

# 機械学習によるスペクトル解析と全反射蛍光X線分析への応用

堂井 真\*, 菊田 真也\*\*

## 1. はじめに

全反射蛍光X線 (Total Reflection X-ray Fluorescence, TXRF) 分析法ではX線を極めて低角度 (約 $0.1^\circ$ ) で試料に入射させることで、入射X線が試料表面において全反射するのでバックグラウンドを低く抑えることができ、試料表面から発生する蛍光X線を感度良く測定して分析する手法である<sup>(1),(2)</sup>。TXRF分析はX線を使用しているため非破壊で測定可能である。さらに、試料表面が平坦であれば特に前処理することなしに測定することができる。このような特長からTXRF分析は産業・環境用途だけでなく半導体製造工程のウェーハなどの汚染評価にも広く使用されている<sup>(3)</sup>。半導体製造工程は厳しい汚染管理がされており、TXRF分析の高感度化の要求は年々高まっている。

半導体製造工程でのウェーハ上の汚染物質は均一に存在することは稀で、パーティクル汚染の様に不規則に散在することが多い。汚染量がウェーハの面内平均では管理濃度以下であっても、局所的に $10^{10}$  atoms/cm<sup>2</sup> オーダー以上の高濃度汚染物質が存在することがあり、歩留まり低下の原因となる。そのような場合に1点あたりを短時間で測定し、かつウェーハ上全面を高速マッピングする方法 (SWEEPING-TXRF)<sup>(4),(5)</sup> が有効であり、この方法が、半導体汚染管理の現場では広く用いられている。SWEEPING-TXRFによりウェーハ上の汚染プロファイルを迅速かつ詳細に分析する事が出来る。さらに、汚染事故が発生した際には原因の追究に役立つ。加えて、ウェーハエッジまでくまなく測定が可能である。図1にSWEEPING-TXRFの測定例を示す。これはウェーハ洗浄工程後の汚染確認にTXRF測定を行った例である。SやClがウェーハ全面に検出されているのは洗浄に使用した硫酸系や塩素系の薬液の残差である。K, Ca, Ti, Cr, Znがウェーハエッ

ジを含む複数個所で検出されていることに加えて、Feがほぼウェーハ全面で検出されている様子が如実にわかる。この例に限らずTXRFは半導体製造工程管理に幅広く用いる事が出来る。SWEEPING-TXRFでは局所的な結果が得られるだけでなく、ウェーハ全面の平均結果も同時に得る事が出来る。局所的な結果と平均結果を組み合わせることで、汚染付着形態を詳細に分析でき、より厳密で多面的な汚染管理を実現できる点が大きな特長である。

通常TXRF測定ではウェーハ上の付着物質とその量に応じて、付着物質から発生した蛍光X線を含んだエネルギースペクトルが得られる。その中には複数のピークが含まれており、それらをピーク分離し単一ピークごとにフィッティングすることで分析を行っている。ピーク強度が十分大きい場合は問題ないが、測定時間が短い、あるいは、汚染量が極めて少なかったりする場合、本当にピークとして存在するのかノイズなのかの判断が難しい場合がある。特にSWEEPING-TXRFでは、ウェーハ全面測定が可能であるが、スループットの観点から1点あたりの測定時間は5秒から10秒の短時間を要求される。そのため従来のピークフィットによる手法ではノイズによるピークの見落としや誤検出が起こりやすいという側面がある。

近年、計算機性能の向上、膨大なデータを含むライブラリの整理、各種センサー類の普及によるデータ取得の容易化などのおかげで人工知能 (AI) に関する技術が急速に発展している。AIは膨大なデータの中から注目すべき特徴量を自動で抽出して分析することが最大の特徴である。画像処理の分野では手書き文字認識に代表される画像認識<sup>(6)</sup>や、低解像度の画像を高解像度の物に変換する超解像度化<sup>(7)</sup>などが盛んに研究されており、多くの実績を上げている。

