Diffracted evanescent wave model for enhanced and suppressed optical transmission through subwavelength hole arrays

Henri J. Lezec[1] **and Tineke Thio**[2]

[1]*ISIS, Université Louis Pasteur, 8 allée Gaspard Monge, 67083 Strasbourg, France*
[2]*Arinna LLC, Princeton NJ 08540, USA; http://www.arinnaresearch.com*

lezec@alum.mit.edu

Abstract: Transmission enhancements of order 1000 have been reported for subwavelength hole arrays in metal films and attributed to surface plasmon (SP) resonance. We show that the properly normalized enhancement factor is consistently less than 7, and that similar enhancements occur in nonmetallic systems that do not support SPs. We present a new model in which the transmission is modulated not by coupling to SPs but by interference of diffracted evanescent waves generated by subwavelength features at the surface, leading to transmission suppression as well as enhancement. This mechanism accounts for the salient optical properties of subwavelength apertures surrounded by periodic surface corrugations.

OCIS codes: (240.6690) Surface waves, (050.1940) Diffraction, (050.1220) Apertures

References and links

1. T.W. Ebbesen, H.J. Lezec, H.F. Ghaemi, T. Thio, and P.A. Wolff, "Extraordinary optical transmission through sub-wavelength hole arrays," Nature (London) **391**, 667–669 (1998).
2. T. Thio, K.M. Pellerin, R.A. Linke, T.W. Ebbesen, and H.J. Lezec, "Enhanced light transmission through a single subwavelength aperture," Opt. Lett. **26**, 1972–1974 (2001).
3. H. Raether, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, (Springer-Verlag, Berlin, 1988).
4. H.F. Ghaemi, T. Thio, D.E. Grupp, T.W. Ebbesen, and H.J. Lezec, "Surface plasmons enhance optical transmission through subwavelength holes," Phys. Rev. B **58**, 6779–6782 (1998).
5. Q. Cao and P. Lalanne, "Negative role of surface plasmons in the transmission of metallic gratings with very narrow slits," Phys. Rev. Lett. **88**, 057403 (2002).
6. H. Cao and A. Nahata, "Resonantly enhanced transmission of terahertz radiation through a periodic array of subwavelength apertures", Opt. Express **12**, 1004–1010 (2004).
7. J. Gmez-Rivas, C. Schotsch, P. Haring Bolivar, and H. Kurz, "Enhanced transmission of THz radiation through subwavelength holes," Phys. Rev. B **68**, 201306(R) (2003).
8. J. M. Steele, C. E. Moran, A. Lee, C. M. Aguirre, and N. J. Halas, "Metallodielectric gratings with subwavelength slots: Optical properties," Phys. Rev. B **68**, 205103 (2003).
9. H. Lochbihler, "Surface polaritons on gold-wire gratings," Phys. Rev. B **50**, 4795–4801 (1994).
10. T. Thio, H.F. Ghaemi, H.J. Lezec, P.A. Wolff and T.W. Ebbesen, "Surface-plasmon-enhanced transmission through hole arrays in Cr films," J. Opt. Soc. Am. B **16**, 1743–1748 (1999).
11. M. Sarrazin and J.-P. Vigneron, "Optical properties of tungsten thin films perforated with a bidimensional array of subwavelength holes," Phys. Rev. E **68**, 016603 (2003).
12. F.J. García-Vidal, H.J. Lezec, T.W. Ebbesen, and L. Martín-Moreno, "Multiple paths to enhance optical transmission through a single subwavelength slit," Phys. Rev. Lett. **90**, 213901 (2003).
13. P. Kramper, M. Agio, C.M. Soukoulis, A. Birner, F. Müller, R. B.Wehrspohn, U. Gösele, and V. Sandoghdar, "Highly directional emission from photonic crystal waveguides of subwavelength width," Phys. Rev. Lett. **92**, 113903 (2004).

#4585 - $15.00 US
Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

14. H.J. Lezec, A. Degiron, E. Devaux, R.A. Linke, M. Martín-Moreno, F.J. García-Vidal and T.W. Ebbesen, "Beaming light from a sub-wavelength aperture," Science **297**, 820–822 (2002).
15. M.J. Lockyear, A.P. Hibbins, J.R. Sambles, and C.R. Lawrence, "Surface-topography-induced enhanced transmission and directivity of microwave radiation through a subwavelength circular metal aperture," Appl. Phys. Lett. **84**, 2040–2042 (2004).
16. A. Krishnan, T. Thio, T. J. Kim, H. J. Lezec, T. W. Ebbesen, P. A. Wolff, J. Pendry, L. Martín-Moreno, and F. J. García-Vidal, "Evanescently coupled resonance in surface plasmon enhanced transmission," Opt. Commun. **200**, 1–7 (2001).
17. E. Hecht, *Optics*, (Addison-Wesley, Reading MA, 1998).
18. J.D. Kraus and K.R. Carver, *Electromagnetics*, (McGraw-Hill, New York, 1973).
19. H. A. Bethe, "Theory of diffraction by small holes," Phys. Rev. **66**, 163–182 (1944); C. J. Bouwkamp, "On Bethe's theory of diffraction by small holes," Philips Res. Rep. **5**, 321–332 (1950).
20. D.E. Grupp, H.J. Lezec, K.M. Pellerin, T.W. Ebbesen, and T. Thio, "Fundamental role of metal surface in enhanced transmission through subwavelength apertures," Appl. Phys. Lett. **77**, 1569–1571 (2000).
21. W.L. Barnes, W.A. Murray, J. Dintlinger, E. Devaux and T.W. Ebbesen, "Surface plasmon polaritons and their role in the enhanced transmission of light through periodic arrays of subwavelength holes in a metal film", Phys. Rev. Lett. **92**, 107401 (2004).
22. M.K. Kowarz, "Homogeneous and evanescent contributions in scalar near-field diffraction," Appl. Opt. **34**, 3055–3063 (1995).
23. A.P. Hibbins, J.R. Sambles, C.R. Lawrence, "Gratingless enhanced microwave transmission through a subwavelength aperture in a thick metal plate," Appl. Phys. Lett. **81**, 4661–4663 (2002).
24. H.J. Lezec and T. Thio, "Multiplicative two-surface enhanced transmission through subwavelength hole arrays", submitted to JOptA.
25. A. Hessel and A.A. Oliner, "A new theory of Wood's anomalies on optical gratings," Appl. Opt. **4**, 1275–1299 (1965).
26. J. A. Porto, F. J. Garcia-Vidal, and J. B. Pendry, "Transmission resonances on metallic gratings with very narrow slits," Phys. Rev. Lett. **83**, 2845–2848 (1999).
27. L. Martín-Moreno, F. J. García-Vidal, H. J. Lezec, A. Degiron, and T. W. Ebbesen, "Theory of highly directional emission from a single subwavelength aperture surrounded by surface corrugations," Phys. Rev. Lett. **90**, 167401 (2003).
28. M.M.J. Treacy, "Dynamical diffraction in metallic optical gratings," Appl. Phys. Lett. **75**, 606–608 (1999).
29. M.M.J. Treacy, "Dynamical diffraction explanation of the anomalous transmission of light through metallic gratings," Phys. Rev. B **66**, 195105 (2002).
30. A. A. Oliner and D. R. Jackson, "Leaky-surface plasmon theory for dramatically enhanced transmission through a sub-wavelength aperture, Part I: Basic features," Proc. IEEE Antennas and Propagation Symposium, Columbus OH, USA, June 2003 (AP-S Digest Vol. 2, pp. 1091–1094).
31. J.M. Vigoureux, "Analysis of the Ebbesen experiment in the light of evanescent short range diffraction," Opt. Commun. **198**, 257–263 (2001).
32. M. Sarrazin, J.-P. Vigneron and J.-M. Vigoureux, "Role of Wood anomalies in optical properties of thin metallic films with a bidimensional array of subwavelength holes," Phys. Rev. B **67**, 085415 (2003).
33. C. Genet, M.P. van Exter, and J.P. Woerdman, "Fano-type interpretation of red shifts and red tails in hole array transmission spectra," Opt. Commun. **225**, 331–336 (2003).
34. E. Altewischer, M.P. van Exter, and J.P. Woerdman, "Plasmon-assisted transmission of entangled photons," Nature (London) **418**, 304–306 (2002).
35. A. Dogariu, T. Thio, L.J. Wang, T.W. Ebbesen, and H.J. Lezec, "Delay in light transmission through small apertures," Opt. Lett. **26**, 450–452 (2001).
36. A. Degiron, H.J. Lezec, W.L. Barnes, and T.W. Ebbesen, "Effects of hole depth on enhanced light transmission through subwavelength hole arrays," Appl. Phys. Lett. **81**, 4327–4329 (2002).
37. T.J. Kim, T. Thio, T.W. Ebbesen, D.E. Grupp and H.J. Lezec, "Control of optical transmission through metals perforated with subwavelength hole arrays," Opt. Lett. **24**, 256–258 (1999).
38. A. Nahata, R.A. Linke, T. Ishi, and K. Ohashi "Enhanced nonlinear optical conversion from a periodically nanostructured metal film," Opt. Lett. **28**, 423–425 (2003).
39. J. Fujikata, T. Ishi, H. Yokota, K. Kato, M. Yanagisawa, M. Nakada, K. Ishihara, K. Ohashi, T. Thio and R.A. Linke, "Surface plasmon enhancement effect and its application to near-field optical recording," presented at the Magneto-Optical Recording International Symposium 2004 (MORIS 2004), Yokohama, Japan (May 2004).
40. S. Shinada, J. Hashizume, and F. Koyama, "Surface plasmon resonance on microaperture vertical-cavity surface-emitting laser with metal grating," Appl. Phys. Lett. **83**, 836–838 (2003).
41. B.A. Munk, *Frequency Selective Surfaces: Theory and Design*, (Wiley, New York, 2000).
42. C.Winnnewisser, F.T. Lewen, M. Schall, M. Walther, and H. Helm, "Characterization and Application of Dichroic Filters in the 0.1-3-THz Region," IEEE Trans. Microwave Theory Tech. **48**, 744–749(2000).

#4585 - $15.00 US Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

43. T. Timusk and P.L. Richards, "Near millimeter wave bandpass filters," Appl. Opt. **20**, 1355–1360 (1981).
44. P.G. Huggard, M. Meyringer, A. Schilz, K. Goller, and W. Prettl, "Far-infrared bandpass filters from perforated metal screens," Appl. Opt. **33**, 39–41 (1994).

1. Introduction

The optical transmission of a subwavelength aperture in a metal film is enhanced when it is placed in a periodic array [1], or when it is surrounded by a periodic corrugation on its illuminated surface [2]. For arrays, very large transmission enhancements, up to a factor 1000, were reported and subsequently quoted. The currently widely-held view is that the transmission enhancement results from resonant excitation of surface plasmon polaritons (SP) [3] at the metal surface, which leads to a large enhancement of the electromagnetic fields at the aperture entrances and hence to enhanced transmission [1, 4]. In that model, the incident light interacts with SP modes through grating coupling at the periodic surface structure.

