1. Introduction
   - Device Feature Size Continues to Shrink
   - Lithography Hotspot Detection
   - Conventional Methods on Hotspot Detection
   - Rethinking

2. Feature
   - Conventional Feature Extraction
   - Rethinking Feature Selection
   - Matrix based Concentric Circle Sampling

3. Model
   - Learning Model Background
   - Hotspot-oriented Model

4. Solver & Analysis
   - Properties of the Objective Function
   - Numerical Optimization
   - Theoretical Analysis

5. Results
   - Experimental Results
Outline

1 Introduction
   - Device Feature Size Continues to Shrink
   - Lithography Hotspot Detection
   - Conventional Methods on Hotspot Detection
   - Rethinking

2 Feature
   - Conventional Feature Extraction
   - Rethinking Feature Selection
   - Matrix based Concentric Circle Sampling

3 Model
   - Learning Model Background
   - Hotspot-oriented Model

4 Solver&Analysis
   - Properties of the Objective Function
   - Numerical Optimization
   - Theoretical Analysis

5 Results
   - Experimental Results

Hang Zhang, Fengyuan Zhu, Haocheng Li, Evangeline F.Y. Young, Bei Yu
The Chinese University of Hong Kong
The International Symposium on Physical Design 2017
Moore’s Law to Extreme Scaling

Moore’s Law

- Processor Technology (μm)
- Number of Transistors per Integrated Circuit

- Intel Microprocessors
- Invention of the Transistor
- Doubles every 2.1 yrs

Hang Zhang, Fengyuan Zhu, Haocheng Li, Evangeline F.Y. Young, Bei Yu
The Chinese University of Hong Kong
The International Symposium on Physical Design 2017
Device Feature Size Continues to Shrink

Shrinking Device Feature Size

Hang Zhang, Fengyuan Zhu, Haocheng Li, Evangeline F.Y. Young, Bei Yu
The Chinese University of Hong Kong
The International Symposium on Physical Design 2017
Lithographic Mechanism

- Light pass through photo masks (mask scale << light wavelength);
- Light diffraction and light interference will happen;
- May cause performance degradation, or even yield loss.
Motivation

- What you design ≠ what you get;
- **DFM**: MPL, OPC, SRAF;
- **Still hotspot**: low fidelity patterns;
- **Simulations**: extremely time intensive.

---

Hang Zhang, Fengyuan Zhu, Haocheng Li, Evangeline F.Y. Young, Bei Yu
The Chinese University of Hong Kong
The International Symposium on Physical Design 2017
Pattern Matching based Hotspot Detection

Conventional Methods on Hotspot Detection

Fast and reasonably accurate; Two-stage filtering, fuzzy pattern matching; [Yu +, ICCAD’14] [Wen +, TCAD’14]; Hard to detect unseen pattern.

Hang Zhang, Fengyuan Zhu, Haocheng Li, Evangeline F.Y. Young, Bei Yu
The Chinese University of Hong Kong
The International Symposium on Physical Design 2017
Pattern Matching based Hotspot Detection

Conventional Methods on Hotspot Detection

- Pattern Matching based Hotspot Detection

- Library

- Pattern matching

- Detected

- Undetected

- Hotspot

- Cannot detect hotspots not in the library

---

Hang Zhang, Fengyuan Zhu, Haocheng Li, Evangeline F.Y. Young, Bei Yu

The Chinese University of Hong Kong

The International Symposium on Physical Design 2017
Pattern Matching based Hotspot Detection

- Fast and reasonably accurate;
- Two-stage filtering, fuzzy pattern matching;
- [Yu+, ICCAD’14][Wen+, TCAD’14];
- Hard to detect unseen pattern.
Machine Learning based Hotspot Detection

Conventional Methods on Hotspot Detection

Can predict new patterns, and are more flexible; Support vector machine, boosting, deep neural network... [Ding +, ASPDAC'12][Yu +, TCAD'15][Zhang +, ICCAD'16][Matsunawa +, SPIE'16]; Hard to balance accuracy and false-alarm.

