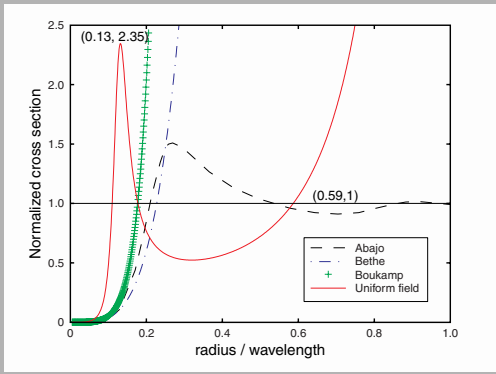
Abstract: Diffraction of normally incident light by a subwavelength circular aperture is calculated analytically. The aperture is opened on a perfectly conducting planar screen with infinitesimal thickness. In our model, the aperture is replaced by uniform magnetic currents and charges. The model allows one to obtain the normalized cross section for the aperture radius up to half of the wavelength, which exceeds the 0.2 wavelength limit of the Bethe-Bouwkamp's dipole model [1–3]. Also, in addition to reproducing the $(ka)^4$ dependence, which is characteristic of the dipole mode, our uniform field model explains the transmission enhancement obtained in Abajo's numerical simulation [4].



Normalized transmission cross section as a function of the normalized radius

Light diffraction by a subwavelength circular aperture

Che-Wei Chang, * *A.K. Sarychev, and V.M. Shalaev*

School of Electrical & Computer Engineering, Purdue University, West Lafayette, IN 47907, United States

Received: 25 January 2005, Accepted: 29 January 2005
Published online: 17 February 2005

Key words: apertures; diffraction theory; scanning microscopy; superresolution

PACS: 42.25.Fx, 42.50.St, 42.70.Ln

1. Introduction

The non-geometric behavior of light passing through an aperture has been studied since the fifteenth century. It was observed that the transmitted light diverges from the original path if the aperture shrinks to approximately ten wavelengths. This diffraction phenomenon becomes much more complex when the aperture decreases further to the subwavelength range. For a circular aperture on a planar conducting screen, the dipole model [1–3] predicts that the normalized cross section, which is defined as the ratio of the total transmitted power to the total incident power over the aperture area, is proportional to $(ka)^4$ (a is the radius of the aperture, and k is the wavenumber of the incident wave). However, recent numerical simulations [4] show that this model works only for radii smaller than 0.2 wavelengths.

Recently, the enhanced transmission was observed for a subwavelength aperture array [5], C-shaped [6], and H-shaped apertures [7]. The radiation intensity with focused pattern was observed for an aperture with a periodic pattern on the exit side [8]. Currently, there is no a single analytical model that could explain all these new phenomena. The dipole model assumes that the radius of the aperture is much smaller than the wavelength and thus has its limitations. In this paper, we assume that the magnetic current is uniform within the aperture. This simple model is capable of describing the normalized cross section for larger apertures.

2. Mathematical formulation

2.1. Problem definition

The free space is separated into two regions by a planar, perfectly conducting screen located at $z = 0$. A circular aperture with radius a is opened at the origin. The thickness of the screen is set to be infinitesimal. We focus our consideration on relation between the aperture radius and

* Corresponding author: e-mail: chang22@purdue.edu

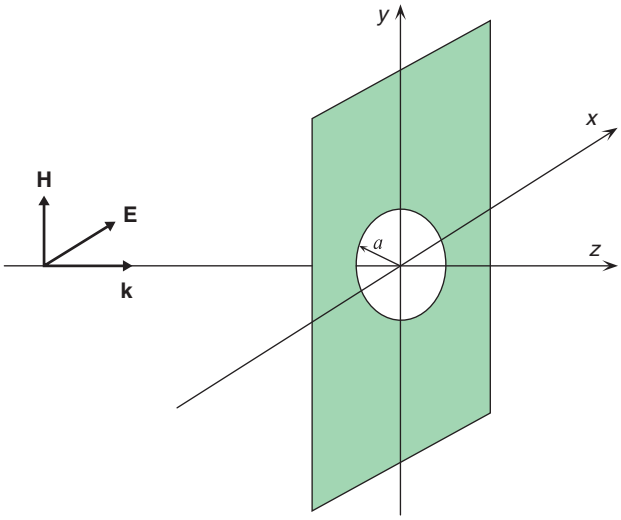


Figure 1 (online color at www.lphys.org) A circular aperture with radius a is centered on a perfectly conducting screen. A linearly polarized plane wave with wavelength λ is incident normally from the left-hand side of the screen

