Int. J. Solids Structures Vol. 30, No. 21, pp. 2983–3012, 1993
Printed in Great Britain

0020–7683/93 $6.00+ .00

STRESS SINGULARITIES ALONG A CYCLOID ROUGH SURFACE

CHENG-HSIN CHIU and HUAJIAN GAO
Division of Applied Mechanics, Stanford University, Stanford, CA 94305, U.S.A.

(*Received* 18 *September* 1992; *in revised form* 15 *March* 1993)

Abstract—The stress concentration along a rough surface is of importance for understanding the nucleation of misfit dislocations and cracks in heteroepitaxial thin films and for general flaw initiation at material surfaces exposed to environmental corrosion. In order to seek a basic understanding on this issue, a cycloid wavy surface subject to a uniform bulk stress is adopted as a model problem. The elastic stress and displacement fields are determined using Muskhelishvili's conformal mapping method. It is shown that a cusped cycloid surface generates a crack-like singular stress field within a thin surface layer; under uniform tension this singularity is found to have identical strength to a row of periodic parallel cracks. The path-independent J-integral requires that the average surface strain energy density be identical to that in the bulk, implying that any stress magnification along a rough surface must be compensated by unloading along some complementary portions of the surface. Along the cusped cycloid surface, the strain energy distribution becomes Dirac singular at the cusp tips, the rest of the surface being completely stress-free. Thus the effect of cycloid cusps is to redistribute and concentrate all the surface strain energy per wavelength at a single (cusp) point, and because of that we can claim that the cycloid surface is the most efficient stress concentrator at a fixed wavelength. Even at a moderate bulk stress level this concentration may be sufficient to cause failure or nucleation of defects.

The full evolution of a rough surface under stress and other corrosion mechanisms must be solved by numerical methods. Our analytic solutions for a cycloid surface are of significant value for guiding numerical computations and gaining insights into essential features of the evolution process. From a global point of view, we show that a cusped cycloid surface becomes energetically favorable once the surface wavelength exceeds a critical value determined from the competition between the strain energy and the surface energy. From a local point of view, analysis of surface diffusion behaviors along an almost cusped cycloid surface indicates that the cusps are stable once they develop. The critical condition for formation of a cusped cycloid corresponds to the Griffith energy balance being exactly satisfied at the cusp tips while the chemical potential remains nearly constant along the rest of the surface. This implies spontaneous Griffith brittle fracture at tension cusps if plastic relaxation is not present to relieve the stress singularity.

INTRODUCTION

The main focus of this paper is to study the stress concentration along a rough surface which can be mathematically described by a cycloid with periodic cusps, as shown in Fig. 1(a) and in eqn (1) to follow. The major result to be reported is simple and yet somewhat surprising. It is found that the cusped cycloid surface of Fig. 1(a) is subjected to the same stress singularity and the same stress intensity factor as a row of periodic parallel cracks shown in Fig. 1(b). In other words, application of fracture mechanics would predict that these two structures should fail at the same stress level, even though there are no apparent "cracks" associated with the cycloid surface!

The present work has been primarily motivated by an effort to understand the surface nucleation of misfit dislocations in heteroepitaxial thin films which are often subjected to very large stresses, typically in the GPa range, due to lattice mismatch (Nix, 1989). In a series of recent papers, Gao (1991a, b, c) used first-order boundary perturbation techniques to analyse a stressed solid with slightly undulating surfaces and interfaces that differ moderately from straightness; he pointed out that even a slight surface undulation can generate significant stress concentration to cause deformation and fracture before the bulk stress reaches a critical level. On the other hand, the elastic strain energy stored in the solid has been identified as providing an intrinsic thermodynamic driving force that tends to destabilize an initially flat surface and thus to promote the development of surface roughness; this conclusion can be reached either by studying linearized kinetic equations along a slightly wavy surface (Srolovitz, 1989; Spencer *et al.*, 1991) or by showing that the strain energy is always reduced when an initially flat surface is slightly perturbed in an arbitrary manner

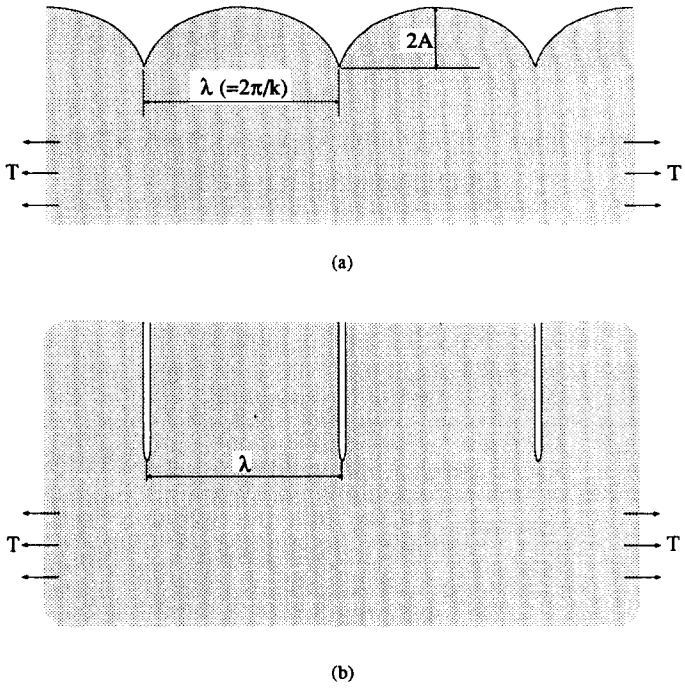

Fig. 1. Schematic diagrams of (a) a semi-infinite solid with a cusped cycloid surface and (b) an infinite solid with periodic parallel cracks.

(Gao, 1991b, c; Grinfeld, 1993). Comparing the destabilizing effect of elastic energy with the stabilizing force of surface energy at short wavelengths and that of gravitational energy at long wavelengths, Gao (1991b, c) concluded that once the bulk stress is sufficiently large, an initially flat surface becomes unstable against a range of perturbations bounded by two critical wavelengths; in highly stressed thin film structures, the most unstable wavelength for surface instabilities is very small so that the variation of gravitational energy becomes negligible and the instability can be simply regarded as a competition between the elastic energy and the surface energy.

The previous linearized stability studies provided a likely explanation for the onset of the island-like surface morphology, also known as the Stranski-Krastonov pattern (Matthews, 1975; Van der Merwe, 1979), during growth or annealing of a heteroepitaxial thin film. However, these studies relied on perturbation solutions of limited validity range, hence not suitable for describing the full range of surface evolution after instability has occurred. More accurate quantitative methods and solutions will be required to calculate stress distribution along a rough surface of large slope variations and to simulate the dynamic processes involved in the surface roughening. Solutions of analytic nature are especially valuable because they can be used to guide and verify numerical computations. For modelling purposes, the cusped cycloid surface shown in Fig. 1(a) appears to be particularly appealing because it closely resembles an island morphology and, more importantly, it admits simple analytic solutions, as will be shown in the next section.

STRESS ANALYSIS OF A CYCLOID SURFACE

The cycloid mapping function

The parametric function:

$$x = \xi + A \sin k\xi, \quad y = A \cos k\xi, \tag{1}$$

for $-\infty < \xi < \infty$, defines a periodic cycloid curve that can be used to describe a family of

two dimensional wavy surfaces having A as the amplitude and k as the wavenumber; the surface wavelength λ is given by $\lambda = 2\pi/k$. For small Ak, the cycloid behaves like a cosine curve $y = A\cos kx$ but when $Ak \to 1$, it forms an island morphology with periodic cusps at $x = (2n+1)\pi/k$ and $y = -A$ (n = integer). Figure 2 plots the cycloid shapes for $Ak = 0.3$, 0.7 and 1.0. When $Ak > 1$, the cycloid curve loses physical significance due to self-crossing.

In this paper we shall only consider cycloid surfaces that are subjected to a uniform lateral bulk stress of magnitude T, as in Fig. 1(a). In heteroepitaxial thin films this corresponds to the stress induced in the film by lattice mismatch. The solution procedure involves use of the Muskhelishvili (1953) complex variable method with the mapping function:

$$z = \omega(\zeta) = \zeta + iA\exp(-ik\zeta), \tag{2}$$

where $i = \sqrt{-1}$, $z = x+iy$ and $\zeta = \xi+i\eta$. The above function transforms the cycloid surface of eqn (1) in the z-plane into a perfectly straight line along the ξ-axis in the ζ-plane and maps, accordingly, the physical domain beneath the surface conformally into the lower half of the ζ-plane, i.e. $\eta < 0$.

At this point, it is not without interest to note that the cycloid is a classic geometric curve that has been studied in the context of calculus of variation by Bernoulli, Euler, Newton, L'Hospital, etc. (Gelfand and Fomin, 1963). The cycloid curve is generated by rolling a circle of radius A along a straight line, which can be viewed as a special case of a more general class of curves known as the hypocycloids that are generated by rolling a small circle along the inside rim of a large circle. Similarly, by rolling a circle along the outside rim of another circle, one obtains yet another class known as the epicycloids. These geometries have received limited attention in the past, particularly by some of the early Russian researchers in the 1940s. For example, a noncusped hypocycloid hole was studied by Shapiro in 1941 (Muskhelishvili, 1953; Savin, 1961). Unfortunately, the formation of cusp-like stress singularities due to surface diffusion, as we discuss in this paper, appears to have been overlooked and perhaps deliberately avoided by the early Russian researchers who were not equipped with the modern fracture mechanics concepts on stress singularities. The cycloid mapping function in eqn (2) can be obtained as a special case of the hypocycloid mapping (by letting the radius of the larger circle approach infinity) or epicycloid mapping functions discussed by Muskhelishvili (1953). The hypocycloid and epicycloid cusps and their formation will be studied in a later paper.

The complex potentials

For two dimensional elasticity problems, following Muskhelishvili (1953), the displacements and stresses can be represented by two complex potentials as:

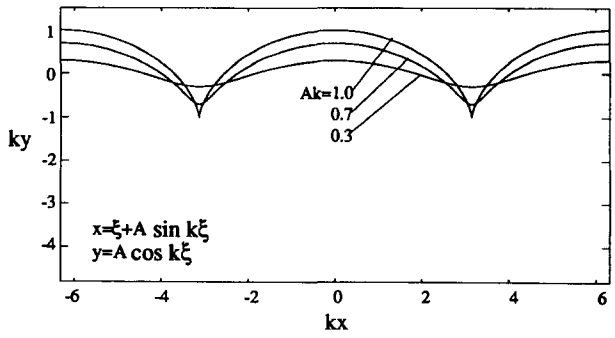

Fig. 2. Cycloid surfaces with $Ak = 0.3$, 0.7 and 1.0.

$$2\mu(u_x+iu_y) = (3-4\nu)\phi(z)-z\overline{\phi'(z)}-\overline{\psi(z)}, \tag{3}$$

$$\sigma_{xx}+\sigma_{yy} = 2[\phi'(z)+\overline{\phi'(z)}], \tag{4}$$

$$\sigma_{yy}-\sigma_{xx}+2i\sigma_{xy} = 2[\bar{z}\phi''(z)+\psi'(z)], \tag{5}$$

where μ denotes the shear modulus and ν the Poisson ratio. For simplicity, we assume plane strain conditions here and throughout this paper. One may obtain the solutions for plane stress by replacing ν in the results for plane strain with $\nu/(1+\nu)$. The resultant force, $f = f_x+if_y$, acting on a segment from a fixed point to z is given by:

$$\phi(z)+z\overline{\phi'(z)}+\overline{\psi(z)} = if+\text{constant}, \tag{6}$$

where the constant can be taken arbitrarily without affecting the stress state.

Complex potentials in the mapping coordinates

The cycloid mapping function, eqn (2), is equivalent to adopting a set of curvilinear ξ- and η-coordinates to describe the elastic field; these are shown in Fig. 3 for $Ak = 1$. It will be helpful to express the stresses and displacements in terms of the mapping coordinates ξ and η. The angle α between the x-axis and the ξ-axis is:

$$e^{i\alpha} = \frac{\omega'(\zeta)}{|\omega'(\zeta)|}, \quad e^{i2\alpha} = \frac{\omega'(\zeta)}{\overline{\omega'(\zeta)}}. \tag{7}$$

For convenience the same symbols for the complex potentials will be used in the ζ-plane, i.e. $\phi[\omega(\zeta)] = \phi(\zeta)$ and $\psi[\omega(\zeta)] = \psi(\zeta)$. With the following convention:

$$\frac{d\phi(z)}{dz} = \frac{1}{\omega'(\zeta)}\frac{d\phi(\zeta)}{d\zeta} = \Phi(\zeta), \quad \frac{d\psi(z)}{dz} = \frac{1}{\omega'(\zeta)}\frac{d\psi(\zeta)}{d\zeta} = \Psi(\zeta), \tag{8}$$

one can write the stress field as:

$$\sigma_{\xi\xi}+\sigma_{\eta\eta} = 2[\Phi(\zeta)+\overline{\Phi(\zeta)}], \tag{9}$$

$$\sigma_{\eta\eta}-\sigma_{\xi\xi}+2i\sigma_{\xi\eta} = \frac{2}{\overline{\omega'(\zeta)}}[\overline{\omega(\zeta)}\Phi'(\zeta)+\omega'(\zeta)\Psi(\zeta)]. \tag{10}$$

The displacement field and the resultant force are obtained from:

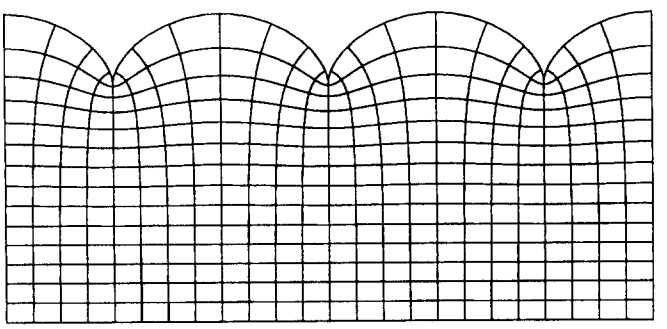

Fig. 3. The curvilinear ξ–η coordinates for a cusped cycloid surface.

$$2\mu(u_x+iu_y) = (3-4\nu)\phi(\zeta)-\omega(\zeta)\overline{\Phi(\zeta)}-\overline{\psi(\zeta)}, \tag{11}$$

$$\phi(\zeta)+\omega(\zeta)\overline{\Phi(\zeta)}+\overline{\psi(\zeta)} = if+\text{constant}. \tag{12}$$

The final solution

The cycloid surface problem can be solved analytically using the conformal mapping method of Muskhelishvili (1953). Let the solution be written in the form:

$$\phi(\zeta) = \phi_\infty(\zeta)+\phi_s(\zeta), \quad \psi(\zeta) = \psi_\infty(\zeta)+\psi_s(\zeta), \tag{13}$$

where ϕ_∞ and ψ_∞ are the potentials of the uniform stress state T,

$$\phi_\infty = \frac{Tz}{4} = \frac{T}{4}[\zeta+iA\exp(-ik\zeta)], \tag{14}$$

$$\psi_\infty = -\frac{Tz}{2} = -\frac{T}{2}[\zeta+iA\exp(-ik\zeta)]. \tag{15}$$

The functions ϕ_s and ψ_s, which represent "disturbances" due to the rough surface, are analytic in the lower half ζ-plane and will be determined from the traction-free boundary condition.

