Original Paper

A background subtraction program for photo- and Auger electron spectra

Masatoshi JO
National Institute of Advanced Industrial Science and Technology (AIST),
National Metrology Institute of Japan
AIST Tsukuba Central 5-1, Higashi 1-1-1, Tsukuba, Ibaraki 305-8565, Japan
m-jo@aist.go.jp

(Received: November 27, 2007; Accepted: April 08, 2008)

A Windows program designed for background analysis of AES and XPS has been developed and is now available on the web. An overview of this program will be given.

1. Introduction

Since its first report [1] the background optimization method has been progressively and continuously improved [2]. In its earlier stages, it could treat only XPS spectra, in which a pair of peaks of known intensity ratio and fixed boundaries was needed. The present version no longer requires such assumptions. It can analyze both XPS and AES, since the escape process of the electrons are same. It can also process the peaks whose intensities and/or boundaries are not known beforehand. In this report, an overview, structure and functionalities, of the program, which is now available on the web [3], will be given.

2. Principle

In this section explained are the essence of employed basic assumptions and their implementation to the optimization problem [1, 2]. The total signal intensity $J_{total}(E)$ of XPS and AES, where E is the electron kinetic energy (KE), are generally decomposed as follows (Fig.1(a) and (b)),

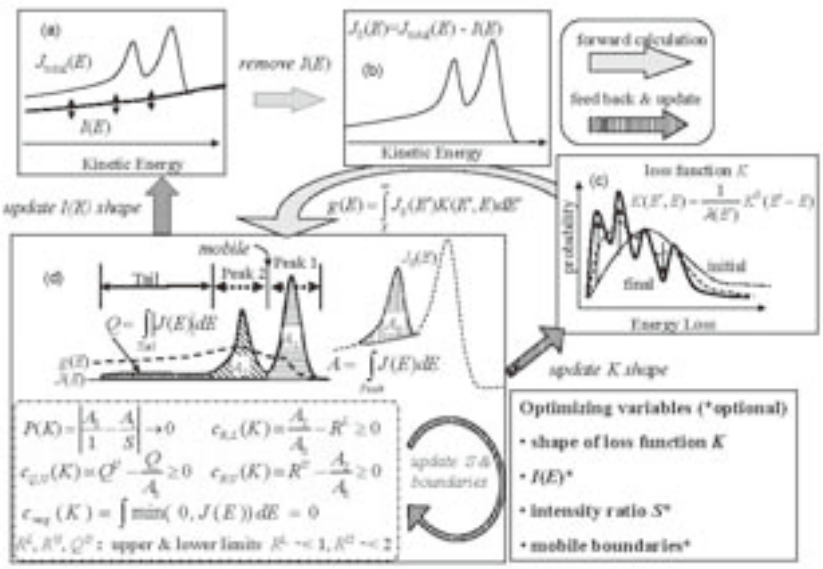

Fig 1. Outline of Algorithm

$$J_{total}(E) = J_0(E) + I(E) \quad (1)$$

$J_0(E)$ is the contribution from the peak in concern, while $I(E)$ is the rest part which is irrelevant to $J_0(E)$. $I(E)$ comes from all other origins on the spectrum, such as true secondary electrons, losses of higher energy peaks, etc. $J_0(E)$, in Fig.1(b), can be further divided into two parts, the inelastic background $g(E)$ and background-subtracted spectrum $J(E)$ by

$$g(E) = \int_E^\infty J_0(E')K(E',E)dE' \quad (2)$$

and

$$J(E) = J_0(E) / \lambda(E) - g(E) \quad (3)$$

as shown in Fig.1(d).

$$K(E'\text{-}E),\ K(E',E) = \frac{1}{\lambda(E')} K^0(E'-E) \quad (4)$$

called the loss function, is the probability of an electron of KE E' losing an amount of (E' - E) per unit time interval and unit travel distance (Fig.1(c)). $\lambda(E')$ is the inelastic mean-free path of an electron of KE = E'. $1/\lambda(E')$ carries the initial KE dependence, whereas K^0 is a normalized part that includes only the energy loss (E'-E). In the present study, it is the shape of K^0 that is determined by the optimization whereas $\lambda(E')$ is provided separately.

