

リソグラフィホットスポット検出における 特徴量評価用データセット生成手法の改良

山本 真大^{1,a)} 稲木 雅人^{1,b)} 永山 忍^{1,c)} 若林 真一^{1,d)} 児玉 親亮^{2,e)}

概要: リソグラフィプロセスにおいて、異常な短絡や開放を引き起こす確率の高い回路パターンであるホットスポットは、設計の段階で検出し除去することが望ましい。そこで近年、高速なホットスポット候補の検出手法として機械学習を用いた手法が複数提案されている。しかし、検出手法の高精度な比較評価ができる評価用データセットが、公開されたデータとして存在しない。そこで既存研究として、既存データセットに修正を加えて光学シミュレーションを適用することで訓練・テストデータの追加生成を行い、データセットを増強する手法が提案されている。本研究では、この光学シミュレーション後のラベル付け工程において、短絡や開放が生じているにもかかわらず非ホットスポットとラベル付けされる等の問題を発見し、このラベル付け工程を改良してデータセットの信頼性を向上させた。さらに、このデータセットを用い、検出手法の比較実験を行った。

Improvement of Dataset Generation Method for Evaluating Features in Lithography Hotspot Detection

MASAHIRO YAMAMOTO^{1,a)} MASATO INAGI^{1,b)} SHINOBU NAGAYAMA^{1,c)} SHIN'ICHI WAKABAYASHI^{1,d)}
CHIKAAKI KODAMA^{2,e)}

Abstract: In the lithography process, a hotspot refers to a circuit pattern that has a high probability of causing undesired open/short-circuits, which should be detected and removed at the design stage. In recent years, for fast hotspot detection, some machine learning-based methods have been proposed. However, there is no open benchmark dataset that can accurately evaluate and compare hotspot detection methods. Therefore, in a study, a method that enhances an existing open benchmark dataset was proposed. It modifies the dataset and applies optical simulation to generate additional training/test data. In this study, we found some problems in its simulation-based labeling process, such as labeling a pattern causing open/short-circuits as a non-hotspot. We improved the process and thus the reliability of the generated dataset. In addition, using the enhanced dataset, we conducted experiments to evaluate and compare some hotspot detection methods.

1. はじめに

年々、より高性能な半導体デバイスが求められるようになっており、それを実現するため半導体の微細化が進んでいる。半導体製造工程の一つであるリソグラフィでは、露

光装置から回路パターンの原板であるフォトマスクにレーザー光を照射し、感光性物質が塗布された半導体ウェハにパターンを転写する。この工程において、回路の微細化に伴い、フォトマスクを通過した光の回折による転写パターンの忠実度低下の影響が大きくなっている [1]。転写の忠実度低下により短絡や開放といった不具合を生じる確率が高いパターンをホットスポットと呼ぶ。ホットスポットは半導体製品の不具合につながるため、設計の段階で検出できるだけ除去する必要がある。

設計の段階でホットスポットを精密に検出する方法としてリソグラフィシミュレーションがあるが、レイアウト全

¹ 広島市立大学 大学院 情報科学研究科

〒 731-3194 広島県広島市安佐南区大塚東 3-4-1

² キオクシア株式会社 〒 247-8585 横浜市栄区笠間 2-5-1

a) m.yamamoto@lcs.info.hiroshima-cu.ac.jp

b) inagi@hiroshima-cu.ac.jp

c) s.naga@hiroshima-cu.ac.jp

d) wakaba@hiroshima-cu.ac.jp

e) chikaaki1.kodama@kioxia.com

体に適用すると非常に時間を要する。事前にホットスポットとなる可能性の高いパターンを高速に発見できれば、シミュレーションの適用範囲や適用回数が削減でき、設計期間の短縮に繋がる。

そこで近年、高速にホットスポットを検出する手法として、機械学習に基づく手法が研究されている [2][3][4][5]。これらの手法では、訓練データとして既知のホットスポットと非ホットスポットのパターンを学習させ、学習したホットスポットの特徴を基にホットスポットを検出する。機械学習に基づく手法は、学習したデータの特徴をもつ未知のホットスポットを検出できるという利点がある。

機械学習に基づく手法においてホットスポット検出精度を高めるには、ホットスポットの特徴をよく捉えられる特徴量の設計が重要である。ホットスポット検出のための既存の特徴量のうち、ビットマップ画像に基づくものとして、レイアウトパターン上の局所領域ごとの配線の濃度（面積割合）を考慮した DBLF (Density Based Layout Feature) [6]、レイアウト画像の平滑化により光の回折を近似的に捉える HOLP (Histogram of Oriented Light Propagation) [5]、隣接配線間の距離と配線幅を考慮した LiWS (Line Width and Separation) [7] などが提案されている。

以上のようなホットスポット検出手法の評価では、ホットスポット判定手法の評価用回路パターンデータセットである ICCAD2012 contest データセット [8]（以下、既存データセット）が広く使われているが、どのような条件でホットスポット/非ホットスポットのラベル付けがされたか不明であること、訓練・テストデータ数が一般的な機械学習向けのデータセットに比べて少ないこと等の課題がある。

そこで、既存データセットの課題を解決し、より正確にホットスポット検出するため、片岡らは a) ICCAD2013 contest にて提供された光学シミュレータ [9] でラベル付けし直すことでラベル付け条件の統一を行い、b) 既存データセットの回路パターンの一部に変更を加えてラベル付けし直すことでデータの追加生成を行った [10]。これにより、ラベル付け条件の統一とデータ数不足の緩和を行うことができた。