normalized cross section. A plane wave with wavelength λ is incident normally from the left-hand side of the screen. The electric field of the incident wave is linearly polarized in the x direction. The configuration is sketched in Fig. 1. In the following discussion, indices 1 and 2 refer to the left- and the right-hand sides, respectively.

2.2. Fields radiated by magnetic currents and charges

From the boundary conditions:

$$\mathbf{J}_s = \mathbf{n}_{12} \times (\mathbf{H}_2 - \mathbf{H}_1)\,, \tag{1}$$

$$\rho_s = D_{n1} - D_{n2}\,, \tag{2}$$

$\mathbf{H}_2$ and D_{n2} can be defined by electric currents and electric charges while the fields in region 1 are replaced by the null fields. This represents the equivalence principle [9–13]. However, the boundary conditions for the tangential electric field and the normal magnetic flux density do not include field sources. Therefore, we must add some fictitious terms, magnetic currents and magnetic charges, in Maxwell's equations to describe the field discontinuities. The modified versions of Maxwell's equations are as follows

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} - \mathbf{K}\,, \tag{3}$$

$$\nabla \times \mathbf{H} = \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} + \mathbf{J}\,, \tag{4}$$

$$\nabla \cdot \mathbf{D} = \rho_e\,, \tag{5}$$

$$\nabla \cdot \mathbf{B} = \rho_m\,, \tag{6}$$

$$\nabla \cdot \mathbf{J} = -\frac{1}{c}\frac{\partial \rho_e}{\partial t}\,, \tag{7}$$

$$\nabla \cdot \mathbf{K} = -\frac{1}{c}\frac{\partial \rho_m}{\partial t}\,, \tag{8}$$

$$\epsilon_0 = 1\,, \tag{9}$$

$$\mu_0 = 1\,. \tag{10}$$

Since $\mathbf{K}$ and ρ_m are only symbols representing the tangential electric field and the normal magnetic flux density, these additional terms do not change the fields in region 2 when we apply the equivalence principle. The reason for setting the fields in region 1 as null is that the aperture can be closed by a perfect electric conductor, without changing the fields in both sides. In such case, $\mathbf{J}$ is short-circuited, and only magnetic currents and magnetic charges contribute to the fields in region 2. The magnetic charges are distributed along the circumference of the aperture to terminate the magnetic currents. In the following steps, the filled screen is removed by applying the method of images, and thus the magnetic currents are doubled.

$$\mathbf{K}_s = -2(\mathbf{n} \times \mathbf{E})\,. \tag{11}$$

To calculate the fields radiated by $\mathbf{K}$ and ρ_m, it is useful to introduce the electric vector potential $\mathbf{A}_E$ in this electric-sources-free problem. The electric vector potential is defined as

$$\mathbf{E} = \nabla \times \mathbf{A}_E\,. \tag{12}$$

By substituting this definition in Eq. (4) and moving the curl operator out, we find the magnetic field as

$$\mathbf{H} = \frac{1}{c}\frac{\partial \mathbf{A}_E}{\partial t} - \nabla\phi_H\,. \tag{13}$$

