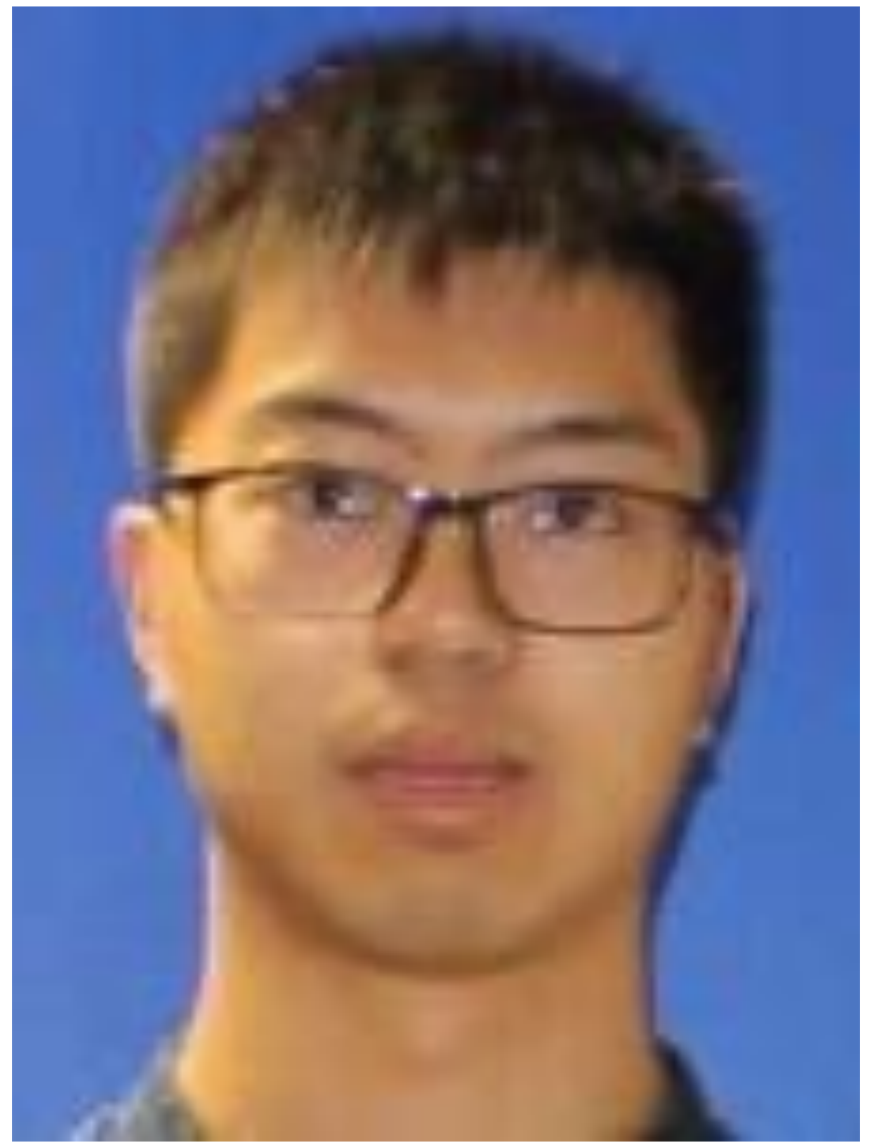
ISPD'24 Contest: GPU/ML-Enhanced Large Scale Global Routing Contest

