Analytical theory of extraordinary transmission through metallic diffraction screens perforated by small holes

R. Marqués,1,* F. Mesa,2 L. Jelinek,1 and F. Medina1

1 Departamento de Electrónica y Electromagnetismo. Universidad de Sevilla
Avd. Reina Mercedes, 41012 Sevilla (Spain)
2 Departamento de Física Aplicada 1, ETSII, Universidad de Sevilla
Avd. Reina Mercedes, 41012 Sevilla (Spain)
* Corresponding author: marques@us.es

Abstract: In this letter, the problem of extraordinary (ET) transmission of electromagnetic waves through opaque screens perforated with subwavelength holes is addressed from an analytical point of view. Our purpose was to find a closed-form expression for the transmission coefficient in a simple case in order to explore and clarify, as much as possible, the physical background of the phenomenon. The solution of this canonical example, apart from matching quite well with numerical simulations given by commercial solvers, has provided new insight in extraordinary transmission as well as Wood’s anomaly. Thus, our analysis has revealed that one of the key factors behind ET is the continuous increase of excess electric energy around the holes as the frequency approaches the onset of some of the higher-order modes associated with the periodicity of the screen. The same analysis also helps to clarify the role of surface modes –or spoof plasmons– in the onset of ET.

© 2009 Optical Society of America

OCIS codes: (050.0050) Diffraction and gratings; (050.0050) Diffraction theory.

References and links
1. Introduction

A few years ago Ebbesen et al. [1] reported a phenomenon of extraordinary transmission (ET) through metallic screens periodically perforated with sub-wavelength holes. This physical effect was originally attributed to the excitation of surface plasmons on the diffraction screen [3, 4, 5], in apparent contradiction with Bethe’s theory for small apertures [2]. In a first period, this phenomenon was mainly related to the plasma-like behavior of metals at optical frequencies. However ET has also been observed at millimeter wave frequencies [6], when metals can no longer be considered solid plasmas (but rather quasi-perfect conductors). This last experimental observation can be explained by means of diffraction theories [7, 8] which emphasize the role of screen periodicity in ET, being the specific behavior of the opaque material secondary to the phenomenon. Subsequently, the surface plasmon concept was rescued to explain ET after considering that plasmon-like surface waves (which some people call spoof plasmons) can be supported by structured metallic surfaces [9, 10] even in the perfect conductor limit. More recently, some of the authors of the present paper proposed a comprehensive equivalent circuit model based on the theory of discontinuities in hollow waveguides [11, 12]. This model accounts for the most salient features observed in ET experiments as well as for fine details of the transmission spectrum obtained from exhaustive numerical computations. For a detailed and comprehensive discussion on the topic, the reader might consult [12] (and references therein) as well as the excellent reviews by C. Genet et al. [13] or by F.J. García de Abajo [14].

The above works, among others, show that ET can be addressed from many different perspectives: surface waves excitation, diffraction theory, equivalent circuit models,..., with each one providing a different approach to the problem. The main aim of this letter is to gain insight into the physics behind ET through the development of an accurate analytical solution of a canonical example. A first and very valuable attempt to develop this analytical solution was recently developed in [15]. However, it will be shown that although many qualitative conclusions of this analysis are correct, the numerical results presented in the paper are inaccurate probably due to inappropriate approximations in the derivations. We feel that the present analytical solution not only provides accurate numerical results but also shed new light on the problem and makes apparent the interconnection between the previous perspectives.

