

# スペクトルデータの自動解釈プログラム

Automatic interpretation of ESCA spectra

山崎 正人 (電総研)

by Masato YAMAZAKI (Electrotechnical Laboratory)

## ABSTRACT

In this report, we firstly present a program that automatically interprets an ESCA (Electron Spectroscopy for Chemical Analysis) spectrum. Then we give some considerations toward the advanced version of this program.

The interpretation process of the program is carried out effectively because it uses top-down or knowledge-driven method. In the advanced version, we will intend to obtain more specialized and quantitative results from the same data.

## §1. はじめに

本稿では、まず、筆者らが開発したスペクトルデータの自動解釈プログラムの first version について紹介する。つぎに、さらに深い解釈を行なうための second version に要求される機能や問題点について検討を加え、基本的な設計方針を与える。

対象とする装置は ESCA (または XPS : X線光電子分光装置) と呼ばれ、近年、金属や半導体表面の組成や結合状態の解明に重要な分析器の一つである。しかし ESCA のデータに関しては、多くの複雑な物理現象がスペクトルに反映されているため専門家にとってもその充分な解釈は難しい。さらにこのスペクトルにはバックグラウンドと呼ばれる処理のむずかしい大きなノイズ成分が含まれているため定量的な解析はさらに困難である。

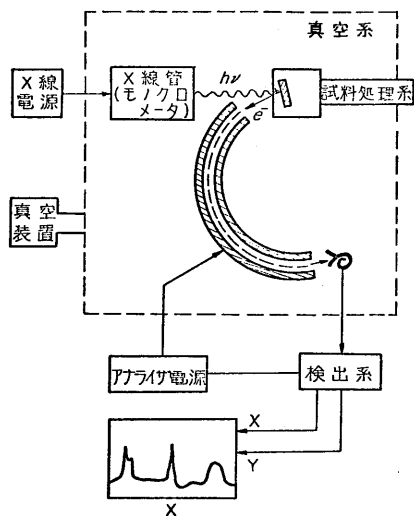
§2 では我々のプログラムの第1版 (ESCA I と呼ぶことにする) について説明する。ここでは、試料の組成分析を行なうことに相当する簡単な問題をとり上げ、解釈を行なうために導入した知識を有効に利用したトップダウン的な方法により効率よく正確な処理が可能であることを示す。ついで §3 では機能を強化した第2版 (ESCA II と呼ぶ) の設計のための種々の検討を行なう。ここではバックグラウンドの除去法、ピーク位置およびピーク面積を正確に定量的に把握するための方法、また化学シフトの評価法に関して検討を加え、合わせて処理の過程を簡潔なルールの集合として表現することによって機能の追加や変更が容易な形で実現を試みる。§4 はまとめである。

## §2. ESCA スペクトルの自動解釈プログラム (ESCA I)

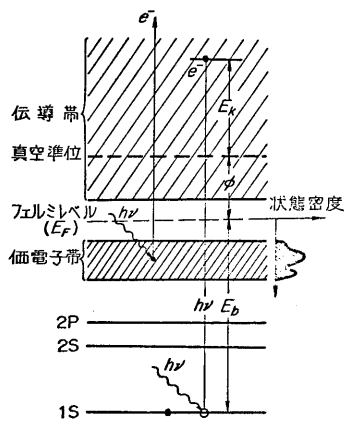
本プログラムは、ESCA によるスペクトルデータを、あらかじめ与えておいた解釈のための知識と照合することによって解釈を行ない、結果として試料の組成分析を行なうことができる。スペクトルの各ピークに解釈を与える段階では、ピーク位置 (結合エネルギー値) の照合だけでは一意的な解釈には到達しない。このため同一元素による複数個のピークの向にある強度の順序関係に注目し、その無矛盾性のチェックを行なうことによって誤った解釈を排除している。また同一元素のピークはトップダウン的 (知識主導形) な方法により効率的に検出し、適確な解釈を行なっている。

## 2.1 ESCAの概要

ESCAは高真空中におかれた試料物質に単色化されたX線を照射し、その結果光電効果によって発生する光電子をエネルギー分析することによって、試料物質中の電子のエネルギー準位と密度を測定する分析器である。これは簡易な組成分析器として使用できるばかりでなく、バンド構造や電荷分布など物性に関する重要な知見を与える。Fig. 1にESCAの原理を示す。(a)は測定器の構成を示し、(b)は光電子励起過程を示している。



