Laplacian Eigenmaps and Bayesian Clustering Based Layout Pattern Sampling and Its Applications to Hotspot Detection and OPC

Tetsuaki Matsunawa¹, Bei Yu² and David Z. Pan³

¹Toshiba Corporation
²The Chinese University of Hong Kong
³The University of Texas at Austin
Outline

• Background
• Pattern Sampling in Physical Verification
• Overall flow
• Laplacian Eigenmaps
• Bayesian Clustering
• Applications
  – Lithography Hotspot Detection
  – OPC (Optical Proximity Correction)
• Conclusion
Background

• **Issue:** Systematic method for pattern sampling is not established

• **Goal:** Pattern sampling automation for process optimization

**Test patterns for:**
Simulation model calibration
Source mask optimization
Wafer verification, etc.

Based on engineer’s knowledge

**1D patterns**  **2D patterns**

**Representative patterns**

**Grouping**
Pattern Sampling

Input Layout

x1 = (0, 1, 0, 1.5, …)
x2 = (2, 0.5, 1, -1, …)
x3 = (1, -1, 0, 0.3, …)
……

Feature extraction

x1 = (0, 1, 0, 1.5, …)
x2 = (2, 0.5, 1, -1, …)
x3 = (1, -1, 0, 0.3, …)
……

Sampling

Dimension Reduction

dimension 1

dimension 2

dimension 1

dimension 2
Pattern Sampling in Physical Verification

- Key techniques: **Dimension reduction** and **Clustering**


Examples of clustering results:

- [I] W. C. Tam
- [II] S. Shim

Classification flow:
- Topology-oriented pattern extraction
- Rasterization
- Fourier transform
Open Questions

• Undefined similarity
  • A criterion for defining pattern similarity to evaluate essential characteristics in real layouts is unclear

• Manual parameter tuning
  • Most clustering algorithms require several preliminary experiments (total number of clusters)
Laplacian Eigenmaps and Bayesian Clustering

• We develop
  – An efficient feature comparison method
    • With nonlinear dimensionality reduction / kernel parameter optimization
  – An automated pattern sampling using Bayesian model based clustering
    • Without manual parameter tuning
Problem formulation: Layout Pattern Sampling

• **Problem:** Given layout data, a classification model is trained to extract representative patterns.

• **Goal:** To classify the layout patterns into a set of classes minimizing the Bayes error.

![Diagram showing layout data input, classification model, and unique pattern set output.]

\[
y = f(x)
\]
Bayes Error (BE)

- To quantify the clustering performance
  - Define a quality of clustering distributions based on Bayes’ theorem

\[
BE = \int \min\{1 - p(\omega_k|x)\} p(x) dx
\]

\[P(\omega|x): \text{conditional probability in class } \omega\]
\[P(x): \text{prior probability of data } x\]

Comparison between BE and Within-Class Scatter/Between-Class Scatter

Bayes Error = 0.02, WCS/BCS = 0.07

Bayes Error = 1.68, WCS/BCS = 0.07
Overall Flow

(1) Sampling phase

GDSII Layout

DRC

Locating Feature Points

Layout Feature Extraction

Feature A

Feature B

Feature C

Dimensionality Reduction

Low-dimensional vectors A

Low-dimensional vectors B

Low-dimensional vectors C

Clustering

Layout dataset A

Layout dataset B

Layout dataset C

Ranking

Ranked dataset (Feature A, B or C)

Ranked dataset (Feature A, B or C)

Ranked dataset (Feature A, B or C)

Sample Plan Application

Model training for
- Hotspot detection,
- Mask Optimization,
- Process Simulation,
- Wafer Inspection, etc.

(2) Application phase
Feature Point Generation & Feature Extraction

GDS

Locating feature points

Feature extraction

Density based encoding

Diffraction order distribution
Why dimension reduction and Bayesian clustering?

Required feature comparison for optimal feature selection
- The optimal characteristics for layout representation vary in different applications

How to compare diverse layout feature types?
- #of dimensions differs with different types of features

Hard to achieve completely automatic clustering
- Hypothetical parameters are required for typical clustering task

Dimension Reduction
Automatic Clustering

Comparable data

Feature A
Feature B
Feature C

High dimension
Low dimension
Laplacian Eigenmaps

To reduce dimensions while preserving complicated structure

Solve an eigenvalue problem:

\[ L\psi = \gamma D\psi \]

Laplacian matrix

\[ L = D - W \]

Diagonal matrix

\[ D = \text{diag} \left( \sum_{i' = 1}^{n} W_{i,i'} \right) \]