There are two alternative ways to express the boundary condition. One requires that $\sigma_{\eta\eta}+i\sigma_{\xi\eta}$ vanish on the surface, which by eqns (9), (10) and (13)–(15) result in:

$$\Phi_s(\xi)+\overline{\Phi}_s(\xi)+\frac{\overline{\omega}(\xi)}{\overline{\omega'}(\xi)}\Phi_s'(\xi)+\frac{\omega'(\xi)}{\overline{\omega'}(\xi)}\Psi_s(\xi) = \frac{TAk}{2}\frac{\exp(-ik\xi)-\exp(ik\xi)}{1+Ak\exp(ik\xi)}. \tag{16}$$

This type of boundary condition can also account for surface tractions. Another approach, which appears to be more convenient for the problem at hand, is to let the resultant force f vanish along the boundary. Substituting eqns (13)–(15) into (12), letting $\zeta = \xi$, and ignoring an arbitrary constant, one obtains:

$$\phi_s(\xi)+\omega(\xi)\overline{\Phi}_s(\xi)+\overline{\psi}_s(\xi) = -\frac{iTA}{2}[\exp(-ik\xi)+\exp(ik\xi)]. \tag{17}$$

These two boundary equations are equivalent in that each can be used to deduce the other by integration or differentiation.

The original elasticity problem is now reduced to one of solving for functions $\phi_s(\zeta)$ and $\psi_s(\zeta)$ which are analytic in the lower half ζ-plane and satisfy the boundary condition (17). Equating the functions which are analytic in $\text{Im}[\zeta] \leqslant 0$ on both sides of eqn (17), and then equating those analytic in $\text{Im}[\zeta] \geqslant 0$ yield the solution,

$$\phi_s(\zeta) = -\frac{iTA}{2}\exp(-ik\zeta), \tag{18}$$

$$\psi_s(\zeta) = \frac{iTA}{2}\left[\exp(-ik\zeta)-\frac{Ak+ik\zeta\exp(-ik\zeta)}{1+Ak\exp(-ik\zeta)}\right]. \tag{19}$$

In the above procedure, a subtle point worth mentioning concerns the term $\omega(\zeta)\overline{\Phi}_s(\zeta)|_{\zeta=\xi}$ in eqn (17), in which $\omega(\zeta)$ is analytic in $\text{Im}[\zeta] \leqslant 0$ but $\overline{\Phi}_s(\zeta)$ is analytic in $\text{Im}[\zeta] \geqslant 0$. The right hand side of eqn (17) consists of two terms, one analytic in $\text{Im}[\zeta] \leqslant 0$ and the other analytic in $\text{Im}[\zeta] \geqslant 0$, which suggests that $\omega(\zeta)\overline{\Phi}_s(\zeta)$ should either have a clearly defined analytic region or be able to be split into functions which have this property.

One may thus proceed by assuming $\omega(\zeta)\overline{\Phi}_s(\zeta)$ to be analytic in $\mathrm{Im}\,[\zeta] \geqslant 0$, and then verify this assumption from the solutions obtained.

The final solution to the cycloid surface problem is:

$$\phi(\zeta) = \frac{T}{4}[\zeta - iA\exp(-ik\zeta)], \tag{20}$$

$$\psi(\zeta) = -\frac{T}{2}\left[\zeta + iA\frac{Ak + ik\zeta\exp(-ik\zeta)}{1 + Ak\exp(-ik\zeta)}\right]. \tag{21}$$

Substituting these expressions into eqns (3)–(5), one finds that the stresses and displacements can be written more explicitly as:

$$\sigma_{xx} + \sigma_{yy} = T + 4\,\mathrm{Re}\,[\Phi_s(\zeta)], \tag{22}$$

$$\sigma_{yy} - \sigma_{xx} + 2i\sigma_{xy} = -T + 2\left[\frac{\bar{\zeta} - iA\bar{P}}{1 + AkP}\Phi_s'(\zeta) + \Psi_s(\zeta)\right], \tag{23}$$

$$u_x + iu_y = \frac{T}{8\mu}\left[(3 - 4\nu)(\zeta - iAP) + \frac{2(\bar{\zeta} - iA^2k) - (\zeta + iAP)(1 - Ak\bar{P})}{1 + Ak\bar{P}}\right], \tag{24}$$

where

$$P = \exp(-ik\zeta), \quad \bar{P} = \exp(ik\bar{\zeta}), \tag{25}$$

$$\Phi_s(\zeta) = -\frac{TAk}{2}\frac{P}{1 + AkP}, \quad \Phi_s'(\zeta) = \frac{iTAk^2P}{2(1 + AkP)^2}, \tag{26}$$

$$\Psi_s(\zeta) = \frac{TAk}{2}\left[\frac{P}{1 + AkP} + \frac{Ak + P}{(1 + AkP)^2} - \frac{Ak + ik\zeta P}{(1 + AkP)^3}\right]. \tag{27}$$

An arbitrary rigid body motion can be superimposed on the displacement field without affecting the stresses. Equations (22)–(24) show that the disturbance due to the surface roughness attenuates exponentially in the depthwise direction (i.e. the negative y-direction) with a characteristic length of $1/k = \lambda/2\pi$. In other words, only within a small boundary layer beneath the surface is the elastic field significantly affected by the surface roughness; away from the layer the uniform stress field is quickly recovered. For example, the stress field is found to be within 1% of the uniform state T at a depth of one wavelength ($\lambda = 2\pi/k$) measured from a cycloid valley for all cycloid amplitudes, including the cusp case $A = 1/k$. Similar observations have been made earlier for slightly undulating surfaces by Gao (1991a) using perturbation methods.

Wang *et al.* (1991)† recently studied the problem of a concentrated point force acting on a cycloid surface. In comparison, our problem involves a laterally applied uniform stress, which is algebraically simpler than the point force problem. The solutions by Wang *et al.* apparently do not have the correct asymptotic behavior near the point force. These authors concluded in their paper "At near field the stress concentrates in the order $1/r^2$, where r is the distance from the load". This is wrong because no matter what the shape of the surface is, the asymptotic stress field very close to a point force must behave as $1/r$, a fundamental feature of elastic Green's functions. Our preliminary investigation seems to suggest that the correct solution for the point force problem should involve an infinite series, a feature not existing in Wang *et al.*'s solutions. We suspect that their solution contains a higher order singularity such as a force dipole at the force location.

† We are grateful to Prof. M. T. Hanson of University of Kentucky for pointing out this reference to us.

The stress and displacement fields near a cycloid surface

The stress and displacement fields for the cycloid surface problem have been determined explicitly in eqns (22)–(24) in terms of the mapping variable ζ. Let us examine some basic features of these solutions. First, the hoop stress $\sigma_{\xi\xi}$ along the surface can be obtained directly from eqns (22) and (26) (the traction components $\sigma_{\eta\eta}$ and $\sigma_{\xi\eta}$ vanish), giving:

$$\sigma_{\xi\xi} = \frac{T(1-A^2k^2)}{1+2Ak\cos k\xi + A^2k^2}. \tag{28}$$

The maximum stress concentration occurs at a valley and the minimum occurs at a peak, with magnitudes:

$$(\sigma_{\xi\xi})_{\max} = T(1+Ak)/(1-Ak), \quad (\sigma_{\xi\xi})_{\min} = T(1-Ak)/(1+Ak). \tag{29}$$

As $Ak \to 1$, the cycloid surface develops periodic cusps where the hoop stress diverges to infinity, signaling a stress singularity. The nature of this singularity will be discussed in the next section.

For a better understanding of the stress distribution near a cycloid surface, the stress and deformation fields are displayed in Figs 4–6 for three representative cases: $Ak = 0.3$, 0.7 and 0.99. Part (a) of these figures plots the contours of the maximum principal stress $\sigma_{\max}$ (solid lines) and those of the maximum shear stress $\tau_{\max}$ (dashed lines). Due to the symmetry of the problem, only one half period of the cycloid is displayed. The stresses are calculated assuming plane strain conditions and the value of each contour is normalized by the reference value in the uniform stress state, which is T for $\sigma_{\max}$ and $T/2$ for $\tau_{\max}$. In all cases investigated, there are three distinct zones for $\sigma_{\max}$ and $\tau_{\max}$, which are divided by two contours of unit value (i.e., no stress change). The first zone, lying between the surface and the first unity-contour, is a region of stress relaxation for both $\sigma_{\max}$ and $\tau_{\max}$. In this region, $\sigma_{\max}$ and $\tau_{\max}$ attain their minimum values at the cycloid peak, and then increase gradually away from the surface. Comparison of Figs 4–6(a) indicates that the location of each stress contour remains almost unchanged as the wave-amplitude A increases. In other words, the augmented part of the solid due to the increase in A does not seem to affect the original stress distribution much at distances away from the valley; the stresses in the augmented part continue to decrease until they completely vanish as $Ak \to 1$. The relaxation zone of the shear stress is smaller than that of the principal stress.

The second region, between the two unity-contours, is a region of stress magnification. The stresses reach the maximum at the valley and attenuate exponentially away from there. For the three cases considered in Figs 4–6(a), the maximum stress concentration factor, $\sigma_{\max}/T$ or $2\tau_{\max}/T$, is of magnitude 1.86, 5.67 and 199, respectively. Although $\sigma_{\max}$ and $\tau_{\max}$ increase rapidly with A, the location and magnitude of each stress contour do not change much, so that zones of high stress concentration are localized to the close vicinity of the valley which becomes a cusp as $Ak \to 1$. The domain of significant shear stress concentration is smaller than that of the maximum principal stress.

The third region is found below the second unity-contour where a weak stress relaxation is observed. In this region, both $\sigma_{\max}$ and $\tau_{\max}$ can be treated as being uniform for any practical purposes.

Part (b) of Figs 4–6 compares the deformed and undeformed configurations of a rectangular mesh near the cycloid surface. These figures are drawn in realistic aspect ratios with a 5% bulk strain. The displacements are assumed to be zero at the valley and plane strain is assumed in the calculation. Observe that for the case $Ak = 0.99$ the surface segments are almost unstretched but are subjected to rigid body rotations.

STRESS SINGULARITY AT A CYCLOID CUSP: THE CASE $Ak = 1$

From the point of view of mechanical failure, the limiting case $Ak = 1$ shown in Fig. 1(a) is of special interest because a singular stress field is induced by the cycloid cusps along the surface.

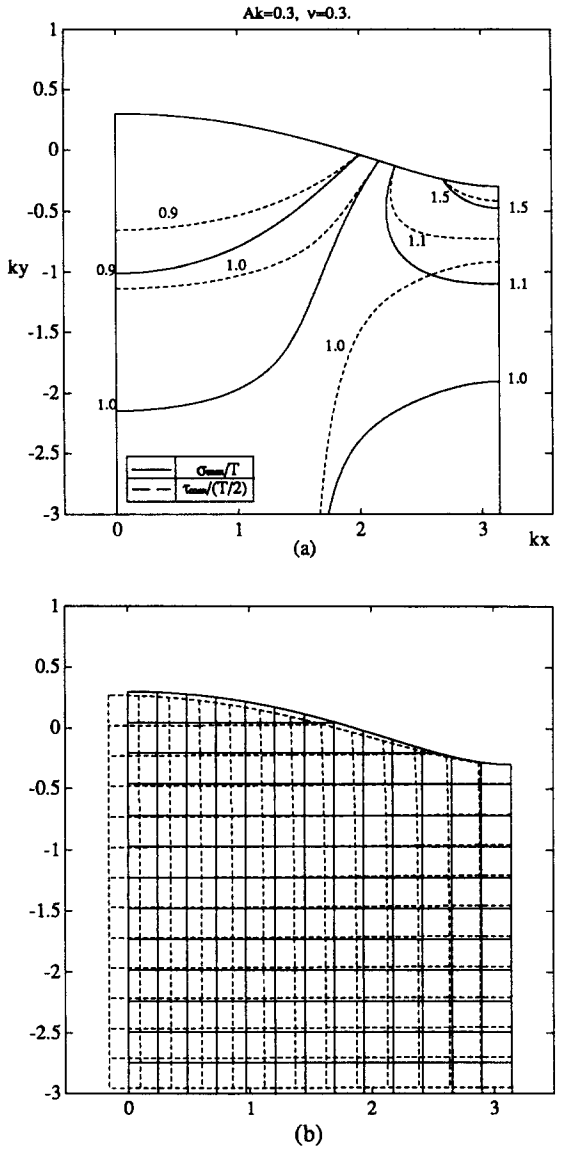

Fig. 4. For a cycloid surface with $Ak = 0.3$ (a) the contour plots of the maximum principal stress σ_{max} and the maximum shear stress τ_{max}, (b) undeformed (solid lines) and deformed (dashed lines) meshes in the solid under 5% bulk strain.

