f=u
99-HE

IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING

A publication of the IEEE Computer Society

MARCH / APRIL 1999 VOLUME 11 NUMBER 2 ITKEEH (ISSN 1041-4347)

REGULAR PAPERS

CORRESPONDENCE

Dr. Naphtali Rishe
Florida International University
School of Computer Science
Southwest 8th St. and 107th Avenue
University Park
Miami, FL 33199

IEEE COMPUTER SOCIETY

http://computer.org • tkde@computer.org

February 1998 VOLS. 313–314 ISSN 0040-6090

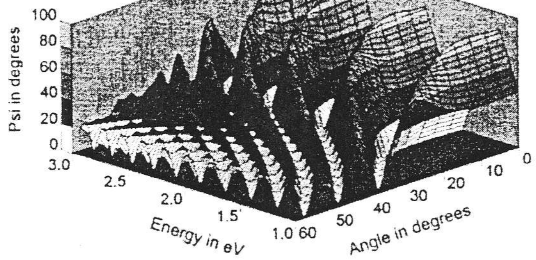
v.313-314

6003

f = papers
98-TP

thin solid films

International Journal on the Science and Technology of Condensed Matter Films



Spectroscopic Ellipsometry

Guest Editors:
R.W. Collins
D.E. Aspnes
E.A. Irene

ELSEVIER

http://www.elsevier.nl/locate/tsf

Editor-in-chief:
J. E. Greene

ELSEVIER

Thin Solid Films 313–314 (1998) 119–123

thin solid films

Toward a priori selection of ellipsometry angles and wavelengths using a high performance semantic database

F.K. Urban III*, David Barton

Department of Electrical and Computer Engineering, ECS 347, Florida International University, Miami, FL 33199, USA

Abstract

Modern ellipsometers only barely resemble those of only a decade ago. In situ measurements are commonly spectroscopic and ex situ investigations frequently add an additional variable, incidence angle. In either case, the number of measurements for an analysis can be 10 000 or more. Experience in a recent international round robin experiment showed no particular agreement in the selection of angles and wavelengths. Selecting a fewer number of 'good' angles of incidence and wavelengths is important because at the same time it would reduce the measurement time, reduce computational load and improve the solutions. Currently we choose to say 'good' points have high 'resolution' (a change in the desired parameter results in a measurable change in measured parameters) and a solvable 'condition' (the inverse Hessian matrix is well conditioned for variably damped least squares). In this work, values of resolution and condition are simulated for approximately 1000 points in wavelength–angle space for a large number of sample configurations and materials (over 10 000). These results are stored in a high performance semantic database which can be more than a Gigabyte. Three-dimensional images of the data and movies over thickness, for example, can be pulled from the database in real time allowing selection of incidence angles and wavelengths for an analysis. © 1998 Published by Elsevier Science S.A.

Keywords: Ellipsometry angles; Wavelengths; Resolution; Condition; High performance semantic database

1. Introduction

In the last decade the number of measurements taken for investigation of a single 'spot' on a reflecting surface has increased by many orders of magnitude. And as spatial resolution and temporal resolution continue to improve, the number of measurements per investigation will increase many orders of magnitude more. The increased data are leading to improved understanding of reflecting surfaces and of the ellipsometry technique itself. For example, it is now being aggressively commercialized as a real-time in situ process control sensor for thin film deposition and etching. The increased data also help to distinguish between complicated candidate reflecting surface models (including multiple layers, surface and interface roughness and media anisotropy).

The fact that many problems of ellipsometry can be solved with just a few measurements then raises the question of how much data is enough. Although the large data sets have had an important role in advancing the technique, perhaps they can be 'too big' [1]. An oversized data set is a burden both during measurement and for numerical analysis. Besides taking longer to acquire and compute, the over-large data set raises another, more subtle danger; not all data contribute in a positive way to the solution.

Resolution is defined in the usual way as the power of the technique to sense changes in the parameters of interest. Condition is defined as the mathematical

---

* Corresponding author.