For simplicity, we will first analyze a zero thickness perfect conducting plate with square holes placed in a square periodic array. Although the present analysis can readily be extended to more complex geometries, the straightforward physical interpretation of this simple structure allows for a better understanding of the physical effects. Following the rationale in [12, 15], the first step in our analysis is the transformation of the diffraction problem into the problem of a small diaphragm inside a TEM waveguide. Then, the problem is solved employing well-known results of diffraction and waveguide theories, and the accuracy of the numerical results is validated by careful electromagnetic simulations. However, as it was already mentioned, the main purpose of this work is not to provide a tool for calculations (something that can easily be obtained from numerical techniques already implemented in common com-
mercial electromagnetic solvers) but to give a different physical insight into the physics of ET. To better achieve this goal, an equivalent circuit (rather than a circuit model) is proposed. The equivalent circuit will implicitly contain all the information already provided by the analytical solution, but it has the additional advantage of making its physical interpretation much easier. The role of the different waveguide modes in the onset of ET will be analyzed and connected with the frequency dependence of the different elements of the proposed equivalent circuit. The same analysis will also be applied to elucidate the role of surface waves (or spoof plasmons) in the onset of ET. Finally the present proposal will allow us to link the reported results—which come basically from a diffraction theory analysis—to the circuit theory approach proposed in [11, 12].

Fig. 1. Perfect conductor screen perforated with square holes: front view (a) and two lateral cuts (b). Front (c) and lateral (d) views of the structure unit cell or equivalent waveguide. (e) Equivalent circuit for the discontinuity in the waveguide. It has been assumed that $t \to 0$.

The structure under study is shown in Figs. 1(a)-(b). For normal incidence of a $y$-polarized electromagnetic plane wave, this structure is equivalent to a TEM waveguide with perfect elec-
lectric conducting plates at the upper and bottom interfaces, perfect magnetic conducting plates at both lateral sides, and a square diaphragm located, say, at $z = 0$ (see Figs. 1(c)-(d)). Due to the symmetry of this structure, and assuming an incident field of amplitude equal to unity, the field component $E_y$ at $z = 0^+$ can be expanded into the following Fourier series:

$$E_y(x, y) = T + \sum_{n=1}^{N} A_{n0}^{TE} f_{n0}(x, y) + \sum_{m=1}^{M} A_{om}^{TM} f_{0m}(x, y) + \sum_{n,m=1}^{N,M} (A_{nm}^{TE} + A_{nm}^{TM}) f_{nm}(x, y),$$

(1)

where $T$ is the transmission coefficient, $A_{n0}^{TE}$, $A_{om}^{TM}$ are the coefficients of the (below cutoff) TE and TM waveguide modes excited at the discontinuity, and $f_{nm}(x, y) = \cos(2n\pi x/a) \cos(2m\pi y/a)$. Using waveguide theory [16], the electric field component $E_x$ can be written as

$$E_x(x, y) = \sum_{n,m=1}^{\infty} \left( \frac{m}{n} A_{n0}^{TE} - \frac{n}{m} A_{om}^{TM} \right) g_{nm}(x, y)$$

(2)

with $g_{nm}(x, y) = \sin(2n\pi x/a) \sin(2m\pi y/a)$.

The coefficients of the expansion (1) can be obtained from the orthogonality properties of functions $f_{nm}$. However, for small holes and not very large values of $n$ and $m$ (taking into account that $E_z$ must be zero at the metallic screen), the following approximation applies [15]:

$$\iint_{w} E_x f_{nm} \, dx \, dy = \iint_{h} E_x f_{nm} \, dx \, dy \approx \iint_{w} E_x \, dx \, dy = \iint_{h} E_x \, dx \, dy = a^2 T,$$

(3)

where subindex $w$ and $h$ stands for the waveguide and hole sections respectively, and $a^2$ is the waveguide section. Thus, it is finally found that

$$A_{n0}^{TE} \approx 2T; \quad A_{om}^{TM} \approx 2T; \quad A_{nm}^{TE} + A_{nm}^{TM} \approx 4T.$$

(4)

For small holes, $E_z$ should be almost zero at the hole (and zero on the metallic screen). Therefore, from (2) and (4):

$$A_{n0}^{TE} \approx 4T \frac{n^2}{n^2 + m^2}; \quad A_{om}^{TM} \approx 4T \frac{m^2}{n^2 + m^2}.$$

(5)

The transmission coefficient $T$ can now be obtained after imposing the appropriate boundary conditions for the transverse magnetic field. Since the scattered field is produced by the electric currents induced in the diffraction screen, which are confined to the $z = 0$ plane, it is deduced from symmetry that all the tangential components of the scattered magnetic field must vanish at the aperture. This conclusion comes out from the fact that such induced currents are vectors invariant by reflection in the $z = 0$ plane, whereas the scattered magnetic field is a pseudo-vector, whose tangential components must change of sign after reflection in such plane.