しかし、片岡らの手法（以下、既存手法）のラベル付けには以下のような問題があった。既存手法のラベル付けでは、レイアウト画像の島の数（配線層における配線の数）とシミュレーション画像の島の数（結像された配線の数）が一致ならホットスポット、不一致なら非ホットスポットとしている。この場合、短絡と開放が同数発生すると元のレイアウト画像の島の数とシミュレーション画像の島の数が一致し、非ホットスポットとラベル付けされてしまう。また別の場合として、シミュレーション後の配線が元の配線より大きく欠けてしまうような場合でも、島の数さえ一致すれば非ホットスポットとラベル付けされてしまう。

本研究では以上のような場合のラベル付けをより正確に

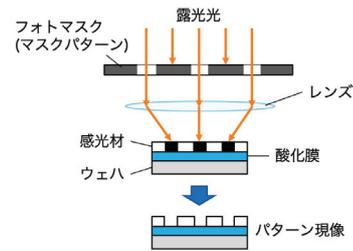


図 1 リソグラフィ [10]

行うために、島の数の一致・不一致の他に、島同士の対応関係と面積を考慮したラベル付けへと改良した。これにより、目視で付与したラベルとの一致度が、既存のラベル付けと比べて大きく向上した。また、改良したラベル付けを行ったデータで特徴量 DBLF, HOLP, LiWS の比較評価を行った。

本稿では、第 2 章でまずリソグラフィとホットスポットについて説明する。第 3 章では、ホットスポット検出問題を定義し、機械学習を用いた検出手法とそのための特徴量について説明する。第 4 章では、既存の配線パターンデータ生成手法について説明する。第 5 章では、第 4 章のラベル付けの問題点を挙げ、それを基に改良したラベル付けを提案する。第 6 章で、提案ラベル付け手法の評価実験の結果を述べ、最後に、第 7 章でまとめを述べる。

2. リソグラフィとホットスポット

リソグラフィプロセス（以下、リソグラフィ）とは、フォトマスクと呼ばれる回路の原板にレーザー光を当て、回路パターンをシリコンウェハ上に転写する（図 1）工程である。シリコンウェハには、予めフォトレジストと呼ぶ感光性樹脂が塗布されており、フォトレジストが光と反応することで回路パターンが転写される。

この工程において、フォトマスクを光が通過した際に回折が発生し、転写される回路パターンの忠実度が低下する。そこで、設計した回路パターン通りに製造できるように OPC (Optical Proximity Correction) [1] や SRAF (Sub-Resolution Assist Feature) [11] などの技術が開発されてきたが、これらを適用してもなお配線の短絡や開放を生じてしまうパターンが存在し、そのようなパターンはホットスポットと呼ばれる（図 2）。リソグラフィシミュレーションを用いれば設計段階でホットスポットを検出できるが、レイアウトパターン全体に対して行う際の計算時間は非常に大きい。そこで、高速にホットスポットの候補を検出できれば、いち早く修正を行うことができ、シミュレーションを行う範囲や回数の削減、設計期間の短縮につながる。

3. 機械学習に基づくホットスポット検出

本節では、まずホットスポット検出問題を定義し、次に機械学習を用いた検出手法の概要と、学習のために用いら

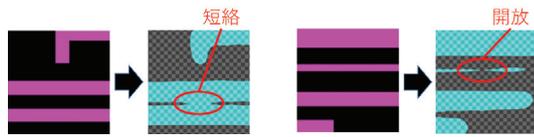


図 2 リソグラフィにおける短絡・開放の発生

れる特徴量について説明する。

3.1 ホットスポット検出問題

本研究で扱うホットスポット検出問題は、レイアウトパターン上から既知および未知のホットスポットパターンを発見する問題である。入力として、既知のホットスポットパターンと非ホットスポットパターン、そして LSI レイアウトパターン全体が与えられる。出力はレイアウトパターン上のホットスポット候補の場所である。

3.2 機械学習による検出方法

機械学習を用いたホットスポットの検出は大きく分けて学習フェーズ、テストフェーズの 2 段階からなる。なお本研究では、レイアウトパターンを配線領域が白（画素値：1）、空白領域が黒（画素値：0）のビットマップ画像として扱う手法を対象とする。

機械学習によるホットスポット検出の流れを図 3 に示す。学習フェーズでは、図 3(a) の様に訓練データとして、複数の既知のホットスポットパターンのレイアウト画像、および複数の非ホットスポットパターンのレイアウト画像が与えられる。まず、訓練データの各レイアウト画像から特徴抽出（図 3(b)）を行い特徴量を得る。各レイアウト画像から抽出した特徴量と、その画像のラベル { ホットスポット：HS, 非ホットスポット：N-HS } を学習モデルに入力する。これにより学習が行われ、ホットスポット検出器が構築される（図 3(c)）。

テストフェーズでは、図 3(d) の様にテストデータとしてレイアウトパターン全体の画像が与えられる。なお、レイアウトパターン上で注目する領域のことを検出窓と呼ぶ。検出窓は訓練データ 1 つと同じサイズである。この検出窓を走査し（図 3(e)）、各位置で特徴を抽出し（図 3(b)）、その後得られた特徴量を検出器に入力する。最終的に、検出器から予測ラベルを出力させ（図 3(f)）、ホットスポット候補の場所を出力する（図 3(g)）。