The magnetic scalar potential ϕ_H is added because $\nabla \times \nabla\phi_H = 0$. Next, with Eq. (3) and the Lorentz gauge,

$$\nabla \cdot \mathbf{A}_E = \frac{1}{c}\frac{\partial \phi_H}{\partial t}\,, \tag{14}$$

we derive the wave equation in terms of $\mathbf{A}_E$:

$$\nabla^2 \mathbf{A}_E + k^2 \mathbf{A}_E = \mathbf{K}\,. \tag{15}$$

Similarly, the wave equation for ϕ_H can be derived from Eq. (6) and Eq. (13)

$$\nabla^2 \phi_H + k^2 \phi_H = -\rho_m\,. \tag{16}$$

The solutions to wave Eqs. (15) and (16) are given by the convolution of the source terms in the right-hand side with the Green's function, G:

$$\mathbf{A}_E(r) = -\frac{1}{4\pi}\int G(|r - r'|)\mathbf{K}'\mathrm{d}v'\,, \tag{17}$$

$$\phi_H(r) = \frac{1}{4\pi}\int G(|r - r'|)\rho'_m\mathrm{d}v'\,, \tag{18}$$

$$G(r) = \frac{\mathrm{e}^{ikr}}{4\pi r}\,. \tag{19}$$

Finally, by substituting Eqs. (11) and (8) into Eqs. (17) and (18), we find the fields in region 2 as

$$\mathbf{A}_E = \frac{1}{2\pi}\int G(|r-r'|)(\mathbf{n}\times\mathbf{E}')\,\mathrm{d}s'\,, \tag{20}$$

$$\phi_H = \frac{i}{2\pi k}\int G(|r-r'|)\nabla'\cdot(\mathbf{n}\times\mathbf{E}')\,\mathrm{d}s'\,. \tag{21}$$

2.3. Boundary conditions

The tangential electric field is the only unknown in Eqs. (20) and (21), and this field can be related to the incident field in region 1 via the boundary conditions. The total field in region 1 contains three parts: the incident field ($\mathbf{E}_i$, $\mathbf{H}_i$), the reflected field from the screen without the aperture ($\mathbf{E}_r$, $\mathbf{H}_r$), and the scattered field from the magnetic currents and magnetic charges ($\mathbf{E}_1$, $\mathbf{H}_1$). The total field in region 2 comes only from the magnetic currents and magnetic charges, $\mathbf{E}_2$ and $\mathbf{H}_2$. The boundary conditions require that

$$\mathbf{E}_{it} + \mathbf{E}_{rt} + \mathbf{E}_{1t} = \mathbf{E}_{2t}\,, \tag{22}$$

$$\mathbf{H}_{it} + \mathbf{H}_{rt} + \mathbf{H}_{1t} = \mathbf{H}_{2t}\,. \tag{23}$$

Now, by applying $\mathbf{E}_{it} = -\mathbf{E}_{rt}$ and $\mathbf{H}_{it} = \mathbf{H}_{rt}$, Eqs. (22) and (23) can be reduced to the following equations

$$\mathbf{E}_{1t} = \mathbf{E}_{2t}\,, \tag{24}$$

$$2\mathbf{H}_{it} + \mathbf{H}_{1t} = \mathbf{H}_{2t}\,. \tag{25}$$

Since the directions of the scattered fields in regions 1 and 2 are opposite, $\mathbf{k}_1 = -\mathbf{k}_2$, we obtain that

$$\mathbf{H}_{1t} = \mathbf{k}_1 \times \mathbf{E}_{1t}\,, \tag{26}$$

$$\mathbf{H}_{2t} = \mathbf{k}_2 \times \mathbf{E}_{2t} = -\mathbf{H}_{1t}\,. \tag{27}$$

By combining Eqs. (27) and (25), we find that

$$\mathbf{H}_{2t} = \mathbf{H}_{it}\,. \tag{28}$$

2.4. Problem solution

When we use for $\mathbf{A}_E$, ϕ_H in Eq. (13) formulas (20) and (21), the diffracted magnetic field becomes a function of the tangential electric field in the hole. Thus, $\mathbf{n}\times\mathbf{E}'$ can be solved with the boundary condition given by Eq. (28). The calculation can be simplified by computing the field at the origin. Then, the assumption that $\mathbf{K}$ is constant in the aperture is applied. Details for this calculation are given in the Appendix. The result is

$$\mathbf{E}' = \frac{\mathbf{H}_{it}}{1 + \frac{\exp(ika)}{2}(\frac{i}{ka} - 1)}\,\widehat{\mathbf{x}}\,. \tag{29}$$