PII S0040-6090(97)00783-9

condition of the inverse Hessian matrix (second derivative matrix) used in the variably damped least squares (VDLS or Levenberg–Marquardt) solution method. This matrix is square with the number of rows and columns equal to the number of variables to be solved. If the matrix is singular, the condition is infinite and no solution is possible (non-independent equations). It is also very important to check to see how close the matrix is to being singular. If the matrix is close to being singular, the condition is not infinite but is very large compared to measurement error and computer truncation error. Such a system of equations is ill-posed and, although it can be solved, meaningful solutions do not result (for example 2 ± 4). Inclusion of poorly conditioned systems of equations (i.e. a large condition) cannot help solution speed and accuracy.

Previous work considered only the condition of the simplest system of equations, film thickness and index real part, computed at each angle and wavelength [1]. The work here considers the more complicated situation of parameterizing the optical properties and solving for the parameters which describe the properties as a function of light wavelength.

2. Experimental

Samples were prepared at Pacific Northwest National Laboratory in a round robin experiment [2]. Raw data collected in the measurements were in good agreement although there were a wide range of different measurement angles and wavelengths selected by each group. However, the modeling approaches were different as each group selected their own favorite model. Consequently the reported film thickness and optical properties were only in reasonable agreement.

Computational experiments were run on an IBM Personal Computer using Matlab for its ease of visualization [3]. All computations were made for angle–wavelength space over 26 wavelengths (340, 360,... 840 nm) and 38 incidence angles (10, 12.... 84°) for a total of 988 points. Resolution plots were approximated numerically in two ways for comparison. Both used the relation $dy/dx \cong [y(x_1) - y(x_1 + \Delta x)]/(-\Delta x)$ in which y is either Ψ or Δ and x is the reflecting surface variable, such as thickness or index of refraction. In the first method Δx was 1 nm for thickness and 0.01 for the index of refraction and extinction coefficient. In the second method Δx was set to 2% of the value of each variable. Both allowed visualization of the regions in angle–wavelength space for which the resolution and condition were best, the main difference being the avoidance of division by small numbers in the first method. Pseudocolor plots were generated and movies over film thickness, from 0 to loss of information or 500 nm, whichever came first. The movies can be stopped at each thickness for examination. They cannot be reproduced here but are available elsewhere[1].

3. Theory

In the round robin experiment the model fitting typically was accomplished by the common variably damped least squares method (Levenberg–Marquardt). Briefly this solving algorithm finds parameters, such as film thickness (d) and optical properties (n and k) which minimize the mean square error (M.S.E.) between measured and computed data, such as ellipsometric, reflectance and transmittance values shown below:

$$\chi^2 = \frac{1}{N-M-1} \sum_i \left[\frac{(\Psi_{mi} - \Psi_{ci})}{\sigma_{\Psi i}} \right]^2 + \left[\frac{(\Delta_{mi} - \Delta_{ci})}{\sigma_{\Delta i}} \right]^2 \tag{1}$$

The summations are over all measured data from 1 to i. The value of N is the total number of ellipsometer measurements and M is the total number of ellipsometry fitting parameters. A more detailed discussion of these methods and determination of the attendant errors can be found in the literature [4].

The usual solution process is to first consider that the function to be fitted can be approximated by a quadratic form:

$$\chi^2(\mathbf{a}) \approx \gamma - \mathbf{d} \cdot \mathbf{a} + \tfrac{1}{2} \mathbf{a} \cdot \mathbf{D} \cdot \mathbf{a} \tag{2}$$

where **a** is the vector of unknown reflecting surface parameters, **d** is an M-vector and **D** is an $M \times M$ matrix (the second derivative matrix or Hessian matrix). Close to the solution we can find adjustments to the solution vector, **a**, which will result in one jump to the solution (well-approximated region)

$$\mathbf{a}_{\mathrm{min}} = \mathbf{a}_{\mathrm{cur}} + \mathbf{D}^{-1} \cdot \left[-\nabla \chi^2(\mathbf{a}_{\mathrm{cur}}) \right] \tag{3}$$

If the estimate, $\mathbf{a}_{\mathrm{cur}}$, is far from the solution (χ^2 is large) then the approximation is not as good and Eq. (3) will not go to the solution immediately but instead the curve must be followed by taking a reduced step in the direction of decreasing χ^2 as follows:

---

[1] To be maintained as long as possible at the author's web site currently at URL http://www.eng.fiu.edu/ECE/, select Faculty and Frank Urban. Instructions for viewing and access available at the site.

$$\mathbf{a}_{\text{next}} = \mathbf{a}_{\text{cur}} - \text{constant} \times -\nabla\chi^2(\mathbf{a}_{\text{cur}}) \tag{4}$$

In all cases, near the solutions it is necessary to solve Eq. (3) which means the condition of the Hessian matrix, $\mathbf{D}$ (same as the condition of its inverse $\mathbf{D}^{-1}$), must be 'sufficiently small' but not too small to be reliably solved. In principal, the limiting usable values of condition depend on the truncation error of the particular computer and on experimental error in a particular measurement, whichever is limiting. For the discussion following we concentrate on the relative condition.

The perceptive reader will also notice that non-zero solutions to Eq. (3) (that moves toward a χ^2 minimum) will also require the gradient of χ^2 to be finite. This condition seems to be taken care of by the fact that the resolution is 'good' where the derivatives in the gradient are non-zero so this will not be discussed further here. Also, in this case of ellipsometry there will be $\Psi(\mathbf{a})$ and $\Delta(\mathbf{a})$, two distinct functions rather than one. The availability of these two functions provides choices on how to pose the equation for χ^2 for maximum benefit. For example, it is possible to use multiple measurements but only the equations for Δ or only for Ψ. Details of this method can be found in Numerical Recipes and a description of the non-linear solving algorithm in the literature [5].

The values of film optical properties can be parameterized with respect to light wavelength in various ways. For the work here the following choice was used:

$$n(\lambda) = A_n + B_n/\lambda^2 + C_n/\lambda^4 \tag{5}$$

and

$$k(\lambda) = A_k + B_k/\lambda^2 + C_k/\lambda^4 \tag{6}$$

but the approach will work for any choice. For these equations the determination of the six parameters (A_n, B_n, C_n, A_k, B_k, C_k) determines the optical properties over the selected wavelength range which in the previous publication was covered by 26 values of n and k for a total of 52 numerical values [1].

4. Results and discussion

Fig. 1 is a plot of approximate derivatives of Ψ with respect to A_n, B_n, C_n (from Eq. (5)) and film thickness, d, for a 60-nm thick ZrO_2 film. The approximations are the change in Ψ resulting from a 2% change in each parameter. Because these values vary over orders of magnitude and run between negative and positive numbers, $\log_{10}$ of the absolute value is plotted here. This figure can be read as a plot of where the ellipsometer value Ψ is sensitive to variations in each parameter. Darker areas correspond to greater sensitivity and white corresponds to sensitivity below the measurement limit of 0.01° (perhaps optimistically). Thus the figure shows where there is sufficient resolution since $-2 = [\log_{10}(0.01)]$ for changes of the parameters of 2%. These data are much easier to visualize in color and in top view as can be seen elsewhere (see author's web site[1]). Fig. 2 shows the similar derivatives of Δ with respect to the same parameters. Note that there are more areas of measurable resolution and that the resolution for Δ is generally greater than that for Ψ in the case of this particular reflecting surface. Other conventions for derivative approximations and resolution limits gave different numerical results but preserved the general shape of the derivative surfaces. Derivative with respect to the k parameters were not taken since k was taken to be zero for this analysis by one of the authors (F.K.U.) in the round robin experiment [2].

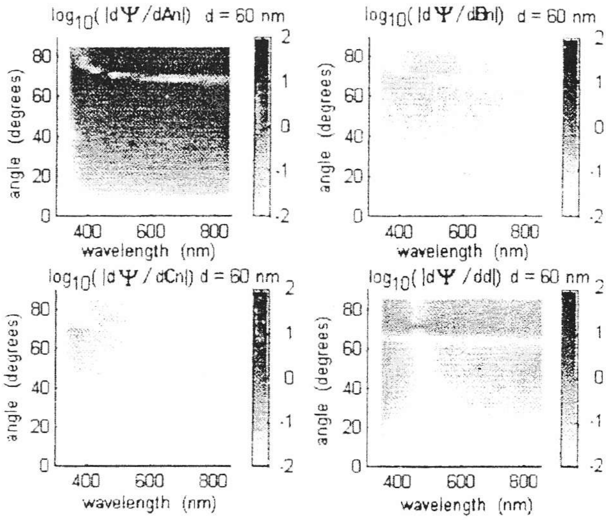


Fig. 1. Log_{10} of the absolute value of the approximate derivatives of Ψ with respect to A_n, B_n, C_n (from Eq. (5)) and film thickness, d, for a 60-nm thick ZrO_2 film. The approximations are the change in Ψ resulting from a 2% change in each parameter.