3.3 レイアウトパターン特徴量

ホットスポット検出のための既存の特徴量として、基本的な特徴量である DBLF[6] と、光の回折を疑似的に考慮した HOLF[5]、配線間距離と配線幅を考慮した LiWS[7] などがある。

DBLF は、検出窓をグリッド状に分割してできた局所領域において、配線が占める割合を考慮した特徴量である。

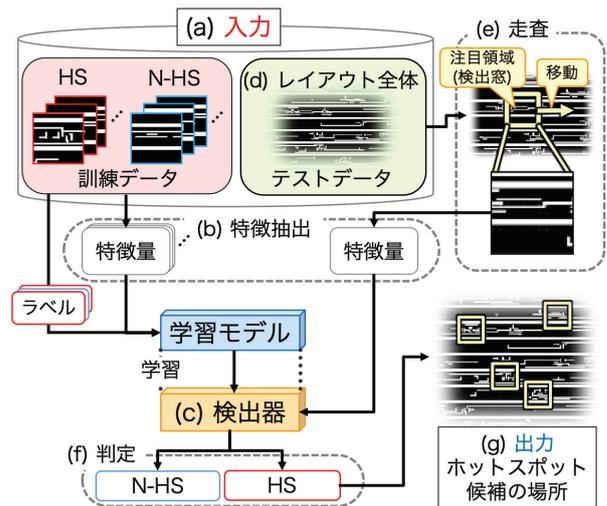


図 3 機械学習に基づくホットスポット検出の流れ [10]

HOLF はレイアウト画像をフィルタを用いて平滑化することで、リソグラフィにおける光の回折を近似的に表現した特徴量である。HOLF では、ガウシアンフィルタを適用し平滑化されたレイアウト画像に対して、DBLF と同様の局所領域内で輝度勾配を求める。そして勾配角度 θ ($0 \leq \theta < 360$) を B 段階で量子化し、勾配強度を重みとしてビン数 B のヒストグラムに投票する。最終的に全ての局所領域で作成したヒストグラムを連結したものが HOLF である。

LiWS は、DBLF や HOLF と同様、検出窓をグリッド状に分割してできた局所領域内で、配線幅および隣接配線間の距離を求める。ここで、 $k \times k$ ピクセルからなる局所領域の各行（各列）において、連続する配線の画素数を配線幅、連続する空白の画素数を配線間距離と考える。LiWS は複数のバリエーションを持ち、最小値ベースと分布ベースの 2 種類に大きく分けられるが、本稿では既存研究で最良の結果となった最小値ベースの LiWS について述べる。最小値ベースの LiWS は、局所領域内の最小配線幅と最小配線間距離のみを考慮している。

4. データセットの問題とデータ生成手法

本節では、まず機械学習に基づくホットスポット検出手法におけるベンチマークとして広く使われている既存の回路パターンデータセットの問題点について説明し、次にその問題を解決・緩和するためのデータ生成手法について説明する。

4.1 既存データセットの問題点

ICCAD2012 contest データセット [8]（以下、既存データセット）は、ホットスポット検出においてベンチマークとして広く使われているデータセットである。

既存データセットでは、レイアウトパターン、ホットス

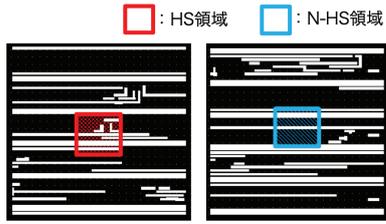


図 4 ICCAD 2012 contest データセット [8] の一部 (GDSII を扱えるエディタで表示) [10]

表 1 ICCAD-2012 contest データセットの内容

回路	HS 領域数	N-HS 領域数	プロセス
A: array_benchmark1	226	319	32nm
B: array_benchmark2	498	4146	28nm
C: array_benchmark3	1808	3541	28nm
D: array_benchmark4	177	3386	28nm
E: array_benchmark5	41	2111	28nm

ポットを内包する矩形 (HS 領域), 内包しない矩形 (N-HS 領域) が与えられる (図 4). HS/N-HS 領域のサイズは $1.2\mu\text{m} \times 1.2\mu\text{m}$ である.

既存データセットは 10 種類の回路からなり, 32nm と 28nm プロセスの回路がある. 本研究では, このうち 5 種類を扱う. 各回路の HS/N-HS 領域の数およびプロセスノードを表 1 に示す.

既存データセットの問題として, データ数が数千程しかなく一般的な機械学習データセットと比較して少ないこと, どのような製造装置や材料の使用を想定した上でラベル付けが成されたか不明瞭であること, 等がある.

4.2 データ生成手法

既存データセットを基に回路パターンの一部を修正し, 光学シミュレーションを用いてラベル付けすることでデータを追加生成する, 片岡らによるデータ生成手法 [10] について説明する. この手法には, データ数を増やすことで機械学習を用いたホットスポット検出手法の精度を向上させること, 高精度な特徴量の評価を行うことという 2 つの目的がある.

