrane from a flow channel. The flow channel allows us to introduce the solutions of interest to the TIRFS. By changing the solutions in the flow channel, this design allows the study of the reversibility of adsorption. Within the cell the polymeric membrane is sandwiched by two thin stainless steel sieve plates for support. The oil phase above the interface is sealed in the cell by a Viton O-
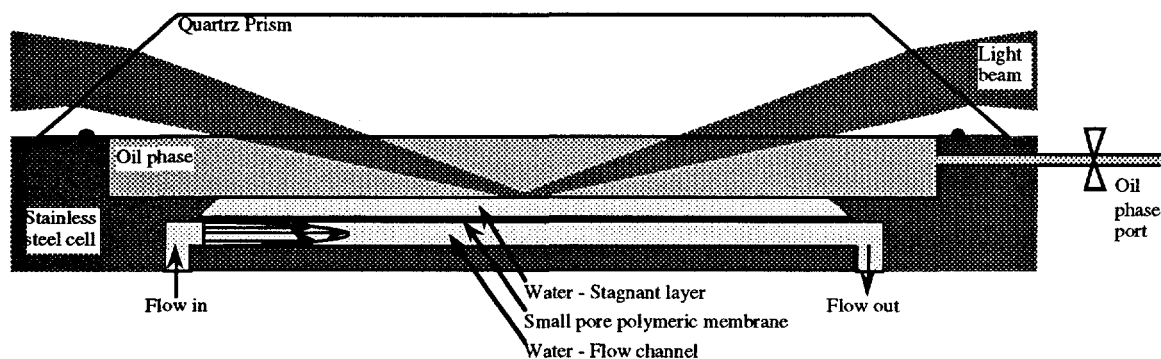

Figure 4. TIRFS Cell Schematic.

ring seal between the cell and a dove quartz prism. Because the oil phase is incompressible and sealed, we are able to eliminate any convective flow through the polymeric membrane thus ensuring pure diffusive transport. To guarantee that the optics stay focused on the interface throughtout the entire experiment the interface is kept flat. This is accomplished by pinning O/W interface at a knife-edge ring and then adjusting the oil phase volume. A flat interface has zero curvature and therefore its shape is not a function of the changing interfacial tension. Once the oil phase is sealed the location of the interface is fixed for the course of the experiment. With this technique, we follow the dynamics of transport of protein to a clean interface and subsequent adsorption from the bulk solution. Use of a monochromator ables the fluorescent spectrum of the adsorbed species to be measured and insight into the chemical environment of the fluorescent probe can be obtained. Studies with a non-adsorbing fluorophore have confirmed the transport mechanism in the TIRFS cell.

Figure 5a shows the loading and washout TIRF for the adsorption of a β-casein/fluorescein conjugate. β-casein dynamic interfacial tension approaches an apparent equilibrium value at long times as discussed earlier. This behavior is unique among the proteins studied and was examined further using the TIRFS to test the adsorption reversibility. Results obtained demonstrate the protein arrival at the interface is consistent with a simple diffusion model. The loading was monitored for a dimensionless time, τ, 1.5 as shown in the figure where τ is defined as:

$$\tau = t \cdot D / h^2 \quad \text{where;}$$
$$t = \text{time}$$
$$D = \text{diffusion coefficient}$$
$$h = \text{thickness of the stagnant layer.}$$

A washout experiment testing reversibility shows that a significant fraction of the adsorbed protein does not desorb over a period of several hours. This result suggests that β-casein is

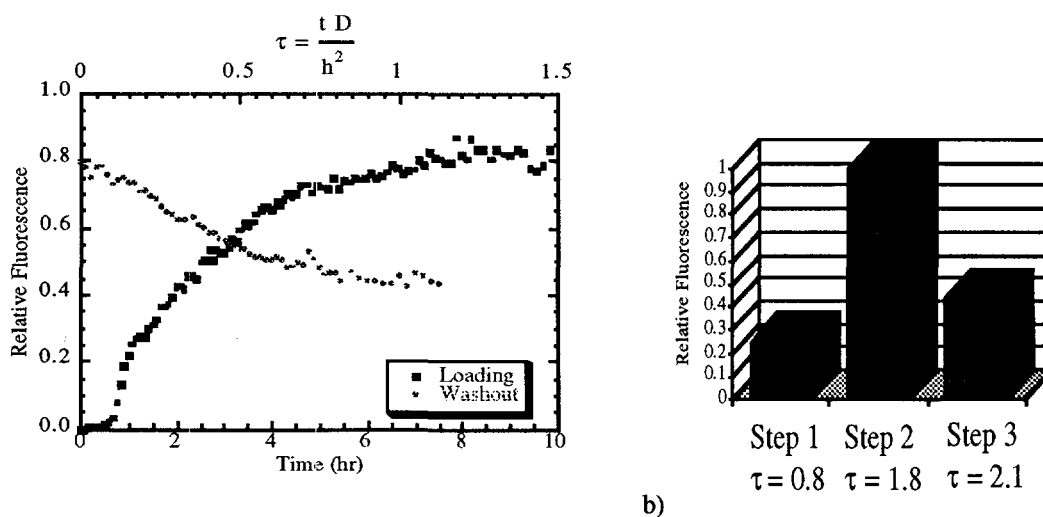a)   b)

Figure 5. Dynamic TIRFS of β-Casein. Fluorescence excitation at 490nm and emmission measured at 520nm. Experiments were carried out at room temperature. Aqueous solution is pH 7.0, 0.1M sodium phosphate buffer solution with 2mM sodium azide. See text for description of the experiments.

irreversibly adsorbed or at least shows very slow desorption, contrary to the equilibrium picture infers from the interfacial tension data.

A second experiment was designed to investigate β-casein adsorption and possible exchange with the bulk solution. β-casein/fluorescein conjugate labeled with a fluorescein to protein ratio of 1:20 was loaded into the cell for dimensionless time (τ) of 0.8. At this time β-casein/fluorescein conjugate labeled with a fluorescein to protein ratio of 1:4.2 was loaded into the cell for dimensionless time of 1.8. Then washout of the protein with pure buffer solution was carried out for dimensionless time of 2.1. The relative fluorescence at the end of each step is shown in Figure 5b. From this data it is obvious some of the adsorbed non-labeled β-casein was exchanged with labeled β-casein from the second loading step. This observation appears to be contrary to the conclusions reached from the experimental data shown in Figure 5a.

The Protein Adsorption Process.

There is insufficient data at present to make definitive conclusions based on the TIRFS data, but our description of protein adsorption follows. The long time interfacial tension decay shown in Figures 1 and 2 is due to denaturation and entanglement of proteins in the adsorbed state and possible continued incorporation of more proteins from the bulk into the adsorbed layer. The decrease in free energy of the adsorbed layer observed in the tension decrease is due to the entropic gain from greater mobility of the protein polymer chain in the denatured state. β-casein is a disordered protein in bulk solution and apparently gains little additional entropic freedom in the adsorbed state but at long times may become increasingly entangled with other β-casein molecules at the O/W interface. During the TIRFS experiment shown in Figure 5a, the exposure of β-casein to the O/W interface was sufficient to cause significant entanglement. For the experiment shown in Figure 5b, the first loading step with dimensionless time 0.8 was insufficient for significant entanglement between adsorbed protein. The second loading step was of long enough duration to allow for entanglement of β-casein.

ADSORPTION OF POLY-AMINO ACIDS

The goal of this work is to probe the effects of amino acid side chain hydrophobicity, surface charge and bulk conformation on protein adsorption. Synthetic co-polymers of glutamic acid are used as model proteins. However, they are comprised of only two monomers, glutamic acid and a test amino acid. This simplifies the isolation of fundamental driving forces. These systems provide one of the few methods to experimentally attribute adsorption phenomena to particular side chains.

A systematic approach was used here to observe the dynamics of poly-amino acid copolymer adsorption at the heptane/water interface. A series of glutamic acid copolymers with varying hydrophobicities of the second amino acid were studied, with the glutamic acid homopolymer as a control. Dynamic interfacial tension was measured using the pendant drop method. Significant pH effects were observed and correlated with bulk conformational changes observed with circular dichroism (CD).
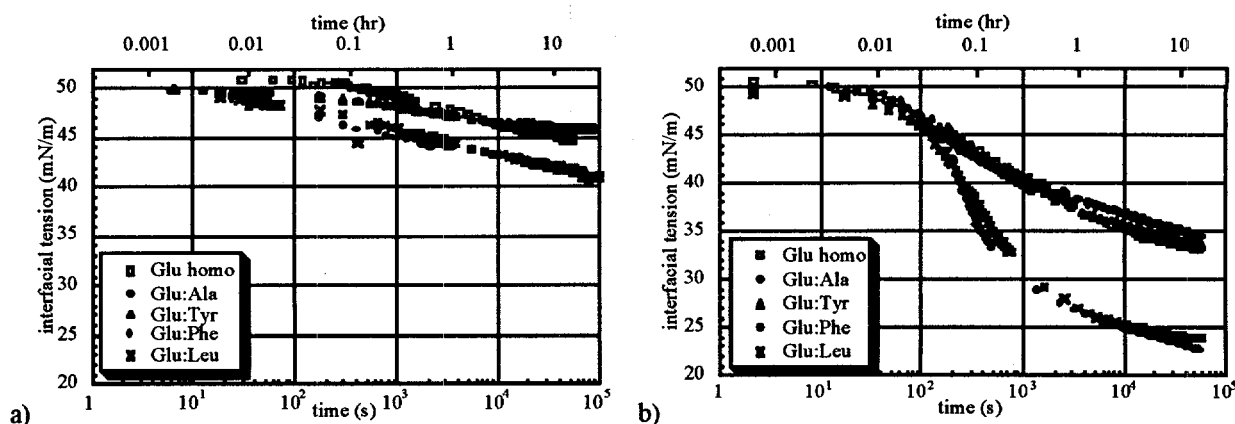
159

Figure 6. Dynamic Interfacial Tension of Poly-Glutamate Random Co-Polymers. Tension of heptane/water interface at room temperature. Water is a 0.1M sodium phosphate buffer solution with 15mg/l polymer concentration. a) pH 7.1 b) pH 5.3.

The series of copolymers used in this work were glutamic acid with alanine, leucine, phenylalanine and tyrosine as co-monomer. This system was chosen because of the solubility engendered by the charged glutamate side chain as well as the commercial availability of several copolymers of similar mole ratios and degrees of polymerization. Chemical characteristics of these residues encompass a representative range of chain length, aliphaticity and hydrophobicity of the amino acids.

A copolymer concentration of 15 mg/l was used ensure that diffusion effects were observable. Copolymers were examined at pH 7.1 and pH 5.3 to observe charge effects on adsorption. The pKa of glutamic acid is approximately 4.5 for the individual amino acid and 4.7 in proteins. All synthetic poly-amino acids were found to be insoluble at pH between 4.0 and 5.0.

To summarize the dynamic interfacial tension results, Figure 6 shows the combined data for the glutamic acid copolymer systems at pH 7.1 and 5.3. General trends at pH 7.1 indicate minimal adsorption with maximum surface pressures of only 9 mN/m attained for any of the polypeptides. Glutamate residues are primarily charged at this pH, resulting in peptide surface charges greater than -100. The high charge of the polymer is shown to result in a random coil like conformation in the bulk solution by CD spectroscopy. The charged poly-amino acid avoids the non-polar oil phase. A trend attributed to side chain structure is apparent, however. Similar final pressures of 4-5 mN/m are seen for the homopolymer, Glu:Ala and Glu:Tyr, while Glu:Leu and Glu:Phe are able to reach surface pressures of 8-9 mN/m. Even at the high pH, effects of hydrophobicity and structure of the second amino acid are manifested in this small difference in pressure. The polypeptide is able to arrange itself in such a manner that the leucine and phenylalanine groups encounter the heptane phase. The formation of a gel-like adsorbed layer observed for natural proteins was not observed for the poly-amino acids at pH 7.1.

At pH 5.3, the poly-amino acids are closer to the isoelectric point of glutamic acid and hence are less charged. The change in charge lead to α-helix conformation of the poly-amino acids in the bulk solution as measured by CD. The effect on adsorption to the heptane interface is apparent as interfacial tension is reduced much further for all peptides as compared to pH

160

7.1. The two trends seen at pH 7.1 are enhanced at the lower pH. Glutamate homopolymer, Glu:Ala and Glu:Tyr exhibit similar tension reductions and final surface pressures of ~16-19 mN/m. Glu:Leu and Glu:Phe show significantly different behaviors with tension curves similar to those for globular proteins. Ultimate tension reduction for these systems are ~27 mN/m. The lower electrostatic barrier at pH 5.3 allows the effects of second side chain structure on adsorption to be easily seen. Long chain, hydrophobic residues strongly adsorb (Leu, Phe), provided the chain structure is not terminated with a non-oily moiety (Tyr). Short chains (Ala) and charge or poly (Glu homo) residues also have comparatively low affinities for the oil interface.

CONCLUSIONS

We have applied pendent drop tensiometry, TEM and TIRFS to study protein adsorption. A number of proteins have been examined as well as a series of glutamic acid random co-polymers as model proteins. The dynamic interfacial tension shows three regimes. At early times the proteins show a lag phase. At later times, the proteins diffuse to the interface and adsorb. Once the proteinsadsorb the begin to denature and undergo a very long time process of agglomeration or entanglement. Dynamic TIRFS demonstrates that strong association between adsorbed proteins occurs over a long time periods. At long times, adsorbed proteins form a gel-like skin at the interface that show structuring in TEM micrographs. The poly-amino acid co-polymers demonstate much of the dynamic interfacial tension properties observed for natural proteins.

ACKNOWLEDGEMENT

REFERENCES

1. J. TRAMPER, H. C. VAN DER PLAS AN DP. LINKO, eds. *Biocatalysts in Organic Synthesis* (Proceedings of an International Symposium on this Subject), Elsevier, Amsterdam, 1985.

2. C. LAANE, J. TRAMPER AND M. D. LILLY, eds. *Biocatalysts in Organic Media* (Proceedings of an International Symposium on this Subject), Elsevier, Amsterdam, 1987.

3. J. M. ANDREAS, E. A. HAUSER AND W. B. TUCKER, "Boundary Tension by Pendant Drop," *J. Phys. Chem.* 42, 1001 (1938).

4. P. CHENG, D. L. L. BORUVKA, Y. ROTENBERG AND A. W. NEUMANN, "Automation of Axisymmetric Drop Shape Analysis for Measurements of Interfacial Tensions and Contact Angles," *Colloids Surf.* 43, 151 (1990).

5. D. E. GRAHAM AND M. C. PHILLIPS, "Proteins at Liquid Interfaces. I. Kinetics fo Adsorption and Surface Denaturation," *J Colloid Interface Sci.* 70, 415 (1979).

6. T. P. BURGHARDT AND D. AXELROD, "Total Internal Reflection/Fluorescence Photobleaching Recovery Study of Serum Albumin Adsorption Dynamics," *Biophys. J.* 33, 455 (1981).

7. R. D. TILTON, C. R. ROBERTSON AND A. P. GAST, "Lateral Diffusion of Bovine Serum Albumin Adsorbed at the Solid-Liquid Interface," *J. Colloid Interface Sci.* 137, 192 (1990).

APPLICATION OF WELDING SCIENCE TO WELDING ENGINEERING: A LUMPED PARAMETER GAS METAL ARC WELDING DYNAMIC PROCESS MODEL

P. E. Murray, H. B. Smartt, J. A. Johnson

Idaho National Engineering and Environmental Laboratory
Lockheed Martin Idaho Technologies
Idaho Falls, Idaho 83415-2210

ABSTRACT

We develop a model of the depth of penetration of the weld pool in gas metal arc welding (GMAW) which demonstrates interaction between the arc, filler wire and weld pool. This model is motivated by the observations of Essers and Walter [1] which suggest a relationship between droplet momentum and penetration depth. A model of gas metal arc welding [2] was augmented to include an improved model of mass transfer and a simple model of accelerating droplets in a plasma jet to obtain the mass and momentum of impinging droplets. The force of the droplets and depth of penetration is correlated by a dimensionless linear relation used to predict weld pool depth for a range of values of arc power and contact tip to workpiece distance. Model accuracy is examined by comparing theoretical predictions and experimental measurements of the pool depth obtained from bead on plate welds of carbon steel in an argon rich shielding gas. Moreover, theoretical predictions of pool depth are compared to the results obtained from the heat conduction model due to Christensen et al. [3] which suggest that in some cases the momentum of impinging droplets is a better indicator of the depth of the weld pool and the presence of a deep, narrow penetration.

INTRODUCTION

Gas metal arc welding uses a consumable metal electrode which melts in the presence of an electric arc. Droplets are expelled from the electrode and transferred to the workpiece thus producing the weld. Energy transfer from the arc to the workpiece is augmented by the energy of impinging droplets. Transfer of droplets affects many aspects of the welding process including the size, shape and depth of penetration of the weld pool. Penetration is caused by droplets with sufficient momentum to carry energy deep into the pool thus enhancing convective mixing [1]. In constrast, penetration in nonconsumable electrode processes such as gas tungsten arc welding is primarily affected by flow in the weld pool caused by surface tension and buoyancy forces [4].

Theoretical models have been used to predict the size and shape of the cross sectional area of fused metal. Notable examples include the model by Christensen et al. [3] based on Rosenthal's [5] model of a moving point source of energy, and the extension to a moving distributed energy source by Eagar and Tsai [6]. These models assume steady state energy transfer to the workpiece by conduction; the effects of convective heat transfer in the weld pool and the deposition of filler metal are not included. To overcome these limitations, numerical models of heat transfer have been recently developed. These include heat transfer models due to Tekriwal and Mazumder [7] and Pardo and Weckman [8], and fluid flow models due to Tsao and Wu [9] and Kim and Na [10]. The affect of filler metal deposition in these recent models was included

by fixing the mass, velocity and rate of impinging droplets rather than direct simulation of droplet dynamics.

Experiments reported by Essers and Walter [1] established that the momentum of impinging droplets effects the depth of the weld pool. Therefore accurate prediction of the depth of penetration requires accurate prediction of the mass, velocity and rate of impinging droplets. Droplets are expelled from the filler wire by electromagnetic, gravitational, inertial and drag forces acting on a droplet forming at the tip of the wire [11]. The size and rate of droplets expelled from the filler wire determines the mode of metal transfer which affects many aspects of the process [12]. The velocity of droplets striking the pool depends on the arc length and the acceleration caused by the plasma jet. Therefore successful prediction of the depth of penetration of the weld pool requires a model that includes the effects of heat transfer in the arc, mass transfer from the filler wire, plasma jet flow in the arc, heat transfer and fluid flow in the pool, and dynamics of impinging droplets. A model of gas metal arc welding that includes all these effects has not yet been developed.

EXPERIMENTS

Bead on plate welds were made on 3/16 and 3/8 inch thick carbon steel using an automated gas metal arc welding apparatus. A constant voltage power supply was used with a slope approximately equal to a 2.5 V decrease per 100 A increase in current. The shielding gas was a mixture of 98% Argon and 2% O_2 flowing at 50 scfh. The travel speed was fixed at 15 mm/s. The filler wire was carbon steel type AWS ER70S-6 and its diameter was fixed at 1.14 mm. The CTWD, which denotes the contact tip to workpiece distance, was set to 13, 19 and 25 mm. Eighteen separate welds were made by varying the wire feed speed from 120 mm/s to 300 mm/s for each value of CTWD, which produced a variation in arc power from 6 to 10 MW. The samples were prepared by wet chemical etching and examined using a microscope. The depth of the weld pool was measured from the original surface of the workpiece to the bottom of the pool. Therefore the reinforcement height was not included in the measured depth.

