

混合分布モデルを利用した 半導体露光装置の位置合わせにおける異常値検出 Outlier Rejection with Mixture Models in alignment for Lithography

中島 伸一*
Shinichi Nakajima

金谷 有歩†
Yuhō Kanaya

馬込 伸貴‡
Nobutaka Magome

Abstract: Exposure tools such as *Stepper* and *Scanner* are the key component for lithography area of semi-conductor device manufacturing. The measurement for alignment is required very high accuracy of according to the development of device shrinkage then outlier in measurement becomes an important issue. This report introduces an algorithm utilizing normal mixture models to reduce the influence of outliers. The Maximum Penalized Likelihood (MPL) method is also used to determine the freedom of the models minimizing the *rework* ratio in the production field.

1 はじめに

本報告では、半導体デバイスの製造に使用される半導体露光装置の、位置合わせにおける異常値判別アルゴリズムについて紹介する。

第2章では背景として、リソグラフィー技術、半導体露光装置とその位置合わせについて概説し、第3章では混合正規分布と本適用において課した制約について、第4章ではパラメータ推定法について、第5章ではモデル選択法について述べる。第6章ではシミュレーションによる評価結果を示し、第7章をまとめとする。

2 半導体露光装置について

半導体デバイスは以下のような手順で製造される。

Si 単結晶の棒を結晶化技術により作成し、スライスする（これは Si ウェハと呼ばれ、現在主流のものは直径 12 インチ（約 300mm）である。）ウェハ上に感光剤（レジスト）を塗布し、半導体露光装置によって、マスク（レチクル）上に描画された回路パターンをウェハ上に縮小投影する。1 回の露光で転写されるのは約 20mm 角

の領域（ショット）であるが、ウェハを載せたステージをステップ&リピート（ステッパの名前の由来である）して露光を繰り返すことにより、約 100 ショットの同じパターンがウェハ上に転写される。

露光が完了したウェハは現像された後、エッチング、金属蒸着、ドーピング、研磨などの加工（プロセス）が施される。感光剤塗布、露光、加工の一連の工程を数十回繰り返すことにより、ウェハ上にはトランジスタ、コンデンサ、配線などが立体的に形成されていく。

露光によって形成されるパターンの典型的な最小線幅は現在 150nm 程度であるため、層間の位置ずれは数十 nm 程度以下でなければならず、高い重ね合わせ（アライメント）精度が要求される。

位置合わせを行うために、マスク上には少なくともひとつの位置合わせ用マークが描画されており、半導体デバイスの回路パターンとともに転写、加工され、ウェハ上に形成される。露光装置は、ウェハ上のいくつかのショットについて前工程で形成されたマークを計測することにより、ショット配列を推定する。通常、10 個前後のマークが計測される。

ウェハは、配列推定のための計測を行う時点ですでにラフ位置合わせされているため、補正すべき量は微小である。そのため、ウェハ座標に関して線形な成分、すなわち x 方向、 y 方向の 2 軸に関する並進、回転、倍率のみが推定される。配列のモデル式は以下で表現される。

$$\begin{pmatrix} X + \Delta x \\ Y + \Delta y \end{pmatrix} = \begin{pmatrix} 1 + S_X & -R_Y \\ R_X & 1 + S_Y \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} O_X \\ O_Y \end{pmatrix} \quad (1)$$

*東京工業大学総合理工学研究科, 〒 226-8503 神奈川県横浜市緑区長津田 4259

Tokyo Institute of Technology, 4259 Nagatsuda, Midori-ku, Yokohama-shi, Kanagawa, 226-8503 Japan

(株)ニコン, 〒 360-8559 埼玉県熊谷市大字御稜威ヶ原 201-9
Nikon Corporation, 201-9 Miizugahara, Kumagaya-shi, Saitama, 360-8559 Japan

tel. 048-533-4709, e-mail nakajima.s@cs.pi.titech.ac.jp

† (株)ニコン, Nikon Corporation, kanaya.uho@nikon.co.jp

‡ (株)ニコン, Nikon Corporation, magome@nikon.co.jp

S_X 及び S_Y は各軸方向の倍率, R_X 及び R_Y は各軸の回転, O_X 及び O_Y は各軸方向の並進に対応するパラメータである. 露光装置は, ウェハ上のマーク計測結果からこれらのパラメータを最小二乗法によって推定し, 推定配列に従って重ね合わせ露光を行う.

