VLSI Layout Hotspot Detection Based on Discriminative Feature Extraction

Hang Zhang, Haoyu Yang, Bei Yu, Evangeline F.Y. Young

Department of Computer Science and Engineering,
The Chinese University of Hong Kong

Oct 28, 2016
Outline

Problem Background

Conventional Methods

Our Method

Results
Lithography Hotspot Detection Background

- What you see ≠ what you get
- DFM: OPC, RET, MPL
- Still hotspot: low fidelity
- Simulations: extremely CPU-intensive

Ratio of lithography simulation time (normalized by 40nm node)
Technology node
Required computational/measurement!
What you see ≠ what you get
Lithography Hotspot Detection Background

- What you see \( \neq \) what you get
- **DFM:** OPC, RET, MPL
What you see ≠ what you get

- **DFM**: OPC, RET, MPL
- **Still hotspot**: low fidelity
What you see ≠ what you get

- **DFM**: OPC, RET, MPL
- Still hotspot: low fidelity
- Simulations: extremely CPU intensive
What you see \neq what you get
- **DFM**: OPC, RET, MPL
- Still hotspot: low fidelity
- Simulations: extremely CPU intensive

**Required computational time reduction!**

**Lithography Hotspot Detection Background**
What you see ≠ what you get

DFM: OPC, RET, MPL

Still hotspot: low fidelity

Simulations: extremely CPU intensive

Pattern Matching or Machine Learning?
Key Issue in Hotspot Detection?

Definitely layout pattern feature extraction. We need discriminative pattern information to detect hotspot.
Key Issue in Hotspot Detection?

- **The first problem** in machine learning based hotspot detection is?
Key Issue in Hotspot Detection?

- The first problem in machine learning based hotspot detection is?
  - Definitely layout pattern feature extraction.
Key Issue in Hotspot Detection?

- The first problem in machine learning based hotspot detection is?

- Definitely layout pattern feature extraction.

- We need discriminative pattern information to detect hotspot.
Outline

Problem Background

Conventional Methods

Our Method

Results
Conventional Feature Extraction

Fragment
[ASPDAC’12][JM3’15]

▶ Fragment feature is very complicated, which leads to over-fitting.
▶ High order local correlation (HLAC) is only efficient in some image processing task.
▶ Density based feature loses some important pattern information.
Fragment feature is very **complicated**, which leads to over-fitting.
Fragment feature is very complicated, which leads to over-fitting.

High order local correlation (HLAC) is only efficient in some image processing task.
Conventional Feature Extraction

- Fragment feature is very complicated, which leads to over-fitting.
- High order local correlation (HLAC) is only efficient in some image processing task.
- Density based feature loses some important pattern information.
Feature in [TCAD’15]

(a) Internal feature

(b) External feature

(c) Diagonal feature

(d) Segment feature

Pros: easy and fast to extract.

Cons: still complicated, hard to detect new patterns.
Feature in [TCAD’15]

- (a) Internal feature
- (b) External feature

Horizontal tiled pattern

(c) Diagonal edge

Pros: easy and fast to extract.
Cons: still complicated, hard to detect new patterns.
Feature in [TCAD’15]

- (a) Internal feature
- (b) External feature
- (c) Diagonal feature
Feature in [TCAD’15]

- (a) Internal feature
- (b) External feature
- (c) Diagonal feature
- (d) Segment feature
Feature in [TCAD’15]

- (a) Internal feature
- (b) External feature
- (c) Diagonal feature
- (d) Segment feature

Pros: easy and fast to extract.
Feature in [TCAD’15]

- (a) Internal feature
- (b) External feature
- (c) Diagonal feature
- (d) Segment feature

- Pros: easy and fast to extract.
- Cons: still complicated, hard to detect new patterns.
Density based Feature [SPIE’15]


g

▶ Side length \( l \) and grid number \( g \).
▶ \( a_{ij} \) denotes the density value of the grid in the \( i \)th row and the \( j \)th column.
▶ Feature vector: \( \mathbf{X} = \{a_{11}, a_{12}, ..., a_{54}, a_{55}\} \)

▶ Pros: simple and efficient compared to previous methods.
▶ Cons: Severe layout pattern information loss.
Density based Feature [SPIE’15]

- Side length \( l \), grid number \( g \).

\( a_{ij} \) denotes the density value of the grid in \( i \)th row and the \( j \)th column.

Feature vector: \( X = \{a_{11}, a_{12}, ..., a_{54}, a_{55}\} \)

Pros: simple and efficient compared to previous methods.

Cons: Severe layout pattern information loss.
Density based Feature [SPIE’15]

- Side length $l$, grid number $g$.
- $a_{ij}$ denotes the density value of the grid in $i^{th}$ row and the $j^{th}$ column.

Feature vector: $X = \{a_{11}, a_{12}, ..., a_{54}, a_{55}\}$.

Pros: simple and efficient compared to previous methods.