Rongjian Liang, Anthony Agnesina, Wen-Hao Liu, Mark Ren

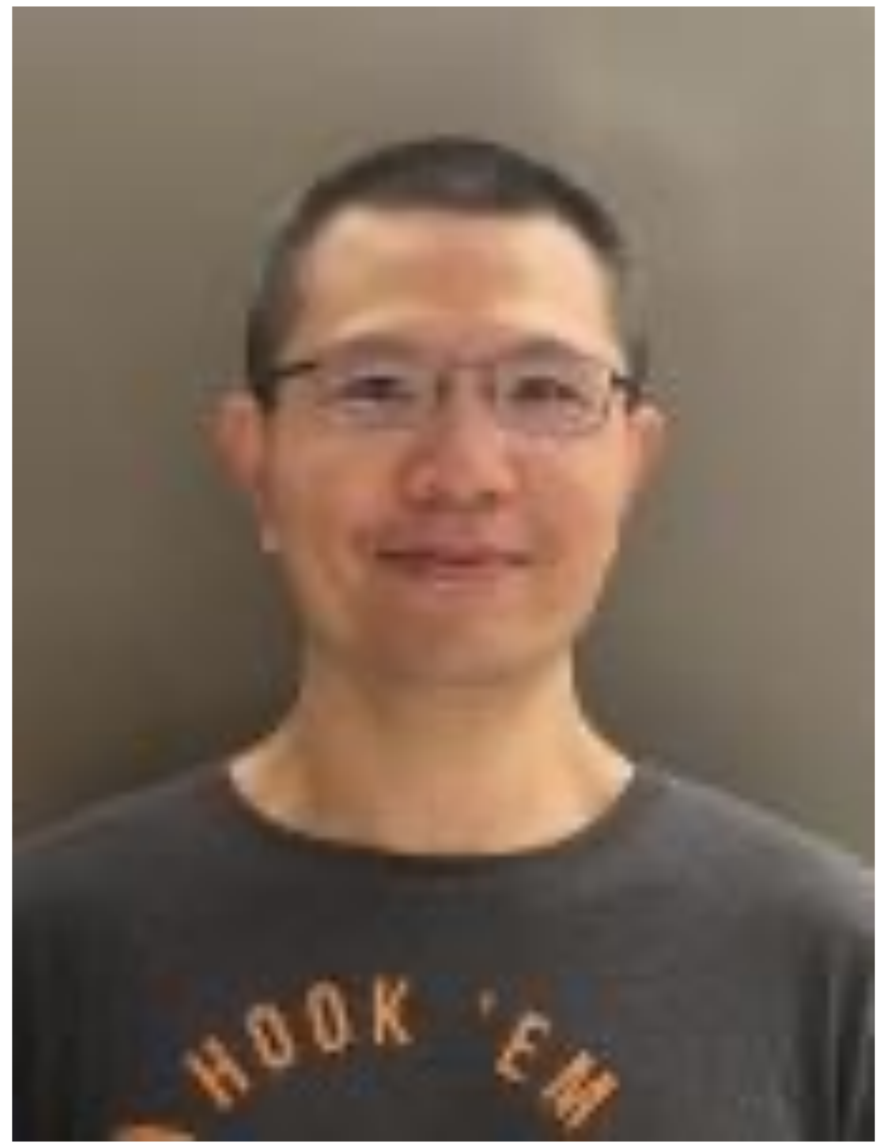
NVIDIA

Contest Organizers

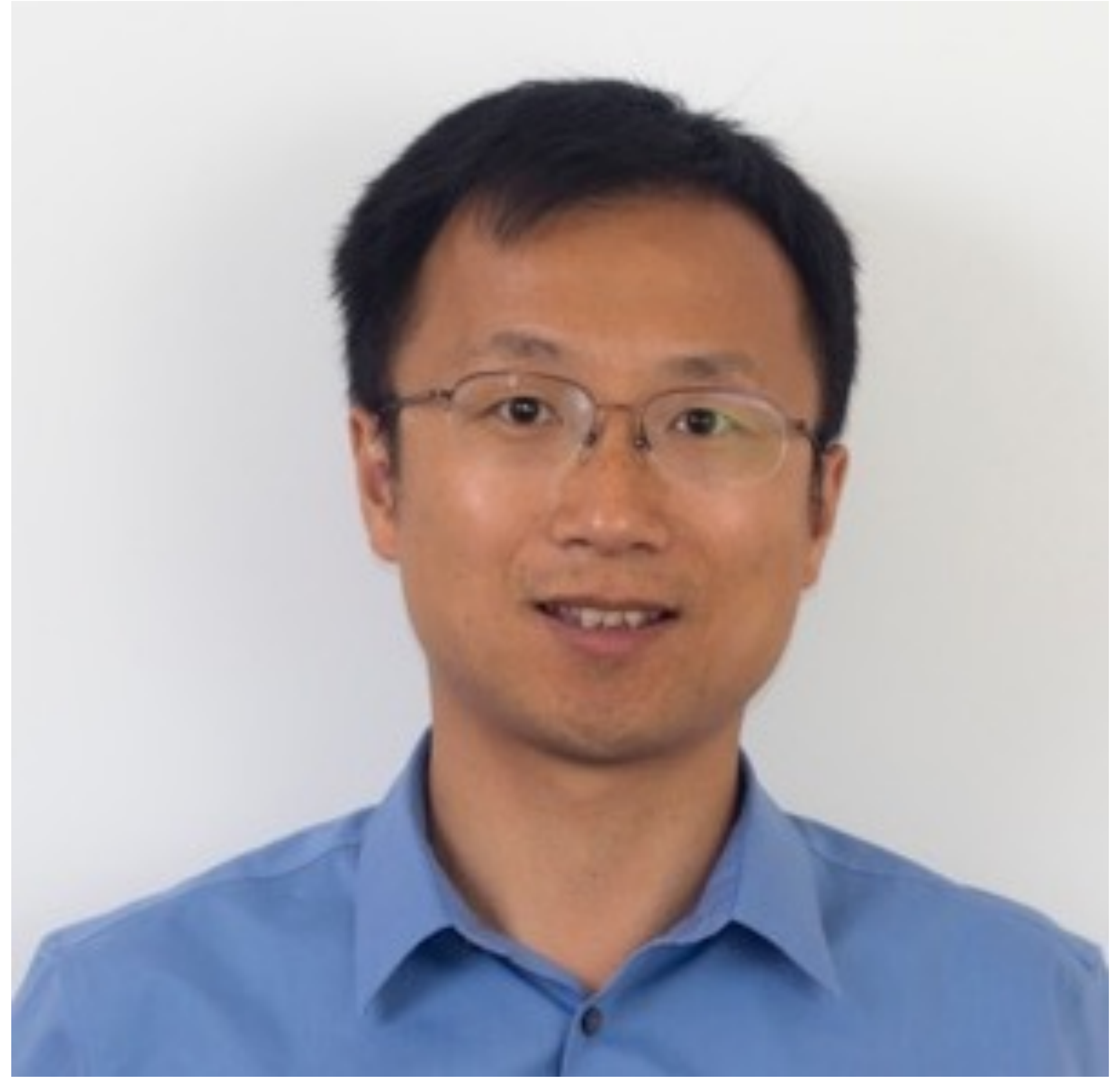


Rongjian Liang
Problem formulation & Evaluation

Anthony Agnesina
Benchmarks



Wen-Hao Liu
Evaluator



Haoxing (Mark) Ren
High-level guidance

Motivations

- Global routing techniques have many **applications across various stages** of modern VLSI design flow
- **Scalability** of global routing techniques is a serious challenge
- **GPU acceleration and ML techniques** have great potential to address the scalability challenge

Table 1: Application of global routing across various VLSI design stages.

Global Router Usage	Design Size	Runtime Requirement	Congestion Resolving Effort	Global Router Features	Goals
Logic Synthesis	20M—100M	fast	low	routability estimation, timing estimation	guide physical-friendly netlist