Hang Zhang, Fengyuan Zhu, Haocheng Li, Evangeline F.Y. Young, Bei Yu
The Chinese University of Hong Kong
The International Symposium on Physical Design 2017
Machine Learning based Hotspot Detection

Conventional Methods on Hotspot Detection

Extract layout features → Hotspot detection model → Classification

Non-Hotspot

Hotspot

Hard to trade-off accuracy and false alarms

Can predict new patterns, and are more flexible;
Support vector machine, boosting, deep neural network...

[Ding +, ASPDAC'12]
[Yu +, TCAD'15]
[Zhang +, ICCAD'16]

Hard to balance accuracy and false alarms.

Hang Zhang, Fengyuan Zhu, Haocheng Li, Evangeline F.Y. Young, Bei Yu
The Chinese University of Hong Kong
The International Symposium on Physical Design 2017
**Machine Learning based Hotspot Detection**

- Can predict **new** patterns, and are more **flexible**;
- Support vector machine, boosting, deep neural network...
- [Ding+, ASPDAC’12][Yu+, TCAD’15][Zhang+, ICCAD’16][Matsunawa+, SPIE’16];
- **Hard** to balance accuracy and false-alarm.

Hang Zhang, Fengyuan Zhu, Haocheng Li, Evangeline F.Y. Young, Bei Yu  
The Chinese University of Hong Kong  
The International Symposium on Physical Design 2017
Rethinking Conventional Methods

- **Conventional**: vector based feature and learning model;
- **Time consuming** steps: 1) feature extraction, 2) feature selection;
- **Destroying** the hidden structural correlations in the layout patterns.
Rethinking Conventional Methods

- **Conventional:** vector-based feature;
- **Time consuming** steps: 1) feature extraction, 2) feature selection;
- **Destroying** the hidden structural correlations in the layout patterns.
Rethinking Conventional Methods

- Conventional: vector-based feature;
- Time consuming steps: 1) feature extraction, 2) feature selection;
- Destroying the hidden structural correlations in the layout patterns.

**Matrix based Concentric Sampling (MCCS)**

1) Matrix Based: preserve the hidden structural correlations;
2) No feature selection: enable parallel computation;
3) Very simple feature: fast to extract.

**Bilinear Lithography Hotspot Detector**

1) Matrix based: capture the hidden structural correlations;
2) Low-complexity model: avoid over-fitting;
3) Fast to train.
Outline

1 Introduction
   - Device Feature Size Continues to Shrink
   - Lithography Hotspot Detection
   - Conventional Methods on Hotspot Detection
   - Rethinking

2 Feature
   - Conventional Feature Extraction
   - Rethinking Feature Selection
   - Matrix based Concentric Circle Sampling

3 Model
   - Learning Model Background
   - Hotspot-oriented Model

4 Solver&Analysis
   - Properties of the Objective Function
   - Numerical Optimization
   - Theoretical Analysis

5 Results
   - Experimental Results
Conventional Feature Extraction

**Geometry based Feature**

- Hard to be adaptive to different layout designs
- Too many parameters to tune
- Sometimes very complex and may be the cause of over fitting

Hang Zhang, Fengyuan Zhu, Haocheng Li, Evangeline F.Y. Young, Bei Yu
The Chinese University of Hong Kong
The International Symposium on Physical Design 2017
We conducted two experiments related to feature extraction through deep neural network training and hotspot detection. Two 28nm node industrial layouts in metal layer, layout A and layout B, are used. The areas of layouts A and B are 100μm² and 90μm², respectively.

### 4.1 Automatic feature extraction

Figure 7 shows the DNN structure we designed, where there are a total of four hidden layers and the output layer includes two units to produce the probabilities of hotspot and non-hotspot. It should be noted that in practice a process of trial and error is required to specify an appropriate DNN structure. Then we train the DNN using layout A based on the optimization strategy discussed in Section 3. The numbers of trials of backpropagation in pre-training and fine-tuning are set to 100 and 1000, respectively. The performance of model training is shown in Figure 8 where (a) is the result of the pre-training and (b) indicates the fine-tuning result. This result shows that our DNN can be successfully trained since suitable convergence is observed.

