Determination of three-dimensional structure in photon scanning tunnelling microscopy

P Scott Carnev¹ and John C Schotland²

- ¹ Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61821, USA
- ² Departments of Electrical Engineering and Radiology, Washington University, St. Louis, MO 63130, USA

Received 30 November 2001, in final form 25 April 2002 Published 14 August 2002 Online at stacks.iop.org/JOptA/4/S140

Abstract

We present an analytic solution to the problem of determining three-dimensional object structure, described by the spatial dependence of the susceptibility, from data accessible through photon scanning tunnelling microscopy (PSTM) experiments. An analysis is presented of the scattering of evanescent waves with detection of the scattered field by a probe in the near zone of the scatterer. The results provide tomographic imaging capability in PSTM.

Keywords: Tomography, microscopy, near-field optics, tunnelling, inverse scattering

1. Introduction

Inverse scattering, tomography and structure determination have, over the years, attracted a great deal of interest in the electromagnetic and optics research communities. The reasons for this interest are largely practical and driven by the need to gain a deeper understanding of scattering, transmission and absorption experiments. For instance, von Laue's insights into the scattering of x-rays by crystals made possible the determination of crystal structure, and hence modern crystallography a reality [1]. The early measurements of x-ray absorption in the human body had obvious promise for imaging, but it was not until the work of Hounsfeld and Cormack brought medical imaging out of the era of projection radiography and into the era of computed tomography (CT) that the promise was fully realized and the measurements understood [2]. Likewise, in fewer dimensions, the analysis and design of power transmission lines was greatly improved by the development of transmission line tomography whereby transmitted signals may be used to reconstruct the structure of the line [3].

Near-field optical microscopy has developed dramatically in recent years [4–6]. The first proposal of a method to circumvent the Rayleigh–Abbe resolution limit was put forward by Synge [8] in 1928. Synge proposed that a thin

sample be illuminated through a subwavelength aperture. By recording the transmitted light as a function of aperture position, a subwavelength resolved image of the sample may be acquired. Today this method is known as near-field scanning optical microscopy (NSOM) [9–13] or scanning near-field optical microscopy (SNOM); it is practised in many variations, including the reciprocal arrangement in which the sample is illuminated by a source in the far zone of the sample and light is collected through a small aperture. The role of the small aperture is now played by a tapered optical fibre, a technique not known to Synge.

Essential to the NSOM modality is the presence of inhomogeneous, or evanescent, modes of the illumination field. Specifically, the illuminating field consists of a superposition of plane waves including the high spatial-frequency evanescent plane waves. These waves are super-oscillatory parallel to some reference plane and are exponentially decaying away from the plane. The super-resolving capabilities of NSOM may be attributed to the high spatial-frequency of the evanescent waves. Instead of generating these modes at the small aperture in NSOM, they may be generated at the interface of two media by total internal reflection, as is done in total internal reflection microscopy (TIRM). Because the evanescent waves are exponentially decaying, TIRM is used to limit the depth of penetration

of the illuminating field. However, the great potential of super-resolved imaging made possible by the high spatial-frequencies in the probe field has remained untapped in TIRM [14].

At the intersection of NSOM and TIRM modalities are the so-called photon scanning tunnelling microscopy (PSTM) methods. In these techniques the object is illuminated by an evanescent wave generated at the face of a prism or slide (as in TIRM), and the scattered field is detected via a tapered fibre probe in the near zone of the sample (as in NSOM). Also practised is the reciprocal arrangement in which the source and detector are interchanged.

In all of the above mentioned modalities, the connection between the measured field or signal and the sample properties has proven to be problematic. To clarify the meaning of the measurements and to provide a three-dimensional imaging capability, it is desirable to find a solution to the inverse scattering problem (ISP). Results in this direction have been reported for the case of surface profile reconstruction in homogeneous media [15–18]. There has also been recent progress for the case of three-dimensional inhomogeneous media for the TIRM and NSOM modalities [14, 19]. In those works, an analytic solution to the ISP was presented for each modality and the results were illustrated by numerical simulations. It is thus natural to consider the analogous question for PSTM, which is the subject of this paper.