Physical Planning	20M—100M	fast	low	routability estimation, timing estimation	partition design, I/O planning, timing budgeting
Placement	1M—5M	fast	low	routability estimation, timing estimation	guide routability- and timing-aware placement
Optimization	1M—5M	medium	medium	timing-driven topology, buffering-friendly topology, incremental update	guide routability-aware optimization
Guide Detailed Routing	1M—5M	long	high	timing-driven topology, power-driven	guide high-quality DR result

Problem Formulation

- Input
  - A 3D routing space defined using a GCell grid graph (.cap file)
  - Net information (.net file)
- Output
  - Routing guide on the Gcell grid graph
- Objectives:
  - A concrete path for each net
  - Minimize total wirelength
  - Minimize via count
  - Minimize overflow

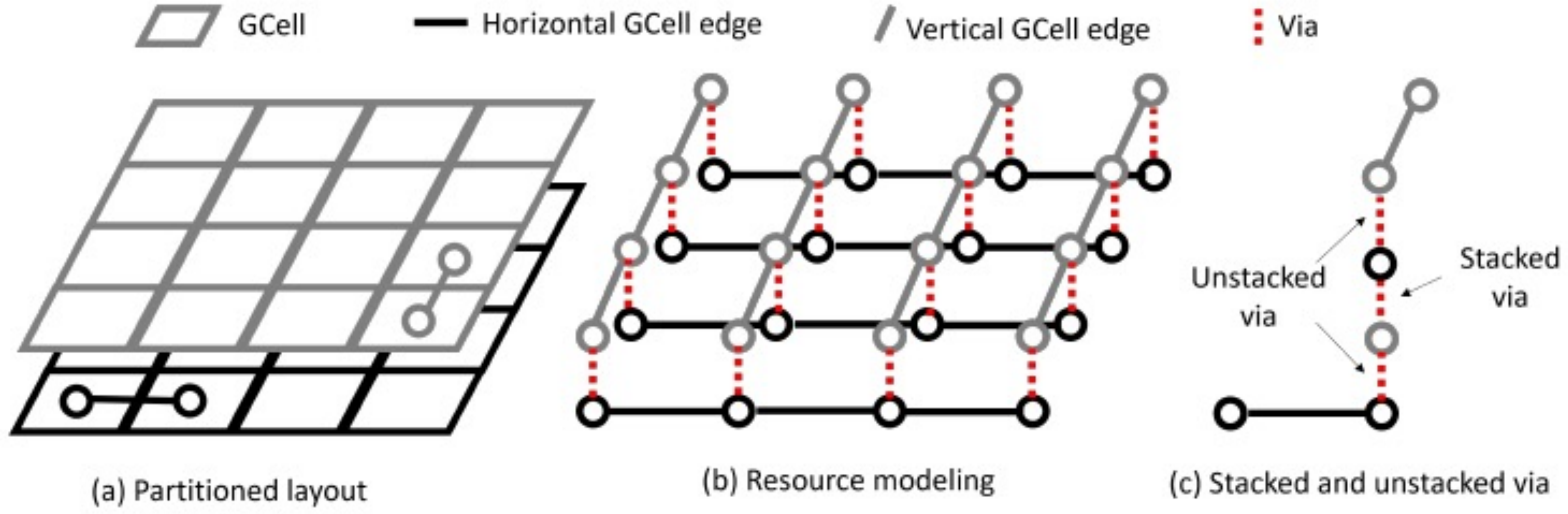


Figure 1: Illustration of a GCell grid graph.