Depth-variation of the normal stress σ_{xx}

To understand the nature of the stress singularity at a cycloid cusp, let us first examine the depth-variation of the normal stress σ_{xx} with distance from a cusp tip.

One may verify from eqn (2) that the cycloid cusps are located along vertical lines $x = \xi = (2n+1)\pi/k = (2n+1)\lambda/2$ (n = integer) in both z- and ζ-planes. Substituting $\zeta = \pi/k + i\eta$ into eqns (22) and (23), and taking $Ak = 1$, the depth-variation of the σ_{xx} below a cusp tip is determined as:

$$\sigma_{xx}(\pi/k, \eta) = \frac{T}{2} + \frac{T}{2(1-e^{k\eta})^3}\{1+2(1+k\eta)\,e^{k\eta} - 5\,e^{2k\eta} + 2\,e^{3k\eta}\}, \tag{30}$$

where the mapping variable η is related to the physical coordinate y, and to the distance d measured from the cusp tip, by:

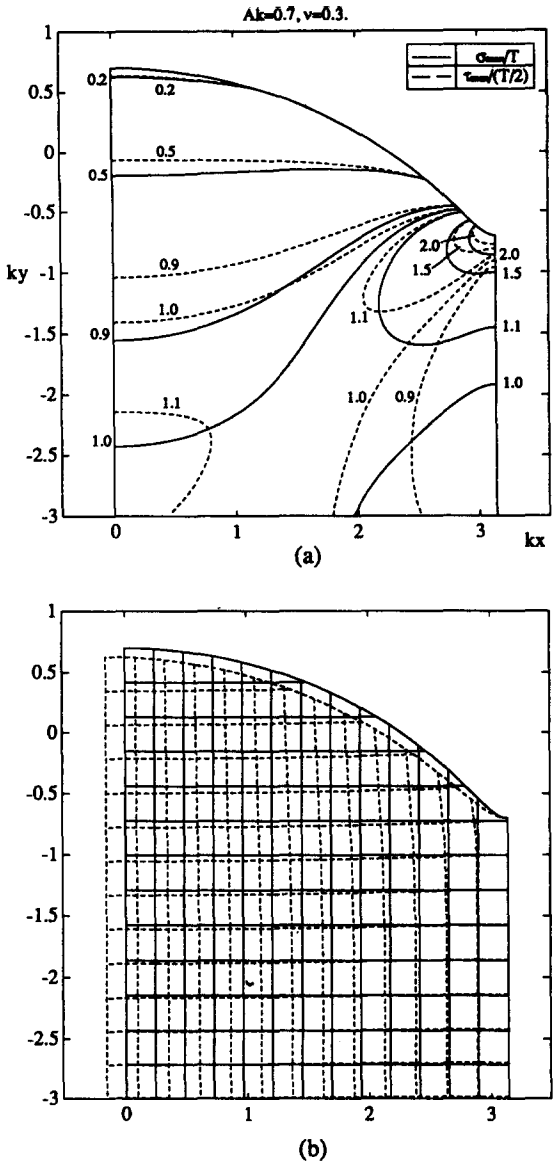

Fig. 5. For a cycloid surface with $Ak = 0.7$ (a) the contour plots of the maximum principal stress σ_{max} and the maximum shear stress τ_{max}, (b) undeformed (solid lines) and deformed (dashed lines) meshes in the solid under 5% bulk strain.

$$y = \eta - e^{k\eta}/k, \quad d = -1/k - \eta + e^{k\eta}/k. \tag{31}$$

From eqns (30) and (31), it can be shown that, as $kd \to 0$, σ_{xx} behaves asymptotically as:

$$\sigma_{xx} \sim -\frac{T}{k\eta} = \frac{T}{\sqrt{2kd}} = \frac{T\sqrt{\lambda/2}}{\sqrt{2\pi d}}. \tag{32}$$

Thus the stress field near a cusp tip is crack-singular. The structure of this singular field is mode I with stress intensity factor:

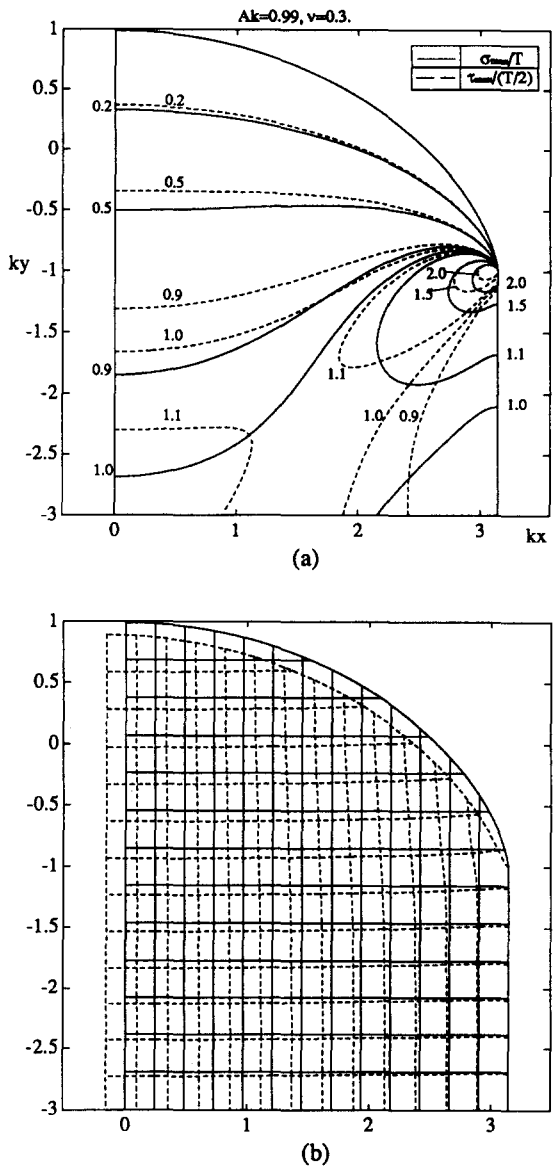

Fig. 6. For a cycloid surface with $Ak = 0.99$ (a) the contour plots of the maximum principal stress σ_{max} and the maximum shear stress τ_{max}, (b) undeformed (solid lines) and deformed (dashed lines) meshes in the solid under 5% bulk strain.

$$K_I = T\sqrt{\lambda/2}, \tag{33}$$

which is identical to that of a row of parallel cracks shown in Fig. 1(b) (Tada *et al.*, 1985).

Therefore, under uniform tension a cusped cycloid surface is subjected to the same stress singularity and stress intensity factor as an infinite row of periodic parallel cracks. Application of fracture mechanics would predict that these two structures fail at the same stress level, even though there may be no apparent "cracks" associated with the cusped rough surface! Strictly speaking, this equivalence is restricted to the uniform tension case. A compressive bulk stress T will not induce a stress intensity factor for the periodic cracks but can induce a negative mode I stress intensity factor for the cusps since the material surfaces near a cusp are not in contact, at least not as complete as the crack surfaces. For other loading cases such as the antiplane shear case considered in Appendix A, the stress

field near a cycloid cusp is still crack-singular; however, the stress intensity factor deviates from the corresponding periodic crack case.

As discussed before, the stress concentration attenuates exponentially with distance away from a cycloid cusp. For $Ak = 1$ and $kd \gg 1$, it can be shown that:

$$\sigma_{xx} \sim T[1 - kd \exp(-kd)]. \tag{34}$$

Figure 7 plots the depthwise variation of σ_{xx} with distance from the cusp (solid line), together with the stress variation near the periodic cracks (dashed line); the asymptotic singular field $\sigma_{xx} = K_I/\sqrt{2\pi d}$ is shown as a dotted line for comparison.

Although the stress field associated with the periodic cracks has been studied by Koiter (1961) and others, the solutions involve sophisticated integral transforms which require a considerable numerical effort in extracting the complete stress field. We are currently devising boundary-integral based procedures for numerically simulating the evolution of a rough surface. Thus to our advantage, a numerical scheme based on the method of Erdorgan *et al.* (1973) on singular integral equations and that of Fleck (1991) on periodic crack problems was adopted in our calculation. Due to length restriction, numerical details are ignored in this paper. It suffices to say that the essence of our procedure is to distribute a continuous array of dislocations along the crack contour or a rough surface contour so that the traction there is completely eliminated. Periodic cracks of finite length $L = 48/k$ and $L = 60/k$ are used to simulate the semi-infinite crack array. The calculated stress intensity factor is within $10^{-3}\%$ of the exact solution. The depthwise distribution of σ_{xx} for the cycloid surface is found to be close to that for the periodic cracks, with the largest relative difference being about 2% at $kd = 0.37$ (Fig. 7). It is also noted that, once kd exceeds 0.26, σ_{xx} deviates more than 10% from the asymptotic K-field. In other words, the 90% K-dominance zone is $0.26/2\pi$, or about 4%, of the spacing λ between cusps, suggesting a small zone of stress singularity, which is consistent with the behavior observed earlier in Figs. 4–6.

The result plotted in Fig. 7 suggested that the normal stress distribution directly beneath a cycloid cusp is practically the same as that beneath a crack tip in the periodic crack case. A further question is whether such corrrespondence could be established everywhere in the cusped solid. For that purpose, we calculated the full stress field for the periodic crack

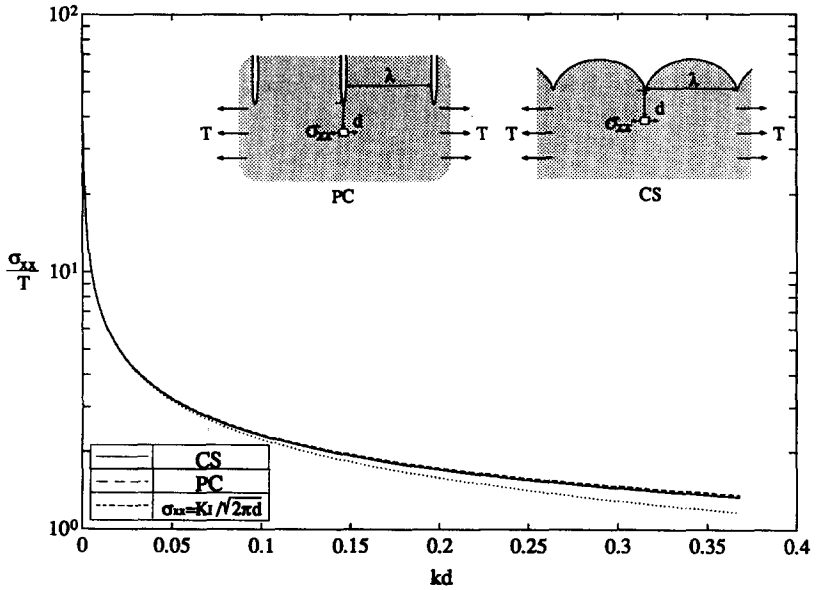

Fig. 7. The variation of the normal stress σ_{xx} beneath a cycloid cusp and a periodic crack tip, the asymptotic singular field for these variations.

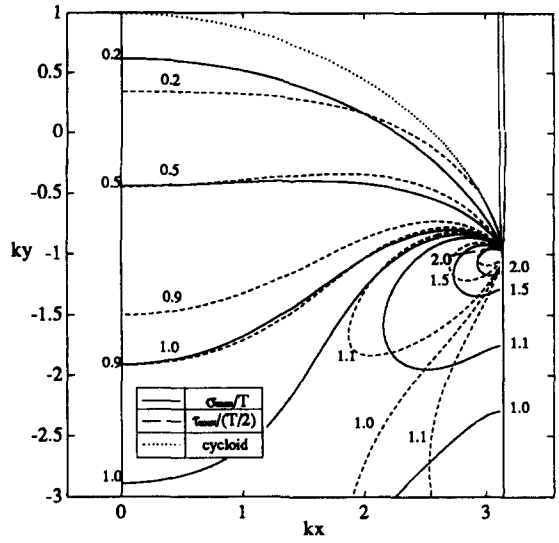

Fig. 8. The contour plots of the maximum principal stress σ_{max} and the maximum shear stress τ_{max} in the periodic crack case.

case. Figure 8 plots the contours of the maximum principal stress and the maximum shear stress within one half period of the periodically cracked body. These stress contours are close, but not identical, to the cycloid stress contours in Fig. 6(a). We also found that the strain energy density along the cycloid contour line in the periodically cracked body (shown as a dotted line in Fig. 8) is very small, within 2% to 3% of the bulk strain energy density w_0, suggesting that the cycloid line is indeed almost stress-free. These results lead to the conclusion that the full stress field induced by a cusped cycloid surface is close, although not identical, to that in the periodic crack case.

Proof of the asymptotic crack-tip field near a cycloid cusp

It can be directly verified that the singular stress field near a cycloid cusp has an identical structure to the standard crack-tip field, with a universal character that the angular variation of the stress field is independent of the global loading and geometry. The stress singularity is completely determined once the stress intensity factors are given.