Cons: Severe layout pattern information loss.
Density based Feature [SPIE’15]

- Side length $l$, grid number $g$.
- $a_{ij}$ denotes the density value of the grid in $i^{th}$ row and the $j^{th}$ column.
- Feature vector: $X = \{a_{11}, a_{12}, ..., a_{54}, a_{55}\}$
Density based Feature [SPIE’15]

▶ Side length $l$, grid number $g$.
▶ $a_{ij}$ denotes the density value of the grid in $i^{th}$ row and the $j^{th}$ column.
▶ Feature vector:
\[ X = \{a_{11}, a_{12}, ..., a_{54}, a_{55}\} \]

▶ Pros: simple and efficient compared to previous methods.
Density based Feature [SPIE’15]

- Side length \( l \), grid number \( g \).
- \( a_{ij} \) denotes the density value of the grid in \( i^{th} \) row and the \( j^{th} \) column.
- Feature vector:
  \[
  X = \{a_{11}, a_{12}, \ldots, a_{54}, a_{55}\}
  \]

- Pros: simple and efficient compared to previous methods.
- Cons: Severe layout pattern information loss.
Outline

- Problem Background
- Conventional Methods
- Our Method
- Results
Learning Framework

- Training stage → models.
- Testing stage
- Learning models: Decision-tree, ANN, SVM...
Major Drawbacks of Conventional Density Based

For both patterns, we can only get the same feature vector. However, their contributions to the hotspot formation are different.
For both patterns, we can only get the same feature vector. However, their contributions to the hotspot formation are different.
For both patterns, we can only get the same feature vector.
Major Drawbacks of Conventional Density Based

- For both patterns, we can only get the same feature vector.
- However, their contributions to the hotspot formation are different.
Local Grid Density Differential (LGDD)

- Locally average the density value of a specific area in a grid.
- We apply triangle area in this paper.
Local Grid Density Differential (LGDD)

- locally average the density value of a specific area in a grid.
Local Grid Density Differential (LGDD)

- locally average the density value of a specific area in a grid.
- We apply triangle area in this paper.
Examples of LGDD

- The area value of the blue region in the shadow part.
Definitions for Evaluations

- **Accuracy**: The rate of correctly predicted hotspots among the set of actual hotspots.

- **Extra**: The number of falsely detected hotspots.
Effect of LGDD

- Performance comparison between LGDD and conventional density based feature.
Effect of LGDD

Performance comparison between LGDD and conventional density based feature.

The impact on accuracy.

The impact on accuracy.
Effect of LGDD

- Performance comparison between LGDD and conventional density based feature.

- The impact on accuracy.

- The impact on extra.
Stride Analysis

Stride is the spacing between two adjacent grids (horizontally and vertically).

Density based feature is a special case with \( w = s \) in our stride analysis.
Stride Analysis

Stride is the spacing between two adjacent grids (horizontally and vertically).
Stride Analysis

- Stride is the **spacing** between two adjacent grids (horizontally and vertically).

- Density based feature is a special case with $w = s$ in our stride analysis.
Effect of Strides

- Performance comparison among different strides.

- The performance raises when shrinking the stride.

- However, after a threshold, the smaller of the stride, the worse of the performance.
Effect of Strides

- Performance comparison among different strides.

- The impact on accuracy.

- The performance raises when shrinking the stride. However, after a threshold, the smaller of the stride, the worse of the performance.

- The impact on accuracy.
Effect of Strides

- Performance comparison among different strides.

- The impact on accuracy.
  - The performance raises when shrinking the stride.
  - However, after a threshold, the smaller of the stride, the worse of the performance.

- The impact on extra.
Learning Model

Adaboost classifier

\textbf{Require}: $X = (x_1, \ldots, x_n)$, $Y = (y_1, \ldots, y_n)$, $T$.
\begin{align*}
1: & \quad \text{for } i \leftarrow 1 \text{ to } n \text{ do:} \\
2: & \quad D_1(i) = \frac{1}{n}; \\
3: & \quad \text{for } t \leftarrow 1 \text{ to } T \text{ do:} \\
4: & \quad h_t \leftarrow \text{base classifier with small error } \epsilon_t; \\
5: & \quad \epsilon_t \leftarrow P(h_t(x_i) \neq y_i) = \sum_{i=1}^{n} D_t(i) I(h_t(x_i) \neq y_i); \\
6: & \quad \alpha_t \leftarrow \frac{1}{2} \log\left(\frac{1-\epsilon_t}{\epsilon_t}\right); \\
7: & \quad Z_t \leftarrow 2[\epsilon_t(1-\epsilon_t)]^{\frac{1}{2}}; \\
8: & \quad \text{for } i \leftarrow 1 \text{ to } n \text{ do:} \\
9: & \quad D_{t+1}(i) \leftarrow \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}; \\
10: & \quad f \leftarrow \text{sign}(\sum_{t=1}^{T} \alpha_t h_t); \\
11: & \quad \text{return } f
\end{align*}

- Decision-Tree as weak learner, more details in the paper.
Outline

Problem Background

Conventional Methods

Our Method

Results
Benchmark Examples

- ICCAD benchmark.
- Industrial benchmark.
Effect of Our Methods

Table: Comparison with conventional density based method

<table>
<thead>
<tr>
<th></th>
<th>Density Based</th>
<th>Our Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Extra#</td>
<td>Accuracy</td>
</tr>
<tr>
<td>ICCAD-1</td>
<td>0</td>
<td>99.50%</td>
</tr>
<tr>
<td>ICCAD-2</td>
<td>0</td>
<td>97.18%</td>
</tr>
<tr>
<td>ICCAD-3</td>
<td>0</td>
<td>97.50%</td>
</tr>
<tr>
<td>ICCAD-4</td>
<td>4</td>
<td>82.49%</td>
</tr>
<tr>
<td>ICCAD-5</td>
<td>0</td>
<td>95.12%</td>
</tr>
<tr>
<td>Industry</td>
<td>70</td>
<td>95.96%</td>
</tr>
<tr>
<td>Average</td>
<td>12.3</td>
<td>94.63%</td>
</tr>
</tbody>
</table>

▶ Consider both LGDD and stride analysis.
▶ Increase accuracy from 94.63% to 95.38%.
▶ Reduce the extra number from 12.3 to 6.
Thanks and Questions?