The remainder of this paper is organized as follows. In section 2 we present an analysis of the forward problem. That is, we obtain an expression for the electric field scattered from an unknown sample in a PSTM experiment. We include the effects of the air–dielectric boundary at the face of the prism as well as all relevant polarization effects by making use of the appropriate Green's tensor for electromagnetic scattering in a half-space. In section 3 we derive a solution to the inverse scattering problem whereby we determine the spatial dependence of the susceptibility of the sample from measurements of the scattered field. We account for sampling of the scattered field as well as finite and discrete sets of illuminating fields. Finally, we discuss regularization schemes for the inversion formula thus obtained.

2. Forward problem

We begin by considering an experiment in which a monochromatic field is incident on a dielectric medium with susceptibility $\eta(r)$. The field incident on the sample will be taken to be a plane wave which may be evanescent or homogeneous. One half-space, taken to be $z \ge 0$, will have the vacuum index of refraction while the z < 0 half-space will have an index of refraction n (see figure 1). We will consider nonmagnetic materials and so will limit our attention to the electric field E. The field satisfies the equation

$$\nabla \times \nabla \times E(r) - n^2(z)k_0^2 E(r) = 4\pi k_0^2 \eta(r) E(r) \quad (1)$$

where k_0 is the free space wavenumber and n(z) is the z-dependent background index of refraction as described above. We will consider the field to consist of two parts:

$$E = E^i + E^s. (2)$$

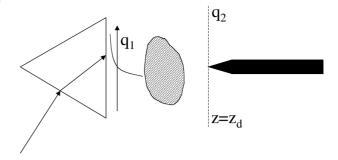


Figure 1. Illustrating the measurement scenario. Evanescent waves are generated at the prism face by total internal reflection (TIR). The TIR is then partly frustrated by the presence of the scatterer and collected by a probe in the near zone.

The incident field E^i obeys the free space wave equation

$$\nabla \times \nabla \times \mathbf{E}^{i}(\mathbf{r}) - n^{2}(z)k_{0}^{2}\mathbf{E}^{i}(\mathbf{r}) = 0.$$
 (3)

The scattered field E^s obeys

$$\nabla \times \nabla \times \mathbf{E}^{s}(\mathbf{r}) - n^{2}(z)k_{0}^{2}\mathbf{E}^{s}(\mathbf{r}) = 4\pi k_{0}^{2}\eta(\mathbf{r})\mathbf{E}(\mathbf{r}).$$
 (4)

Equation (4) may be recast as the integral equation

$$E_{\alpha}^{s}(\mathbf{r}) = k_0^2 \int d^3 r' G_{\alpha\beta}(\mathbf{r}, \mathbf{r}') E_{\beta}(\mathbf{r}') \eta(\mathbf{r}')$$
 (5)

where the summation convention over repeated indices applies and will throughout. The Green's tensor $G_{\alpha\beta}(r,r')$ satisfies the equation

$$\nabla \times \nabla \times G(\mathbf{r}, \mathbf{r}') - n^2(z)k_0^2G(\mathbf{r}, \mathbf{r}') = 4\pi\delta(\mathbf{r} - \mathbf{r}')\mathbf{I}$$
 (6)

where I is the unit tensor. The Green's tensor must also satisfy the boundary conditions

$$\hat{z} \times G(r, r')|_{z=0^{+}} = \hat{z} \times G(r, r')|_{z=0^{-}}$$
 (7)

$$\hat{z} \times \nabla \times G(r, r')|_{z=0^+} = \hat{z} \times \nabla \times G(r, r')|_{z=0^-}.$$
 (8)

Making use of a plane wave decomposition, it may be found that [6, 20]

$$G_{\alpha\beta}(\mathbf{r}, \mathbf{r}') = \frac{\mathrm{i}}{2\pi} \int \frac{\mathrm{d}^2 q}{k_z(\mathbf{q})} S_{\alpha\gamma}^{-1}(\mathbf{q}) g_{\gamma\delta}(\mathbf{q}, z') S_{\delta\beta}(\mathbf{q})$$

$$\times \exp[\mathrm{i}k(\mathbf{q}) \cdot (\mathbf{r} - \mathbf{r}')] \tag{9}$$

where $k(q)=(q,k_z(q))$, and $k_z(q)=\sqrt{k_0^2-q^2}$. The plane wave modes appearing in equation (9) are labelled by the transverse part of the wavevector, q. The modes for which $|q|< k_0$ are homogeneous, or propagating. When $|q|>k_0$ the plane wave is evanescent, decaying exponentially with increasing values of z. These waves are super-oscillatory in the transverse plane and thus provide a means to probe the high spatial-frequency structure of the sample. Expressions for the matrices S(q) and g(q,z) are given in the appendix.