Without loss of generality, let us consider a cycloid cusp located at $x = \lambda/2$ and $y = -A$, corresponding to $\xi = \lambda/2$ and $\eta = 0$ in the ζ-plane. Set local coordinates:

$$z_c = z-(\lambda/2-iA), \quad \zeta_c = \zeta-\lambda/2, \tag{35}$$

at the cusp tip. Substituting $Ak = 1$ into the mapping function of eqn (2), one may show that:

$$z_c = ik\zeta_c^2/2+0(\zeta_c^3). \tag{36}$$

According to Muskhelishvili (1953), the complex potentials in the local coordinate frame, $\phi_c(\zeta_c)$ and $\psi_c(\zeta_c)$, are related to those in global coordinates, $\phi(\zeta)$ and $\psi(\zeta)$, as:

$$\phi_c(\zeta_c) = \phi(\zeta_c+\lambda/2), \quad \psi_c(\zeta_c) = \psi(\zeta_c+\lambda/2)+(\lambda/2+iA)\frac{\phi'(\zeta_c+\lambda/2)}{\omega'(\zeta_c+\lambda/2)}. \tag{37}$$

Substituting eqns (20) and (21) for $Ak = 1$ into eqn (37) and expanding the result in powers of ζ_c yields:

$$\phi_c(\zeta_c) = \text{constant} + \frac{T}{2}\zeta_c + 0(\zeta_c^2)$$

$$\psi_c(\zeta_c) = \text{constant} - \frac{T}{4}\zeta_c + 0(\zeta_c^2). \tag{38}$$

Dropping the constants (which do not affect the stress field) and higher order terms, then using eqn (36) and $K_I = T\sqrt{\lambda/2}$, we obtain:

$$\phi_c(z_c) = \frac{K_I}{\sqrt{2\pi}} e^{-i\pi/4}\sqrt{z_c} + 0(z_c),$$

$$\psi_c(z_c) = -\frac{K_I}{2\sqrt{2\pi}} e^{-i\pi/4}\sqrt{z_c} + 0(z_c). \tag{39}$$

These are just the complex potentials for the mode *I* crack-tip field for a crack pointing in the negative y_c direction.

In summary, the stress field near a cycloid cusp for $Ak = 1$ is found to be crack-singular, i.e. of singularity order $d^{-1/2}$. Under uniform tension the stress intensity factor at a cycloid cusp is found to be identical to that of an array of periodic parallel cracks. This provides an interesting connection between rough surfaces and sharp cracks; the latter has been widely studied in the framework of Griffith–Irwin fracture mechanics.

J-integral analysis

The stress singularity near a cycloid cusp can be further understood from the path-independent *J*-integral of fracture mechanics (Rice, 1968). Following eqn (28), the hoop stress vanishes everywhere along the surface as $Ak \to 1$, except at the tip of a cusp where it diverges. For cracks oriented along the *y*-direction, the *J*-integral is defined as:

$$J = \int_\Gamma \left(w\,\mathrm{d}x - n_j\sigma_{ij}\frac{\partial u_i}{\partial y}\,\mathrm{d}s \right), \tag{40}$$

where w is the strain energy density and n_i the normal vector along the integration contour Γ. Let Γ be a closed contour composed of several line segments: $\Gamma = \Gamma_1 + \Gamma_\infty + \Gamma_2 + \Gamma_c + \Gamma_\varepsilon$, as shown in Fig. 9; Γ_1 and Γ_2 make no contribution to J because they are two vertical straight lines at adjacent cycloid peaks where symmetry dictates that the shear stress vanish and the displacement normal to the segments be constant; Γ_c denotes a surface segment which is completely stress-free, also not contributing to J. The only contributions to J come from the loop Γ_ε around the cusp-tip, which gives the energetic force (energy release rate) G on the singularity, and from the segment at infinity which gives the strain energy stored in a strip of length λ and unit width. The conservation property of J leads to:

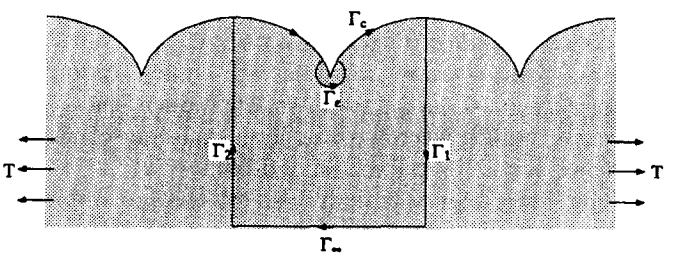

Fig. 9. The *J*-integral contour used to determine the energy release rate G at a cusp on the cycloid surface.

$$G = \lambda w_0 \tag{41}$$

where $w_0 = T^2(1-\nu)/4\mu$ is the strain energy density in the bulk. Application of the Irwin relation $G = K_I^2(1-\nu)/2\mu$ immediately verifies the stress intensity factor given in eqn (33).

The J-integral interpretation of the cycloid stress singularity is important because it implies that a failure criteria based on J or K would remain approximately valid even if there are some small scale geometric variations or nonlinear effects at the cycloid cusp. In similar spirit to a discussion by Rice (1968) for small scale yielding at a crack tip, one may argue that, since the J-integral remains conserved even for nonlinear elastic material behaviors, the small scale local variations in geometry and material conditions near the cusp will not affect the remote bulk stress field, hence they have little effect on the basic energetic force $w_0\lambda$ acting on the cusp.

For a noncusped cycloid surface, the same J-contour in Fig. 9 provides an interesting condition on the surface strain energy distribution. The infinitesimal loop Γ_ε becomes unnecessary in the noncusp cases. The two vertical lines Γ_1 and Γ_2 still make no contribution to J due to symmetry. The contributions to J now come from the surface segment Γ_c, which gives $\int_0^\lambda w\,\mathrm{d}x$, and from the segment at infinity, which gives $w_0\lambda$. Conservation of J requires:

$$\langle w\,(\text{surface})\rangle = \frac{1}{\lambda}\int_0^\lambda w\,\mathrm{d}x = w_0, \tag{42}$$

which means that the average surface strain energy density, $\langle w(\text{surface})\rangle$, is a constant independent of the surface geometry. It is not hard to see that this relation in fact holds for any periodic rough surfaces. In that case, the individual contributions to J from the two vertical lines Γ_1 and Γ_2 may not vanish, but they always cancel each other out because of the periodicity and their opposite sense. The condition $\langle w\,(\text{surface})\rangle = w_0$ can be further generalized to aperiodic surfaces if the period λ is allowed to approach infinity, somewhat reminiscent of the procedure of extending a Fourier series to a Fourier integral. If the line integral over x in eqn (42) is replaced by an area integral, the same condition can also be established for three dimensional nonplanar rough surfaces. Therefore, we conclude that:

$$\langle w\,(\text{surface})\rangle = w_0, \tag{43}$$

applies to any rough surfaces having a uniform bulk strain energy density w_0.

The J-integral condition in eqn (43) suggests that any stress magnification (e.g. due to a surface valley or cusp) along a rough surface must be compensated by relaxation along some complementary portions of the surface and vice versa. Clearly, the optimum surface shape for least stress concentration is a perfectly planar one. Along a cusped cycloid surface, the strain energy distribution bcomes Dirac singular at the cusp tips, the rest of the surface being completely unloaded. Therefore, the effect of cycloid cusps is to redistribute and concentrate all the surface strain energy per wavelength at a single (cusp) point. In that sense, the cycloid surface is the most efficient stress concentrator, or the optimum shape for maximum stress concentration, at a fixed surface wavelength. Since the periodic parallel cracks induce the same stress intensity factor as the cusped cycloid, it seems that any cusped surfaces between the two configurations should induce similar, but no larger, stress concentration. Even at a moderate bulk stress level this type of singular concentration may be sufficient to cause failure or nucleation of defects.

The J-integral condition, eqn (43), may also be used to improve the accuracy of numerical computations. For example, renormalizing a numerically computed surface stress distribution using eqn (43) may lead to increased accuracy or perhaps a more stable algorithm for simulating the surface evolution.

ENERGY VARIATIONS ASSOCIATED WITH EVOLUTION OF A CYCLOID SURFACE

Several recent papers (Srolovitz, 1989; Gao, 1991b, c; Grinfeld, 1993; Spencer *et al.*, 1991) have been directed at understanding the thermodynamic driving force and energy variations for the development of surface roughness in a stressed solid. In these studies, linearized perturbation methods are used to show that elastic stains tend to destabilize an initially flat surface because the strain energy is always reduced by surface roughening. The full problem of evolution of a rough surface after initial instability will have to be solved using numerical methods. Analytical studies, if their limitations are well undestood, will be helpful to check and guide numerical computations, and may also provide valuable insights into the role of major competing factors such as the strain and surface energies during surface evolution. In the following, we present a study of global energy variations associated with evolution of a cycloid surface, in which the surface will be restricted to a cycloid shape and the surface "evolution" is achieved by varying the amplitude parameter A. In particular, we are interested in finding out whether the development of a cusped cycloid surface is energetically favorable. The analysis procedure follows closely that developed earlier by Gao (1991b, c).

The strain energy variation for a shape change of an elastic body was studied by Rice and Drucker (1967). The energy change associated with a moving interface beteen dissimilar materials was studied by Eshelby (1970). The implication of these studies is that the strain energy change associated with an infinitesimal configurational perturbation $\delta\mathbf{x}$ along the cycloid surface can be written as:

$$\delta U = \int w(s)\mathbf{n}\cdot\delta\mathbf{x}\,\mathrm{d}s, \tag{44}$$

where $w(s)$ denotes the strain energy density distribution and $\mathbf{n}$ denotes the outward normal vector of the surface.

We shall assume that the total mass is conserved in the development of surface roughness, as in a diffusional mass transport. Shifting the x-axis to the "neutral" position, the mass-conserved cycloid surface profile can be written as:

$$x = \xi + A\sin k\xi, \quad y = A\cos k\xi - \frac{A^2k}{2}. \tag{45}$$

This profile has the property that the net volume change, which is proportional to the integral $\int_0^\lambda y\,\mathrm{d}x$, always remains zero for any cycloid amplitude A.

Imposing an infinitesimal amplitude variation δA, the points along the mass-conserved cycloid surface of eqn (45) will move by:

$$\delta\mathbf{x} = [\sin k\xi\,\mathbf{e}_x + (\cos k\xi - Ak)\mathbf{e}_y]\,\delta A, \tag{46}$$

where $\mathbf{e}_x$ and $\mathbf{e}_y$ are unit vectors in the x and y directions. Also, one may show that:

$$\mathbf{n}\,\mathrm{d}s = \left(-\frac{\partial y}{\partial\xi}\mathbf{e}_x + \frac{\partial x}{\partial\xi}\mathbf{e}_y\right)\mathrm{d}\xi = [Ak\sin k\xi\,\mathbf{e}_x + (1+Ak\cos k\xi)\mathbf{e}_y]\,\mathrm{d}\xi, \tag{47}$$

and

$$w(\xi) = w_0\frac{(1-A^2k^2)^2}{(1+2Ak\cos k\xi + A^2k^2)^2}, \tag{48}$$

where $w_0 = T^2(1-\nu)/4\mu$ denotes the bulk strain energy density.

Substituting the above expressions for w, $\delta\mathbf{x}$, and $\mathbf{n}\,\mathrm{d}s$ into eqn (44), dividing both sides by δA, and then letting δA approach zero, one finds that the rate of change of the strain energy in one period of the cycloid with the amplitude A as:

$$\frac{\partial U}{\partial A} = w_0 \int_0^{2\pi/k} \frac{(1-A^2k^2)^3 \cos k\xi}{(1+2Ak\cos k\xi + A^2k^2)^2}\, d\xi = -4w_0\pi A. \tag{49}$$

Therefore, the strain energy is always reduced by enlarging A toward the cusp limit $A = 1/k = \lambda/2\pi$. This provides a driving force for the development of surface roughness under stress corrosion. The total strain energy releasd in developing a cycloid surface of amplitude A from a perfectly flat surface is obtained by integrating the energy derivative, eqn (49), giving:

$$\Delta U = \int_0^A \frac{\partial U}{\partial A}\, dA = -2w_0\pi A^2. \tag{50}$$

The above results are formally coincident with those derived by Gao (1991b) for a slightly undulating cosine surface using linear perturbation techniques. The reason is that a cycloid is asymptotically identical to a cosine curve at small amplitude ($Ak \ll 1$), combined with the fact that the exact strain energy reduction rate of a cycloid surface is linear in A. Gao (1991c) also derived explicit solutions for energy variation associated with slightly roughened surfaces in anisotropic elastic solids, in which case the cycloid-surface solution has not yet been developed.

Working against the strain energy, the surface energy increases with the surface roughness and suppresses instabilities at short wavelengths. One may also include the effect of gravity. Gao (1991b) showed that if the surface is pointing upwards opposite to the gravitational force, the gravity tends to suppress instabilities at long wavelengths. In the case of highly stressed solids such as thin films, the gravity effect can be ignored since the most unstable instability wavelength is typically of micron or submicron scale.

Neglecting material anisotropy, the surface energy is a product of the surface area and the surface energy density γ, which for one period of the cyclolid surface is:

$$E_s = \gamma \int_0^{2\pi/k} \sqrt{1+2Ak\cos k\xi + A^2k^2}\, d\xi. \tag{51}$$

Differentiating eqn (51) with respect to A gives the rate of change of the surface energy with the cycloid amplitude A, which may be written as:

$$\delta E_s = \frac{\partial E_s}{\partial A}\delta A = \gamma \delta A \int_0^{2\pi/k} \frac{k(\cos k\xi + Ak)}{\sqrt{1+2Ak\cos k\xi + A^2k^2}}\, d\xi. \tag{52}$$

At this point, let us note in passing that an alternative variational formula for surface energy change similar to the strain energy formula, eqn (44), is (Alexander and Johnson, 1985):

$$\delta E_s = \int \gamma\kappa(s)\mathbf{n}\cdot\delta\mathbf{x}\, ds. \tag{53}$$

These two eqns, (52) and (53), are equivalent until the cusp is formed. In the cusp case, the variational eqn (53) breaks down but the more direct formula, eqn (52), remains valid, as shown in Appendix B. The variational formula, eqn (53), is however, more helpful in deriving the chemical potential for surface kinetics equations.