3 混合正規分布モデル

前章で述べた方法で半導体デバイスを大量に生産する状況において, 位置合わせ用マークが常に正常に形成されるとは限らない. マーク計測精度は, これを形成するための前工程の加工方法などに強く依存する. 加工方法によっては, マークの損壊などにより, 計測データに異常値が含まれる場合がある. そこで, 異常値の影響を除外すべく, 混合正規分布を使用する.

$$\phi(x_i) = \sum_{k=1}^G \pi_k N_k(x_i; \mu_k, \sigma_k^2) \quad (2)$$

ここで, x_i は観測値, $N_k(x_i; \mu_k, \sigma_k^2)$ は第 k 番目の正規分布の確率密度関数 (平均値 μ_k , 分散 σ_k^2) であり, G は成分数である. また, π_k は観測値が第 k 番目の成分に属する確率であり,

$$\pi_k \geq 0, \quad \sum_{k=1}^G \pi_k = 1 \quad (3)$$

を満たす. なお, 露光装置における配列推定を実際に行う際には, 観測値は 2 変数 ($\Delta x, \Delta y$) であり, 平均値 μ_k は, 式 (1) で表現されるウェハ上の座標の関数であるが, 以下では単純化して 1 次元の平均値推定問題として取り扱う. 式 (1) がパラメータに関して線形であり, ($\Delta x, \Delta y$) の相関も特に考慮する必要がないため, パラメータ数が異なることを除けば, 以下の議論は実問題にそのまま適用される.

n 個の観測値に対するこの分布の対数尤度は

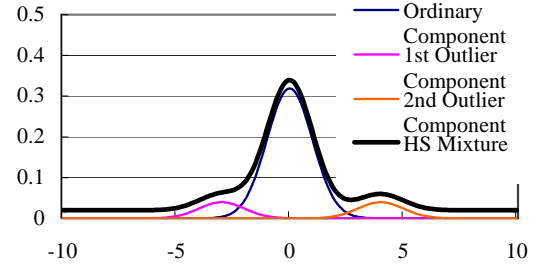
$$L = \sum_{i=1}^n \log \left(\sum_{k=1}^G \pi_k N_k(x_i; \mu_k, \sigma_k^2) \right) \quad (4)$$

であるが, 対数の中に和を含むので扱いにくい. もしも, 個々の観測値がどの成分に属するかが既知であったとするなら, 対数尤度は以下のような取り扱い易い形で表現できる.

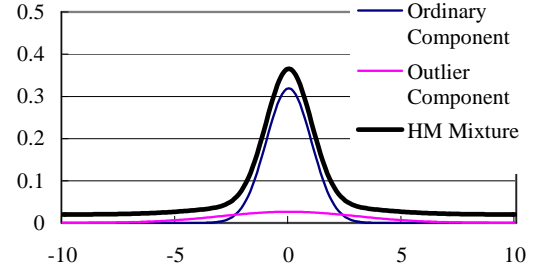
$$L_c = \sum_{k=1}^G \sum_{i=1}^n z_{ik} \log (\pi_k N_k(x_i; \mu_k, \sigma_k^2)) \quad (5)$$

ただし, 個々の観測値の所属ラベルを

$$z_{ik} = \begin{cases} 1 & (\text{if } x_i \text{ belongs to component } k) \\ 0 & (\text{otherwise}) \end{cases} \quad (6)$$



(a) Homo-Schedastic Model



(b) Homo-Mean Model

図 1: 2 種の制約における混合正規分布の確率密度関数の例 (見やすさのため, 混合分布はわずかに上方にシフトさせて描画している.)

で表した.