The SP model is not consistent with all experimental observations. For example, the SP model predicts transmission peaks at wavelengths which are about 15% smaller than experimental values, and which in fact provide a better match to experimental minima [5]. This holds true whether the hole arrays are designed to operate in the visible or the THz region [6, 7]. It has been argued that the surface corrugations modify the SP dispersion relation significantly, but experiments on slit arrays, the limiting case of very deep surface corrugations, show that the SP dispersion is within 1% of the smooth-surface case [8, 9]. Moreover, large transmission enhancements are observed for hole arrays in Cr, in a wavelength region (near IR) for which it is only marginally metallic and hence not expected to support SP modes [10]. Similarly, transmission spectra very similar to those observed in metallic systems are obtained by numerical calculations for both hole arrays in tungsten, which is nonmetallic in the visible region [11], as well as for single apertures in a corrugated perfect metal [12], even though perfect metal surfaces do not support SPs. Finally, beaming effects from a waveguide in a non-metallic photonic crystal [13] have been observed with characteristics similar to those of metallic systems [14, 15].

In this paper we re-evaluate the magnitude of the transmission enhancement as well as the relevance of the SP model. We show that the quoted enhancement factor of 1000 for optical transmission through subwavelength hole arrays is misleading, and that in fact placing a hole in an array leads to enhancement of its transmission coefficient by at most a factor 7 at selected wavelengths. This follows from a careful comparison to the transmission coefficient of a real, identical, isolated hole in the same film. Equally unexpected is the observation of transmission suppression of equivalent magnitude at other wavelengths. Such a suppression is not consistent with a model based on resonant coupling to SP modes, which predicts only enhancement compared to the case of an isolated hole. Furthermore, we confirm experimentally that hole arrays in non-metallic films have both transmission and reflection spectra similar to those of metallic films. Finally, measurements on a metallic device specifically tailored to reveal any contribution of SPs to the enhanced transmission process yield a negative result.

We propose a new model for the transmission enhancement and suppression. We consider the complete diffracted field of a single subwavelength surface feature, which launches evanescent waves with a large distribution of in-plane k-vectors (only one of which matches the SP mode at a given frequency). We show that the superposition of these individual components forms a composite evanescent wave which propagates with a well-defined wavevector, an amplitude decaying with distance, and a finite phase shift with respect to the source. It is the interference of this wave with light exciting directly the opening of a nearby aperture which gives rise to transmission enhancement and suppression at that aperture.

#4585 - $15.00 US Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

2. Re-consideration of existing claims

2.1. Magnitude of transmission enhancement

For the first time, we compare the transmission properties of a hole in an array to those of an identical single, isolated hole fabricated on the same film, varying the number of holes to determine the optimal array size for transmission enhancement. In order to properly characterize the total amount of light transmitted in each case, we take into account the total radiated diffraction pattern and its intersection with the collection system. We have measured the transmission spectra of apertures in both metallic and non-metallic films, either free-standing or deposited on glass substrates. The films are optically thick (opaque in the absence of apertures). Cylindrical apertures and surface corrugations were milled in the films using a FEI DB-235 focused-ion beam (FIB) system (Ga^+ ions, 30 keV). A low beam current ($\sim$10–30 pA) was used in order to achieve sharp surface features defined with a lateral precision on the order of 10 nm and holes with near-vertical sidewalls. In the case of the films on glass substrates, index-matching fluid was applied to the metal-air interface; matching the refractive index of the dielectric media on either side of the film maximizes the transmission efficiencies while simplifying the transmission spectra [16]. The samples were illuminated at normal incidence with white light collimated to $\pm 3^\circ$. Transmitted light was collected with a $\times$40 objective (N.A.=0.6), and spectrally analyzed using an Acton monochromator coupled to a cooled CCD camera (Princeton Instruments). In the analysis of the transmission spectra, we define $Q_R(\lambda)$ as the total power emitted into free-space radiative modes on the exit side of the hole array or single hole, and $Q_C(\lambda)$ as the power collected by the objective. The collection efficiency of the system is then given by $\gamma(\lambda) = Q_C(\lambda)/Q_R(\lambda)$. Light emerging from a single subwavelength aperture is diffracted nearly isotropically, resulting in a typical value for γ of only $\sim$ 20% for light emerging into air (and only $\sim$ 10% for light emerging into a glass substrate of finite thickness, which traps the most strongly diffracted rays by total internal reflection). In contrast, a large array of apertures acts as a transmission grating which emits a set of well-collimated beams (diffraction orders) with wavelength-dependent angles. The zero-order beam, collinear with the incident beam, is always collected, so that γ approaches 100% when λ exceeds the lattice constant.

We consider an $N \times N$ square array of holes of diameter d and period P (where $N = 1$ corresponds to a single, isolated hole), illuminated by a plane wave of constant intensity I_0. The power from the incident beam impinging on the opening of each aperture is $Q_0 = I_0 \pi d^2/4$. For each value of N, we calculate $\gamma_N(\lambda)$ by modeling the far field transmitted radiation pattern as the product of the Fraunhofer distribution function for emission from a circular aperture of diameter d, with the diffraction pattern of an $N \times N$ array of point emitters [17]. The total per-hole transmission coefficient, which takes into account all the transmitted power radiated into free space, is then derived from a measurement of $Q_C(\lambda)$ according to the formula $T_{R,N}(\lambda) = (1/N^2)Q_R(\lambda)/Q_0 = (1/N^2)(1/\gamma_N(\lambda))Q_C(\lambda)/Q_0$. We define an array enhancement factor $G_N(\lambda) = T_{R,N}(\lambda)/T_{R,1}(\lambda)$, which represents the per-hole gain in total radiative transmission achieved by taking a single, isolated hole and placing it in an $N \times N$ array.

Figure 1(d) displays the as-collected transmission spectrum $(1/N^2)Q_C(\lambda)/Q_0$ for an $N \times N$ array of cylindrical holes ($P = 410$ nm, $d = 150$ nm) in an Ag film (thickness $t = 175$ nm) on fused-silica glass , overcoated with index-matching fluid ($n = 1.46$). As N is increased from 1 to 19, distinct peaks develop in the transmission spectrum; full saturation in the shape of the spectrum is achieved by $N = 15$. The longest-wavelength maximum occurs at a wavelength substantially larger than the lattice constant ($\lambda_{max}/P = 1.18$). The per-hole total transmission coefficient, $T_{R,N}(\lambda)$, is shown in Fig. 1(e). As N is increased beyond 1, $T_{R,N}(\lambda)$ develops a peak situated beyond the effective index-corrected period [4], $nP = 600$ nm, similar to the trend observed for $Q_C(\lambda)/Q_0$. This peak reaches full saturation for $N \geq 9$. Throughout the displayed wavelength range, the hole can be considered to be strictly subwavelength, since its diameter

#4585 - $15.00 US Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

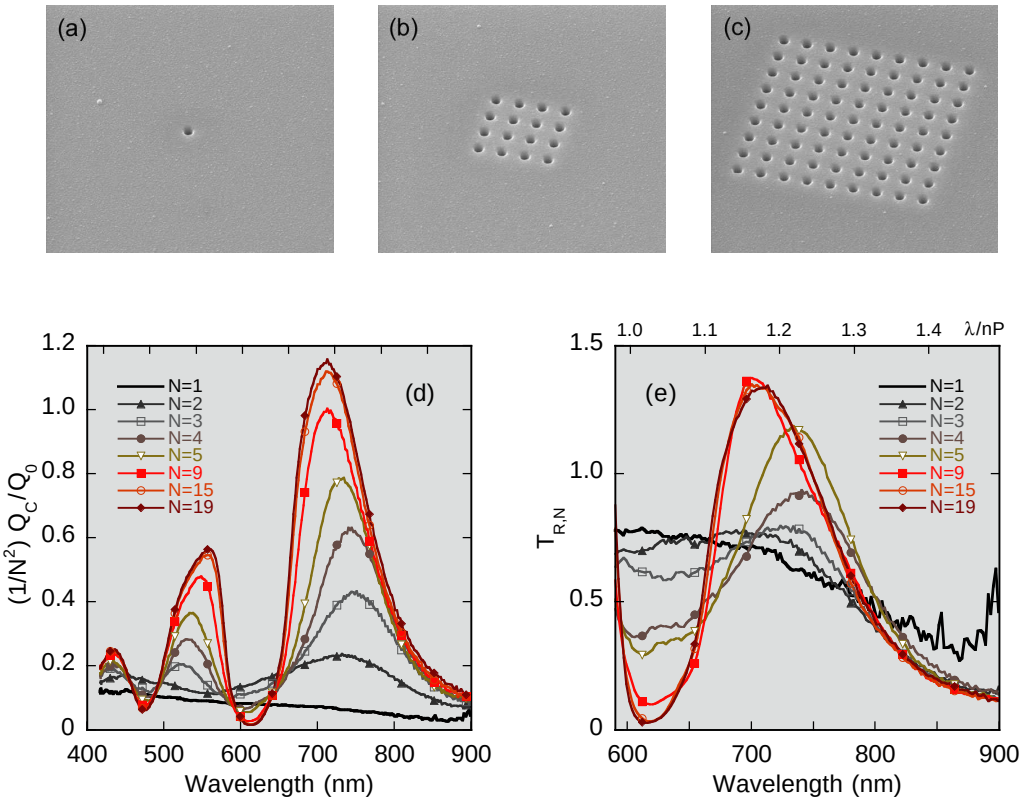

Fig. 1. Optical characterization of an $N \times N$ array ($P = 410$ nm) of cylindrical holes ($d = 150$ nm) in a silver film ($t = 175$ nm) on glass, overcoated with index-matching fluid. (a,b,c) SEM micrographs of structures for N=1, 4, and 9, respectively (horizontal field of view FOV $= 5$ μm, observation angle with respect to normal $a = 45^{\circ}$); (d) as-collected transmission spectra for selected N; (e) corresponding per-hole transmission coefficient $T_{R,N}(\lambda)$ for selected N.

satisfies $d < \lambda_{max}/(1.8n)$, the cutoff condition for a cylindrical metallic waveguide [18].

Some key observations emerge from the data of Fig. 1(e). Compared to $T_{R,1}(\lambda)$, $T_{R,N>1}(\lambda)$ is *suppressed* at some wavelengths, in addition to being enhanced at other wavelengths. Furthermore, array-induced enhancement and suppression occur with comparable amplitudes with respect to a baseline defined by the transmission coefficient of the isolated hole. Finally, at saturation, the array-enhancement factor remains modest, reaching a maximum value of $G_N = 2.0$ (at $\lambda = 700$ nm, for $N \geq 9$).

