SECTION 12
DATA MASSAGE ALGORITHMS

Contents

12.0 Introduction 12-1
12.1 Area 12-2
12.2 Atomic Concentration 12-3
12.3 Atomic Concentration Plot 12-3
12.4 Baseline Subtract 12-4
12.5 Curve Fit 12-4
12.5.1 Standard Peak Shapes 12-4
12.5.2 Peak Shape Parameters 12-4
12.5.3 Multiplet Locking 12-5
12.5.4 Parameter Limits and Value Locking 12-5
12.5.5 Background Subtraction 12-5
12.5.6 Manual and Automatic Parameter Adjustment 12-6
12.5.7 Non-linear Least Squares Optimization 12-6
12.5.8 Asymmetric Gaussian-Lorentzian function 12-8
12.5.9 Optimization Algorithm 12-8
12.5.10 Dependency Locking 12-10
12.5.11 Special Considerations 12-11
12.6 Deconvolution 12-12
12.7 Differentiation 12-14
12.8 Height/FWHM 12-16
12.9 Normalize 12-17
12.10 Normalize by E 12-17
12.11 Smoothing 12-18
12.12 X-Ray Satellite Subtraction 12-20

12.0 Introduction

Section 12 discusses the principles of data reduction and massage. Mathematical algorithms are provided for selected data massage routines and briefly described. For a more complete discussion of the algorithms, refer to the technical articles cited and listed in the footnotes.

12.1 Area

The area under a peak provides a quantitative measure of signal strength. While the peak height may vary due to the chemical state of the element or the operating conditions of the analyzer, the area will remain consistent.

Four steps are involved in calculating the peak area for ESCA acquired data:

- The raw data (in counts), which defines the curve, is used to calculate the absolute area under the curve.
- The background area is calculated and subtracted from the absolute area.
- The area is normalized for time.
- The area is normalized for the unit step.

If the area is to be calculated for ISS acquired data, the area must also be normalized for the ion energy.

Calculating Area

1. All data curves are described by a discrete number of data points. The absolute area under the curve is calculated using the equation described below. Figure 12-1 represents area integration by parabolic interpolation. The intervals are chosen symmetrically about the data points. Because the number of data points used in fitting each parabola is odd, the intervals must be chosen straddling data points, rather than between them.

   Since the first interval A_1, is only half as wide as the others, the area is determined by the equation:

   $A_1 = [ax + 1/2bx^2 + 1/3cx^3]^{1/2}_0$

   $= 1/24(8y_1 + 5y_2 - y_3)$

   The area in the next interval is determined by the equation:

   $A_2 = [ax + 1/2bx^2 + 1/3cx^3]^{3/2}_{1/2}$

   $= 1/24(y_1 + 22y_2 + y_3)$

   This equation is heavily weighted in favor of the middle data point. The area of the following intervals A_i are evaluated in the same way as that of A_2 with the equation listed above incremented for each successive interval.

   $A_1 = 1/24(y_{i-1} + 22y_1 + y_{i+1}) \qquad i=2, N-2$

   When the areas are added together, the contribution from each data point in the middle of the range is the same, but the contribution from the first and last few data points is not equal.

   $\text{Area(parabolic)} = 3/8(y_1 + y_N) + 7/6\ (y_2 + y_{N-1})$

   $+ 23/24\ (y_3 + y_{N-2}) + \sum_{i=4}^{N-3} y_1$

   The sum of the contributions from the six end-data points has the weight corresponding to an average over five data points. This follows that there are only N-1 intervals between data points.[1]

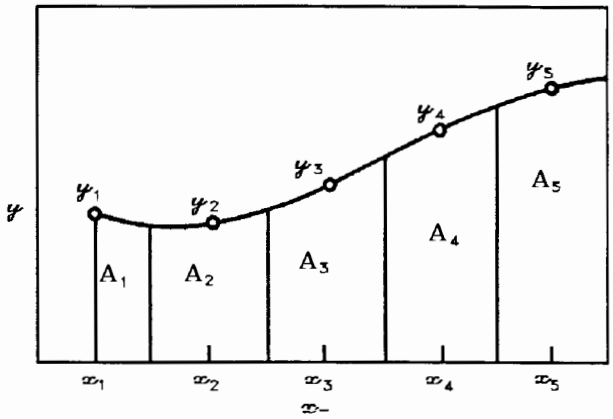

Figure 12-1. Area Integration by Parabolic Interpolation.

2. The area of the background is calculated using the formula for calculating the area of a trapezoid:

   $1/2\ [(b + e) * (n - 1)]$

   where:

   b=the number of counts for the beginning end point
   e=the number of counts for the ending end point
   n=total number of data points.

   The area of the background is then subtracted from the absolute area under the curve.

1 Phillip R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences* (McGraw-Hill Book Company, New York, 1969), pp. 268-270.

3. The area is normalized for time by dividing the area by the time/step (in seconds) used to acquire the data.

4. The area is normalized for the size of the unit step by multiplying the area by the step size. In ESCA applications the step size is expressed as the number of volts/step. The resulting normalized area is expressed in counts eV/second. For ISS acquired data, step size is expressed as a ratio of the kinetic energy of the electron to the ion beam energy (eV/E_o). The resulting normalized area is expressed as counts (eV/E_o)/seconds.

5. If ISS was the technique used to acquire the data, the area must be normalized for the ion energy. In ISS, the step size is expressed as a ratio of eV/E_o. To normalize for the ion energy, the area must be multiplied by the ion energy (E_o). The resulting normalized area is expressed in counts eV/second.

12.2 Atomic Concentration

The atomic concentration (AC) calculation provides a ratio of a single element to the sum of the other elements present in the acquired data. AC is expressed as a percentage in a tabular form and is based on either the height of the peak or the area under the peak. Table 12-1 illustrates a typical five region AC table. Each element is listed as a percentage of the whole.

TABLE 12-1. ATOMIC CONCENTRATION.

Element	Percent	Sensitivity Factor
O1	5.17	0.40
Tb1	18.63	0.10
Fe1	23.72	0.13
Co1	14.82	0.23
Si2	37.65	0.04

Peak height is calculated by taking the raw data in counts, performing a five-point smooth on the data, and performing a linear or integrated baseline subtraction. The resulting maximum number of counts is the peak height.