画像処理に関してはAIを活用した事例が多く存在するものの、2次元の画像ではなく1次元のスペクトルにAI技術を適用した例は少ないように見受けられる。

\*株式会社リガク

\*\* 共著者の菊田真也氏は本原稿執筆中に急逝しました。本研究における菊田氏の貢献に感謝すると同時に、故人のご冥福を心よりお祈り申し上げます。堂井

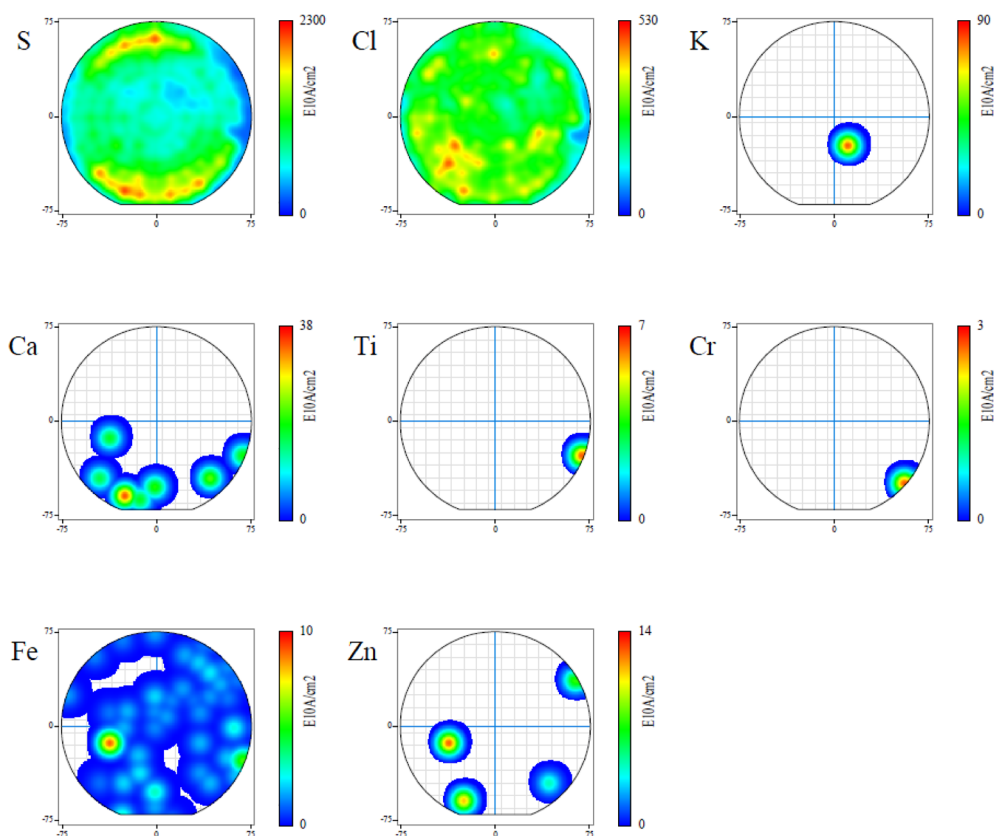


図1. SWEEPING-TXRF 測定例.

そこで本稿では、TXRF 分析のデータ処理、特に短時間測定で得たスペクトルからの汚染元素定量に関して機械学習の手法を応用し、その効果を紹介する。

## 2. 手法

### 2.1. 機械学習における教師データと学習方法

機械学習とは大量のデータの相関関係を学習させることで、何らかのパターンやルールを発見させ、それを様々な物ごとに利用することで種々の判別や予測を行う技術である。TXRF においては測定で得られたエネルギースペクトルから含まれる元素とその量を推定させることが目的となる。今回は機械学習の中でも教師あり学習に分類される技術を適用した。すなわち、測定スペクトルと含まれる元素とその量があらかじめ既知のデータセット（教師データ）を使ってこれらの対応関係や法則性を学習させ、学習結果を教師データとは独立に測定したスペクトル（検証用データ）に適用し、それらに含まれる元素とその量を推定させる。既知とする含有元素とその量は、同一測定点を長時間測定したデータから従来のピーク検索とピークフィット（詳細なアルゴリズムについては割愛する）によって求めた。

機械学習演算ライブラリには TensorFlow<sup>(8)</sup> を使用し、機械学習のモデルには畳み込み層が 1 層、全結合層が 6 層の畳み込みニューラルネットワークモデル