The general characteristics of $T_{R,N}(\lambda)$, and $G_N(\lambda)$ are further explored by optical transmission measurements on hole arrays for which P and d are varied from device to device. The arrays are fabricated in a free-standing film of Ag ($t = 340$ nm), with the number of holes fixed at 9×9 in order to ensure fully developed modulation spectra. Figure 2(a) compares $T_{R,N}(\lambda)$ to $T_{R,1}(\lambda)$ for two arrays with periods $P = 500$ nm and $P = 600$ nm, respectively (and a constant hole diameter $d = 250$ nm). In both cases, $T_{R,N}(\lambda)$ displays both suppression and enhancement relative to the single-hole transmission coefficient, $T_{R,1}(\lambda)$. As P is increased from 500 nm to 600 nm, the spectral features shift to longer wavelengths, their amplitudes decreasing following the trend of $T_{R,1}(\lambda)$. Similar to the trend observed in Fig. 1, the longest-wavelength peak

#4585 - $15.00 US
Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

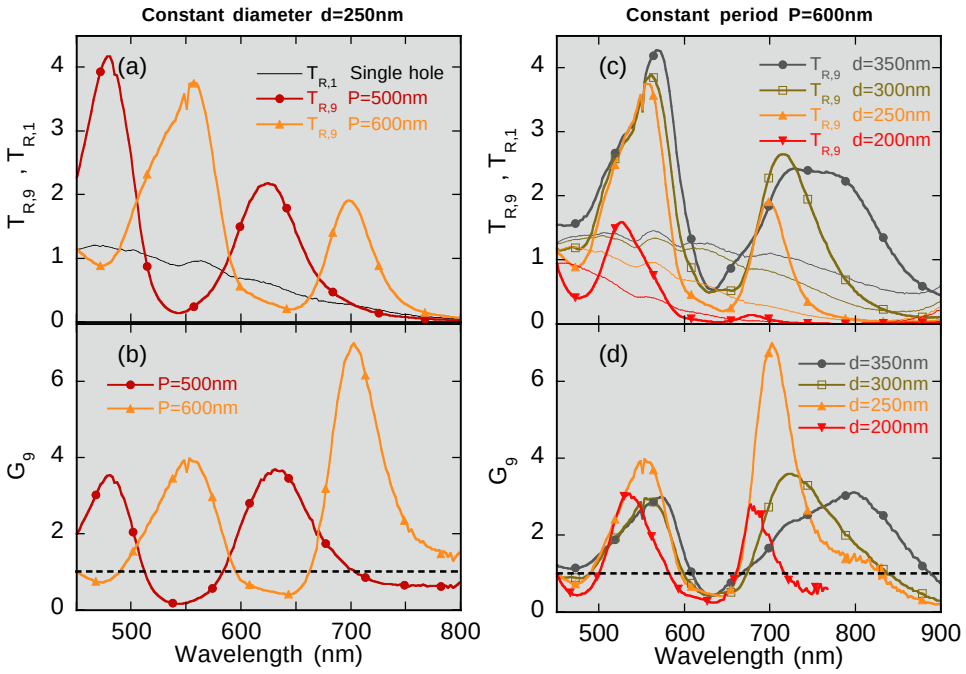

Fig. 2. Optical characterization of a 9×9 array of cylindrical holes in a suspended Ag film as a function of array period P and hole diameter d. (a) Per-hole transmission coefficient $T_{R,9}(\lambda)$ for selected P (at constant d) and transmission coefficient $T_{R,1}(\lambda)$ of an equivalent isolated hole; (b) resulting array enhancement factor $G_9(\lambda) = T_{R,9}/T_{R,1}$ for selected P; (c) per-hole transmission coefficients $T_{R,9}(\lambda)$ (thick lines) and $T_{R,1}(\lambda)$ (thin lines) for selected d (at constant P); (d) resulting array enhancement factor $G_9(\lambda)$ for selected d.

occurs at a value λ_p which is substantially red-shifted with respect to the lattice constant P ($\lambda_p/P = 1.25$ and 1.17 for $P = 500$ nm and 600 nm, respectively). The enhancement factor $G_N(\lambda)$ displays corresponding periodic modulations below and above unity (Fig. 2(b)). The maximum values of G_N achieved in the wavelength region $\lambda > P$ are $G_N(\lambda_p) = 3.9$ and 6.9, for $P = 500$ nm and 600 nm, respectively. Figure 2(c) compares $T_{R,N}(\lambda)$ to $T_{R,1}(\lambda)$ for four distinct 9×9 arrays in which the hole diameter d is varied at constant period $P = 600$ nm. As d is increased, $T_{R,N}(\lambda)$ and $T_{R,1}(\lambda)$ increase in a commensurate manner across the spectral range, illustrating once again the dominance of the single-hole transmission function in setting the overall level of transmission. This is further illustrated in the relative invariance of the maximum peak enhancement factor $G_N(\lambda_p)$ (Fig. 2(d)). As d is increased from 200 nm to 350 nm, $T_{R,1}(\lambda_p)$ increases by a factor of ~ 20 but $G_N(\lambda_p)$ varies by less than a factor of 2. A maximum enhancement factor $G_N(\lambda_p)$=6.9 is obtained for $d/P \simeq 3$; as d/P is increased beyond this optimal ratio, the enhancement peak is broadened as well as shifted to longer wavelengths.

We have explored hole arrays with a wide range of parameters, such as hole diameter, hole depth, film thickness, and film material; in all instances $T_{R,N}(\lambda)$ is enhanced by less than a factor 7 compared to $T_{R,1}(\lambda)$. Such a small enhancement is rather less than the often-quoted [1] enhancement factor of order 1000. We account for the large discrepancy as follows: In Ref. [1] the maximum per-hole transmission coefficient $T_{R,N}(\lambda_p)$ measured for a hole array at its longest peak wavelength λ_p, was compared not to the experimental transmission coefficient $T_{R,1}(\lambda_p)$, of a comparable real, single hole (as in Figs. 1 and 2), but to the theoretical Bethe-Bouwkamp

#4585 - $15.00 US
Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

prediction for the transmission coefficient of a hole in an perfect metal film with infinitesimal thickness, $T_{BB}(\lambda) = 23(d/\lambda)^4$ [19]. The enhancement factor reported in Ref. [1], equivalent in the nomenclature of the present paper to $G_{BB}(\lambda_p) = T_{R,N}(\lambda_p)/T_{BB}(\lambda_p)$, is proportional to λ^4/d^6 (including the hole filling-fraction), and is thus extremely sensitive to d. The correct comparison to a waveguide beyond cutoff has an even stronger, exponential dependence on d [18]. In either case, the comparison to a theoretical prediction so sensitive to d is perilous if d is not well defined. The hole array of Fig. 1 in Ref. [1] was milled with the same FIB parameters, the same targeted hole diameter and in a film of the same thickness ($t = 200$ nm), as the hole array of Fig. 1 of Ref. [4]. Examination of the latter figure suggests that holes fabricated under such conditions are not cylindrical, as incorrectly stated in Ref. [1], but have an entrance diameter which is more than twice as large as the quoted diameter of $d = 150$ nm. The tapered hole profile results from the use of a relatively large-diameter FIB (with a Gaussian intensity profile with FWHM $\simeq 150$ nm, not $\simeq 5$ nm as quoted in [1]), and a large beam current (1 nA) chosen to enable milling of large arrays. We conclude that the quoted enhancement factor ~ 1000 represents a severe overestimate of the actual enhancement factor $G_{R,N}(\lambda_p)$ of the hole array of Fig. 1, Ref. [1]. To demonstrate this quantitatively, we calculate $G_{BB}(\lambda_p)$ for the data of Fig. 2(c), for which d is well defined. We obtain $G_{BB}(\lambda_p)$= 0.8, 5.2, 3.8 and 2.3, for $d = 200$, 250, 300 and 350 nm respectively. These values are close to the experimentally measured values $G_{R,N}(\lambda_p)$=2.8, 6.9, 3.6 and 3.2 (see Fig. 2(d)), not orders of magnitude higher. We emphasize that the Bethe-Bouwkamp theory is of limited relevance when considering holes in a real metal film of finite thickness, which are best described as waveguides [18] (whether or not beyond cutoff) .

2.2. Transmission enhancement in nonmetallic systems

The observation of both enhancement and suppression, moreover of comparable amplitude, in the transmission spectra of hole arrays in metallic films suggests that a lateral interference phenomenon is the driving force in shaping the transmission spectra. The latter phenomenon cannot be explained by the surface-plasmon model which invokes resonant coupling to SP modes, and predicts only enhancement with respect to the single-hole case. The notion that SPs are not essential to the transmission spectra of hole arrays is supported by numerical calculations for hole arrays in tungsten [11]. This material is nonmetallic in the visible region (where the real part of the dielectric constant is $\varepsilon^{'} \simeq +5$) and therefore does not support SP modes [3].

This result is confirmed by experimental transmission measurements on a hole array fabricated in a free-standing W film (Fig. 3(a)). The spectral dependence of the transmission coefficient is identical to that of a hole array fabricated in a suspended Ag film, up to a multiplicative factor (where the spectrum in the Ag case shows no additional features which can be assigned to SP modes). Moreover, the same observations hold true for a hole array fabricated in amorphous silicon (a-Si), which in the as-deposited form is a dielectric at all frequencies (Fig. 3(a)). The overall transmission efficiencies of arrays in non-metallic films, such as W or a-Si, are smaller than those of arrays in metallic films such as Ag. We argue that this stems simply from the larger surface absorption of W and a-Si compared to Ag. Similarly, hole arrays in highly absorbing Ni also exhibit substantially reduced transmission efficiencies compared to those of identical arrays in Ag [20]. In all cases however, whether fabricated in a film that is metallic or non-metallic, close to an ideal metal or highly absorbing, hole arrays are characterized by transmission spectra with remarkable correspondence in shape, which suggests a common modulation mechanism unrelated to SPs.

The reflectivity of non-metallic systems corresponds closely to that of metallic systems. Figure 3(b) compares the spectrum of the first-order reflected beam (R[1]) of an array of holes in Ag to that of an array of deep dimples in the surface of an undoped Si wafer. In both reflec-

#4585 - $15.00 US Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

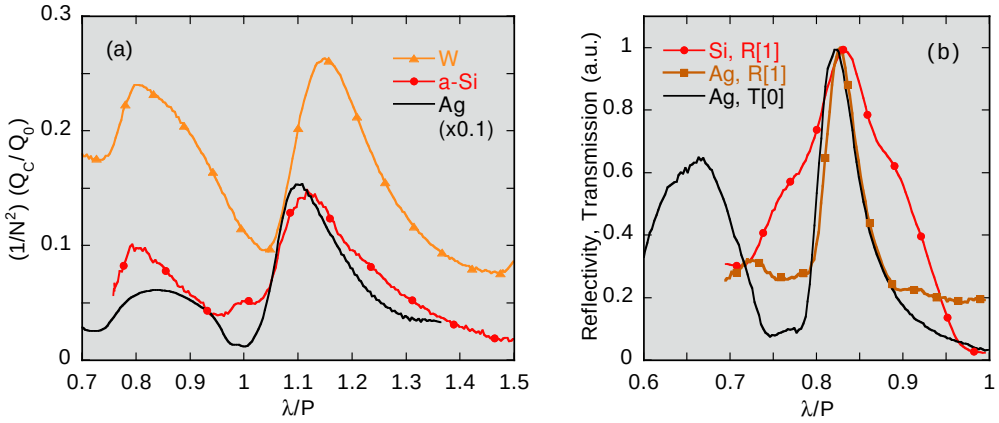

Fig. 3. (a) As-collected transmission spectra of 15×15 hole arrays in suspended films of: tungsten (W, $t=400$ nm, $d=300$ nm, $P=600$ nm); amorphous silicon (a-Si, $t=200$ nm, $d=250$ nm, $P=550$ nm); and silver (Ag, $t=300$ nm, $d=250$ nm, $P=750$ nm); (b) zero-order transmission (Ag, T[0]) and first-order reflectivity (Ag, R[1]) of a 25×25 array ($P=820$ nm) of holes ($d=250$ nm) in suspended Ag ($t=300$ nm), compared to first-order reflectivity (Si, R[1]) of a 25×25 array ($P=820$ nm) of cylindrical dimples ($d=300$ nm, depth $h\simeq1$ μm) in undoped Si.

tion spectra a maximum occurs at the same wavelength. We attribute the difference between the Ag and Si reflection spectra to a larger optical penetration range in the second case, leading to a larger effective dimple diameter and hence wider spectral features (compare to the data of Fig. 2(c)). The peak wavelength of R[1] for both Ag and Si arrays coincides with a maximum in the zero-order transmission spectrum, T[0], of the same Ag array. In contrast, the zero-order reflection spectrum has a minimum at this wavelength [21]. If this minimum is due to an absorption by SP modes as claimed in Ref. [21], it would also lead to a minimum in the first-order reflection spectrum, contrary to what is observed (Fig. 3(b)). In section 4.2 we offer an alternate interference-based explanation for the observed coincidence of zero-order reflection minima with transmission maxima.