Peak area is calculated by taking the raw data in counts, performing a five-point smooth on the data, calculating the absolute area, calculating/subtracting the background area, and normalizing the area for step size.

The formula for calculating the atomic concentration percentage is given by the following equation.

$$\% \text{ AC for element x} = \frac{\frac{I_x}{S_x T_x}}{\sum_{i=1}^{n} \frac{I_i}{S_i T_i}} \times 100$$

where:

n=number of regions
i=peak height (or area) of the five-point smoothed, baseline-corrected, normalized data
S=peak height (or area) sensitivity factor*
T=total acquisition time per data point

*Sensitivity factors for both peak height and peak area are provided in Table 13-1.

12.3 Atomic Concentration Plot

An atomic concentration plot (ACP) provides a ratio of a single element to the sum of the other elements present in the acquired data. ACP is displayed as a graph illustrating the percentage of a particular element to the other elements as a function of sputter time and surface depth. ACP is based on either the height of the peak or the area under the peak, depending on how the data was acquired.

During depth profile data acquisition, peak height is calculated by taking the raw data in counts, performing a five-point smooth on the data, and doing a linear baseline subtraction. Peak area is calculated by taking the raw data in counts, performing a five-point smooth on the data, calculating the absolute area, calculating and subtracting the background area, and normalizing the area for step size.

The ACP function normalizes this data for time and calculates the percentages using the following formula:

$$\% \text{ AC for element } x = \frac{\frac{I_x}{S_x T_x}}{\sum_{i=1}^{n} \frac{I_i}{S_i T_i}} \times 100$$

where:

n=number of regions
i=peak height (or area) of the five-point smoothed, baseline-corrected, normalized data
S=peak height (or area) sensitivity factor*
T=total acquisition time per data point

*Sensitivity factors for both peak height and peak area are provided in Table 13-1.

12.4 Baseline Subtract

This Shirley Background Subtraction routine takes into account the scattering of low energy electrons. In the baseline subtract routine, the operator selects endpoints to define a baseline. The endpoints will then be averaged with data points on either side. These averaged values define the baseline. The baseline value is then calculated and subtracted at each data point. The resulting curve is shifted so the minimum data value is at zero intensity, or negative excursions can be truncated to zero.

12.5 Curve Fit

The information in this section is proprietary and is provided to help the operator understand the PHI curve fitting routine. It is not intended to be a complete description.

With the curve fitting routine, the operator can develop a synthetic peak structure that is composed of a number of peaks of a standard type, superimpose this peak structure on an experimentally obtained data set, and adjust the parameters defining the synthetic peak structure, either manually or by letting the computer perform a non-linear least squares fit, in order to simulate the experimental data. Energy positions and shifts can then be more readily determined, often enabling more accurate and more reliable chemical state identification.

Some of the key features of the PHI Curve Fit routine include the following:

1. The ability to select the peak type from three standard peak shapes: Gaussian, Gaussian-Lorentzian, or Asymmetric Gaussian-Lorentzian.

2. The ability to lock the peak separations and/or peak area ratios of two or more peaks together (Multiplet Locking).

3. The ability to limit the range over which each peak parameter will be allowed to change or to lock any peak parameter to a fixed value during the automatic curve fit.

4. The ability to perform either a linear, an integrated, or no background correction to the experimental data set.

5. The ability to alter the parameters defining the synthetic peak structure either manually through interactive menus or automatically by allowing the computer to perform a non-linear least squares fit.

These features will be defined in more detail in the following sections.

12.5.1 Standard Peak Shapes

The synthetic peak structure the operator defines is the composite of one to twelve separate peaks. Each peak can be one of three standard peak shapes; Gaussian, Gaussian-Lorentzian, or Asymmetric Gaussian-Lorentzian. The choice of peak shape applies to each of the bands of a given curve fit. It is important to note here that, although the Asymmetric Gaussian-Lorentzian combination could, in theory, be used for every curve fit, the speed of both the interactive setup and automated fit will be enhanced by using the simplest peak shape possible.

12.5.2 Peak Shape Parameters

The following is a summary of the parameters that define the different peak shapes:

1. Energy Position—The energy of the peak's center

2. Height—The height of the peak at its center

3. Full Width at Half Maximum (FWHM)—The width of the peak at one half the height. One half of FWHM is known as Half Width at Half Maximum.

4. Percent Gaussian—The mix of Gaussian and Lorentzian contributions to the peak shape. The range of this parameter is 0% to 100% with 0% consisting of a pure Lorentzian curve and 100% consisting of a pure Gaussian curve.

5. Tail Length—The distance in units of Half Width at Half Maximum (HWHM) at which the tail becomes insignificant (insignificant is defined as less than 1% of height).

6. Tail Scale—Used to increase or decrease the intensity of the tail shape which is defined by the Tail Length.

A Gaussian curve is defined by parameters 1, 2, and 3; a Gaussian-Lorentzian combination is defined by parameters 1, 2, 3, and 4; and an Asymmetric Gaussian-Lorentzian combination is defined by parameters 1, 2, 3, 4, 5, and 6.

12.5.3 Multiplet Locking

The operator may define certain bands to be dependent on other bands. There are two types of dependencies: position locking and area ratio locking.

Position Locking
With position locking, the dependent peak's position is anchored to the independent peak's position by an energy separation. As the independent peak's position is adjusted, the dependent peak's position will also change to maintain this constant separation.

Area Ratio Locking
With area ratio locking, the dependent peak's area will be maintained at a specified ratio to the independent peak's area. Since all the parameters except position are used in every area calculation, any of the other five parameters may be altered to maintain this constant area ratio.

Position locking and area ratio locking are entirely independent. The selection of which dependency to use is made on a case-by-case basis. Multiple dependencies may also be defined. For example, an independent band may have more than one dependent band associated with it. Also, a dependent band in one dependency may be used as the independent band in another dependency. In the following two cases, the arrows show dependence.

