Superficies y Vacío 1, 29-43(1989)

INTERACTION OF LIGHT WITH CORRUGATED FILMS

P. Halevi

Departamento de Física del Instituto de
Ciencias, Universidad Autónoma de Puebla,
Apdo. Post. J-48, Puebla, Pue. 72570

ABSTRACT

First I will review some applications of corrugated thin films interacting with light. Then I will outline the theoretical approach that leads to the possible wave-mode solutions: dispersion, damping, and field amplitudes. The model employed is quite general, being valid for arbitrary corrugation profiles at the two interfaces of the film (with the same periods, however). The light may have either transverse electric or transverse magnetic polarization. The three media that comprise the thin film are assumed to be isotropic and spatially nondispersive (local). A crucial simplification arises by invoking the Rayleigh hypothesis. Apart from this approximation the general results are exact, and may be applied to a corrugated surface with a transition layer, as well as to a film that is corrugated on one or on both sides.

The general formulas are subjected to a perturbative treatment, valid for sufficiently small corrugation height. There are two cases of interest, depending on the frequency (ω) and propagation constant (α) region. If there is no interaction of wave modes (of the plane film) in the region of interest then only small corrections arise. On the other hand, dramatic changes in the wave properties occur in the neighbourhood of mode interactions (intersections of unperturbed dispersion curves). In principle, four different cases may arise: repulsion of ω versus α dispersion curves; ω-gap, α-gap; and simultaneous ω- and α-gap.

I will analyze these four possibilities and present the experimental evidence available. Of special interest is the momentum ($\hbar\alpha$) -gap. Within a gap, by definition, there can be no modes. Nevertheless, closely related complex-α "gap-modes" have led to the distributed-feedback laser. Here I will also discuss complex-ω gap modes and raise an intriguing question: are these more than merely a mathematical curiosity?

INTRODUCTION

In the past 20 years the interaction of light with corrugated films has been extensively studied in the fields of Optics and Solid State Electronics. First I will briefly review some of the more important applications. The major part of this paper deals with recent developments related to energy and momentum gaps, mainly in the context of surface plasmon polaritons.

a) Enhanced interaction of light with absorbates.

It is well known that the electric field at a charged surface is greater the smaller the local radius of curvature. In addition, the amplitude of a surface polariton wave attains maximum value at the nominal surface. These two factors explain the large value of the electric field of a polariton at a corrugated or randomly rough surface. Light scattered from an adsorbed molecule may undergo enhancement by a factor $\sim 10^5$; this is the "giant Raman scattering".[1] This type of interaction may result in applications in the field of Photochemistry.

b) Grating couplers for waveguides.

If light is incident from air at a plane-parallel (smooth) film only waveguide modes of relatively small propagation vector $\alpha(\omega)$ may be excited. This is because, by momentum conservation, one must have $\alpha(\omega)=(\omega/c)\sin\theta$, where ω/c is the wavevector of the incident photon and θ is the angle of incidence. If the surface of incidence is corrugated then the "reciprocal lattice vector" $(2\pi/d)m$ (where d is the period of the corrugation and m is an integer) also contributes to the momentum balance, which then becomes

$$\alpha(\omega) = (\omega/c)\sin\theta + (2\pi/d)m \qquad (1)$$

Be satisfying this condition light may be resonantly coupled into a TE or TM mode of an optical waveguide; a 40% efficiency has been achieved.[2] Grating couplers are also used to excite surface polaritons, which always have $\alpha(\omega) > \omega/c$.[3,4]

c) Grating-enhanced emission of light from tunnel junctions.

A tunnel junction is composed from two different metallic films separated by an insulating layer, for example Al-Al_2O_3-Ag. If a bias voltage is applied between the two metals then electrons may tunnel through the oxide barrier. As a result of inelastic processes the electrons excite surface plasmon modes characteristic of the junctions. As in b) momentum must be conserved, so the surface polaritons may decay into light only provided that eq. (1) is satisfied. This effect was discovered (for roughened surfaces) by Lambe and McCarthy,[5] and it has been also observed for corrugated surfaces. The emitted radiation has the following properties: (1) the frequency has a cutoff, because the energy of an emitted photon, $\hbar\omega$, cannot be larger than the bias voltage $|eV|$; (2) the light is TM polarized - a direct result of the polarization of surface polaritons; (3) for a given order of diffraction m, as the angle of observation θ increases the emission peaks at smaller ω; these devices emit very dim light.

d) Light emission from Schottky diodes.

The mechanism of this device is similar to that in c), however the configuration is simplified to a corrugated metal/semiconductor interface between which a bias is applied. Light emission was recently observed for an Ag/n-GaAs junction.[7]

e) Thin film lasers for integrated optics.

These minuscule light sources (~1 mm in length, ~0.1 mm in height) are composed of a stack of several semiconductor thin films. A bias voltage applied across the sample accelerates electrons and holes in opposite directions, into a certain "active" layer. The electrons enter into the conduction band and the holes occupy the valence band. These charge carriers are confined to the active layer by means of potential energy differences at the interfaces between this thin film and the adjacent ones. Subsequently the electrons and holes recombine with the emission of photons whose energy is equal to the band gap of the semiconductor. This leads to amplification of the guided wave, confined at least in part to the active layer. In order to achieve laser action the wave must oscillate back and forth, gaining energy successively. One way of achieving reflections at the two ends of the structure is by creating a corrugated interface between two of the films and operating the device in the immediate vicinity of the "Bragg condition" $\alpha(\omega) = \pm\pi/d$. Then the "incident" (+) and "diffracted" (-) waves have approximately equal amplitudea, and other orders of diffraction may be ignored.[8] Two well known types of thin film lasers are the "distributed feedback" laser[9] and the "Bragg reflector"[10] laser. In the first geometry the corrugated interface extends over the entire length of the device, while in the latter case the corrugated section (s) is (are) limited.

These examples of applications show that we are dealing with a mature field, while, on the other hand, exciting new developments are still taking place. Now I will give an outline of the theoretical work that leads to energy and momentum minigaps - a topic of considerable recent interest.

GENERAL CONSIDERATIONS

We are concerned with a periodically corrugated thin film, as shown in Fig. 1. The one-dimensional corrugation profiles, f(x) and g(x), have an arbitrary shape, the only restriction being that they must have the same period, d. These profiles are defined in such a way that $\int f(x)dx$ and $\int g(x)dx$ vanish. In general the dielectric functions $\epsilon_j(\omega)$ (j = 1,2,3) are complex, thus - depending on the sign of $Im\epsilon_j(\omega)$ - allowing for either attenuation or

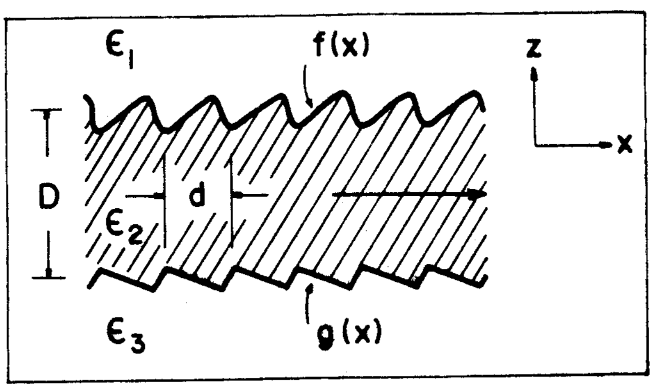

Fig. 1 The profile of a thin film, periodically corrugated at both interfaces. Guided waves propagate in the x-direction and decay exponentially away from at least one interface.

growth of the wave. The dielectric functions are assumed to be independent of the wavevector, which is to say that effects of spatial dispersion are excluded.[11] Also, all the media are taken to be nonmagnetic, $\mu_j = 1$.

