Novel Transformer Model Based Clustering Method for Standard Cell Design Automation

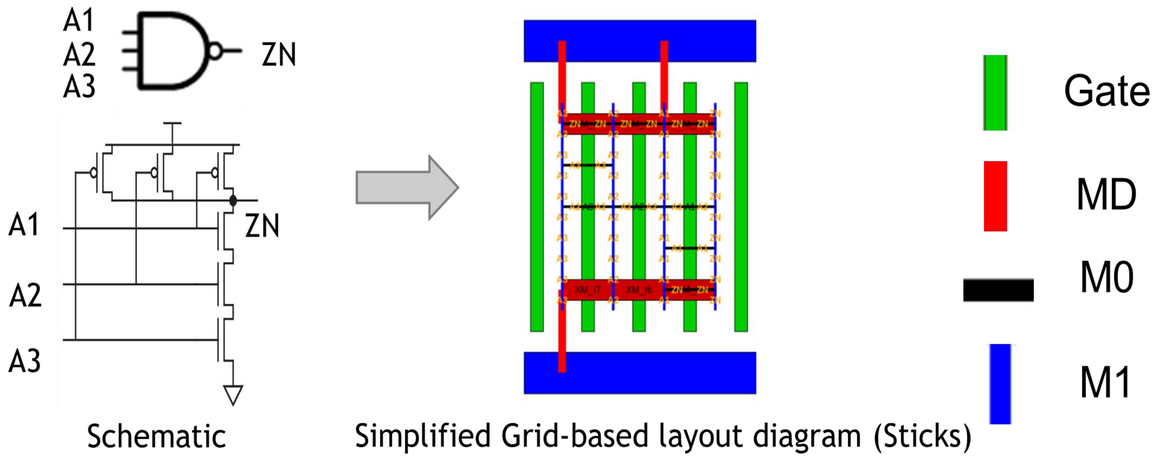
Chia-Tung (Mark) Ho, Ajay Chandna, David Guan, Alvin Ho, Minsoo Kim,
Yaguang Li and **Haoxing (Mark) Ren** | **ISPD 2024**

OUTLINE

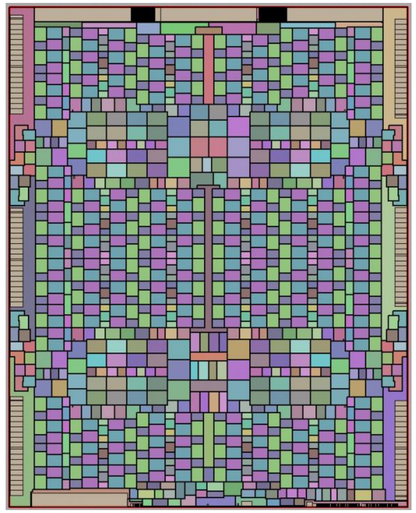
- **Introduction**
- Novel Transformer Model Based Clustering Method
- Experimental Results
- Conclusion