1. Peak 1 ← Peak 2
   Peak 1 ← Peak 3

2. Peak 1 ← Peak 2 ← Peak 3

12.5.4 Parameter Limits and Value Locking

Two additional constraints may be imposed upon each of the six parameters listed previously: range and invariance. First, the range over which the parameter is allowed to be adjusted is set by the operator. Therefore, the operator has the option on how tightly to constrain the parameter.

The second constraint is invariance. By setting the invariance, the parameter value is essentially locked to its estimated value. The invariance can be set for a selected number of iterations. After this number of invariant iterations is reached, the computer will allow the parameter to change during the optimization. If the number of iterations during which the parameter will be invariant is set equal to the maximum number of iterations, that parameter's value is essentially set for the entire curve fit.

12.5.5 Background Subtraction

In the PHI curve fitting procedure, the operator has the option of performing background subtraction on the experimental data in either the linear or integrated mode, or the operator may choose to perform no background subtraction. The window within the data set upon which a curve fit is to be performed may also be selected. Any background subtraction that is performed will be done within this window.

Linear
If linear background subtraction is selected, the computer draws a line between the left and right endpoints of the window. The area below this line (the background) will be subtracted.

Integrated
If integrated is chosen, the computer performs a right-to-left integration between the two endpoints of the window in order to form an integrated background curve. This new curve is then adjusted so that the amplitude of each endpoint corresponds to the amplitude of the corresponding data points in the original curve. The adjusted curve is then subtracted from the original curve. Only positive amplitudes of the final curve are recorded and displayed.

12.5.6 Manual and Automatic Parameter Adjustment

The operator can alter the parameters defining the synthetic peak structure either manually through interactive menus or automatically by allowing the computer to perform a non-linear least squares fit.

Manual Mode
Several options aid the operator in adjusting parameters. These options include an interactive display mode where the peaks and the composite structure will automatically be updated whenever a parameter is changed, cursors to visually display the parameters, and the ability to slew most of the parameters with either cursor control keys or the mouse.

Another key feature of the manual mode is that the computer will automatically adjust parameters to maintain the multiplet locking dependencies (energy separation and area ratio) described earlier in Multiplet Locking.

The remainder of this section will be devoted to describing the non-linear least squares algorithm used during the automated parameter optimization performed by the computer.

12.5.7 Non-linear Least Squares Optimization

Following is a description of the modified Gauss-Newton non-linear least squares optimization procedure used to perform the automated curve fitting in PHI's curve fit software. This description addresses the implementation of this procedure within the PHI software.[2]

To begin, a set of experimental data points Y_i (i = 1, . . ., n) have been obtained over some domain represented by the set of points X_i (i = 1, . . ., n). In this case, the domain is the electron binding energy. X_i represents the energy step within the spectrum, and Y_i is the electron intensity in step i. The data is then in the form of a list of n points: (X_i, Y_i) where i = 1, . . ., n.

2 For a description of the theoretical aspects of this procedure, please see "Applications of Stepwise Regression to Non-Linear Estimation" by R. I. Jennrich and P. F. Sampson, *Technometrics*, Vol. 10, No. 1, February 1968.

Gaussian Formula

$$B(X_i,Q) = H \exp\left(-\ln 2\left[\frac{2(X_i - PP)}{FWHM}\right]^2\right)$$

Gaussian-Lorentzian Formula

$$B(X_i,Q) = H\left\{\left\{PG \exp\left(-\ln 2\left[\frac{2(X_i - PP)}{FWHM}\right]^2\right)\right\} + \left(\frac{(1 - PG)}{\left[1 + \frac{2(X_i - PP)}{FWHM}\right]^2}\right)\right\}$$

Asymmetric Gaussian-Lorentzian Formula

$$B(X_i,Q) = H\left\{\left\{PG \exp\left(-\ln 2\left[\frac{2(X_i - PP)}{FWHM}\right]^2\right)\right\} + \left(\frac{(1 - PG)}{\left[1 + \frac{2(X_i - PP)}{FWHM}\right]^2}\right) + TAIL\right\}$$

where TAIL is defined as:

$$TS\left\{1 - \exp\left(-\ln 2\left[\frac{2(X_i - PP)}{FWHM}\right]^2\right)\right\} \exp\left(\left[\frac{-6.9}{TL}\right]\left[\frac{2(X_i - PP)}{FWHM}\right]\right)$$

for $X_i > PP$ (the high binding energy side of the peak)

and 0 for $X_i \leq PP$ (the low binding energy side of the peak).

For All Formulas:

$B(X_i,Q)$--the functional value at energy point X_i,

X_i--binding energy value for data point i,

PP--the binding energy of the peak's center,

H--the height of the peak at its center,

FWHM--the full width at half maximum of the peak,

PG--percent gaussian, (This parameter will actually lie in the range 0.0 to 1.0, where 0.0 is 0% and 1.0 is 100% Gaussian.)

TL--Tail Length in Half Width at Half Maximum, and

TS--Tail Scale.

Figure 12-2 General Curve Fit Formulas

A function $F(X_i, P)$ is defined to be fitted to this data. P is defined as a parameter set which will contain a number of parameters P_j (j=1, 2, . . . , m) that define a synthetic peak structure. The purpose of the program is to choose values for the parameters that give the best representation of the experimental data. "Best representation" is defined as the set of parameter values P_j which minimizes the error mean square (EMS) value in the following equation:

$$EMS = \frac{1}{(n - m)} \sum_{i=1}^{n} W_i [F(X_i,P) - Y_i]^2 \text{ where}$$

- EMS—error mean square,
- n—number of data points in the data set,
- m—number of parameters defining the synthetic peak structure,
- $F(X_i,P)$—the function as defined above,
- Y_i—the experimental data intensities, and
- W_i—a weighting factor. (Defined as $1 / Y_i$.)

Actually, the function $F(X_i,P)$ is of the form:

$$F(X_i,P) = \sum_{k=1}^{n_b} B_k(X_i,Q) \text{ where:}$$

- n_b—number of bands used to define the synthetic peak structure
- B_k—the function defining a single band or peak
- Q—the parameters used in the function to define the band

The parameter set P will therefore be the union of the parameter sets Q for each of the n_b bands.