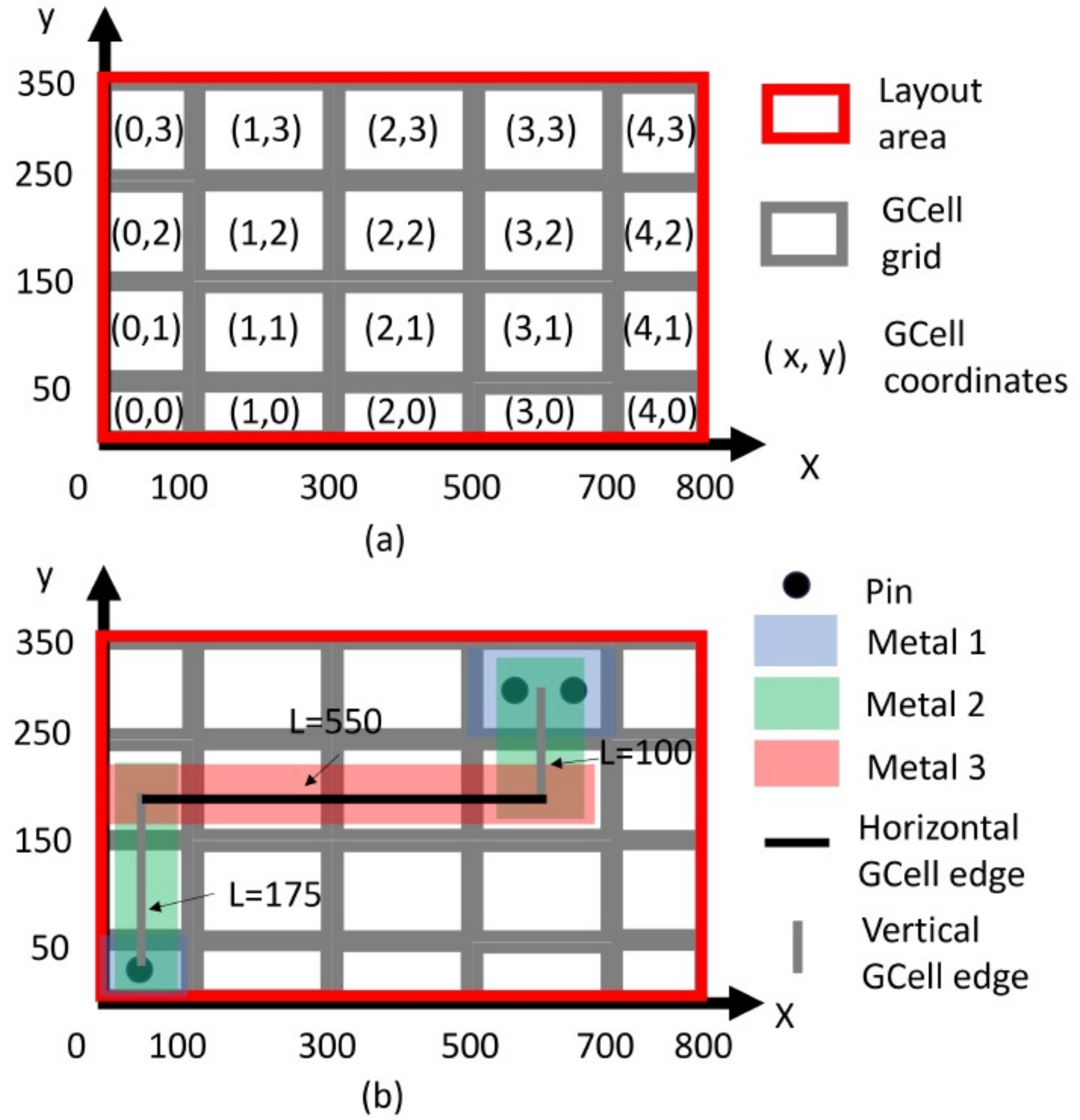


Figure 2: Example of GCell definition and global routing solution. (a) shows an example of GCell definition; (b) depicts a global routing solution (adapted from [6]).