![Network structure from Matsunawa+, SPIE'16](Network structure.png)

- **Pros**: automatic layout feature extraction; easy to adapt
- **Cons**: expensive cost in training (may cause even several hours)

Hang Zhang, Fengyuan Zhu, Haocheng Li, Evangeline F.Y. Young, Bei Yu The Chinese University of Hong Kong
Rethinking MCMI

- Maximal Circular Mutual Information (MCMI) [Zhang+, ICCAD’16];
- Preserve the effects of light propagation;
- Searching for the local correlations within each circle.
Rethinking Feature Selection

Rethinking MCMI

- Maximal Circular Mutual Information (MCMI) [Zhang+, ICCAD’16];
- Preserve the effects of light propagation;
- Searching for the local correlations within each circle.

Questions:

Can we utilize the global correlations among these sampled circles?
Two follow up questions:
1. Can we preserve these correlations using our feature?
2. Can we capture these correlations using our machine learning model?
Matrix based Concentric Circle Sampling (MCCS)

- $r_s$: is the radius of the sampling area;
- $r_{in}$: controls the sampling density;
- $l$: controls the clip size;
- $n_p$: is the number of points sampled on a circle.
Matrix based Concentric Circle Sampling (MCCS)

- Points from one circle form a vector;
- Each vector forms one row of the feature matrix;
- Under the condition that $l = 1200\,nm$, $r_{in} = 60\,nm$, $n_p = 16$, the dimension of the feature matrix is $33 \times 16$ ($33 = 6 + 27$).
Matrix based Concentric Circle Sampling (MCCS)

- Preserve the hidden structural information;
- Each circle forms a row: light propagation;
- There exist linear combinations among these rows and columns: light diffraction and interference.
- Linear combinations of the rows: correlations among circles;
- Linear combinations of the columns: correlations among lines of points.

Hang Zhang, Fengyuan Zhu, Haocheng Li, Evangeline F.Y. Young, Bei Yu
The Chinese University of Hong Kong
The International Symposium on Physical Design 2017
Matrix based Concentric Circle Sampling (MCCS)

Questions:
Can we utilize the global correlations among these sampled circles?
Two follow up questions:
1. Can we preserve these correlations using our feature?
2. Can we capture these correlations using our machine learning model?
Questions:

Can we utilize the global correlations among these sampled circles?

Two follow up questions:
1. Can we preserve these correlations using our feature? **YES**
2. Can we capture these correlations using our machine learning model?
Outline

1. Introduction
   - Device Feature Size Continues to Shrink
   - Lithography Hotspot Detection
   - Conventional Methods on Hotspot Detection
   - Rethinking

2. Feature
   - Conventional Feature Extraction
   - Rethinking Feature Selection
   - Matrix based Concentric Circle Sampling

3. Model
   - Learning Model Background
   - Hotspot-oriented Model

4. Solver&Analysis
   - Properties of the Objective Function
   - Numerical Optimization
   - Theoretical Analysis

5. Results
   - Experimental Results

Hang Zhang, Fengyuan Zhu, Haocheng Li, Evangeline F.Y. Young, Bei Yu
The Chinese University of Hong Kong
The International Symposium on Physical Design 2017
Learning Model Background

Notations

- scalar: $x$
- vector: $x$
- matrix: $X$
- rank $r$: $X \in \mathbb{R}^{p \times q}$ and $r \leq \min(p, q)$
- nuclear norm: $\|X\|_* = \sum_{i=1}^{n} \sigma_i$
- weighted nuclear norm: $\|X\|_{W,*} = \sum_{i}^{n} w_i \sigma_i$
- $(i, j)=$ entity: $X_{i,j}$
- trace: $\text{tr}(\cdot)$
- $(a)_+ = \max(0, a)$
- $\langle A, B \rangle = \sum_{i,j} A_{i,j} \cdot B_{i,j}$
- Frobenius norm: $\|X\|_F = \sqrt{\sum_{i,j} X_{i,j}^2}$
- Spectral Elastic Net: $\frac{1}{2} \text{tr}(W^T W) + \lambda \|W\|_*$

Hang Zhang, Fengyuan Zhu, Haocheng Li, Evangeline F.Y. Young, Bei Yu
The Chinese University of Hong Kong
The International Symposium on Physical Design 2017
Modern techniques are producing datasets with complex hidden structures;

These features can be naturally represented as matrices instead of vectors.