The PHI curve fitting routine allows for three different forms of $B(X_i,Q)$:

- Gaussian
- Gaussian-Lorentzian
- Asymmetric Gaussian-Lorentzian

The general formulas for each of these functions are given in Figure 12-2. Both the Gaussian and Gaussian-Lorentzian formulas should sufficiently define these functions. The Asymmetric Gaussian-Lorentzian combination, however, will be further clarified in the following description.

12.5.8 Asymmetric Gaussian-Lorentzian function

The Asymmetric Gaussian-Lorentzian function is a combination of the regular Gaussian-Lorentzian function and an exponential tail function. This exponential tail function is only applied to the higher binding energy side of the peak, thereby giving the peak its asymmetric shape.

This exponential tail function, whose shape is determined by the Tail Length parameter, is the convolution of an exponential function and a Gaussian peak shape. It should be noted that a pure Gaussian shape is used in the determination of the tail function, even though the symmetric portion of this peak is a Gaussian-Lorentzian combination.

The shape of the exponential tail is determined by the Tail Length parameter. A scaling parameter, Tail Scale, is then applied to the tail to properly size the tail to the symmetric portion of the curve.

The area calculation of the Asymmetric Gaussian-Lorentzian peak must be determined numerically. The integral defining the area under the exponential tail cannot be solved analytically. Again, the symmetric portion of the peak is a standard Gaussian-Lorentzian peak. The area under this portion can be determined by integrating (analytically) the Gaussian-Lorentzian formula and evaluating this integral. The tail curve produced by a given Tail Length can be calculated over the window in which the tail is significant, then integrated numerically. In this case, a summation technique is used.

12.5.9 Optimization Algorithm

The following will describe the step-by-step procedure used by the PHI optimization algorithm. The part of the procedure where the dependencies (both peak location and peak area) are resolved will be noted. A full description of these dependency-locking steps will be given after the description of the general optimization algorithm is completed.

The following is a definition of terms that will be used in this description. Most were defined previously.

- (X_i,Y_i)—the set of experimental data points Y_i over the energy steps X_i,
- n—the number of data points in the above set,

- F(X,P)—the function that will be fit to the (X,Y) data set, (This function is defined in terms of the parameter set P.)
- P—the parameter set used to define F(X,P), (P is the union of the parameter sets Q, each of which define a single band or peak shape.)
- m—the number of parameters in the parameter set P,
- $\partial Y_i/\partial P_j$—the partial derivative of F(X,P) with respect to parameter P_j evaluated at X_i,
- EMS— $\frac{1}{(n - m)} \sum_{i=1}^{n} W_i [F(X_i,P) - Y_i]^2$, and
- W_i—a weighting factor. (Defined as $1 / Y_i$).

This routine obtains a weighted-least-squares fit of the function $F(X_i, P_j)$ to the data values (X_i, Y_i) by means of stepwise Gauss-Newton iterations of parameter P_j. Within each iteration, parameters are selected for modification in a stepwise manner. The parameter selected at a given step is the one that, differentially, makes the greatest reduction in the error sum of squares.

If necessary to avoid matrix singularity problems, only a subset of the parameters may be modified during a given iteration. In addition, parameters whose modification in a given iteration would lead to violations of the range defined for it are not modified. In effect, the iteration is performed "on the boundary" when it appears that the best fit occurs outside the parameter range.

Beginning with an initial set of parameter estimates for the parameter set P, the program minimizes the error mean square (EMS) by means of stepwise Gauss-Newton iterations. Each iteration will consist of the following steps:

Step 1
Let Z be the matrix defined as follows:

$Z_{ij} = \partial Y_i/\partial P_j$ for i=1, . . ., n and j = 1, . . ., m

$Z_{iq} = Y_i - F(X_i, P)$ for i = 1, . . ., n and q = m + 1

At this point the first stage of the peak separation dependency locking is performed. This will be described in more detail later. At this point, the invariance is also applied to any parameters that will be held constant during this iteration. Assuming parameter P_r is to be held constant, the invariance is accomplished by setting $Z_{ir} = 0$ for i = 1, . . ., n.

The matrix A is then formed as follows:

$$A_{kl} = \sum_{i=1}^{n} Z_{ik} Z_{il} W_i$$

for k,l = 1, . . ., q. The error mean square (EMS) is then determined as $EMS = A_{qq}/(n - m)$. Matrix A is then normalized.

Step 2.1
The diagonal elements of matrix A are pivoted on in a stepwise manner. At each step, index p of the pivot element maximizes A_{pq}^2/A_{pp} for all p satisfying the following:

$1 \le p \le m$ and p has not previously been used as a pivot index.

A_{pp}/A'_{pp} T where T is the tolerance specified, and $A'_{pp} = A_{pp}$ before any pivoting began. For PHI applications, T=0.00001.

If the tolerance drops below T, some diagonal elements in addition to A_{qq} may not be used as pivots. This pivoting continues until no index p is found that satisfies both of the above conditions.

Step 2.2
Next, pivots corresponding to boundary values and resulting in boundary violations are undone. Let LL_p and UL_p denote the lower and upper limits of the range defined for parameter P_p. Next, the value of the index p is searched over all p that have been pivoted and have not been unpivoted, which minimizes the value R_p defined as follows:

$R_p = \text{maximum}[(LL_p - P_p)/A_{pq} , (UL_p - P_p)/A_{pq}]$

If $R_p \le 10^{-6}$, then matrix A is unpivoted around its p^{th} diagonal element. After unpivoting, return to step 2.1 to search for additional pivot operations that may have become valid due to the unpivot just performed. This process of pivoting (Step 2.1) and unpivoting (Step 2.2) is continued until $R_p > 10^{-6}$. At this time, there are no indexes upon which either pivoted or unpivoted operations are valid.

The value S is then defined as follows:

S = minimum [1, the final value of R_p from Step 2.2]

Step 3
The second stage of the peak separation dependency locking is then performed at this point. A detailed description of this procedure is described in the Dependency Locking section.