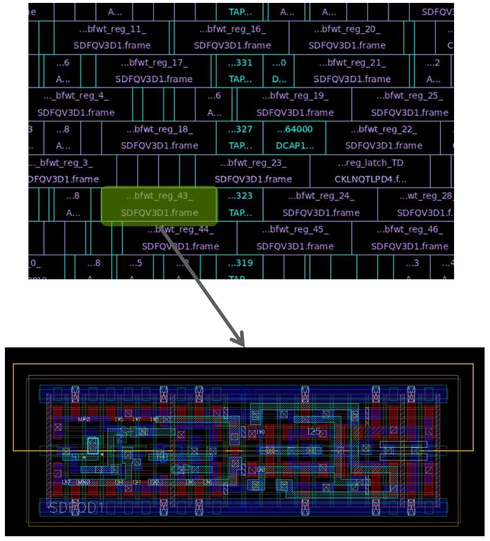
STANDARD CELL LAYOUT AUTOMATION

- 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

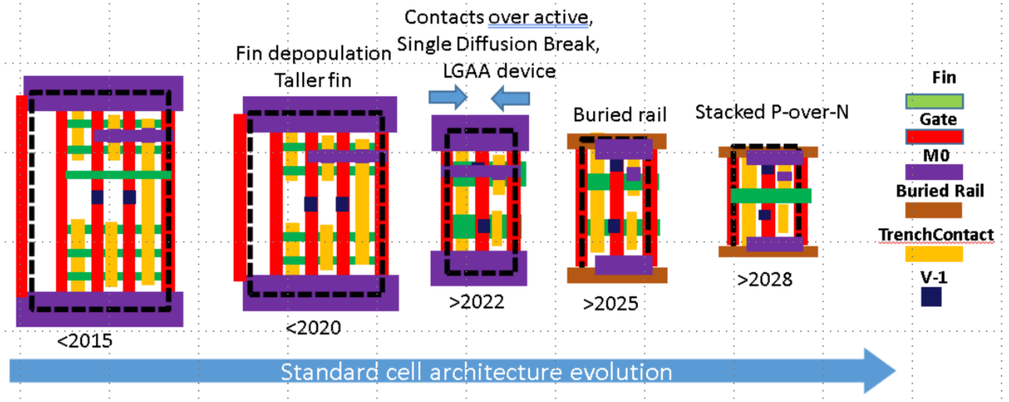




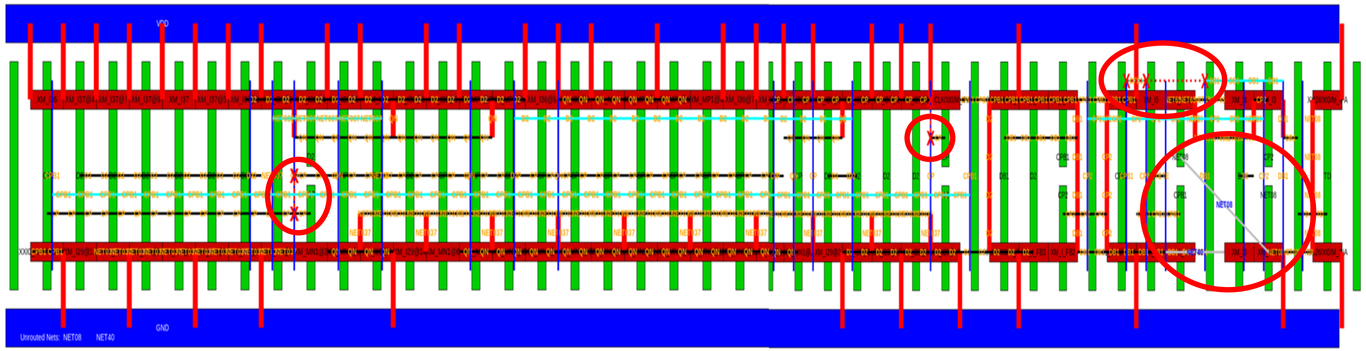
GA100 – 1.7B standard cells

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?



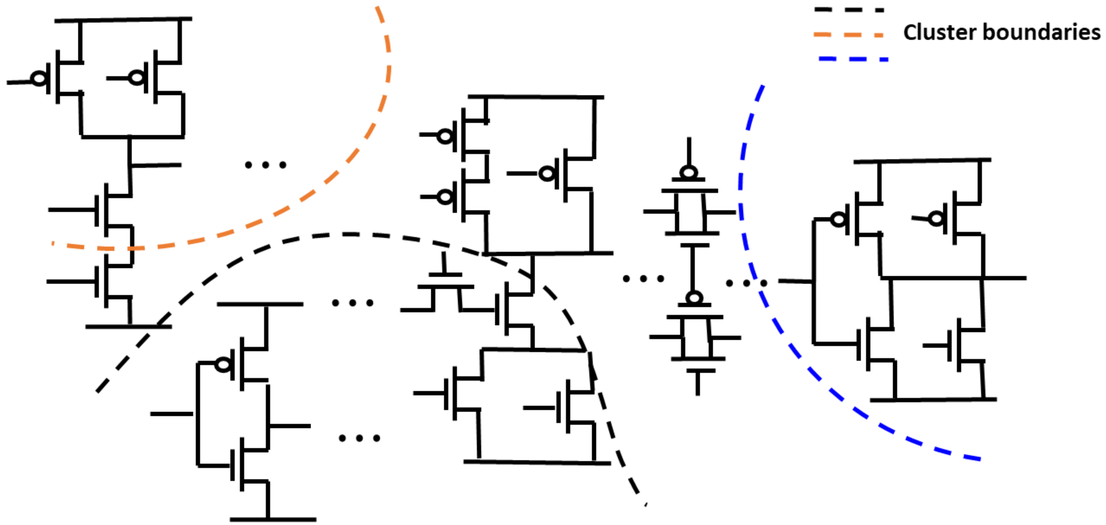
IRDS Roadmap 2022

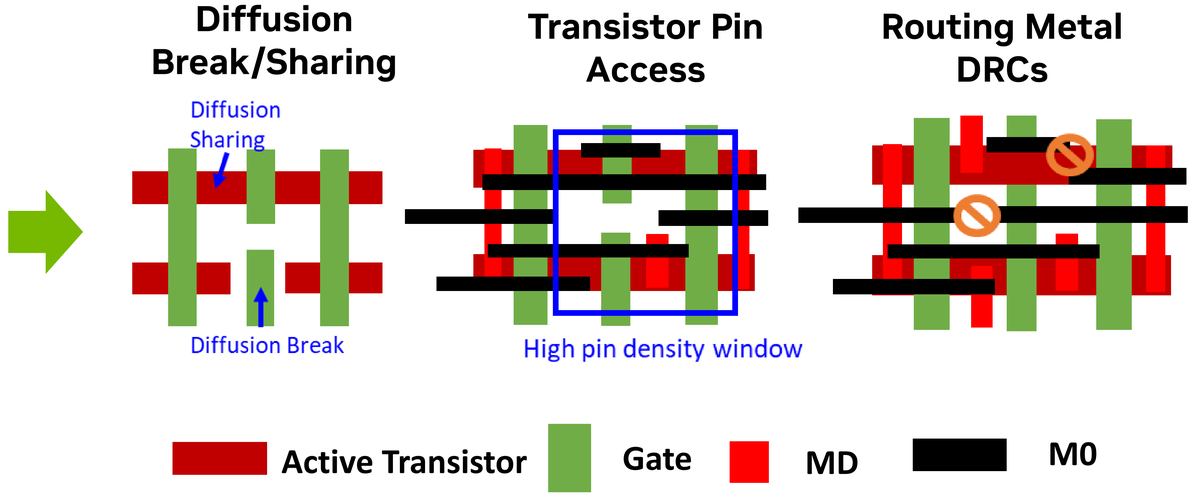


Routability Challenges of a Latch Design in 5nm in stick format

LAYOUT-AWARE DEVICE CLUSTRING

- 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**





OUTLINE

- Introduction
- **Novel Transformer Model Based Clustering Method**
- Experimental Results
- Conclusion