PROCESS MODEL

The model of droplet growth and detachment used in this work evolved from a prior model developed by Shaw [13] that simulates water dripping from a faucet. This non-linear, second-order, spring-mass-damper system may also be used, with suitable modifications, to simulate droplet dynamics in GMAW:

$$m\ddot{x} + b\dot{x} + kx = F_{tot},$$

where x is the droplet displacement, $\dot{x}$ is the droplet velocity, and $\ddot{x}$ is the droplet acceleration. The time-varying mass of the droplet is m, b is the damping coefficient, k is the spring constant, and F_{tot} is the sum of the external forces acting on the droplet. The spring constant and damping coefficient represent the surface tension and internal viscous forces of the liquid neck attaching the droplet to the solid electrode.

In addition to the force balance, the model includes information about the electrical properties of the welding circuit. The dynamic response of the electrical circuit is found from Ohm's Law augmented by an arc voltage model as shown by Shepard [14]:

164

$$\frac{dI}{dt} = \frac{V_{sp} - R_e I - R_s I - V_{arc}}{L_s},$$

where I is the current, V_{sp} is the power supply voltage setpoint, R_e is the total electrode and droplet resistance, R_s and L_s are the source resistance and inductance, and V_{arc} is the voltage drop across the arc.

The resistance of the solid electrode and droplet is given by:

$$R_e = \rho_e (l_e + 0.5(x + r_d)),$$

where ρ is the resistivity of the electrode material and l_e is the stick-out of the electrode. The length of the drop attached to the electrode includes the radius of the droplet, r_d, and the elongation represented by the spring length, x. The voltage drop across the arc is given by Shepard [15]:

$$V_{arc} = V_o + R_a I + E_a (CTWD - l_e),$$

where V_o is the arc voltage constant, R_a is the arc resistance, E_a is the arc length factor, and CTWD is the contact tube-to-workpiece distance.

The electrode melting rate was shown by Lesnewich [16] to be a function of resistive and anodic heating:

$$M_e = C_1 (I^2 \rho_e l_e) + C_2 (I),$$

where C_1 and C_2 are constants.

The model incorporates two criteria for drop detachment: balance of forces and pinch instability. The first criterion is based on an imbalance of forces acting on the drop. The sum of the external forces on the drop includes four forces that act to produce the dynamics of droplet motion [17]. These are (1) gravity, (2) the Lorentz force due to the interaction of the electric field and its self-induced magnetic field, (3) the aerodynamic drag caused by the arc plasma flowing past the drop, and (4) a force due to electrode melting which adds momentum to the drop. A force due to the surface tension of the liquid metal acts to prevent the drop from detaching. The second detachment criterion is based on pinch instability theory [18]: an electromagnetic pinch force causes a fluid instability and detachment in case the radius of the drop exceeds a critical radius. These criteria are discussed in detail in Reutzel et al. [2].

When either of the detachment criteria are satisfied, a drop detaches and the volume of the detached droplet, Vol_{detach}, is a function of drop velocity and drop volume, Vol_d:

$$Vol_{detach} = Vol_d \frac{1}{2} \left(\frac{1}{1 + e^{-C_s x}} + 1 \right),$$

where C_s is an empirically determined constant. This relation ensures that an increasing droplet velocity leads to an increasing volume of detachment [13].

Calculation of weld pool depth is based on the assumption that the depth of the weld pool depends on the force of impinging droplets. Therefore we consider the case in which the droplets have sufficient momentum to penetrate the weld pool and cause a deep, narrow penetration at the base of the pool. Otherwise the depth of the pool depends on heat transfer in the workpiece, and in this case models by Christensen et al. [3] and Eagar and Tsai [6] may be used to obtain an approximation to the pool depth.

We introduce a characteristic length scale, $\mu/(\rho V)$, where μ is the viscosity of the pool, ρ is the density of the pool, and V is the velocity of impinging droplets. Using Stokes' Law for the force on a sphere immersed in a fluid, we introduce a characteristic force, μRV, where R is the radius of impinging droplets. Assuming a linear relation between dimensionless force and dimensionless depth, we obtain:

$$\frac{D}{\left(\dfrac{\mu}{\rho V}\right)} = A\frac{F}{\mu RV} + B,$$

where D is the depth of penetration, F is the force of impinging droplets, and A and B are constants. The force of impinging droplets is given by:

$$F = \dot{M}V,$$

where $\dot{M}$ is the melting rate of the filler wire. We found that the constants $A = 46$ and $B = 9500$ enable us to correlate force and penetration depth for a range of values of arc power and CTWD.

The velocity of impinging droplets is given by:

$$V = \sqrt{V_0^2 + 2a(L - x)},$$

where V_0 is the velocity of the droplet at the time of detachment from the filler wire, L is the arc length, x is the position of the centroid of the droplet at the time of detachment, and a is the acceleration of the droplet in flight, given by:

$$a = g + \frac{3}{8}C_d\frac{\rho_{gas}V_{gas}^2}{\rho_{drop}R},$$

where g is the gravitational acceleration, C_d is the coefficient of drag, ρ_{gas} is the density of the plasma jet, V_{gas} is the velocity of the plasma jet, and ρ_{drop} is the density of the expelled droplet. We assume that the velocity of the plasma jet is equal to 100 m/s and the coefficient of drag is equal to 0.44 [12].

To compare our model to previous models, we choose the heat conduction model by Christensen et al. [3]:

$$\frac{D}{\left(\dfrac{2\alpha}{U}\right)} = \frac{\varsigma}{1+\varsigma}\sqrt{1+2\varsigma},$$

where α is the thermal diffusivity of the workpiece, U is the travel speed, and ς is the radius of the weld pool obtained from:

$$\frac{\eta VIU}{4\pi\alpha^2\Delta H} = \varsigma\, e^{\frac{\varsigma}{1+\varsigma}},$$

where η is the arc efficiency which is set to 0.80, V is the arc voltage, I is the current, and ΔH is the change in enthalpy associated with raising the workpiece temperature to the melting point.

RESULTS AND SUMMARY

The measured depth of the weld pool as a function of arc power for various values of CTWD is shown in Figures 1-3. These results show that increasing CTWD at constant power leads to an
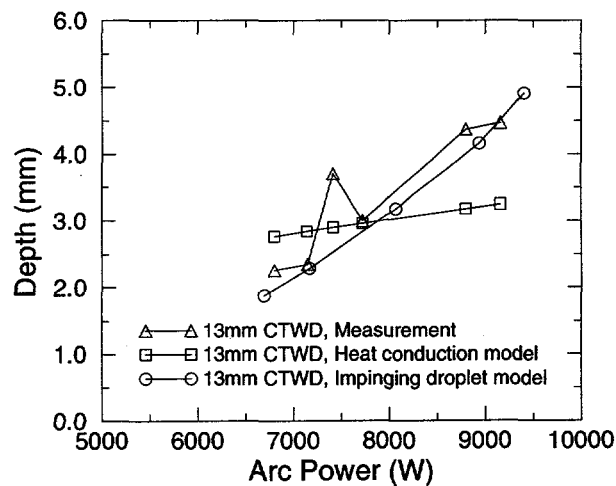

Fig. 1. Measured and simulated depth of the weld pool as a function of arc power for a contact tip to workpiece distance equal to 13mm.

increasing depth of penetration. A larger CTWD is accompanied by an increase in electrical resistance which reduces the current. Therefore the wire feed speed must increase in order to maintain constant power while increasing the CTWD. This leads to an increase in the melting rate of the wire which leads to an increase in the momentum of impinging droplets. Another factor is that a change in CTWD at constant power causes a change in arc length. Since the droplets are accelerated by the plasma jet, a larger arc length increases the momentum of impinging droplets. Therefore the depth of penetration may also increase due to an increase in the arc length.

Comparisons of the measured and simulated depth of the weld pool as a function of arc power for three values of CTWD are shown in Figures 1-3. These results show that in some cases the simulated depth of penetration of impinging droplets is a suitable indicator of the pool depth. In constrast, the simulated depth based on the heat conduction model due to Christensen et al. [3] is not a suitable indicator of the pool depth. Further validation is needed to ensure that the model may be used to correlate droplet momentum and penetration depth for a wider range of welding process variables, including variable travel speed, filler wire size, material type and shielding gas. The results obtained in this study suggest that in some cases the momentum of impinging droplets has a greater effect on the depth of the weld pool than heat transfer from the arc.
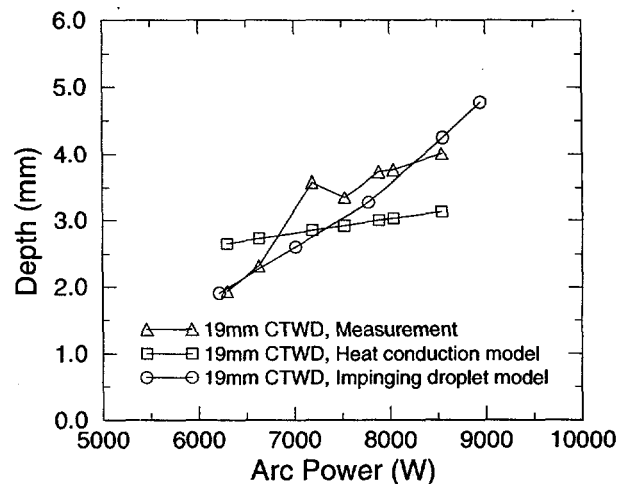

Fig. 2. Measured and simulated depth of the weld pool as a function of arc power for a contact tip to workpiece distance equal to 19mm.
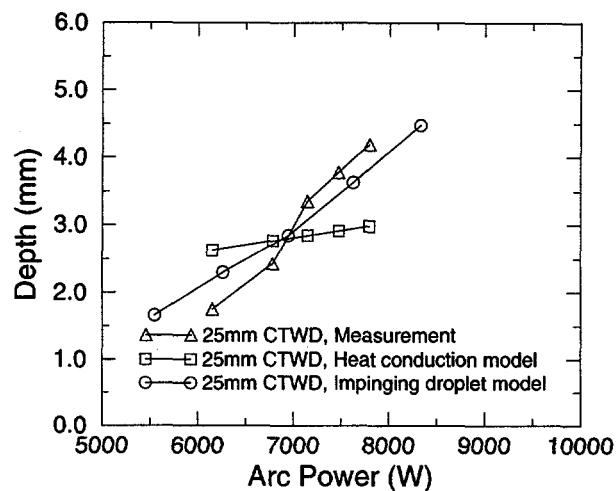
Fig. 3. Measured and simulated depth of the weld pool as a function of arc power for a contact tip to workpiece distance equal to 25mm.

REFERENCES

1. W. G. ESSERS. and R. WALTER, "Heat Transfer and Penetration Mechanisms With GMA and Plasma-GMA Welding," *Welding J.*, Vol. 60, pp. 37s-42s, 1981.

2. E. W. REUTZEL, C. J. EINERSON, J. A. JOHNSON, H. B. SMARTT, T. HARMER, and K. L. MOORE, "Derivation and Calibration of a Gas Metal Arc Welding (GMAW) Dynamic Droplet Model,", 4th International Conference on Trends in Welding Research, Gatlinburg, Tennessee, June 5-8, 1995.

3. N. CHRISTENSEN, V. de L.DAVIES, and K.GJERMUNDSEN, "Distribution of Temperatures in Arc Welding," *British Welding J.*, Vol. 2, pp. 54-75, 1965.

4. K. C. MILLS, and B. J. KEENE, "Factors Affecting Variable Weld Penetration," *Int. Mater. Rev.*, Vol. 35, pp. 185-216, 1990.

5. D. ROSENTHAL, "The Theory of Moving Sources of Heat and its Application to Metal Treatments," *Trans. ASME*, Vol. 43, pp. 849-866, 1946.

6. T. W. EAGAR and N.-S. TSAI, "Temperature Fields Produced by Traveling Distributed Heat Sources," *Welding J.*, Vol. 62, pp. 346s-355s, 1983.

7. P. TEKRIWAL and J. MAZUMDER, "Finite Element Analysis of 3 Dimensional Heat Transfer in GMA Welding," *Welding J.*, Vol. 67, pp. 150s-156s, 1988.

8. E. PARDO and D. C. WECKMAN, "Prediction of Weld Pool and Reinforcement Dimensions of GMA Welds Using a Finite Element Model," *Metall. Trans. B*, Vol. 20, pp. 937-947, 1989.

9. K. C. TSAO and C. S. WU, "Fluid Flow and Heat Transfer in GMA Weld Pools," *Welding J.*, Vol. 67, pp. 70s-75s, 1988.

10. J.-W. KIM and S.-J. NA, "A Study on the Effect of Contact Tube to Workpiece Distance on Weld Pool Shape in Gas Metal Arc Welding," *Welding J.*, Vol. , pp. 141s-152s, 1995.

11. J. H. WASZINK and L. H. J. GRAAT, "Experimental Investigation of the Forces Acting on a Drop of Weld Metal," *Welding J.*, Vol. 62, pp. 109s-116s, 1983.

12. Y.-S. KIM, and T. W. EAGAR, "Analysis of Metal Transfer in Gas Metal Arc Welding," *Welding J.*, Vol. 72, pp. 269s-278s, 1993.

13. R. SHAW, "The Dripping Faucet as a Model Chaotic System," The Science Frontier Express, Aerial Press, Inc., 1984.

14. M. E. SHEPARD and G. E. COOK, "A Non-linear Time Domain Simulation of Self Regulation in Gas Metal Arc Welding," 3rd International Conference on Trends in Welding Research, Gatlinburg, Tennessee, June 1-5, 1992.

15. M. E. SHEPARD, Modeling of Self-Regulation in Gas Metal Arc Welding, Ph.D. Dissertation, Vanderbilt University, 1991.

16. A. LESNEWICH, "Control of Melting Rate and Metal Transfer in Gas-Shielded Metal Arc Welding, Part I: Control of Electrode Melting Rate," *Welding J.*, Vol. 37, pp. 343s-353s, 1958.

17. A. D. WATKINS, H. B. SMARTT, and J. A. JOHNSON, "A Dynamic Model of Droplet Growth and Detachment in GMAW," 3rd International Conference on Trends in Welding Research, Gatlinburg, Tennessee, June 1-5, 1992.

18. J. F. Lancaster, The Physics of Welding, Pergamon Press, Oxford, 1984.

# STATUS OF RESEARCH AIMED AT PREDICTING STRUCTURAL INTEGRITY

W. G. Reuter

Lockheed Martin Idaho Technologies Company
P.O. Box 1625, Idaho Falls, Idaho, USA 83415-2218

ABSTRACT

Considerable research has been performed throughout the world on measuring the fracture toughness of metals. The existing capability fills the need encountered when selecting materials, thermal-mechanical treatments, welding procedures, etc., but cannot predict the fracture process of structural components containing cracks. The Idaho National Engineering and Environmental Laboratory and the Massachusetts Institute of Technology have been collaborating for a number of years on developing capabilities for using fracture toughness results to predict structural integrity. Because of the high cost of fabricating and testing structural components, these studies have been limited to predicting the fracture process in specimens containing surface cracks. This paper summarizes the present status of the experimental studies of using fracture toughness data to predict crack growth initiation in specimens (structural components) containing surface cracks. These results are limited to homogeneous base materials.

INTRODUCTION

The concern addressed in this paper is to identify the ability and limitations of using a single fracture toughness parameter (K, J, or δ), which is assumed to uniquely quantify the displacement, strain, and stress fields at the crack tip, to predict structural integrity. In predicting the fracture process (crack growth initiation, stable crack growth, and catastrophic failure) for structural components, it is necessary to have some method of measuring the fracture toughness of the component and the ability to relate these measurements to the behavior of the structural component. ASTM test standards exist for measuring plane strain fracture toughness (K_{Ic}) [1] and the critical value of J (J_{Ic}) near the onset of stable crack extension.[2] The critical value for crack tip opening displacement (δ) measured per E1290[3] may be substituted for J. K_{Ic} is limited to linear-elastic behavior whereas J_{Ic} is used for linear-elastic and elastic-plastic conditions. A single fracture parameter is used in many prediction procedures; the most commonly used of these procedures is the failure analysis diagram (FAD). Figure 1 is an example of a FAD. The region of interest, in this paper, is the vertical axis on the left side identified as the Toughness Ratio, which is the applied stress intensity factor (K) divided by the material's fracture toughness (K_{Ic}).
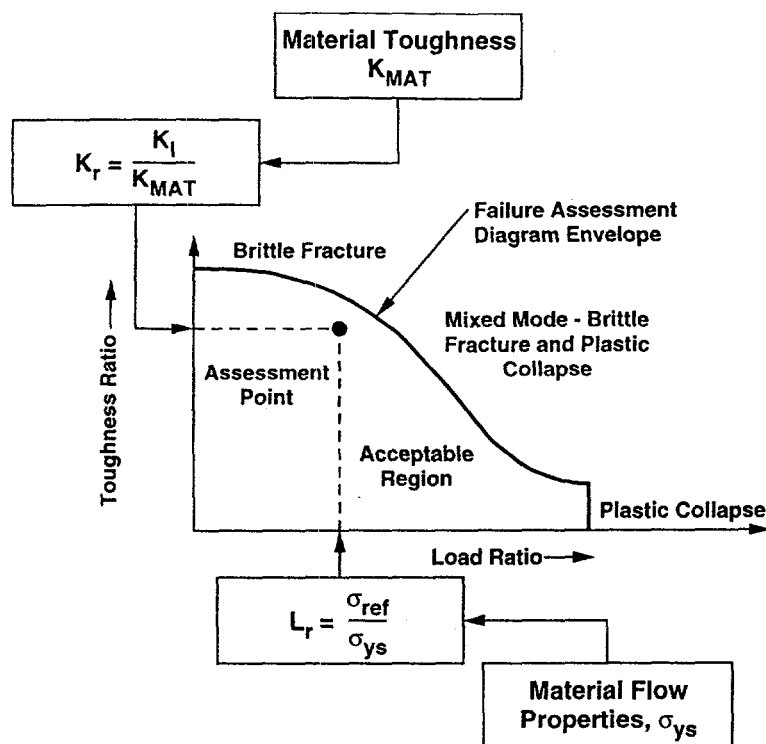
Figure 1. Failure assessment diagram.

Three potentially significant differences exist among a) test specimens used in ASTM E399, E813, and E1290, b) structural components, and c) specimens containing surface cracks (see Table 1). First, the difference in size can often lead to an elastic-plastic or fully plastic condition in the test specimen while the structure may exhibit a linear-elastic behavior. Second, there may be a significant difference between the size of the crack in the test specimen and in the structural component. This difference can lead to a nonconservative estimate of the structural lifetime because of statistical effects. Finally, the longer crack is more likely to intercept an embrittled region (lower fracture toughness) than the shorter crack. Because surface-cracked specimens are more similar to structural components than ASTM test specimens, it is expected that tests of surface-cracked specimens more closely simulate the response of a structure. Therefore, the use of specimens containing surface cracks to simulate the behavior of a structure is a logical step.

The main difference between a through crack and a surface (part-through) crack is that the through crack is frequently treated as a two-dimensional crack problem with the crack driving force reasonably constant along the straight crack front. Crack growth initiation occurs whenever the crack driving force at any location along the crack front equals the fracture toughness. Since the crack driving force is generally constant along the straight crack front, it follows that initiation of crack growth occurs at the same "instant." (It is sometimes observed that the crack growth occurs in a tunnel fashion where the crack growth is retarded at the specimen free surfaces. This situation is of no concern to subject matter being addressed.) The surface crack is a three-dimensional crack (see Figure 2), and the crack driving force varies around the perimeter of the crack front. Currently, the standard approach is to assume that when the crack driving force at any location around the crack perimeter equals the fracture toughness of the material, then crack growth initiation occurs. Another difference is that through-crack test specimens are normally removed from a plate, whereas surface-crack specimens (and structures) can be plates, cylindrical sections, etc.