Eg. 1: the two-dimensional digital images, with quantized values of different colors at certain rows and columns of pixels;

Eg. 2: electroencephalography (EEG) data with voltage fluctuations at multiple channels over a period of time.
Most existing learning models are vector based;

People propose bilinear classifiers that can tackle data in matrix form: [Wolf+, CVPR’07][Pirsiavash+, NIPS’09][Luo+, ICML’15];

However, these methods have their own drawbacks.
Drawbacks I

- [Wolf+, CVPR’07] uses the sum of \( k \) rank-one orthogonal matrices to model the classifier matrix;
- [Pirsiavash+, NIPS’09] assumes the rank of the classifier matrix to be \( k \);
- Both methods describe the correlations of data in different ways, but they require the rank \( k \) to be pre-specified.
### Learning Model Background

#### Drawbacks II

- [Luo+, ICML’15] could determine the rank automatically, however:
- when using the nuclear norm, it assigns same weights to all singular values;
- it aims at capturing the grouping effects (No such effects in our problem) by spectral elastic net term.

---

**Table 1. Summary of four data sets**

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EEG alcoholism</td>
<td>The EEG data set consists of 200 males and 200 females. Each sample is a 200×200 × 2 gray image, and there exist people in the image. We normalize the samples.</td>
</tr>
<tr>
<td>INRIA person</td>
<td>The INRIA person data set arises to examine EEG correlates of genetic predisposition to alcoholism. It contains 1821 samples in total.</td>
</tr>
<tr>
<td>Student face</td>
<td>The student face data set consists of 200 males and 200 females. Each sample is a face image, and there exist people in the image. We normalize the samples.</td>
</tr>
<tr>
<td>INRIA emotion</td>
<td>The INRIA emotion data set consists of 200 males and 200 females. Each sample is a face image, and there exist people in the image. We normalize the samples.</td>
</tr>
</tbody>
</table>

---

**Figure 1.** The top row displays the values of normalized regression matrix of B-SVM, R-GLM and SMM respectively. The bottom row displays the values of normalized regression matrix of B-SVM, R-GLM and SMM respectively.

**Figure 2.** Classification accuracy on synthetic data with different aspects. Combining with the negative samples, we directly use the pixels as input features without any advanced visual features.
There are several issues for our hotspot detection problem.

- Can we address them?

### Needs for our New Model

1. Reduce the impact of outliers;
2. The grouping effects should be discarded;
3. The rank $k$ should be automatically determined;
4. Less weights should be assigned to larger singular values.
Objective Function of our Model

Needs for our New Models

1. Reduce the impact of outliers;
2. The grouping effects should be discarded;
3. The rank $k$ should be automatically determined;
4. Less weights should be assigned to larger singular values.
Objective Function of our Model

Needs for our New Models
1. Reduce the impact of outliers;
2. The grouping effects should be discarded;
3. The rank $k$ should be automatically determined;
4. Less weights should be assigned to larger singular values.

Final Objective Function

$$\arg \min_{W,b} \lambda \|W\|_{W,*} + C \sum_{i=1}^{n} \{1 - y_i[\text{tr}(W^TX_i) + b]\}^+. \quad (1)$$
Objective Function of our Model

Needs for our New Models

1. Reduce the impact of outliers;
2. The grouping effects should be discarded;
3. The rank $k$ should be automatically determined;
4. Less weights should be assigned to larger singular values.

Final Objective Function

$$\arg \min_{W,b} \lambda ||W||_W,\ast + C \sum_{i}^{n} \{1 - y_i[\text{tr}(W^T X_i) + b]\}^+.$$
Objective Function of our Model

Needs for our New Models

1. Reduce the impact of outliers;
2. The grouping effects should be discarded;
3. The rank $k$ should be automatically determined;
4. Less weights should be assigned to larger singular values.
Objective Function of our Model

Needs for our New Models

1. Reduce the impact of outliers;
2. The grouping effects should be discarded;
3. The rank $k$ should be automatically determined;
4. Less weights should be assigned to larger singular values.