In summary, the two key observations that (a) placing the hole in a periodic array leads to suppression as well as enhancement of the transmission, and that (b) hole arrays in Ag show the same transmission spectra as those in non-metallic films which do not support SP modes, are in direct contradiction to the SP-based model for enhanced transmssion through subwavelength hole arrays.

3. New model: diffraction and interference of evanescent waves

We consider the diffraction of the illuminating plane wave by a subwavelength feature on the surface of a film. In particular, we show that the evanescent component of the diffracted field, which includes contributions from all in-plane k-vectors larger than k_0, can be described simply as a composite evanescent wave with propagation characteristics that are distinct from those of an SP. We construct a simple analytical model in which interference of this composite wave with the light impinging directly on the entrance or exit surface of an aperture leads to transmission enhancement and suppression. A number of experiments are described which support the model.

#4585 - $15.00 US
Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

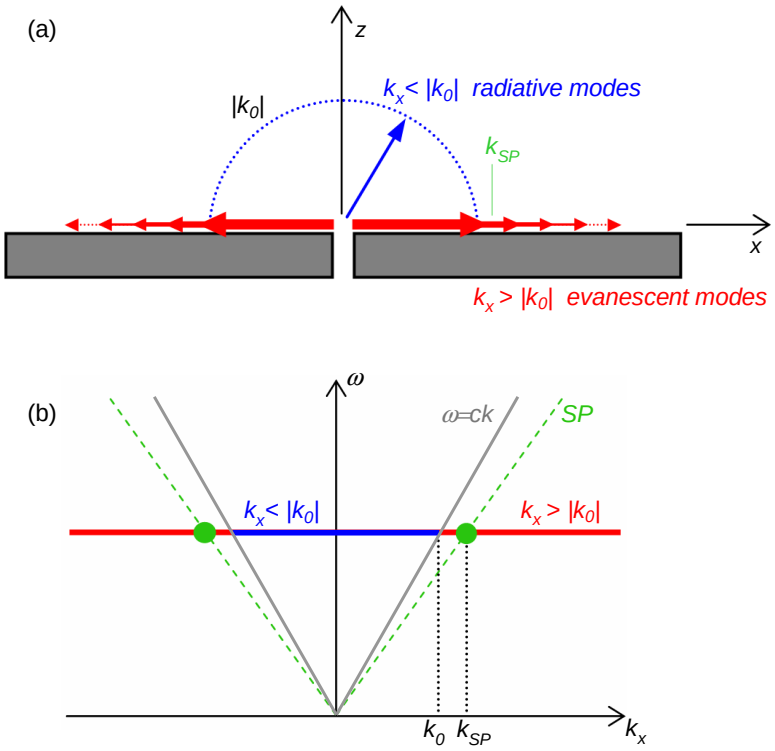

Fig. 4. Geometry of optical scattering by a hole in a screen in (a) real space and (b) k-space.

3.1. *Diffraction by subwavelength surface features*

We begin by considering light of frequency $\omega = ck_0 = 2\pi c/\lambda_0$ emerging from a subwavelength aperture of width d (see Fig. 4(a)). The light is diffracted partly into a continuum of radiative ("homogeneous") modes of which the in-plane component of the wave vector is $k_x < k_0$ (blue circle in Fig. 4(a) and blue horizontal line in Fig. 4(b)), and partly into a continuum of evanescent ("inhomogeneous") modes propagating along the surface with real $k_x > k_0$ and imaginary $k_z = i(k_x^2 - k_0^2)^{1/2}$ (red arrows in Fig. 4(a) and red horizontal lines in Fig. 4(b)). As d/λ_0 is reduced, the fraction of the total power emerging from the aperture which is diffracted into evanescent modes grows. Of this broad distribution of evanescent modes, a single mode (indicated by the green line in Fig. 4(a) and by the green dot in Fig. 4(b)) matches the wave vector of a surface-plasmon polariton. The SP is a TM guided mode [3] supported by the metal–dielectric interface only when $\varepsilon_m' < -\varepsilon_d'$, where ε_m' and ε_d' are the real parts of the permittivities of the metal and the dielectric, respectively. The SP in-plane wave vector is given by $k_{SP} = n_{SP}(2\pi/\lambda_0)$ (dashed green line in Fig. 4(b)), where $n_{SP} = (\varepsilon_m'\varepsilon_d'/(\varepsilon_m' + \varepsilon_d'))^{1/2}$. For a good metal such as silver, the SP dispersion curve lies very close to the light line (e.g. $n_{SP} = 1.03$ at $\lambda_0 = 650$ nm for the Ag–air interface), forming a mode with very low density of states.

The condition $k_{SP} > k_0$ implies that light incident on a flat metal surface cannot couple directly to an SP at any finite frequency. Diffraction scattering from a subwavelength surface feature, such as an aperture, does provide the necessary momentum to couple to an SP (see Fig. 4(b)). However, the fraction of the total evanescent power coupled into the SP is very low (in other words the SP does not constitute a preferred channel for diffraction compared to the other diffracted lossy modes which are not long-range guided modes of the surface). A

#4585 - $15.00 US
Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

subwavelength aperture is therefore an inefficient device for coupling to SPs.

The apertures in the hole arrays considered in the present study are in close, wavelength-scale proximity to each other. As a result, each subwavelength aperture acts as a probe of the power diffracted into the continuum of evanescent surface modes by the neighboring holes. It is thus imperative to consider the effect of the total evanescent field diffracted by a given hole.

3.2. Analytical characterization of the total, in-plane diffracted evanescent field

To understand the fundamental characteristics of the total evanescent field diffracted by a subwavelength aperture, we consider the concrete example of light diffracted a single slit of width d in an opaque screen. The electric field amplitude of the total diffracted field at the surface is given by [22]:

$$E(x,z=0) = -\frac{E_0}{\pi}\left[\mathrm{Si}\left(k_0(x+\frac{d}{2})\right) - \mathrm{Si}\left(k_0(x-\frac{d}{2})\right)\right], \tag{1}$$

where x is the distance from the center of the slit or groove, E_0 is the source field amplitude, and $\mathrm{Si}(\alpha)$ is the sine integral $\mathrm{Si}(\alpha) = \int_0^\alpha (\sin(t)/t)dt$. This expression has a simple physical interpretation for $d < \lambda/2$ and in the full lateral region beyond the immediate vicinity of the aperture. Indeed, for $x > \lambda/2$ we find that Eq. (1) is accurately approximated by:

$$E(x) = \frac{E_0}{\pi}\frac{d}{x}\cos(k_0 x + \frac{\pi}{2}). \tag{2}$$

The total diffracted evanescent field can thus be described as a running wave propagating with a well-defined wave-vector lying in the plane of the surface. We designate this wave as a "composite diffracted evanescent wave" (CDEW), where "composite" refers to the summmation of all diffracted inhomogeneous modes performed to obtain Eq. (1). We emphasize that the CDEW is launched at any frequency, independently of the presence of neighboring holes, with a wavevector magnitude equal to that of the source, $k_{\mathrm{CDEW}} = k_0 = 2\pi/\lambda_0$. The amplitude of the CDEW decays with lateral distance as $1/x$, and its phase is shifted by $\phi = +\pi/2$ with respect to the light emerging from the slit. A CDEW thus has physical characteristics that clearly distinguish it from an SP (which is launched with a phase equal to that of the source and propagates with a constant amplitude).

In the well-known Fraunhofer formalism, the far-field angular intensity distribution of light diffracted by a slit of width d is given by $I(\beta) = (sin(\beta)/\beta)^2$ where $\beta = k_x d/2$ [17]. This function constitutes an effective "density of states" for k_x in the range of radiative modes $0 < k_x < k_0$. It is interesting to examine the result of extrapolating $I(\beta)$ to encompass as well the range of propagating evanescent modes with $k_x > k_0$. Using this *ad-hoc* extension of the density-of-states function, we can estimate the superposition of all diffracted evanescent modes along the surface. A numerical integration of the spectrum of weighted modes $|I(\beta)|^{1/2}\exp(ik_x x)$ over the interval $k_0 < k_x < \infty$ yields exactly the expression given by Eq. (2), when $x > d$. Whether this is largely a coincidence or carries deeper meaning is not yet fully understood.

We note that the preceding analysis applies equally well to the case of a subwavelength aperture-opening or groove on the illuminated side of the film. Such a feature diffracts an incident plane wave of field amplitude E_0 into a CDEW according to Eq. (2). Thus an aperture in an opaque film will launch two distinct CDEWs, one on each side of the film.

3.3. Experimental characterization of the composite diffracted evanescent field

Since the form of the CDEW is crucial to the enhanced transmission phenomenon, we perform a number of experiments to verify Eq. (2). In a first experiment (Fig. 5), we check the phase- and distance-dependence of the CDEW by characterizing the in-plane optical interference of

#4585 - $15.00 US Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

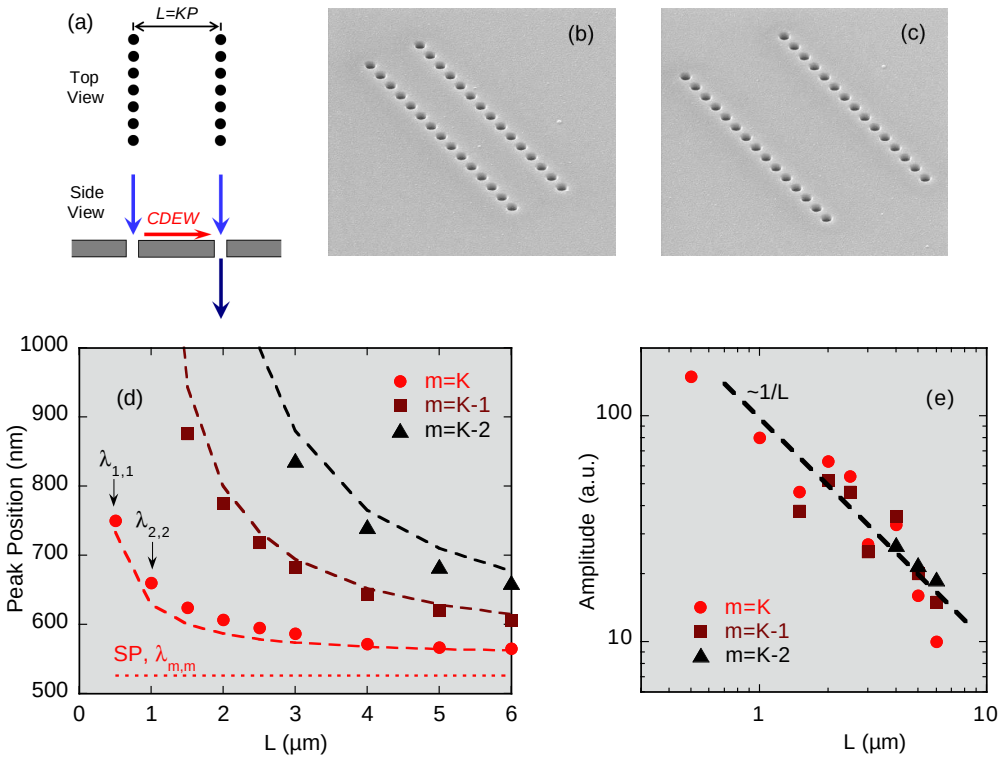

Fig. 5. Characterization of CDEW amplitude and phase by in-plane interferometry. (a) Device geometry: double row of cylindrical holes ($d=250$ nm, $P=500$ nm) in a suspended Ag film ($t=300$ nm) with row–row spacing L varied in integer multiples of P; (b and c) SEM micrographs of devices with $L=1.5$ and 3 μm, respectively (FOV $=7.6$ μm, $a=52^{\circ}$); (d) wavelength positions of interference maxima $\lambda_{m,K}$ (where m is interference order), plotted as a function of L: experimental data (symbols), CDEW model (dashed lines), and SP model (dotted line, shown only for case $m=K$); (e) CDEW amplitude C as function of traveled distance L: experimental data (symbols) and $1/L$ fit (dashed line).