We take the incident field to be an evanescent plane wave with polarization e:

$$E_{\alpha}^{i}(\mathbf{r}) = e_{\alpha} \exp[i\mathbf{k}(\mathbf{q}) \cdot \mathbf{r}]. \tag{10}$$

Note that if the evanescent wave is generated by a prism with index of refraction n then $k_0 \le |\mathbf{q}| \le nk_0$. Within the accuracy

of the first Born approximation, we find that the scattered field is given by the expression

$$E_{\alpha}^{s}(\mathbf{r}) = k_0^2 \int d^3r' G_{\alpha\beta}(\mathbf{r}, \mathbf{r}') e_{\beta} \exp[i\mathbf{k}(\mathbf{q}) \cdot \mathbf{r}'] \eta(\mathbf{r}'). \quad (11)$$

We will consider the inversion of this integral equation to obtain η in section 3.

3. Inverse problem

In this section we consider the inverse scattering problem (ISP) of reconstructing the susceptibility function $\eta(r)$ from measurements of the scattered field. We assume that the sample is illuminated by an incident evanescent wave, although incident homogeneous waves may also be employed. The scattered field is detected in the near zone by means of an idealized point detector. Note that the PSTM measurement is intrinsically holographic, allowing the extraction of the complex amplitude of the scattered field from the measured signal [6, 7]. The detected field on the plane $z=z_d$ is assumed to be sampled on a square lattice with lattice spacing a.

It will prove useful to define a data function $\Phi_{\alpha}(q_1, q_2)$ as the two-dimensional Fourier transform of the scattered field at the point $r = (\rho, z_d)$. Explicitly,

$$\Phi_{\alpha}(q_1, q_2) = \sum_{\rho} e^{iq_2 \cdot \rho} E_{\alpha}^s(\rho, z_d; q_1)$$
 (12)

where the dependence on the transverse part q_1 of the incident wave has been indicated, the sum over ρ is carried out over all lattice vectors and q_2 belongs to the first Brillouin zone (FBZ) of the lattice. Making use of equations (9) and (11) and the identity

$$\sum_{\rho} \exp[i\mathbf{q} \cdot \rho] = \left(\frac{2\pi}{a}\right)^2 \sum_{p} \delta(\mathbf{q} - p) \tag{13}$$

we find that

$$\Phi_{\alpha}(\mathbf{q}_1, \mathbf{q}_2) = \int d^3r \sum_{\mathbf{p}} w_{\alpha}(\mathbf{p} - \mathbf{q}_2, z) \exp\{i(\mathbf{q}_1 + \mathbf{q}_2 - \mathbf{p}) \cdot \boldsymbol{\rho} + i[k_z(\mathbf{q}_1) - k_z(\mathbf{p} - \mathbf{q}_2)]z\}\eta(\mathbf{r}).$$
(14)

Here

$$w_{\alpha}(q,z) = \frac{i2\pi k_0^2 e^{ik_z(q)z_d}}{k_z(q)a^2} S_{\alpha\gamma}^{-1}(q) g_{\gamma\delta}(q,z) S_{\delta\beta}(q) e_{\beta}$$
 (15)

and p denotes a reciprocal lattice vector.