観測値とラベルのセット $y_i = (x_i, (z_{i1}, \dots, z_{iG}))$ が得られた時, これを完全データと呼び, 式 (5) を完全データの対数尤度と呼ぶ. 一方, 式 (4) を不完全データの対数尤度, あるいは混合分布の周辺対数尤度と呼ぶ. ラベルは観測されないため, 後者がこの分布の対数尤度であるが, 第 4 章で述べるようにその計算に式 (5) が利用される.

次章で, このモデルのパラメータ推定法について議論するが, 混合正規分布においては, すべての成分の平均と分散を自由に最尤推定すると, 平均値がある観測値に一致し, 分散を 0 とするような成分により尤度が発散してしまう. そのため, 以下のような 2 種類の制約のいずれかを課したモデルを使用する.

3.1 共通分散モデル (Homo-Schedastic Model)

異常値は, 正常値と同じ母集団に属する統計的誤差に大きいオフセットが重畳されたものと考える. すなわち分散は全成分共通であり, 確率密度関数は

$$\phi(x_i) = \sum_{k=1}^G \pi_k N_k(x_i; \mu_k, \sigma^2) \quad (7)$$

と表される. このモデルでは, 異常値に対応する成分はひとつの異常な観測値に対応する. よって成分数は $G = 1 + (\text{異常値数})$ である (図 1 (a))

3.2 共通平均モデル (Homo-Mean Model)

このモデルでは、正常値と異常値それぞれに対応する2成分が存在し、これらが共通の平均値を持つと仮定する。確率密度関数は

$$\phi(x_i) = \sum_{k=1}^G \pi_k N_k(x_i; \mu, \sigma_k^2) \quad (8)$$

である。このモデルでは、成分数 G は1あるいは2である (図1 (b))

4 パラメータの推定

パラメータ推定法として、以下に挙げる4つの方法を検討した。紙面の都合上、第1の方法についてのみ詳細に述べ、他の方法は概念のみ記述する ([1] に詳しく記述している。)

4.1 ラベル決定 (Deterministic Labeling 法)

尤度が最大となるラベル付けを行う。異常値の個数についても、第5章に述べる評価規範によって最適なものを選択する。具体的なアルゴリズムを、前章の2種の制約モデルについて以下に示す。

4.1.1 HS モデルの場合

HS モデルでは、異常値に対応する成分にはそれぞれ唯一の観測値が属すると仮定し、以下の手順でパラメータを決定する。

1. $s = 1, G = s$ とし、すべての観測値が第1成分 (正常値) に属するとする。すなわち

$$z_{ik} = \delta_{k1} \quad (1 \leq i \leq n) \quad (9)$$

ただし、 δ_{ab} はクロネッカーのデルタである。

2. z_{ik} のもとで、パラメータ π_k, μ_k, σ の最尤推定値を求める。

$$\hat{\pi}_k = \frac{1}{n} \sum_{i=1}^n z_{ik} \quad (10)$$

$$\hat{\mu}_k = \frac{1}{n\hat{\pi}_k} \sum_{i=1}^n z_{ik} x_i \quad (11)$$

$$\hat{\sigma} = \frac{1}{n} \sum_{i=1}^n z_{i1} (x_i - \hat{\mu}_1)^2 \quad (12)$$

3. 完全データ尤度を計算する。

$$\begin{aligned} L_c &= \sum_{k=1}^G \sum_{i=1}^n z_{ik} \log(\hat{\pi}_k N_k(x_i; \hat{\mu}_k, \hat{\sigma}^2)) \\ &= n\hat{\pi}_1 \log(n\hat{\pi}_1) - n \log \hat{\sigma} + \text{const.} \end{aligned} \quad (13)$$

4. 第1成分に属する観測値の中で残差が最大のものを選択し、次の異常値候補とする。その観測値に対応する異常値成分を追加する。

$$z_{jk} = \delta_{k(G+1)} \quad (14)$$

ただし

$$j = \operatorname{argmax}_{\{i; z_{i1}=1\}} |x_i - \hat{\mu}_1| \quad (15)$$

5. $s = s + 1, G = s$ を代入し、2. ~ 5. を $s < s_{\max} + 1$ の間繰り返す。
6. 第5章で述べる評価規範が最適となるモデル s を最尤なモデルとして選択する。