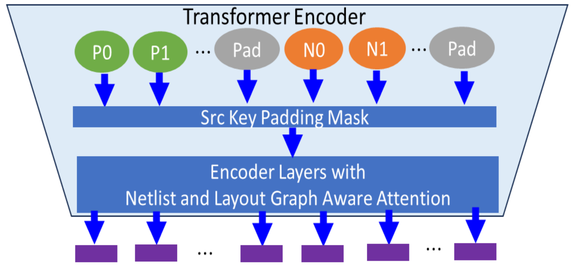
NOVEL TRANSFORMER MODEL BASED CLUSTERING METHOD

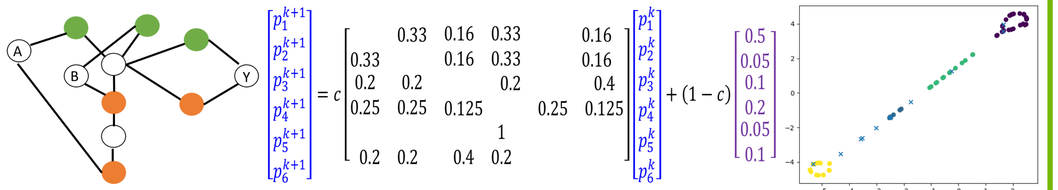
Framework Overview

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

Layout Specification
(Cell Architecture)

Novel Transformer Model Based Clustering



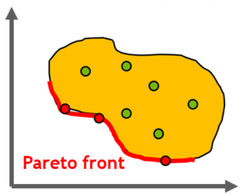


$$\begin{bmatrix} p_1^{k+1} \\ p_2^{k+1} \\ p_3^{k+1} \\ p_4^{k+1} \\ p_5^{k+1} \\ p_6^{k+1} \end{bmatrix} = c \begin{bmatrix} & 0.33 & 0.16 & 0.33 & & 0.16 \\ 0.33 & & 0.16 & 0.33 & & 0.16 \\ 0.2 & 0.2 & & 0.2 & & 0.4 \\ 0.25 & 0.25 & 0.125 & & 0.25 & 0.125 \\ & & & 1 & & \\ 0.2 & 0.2 & 0.4 & 0.2 & & \end{bmatrix} \begin{bmatrix} p_1^k \\ p_2^k \\ p_3^k \\ p_4^k \\ p_5^k \\ p_6^k \end{bmatrix} + (1-c) \begin{bmatrix} 0.5 \\ 0.05 \\ 0.1 \\ 0.2 \\ 0.05 \\ 0.1 \end{bmatrix}$$

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

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

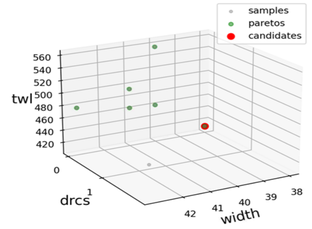
Multi-Obj Opt.