Final Objective Function

$$\arg\min_{W,b} \frac{1}{2} \text{tr}(W^TW) + \lambda \|W\|_{W,*} + C \sum_{i}^{n} \{1 - y_i[\text{tr}(W^TX_i) + b]\}^+. $$
Hotspot-oriented Model

Objectives Function of our Model

Needs for our New Models
1. Reduce the impact of outliers;
2. The grouping effects should be discarded;
3. The rank $k$ should be automatically determined;
4. Less weights should be assigned to larger singular values.

Final Objective Function

$$\text{arg min}_{W, b} \lambda ||W||_W, + C \sum_i^n \{1 - y_i[\text{tr}(W^\top X_i) + b]\}_+.$$
Objective Function of our Model

Final Objective Function

\[
\arg\min_{W,b} \lambda \|W\|_{W,*} + C \sum_{i}^{n} \{1 - y_i[\text{tr}(W^\top X_i) + b]\} +.
\] (2)

Questions:

Can we utilize the global correlations among these sampled circles?

Two follow up questions:

1. Can we preserve these correlations using our feature? **YES**
2. Can we capture these correlations using our machine learning model?
Objective Function of our Model

Final Objective Function

$$\arg \min_{W,b} \lambda ||W||_{W,*} + C \sum_{i}^{n} \{1 - y_i [\text{tr}(W^T X_i) + b] \}^+. \quad (2)$$

Questions:
Can we utilize the global correlations among these sampled circles?
Two follow up questions:
1. Can we preserve these correlations using our feature? **YES**
2. Can we capture these correlations using our machine learning model? **YES**
Outline

1 Introduction
   - Device Feature Size Continues to Shrink
   - Lithography Hotspot Detection
   - Conventional Methods on Hotspot Detection
   - Rethinking

2 Feature
   - Conventional Feature Extraction
   - Rethinking Feature Selection
   - Matrix based Concentric Circle Sampling

3 Model
   - Learning Model Background
   - Hotspot-oriented Model

4 Solver&Analysis
   - Properties of the Objective Function
   - Numerical Optimization
   - Theoretical Analysis

5 Results
   - Experimental Results
Resolve Issues

- **Hinge loss**: non-smooth;
- **Weighted nuclear norm**: non-smooth, maybe non-convex\([Gu+, IJCV’16]\), which depends on the weight order;
Resolve Issues

- **Hinge loss**: non-smooth;
- **Weighted nuclear norm**: non-smooth, maybe non-convex [Gu+, IJCV’16], which depends on the weight order;
- We resort to Alternating Direction Method of Multipliers (ADMM) [Boyd+, FTML’11][Goldstein+, SIAM’14].
Resolve Issues

- **Hinge loss**: non-smooth;
- **Weighted nuclear norm**: non-smooth, maybe non-convex\([Gu_+, IJCV'16]\), which depends on the **weight order**;
- We resort to Alternating Direction Method of Multipliers (ADMM)\([Boyd_+, FTML'11][Goldstein_+, SIAM'14]\).

**Equivalent Objective Function With Auxiliary Variable \(S\)**

\[
\arg\min_{W,b,S} \lambda \|S\|_{W,*} + C \sum_{i}^{n} \{1 - y_i [\text{tr}(W^\top X_i) + b]\}_+,
\]

s.t. \(S - W = 0\),
Resolve Issues

Equivalent Objective Function With Auxiliary Variable $S$

\[
\arg \min_{W,b,S} \lambda \|S\|_{W,*} + C \sum_{i}^{n} \{1 - y_{i} [\text{tr}(W^{T}X_{i}) + b]\}_{+},
\]

s.t. $S - W = 0,$

In this way, the original optimization problem is split into two sub-problems with respect to $\{W, b\}$ and the auxiliary variable $S$. 

Hang Zhang, Fengyuan Zhu, Haocheng Li, Evangeline F.Y. Young, Bei Yu
The Chinese University of Hong Kong
The International Symposium on Physical Design 2017
Then we apply Augmented Lagrangian Multiplier to develop an efficient ADMM method as follows:

\[
L(W, b, S, \Lambda) = \lambda \|S\|_{W,*} + C \sum_{i}^{n} \{1 - y_i [\text{tr}(W^T X_i) + b]\} + \\
+ \text{tr}[\Lambda^T (S - W)] + \frac{\rho}{2} \|S - W\|_F^2,
\]
Subproblems