In general $\epsilon_1 \neq \epsilon_3$; to this we will refer as the "asymmetric configuration". For instance, it corresponds to a thin film on a substrate. The special case $\epsilon_1 = \epsilon_3$ is also of interest. This "symmetric configuration" is realized, for example, in the case of an unsupported film (bounded by vacuum on both sides). In addition, the formalism is readily applicable to a corrugated surface by taking $\epsilon_2 = \epsilon_3$.

The corrugation profiles may also assume particular forms of interest. If $f(x) = g(x) \equiv 0$ then we are left with a plane-parallel or smooth film.[12] The case $g(x) \equiv 0$, that is only one corrugated interface, has been studied by Mata-Mendez and the author.[13,14] In practice both interfaces frequently assume identical profiles; then $f(x) \equiv g(x)$. The generalization to arbitrary f(x) and g(x), as in Fig. 1, has been performed recently by Lopez-Carrillo and the author.[15]

Our calculations are valid for either TE or TM modes. In the first case of polarization the electric field of the wave points in the y-direction, that is, it is parallel to the grooves of the corrugation. For TM modes it is the magnetic field of the wave that is parallel to the grooves. Depending on the dielectric properties of the three media and on the polarization, the propagating modes may have volume (bulk) or surface character. Volume modes have a constant amplitude inside the film; frequently they are called "waveguide modes". On the other hand the amplitude of surface modes decreases exponentially as we move away from the interfaces toward the interior of the film. In either case - bulk and surface modes - the wave is "guided". By this I mean that its amplitude decreases exponentially outward, at least on one side of the film. If the wave is guided by only one interface than it loses energy by radiation at the other interface. Such a mode is termed "radiative". Both collisional and radiative damping cause modifications in the properties discussed above.

OUTLINE OF THEORETICAL WORK

In the absence of corrugation all the components of the electromagnetic field are proportional to the factor

$$e^{i(\vec{q}_j \cdot \vec{r} - \omega t)} = e^{i(\alpha x + \beta_j z - \omega t)}, \quad j=1,2,3 \quad (2)$$

Here $\vec{q}_j$ is the wavevector and α and β_j are its components in the direction of propagation (x) and the direction perpendicular to the film (z). I have chosen, without limitation of generality, the y-component of $\vec{q}_j$ as zero. This corresponds to selecting the xz plane as the "sagittal" plane, that is, in the related problem of excitation this should be the plane of incidence. The "propagation constant" α has the same value in the three layers that comprise the thin film. Of course this must be so because otherwise it would be impossible to satisfy any continuity relation for the electromagnetic fields at the two interfaces for arbitrary x. The wave equation, applied to a medium of dielectric function $\epsilon_j(\omega)$, implies that the equation

$$\alpha^2 + \beta_j^2 = (\omega/c)^2 \epsilon_j(\omega), \quad j=1,2,3 \quad (3)$$

must be satisfied. Thus β_j depends on the dielectric properties of medium j. For a bounded wave β_j is an imaginary quantity; then the right side of eq.(2) is proportional to $\exp(\pm|\beta_j|z)$. With the sign properly chosen the amplitude of the wave decreases exponentially away from an interface, and the quantity $1/|\beta_j|$ is termed "attenuation length".

In the presence of a periodic corrugation eq. (2) is no more a solution of the wave equation applied to our thin film. The reason for this is that the wave is modulated by the periodicity of the system. Let us denote by $U_j(x,z)$ the y ("transverse") component of the electromagnetic field in layer j (the temporal factor $\exp(-i\omega t)$ is disregarded). For TE modes U_j is the electric field, while for TM modes it is the magnetic field. The modulation implies solutions of the form

$$U_j(x,z) = u_j(x,z)e^{i\alpha x}, \quad j=1,2,3 \quad (4)$$

$$u_j(x+d,z) = u_j(x,z) \quad (5)$$

That is, $u_j(x,z)$ are periodic functions in the x-direction, and the period is equal to that of the corrugation. This being the case, the functions $u_j(x,z)$ may be expanded in Fourier series. Then the solutions must have the form

$$U_j(x,z) = \sum_{m=-\infty}^{\infty} F_m^{(j)}(z)e^{i\alpha_m x}, \quad j=1,2,3 \quad (6)$$

where

$$\alpha_m = \alpha + (2\pi/d)m, \quad m=0,\pm1,\pm2,\ldots \quad (7)$$

Eq. (6) satisfies the Bloch (Floquet) theorem

$$U_j(x+d,z) = U_j(x,z)e^{i\alpha d} \quad (8)$$

The meaning of eq. (7) is that α is not the only permitted propagation vector, as it was the case of a smooth film. It may be shifted by an arbitrary "reciprocal lattice vector", $(2\pi/d)m$. The decomposition of the wave into partial waves "m" is equivalent to the phenomenon of diffraction. The number of partial waves that contribute to the sum in eq. (6) depends on the corrugation profiles f(x) and g(x). If we denote their Fourier coefficients by $\hat{f}(m)$ and $\hat{g}(m)$, they are nonvanishing only for certain values of m. For instance, in case of a sinusoidal profile the contributing diffraction orders are $m = \pm1$. The sharper the asperities are the more higher harmonics m one has.

Next I turn to the question of determination of the function $F_m^{(j)}(z)$. Because $U_j(x,z)$ must satisfy the Helmholtz equation,

$$\nabla^2 U_j + (\omega/c)^2 \epsilon_j U_j = 0 \quad (9)$$

we get

$$\sum_{m=-\infty}^{\infty} \left[\frac{d^2F_m^{(j)}}{dz^2} + (\beta_m^{(j)})^2 F_m^{(j)}(z)\right] e^{i\alpha_m x} = 0 \quad (10)$$

$$\beta_m^{(j)} = (\epsilon_j \omega^2/c^2 - \alpha_m^2)^{1/2} \quad (11)$$

First consider the regions of z <u>outside the "selvedges"</u>, by which I mean the two corrugated sections. Because eq. (10) must be satisfied for arbitrary x, we conclude that

$$d^2F_m^{(j)}/dz^2 + (\beta_m^{(j)})^2F_m^{(j)}(z) = 0 \qquad (12)$$

This equation has the solution

$$F_m^{(j)}(z) = A_m^{(j)}e^{i\beta_m^{(j)}z} + B_m^{(j)}e^{-i\beta_m^{(j)}z}$$

$$j = 1,2,3 \qquad (13)$$

with as yet undetermined coefficients $A_m^{(j)}$ and $B_m^{(j)}$. Unfortunately the same argument cannot be applied within the selvedge regions because, for a given z, satisfying eq. (10) for arbitrary x would imply crossing from one medium to another, thus ϵ_j would effectively depend on x. Inside the selvedge regions the exact solutions are much more complicated than eq.(13). The calculation is based on Green's theorem,[16-18] which leads to a reformulation of the problem in terms of the "extinction theorem"[19]: ficticious sources, defined in terms of the field U and its normal derivative on a closed surface, cause the total field inside the surface to vanish (be "extinguished").

I will not pursue the rigorous theory of diffraction because of its mathematical complexity. We will proceed on the basis of the "Rayleigh hypothesis", which is really an assumption or an approximation, rather than a "hypothesis". The supposition is that the electromagnetic fields in the selvedge region are given by the same expressions as the fields outside this region. In order words, eq. (13) is assumed to be valid for all z. Intuitively I would expect that the approximation is good provided that the volume of the selvedge region is small compared to the volume of the film, that is

$$\int_0^d |f(x)|dx + \int_0^d |g(x)|dx << Dd$$

This implies small corrugation heights in comparison to the film thickness. For a sinusoidally corrugated surface Petit[20] found $h/d < 0.142$ as the condition of validity of the hypothesis. In practice acceptable results are obtained even when this condition is violated,[18] and the general concensus seems to be that the Rayleigh hypothesis works well even for rather deep non-harmonic corrugation profiles.

