

KNOWLEDGE-DRIVEN INTERPRETATION OF ESCA SPECTRA

Masato Yamazaki

Automatic Control Division
Electrotechnical Laboratory
Nagata-cho, Tokyo 100, Japan

Hideo Ihara

Materials Division
Electrotechnical Laboratory
Tanashi-shi, Tokyo 188, Japan

A system for automatically interpreting raw spectral data of an ESCA (Electron Spectroscopy for Chemical Analysis) is described. This system can remove background noise from the spectrum, identify each spectral peak, assign each peak to a corresponding atomic orbital and can obtain the chemical composition of a sample. A knowledge-driven method enables the system to identify very small peaks and rider peaks, and to obtain the most reliable assignment to each peak on the basis of specific chemical knowledge. Utilization of knowledge in low level processing is our main interest, while higher-level logical inference is emphasized in other knowledge-based systems, for example DENDRAL and MYCIN.

1. INTRODUCTION

An ESCA spectrometer is an important equipment which gives much information about chemical composition, bond nature and electronic structure of a sample. Considerable efforts have been done for removing large background noise[1] and for analysing bond nature from this spectrum. In this version of our system we intend to perform the composition analysis by a wide-range ESCA spectrum. It would be the first attempt to deal directly with the wide-range spectrum and to obtain an extensive result automatically.

Chemical knowledge is effectively utilized for identifying a peak and giving rational assignment to each peak. This paper briefly explains the automatic interpretation system and demonstrates an experimental result.

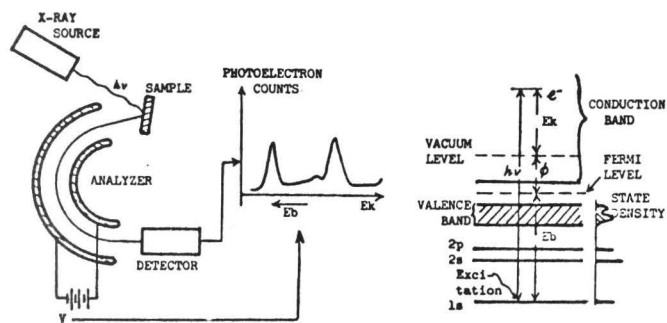


Fig.1 ESCA spectrometer and the photoelectric effect

2. ESCA SPECTROMETER

The ESCA spectrometer is illustrated in Fig.1. A monochromatic X-rays ($E_x = h\nu = 1486.6\text{eV}$) is used to irradiate a sample in a vacuum. The ejected photoelectrons have a variety of kinetic energy, depending on which element and which level they come from. The binding energy E_b relative to the Fermi level is calculated by the following equation:

$$E_b = h\nu - E_k - \phi \quad (1)$$

where E_k is the measured kinetic energy and ϕ is the work function of a sample.

Both an element and orbitals can be determined by the particular positions of a set of peaks. Relative quantity of each element can be calculated by measured peak area, or the integrated intensity, and theoretical cross section value.

3. INTERPRETATION PROCESS

3.1 Outline

Fig.2 shows in outline the identification and assignment process. Firstly, in the preprocessing stage, background noise of low frequency is removed. Secondly, in the knowledge-driven identification stage, obvious peaks and peaks related to them are identified, then peak-assignment pairs (P-A pairs) are made from these peaks and the corresponding possible assignments. This stage is carried out efficiently by making use of chemical knowledge. Thirdly, in the bottom-up identification stage, all the

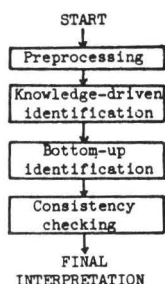


Fig. 2. Outline of the process

remaining peaks that are not relevant and are not related to other peaks are identified by a conventional statistical test. Finally, in the consistency checking stage, the Reliability and other properties of the P-A pairs are determined and the final interpretation is obtained.