two rows of holes ($d=250$ nm), for which the row–row distance $L=KP$ is varied in fixed increments $P=500$ nm with $K=1,2,\ldots 10$. Both rows (see Fig. 5(a)) are illuminated with a plane-wave (light blue arrows) but light is captured and spectrally analyzed from only one of the rows (dark blue arrow). Light incident on the left-hand row of holes is partly diffracted into a CDEW (red arrow), which propagates to the second row where it interferes with the light that is directly incident on that row (by symmetry, the inverse process from right to left occurs as well). This interference results in intensity variations (which depend on λ and L) at the entrance to the holes of the right-hand row, and to spectral modulation of the transmitted power Q_C/Q_1 (where Q_C is the collected power emerging from a given row, and Q_1 that of a reference device consisting of a single, isolated row). Following the prediction of Eq. (2), assume that the wave traveling from one row to the other is given by a function of the form $E(x)=C(x)\cos(n_{eff}(2\pi/\lambda)x+\phi)$, where ϕ is the phase shift with respect to the incident illumination upon launching, $C(x)$ is a monotonically varying function of distance, and n_{eff} is an adjustable parameter corresponding to an effective index of refraction experienced by the CDEW. Constructive interference, leading to transmission maxima, is then predicted to occur at wavelengths $\lambda_{m,K}(L)=n_{eff}L/(m-\phi/2\pi)=n_{eff}KP/(m-\phi/2\pi)$, where $m=1,2,\ldots,K-$

#4585 - $15.00 US
Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

$1, K$ is the interference order. The generation of a CDEW, and its interference with the incident plane wave is evinced by the presence of pronounced interference fringes in the spectrum of $Q_C(\lambda)/Q_0(\lambda)$. The peak positions of these modulations are plotted as solid symbols in Fig. 5(d) as a function of $L = KP$ and identified by interference order m. The observed peak wavelengths $\lambda_{m,k}$ show an excellent correspondence to the behavior predicted by the CDEW model of Eq. (2) using fitting parameters $\phi = \pi/2$ and $n_{eff} \simeq 1.1$ (dashed lines in Fig. 5(d)). The peak position for the data set $m = K$ shows an upturn at small L which constitutes direct and unequivocal evidence of a phase shift of the CDEW with respect to the incident light. In contrast, an SP would be launched in phase with its source ($\phi = 0$); for the highest interference order, $m = K$, the expected SP interference peak position $\lambda_{m,m} = n_{eff}P = n_{\mathrm{SP}}P$ is independent of L, as indicated by the dotted line in Fig. 5(d).

The effective refractive index n_{eff} experienced by the CDEW represents a correction to Eq. (2) which was derived for a generic opaque screen. A-priori, n_{eff} is expected to represent the weighted effect of the metal and of the adjacent dielectric medium (of index n_d) sampled by the evanescent field of the CDEW to either side of the interface. For an SP (one of the constituent components of the CDEW) the effective index $n_{eff} = n_{\mathrm{SP}}$ is indeed material- and wavelength-dependent, and slightly larger than n_d; at an Ag–air interface for example, n_{SP} decreases monotonically from 1.04 to 1.02 as λ increases from 200 to 800 nm. In contrast, we find that the CDEW consistently takes on a substantially higher value $n_{eff} \simeq 1.1 n_d$. This value for n_{eff} is consistently observed at an Ag–air interface in a variety of geometries, over a wide wavelength range (500 to 850 nm), and for propagation over distances of at least several microns (see below and section 3.5). The CDEW is composed of a large range of k_x, with corresponding k_z. When $k_x = k_0$ for example, $k_z = 0$ and the evanescent wave is barely confined to the surface; for large values of k_x on the other hand, $|k_z|$ is large and the evanescent wave is tightly confined to the surface. As a result, it is not surprising that the effective medium experienced by the CDEW should be different from that experienced by an SP. Somewhat more unexpectedly, applying the CDEW model to experimental results reported for an enhanced-transmission structure in the microwave region [23], for which metals display near-perfect behavior, also leads to a best-fit value of $n_{eff} \simeq 1.1$ (see section 3.5). The insensitivity of n_{eff} to parameters such as material properties and wavelength range hints at a more fundamental process governing the apparent retardation of the CDEW, possibly related to the initial distribution and subsequent evolution (as a function of distance) of its constituent evanescent modes. Clearly, further study of this issue is required. In the meantime, we simply include n_{eff} as a phenomenological correction to Eqs. 1 and 2 where k_0 is replaced by $k_{eff} = n_{eff}k_0$.

Figure 5(e) displays the variation in CDEW amplitude as a function of traveled distance L. This function $C(L)$ is inferred from the magnitude of the peak-to-valley modulation of the detected interference signal at the second row of holes as a function of wavelength, for different row-row separations. The observed CDEW amplitude is found to follow a $1/L$ dependence over more than one order of magnitude, consistent with Eq. (2). In contrast, as a guided mode of a metal–dielectric interface, an SP can travel relatively long distances along the interface, limited only by surface absorption. The amplitude of an SP is expected to follow $C(L) \sim \exp(-x/2L_{SP})$ where $L_{SP}(\lambda_0)$ is the characteristic SP propagation length. For example, for a Ag–air interface, L_{SP} increases from 18 to 63 μm as λ_0 is varied from 600 nm to 800 nm. For λ in this range, a plot of $C(L)$ on the log-log plot of Fig. 5(e) would correspond to a straight horizontal line.

We remark that a related set of experiments involving the interference between a pair of subwavelength slits in a silver film yields equivalent detection of a CDEW and identical predictions for its phase ($\phi = \pi/2$) and amplitude (inversely proportional to traveled distance).

As noted above, each row of the double-row geometry acts both as a source and detector of CDEWs. As a result, even though the CDEW decays fairly rapidly as a function of distance, we

#4585 - $15.00 US Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

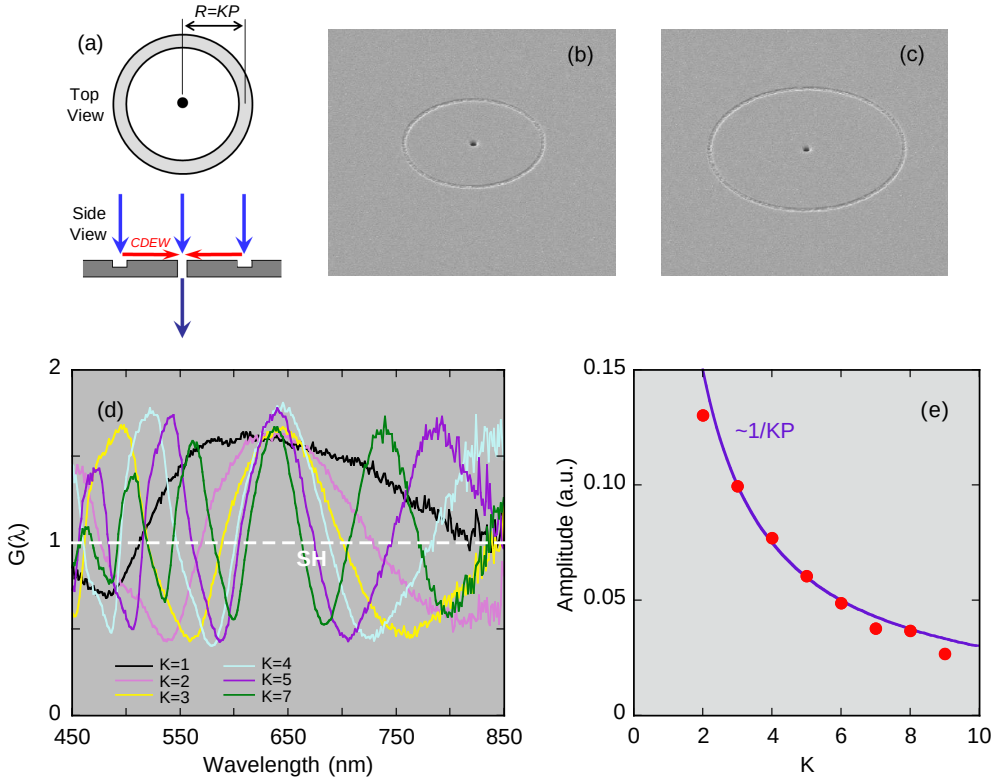

Fig. 6. Characterization of distance-dependence of CDEW amplitude by interferometry in a circular configuration. (a) Device geometry: single hole ($d = 300$ nm) in a silver film ($t = 250$ nm) on glass surrounded by a single circular groove (width $w = 150$ nm, depth $h = 100$ nm) with radius R varied in integer multiples of $P = 600$ nm; (b and c) SEM micrographs of devices with $R = 3$ and 4.2 μm, respectively (FOV = 12.5 μm, $a = 52^{\circ}$); (d) enhancement factor $G(\lambda)$ plotted for increasing normalized radius $K = R/P$; (e) CDEW amplitude C as a function of normalized traveled radial distance $K = R/P$: experimental data (symbols) and $1/R$ fit (solid line).

cannot rule out the influence of a higher-order in-plane process, such as multiple scattering, on the amplitude results of Fig. 5(e). To avoid this possibility, we devise a CDEW interferometer with an asymmetric configuration in which the element used for CDEW detection represents a minimal perturbation as a source of CDEWs: a single hole at the center of a single circular groove (see Fig. 6). Since the area of the hole is very small compared to that of the ring, the CDEW generated by the hole contributes a negligible fraction of the total CDEW field inside the ring, and the hole can therefore be considered to be a true probe of the CDEW launched by the groove.

In the experimental implementation of such a device (Fig. 6(a)), the radius of the ring groove is varied in fixed steps of P according to $R_K = KP$ (where $P = 600$ nm and $K = 1, 2, \ldots 9$). The transmission enhancement factor of the central hole, $G(\lambda) = Q_C(\lambda)/Q_{\mathrm{SH}}(\lambda)$ (where $Q_{\mathrm{SH}}(\lambda)$ is the collected signal from an equivalent reference hole not surrounded by a ring) is measured for plane-wave illumination of the grooved side of the device. The resulting plot of $G(\lambda)$ for various values of K (Fig. 6(d)) reveals a set of spectra, all strongly modulated as a function of wavelength, with an amplitude that is essentially independent of K. The transmission modu-

#4585 - $15.00 US
Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

lation can be interpreted as evidence of interference at the hole entrance: the groove diffracts the incident light partly into a CDEW (red arrows in Fig. 6(a)) which propagates to the central hole where it interferes with the light directly impinging on the hole (light blue arrow). The transmitted light intensity (dark blue arrow) is proportional to the interference signal at the hole entrance. If only the ring were illuminated, no such interference would occur and the normalized spectra of Fig. 6(d) would be flat.

