New Approach for Correction of Distortions in Spectral Line Profiles in Auger Electron Spectroscopy

A. G. B. M. Sasse,* H. Wormeester and A. van Silfhout

University of Twente, Faculty of Applied Physics, PO Box 217, 7500 AE, Enschede, The Netherlands

A new deconvolution method for Auger electron spectroscopy is presented. This method is based on a non-linear least squares minimizing routine (Levenberg-Marquardt) and global approximation using splines, solving many of the drawbacks inherent to the Van Cittert and Fourier transform based deconvolution methods. The deconvolution routine can be run on a personal computer. The application of this method goes beyond the electron spectroscopies and can be considered as a general deconvolution method.

1. INTRODUCTION

To obtain fundamental information from the electrons, the use of the first derivative ($dN(E)/dE$) of the kinetic energy distribution of the Auger process, background subtraction and an accurate unfolding of the broadening of the spectrum are a prerequisite.¹⁻¹⁰ This leads to the unbroadened Auger electron distribution $D(E)$.

Several papers¹¹⁻¹⁵ have discussed how to correct these distortions in the measured $N(E)$ spectra. These corrections are mainly based on background subtraction as outlined by Ramaker *et al.*,¹⁴ and deconvolution of the spectrum with the Van Cittert,¹⁶⁻²⁰ or Fourier transform based methods.^{11,19,20}

Although the Van Cittert method is commonly used by many investigators in research and industrial laboratories, some drawbacks limit its use. One of the main disadvantages concerns the cut-off oscillations at the edges. The width of the broadening function should therefore be sufficiently small.¹⁸ Noisy spectra must be smoothed. Moreover, the method lacks a clear-cut set of analytical conditions which ensures convergence to the true function.^{9,17,18}

Recently,^{20,21} new improvements have been reported on Fourier transform based deconvolution. However, problems such as noise amplification, divergence as a result of division by small values or zero and the choice of the frequency cut-off still remain. We are aware that these difficulties can be overcome to a large extent, but result in a less straightforward method.

The aim of this paper is to present a straightforward deconvolution method, which does not suffer from the drawbacks of using Van Cittert or Fourier transform based deconvolution routines. Our deconvolution method is based on an iterative non-linear least squares minimization and global approximation. Global approximation determines automatically the line regions over the entire spectrum. This is obtained by constructing a trial function composed of splines.²² The

iterative minimizing procedure is based on a modified Levenberg-Marquardt routine (LM),²³ which minimizes the sum of the squares of the difference between the convolution of the trial function with the broadening function and the measured spectrum.

Before the Auger spectrum can be deconvoluted the distortion due to the cascade of collisions of the secondary electrons (secondary background) must be corrected. The background correction is outlined in Section 2. In Section 3 we explain the LM method and the approximation with splines. The application of our method to the deconvolution problem is presented in Section 4. Firstly, a deconvolution of noise containing artificial data, which is a good example of the potency of our method, is shown. Secondly, the application of the deconvolution technique to the Auger spectra of clean and heavily oxidized silicon is presented. The experimental set-up, surface cleaning procedure, and gas handling have been described elsewhere.^{24,25} Concluding remarks are given in Section 5.

2. SECONDARY ELECTRON BACKGROUND

Secondary electron background correction in electron-induced Auger spectra has been performed in several ways,^{11-15,26,27,28,29} the results showing that the best correction for the secondary background is obtained by fitting a theoretical approximation on the low and high energy part of the measured spectrum, which is characteristic for the secondary electrons. Wolff³⁰ derived such an approximation for metals, which was modified by Sickafus¹⁵ and Seah.²⁶

$$B(E) = c_1 + \frac{c_2}{(E - E_\phi)^{c_3}} \quad (1)$$

$B(E)$ is the theoretical secondary background with the c values and E_ϕ as adjustable parameters. Ramaker *et al.*¹⁴ reported that this formula can also be used for semiconductors and insulators. The value of c_3 will change in passing through the Auger transition energy region, as shown by Sickafus.¹⁵ In the high energy