Therefore, the total tangential magnetic field in the hole must be equal to the incident field [16], that is, $H_x = -Y_0 = -\nabla \times \mathbf{E}_0$ and $H_z = 0^+$. Once the tangential magnetic field at the aperture has been evaluated, upon substitution of (4)-(5) in (1) the transmission coefficient $T$ can be obtained after solving the following equation:

$$b^2 Y_0(1 - T) = ab T \sum_{n=1}^{N} \frac{2}{n\pi} Y_{n0}^{TE} \sin\left(\frac{n\pi b}{a}\right) + ab T \sum_{m=1}^{M} \frac{2}{m\pi} Y_{02m}^{TM} \sin\left(\frac{m\pi b}{a}\right) + a^2 T \sum_{n,m=1}^{N,M} \frac{4}{nm\pi^2} \left( Y_{n2m}^{TE} \frac{n^2}{n^2 + m^2} + Y_{2n2m}^{TM} \frac{m^2}{n^2 + m^2} \right) \sin\left(\frac{n\pi b}{a}\right) \sin\left(\frac{m\pi b}{a}\right),$$

(6)

It is worth to mention that a completely different value for $H_z$ and $H_y$ was assumed in [15], which is probably the sources of the numerical discrepancies with our analysis (see Fig. 2).
where $Y^\text{TE}$ and $Y^\text{TM}$ are the TE and TM modal admittances given by [16]

$$Y^\text{TE}_{2n,2m} = i Y_0 \sqrt{\left(\frac{n \lambda}{a}\right)^2 + \left(\frac{m \lambda}{a}\right)^2 - 1}$$

(7)

$$Y^\text{TM}_{2n,2m} = -i Y_0 / \sqrt{\left(\frac{n \lambda}{a}\right)^2 + \left(\frac{m \lambda}{a}\right)^2 - 1}.$$  

(8)

Since the maximum “resolution” of Eq. (1) is limited by the minimum wavelength, $\lambda_m = a/n$, the upper limits of the series in (6) can be determined by imposing a “resolution” equal to the hole size $b$. This leads to $N, M \approx a/b$, which completes the determination of $T$ from (6). Results for the transmission coefficient for several values of $a/b$ computed from (6) are shown in Fig. 2 together with data coming from full-wave electromagnetic simulations using the commercial software CST Microwave Studio. Both set of results agree quite well not only qualitatively but also quantitatively. The figure also shows other previous analytical results on the same structure [15].

![Fig. 2. Transmission coefficient of the structure shown in Fig. 1 for different values of the ratio $a/b$ versus the ratio $(f_W - f)/f_W$, where $f_W = c/a$ is the Wood’s anomaly frequency, with $c$ being the light velocity in free space. Solid Lines correspond to data from (6). Dotted lines correspond to data from CST. For comparison purposes, the numerical value for the ET frequency provided in [15] for $a/b = 7.07$ (i.e. holes covering a 2% of the total area) is shown with an arrow.]

Now, the equivalent circuit shown in Fig. 1(e) is proposed for the waveguide discontinuity problem shown in Figs. 1(c) and (d). The transmission coefficient for this equivalent circuit configuration can be found from the solution of the following equation:

$$Y_0(1 - T) = T \left(-i \frac{\omega C}{2} - \frac{1}{i \omega 2L}\right).$$

(9)

where $\omega = 2 \pi f$ is the angular frequency. The above equation clearly shows that total transmission is obtained at frequency $\omega_0 = \sqrt{1/(LC)}$. Considering now that the evanescent TE(TM) mode admittances are imaginary and positive(negative) [16], a direct comparison between (9) and (6) leads to the following expressions for the capacitive, $B_C$, and inductive, $B_L$, susceptances
appearing in (9):