A closed-form solution to the evanescent wave field scattered by an annular feature is not available, and we model the groove as a collection of scatterers with fixed linear density along the ring perimeter, each of which generates an evanescent wave field $E_{\mathrm{CDEW}}(x) \propto C\cos(k_{eff}x+\phi)$. Integrating this expression around the perimeter of the ring (which is appropriate for the unpolarized incident light), gives the total field at the central aperture as $E = E_0 + \oint E_{\mathrm{CDEW}}(KP)$, where E_0 is the field of the light impinging on the aperture. We fit this expression to the family of experimental curves of Fig. 6(d) (over the range 500 nm $< \lambda <$ 850 nm) with the constants k_{eff}, ϕ, and the function C as fitting parameters. Once again we find $k_{eff} = 1.1k_0$, consistent with the interference experiment based on the double row of holes (Fig. 5). In a departure from the case of a CDEW launched from a linear array of holes or a linear groove however, the apparent phase of the CDEW detected at the central hole takes on the value $\phi = -\pi/2$ (instead of $+\pi/2$). While the cause of this sign change is not understood at present, we suspect it results from the unique circular symmetry of the problem. Most importantly, the fit reveals that the amplitude of E_{CDEW} follows $C \simeq 1/R = 1/KP$ (Fig. 6(e)), which is once again consistent with Eq. (2). The present measurement provides a high signal-to-noise characterization of the amplitude decay profile of the CDEW over substantial distances. Indeed, multiplying the annular groove radius by a factor K increases its perimeter and hence the total strength of the CDEW field source propagating towards the hole by a factor K. This exactly compensates for the $1/K$ reduction factor in the amplitude of the CDEW arriving at the hole due to the increased distance to the hole. As a result, the amplitude of the CDEW interfering with the light impinging directly on the hole is independent of K, resulting in interference spectra (Fig. 6(d)) which display constant amplitude for all K.

The single annular groove is of course expected to launch an SP, however with an amplitude expected to increase linearly with K at the hole (as a result of the corresponding linear increase with K of the integrated SP source field strength along the perimeter, together with a constant amplitude as a function of distance of the SP generated per unit length). The constant amplitude of the spectral modulation observed in Fig. 6(d) therefore rules out any significant surface-plasmon contribution on the scale of the detected signal.

We emphasize that the CDEW is not a long-range guided mode of the surface, as is made explicit in the decay of its amplitude with propagation distance, whereas the SP launched by a subwavelength groove (green dot in Fig. 4(b)) is a long-range propagating mode. However, Fig. 5(e) and Fig. 6(e) demonstrate that within at least 6 μm from the source (for an Ag–air interface at optical frequencies) the contribution of the CDEW overwhelms that of the SP. Such a distance corresponds to about 10 times the lattice constant of typical hole arrays, such as those studied in Section 2.1. We therefore expect the CDEW to completely dominate the transmission characteristics of hole arrays, and the role of the SP to be insignificant.

3.4. Mechanism for transmission enhancement and suppression

In the previous section we have shown that the optical transmission properties of an aperture can be strongly affected by the CDEW diffracted by another individual subwavelength feature (hole or groove) located up to several microns away along the surface. In this section we consider the case of a film with one or more subwavelength apertures and surface corrugations on one or both sides, illuminated on its entire front side with a plane wave.

#4585 - $15.00 US Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

For a given aperture, the illumination intensity at the entrance is modified by a function $A_1(\lambda)$ due to the interference of CDEWs launched by neighboring features. The transmission through the hole is characterized by its intrinsic transmission, defined as $T_H = (Q_R + Q_E)/Q_0$ measured in the case of the single, isolated hole, where Q_E is the total power emitted into evanescent modes on the exit side of the hole, in a diffraction process similar to that described in Section 3.1 and illustrated in Fig. 4. At the aperture exit, the intensity is in turn modified by a function $A_2(\lambda)$ which results from the interference by CDEWs diffracted by neighboring holes (first order process), as well as back-scattering of CDEWs, generated by the aperture itself (higher-order processes). There is strong evidence [15, 23] that the two modulation functions $A_1(\lambda)$ and $A_2(\lambda)$ are independent of one another for optically thick films, and that neither modifies $T_H(\lambda)$. New results on an array of well-defined cylindrical holes [24] show that $T_H(\lambda)$ is indeed independent of both $A_1(\lambda)$ and $A_2(\lambda)$. Therefore, the overall transmission function as measured by an objective of numerical aperture NA, $T_C(\lambda)$, can be found by cascading multiplicatively the front-surface modulation function $A_1(\lambda)$, the intrinsic transmission function through the hole $T_H(\lambda)$, the back-surface modulation function $A_2(\lambda)$, and the function $f_C(\lambda) = Q_C/(Q_R + Q_E)$ representing the fraction of the total power transmitted to the back surface of the film (including both radiative and evanescent surface modes) which is collected by the lens. $T_C(\lambda)$ is then written:

$$T_C(\lambda) = A_1(\lambda; n_1, P_1, d_1) T_H(\lambda; n_H, d, t) A_2(\lambda; n_2, P_2, d_2) f_C(\lambda; NA, P_2, d_2), \tag{3}$$

where n_1 and n_2 are the refractive indices of the dielectric media on front and back sides of the film, respectively, n_H is the refractive index of the medium inside the aperture, and t is the film thickness. P_1 and P_2 are the lattice constants and d_1 and d_2 the lateral dimensions of the periodic features on front and back sides, respectively.

In the limit of very large arrays and $\lambda > P$, the single outgoing beam is collimated; by time reversal invariance $A_1(\lambda) = A_2(\lambda)$ so that the transmission in that region of the spectrum is independent of which side of the film is illuminated, even if the film is not symmetric, as is the case for a perforated film supported by a substrate [4, 1].

Equation (3) indicates that $T_C(\lambda)$ will display a local maximum at wavelengths for which either $A_1(\lambda)$ or $A_2(\lambda)$ has a maximum. A maximum in $A_1(\lambda)$ corresponds to constructive interference of front-side CDEWs with the light directly incident on the hole, i.e. to a maximum in illumination intensity of the hole. Since the hole entrances are radiative scattering centers, this also leads to an intensity maximum in the reflected diffraction orders (apart from the special case of the zero-order reflected beam discussed in Sections 2.2 and 4.2). A maximum in $A_2(\lambda)$ corresponds to constructive interference, at the hole exits, of back-side CDEWs with light emerging directly from the holes. Since the hole exit openings also represent radiative scattering centers, this also corresponds to a condition of enhanced desorption of power carried by back-side CDEWs. This leads to an intensity maximum in all the diffracted orders, including the zero-order beam which is always intercepted by the objective. Conversely, local minima in $T_C(\lambda)$ are associated with local minima in $A_1(\lambda)$ or $A_2(\lambda)$, corresponding to destructive interference of CDEWs with the incident or emerging light, at the hole entrances or exits, respectively. This leads to suppressed illumination at the hole entrance or suppressed desorption at the hole exit, respectively.

In summary, CDEWs are excited by diffraction at all wavelengths at both the entrance and the exit surface of a given aperture, and thus represent a power-loss channel along the respective surfaces. Some of this lost power can be recovered into the transmitted signal when the CDEW is intercepted by neighboring holes in phase with their incident or transmitted illumination, on front or back surfaces, respectively.

#4585 - $15.00 US Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

3.5. *Comparison of model to data*

We apply the analysis outlined above to the case of a single subwavelength slit, flanked on its illuminated side by $\pm N$ parallel subwavelength-width grooves, periodically spaced by P (such as shown in Fig. 7(a)). Under plane-wave illumination, each groove launches a CDEW which interferes with the field of the direct illumination at the slit, resulting in a spectral modulation at the slit entrance given by:

$$A_1(\lambda) = \left(1 + 2\sum_{j=1}^{N} \frac{\alpha d}{jP} \cos\left(\frac{2\pi}{\lambda} n_{eff} jP + \frac{\pi}{2}\right)\right)^2, \tag{4}$$

where α is the amplitude of the CDEW launched at each groove relative to the amplitude of the incident light, and n_{eff} the effective index of refraction experienced by the CDEW along the interface. We recall that the data of Fig. 5 shows that Eq. (4) is valid also for a linear row of holes.

The radiative transmission coefficient $T_R(\lambda)$ is given by Eq. (3) with the numerical aperture NA set to 1. Since $A_2(\lambda) = 1$ for a flat output surface, and both f_C and T_H are slowly varying monotonic functions for a slit, we deduce that $T_R(\lambda) \propto T_C(\lambda) \propto A_1(\lambda)$. Figure 7(d) shows the theoretical function $A_1(\lambda)$ plotted as a function of $\lambda/(n_{eff}P)$ for $N = 0, \pm 1, \ldots \pm 5$. In this figure, the CDEW amplitude is treated as an adjustable parameter; the value $\alpha = 0.7$ yields a best fit to experimental observations.

For $N = \pm 1$, $A_1(\lambda)$ has a maximum at $\lambda = 4P/3$. The substantial red-shift of this value compared to P is a direct result of the characteristic CDEW phase shift of $\phi = \pi/2$. We note that a similar red shift is also observed in the transmission spectra of a device consisting of a single slit in a thick sheet of Al alloy, patterned with a single groove to either side (i.e. $N = \pm 1$) and illuminated with microwave radiation [23]. A transmission peak is observed at a wavelength $\lambda = 1.46P$, consistent with the CDEW model which, with $n_{eff} = 1.1$, predicts $\lambda = n_{eff}(4P/3) = 1.47P$.

As N is increased, the spectral peaks in Fig. 7(d) develop a pronounced asymmetrical profile, with a sharp rise on the left (through a fixed "pivot point" at $\lambda_{pp} = 0.95P$) and a more gradual decay to longer wavelengths, similar to the asymmetric peak profiles typically observed in the transmission spectra of hole arrays.

The as-measured experimental transmission spectrum $Q_C(\lambda)/Q_0$ for such a device is plotted in Fig. 7(b) for select values of N. Features similar to those of the analytical model of Fig. 7(d) are obtained as N is increased from 1 to 5, including a shift of the peak wavelength towards P , as well as the occurrence of a pivot point at $\lambda = 1.04P = n_{eff}\lambda_{pp}$, where the effective index is once more found to be $n_{eff} = 1.1$.

In contrast, a simple analytical surface-plasmon model for $A_1(\lambda)$ similar to Eq. (4) (where the SP is characterized by a phase $\phi = 0$, an amplitude which is constant as a function of distance, and an effective index $n_{eff} = n_{\mathrm{SP}}$) predicts a sharpening of the peak with increasing N, at a peak wavelength $\lambda = n_{\mathrm{SP}}P$ which is independent of N (Fig. 7(c)). Clearly the CDEW model of Fig. 7(d) gives a better correspondence to the experimental observations of Fig. 7(b). A complete solution of the Maxwell equations [12] for this structure yields results very similar to those of Fig. 7(d).