求める値 (ショット配列) は、最終的に選ばれたモデルの正常値の平均値 $\hat{\mu}_1$ である。また本報告においては、最大 $n/2$ 個の異常値に対応すべく、 $s_{\max} = n/2 + 1$ とした。

4.1.2 HM モデルの場合

HM モデルでは、成分数 G は1あるいは2である。以下の手順でパラメータを推定する。

1. 最初のモデル ($s = 1$) は、HS モデルの場合と同じである。
2. $s = 2, G = 2$ として、この時点ではすべての観測値を第1番目に属するとする。
3. ($s - 1$) 番目のモデルの第1成分 (正常値) に属する観測値の中で最も残差の大きいものを、 s 番目のモデルにおける異常値候補として、第2成分 (異常値) に移動させる。すなわち

$$z_{i1}^{\{s\}} = z_{i1}^{\{s-1\}} - \delta_{ij} \quad (16)$$

$$z_{i2}^{\{s\}} = z_{i2}^{\{s-1\}} + \delta_{ij} \quad (17)$$

ただし

$$j = \operatorname{argmax}_{\{i; z_{i1}^{\{s-1\}}=1\}} |x_i - \mu^{\{s-1\}}| \quad (18)$$

4. z_{ik} のもとで、パラメータ π_k, μ, σ_k の最尤推定値を求める。

$$\hat{\pi}_k = \frac{1}{n} \sum_{i=1}^n z_{ik} \quad (19)$$

$$\hat{\mu} = \frac{\sum_{k=1}^G \sum_{i=1}^n \frac{z_{ik} x_i}{\sigma_k^2}}{\sum_{k=1}^G \sum_{i=1}^n \frac{z_{ik}}{\sigma_k^2}} \quad (20)$$

$$\approx \frac{1}{n \hat{\pi}_1} \sum_{i=1}^n (z_{i1} x_i) \quad (21)$$

$$\hat{\sigma}_k^2 = \frac{1}{n \hat{\pi}_k} \sum_{i=1}^n z_{ik} (x_i - \hat{\mu})^2 \quad (22)$$

5. 完全データに対する対数尤度を求める．

$$L_c = \sum_{k=1}^G \sum_{i=1}^n z_{ik} \log(\hat{\pi}_k N_k(x_i; \hat{\mu}, \hat{\sigma}_k^2)) \\ \approx \sum_{k=1}^G \{n \hat{\pi}_k \log(n \hat{\pi}_k) - n \hat{\pi}_k \log \hat{\sigma}_k^2\} + \text{const.} \quad (23)$$

6. $s = s + 1$ を代入し, 3. ~ 6. を $s < s_{\max} + 1$ の間繰り返す．

7. 第 5 章で述べる評価規範が最適となるモデル s を最尤なモデルとして選択する．

正常値及び異常値に共通な平均値 $\hat{\mu}$ が, 求める値 (シヨット配列) に対応する．式 (21), 式 (23) における近似は, 第 2 成分 (異常値) の分散が, 第 1 成分 (正常値) の分散より十分大きい ($\sigma_1 \ll \sigma_2$) という仮定に基づく．この仮定が十分満たされないようなモデルの尤度を過小評価することになるが, そのようなモデルは尤度そのものが小さいので, 最終結果に大きな影響を及ぼすことはない．なおこの仮定により, 求める配列は異常値と判別された観測値からの影響を受けない．

4.2 EM アルゴリズム (EM 法)

第 2 の方法では, 不完全データが得られる場合の最尤推定に広く用いられる EM アルゴリズム [5] を使用する．EM アルゴリズムは, 式 (5) において z_{ik} をその暫定的な事後確率で置き換えたものを逐次最小化していく方法であるが, その最小化 (M) ステップは, 式 (10), 式 (11) 及び式 (12), あるいは式 (19), 式 (20) 及び式 (22) と似た簡単な計算で実現される．[1]

4.3 EM アルゴリズムを用いたラベル決定法 (DLEM 法)

第 3 の方法では, EM アルゴリズムによって分布を推定した後, 観測値のラベルを事後確率 (z_{ik} の期待値) に基づいて決定する．その上で正常値のみから平均値 (配列) を推定する．