(a) ESCAの概念図



(b) 光電子励起過程

し、(b)は光電子励起過程を示している。

放出された光電子の運動エネルギー  $E_k$  を測定することにより、照射した単色X線の光子エネルギー  $h\nu$  と試料物質の仕事関数  $\phi$  を用いて、電子のエネルギー準位（すなわちフェルミレベルを基準としたときの結合エネルギー） $E_b$  は次式により求めることができる。

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

Fig. 1 ESCAの原理

ESCAのスペクトルはこの結合エネルギーに対する電子の状態密度の分布を表わしている。

価電子帯から励起された電子のスペクトルによつては固体のバンド構造や電子状態密度に関する知見が得られ、また内殻準位から励起された電子のスペクトルによつては成分元素の同定や定量ばかりでなく電荷分布やイオン性、共有結合性などの結合状態に関する情報が得られる。このようにESCAのスペクトルにふくまれる情報は多岐にわたるが、ここでは内殻電子によるピークを各原子軌道に帰属させる問題を中心に考える。

## 2.2 処理方式

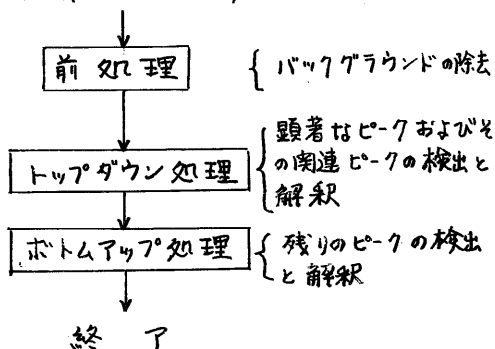
### 2.2.1 処理の流れ

処理はFig. 2に示すように3つのプロセスより成る。前処理ではESCA特有の大きなバックグラウンドノイズ（Fig. 4 および Fig. 13 参照）の除去を行なう。トップダウン処理はここでの中心となるプロセスであり、顕著なピーク（通常ある元素の主ピークである）とその関連ピーク（主ピークと同一元素によるピーク）を検出し、解釈を行なう。ボトムアップ処理では主として小さく、かつ孤立したピークの検出と解釈を行なう。主要なピークに関する情報だけが重要な場合には、このプロセスは不要であり、処理は単純化され処理時間も短縮される。

Fig. 3はFig. 2の処理の流れに対応するプログラムの最上位関数である。(使用言語はLISP 1.5である) 次にこれを簡単に説明する。

行番号	意味
3	INPUTで指定される機器よりデータアレイ(3行1024列のアレイで名前はDATA)の一行目に1024個のデータが入力される。
4~5	前処理を施し結果をアレイの2行目にセットする。
6~7	トップダウン処理
8~10	トップダウン処理後残された小さなピークを検出し頂点の座標をアレイの1行目にセットする。ピークの数をJにセットする。
11~12	ボトムアップ処理

ESCAスペクトルデータ



```

1 (SOLVE (LAMBDA () (PROG (J)
2 * PREPROCESSING PART *
3 (READDATA INPUT 1 1024)
4 (FILTER 1 2 FILTCOEF)
5 (BIN-OP 1 2 2 1024 (FUNCTION DIFFERENCE))
6 * TOP DOWN INTERPRETATION *
7 (TOPDOWN-INTPRT 2 1024 100 30 20 3 3)
8 * FIND LOWER PEAKS *
9 * AND SET J WITH THE NUMBER OF PEAKS *
10 (SETQ J (FINDPEAKS 2 1 1024 30) )
11 * BOTTOM UP INTERPRETATION *
12 (BOTTOMUP-INTPRT 1 J 30 20 3 3)
13 (RETURN T) )))
  