Suppose that the 'true' loss function K is known. Then the spectrum $J(E)$ after subtracting $g(E)$ using K is also true, by definition. For this true $J(E)$, it is quite reasonable to expect that four assumptions described below hold, if $J(E)$ contains at least two distinct regions (denoted as Peak1 and Peak2 in Fig.1(d)) each of which includes at least one peak as a usual meaning, and at least one region at the lower-KE side of Peaks (denoted as Tail) where no appreciable intensity is expected. In the expressions below, there may be variables other than the shape of K, which are omitted for brevity.

1) Let the integrated peak areas of Peak1 and Peak2, be A_1 and A_2. Then these would be proportional to certain values that are close to the electron emission probabilities of isolated atoms. Let the true area of Peak1 is S times larger than that of Peak2, then the following function P should be 0 if true K is obtained,

$$P(K) = \left| \frac{A_2}{1} - \frac{A_1}{S} \right| \to 0 \quad (5)$$

Thus P is used as the objective function to be minimized. The pair of Peak1 and Peak2 is denoted as 'Pair'.

2) The Peak areas also must not be too small or too large compared with themselves calculated using simpler procedures such as straight line method, as long as the peaks originate from the actual no-loss peaks. In Fig.1(d), A_2 is compared to the area A_0 of the same energy interval but calculated by the simpler method (right-upper inset). A_0 is used as a measure of typical no-loss peak intensity estimated directly from the data. The definition of A_0 suggests that A_2/A_0 would be mostly around 1 and not become much larger than 2. This leads to two new constraints with constants R^L and R^U as follows, which are lower and upper limits of A_2/A_0, respectively.

$$c_{R,L}(K) \equiv \frac{A_2}{A_0} - R^L \geq 0$$

$$c_{R,U}(K) \equiv R^U - \frac{A_2}{A_0} \geq 0 \quad (6)$$

From its definition, it is not likely that R^L >>1, and/or R^U >>2.

3) In a similar manner, the area (integrated absolute intensity) of Tail, denoted as Q,

$$Q = \int_{Tail} |J(E)| dE \quad (7)$$

must not be too large by definition, if it is normalized by A_0, . Upper limit of Q/A_0 is denoted as Q^U in the figure. Absolute value of $J(E)$ is used in the integration to avoid the accidental cancellation of large positive and negative intensity. Note that $J(E) \equiv 0$ is not needed because there may be low but continuous intrinsic contribution outside the Peak, in general. The condition is expressed as

$$c_{Q,U}(K) \equiv Q^U - \frac{Q}{A_0} \geq 0 \quad (8)$$

The magnitude of Q^U is typically comparable to that of R^U, but also depends on the width of Tail.

4) The signal of $J(E)$ must not, of course, be less than zero anywhere, i.e.,

$$c_{neg}(K) \equiv \int \min(0, J(E)) dE = 0 \quad (9)$$

What is done in the background analysis is to deter-

mine good $g(E)$ and $J(E)$ that satisfy these assumptions, by estimating the *shape* of K^0. How to express the shape of K^0 by optimizing variables is described in ref. [1]. Obviously, above assumptions are not enough to determine K, because, in general, none of the proper intensity ratio of Peaks S, proper extents of Peak and Tail, or proper shape of $I(E)$, is available beforehand the analysis. To overcome this serious ambiguity, the present method has options to treat these unknowns as optimizing variables. For intensity ratio S and Peak/Tail boundaries, how to do this has been presented in the previous report [2]. Parameterizing the shape of $I(E)$ is similar to that for K^0 [1]. Typical numbers of variables for K^0, S, boundaries and $I(E)$ are several tens to several hundreds, 0 or 1, 0 ~ 2, and 0 ~ 10, respectively. Of course, if reliable values of these are known, one can use them in the problem and reduce the ambiguity.

In the present study, function P is chosen as the objective function to minimize, whereas c's in expressions (6, 8, 9) are used as the constraint conditions. It is noted that P and constraints in (6) mainly determines the shape of lower energy-loss part of K0, i.e. 0 to the total width of (Peak1 + Peak2), whereas condition (8) carries the information of K0 up to the width of (Peak1 + Peak2 + Tail). Problem thus constructed is solved by a successive quadratic programming (SQP) method written by Fukushima [4]. Optimization process resembles a motion of a point sliding down on a potential surface determined by the objective function z = P(x) and surfaces by c(x) = 0 [5]. The point feels repulsive force when it enters the forbidden region c < 0. The current position x of the point represents all the optimization variables. The position where the point eventually comes to rest, which is typically one of local minima, is determined by a balancing of forces exerted on the point, i.e. 'gravity' to minimize the objective function, and 'forces' from the surfaces that the point is simultaneously in contact with. This force balancing is called as Kuhn-Tucker (K-T) condition [5]. One may easily notice that, in eq.(5), P readily becomes zero, if S = A1/A2 for arbitrary values of A2 and A1. However this point is not selected as the solution because K-T condition is not satisfied.