まず, 既存データセットを HS/N-HS 領域のパターンで切り出し, ICCAD 2013 contest にて提供された光学シミュレータ [9] の入力ファイル形式で保存する. 既存データセットとこのシミュレータは各種パラメータが異なり, 既存データセットをそのまま入力してもほとんどのケースでパターンが正しく結像されない. 本来は ICCAD 2012 contest データセットと同じパラメータでシミュレーションを行うべきであるが, 両者とも詳細なパラメータが不明なため, このシミュレータで一定割合以上のパターンが正しく結像されるスケールまでパターンを拡大する.*1

*1 先端プロセスで使用される光学的・リソグラフィ的パラメータの適用は今後の課題とする.

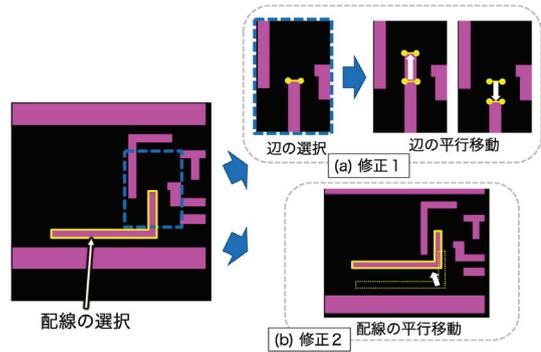


図 5 パターン修正のアイデア [10]

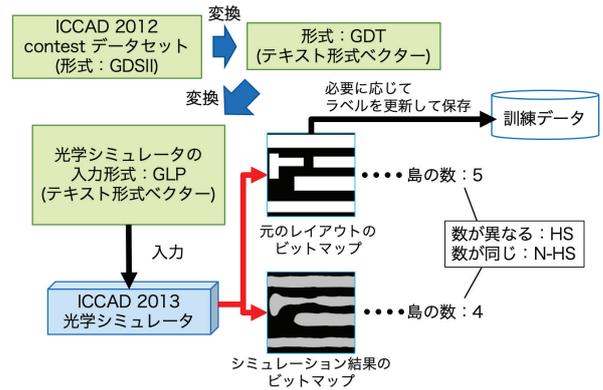


図 6 光学シミュレーションとラベル付けの流れ [10]

次に, 切り出してきたパターンを以下の 2 種類の方法で修正する.

修正 1 HS/N-HS 領域の中心寄りの配線を 1 つ選択し, 一辺を平行移動させ変形する (図 5(a)).

修正 2 HS/N-HS 領域の中心寄りの配線を 1 つ選択し, 配線自体を平行移動する (図 5(b)).

なお, 移動方向はランダム, 移動距離は, 任意の 2 つの値 a, b ($a < b$) の範囲内でのランダム値である.

パターンの修正後, 光学シミュレーションを適用しラベル付けを行う. 光学シミュレーションでは元のレイアウト画像 I_{mask} とシミュレーション後の画像 I_{sim} がビットマップ形式で出力される. その後, I_{mask} および I_{sim} の島の数をカウントし, 両者の島の数が一致なら非ホットスポット, 不一致ならホットスポットとラベル付けする. ここで島とは, 画像の外周もしくは画素値 0 の領域に囲まれた, 画素値 1 の領域とする. 光学シミュレーションとラベル付けの流れを図 6 に示す.

5. 既存のラベル付けの問題点と改良

本研究では, 既存のデータ生成手法のラベル付け工程において, ラベル付けが正しく行われなかった場合を発見した. 本節では, まず発見したラベル付けの問題点について述べ, 次にこれらの問題点を解消・緩和した改良ラベル付け手法を提案する.

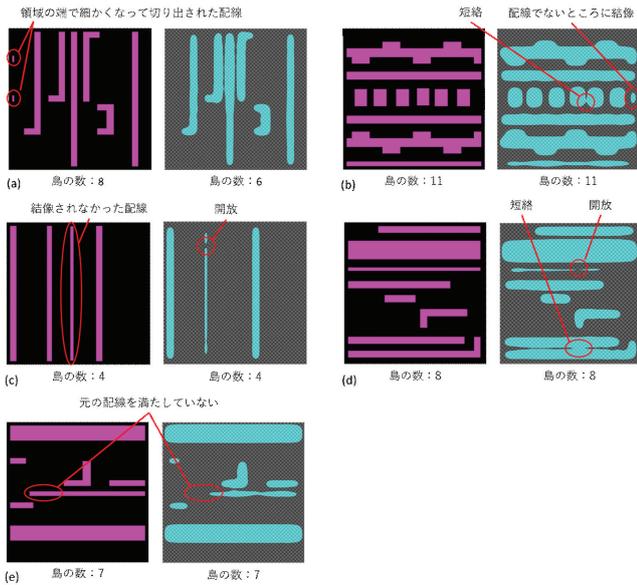


図 7 正しくラベル付けされていない場合：(a)~(e)それぞれの左図はレイアウト画像，右図はシミュレーション画像

5.1 既存のラベル付けの問題点

既存手法 [10] のラベル付け (以下, 既存ラベル付け手法) はレイアウト画像とシミュレーション画像の島の数の一致・不一致のみでホットスポットか非ホットスポットかを判定している。しかし, 島の数のみで判定を行うと, 以下のような場合にラベル付けが正しく行われない。