Table 1. Comparison between specimens and structures.

Configuration	Specimen Size	Crack Configuration	Constraint[a]
Test specimens	Small, flat plates	Straight, through-thickness, crack lengths are generally short	High
Structural components	Large, flat plates or cylindrical sections	Curved, part-through thickness, crack lengths may be large. Often experiences multiaxial loading.	Variable
Surface cracked specimens	Flat plates or cylinders	Curved, crack lengths range from short to medium. Substantial variation in a/t and a/2c. Generally exposed to uniaxial loading.	Variable

[a]Constraint is defined as the ratio of the hydrostatic stress to the equivalent von Mises stress.

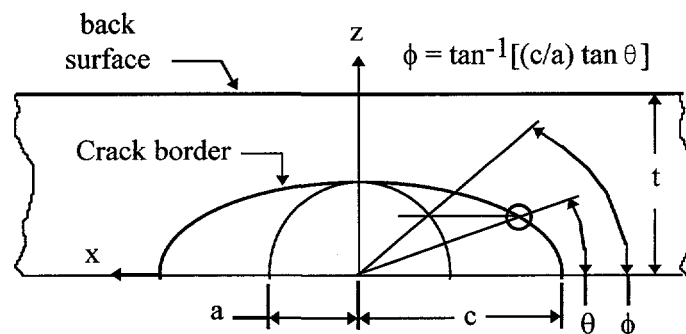$$\phi = \tan^{-1}[(c/a)\tan\theta]$$

Figure 2. Surface crack geometry.

For test specimens that are configured to obtain a maximum constraint, the measured fracture toughness is generally the lowest value that may be obtained. For structural components that exhibit a lower constraint (and generally a higher fracture toughness) due to shallow cracks or an adjacent free surface, use of the lowest value fracture toughness may be too conservative, i.e., requiring unnecessary work stoppage or repairs. This paper summarizes experimental research that provides answers to many questions of transferability of test data to structural behavior.

APPROACH

Fracture toughness measurements provided in ASTM E 399, E 813, and E 1290 are based on conditions associated with initiation of crack growth. For linear-elastic fracture mechanics (LEFM) conditions, crack growth initiation is often, but not always, synonymous with catastrophic failure. Therefore, an approach based on crack growth initiation appears to be useful for structures that exhibit linear-elastic behavior. An alternative approach is to use a procedure based on crack arrest. ASTM E 1221[4] provides test methods for measuring plane-strain crack-arrest fracture toughness. However, it is very difficult to transfer this concept to a structure because it is necessary to quantify the

compliance of the component as it plays a key role in the crack arrest process. Therefore, crack-arrest fracture toughness will be ignored in this paper.

For elastic-plastic [nonlinear elastic fracture mechanics (NLEFM)] and fully-plastic conditions, it is generally observed that substantial stable crack growth may occur after initiation of crack growth. Therefore, to more accurately predict the fracture process for these conditions, it is necessary to have test results that provide information on stable crack growth. The J-Δa results obtained using ASTM E 1152[5] provide information on stable crack growth. But, from an analytical viewpoint, the wake of a growing crack includes cold-worked, plastically deformed material, where Hutchinson-Rice-Rosengren (HRR) solutions are no longer valid, i.e., no longer unique.[a] The limit of crack growth for J validity has been studied by Xia et al.,[6] who concluded that no approach can be based on a single parameter resistance curve. To evaluate if this is a practical problem, however, requires experimental verification. Dadkhah and Kobayashi[7] and May and Kobayashi[8] performed experiments in which they observed that J no longer provided the HRR fields at the crack tip when Δa exceeded some amount of crack growth. This strongly suggests that J no longer represents the crack tip stress fields when crack growth, Δa, is more than two to three times crack-tip opening displacement (CTOD). Based on these statements, the following discussion is limited to using fracture toughness results obtained per existing ASTM Test Standards to predict conditions for initiation of crack growth in specimens containing surface cracks. These specimens have been fabricated from homogeneous materials.

Because of the complexity of the issues being considered, this paper will examine linear-elastic fracture mechanics and nonlinear elastic fracture mechanics as two separate topics. However, the specimens used to measure fracture toughness were removed from the same piece of material as specimens containing surface cracks.

LEFM CONDITIONS

The ability to predict the maximum flaw size that may be allowed in a structural component is based on knowing the applied stress, the fracture toughness (measured per E399), and having an applicable equation. For these tests, the maximum applied stress intensity factor (K_{max}) was calculated using the failure load, the actual crack size, and the Newman-Raju[9] equations. The ability to predict crack growth initiation was quantified by calculating the ratio of the calculated maximum applied stress intensity (K_{max}) to K_{Ic}, the same ratio used in the FAD diagram.

Results

The following results were obtained from three materials, Ti-15-3, a monolithic SiC, and D6-aC (a high-strength steel). The Ti-15-3 was heat-treated to a yield strength (σ_{ys}) of 1,452 MPa, with resulting plane strain fracture toughness (K_{Ic}) = 41.4 MPa$\sqrt{m}$. The specimens containing surface cracks had a fatigue precrack starter notch fabricated by electric discharge machining (EDM), and the specimens were then load cycled either in tension or in bending to grow the desired fatigue precrack. The fatigue precracks had a crack depth-to-thickness ratio (a/t) ranging from 0.05 to 0.94 and crack depth-to-length ratios (a/2c) ranging from 0.01 to 0.47. The specimens were tested by monotonic loading in either tension or bending. The test results consisted of load versus acoustic emission, load versus crack mouth opening displacement, and load versus displacement (quantified using moiré interferometry) data. The specimens containing surface cracks failed catastrophically with little or no

---

[a]Personal communication with F. McClintock, September 11, 1996.

stable crack growth. Reuter et al.[10] discussed these results and noted that K_{max}/K_{Ic} ranged from 1.02 to 1.64 for specimens loaded in tension, and from 1.02 to 2.07 for specimens loaded in bending.

Standardized test procedures are not available for measuring the plane strain fracture toughness of ceramics. The procedures provided in ASTM E 399 were used, except that the SiC specimens contained EDM notches as opposed to fatigue precracks. The defects in the specimens containing surface cracks were also made with EDM and had the same notch root radius. It was assumed that comparisons between the EDM-notched specimens would be as valid as comparisons between specimens containing fatigue precracks. The specimens were tested by monotonic loading in bending. The test results were the same as those collected for Ti-15-3. The specimens failed catastrophically with no stable crack growth. Reuter et al.[10] presented these results and noted that K_{max}/K_{Ic} ranged from 0.99 to 1.41 for one series of SiC specimens and from 0.94 (1.00)[b] to 1.93 (1.39) for a second series of SiC specimens.

The tests described above were performed at a single facility and it was desired to broaden the scope of material to include a high-strength steel and multiple test facilities. Therefore, an International Cooperative Test Program was organized to test specimens fabricated from D6-aC, a high-strength steel.[11] The material was heat-treated to σ_{ys} = 1,587 MPa, with resultant K_{Ic} = 54 MPa$\sqrt{m}$. The surface crack configuration had an a/2c ratio ranging from 0.08 to 0.60 (0.56) for tensile loading and from 0.08 to 0.51 for bending loads; the a/t ratio ranged from 0.23 (0.28) to 0.89 (0.83) for tensile loads and from 0.22 to 1.0 (0.84) for bending loads. The fracture toughness results per ASTM E399 were the same as those collected for Ti-15-3 in that they exhibited a nominal elastic behavior. For the surface-cracked specimens, the behavior ranged from general elastic to substantial crack growth prior to attainment of the maximum load. Use of the initial precrack size and shape and the maximum load at failure to calculate K_{max} for comparison with K_{Ic} was inappropriate. Therefore, the test plan was modified to detect the onset of crack growth initiation using d.c. potential drop and acoustic emission. A change of 5% in d.c. potential drop was defined as crack growth initiation. The K_{max}/K_{Ic} ratio ranged from 0.91 (0.98) to 1.61 for tensile loading and from 1.14 to 1.81 for bending loads, see Figure 3.

Discussion

In a vast majority of these 99 tests, $K_{max}/K_{Ic} \geq 1.0$.[c] This illustrates that the use of the Newman-Raju[9] equation and the measured K_{Ic} result in conservative estimates of failure for specimens (structures) containing surface cracks. But a number of instances (8 out of 19 metal specimens tested in bending) were observed in which considerable conservatism occurred ($K_{max}/K_{Ic} > 1.50$) when K_{max} occurred at the free surface (ϕ = 0 degrees). These results are acceptable for many applications, but it might be necessary to better understand the conditions controlling fracture. The primary questions of interest are a) What parameter other than the calculated crack driving force (K_{app}) is responsible for initiation of crack growth? and b) Is the local K or modification (average, specific locations, etc.) responsible for initiation of crack growth?

In Table 2, test results of the SiC specimens[10] showed that 10 specimens had K_{max} occurring at the free surface and that the ratio K_{max}/K_{Ic} ranged from 0.99 to 1.41 (1.28), suggesting that the substantial conservatism was not observed in this material. For the Ti-15-3 specimens tested,[10] 10 specimens had K_{max} occurring at the free surface and the ratio K_{max}/K_{Ic} ranged from 1.18 to 2.07

---

[b]Items in parentheses denote the value of the next closest neighbor.

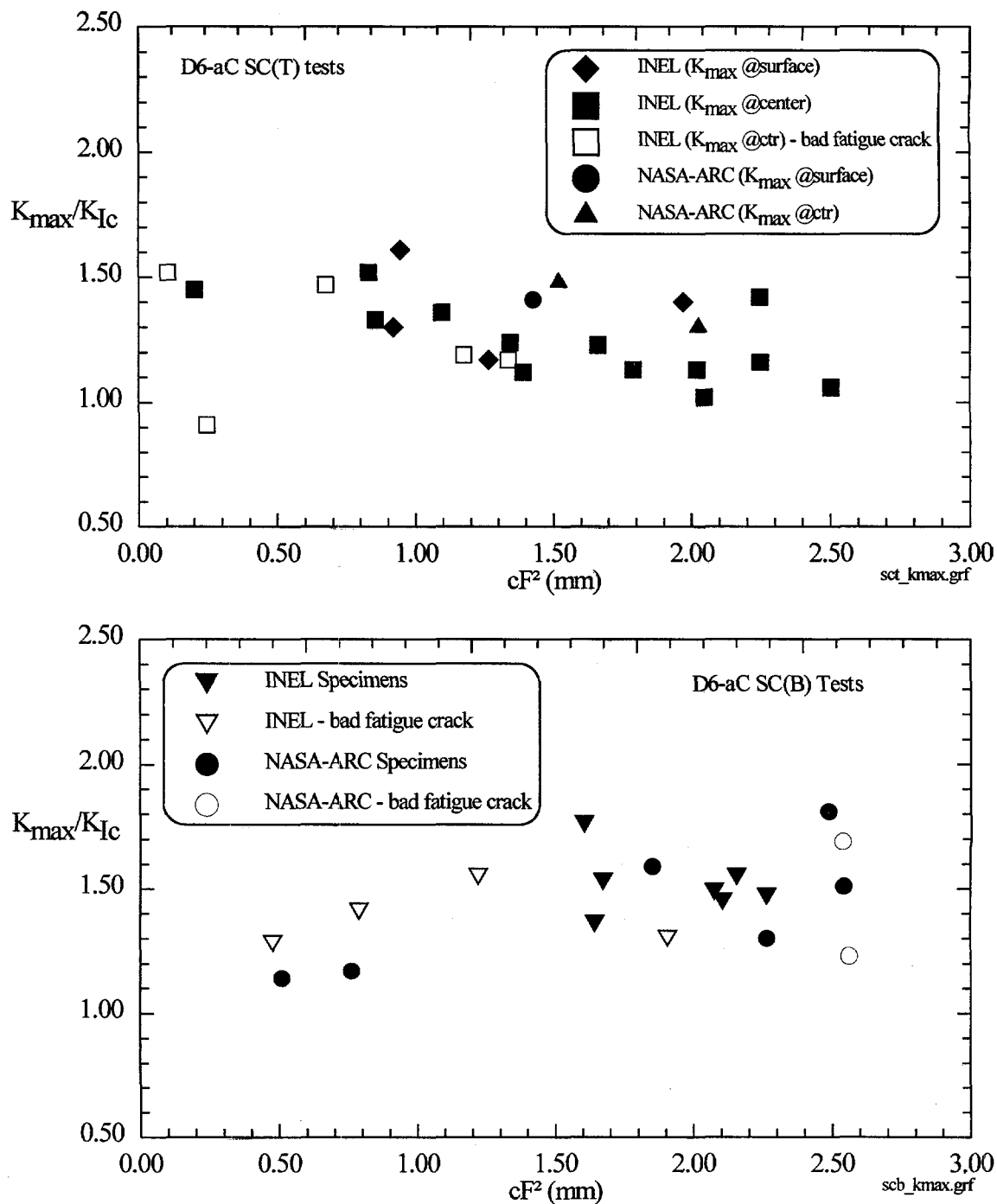[c]This ratio was greater than 1.0 for 97 of 99 specimens tested.

Figure 3. K_{max}/K_{Ic} versus cf^2 for load corresponding to crack growth initiation (d.c. potential drop).

(1.83), suggesting that substantial conservatism was observed in this material. Reuter et al.[10] examined the use of K_{max}, $K_{average}$ (root-mean square of K along the crack front), K_{ave} (average of local K values calculated at all locations along the crack front), and K_{ϕ} (at a specific location). It was concluded that K_{max} was the most conservative single parameter fracture criterion if attainment of K_{Ic} is considered a sufficient condition for fracture. A major concern is how long a crack length segment is required where $K=K_{Ic}$ before crack growth initiation occurs. This has not been answered yet. The use

of $K_{average}$ and $K_{\phi=30°}$ was based on results of Sommers and Aurich[12] for elastic-plastic conditions where it was observed that the maximum crack driving force (J) occurred at ϕ = 30 degrees. This suggested that the maximum CTOD (δ) also occurred at ϕ = 30 degrees. Reuter and Lloyd[13] showed that δ was not a maximum at ϕ = 30 degrees for specimens tested in tension that exhibited elastic-plastic behavior. They observed that the relative magnitude of δ followed the calculated relation for K at applied stresses where crack growth initiation was detected.

As noted earlier, many D6-aC steel specimens containing surface cracks exhibited substantial crack growth. A combination of electric potential change (DCP) and acoustic emission (AE) monitoring was used to detect initiation of crack growth and the associated applied load. The load was reduced after initiation was detected, and cyclic loading was applied to decorate the location and extent of crack growth. Several specimens loaded in tension were examined and it was observed that crack growth initiation occurred at ϕ = 90 degrees, with the majority of crack growth within ±15 degrees of ϕ = 90 degrees, and with no crack growth at ϕ = 30 degrees (see Figure 4). Several of these specimens, loaded in tension, were loaded multiple times, which consisted of (1) fatigue precracking, (2) monotonic loading to obtain stable crack growth, (3) decreased load, (4) cyclic loading to outline the region of the stable crack growth, and (5) unloading. Steps 2, 3, 4, and 5 were repeated until the crack penetrated the back surface (see Figure 4).

For three specimens with thickness (t) = 6.4 mm, K_{init} increased for each cycle of monotonic loading, i.e., with increasing a/t (see Figure 5). This suggests a) a loading history effect, i.e., plastic zone development, b) a constraint effect due to the proximity of the crack tip to the back surface, or c) the Newman-Raju[9] stress intensity distribution limit a/t < 0.80 for accurate results has been exceeded. At this time, none of the three possible explanations has been ruled out.

Table 2. Summary of the ratio K_{max}/K_{Ic} for the materials tested.

Material	Test Conditions	Range of K_{max}/K_{Ic}
Ti-6Aal-4V	Tension	1.02 to 1.64
	Bending	1.02 to 2.07
SiC (Two different materials)	Bending	0.99 to 1.41
	Bending	0.94 to 1.93
D6-aC	Tension	0.91 to 1.61, see Fig. 3
	Bending	1.14 to 1.81, see Fig. 3

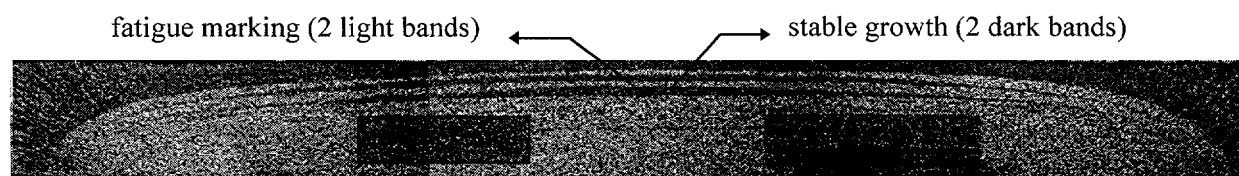fatigue marking (2 light bands) ⟵   ⟶ stable growth (2 dark bands)

Figure 4. D6-aC Specimen #26 - Fracture surface showing two loading sequences.
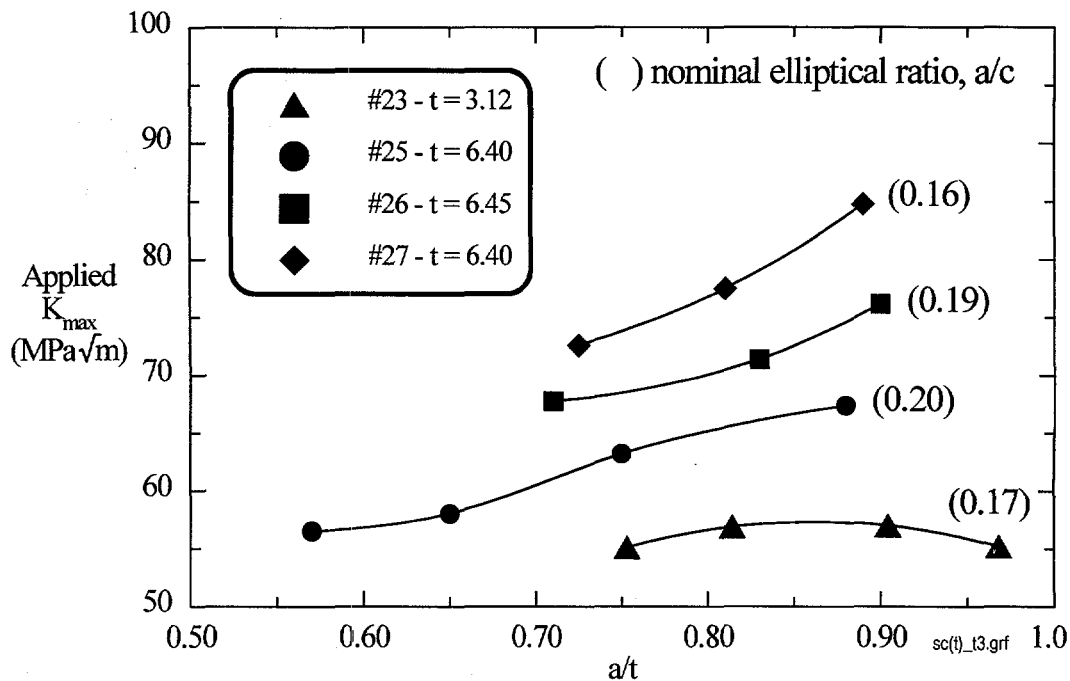
Figure 5. Crack initiation stress intensity factors for varying crack depths.