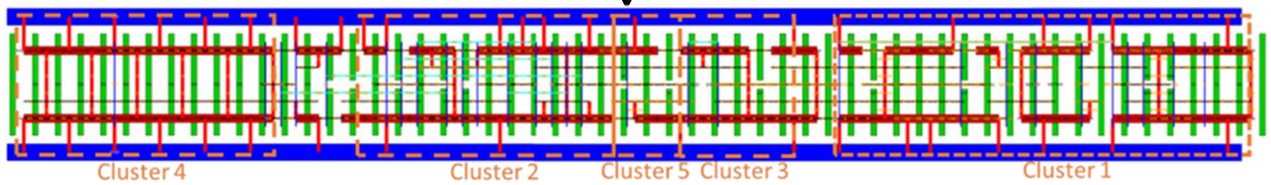


Objective space:
CW, TWL, Congestion

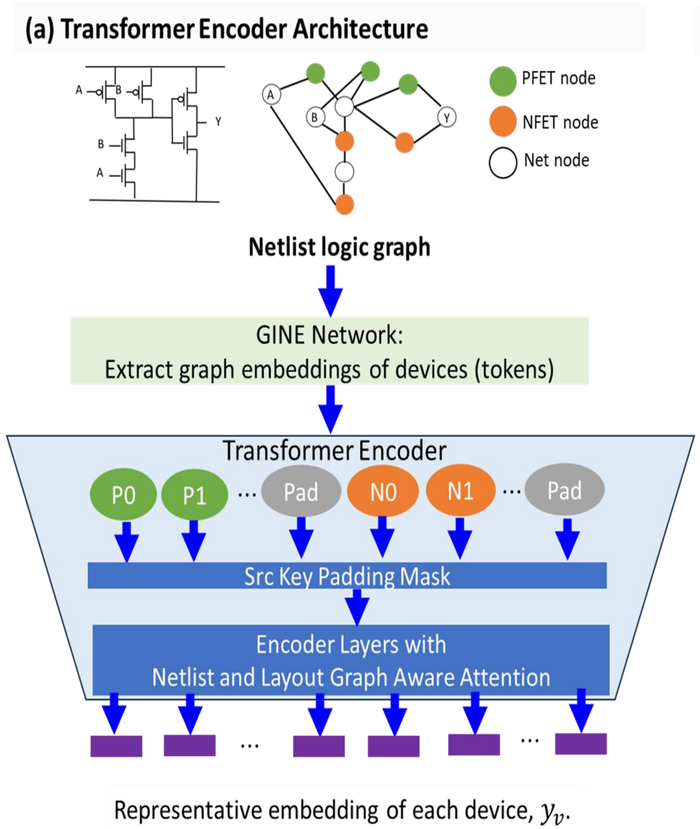
Routability-Driven SA placer

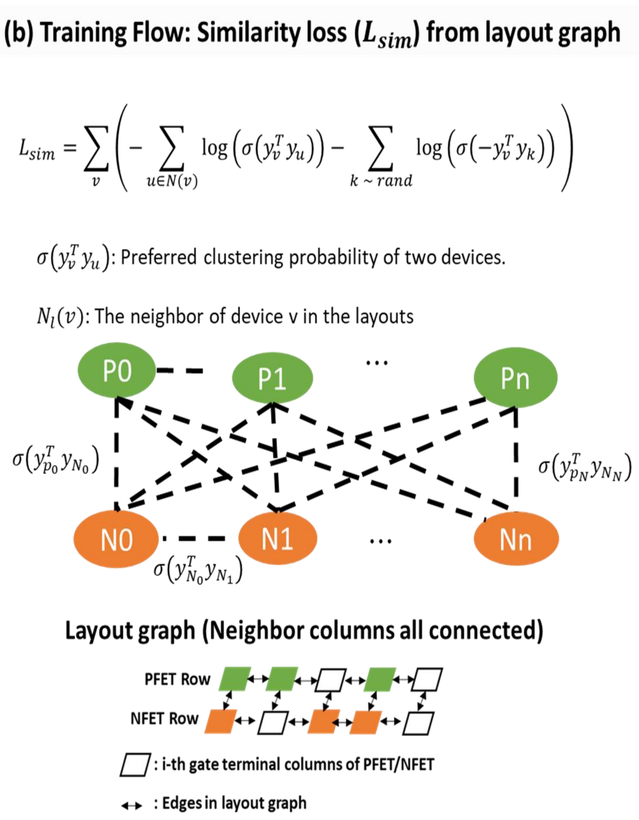
GA based Router + RL DRC fixing





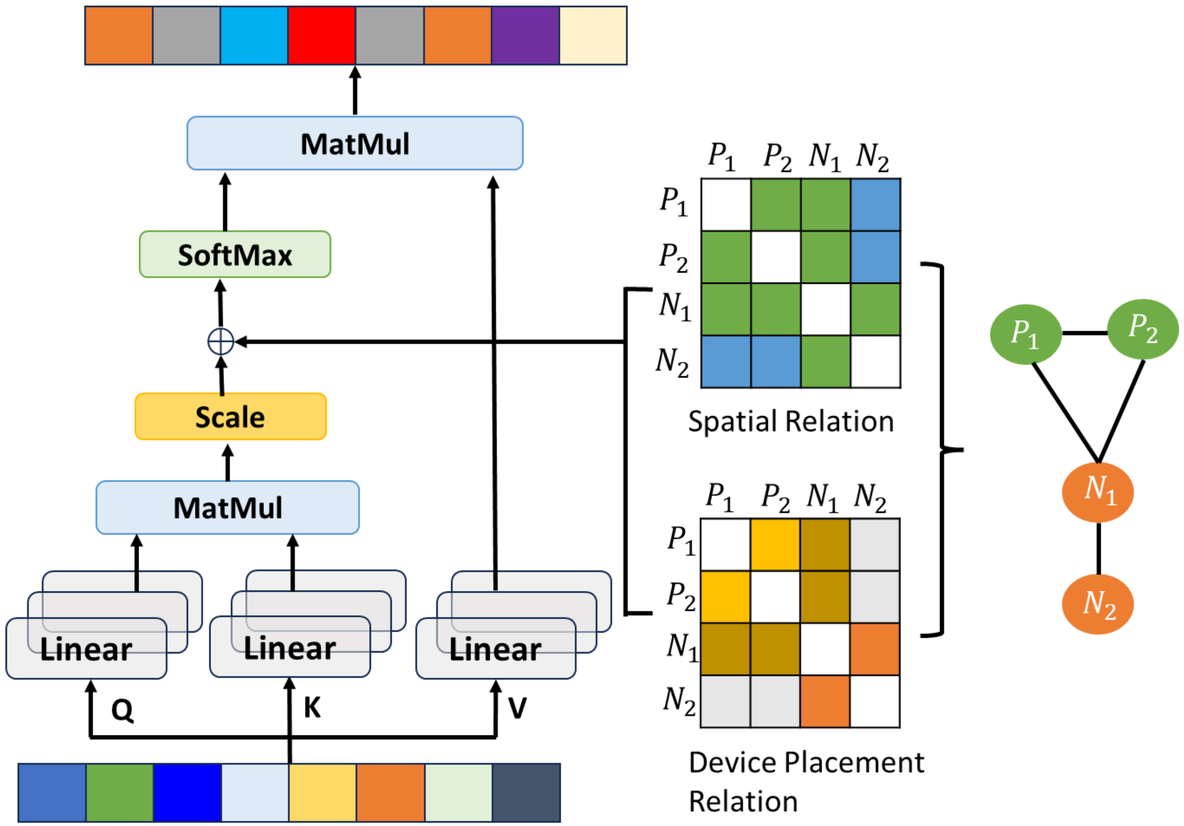
TRANSFORMER ENCODER STRUCTURE





- 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



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

Spatial Relation Attn bias

Device Placement Attn bias

- **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.