- (1) 領域の境界付近の配線がわずかな面積で切り出された場合 (図 7(a))
- (2) レイアウト画像上に配線がないところに結像され (島が 1 増加), かつ別の 2 つの配線で短絡が生じる (島が 1 減少) 場合 (図 7(b))
- (3) 1 つの配線が結像されず (島が 1 減少), かつ別の 1 つの配線で開放が生じる (島が 1 増加) 場合 (図 7(c))
- (4) 短絡と開放が同時に生じる場合 (図 7(d))
- (5) シミュレーション後の配線が元の配線の面積を十分に有していない場合 (図 7(e))

(1) の場合, 該当の配線は結像されず, したがって島の数は一致せず, HS とラベル付けされる。しかし, 領域の境界上で切り出された配線は本来はもっと大きな面積の配線であり, 切り出された部分が結像しなくとも必ずしもホットスポットとは言えない。(2) ~ (4) の場合は, 配線の接続性から考えて明らかにホットスポットであるにもかかわらず, 非ホットスポットとラベル付けされる。(5) の場合は, 目視では明らかに不適切な結像の状況でありホットスポットと考えられるが, これも非ホットスポットとラベル付けされる。

5.2 合成画像を用いたラベル判定条件の追加

5.1 節で述べたラベル付けが正しく行われていない場合の (1) は, わずかに領域内部 (検出窓内部) に入った境界

部分の配線はシミュレーションの結果結像されないのは当然であるにもかかわらず, これを判定に含むことで正しくラベル付けできないことが分かる。より正確にラベル付けするために, 提案手法ではこのような配線はラベル付けの判定から除外する。具体的には, 境界から一定以上内側に入っている配線のみを切り出すことにする。本研究では, 境界から少なくとも最小配線幅だけ内側に入り込んでいる配線のみを, 境界で切り出した。なお, ここでの最小配線幅はパターン修正前のデータセットにおけるものである。

5.1 節で述べたラベル付けが正しく行われていない場合の (2), (3), (4) は, 元の配線とシミュレーション後の配線が 1 対 1 の関係でないこと, (5) は 1 対 1 の関係ではあるが面積が不足していることが原因により, レイアウト画像とシミュレーション画像の島の数のみの判定では正しくラベル付けできない。より正確にラベル付けするためには, 島の数の他に島の対応関係と面積を考慮する必要がある。そのために, レイアウト画像とシミュレーション画像の合成画像を用いて判定する手法を提案する。

レイアウト画像の配線部分を画素値 α ($\neq 0$), シミュレーション画像を画素値 β ($\neq 0$) のグレースケール画像に変換する (ただし $\alpha \neq \beta$)。グレースケール画像は 2^2 階調以上の階調とする。合成画像は, これらを加算処理したものである。合成画像では, 画素値 $\alpha + \beta$ が連続する部分を島と呼ぶ。合成画像を用いた判定方法を以下で説明する。

(2) と (3) の場合の判定について説明する。(2) と (3) の場合は, 配線の消失や本来存在しない配線の結像が容易に判定できる。そのような箇所があれば, それはホットスポットと判定する。合成画像で示すと図 8(a), (b) である。

(4) の場合の判定について説明する。合成画像の島は, レイアウト画像の配線がそれと同じ位置に結像されたことを示している。(4) の場合は, 図 8(c) に示すように, ある 2 つの配線で短絡が生じていてそれに対応する島の数は 2 つとなり, ある 1 つの配線で開放が生じていてそれに対応する島の数は 2 つとなる。つまり, レイアウト画像の島と比較し合成画像の島の数は 1 つ増える。3 つ以上の配線の短絡, 1 つの配線で開放が 2 つ以上生じたときも同様な議論が行える。よって, 合成画像とレイアウト画像の島の数を比較し, 合成画像の島の数が増加していればホットスポットとする。ここまででホットスポットとラベル付けされていないデータは, レイアウト画像の島 (配線) とシミュレーション画像の島 (結像) が 1 対 1 対応している。

(5) の場合の判定について説明する。合成画像により, 図 8(d) に示すように配線端部の消失が分かる。提案手法では, これを次のように近似的に判定する。元の各配線に対して結像されなかった配線部分の面積をエラー面積とし, エラー面積の割合が γ を超えているような配線が 1 つでも存在したときにホットスポットとする。本研究では, 2 割 ($\gamma = 0.2$) とした。これは, データセットごとにランダム