3.2 Knowledge used

The chemical knowledge which will be used in the following sections consists of the expected binding energy and cross section values[2]. The former corresponds to peak position, and the latter to the peak area. These values are stored in ESCATBL as follows:

$$\text{ESCATBL} = (\text{KNLG}_1, \text{KNLG}_2, \dots, \text{KNLG}_n) \quad (2)$$

$$\text{KNLG}_i = (\text{ASS}_{i1}, \text{ASS}_{i2}, \dots, \text{ASS}_{im}) \quad (3)$$

$$\text{ASS}_{ij} = (\text{ATOM}_i, \text{LEVEL}_{ij}, \text{EBE}_{ij}, \text{CRS}_{ij}) \quad (4)$$

In this formula, EBE_{ij} , the expected binding energy, and CRS_{ij} , the theoretical cross section value, are represented for each LEVEL_{ij} of ATOM_i . Fig.3 shows part of this knowledge table.

The CRS_{ij} are normalized by setting the value of Carbon 1s to 1.000. Let S_{ai} be that of level i of the element A, and let S_{bj} be that of level j of the element B. The composition ratio is calculated as

$$A : B = S_{a1}/\text{CRS}_{a1} : S_{b1}/\text{CRS}_{b1} \quad (5)$$

```

ESCATBL (
  (H E1S 13.595 0.0002)
  (HE E1S 24.580 0.0082)
  :
  (O E2P3/2 13.614 0.0128)
  (O E2P1/2 7.0 0.0065)
  (O E2S 24.0 0.1405)
  (O E1S 532.0 2.93)
  :
  (PB E6P3/2 1 0.0291)
  (PB E6P1/2 1 0.0148)
  :
  (PB E4P1/2 764 2.12)
)
  
```

Fig.3 Segment of the knowledge

3.3 Knowledge-driven identification stage

Obvious peaks and peaks related to them are identified, and possible P-A pairs are obtained as the output of this stage. The knowledge-driven method used here consists of the following four steps (see Fig.4):

STEP1: The process begins by finding large peaks. A list HPEAKS of large peaks that are higher than a threshold TH is constructed by scanning through the preprocessed data DP

$$\text{HPEAKS} = (\text{PEAK}_1, \dots, \text{PEAK}_n) \quad (6)$$

This TH is determined by an operator for terminating this stage.

STEP2: The highest peak is selected from the HPEAKS. All assignments which are consistent with the existence of this peak are chosen, and P-A pairs are made and stored in PALIST. A peak k is characterized by a triplet as

$$\text{PEAK}_k = (\text{HGT}_k, \text{POS}_k, \text{BE}_k) \quad (7)$$

where HGT_k means peak height, POS_k peak position, and BE_k the binding energy value corresponding to this position. From Eqs.(4) and (7), a P-A pair is represented as follows:

$$\begin{aligned} \text{P-A} &= (\text{PEAK}_k, \text{ASS}_{ij}) \\ &= ((\text{HGT}_k, \text{POS}_k, \text{BE}_k), (\text{ATOM}_i, \text{LEVEL}_{ij}, \text{CRS}_{ij})) \end{aligned} \quad (8)$$

Comparing BEs with EBEs, matching tolerance should be determined by estimated instrumental error and the possible chemical shift.

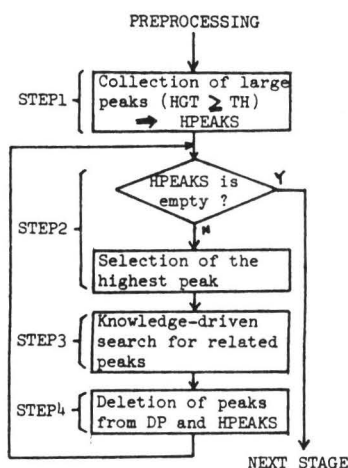


Fig. 4. Knowledge-driven identification stage

STEP3: For every candidate element the related peaks are knowledge-drivenly searched for, and all possible assignments are chosen. The P-A pairs are then made and added to the PALIST.

STEP4: The process deletes all peaks mentioned above from HPEAKS and DP. It then returns to STEP2, and repeats STEP2 through STEP4 until there remains no peak higher than the TH.