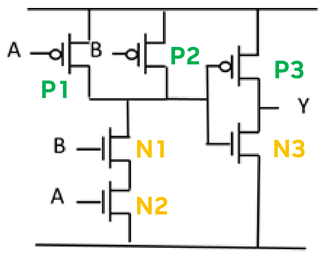
NETLIST & LAYOUT AWARE PERSONALIZED PAGE RANK VECTOR CLUSTERING

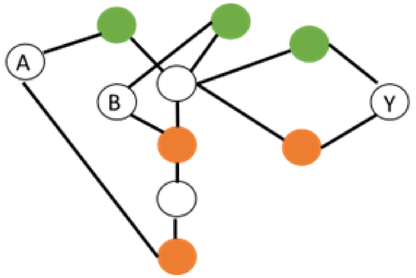
For each device, calculate personalized page rank vector

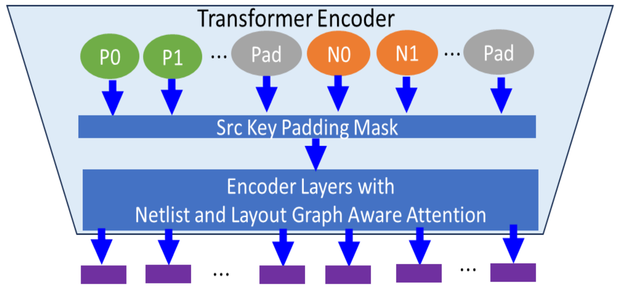
$$p_i^{k+1} = c\mathbf{B}p_i^k + (1-c)\sigma(y_i^T\mathbf{Y});\ c\text{: jumping probability}$$

$$\begin{bmatrix} p_1^{k+1} \\ p_2^{k+1} \\ p_3^{k+1} \\ p_4^{k+1} \\ p_5^{k+1} \\ p_6^{k+1} \end{bmatrix} = c \begin{bmatrix} & 0.33 & 0.16 & 0.33 & & 0.16 \\ 0.33 & & 0.16 & 0.33 & & 0.16 \\ 0.2 & 0.2 & & 0.2 & & 0.4 \\ 0.25 & 0.25 & 0.125 & & 0.25 & 0.125 \\ & & & 1 & & \\ 0.2 & 0.2 & 0.4 & 0.2 & & \end{bmatrix} \begin{bmatrix} p_1^k \\ p_2^k \\ p_3^k \\ p_4^k \\ p_5^k \\ p_6^k \end{bmatrix} + (1-c) \begin{bmatrix} 0.5 \\ 0.05 \\ 0.1 \\ 0.2 \\ 0.05 \\ 0.1 \end{bmatrix}$$

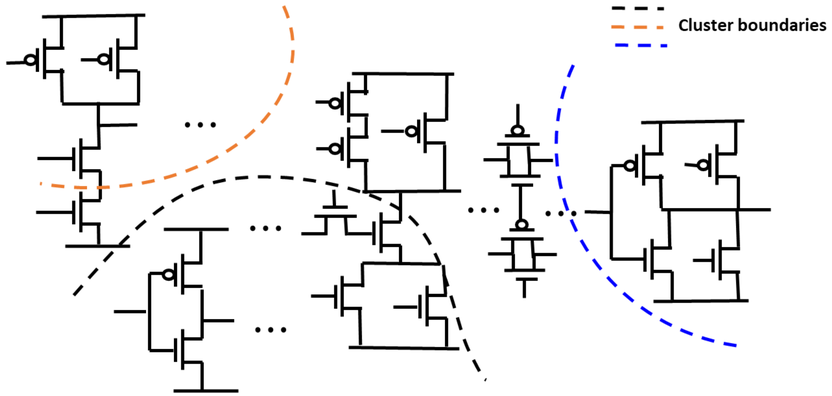
Predicted preference probability vector of i-th device ($\sigma(y_i^T\mathbf{Y})$)