4.4 ベイズ推定的方法 (Simplified Bayes 法)

第 4 の方法では, いくつかの異常値数に対し, 対応するすべての可能なラベル付けについてパラメータを推定し, それらをモデルの尤度で重み付け平均したものを最終的な推定値とする．

5 モデル選択

異常値を判別するにあたって, パラメータの自由度に対応する成分数 G を決定することが必要となる．正則モデルにおいて導出された AIC[3] や BIC[4] は, 混合正規分布のような非正則なモデルにおいて正当性を持たないが, そのことを認識した上でいくつかの評価報告がなされている [2]．我々は, 形式的にこれら情報量規準の拡張とみなせる, ペナルティー付き最尤推定法 (Maximum Penalized Likelihood) 法 [6] を用いる．ペナルティーの課し方は, パラメータ数に比例する形に限定する．すなわち,

$$PL = L - d \cdot \beta(n) \quad (24)$$

を最大にするモデルを選択する． L は対数尤度, d はモデルのパラメータ数, $\beta(n)$ はペナルティーの強さに対応する係数であり, $\beta(n) = 1$ で AIC, $\beta(n) = (\log) / 2$ で BIC をそれぞれ最小化する方法と等価になる．本適用では, 次章で述べるシミュレーション結果に基づいて, 不良品率に対応する評価関数を最小化する $\beta(n)$ を採用する．

6 評価

本適用の評価データとして, 実際の半導体デバイス量産時のデータを用いることが望ましい．しかし通常, 異常値の出現頻度は低く, さらには異常値の発生が不良品の製造に直結する場合, この状況は放置されることなく, 別の手段で改善がなされることが多い．例えば, マークが高い頻度で損壊するような状況では, その工程におけるウェハの加工温度を高精度で制御するなどの対策が施される．本適用は, そのような人的労力を低減することを目的としているので, 評価に用いるために十分な数のデータを蓄積することは現状では困難である．そこで本報告では, シミュレーションによって異常値を含む観測値を発生させ, アルゴリズムの評価を行った．

6.1 発生させる観測値の分布

平均値 0, 分散 V^2 の正規分布を正常値の分布とし, $[-T, T]$ の一様分布を異常値の分布とする．異常値の発生率を Q とすると, このデータの確率密度関数は以下

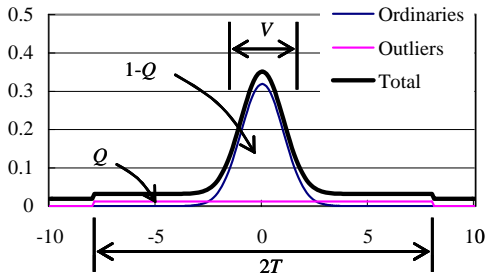


図 2: シミュレーションデータの確率密度関数.

parameters		values			
Q	異常値発生率	0	0.001	0.01	0.1
T	異常値振幅	—	10	100	1000
V	正常値標準偏差	—	5	10	20

表 1: 「状況」パラメータの値. すべての意味のある組み合わせに対応する「状況」が等確率に生起すると仮定する. T 及び V の単位を nm とすると, 現実の露光装置の位置合わせにおけるスケールとほぼ一致する.

のようになる (図 2)

$$\psi(x) = (1 - Q) N(x; 0, V^2) + \frac{Q}{2T} I(|x| < T) \quad (25)$$

ただし, $I(\text{condition})$ は, condition が真の時 1, 偽の時 0 を取る関数である. 本報告では Q, V 及び T を「状況」パラメータと呼ぶ.

「状況」パラメータは, 第 3 章のはじめに述べたように, 前工程で形成されるマークの加工方法に依存すると考えてよい. 半導体露光装置には通常, いろいろなデバイス品種, 工程のウェハを, ロット (数十枚) 単位で不規則に処理するようなタスクが与えられる. 従って, 「状況」パラメータがランダムに変化する状況を想定すれば良い. しかし, 「状況」パラメータの分布は全く未知であるため, 本報告ではこれらの分布が現実起こりうると思われる範囲で一様であると仮定し, 具体的には表 1 に示した値の, 意味のあるすべての組み合わせが等確率で起こるものとする. T 及び V の単位を nm とすると, 現実の露光装置のアライメントにおけるスケールとほぼ合致する.