Chao and Reuter[14] examined several specimens loaded in bending and observed that initiation did not occur at the free surface even though K_{max} was calculated to occur at $\phi = 0$ degrees (see Figure 6). Of 19 specimens tested in bending, the crack growth initiation sites were located for 15 specimens. For these specimens, the crack initiation sites occurred at ϕ ranging from 45 to 9 degrees, with an average of $\phi = 21$ degrees (close to 30 degrees). The reason for $\phi = 21$ degrees is not yet understood.

These results show that the Toughness Ratio used in the FAD (see Figure 1) can be conservative if based on using applicable equations[9] for flat plate specimens (structures) containing surface cracks. (It is necessary to perform a similar series of tests for cylindrical sections.) A potential concern is that the conservatism may be excessive for specimens experiencing bending loads where K_{max} occurs at the free surface. Therefore, it is necessary to perform numerical analyses to develop an understanding of what controls the location of crack-growth initiation.

ELASTIC-PLASTIC CONDITIONS

For conditions in which K is no longer applicable, it is necessary to use either J or δ as the critical fracture toughness parameter. Unfortunately, the boundary separating LEFM from elastic-plastic fracture mechanics is not well defined.

Results

Specimens fabricated from ASTM A710 steel were tested at a temperature (22°C) corresponding to about midway in the ductile-brittle transition region. Reuter et al.[15] performed fracture toughness tests of C(T) (compact tension), SE(B) (single-edge notch bend), M(T) (center-cracked plate), and SC(T) (surface-cracked plates loaded in tension) specimens. Both multiple-specimen techniques and single-specimen techniques were used in the test procedure. For multiple specimens, several replicate specimens were loaded, each to a different value of load or displacement, and the specimens unloaded. Each of these specimens were sectioned, polished, and examined metallographically to measure the
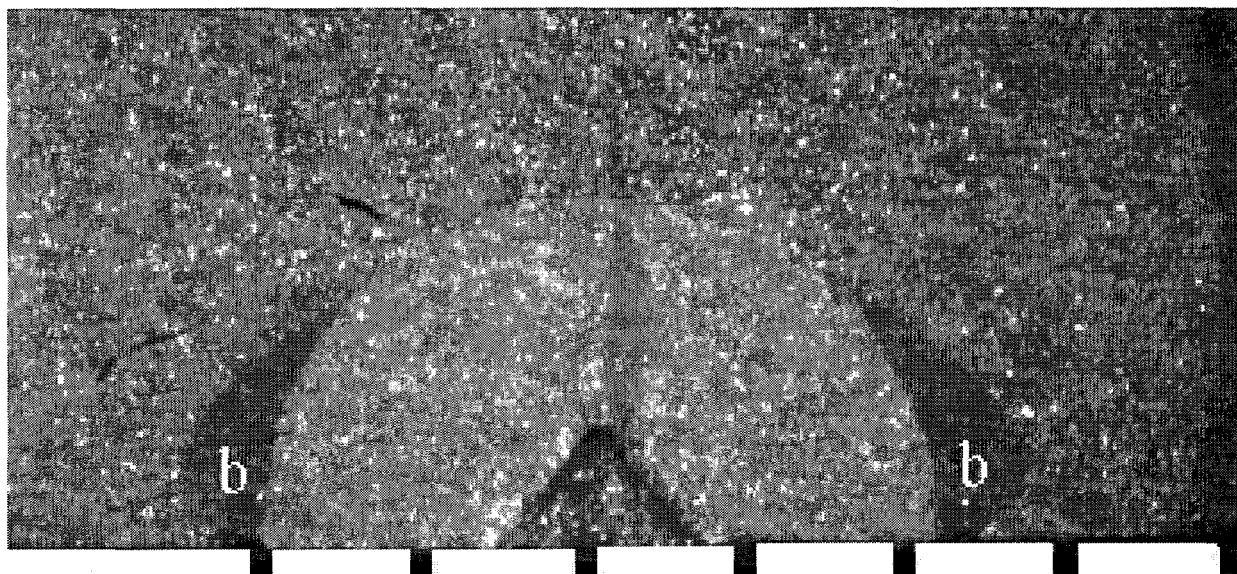
Figure 6. D6-aC Specimen #10 - Fracture surface showing stable growth near surface at "b." The dark region adjacent to the semicircular precrack is where crack growth occurred. (mm scale bars)

extent of crack growth and the corresponding value of δ. These results were used to obtain a plot of δ versus Δa, which was then extrapolated to $\Delta a = 0$ to estimate δ for crack growth initiation. For the single-specimen evaluation, a specimen was loaded until crack growth was detected, the load reduced, and either the specimen experienced cyclic loading to failure or it was loaded in liquid nitrogen to cause cleavage fracture, either of which will decorate the magnitude of stable crack growth. The two fracture surfaces were then examined using microtopography to measure δ at crack growth initiation. See Reuter and Lloyd[13] for additional discussion of the microtopography technique.

In addition, the constraint (hydrostatic stress normalized by dividing by the equivalent stress based on the von Mises yielding criteria) was calculated for each of the above specimen configurations, and a relation was observed between δ (crack growth initiation) and constraint. This relation was expanded and more completely developed in Hancock et al.[16]

In Reuter and Lloyd,[13] a series of tensile tests were performed in which measurements of δ and crack tip opening angle were made as a function of load at several locations around the perimeter of the surface cracks. These specimens were loaded in tension and had a a/2c ratio of 0.1 or 0.5. For specimens with a/2c = 0.1, crack growth initiation occurred at $\phi = 90$ degrees and disappeared at $\phi = 0$ degrees. As the applied stress (σ) to σ_{ys} ratio approached 0.96, δ_{pl} (plastic component of δ) followed the prediction of Newman and Raju.[9] As $\sigma/\sigma_{ys} > 0.96$, there was a larger increase in δ_{pl} near the free surface than occurred elsewhere.

For a/2c = 0.5, crack growth initiation appeared to occur at $\phi = 90$ degrees and disappeared at $\phi = 0$ degrees. (Some crack growth was visible around much of the surface crack perimeter, but none at $\phi = 0$ degrees.) As σ/σ_{ys} approached 1.22, δ_{pl} followed the prediction of the Newman and Raju[9] equation. When $\sigma/\sigma_{ys} > 1.3$, a substantial increase in δ/δ_{el} (δ_{el} is an elastic component of δ) occurs, particularly at $\phi = 90$ degrees.

Discussion

The results in Reuter et al.[15] suggest that the use of δ measured per E1290 may provide too severe conservatism when predicting the condition for crack growth initiation of a structural component containing a shallow crack. If the relation between δ and constraint is known, it is possible to use a more realistic value for δ. The corresponding value of δ for the surface crack specimen was 1.4 times the value obtained in the more highly constrained SE(B) specimen (see Figure 7).

From a metallurgical point of view, the relationship between δ or J and constraint is unique because of the microstructure, trace elements, and thermal mechanical procedures used to develop the material to its final microstructure. Therefore, depending on the consequences of failure, it is strongly suggested that specimens/crack depth that provide the same constraint expected in the structure be tested to verify the magnitude of δ responsible for crack growth initiation.

Figure 7. Crack tip opening displacement (at initiation) for different amounts of constraint.

The substantial changes in δ as a function of σ/σ_{ys} observed in Reuter and Lloyd[13] are probably due to loss of constraint. Parks and Wang[17] predict loss of HRR field dominance to occur at about $\sigma/\sigma_{ys} = 0.85$ and 1.04 for $a/2c = 0.1$ and 0.5, respectively. But constraint was still high at the load at which crack growth initiation occurred, and it was observed that crack growth initiation occurred at $\phi = 90$ degrees.

This suggests that since the surface-cracked specimen (structure?) and the middle-cracked tensile specimen (structure?) will have a constraint equal to or less than that of the test specimen designed per ASTM standards, that J/J_{Ic} or $\delta/\delta_{in} \geq 1.0$. Therefore, the Toughness Ratio in the FAD diagram of Figure 1 will be conservative for flat plate specimens loaded in tension.

CONCLUSIONS

Structural components were not available to determine the ability of using fracture toughness data to predict structural integrity, so specimens containing surface cracks were used instead. This is a logical step as the surface cracks used in the specimens encompass many of the configurations found in structural components but they generally experience only uniaxial loading..

These results cannot be used to evaluate crack size effects associated with differences between the specimen size and the structure size because the surface-cracked specimens were limited in size.

For LEFM conditions, crack growth initiation did not occur when $K_{max} = K_{Ic}$. The use of K_{Ic} and the Newman-Raju equation resulted in conservative predictions of crack growth initiation. This was true for surface-cracked specimens loaded in tension or in bending. For specimens tested in bending, the conservatism may be substantial when K_{max} occurs at the free surface, especially metallic specimens— the magnitude of the conservatism was reduced considerably for monolithic ceramics. These conclusions are limited to conditions in which crack growth initiation is detected. If stable crack growth
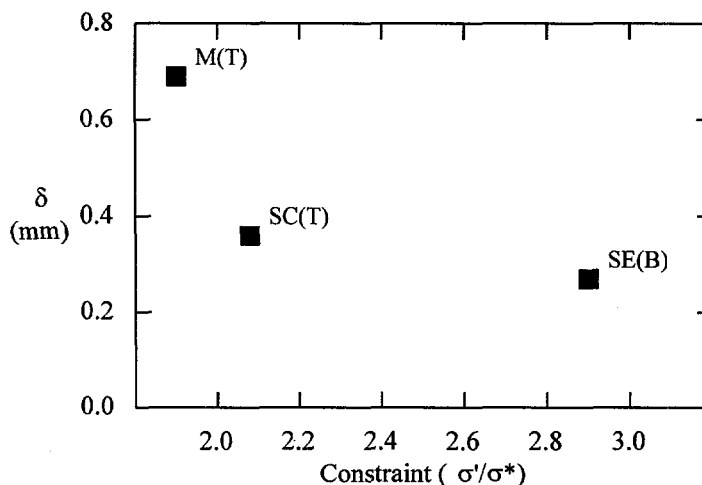
occurs, then the use of the maximum applied stress and the original crack size used in conjunction with the Newman-Raju equation to calculate K_{max} has no relevance to K_{Ic}.

At this time, it is not possible to answer if an average K (a K at a specific location or an additional parameter with a local K value) is responsible for initiation of crack growth for surface-cracked specimens. There is no basis to suggest that structural components will behave differently from the surface-cracked specimens tested. (Sufficient tests to evaluate the sensitivity of the fracture initiation process to multiaxial loading have not been performed.)

For NLEFM conditions, it is possible to predict crack growth initiation for specimens containing surface cracks if the relationship between δ and constraint has been quantified. This relation is known to vary as a function of material type and is expected to vary as a function of heat-to-heat variations within a given material. Therefore, once this relation is known for the specific material of interest, it is then necessary to quantify the constraint for the specific crack in the structural component. It is then possible to identify the critical value of δ associated with the specific constraint for predicting structural integrity.

The results for LEFM, where specimens are loaded in tension or in bending, and for elastic-plastic conditions where specimens are loaded in tension, the Toughness Ratio of the FAD, is conservative. The magnitude of the conservatism may become substantial for LEFM conditions where bending loads are encountered and K_{max} occurs at the free surface.

The above conclusions are limited to flat plates in tension or in bending for LEFM and flat plates in tension for elastic-plastic conditions and do not address cylinders.

ACKNOWLEDGMENT

REFERENCES

1   ASTM E399-90. Standard Test Method for Plane-Strain Fracture Toughness of Metallic Materials. *Annual Book of ASTM Standards*, Section 3, Vol. 03.01.

2   ASTM E813-89. Standard Test Method for J_{Ic}, A Measure of Fracture Toughness. Ibid.

3   ASTM E1290-89. Standard Test Method for Crack-Tip Opening Displacement (CTOD) Fracture Toughness Measurement. Ibid.

4   ASTM E1221-88. Standard Test Method for Determining Plane-Strain Crack-Arrest Fracture Toughness, K1a of Ferritic Steels. Ibid.

5   ASTM E1152-87. Standard Test Method for Determining J-R Curves. Ibid.

6   Xia, L., C. F. Shih, and J. W. Hutchinson (1995). A Computational Approach to Ductile Crack Growth Under Large Scale Yielding Conditions. *J. Mech. Phys. Solids 43*, 389-414.

7   Dadkhah, M. S. and A. S. Kobayashi (1994). Two-Parameter Crack Tip Stress Field Associated With Stable Crack Growth in Thin Plate. *Fracture Mechanics, 24th Vol.*, ASTM STP 1207, J. D. Landes, D. E. McCabe, and J. A. M. Boulet (eds.), 48-61.

8   May, G. B. and A. S. Kobayashi (1995). Plane stress stable crack growth and J-integral/HRR field. *Int'l J. Solids and Struct.* 37/617, 857-881.

9   Newman, J. C. and I. S. Raju (1978). Analysis of Surface Cracks in Finite Plates Under Tension or Bending Loads. NASA Technical Paper 1578, National Aeronautics and Space Administration, Washington, D.C.

10   Reuter, W. G., J. C. Newman, Jr., B. D. Macdonald and S. R. Powell (1994). Fracture Criteria for Surface Cracks in Brittle Materials. *Fracture Mechanics: Twenty-Fourth Volume*, ASTM STP 1207, J. D. Landes, D. E. McCabe, and J. A. M. Boulet (eds.), American Society for Testing and Materials, Philadelphia, PA, 1994, 617-635.

11   Reuter, W. G., N. Elfer, D. A. Hull, J. C. Newman, Jr., D. Munz and T. L. Ponontin (1996). Test Results and Preliminary Analysis for International Cooperative Test Program on Specimens Containing Surface Cracks. ASTM-sponsored *28th National Symposium on Fatigue and Fracture, Albany, NY, June 25-27, 1996.*

12   Sommers, A. and D. Aurich (1991). On the Effect of Constraint on Ductile Fracture. *Defect Assessment in Components - Fundamentals and Applications*, J. G. Bland and K.-H. Schwalbe (eds.), ESIS/EGF Publication 9, 141-174.

13   Reuter, W. G. and W. R. Lloyd (1990). Measurement of CTOD and CTOA Around Surface-Crack Perimeters and Relationships Between Elastic and Elastic-Plastic CTOD Values. *Surface-Crack Growth, Experiments, and Structures*, ASTM STP 1060, W. G. Reuter, J. H. Underwood, and J. C. Newman, Jr. (eds.), American Society for Testing and Materials, Philadelphia, PA, 152-176.

14   Chao, Y. J. and W. G. Reuter (1996). Fracture of Surface Cracks. ASTM-sponsored *28th National Symposium on Fatigue and Fracture, Albany, NY, June 25-27, 1996.*

15   Reuter, W. G., S. M. Graham, W. R. Lloyd and R. L. Williamson (1991). Ability of Using Experimental Measurements of δ to Predict Crack Initiation for Structural Components. *Defect Assessments in Components - Fundamentals and Applications*, J. G. Bland and K.-H. Schwalbe (eds.), 1991 Mechanical Engineering Publication, London, 175-188.

16   Hancock, J. W., W. G. Reuter, and D. M. Parks (1993). Constraint and Toughness Parameterized by T. *Constraint Effects in Fracture, ASTM STP 1171*, E. M. Hackett, K.-H. Schwalbe, and R. H. Dodds (eds.), American Society for Testing and Materials, Philadelphia, PA, 21-40.

17   Parks, D. M. and Y. Y. Wang (1988). Elastic-Plastic Analysis of Part-Through Surface Cracks. Proceedings of the Symposium on Analytical, Numerical, and Experimental Aspects of Three-Dimensional Fracture Process, American Society of Mechanical Engineers, Berkeley, CA, June 1988.

Multisensor Based Robotic Manipulation in an Uncalibrated Manufacturing Workcell

Bijoy K. Ghosh, Di Xiao, Ning Xi and Tzyh-Jong Tarn
Campus Box 1040, Washington University
One Brookings Drive
Saint Louis, MO 63130-4899, USA
(314) 935 6039 (Phone)
(314) 935 6121 (Fax)
ghosh@zach.wustl.edu

Abstract

The main problem that we address in this paper is how a robot manipulator is able to track and grasp a part placed arbitrarily on a moving disc conveyor aided by a single CCD camera and fusing information from encoders placed on the conveyor and also from encoders on the robot manipulator. The important assumption that distinguishes our work from what has been previously reported in the literature is that the position and orientation of the camera and the base frame of the robot is apriori assumed to be unknown and is 'visually calibrated' during the operation of the manipulator. Moreover the part placed on the conveyor is assumed to be non-planar, i.e. the feature points observed on the part is assumed to be located arbitrarily in $\mathbb{R}^3$. The novelties of the proposed approach in this paper includes a (i) *multisensor fusion scheme* based on complementary data for the purpose of part localization, and (ii) *self-calibration* between the turntable and the robot manipulator using visual data and feature points on the end-effector. The principle advantages of the proposed scheme are the following. (i) It renders possible to reconfigure a manufacturing workcell without recalibrating the relation between the turntable and the robot. This significantly shortens the setup time of the workcell. (ii) It greatly weakens the requirement on the image processing speed.

1 Introduction

Sensor-guided tracking plays an important role in today's flexible manufacturing systems. Using sensors, a manufacturing system can compensate for changes in environments and uncertainties in its model. Vision becomes an especially useful sensor in Robotics. Use of visual information provides a way of overcoming some difficulties of uncertain models and unknown environment and hence extends the domain of applications of robots without explicit intervention or reprogramming.

Control of robot manipulators with vision in the feedback loop has an exciting history starting probably with the pioneering work of Hill and Park [1], Weiss, Sanderson and Neuman [2]. Subsequent work in the area has focused on "Visual Servoing" wherein the emphasis is to visually locate the position and orientation of an object and to control a robot manipulator to grasp and manipulate the object. If the object is not stationary, then the process of locating the object and repositioning the robot through control must be repeated iteratively until the task has been accomplished. This leads naturally to real time vision based feedback and control problems that have been subsequently studied in [3], [4], [5], [6], [7] and probably many others. As a result, many important tasks, such as Bolt Insertion, Conveyor Belt Picking, Weld Seal Tracking, Part Mating, Road Vehicle Guidance, Juggling, Fruit Picking etc, to name a few, have been accomplished with the aid of computer vision.

However, one has to overcome many difficulties in order to utilize visual information. First, visual data is not always reliable. Vision systems could occasionally fail to generate any useful information but noise due to variation of illumination, overlapping of different workpieces or accidental obstruction of the camera. Secondly, image processing algorithms are always time-consuming. Therefore, direct use of visual information for robot control purpose will lead to a poor accuracy or even, in the case of dynamic visual servoing, can cause stability

problem. In addition, most of previous work in visual servoing area required precise calibration before control schemes were implemented. As is well-known, calibration is also a time-consuming job and makes visual servo control expensive or even impossible in some cases.

We propose in this paper, a new visually guided planning and control of a robot manipulator for precise positioning and tracking in an uncertain and dynamic environment. The proposed scheme is robust against precise position of the camera and utilizes a scheme called *virtual rotation* wherein the observed image is transformed to what it would appear if the camera has been pointed vertically down. Additionally the proposed scheme is also robust against uncertainty in precise position of the robot manipulator.

In this paper, the real time part localization does not entirely depend on the data from the CCD camera. Information from encoders attached to the rotating turntable is also used in a new "Multiple Sensor Integration" scheme, which greatly reduces the requirement on the image processing speed while obtaining a good motion estimation. In addition, the proposed scheme introduces the idea of considering complimentary data (instead of redundant ones as in traditional multisensor fusion schemes) in a fusion process.