* Author to whom correspondence should be addressed.

region (55–100 eV) these changes are most pronounced. In this energy region, however, the secondary background $B(E)$ can, to a first order, be approximated by a straight line.²⁷ A single value of c_3 therefore suffices (1). Our fit routine operates in such a way that the areas above and below the fit-curve are made equal.²⁷ The absence of structure on the high energy side assures that any point taken here to determine c_1 does not alter the solution of our fit. This eliminates one degree of freedom in eqn (1). The values of c_3 and E_ϕ are 2.0 and 15 eV, respectively.⁴

3. DECONVOLUTION SCHEME

3.1. Modified Levenberg–Marquardt method

As an iterative minimizing method we used in this work a modified Levenberg–Marquardt routine (LM).^{23,31} The LM method states that the direction of search for the minimum lies between the Gauss–Newton method and the steepest descent method. Furthermore, it combines the speed of the Gauss–Newton with the absolute convergence of the steepest descent method. The method minimizes the least squares sum $F(x, c^{(m)}) = f^T f$ in an iterative way.

$$f(x, c^{(m)}) = y(x, c^{(m)}) - y(x)^* \quad (2)$$

$$c^{(m+1)} = c^{(m)} + \delta^{(m)} \quad (3)$$

and

$$(J^T J + \lambda^{(m)} D) \delta^{(m)} = -J^T f(x, c^{(m)}) \quad (4)$$

c = the set of spline coefficients

$y(x, c^{(m)})$ = trial spectrum

$y(x)^*$ = measured spectrum

J = Jacobian defined as

$$J_{ij}(x) = (f_i(x, c_j + \delta_j) - f_i(x, c_j)) / \delta_j$$

D = diagonal matrix of $J^T J$

$\lambda^{(m)}$ = sequence of non-negative real constants

m = iteration number

For ($\lambda \ll 1$) the LM iteration behaves like a Gauss–Newton method, and for ($\lambda \gg 1$) like a steepest descent method. The magnitude of this constant (λ) is properly chosen by a sequence developed by Fletcher.^{23,31} The parameter (λ) ensures that the LM iteration follows the fast Gauss–Newton method as closely as possible without losing convergence. It is obvious that the choice of (λ) is of ultimate importance for the performance of this method.

3.2. Splines

Global approximation of the unbroadened Auger electron energy distribution $D(x)$, can be achieved by constructing the trial function $D(x, c)$ on a basis of piecewise continuous functions, which are connected to each other in such a way that the sum of these functions and their derivatives are continuous to a certain order ($k - 1$) over the entire spectrum, with k the order of the piecewise continuous function. A very convenient definition of these piecewise continuous functions are splines.^{22,31} These splines, $B_{i,k}(x)$, defined on a subinter-

Figure 1. Three splines of fourth order and three left and right boundary splines. The summation of splines with all $c_i = 1$ representing a block function is shown.

val (x_i, x_{i+k}) of the interval (x_1, x_N), can be constructed by a recursive sequence of variable order k .

$$B_{i,k}(x) = \frac{x - x_i}{x_{i+k-1} - x_i} B_{i,k-1}(x) + \frac{x_{i+k} - x}{x_{i+k} - x_{i+1}} \times B_{i+1,k-1}(x) \quad (5)$$

with

$$B_{i,1}(x) = \begin{cases} 1; & x_i \leq x < x_{i+1} \\ 0; & \text{elsewhere} \end{cases} \quad \begin{matrix} k = 2, 3, 4, \dots \\ i = 1, \dots, N \end{matrix}$$

The number of intervals influenced by a spline, defined as support, is equal to the order of the spline. With the introduction of the boundary spline a larger class of functions can be represented (Fig. 1). High frequency oscillations (like noise) are suppressed when higher order splines ($k \geq 3$) are used. Since the splines are normalized (6), a geometrical interpretation of the coefficients c_i (7) is straightforward.

$$\sum_i B_{i,k}(x) = 1 \quad (6)$$

This definition of piecewise continuous functions is very convenient for constructing simple computer algorithms.