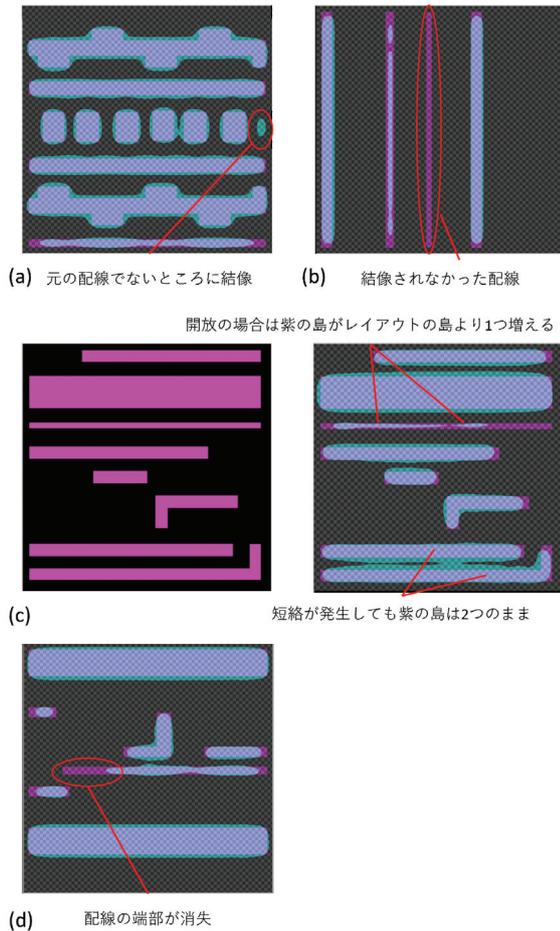


図 8 合成画像による判定

にデータを数十個抽出して確認したところ、エラー面積が2割以上のものは明らかに配線の端部が欠損しているなどホットスポットであることが明白であったためである。

提案ラベル付け手法のフローチャートを図9に示す。なお本研究は、(3)が上手くラベル付けできているかを図示により分析するためグレースケールの合成画像(加算処理)で行ったが、ラベル付けだけであれば元の画像と2値の合成画像(乗算処理)で可能である。

6. 実験

この節では、提案ラベル付け手法と既存ラベル付け手法とでラベル付けの精度を比較する。また、提案ラベル付け手法により生成した訓練・テストデータを用い特徴量の評価を行う。

6.1 ラベル付け精度の比較実験

元のレイアウトパターンを修正して生成したデータに既存手法でラベル付けした場合と提案手法でラベル付けした場合とで誤判定になった割合を比較した。ベンチマークデータ arraya_benchmark1 では HS/N-HS 領域 $1.2\mu\text{m} \times 1.2\mu\text{m}$ で切り出したレイアウトパターンを1.6倍に拡大、arraya_benchmark2~arraya_benchmark5 では HS/N-HS 領域

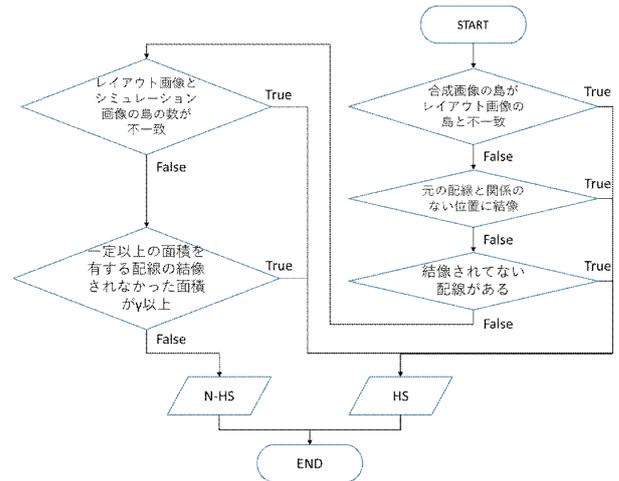


図 9 光学シミュレーション適用までの流れ

表 2 既存手法と提案手法の誤ラベルの割合

回路	誤ラベルの数		誤ラベルの割合 (%)	
	既存	提案	既存	提案
A: array_benchmark1	69	1	12.7	0.183
B: array_benchmark2	202	3	4.3	0.065
C: array_benchmark3	175	4	3.3	0.075
D: array_benchmark4	62	2	1.7	0.056
E: array_benchmark5	86	2	4.0	0.092

のさらに内側 $1.0\mu\text{m} \times 1.0\mu\text{m}$ で切り出したレイアウトパターンを1.9倍に拡大してから修正パターンを生成した。ラベル付け精度の比較に用いたデータは array_benchmark1~array_benchmark5 の配線の1辺を平行移動させたパターン(修正1)のデータである。配線の1辺を平行移動させたパターンを比較に用いた理由は、配線幅が広くなったり狭くなったりしたことで短絡や開放を引き起こす場合が多く、5.1節で説明した正しくラベル付けされていない例が多く現れたためである。なお誤判定は、目視でシミュレーション画像を確認し、明らかにホットスポットであるにもかかわらず非ホットスポットとラベル付けされている場合や、その逆の場合とした。そのためラベルだけ付け直したパターン(修正なし)および配線自体を平行移動させたパターン(修正2)は時間の都合により未評価である。実験結果を表2に示す。実験結果から、提案手法により誤判定が大幅に削減されることが確認できた。

本実験により新たに判った誤判定が生じる場合として、コの字型の配線パターンがある。図10のように、コの字型の配線の垂直部分が少しだけレイアウト画像に入った場合、垂直部分は結像されないことがあり、その場合シミュレーション後の島の数が2つになる。このように少しだけ領域内に入る配線は5.1節の(1)で考慮しているが、この例に対しては不十分であり、今後の課題となる。また5.1節の(5)の場合に類似した問題として、結像後の面積が過剰な場合への対応も今後の課題として挙げられる。