```

Fig. 2 処理の流れ

Fig. 3 最上位関数 SOLVE

### 2.2.2 前処理

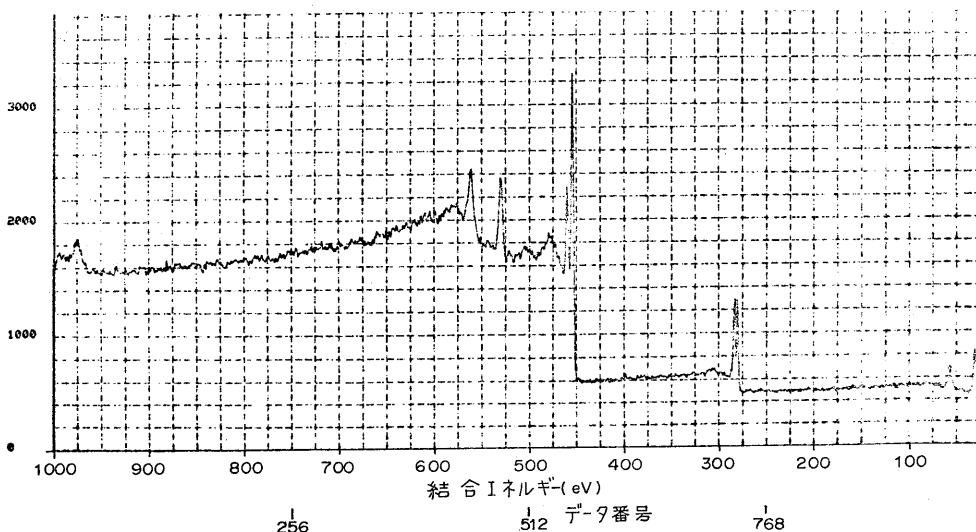
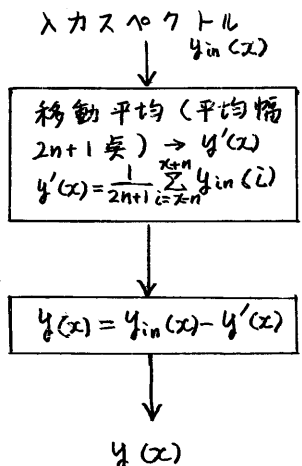


Fig. 4 入力データ

Fig. 4は耐熱材料の一種であるチタンの炭化物(TiC)のESCAによるスペクトルである。図は横軸方向1024英、各英10進4桁の数字で与えられたスペクトルを折れ線グラフの形で示している。各ピークは試料の表面附近にある原子の種々の電子軌道に存在する電子に対応し、横軸はその結合エネルギーの値を与える。また、大きなピークが存在する左側には、主として電子の非弾性散乱に由来する大きなバックグラウンドノイズが存在し、直流レベルにも段差が見られるのがわかる。この種のノイズは元素の種類や結合状態により形が変化するため一律の方法で除去するのは難しい。だが、一方、これを合理的に除去することがピーク位置およびピーク強度(ピークの面積)を正確に評価することはできない。ここでは帰属の決定が目的なので、この目的に合った近似的な方法でバックグラウンドの除去を行なっている。Fig. 5に前処理の過程を示す。



この処理は一種の High-pass filtering であり、過帯域幅は移動平均の幅  $(2n+1)$  を決定する  $n$  の値に反比例する。すなわち、左図により

$$y(x) = y_{in}(x) - \frac{1}{2n+1} \sum_{i=x-n}^{x+n} y_{in}(i) \quad \dots (2)$$

ここで  $x$  を連続量と見て空間周波数領域で考えると

$$Y(\omega) = Y_{in}(\omega) - Y_{in}(\omega) \times \frac{\sin \omega T}{\omega T} \quad \dots (3)$$