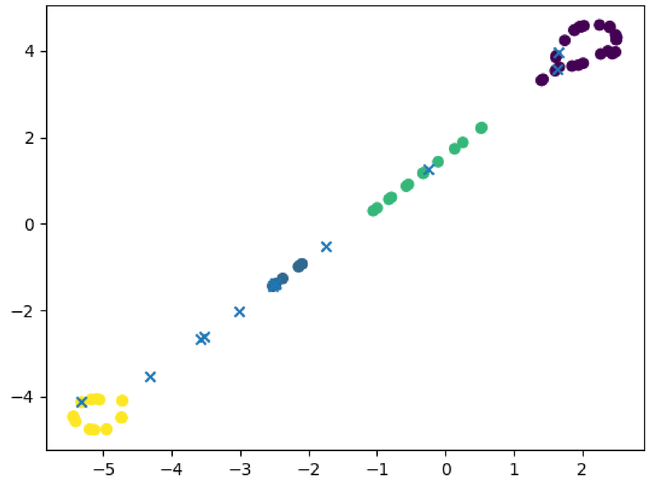






DBScan Clustering from p_i



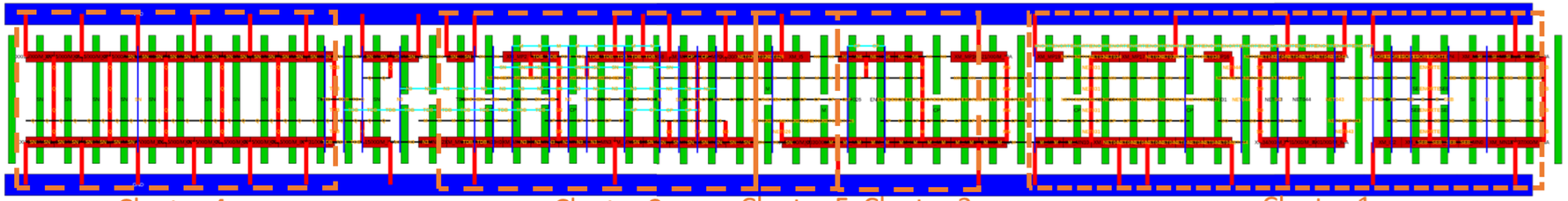


PCA projected cluster result

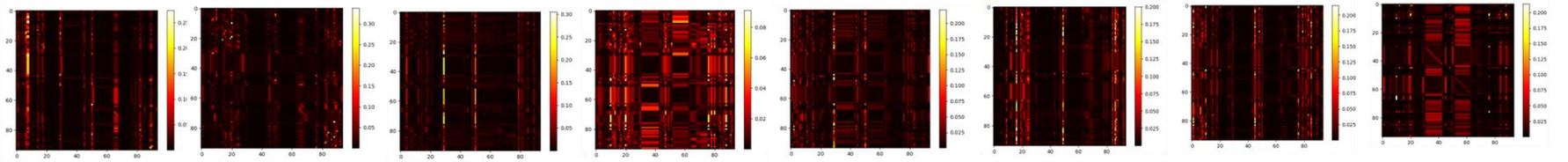
GENERATED LVS/DRC LAYOUT AFTER CLUSTERING

Generated LVS/DRC Clean Latch Design (~ 100 devices)

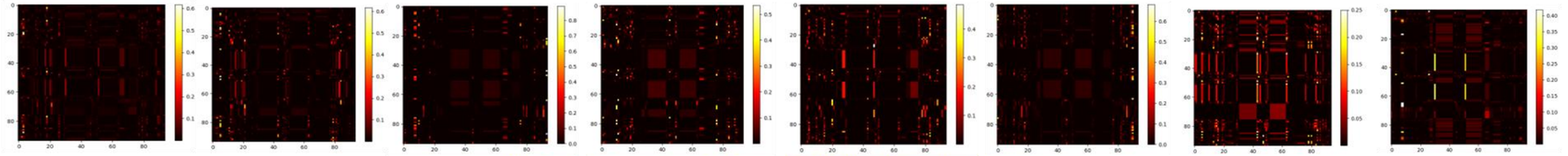
Manual Cell Width = 58 / Generated Cell Width = 56 TWL = 671



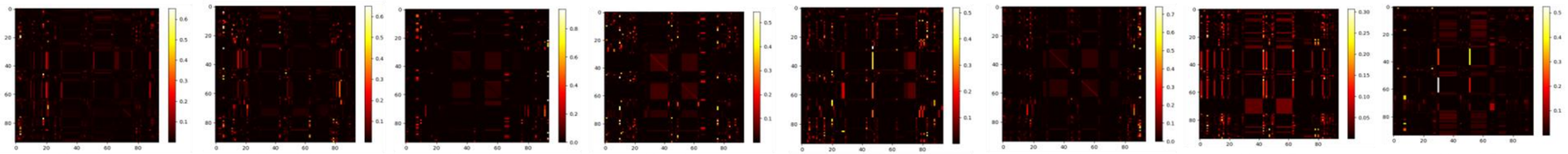
Netlist & Layout Graph Aware Multi-Head Attention Layer 1



Netlist & Layout Graph Aware Multi-Head Attention Layer 2



Netlist & Layout Graph Aware Multi-Head Attention Layer 3

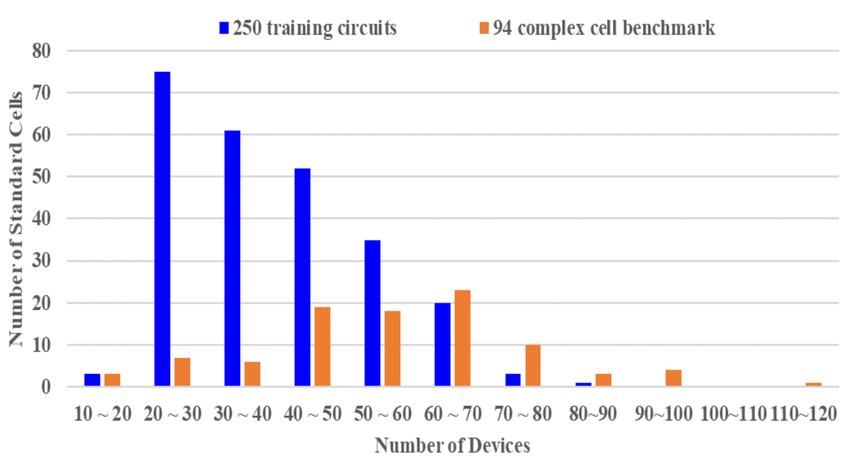


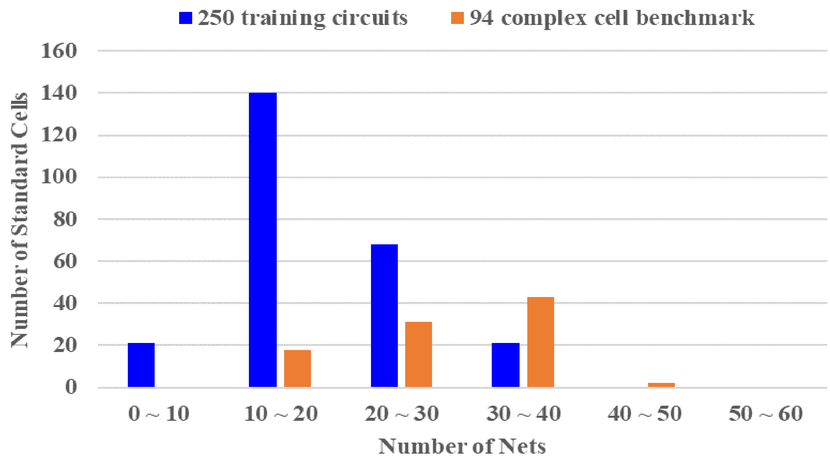
OUTLINE

- 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)