In eq. (13) $\beta_m^{(j)}$ is the z-component of the wavevector, as given by eq. (11) for the corresponding order of diffraction m. Thus the wavevector of a partial wave m is $\hat{x}\alpha_m \pm \hat{z}\beta_m^{(j)}$. According to eq. (6) the solutions are linear combinations of all the waves $m = 0, \pm1, \pm2, \ldots$ One case of interest is a wave which is bound to one of the interfaces, that is its amplitude is exponentially decreasing outward. An exponentially increasing wave is, of course, nonphysical. Another situation that we will deal with is a radiating mode. Again, the radiation must be directed outward, otherwise we would be confronted with a nonexistent source of radiation at infinity. In both cases there exists only one term in eq. (13) for j = 1 and j = 3. Then I choose

$$B_m^{(1)} = A_m^{(3)} = 0 \qquad (14)$$

For a bounded mode, we must have $\mathrm{Im}\beta_m^{(1,3)} > 0$, assuring that $U_1 \to 0$ for $z \to \infty$ and $U_3 \to 0$ if $z \to -\infty$. Similarly, for a radiative mode, if we choose $\mathrm{Re}\beta_m^{(1,3)} > 0$ then the real part of the wavevector is directed away from the film.

We are now left with the still undetermined constans $A_m^{(1)}$, $A_m^{(2)}$, $B_m^{(2)}$, and $B_m^{(3)}$ in eq. (13). Two of these, $A_m^{(1)}$ and $B_m^{(3)}$, may be eliminated by applying the boundary conditions at the two interfaces, $z = (1/2)D + f(x)$ and $z = (-1/2)D + g(x)$. The conditions are the continuity of the components, parallel to the interface, of the $\vec{E}$ and $\vec{H}$ fields. For both TE and TM polarizations they may be expressed as the continuity of $U_j(x,y)$ and of the quantity $(1/V_j)\partial U_j/\partial n$, where

$$V_j = \begin{cases} 1 \text{ for TE polarization} \\ \epsilon_j \text{ for TM polarization} \end{cases} \qquad (15)$$

and $\partial/\partial n$ is the derivative normal to the (corrugated) interface. Explicitly, the boundary conditions are

$$U_1\left[x,(1/2)D+f(x)\right] = U_2\left[x,(1/2)D+f(x)\right]$$

$$\frac{1}{V_1}\left[\frac{\partial U_1}{\partial n}\right]_{z=(1/2)D+f(x)} = \frac{1}{V_2}\left[\frac{\partial U_2}{\partial n}\right]_{z=(1/2)D+f(x)}$$

$$U_3\left[x,(-1/2)D+g(x)\right] = U_2\left[x,(-1/2)D+g(x)\right]$$

$$\frac{1}{V_3}\left[\frac{\partial U_3}{\partial n}\right]_{z=(-1/2)D+g(x)} = \frac{1}{V_2}\left[\frac{\partial U_2}{\partial n}\right]_{z=(-1/2)D+g(x)} \tag{16}$$

The elimination procedure is due to Toigo et al.[20] Finally we are left with two equations that involve only the field amplitudes in the thin film, $A_m^{(2)}$ and $B_m^{(2)}$:[15]

$$\sum_{n=-\infty}^{\infty}\left[M_{11}(m,n)A_n^{(2)} + M_{12}(m,n)B_n^{(2)}\right] = 0$$

$$\sum_{n=-\infty}^{\infty}\left[M_{11}(m,n)A_n^{(2)} + M_{22}(m,n)B_n^{(2)}\right] = 0 \tag{17a}$$

The matrices M_{ij} are complicated functions of the parameters of the problem and I will not write them out explicitly here. In a compact form eq. (17a) may be rewritten as

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}\begin{pmatrix} A^{(2)} \\ B^{(2)} \end{pmatrix} = 0 \tag{17b}$$

The vectors $A^{(2)}$ and $B^{(2)}$ have infinite dimensionality, and the matrices M_{ij} have an infinite number of rows and columns. The condition for a non-trivial solution of eq. (17) is that

$$\text{Det } M \equiv \begin{vmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{vmatrix} = 0 \tag{18}$$

This equation relates the frequency ω and the propagation constant α, so it is the dispersion relation. It also depends on the parameters ε_j, D, and d, and on the Fourier coefficients of the corrugations $\hat{f}(m)$ and $\hat{g}(m)$.

DISPERSION RELATION: SURFACE PLASMONS

In principle the dispersion relation eq. (18) determines the essential properties of electromagnetic modes in a corrugated film: $\omega(\alpha)$ or $\alpha(\omega)$, the phase velocity ω/α and the group velocity $d\omega/d\alpha$, and the dissipative loss or gain - as the case may be - given by $\text{Im}\,\omega(\alpha)$ or by $\text{Im}\,\alpha(\omega)$. In addition, once the dispersion relation has been solved one may calculate the field amplitudes $A_n^{(2)}$ and $B_n^{(2)}$ from eq. (17a) in terms of one given amplitude, say $A_o^{(2)}$. Then in turn, the amplitudes $A_n^{(1)}$ and $B_n^{(3)}$ may be also calculated. The undetermined amplitude, $A_o^{(2)}$, is arbitrary, and this corresponds to the fact that the total energy of the wave is also arbitrary. No need to say, in practice the infinite number of equations (17a) must be curtailed to a minimum number that gives sufficiently good convergence.

Now I will focus the discussion on surface plasmons coupled to light - "surface plasmon polaritons".[3,4] However, much of the discussion will hold good for other excitations. In the case of a plane metallic surface these electromagnetic modes propagate only for TM polarization and frequencies such that $\varepsilon(\omega) < -1$. Let us represent the metal by the simple Drude dielectric function,

$$\varepsilon(\omega) = 1 - \omega_p^2/\omega^2 \tag{19}$$

where $\omega_p = (4\pi ne^2/m)^{1/2}$ is the plasma frequency, defined in terms of the density n of charge carriers and their effective mass m. The condition $\varepsilon(\omega) < -1$ implies that the limiting frequency is $\omega_p/\sqrt{2}$, as shown in Fig. 2(a). Explicitly, the dispersion relation is[21]

$$\alpha(\omega) = \frac{\omega}{c}\left[\frac{\varepsilon(\omega)}{1+\varepsilon(\omega)}\right]^{1/2} \tag{20}$$

It is clear that the dispersion curve is located to the right of the vacuum light-line, that is $\alpha(\omega) > \omega/c$.