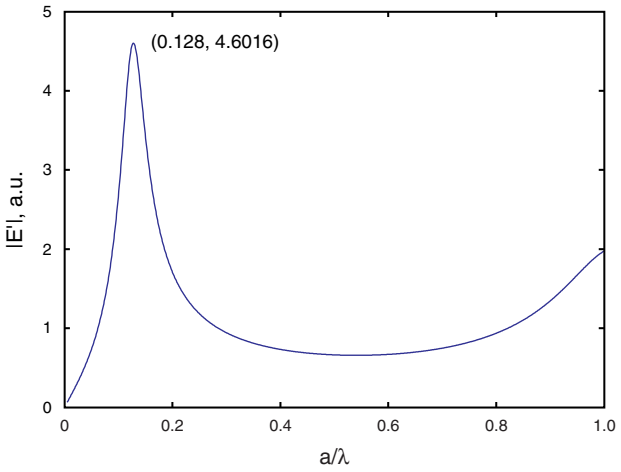


Figure 2 (online color at www.lphys.org) Electric field in the center of the aperture

The $\mathbf{E}'$ field is plotted in Fig. 2. For the fields in the far zone ($r \gg \lambda$), we have

$$\mathbf{E} = \frac{ika^2\exp(ika)}{2r}\hat{\mathbf{r}}\times[\mathbf{n}\times\mathbf{E}']\,, \tag{30}$$

$$\mathbf{H} = \frac{ika^2\exp(ika)}{2r}\hat{\mathbf{r}}\times\hat{\mathbf{r}}\times[\mathbf{n}\times\mathbf{E}']\,. \tag{31}$$

Hereafter, the subscripts for the fields, referring to the region of space, are omitted since only the fields in the right-hand side region will be discussed.

2.5. Normalized cross section

In order to compare our results with the Bethe formula, the normalized cross section is calculated. First, the Poynting vector is given by

$$\mathbf{S} = \frac{c}{8\pi}\mathrm{Re}(\mathbf{E}\times\mathbf{H}^*) = \tag{32}$$

$$= \frac{ck^2a^4}{32\pi r^2}|\mathbf{E}'|^2(\cos^2\theta + \sin^2\theta\cos^2\varphi)\hat{\mathbf{r}}\,.$$

Here θ denotes the angle between $\hat{\mathbf{z}}$ and $\hat{\mathbf{r}}$, and φ gives the angle between the plane of $\hat{\mathbf{z}}$ and $\hat{\mathbf{x}}$ and the plane of $\hat{\mathbf{z}}$ and $\hat{\mathbf{r}}$. The power density is not uniform, neither in θ direction nor in the φ direction. Thus, the total power through the aperture should be integrated over the semi-sphere.

$$S_{tot} = \int_0^{2\pi}\int_0^{\frac{\pi}{2}} |\mathbf{S}|r^2\sin\theta\mathrm{d}\theta\mathrm{d}\varphi = \frac{ck^2a^4}{24}|\mathbf{E}'|^2\,. \tag{33}$$

Finally, the normalized cross section is found as

$$\sigma = \frac{S_{tot}}{S_{inc}} = \tag{34}$$

$$= \frac{2}{3}(ka)^2 \frac{1}{5 - 4\cos ka - \frac{4\sin ka}{ka} + \frac{1}{(ka)^2}} \simeq$$

$$\simeq \frac{2}{3}(ka)^4[1 + 3(ka)^2 + \frac{19}{3}(ka)^4], \quad (35)$$

which is valid at $ka \ll 1$.

3. Discussion

In order to compare the uniform model with other models, the normalized cross section is shown as a function of the normalized radius, a/λ, in Fig. 3. Note that the screen thickness in the Abajo's boundary element method [4] is 0.1λ instead of zero as in other models.