Step 3.1
New values for the parameter set P are then calculated using the final values of the matrix A and the value S (determined in step 2.2), which is used as a scaling factor. The parameter set P is updated as follows:

$P_j = P_j + S'A_{jq}$ for j's used as a pivot index and not used as an unpivot index.

or

$P_j = P_j$ for j's not used as a pivot index or that were unpivoted.

These new parameter values, P_j, are then adjusted to resolve any peak area dependencies that might be present. A detailed description of this procedure will come later.

Step 3.2
A new value of the error mean square (EMS) is then calculated. If this new EMS is larger than the EMS computed in step 1, a new scaling factor S is computed as S = S / 4. Next, return to Step 3.1.

This "step reduction" is continued until the value of the new EMS is less than that computed in step 1 or up to a maximum of three times.

Step 4
A check is now made to see if the routine has reached convergence. There are two possible methods by which convergence can be reached. First, if three attempts of the "step reduction" (described in Step 3.2) were tried, and the value of the new EMS is still greater than or equal to the value obtained in step 1, the routine will be considered to have converged to the "best fit".

Second, if a new smaller EMS was successfully computed, the relative change in the EMS is calculated as follows:

$$\text{Relative Change} = \frac{|\text{EMS(Step 3.2)} - \text{EMS(Step 1)}|}{\text{EMS(Step3.2)}}$$

This Relative Change is then compared to the convergence criterion, which is defined as 0.001 for the PHI curve fit routine. There are two possible cases that could occur as follows:

If the Relative Change ≥ 0.001, then the routine will loop back to Step 1, and the process is repeated.

If the Relative Change < 0.001 for the first time, then the routine will loop back to Step 1, and another iteration is performed. If the Relative Change < 0.001 for the second time in a row, then the routine will be considered to have converged to the "best fit". The Relative Change must be < 0.001 twice in a row to ensure that the routine has not converged to a local minimum.

The routine will also terminate without convergence if the maximum number of iterations allowed are executed. Finally, the operator has the option of stopping the routine at the end of the current iteration if desired.

Upon termination of the curve fit, the system will display the final error mean square value, the cause of termination, the number of parameters that were constrained due to boundary conditions, and the number of area ratio dependencies that were violated (this will be explained in the following section).

12.5.10 Dependency Locking

Peak location and peak area ratio dependency locking will now be described. The two types of dependencies are different and are handled at different points in the optimization routine as could be seen in the description of this optimization in the preceding section. Therefore, each will be described separately.

It should be noted that both dependencies introduce additional constraints into the optimization process. This will probably produce a fit which is not as optimal as that which would be produced in an unconstrained system.

Peak Separation Dependency Locking

This locking procedure involves two stages. The first occurs at Step 1 in the optimization algorithm where the Z matrix of partial derivatives is calculated. Here the partial derivative for the dependent peak's location parameter is added to the partial derivative for the independent peak's location parameter. The dependent peak's partial derivative is then set to zero. For this example, assume P_d is the dependent band's peak location parameter, and P_e is the independent band's peak location parameter. This first step is defined as:

$Z_{ie} = Z_{ie} + Z_{id}$ for i = 1, . . ., n and
$Z_{id} > 0$ for i = 1, . . ., n.

The second stage occurs during Step 3 of the optimization routine. Here the value of the matrix element A_{dq}, which will be used to calculate the new parameter value for parameter P_d, is assigned the value of A_{eq}, that is, $A_{dq} = A_{eq}$ for each dependency defined in the system.

Peak Area Ratio Dependency Locking

Peak area ratio dependency locking is accomplished during Step 3.1 of the optimization routine. For each peak area ratio dependency that is defined, an iterative search is carried out to find the parameters that when adjusted will cause the areas of the dependent and independent peaks to be in the proper ratio while having the least impact on the EMS.

For each peak shape type: Gaussian, Gaussian-Lorentzian, and Asymmetric Gaussian-Lorentzian, all parameters, with the exception of the peak location, are parameters in the peak area calculation. Therefore all parameters except peak location are possible candidates for adjustment when attempting to resolve the dependent and independent areas to the proper ratio. However, no parameter will be adjusted so that the range defined for that parameter is violated.

The case can arise that all parameters are disregarded due to boundary violations. This can arise if the limits on the parameters for the dependent peak are tight. In general, either tight limits or area ratio locking should be used but not both. At least some of the parameters involved in the area calculations must have fairly loose limits defined. However, if the case does arise that all parameters are disregarded due to limit violations, the routine then attempts to find the parameter that when adjusted will produce the closest match to the given area ratio. A note is made of this ratio violation, and the operator will be informed.

12.5.11 Special Considerations

Under most circumstances the PHI curve fit routine displays very good convergence behavior. However, the algorithm is not guaranteed to find the absolute minimum least squares solution. Rather, starting from the estimates for the parameters that the operator provides, the algorithm "searches" the least squares solution space for a minimum. It is possible that this "search" may converge to a local minimum rather than the absolute minimum desired.

If the algorithm appears to be displaying the behavior described above, the following steps can be taken to help prevent this situation from occurring or to correct it if it does occur:

1. Minimize the number of parameters present in the system. As the number of parameters that define the solution space increases, so does the probability of the occurrence of local minimums. Reducing the number of parameters can be accomplished by:

   a) Using a simpler peak shape. For example, using a Gaussian rather than an Asymmetric Gaussian-Lorentzian combination will reduce the parameter count from six to three per peak.

   b) Reduce the number of peaks used to define the synthetic peak shape. An "over relaxed" fit that tends to converge to unsatisfactory local minimums is often an indication that too many peaks are being used.

2. Select proper initial estimates for the parameters. Starting from a reasonably good fit will help increase the probability of convergence to a "good" fit. Also, if an unsatisfactory fit is obtained, selecting a different set of initial estimates and performing the fit again will generally resolve the situation. Choosing this new set of initial estimates to be on the "other side" of the solution sought is very useful. For example, if the old initial estimates produced a synthetic curve that generally lies below the experimental data, select new estimates that produce a synthetic curve that lies above the experimental data.