\[
B_C = -\omega C/2 = \frac{1}{i} \left\{ \frac{a}{b} \sum_{m=1}^{M} \frac{2Y_{0,2m}^{TM}}{m\pi} \sin \left( \frac{m\pi b}{a} \right) + \left( \frac{a}{b} \right)^2 \sum_{n,m=1}^{N,M} \frac{4Y_{2n,2m}^{TM}}{nm\pi^2} \frac{m^2}{n^2 + m^2} \sin \left( \frac{n\pi b}{a} \right) \sin \left( \frac{m\pi b}{a} \right) \right\} \quad (10)
\]

\[
B_L = \frac{1}{2\omega L} = \frac{1}{i} \left\{ \frac{a}{b} \sum_{n=1}^{N} \frac{2Y_{0,2m}^{TE}}{n\pi} \sin \left( \frac{n\pi b}{a} \right) + \left( \frac{a}{b} \right)^2 \sum_{n,m=1}^{N,M} \frac{4Y_{2n,2m}^{TE}}{nm\pi^2} \frac{m^2}{n^2 + m^2} \sin \left( \frac{n\pi b}{a} \right) \sin \left( \frac{m\pi b}{a} \right) \right\} . \quad (11)
\]

This result is somehow expected since it is well known that evanescent TM(TE) modes present a capacitive(inductive) behavior as they store mainly electric(magnetic) energy. The above transformations makes it apparent that the equivalent circuit of Fig. 1(e) is not a simple model but merely another way to express the previously obtained analytical results in a circuit-like fashion. It is then apparent the connection between the proposed equivalent circuit and the diffraction theory proposed in those papers.

The frequency dependence provided by the modal admittances appearing in (10) and (11) gives rise to some relevant facts. Near the Wood’s anomaly wavelength: \( \lambda = a \) (in our case, it also corresponds to the cutoff frequency of the TM_{02} mode), the admittance of the TM_{02} mode suddenly grows to infinity, which makes the term associated with this latter mode be dominant in the capacitive susceptance series (10). Under the same circumstances, however, the admittance of the TE_{20} mode goes to zero, which means that there is no singularity in the inductive susceptance \( B_L \). As it is well know from Bethe’s theory, for normal incidence and frequencies far and below Wood’s anomaly, a small hole has an inductive behavior (it can be modeled by an equivalent magnetic dipole). Nevertheless, as frequency approaches Wood’s anomaly, we have just seen that the absolute value of the associated capacitive susceptance \( B_C \) grows to infinity and, at certain frequency, it will cancel out the inductive susceptance associated with the hole (namely, \( B_C + B_L = 0 \)) and will give rise to total transmission. It is also interesting to note that, within the same previous frequency range, all admittances except \( Y_{0,2}^{TM} \) have a smooth frequency dependence. In that case, the inductive susceptance \( B_L \) (11) is found to be roughly proportional to \( (a/b)^2 \). However, the capacitive susceptance, which is dominated by \( Y_{0,2}^{TM} \), will be proportional to \( a/b \). This means that \( |B_C|/B_L \approx b/a \) as \( \lambda \to a \), which implies that the smaller the hole, the smaller the absolute value of \( B_C \) is with regard to \( B_L \). In other words, the smaller the hole, the closer ET is to Wood’s anomaly. An additional observation can be made after considering that the absolute value of the capacitive susceptance still increases for frequencies above ET until it becomes infinity at Wood’s anomaly. At this last frequency, the \( LC \) tank in the equivalent circuit of Fig. 1(e) becomes a short circuit and total reflection will appear. Therefore the equivalent circuit of Fig. 1(e) along with the transformations (10) and (11) explains satisfactorily both ET and Wood’s anomaly in periodically perforated zero thickness screens.