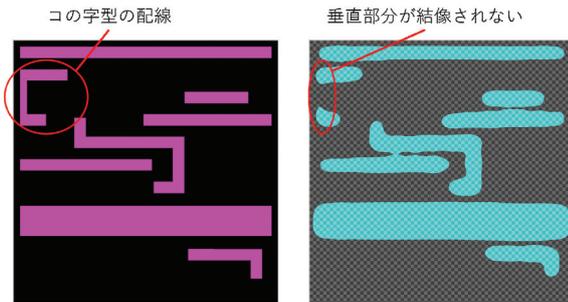


図 10 境界付近のコの字型の配線パターン

表 3 光学シミュレーションに用いるデータのラベル数

回路	ReLabel		EdgeMove		WireMove	
	HS	N-HS	HS	N-HS	HS	N-HS
A	338	207	360	185	318	227
B	1795	2849	1974	2670	2336	2308
C	3415	1934	3473	1876	3805	1544
D	1305	2258	1406	2157	1543	2020
E	677	1475	715	1437	716	1436

6.2 訓練データの追加生成を伴う特徴量評価実験

より正確に特徴量の性能を比較すること、そして訓練データ数の増加がホットスポット検出精度の向上につながるか検証することを目的として、ホットスポット判定実験を行った。具体的には、ラベルを付け直した既存データセットのみを用いた場合と、訓練データの追加生成を伴った場合の、2つのホットスポット判定実験である。

使用した特徴量は、DBLF、HOLP、LiWS である。実験には Intel Xeon E5-2620 v4 2.1GHz、メモリ 128GB の LinuxPC (CentOS 6.10) を用いた。実装言語は Python 3.6.8、学習モデルは Scikit-learn ライブラリの AdaBoost クラスを用いた。なお、Adaboost の弱学習器は決定木を用いた。また、DBLF 以外のすべての特徴量は DBLF のベクトルに結合し、DBLF の情報を包含する形で用いた。特徴量同士の基本的な性能を比較するため、ホットスポット検出ではなく、その主要な工程であるホットスポット判定のみを行い、判定精度を評価した。

本実験では、ラベルを付け直した既存データセット（以下、ReLabel）のみを用いた場合は 70% を訓練データに用い、残り 30% をテストデータとしている。また、訓練データの追加生成を伴った場合は、ReLabel に配線の 1 辺を平行移動させたデータ（以下、EdgeMove）と配線自体を平行移動させたデータ（以下、WireMove）を追加し、訓練データを 3 倍としている。ReLabel、EdgeMove、WireMove のラベルの内訳を表 3 に示す。

評価指標は、誤検出の少なさと見逃しの少なさの複合的な指標である F1 スコアを用いた [12]。F1 スコアは高いほど良い。なお、検出窓の分割数 $N = 10$ とした。また、HOLP のヒストグラムのビン数 B は [5] で示されている通りに $B = 8$ とした。また、公平な特徴量の比較のため、各

表 4 追加生成ありとなしの場合の F1 スコア

データセット (回路)	追加	DBLF	HOLP	LiWS
A: array_benchmark1	なし	0.931	0.929	0.941
	あり	0.927	0.923	0.937
B: array_benchmark2	なし	0.688	0.697	0.724
	あり	0.679	0.676	0.713
C: array_benchmark3	なし	0.858	0.867	0.875
	あり	0.859	0.868	0.883
D: array_benchmark4	なし	0.816	0.832	0.860
	あり	0.768	0.781	0.829
E: array_benchmark5	なし	0.781	0.794	0.812
	あり	0.738	0.743	0.796

特徴量で F1 スコアが最良となるようなハイパーパラメータを、以下に示す全ての組み合わせで探索した [13]。

弱学習器数 2, 4, 6, 8, 10, 20, 40, 60, 80, 100, 200, 400, 600, 800, 1000

決定木の深さ 2, 3, 4

学習率 0.95, 0.96, 0.97, 0.98, 0.99

各データセットにおける訓練データの追加生成ありとなしの場合の各特徴量の F1 スコアの最高値を示したものを表 4 に示す。また、データ数が一番少ない array_benchmark1 の訓練データの追加生成ありとなしの場合の、弱学習器数を横軸、最高値となった F1 スコア縦軸としたグラフを図 11 と図 12 に示す。

実験結果から、追加生成ありとなしどちらの場合も、F1 スコアの最高値が最も高かったのはどのデータセットにおいても LiWS となった。追加生成なしの場合と比べて、追加生成を伴った場合に F1 スコアの最高値が向上したのは、array_benchmark3 を用いた場合のみで、それ以外のデータセットではどの手法も低下した。

追加生成を伴った場合に F1 スコアが向上したのが array_benchmark3 を用いた場合のみとなった原因として、他のベンチマークデータは外れ値となるデータが多く追加生成されてしまい、それらの外れ値を学習してしまったという可能性が考えられる。外れ値が多いということは、問題例が複雑である、どの特徴量でも捉えられていない特徴がある、ラベル付けの精度が不十分であるといった可能性がある。原因の特定および特徴量・ラベル付け手法の改良は今後の課題である。

7. まとめ

本研究では、機械学習に基づくリソグラフィホットスポット判定手法のための学習・テストデータ生成手法においてラベル付け工程の問題点を発見し、それらの問題点を解決・緩和するための改良手法を提案した。また、提案手法により生成したデータを用いてホットスポット判定手法の評価実験を行った。今後の課題として、先端プロセス条件下での実験、より多くのパターンでのラベル付け精度の