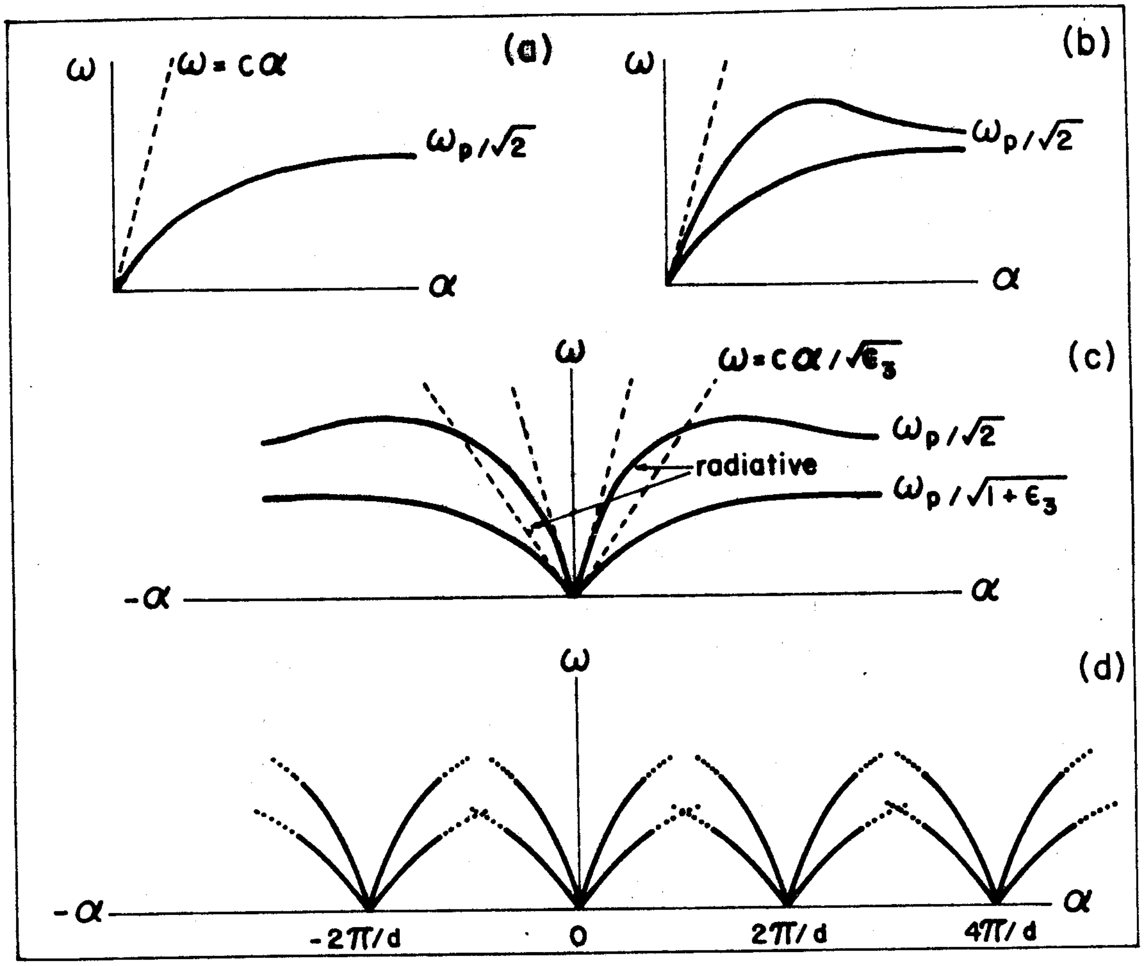

Fig. 2 Surface-plasmon polariton dispersion relations for a metal characterized by the dielectric function $\varepsilon(\omega) = 1 - \omega_p^2/\omega^2$. (a) Metallic surface; (b) metallic thin film in vacuum; (c) thin film on a dielectric substrate; (d) corrugated film on a substrate in the limit of vanishing corrugation height. (For clarity only the lower part of the modes are displayed). There is an infinite number of partial modes m.

Next I turn to a metallic thin film (still smooth) in vacuum - an unsupported film. There are now two modes, Fig. 2(b), both to the right of the light-line and terminating at $\omega_p/\sqrt{2}$. The (transverse) magnetic field of the wave, H(z), is even (odd) for the upper (lower) mode. In the limit $D \to \infty$ both modes approach each other untill they coincide with the single mode of Fig. 2(a). On the other hand, in the limit $D \to 0$ the upper mode approaches the vacuum light-line while the lower mode disappears. The maximum of the upper mode only develops for a sufficiently thin film.

In the more complicated case of a metallic film on a dielectric substrate, Fig. 2(c), the upper mode is associated with the vacuum side, and is very similar to the upper mode in Fig. 2(b). The lower mode is related to the film/substrate interface. It lies to the right of the light-line in the dielectric, $\omega = c\alpha/\sqrt{\varepsilon_3}$, and it approaches the limit $\omega_p/\sqrt{1+\varepsilon_3}$. It should be stressed, however, that for intermediate frequencies $\alpha(\omega)$ for either mode depends on both ε_1 and ε_3 and does not really "belong" to one or the other interface. The waves may propagate in the positive ("to the right") or in the negative ("to the left") direction of the x axis. Thus α may be positive or negative for the same ω, as shown in the figure.

The region between the two light-lines in Fig. 2(c) is of special interest. For any α therein,

$$\omega/c < \alpha < \sqrt{\epsilon_3}\omega/c \qquad (21)$$

Then it follows from eq. (3) that β_1 is an imaginary quantity and that β_3 is real, provided that the ϵ_j are real. This is to say that the wavefields decay exponentially away from the film on the vacuum side, however they have real wave character in the dielectric substrate. As a consequence of this radiation the upper surface mode is leaking energy onto the substrate side, and α becomes complex (then the coordinate in Fig. 2(c) should be really marked Reα). Provided that the film is not too thin $\left[\exp(-2\mathrm{Im}\beta_2 D) << 1\right]$ the radiative loss in small (Imα << Reα). The dispersion relation for the (plane) thin film system (TM polarization) is

$$\left(1 - \frac{\epsilon_2}{\epsilon_3}\frac{\beta^{(3)}}{\beta^{(2)}}\right)\left(\beta^{(2)} - \frac{\epsilon_2}{\epsilon_1}\beta^{(1)}\right)e^{i\beta^{(2)}D}$$

$$-\left(1 + \frac{\epsilon_2}{\epsilon_3}\frac{\beta^{(3)}}{\beta^{(2)}}\right)\left(\beta^{(2)} + \frac{\epsilon_2}{\epsilon_1}\beta^{(1)}\right)e^{-i\beta^{(2)}D}$$

$$\equiv \gamma(\omega,\alpha) = 0 \qquad (22)$$

This equation governs the propagation of volume (waveguide) as well as surface waves. It is a simple matter to deduce the dispersion relation for an interface (1/2 and 3/2) in the limit $D \to \infty$ or (1/3) in the limit $D \to 0$. Eq. (22) is a transcendental equation and must be solved numerically in general.

The next step is to introduce an extremely small corrugation, of negligible height. While we may take $f(x) \to 0$ and $g(x) \to 0$, we still do take in account the all-important effects of periodicity. This is the "empty-lattice approximation" of Solid State Physics, where the limit $V(\vec{r}) \to 0$ is taken for the potential experienced by an electron. It is well known that, in a periodic system, the wavevector is defined only up to a reciprocal lattice vector. Applied to our case this, in fact, is the content of eqs. (6), (7). A propagation constant $(\alpha + 2\pi m/d)$ for $m \neq 0$ is meaningful as α itself. Then Fig. 2(c) must be repeated periodically an infinite number of times, as drawn schematically in Fig. 2(d). Therefore the dispersion relation in the empty-lattice approximation is

$$\gamma(\omega,\alpha + 2\pi m/d) \equiv \gamma(\omega,\alpha_m) = 0, \quad m = 0, \pm 1, \pm 2, \ldots \qquad (23a)$$

or

$$\ldots\gamma(\omega,\alpha_{-2})\gamma(\omega,\alpha_{-1})\gamma(\omega,\alpha_o)\gamma(\omega,\alpha_1)\gamma(\omega,\alpha_2)\ldots = 0 \qquad (23b)$$

As a matter of fact, in the limit $\hat{f}(m)$ 0, $\hat{g}(m) \to 0$, the determinant in eq. (18) reduces to the product of its diagonal elements. Then eq. (18) becomes identical with eq. (23b).