(図2) を構築し学習を行った。すなわち、測定スペクトルの各エネルギー（各チャンネル）の X 線強度を入力層に与え、出力層より対象元素の推定値を得る。最後の出力層を除いた各層の活性化関数には ReLU (Rectified Linear Unit,  $f(x)=\max(0, x)$ )<sup>(9)</sup> を使用した。最終的出力は各元素の X 線強度であるので、負になることはない。そこで、最後の層の活性化関数には次の式(1) で与えられる関数を使用した。

$$f(x)=\log(1+\exp(x)) \quad (1)$$

この関数は、変数  $x$  が大きい時は  $f(x)=x$  に漸近的に近づき、 $x$  が負になれば  $f(x)=0$  に正値から近づく。

学習は長時間測定データを解析して得た対象元素の定量値（ここでは敢えて「真値」と呼ぶ）と機械学習による推定結果を比較して行う。ニューラルネットワークを構成するパラメータの最適化は、損失関数を最小化するように進める。損失関数としては以下で与えられる真値との差の 2 乗和を使用する。

$$\text{Loss} = \sum w_i (I_i - I_i^0)^2 \quad (2)$$

$I_i^0$  は真値で、 $I_i$  は機械学習による推定値である。

真値からのずれが Gauss 分布の時は、この損失関数を最小とする  $I_i$  の分布が最尤推定量を与える（最尤法）。しかし、実際に検出される X 線強度の変動は小さいときは Poisson 分布であり、数え落とし補正（デッド

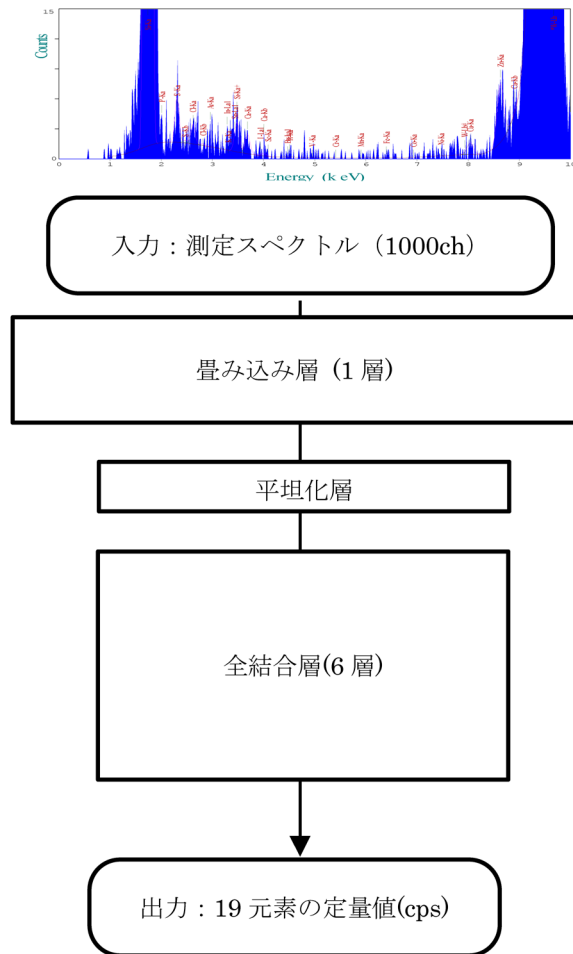


図2. 使用した畳み込みニューラルネットワークモデル。

タイム補正)が必要となる程X線強度が大きい時はPoisson分布からもずれる<sup>(10),(11)</sup>ので上記損失関数の最小化は必ずしも最尤推定とはならないが、取り扱いが簡単であるのでここでは上記残差2乗和を使用する。

次に上記式の重み $w_i$ について考察する。X線強度は統計的揺らぎを持っており、その標準偏差は(Poisson過程の場合)強度の平方根に比例する。

$$\sigma_i \propto \sqrt{I_i^0} \quad (3)$$

従って、残差の2乗 $(I_i - I_i^0)^2$ は $\sigma_i^2$ の不定性を持っているので、各項の重みとしては、 $w_i = 1/\sigma_i^2$ とするのが適当である。

$$\begin{aligned} \text{Loss} &= \sum w_i (I_i - I_i^0)^2 \\ &= \sum \frac{(I_i - I_i^0)^2}{\sigma_i^2} \propto \sum \frac{(I_i - I_i^0)^2}{I_i^0} \end{aligned} \quad (4)$$