This paper also introduces a new 'self-calibration' scheme in order to ascertain the position and orientation of the robot manipulator in the workcell. The scheme utilizes a set of apriori chosen points on the end-effector to compute the coordinate transformation between the base of the manipulator and a certain fixed coordinate attached to the workcell. (iii) Although our approach is feature based, we do not propose to visually estimate structure and motion. The estimation scheme presented in this paper differs considerably from standard results in motion and structure estimation from feature correspondences (see [10]). As a matter of fact, on one hand, the motion of the moving part in a fixed disc frame is obtained by fusing visual information from the camera with readings from an encoder attached to the rotating disc conveyor. On the other hand, the structure information in terms of the orientation of the end-effector, is measured with respect to the base coordinate frame of the robot using the encoders attached to the robot. One of the main issues addressed in this paper is "how to fuse information obtained by the visual and the encoder sensors". The main conclusion is that the procedure of 'multiple sensor integration' already introduced in [8] leads to a unique self-calibration scheme which is reliable and has been tested via experiments (iv) New planning and control schemes in task space are also discussed.

The paper is structured as follows. We emphasize the real-time part localization scheme of a three dimensional part on the turntable in sufficient details. In so doing, we show that even when the height of the part is unknown, the relative position of the part can be computed visually. The self calibration scheme, described in this paper, has already been reported in [8]. Thus the details of this scheme are omitted.

2   Multisensor Based Visual Sensing

The sensor fusion scheme discussed in this paper combines three sources of sensory data to obtain measurements that individual sensors cannot obtain. The scheme is described as follows. From visual information, we recover the relative position and orientation of the part with respect to the reference line joining the disc center and the reference point on the turntable. Since the relative pose of the part does not change over time, from reading the encoder of the turntable the real-time pose of the part with respect to the fixed disc frame can be obtained. Combining the visual information with data from encoders of the robot and the turntable, we also can determine the relation between the fixed disc frame and the base frame of the robot. Finally, we compute the real-time pose of the part in the base frame of the robot.

In this paper, the position and orientation of the part are determined based on feature points. Fundamentally, in our case, the problem of determining the pose of the part can be reduced to that of determining the position of a point in some coordinate frame under the assumptions that the plane of the turntable and XY-plane of the base frame of the robot are parallel and that the part to be manipulated has a simple known shape.

Let $OXYZ$ be any Cartesian coordinate frame with its origin O at the disc center and its Z-axis perpendicular to the turntable. After virtual rotation of the camera, the position of a point p in 3-D space can be represented by

$$X_p = \frac{aX_{ref} - bY_{ref}}{l^2} + \frac{(\overline{a} - a)X_{ref} - (\overline{b} - b)Y_{ref}}{lLf}h; \tag{2.1}$$

$$Y_p = \frac{bX_{ref} + aY_{ref}}{l^2} + \frac{(\overline{b} - b)X_{ref} + (\overline{a} - a)Y_{ref}}{lLf}h, \tag{2.2}$$

with

$$a = \overrightarrow{C_oP_o} \cdot \overrightarrow{C_oR_o}; \qquad b = \overrightarrow{C_oP_o} \times \overrightarrow{C_oR_o}; \tag{2.3}$$

$$\overline{a} = \overrightarrow{C_oO_o} \cdot \overrightarrow{C_oR_o}; \qquad \overline{b} = \overrightarrow{C_oO_o} \times \overrightarrow{C_oR_o}; \tag{2.4}$$

where $\overrightarrow{AB}$ denotes the vector from A to B while $\cdot$ and $\times$ are operators of dot product and cross product, respectively. C_o, R_o and P_o are the transformed images of the disc center, the reference point and the point p via perspective projection, respectively. O_o is the intersection of the optical axis and the transformed image plane. h stands for the distance of the point p from the turntable. $(X_{ref}, Y_{ref}, 0)$ are the coordinates of reference point in the frame $OXYZ$ and (X_p, Y_p, h) the coordinates of the point p in the same frame. L denotes the length of the reference line while l is the length of the transformed image of the reference line; f is the focal length of the camera. Note that a, b, $\bar{a}$, $\bar{b}$ and l can be easily computed in terms of the image coordinates of C_o, R_o and P_o. In our configuration, L and f are known constants. For the fixed camera, l is also a constant. Hence, if the coordinates (X_{ref}, Y_{ref}) of the reference point are given in the frame $OXYZ$, the position of the point p in the same frame is an affine function of h, the distance of the point from the turntable.

2.1 Determination of relative position of a point

Suppose that X-axis of the fixed disc frame $O_d X_d Y_d Z_d$ coincides with the initial position of the reference line joining the disc center and the reference point and that X-axis of the attached disc frame $O_a X_a Y_a Z_a$ is along with the real-time position of the reference line. Therefore, the two frames are exactly the same initially. The coordinates of the reference point on the turntable in the attached disc frame are $(L, 0, 0)$.

From (2.1) and (2.2), the relative position of a point which has distance of h away from the turntable, with respect to the attached disc frame, can be written as

$$({}^aX_p, \ {}^aY_a, \ {}^aZ_p) = (\frac{aL}{l^2} + \frac{(\bar{a} - a)}{lf}h, \ \frac{bL}{l^2} + \frac{(\bar{b} - b)}{lf}h, h).$$

with

$$\begin{cases} a = (x_r - x_c)(x_p - x_c) + (y_r - y_c)(y_p - y_c); \\ b = (x_p - x_c)(y_r - y_c) - (y_p - y_c)(x_r - x_c); \\ \bar{a} = x_c(x_c - x_r) + y_c(y_c - y_r); \\ \bar{b} = y_c(x_r - x_c) - x_c(y_r - y_c); \end{cases} \qquad (2.5)$$

where (x_c, y_c), (x_r, y_r) and (x_p, y_p) are transformed image coordinates of the disc center, the reference point and the point p, respectively.

Especially, for the points on a part of negligible height, we obtain their coordinates in the attached disc frame $(\frac{aL}{l^2}, \frac{bL}{l^2}, 0)$.

Suppose that the distance of a point on the part from the turntable is unknown. we cannot hope to determine the relative position of the point via a single image. However, since the part is moving as the turntable rotates and the relative position of the part with respect to the attached disc frame does not change over time, hence we can also determine the relative position of the part from multiple images taken at different time instants.

Let us consider a point p on the part. Suppose two different images are taken at time t_1 and t_2, respectively. From the image taken at t_i ($i = 1, 2$), we have the same relative position of the point with respect to the attached disc frame, namely,

$$({}^aX_p, \ {}^aY_p, \ {}^aZ_p) = (\frac{a_i L}{l^2} + \frac{(\bar{a}_i - a_i)}{lf}h, \ \frac{b_i L}{l^2} + \frac{(\bar{b}_i - b_i)}{lf}h, h),$$

where a_i, b_i, $\bar{a}_i$ and $\bar{b}_i$ are computed by means of Eqn.(2.5) with $(x_c(t_i), y_c(t_i))$, $(x_r(t_i), y_r(t_i))$ and $(x_p(t_i), y_p(t_i))$ being transformed image coordinates of the disc center, the reference point and the point p at time t_i, respectively. Therefore, it is seen that

$$\underbrace{\begin{bmatrix} \bar{a}_1 - a_1 + a_2 - \bar{a}_2 \\ \bar{b}_1 - b_1 + b_2 - \bar{b}_2 \end{bmatrix}}_{A} h = \underbrace{\begin{bmatrix} a_2 - a_1 \\ b_2 - b_1 \end{bmatrix}}_{B} \frac{fL}{l}. \qquad (2.6)$$

from which one can get a least squire solution, namely $h = (A^T A)^{-1} A^T B$.

As a result, the relative position of the point p can easily be determined. It should be pointed out that so far we have only used the visual information for determination of the relative position of a point on the part with respect to the attached disc frame via one or two images.

185

2.2 Determination of the relation between the fixed disc frame and the base frame

In order to determine the relation between the fixed disc frame and the base frame of the robot, we need to describe a set of points in both frames since the two frames are related by

$$^bP = {}^bR_d\, {}^dP + {}^bT_d,$$

where ^bP and ^dP are the coordinates of a point in the base frame and the fixed disc frame, respectively. We assume that the plane of the turntable and XY-plane of the base frame are parallel. It is seen that there is one unknown in the rotation matrix ^bR_d. ^bT_d has three unknown elements. For a point q, if we know its coordinates in the both frames, i.e. ^bP_q and ^dP_q, then from the last equality we have three equations for the four unknown variables. Obviously, in order to get a unique solution to the relation, we need to know at least two points in the both frames. Fortunately, from reading encoders of the robot, the coordinates of points on the end-effector with respect to the base frame of the robot can be readily obtained. In what follows, we describe the points on the end-effector in the fixed disc frame with the aid of the single camera.

Suppose that an image was taken at time t. At the very moment the coordinates of the reference point in the fixed disc frame are $(L\cos(\theta(t)), L\sin(\theta(t)), 0)$.

After computing the corresponding a, b, $\bar{a}$ and $\bar{b}$ via the image data, the coordinates of a point on the end-effector in the fixed disc frame can be easily obtained by substituting the results into (2.1) and (2.2), as long as the distance of the point from the turntable is known.

Suppose that the distance between the plane of the turntable and the XY-plane of the base frame is unknown apriori. Now, we need to determine the distance of the point from the turntable in order to describe the point in the fixed disc frame.

In [9] we proposed an algorithm to estimate the distance by considering two basic cases: two points or three points on the end-effector are observed by the camera. Of course, more than three points may be employed to increase robustness against the possible noise in data. To save space, we do not give details here. For those two cases, by combining visual information with information from encoders of the robot, we have a set of quadratic equations of one unknown. The main results we obtained in [9] include the following. In case of two points, there are at most two solutions. In many cases, the unique solution can be determined by taking advantage of physical constraints. In the case where three points on the end-effector are observed, the solution can be uniquely determined. Moreover, in this case, we do not need know the individual correspondence between the points and their images. A by-product is that we can determine the actual individual correspondence as well as the unique solution.

It should be pointed out that similar problem (so-called 3D to 2D problem) has been studied in the computer vision field [10]. As you have seen above, however, the results obtained in [9] are much stronger both regarding the uniqueness of the solution and regarding the correspondence between points and their images because in our cases additional information coming from the robot can be utilized.

2.2.1 Computation of the relation between the frames

Having described at least two points in both the fixed disc frame and the base frame of the robot, we obtain

$$\begin{bmatrix} ^bx_i \\ ^by_i \\ ^bz_i \end{bmatrix} = {}^bR_d \begin{bmatrix} ^dx_i \\ ^dy_i \\ ^dz_i \end{bmatrix} + {}^bT_d, \quad i = 1, 2, \cdots$$

where $(^bx_i, {}^by_i, {}^bz_i)$ and $(^dx_i, {}^dy_i, {}^dz_i)$ are the coordinates of the i-th point with respect to the base frame and the fixed disc frame, respectively. Recall the assumption that the plane of the turntable and XY-plane of the base frame are parallel, we know that ^bR_d has the following structure

$$^bR_d = \begin{bmatrix} r_{11} & r_{12} & 0 \\ -r_{12} & r_{11} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.7}$$

186

In this case, we can determine ^bR_d and ^bT_d with knowing two points in the two frames. As a matter of fact, it is seen that

$$\begin{bmatrix} ^bx_2 - {}^bx_1 \\ ^by_2 - {}^by_1 \\ ^bz_2 - {}^bz_1 \end{bmatrix} = {}^bR_d \begin{bmatrix} ^dx_2 - {}^dx_1 \\ ^dy_2 - {}^dy_1 \\ ^dz_2 - {}^dz_1 \end{bmatrix} \tag{2.8}$$

Let

$$\begin{bmatrix} ^be_1 \\ ^be_2 \\ ^be_3 \end{bmatrix} = \begin{bmatrix} ^bx_2 - {}^bx_1 \\ ^by_2 - {}^by_1 \\ ^bz_2 - {}^bz_1 \end{bmatrix}, \quad \begin{bmatrix} ^de_1 \\ ^de_2 \\ ^de_3 \end{bmatrix} = \begin{bmatrix} ^dx_2 - {}^dx_1 \\ ^dy_2 - {}^dy_1 \\ ^dz_2 - {}^dz_1 \end{bmatrix}$$

Eqn.(2.8) represents linear equations of r_{11} and r_{12}. As long as the line joining the two points is not parallel to z-axis of the base frame, the equations always have a unique solution. However, such a solution may not satisfy the constraint $r_{11}^2 + r_{12}^2 = 1$ due to the possible noise in observed data and computation errors. In other words, ^bR_d obtained in this way may not be orthogonal and is therefore meaningless.

Actually, the problem of determining ^bR_d can be viewed as the optimization problem of determining ^bR_d with the structure in (2.7) such that

$$\left\| \begin{bmatrix} ^be_1 \\ ^be_2 \\ ^be_3 \end{bmatrix} - {}^bR_d \begin{bmatrix} ^de_1 \\ ^de_2 \\ ^de_3 \end{bmatrix} \right\|_2$$

is minimized subject to the constraint $r_{11}^2 + r_{12}^2 = 1$. Solving this optimization problem yields

$$\begin{bmatrix} r_{11} \\ r_{12} \end{bmatrix} = \frac{1}{\sqrt{^be_1^2 + {}^be_2^2}\sqrt{^de_1^2 + {}^de_2^2}} \begin{bmatrix} ^be_1\,{}^de_1 + {}^be_2\,{}^de_2 \\ ^be_1\,{}^de_2 - {}^be_2\,{}^de_1 \end{bmatrix}$$

With knowing ^bR_d, we have

$$^bT_d = \begin{bmatrix} ^bx_1 \\ ^bx_2 \\ ^bx_3 \end{bmatrix} - {}^bR_d \begin{bmatrix} ^dx_1 \\ ^dx_2 \\ ^dx_3 \end{bmatrix}.$$

Now, we can determine any observed point on the part with respect to the base frame of the robot. Based on the recovered coordinates of the points on the part, we can easily know the position of the centroid of the part and the orientation of the part.

3   Experiments

An experimental system has been set up in the Center for Robotics and Automation at Washington University. It consists of one PUMA 560 manipulator and a turntable with the diameter of 0.9m. On the turntable the center and a reference point are marked so that they can clearly be seen by the vision system. The distance between the center and the reference point is 0.295 m. The computer vision system consists of a CCD camera with image resolution of 256×256 and the Intelledex Vision processor based on a 16 MHz Intel 80386 CPU. The focal length of the camera is 0.0125 m. The vision system interfaces to the host computer, a SGI 4D/340 VGX. Visual measurements are sent to SGI by a parallel interface. The robot is controlled by UMC controller that also interfaces to SGI through memory mapping.

In the experiments, a non-planar part was randomly placed on the turntable. The precise locations of camera and turntable in the base frame of the robot are both unknown, though we assume that the plane of turntable and XY-plane of the base frame of the robot are parallel. The robot has been successfully controlled to pick up the part and then drop it to a prespecified place. In this paper, we focus on the proposed multisensor-based calibration-free schemes for localization of the part in the base frame of the robot. The experimental results regarding robot control are not presented here.

In the experiments, two tasks have been completed. One is to control the robot to pick up a part after the turntable stops rotation. We call it static picking up task. The other one is to require the robot to track and pick up a moving part while the turntable is rotating. We call it tracking and picking up task. We observed two points on the end-effector. Recall that if the line joining the two points is approximately parallel to the turntable then a unique solution for the distance of the end-effector from the turntable can be determined. Therefore, we can determine the position and orientation of the part in the base frame of the robot.

The experimental results for static picking up task are indicated in Fig. 4. The actual and estimated trajectory of the centroid of the part in base frame of the robot in Fig. 4 (a) , while Fig. 4 (b) provides the actual and estimated

orientation of the part with respect to the base frame of the robot. Fig. 4 clearly shows the procedure during which the robot end-effector is tracking, picking up a moving part and drop it to a pre-specified location. In the figure, the part is undergoing a circular motion in a horizontal plane and the robot end-effector moves from some place above the plane, approaches to the part, grasps it and then goes straightforward to the pre-specified place.

In the experiment for the static picking task, we have taken one hundred images. For each image, we can recover the position and orientation of the part in the base frame by using the proposed multisensor fusion scheme. To deal with possible noise in observed data, we choose average of all the results as the final estimation. Note that the relative position and orientation of the part with respect to the turntable are time-invariant. Once the position of the turntable in the base frame and the relative pose of the part with respect to the turntable are determined, one can easily know the trajectory of the part in the base frame via encoder reading. In the experiment for the tracking and picking task, fifty images have continuously been taken. In this case, we take the last computation results of the pose of the part as our estimate and feedback the values to the planner. This is why no "shaking" part can be seen in Fig. 4. Because there are some estimation errors in the position of the center of the turntable, even after completion of the estimation the trajectory error of the part is a periodic function of time due to the rotation of the turntable. It should be pointed out that the estimation errors are within 0.01 (m). In our experiments, if the position errors of the part are within 0.01 (m) and the orientation errors are within 5 degrees, the picking tasks can be successfully accomplished.

The above experiments have verified the proposed estimation scheme. It does not just provide a Calibration-Free Vision for part localization, but also significantly reduces the requirement for processing speed of the vision system. In these experiments, a very primitive vision system with a low speed CPU was used. Nevertheless, it is capable of providing the position and orientation estimations for a fast moving part.

4 Conclusion

In this paper we developed a sensor fusion scheme estimate the position and orientation of parts in an uncalibrated environment in order to manipulate the parts in a typical manufacturing workcell that is composed of a robot manipulator, a rotating turntable and a camera system. Even though the visual computations are performed in low rate, part position and orientation information can still be updated in same rate of feedback loop using an additional encoder sensor. We also demonstrate a practical tracking algorithm which takes into account the fact that the torque that the robot control system can supply is bounded. The proposed algorithm is primarily based on error feedback with an extra error reduction term added in order to force the required torque requirement to remain within acceptable bounds. The main features of our scheme are (i) requirement of speed for image processing is reduced and hence cost of the system is low; (ii) the scheme can handle uncalibrated environment and therefore has certain flexibility, allowing possible reconfiguration of a manufacturing workcell without recalibrating the relation between the turntable and the robot and also (iii) the scheme can be used in various tasks, e.g. peg-in-hole in a horizontal plane and the like. The experimental results clearly demonstrate the advantages of the proposed scheme.

References

[1] J. Hill and W. T. Park , Real time control of a robot with a mobile camera, Proc. of 9th ISIR, Washington D.C., March 1979, pp.233-246.

[2] L. Weiss, A. Sanderson, and C. Neuman, Dynamic sensor based control of robots with visual feedback, *IEEE J. of Robotics and Automation*, **RA-3**(5):404–417, 1987.

[3] B. H. Yoshimi and P. K. Allen, Alignment using an uncalibrated camera system, *IEEE J. of Robotics and Automation*, **RA11**(4):516–521, 1995.

[4] E. D. Dickmanns and V. Graefe, Dynamic monocular machine vision, *Machine Vision and Applications*, 1:223–240, 1988.

[5] J. T. Feddema and O. R. Mitchell, Vision-guided servoing with feature-based trajectory generation, *IEEE J. of Robotics and Automation*, **RA-5**(5):691–700, 1989.

[6] N. Papanikolopoulos, P. K. Khosla, and T. Kanade, Visual tracking of a moving target by a camera mounted on a robot: a combination of control and vision, *IEEE J. of Robotics and Automation*, **RA-9**(1):14–35, 1993.

[7] P. I. Corke, Visual control of robot manipulators–a review, In K. Hashimoto, editor, *Visual Servoing*, pages 1–32. World Scientific, 1994.

[8] B. K. Ghosh, T. J. Tarn, N. Xi, Z. .Y Yu and Di Xiao, Visually controlled manipulation of parts in a robotic manufacturing workcell, the Proc. of the 1996 IEEE Trans. Robotics and Automation, Minnesota, U.S.A., April, 1996.