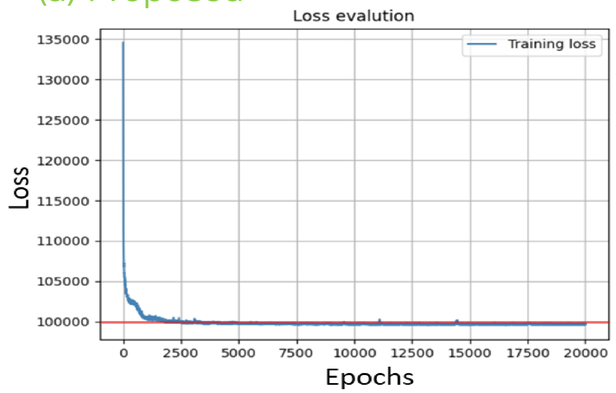
Clustering Quality Study

- 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

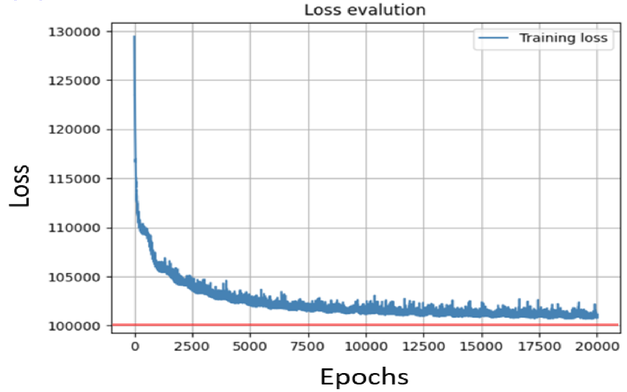
$$Silhouette\ score = \frac{(b-a)}{\max(a,b)}\ ;\ a = mean\ intraclass\ distance\ ;\ b = mean\ nearest\ cluster\ distance$$

Rep. Node Vectors	Model	Avg. Silhouette	Impr. (%)
Pred. Preference	GINE	0.21	200%
	Transformer based	0.42	50%
PPR with Pred. Preference	N/A	0.22	186%
	GINE	0.37	70%
	Transformer based (proposed)	**0.63**	-