3.4 Consistency Checking stage

We explain three typical checking routines.

(1) Reliability test for P-A pairs: Let A be an element, suppose it has m levels and $CRS_1, \dots, CRS_m$ be the cross section values, reliability factor RF is then defined as

$$RF = \frac{\sum_{i=1}^m \alpha_i CRS_{ai}}{\sum_{i=1}^m CRS_{ai}} \quad (9)$$

where $\alpha_i = \begin{cases} 1, & \text{if such P-A pair is found} \\ -2, & \text{if such P-A pair is not found} \end{cases}$

Only reliable set of P-A pairs are selected based on this criterion.

(2) Calculation of the composition formula: Estimating peak areas on the original spectrum, calculate Eq.(5).

(3) Test of chemical shift: The chemical shift, a small deviation of a peak position, should be approximately equal for the same element and the same chemical state. So, if only a few shifts are clearly different from others under these conditions, a rational explanation of the fact would be that the identification stage has failed to find a correct peak and has found instead neighbouring one. There are two cases; 1) the correct peak is too small to identify and 2) the correct peak is a rider peak.

4. EXAMPLE

Fig.5 shows a 1024 point spectrum of a super conductive materials Pb Mo S. In the pre-processing stage step-like background noise was removed (Fig.6). Next, knowledge-driven identification began by finding the highest peak 'a'!

```
RF = 0.93
((565 857 163.1) ( S E2p3/2 165 1.11)) - g
((565 857 163.1) ( S E2p1/2 165 0.567)) - g
RF = 0.86
((158 479 532.2) ( O E1s 532.0 2.93)) - h
RF = 0.93
((88 987 36.1) ( MO E4p3/2 35 0.86)) - f
((88 987 36.1) ( MO E4p1/2 35 0.445)) - f
((1003 789 229.5) ( MO E3d5/2 227 5.62)) - a
((491 786 232.4) ( MO E3d3/2 230 3.88)) - b
((302 620 394.5) ( MO E3p3/2 392 5.94)) - c
((162 602 412.1) ( MO E3p1/2 410 3.04)) - c
((53 505 506.8) ( MO E3s 505 2.34)) - e
RF = 0.68
((61 1004 19.5) ( PB E5d5/2 19 1.58)) - i
((32 1002 21.5) ( PB E5d3/2 22 1.11)) - j
((643 882 138.6) ( PB E4f7/2 138 12.73)) - k
((432 877 143.5) ( PB E4f5/2 143 10.01)) - l
((162 602 412.1) ( PB E4d5/2 413 13.02)) - d
((98 578 435.5) ( PB E4d3/2 435 8.87)) - m
((42 363 645.5) ( PB E4p3/2 645 6.33)) - n
```

Pb : Mo : S = 1 : 7.4 : 9.5

Fig.7 Result obtained

Candidate assignments were selected as Sulfur 2s, Molybdenum 4d3/2, and Molybdenum 4d5/2. In STEP3 other peaks of Sulfur and Molybdenum were searched for, all the possible assignments were chosen for these peaks, and P-A pairs were made. Thirty-three pairs were produced on 14 peaks in this example. Reliability factor was then calculated for each element. S, Mo, Pb and O were verified and Ti, W, Ne, C and F were rejected. Final assignments and the composition ratio are shown in Fig.7. Difference of about 20% from the correct composition ratio (= 1:6:8) is thought to be due to inaccuracy of evaluated peak areas.

5. CONCLUSION

We have presented a system for interpreting ESCA spectra on the basis of chemical knowledge. Core-level peaks are identified and given rational assignments. The chemical composition of a sample is obtained. An example have shown effectiveness of our method in some difficult cases, for example, in finding a small peak, detecting a rider peak and choosing the most reasonable assignment from many alternatives.

REFERENCES

- [1] J.N. Chazalviel, M. Campagna, G.K.Wertheim, and P.H. Schmidt, Physical Review B, Vol.14, No.10, 4586, Nov. 1976
- [2] J.H. Scofield, J. of Elec. Spec. Rel. Phen., No.8, pp.129, 1976