この損失関数の最小値を与える最適パラメータの探索には、確率的最適化の一手法であるAdamax法<sup>(12)</sup>を使用した。

教師データと検証用データは、リガク製全反射蛍光X線分析装置TXRF 3760とTXRF-V310<sup>(13)</sup>を用いて

様々なSiウェーハを測定した結果より作製した。

## 2.2. 対象元素と試料

リガク製TXRF 3760とTXRF-V310では三つの励起X線源によりNaからUまで分析可能であるが、今回、励起X線源はW-L $\beta$ 線のみを考え、対象元素は比較的よく注目されるSi, P, S, Cl, K, Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Sn, Baとした。また、励起X線由来のW-L $\beta$ 線強度も学習対象とした。測定試料は上述のようにSiウェーハである。これら以外の元素や試料について機械学習を適用する場合には、注目する元素や試料を含んだ教師データを準備する必要がある。

なお、対象元素の内Siは基板由来であり、W-L $\beta$ 線は励起X線由来であるため、他の元素の蛍光X線強度よりも非常に大きくなる。学習時にこれらのX線強度を同様に扱くと数値の大きなSiとW-L $\beta$ に学習時のパラメータが引っ張られる可能性があるため、SiとW-L $\beta$ のX線強度は学習時には1/1000にした。

## 3. 結果と考察

### 3.1. TXRF 3760で取得したデータによる機械学習とその効果

TXRF 3760で取得した15,277点の5秒測定のスpectrumとそれを従来法で定量した各元素のX線強度、および同一点を60秒測定したスpectrumから従来法で定量した各元素のX線強度を元に、15,277点からランダムに抽出した13,749点を教師データとし、残り1,528点を検証用データとした。

まず、5秒測定、60秒測定の検証用データのspectrumを従来法で解析した時の各元素のX線強度相関グラフを図3(a)に示す。横軸が60秒測定の結果で縦軸が5秒測定の結果である。多くの元素で5秒測定でも60秒測定の結果に近い値が得られているが、X線強度が小さい場合には相関が悪くなり5秒測定結果は不定性が大きくなっている。次に、教師データで機械学習したモデルに5秒測定の検証用データのspectrumを読み込ませて得た結果と、60秒測定の従来法の結果の相関グラフを図3(b)に示す。X線強度が強い場合には従来法の結果(図3(a))とよく似た相関がみられる。更に、X線強度が小さい場合でもばらつきは小さくなっている様に見える。

これらのグラフの相関係数を表1に、またそれらの値をプロットしたグラフを図4に示す。図4の横軸は従来法で得た相関係数の値で、縦軸は機械学習モデルで得た値である。大部分の元素において機械学習モデルで相関係数が向上している。

TXRFの解析においては、X線強度が強い時の正確度も重要だが、X線強度が弱い時でも誤定量しないことがより重要な性能となる。そこで、60秒測定結果では1 cps以上であるのに5秒測定spectrumを解析