Subproblem 1 to Solve $S$

$$\arg \min_S \lambda \|S\|_{\mathcal{W},*} + \text{tr}(\Lambda^T S) + \frac{\rho}{2} \|W - S\|^2_F. \quad (6)$$
Subproblems

Subproblem 1 to Solve $S$

$$\arg \min_S \lambda \|S\|_{\mathcal{W},*} + \text{tr}(\Lambda^T S) + \frac{\rho}{2} \|W - S\|_F^2. \quad (6)$$

- We use the shrinkage thresholding method to solve this subproblem.
Subproblems

Subproblem 2 to Solve \((W, b)\)

\[
\begin{align*}
\arg \min_{W, b} & \quad C \sum_{i}^{n} \left\{ 1 - y_i \left[ \text{tr}(W^\top X_i) + b \right] \right\}_+ \\
& + \text{tr}[\Lambda^\top (S - W)] + \frac{\rho}{2} ||S - W||_F^2, \\
\end{align*}
\]

We use the KKT conditions and then the box constraint quadratic programming method to solve this subproblems.
Theoretical Analysis

- We analyze the excessive risk of the proposed classifier theoretically;
- We prove the consistency and correctness of our model;
- Excess risk means the difference between the empirical risk and the expected risk (Definitions in the next slide).
Lemma 1

The dual norm of the weighted nuclear norm $||W||_{\mathcal{W},*}$ is

$$||W||_{\mathcal{W},*} = \max_i \frac{1}{w_i} \Sigma_{ii}$$

(8)

where $W = U \Sigma V^T$ through SVD.

* please read the paper for more details of the proof
Theoretical Analysis

Theorem 1

With Lemma 1, we can come up with the excessive risk bound for our model:

**Theorem 1**

With probability at least $1 - \delta$, the excess risk of our method, for each data $X_i \in \mathbb{R}^{d_1 \times d_2}$, is bounded as

$$R(\hat{W}) - R(W^*) \leq \frac{2BL}{\sqrt{n}} \max_i \left( \frac{1}{w_i} \right) \cdot (\sqrt{d_1} + \sqrt{d_2}) + \sqrt{\frac{\ln(1/\delta)}{2n}}. \quad (9)$$

* please read the paper for more details of the proof
Outline

1. Introduction
   - Device Feature Size Continues to Shrink
   - Lithography Hotspot Detection
   - Conventional Methods on Hotspot Detection
   - Rethinking

2. Feature
   - Conventional Feature Extraction
   - Rethinking Feature Selection
   - Matrix based Concentric Circle Sampling

3. Model
   - Learning Model Background
   - Hotspot-oriented Model

4. Solver & Analysis
   - Properties of the Objective Function
   - Numerical Optimization
   - Theoretical Analysis

5. Results
   - Experimental Results
Experimental Results

- Verified in ICCAD-2012 contest benchmark;
- 2x speed-up in M-CPU (s);
- 19x speed-up in CPU (s);
- Increase detection accuracy from 95.13% to 98.16%.

Table 1: Comparisons with three classical methods

<table>
<thead>
<tr>
<th></th>
<th>VCCS-SVM</th>
<th>VCCS-Adaboost</th>
<th>DBF-Adaboost</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-CPU(s)</td>
<td>Accuracy</td>
<td>FA#</td>
<td>M-CPU(s)</td>
<td>Accuracy</td>
</tr>
<tr>
<td>Case 1</td>
<td>1.09</td>
<td><strong>100.00%</strong></td>
<td>0</td>
<td>1.37</td>
</tr>
<tr>
<td>Case 2</td>
<td>1.81</td>
<td>94.78%</td>
<td>4</td>
<td>5.44</td>
</tr>
<tr>
<td>Case 3</td>
<td>3.26</td>
<td>95.52%</td>
<td>94</td>
<td>4.73</td>
</tr>
<tr>
<td>Case 4</td>
<td>1.74</td>
<td>80.23%</td>
<td>31</td>
<td>9.45</td>
</tr>
<tr>
<td>Case 5</td>
<td>1.30</td>
<td>95.12%</td>
<td>0</td>
<td>2.27</td>
</tr>
<tr>
<td>avg.</td>
<td>1.84</td>
<td>93.13%</td>
<td>25.8</td>
<td>4.65</td>
</tr>
<tr>
<td>ratio</td>
<td>2.46</td>
<td>-</td>
<td>-</td>
<td>6.21</td>
</tr>
</tbody>
</table>
Experimental Results

- 4x sped-up in CPUs (s);
- Increase the accuracy to 98.16%;
- Reduce the false alarms by around 15%.