The deconvoluted spectrum $D(x)$ can be approximated by N_S splines.

$$D(x, c) = \sum_i^{N_S} c_{i,k} B_{i,k}(x) \quad (7)$$

According to the definition of the convolution; $S = D * V$, with V the broadening function and D the deconvoluted spectrum, we can write.

$$\begin{aligned} S(x, c) &= \int D(x - \Delta) V(\Delta) d\Delta \\ &= \int \sum_i c_{i,k} B_{i,k}(x - \Delta) V(\Delta) d\Delta \end{aligned} \quad (8)$$

The function $F(x, c)$, which is minimized in the LM method is defined as

$$\begin{aligned} f(x, c) &= S(x, c) - S^* \\ F(x, c) &= f^T f \end{aligned} \quad (9)$$

where S^* is the measured signal. As input for the LM method, the array of points composing the spectrum, is defined as a vector. In the LM method a set of c_i is calculated, from which a $D(x, c)$ can be reconstructed (7). This is the derived approximation of the real $D(x)$.

Figure 2. The iteration scheme of the deconvolution calculation.

The scheme of our deconvolution routine is displayed in Fig. 2. The influence of the choice of the parameters in this routine on the result of the calculation is tested elaborately and showed to be negligible, if they are not tuned on extreme values. The number of points per interval should be larger than 5 and $N/k > 5$. However, there is no principal limitation of k . A convenient choice is given in the caption of Fig. 5.

4. APPLICATION TO AUGER ELECTRON SPECTROSCOPY

Mularie and Peria¹³ have shown a method to obtain a broadening function $V(E)$ that is believed to correct for distortions of the Auger line-shape as loss processes, transport and instrumental broadening. The Auger line-shape $D(E)$ can either be derived from deconvolution of the energy distribution $S(E)$ with the broadening function $V(E)$ or from its derivatives. This mathematical property of the convolution can be easily derived from the definition of the convolution integral.

$$S(E) = D * V = V * D = \int_{-\infty}^{+\infty} V(E - \Delta) D(\Delta) d\Delta \quad (10)$$

When $D(E)$ and $V(E)$ are well behaved functions, which only differ from zero on a well-defined interval $[0, E_0]$, (10) may be written as.

$$\frac{\partial S(E)}{\partial E} = \int_0^{E_0} D(\Delta) \frac{\partial V(E - \Delta)}{\partial E} d\Delta \quad (11)$$

The Auger spectra and the broadening function are

measured in the derivative mode as described in Refs 24 and 25. By deconvoluting (11) instead of (10), the Auger energy distribution $S(E)$ and the broadening function $V(E)$ do not need to be integrated. Ramaker *et al.*¹⁴ have shown that small differences in the background subtraction leads to an accumulation of the errors on one energy side of the integrated spectrum. Deconvolution in the derivative mode therefore gives a more accurate fit over the entire energy region of $S(E)$, i.e. $\partial S/\partial E$.

The virtue of our deconvolution procedure is its global nature, which prevents noise amplification. This in contrast to the Van Cittert method, where, owing to noise, convergence of the residue does not automatically imply convergence of the deconvoluted spectrum. The restrictions on $S(E)$ and $V(E)$, such as non-rapid fluctuating functions and the width of $V(E)$, are no longer necessary.^{9,17,18}

Furthermore, we do not suffer from instabilities inherent to the Fourier transform based deconvolution routine^{19,20} such as the divergence as a result of division by small numbers or zero, aliasing, etc. These Fourier transform based methods suffer more from noise amplification than the Van Cittert method and additional repeated smoothing is necessary, resulting in a less straightforward method.

The number of mathematical solutions of the deconvolution problem, which can be obtained and their physical relevance is discussed in the Appendix. Therein, it is shown that $D(E)$ can be uniquely determined if all frequencies of the Fourier transform of the real solution $D(E)$ are contained within the frequency spectrum of the broadening function. This condition is usually fulfilled for Si-L_{2,3} VV Auger spectra.