Distinctions Compared to Previous Routing Contest

1. **Large scale testcases** (up to 50M cells) and tight runtime/memory budgets (< 50000s and 200GB RAM)
2. **GPU** server to encourage the usage of GPU acceleration and ML techniques
3. **Simplified input/output formats**
4. Two sets of evaluation metrices to advance the frontier of global routing techniques applicable across various stages

Benchmark Suit Characteristics

- The benchmarks derived from [1]
- The benchmarks are synthesized using the nangate45nm technology and cell libraries
- The benchmarks are divided into two sets
  - One placed layout from each design for testing (public)
  - Another placed layout from each design for final ranking (blind)
- Simplifications for the contest
  - Power and ground nets are removed
  - Clock tree routing is not considered

[1] C.-K. Cheng, A. B. Kahng, S. Kundu, Y. Wang and Z. Wang, "Assessment of Reinforcement Learning for Macro Placement", Proc. ISPD, 2023

Benchmark Suit Characteristics (Cont.)

Table 2: Design statistics of the ISPD 2024 benchmark suite. Technology is the NanGate 45nm process. Approximate numbers are reported — we provide variations in terms of netlist/placement/floorplan for public testing vs. blind evaluation/ranking. The number of routing layers is 10. The GCell is typically a square between 16 and 32 standard rows in size.

Design	# std cells	# macros	# nets	# pins	density (%)	GCell grid
public						
Ariane_sample	122K	133	129K	420K	51	844×1144
MemPool-Tile_sample	129K	20	136K	500K	51	475×644
NVDLA_sample	166K	128	177K	630K	51	1240×1682
BlackParrot_sample	715K	220	770K	2.9M	68	1532×2077
MemPool-Group_sample	3.1M	320	3.3M	10.9M	68	1782×2417
MemPool-Cluster_sample	9.9M	1296	10.6M	40.2M	68	3511×4764
TeraPool-Cluster_sample	49.7M	4192	59.3M	213M	68	7891×10708
blind						
Ariane_rank	121K	133	128K	435K	68	716×971
MemPool-Tile_rank	128K	20	136K	483K	68	429×581
NVDLA_rank	164K	128	174K	610K	68	908×1682
BlackParrot_rank	780K	220	825K	2.9M	68	1532×2077
MemPool-Group_rank	3.0M	320	3.2M	10.9M	68	1782×2417
MemPool-Cluster_rank	9.9M	1296	10.6M	40.2M	51	4113×5580
TeraPool-Cluster_rank	49.7M	4192	59.3M	213M	51	9245×12544

Evaluation Metric

- **Main track**

$$scaled\_score = original\_score \cdot (1 + runtime\_factor + np)$$

$$original\_score = w_1 \cdot \text{TotalWL} + w_2 \cdot \text{ViaCount} + \text{OverflowScore},$$

$$\text{OverflowCost}(c, d, l) = \text{OFWeight}[l] \cdot e^{s(d-c)} \qquad s = \begin{cases} 0.5 & \text{if } c > 0, \\ 1.5 & \text{if } c = 0, \end{cases}$$

$$T = 0.02 \cdot \log_2 \left(\frac{GRouter\_Wall\_Time}{Median\_Wall\_Time} \right)$$

$$runtime\_factor = \min(0.2, \max(-0.2, T))$$

- **Special honor track**

$$T = 0.05 \cdot \log_2 \left(\frac{GRouter\_Wall\_Time}{Median\_Wall\_Time} \right)$$

$$runtime\_factor = \min(0.5, \max(-0.5, T))$$

Further emphasize runtime scalability to facilitate the deployment of global routing techniques in early design stages

Ranking

- Weight-sum of the ranking of each benchmark
- The weight is proportional to **the cube root of the #nets** of the benchmark

Design	#nets	weight
Ariane_rank	128K	0.05
MemPool-Tile_rank	136K	0.05
NVDLA_rank	174K	0.05
BlackParrot_rank	825K	0.09
MemPool-Group_rank	3.2M	0.15
MemPool-Cluster_rank	10.6M	0.22
TeraPool-Cluster_rank	59.3M	0.39

Evaluation Platform

- RAM: 200 GB
- CPU Cores: 8 cores
- GPUs: 4 NVIDIA A100 GPUs

Participation Statistics

- 52 initial registrations
  - Asia: 33 teams
  - Europe: 1 team
  - North America: 5 teams
  - South America: 1 team
  - Others (unknown affiliations): 12 teams

- 18 final submissions

Prizes

Prizes are based on the ranking in the main track

- 1st place: $1000 + one NVIDIA GPU of similar value
- 2nd place: $500 + one NVIDIA GPU of similar value
- 3rd place: $250 + one NVIDIA GPU of similar value