Next, it will be studied the behavior of the corresponding electromagnetic field at frequencies near Wood’s anomaly. According to (4) and (5), the amplitudes of the different evanescent
modes (measured as the amplitudes of the electric field component $E_y$) excited around the
diffraction screen are roughly of the same order. However, given that the absolute value of the admittance of the TM$_{02}$ mode is much larger than the admittance of any other mode, the near field will be dominated by the magnetic field component $H_x$ of this mode, as well as for its associated axial electric field $E_z$. Since the TM$_{02}$ mode is near cutoff, its associated attenuation constant is small and its near field configuration will extend relatively far from the diffraction screen. This near field picture is valid not only at ET (total transmission) but also at Wood’s anomaly (total reflection), and therefore it cannot explain by itself ET. Only when the combined effect of all the remaining evanescent modes is also considered, ET can be properly explained. That is, only when the admittance of the TM$_{02}$ becomes exactly so large (and its attenuation constant exactly so small) that the excess of electric energy stored in this mode equals the excess of magnetic energy stored by all the remaining evanescent modes, ET will occur. We feel that this is one of the most important conclusions of our analysis: ET is closely related to Wood’s anomaly because in both cases the energy stored in the evanescent TM$_{02}$ is much larger than the energy stored in any other evanescent mode. However, whereas Wood’s anomaly appears when this energy becomes infinite, ET appears when the excess of electric energy associated with this TM$_{02}$ mode cancels out the excess of magnetic energy associated with the remaining evanescent modes excited at the hole.

The previous analysis can also help to clarify the role of surface waves or *spoof plasmons* [9, 10] in the onset of ET. For this purpose we should consider the solutions of the equivalent circuit of Fig. 1(e) in the absence of excitation. These solutions provide the frequency at which the surface wave supported by the periodic structure has a propagation constant with zero real part so that phase matching with the normally impinging TEM wave is possible and surface plasmons are then excited. In such case only outgoing TEM waves can exist, and the left- and right-hand side transmission lines, which account for free space, can be substituted by resistances accounting for radiation in free space: $R_0 = 1/Y_0(= Z_0) = \sqrt{\mu_0/\varepsilon_0} \approx 377\Omega$. The frequency of excitation of surface plasmons with the appropriate wavevector, $k = 2\pi/a$, can now be identified with the frequency of resonance of the loaded LC resonator shown in Fig. 3. This frequency of resonance, $\hat{\omega} = \omega' - i\omega''$, must be complex in order to account for radiation losses, and can be computed as the solution to the following implicit equation:

$$-i\omega C(\omega) - \frac{1}{i\omega L(\omega)} + \frac{2}{R_0} = 0$$

(note that both $C$ and $L$ depends on frequency via the TM and TE admittances). The real part of the complex frequency is actually the frequency of excitation of the surface wave (for $k = 2\pi/a$), and its imaginary part gives the *lifetime* of the wave through $\tau = 1/\omega''$. Clearly, if $R_0$ is much larger than $|B_C|$ as well as $B_L$, the frequency of excitation of the surface waves will be very close to the frequency of ET (although both frequencies will never coincide). Table 1 shows a comparison between these resonance frequencies and the ET frequencies for the cases analyzed in Fig. 2. This table shows that the higher the ratio $a/b$ is, the closer both frequencies appear.
Table 1. Normalized resonance and extraordinary transmission frequencies, \((f_W - f)/f_W\), for the cases studied in Fig. 2.

<table>
<thead>
<tr>
<th>(a/b)</th>
<th>Norm. Reson. freq.</th>
<th>Norm. ET freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>7.3472E-03 -i9.4994E-04</td>
<td>7.4472E-03</td>
</tr>
<tr>
<td>5</td>
<td>1.5265E-03 -i8.9025E-05</td>
<td>1.5266E-03</td>
</tr>
<tr>
<td>6</td>
<td>4.4240E-04 -i1.3641E-05</td>
<td>4.4240E-04</td>
</tr>
<tr>
<td>7</td>
<td>1.5887E-04 -i2.9409E-06</td>
<td>1.5887E-04</td>
</tr>
<tr>
<td>8</td>
<td>7.5485E-05 -i8.0277E-07</td>
<td>7.5485E-05</td>
</tr>
</tbody>
</table>

In fact, both frequencies have more than five identical significant digits for \(a/b > 5\). Also, the imaginary part of \(\tilde{\omega}\) becomes almost negligible for high values of this ratio, which shows that this frequency exactly coincides with the frequency of ET in the limit of very small holes \((a/b \to \infty)\). However, for smaller values of this ratio, both frequencies, although close, show a significant deviation. Accordingly, the imaginary part of \(\tilde{\omega}\) becomes more significant as the ratio \(a/b\) increases. Larger holes would yield even higher differences between the frequencies associated with the surface plasmon and the extraordinary transmission.