Let us focus on only two orders of diffraction, m = p and m = q. The dispersion relations corresponding to these are $\gamma(\omega,\alpha_p) = 0$ and $\gamma(\omega,\alpha_q) = 0$. They are centered, respectively, at $\alpha = -2\pi p/d$ and $\alpha = -2\pi q/d$, and plotted in Fig. 3. The cases of an asymmetric (a) and of a symmetric (b) metallic film are both shown. Obviously intersections occur at the midpoint, $\alpha = -\pi(p+q)/d$. This is a Brillouin Zone (BZ) boundary. For a corrugated <u>surface</u> all intersections are necessarily of this type, as is obvious from a glance at Fig. 2(a). However, for a thin film - symmetric, as well as asymmetric - we observe intersections that do not coincide with a BZ boundary. Some of these intersections in Fig. 3 (encircled) occur outside the light-cones, in the non-radiative region. In Fig. 3(a) there are also intersections (unmarked) in the radiative region between the two light-lines. These will not be considered.

Now, if we allow for however small, finite corrugation height, important qualitative changes will occur. For $\hat{f}(m) \neq 0$ or $\hat{g}(m) \neq 0$, the intersections of curves in Fig. 3 cannot occur. The reason for this is that an intersection of two curves would correspond to a separate dispersion relation for each of them, that is to a factorizable dispersion equation. However, the fact is that it is not possible to factorize eq. (18) for finite $\hat{f}(m)$ and/or $\hat{g}(m)$. The conclusion is that the dispersion characteristics must radically change so as to remove the degeneracy associated with a mode-intersection. This situation is addressed perturbationally in the following section.

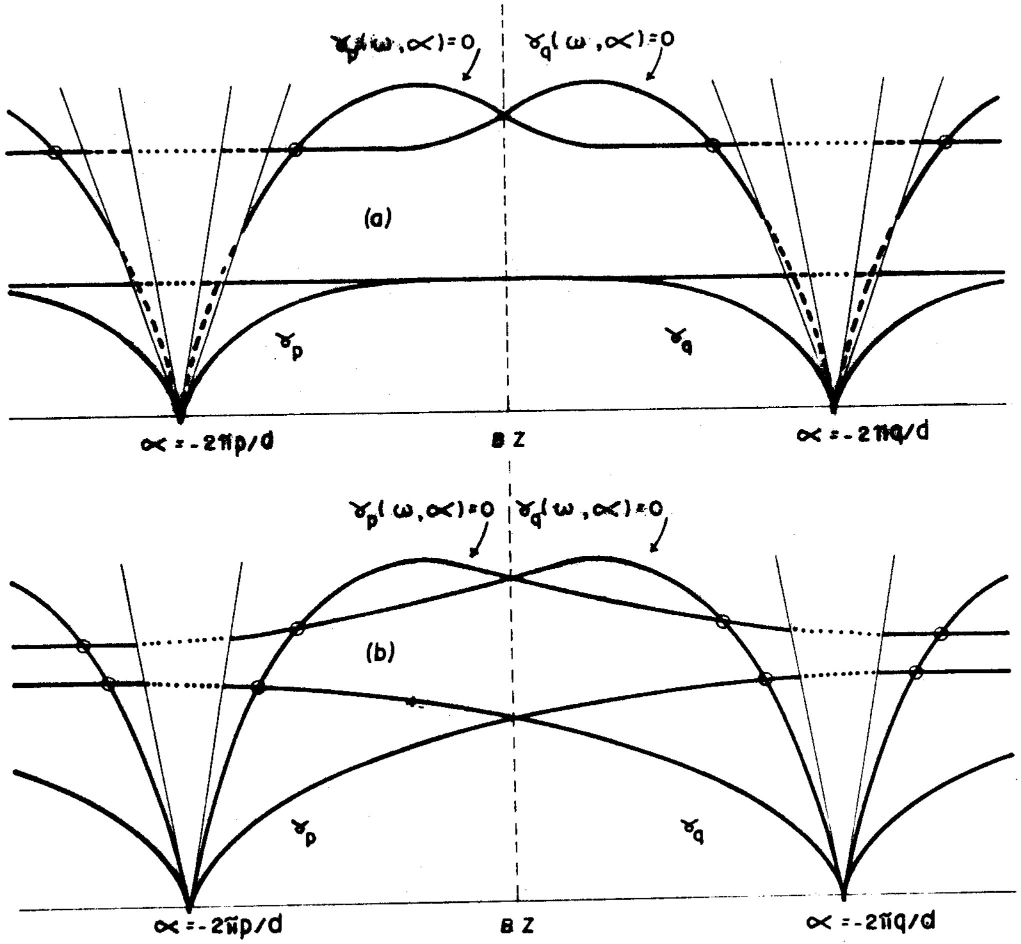

Fig. 3 (a) Detailed rendering of two of the partial modes of Fig.2(d): $m = p$ and $m = q$. (b)Same as (a), however now the metallic film is bounded by identical media. The mode-intersections are removed for a finite height of corrugation. From Halevi and Mata-Mendez.[14]

PERTURBATIONAL RESULTS

It is assumed that the amplitudes of the periodic functions $f(x)$ and $g(x)$ are much smaller than other characteristic lengths in our problem. Starting with the general dispersion equation, eq. (18), the explicit dispersion relation $\omega(\alpha)$ or $\alpha(\omega)$ is calculated in lowest order in the Fourier coefficients $\hat{f}(m)$ and $\hat{g}(m)$. This amounts to a first-order perturbational theory.

The calculation turns out to be sharply different far away from a mode intersection and in its immediate vicinity. There is not much to be said about the first situation; far away from an intersection the corrections to $\omega(\alpha)$ or to $\alpha(\omega)$ are linear in $|\hat{f}(m)|^2$, $|\hat{g}(m)|^2$, and $\hat{f}(m)\hat{g}(-m)$, where m takes all the values from $-\infty$ to $+\infty$.[15] Therefore, in these regions, the dispersion curves differ very little from those of a smooth film, Fig. 2(b) or (c). Even in the relatively simple situation that one interface of the film is plane, $g(x) \equiv 0$,[14] it is very difficult to determine the sign of the shift. Inagaki et. al.[22] measured positive shifts $\Delta\alpha$, however they did not confirm the proportionality to the height of the corrugation squared.

Now let us turn our attention to the neighbourhood of an intersection of two zero-order modes p and q. Not surprisingly, their amplitudes are large in comparison to those of other modes. Thus the dispersion relation is completely determined in terms of the diffraction orders m = p and m = q, this provided that $\hat{f}(p-q)$ and $\hat{g}(p-q)$ do not vanish both. If these expressions happen to vanish then second-order perturbation theory is necessary. This complicated calculation has not been attempted. Here I will limit the discussion to the relatively simple case of a thin film that is corrugated only at one side: $g(x) \equiv 0$ and $\hat{f}(p-q) \neq 0$.[14] Denoting by $(\tilde{\omega},\tilde{\alpha})$ the point of intersection of the unperturbed curves, the shifts $(\omega-\tilde{\omega})$ as a function of $(\alpha-\tilde{\alpha})$, and of $(\alpha-\tilde{\alpha})$ as a function of $(\omega-\tilde{\omega})$ are

$$\omega-\tilde{\omega} = A(\alpha-\tilde{\alpha}) \pm \left[B+C^2(\alpha-\tilde{\alpha})^2\right]^{1/2} \quad (24a)$$

$$\alpha-\tilde{\alpha}=A(A^2 - C^2)^{-1} \pm \left[B(A^2-C^2)^{-1} + C^2(A^2-C^2)^{-2}(\omega-\tilde{\omega})^2\right]^{1/2} \quad (24b)$$

The constants A, B, and C are complicated, but explicit functions of $\tilde{\alpha}_p,\tilde{\alpha}_q,\tilde{\beta}_p^{(j)}$, $\tilde{\beta}_q^{(j)}$, and the derivatives of $\gamma(\tilde{\omega},\tilde{\alpha}_p)$ and $\gamma(\tilde{\omega},\tilde{\alpha}_q)$ with respect to $\tilde{\omega}$ and $\tilde{\alpha}$. It is important to stress that A, B and C are entirely given in terms of the unperturbed quantities in the empty-lattice approximation; the corrugation only enters through B and, in fact,

$$B \propto |\hat{f}(p-q)|^2 \quad (25)$$

As mentioned before, in Fig. 3 only intersections in the nonradiative zone are considered. Thus radiative loss is absent.