Not surprisingly, Figs. 1 and 2 show that the ranges of 'good' resolution are not the same for the eight derivatives and seven parameters. Dividing the angle–wavelength space into sensitive (derivative > 0.01°) and insensitive (< 0.01°) and overlaying leads to Fig. 3. This figure shows the points for which all derivatives are sensitive. Each point is labeled with an 'x'. It is interesting that wavelengths longer than 540 nm are not useful to the overall problem. It is less surprising that the usable range of incidence angles runs between 46° and 82° using this algorithm for selecting the overall useful region. There is a gap at 460 nm for lower angles and at 380 nm for larger angles.

In principle, measurements made at the labeled

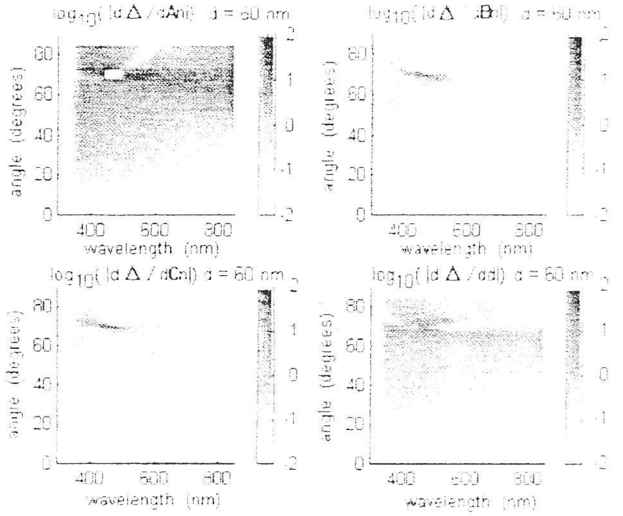


Fig. 2. Log_{10} of the absolute value of the approximate derivatives of Δ with respect to A_n, B_n, C_n (from Eq. (5)) and film thickness, d, for a 60-nm thick ZrO_2 film. The approximations are the change in Ψ resulting from a 2% change in each parameter.

points would have sensitivity to all desired parameters. However, not all of these points provide useful information to the numerical processing solution of the problem. In principle, a minimum of only four of these measurements would be required to determine up to eight real number unknowns (we have seven). So the task is to remove measurement points that do not contribute to improving the condition of the Hessian. Some measurements actually make it worse. This was accomplished by removing the measurements one by one and checking to see if their presence contributed significantly to the condition. It was noted that removal of many of the measurements commonly raised (worsened) the condition a few percent, an amount not likely to bother the accuracy of the solutions in a meaningful way. Other removal resulted in a marked decrease (improvement) in the condition indicating that they clearly should be removed. The removal algorithm could be researched in itself, consisting of all kinds of combinatorial possibilities. For simplicity, the following algorithm was adopted. Points were removed one at a time and the condition checked. If the condition got better on removal (went down) or worsened only slightly (went up 10% or less) then the point was eliminated from the analysis. This process was continued until all points were considered once. While this did not result in selecting the four 'best points' it did result in the selection of far fewer points, 12 out of 114. After the removal, the condition was approximately the same (3.6×10^8 before vs. 4.5×10^7 after) while the measurement and computational load dropped by an order of magnitude.