As noted above, poor convergence behavior may or may not be encountered during any particular curve fit session. Since the general convergence behavior of the system is good, care should be taken not to spend time producing a "perfect" manual fit rather than letting the computer perform the optimization unless poor convergence is encountered.

12.6 Deconvolution

The PHI X-Ray Line Deconvolution Routine is based on a deconvolution technique known as van Cittert's Method.[3] It is an iterative method that has been shown to mathematically approximate a Fourier transform deconvolution technique. When applied to experimental ESCA data, it functions to reduce the line width by reducing or removing the contribution of the x-ray source line width present when exciting the specimen. Deconvolution sharpens the peaks, making it easier to identify the exact energy of photoelectron peaks and to detect and identify shoulders on the peak.

Since this technique also serves to deconvolute any noise present in the data, optimum results are obtained by signal averaging the raw data before using the deconvolution routine, and smoothing the deconvoluted data (generally with an 11 point smooth) once it has been processed. The raw ESCA data used should be represented by steps that are no greater the 0.5 eV. Otherwise the steps will be too widely spaced and too much data extrapolation will be required to obtain satisfactory results.

Deconvolution is performed by passing the data spectrum through a mathematical filter representing the inverse of the x-ray line shape and the electron spectrometer line shape. Since the x-ray line shape is fixed for a given anode, it is more convenient to filter out the x-ray line, leaving the electron spectrometer line shape without the contributing effects of the x-ray line.

3 H. C. Burger and P. H. van Cittert, *Z. Phys.*, **79** (1932) 722.

Deconvolution is performed using a model which is 63 data points wide. The formula for the model is:

$$Ci=[1/(\text{energy/Lorentzian line width})^2 + 1] + [2/(\text{energy separations/Lorentzian line width})^2 + 1]$$

where:

energy=center mass + (i*volts/step)
Lorentzian line width=0.28
Mg center mass=0.1357177
Al center mass=0.12
Mg separation=0.2
Al separation=0.43

These constants are then normalized (divided by) the summation of all constants:

$$Ci' = \frac{Ci}{\sum_{j=-31}^{31} C_j}$$

where:

C_i=constant for the i^{th} point in the model
i=any given point in the model
C_i'=normalized constant

The constants obtained with this model are then used in the following way. Two data sets are established: the raw data set and the working data set. During the first iteration, the raw data set is equal to the working data set.

For each data point in the working data set (moving right to left) the constants from the model are convoluted with the surrounding data set (63 points) to obtain a new value for the data point. Near the end points, the end data value is used as the value for all missing data points.

This new data value is subtracted from the corresponding data point in the raw data set. This difference is then added to the data point in the working data set.

In this operation, the working data set is altered even while it is being used for the convolution. At each data point, the 32 leftmost data points used in the convolution are from the working data set of the previous iteration (or the raw data set in the first iteration) while the 31 rightmost data points are the already altered points from the current iteration.

Each new iteration is based on the shape of the preceding iteration and is a closer approximation of the specimen's characteristic peaks. The sequence may be continued until the difference in iterations is smaller than the resolution of the system. This, however, results in the use of excess computer time, and there will appear to be a continual increase in the relative amplitude of any noise contained in the original data. Such noise can be later reduced with the smooth routine.

This sequence may be continued until the difference in iterations is smaller than the resolution of the system. Typical iteration values are 1, 2, or 3. The final data will not be distorted should the sequence include more repetitions than that required for the maximum precision of the system.

For more information on deconvolution, refer to: "Computer Resolution Enhancement for Photoelectron Spectroscopy," John J. Uebbing, Varian Associates, October 17, 1972.

12.7 Differentiation

The differentiation routine is used as a means of suppressing the background in cases in which the peak-to-background ratio is small. In PHI systems, differentiation is accomplished using the Savitzky-Golay convolution algorithm.

The concept of convolution is an expanded calculation of a moving average, in which a convolution function involving an odd number of data points is moved along the data curve. The data point being modified is at the center of the window. The new data value is the resultant when the values in the window are multiplied by the corresponding convolution coefficients (Table 12-2). The coefficients are derived by a method known as least squares.

NOTE

In general, the smaller the number of points , the sharper the differentiated peak shape and the greater the resultant noise. The larger the number of points, the broader the differentiated peak shape, but the noise level will be reduced. Usually, a number between 7 and 13 will provide good peak definition and a reasonable noise level.

For example, if the operator selects a five point differentiation, the third data point is the center point. Two data points before and two points after the center point will be used in determining the weighted average of the center point. Each data point value will be multiplied by its corresponding convolution coefficient from Table 12-2. These results are then summed and divided by the appropriate normalizing factor. When the calculation is complete, the five point "window" will be advanced by one and the process repeated.