Contest Results – Main Track

	ariane		bsg		nvdla		mempool_tile		mempool_group		mempool_cluster		tera_cluster		
Team	score	rank	score	rank	score	rank	score	rank	score	rank	score	rank	score	rank	weighted_rank
A	22326149	**3**	1.11E+08	**6**	42785372	**3**	13808883	**3**	3.71E+08	**2**	1.72E+09	**1**	1.202E+10	**1**	**1.9**
B	22126529	**2**	1.07E+08	**1**	42673699	**2**	13750400	**2**	3.77E+08	**3**	1.73E+09	**2**	1.204E+10	**2**	**2.06**
C	21577912	**1**	1.07E+08	**2**	42129054	**1**	13219650	**1**	3.67E+08	**1**	1.74E+09	**3**	1.208E+10	**3**	**2.31**
D	22718430	**5**	1.1E+08	**4**	43474248	**4**	14025624	**5**	3.9E+08	**4**	1.81E+09	**4**	1.256E+10	**4**	**4.1**
E	22501510	**4**	1.1E+08	**3**	43763726	**5**	14094144	**6**	3.95E+08	**5**	1.85E+09	**5**	1.4E+10	**6**	**5.21**

Contest Results – Special Honor

	ariane		bsg		nvdla		mempool_tile		mempool_group		mempool_cluster		tera_cluster		
Team	score	rank	score	rank	score	rank	score	rank	score	rank	score	rank	score	rank	weighted_rank
B	21068226	**3**	1.01E+08	**2**	40538542	**2**	13238490	**3**	3.53E+08	**2**	1.59E+09	**1**	1.14E+10	**1**	**1.49**
C	20100380	**1**	1.01E+08	**1**	40445341	**1**	12309656	**1**	3.41E+08	**1**	1.67E+09	**3**	1.16E+10	**2**	**1.83**
A	20845927	**2**	1.07E+08	**3**	40664270	**3**	12831966	**2**	3.53E+08	**3**	1.64E+09	**2**	1.17E+10	**3**	**2.68**
D	22054900	**5**	1.08E+08	**5**	43566415	**4**	13701731	**4**	3.8E+08	**4**	1.75E+09	**4**	1.31E+10	**4**	**4.14**
E	22099273	**6**	1.07E+08	**4**	44139231	**7**	13920117	**6**	3.93E+08	**5**	1.85E+09	**5**	1.4E+10	**6**	**5.5**