For single-hole devices with output corrugations, first-order decoupling dominates the beaming process [12, 27], and we still have $A_2(\lambda) = 1$. However, in the case of a hole array, additional sources are present on the exit surface. When the dielectric media are matched, $A_1(\lambda) = A_2(\lambda) \equiv A(\lambda)$ and T_C is simply given by $T_C(\lambda) = f_C(\lambda)A^2(\lambda)T_H(\lambda)$. Figure 8 compares the function $[A(\lambda/P)]^2$ derived from the experimental data for a 9×9 array, to that predicted from the CDEW model (adapted to a two-dimensional configuration by taking into ac-

#4585 - $15.00 US Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

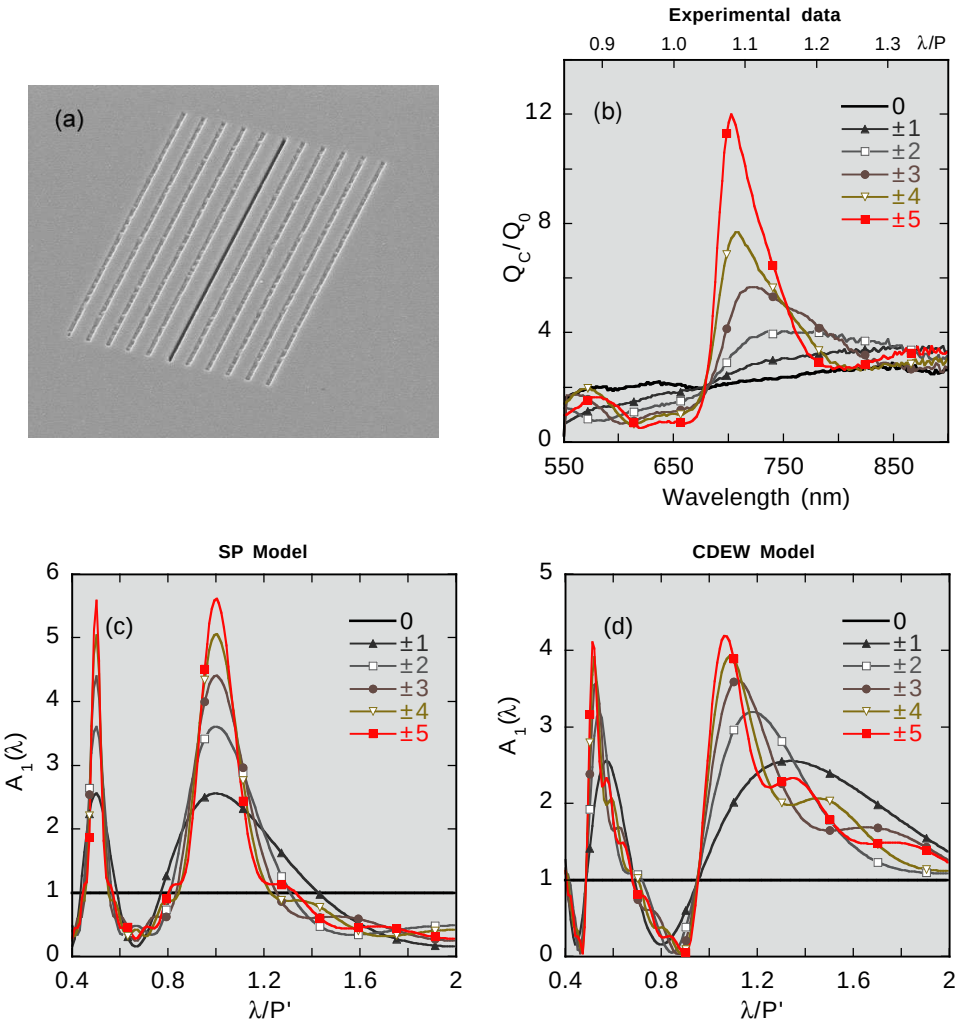

Fig. 7. Spectral effect of increasing the number of surface-wave sources: experimental results compared to simple analytical predictions for single slit surrounded by $\pm N$ periodic grooves. (a) SEM micrograph (FOV = 12.7 μm, $a = 45°$) of experimental device consisting of a single slit (width $w = 100$ nm, length $l = 10$ μm) in a silver film ($t = 340$ nm) on glass, surrounded by ± 5 grooves ($P = 650$ nm, $w = 100$ nm, $h \simeq 50$ nm); (b) experimental transmission spectra, N varied from 1 to 5 with other parameters constant as in (a); (c) intensity modulation at slit entrance predicted by SP model; (d) intensity modulation at slit entrance predicted by CDEW model.

count the effect of the various lattice planes up to order (3,1)). This two-dimensional CDEW-based model provides a good fit to the data in terms of both peak position and peak width. Since the wavelength-dependence of the factor α (which represents the fraction of light diffracted into a CDEW at each groove) is not taken into account, discrepancies develop at shorter wavelengths (where d/λ is effectively larger, resulting in less diffraction into evanescent modes and a reduced value for α).

#4585 - $15.00 US
Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

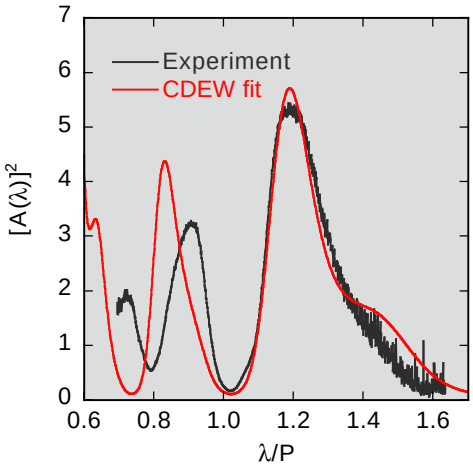

Fig. 8. Cascaded two-surface modulation function $[A(\lambda)]^2$ for a 9×9 hole array facing identical dielectric media on both sides: function derived from experimental data of Fig. 1 compared to prediction of CDEW model.

3.6. Search for surface plasmon contribution

While our new model accounts for all the features in the transmission spectra, we do not deny that surface plasmons are launched whenever a plane wave is incident on a metallic surface with a periodic structure, as discussed above in Sections 3.1 and 3.3. But one must ask, what is the relative contribution of the SPs to the transmission spectra? Unfortunately, this question is difficult to answer in the case of a periodic array of holes, where the peak wavelength associated with any SP resonance coincides [5] with the deep Rayleigh minimum associated with the Wood's anomaly [25]. It has been argued that the SP dispersion may be modified significantly by the very deep corrugation represented by a hole array to the extent that the transmission peak is red-shifted significantly compared to the smooth-surface limit, but we are aware of no theoretical work to back this *ad hoc* statement. On the contrary, experimental transmission spectra of a periodic array of slits show SP features at wavelengths very close to the smooth-surface dispersion [8, 9].

We believe that the evidence put forth in the preceding sections is sufficient to refute the SP model. Nevertheless, we set out to find out what contribution, if any, the SPs make to the transmission. We do this under circumstances which are optimized to reveal the SPs. We use Ag films, which have a dielectric constant with large real part $|\varepsilon_m^{'}|$ and small imaginary part $\varepsilon_m^{''}$ in the visible region, and choose a geometry with axial symmetry in which the Rayleigh anomaly is absent. Such a geometry is provided by a single hole surrounded by concentric ring grooves, with radii $R_K = (K+f)P$, where $K = 4, 5, \ldots 19$ and $f = 0, 0.1, \ldots 1.0$ (see Fig. 9). The innermost rings are intentionally omitted to reduce the sensitivity of the transmission spectra to the structure in the immediate vicinity of the hole; a total of 16 rings are spaced by $P = 600$ nm in order to create periodic structure in the radial direction, enough to develop a well defined-SP mode. Since the SP is a long-range propagating mode, any SP resonance on the grating is observable via the hole (which acts like an NSOM built into the surface). The effect of varying f is to shift the observation point with respect to the grating, without however changing the lattice constant of the grating.

Figure 9(b) shows the gain spectrum of the device, $G(\lambda) = Q_{C,16}/Q_{C,0}$ (where the normalization is to the transmission spectrum of an equivalent single hole, $Q_{C,0}$); in the interest of

#4585 - $15.00 US
Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

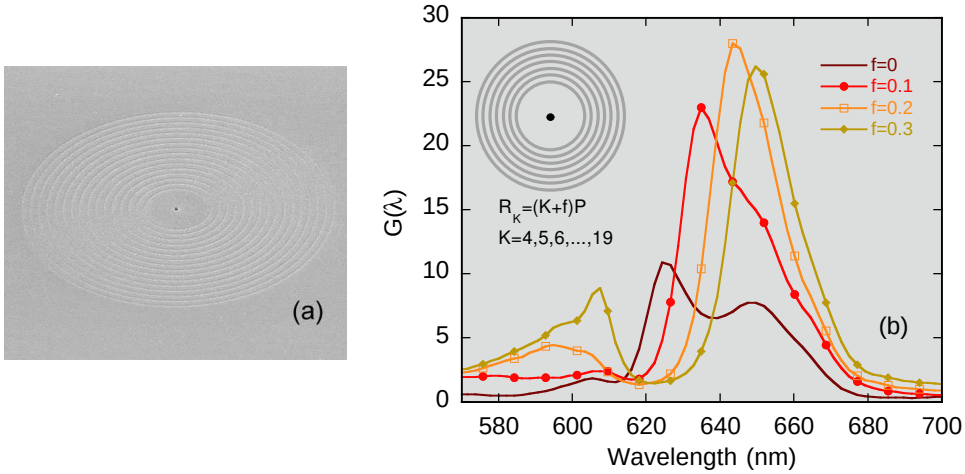

Fig. 9. Surface-wave detector consisting of a single hole ($d = 300$ nm) surrounded by 16 ring grooves ($w = 150$ nm, $h = 100$ nm) in a silver film ($t = 250$ nm) on glass. Grooves form a concentric grating with period $P = 600$ nm, starting at radius $4P$. (a) SEM micrograph of device (FOV $= 30.4$ μm, $a = 52°$); (b) effect on gain spectrum $G(\lambda)$ of increasing hole-grating radial distance by increments of $P/10$.

clarity, spectra for only four values of f are displayed. Two distinct peaks can be observed in the spectrum for $f = 0$ (at wavelengths of 625 and 650 nm, respectively), which have previously been attributed to distinct SP modes of the grating [2]. However, the spectrum evolves dramatically, with peaks varying in position, amplitude and number, as the observation point is shifted by minute, 60-nm increments. This indicates that the peaks are not a signature of an inherent resonance of the grating, but simply represent artifacts of the measurement at the hole, where the surface wave launched by the grating interferes with light incident directly on the hole. Based on its interference properties at the hole, this surface wave presents a rapidly varying phase within a broad envelope function ranging from $\lambda = 590$ to 680 nm, peaking at $\lambda_p = 648$ nm $= n_{eff}P$, where $n_{eff} = 1.08$, close to the value of 1.1 observed for n_{eff} in the previous sections. In addition, the highest value for $G(\lambda)$ is observed close to the value $f = 0.25$, consistent with the observation that the CDEW arriving at the central hole has a phase shift $\phi = -\pi/2$ with respect to the incident light (see the discussion of Fig. 6). The data of Fig. 9 is therefore entirely consistent with the CDEW model.

In comparison, the expected signature of an SP resonance is a fixed, single transmission peak at 630 nm [3, 8, 9], with a FWHM of about 25 nm. Thus, the strong dependence on f of the peak positions in the data of Fig. 9(b) are not consistent with the SP-resonance model. We therefore conclude that even in such a structure which is optimized for detecting surface plasmons, the SP contribution to the transmission spectrum is negligible compared to that of the CDEWs.