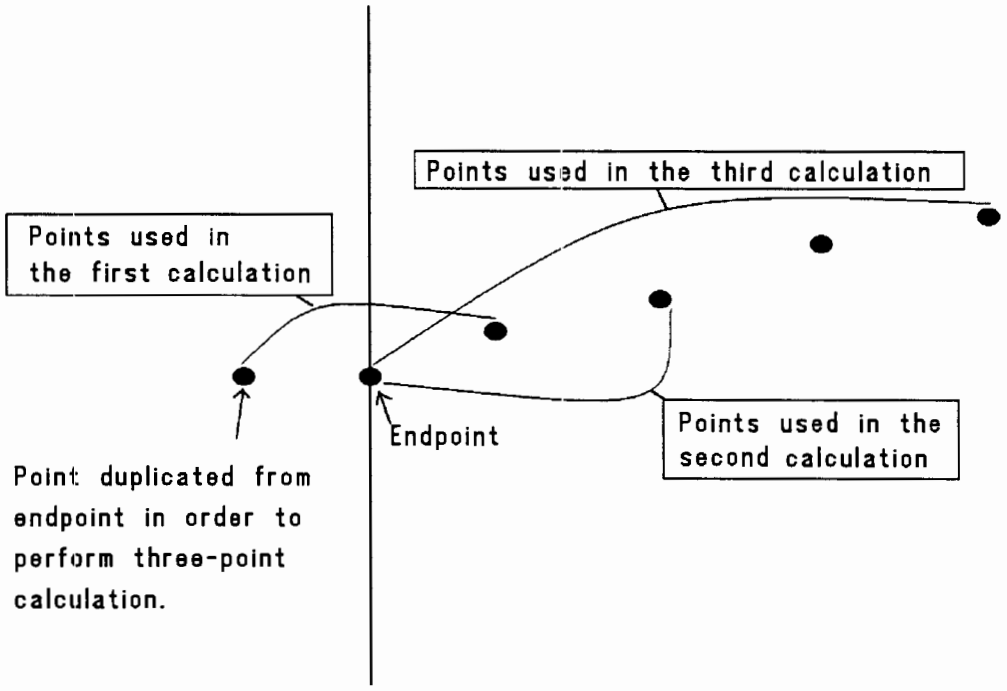

Figure 12-3

Because there are data points on only one side of each endpoint, the differentiation process changes. Only three points are used to differentiate an endpoint and its neighboring point, even if a higher number is specified. To calculate the differentiation for the endpoint, the neighboring point, the endpoint, and a duplicate of the endpoint are used. For the neighboring point calculation, the endpoint, the neighboring point, and the third point are used. For each subsequent data point, the number of points used in the differentiation routine is increased by two until the number entered for **a N-Point Differentiate** *is reached. This is true for points at each end of the spectrum or region.*

TABLE 12-2. CONVOLUTING COEFFICIENTS

POINTS	25	23	21	19	17	15	13	11	9	7	5	3
-12	-12											
-11	-11	-11										
-10	-10	-10	-10									
-09	-9	-9	-9	-9								
-08	-8	-8	-8	-8	-8							
-07	-7	-7	-7	-7	-7	-7						
-06	-6	-6	-6	-6	-6	-6	-6					
-05	-5	-5	-5	-5	-5	-5	-5	-5				
-04	-4	-4	-4	-4	-4	-4	-4	-4	-4			
-03	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3		
-02	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	
-01	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
00	0	0	0	0	0	0	0	0	0	0	0	0
01	1	1	1	1	1	1	1	1	1	1	1	1
02	2	2	2	2	2	2	2	2	2	2	2	
03	3	3	3	3	3	3	3	3	3	3		
04	4	4	4	4	4	4	4	4	4			
05	5	5	5	5	5	5	5	5				
06	6	6	6	6	6	6	6					
07	7	7	7	7	7	7						
08	8	8	8	8	8							
09	9	9	9	9								
10	10	10	10									
11	11	11										
12	12											
NORM	1300	1012	770	570	408	280	182	110	60	28	10	2

The mathematical description of this process is:

$$Y_j^* = \frac{\sum_{i=-m}^{i=m} C_i Y_j + 1}{N}$$

where:

j=running index of the ordinate data in the original data table

Y_j=original data point

Y_j^*=recalculated data value

m=(2n-1)/2, where n is the number of points specified by the operator

N=normalization factor

C=convoluting integer from Table 12-2.

12.8 Height/FWHM

This routine is used to determine the peak height between a portion of a curve and a defined baseline. Included in the routine are calculations of the FWHM for the peak, the signal to background ratio, and the signal to noise ratio.

- The peak is determined by subtracting the defined baseline counts from the actual counts of the displayed curve.

- The FWHM is determined by finding the first occurrence on both sides of the peak energy of a counts value that is less than, or equal to, half of the peak counts. The exact energy corresponding to the half maximum is then determined by interpolating between the surrounding energies. The full width is then the difference between the energies of these occurrences.

- The background is the average of all points that fall within the specified energy range.

- The noise is calculated by the root mean square method:

  where N is the total number of points in the specified energy range, y_i is the actual value of each point, and $\bar{y}_i$ is the value of a second order polynomial fit.

$$\text{Noise} = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2}{(N - 1)}}$$

- The signal is the total counts that occur at the peak location, including background.

12.9 Normalize

The normalization routine operates in either of the two modes described below.

One-Point Normalization

In a one-point normalization, the data curve designated as "B" is rescaled so the amplitudes of data curves A and B are equal at the energy specified by the operator.

$B_n = B_i(A_p/B_p)$

B_n is the normalized value of data point B_i.

A_p and B_p are the value of data curve A and the original value of data curve B, respectively, at the reference energy.

Three-Point Normalization

In a three-point normalization, the data curve designated as "B" is rescaled so the difference in amplitude between the two energies specified by the operator is the same for data curves A and B.

$$B_n = (B_i - B_b)\frac{(A_p - A_b)}{(B_p - B_b)} + E_b$$

B_n is the normalized value of data point B_i.

A_p and B_p are the peak reference energy data values for curves A and B, respectively.

A_b and B_b are the baseline reference energy data values for curves A and B, respectively.

E_b is a baseline reference value used to offset B to A.

After normalization, the computer prints a normalization factor, which is either the ratio A_p/A_b or the ratio $(A_p-A_b)/(B_p-B_b)$.

12.10 Normalize by E

The normalize energy routine is used to compensate for the effect of the analyzer transmission function on the measured signal. A cylindrical mirror analyzer operated in non-retarding mode has a transmission function N(E)*E. This routine divides each data value by its corresponding energy to produce an N(E) curve.

12.11 Smoothing

Smoothing reduces the noise evident in a spectrum, improving the visual presentation of the data and easing data interpretation. In the PHI 5000 Series Systems, smoothing is accomplished using the Savitzky-Golay convolution algorithm.[4]

The concept of smoothing is an expanded calculation of a moving average. Refer to Figure 12-4, which shows a five-point smooth. The C's represent any set of integers. There is an associated normalizing or scaling factor. The procedure is to multiply C_{-2} times the number opposite it, the C_{-1} by its number, etc., sum the results, and divide by the normalizing factor, if appropriate. The result is the desired function evaluated at the point indicated by C_0.

1800.0			705
1799.8			712
1799.6			717
1799.4			718
1799.2			721
1799.0			722
1798.8	x_o^{-2}	C_{-2}	725
1798.6	x_o^{-1}	C_{-1}	730
1798.4	x_o	C_0	735
1798.2	x_o^{+1}	C_1	736
1798.0	x_o^{+2}	C_2	741
1797.8			746
1797.6			750

Figure 12-4. Smoothing Operation. (Abscissa points at left, tabular data at right. In box are the convolution integers, Ci).