$(T = n + 0.5)$

(3)式より  $\omega T \geq \pi$  の信号分はよく通過し、

$\omega T < \pi$  の信号が減衰することが示される。

$n$  の値はバックグラウンドが充分減衰し、かつ不自然なアレジューツが出ないように決める。

Fig. 4のデータに  $n=4$  として前処理を行なったときの結果を Fig. 6 に示す。

Fig. 5 前処理

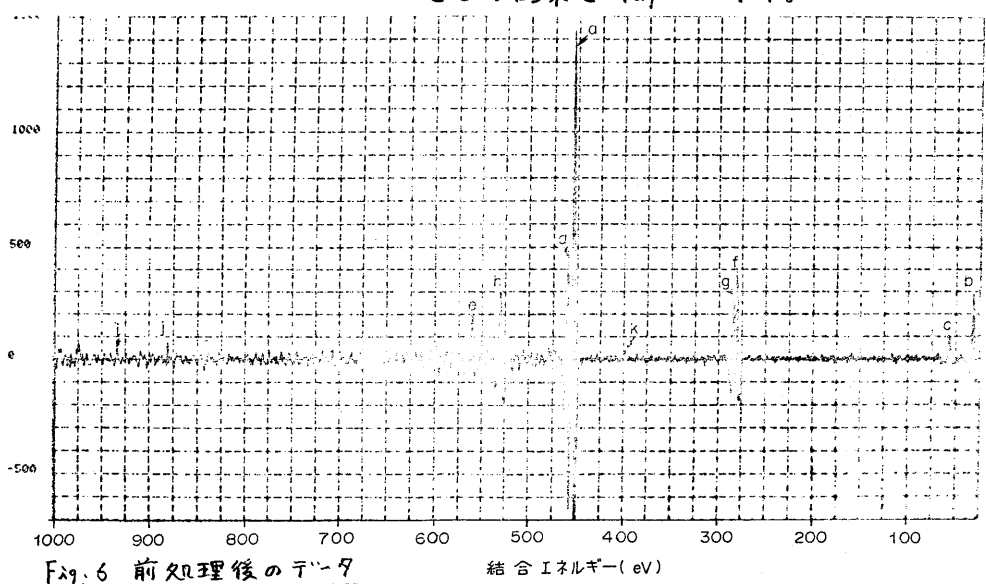


Fig. 6 前処理後のデータ 結合エネルギー(eV)

## 2.2.3 トップダウン処理

ここでは、顕著なピークとその関連ピークが知識駆動形のトップダウンの方法により効率よく検出され、同時に帰属が決定される。Fig. 7はそのフローチャートであり、Fig. 8は対応する関数 TOP DOWN INTPRT である。

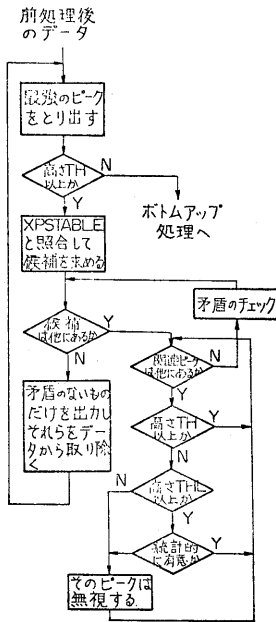


Fig. 7 トップダウン処理  
のフローチャート

```

1 (TOPDOWN-INTPRT (LAMBDA (IN N TH THL WN DE SF)
2 (PROG (A L TL LL LA LLA M)
3 (PRINT (QUOTE $$$ ** TOP DOWN INTERPRETATION START **$))
4 * FIND THE HIGHEST PEAK AND INTERPRET IT *
5 LO (SETQ A (FIND-HIGHEST IN N TH) )
6 (COND ((NULL A) (GO EX) ))
7 (SETQ L (SRCHATOMS A XPSTABLE DE) )
8 (COND ((NULL (CAR (CDDR L))) (GO LO))
9 (T (PUSH 1 L) )
10 * SEARCH THE RELATED PEAKS *
11 (SETQ TL (MAKE-CTABLE XPSTABLE L) )
12 P1 (COND ((NULL TL) (GO LO)) )
13 (SETQ LL (CDAR TL) )
14 (SETQ A (CAAR TL) )
15 P2 (COND ((NULL LL) (CHECK 1 IN THL) (CLEAR 1)
16 (SETQ TL (CDR TL)) (GO P1) ) )
17 (SETQ LA (CAR LL) )
18 (SETQ LLA (LIST A LA) )
19 (SETQ PL (SRCHPEAKS IN N (TRI (CAR LA)) DE THL) )
20 P4 (COND ((NULL PL) (GO P7) ))
21 (COND ((GREATERP (CAAR PL) TH) (GO P5))
22 ((LESSP (CAAR PL) THL) (GO P6))
23 ((STCHAPPRV IN (CAR PL) N WN SF) (GO P5))
24 (T (GO P6) )
25 P5 (SETQ M (LIST (CAR (CDAR PL)) (TR (CAR (CDAR PL)))) )
26 (PUSH 1 M)
27 P6 (SETQ PL (CDR PL) )
28 (GO P4)
29 P7 (SETQ LL (CDR LL) )
30 (GO P2)
31 EX (RETURN T) )))