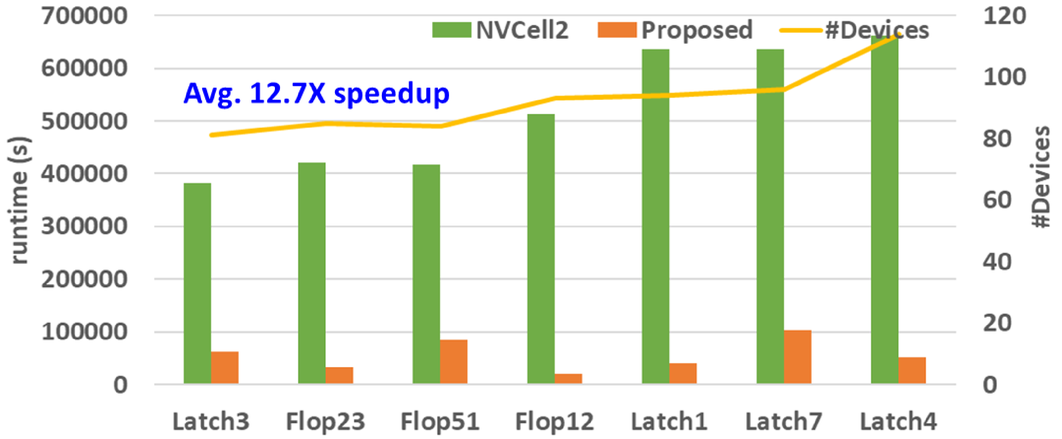
(a) Proposed



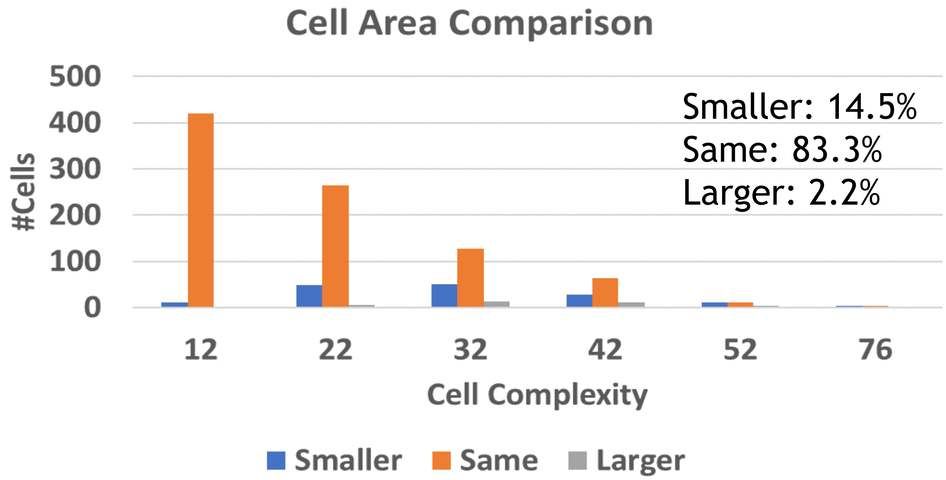
(b) PPR with Pred. Preference + GINE



RESULTS of 5nm



Achieved 12.7X speedup on average



	Success Rate (%)	Cell Width Comparison		
		Smaller	Same	Larger
NVCell (DAC 2021)	0%	N/A	N/A	N/A
NVCell2 (ISPD 2023)	87.2%	22.3%	37.2%	27.7%
Proposed	100%	29.8%	58.5%	11.7%

On a difficult routing benchmark (94 cells)

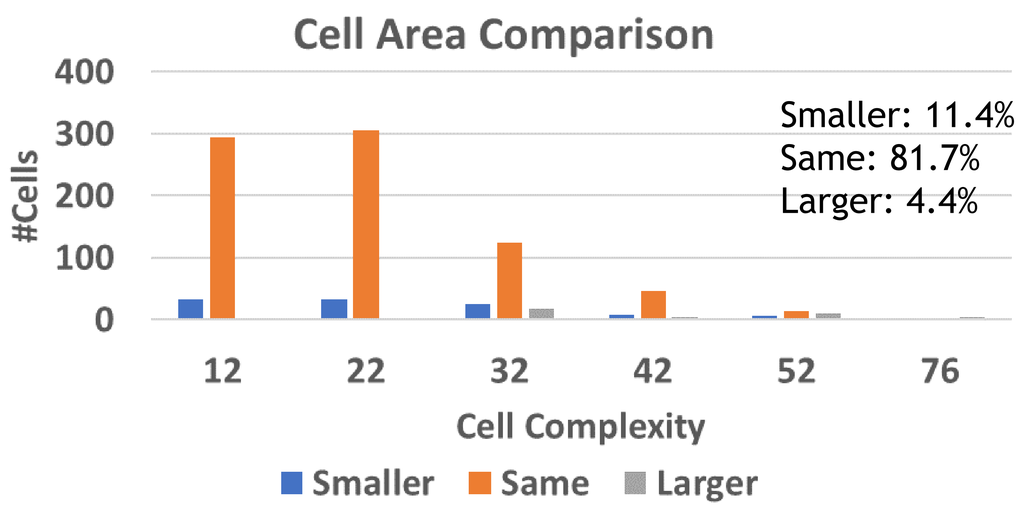
	Success Rate (%)	Cell Width Comparison		
		Smaller	Same	Larger
NVCell (DAC 2021)	91.2%	11.8%	77.6%	1.8%
NVCell2 (ISPD 2023)	98.8%	13.7%	80.1%	4.3%
Proposed	100%	14.5%	83.3%	2.2%

On entire cell library (1078 Cells)

RESULTS of 3nm

Without Retraining Model

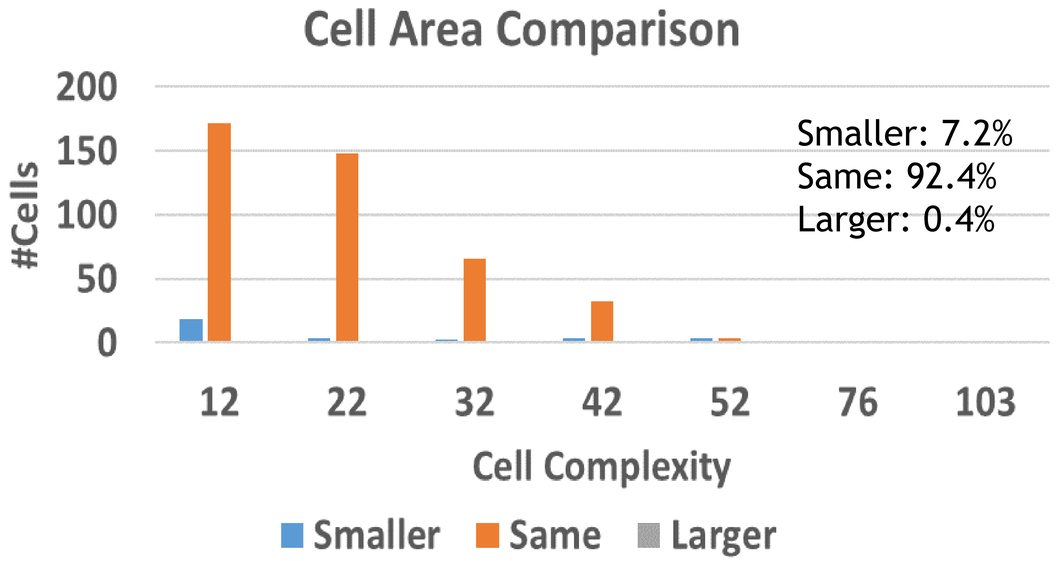
5 Routing Tracks Cell Architecture (938 Cells)



	Success Rate (%)	Cell Width Comparison		
		Smaller	Same	Larger
Proposed	100%	11.4%	81.7%	4.4%

Total Cell: 938 Cells

4 Routing Tracks Cell Architecture (458 Cells)



	Success Rate (%)	Cell Width Comparison		
		Smaller	Same	Larger
Proposed	100%	7.2%	92.4%	0.4%

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