In addition, I will assume that $\mathrm{Im}\,\epsilon_j << |\mathrm{Re}\,\epsilon_j|$, thus neglecting collisional damping as well. This situation is characterized by real $\tilde{\alpha}_p$, $\tilde{\alpha}_q$, and imaginary $\tilde{\beta}_p^{(j)}$, $\tilde{\beta}_q^{(j)}$. Then A, B, and C turn out to be real, which greatly facilitates the interpretation of eq. (24). Relatively far from the intersection, where $(\alpha-\tilde{\alpha})^2 >> |B|/C^2$, eq. (24a) has the asymptotic limit.

$$\omega - \tilde{\omega} = (A \pm C)(\alpha - \tilde{\alpha}) \quad (26)$$

An examination of the quantities (A+C) and (A-C) reveals that they are simply the slopes of the curves $\gamma(\tilde{\omega}, \tilde{\alpha}_p) = 0$ and $\gamma(\tilde{\omega}, \tilde{\alpha}_q) = 0$ at the intersection point. The expression $(A^2 - C^2)$ is equal to the product of the slopes, and this determines the behavior of the square root in eq. (24b). If the slopes of the asymptotes have opposite (equal) signs then $(A^2 - C^2)$ is negative (positive). As for the behavior of the square root in eq. (24a), it depends on the sign of B. Eq. (25) suggests that this sign is associated with the nature of the interaction between the modes p and q. Unfortunately it is extremely difficult to derive general criteria for the sign of the proportionality constant in eq. (25). All that I can say with certainty is that, for an intersection at a BZ boundary (a "symmetric" intersection), B is positive.

For the sake of generality I will consider the case $B < 0$, as well as $B > 0$. Combined with the cases $(A^2 - C^2) > 0$ and $(A^2 - C^2) < 0$ eq. (24) then gives four possibilities, shown in Fig. 4.

For $B > 0$ and energy ($\hbar\omega$) gap obtains if the slopes of the zero-order curves have different signs (Fig. 4a), and a repulsion of the curves occurs for equal signs of the slopes (Fig. 4b). In general the gap is "indirect", by which I mean that the minimum of the upper branch does not lie exactly above the maximum of the lower branch. The size of the gap is proportional to $|\hat{f}(p-q)|$, so this is a larger effect than the $|\hat{f}(p-q)|^2$ perturbation obtained far away from an intersection region. Still, typical values lie between 10 and 100 meV, so we refer to "minigaps".

Energy minigaps for a corrugated metallic surface (at BZ boundaries) have been observed as long as 21 years ago.[23] They have been also found in other periodic structures: a two-dimensional electron gas with spatially modulated charge density,[24] and a superlattice.[25] Chen et. al.[26] and Koteles et. al.[27] performed detailed optical studies of the minigap region, including effects of the corrugation profile and deviation from perfect periodicity. According to calculations by Weber and Mills,[28] Attenuated-Total-Reflectance (ATR) spectroscopy yields energy gaps provided that the frequency is scanned at a constant angle of incidence; however an angular scan for a fixed frequency gives a distorted picture of the dispersion.

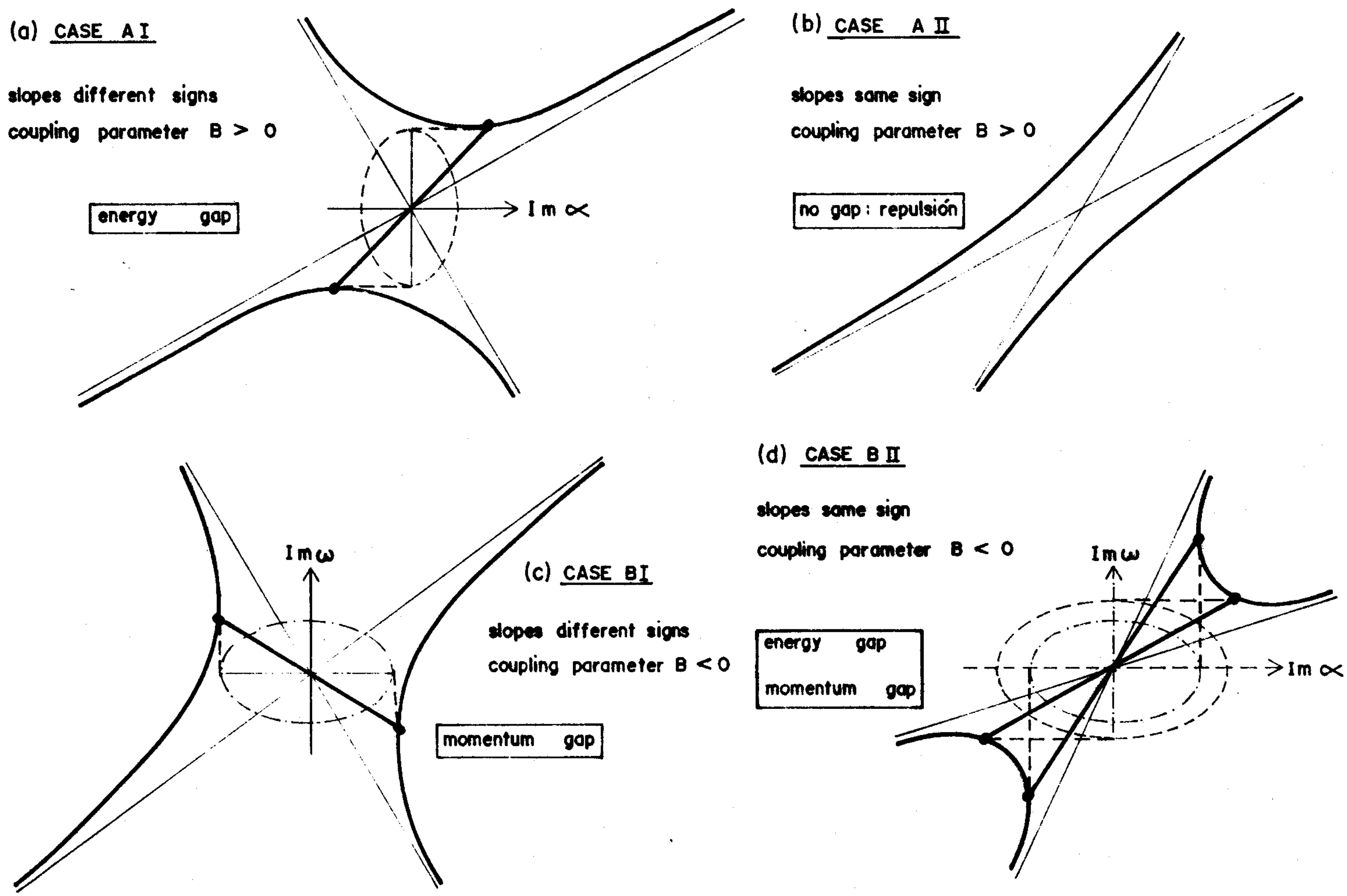

Fig. 4 The four possible outcomes of mode interaction. (a)energy gap (non-direct, in general); (b)mode-repulsion; (c)momentum gap (non-direct, in general); (d)combined energy and momentum gaps. The inclined straight lines are asymptotes (non-perturbed solutions). In cases (a), (c), and (d) solutions within the gaps are possible; these have a complex propagation constant α or a complex frequency ω. From Halevi and Mata-Mendez.[14]

The same authors[29] also claim that coupling of the upper and lower modes in Fig. 3 is probably too weak to be observed. Also worth mentioning is an interesting series of experiments in which Gruhlke et. al[30] utilized the coupling phenomenon in order to obtain radiation at one side of a thin metallic film from fluorescent molecules on the other side.