[9] Bijoy K.Ghosh, Di Xiao, Ning Xi and Tzyh Jong Tarn, Calibration-free vision based control of a robotic manipulator, The 34th Annual Allerton Conference on Communication, Control and Computing, Allerton, Illinois, USA, Oct.2-4,1996.

[10] T. S. Huang, A. N. Netravali, Motion and structure from feature correspondences: A Review, Proc. of the IEEE, Vol.82, No.2, Feb. 1994, pp.252-268.

[11] R. I. Hartley, Estimation of relative camera positions for uncalibrated cameras, Proc. of 2nd E.C.C.V., Santa Margherita Ligure, Italy, May 1992, pp.579-587.

[12] M. A. Fischler and R. C. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM, Vol.24, NO.6, June 1981, pp.381-355.

[13] Y.Wu, S. S. Iyengar, R. Jain, S.Bose, A new generalized computational framework for finding object orientation using perspective trihedral angle constraint, IEEE Trans. on PAMI, Vol.16, No.10, Oct. 1994, pp.961-975.

[14] K.Kanatani, Group-Theoretical Methods in Image Understanding, Springer-Verlag, 1990.

[15] P.Allen and R.Bajcsy, Object recognition using vision and touching, Proc. of the 9th Joint Conf. on Artificial Intelligence, Los Angeles, Aug. 1985.

[16] A. J. Koivo and N. Houshangi, Real-time vision feedback for servoing robotic manipulator with self-tuning controller. IEEE Trans. on Systems, Man and Cybernetics, Vol.21, No.1, Jan/Feb. 1991, pp.134-142.

(a) Actual and estimated traj. of part in base frame
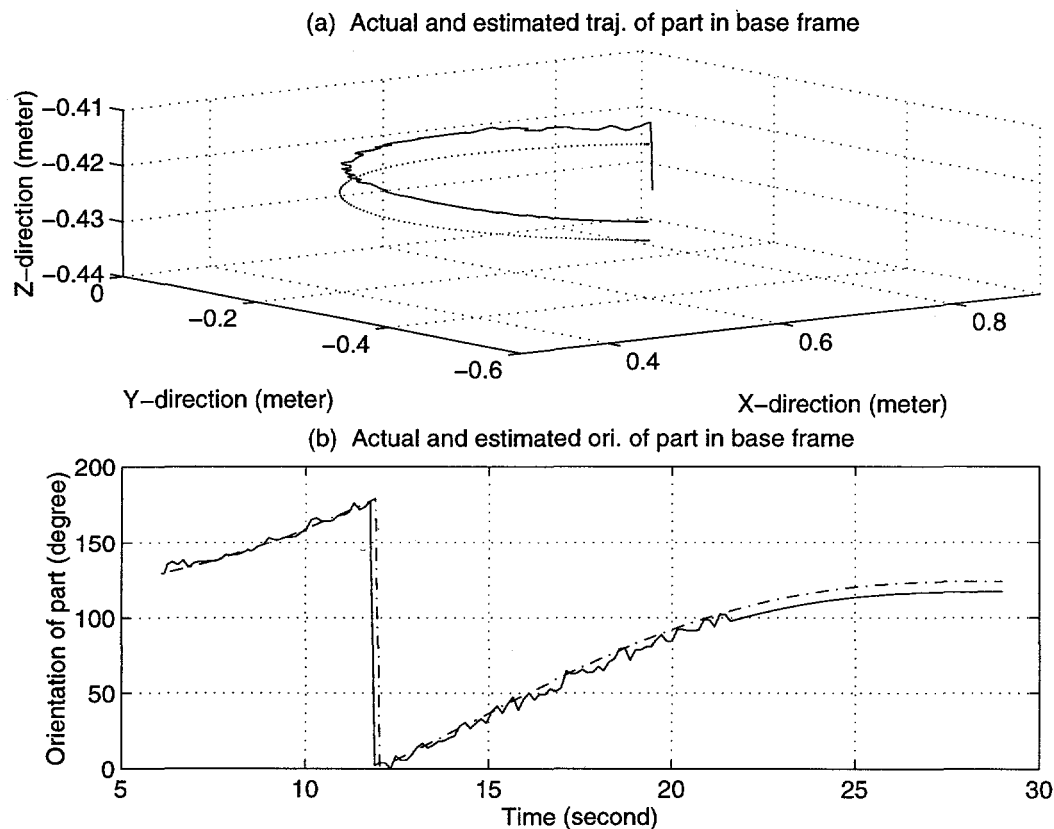
(b) Actual and estimated ori. of part in base frame

Figure 4.1: The actual and estimated pose of the part in the base frame of the robot for the static picking task.
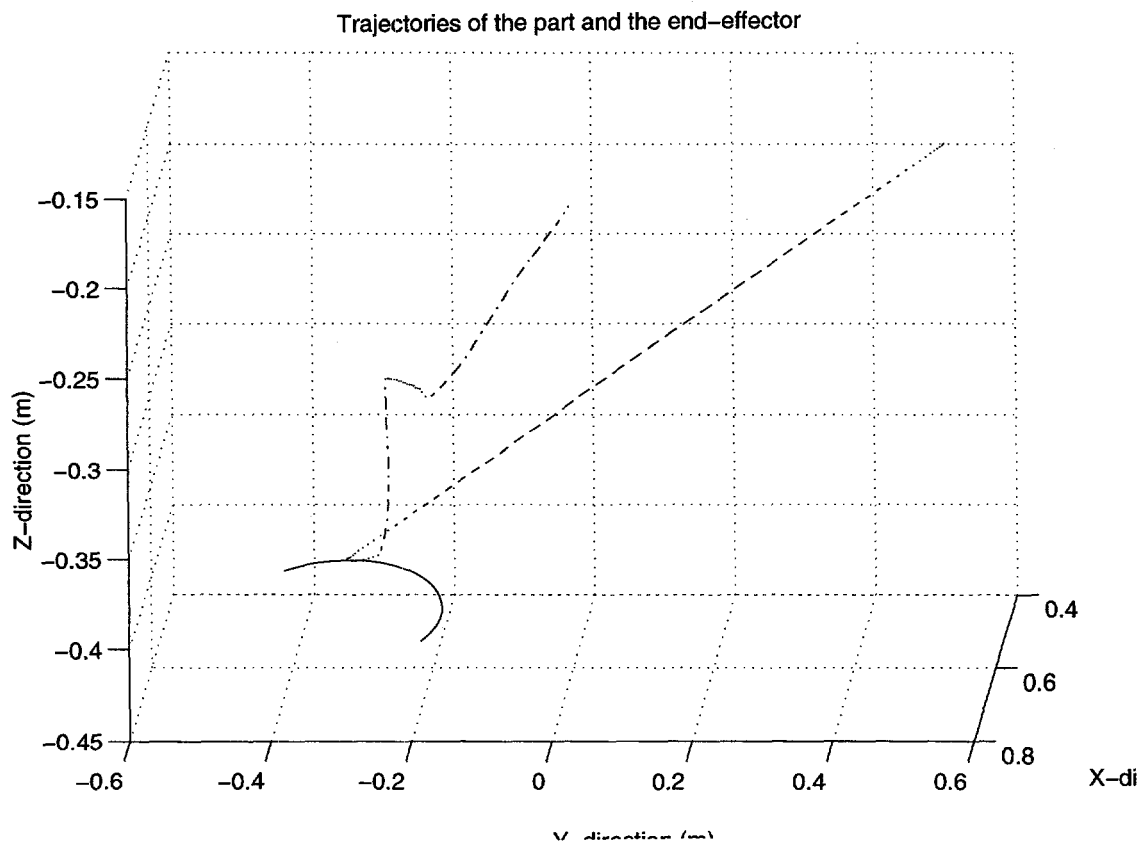
Figure 4.2: The trajectories of the moving part and the end-effector of the robot in the tracking and picking task.

BIFURCATION AND NECK FORMATION AS A PRECURSOR TO DUCTILE FRACTURE DURING HIGH RATE EXTENSION

L. B. Freund and N. J. Sörensen*

Division of Engineering, Brown University, Providence, RI 02912

Abstract

A block of ductile material, typically a segment of a plate or shell, being deformed homogeneously in simple plane strain extension commonly undergoes a bifurcation in deformation mode to nonuniform straining in the advanced stages of plastic flow. The focus here is on the influence of material inertia on the bifurcation process, particularly on the formation of diffuse necks as precursors to dynamic ductile fracture. The issue is considered from two points of view, first within the context of the theory of bifurcation of rate-independent, incrementally linear materials and then in terms of the complete numerical solution of a boundary value problem for an elastic-viscoplastic material. It is found that inertia favors the formation of relatively short wavelength necks as observed in shaped charge break-up and dynamic fragmentation.

INTRODUCTION

During extension of a ductile plate or radial expansion of a ductile shell, deformation is often found to proceed more or less homogeneously within the plastic range until it is interrupted by the formation of localized necks or shear bands. These regions of localized deformation commonly evolve into ductile fractures. The focus here is on the formation of such necks during high rate extension, and on the influence of inertia on the stress level necessary for neck formation.

To this end, a rectangular block of an incompressible elastic–plastic material deforming at high rate under plane strain conditions is considered; see Figure 1. The block can be viewed as a segment of a plate or shell when symmetry is enforced. Opposite ends of the block are subjected to a uniform normal velocity V_0 in the 1–direction. Otherwise, the faces of the block are free of traction. The goal is to establish conditions on loading, geometry and material properties under which the homogeneous deformation can give way to a non–homogeneous deformation.

Up to the instant of bifurcation, the rate of deformation field throughout the material is essentially uniform. Material coordinates X_k referred to an underlying cartesian basis are introduced in the reference configuration for the block, which is taken to occupy the region $-\ell_1 \leq X_1 \leq \ell_1$, $-\ell_2 \leq X_2 \leq \ell_2$. The dominant material velocity field referred to these coordinates is

$$v_1^0(X_1, X_2, t) = \frac{X_1 V_0}{\ell_1}, \qquad v_2^0(X_1, X_2, t) = -\frac{X_2 V_0}{\ell_1 \left(1 + V_0 t/\ell_1\right)^2} \tag{1}$$

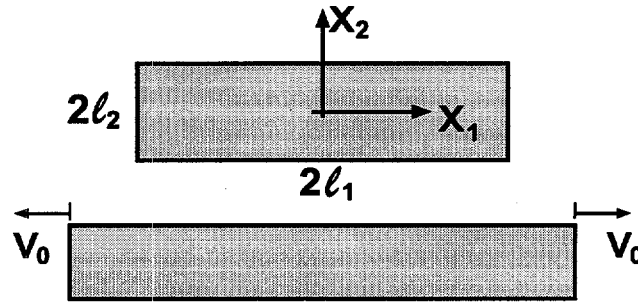*Now at Department of Mechanical Engineering, Lund University, Sweden.

Figure 1: A reference configuration (top) for a block of material, and a subsequent configuration (bottom) achieved by homogeneous plane strain extension.

The instantaneous stretching rate in the 1–direction is V_0/ℓ_1.

It is noteworthy that the deformation (1) cannot be maintained by a homogeneous state of uniaxial stress. The pathlines of material particles are curves in space, which implies a nonuniform distribution of particle acceleration throughout the block. The equations enforcing the balance of linear momentum lead to the conclusion that the stress state is a homogeneous tension in the 1–direction, say σ, plus a hydrostatic pressure which varies in the 2–direction. In terms of components of Cauchy stress σ_{ij}, the distribution has the form $\sigma_{11} = \sigma - p_I$, $\sigma_{22} = -p_I$ and $\sigma_{12} = 0$. In terms of material coordinates in the reference configuration, the pressure p_I has the parabolic distribution

$$p_I(X_2) = \frac{\rho V_0^2}{\ell_1^2} \frac{\ell_2^2 - X_2^2}{(1 + V_0 t/\ell_1)^4} \qquad (2)$$

where ρ is the material mass density.

A number of studies of the conditions for bifurcation of the elastic–plastic block under quasistatic imposed extension or compression have been reported. Hill and Hutchinson [1] identified the regimes of behavior and ranges of moduli for which the governing equations are elliptic, parabolic or hyperbolic. They also calculated the spectrum of bifurcation stresses for symmetric and antisymmetric diffuse deformation modes. Young [2] carried out a similar analysis for plane strain compression. Needleman [3] extended the analysis for plane strain tension/compression to solids characterized by a flow rule with the plastic deformation rate not being normal to the flow surface. Recently, Benallal and Tvergaard [4] examined the case of bifurcation of a block for a particular strain gradient plasticity theory. All of these analyses have been based on Hill's [5,6] bifurcation theory for quasistatic deformation. Results on inertial influence on neck formation in tension have been reported in [7–10].

BIFURCATION ANALYSIS

In this section, attention is on the possible onset of a spatially *nonuniform* deformation field as a bifurcation from the velocity field (1) at a configuration which represents some fully developed stage of deformation. A useful simplification of the boundary value problem formulation is obtained by choosing the current configuration, whose state is being interrogated, as the reference configuration and by choosing the time scale so that the system is in this configuration at $t = 0$, as has been done in (1). Thus, the velocity field (1) evaluated at $t = 0$ is that of both the current and reference configurations. For this choice of reference configuration, the nominal stress s_{ij} and the Cauchy stress σ_{ij} are identical (but their *rates* are different, in general). All equations can

be satisfied by continuing homogeneous deformation beyond the reference configuration. However, the goal is to seek conditions for which the governing equations can also be met for a continuing deformation which is *not* homogeneous beyond the reference configuration.

Following Hill and Hutchinson [1], the balance of rate of momentum in terms of rate of nominal stress reduces to

$$(\dot{s}_{11} - \dot{s}_{22})_{,12} + \dot{s}_{21,22} - \dot{s}_{12,11} = \rho (v_{1,1} - v_{2,2}) \tag{3}$$

upon elimination of the mean pressure. Likewise, the behavior of an initially isotropic and rate–independent material can be represented by the incrementally linear relationships

$$\dot{s}_{11} - \dot{s}_{22} = \left[2\mu^* - \tfrac{1}{2}\sigma\right] (v_{1,1} - v_{2,2}) \tag{4}$$

$$\dot{s}_{12} = \left[\mu + \tfrac{1}{2}\sigma\right] v_{2,1} + \left[\mu - \tfrac{1}{2}\sigma + p_I\right] v_{1,2} \tag{5}$$

$$\dot{s}_{21} = \left[\mu - \tfrac{1}{2}\sigma + p_I\right] v_{2,1} + \left[\mu - \tfrac{1}{2}\sigma\right] v_{1,2} \tag{6}$$

where σ is the uniform deviatoric stress acting in the 1–direction and p_I is the hydrostatic pressure given in (2). This material is characterized by two parameters, the instantaneous tangent modulus $4\mu^*$ for ongoing extension and the instantaneous tangent modulus μ for shearing of principal directions. For elastic–plastic materials the parameters are limited by $0 < \mu/\mu^* < \tfrac{1}{2}$.

The perturbation fields on the uniform background deformation which must satisfy these equations are subject to the boundary conditions

$$v_1 = 0 \quad \text{at} \quad X_1 = \pm \ell_1 \tag{7}$$

$$\dot{s}_{12} = 0 \quad \text{at} \quad X_1 = \pm \ell_1, \qquad \dot{s}_{21} = 0 \quad \text{and} \quad \dot{s}_{22} = 0 \quad \text{at} \quad X_2 = \pm \ell_2 \tag{8}$$

The condition (7) is the kinematic condition on the velocity field, and the conditions (8) are the natural boundary conditions.

The equations governing the linear rate problem are equivalent to the variational statement that the functional

$$\Phi[v_i] = \int_{t_1}^{t_2} \int_{A_X} \left[\tfrac{1}{2}\dot{s}_{ij}v_{j,i} - \tfrac{1}{2}\rho\dot{v}_i\dot{v}_i\right] dX_1\, dX_2\, dt \equiv W[v_i] - K[v_i] \tag{9}$$

is stationary under variations of the instantaneous velocity field, where A_X is the area of the plane occupied by the block in its reference configuration and the times t_1, t_2 are arbitrary. It is important to note that, for purposes of the variational statement, the nominal stress rate depends on velocity only through its linear dependence on *current* velocity gradient.

It is obvious that the kinetic energy–like functional $K[v_i]$ is positive for all admissible velocity fields. In analogy with the linear theory of elasticity, the uniqueness of the incrementally linear rate problem is therefore assured if the stress work–like functional $W[v_i] > 0$. The coefficients in the constitutive equations (4)–(6) are different from those in the quasistatic case because of the pressure term p_I. Thus, the failure of the inequality can be expected to occur under different conditions in the dynamic case, in general, than in the quasistatic case. Hutchinson's [11] sufficiency criterion for a quasistatic bifurcation of a plastic solid with a smooth yield surface under multiaxial stress states would be unaffected by the additional hydrostatic pressure due to inertia because only deviatoric stress measures enter this criteria. Thus, a possible dynamic bifurcation from the uniform velocity field (1) differs from the quasistatic bifurcation because of the different constitutive coefficients or because other types of restrictions arise from the high velocity V_0 in the background motion (1).
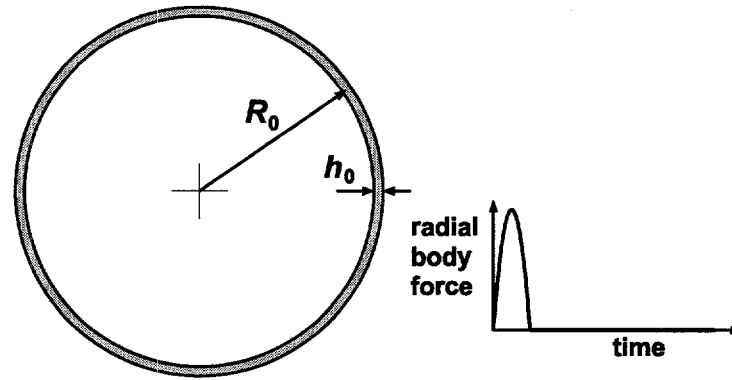
Figure 2: Initial configuration of the ring deformed in plane strain due to action of an impulsive radial body force with time history as indicated.

Obviously, the coefficients in the constitutive equations (4)–(6) reduce to the coefficients for the quasistatic case for rectangular blocks with a large aspect ratio because the effect of lateral inertia as given by (2) vanishes with the square of the aspect ratio of the block. It is noteworthy that the classification into elliptic, parabolic and hyperbolic regimes of the dynamic incremental field equations *without* the hydrostatic pressure due to lateral inertia is identical to the classification reported by Hill and Hutchinson [1] for the quasistatic case. The neck bifurcation modes are thus also relevant for the dynamic extension of a block with a large aspect ratio, as long as these modes are otherwise admitted by the dynamic nature of the problem.

NECK FORMATION IN DYNAMICALLY EXPANDING RINGS

In this section, results of simulation of steel rings expanding radially under plane strain conditions due to dynamic loading are summarized; see Figure 2. For all results reported, the rings have an initial radius of $R_0 = 0.07\,\text{m}$ and an initial thickness-to-radius ratio of $h_0/R_0 = 1/35, 1/70$ or $1/140$. The loading is a uniformly distributed radial body force of large amplitude and short duration. The time history of the body force follows a half sine curve as indicated in Figure 2. For all results reported here, the magnitude of the body force variation is $10^8\,\text{N/kg}$ and the duration is $10\,\mu\text{s}$. Momentum imparted to the material by this impulse drives the deformation thereafter due to material inertia. The wall thickness is assumed to have an imperfection in the form of a periodic variation in half–thickness $\Delta h_0(\theta) = \xi \cos 8\theta$ where θ is the angular distance along the ring. The resulting deformation is assumed to be periodic and only a $22.5°$ segment of the ring is analyzed.