The implemented computer program was tested with artificial noise containing data. Good agreement between the outcome of the deconvolution calculation and the initial artificial data was achieved. The results, depicted in Figs 3 and 4 were obtained after seven iterations and contain 200 points. The presumptions of the functions for applying the Van Cittert method are not fulfilled¹⁷ in the data depicted in Fig. 4.

The broadening function $V(E)$, which arises from the finite resolution of the analyser, intrinsic loss effects (inelastic scattering etc.), has been obtained following the procedure as described by Mularie and Peria.¹³ We noticed in our tests that deviations in $\partial S/\partial E$ did not yield large changes in our $D(E)$. Deviations in $\partial V/\partial E$, however, have a significant effect on the final result. This difference in response is due to a mathematical property of the convolution. Therefore, $\partial V/\partial E$ should be

Figure 3. Test of the deconvolution method with artificial data: (a) Top: $S(E)$; convolution of the deconvoluted result $D(E)$ (top curve Fig. 3(c)) of the noise contained spectrum $S^*(E)$ with the broadening function $V(E)$. Bottom: $S^*(E)$ input signal of the deconvolution routine ($S^*(E) = D^*(E) * D^*(E)$ with 5% noise added). (b) Broadening function $V(E)$ ($=D^*(E)$ with 5% noise added). (c) Top: $D(E)$, result of the deconvolution. Bottom: $D^*(E)$, original function. The residue of the minimizing is 3.8×10^{-4} per point (200 data points). The parameters are of the same order as used in Fig. 5. and are not critical.

Figure 4. Test of a deconvolution with artificial data: (a) Top: $S(E)$; convolution of the deconvoluted result $D(E)$ (top curve Fig. 4(c)) of the noise contained spectrum $S^*(E)$ with the broadening function $V(E)$. Bottom: $S^*(E)$ input signal of the deconvolution routine ($S^*(E) = D^*(E) * D^*(E)$ with 10% noise added). (b) Broadening function $V(E)$ ($=D^*(E)$ with 10% noise added). (c) Top: $D(E)$, result of the deconvolution. Bottom: $D^*(E)$, original function. The residue of the minimizing is $<5 \times 10^{-4}$ per point (200 data points). The parameters are of the same order as used in Fig. 5.

measured as accurately as possible. The increase of the noise, inherent to a more sensitive measurement of $\partial V/\partial E$, is insignificant, because of the global nature of our newly developed method. On $\partial V/\partial E$ the same background correction procedure has been used as on the Auger spectra.

To illustrate its performance for real measured spectra our newly developed deconvolution method has been applied to the actual Auger Si-L_{2,3} VV spectra of clean and oxidized silicon. The results are shown in Fig.

Figure 5. (a) Clean Si secondary background corrected. (b) $D(E)$ clean Si. (c) oxidized Si background corrected. (d) $D(E)$ of oxidized Si. Some characteristic parameters used in our deconvolution method: $k = 3$, $c_i = 0.2$, $\lambda_0 = 100$, $N_s = 27$, $N = 30$, data points = 200, $N_{1,r} = 0$; Residue of the oxidized Si < 0.1 , residue of clean Si < 0.042 .

5. The overall shape is in a good agreement with the results of Knotek and Houston.³² In a forthcoming publication these and other spectra will be the subject of extensive discussion.

5. CONCLUDING REMARKS

In this paper we have presented a new method of deconvolution. The method uses splines in combination with a powerful iterative minimizing routine. Since the method has a global character noise amplification, inherent to the Van Cittert and Fourier transform based algorithms, has been suppressed. Also other drawbacks such as division by small numbers or values of zero in FFT based routines and the presumptions of the functions used in the Van Cittert method are eliminated. This allows us to use relatively large fluctuating and noise containing functions as input for our deconvolution method. Data processing can therefore be performed in the sensitive derivative mode, thus avoiding integration-induced errors. Our method can be considered to be essentially free of artefacts. Furthermore splines have convenient numerical properties.