In the situations mentioned above the unperturbed modes have slopes of opposite signs - as they should in order to produce energy gaps. For slopes of equal signs Pockrand[31] found that the coupling of surface-plasmon polaritons at the two interfaces of a corrugated silver film produced a repulsion - as we may expect from Fig. 4b.

Auto et. al.[32] performed an exact numerical calculation of the normal modes for a thin metallic film corrugated on one side; the Rayleigh hypothesis was employed. The dispersion relation is conveniently plotted in one half of the "reduced-zone scheme", $0 \leq \alpha \leq \pi/d$. That is, for $\alpha > \pi/d$, the lower and upper surface-plasmon polaritons are both folded, accordion-style, into the first BZ. Then energy gaps appear at the boundary $\alpha = \pi/d$ of this zone.

MOMENTUM GAPS

If $B < 0$ a momentum ($h\alpha$) gap occurs for different signs of the slopes (Fig. 4c), and a simultaneous energy and momentum gap is predicted for the same sign of slopes (Fig. 4d). Again, these minigaps are proportional to $|\hat{f}(p-q)|$, and are in general indirect.

The possibility of momentum gaps is particularly exciting. As far as concerns the calculation reviewed here the crucial question is: is it possible -and how- to realize a situation characterized by a negative coupling constant B?

In the early 70's Kogelink and Shank[33] studied a different periodic geometry, namely a waveguide structure in which either the index of refraction or the gain constant (for a medium exhibiting gain, rather than absorption) was modulated. In the first case the mode coupling results in an energy gap, while in the second case a momentum gap is obtained. C. Vasallo[34] pointed out that two optical waveguides that are coupled in a limited region of space (so that there is an overlap of the electromagnetic fields) give rise to precisely the four possibilities depicted in Fig. 4.

Returning to corrugated structures, Kroó et. al.[35] analyzed light emitted from metal-oxide-metal tunnel junctions, according to the same mechanism already surveyed in part (c) of the Introduction. Momentum gaps appeared at the center of the BZ (the "Γ point"), as reproduced in Fig. 5. A subsequent study by Heitman et. al.[36] revealed that the nature of the anomaly - energy or momentum gap - depends on the corrugation profile and on the form of the excitation. Thus, according to the authors[36] the coupling to the photon field is essential for the existence of momentum gaps in the spectrum of corrugated tunnel junctions. Recent calculations of reflectivity by Celli et. al.[37] and by Tran et. al.[38] have indeed confirmed the importance of the corrugation profile, stressing the necessity of weak coupling between the two interacting surface-plasmon polariton modes and the existence of higher harmonics in f(x).

Additional study - both experimental and theoretical - is need before a coherent understanding of momentum gaps is obtained. It is also worth pointing out that, to my best knowledge, a simultaneous energy and momentum gap (Fig. 4d) has not yet been observed.

GAP MODES

The very definition of a "gap" implies that there are no solutions - no electromagnetic modes - within the gap. Actually, this is true only as long as the definition is restricted to convergent solutions. Nevertheless, in certain situations it is useful to consider nonconvergent solutions; to these I refer as "gap modes".

An examination of eq. (24b) reveals that, for real ω, inside the energy gap there is a solution with complex α. The real part of α is given by the heavy line connecting the minimum of the upper band and the maximum of the lower band in Fig. 4a. Thus the dispersion is linear. The imaginary part of α is drawn with dashed lines; $\mathrm{Im}\alpha(\omega)$ is an elipse. The maximum value of $\mathrm{Im}\alpha$ is proportional to $|\hat{f}(p-q)|$, so usually $|\mathrm{Im}\alpha| \ll |\tilde{\alpha}|$. Now, for a given ω there are two solutions for $\mathrm{Im}\alpha$:

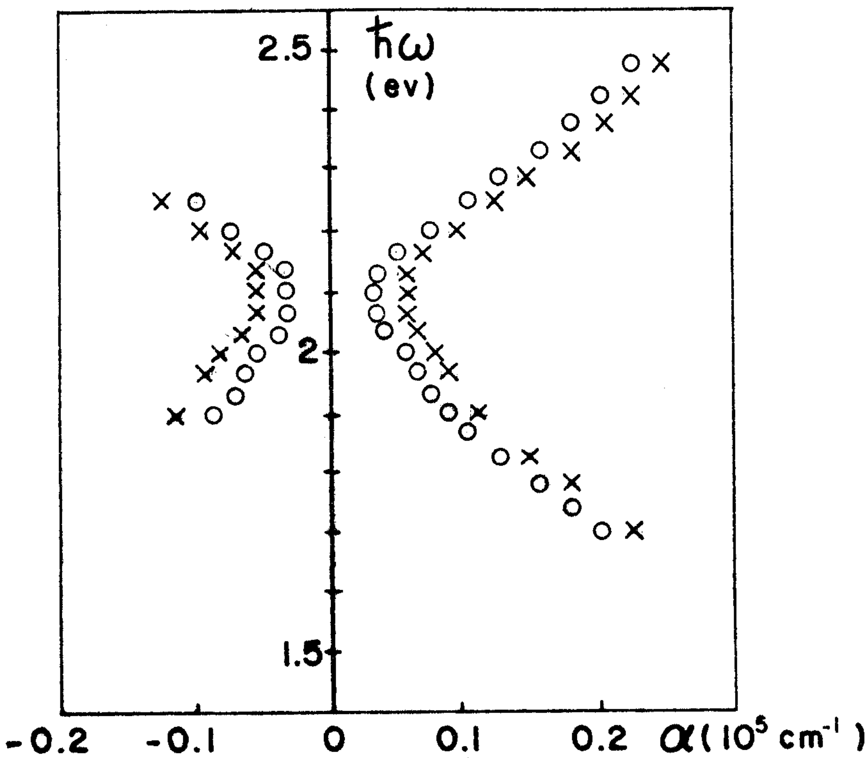

Fig. 5 Dispersion relation of surface plasmon for silver topped metal-oxide-metal diodes with period d = 555 nm and with different corrugation amplitudes h (o-h ≃ 42 nm, x-h ≃ 55 nm). Tunnel emission intensity was measured in ω = const scans at a +2.5 V bias and at room temperature. It is seen that the momentum gap increases with h. From Kroó, Szentirmai, and Félszerfalvi.[35]

$$\mathrm{Im}\alpha = \pm \left[\frac{B}{(C^2 - A^2)^{1/2}} - \frac{C^2}{C^2 - A^2} (\omega - \tilde{\omega})^2 \right]^{1/2} \quad (27)$$

Substituted in the factor $\exp(i\alpha x)$ the "+" solution diverges for $x \to -\infty$, and the "-" solution diverges for $x \to +\infty$. Then it is clear that these gap-modes are not physical for periodically corrugated films of infinitely large extension in the direction perpendicular to the grooves.