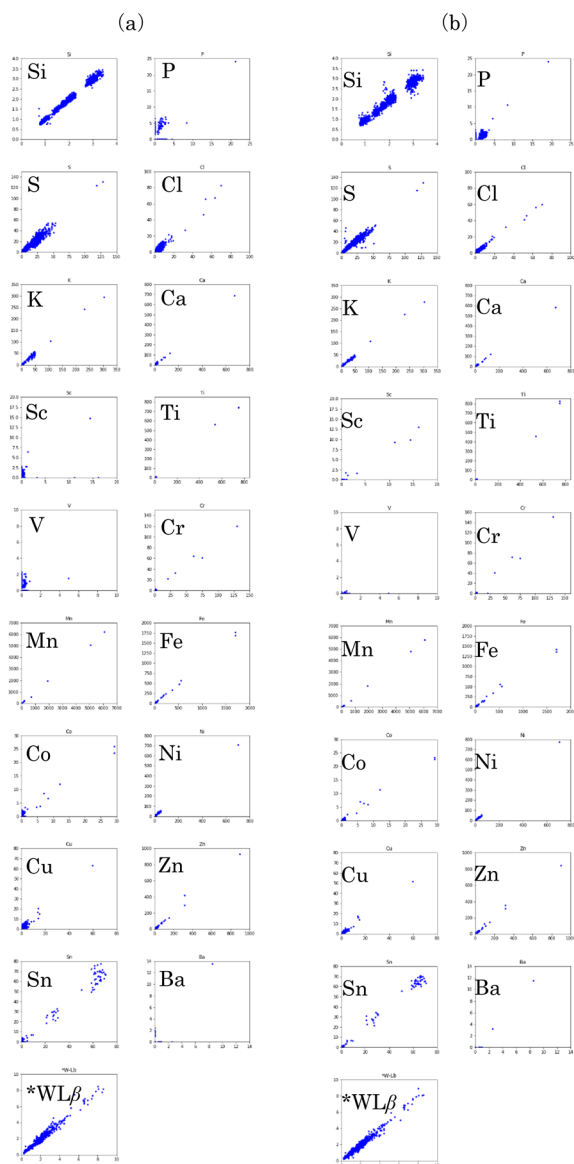


図3. 60秒測定結果（横軸）と5秒測定結果（縦軸）の相関。  
(a) 従来法による解析, (b) 機械学習による推定.

した結果が1 cps以下だった件数（未検出件数）と、60秒測定結果では1 cps以下であるのに5秒測定スペクトルを解析した結果が1 cps以上であった件数（過剰検出件数）の合計を誤検出した件数として表2にまとめた。従来法に比べて機械学習モデルによる解析では誤検出した件数が大幅に減少している。

### 3.2. TXRF310データへの適用

上記3.1節ではTXRF 3760で取得した教師データで学習し、同一装置で取得した検証用データでその効果を確認した。ここでは、TXRF 3760のデータで学習した機械学習モデルを別装置のTXRF-V310のデータへの適用を考察する。学習したモデルを他へ応用する場合、一般に (1) そのまま適用する, (2) 学習モデルの全てのパラメータを適用対象の少数の教師データを使って再学習させる（ファインチューニング）, (3) 学習モデルの一部の層に関与するパラメータのみ

表1. 60秒測定結果に対する5秒測定結果の相関係数 (TXRF 3760)。

| 元素    | 従来法   | 機械学習  |
|-------|-------|-------|
| Si    | 0.993 | 0.974 |
| P     | 0.442 | 0.795 |
| S     | 0.954 | 0.960 |
| Cl    | 0.915 | 0.975 |
| K     | 0.994 | 0.997 |
| Ca    | 0.997 | 0.999 |
| Sc    | 0.342 | 0.973 |
| Ti    | 1.000 | 1.000 |
| V     | 0.088 | 0.085 |
| Cr    | 0.991 | 0.987 |
| Mn    | 1.000 | 1.000 |
| Fe    | 0.999 | 0.997 |
| Co    | 0.886 | 0.986 |
| Ni    | 0.998 | 0.999 |
| Cu    | 0.867 | 0.967 |
| Zn    | 0.995 | 0.998 |
| Sn    | 0.993 | 0.998 |
| Ba    | 0.855 | 0.978 |
| *W-Lβ | 0.986 | 0.985 |

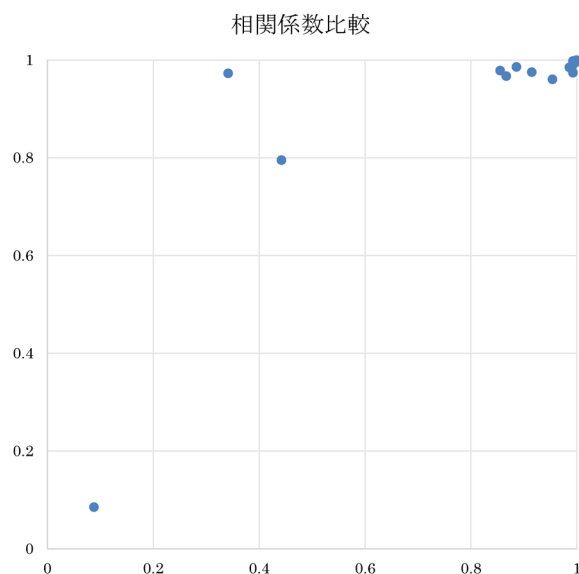


図4. 各元素の相関係数プロット。横軸は従来法、縦軸は機械学習モデルによる結果。

を適用対象の少数の教師データを使って再学習させる。この場合、他の層に関与するパラメータは固定して扱う（転移学習）, の3通りがある。今回は (1) と (2) の結果を示す。

ここではTXRF-V310で取得した1634点を検証用データとし、同一点を5秒測定、60秒測定して従来法で得た結果の相関および誤検出件数を図5(a), および、表3, 表4の従来法の欄に示す。これに対して、①TXRF3760で学習したモデルをそのままTXRF-