The sum of eqns (49) and (52) gives the total energy variation with respect to A,

$$\frac{\partial E_{\text{tot}}}{\partial A} = -4w_0\pi A + \gamma \int_0^{2\pi/k} \frac{k(\cos k\xi + Ak)}{\sqrt{1+2Ak\cos k\xi + A^2k^2}}\,\mathrm{d}\xi. \tag{54}$$

At the initial instability when Ak is small, it can be shown that:

$$\left.\frac{\partial E_{\text{tot}}}{\partial A}\right|_{Ak\ll 1} \sim (-4w_0 + k\gamma)\pi A. \tag{55}$$

The critical wavenumber for initial instability is then $k_{\text{cr}} = 4w_0/\gamma$, corresponding to $\lambda_{\text{cr}} = \pi\gamma/2w_0$, which is consistent with results presented in Srolovitz (1989), Gao (1991b), Grinfeld (1993) and Spencer *et al.* (1991). Gao (1991b) also argued that, since the driving force per unit surface area for diffusion is proportional to $k\,\partial E_{\text{tot}}/\partial A$ and a diffusion process would involve transport of mass in an area of order A/k over a distance $1/k$, the most unstable instability mode which grows fastest at the initial stage should correspond to maximizing the quantity $k^3\partial E_{\text{tot}}/\partial A$, giving $k = k_{\text{diffusion}} = 3w_0/\gamma$. Under conditions such as evaporation or condensation, the fastest growing mode should be the one releasing most energy per unit area, corresponding to maximizing $k\partial E_{\text{tot}}/\partial A$, which yields $k = k_{\text{condensation}} = 2w_0/\gamma$. These results confirm those obtained by Srolovitz (1989) via kinetic equations.

When the cusp is formed, eqn (54) reduces to:

$$\left.\frac{\partial E_{\text{tot}}}{\partial A}\right|_{Ak=1} = 4\left(\gamma - \frac{\pi w_0}{k}\right). \tag{56}$$

The critical wavenumber k_{lcr} for the development of a cusped cycloid surface is defined by letting eqn (56) vanish, giving:

$$k_{\text{lcr}} = \pi w_0/\gamma. \tag{57}$$

Note $k_{\text{cr}} > k_{\text{lcr}} > k_{\text{diffusion}} > k_{\text{condensation}}$, indicating that the fastest growing modes, by diffusion (during annealing) or by condensation (during growth), at the initial instability have the potential to develop into a cusped rough surface. Figure 10 plots $(k/w_0)\,\partial E_{\text{tot}}/\partial A$ versus Ak, with k taken as the three special values: $k = k_{\text{lcr}}$, $k_{\text{diffusion}}$ and $k_{\text{condensation}}$.

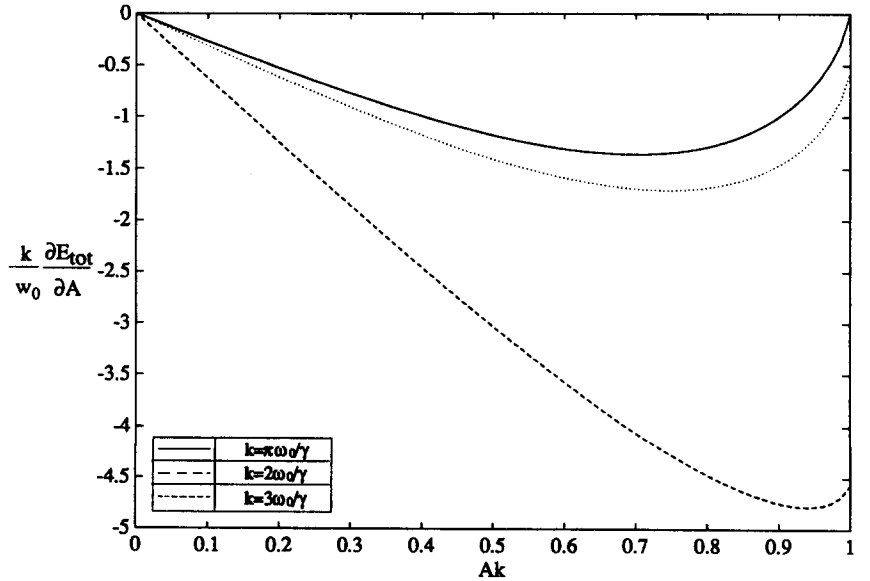

Fig. 10. The variation of energy derivative $(k/w_0)\,\partial E_{\text{tot}}/\partial A$ associated with evolution of a cycloid surface.

As the surface changes from a perfectly flat shape into a fully cusped cycloid, the total surface energy change should be:

$$\Delta E_s = \gamma\left[\int_0^{\lambda}\sqrt{1+2Ak\cos k\xi + A^2k^2}\,\mathrm{d}\xi - \lambda\right]_{Ak=1} = \gamma(8-2\pi)A. \tag{58}$$

Combining this result with eqn (50), the overall energy change in the transition is:

$$\Delta E_{\text{tot}} = A[\gamma(8-2\pi) - 2w_0\pi/k]. \tag{59}$$

The strain energy dominates for sufficiently small k, so that $\Delta E_{\text{tot}} \sim -2w_0A\pi/k = -2w_0\pi/k^2$ as $k \to 0$. Once $k < 2w_0\pi/\gamma(8-2\pi) = 3.66w_0/\gamma$, the energy of the cusped cycloid is already smaller than a perfectly flat surface. For k_{lcr} $(= \pi w_0/\gamma) < k < 3.66w_0/\gamma$, the cusped cycloid is not the most favorable configuration among all cycloid shapes. Only when $k < k_{\text{lcr}}$ does the cusped cycloid become the most preferred state among the cycloid family. In that case, are there any other surfaces still more favorable than a cusped cycloid? This question will have to be answered by a numerical study.

For initial stability analysis the two dimensional plane strain results can be directly applied to the three dimensional configuration of a planar surface under biaxial loading. This is because a three dimensional perturbation wave can always be decomposed into two two-dimensional perturbation waves, in the spirit:

$$y = A\cos k_x x\cos k_3x_3 = (A/2)\cos kx' + (A/2)\cos kx'', \tag{60}$$

where x_3 denotes the coordinate axis normal to the x–y plane, $k = \sqrt{k_x^2+k_3^2}$, $kx' = k_xx+k_3x_3$ and $kx'' = k_xx-k_3x_3$. In this manner the three dimensional perturbation problem is decomposed as a linear superposition of two two-dimensional plane strain problems in the x'–y and the x''–y planes. For biaxial loading, it can be shown that the critical wavelengths such as k_{cr}, $k_{\text{diffusion}}$, $k_{\text{condensation}}$ derived from plane strain analysis remain valid. Gao (1991a) showed that the three dimensional effects tend to slightly relax the stress concentration at the valleys (by 5–12% for Poisson ratio ν ranging from 0.25–0.35).

The cycloid surface we have studied results in downward-pointing cusps as $Ak \to 1$. This shape has been chosen for its similarity to the island morphology and more importantly, for its convenience in analytical studies. There is another class of cycloid surfaces which results in upward-pointing cusps as $Ak \to 1$. For lack of a better terminology, we refer to this class as crest-like cycloid surfaces. The elasticity problem for a crest-like cycloid can not be solved analytically because the corresponding mapping function involves a term $\exp(ik\zeta)$ which is not analytic, hence not conformal, in the material region. We have numerically studied this problem and our results presented in Appendix C show that the crest-like cycloids are less interesting because the strain energy release rate in forming upward-pointing cusps is substantially less than that for the cusp-like case. A simple singularity analysis also shows that the stresses at an upward-pointing cusp is identically zero. Perturbation analyses of the stress field (Gao, 1991a) of slight undulating surfaces have indicated that a protruding part over a flat surface is unloaded very quickly. The surface energy dominates along these portions, and this eliminates any possibility of forming a cusp. In other words, upward-pointing cusps are intrinsically unstable.

SURFACE DIFFUSION ALONG AN ALMOST-CUSPED CYCLOID SURFACE

In the previous section, it has been shown that, from a global point of view, a cusped cycloid surface with crack-like singularities becomes energetically favorable below a critical wave number k_{lcr} (equivalently, beyond a critical wavelength λ_{lcr}). The cusps may in turn play a significant role in nucleating defects leading to plastic deformation or brittle fracture, as will be discussed later. Another question that needs to be resolved at this point is the stability of cycloid cusps, i.e. whether such cusps will be smoothed out, hence eliminated,

by mass transport mechanisms such as evaporation, condensation and surface and bulk diffusion.

We will be particularly interested in understanding the stability of a cycloid cusp against surface diffusion. For that purpose we consider a nearly-cusped cycloid surface with $Ak \approx 1$. At this nearly-cusped state, let the surface diffusion be suddenly set on stage (a thought experiment). The question to be posed is "Will the induced surface movement at this onset of diffusion tend to promote or eliminate the cusp formation?" In other words, will the diffusion tend to increase or decrease the local curvature, κ_v, at a nearly-cusped cycloid valley? Since the curvature at a valley is negative, the cusps are said to be stable if $\delta\kappa_v < 0$, corresponding to κ_v increasing in absolute magnitude (sharper); oppositely, the cusps are unstable if $\delta\kappa_v > 0$, corresponding to κ_v decreasing in magnitude. In the latter case, a nearly-cusped valley tends to be smoothed out by diffusion.

It can be shown that the curvature variation $\delta\kappa(s)$ at a surface point s associated with an infinitesimal surface movement $\delta u_n(s)$ is:

$$\delta\kappa(s) = -\kappa^2(s)\delta u_n(s) - \delta u_n''(s), \tag{61}$$

where s is the arc length position, and the direction of δu_n is along the outward normal of the surface. This formula can be used to calculate $\delta\kappa_v > 0$ once $\delta u_n(s)$ is known.

Chemical potential along a cycloid surface

The surface movement $\delta u_n(s)$ is controlled by the distribution of the chemical potential $\chi(s)$. The variational formulae in eqns (44) and (53) suggest that $\chi(s)$ may be written as (Herring, 1950; Srolovitz, 1989; Spencer *et al.*, 1991):

$$\chi(s) = \chi_0 + \gamma\Omega\kappa(s) + \Omega w(s), \tag{62}$$

where χ_0 is a reference value, $\kappa(s)$ the surface curvature, $w(s)$ the surface strain energy density and Ω the atomic volume. Along a cycloid surface, the curvature κ can be determined as a function of the variable ξ from:

$$\kappa = \frac{x''(\xi)y'(\xi) - x'(\xi)y''(\xi)}{(x'(\xi)^2 + y'(\xi)^2)^{3/2}}, \tag{63}$$

which after inserting eqn (45) gives:

$$\kappa = \frac{Ak^2(Ak + \cos k\xi)}{[1 + 2Ak\cos k\xi + A^2k^2]^{3/2}}. \tag{64}$$

The local curvature at the cycloid valley is:

$$\kappa_v = -\frac{Ak^2}{(1 - Ak)^2}. \tag{65}$$

The chemical potential now has the expression:

$$\chi = x_0 + \Omega\left[\frac{\gamma Ak^2(Ak + \cos k\xi)}{(1 + 2Ak\cos k\xi + A^2k^2)^{3/2}} + \frac{w_0(1 - A^2k^2)^2}{(1 + 2Ak\cos k\xi + A^2k^2)^2}\right], \tag{66}$$

which reduces to the first-order formula by Srolovitz (1989) in the case of small Ak. At a valley where $\cos k\xi = -1$, χ has the value χ_v,

$$\chi_v - \hat{\chi}_0 = \Omega\kappa_v(\gamma - 4w_0/k) = \Omega\kappa_v\gamma(1 - k_{cr}/k), \tag{67}$$

where $\hat{\chi}_0 = \chi_0 + \Omega w_0$. In the above equation, $k_{cr} = 4w_0/\gamma$ is just the critical wavenumber for

the initial surface instability, as in eqn (55). Interestingly, eqn (67) is valid for all Ak values, from $Ak \ll 1$ up to the cusp limit $Ak \to 1$; the strain energy competes with the surface energy, each contributing a term proportional to the local curvature κ_v. The strain energy prevails if $k < k_{cr}$ and the surface energy prevails if $k > k_{cr}$. Thus the surface instability can be interpreted as a signal for the beginning of strain energy dominance in chemical potential at a cycloid valley. Since $k_{lcr} < k_{cr}$, we conclude that once a cusped cycloid surface becomes energetically favorable in the global sense, the local dominance of the strain energy is also established.

Consider a nearly-cusped cycloid surface, $Ak \approx 1$, in which case the strain energy density is almost Dirac singular at the cusp location. The surface energy dominates everywhere along the surface except very near the cusp. In the close vicinity of the cycloid valley, χ has the asymptotic expansion:

$$\chi(s) - \hat{\chi}_0 = \Omega\kappa_v \left\{ (\gamma - 4w_0/k) - (3\gamma - 16w_0/k)\frac{(\kappa_v s)^2}{2} + (57\gamma - 352w_0/k)\frac{(\kappa_v s)^4}{4!} + o[(\kappa_v s)^6] \right\}, \tag{68}$$

where s is the surface arc length measured from the valley. Note that the coefficients of odd powers in $\kappa_v s$ vanish due to symmetry, and those of even powers are obtained from the zeroth, second, and fourth order derivatives of $\chi(s)$ at $s = 0$. The zeroth order term is valid for all Ak values, but the higher order terms are valid only for $Ak \approx 1$. The strain energy competes with the surface energy in all expansion terms of $\chi(s)$.