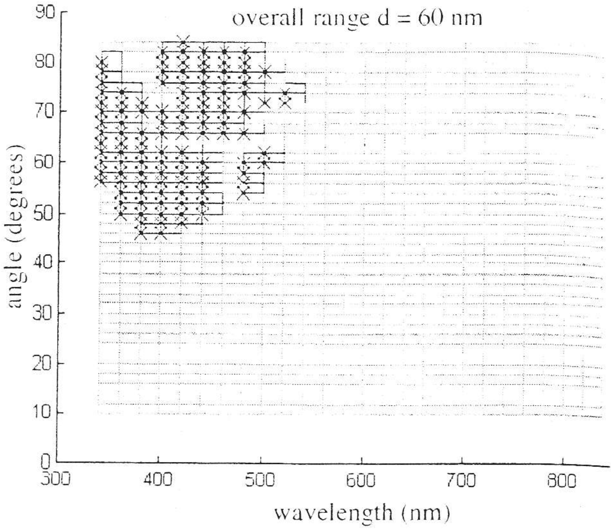


Fig. 3. Plot of points for which 2% changes in all parameters give rise to absolute changes in Ψ and Δ which exceed 0.01°.

Fig. 4 shows the points after removal for the 60-nm thick ZrO_2 film on fused silica. The darker rectangles correspond to the region of overall good resolution. Points remaining following the removal process are circled as before. A similar plot for an 80-nm thick film of the same materials shows quite different set of remaining points. This suggests that if a film were expected in the 60–80 nm range the selected points should include either both sets shown or some logical subset of them.

5. Conclusions

In the following, the major conclusions of this work are enumerated.

1. Expressing film optical properties in parameterized equation form, the resolution of ellipsometry measurements to the equation parameters and film thickness varies over orders of magnitude throughout the angle–wavelength space normally used for variable angle spectroscopic measurements.
2. Numerical approximations of the derivatives of the measured parameters Ψ and Δ can be plotted over this space to clearly show regions of desirable resolution.
3. The systems of ellipsometry equations for parameterized models are such that parameter independence is not guaranteed. The work here demonstrates a method for evaluating the condition of the Hessian and selecting the best angles and wavelengths for computational accuracy.

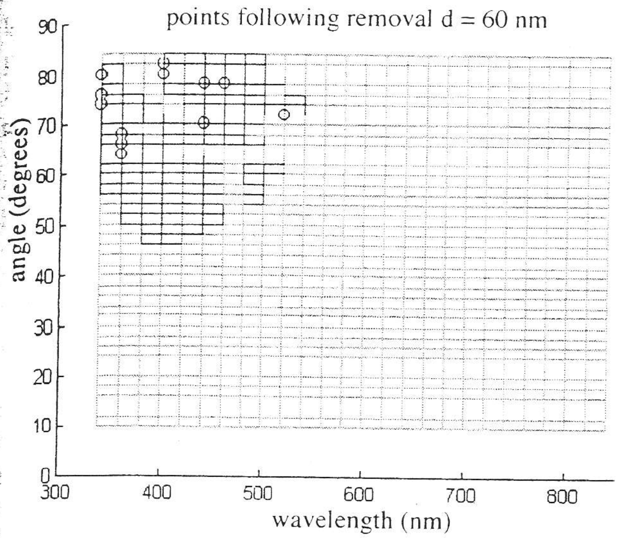


Fig. 4. Plot of the remaining points resulting in a good condition of the Hessian following removal of points which either made the condition worse or had little effect on it (< 10%). Film thickness is 60 nm.

4. It is possible to reduce the number of measurements to on the order of 10 and in principle to four for a system of seven unknowns.
5. The work here suggests that smaller, better ellipsometry data sets can be identified which will result in less measurement time and more accurate results.

Acknowledgements

This research was supported in part by NASA (under grant NAGW-080), ARO (under grants DAH04-96-1-0049 and DAAH04-96-1-0278), NSF (under grant CDA-9313624 and others) and the State of Florida.

References

[1] F.K. Urban, III, D. Barton, Thin Solid Films (in press, 1997).
[2] F.K. Urban, III, P.R. Athey, J. Elman, G. Exarhos, G.E. Jellison, Jr., C. Pickering, J.A. Woollam, M. Yamamoto, J. Vac. Sci. Technol. (accepted, 1997).
[3] Matlab, High-Performance Numeric Computation and Visualization Software, Reference Guide, The MathWorks, Natick, MA, 1995.
[4] F.K. Urban III, J.C. Comfort, Thin Solid Films 253 (1997) 262.
[5] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes the Art of Scientific Computing, Cambridge University Press, New York, 1990.