表2. TXRF3760誤検出件数 (全1,528点中).

| 元素 | 従来法 | 機械学習 |
|----|-----|------|
| P  | 775 | 450  |
| S  | 68  | 18   |
| Cl | 306 | 114  |
| K  | 241 | 23   |
| Ca | 256 | 38   |
| Sc | 153 | 2    |
| Ti | 110 | 3    |
| V  | 140 | 1    |
| Cr | 103 | 7    |
| Mn | 120 | 16   |
| Fe | 354 | 111  |
| Co | 165 | 6    |
| Ni | 99  | 27   |
| Cu | 431 | 93   |
| Zn | 282 | 94   |
| Sn | 182 | 19   |
| Ba | 15  | 0    |

V310の検証用データに適用した場合 (機械学習①), ②TXRF-V310で検証用データとは別に用意した教師データ1,000点で再学習したのち, 検証用データに適用した場合 (機械学習②), ③TXRF-V310で検証用データとは別に用意した教師データ100点で再学習したのち, 検証用データに適用した場合 (機械学習③), それぞれを図5(b)(c)(d), および, 表3, 表4の機械学習①, ②, ③の欄に示す.

相関関数については, そのまま適用 (機械学習①), 100点で再学習した場合 (機械学習③) とも従来法より全般的に劣った結果となった. 1,000点で再学習した場合 (機械学習②) は, 元素にもよるが, 全体としては従来法に近い値を得た.

次に, 誤検出件数については, 1,000点で再学習させた場合 (機械学習②) は全元素で確実に改善している. 100点で再学習した場合 (機械学習③) でも従来法と比べて全体として見劣りしない.