Until now we have considered infinitely thin screens. However, the proposed theory can easily be extended to diffraction screens of finite thickness \(t\). In this case the circuit model of Fig. 1(c) must be modified in order to include the evanescent waveguide formed by the hole. If the hole is small, it will be assumed that only the dominant TE\(_{10}\) mode is significantly excited inside the hole, and hence the effect of higher order modes are neglected. Thus the hole is modeled as an evanescent transmission line with admittance equal to the admittance of this TE\(_{10}\) mode. For square holes, this admittance (defined as the average current through the hole divided by the average voltage across the hole) coincides with the wave admittance of the aforementioned TE\(_{10}\) mode: \(Y_{TE_{10}} = iY_0\sqrt{\lambda/2b} - 1\). Moreover, only a fraction of the current flowing through the diffraction screen will go through the holes. This fraction can be roughly estimated as the fractional length along the \(x\)-axis (the axis perpendicular to the current) occupied by the holes; namely, \(b/a\). Thus, from power conservation, the admittance \(Y'\) seen in the diffraction screen at the input of the hole can be obtained from

\[
P = \frac{1}{2} \frac{|I|^2}{Y'} = \frac{1}{2} \left(Y_{TE_{10}}\right)^{-1} \left|\frac{b}{a}\right|^2 I^2 \tag{13}
\]

which gives \(Y' = (a/b)^2 Y_{TE_{10}}\). The circuit element providing this admittance transformation is an ideal transformer with turns ratio \(n = a/b\). Therefore, this ideal transformer must be included between the resonant tank modeling the step discontinuity and the transmission line modeling the hole. The resulting equivalent circuit is shown in Fig. 4.

Fig. 4. Equivalent circuit for the structure of Figs. 1(a)-(b) with finite thickness \((t \neq 0)\).

In Fig. 5 the computed results for a screen of thickness \(t = a/7\) are shown and compared with
the results obtained for a zero-thickness screen. For comparison purposes the results obtained from numerical simulation using CST Microwave Studio are also shown. As it can be seen, there is a good qualitative agreement between theory and simulations. This agreement includes the presence of two transmission peaks, a well known effect in moderate thicknes screens (see [14] and references therein). Taking into account the logarithmic scale, the quantitative agreement between theory and simulations is also quite good (more than four digits in the frequency of resonance). The source of the small numerical disagreement could be attributed to the assumed approximate value of the transformer ratio.

In summary, an analytical solution for ET through thin diffraction screens has been presented. Our analysis, based on the equivalence with a waveguide discontinuity problem, shows that ET and Wood's anomaly can both be explained from the peculiar behavior of the evanescent TM_{02} mode excited at the holes. Since this behavior is imposed by the screen periodicity, the analysis shows that it is this periodicity, rather than the physical nature of the screen, which is on the grounds of ET. Our analysis is also in agreement with the circuit theory of ET recently proposed by some of the authors. The analysis can also be applied to elucidate the role played by surface waves (or spoof plasmons) in ET. It has been shown that a radiating surface wave can be excited at a frequency very close to ET frequency but not exactly at the same frequency. This result suggests that radiating surface waves can play a significant role in the transitory states at the onset and at the end of a monochromatic ET steady state. Finally, the analysis was extended to diffraction screens of finite thickness, thus showing that the present theory also applies to ET in thick screens.

Acknowledgments

This work has been supported by the Spanish Ministerio de Educación y Ciencia and European Union FEDER funds (projects TEC2007-65376, TEC2007-68013-C02-01, and CSD2008-00066), and by Junta de Andalucía (project TIC–253).