The simulations were based on a dynamic finite element procedure (e.g., see Knoche and Needleman 1993) with an internal variable dilatant viscoplastic material model used to describe the ductile failure by void nucleation, growth and coalescence (cf. Gurson [12], Tvergaard [13]). In the present implementation of the Gurson model, only strain–controlled void nucleation was used. Also, thermal softening by adiabatic heating was taken into account.

The material data used in the simulations correspond to a ductile pressure vessel steel (denoted pvs in the following) which has been characterized by Naus et al. [14]. Some of the relevant data for the pvs steel ring as used here are the strain hardening exponent $N = 0.1$, the strain rate hardening exponent $m = 0.002$, the room temperature yield strength $\sigma_0 = 426\,\text{MPa}$, the material mass density $\rho = 7850\,\text{kg/m}^3$, the initial temperature $20°\text{C}$, the coefficient of thermal expansion $\beta = 1.1 \times 10^{-5}/°\text{C}$, the fraction of plastic work converted into heat $\chi = 0.9$, the elastic
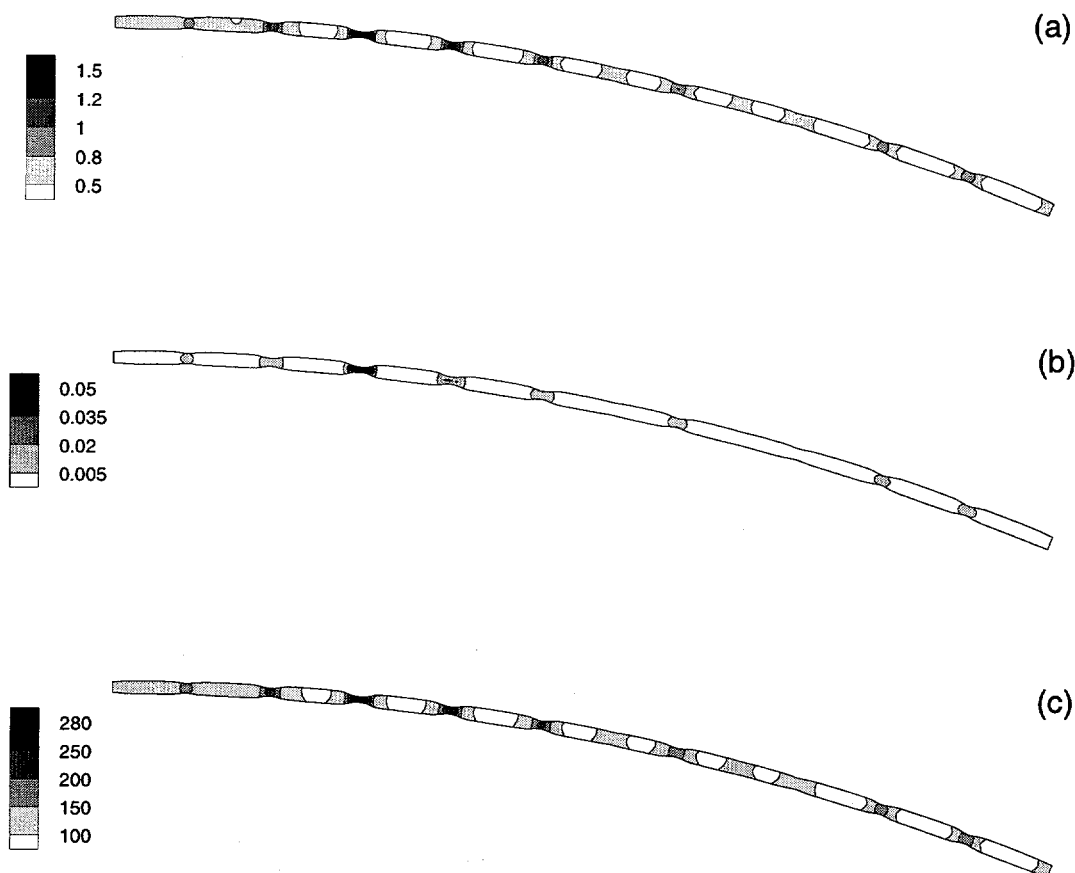
Figure 3: Level curves of (a) effective plastic strain, (b) void volume fraction, and (c) temperature in °C for a pvs–steel ring with $\frac{h_0}{R_0} = \frac{1}{70}$ deformed to $\frac{\Delta R}{R_0} = 0.65$. The ring is initially thinnest at the extreme left end and thickest at the right end in the figure ($\xi = 0.01$).

modulus $E = 206.9\,\mathrm{GPa}$, the Poisson ratio $\nu = 0.3$ and the volume fraction of strain–controlled void nucleating particles $f_v = 0.002$.

A study of dynamic necking in rings has been conducted recently by Han and Tvergaard [8] using a rate–independent plasticity model. The present computations have been carried out along the same lines; for details on boundary conditions the reader is referred to this study. However, the phenomenon of multiple necking observed in [8] for a rate–independent material and found here for a slightly rate–dependent material are very similar, suggesting that strain–rate–hardening effects play a secondary role compared with the imperfection sensitivity of the viscoplastic material in the process of neck formation.

Typically, the average strain rate in these calculations has increased to a value somewhat less than $10^4/\mathrm{s}$ at the time the impact loading on the ring ceases, and it then decays as the rate of expansion of the ring gradually diminishes. For the case of $h_0/R_0 = 1/70$, contour plots of several fields are shown in Figure 3 at a nominal hoop strain of 65%. It is evident that roughly periodic necking has emerged with spacing between necks being much less than the wavelength of the initial imperfection in wall thickness.

In order to visualize the evolution of the situation depicted in Figure 3, an average effective plastic strain rate over a cross–section, normalized by the average strain rate $\dot{R}/R$, provides a useful

instantaneous measure of the deviation from uniform deformation. Results in the form of contours of this normalized strain rate $\dot{e}_s(\theta, t)$ in the θ, t–plane are shown in Figure 4 for three impact simulations representing rings with different thickness–to–radius ratios. Viewed in this way, the darker (lighter) shading shows plastic strain rates which are higher (lower) than the average strain rate. For all cases, the strain rate distribution around the circumference is found to be nearly uniform at the time when the impact loading is completely removed, that is, after $10\mu s$.
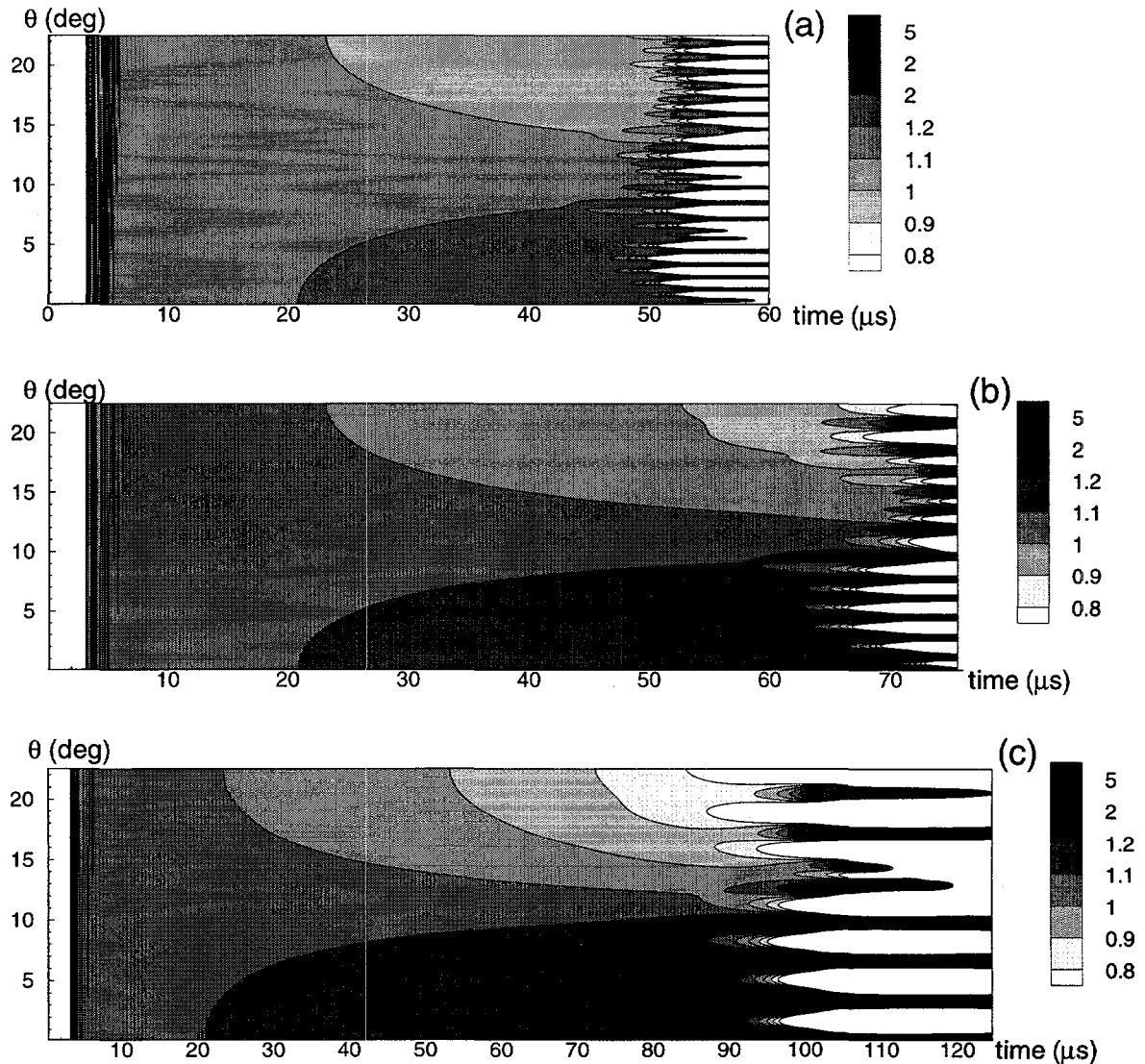

Figure 4: Level curves of normalized cross–sectional effective plastic strain rate $\dot{e}_s$ over the θ, t–plane for rings of the pvs–steel with the initial thickness–to–radius ratio being equal to (a) $\frac{1}{140}$, (b) $\frac{1}{70}$ and (c) $\frac{1}{35}$.

After the applied load is completely removed, inertia continues to drive the deformation which is nearly uniform for the three cases in Figure 4 until the time is approximately two times the impact time. In all cases, a slightly higher strain rate is found at the thinnest point of the ring shortly after $20\,\mu s$ and this higher strain rate region grows in size thereafter. Almost simultaneously, a strain rate slightly smaller than average is visible at the thicker part of the ring. Even with higher

196

resolution in the contour plots than can be represented here, this pattern seems to be common to all cases.

In the context of the bifurcation results for a quasistatic plane strain tension test, this situation can be described qualitatively as slow growth of the long wavelength mode which, in the quasistatic case, would lead to a neck developing symmetrically near $0°$ and extending to a few degrees to either side of this point. However, in addition to the slowly growing long–wavelength mode, the three cases in Figure 4 show rather different behavior from the quasistatic case in that multiple, closely spaced high/low strain rate zones are seen to develop relatively rapidly for the dynamic cases analyzed here. The order of appearance of these high/low strain rate zones is not the same for the three cases. Furthermore, these short wavelength, highly non–uniform plastic strain rate patterns do not all develop into necks with equal strain intensity. However, this mode appears to be the *critical mode* in the sense that some of these *neck–precursors* do develop into macroscopic necks that are visible in contour plots like those in Figure 3. Some of the neck–precursors corresponding to the critical mode shown Figure 4 die out and the strain levels in the necks in Figure 3 vary accordingly.

SUMMARY

Although the phenomenon of interest has been studied only over a limited range of system parameters, some noteworthy features or trends have emerged. Among these are:

(i) even though the eventual array of necks observed seemed to have little correlation with the original imperfection in wall thickness, it is noteworthy that the necks did not appear at all up to overall strains of unity if no imperfection was present. This suggests that some nonuniformity of the deformation is probably important in precipitating the distribution of necks.

(ii) strain-rate-sensitivity effects seem to play a secondary role compared with the imperfection sensitivity of the viscoplastic material in the process of neck formation.

(iii) the role of the dilatant plasticity model is of minor importance in the process. The constitutive features which essentially control the failure evolution accounted for here are apparently the adiabatic softening and the basic viscoplastic material response.

(iv) for the three cases studied, the spacing of the array of necks varied directly as the initial wall thickness of the ring. At the time of neck formation, the spacing is roughly three to four times the wall thickness.

(v) although most calculations were done for strain hardening parameter $N = 0.1$, a few cases with larger or smaller values were considered. A significant decrease in N results in similar neck formation but at an earlier time, while a signifiant increase in N suppresses formation of necks, particularly in the portion of the wall that is initial thicker.

ACKNOWLEDGMENT

REFERENCES

1. Hill, R. and Hutchinson, J. W. (1975), Bifurcation phenomena in the plane strain tension test, *Proc. Roy. Soc. (London)* **A23**, 239-264.

2. Young, N. J. B. (1976), Bifurcation phenomena in the plane strain compression test, *J. Mech. Phys. Solids* **24**, 77-91.

3. Needleman, A. (1979), Non–normality and bifurcation in plane strain tension and compression, *J. Mech. Phys. Solids* **27**, 231–254.

4. Benallal, A. and Tvergaard, V. (1995), Nonlocal continuum effects on bifurcation in the plane strain tension–compression test, *J. Mech. Phys. Solids* **43**, 741–770.

5. Hill, R. (1958), A general theory of uniqueness and stability in elastic–plastic solids, *J. Mech. Phys. Solids* **6**, 236–249.

6. Hill, R. (1961), Bifurcation and uniqueness in non–linear mechanics of continua, *Problems of Continuum Mechanics* **n**, 155–164.

7. Fressengeas, C. and Molinari, A. (1994), Fragmentation of rapidly stretching sheets, *Europ. J. Mech. A/Solids* **13**,251–268.

8. Han, J. and Tvergaard, V. (1995), Effect of inertia on the necking behavior of ring specimens under rapid radial expansion, *Europ. J. Mech. A/Solids* **14**, 287–307.

9. Needleman, A. (1991), The effect of inertia on neck development, *Topics in Plasticity* **n**, 151–160.

10. Tugcu, P., Neale, K. W. and Lahoud, A. E. (1990), Inertial effects on necking in tension, *Int. J. Solids Struc.* **11**, 1275–1285.

11. Hutchinson, J. W. (1974), Plastic buckling, in *Advances in Applied Mechanics*, edited by C. S. Yih, Academic Press, New York, **14**, 67–144.

12. Gurson, A. L. (1977), Continuum theory of ductile rupture by void nucleation and growth – Part I. Yield criteria and flow rules for porous ductile media, J. Engr. Matl. Tech., **99**, 2–15.

13. Tvergaard, V. (1990), Material failure by void growth to coalescence, in *Advances in Applied Mechanics*, edited by J. W. Hutchinson and T. Wu, Academic Press, New York, **27**, 83–151.

14. Naus, D. J., Keeney–Walker, J., Bass, B. R., Fields, R. J., DeWit, R. and Low III, S. R. (1989), High temperature crack–arrest behavior in 152–mm thick SEN wide plates of quenched and tempered A533 grade B class 1 steel, NUREG/CR–5330 (ORNL/TM–11083), Oak Ridge National Laboratory, National Institute of Standards and Technology.

MODELING OF HIGH HOMOLOGOUS TEMPERATURE DEFORMATION BEHAVIOR FOR STRESS AND LIFE-TIME ANALYSES

Erhard Krempl

Mechanics of Materials Laboratory
Rensselaer Polytechnic Institute
Troy, NY 12180-3590

ABSTRACT

Stress and life-time analyses need realistic and accurate constitutive models for the inelastic deformation behavior of engineering alloys at low and high temperatures. Conventional creep and plasticity models have fundamental difficulties in reproducing high homologous temperature behavior. To improve the modeling capabilities "unified" state variable theories were conceived. They consider all inelastic deformation rate-dependent and do not have separate repositories for creep and plasticity. The viscoplasticity theory based on overstress (VBO), one of the unified theories, is introduced and its properties are delineated. At high homologous temperature where secondary and tertiary creep are observed modeling is primarily accomplished by a static recovery term and a softening isotropic stress. At low temperatures creep is merely a manifestation of rate dependence. The primary creep modeled at low homologous temperature is due to the rate dependence of the flow law. The model is unaltered in the transition from low to high temperature except that the softening of the isotropic stress and the influence of the static recovery term increase with an increase of the temperature.

INTRODUCTION

Design stress analyses of highly loaded components are now performed using finite element programs. For most structural components linear elasticity is appropriate. However, when inelastic material behavior is experienced this model is not adequate. This can happen at low

homologous[1] temperature when the elastic limit is exceeded by an overload event or is allowed to be exceeded to ensure an economic use of the material. At high homologous temperatures a true elastic limit may not exist. Stress levels in the quasi elastic region may lead to significant creep deformation and even to creep rupture. For these and other cases material models that capture the inelastic deformation behavior are needed. Once developed, they need to be implemented in constitutive equation subroutines of finite element programs.

For the inelastic deformation behavior of metals and alloys rate(time)-independent plasticity and combined plasticity and creep are generally used at low and high homologous temperature, respectively. Plasticity and creep models were developed independently and were combined as soon as the finite element programs were ready to accept these nonlinear models and as soon as the technology demanded their use. This was the case in the 1970s when the consumer oriented society started to demand safety and predictability of performance of engineering structures. When the predictions of these models were compared with experimental results major deficiencies were detected in modeling of cyclic loading, of sequences of monotonic loading and of intermittent creep.

"Unified" models were then developed that consider all inelastic deformation as rate-dependent. This approach is in agreement with the notions of materials science where plastic deformation is considered to be a rate process. No separate repository for creep is introduced in the "unified" constitutive equations. Creep and relaxation are manifestations of rate-dependence and of diffusion processes when the boundary conditions in homogeneous states of deformation are zero stress rate and zero total strain rate, respectively.

The conventional plasticity-creep formulation assumes that the low homologous temperature deformation is rate-independent and that rate(time)-dependent deformation commences suddenly when the so-called creep range is reached. Then the model has to be changed from plasticity to combined plasticity and creep.

The separation of creep and plastic deformation models does not have a physical basis. Inelastic deformation is caused by changes in the defect structure such as dislocation motions for monotonic (increasing stress) and creep (constant stress) conditions.

Even if the loading on the boundary of engineering structures is kept constant in time, the constant stress condition generally does not exist inside a structure. The inhomogeneous stress field causes a redistribution of the stresses with time due to the stress distribution and a creep event with constant stress does generally not exist inside the structure.

The combined plasticity-creep formulation can also lead to contradictions. Creep tests performed at stress levels within the "quasi elastic" region of a stress-strain diagram can lead to significant creep deformation and even to creep rupture at high temperature. The idealization of plasticity requires that the yield surface encloses the purely elastic region as defined by the linearity of the stress-strain diagram. So, creep deformation and even creep rupture are taking place inside the yield surface!

"Unified" models were invented to circumvent these conceptual and technological problems.

The purpose of this paper is to introduce a "unified" model, the viscoplasticity theory based on overstress (VBO), and to show some of its qualitative properties for high and low homologous temperature.

---

[1] The dimensionless temperature ratio of test temperature over melting temperature both measured in °K. Engineering structures seldom operate at a homologous temperature greater than 0.6 and operating temperatures below 0.3 are usually considered low homologous temperature.