Table 2: Comparisons with three state-of-the-art hotspot detectors

<table>
<thead>
<tr>
<th></th>
<th>TCAD’14</th>
<th></th>
<th></th>
<th>TCAD’15</th>
<th></th>
<th></th>
<th>ICCAD’16</th>
<th></th>
<th></th>
<th>Ours</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CPU(s)</td>
<td>Accuracy</td>
<td>FA#</td>
<td>CPU(s)</td>
<td>Accuracy</td>
<td>FA#</td>
<td>CPU(s)</td>
<td>Accuracy</td>
<td>FA#</td>
<td>CPU(s)</td>
<td>Accuracy</td>
<td>FA#</td>
<td>CPU(s)</td>
<td>Accuracy</td>
</tr>
<tr>
<td>Case 1</td>
<td>11</td>
<td>100.00%</td>
<td>1714</td>
<td>38</td>
<td>94.69%</td>
<td>1493</td>
<td>10</td>
<td>100.00%</td>
<td>788</td>
<td>4</td>
<td>100.00%</td>
<td>783</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case 2</td>
<td>287</td>
<td>99.80%</td>
<td>4058</td>
<td>234</td>
<td>98.20%</td>
<td>11834</td>
<td>103</td>
<td>99.40%</td>
<td>544</td>
<td>17</td>
<td>99.40%</td>
<td>700</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case 3</td>
<td>417</td>
<td>93.80%</td>
<td>9486</td>
<td>778</td>
<td>91.88%</td>
<td>13850</td>
<td>110</td>
<td>97.51%</td>
<td>2052</td>
<td>49</td>
<td>97.78%</td>
<td>2166</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case 4</td>
<td>102</td>
<td>91.00%</td>
<td>1120</td>
<td>356</td>
<td>85.94%</td>
<td>3664</td>
<td>69</td>
<td>97.74%</td>
<td>3341</td>
<td>14</td>
<td>96.05%</td>
<td>2132</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case 5</td>
<td>49</td>
<td>87.80%</td>
<td>199</td>
<td>20</td>
<td>92.86%</td>
<td>1205</td>
<td>41</td>
<td>95.12%</td>
<td>94</td>
<td>9</td>
<td>97.56%</td>
<td>52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>avg.</td>
<td>173.2</td>
<td>94.48%</td>
<td>3315.4</td>
<td>285.2</td>
<td>92.71%</td>
<td>6409.2</td>
<td>66.6</td>
<td>97.95%</td>
<td>1363.8</td>
<td>18.4</td>
<td>98.16%</td>
<td>1166.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ratio</td>
<td>9.40</td>
<td>-</td>
<td>2.84</td>
<td>15.50</td>
<td>-</td>
<td>5.49</td>
<td>3.62</td>
<td>-</td>
<td>1.17</td>
<td>1.0</td>
<td>-</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

**Novel Insights in Hotspot Detection Problem**

- **Novel matrix feature** with hidden structural information preserved;
- **Novel Bilinear Machine Learning Model**;
- **Theoretical analysis** proves the correctness and consistency of the model.

**Future Work**

- Customized computing system for further speedup
- Transfer learning for further performance improvement
Conclusions

Future Work

- Adjust our methods to new layout designs
- Extend our method to OPC and MPL

We are looking forward to cooperation:

- Industrial benchmarks for HSD
- Industrial benchmarks for OPC, MPL
Thank you

Hang Zhang (hzhang@cse.cuhk.edu.hk)
Fengyuan Zhu (fyzhu@cse.cuhk.edu.hk)
Haocheng Li (hcli@cse.cuhk.edu.hk)
Evangeline F. Y. Young (fyyoung@cse.cuhk.edu.hk)
Bei Yu (byu@cse.cuhk.edu.hk)

香港中文大學
The Chinese University of Hong Kong