Contest Results – Summary

Team	Main track ranking	Special honor ranking
A	1	3
B	2	1
C	3	2
D	4	4
E	5	5

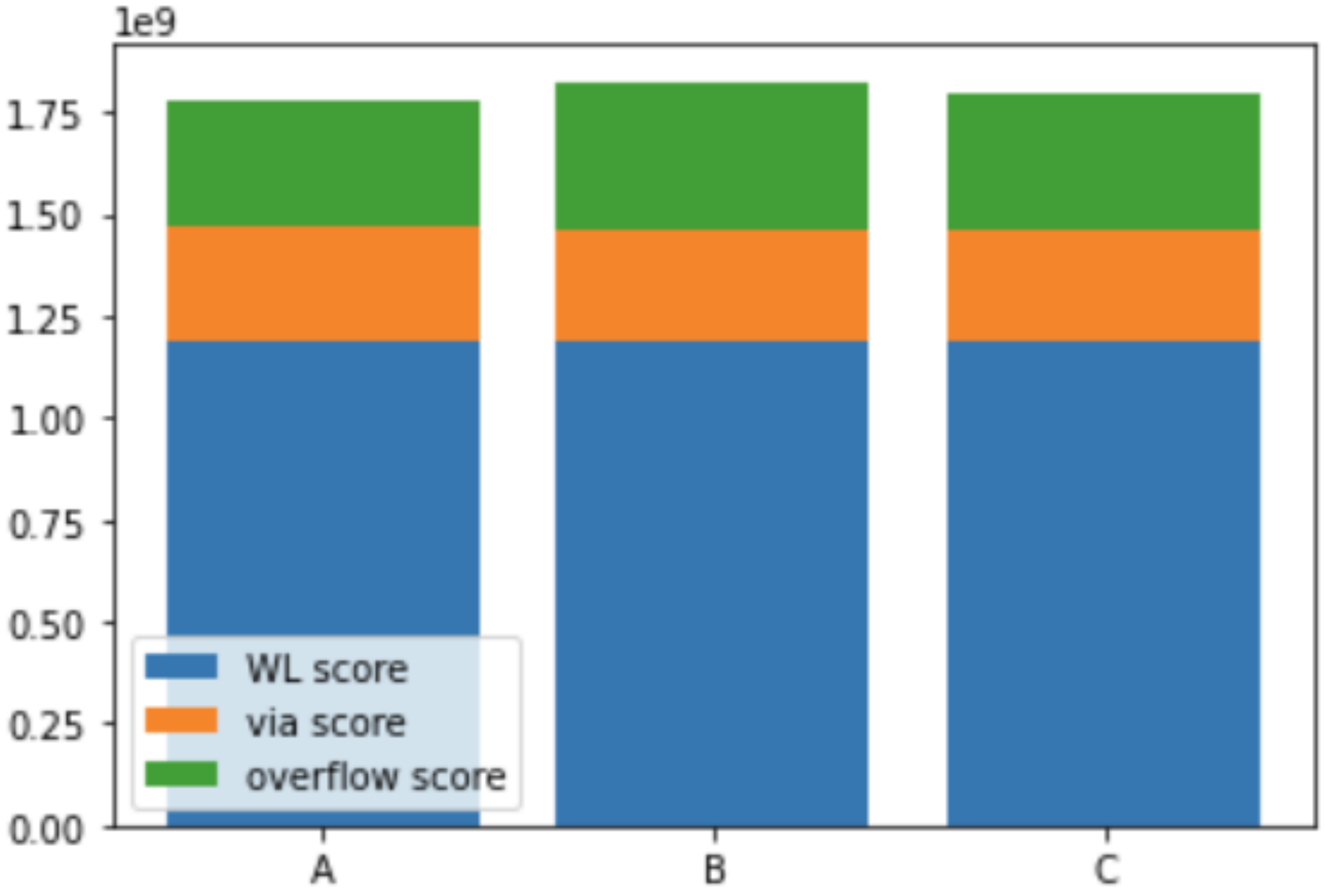
Results Analysis – GPU Usage

Team	Use GPU?
A	Y
B	Y
C	N

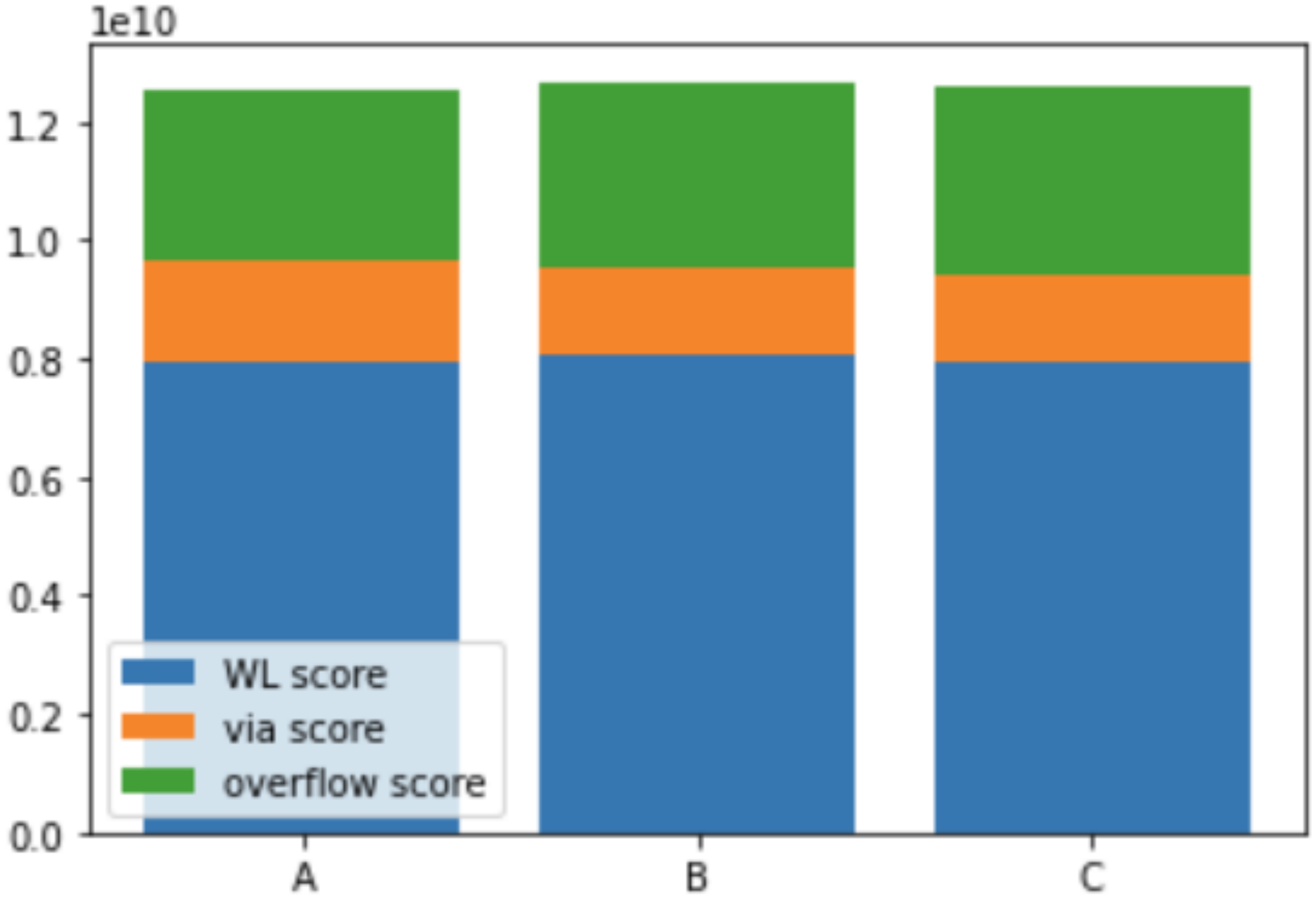
Results Analysis - Deterministic

- Among 18 teams in the final submission, 14 teams have deterministic solutions
- Among top-3 teams in the final submission, all of them have deterministic solutions