Motion of a nearly-cusped cycloid surface at the onset of surface diffusion

The surface evolution can be affected by several mass transport processes such as surface and bulk diffusion, condensation and evaporation. Here we only consider the surface diffusion mechanism, in which case the induced surface movement in the normal direction, u_n, is governed by (Srolovitz, 1989; Spencer *et al.*, 1991):

$$\frac{\partial u_n}{\partial t} = \frac{D_s \Omega \delta}{k_B T_K} \frac{\partial^2 \chi}{\partial s^2}, \tag{69}$$

where D_s is the surface diffusivity, δ is the number of atoms (sites) per unit area, Ω is the atomic volume, k_B is the Boltzmann constant, and T_K the Kelvin temperature. The normal direction is defined to be positive when pointing outward so that a negative value of $\partial u_n/\partial t$ at the valley means that the surface will move further downward (to form a cusp) and a positive value means that the surface will move upward (to smooth out a cusp).

At the onset of surface diffusion, the second order derivative $\partial^2\chi/\partial s^2$ of the chemical potential along a nearly-cusped cycloid surface can be obtained from eqn (66) using:

$$\frac{d\xi}{ds} = \frac{1}{(1 + 2Ak\cos k\xi + A^2k^2)^{1/2}}. \tag{70}$$

For the cycloid surface, eqn (69) becomes:

$$\frac{\partial u_n}{\partial t} = \frac{D_s \delta \Omega^2 A k^3}{k_B T_K} \left\{ \frac{4w_0(1 - A^2k^2)^2[7Ak + (1 + A^2k^2)\cos k\xi - 5Ak\cos^2 k\xi]}{(1 + 2Ak\cos k\xi + A^2k^2)^5} \right.$$

$$\left. - \gamma k \frac{[7Ak - 11A^3k^3 + (1 - 5A^2k^2 - 2A^4k^4)\cos k\xi + 6(-Ak + A^3k^3)\cos^2 k\xi + 2A^2k^2\cos^3 k\xi]}{(1 + 2Ak\cos k\xi + A^2k^2)^{9/2}} \right\}. \tag{71}$$

For $Ak \approx 1$, examination of the above expression indicates that the induced surface motion by diffusion is dominated by the surface energy everywhere along the surface except very near the valley where one finds the following expansion:

$$\frac{\partial u_n}{\partial t} = \frac{D_s \delta \Omega^2 \kappa_v^3}{k_B T_K} \left\{ \left(-3\gamma + \frac{16 w_0}{k} \right) + \left(57\gamma - \frac{352 w_0}{k} \right) \frac{(\kappa_v s)^2}{2} + o[(\kappa_v s)^6] \right\}. \tag{72}$$

Thus, the magnitude of the surface motion near the cusp is proportional to κ_v^3 and the direction of the surface movement is determined by the competition between the strain energy and the surface energy. The strain energy dominates and causes the cycloid valley to move downward if $k < 16w_0/3\gamma$. Again, since $k_{lcr} < 16w_0/3\gamma$, once the formation of a cycloid cusp becomes energetically favorable in the global sense, the behavior of the local surface movement also exhibits a cusp-forming tendency. On the other hand, if $k > 16w_0/3\gamma$, the surface energy will suppress the strain energy and result in upward surface movement to eliminate the cusp.

Figure 11 plots the variation of the normalized $\partial u_n/\partial t$ with arc length s for $Ak = 0.999$. The origin $s = 0$ corresponds to the valley. Four cases where $k = 16w_0/3\gamma$, $\pi w_0/\gamma$, $4w_0/\gamma$, and $2w_0/\gamma$ are chosen to visualize the influence of the wavenumber on the motion of a nearly-cusped cycloid surface. As long as $k < 16w_0/3\gamma$, there is always a region near the valley where the surface moves downward.

For numerical simulation of the formation process of surface cusps, it may be interesting to note from Fig. 11 that at the nearly-cusped state, the surface motion is highly concentrated at the valleys with large magnitude. Within one wavelength of the nearly-cusped cycloid, the size of the significant diffusion zone is very small, on the order of $ks = 10^{-6}$, in comparison with the total arclength $ks_{tot} = 8$. Significant diffusion activity is concentrated to a zone of order of the radius of local curvature at the valley, i.e. $\sim 1/\kappa_v$.

Stability of a cycloid cusp against surface diffusion

The quantity $\partial u_n''(s)/\partial t$ at the valleys of a near-cusped cycloid can be obtained by differentiating eqn (72) with s twice, giving

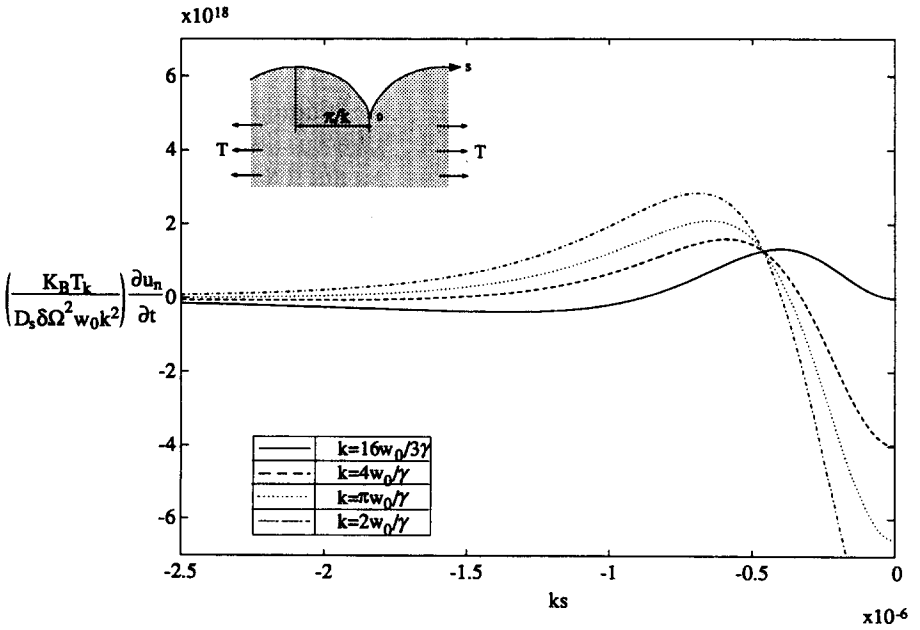

Fig. 11. The surface motion at a nearly-cusped cycloid valley with $Ak = 0.999$ at the onset of diffusion for $k = 16w_0/3\gamma$, $4w_0/\gamma$, $\pi w_0/\gamma$ and $2w_0/\gamma$.

$$\left.\frac{\partial u_n''(s)}{\partial t}\right|_{s=0} = -\frac{D_s\delta\Omega^2}{k_B T_K}\kappa_v^5\left(\frac{352w_0}{k} - 57\gamma\right), \tag{73}$$

where κ_v is the local curvature given in eqn (65). Combining eqns (72), (73) and (61) leads to the rate of curvature change at the nearly-cusped valleys at the onset of surface diffusion. The result is:

$$\frac{1}{\kappa_v^5}\frac{\partial\kappa_v}{\partial t} = \frac{6D_s\delta\Omega^2}{k_B T_K}\left(\frac{56w_0}{k} - 9\gamma\right). \tag{74}$$

This implies that the absolute magnitude of κ_v increases when $k < 56w_0/9\gamma$. In that case, the cycloid cusps are stable against surface diffusion in that a nearly-cusped state tends to evolve toward the fully-cusped state. On the other hand, when $k > 56w_0/9\gamma$, the cusps are unstable in that the diffusion reduces the magnitude of the local curvature, and hence tends to eliminate the cusp.

In summary, since $k_{lcr} < 16w_0/3\gamma < 56w_0/9\gamma$, one concludes that once the cusped cycloid surface becomes energetically favorable in the global sense, the strain energy completely dominates the surface energy in the local diffusion behavior near the cusps; the strain energy is capable of preventing the surface energy from smoothing out and eliminating the cusps once they develop.

Rigorously, the curvature change rate $\partial\kappa_v/\partial t$ given in eqn (74) is only valid at the onset of surface difusion. After the diffusion starts, the curvature change rate may not show such a simple behavior. However, note that the expression on the right hand side of eqn (74) only involves material properties and global quantities such as γ, w_0, k, etc. For a rough estimate it is not completely without merit to make the assumption that eqn (74) remains approximately valid after the surface diffusion begins to drive the local curvature into a cusp. If that assumption is approximately valid, then the evolution of κ_v can be obtained by integrating eqn (74), giving:

$$\kappa_v(t) = \kappa_v^0(1 - t/t^*)^{-1/4}, \tag{75}$$

where κ_v^0 is the local curvature at the onset of diffusion and,

$$t^* = \frac{k_B T_K}{24D_s\delta\Omega^2}\frac{k}{(\kappa_v^0)^4(56w_0 - 9\gamma k)}. \tag{76}$$

At $t = t^*$, the surface reaches the perfectly cusped state in which $\kappa_v = -\infty$. Only a finite amount of time is needed in this "final stage" of cusp formation. Further, assuming that the surface motion at the valley in eqn (72) is also approximately valid after diffusion occurs, the surface movement u_n of the valley during the final stage of cusp formation can be determined by substituting eqn (75) into (72) and integrating (72) with t. The result is:

$$u_n|_{s=0} = \frac{1}{6\kappa_v^0}\left(\frac{3\gamma - 16w_0/k}{9\gamma - 56w_0/k}\right)\left[1 - \left(1 - \frac{t}{t^*}\right)^{1/4}\right]. \tag{77}$$

The local surface velocity at the valley behaves as k_v^3, hence approaching infinity as the cusp is formed, $k_v \to -\infty$. However, eqn (77) suggests that the total surface movement remains finite, provided that our assumptions are valid.

FURTHER DISCUSSIONS

A nearly cycloidal surface with constant chemical potential

Analysis of global energy variations indicated that a cusped cycloid surface is energetically favorable beyond a critical wavelength. In the previous section we have further

shown that, once the global strain energy dominance is achieved, the cycloid cusps are also locally stable in that once the cusp is formed, the strain energy prevents the surface energy from smoothing out the cusp via mass transport. The evolution of a material surface by diffusion or other kinetic processes is controlled by the chemical potentials $\chi(s)$ given in eqn (62). Along a cusped cycloid surface, the strain energy density $w(s)$ vanishes everywhere except at the cusp tips. Constant chemical potential requires modifiying the surface toward one of constant curvature. Normally, changing the surface curvature will cause redistribution of stresses and may induce a nonzero $w(s)$. However, boundary perturbation formulations (Gao, 1991b) suggest that the stress variation associated with perturbation of a stress-free surface segment is zero within first order accuracy. Thus the strain energy along a nearly cycloidal surface segment having cusps is also approximately stress-free as long as the cusp vicinity is unperturbed. Assuming mass-conservation, the closest matching circular arc for the cusped cycloid has radius of 0.5433λ. As shown in Fig. 12, this arc matches the cusped cycloid very well, with maximum deviation not exceeding 4%. Thus, a nearly cycloidal surface with constant chemical potential can be found between a cusped cycloid and its closest matching circular arc. Near the cusp we expect that the cusp configuration should remain because of the dominance of strain energy. Away from the cusp we expect that the surface should be close to the matching circular arc because the segment is stress-free and the surface energy should dominate.

At the critical wavenumber $k_{\text{lcr}} = \pi w_0/\gamma$ for cusp formation, the energetic force on the cusp is:

$$G = w_0\lambda_{\text{lcr}} = 2\pi w_0/k_{\text{lcr}} = 2\gamma. \tag{78}$$

But this is just the Griffith condition for energy balance at a crack-tip. Therefore, a nearly cycloidal surface at k_{lcr} can satisfy the Griffith energy balance exactly at the cusps while maintaining constant chemical potential along the rest of the surface. If the Griffith condition, eqn (78), is postulated as a necessary condition for the formation of any cusp, then during initial surface instability the perturbations fall into three distinct categories. Perturbations with wavelengths smaller than λ_{cr} tend to decay; those with $\lambda_{\text{cr}} < \lambda < \lambda_{\text{lcr}}$ grow but will not lead to cusp formation because the Griffith condition requires the minimum spacing between cusps to be $\lambda_{\text{lcr}} = 2\gamma/w_0$. Finally, perturbations with wavelengths exceeding λ_{lcr} have the potential of growing into a cusped configuration. The final spacing between the cusps will have values around $\lambda_{\text{diffusion}}$ under prevailing surface diffusion conditions and around $\lambda_{\text{condensation}}$ under vapor condensation.

Nucleation of misfit dislocations from thin film surfaces

A basic understanding of the process of plastic deformation by misfit dislocations in thin heteroepitaxial films has been provided through the pioneering works of Van der Merwe (1963) and Matthews (1975), and through recent theoretical and experimental efforts (Freund, 1987; Nix, 1989; Tsao, 1993). Based on this understanding, it has been possible

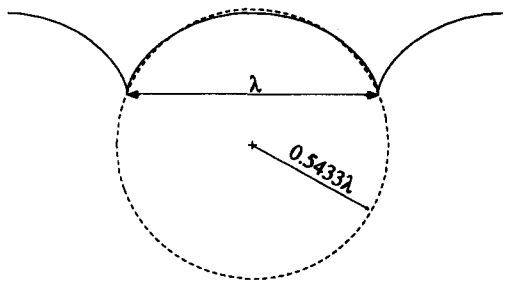

Fig. 12. A cusped cycloid surface of wavelength λ and its closest matching circular arc of radius 0.5433λ.

to predict and understand the thermodynamic and kinetic factors that control the movements of individual dislocations in heteroepitaxial films. Despite this progress, it is still not possible to predict or understand the overall kinetics of strain relaxation which involves the nucleation and multiplication of a large number of dislocations in these films. It is now well known that dislocations form spontaneously in strained layers grown epitaxially onto dislocation-free substrates once the layer thickness exceeds a critical value. Most evidence suggests that some kind of nucleation must be occurring at the surface of the strained film. However, if the film surface of a strained epitaxy layer is assumed to be flat, then it appears that the activation energy barrier for dislocation formation is too high for a practical nucleation process. The development of surface stress concentration, perhaps by a process leading to a cusped rough surface, may be critical in explaining the observed dislocation kinetics.