It is clear that the uniform field model reflects properly the important characteristics suggested by the dipole model and the boundary element method. First, in the uniform field model, the normalized cross section for a small radius is proportional to $(ka)^4$ if the higher-order terms in Eq. (35) are neglected. This is consistent with the Bethe-Bouwkamp dipole model. The Bouwkamp's curve approaches the uniform field model if more correction terms are added. Only two correction terms are included in Bouwkamp's model in Fig. 3:

$$\sigma = \frac{64}{27\pi^2}(ka)^4 \times \quad (36)$$

$$\times[1 + \frac{22}{25}(ka)^2 + \frac{7312}{18375}(ka)^4 + \ldots].$$

Bouwkamp

Second, the uniform field model has one local maximum as the boundary element method does. As seen in Fig. 3, the local maximum occurs at 0.13λ; this corresponds to the maximum electric field in Fig. 2. Although in Fig. 3 the peak of the boundary element method is at 0.27λ, Abajo reported that the peak occurs at a smaller radius and the enhancement is further increased as the screen thickness approaches zero.

The normalized cross section drops down below 1 as the radius exceeds 0.2λ. Then, the normalized cross section is equal to 1 again at 0.6λ and then it goes to infinity. The Abajo calculation claims that the normalized cross section approaches 1 for radii larger than one wavelength. We suggest that the uniform field model provides a reasonable description for radii smaller than a half of the wavelength.

4. Conclusion

In conclusion, the normalized cross section for a circular aperture in a perfectly conducting planar screen is calculated analytically. A linearly polarized plane wave is incident normally onto the screen and the thickness of the screen is set to be zero. In our calculations, the magnetic current density in the aperture is assumed to be constant. This uniform field model describes properly the most important features of the two other models. First, it reduces to the $(ka)^4$ equation as Bethe-Bouwkamp's dipole model does for radii smaller than 0.13λ. Second, this analytic model demonstrates the transmission enhancement similar to Abajo's numerical result. In the future, the uniform field model will be applied to the subwavelength slit problem with TM polarization. A proper modification of the uniform field assumption could improve the result so that the normalized cross section approaches one when radius is larger than a wavelength. The model can be also improved by using the Taylor expansion for the electrical field in the center of the hole. Then, the integration of terms in the Taylor series (as was done here for the case of the constant field) would give a set of linear equations that fully solve this problem for a single hole. Note that for each step of the procedure, one would obtain analytical expressions for both the far fields and local fields.

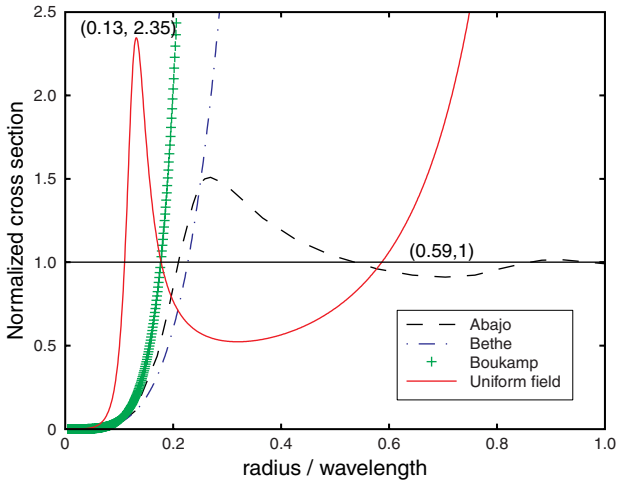


Figure 3 (online color at www.lphys.org) This figure shows the normalized transmission cross section as a function of the normalized radius. The red solid line is the uniform field model proposed in this paper. The blue dot-dash line is Bethe's dipole model. The green plus line is Boukamp's modified dipole model with two higher-order terms, Eq. (36). The black dash line is Abajo's boundary element method. In Abajo's simulation, the thickness of the screen is 0.1λ