```

Fig. 8 関数 TOP DOWN INTPRT

次にこの処理過程を、基本となる処理ルールの集合として表わしてみよう。

- ①. ピークの高さに関して2種の閾値 TH, および THL を設定する。TH はピークの無条件有意レベルを示し、これを越えるピークは顕著なピークと見なす。THL はノイズレベルを示し、これ未達のピークはピークとしては考えない。処理の停止条件を与えている。
- ②. TH未達 THL以上のピークについては近傍のその時点でノイズ分と考えられるデータに対して統計的に有意か否かの判定を行ない、その結果によってピークとして認定するか否かを決定する。
- ③. ピークの解釈は、その時点で残っている最強のピークから出発する。
- ④. 解釈は、結合エネルギーの値に関して、与えられた知識とのマッチング(許される誤差は化学シフトにもとづく本質的なもの、およびキャリブレーション、量子化誤差、計算誤差等を考えとあらかじめ与えておく)を基本とする。ここで用いた知識の一部を Fig. 9 に示す。

```

CSET (XPSTABLE (
(H (3.595 E1S 1))
(HE (24.580 E1S 1))
(C (11.264 E2P1/2 2) (18.35 E2S 2) (284 E1S 1))
---
(CA (6.111 E4S 3) (25.0 E3P1/2 2) (44 E3S 2) (346 E2P3/2 1) (350 E2P1/2 2)
(438 E2S 2) (4038 E1S 2))
---
(TI (6.83 E4S 3) (4 E3D3/2 3) (35 E3P1/2 2) (60 E3S 2) (456 E2P3/2 1)
(462 E2P1/2 2) (564 E2S 2) (4966 E1S 2))
---

```

)) Fig. 9 使用した知識

- ⑤. 一つのピークに対して解釈が定まった場合(一意とは限らないが), その同一元素による関連ピークを先に探す。ここで知識主導形の探索が行われる。
- ⑥. 同一元素と考えられる一連のピークに対する処理が終了したところでの強度順位の情報により誤った解釈をふるい落とす。ここでは強度順位として強→弱に1, 2, 3の3段階を与える
  - i) 2のものがあった1のものがない, または ii) 3のものがあった2のものがないの2つの場合にその一連の解釈を排除している。
- ⑦. 解釈が一意に定まったピークは作業用データエリアから除去する。
- ⑧. トップダウン処理の停止条件: 最強のピークから始まった一連の処理が終了すると, 次の最強ピークを探すが, このときTH以上のものが見あたらぬ場合トップダウン処理は終了し, ボトムアップ処理に移行する。

### 2.2.4 ボトムアップ処理

ボトムアップ処理に残されたピークはすべてTH未満である。小さくかつ顕著なピークと関連を持っていないことから, これらのピークは微量に含まれた元素の主ピークである可能性が高い。したがってここではTHL以上のすべてのピークを検出し, 統計的な検定によって有意と判定されたピークについて順に上記④により解釈を行なう。すべてのピークについて解釈がなされた時点で同一元素にassignされた複数の解釈があれば, それらについて上記⑥と同様のチェックを行なう。しかし主ピークが強直して存在する場合には多義の解釈が十分に考えられ, この場合にはここで与えた知識の範囲内ではこれ以上しぼり切れない。

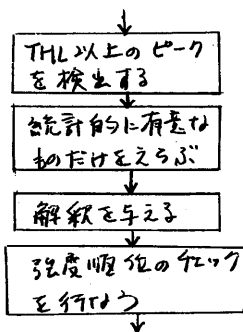


Fig. 10 ボトムアップ処理

Fig. 10 は処理のフローチャート, Fig. 11 はプログラムである。

```

1 (BOTTOMUP-INTPRT (LAMBDA (IN J WN DE SF)
2 (PROG (I A L)
3 (PRINT (QUOTE $SZ ** BOTTOM UP INTERPRETATION START **))
4 (SETQ I 0)
5 L0 (SETQ I (ADD1 I))
6 (COND ((GREATERP I J) (GO EX)))
7 (SETQ A (DATA IN I))
8 (COND ((STCHAPPRV 2 A 1024 WN SF) (GO L1)))
9 (T (GO L0)))
10 L1 (SETQ L (SRCHATOMS A XPSTABLE DE))
11 (COND ((NULL (CAR (CDDR L)))) (GO L0))
12 (T (PUSH 1 L)))
13 (GO L0)
14 EX (C/CHECK 1 2 THL)
15 (CLEAR L)
16 (RETURN T)))