Kernel: k-nearest neighbors

\[ W_{i,i'} = \begin{cases} 
1 & \text{if } x_i \in k\text{NN}(x_{i'}) \\
0 & \text{otherwise}
\end{cases} \]

Comparison with linear/nonlinear algorithm

Original data (3D)

Linear (2D)

Nonlinear (2D)

Principal Component Analysis

Laplacian Eigenmaps
Kernel Parameter Optimization

- Optimization through estimating density-ratio $\hat{r}(x) = w\Phi(x)$ between given feature vectors $P(x)$ and embedded feature vectors $P'(x)$

\[
\max_w \sum_{i=1}^{n'} \log \left( w^T \phi(x'_i) \right) \\
\text{Subject to } \sum_{i=1}^{n} w^T \phi(x_i) = n \text{ and } w \geq 0
\]

This is convex optimization, so repeating gradient ascent and constraint satisfaction converges to global solution

\[
r(x) = \frac{P'(x)}{P(x)}
\]
Bayesian Clustering

- Clustering automation without arbitrary parameter tuning
- Bayesian based method: express a parameter distribution as an infinite dimensional distribution

$$p(\mathbf{x}|\alpha, p(\Theta)) = \sum_{k=1}^{\infty} \pi_k \mathcal{N}(\mu_k, \sigma_k)$$

Prior probability:

$$p(z_n = k|x_n, z_1, \ldots, z_{n-1}) \propto \begin{cases} 
  p(x_n|k) \frac{n_k}{\alpha + n - 1} & (k = 1 \ldots K) \\
  p(x_n|k^{\text{new}}) \frac{\alpha}{\alpha + n - 1} & (k = K + 1) 
\end{cases}$$
Experiments

• Pattern sampling
  – Comparison of conventional methods
    • Dimensionality reduction
      – Principal Component Analysis (PCA) vs. Laplacian Eigenmaps (LE)
    • Clustering
      – K-means (Km) vs. Bayesian clustering (BC)

• Applications to
  – Lithography Hotspot Detection
  – OPC
Effectiveness of Pattern Sampling

- Representative patterns could be automatically selected

**Clustering results:**

<table>
<thead>
<tr>
<th>Method</th>
<th>Density</th>
<th>Diffraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA+Km</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LE+Km</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA+BC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LE+BC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Misclassification error rate:**

<table>
<thead>
<tr>
<th>Method</th>
<th>Density</th>
<th>Diffraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA+Km</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LE+Km</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA+BC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LE+BC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Ratio:**

- PCA+Km: 1.0
- LE+Km: 5.6
- PCA+BC: 0.7
- LE+BC: 0.5
Application to Lithography Hotspot Detection

• To detect hotspot in short runtime

• Experiments
  – Detection model training with different patterns
    • PCA+Km, LE+Km, PCA+BC, LE+BC
    • Learning algorithm is fixed to Adaptive Boosting (AdaBoost)
  – Metrics: detection accuracy and false alarm
Effectiveness of Hotspot Detection

- Comparison with conventional clustering method
- Result: Proposed framework achieved the best false-alarm

Detection accuracy:
#correctly detected hotspots / #total hotspots

False alarm:
#correctly detected hotspots / #falsely detected hotspots

![Graphs showing detection accuracy and false alarm for different methods: PCA+Km, LE+Km, PCA+BC, LE+BC.](image-url)
Application to Regression-based OPC

• To predict edge movements in short runtime

Regression based method

Conventional model-based OPC (time consuming)

- Prediction model training with different patterns
  - PCA+Km, LE+Km, PCA+BC, LE+BC
  - Learning algorithm is fixed to Linear regression
- Metric: RMSPE (Root Mean Square Prediction Error)
Effectiveness of OPC regression

- Proposed framework achieved the best prediction accuracy

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSPE (nm)</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA+Km</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>LE+Km</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>PCA+BC</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>LE+BC</td>
<td>0.8</td>
<td></td>
</tr>
</tbody>
</table>

Prediction accuracy:
RMSPE: Root mean square prediction error

Ratio:
- PCA+Km: 1.0
- LE+Km: 1.1
- PCA+BC: 0.9
- LE+BC: 0.8
Conclusion

- We have introduced a new method to sample unique patterns.
- By applying our dimension reduction technique, dimensionality- and type-independent layout feature can be used in accordance with applications.
- The Bayesian clustering is able to classify layout data without manual parameter tuning.
- The experimental results show that our proposed method can effectively sample layout patterns that represent characteristics of whole chip layout.