5. Appendix

This section describes how to obtain $\mathbf{A}_E$ and ϕ_H at the origin. The calculation is simplified by computing the field at the origin. Then, the assumption that $\mathbf{K}$ is constant in the aperture is applied.

$$\mathbf{A}_E\Big|_{\mathbf{r}=0} = \left[\frac{1}{2\pi}\int G(|r - r'|)(\mathbf{n} \times \mathbf{E}')\mathrm{d}s'\right]\Big|_{\mathbf{r}=0} = \quad (37)$$

$$= \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty}\frac{\exp(ik\rho)}{\rho}(E'\hat{\mathbf{y}})\rho \mathrm{d}\rho\mathrm{d}\varphi = \tag{38}$$

$$= \frac{1}{ik}[\exp(ika)-1]E'\hat{\mathbf{y}}\,, \tag{39}$$

$$\phi_H = \frac{i}{2\pi k}\int G(|r-r'|)\nabla'\cdot(\mathbf{n}\times\mathbf{E}')\mathrm{d}s' = \tag{40}$$

$$= -\frac{i}{2\pi k}\int(\nabla' G(|r-r'|))\cdot(\mathbf{n}\times\mathbf{E}')\mathrm{d}s' = \tag{41}$$

$$= -\frac{i}{2\pi k}\oint\limits_{\rho'=a} G(|r-r'|)(\mathbf{n}\times\mathbf{E}')\cdot\hat{\mathbf{n}}\mathrm{d}\ell' = \tag{42}$$

$$= -\frac{i}{2\pi k}E'a\int_0^{2\pi} G(|r-r'|)\Big|_{\rho=a}\sin\varphi'\mathrm{d}\varphi'\,, \tag{43}$$

$$\nabla\phi_H = -\frac{i}{2\pi k}E'a\int_0^{2\pi}\sin\varphi'\times \tag{44}$$

$$\times\left[G(|r-r'|)\left(ik-\frac{1}{|r-r'|}\right)\frac{\mathbf{r}-\mathbf{r}'}{|r-r'|}\right]\Bigg|_{\rho'=a}\mathrm{d}\varphi'\,,$$

$$\nabla\phi_H\Big|_{\mathbf{r}=0} = \frac{i}{2\pi k}E'a\times \tag{45}$$

$$\times\int_0^{2\pi}\sin\varphi'\frac{\exp(ika)}{a}\left(ik-\frac{1}{a}\right)\frac{a\hat{\rho}}{a}\mathrm{d}\varphi' =$$

$$= \frac{i}{2k}E'\exp(ika)\left(ik-\frac{1}{a}\right)\hat{\mathbf{y}}\,. \tag{46}$$

References

[1] H.A. Bethe, Phys. Rev. **66**, 163 (1944).
[2] C.J. Bouwkamp, Philips Research Reports **5**, 321 (1950).
[3] C.J. Bouwkamp, Reports on Progress in Physics **XVIII**, 35 (1954).
[4] F.J. García de Abajo, Opt. Express **10**, 1475 (2002).
[5] T.W. Ebbesen, H.J. Lezec, H.F. Ghaemi, et al., Nature **391**, 667 (1998).
[6] X. Shi and L. Hesselink, Jpn. J. Appl. Phys. **41**, 1632 (2002).
[7] E.X. Jin and X. Xu, in: Proceedings of IMECE'03 (Washington, D.C., 2003), pp. 1-6.
[8] T. Thio, K.M. Pellerin, and R.A. Linke, Opt. Lett. **26**, 1972 (2001).
[9] J.A. Stratton, Electormagnetic Theory (McGraw-Hill, New York, 1941).
[10] R.F. Harrington, Time-Harmonic Electromagnetic Fields (McGraw-Hill, New York, 1961).
[11] R.F. Harrington, Introduction to Electromagnetic Engineering (McGraw-Hill, New York, 1958).
[12] S. Ramo, Fields and Waves in Modern Radio (Wiley, New York, 1953).
[13] R.E. Collin, Antennas and Radiowave Propagation (McGraw-Hill, New York, 1985).