Of course, in practice one deals with structures of finite length, and then the divergence is removed. Then complex, gap solutions become perfectly acceptable. Such solutions are well known in the context of Solid State Physics,[39] and have been seen experimentally for surfaces and junctions.[40] In the case of the thin-film lasers that I reviewed in the Introduction, (e), a finite length for the structure is of fundamental importance: it leads to the threshold condition for the distributed-feedback laser.[8,33] These devices operate in the immediate vicinity of a BZ boundary, so that the gaps considered are "direct". The gap mode in Fig. 4a is a generalization to "indirect" gaps, off a BZ boundary.

Turning to the momentum gap, Fig. 4c, again there are gap solutions, now with real α and complex ω. The heavy line between the extremum points gives the real part of the dispersion relation, and we see that $(\mathrm{Re}\,\omega - \tilde{\omega})$ is proportional to $(\alpha - \tilde{\alpha})$; the slope is equal to A, which is just the algebraic average of the two asymptotic lines. The locus of the imaginary part, $\mathrm{Im}\omega(\alpha)$, is again an ellipse, and $|\mathrm{Im}\omega| \ll \tilde{\omega}$.

There are two permissible signs for $\mathrm{Im}\,\omega$, and this leads to a temporal behavior of the form

$$e^{-i\omega t} = e^{-i(\mathrm{Re}\,\omega)t}\, e^{\pm(\mathrm{Im}\,\omega)t} \quad (28)$$

for α inside the momentum gap. For a monochromatic wave the time is unlimited and this expression diverges for $t \to \pm\infty$. However, it is possible that these gap modes may be realized for waves that are not monochromatic (pulses). Such waves have a limited duration, so that the divergences would be removed.

There is a beautiful symmetry between the complex gap modes of a momentum gap and of an energy-gap. Just as limiting the spatial extent of a wave led to important applications, a limitation of its temporal extent may have useful consecuences.[14]

The last case, Fig. 4d, is particularly interesting because there may be two kinds of gap modes: one that has a real frequency and a complex propagation constant, and another with real propagation constant and a complex frequency. The real- modes posses all the properties of the energy-gap modes, Fig. 4a. The real-α modes behave in the same way as the momentum-gap modes in Fig. 4c. Whether the former or the latter waves propagate would depend on the method of excitation and detection utilized.

REFERENCES

1. W.H. Weber and G.W. Ford, Phys. Rev. Lett. 44, 1774 (1980); A. Hartstein, J.R. Kirtley, and J.C. Tsang, ibid 45, 201 (1980); P.N. Sanda, J.M. Warlaumont, J.E. Demuth, J.C. Tsang, K. Christmann, and J.A. Bradley, ibid 45, 1519 (1980); R. Reinisch and M. Nevière, Opt. Engin. 20,629 (1981); and references therein.
2. M.L.Daks, L. Kuhn, P.F. Heidnich, and B.A. Scott, Appl. Phys. Lett. 16, 523 (1970).
3. A.D. Boardman, editor, Electromagnetic Surface Modes, J. Wiley & Sons, Chichester 1982.
4. V.M. Agranovich and D.L. Mills, editors, Surface Polaritons, North-Holland, Amsterdam 1982.
5. J. Lambe and S.L. McCarthy, Phys. Rev. Lett. 37, 923 (1976); S.L. Mc Carthy and J. Lambe, Appl. Phys. Lett. 30, 427 (1977).
6. J.R. Kirtley, T.N. Theis, and J.C. Tsang, Appl. Phys. Lett. 37, 435 (1980).
7. A. Köck, W. Beinstingl, K. Berthold, and E. Gornik, Appl. Phys. Lett. 52, 1164 (1988).
8. P.K. Tien, Rev. Mod. Phys. 49, 361 (1977).
9. See, for example, H.C. Casey, Jr., S. Somekh, and M. Ilegams, Appl. Phys. Lett. 27, 142 (1975).
10. See, for example, F.K. Reinhart, R.A. Logan, and C.V. Shank, Appl. Phys. Lett. 27, 45 (1975); W.T. Tsang and S. Wang, Appl. Phys. Lett 28, 596 (1976).

11. Recently the hydrodynamic model has been applied to a corrugated metallic surface by S. Wang, R.G. Barrera, and W.L. Mochán, Phys. Rev. B40, 1571 (1989).

12. See, for instance, R.W. Alexander, Jr. R.J. Bell, and C.A. Ward in ref. 3, p.201; G.W. Zhizhin and V.A. Yakovlev in ref. 4, p. 275; K.L. Kliewer and R. Fuchs, Advan. Chem. Phys. 27, 355 (1974).

13. O. Mata-Mendez and P. Halevi, Phys. Rev. B36, 1007 (1987).

14. P. Halevi and O. Mata-Mendez, Phys. Rev. B39, 5694 (1989).

15. C. López-Carrillo and P. Halevi, to be published.

16. A.A. Maradudin, ref. 4, p. 405.

17. D. Maystre, ref. 3, p. 661.

18. D. Maystre, Progr. Opt. 21, 1 (1984).

19. E. Wolf, in Coherence and Quantum Optics, eds. L. Mandel and E. Wolf (Plenum, New York 1973), p. 339.

20. F. Toigo, A. Marvin, V. Celli, and N.R. Hill, Phys. Rev. B15, 5618 (1977).

21. See, for instance, P. Halevi, ref. 3, p. 249.

22. T. Inagaki, M. Motosuga, E.T. Arakawa, and J.P. Goudonnet, Phys. Rev. B31, 2548 (1985); 32, 6238 (1985).

23. R.H. Ritchie, E.T. Arakawa, J.J. Cowan, and R.N. Hamm, Phys. Rev. Lett. 21, 1530 (1968).

24. U. Mackens, D. Heitmann, L. Prager, J.P. Kotthaus, and W. Beinvogl, Phys. Rev. Lett. 53, 1485 (1984).

25. R. Haupt and L. Wendler, Phys. Stat. Solidi B142, 125 (1987).

26. Y.J. Chen, E.S. Koteles, and R.J. Seymour, Solid State Commun. 46, 95 (1983).

27. E.S. Koteles, Y.J. Chen, G.J. Sonek, and J.M. Ballantyne, J. Phys. 45, C5-213 (1984).

28. M.G. Weber and D.L. Mills, Phys. Rev. B34, 2893 (1986).

29. M.G. Weber and D.L. Mills, Phys. Rev. B32, 5057 (1985).

30. R.W. Gruhlke, W.R. Holland, and D.G. Hall, Phys. Rev. Lett. 56, 2838 (1986); Opt. Lett. 21, 364 (1987).

31. I. Pockrand, Opt. Commun. 13, 311 (1975).

32. M.M. Auto, G.A. Farias, and A.A. Maradudin, Surface Sci. 167, 57 (1986).

33. H. Kogelnik and C.V. Shank, Appl. Phys. Lett. 18, 152 (1971); J. Appl. Phys. 43, 2327 (1972).

34. C. Vassallo, Théorie des Guides d'ondes Électromagnétiques tome 1

(Eyrolles, Collection Technique et Scientifique des Télécommunications, Paris 1985).

35. K. Kroó, Zs. Szentirmai, and J. Félszerfalvi, Phys. Lett. A86, 445 (1981).

36. D. Heitmann, N. Kroó, C. Schultz, and Zs. Szentirmay, Phys. Rev. B35, 2660 (1987).

37. V. Celli, P. Tran, A.A. Maradudin, and D.L. Mills, Phys. Rev. B37, 9089 (1988).

38. P. Tran, V. Celli, and A.A. Maradudin, Opt. Lett. 13, 530 (1988).

39. C. Kittel, Introduction to Solid State Physics, 5th ed., Wiley and Sons, New York 1976; see Problem 5 and Fig. 12 in Chap. 7.

40. G.H. Parker and C.A. Mead, Phys. Rev. Lett. 21, 605 (1968); S. Kurtin, T.C. McGill, and C.A. Mead, ibid, 25, 756 (1970).