We assume that $\Phi_{\alpha}(q_1, q_2)$ is known for (q_1, q_2) in the data set $\mathcal Q$ and introduce a function $\chi(q_1, q_2)$ which is unity if $(q_1, q_2) \in \mathcal Q$ and is zero otherwise. Then we define a new function

$$\Psi_{\alpha}(q,Q) = \Phi_{\alpha}(Q - q, q)\chi(Q - q, q) \tag{16}$$

where q, Q range over all space. Making use of these definitions we arrive at the system of equations

$$\Psi_{\alpha}(q, \mathbf{Q}) = \int_{0}^{L} dz \sum_{p} K_{\alpha}(p - q, z; \mathbf{Q}) \tilde{\eta}(\mathbf{Q} - p, z) \quad (17)$$

where

$$K_{\alpha}(q, z; \mathbf{Q}) = w_{\alpha}(p - q, z) \exp\{i[k_{z}(\mathbf{Q} + q - p) - k_{z}(p - q)]z\}\chi(\mathbf{Q} + q - p, q)$$

$$\tilde{\eta}(\mathbf{Q}, z) = \int d^{2}\rho \, e^{i\mathbf{Q}\cdot\rho}\eta(\mathbf{r})$$
(18)

and L denotes the range of $\eta(r)$ in the z direction. For fixed Q, equation (17) defines a one-dimensional integral equation for $\tilde{\eta}(Q, z)$. Following the general method of [21] we find that the pseudoinverse solution of (17) is given by

$$\tilde{\eta}(Q - p, z) = \int d^2q \, d^2q' \, K_{\alpha}^*(p - q, z; Q)$$

$$\times M_{\alpha\beta}^{-1}(q, q'; Q) \Psi_{\beta}(q', Q)$$
(19)

where $M_{\alpha\beta}^{-1}(q, q'; Q)$ is obtained from the overlap integral

$$M_{\alpha\beta}(\boldsymbol{q}, \boldsymbol{q}'; \boldsymbol{Q}) = \int_0^L dz \sum_{\boldsymbol{p}} K_{\alpha}(\boldsymbol{p} - \boldsymbol{q}, z; \boldsymbol{Q}) K_{\beta}^*(\boldsymbol{p} - \boldsymbol{q}', z; \boldsymbol{Q}).$$
(20)

Equation (19) specifies all possible transverse frequency components of $\eta(r)$. Thus we may apply the inverse Fourier transform to obtain

$$\eta(\mathbf{r}) = \int \frac{\mathrm{d}^2 Q}{(2\pi)^2} \int \mathrm{d}^2 q \, \mathrm{d}^2 q' \sum_{\mathbf{p}} \mathrm{e}^{-\mathrm{i}(Q-\mathbf{p}) \cdot \rho} K_{\alpha}^*(\mathbf{p} - \mathbf{q}, z; \mathbf{Q})$$
$$\times M_{\alpha\beta}^{-1}(\mathbf{q}, \mathbf{q}'; \mathbf{Q}) \Phi_{\beta}(\mathbf{Q} - \mathbf{q}', \mathbf{q}') \tag{21}$$

which is the required inversion formula for the ISP.

Several comments on the inversion formula (21) are necessary. First, we restrict the integrations over q, q' to the FBZ. Second, if $\eta(r)$ is transversely band limited, then the sum over p may be truncated. Third, if the transverse coordinate ρ is taken to be a lattice vector, then the term $\exp(ip \cdot \rho)$ in the summation and in (18) is equal to unity. Finally, the solution we have constructed to the ISP is the unique minimum L^2 norm solution of (14) [2].

In order to avoid numerical instability and set the resolution of the reconstructed image to be commensurate with the available data, $M_{\alpha\beta}^{-1}(q,q';Q)$ must be regularized. In particular, we set

$$M_{\alpha\beta}^{-1}(q, q'; Q) = \int d^2 Q' R(\sigma_{QQ'}) f_{Q'}^{\alpha}(q; Q) f_{Q'}^{\beta*}(q'; Q)$$
(22)

where $f_{Q'}^{\alpha}(q;Q)$ and $\sigma_{QQ'}^2$ are the eigenfunctions and eigenvalues of M and R is a suitable regularizer. Many of the commonly used regularization schemes result in a simple filtering of the eigenvalues of M. The effect of such regularization is to limit the contribution of the small singular values to the reconstruction. One way to do this is to simply cut off all σ below some cutoff σ_c . That is, we set

$$R(\sigma) = \frac{1}{\sigma^2} \theta(\sigma - \sigma_c)$$
 (23)

with θ denoting the usual Heaviside step function. Alternatively, the Tikhonov method may be employed.