4. Discussion and conclusion

4.1. Comparison to other models

One of the most attractive features of the CDEW model we have proposed is its simplicity. We do not claim to offer a complete solution to the Maxwell equations as has been done in many numerical calculations [5, 8, 11, 12, 13, 26, 27, 28, 29]. For instance, we have considered only the first-order effect of scattering into a CDEW and subsequent propagation on the surface, and have not included such higher-order processes as multiple scattering. Yet the model is

#4585 - $15.00 US
Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

successful in the sense that it accounts for the salient features of the transmission spectra of hole arrays or single apertures surrounded by a corrugated surface. Indeed, in the simplicity of the model lies its elegance, for the CDEW model reveals the essence of the physics underlying the transmission enhancement (and suppression) in a clear and intuitive way, so that it can be used as a platform for further development, both in understanding the underlying physical principles and for the purpose of device design and optimization. We note in passing that the limited range of the CDEW tends naturally to disfavor multiple scattering. Moreover, the present structures are optimized for operation (high transmission) at the condition $\lambda \simeq P$, which favors in- and out-coupling normal to the surface at the expense of higher-order in-plane processes. This is the exact opposite of the Bragg condition $\lambda = 2P$ required for efficient in-plane confinement (such as is used in the operation of distributed feedback lasers for example).

The model proposed here is similar in spirit to the dynamical diffraction model of Refs. [28, 29] which offer a complete solution to the problem of enhanced-transmission through a one-dimensional array of slits in terms of propagating Bloch waves. We believe that the successful experimental predictions of our intuitive CDEW model vindicate such an approach based entirely on diffraction.

A leaky surface plasmon theory [30] includes scattering by the surface topography, but considers only SPs interacting with and emerging from the subwavelength aperture(s) whereas the CDEW model includes all the evanescent modes. Another theory [31] treats the transmission peaks of hole arrays in terms of second-order scattering from classical diffraction lobes as they become tangential to the surface. However this model predicts only transmission enhancement but not suppression, and inaccurately predicts transmission maxima at positions of experimental minima. Several groups [32, 33] have proposed to interpret the transmission spectra of hole arrays in terms of the Wood's anomaly, which incorporates both the minima as Rayleigh anomalies and the maxima as Fano resonances [25], but only in a phenomenological way. The CDEW model is in essence a Wood's anomaly model that goes beyond the phenomenological approach and offers a concrete and quantitative (albeit to first order) description of the transmission of hole arrays, of which the principles are amply supported by experimental data.

4.2. Consideration of selected data in light of CDEW model

In this section we consider a selected set of data which has been interpreted with the SP model, and re-consider the results in the light of the CDEW model.

A recent observation that the zero-order reflectivity spectrum exhibits minima at wavelengths corresponding to maxima in the zero order transmission spectrum, was interpreted as evidence for absorption into SP modes at those wavelengths [21]. The CDEW model predicts that for a symmetric situation such as a hole array in a free-standing membrane, the radiative emission due to the CDEW is the same on the front and back surfaces, and one would thus expect transmission maxima to coincide with reflection maxima, and to be of roughly the same magnitude (relative to the average light intensity on either side of the film). This is in fact observed (see Fig. 3(b)) in the spectrum of the first order reflected beam. The zero-order reflected beam has contributions not only from the CDEW scattering off the corrugated surface, but also from the specular reflection of the main beam. Since on a metallic surface the specularly reflected light undergoes a phase change of π, it is out of phase with the CDEW contribution which thus causes a reduction of the total zero-order intensity, without the need for absorption. We should mention here that by Babinet's principle, the zero order reflection spectrum of an array of dots is expected to be the same as the zero order transmission spectrum of an hole array in a screen. Preliminary measurements show that this is indeed the case.

An intriguing experiment has demonstrated [34] that when two photons are generated in a quantum entangled state, the entanglement survives even when one of the photons is made

#4585 - $15.00 US Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

to pass through a hole array. This result is surprising if it is assumed that the light incident on the hole array is converted into a resonant SP mode which subsequently "tunnels" through the holes. However, in the CDEW model the scattered light, including the evanescent waves, remains coherent with the incident radiation, and it is not surprising that the entangled state is preserved on passage through the hole array (up to a small effect due to finite losses and geometrical imperfections at the metal surface).

The time delay experienced by a light beam on passing through a hole array corresponds to a surface propagation length of a few lattice constants [35], consistent with the fact that the CDEW in our model decays like the inverse of the propagation distance, so that the scattering of corrugations or holes located beyond than 4–5 nearest neighbors make a negligible contribution to the total field at any given hole. This is in turn consistent with the observation that the transmission modulation saturates for a relatively small 9×9 array size (Fig. 1(b)). This finite-size effect, observed also in very large arrays and a long-standing mystery in the SP model, follows naturally from the CDEW model.

A study on the effect of the hole depth or film thickness t on the transmission of hole arrays (at constant d and P) has shown that the peak transmission at the longest-wavelength peak depends on t [36]. However, when the magnitude of the transmission enhancement $G(\lambda_p)$ is obtained by normalizing to a comparable-diameter comparable-diameter hole, treated as a waveguide, it appears that the transmission enhancement has a much weaker thickness dependence, consistent with the results from Fig. 2. In the same study, a broadening of the transmission peak for the thinnest films has been interpreted as evidence for a coupling between SP modes on the front and back surfaces of the film. In the framework of the CDEW model, we interpret the broadening of the spectra in terms of a finite-size effect: the intrinsic hole transmission T_H becomes larger with decreasing t; this acts as a leakage channel for the CDEW so each hole is reached by CDEWs from fewer neighbors. The effective reduction in N (see Eq. (4)) then results in a reduction of $A_1(\lambda)$, a broadening of the peak, as well as a red shift (compare to the data of Fig. 7), all features which are observed in the data of Ref. [36]. A similar lateral screening effect and effective reduction in N is interpreted to account for the broadening and red-shift of hole-array transmission peaks as the hole diameter is increased, such as observed in Fig. 2(c,d).

Finally, the CDEW model explains how light emerging from a single subwavelength aperture can excite a surrounding area with periodic surface corrugations with high efficiency (and a well-defined phase) to produce an intense, low-divergence beam [14]. The SP resonance model fails to predict the high efficiency of the process, since a subwavelength aperture is a point source which is extremely ill-suited to SP grating coupling; moreover the SP mode launched directly by the aperture by diffraction into the evanescent regime ($k_x > k_0$) is but one of a continuum of traveling modes and thus carries a small fraction of the power emitted by the aperture. The observed beaming conditions are consistent with the fundamental characteristics of the CDEW: beaming at all wavelengths (for example for linear-groove device of Ref. [14] resulting from the launching of a CDEW at all wavelengths; a red shift with respect to the groove period of the optimal wavelength for beaming, resulting from the phase shift $\phi = \pi/2$ of the CDEW for the linear case; and finally a laterally confined emission area corresponding to the limited lateral range of the CDEW.

4.3. Conclusion

In summary, we have presented evidence contrary to the notion that a resonance with SP modes leads to enhancement of the transmission through subwavelength hole arrays. We have proposed a new model which involves the interference of composite evanescent waves generated by diffraction of the incident light at the constituent subwavelength surface structures. We em-

#4585 - $15.00 US Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

phasize once more that the CDEW of our model can be clearly distinguished from an SP mode. The latter is simply not supported by nonmetallic systems, yet nonmetallic samples display the same fundamental transmission characteristics as metallic ones. In addition, a search for an SP contribution in an optimized geometry and material yielded a negative result on the scale of the detected CDEW. Experimental evidence confirms that the physical properties of CDEWs are as described in the proposed model, and are markedly different from those of SPs. The CDEW model is consistent with a large body of data published on hole-array transmission. In particular, the interference aspect of the model directly predicts transmission suppression as well as enhancement, which is observed for properly normalized $T_R(\lambda)$. The transmission enhancement in hole arrays is found to be at most a factor $G \simeq 7$, a value consistent with the two-sided lateral-interference mechanism underlying the CDEW model.

While the small transmission enhancement diminishes the relevance of hole arrays for use in a number of applications, the modulation in the transmission spectrum is still attractive for use in, for instance, display devices: Since the positions of the maxima and minima are tunable, this means that for fixed wavelength the transmission efficiency can be controlled [37]. Display devices based on hole arrays can be light and flexible, and the peak transmission can in principle be close to unity. This high efficiency could result in mobile devices with vastly improved battery lifetimes.

Although $G \leq 7$ for hole arrays, it can be far larger for a single hole with an optimized corrugation [2] (see also Fig. 9). The largest transmission enhancement reported to date is $G = 125$ for a single hole with a set of concentric ring grooves, periodically spaced in the radial direction [38]. In that sample, second harmonic generation was demonstrated with an enhanced efficiency: For a given input power, the intensity in the second harmonic is increased by a factor 10^4 compared to the signal from a bare hole in the same film, since the second harmonic intensity follows the square of the excitation intensity. Such a transmission enhancement is very promising for near-field optical microscopy, in particular real-time high-resolution imaging or microscopy of nonlinear optical processes such as Raman scattering and optical conversion. Concentric-ring devices may also prove to be useful in applications such as read-heads for high-density data storage [39, 40].

The CDEW model has bearing on the long-standing problem of Wood's anomaly, in which the diffraction efficiency of a grating shows large variations as the wavelength is varied. The variations are usually characterized as a combination of deep minima, known as Rayleigh anomalies, and high maxima, labeled Fano resonances [25]. Both of these effects follow naturally from the interference properties of CDEWs, with Rayleigh minima occurring at conditions of destructive interference of the CDEW with the incident light, and Fano-resonance occuring at conditions of constructive interference (the spectral positions and relative magnitudes of these minima and maxima depending sensitively on the details of the surface and its corrugation geometry).

Similarly, the CDEW model offers a quantitative yet intuitive approach to the related issue of dichroic filters, also known as metallic meshes or frequency-selective surfaces [41, 42]. These are perforated metal films of which the transmission spectra exhibit intensity variations with such high contrast ratios that they are useful as filters. In fact, the first reports of the enhanced transmission phenomenon in hole arrays can be traced back to studies on near-millimeter [43] or far-infrared [44] band-pass filters. In the first case [43], a hexagonal array of holes in thick brass ($t = 3$ mm, $d = 1.4$ mm, hole spacing $s = 1.6$ mm) was shown to display a distinct transmission peak at a red-shifted wavelength $\lambda_p = 1.44P$, where $P = 0.87s$ is the smallest spacing between the rows of holes, leading to a per-hole transmission coefficient $T_R(\lambda_p) = 1.17$. In the second case [44], a distinct red-shifted transmission peak, $\lambda_p = 1.12P$ corresponding to a per hole transmission coefficient $T_R(\lambda_p) = 3.2$ was observed for a hexagonal arrays of holes

#4585 - $15.00 US Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004

in Ni ($t = 52$ μm, $d = 37$ μm, $s = 72$ μm). In both cases, the transmission is enhanced to efficiencies larger than unity when normalised to the fractional area occupied by the holes. Such high efficiencies are comparable to values for $T_R(\lambda_p) > 1$ obtained for hole arrays operating in the visible (such as the devices of Fig. 1 or Fig. 2), hinting at the key role of CDEWs, in all frequency ranges, in recovering power lost to evanescent modes by diffraction.

The most efficient hole-array filters are metallic simply because the large index contrast between the film material and the holes results in the most efficient generation of evanescent waves. But a metal surface is not necessary, as demonstrated for instance in beaming effects at the surface of a photonic crystal [13]. The CDEW model is thus not limited to the problem of transmission through subwavelength apertures in metal films, but is relevant to a large class of systems with periodic surface structures.

#4585 - $15.00 US Received 10 June 2004; Revised 30 July 2004; accepted 2 August 2004