Category	Functionality	Comments
File I/O	automatic format conversion from PHI (*.pcs) to NPL	new format can be added on request
	Load/Save original data format	
OLE	automatic and direct transfer to Excel	for various data&info in this analysis
Data Reduction	operation on the data by math formula	see Math formula below
	linear, spline, Shirley background subtraction	
	peak areas, peak widths	
	Savitzky-Golay Filter (smooth & 1st/2nd derivatives)	cubic, up to 25 points
	spline approximation	linear/cubic polynomial
operations on spectrum region	region duplication	creates a new region by copying whole or a part of a spectrum
	data points reduction if they are too dense	ex. 0.01 eV step → 0.1 eV step
spectrum display	vertical scale change of individual spectrum	for both lower & higher boundaries
	zooming of any portion of data	
loss function (K^0)	intialization using Tougaard's 2-parameter universal function	energy step may differ from that of spectrum data
	integration, smoothing, self-convolution	
IMFP	values by Tanuma, Powell, and Penn	stored in text (*.csv) file which can be edited by the user
	TPP-2M by user-supplied parameters	
Peak and Tail conditions	edits coefficients and boundary positions	can be optimizing variables
Batch conditions	Batch constants	R^2, $R^{2'}$, Q^2 and others
cubic/linear spline background	used either for irrelevant background or data approximation	can be optimizing variables
Math formula	data modification using arithmetic expressions including intensities, kinetic energies, IMFP, splines, sin, cos, atn, tan, sqr, exp, log, abs and step	performs point-wise operation on one spectrum j, or operation between two spectra j1 and j2, such like f(E)*j(E) + const(E) or f1(E)*j1(E)+f2(E)*j2(E)+const(E)

f, f1, f2, const: arbitrary functions of kinetic energy and other quantities in the spectrum
j, j1, j2: specta on which mathmatical operation is done

Table 1. Program's functionnalities

Similar to the additional variables described above, proper values of R^L, R^U and Q^U are also not known. However, it is still not possible to treat these as optimizing variables because conditions (6) and (8) are easily and wrongly fulfilled whenever R^U and Q^U are large and R^L is small. This is a serious loss of search path guided by the conditions. The present method proposes a series of calculations each of which has different combination of R^L, R^U and Q^U that is systematically generated instead of dealing with them as optimization variables. This is denoted as 'batch calculation' below. For example, if R^L=0.4, 0.6, 0.8, R^U=0.5, 0.7, 0.9, 1.1, Q^U=0.4, 0.6, 0.8, 1.0, totally 36 different jobs are possible considering that cases $R^U \leq R^L$ are meaningless. In addition to these, there are more a few constants including weight of each constraint, which are omitted in the present discussion. Appropriateness of each result is judged by checking various unfavorable aspects, negative $J(E)$, divergence or rapid oscillation of loss function, etc.

Usual initial value settings of the variables are as follows. For S, A_1^0/A_2^0 is used, where superscript 0 denotes peak area calculated by a simpler method like A_0. For K^0, shape of Tougaard's universal function is used. For mobile boundary, apparent border between peaks is used. For $I(E)$, straight line extrapolated from the higher KE side's base line of the peak is used.

It is noted that, for such complicated problem, obtained 'solution' is generally not the true globally optimal (and unique) one, but a state that is in the neighborhood of one of local minima. If this local minimum is in the neighborhood of true minimum, the solution would be regarded an approximation of the true solution.

