Novel Transformer Model Based Clustering Method for Standard Cell Design Automation

Chia-Tung (Mark) Ho, Ajay Chandna, David Guan, Alvin Ho, Minsoo Kim, Yaguang Li and Haoxing (Mark) Ren | ISPD 2024
OUTLINE

• Introduction

• Novel Transformer Model Based Clustering Method

• Experimental Results

• Conclusion
STANDARD CELL LAYOUT AUTOMATION

- Standard Cells (STD) are building blocks of digital design layout: AND, NOR, Flip-Flop, Adder, etc
- Months manual design turn around time per library to deliver competitive Power, Performance, and Area (PPA)
- Standard cell layout automation benefits - NVCell (DAC2021, ISPD2023)
  - Productivity: Fast turn around time
  - Performance: Explore more design space
  - Performance: More custom cell design
  - Optimization: Design Technology Co-Optimization

![Standard Cell Layout Diagram](image)

GA100 - 1.7B standard cells
CHALLENGES: STANDARD CELL LAYOUT AUTOMATION

- Standard cell layout automation challenges as advancing beyond 5nm
  - Limited in-cell routing resource: less routing tracks (i.e., 5 routing tracks)
  - Design rule complexity: Increasing number and complexity of design rules + strict patterning rules
  - Scalability: > hundreds of transistors cell designs
- Better and Efficient standard cell layout automation framework
  - Routability
  - Scalability
  - High-quality PPA
  - How?

IRDS Roadmap 2022

Routability Challenges of a Latch Design in 5nm in stick format
HIGH QUALITY DEVICE CLUSTERING SHOULD CONSIDER TRANSISTOR LAYOUT CHARACTERISTICS

- Diffusion break/sharing
- Transistor pin access
- Routing metal DRCs

**Novel transformer model-based clustering methodology**

- Reduce complexity
- Narrow down searching space
- Assist finding routable + optimal layouts faster

---

**Diagram:**

- Cluster boundaries
- Diffusion break/sharing
- Transistor pin access
- Routing metal DRCs
- Active Transistor
- Gate
- MD
- M0
OUTLINE

• Introduction

• Novel Transformer Model Based Clustering Method

• Experimental Results

• Conclusion
NOVEL TRANSFORMER MODEL BASED CLUSTERING METHOD

Framework Overview

Schematic of Cell Logics (.sp) 
(Netlist Information)

Layout Specification 
(Cell Architecture)

Novel Transformer Model Based Clustering

\[ p_i^{k+1} = cBp_i^k + (1 - c)\sigma(y_i^T\mathbf{Y}); \ c: \text{jumping probability} \]

NVCell Standard Cell Layout Automation Framework (DAC 2021 + ISPD 2023)

Multi-Obj Opt.

Routability-Driven SA placer

GA based Router + RL DRC fixing

Objective space: CW, TWL, Congestion

Pareto front

Cluster 4  Cluster 2  Cluster 5  Cluster 3  Cluster 1
**Goal:** Learn the relationship between device pairs in the layout graph

**Netlist logic graph:** Topology is from spice netlist
- Nodes: Nets, devices, and pins
- Edges: Connections

**Layout graph:** Neighbor grids are all connected
- Column: gate terminal
- Row: PFET, NFET

**Unsupervised learning from LVS/DRC layouts**

**Global receptive field + Netlist Structure + Device placement relation**
NETLIST & LAYOUT GRAPH MULTI-HEAD ATTENTION

- **Spatial Relation bias**: shortest path distance between devices in the netlist logic graph
- **Device Placement bias**: device layout characteristics, such as diffusion sharing, vertical gate/diffusion connect.

\[
a_{v,u} = \frac{(h_v W_Q)(h_u W_K)^T}{\sqrt{d}} + b_{\phi(v,u)} + b_{\kappa(v,u)}
\]

- Spatial Relation Attn bias
- Device Placement Attn bias
NETLIST & LAYOUT AWARE PERSONALIZED PAGE RANK VECTOR CLUSTERING

For each device, calculate personalized page rank vector

\[ p^{k+1}_i = cBp^k_i + (1 - c)\sigma(y^T_iY); \ c: \text{jumping probability} \]

\[
\begin{bmatrix}
    p^{k+1}_1 \\
    p^{k+1}_2 \\
    p^{k+1}_3 \\
    p^{k+1}_4 \\
    p^{k+1}_5 \\
    p^{k+1}_6 \\
\end{bmatrix} =
\begin{bmatrix}
    0.33 & 0.16 & 0.33 & 0.16 \\
    0.33 & 0.16 & 0.33 & 0.16 \\
    0.2 & 0.2 & 0.2 & 0.4 \\
    0.25 & 0.25 & 0.125 & 0.25 & 0.125 \\
    0.2 & 0.2 & 0.4 & 0.2 \\
\end{bmatrix}
\begin{bmatrix}
    p^k_1 \\
    p^k_2 \\
    p^k_3 \\
    p^k_4 \\
    p^k_5 \\
    p^k_6 \\
\end{bmatrix} + (1 - c)
\begin{bmatrix}
    0.5 \\
    0.05 \\
    0.1 \\
    0.2 \\
    0.05 \\
    0.1 \\
\end{bmatrix}
\]