The deconvolution routine can be executed on a small personal computer (Olivetti M24, 64 Kb; calculation time about a few seconds, 1000 points, 7 iterations) and is easy to implement.³³

Not only the deconvolution problem can be solved by this method but also many problems of the form,

$$D(x) = A^{-1}(S(x)) \quad (12)$$

with A being a well-defined operation on an unknown spectrum $D(x)$ and $S(x)$ a known function. Often it is not easy, or even impossible, to compute A^{-1} explicitly. This class of problems can in principle be solved by applying our iterative scheme. As example, numerical differentiation of noisy spectra could be achieved successfully by implementing an integration procedure instead of convolution in our scheme (Fig. 2).³³ Another example of this scheme is discussed in Ref. 33.

Acknowledgements

We gratefully acknowledge useful conversations with Dr R. H. J. Gmelig Meyeling and Ir A. Weishaupt about the ins and outs of splines. Ir D. G. Lakerveld is acknowledged for implementing the deconvolution method on the personal computer. We are very grateful to Dr E. G. Keim for careful reading of the manuscript before publishing.

REFERENCES

1. P. H. Holloway, *Adv. Electron. Electron Phys.* **54**, 241 (1980).
2. D. E. Ramaker, *Chemistry and Physics of Solid Surfaces IV*, Springer Series in Chemical Physics, Vol. 20, Ed. by R. Vanselow and R. Howe, Springer, Berlin, p. 19 (1982).
3. P. J. Feibelman, E. J. McGuire and K. C. Pandey, *Phys. Rev.* **B5**, 2202 (1977).
4. D. E. Ramaker, J. S. Murday, N. H. Turner, C. Moore, M. G. Lagally and J. E. Houston, *Phys. Rev.* **B19**, 5375 (1979).
5. J. J. Lander, *Phys. Rev.* **91**, 1382 (1953).
6. H. D. Hagstrum and G. E. Becker, *Phys. Rev.* **B4**, 4187 (1971).
7. H. D. Hagstrum and G. E. Becker, *Phys. Rev.* **B8**, 1580 (1973).
8. G. F. Amelio, *Surf. Sci.* **22**, 301 (1980).
9. J. E. Houston, *J. Vac. Sci. Technol.* **12**, 255 (1974).
10. R. Weissmann and K. Muller, *Surf. Sci. Rept.* **1**, 251 (1981).
11. J. A. Taggle, V. Martinez, J. M. Rojo and M. Salmeron, *Surf. Sci.* **79**, 93 (1978).
12. D. E. Ramaker, *Appl. Surf. Sci.* **21**, 243 (1985).
13. W. M. Mularie and W. T. Peria, *Surf. Sci.* **26**, 125 (1971).
14. D. E. Ramaker, J. S. Murday and N. H. Turner, *J. Electron Spectrosc. Rel. Phenom.* **17**, 45 (1979).
15. E. N. Sickafus, *Phys. Rev.* **B16**, 1436 (1977); E. N. Sickafus, *Phys. Rev.* **B16**, 1448 (1977); E. N. Sickafus and C. Kukla, *Phys. Rev.* **B19**, 4056 (1979).
16. P. H. Van Cittert, *Z. Phys.* **69**, 298 (1931).
17. H. H. Madden and J. E. Houston, *J. Vac. Sci. Technol.* **47**, 3071 (1976).
18. H. H. Madden and J. E. Houston, *J. Appl. Phys.* **47**, 3071 (1976).
19. A. F. Carley and R. W. Joyner, *J. Electron. Spectrosc. Rel. Phenom.* **16**, 1 (1979).
20. B. Chornik, R. Sopizet and C. le Gressus, *J. Electron Spectrosc. Rel. Phenom.* **42**, 392 (1987).
21. A. Kosarev and E. Pantos, *J. Phys.* **E16**, 537 (1983).
22. P. M. Prenter, *Splines and Variational Methods*, Wiley, New York (1975).
23. R. Fletcher, *A Modified Marquardt Subroutine for Non-linear Least Squares Fitting*, U.K. Atomic Energy Authority Research Group, Theoretical Physics Division, AERE-R 6799, Harwell, Berkshire (1971).
24. E. G. Keim, L. Wolterbeek and A. van Silfhout, *Surf. Sci.* **180**, 565 (1987).
25. A. G. B. M. Sasse, D. G. Lakerveld and A. van Silfhout, *Surf. Sci.* **195**, L167 (1988).
26. M. P. Seah, *Surf. Sci.* **17**, 132 (1969).
27. R. H. Brockman and G. J. Russell, *Phys. Rev.* **B22**, 6302 (1980).
28. D. C. Peacock and J. P. Duraud, *Surf. Interface Anal.* **8**, 1 (1986).
29. J. A. D. Matthew, M. Prutton, M. M. El Gomati and D. C. Peacock, *Surf. Interface Anal.* **11**, 173 (1988).
30. P. A. Wolff, *Phys. Rev.* **95**, 56 (1954).
31. A. G. B. M. Sasse, D. G. Lakerveld and A. van Silfhout, *J. Vac. Sci. Technol.* **A6**, 1045 (1988).
32. M. L. Knotek and J. E. Houston, *J. Vac. Sci. Technol.* **1**, 899 (1983).
33. H. Wormeester, A. G. B. M. Sasse, and A. van Silfhout, *Comp. Phys. Comm.* **52** (1988).