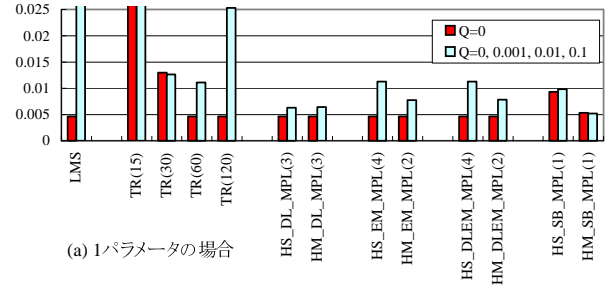
6.2 評価規範

アルゴリズムの良さを判断する基準として, 不良品発生率に関連する以下の値 (*Error Rate*) を用いる.

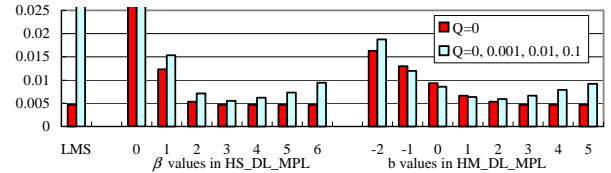
$$ER = p(|\hat{\mu} - \mu_0| > 3\sigma_{\mu}^{\text{ideal}}) \quad (26)$$

ただし

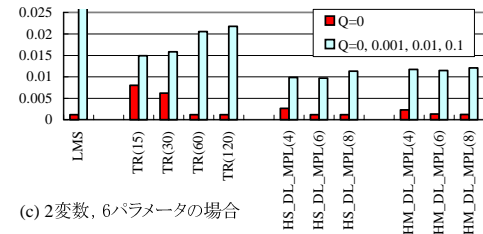
$$\sigma_{\mu}^{\text{ideal}} = V/\sqrt{Q \cdot n} \quad (27)$$



(a) 1パラメータの場合



(b) DL法の β 依存性



(c) 2変数, 6パラメータの場合

図 3: シミュレーション評価結果. 縦軸はすべて ER 値.

であり, これはラベルが既知の状況で最尤推定を行った場合のパラメータ推定誤差の期待値に対応する. ER は, ラベルが既知であるという理想的な推定精度の 3 倍を超えた時, 不良品が発生するという仮定に基づいた不良品率として定義されている.

6.3 評価結果

図 3 (a) に, 第 3 章で述べた 2 種のモデル (制約) 及び第 4 章で述べた 4 種のパラメータ推定法に基づいたアルゴリズムの評価結果を示す. サンプル数 $n = 12$ である. 図には, これら以外に以下の 2 種のアルゴリズムの評価結果を比較のために示した.

LMS 法: 最小二乗法 (単一正規分布による最尤推定法)

TR(h) 法: 閾値による異常値除去法 (Threshold Rejection). LMS で配列を推定した後, 残差が閾値 h を越えるものを異常値として除外し, 残りの観測値によって配列を推定しなおす. これを, 最大残差が h 以下になるまで繰り返す. 半導体露光装置ですでに使用されている従来法である. 閾値 15, 30, 60 はそれぞれ表 1 中の V の値の 3 倍であり, それぞれの「状況」に対して適切な値と考えてよい.