```

Fig. 11 ボトムアップ処理の関数

### 2.3 処理例 1

Fig. 4 のデータを入力として処理を行なった結果を Fig. 12 に示す。図中ピーク a がまず検出されこのタンの主ピークとして解釈される。ついでこのタンの他のピークが知識駆動形で探索され b ~ e のピークが検出され解釈が与えられる。この一連の解釈は強度順位に矛盾がないことからすべて採用される。次に最強のピークとして f が検出され炭素 C の 1s と解釈され関連ピークとして g も C の 1s と解釈される。この2つのピークは同じ C の 1s によるものであるが結合環境の違い (Ti と結合したものと O と結合したもの) による化学シフト量の差から2つに分離したものである。(このような解釈も行なえるよう

```

** TOP DOWN INTERPRETATION START **
(559 458 (TI (456 E2P3/2 1)))-- a
(992 35 (TI (35 E3P1/2 2))) -- b
(966 60 (TI (60 E3S1/2 3))) -- c
(553 464 (TI (462 E2P1/2 2)))-- d
(450 564 (TI (564 E2S 2))) -- e
(737 284 (C (284 E1S 1))) -- f
(734 287 (C (284 E1S 1))) -- g
(482 533 (O (532 E1S 1))) -- h
** BOTTOM UP INTERPRETATION START **
(68 937 NA) -- i
(122 885 NA) -- j
(616 402 (N (402 E1S 1))) -- k
    
```

#### Fig. 12 処理結果 1

にはれば非常に有用である) つぎに h の酸素によるピークが検出され解釈される。これが終了するとあらかじめ決めておいた閾値 TH (100) を越えるピークが存在しなくなるためトップダウン処理は終了する。ボトムアップ処理ではまず THL (30) 以上のピークをすべて検出し、それぞれについて近傍のノイズ分に対する統計的な有意性の判定を行ない有意のものについて解釈を試みる。この結果 a の小さなピークは窒素 N の主ピークと解釈される。しかし、i, j のピークについては該当する解釈が存在しないため NA (not assigned) として出力された。これは先に Fig. 1 で説明した ESCA の光電子励起過程とは別の現象にもとましくオージェピークと推察される。これについても同形式で知識を与えておけば解釈も可能であるが強度順位に関しては別の扱いが必要である。

### 2.4 処理例 2

Fig. 13 は CaB<sub>6</sub> と呼ばれる試料の ESCA スペクトルである。これは横方向 512 点のデータであるが上の例にくらべて多くのピークが見られる。前処理後の結果を Fig. 14 に、さらに解釈を試みた結果を Fig. 15 に示す。この例ではまず a の酸素 B の主ピークが検出され解釈される。つぎにカルシウム Ca の主ピークが検出され解釈される。つぎにこの関連ピーク群 c ~ f が検出される。ここでは正しい解釈が与えられいるが b と c はきわめて近いので解釈が重なるような場合も考えられる。このよう