Predicted preference probability vector of i-th device (\( \sigma(y^T_iY) \))

DBScan Clustering from \( p_i \)
GENERATED LVS/DRC LAYOUT AFTER CLUSTERING

Generated LVS/DRC CleanLatch Design (~100 devices)

Manual Cell Width = 58 / Generated Cell Width = 56 TWL = 671
OUTLINE

• Introduction

• Novel Transformer Model Based Clustering Method

• Experimental Results

• Conclusion
Experimental Setup

- Training 250 training circuits set in 5nm library
  - Apply to larger and more complex circuits
  - Apply to different technology nodes
- **Experiment I:** Clustering Quality Study
- **Experiment II:** Results of 5nm industrial library
  - 94 complex cell benchmark + Entire cell library
- **Experiment III:** Results of 3nm (without retraining the model)
Clustering Quality Study

• Quality of DBScan clustering method with Different Models and Representative Node Vectors (Rep. Node Vectors)
• Silhouette score (Larger is better): clustering result and the actual LVS/DRC clean layout placement

Silhouette score \( = \frac{(b - a)}{\max(a, b)} \); \( a = \text{mean intraclass distance} \); \( b = \text{mean nearest cluster distance} \)

<table>
<thead>
<tr>
<th>Rep. Node Vectors</th>
<th>Model</th>
<th>Avg. Silhouette</th>
<th>Impr. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pred. Preference</td>
<td>GINE</td>
<td>0.21</td>
<td>200%</td>
</tr>
<tr>
<td></td>
<td>Transformer based</td>
<td>0.42</td>
<td>50%</td>
</tr>
<tr>
<td>PPR with Pred. Preference</td>
<td>N/A</td>
<td>0.22</td>
<td>186%</td>
</tr>
<tr>
<td></td>
<td>GINE</td>
<td>0.37</td>
<td>70%</td>
</tr>
<tr>
<td></td>
<td>Transformer based (proposed)</td>
<td>0.63</td>
<td>-</td>
</tr>
</tbody>
</table>

(a) Proposed

(b) PPR with Pred. Preference + GINE
RESULTS of 5nm

Achieved 12.7X speedup on average

<table>
<thead>
<tr>
<th>Success Rate (%)</th>
<th>Cell Width Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Smaller</td>
</tr>
<tr>
<td>NVCell (DAC 2021)</td>
<td>0%</td>
</tr>
<tr>
<td>NVCell2 (ISPD 2023)</td>
<td>87.2%</td>
</tr>
<tr>
<td>Proposed</td>
<td>100%</td>
</tr>
</tbody>
</table>

Smaller: 14.5%
Same: 83.3%
Larger: 2.2%

Achieved 12.7X speedup on entire cell library (1078 Cells)

On a difficult routing benchmark (94 cells)

On entire cell library (1078 Cells)
RESULTS of 3nm
Without Retraining Model

5 Routing Tracks Cell Architecture (938 Cells)

Cell Area Comparison

<table>
<thead>
<tr>
<th>Cell Complexity</th>
<th>Smaller: 11.4%</th>
<th>Same: 81.7%</th>
<th>Larger: 4.4%</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Success Rate (%)

Proposed: 100%

Total Cell: 938 Cells

4 Routing Tracks Cell Architecture (458 Cells)

Cell Area Comparison

<table>
<thead>
<tr>
<th>Cell Complexity</th>
<th>Smaller: 7.2%</th>
<th>Same: 92.4%</th>
<th>Larger: 0.4%</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>103</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Success Rate (%)

Proposed: 100%

Total Cell: 458
OUTLINE

• Introduction

• Novel Transformer Model Based Clustering Method

• Experimental Results

• Conclusion
CONCLUSIONS

• Proposed Novel Transformer Model Based Clustering Method successfully improves the success rate, performance, and area
  • 100% success rate for 5nm
  • 14.5% smaller cell width + 83.3% same cell width
  • Avg. 12.7X speedup on cells with > 80 devices than previous work

• Transferable to different technology nodes

• Competitive PPA result: Power impr. up to 12%, Delay impr. up to 8%, and Area impr. up to 14.29%

• Future works:
  • PPA-driven standard cell layout automation
  • Cluster constraint debugging and optimization for PPA