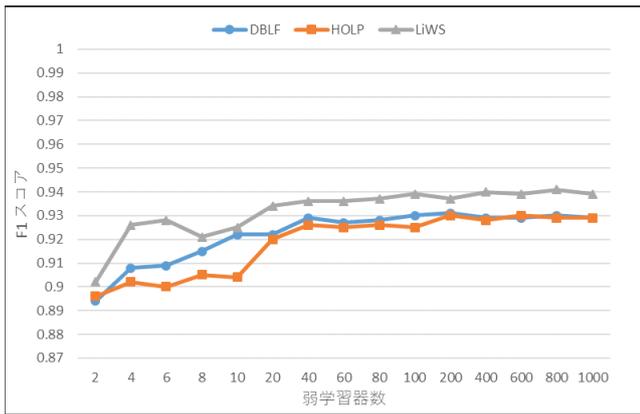


図 11 array_benchmark1 追加生成ありの F1 スコア

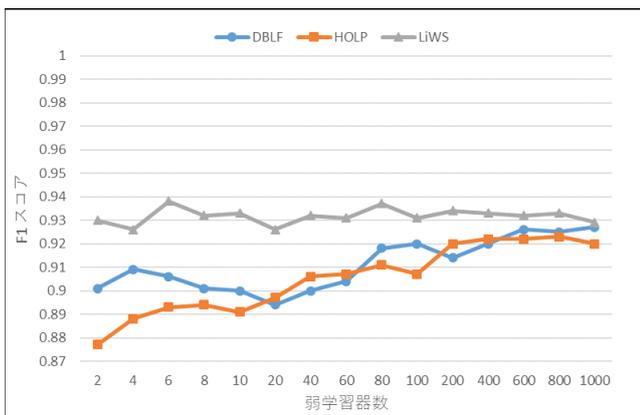


図 12 array_benchmark1 追加生成なしの F1 スコア

評価, EPE (Edge Placement Error) 解析 [14] (レイアウトパターン各配線の辺と露光パターンの結像のエッジの距離によるエラー判定) のラベル付けへの応用, より正確にホットスポットと非ホットスポットを切り分けることが出来る特徴量の設計などが挙げられる。

参考文献

[1] 小谷敏也, 間下浩充, 宇野太賀: 半導体デバイスの微細化を支える OPC 技術と DFM 技術, 東芝レビュー, Vol.67, No.4, pp.11-15 (2012).

[2] Lin, S. -Y., Chen, J. -Y., Li, J. -C., Wen, W. -Y. and Chang, S. -C.: *A novel fuzzy matching model for lithography hotspot detection*, in Proc. ACM/EDAC/IEEE DAC 2013, pp.1-6 (2013).

[3] Matsunawa, T., Gao, J. -R., Yu, B. and Pan, D. Z.: *A new lithography hotspot detection framework based on adaboost classifier and simplified feature extraction*, in Proc. SPIE, vol.9427, 94270S, pp.1-11 (2015).

[4] Yu, Y. -T., Lin, G. -H., Jiang, I. -R. and Chiang, C.: *Machine-learning-based hotspot detection using topological classification and critical feature extraction*, IEEE Trans. on CAD, vol.34, pp.460-470 (2015).

[5] Tomioka, Y. and Matsunawa, T.: *Lithography hotspot detection by two-stage cascade classifier using histogram of oriented light propagation*, in Proc. ASP-DAC 2017, pp.81-86 (2017).

[6] Wu, J. -Y., Pikus, F. G., Torres, A. and Sadowska, M. M.: *Rapid layout pattern classification*, in Proc. ASP-

DAC 2011, pp.781-786 (2011).

[7] 片岡岳, 稲木雅人, 永山忍, 若林真一: 配線幅および配線間距離を考慮した特徴量によるリソグラフィホットスポット検出, 電子情報通信学会 技術研究報告 (VLD2019-51), vol.119, no.282, pp.185-190 (2019).

[8] Torres, J. A.: *ICCAD-2012 CAD contest in fuzzy pattern matching for physical verification and benchmark suite*, in Proc. IEEE/ACM ICCAD 2012, pp.349-350, (2012).

[9] Banerjee, S., Li, Z. and Nassif, S. R.: *ICCAD-2013 CAD Contest in Mask Optimization and Benchmark Suite*, in Proc IEEE/ACM ICCAD 2013, pp.271-274 (2013).

[10] 片岡岳, 稲木雅人, 永山忍, 若林真一, 児玉親亮: リソグラフィホットスポット検出用既存訓練データの修正による訓練データの追加生成, 電子情報通信学会 技術研究報告 (VLD2019-107), vol.119, no.443, pp77-82 (2020).

[11] Matsunawa, T., Yu, B. and Pan, D. -Z.: *Optical proximity correction with hierarchical Bayes model*, J. Micro/Nanolith. MEMS MOEMS, vol.15 (2), no.021009, pp.1-8 (2016).

[12] Raschka, S., Mirjalili, V., 福島真太郎, 株式会社クイープ: Python 機械学習プログラミング 達人データサイエンティストによる理論と実践 第3版, インプレス (2020).

[13] 片岡岳, 稲木雅人, 永山忍, 若林真一: 最適ハイパーパラメータ下での機械学習に基づくリソグラフィホットスポット検出手法の比較検討, 電子情報通信学会 技術研究報告 (VLD2018-71), vol.118, no.334, pp.215-220 (2018).

[14] Yu, Peng. and Pan, D. Z.: *A novel intensity based optical proximity correction algorithm with speedup in lithography simulation*, in Proc IEEE/ACM ICCAD 2007, pp. 854-859 (2007).