以上より, 機械学習法は短時間測定において誤検出

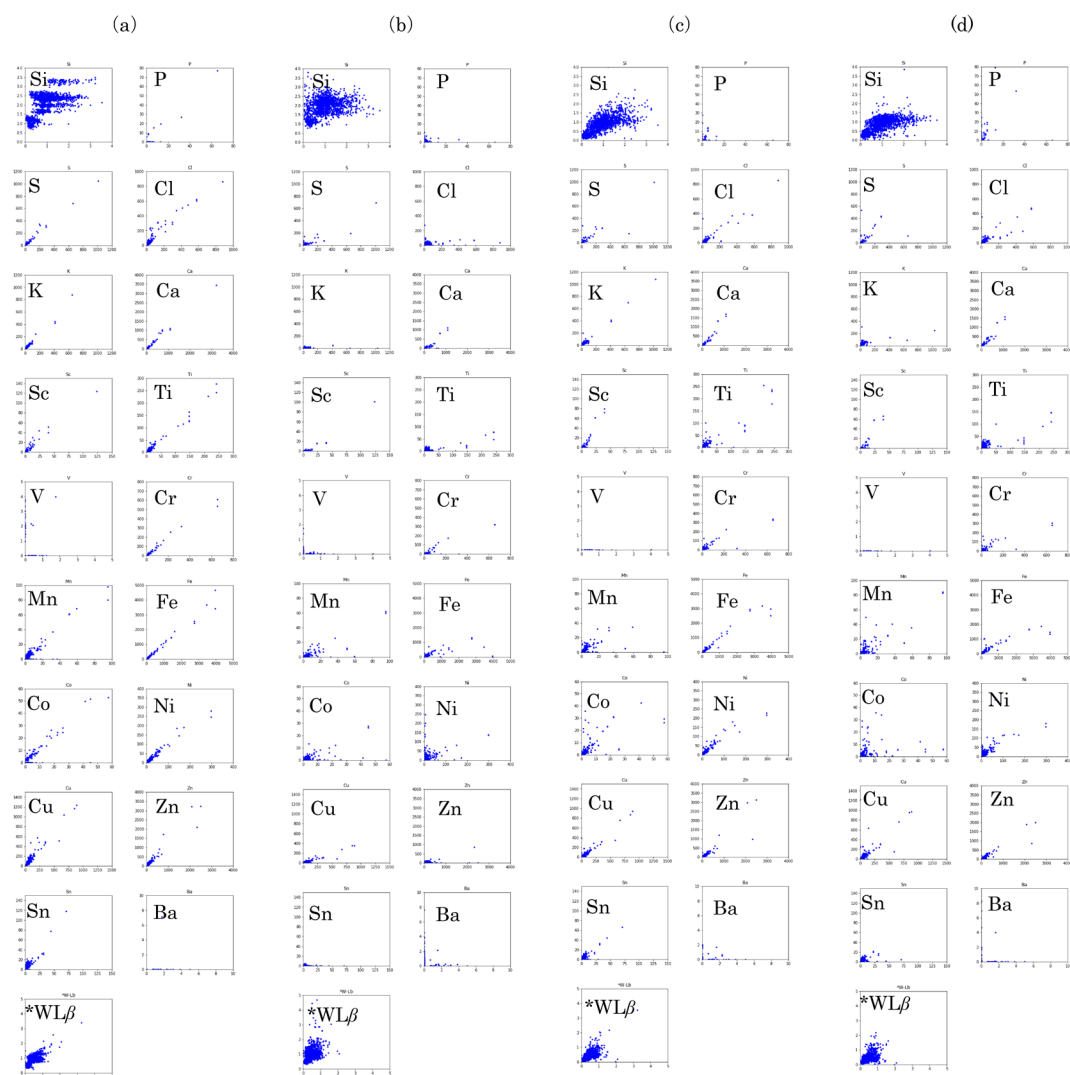


図5. 60秒測定結果 (横軸) と5秒測定結果 (縦軸) の相関. (a) 従来法による解析, (b) 機械学習①, (c) 機械学習②, (d) 機械学習③. 機械学習①, ②, ③については本文説明を参照.

表3. 長時間測定結果に対する5秒測定結果の相関係数 (TXRF-V310).

|              | 従来法   | 機械学習<br>① | 機械学習<br>② | 機械学習<br>③ |
|--------------|-------|-----------|-----------|-----------|
| Si           | 0.590 | 0.135     | 0.685     | 0.581     |
| P            | 0.935 | 0.061     | 0.373     | 0.387     |
| S            | 0.984 | 0.893     | 0.878     | 0.855     |
| Cl           | 0.957 | 0.364     | 0.947     | 0.883     |
| K            | 0.994 | 0.303     | 0.988     | 0.649     |
| Ca           | 0.993 | 0.924     | 0.959     | 0.959     |
| Sc           | 0.965 | 0.945     | 0.970     | 0.952     |
| Ti           | 0.986 | 0.859     | 0.917     | 0.810     |
| V            | 0.694 | -0.003    | 0.015     | -0.002    |
| Cr           | 0.992 | 0.910     | 0.883     | 0.851     |
| Mn           | 0.929 | 0.843     | 0.482     | 0.700     |
| Fe           | 0.992 | 0.654     | 0.970     | 0.905     |
| Co           | 0.904 | 0.566     | 0.629     | 0.078     |
| Ni           | 0.980 | 0.467     | 0.950     | 0.670     |
| Cu           | 0.964 | 0.929     | 0.915     | 0.767     |
| Zn           | 0.969 | 0.549     | 0.932     | 0.943     |
| Sn           | 0.876 | 0.144     | 0.909     | 0.576     |
| Ba           | 0.169 | 0.029     | 0.102     | 0.009     |
| *W-L $\beta$ | 0.667 | 0.471     | 0.621     | 0.535     |

機械学習①, ②, ③については本文説明を参照.

表4. TXRF310 誤検出件数 (全1,634点中).

| 元素 | 従来法 | 機械学習<br>① | 機械学習<br>② | 機械学習<br>③ |
|----|-----|-----------|-----------|-----------|
| P  | 52  | 271       | 54        | 461       |
| S  | 448 | 293       | 177       | 62        |
| Cl | 420 | 397       | 300       | 261       |
| K  | 147 | 176       | 111       | 387       |
| Ca | 112 | 80        | 70        | 128       |
| Sc | 78  | 70        | 47        | 81        |
| Ti | 131 | 247       | 78        | 59        |
| V  | 26  | 8         | 6         | 107       |
| Cr | 156 | 252       | 118       | 6         |
| Mn | 121 | 154       | 71        | 150       |
| Fe | 71  | 102       | 37        | 112       |
| Co | 119 | 110       | 71        | 52        |
| Ni | 156 | 169       | 103       | 103       |
| Cu | 79  | 106       | 58        | 117       |
| Zn | 124 | 103       | 71        | 80        |
| Sn | 309 | 436       | 193       | 87        |
| Ba | 16  | 43        | 20        | 221       |

機械学習①, ②, ③については本文説明を参照.