Our analysis of the stress singularity induced by a cycloid rough surface suggests that the problem of dislocation nucleation from a rough surface may be closely associated with the dislocation emission from a crack tip. The latter problem has been a topic of intensive study during the past two decades. An extensive review of recent progress on this subject was given by Rice (1992). It seems that many existing results on the modelling of the crack-tip dislocation emission process can be borrowed to build a mechanistic theory for the nucleation of misfit dislocations from micro-level singularities along the thin film surface. This will be studied in future work.

There is preliminary evidence that thermally-assisted surface diffusion may have played a vital role in the kinetics of surface roughness development. In a recent effort to understand the kinetics and the mechanisms of strain relaxations in Si–Ge heteroepitaxial films, Gillard *et al.* (1992) measured the elastic misfit strain as a function of temperature using a laser scanning technique. In that experiment, the relaxation due to dislocation processes is controlled by varying the film thickness. Comparison between the elastic strains measured at room temperature and those obtained at 700°C suggests that a portion of the strain relaxation occurs almost instantly during the heat-up procedure. This relaxation can not be satisfactorily explained by dislocation relaxation because some of the samples tested are close to or below the critical thickness for dislocation generation. Part of the answer to such "mysterious" relaxation at high temperature could lie with the development of the surface roughness of the film. Further experimental and theoretical investigations will be needed to fully resolve this issue.

Development of surface cracks in brittle solids

Our present work has also been partly motivated by an interest to understand the basic mechanisms for surface flaw initiation in brittle material systems, such as ceramics, which are being widely studied for high temperature applications. In most practical service conditions, observations and postmortem failure analysis indicate (Stokes, 1972; Varner and Frechette, 1986; Frechette, 1990) that flaw initiation takes place at exposed material surfaces which are subjected to hostile environmental corrosions.

Let us consider an ideally brittle solid with no plastic relaxation. Suppose that a cusped surface has been developed through some kinetic process such as diffusion, annealing or deposition. The following question can be posed "once the cusps are formed, can they further propagate into the solid to develop macroscopic surface cracks?" The answer is obviously positive if our conjecture that the Griffith criterion is a necessary condition for cusp formation is correct and if the solid is under tension. In that case, the condition for crack propagation and that for cusp formation are coincident with the Griffith criterion, that is:

$$G \geqslant 2\gamma. \tag{79}$$

Thus an energetically favorable path for initiation of brittle fracture is achieved without any activation energy barrier. An initially defect-free solid with a perfectly smooth surface can gradually develop a rough surface morphology on the micro-scale through kinetic processes, such as diffusion, creep, fatigue and any other physical or chemical mechanisms.

Given enough time the development of cusps causes increasing concentration of stress and energy, ultimately leading to catastrophic brittle fracture. The history of this failure development seems to correspond to what is commonly known as the static fatigue of brittle solids.

This flaw initiation mechanism could have important applications to the understanding of strength degradation in advanced structural materials including various types of ceramics, polymer, carbon, intermetallics and their composites. For example, dislocation motions are to a large extent inhibited in ceramic materials due to their strong covalent type atomic bonding. The initiation of brittle fracture by development of surface roughness seems to be a viable mechanism for failure in these materials. It will be interesting to use the cycloid surface as a model to study brittle crack initiation from surfaces of a stressed solid under various mechanical conditions such as friction, wear, impact, etc.

Another topic of fundamental interest is to study the ductile vs brittle response to cracks in various materials, including considerations of whether a solid is intrinsically cleavable. Rice and Thomson (1974) proposed that such intrinsic cleavability is determined by the competition between the cleavage decohesion and the crack-tip emission of blunting dislocations. In other words, if the conditions for decohesion of crystal plane ahead of the crack are reached before those for dislocation nucleation, it is then feasible to conclude that the crystal will fail by cleavage. Previous investigations all rely on the assumption of a pre-existing slit crack. The cycloid surface provides an ideal mechanical model to study the issue of brittle crack initiation vs dislocation generation from a rough surface.

Numerical simulation of the kinetics of surface evolution

The analysis presented in this paper has provided some insights into the problem of how a rough surface can be developed from an initially flat surface and how a cusped cycloid surface can generate crack-like stress singularities which may lead to the nucleation of brittle fracture or plastic deformation. However, the kinetics of how a surface evolves under the influence of chemical potential has not been addressed. This evolution will depend on specific physical and chemical processes involved and the material parameters associated with them. One has to resort to numerical techniques to simulate these processes. We have tested and are currently implementing boundary integral type methods for carrying out this task. During the surface evolution, for numerical convergence it is convenient to define a cut-off wavenumber $k_{\text{cut off}}$ so that high frequency numerical noises can be filtered out. When instability occurs, $k_{\text{cut off}}$ can be taken as $k_{\text{cr}} = 4w_0/\gamma$ but it then increases with the maximum strain energy density, $w_{\max}$, along the surface. It can be argued that $k_{\text{cut off}}$ can be taken as:

$$k_{\text{cut off}} = 4w_{\max}/\gamma. \tag{80}$$

During the cusp formation, perturbations of increasingly higher frequencies are calculated. This treatment seems to be important for constructing a properly convergent numerical algorithm. Our numerical results will be reported in a forthcoming paper.

As a final remark, we note that stress-driven surface or interfacial instabilities like those discussed in this paper are associated with a wide class of material instability problems. One of the major objectives in modern materials research is to develop microstructures with desirable physical properties. However, useful microstructures are almost always thermodynamically unstable since for any material system there is only one completely stable structure out of an infinite number of possible unstable or metastable structures. Interfacial instabilities, notably that studied by Mullins and Sekerka (1964) for solidification under constitutional supercooling conditions, have long been associated with such experimental observations as dendritic and cellular crystal growth which leads to formation of grains in polycrystalline solids. Fluid systems also exhibit interfacial instabilities that sometimes bear mathematical analogy to those in solids (Pelce, 1988); examples are the Rayleigh–Taylor instability between accelerating fluids of different densities and the Saffman–Taylor instability in pressure-driven two-phase viscous flow in porous media. These problems provide a rich variety of research opportunities.

CONCLUSION

It is shown in this paper that a cusped cycloid surface can have the same effect of stress concentration as an array of periodic parallel cracks. This connection between rough surfaces and cracks may open new channels for a series of interesting mechanics problems as have been discussed in this paper.

The development of a cusped cycloid surface from an initially flat surface is shown to be energetically favorable once the wavelength of the rough surface exceeds a critical value determined by the competition between the elastic energy and the surface energy. The critical condition for formation of cycloid cusps is found to be just equal to the Griffith energy balance $G = 2\gamma$, i.e. the driving force on the cusp-singularity equals the resistance force due to surface energy. If the plastic relaxation is fully suppressed and the solid is under tension, it is found that the development of cusps results in spontaneous brittle fracture.

Acknowledgement—The work reported was supported by the U.S. Department of Energy under grant DE-FG03-91ER14196 and the U.S. Office of Naval Research under grant N00014-92-J-4094; C. Chiu also acknowledges a graduate fellowship from IBM Almaden Research Center. We are grateful to Professor W. D. Nix, Dr T. W. Wu and two anonymous reviewers for helpful discussions and comments.

REFERENCES

Alexander, J. I. D. and Johnson, W. C. (1985). Thermomechanical equilibrium in solid-fluid systems with curved interfaces. *J. Appl. Phys.* **58,** 816–824.

Erdorgan, F., Gupta, G. D. and Cook, T. S. (1973). Numerical solution of singular integral equations. In *Methods of Analysis and Solutions of Crack Problems* (Edited by G. C. Sih), pp. 368–425. Leyden, Noordhoff.

Eshelby, J. D. (1970). Energy relations and the energy-momentum tensor in continuum mechanics. In *Inelastic Behavior of Solids* (Edited by M. F. Kanninen), pp. 78–115. McGraw-Hill.

Fleck, N. A. (1991). Brittle fracture due to an array of microcracks. *Proc. Royal Society of London*, Vol. A432, pp. 55–76.

Frechette, V. D. (1990). *Failure Analysis of Brittle Materials.* The American Ceramic Society, Westerville, OH.

Freund, L. B. (1987). The stability of a dislocation threading a strained layer on a substrate. *J. Appl. Mech.* **54,** 553–557.

Gao, H. (1991a). Stress concentration at slightly undulating surfaces. *J. Mech. Phys. Solids* **39,** 443–458.

Gao, H. (1991b). A boundary perturbation analysis for elastic inclusions and interfaces. *Int. J. Solids Structures* **28,** 703–725.

Gao, H. (1991c). Morphological instabilities along surfaces of anisotropic solids. In *Modern Theory of Anisotropic Elasticity and Applications* (Edited by J. J. Wu, T. C. T. Ting and D. M. Barnett), pp. 139–150. SIAM, Philadelphia.

Gelfand, I. M. and Fomin, S. V. (1963). *Calculus of Variation.* Prentice-Hall, Englewood Cliffs, New Jersey.

Gillard, V. T., Noble, D. B. and Nix, W. D. (1992). In situ study of isothermal strain relaxation in Si–Ge heteroepitaxial films using substrate curvature measurements. *Materials Research Symp. Proc.*, Vol. 239, pp. 395–400.

Gradshteyn, I. S. and Ryzhik, I. M. (1980). *Tables of Integrals, Series, and Products.* Academic Press, New York.

Grinfeld, M. A. (1993). The Stress Driven Instabilities in Crystals: Mathematical Models and Physical Manifestations, *J. Nonlinear Sci.* **3,** 35–83.

Herring, C. (1950). Diffusional viscosity of a polycrystalline solid. *J. Appl. Phys.* **21,** 437–445.

Koiter, W. T. (1961). An infinite row of parallel cracks in an infinite elastic sheet. In *Problems of Continuum Mechanics* (Muskhelishvili Annual Vol.), pp. 246–259, SIAM, Philadelphia.

Matthews, J. W. (1975). Coherent interfaces and misfit dislocations. In *Epitaxial Growth* (Edited by J. W. Matthews), Part B, Chapter 8. Academic Press, New York.

Mullins, W. W. and Sekerka, R. F. (1964). Stability of a planar interface during solidification of a dilute binary alloy. *J. Appl. Phys.* **35,** 444–451.

Muskhelishvili, N. I. (1953). *Some Basic Problems of the Mathematical Theory of Elasticity.* Noordhoff, Groningen, The Netherlands.

Nix, W. D. (1989). Mechanical properties of thin films. *Metall. Trans. A* **20A,** 2217–2245.

Pelce, P. (Ed.) (1988). *Dynamics of Curved Fronts.* Academic Press, San Diego, CA.

Rice, J. R. (1968). A path-independent integral and the approximate analysis of strain concentration by notches and cracks. *J. Appl. Mech.* **35,** 379–386.

Rice, J. R. (1992). Dislocation nucleation from a crack tip: an analysis based on the Peierls concept. *J. Mech. Phys. Solids* **40,** 239–271.

Rice, J. R. and Drucker, D. C. (1967). Energy changes in stressed bodies due to void and crack growth. *Int. J. Fracture Mech.* **3,** 19–27.

Rice, J. R. and Thomson, R. (1974). Ductile versus brittle behaviour of crystals. *Philosophical Magazine* **29,** 73–97.

Savin, G. N. (1961). *Stress Concentration Around Holes.* Pergamon Press, New York.

Spencer, B. J., Voorhees, P. W. and Davis, S. H. (1991). Morphological instability in epitaxially strained dislocation-free solid films. *Phys. Review Lett.* **67,** 3696–3699.

Srolovitz, D. J. (1989). On the stability of surfaces of stressed solids. *Acta metall.* **37,** 621–625.

Stokes, R. J. (1972). Microscopic aspects of fracture in ceramics. In *Fracture*, Vol. VII (Edited by H. Liebowitz), pp. 158–238. Academic Press, New York.
Tada, H., Paris, P. C. and Irwin, G. R. (1985). *The Stress Analysis of Cracks Handbook*. Del Research Corporation, Hellertown, PA.
Tsao, J. Y. (1993). *Materials fundamentals of molecular beam epitaxy*. Academic Press, Boston.
Van der Merwe, J. H. (1963). Crystal interfaces. Part II. Finite overgrowths. *J. Appl. Phys.* **34**, 123–127.
Van der Merwe, J. H. (1979). The role of lattice misfit in epitaxy. In *Chemistry Phys. Solid Surfaces* (Edited by R. Vanselow), Vol. II, pp. 129–151. CRC Press Inc., Boca Raton, FL.
Varner, J. R. and Frechette, V. D. (Ed.) (1986). *Fractography of Glasses and Ceramics*. The American Ceramic Society, Westerville, OH.
Wang, K. S., Yang, Y. P. and Chu, C. M. (1991). Stress analysis of an elastic body with a wavy surface. *Wear* **145**, 101–112.

APPENDIX A: A CYCLOID SURFACE UNDER ANTIPLANE SHEAR

In the text we focused on the discussion of a cycloid surface subjected to a lateral inplane tension. For comparison purposes, there may be some interest in solving the related and simpler problem of a cycloid surface under remote antiplane shear stress τ.