縦軸は ER 値であり, 異常値が発生しない「状況」($Q = 0$) のみの期待値と, すべての「状況」に対する期待値

が示されている。前節で述べたような「状況」パラメータの組み合わせが等確率に生起する場合を想定したものは後者である。前者については、以下の理由で評価に加えた。

Q は未知であり、デバイス品種、工程に依存する。 Q がほぼ 0 とみなせる「状況」も少なくなく、このとき LMS はほぼ最良の推定法と考えてよい。新アルゴリズムが、実際の製造現場で「デフォルト」アルゴリズムとして使用されるためには、最も LMS 法に取って有利な状況においてもこれに対して大きく見劣りするようなことがあってはならない。そのため、 $Q = 0$ の場合に LMS に対して大きく劣らないことを確認し、後述するペナルティ係数の決定においてもこの点に配慮した。なお、 $Q = 0$ の場合の LMS 法による ER の理論値は約 0.004 である。

DL 法、EM 法、DLEM 法及び SB 法では、モデル選択に MPL 法を用いており、図中の項目名の括弧内の値は式 (24) における β である。SB 法を除いて、この値は ER 値による評価結果が最良となるものを選んで図示した。例として、DL 法の β 依存性を図 3 (b) に示す。HS、HM の両方のモデルにおいて、選択された (a) で図示された $\beta = 3$ が (Q に関する前述の 2 つの観点で) 良好な結果を示していることがわかる。

SB 法に関しては $\beta = 1$ の場合についてのみ評価した。これは、AIC をモデルの対数尤度 (の符号反転の 2 倍) として利用することに相当し、本報告で評価したアルゴリズムの中で最も良い結果を示すことは注目に値する。しかし、処理時間が製品仕様に適合しないため、後述する最後の評価からは除外した。

EM 法及び DLEM 法を DL 法と比較すると、EM アルゴリズムを用いて分布の推定精度を上げて、推定精度の向上につながらないことが分かる。従ってこれらも最後の評価からは除外した。

最後に、実際の露光装置の位置合わせと同じ問題、すなわち式 (1) により、2 つの観測値 ($\Delta x, \Delta y$) を 2 つの説明変数 (X, Y) で回帰する 6 パラメータモデルの推定におけるシミュレーション結果を図 3 (c) に示す。この場合は ER 値として、推定配列から予測される Δx あるい Δy のいずれかが基準値 (式 (27) の 3 倍) を超える確率をとった。HS_DL 及び HM_DL について、それぞれの 3 つの値の結果をプロットした。最適な値はいずれも $\beta = 6$ と判断される。混合正規分布を用いる方法は、従来法である TR 法と比較して優位であるが、その優位性は 1 パラメータの時 (図 3 (a)) より小さい。これは、 $n = 12$ という小サンプルであるにもかかわらず、推定パラメータ数が大きいことに起因すると思われる。実際、

より少数のサンプル数においては同等、あるいは不利に転ずることもあることが確認された。

7 まとめ

本報告では、混合正規分布モデルを用い、ペナルティ付き最尤推定法による自由度選択を行うことにより、半導体露光装置の位置合わせにおいて異常値を検出する方法を紹介した。本報告で論じたような小サンプルからモデルを推定する問題では、EM アルゴリズムが予測精度向上につながらないこと、パラメータの自由度を増やすに従って、TR 法 (これは正常値の分散を既知として異常値を判別していることに相当する) に対する優位性が減少することを確認した。

また、ベイズの推定法の振る舞いが良好であることが確認された。ベイズ推定法を用いた場合には、モデル選択法として確率的複雑さを使用することができる [7]。これらを今後の課題としたい。

参考文献

- [1] S. Nakajima et al., "Outlier rejection with Mixture Models in Alignment," *Proc. of SPIE*, Vol.5040, pp.1729–1741, March 2003.
- [2] G. McLachlan and D. Peel, *Finite Mixture Models*, Wiley, New York, 2000.
- [3] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Trans. on Automatic Control*, Vol. 19, pp.716–723, 1974.
- [4] G. Schwarz, "Estimating the Dimension of a Model," *Annals of Statistics*, Vol.6, No.2, pp.461–464, 1978.
- [5] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum Likelihood for Incomplete Data Via the EM Algorithm," *J. R. Statistical Society*, Vol.39-B, pp.1–38, 1977.
- [6] B. G. Leroux and M. L. Puterman, "Maximum Penalized Likelihood Estimation for Independent and Markov-Dependent Mixture Models," *Biometrics*, Vol.41, pp.545–558, 1992.
- [7] 渡辺澄夫, データ学習アルゴリズム, 共立出版, 東京, 2001.