3. The program

The present program [3], running on Windows, consists of two units, graphical user interface (GUI) and Optimizer. GUI performs all the tasks except optimization. Capabilities implemented in GUI are summarized in Table 1. In addition to those essential for analysis, various features are equipped to aid further analysis. Optimizer is based on the Successive Quadratic Programming method (SQP) written by Fukushima [5], which has been arranged for the present study by the author. The treatable data size is limited by only the installed memory amount. Typical calculation time for one optimization

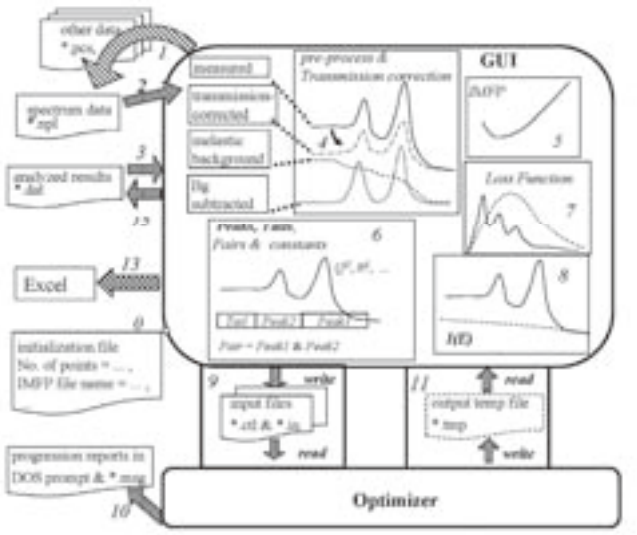

Fig 2. Steps of Analysis

is several seconds to several minutes depending on the actual data size, problem's complexity and CPU speed.

For both XPS and AES, the lowest KE region, where the secondary electron intensity is dominant, is difficult to analyze due either to lack of proper IMFP, or to very steep slope of the data. Also, for AES, analysis of primary beam and its inelastic background is not possible because there is only one peak. In between these, part including more than two photo- or Auger peaks with total energy span of 50 - 100 eV is the typical region suitable for analysis. The width of spectrum restricts calculable maximum energy loss of K^0.

Although a spectrum of better quality, i.e., finer energy step, higher resolution and less noise, would give more satisfactory result, this is not always required. In some cases a portion of wide scan spectrum, which is taken quickly as a survey, e.g. by coarse energy step (= 0.5 - 1 eV) and medium/low resolution, is enough for analysis. It is noted that too many points in the spectrum would make the calculation slow.

There is no restriction on the material's type, as long as the spectrum is properly measured. However a special care should be taken as follows. In the present method, one loss function, satisfying eq.(2), is obtained for each analysis. This implies that the element's depth distribution is assumed to be uniform, as in the case of pure metal. If the distribution is indeed uniform, the background of entire spectrum would be subtracted very well by this single loss function. If not, as in the case of mate-

rial with surface layer, the obtained 'loss function' should be interpreted as a complicated function of depth-dependent versions of loss function, IMFP, and atom density [6]. In this case, the obtained loss function would fail to subtract the background of the peaks whose depth distribution is different from those used in the calculation. This may be used for distinguishing the difference of depth distribution.

Steps for a typical analysis are shown in Fig.2. These can be repeated if desired. In batch mode, Steps 10 - 12 are repeated.

Step 0: Parameters are initialized.

Step 1: Format conversion to npl if necessary.

Steps 2 & 3: Load npl or previously processed file.

Step 4: Analyzer transmission correction is done.

Step 5: IMFP is set.

Step 6: Initialization of Peaks, Tails, and other conditions is done.

Step 7: Initialization of loss function.

Step 8: $I(E)$ is defined and initialized.

Step 9: Create the initial data files and start the Optimizer.

Step 10: The Optimizer shows the progress during calculation.

Step 11: The Optimizer returns the result to GUI.

Step 12: GUI saves the obtained result.

Step 13: The result can be directly transferred to Microsoft Excel for further examination.

4. Summary

Algorithm and features of the program, as of November 2007, for background subtraction have been described. The program is accessible on the internet [3] as a free software. It can be also used as a general processing tool of electron spectroscopic data as shown in Table 1.

5. Acknowledgements

This work was financially supported by New Energy and Industrial Technology Development Organization (NEDO) (*the Material Nanotechnology Program, 2001* - 2007)

6. References

[1] M. Jo, *Surf Sci*., 320,191 (1994).

[2] M. Jo, *Surf. Interface Anal.,* 35, 729(2003)

[3] http://nano.s102.xrea.com/xps/xps_program_form.html

[4] M. Fukushima, *Math. Program,*. 35, 253 (1986).

[5] U. Diwekar, *Introduction to Applied Optimization*, Applied Optimization Vol. 80, Kluwer Academic Publishers (2003) ISBN 1-4020-7456-5.

[6] S. Tougaard and H. S. Hansen, *Surf. Interface Anal.,*. 14, 730(1989).