An antiplane elasticity problem can be stated as determining the displacement $u_3 = u_3(x_1, x_2)$ that satisfies the Laplacian equation:

$$\nabla^2 u_3 = 0. \tag{A1}$$

Here we use x_1, x_2 interchangeably with x, y to denote the inplane coordinate frame. The general solution to (A1) is usually represented by a scalar complex potential $f(z)$ as:

$$\mu u_3 = \mathrm{Im}[f(z)], \quad \sigma_{32} = \mathrm{Re}[f_{,1}(z)], \quad \sigma_{31} = -\mathrm{Re}[f_{,2}(z)], \tag{A2}$$

where $z = x_1 + ix_2$. The resultant force along an inplane curve from a fixed point to z is:

$$p = \int^z \sigma_{i3} n_i \, \mathrm{d}s = -\int^z \mathrm{Re}[f_{,i}(z)] \, \mathrm{d}x_i = -\mathrm{Re}[f(z)] + \mathrm{const.} \tag{A3}$$

The traction-free condition in the antiplane case can thus be stated as:

$$\mathrm{Re}[f(z)] = 0. \tag{A4}$$

Now consider a cycloid surface subject to a remote antiplane shear loading $\tau = \sigma_{31}^{\infty}$. Following the same mapping procedure discussed in the text for inplane loading, let the solution be written as:

$$f(\zeta) = f_{\infty}(\zeta) + f_s(\zeta), \tag{A5}$$

where $f_{\infty}(\zeta)$ is the potential of the uniform field:

$$f_{\infty} = i\tau z = i\tau[\zeta + iA \exp(-ik\zeta)]. \tag{A6}$$

Using the traction-free boundary condition:

$$\mathrm{Re}[f_{\infty}(\xi) + f_s(\xi)] = 0, \tag{A7}$$

and the fact that $f_s(\zeta)$ is analytic in the lower half ζ-plane yields:

$$f_s(\zeta) = A\tau \exp(-ik\zeta). \tag{A8}$$

The final solution for the antiplane cycloid problem is thus:

$$f(\zeta) = i\tau\zeta. \tag{A9}$$

The stress along the surface $\sigma_{3\xi}$ can be written as:

$$\sigma_{3\xi} = -i(\sigma_{32} + i\sigma_{31}) \mathrm{e}^{i\alpha} = -i\frac{f'(\zeta)}{|\omega'(\zeta)|} = \frac{\tau}{\sqrt{1 + 2Ak\cos k\xi + A^2k^2}}. \tag{A10}$$

The maximum stress at the cycloid valley is:

$$(\sigma_{3\xi})_{\max} = \tau/(1 - Ak). \tag{A11}$$

In contrast to the inplane tension case, the surface segment between two cycloid cusps is not stress-free under antiplane shear.

For $Ak = 1$, consider the local coordinate frame ζ_c at a cycloid cusp, the complex potential behaves asymptotically as:

$$f(\zeta) = i\tau\zeta_c + \text{const} = i\frac{2K_{III}}{\sqrt{2\pi}} e^{-i\pi/4}\sqrt{z_c}. \tag{A12}$$

This is just the standard mode III crack-tip field for a crack tip pointing in the negative y_c direction, with stress intensity factor:

$$K_{III} = \tau\sqrt{\lambda/2}. \tag{A13}$$

Interestingly, the stress intensity factor for an antiplane cycloid is formally identical to the inplane tension case, but is different from the corresponding mode III periodic crack problem; the latter has stress intensity factor (Tada *et al.*, 1985):

$$(K_{III})_{pc} = \tau\sqrt{\lambda}. \tag{A14}$$

For a fixed wavelength, the mode III stress intensity factor for the periodic cracks is about 40% ($\sqrt{2}-1$) larger than that for a cusped cycloid.

One can also study the strain energy variation associated with evolution of an antiplane cycloid surface. The analysis presented in the text can be easily extended to this case and interested readers can verify that the result corresponding to eqn (50) in the text is:

$$\Delta U = -w_0\pi A^2, \tag{A15}$$

where

$$w_0 = \tau^2/2\mu. \tag{A16}$$

Under the same bulk strain energy density, observe that the antiplane cycloid only releases half as much strain energy as the inplane case.

APPENDIX B: A VARIATIONAL FORMULA FOR SURFACE ENERGY

Equation (52) in the text allows one to calculate the change in surface energy associated with evolution of a cycloid surface. A variational form of the same formula, as expressed in eqn (53), is:

$$\delta E_s = \int \gamma\kappa(s)\mathbf{n}\cdot\delta\mathbf{x}\, ds. \tag{B1}$$

This formula is equivalent to eqn (52) except for the cusp case $Ak = 1$.

Letting the cycloid surface be rigidly translated a unit distance in the y direction obviously should leave the surface energy unaffected. Using eqns (B1) and (64), this invariance requires that:

$$I(Ak) = \int_0^\lambda \frac{k(Ak+\cos k\xi)(1+Ak\cos k\xi)}{(1+2Ak\cos k\xi + A^2k^2)^{3/2}}\, d\xi = 0. \tag{B2}$$

As will be shown shortly, this integral indeed vanishes as long as $Ak \neq 1$. However, in the cusp case $Ak = 1$, direct integration shows that $I(1) = 2$, violating the translational invariance requirement. Therefore, the variational formula (B1) breaks down in the cusp case. In contrast, the direct formula in eqn (52) of the text is consistent with eqn (B1) for $0 \leqslant Ak < 1$ and also has the advantage of remaining valid for $Ak = 1$.

Using the corresponding expressions for $\kappa(s)$, $\mathbf{n}\, ds$ and $\delta\mathbf{x}$ given in eqns (52), (47), (46) of the text, the variational formula for a cycloid surface can be rewritten as:

$$\delta E_s = \gamma\delta A\int_0^\lambda \frac{Ak^2(1-A^2k^2)\cos k\xi(Ak+\cos k\xi)}{\sqrt{(1+2Ak\cos k\xi + A^2k^2)^3}}\, d\xi. \tag{B3}$$

For $0 \leqslant Ak < 1$, using the integral result in eqn (B2), it is not difficult to show that eqn (52) can be reduced to eqn (53) or eqn (B1) via following steps:

$$\begin{aligned}
\delta E_\gamma &= \gamma\delta A\int_0^\lambda \frac{k(Ak+\cos k\xi)}{\sqrt{1+2Ak\cos k\xi + A^2k^2}}\, d\xi \\
&= \gamma\delta A\int_0^\lambda \frac{k(1+2Ak\cos k\xi + A^2k^2)(Ak+\cos k\xi)}{\sqrt{(1+2Ak\cos k\xi + A^2k^2)^3}}\, d\xi \\
&= \gamma\delta A\int_0^\lambda \frac{Ak^2(1+Ak+\cos k\xi)(Ak+\cos k\xi)}{\sqrt{(1+2Ak\cos k\xi + A^2k^2)^3}}\, d\xi \\
&= \gamma\delta A\int_0^\lambda \frac{Ak^2(1-A^2k^2)\cos k\xi(Ak+\cos k\xi)}{\sqrt{(1+2Ak\cos k\xi + Ak^2)^3}}\, d\xi \\
&= \int_0^\lambda \gamma\kappa(s)\mathbf{n}\cdot\delta\mathbf{x}\, ds.
\end{aligned} \tag{B4}$$

The integral in eqn (B2) can be reduced to a simpler form:

$$I(\varepsilon) = \int_0^{2\pi} \frac{(\varepsilon+\cos t)(1+\varepsilon\cos t)}{(1+2\varepsilon\cos t+\varepsilon^2)^{3/2}}\,dt, \tag{B5}$$

which is identically zero for all $-\infty < \varepsilon < \infty$ except at two points $\varepsilon = \pm 1$, where $I(1) = -I(-1) = 2$. To prove this result, it suffices to consider $\varepsilon > 0$ since one can easily verify that $I(\varepsilon) = -I(-\varepsilon)$. Making a variable change $\theta = t/2$, the integral can be rearranged into the following form:

$$I = \frac{2}{1+\varepsilon}\int_0^{\pi} \frac{1-2\sin^2\theta+m^2\sin^4\theta}{\sqrt{(1-m^2\sin^2\theta)^3}}\,d\theta, \tag{B6}$$

where

$$m = 2\sqrt{\varepsilon}/(1+\varepsilon). \tag{B7}$$

When $\varepsilon \neq 1$, $m < 1$, one may extract the following integral formulae from Gradshteyn and Ryzhik (1980),

$$\int_0^{\pi} \frac{1}{\sqrt{(1-m^2\sin^2\theta)^3}}\,d\theta = 2\frac{E(m)}{m'^2} \tag{B8}$$

$$\int_0^{\pi} \frac{\sin^2\theta}{\sqrt{(1-m^2\sin^2\theta)^3}}\,d\theta = 2\frac{E(m)-m'^2K(m)}{m'^2m^2} \tag{B9}$$

$$\int_0^{\pi} \frac{\sin^4\theta}{\sqrt{(1-m^2\sin^2\theta)^3}}\,d\theta = 2\frac{(1+m'^2)E(m)-2m'^2K(m)}{m'^2m^4}, \tag{B10}$$

where $K(m)$, $E(m)$ are the standard elliptic integrals and

$$m' = \sqrt{1-m^2} = (1-\varepsilon)/(1+\varepsilon). \tag{B11}$$

Combining eqns (B6)–(B11) leads to $I(\varepsilon) = 0$ for $\varepsilon \neq 1$.

APPENDIX C: THE STRAIN ENERGY VARIATION ASSOCIATED WITH EVOLUTION OF A CREST-LIKE CYCLOID SURFACE

In the text we studied cycloid surfaces which result in downward-pointing cusps as $Ak \to 1$. One may consider another class of cycloid surfaces that results in crested cycloids having upward-pointing cusps as $Ak \to 1$. This class is described by:

$$x = \xi - A\sin k\xi, \quad y = A\cos k\xi + \frac{A^2k}{2}, \tag{C1}$$

where the constant $A^2k/2$ is added for mass-conservation as the amplitude A varies. When Ak is somewhere between 0 and 1, eqn (C1) yields island-like shapes with somewhat sharp peaks separated by broad valleys. As $Ak \to 1$, the surface develops into periodic crests at $x = 2n\pi/k$ and $y = 3A/2$ (n is integer), a schematic diagram of which is shown in Fig. 13. The perturbation analysis of slightly undulating surfaces by Gao (1991a) indicated

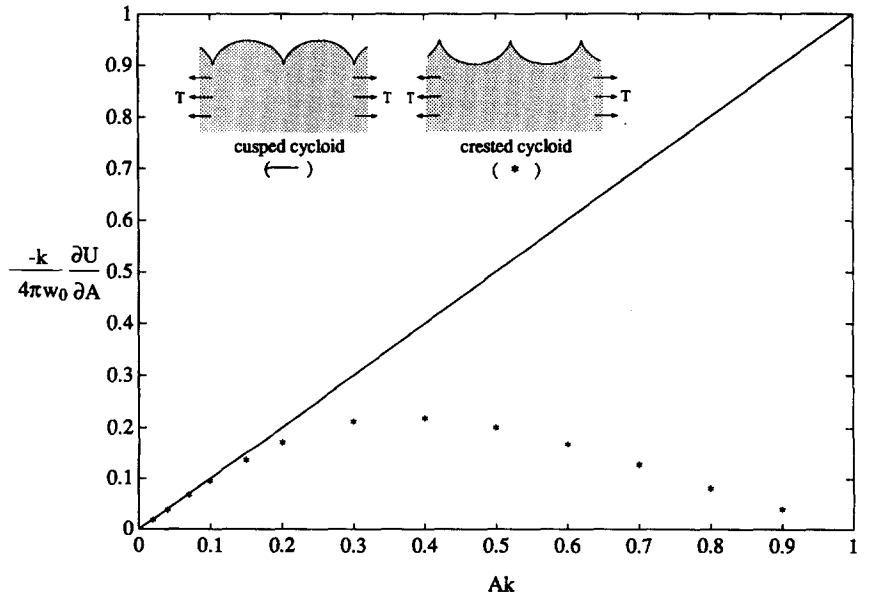

Fig. C1. The strain energy release rate, $(-k/4\pi w_0)\,\partial U/\partial A$, associated with evolution of a cusp-like and a crest-like cycloid surface.

that protruding portions along a flat surface are unloaded very quickly from the bulk stress. Thus the surface energy is expected to dominate along these portions, which precludes the formation of cycloid crests. However, to support our analysis of cusped cycloid in the text we will firmly prove the conjecture that a crest-like surface is energetically less favorable than the cusp case.

The stress problems for the crest cycloids can not be solved analytically because the mapping function involves a term $\exp(ik\zeta)$ which is not analytic, hence not conformal, in the material region. We have implemented a boundary-element type numerical method based on superposition of continuous arrays of dislocations to simulate cracks or traction-free rough surfaces (see discussions in the text). This numerical method has been used to verify the relevant analytic solutions for a cusp-like cycloid surface and can be easily used to calculate solutions for the crest-like case.

Figure 13 plots a nondimensional strain energy release rate per wavelength, $(-k/4\pi w_0)\,\partial U/\partial A$, as a function of the amplitude Ak. The accuracy of numerical solutions can be estimated by the J-integral condition in eqn (43) in the text and by test of convergence as different mesh sizes are adopted. The strain energy release rate is determined from the variational integral in eqn (44). Also plotted for comparison in Fig. 13 is the corresponding energy release rate in the cusp case, which is a straight line given in eqn (49). For small amplitudes $Ak \ll 1$, the energy release rate for both crest-like and cusp-like cycloids reduces to the corresponding perturbation result derived by Gao (1991b) since both types of surfaces asymptotically approach the same cosine curve. As the cycloid amplitude Ak increases, the energy release rate of a cusp-like cycloid increases linearly all the way up to the cusp limit. In contrast, fairly large nonlinear behavior exhibits in the energy release rate for a crest-like cycloid, resulting in a substantial energy difference between the two cases. The $\partial U/\partial A$ for a crest-like cycloid reaches maximum at about $Ak = 0.4$ and then declines as Ak further increases. The overall strain energy reduction can be calculated by integrating $\partial U/\partial A$ and is found to be over four times larger in the cusp case than in the crest case. The surface energy change is identical for both cases since for equal cycloid amplitudes the surface area change is the same. Therefore we conclude that the cusped configuration is enegetically more favorable compared to the crested case. Also, a simple singularity analysis would show that the stresses and the strain density at the crest tip are identically zero, suggesting that local surface diffusion would immediately smooth out any surface crests if they ever exist.