For the next point, the set of integers is moved down, and the process is repeated. The mathematical description of this process is:

$$Y_j^* = \frac{\sum_{i=-m}^{i=m} C_i Y_j + 1}{N}$$

where:

j=running index of the ordinate data in the original data table
Y_j=original data point
Y_j^*=smoothed data point
N=normalization factor
C=smoothing integer from Table 12-3.
m=(number of point smooth - 1) / 2

For more information refer to the technical article listed.

NOTE

Refer to Figure 12-5. Because there are data points on only one side of each endpoint, the smooth process changes. Only three points are used to smooth an endpoint and its neighboring point, even if a higher number is specified. To calculate the smooth for the endpoint, the neighboring point, the endpoint, and a duplicate of the endpoint are used. For the neighboring point calculation, the endpoint, the neighboring point, and the third point are used. For each subsequent data point, the number of points used in the differentiation routine is increased by two until the number entered for **a N-Point Differentiate** *is reached. This is true for points at each end of the spectrum or region.*

4 Abraham Savitzky and Marcel J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Analytical Chemistry*, Vol. 36, No. 8 (July 1964), pp. 1627-1639.

TABLE 12-3. SMOOTHING COEFFICIENTS.

POINTS	25	23	21	19	17	15	13	11	9	7	5	3
-12	-253											
-11	-138	-42										
-10	-33	-21	-171									
-09	62	-2	-76	-136								
-08	147	15	9	-51	-21							
-07	222	30	84	24	-6	-78						
-06	287	43	149	89	7	-13	-11					
-05	342	54	204	144	18	42	0	-36				
-04	387	63	249	189	27	87	9	9	-21			
-03	422	70	284	224	34	122	16	44	14	-2		
-02	447	75	309	249	39	147	21	69	39	3	-3	
-01	462	78	324	264	42	162	24	84	54	6	12	1
00	467	79	329	269	43	167	25	89	59	7	17	2
01	462	78	324	264	42	162	24	84	54	6	12	1
02	447	75	309	249	39	147	21	69	39	3	-3	
03	422	70	284	224	34	122	16	44	14	-2		
04	387	63	249	189	27	87	9	9	-21			
05	342	54	204	144	18	42	0	-36				
06	287	43	149	89	7	-13	-11					
07	222	30	84	24	-6	-78						
08	147	15	9	-51	-21							
09	62	-2	-76	-136								
10	-33	-21	-171									
11	-138	-42										
12	-253											
NORM	5175	805	3059	2261	323	1105	143	429	231	21	35	4

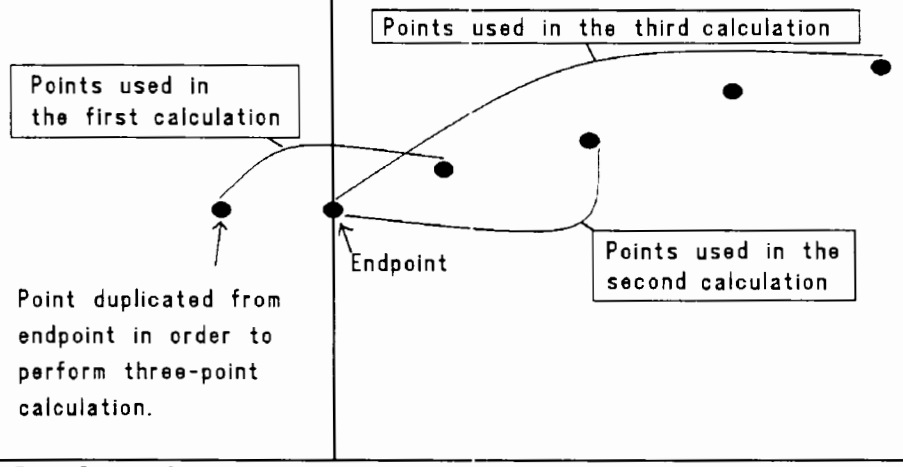

Figure 12-5. See NOTE on facing page.

12.12 X-Ray Satellite Subtraction

The X-Ray Satellite Subtraction routine is used to correct for the effects of x-ray source satellite lines on the measured signals. The routine will correct data measured with Al or Mg anodes.

The satellites are subtracted using the following algorithm:

$D(n)=D(n) - f_3 * D(n-dE_3/s) - f_4 * D(n-dE_4/s) - f_5 * D(n-dE_5/s) - f_6 * D(n-dE_6/s) - f_B * D(n-dE_B/s)$

where:

f_3, f_4, f_5, f_6, f_B are relative height factors for each satellite.

dE_3, dE_4, dE_5, dE_6, dE_B are energy displacements for each satellite.

s is the energy step size used for the acquisition.

D(n) is the intensity at the nth data point, which is associated with an energy equal to the starting energy - n*s.

The relative height factors (%) and energy displacements (eV) are:

Mg anode
$f_3 = 8.0$, $f_4 = 4.1$, $f_5 = 0.55$, $f_6 = 0.45$, $f_B = 0.5$

$dE_3 = 8.4$, $dE_4 = 10.2$, $dE_5 = 17.5$, $dE_6 = 20.0$, $dE_B = 48.5$

Al anode
$f_3 = 6.4$, $f_4 = 3.2$, $f_5 = 0.4$, $f_6 = 0.3$, $f_B = 0.55$

$dE_3 = 9.8$, $dE_4 = 11.8$, $dE_5 = 20.1$, $dE_6 = 23.4$, $dE_B = 69.7$

If the energy displacement for a satellite is not an integral step size, the intensity at that displacement is determined by linear interpolation of the intensities of the two immediate surrounding data points.

If the energy displacement falls outside the energy window over which data was acquired, its intensity is set to that of the first data point after it has had all satellites subtracted. Thus the highest binding energy intensities (those within dE_B, which is 48.5 eV for Mg and 69.7 eV for Al) will be affected by the "satellites" associated with the highest binding energy data point's intensity.