In figure 2 we demonstrate an application of the method described above. We consider two point scatterers on the prism face separated by 0.3λ . We simulate the scattered field and compute the intensity in the measurement plane for three scan heights, z_d , of the probe. We make use of a limited data set.

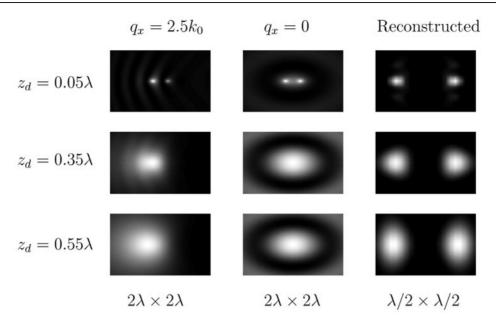


Figure 2. Demonstrating the observable intensities and reconstructed images for two point scatterers separated by 0.3λ with the measurement plane at various distances from the prism face. The left column contains simulated intensity in the measurement plane $z=z_d$ for an illuminating field with transverse wavevector q_x set to the maximum value attainable in a prism with index n=2.5. The middle column shows the simulated intensity with the illuminating wave incident normal to the plane of observation. The right column displays the image reconstructed from multiple views obtained with different illuminating fields. Note that the images of the intensity are shown with a $2\lambda \times 2\lambda$ field of view while the reconstructed image is shown with a $\lambda/2 \times \lambda/2$ field of view.

We use only the TE polarization, i.e. the polarization vector parallel to the prism face. The scattered field is computed on a $4\lambda \times 4\lambda$ window and sampled on a cartesian grid at a spacing of $\lambda/10$ with a total of 41 illuminating plane waves, all with $q_y = 0$ and q_x on equally spaced points in the range $[-2.5k_0, 2.5k_0]$, corresponding to a range attainable with a prism of index n = 2.5. We place on the object a band limit commensurate with the sampling of the field, that is we retain only the p = 0 term of the sum in equation (21). The computed fields are shown in figure 2 in a $2\lambda \times 2\lambda$ field of view while the reconstructed scatterer is shown in a $\lambda/2 \times \lambda/2$ field of view.

The measurement may be seen to be intrinsically holographic, containing all the necessary phase information. That is, the incident field and the scattered field in combination produce an interference pattern in the measurement plane from which the necessary phase information may be observed. The scan made furthest from the prism face with an illuminating evanescent wave may be seen to be completely dominated by the scattered wave due to the exponential decay of the illuminating field. Thus when the illuminating field is highly evanescent and the measurements are made far from the prism face it may be necessary to also make use of a separate reference wave to measure the scattered field. It may be observed that the object structure, which has become unclear in the direct measurements made further from the sample, is still clearly evident in the reconstructions. The increased spread of the points may be attributed to the loss of high spatial frequency components of the scattered field to exponential decay away from the scatterers. It is in principle possible to reconstruct the scatterers with the same resolution at any distance from the measurement plane. However, any regularization imposed on the algorithm will produce a depth-dependent resolution limit due to the exponential decay of the high spatial-frequency

components of the scattered field. With the measurement plane very close to the prism face the resolution becomes comparable to $(nk_0/2\pi+1/\Delta)^{-1}$ where Δ is the spatial sampling interval. With the measurement plane far from the prism, the experiment becomes that described in [14] and the resolution approaches $[(1+n)k_0/2\pi]^{-1}$.

4. Conclusions

We have examined the forward and inverse scattering problems relevant to PSTM. These results enable determination of the three-dimensional structure of the sample under investigation. Crucial to implementation, we have included the effects of polarization, an interface in the near zone of the sample, and have allowed for reconstruction from a sampled or incomplete data set. Future research will focus on nonlinear corrections to the inverse problem and more realistic detector tip models.