APPENDIX

The main objection to deconvolution is the infinite number of solutions which can be generated. In this appendix we show the relation between the mathematical solutions and their physical relevance. The convolution integral is defined as

$$S(y) = \int_{-\infty}^{+\infty} D(y-x)V(x) dx \quad (\text{A1})$$

with D and V being well-defined functions. The deconvolution problem is defined as follows: the functions S and V are known and the convolution operation is defined by (A1), then the function D can be calculated as follows

$$D = A^{-1}(S, V) \quad (\text{A2})$$

with A^{-1} being the inverse operation of the convolution.

Assume N is a well-defined arbitrary function, which is under the constraint of

$$0 = \int_{-\infty}^{+\infty} N(y-x)V(x) dx \quad (\text{A3})$$

All functions $D^*(x) = D(x) + N(x)$ are mathematical solutions of the deconvolution problem. An infinite number of functions N exists. This is easily seen by Fourier transforming (A3).

$$0 = N(\omega)V(\omega) \quad (\text{A4})$$

If we assume that the highest frequency of V is ω_v , then N is a function containing only frequencies $[\omega_v, \infty]$.

Fortunately, the determination of the solution of the deconvolution is less difficult in real physical cases. High frequencies can be characterized as noise. Therefore, the amount of functions N , which contributes to D^* is limited on physical grounds. If the maximum frequency ω_d of D is lower than ω_v , D can be uniquely determined. If $\omega_d > \omega_v$, then (on mathematical grounds) D cannot be determined in any deconvolution method and the information related to the frequency interval $[\omega_v, \omega_d]$ is lost. In Si-L_{2,3} VV Auger electron spectroscopy we can satisfy the condition $\omega_d < \omega_v$ so that D can be determined. Noise amplification in the Van Cittert and Fourier transform based methods is a consequence of the mathematical existence of functions N . This amplification can be suppressed by repeated smoothing during the calculation, however physically relevant information will also be distorted.

The global approximation with splines implies that one fits with a fixed range of frequencies, which must include all frequencies of D . The frequency spectrum in polynome fitting will change during the approximation and introduces additional noise, in contrast to the unchanged frequency spectrum of a spline.

The deconvolution method discussed in this paper uses only the properties of the convolution integral, therefore the theoretical limit can be obtained. This statement implies that the deconvolution method for $\omega_d < \omega_v$ is restricted by numerical limitations only.