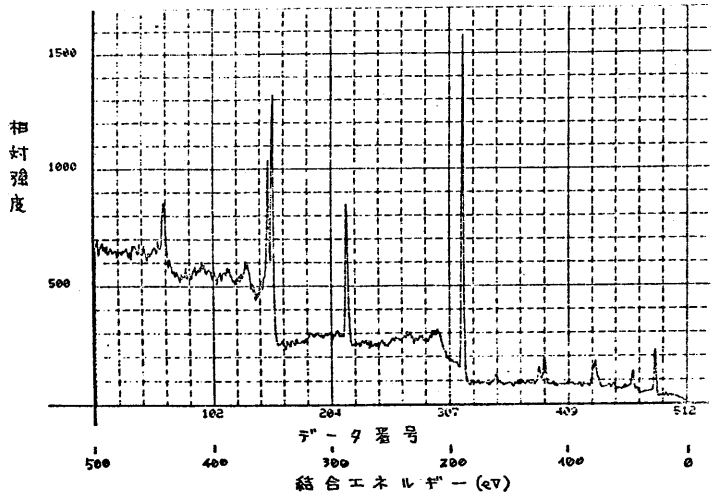
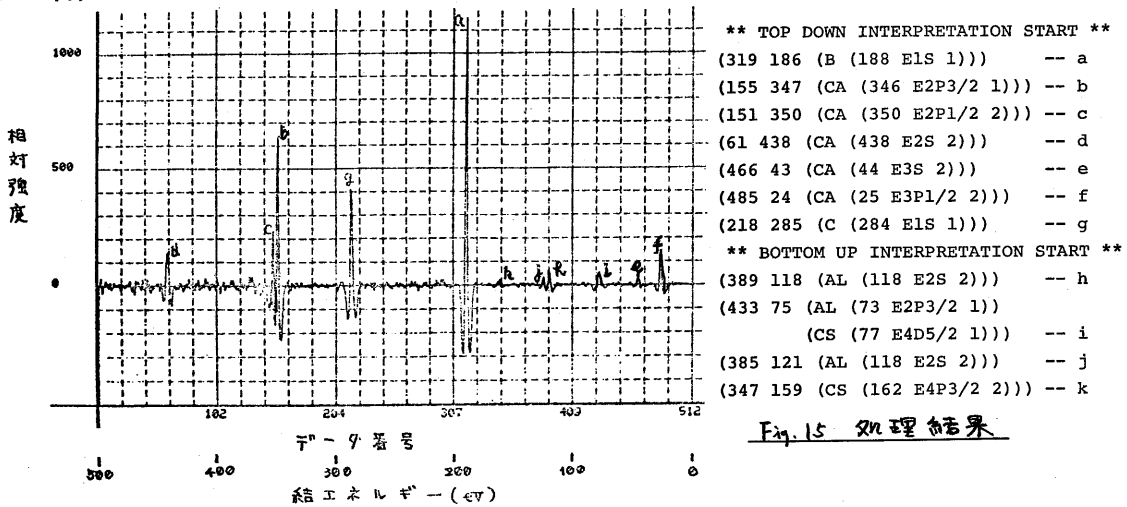


Fig. 13 入力データ 2

は場合に関しては新たなルールの導入が必要であろう。つぎには炭素 C の主ピークが処理され、そので TH (100) 以上のピークがなくなるためトップダウン処理を終える。ボトムアップ処理では THL (30) 以上の有意なピークとして長～長の4個が検出される。このうち長、い、はアルミニウムのピークとして解釈することができ強度順位にも矛盾が見られぬことがわかる。また i のピークはセシウム Cs の主ピークと解釈することもでき、たのピークが同じく Cs と解釈されることと合わせて試料中のセシウムの存在が示唆される。実際、この試料の場合、試料の処理過程などでアルミニウムおよびセシウムが混入する可能性が示され、この解釈は正しいものと考えられる。



以上2つの例についてプログラムの能力を評価すると ESCA の初心者レベルにあると考えられるが、例1での長、い、のピークの検出能力、例2の長～長のピークに関する解釈など人間が不得手とする部分をカバーし得る長も見られる。

### 3.3 機能向上に関する考察

本章では、専門家にとつてもさらに有用なシステムを提供することを目標とした ESCA II に組み込むべき機能とそれを実現するための方法について考察を加える。

#### 3.1 プログラムの構造

前出、Fig. 8 等のプログラムは LISP による度教評価形の構造のため、手続き表現による FORTRAN 上でのプログラムにくらべて構造が明確で処理の基本ルールも比較的明確な形であらわされている。しかし、例1での長～長のピークの検出能力、例2の長～長のピークに関する解釈など人間が不得手とする部分をカバーし得る長も見られる。このシステムではルールの追加、変更の容易さは極めて重要な要素であり、このよう観点から、ルールはプログラムの制御構造に直接うめこまれることとなくデータとして表現されることが望ましい。ESCA II では処理過程の多くの部分をルールの集合として表現し、ルール駆動形の構造を採用したい。ルールは基本的には



## 条件 → ACTION (処理, 結論)

の形をとる。すなわち与えられたデータに関して左辺の条件が満たされると、右辺で指定された処理を行なって一定の結論を得、さらに必要ならば次に行なう処理を指定する。詳細については、Shortliffe 等による MYCIN や言語理解システム等を参考に本システムに適した方法を考えた。