を防ぐ為にはかなり有効な手法と言える. ただし, ある程度の定量性 (長時間測定結果との相関性) を得るには, 1,000点程度の教師データで再学習させる必要がある. 尚, ここでは示していないが10,000点の教師データを用意して再学習すれば, 3.1節で示したTXRF

3760の場合と同様に, 相関係数, 誤検出件数とも従来法よりは大きく改善した.

#### 4. おわりに

本稿ではまず実際の半導体製造工程の現場で使用されているSWEEPING-TXRFについて紹介し, 半導体の技術進歩によりTXRFに対する高感度化の要求が年々高まってきていることを述べた. それに対して, 機械学習を応用することで特に短時間測定での微小ピークに対する検出性能向上の可能性を調査した.

機械学習では教師あり学習の一つである, 畳み込みニューラルネットワーク法を適用した. 教師データには5秒のTXRF測定で得られたスペクトルと, 同一点を60秒測定して得られたデータを解析した結果を組にして, 含まれる元素, 含有量の関係性を学習させた. その上で, 教師データとは独立に用意した検証用データから含まれる元素と含有量を推定させ, その効果を調べた.

まず, 1万点以上の十分な教師データを用意して学習させた場合, 機械学習による推定値は長時間測定で得た結果と良い相関が得られることが分かった. また, 短時間測定での微小ピークで一般に問題となる誤検出の件数は従来法よりも大いに改善した.

次に, 上で学習したモデルを他装置に適用することを試みた. 再学習しない場合や100点の少数教師データで再学習させた場合は, 長時間測定の結果との相関性は従来法に比べてかなり劣っていた. 一方, 誤検出件数については100点の教師データでもある程度の改善が見られた. 教師データ数が1,000点以上あれば, 相関係数でも誤検出件数においても有効であった.

上記の結果より, 教師データが十分無い場合は, 機械学習と従来法を組み合わせた方法が有効な可能性がある. 例えば, 微小ピークにおいては機械学習法を適用して誤検出を減らし, ピークがある程度大きい場合は従来法に切り替える. または, 機械学習の結果を従来法で解析する時の初期値として使用するなどが考えられる. これらのハイブリッド法の有効性については今後さらに調査を進めていく.

今回は, 他装置に適用する時の再学習法として3.2節で述べたファインチューニング法の結果を示した. 一部のネットワークのみ学習させる転移学習法についても今後検証していきたい.

#### 参考文献

- 1) H. Aiginger: Spectrochim. Acta Part B, **46**, 1313 (1991)
- 2) H. Schwenke, P. A. Beaven, J. Knoth: J. Anal. Chem., **365**, 19 (1999)
- 3) D. Hellin, S. D. Gendt, N. Valckx, P. W. Mertens, C. Vinckier: Spectrochim. Acta Part B, **61**, 496 (2006)

- 4) Y. Mori, Adv. X-ray Anal.: **45**, 253 (2002)
- 5) 森良弘, 上村賢一, 河野浩, 山上基行, 清水康裕, 鬼塚義延, 飯塚悦功: X線分析の進歩, **36**, (2005)
- 6) A. Krizhevsky, I. Sutskever, G. E. Hilton: Proceedings of the 25th international conference on neural information processing systems, **1**, 1097 (2012)
- 7) C. Dong, C. C. Loy, K. He, X. Tang: IEEE Trans. Pattern. Anal. Mach. Intell., **38**, 2 (2016)
- 8) TensorFlow<sup>TM</sup>: An open source machine learning library for research and production. <https://www.tensorflow.org/>.
- 9) R. L. T. Hahnloser: On the piecewise analysis of networks of linear threshold neurons. Neural Netw., **11**(4), 691–697. (1998)
- 10) J. W. Muller: Nucl. Instrum. Methods, **112**, 27 (1973)
- 11) K. Omote: Nucl. Instrum. Methods Phys. Res. Sect. A, **293**, 582 (1990)
- 12) Diederik P. Kingma, Jimmy Ba: Adam: A Method for Stochastic Optimization. ICLR (Poster) 2015, arXiv:1412.6980v9
- 13) Rigaku Corporation : <http://www.rigaku.co.jp>