THE "UNIFIED" STATE VARIABLE THEORIES

General Remarks

"Unified" models including VBO are basically continuum models and assume a representative volume element exists. At this level the contributions of the many possible microstructural mechanisms are only recognized through their aggregate effects. In experiments, the specimen representing a macroscopically homogeneous state of deformation (tensile bar or thin-walled tube) is the representative volume element. It serves as an integrator of all mechanisms. The individual contributions are recognized in " smeared out" form as long as they are influencing the specimen behavior.

In VBO and other state variable theories the modeling of the changing microstructure is accomplished by the so-called state variables and their growth laws. There is no one-to-one correspondence between the state variables and certain micromechanisms. The correspondence is rather diffuse. In VBO the state variables are motivated by experimental results and some qualitative considerations, see Krempl [1]. Generally speaking, state variables for modeling work-hardening in monotonic loading, for cyclic hardening (softening) and for the Bauschinger effect are needed.

Inelastic deformation in metals and alloys is primarily affected by dislocation motion and by other changes in the defect structure. In most of the cases the dislocation density increases with inelastic deformation and further movement of dislocations is impeded by their increasing density. As a consequence work hardening is observed macroscopically. In some cases, when cold-worked metals are subjected to cyclic inelastic loading for example, cyclic softening occurs indicative of an easing in the passage of dislocations and a decrease in dislocation density.

At low homologous temperatures diffusion is negligible and the defect structure acquired during inelastic deformation is stable in the absence of mechanical loading. As the temperature increases diffusion processes become important. Defects can now change by "thermal action." Generally, diffusion tends to counteract the hardening effects of inelastic deformation. Hardening due to inelastic deformation and softening due to diffusion occur while external mechanical loads are applied. In the absence of external loads hardening essentially ceases and the effects of diffusion continue until equilibrium is attained[2]. The defect structure is observed to change with time, "static" recovery is said to occur. Subsequent loading shows that the hold periods with zero external load can lead to a softening of the response[3].

Hardening due to inelastic deformation and softening due to diffusion (static recovery) act simultaneously. Depending on the loading and on the temperature, hardening or softening may be pronounced. At low homologous temperatures hardening is dominant. Hardening and softening can also be in equilibrium as happens when the stress is constant and secondary creep can be observed. As the temperature increases diffusion effects become increasingly important and hardening ceases when the melting temperature is approached.

---

[2] Although there are no external mechanical loads acting on the material, residual internal stresses with zero resultant exist and they influence the change in the microstructure. These internal stresses are self equilibrating and do not enter into a continuum formulation.

[3] It is assumed that the material behavior is "normal" and that effects like strain aging and age hardening do not occur.

At low temperatures where diffusion is negligible the growth laws for the state variables are homogeneous of degree one in the rates so that neither time nor the rate of loading has an influence on hardening. The modeling of the diffusion processes in a continuum approach mostly follows the Orowan-Bailey format, via a static recovery term in the growth laws of the state variables. This term is introduced to counteract the hardening and it is not homogeneous of degree one in the rates. With it rate and time of loading have an influence on the growth of the state variables. When the material is unloaded to zero load and is left at zero load, the recovery terms return the state variables to zero with time. Majors and Krempl [2] have, however, shown that the static recovery term alone cannot model cyclic softening and tertiary creep.

Viscoplasticity Theory Based on Overstress (VBO)

Assuming isotropy, small strains, variable temperature and volume preserving inelastic deformation the deviatoric flow law can be written as

$$\dot{e} = \dot{e}^{el} + \dot{e}^{in} = \frac{d}{dt}\left(\frac{1+\nu}{E}s\right) + \frac{3}{2}F[\Gamma]\frac{s-g}{\Gamma} \tag{1}$$

where a bold face quantity denotes tensor and a superposed dot or $\frac{d}{dt}$ designates material time derivative; e and s denote the strain and stress deviators, respectively; g is the equilibrium stress deviator, $s - g$ is the overstress (effective stress) deviator and $\Gamma = \sqrt{\frac{3}{2}tr[(s-g)(s-g)]}$ is the overstress invariant or effective overstress. The elastic modulus is E and ν is the elastic Poisson's ratio. The function F has the dimension of 1/time and is positive, increasing with $F[0]=0$. Experimental evidence for introducing the overstress dependence of the inelastic strain rate is discussed by Krempl [1].

When all rates are zero then $s = g$ is a solution of Eq. (1); g therefore represents the stress that can be sustained at rest or in equilibrium and this property has given g its name.

From Eq.(1) the effective inelastic strain rate can be obtained as $\sqrt{\frac{2}{3}tr(\dot{e}^{in}\dot{e}^{in})} = \dot{p} = F[\Gamma]$.

The deviatoric relations have to be augmented with a statement that the response to a hydrostatic state of stress is elastic

$$tr\dot{\epsilon} = \frac{d}{dt}\left(\frac{1-2\nu}{E}tr\sigma\right) + 3\alpha\dot{T} \tag{2}$$

where σ and ϵ are the stress and strain tensors, respectively. The coefficient of thermal expansion is α and $T - T_0$ is the change from the reference temperature T_0. The writing of the first terms on the right hand side of Eqs. (1) and (2) ensure path independence of elastic deformation when the elastic constants are dependent on temperature, see Krempl [3].

The equilibrium stress enables the modeling of nonlinear rate sensitivity and the separation of viscous, work-hardening and rate-independent contributions to the flow stress, see Krempl [3]. Its growth law with a static recovery term is

$$\dot{g} = \frac{\psi[\Gamma]}{E}\dot{s} + s\frac{\partial}{\partial T}\left(\frac{\psi[\Gamma]}{E}\right) + F[\Gamma]\left(\frac{s-g}{\Gamma} - \frac{g-f}{A}\right)\psi[\Gamma] + \left(1 - \frac{\psi[\Gamma]}{E}\right)\dot{f} - R[\bar{g}]g \tag{3}$$

The positive decreasing function $\psi[\Gamma]$ with the dimension of stress controls the transition from the initial elastic behavior to the fully established inelastic flow. It is bounded by $E > \psi[\Gamma] > E_t$

202

where E_t is the tangent modulus in the inelastic range based on total strain rate. The second term on the right hand side, together with corresponding expressions in the flow law, ensures path independent elastic behavior.

The deviatoric quantity f is another state variable called the kinematic stress. It is the repository for modeling the Bauschinger effect. In addition work-hardening, work-softening or flow at constant stress in monotonic loading is reproduced depending on whether the slope based on inelastic strain rate $\hat{E}_t$ is positive, negative or zero. The slopes in the flow stress region are related by $\hat{E}_t = E_t/(1 - E_t/E)$ and are almost equal for most engineering alloys. The growth law for the kinematic stress is

$$\dot{f} = \hat{E}_t F[\Gamma] \frac{s - g}{\Gamma} \tag{4}$$

It is also possible to make $\hat{E}_t$ a function if there is a variable slope in the flow stress region.

The positive, increasing function $R[\bar{g}]$ with the dimension of reciprocal time constitutes the static recovery term. The equilibrium stress invariant is $\bar{g} = \sqrt{\frac{3}{2} tr(gg)}$ and $R[0] = 0$. The static recovery term reduces the growth of g and it is zero when the equilibrium stress is zero.

The scalar state variable A is called the isotropic stress. It is primarily the repository for modeling of cyclic hardening, cyclic softening and tertiary creep, see Majors and Krempl [2]. An appropriate growth law has to be specified.

At low homologous temperature A represents the rate-independent contribution to the flow stress, see Fig. 1 of Krempl [3]. As the temperature increases diffusion becomes important and the contribution to the flow stress represented by A depends on the duration of the test. It and with it the equilibrium stress will become rate-dependent.

To account for rate-independent cyclic hardening/softening at low homologous temperature the growth law must be formulated to be homogeneous of degree one in the rates. In this formulation time has no effect. Examples of modeling complex situations such as extra hardening in out-of-phase loading, see Krempl and Choi [4] and Choi and Krempl [5].

As the temperature increases a rate-dependent growth law with softening is appropriate to model tertiary creep, see Majors and Krempl [2], Tachibana and Krempl [6] and Maciucescu et al.[7]. The growth law used by Maciucescu et al. [7] is

$$\dot{A} = -\beta \left(\frac{A - A_2}{A_2} \right) p \tag{5}$$

where β is a constant with dimension stress/time and A_2 is the final value of A. If the initial condition is $A_1 > A_2$, softening will be modeled. This is appropriate for high homologous temperature. Hardening would be represented if $A_1 < A_2$. The quantity p is the inelastic strain path length defined above.

Modeling of Deformation Behavior

The above equations represent the three dimensional version of VBO. To ascertain its modeling capabilities the material constants E, E_t, ν, and α and the functions

$F[\Gamma]$, $\psi[\Gamma]$, and $R[\bar{g}]$[4] must be determined from appropriate tests. Included are monotonic and cyclic loading in displacement (strain) or load (stress) control, creep (stress is constant) and relaxation (displacement is constant). Such tests are usually performed in the uniaxial state of stress or, less frequently, in biaxial stress states. To start the determinations the three dimensional equations must be specialized for the appropriate stress state and the test conditions. Then values for the constants are assumed and the coupled, nonlinear, stiff differential equations are integrated numerically. The results are compared with the experiments. New constants are assumed and integration yields new results. This process is continued until a satisfactory match of simulation and experimental data is achieved. Once the constants have been determined the constitutive equations should be specialized for another set of tests which had not been used in the determination of the constants. Then the model is to be specialized for these tests to affect a prediction. If the comparison with experiment is satisfactory, the model can be used with confidence in stress analyses.

Reduction to the Uniaxial State of Stress

Using the definition of a deviator for the strain rate Eqs. (1) and (2) can be written as

$$\dot{\boldsymbol{\varepsilon}} = \frac{d}{dt}\left(\frac{1+v}{E}\boldsymbol{s} + \frac{1-2v}{3E}tr\boldsymbol{\sigma}\boldsymbol{I}\right) + \alpha\dot{T}\boldsymbol{I} + \frac{3}{2}F[\Gamma]\frac{\boldsymbol{s}-\boldsymbol{g}}{\Gamma} \tag{6}$$

where $\boldsymbol{I}$ is the identity matrix. For the uniaxial state of stress the stress deviator is given by

$$[\boldsymbol{s}] = \frac{\sigma}{3}\begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{7}$$

where σ is the uniaxial true stress. Assuming the same relation for the equilibrium stress with G denoting the counterpart of σ, the uniaxial component of the strain rate $\dot{\varepsilon}$ is found to be

$$\dot{\varepsilon} = \frac{d}{dt}\left(\frac{\sigma}{E}\right) + \alpha\dot{T} + F[\Gamma]\frac{\sigma-G}{\Gamma} \tag{8}$$

It is easy to convert the growth laws for the state variables to the uniaxial state of stress. They are

$$\dot{g} = \frac{\psi[\Gamma]}{E}\dot{\sigma} + \sigma\frac{\partial}{\partial T}\left(\frac{\psi[\Gamma]}{E}\right) + F[\Gamma]\left(\frac{\sigma-G}{\Gamma} - \frac{G-\bar{f}}{A}\right)\psi[\Gamma] + \left(1 - \frac{\psi[\Gamma]}{E}\right)\bar{f} - R[|G|]G \tag{9}$$

and

$$\dot{\bar{f}} = \hat{E}_t F[\Gamma]\frac{\sigma-G}{\Gamma} \tag{10}$$

where $\bar{f}$ is the counterpart of σ and G and where $\Gamma = |\sigma - G|$. The growth law for A does not change at all but p is now the accumulated inelastic strain.

Eqs. (5) and (8) through (10) are the uniaxial constitutive equations which must be specialized for the test conditions before they can be used.

With modern servo-controlled testing machines load or displacement can be imposed on the specimen. In most of the tests the rates are constant over certain time intervals and can change instantaneously. For example loading up to a creep stress level can be done with a certain stress or

---

[4] They in turn contain other constants. It is possible to use constants instead of functions thus reducing the number of constants needed, see Maciucescu et al. [7].

strain rate with a change to zero stress rate when the creep test starts. Either engineering or true quantities can be controlled.

It is useful to recall the relations between the true and the engineering quantities (designated by the same symbol but with a $\wedge$). With constant density assumed these relations are $\varepsilon = \ln(1+\hat{\varepsilon})$ and $\sigma = \hat{\sigma}(1+\hat{\varepsilon})$. To simulate the described tests the conditions listed in the last column of the table below must be substituted into the set of constitutive equations with an appropriate value for the constant rates.

Type of Control	Condition	Expression to be substituted
Strain control	$\dot{\hat{\varepsilon}}$ = constant	$\dot{\varepsilon} = \dot{\hat{\varepsilon}}/(1+\hat{\varepsilon}) = \dot{\hat{\varepsilon}}/\exp\varepsilon$
True strain control	$\dot{\varepsilon}$ = constant	$\dot{\varepsilon}$
Stress control	$\dot{\hat{\sigma}}$ = constant	$\dot{\sigma} = \dot{\hat{\sigma}}(1+\hat{\varepsilon}) + \hat{\sigma}\dot{\hat{\varepsilon}} = \dot{\hat{\sigma}}\exp\varepsilon + \sigma\dot{\varepsilon}$
True stress control	$\dot{\sigma}$ = constant	$\dot{\sigma}$

In addition the temperature history must be specified,

Note that for relaxation both the engineering and the true strain rates are zero. There is a difference between the " constant load" creep test (engineering stress rate is zero) and the "constant stress" creep test (true stress rate is zero.). It is known that the experimentally observed constant load and constant true stress creep curves can be different. Therefore, a possibility exists to model these differences with VBO.

Other test conditions such as a cyclic test require the specification of the stress or strain and temperature histories.

With these expressions the coupled nonlinear non-autonomous differential equations must be integrated numerically to simulate the specific tests. For a good model the curves obtained from the numerical test should be identical to the experimental one.

Without specification of the test condition the constitutive equations cannot be solved. The response of the model will depend on the temporal evolution of the input conditions.

Isothermal Creep with Constant True Stress

Loading up to the flow stress region is performed with a strain rate $\dot{\varepsilon}^-$. Then the true stress is kept constant. Applying the uniaxial equations without the recovery term immediately before and after the switch to constant stress enables the calculation of the initial creep rate and of the equilibrium stress rate. They are

$$\dot{\varepsilon}^+ = \dot{\varepsilon}^-(1 - E_t/E) \tag{11}$$

and

$$\dot{G}^+ = \dot{\varepsilon}^- E_t(1 - \psi/E) \tag{12}$$

where $^+$ designates the rate at the beginning of the creep test. The equilibrium stress rate before the start of the creep test is $\dot{G}^- = \dot{\varepsilon}^- E_t$ where we have made use of the long-term asymptotic solutions of VBO, see Krempl [3]. It is seen from Eqs. (11) and (12)that the rates at the beginning of the creep test are reduced from the values just before the test. It is also seen that the equilibrium stress continues to increase. Although it is believed that similar relations hold in the presence of the static recovery term it cannot be proven mathematically.

beginning of the creep test are reduced from the values just before the test. It is also seen that the equilibrium stress continues to increase. Although it is believed that similar relations hold in the presence of the static recovery term it cannot be proven mathematically.

Usually creep is classified as primary (creep rate decreases in magnitude), secondary (creep rate magnitude is constant) and tertiary (creep rate magnitude increases)[5]. To ascertain the properties of the model it is best to differentiate the flow law Eq. (8) which results in

$$\ddot{\varepsilon}^{in} = F'\big[|\sigma_o - G|\big]\big(-\dot{G}\big) = \ddot{\varepsilon}^{creep} \qquad (13)$$

The prime designates total derivative and the subscript $_o$ indicates that the stress is constant.

It can be seen that the sign of the second derivative is determined by the sign of the equilibrium stress rate. For primary creep $\dot{G} > 0$ and $\dot{G} < 0$ is required for tertiary creep. The equilibrium stress rate has to be zero for the modeling of secondary creep.

The purpose of the static recovery term and the softening of the isotropic stress is to bring about the changes in the slope of the equilibrium stress to represent secondary and/or tertiary creep at long times.

By substitution of the inelastic strain rates in the growth law for the equilibrium stress it can be shown that all the terms in the growth law for the equilibrium stress, except the static recovery term, are homogeneous of degree one in the rates, see Krempl [3]. Rate and elapsed time have no influence on their contribution to the growth of G. The influence of the static recovery term which reduces the growth of the equilibrium stress rate increases with test duration. Also the decreasing isotropic stress A reduces the equilibrium stress rate. Both are needed to model secondary and tertiary creep. Examples of modeling are given in Tachibana and Krempl [5] and Maciucescu et al. [7].

At low temperatures diffusion has no influence and the static recovery term is set to zero. It is not possible to reproduce tertiary creep as long as the model is required to reproduce work-hardening or inelastic flow at constant stress in a tensile test, see Krempl [3][6]. Only primary or at the most secondary creep can be modeled at low homologous temperature usually referred to as "Cold Creep." "Cold creep", relaxation and loading rate sensitivity are simply manifestations of rate-dependence under different boundary conditions. This type of behavior was found in engineering alloys at room temperature, see Krempl [8] and Kujawski and Krempl [9].

CONCLUSION

The qualitative properties of VBO in modeling creep behavior were described. The transition from low homologous temperature behavior to that at high homologous temperature is affected by the static recovery term and the softening of the isotropic stress which is a scalar state variable of VBO. They become increasingly important as temperature and duration increase. "Cold" as well as high temperature creep including tertiary creep can be modeled with one constitutive model. It is not necessary to switch to a different set of constitutive equations as temperature increases using this "unified" approach.

---

[5] We refer to magnitude to make the description valid for tension and compression.

[6] Tertiary creep can be modeled only if there is work-softening in monotonic loading.

ACKNOWLEDGEMENT

REFERENCES

1. Krempl, E., "The overstress dependence of the inelastic rate of deformation inferred from transient tests," Materials Science Research International, 1, 3-10 (1995).

2. Majors, P. S. and E. Krempl, "The isotropic viscoplasticity theory based on overstress applied to the modeling of modified 9Cr-1Mo steel at 538 °C," Materials Science and Engineering, A186, 23-34 (1994).

3. Krempl, E., "A small strain viscoplasticity theory based on overstress," in *Unified Constitutive Laws of Plastic Deformation*, A. S. Krausz and K. Krausz, editors, Academic Press, 281-318 (1996).

4. Krempl, E. and S. H. Choi, "Viscoplasticity theory based on overstress: The modeling of ratchetting and cyclic hardening of AISI Type 304 Stainless Steel," Nuclear Eng. and Design, 113, 401-410 (1992).

5. Choi, S. H. and E. Krempl, "Viscoplasticity theory based on overstress: The modeling of biaxial cyclic hardening using irreversible plastic strain," in *Advances in Multiaxial Fatigue, ASTM STP 1191*, D. L. McDowell and R. Ellis, editors, American Society for Testing and Materials, Philadelphia, PA, 259-272 (1993).

6. Tachibana, Y. and E. Krempl, "Modeling of creep and tensile behavior at high homologous temperature using the viscoplasticity theory based on overstress (VBO): Part I-Creep and tensile behavior," Journal of Engineering Materials and Technology, 117, 456-461 (1995).

7. Maciucescu, L., T.-L. Sham and E. Krempl, "VBO: A state variable constitutive equation for a solder alloy," to be published in the proceedings of INTERpack'97 to be held June 15-19, 1997, Island of Hawaii.

8. Krempl, E., "An experimental study of room-temperature rate-sensitivity, creep and relaxation of AISI Type 304 Stainless steel," J. Mech. Phys. Solids, 27, 363-385 (1979).

9. Kujawski, D. and E. Krempl, "The rate(time)-dependent behavior of Ti-7Al-2Cb-1Ta Titanium alloy at room temperature under quasi-static monotonic and cyclic loading," Journal of Applied Mechanics, 103, 55-63 (1981).