### 3.2 解釈の確度の評価法

ESCA I では解釈の候補が複数個存在する場合に、強度順位に関して矛盾のあるものを排除<sup>除外</sup>以外には評価の方法はなかった。しかし現実には関連ピークが数多く存在するようないくつかの解釈と孤立した解釈とでは量に差が生じて当然であるし、定量的な解釈を行なうような場合にはさらに確かな評価が必要となる。ESCA II では解釈の確かさを表現する得点を導入し、この得点を増減するルールによって解釈の確度を評価する。たとえば次のようなルールを考える。

- ① 問題とするピークについて結合エネルギーに関してマッチングがとれれば +10 加える。
- ② 同一元素に関する一連の解釈で強度順位に関して、(i) 3 のものが存在して 2 のものが存在しない、または (ii) 2 のものが存在して 1 のものが存在しない場合、一連の解釈すべてを -10 とする。
- ③ 同一元素の一連の解釈に関して主ピークが存在すればすべての解釈に +20 加え、他のピークは存在する数だけ +10 の加算を行なう。

この結果は矛盾のある解釈は -10 とはり、他は +30 以上となり関連ピークが多いほど得点が高くなる。

### 3.3 その他

#### [1] TH, THL の自動設定

ESCA I では TH, THL を人間が与えていたが、これは相当煩雑なので自動設定としたい。一般のスペクトルを対象にするとこれは難しいが、ESCA のスペクトルに限れば比較的容易であると考えられる。たとえば前処理後のスペクトル全体の r.m.s 値 (2 乗平均平方根値) を TH ととり、トップダウン処理後のスペクトルの r.m.s 値を THL ととり (必然的にトップダウン処理中では THL を使用できなくなるが) などが考えられる。

#### [2] 自動キャリブレーション

まず、現在と同じく、データ番号と結合エネルギーの値の直線関係があると仮定する。ESCA I では変換のための関数 TR およびその逆関数 TRI を与えていたが、オペレータが任意の 2 点についてデータ番号と結合エネルギー値を与えたことによって自動的に変換関数を生成することを考える。

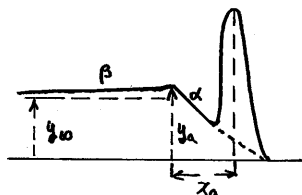
#### [3] ピーク位置の検出精度の向上

ピーク位置はピークの最高点の X 座標の値であるが、化学シフトの値を出すために高精度化したい。このため各ピークごとに頂点附近に曲線 (2 次曲線?)

を割り出し、その曲線の頂点の座標をピーク位置とする。

#### CA7 バックグラウンドの除去

ESCAの定量性を向上させるためにはバックグラウンドの正確な除去が最も重要である。存在した単一ピークに関しては比較的良い方法もあるが我々の場合に直接応用するのは困難である。



左図はピークとこれにともなうバックグラウンドの様子を示している。図中 $\alpha$ の部分はほぼ直線、 $\beta$ の部分は $\gamma_0$ に収束する指数関数として近似することができ、 $x_a$ ,  $\gamma_a$ ,  $\gamma_0$ などを理論的に求めることは困難である。ESCA IIでは何らかのheuristicな方法によってESCA Iよりは正確な除去を行おうべく検討中である。

#### §4. あとがき

以上、我々が開発したESCAスペクトルの自動解析プログラムについて概説し、このプログラムの近い将来への発展方向を示した。

実験室データの処理システムも考えたと重要なことは、人間が不得手とする仕事、すなわち、"他の可能性"をすべをつくしたり、定量的な評価を行うことであろう。手処理システムの形としては、処理方式や機能が容易に拡張、変更できることであろう。簡単な処理であってもheuristicなルールをふくむ場合が多く、またこれらのルールが往々にして不完全であるというのがその理由である。また、我々のシステムはLISPを中心として実現されているが、奥層面から考えると、よりコンパクトな数値計算能力に富む言語の開発、実用化も重要である。

最後に、共同研究者である伊原英雄氏、はじめ材料部、創研部各位に感謝いたします。

#### [参考文献]

- 1) 山崎, 伊原 "トップダウン法によるスペクトルデータの自動解析" 電総研彙報 第40巻10号, 昭和51年10月
- 2) E.H.Shortliffe: Computer-Based Medical Consultations: MYCIN, ELSEVIER, 1976.
- 3) B. Buchanan Georgia Sutherland & E.A. Feigenbaum: HEURISTIC DENDRAL: a Program for Generating Explanatory Hypothesis in Organic Chemistry, Machine Intelligence 4