Appendix

The Green's tensor in the half-space geometry is given by the expression

$$G_{\alpha\beta}(r,r') = \frac{\mathrm{i}}{2\pi} \int \frac{\mathrm{d}^2 q}{k_z(q)} S_{\alpha\gamma}^{-1}(q) g_{\gamma\delta}(q) S_{\delta\beta}(q)$$

$$\times \exp[\mathrm{i}k(q) \cdot (r - r')] \tag{24}$$

where $k(q) = (q, k_z(q))$ and $k_z(q) = \sqrt{k_0^2 - q^2}$. Additionally, S(q) is the matrix which rotates k(q) into the xz plane, or more explicitly

$$S(q) = |q|^{-1} \begin{pmatrix} q_x & q_y & 0 \\ -q_y & q_x & 0 \\ 0 & 0 & |q| \end{pmatrix}$$
 (25)

and

$$g_{xx} = \left(\frac{k_z(q)}{k_0}\right)^2 \{1 + R_2(q) \exp[2ik_z(q)z']\}$$
 (26)

$$g_{yy} = 1 + R_1(q) \exp[2ik_z(q)z']$$
 (27)

$$g_{zz} = \left(\frac{|q|}{k_0}\right)^2 \{1 - R_2(q) \exp[2ik_z(q)z']\}$$
 (28)

$$g_{zx} = \frac{-|q|k_z(q)}{k_0^2} \{1 + R_2(q) \exp[2ik_z(q)z']\}$$
 (29)

$$g_{xz} = \frac{-|q|k_z(q)}{k_0^2} \{1 - R_2(q) \exp[2ik_z(q)z']\}$$
 (30)

all other elements of g being zero. Here $R_1(q)$ and $R_2(q)$ are the reflection coefficients given by

$$R_1(q) = \frac{k_z(q) - k_z'(q)}{k_z(q) + k_z'(q)}$$
(31)

and

$$R_2(q) = \frac{k_z'(q) - nk_z(q)}{k_z'(q) + nk_z(q)}$$
(32)

with
$$k'_z(q) = \sqrt{n^2 k_0^2 - q^2}$$
.

Acknowledgments

We would like to thank the organizers of the European Optical Society Topical Meeting: Electromagnetic Optics 2 for a lively and stimulating meeting. We would also like to thank our wives

who graciously excused us from family duties during most of the US Thanksgiving holiday while we completed this paper.

References

- [1] Friedrich W, Knippling P and von Laue M 1912 Sitb. Math.—Phys. Bayr. Akad. Wiss. 1912 303–22
- [2] Natterer F 1986 *The Mathematics of Computerized Tomography* (Chichester: Wiley)
- [3] Bruckstein A M and Kailath T 1987 SIAM Rev. 29 359
- [4] Courjon D and Bainier C 1994 Near field microscopy and near field optics Rep. Prog. Phys. 57 989
- [5] Girard C and Dereux A 1996 Rep. Prog. Phys. **59** 657
- [6] Greffet J-J and Carminati R 1997 Prog. Surf. Sci. **56** 133
- [7] Porto J A, Carminati R and Greffet J-J 2000 J. Appl. Phys. 88
- [8] Synge E 1928 Phil. Mag. 6 356
- [9] Ash E and Nicholls G 1972 Nature 237 510
- [10] Lewis A, Isaacson M, Harootunian A and Muray A 1984 Ultramicroscopy 13 227
- [11] Pohl D W, Denk W and Lanz M 1984 Appl. Phys. Lett. 44 651
- [12] Betzig E and Trautman J K 1992 Science 257 189
- [13] Dickson R, Norris D, Tzeng Y-L and Moerner W 1996 Science 274 966
- [14] Carney P S and Schotland J C 2001 Opt. Lett. 26 1072
- [15] Greffet J-J, Sentenac A and Carminati R 1995 Opt. Commun.
- [16] Garcia N and Nieto-Vesperinas M 1993 Opt. Lett. 18 2090
- [17] Garcia N and Nieto-Vesperinas M 1995 Opt. Lett. 20 949
- [18] Carminati R, Greffet J-J, Garcia N and Nieto-Vesperinas M 1996 Opt. Lett. 21 501
- [19] Carney P S and Schotland J C 2000 Appl. Phys. Lett. 77 2798
- [20] Maradudin A and Mills D 1975 Phys. Rev. B 11 1392
- [21] Carney P S, Markel V A and Schotland J C 2001 Phys. Rev. Lett. 86 5874