Contest Results – Score Breakdown (Main Track)



Mempool_cluster

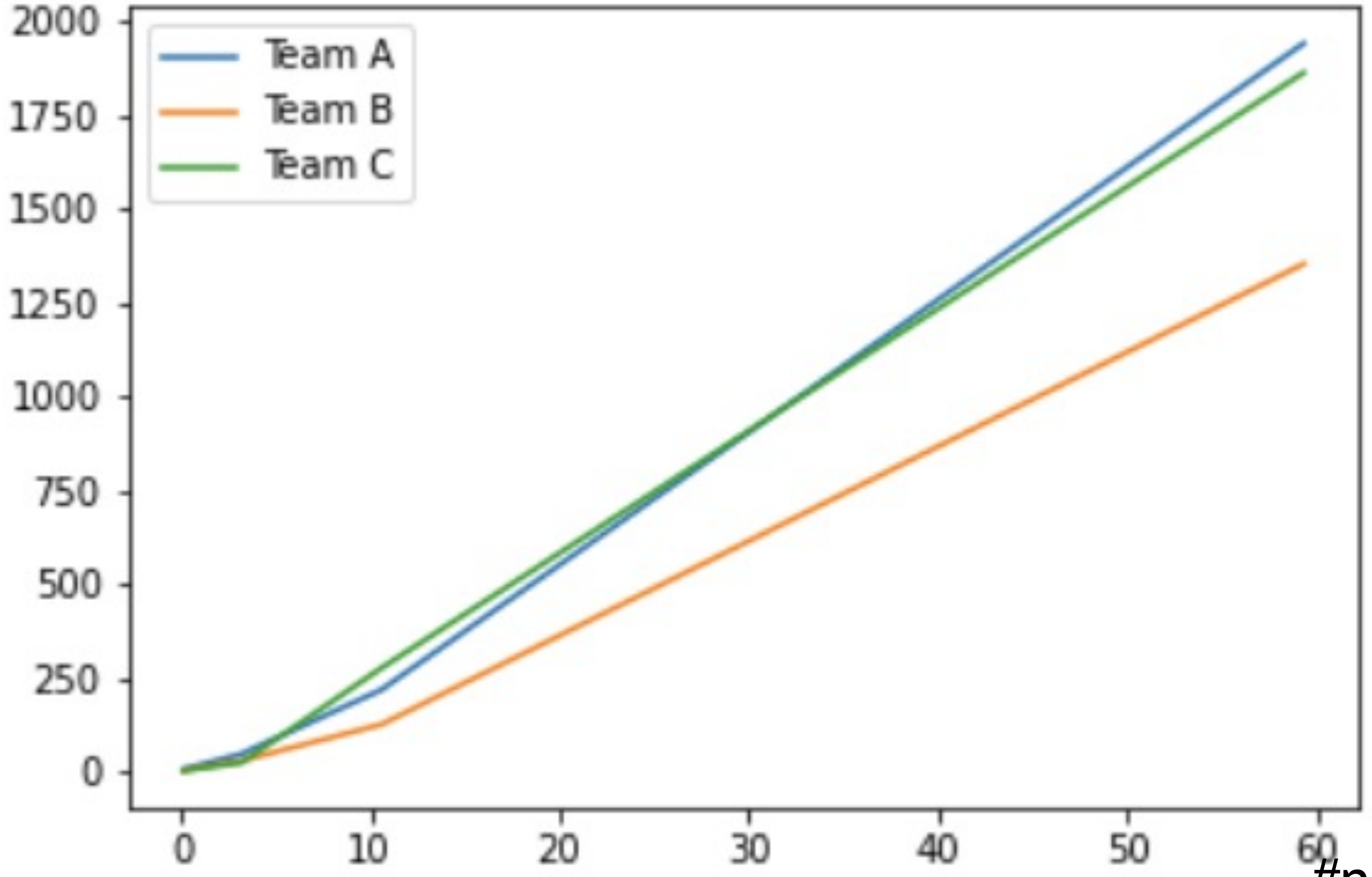


Tera_cluster

WL score > overflow score > via score
QoR performance: Team A > Team C > Team B

Results Analysis – Runtime Scalability (Main Track)

Runtime/s

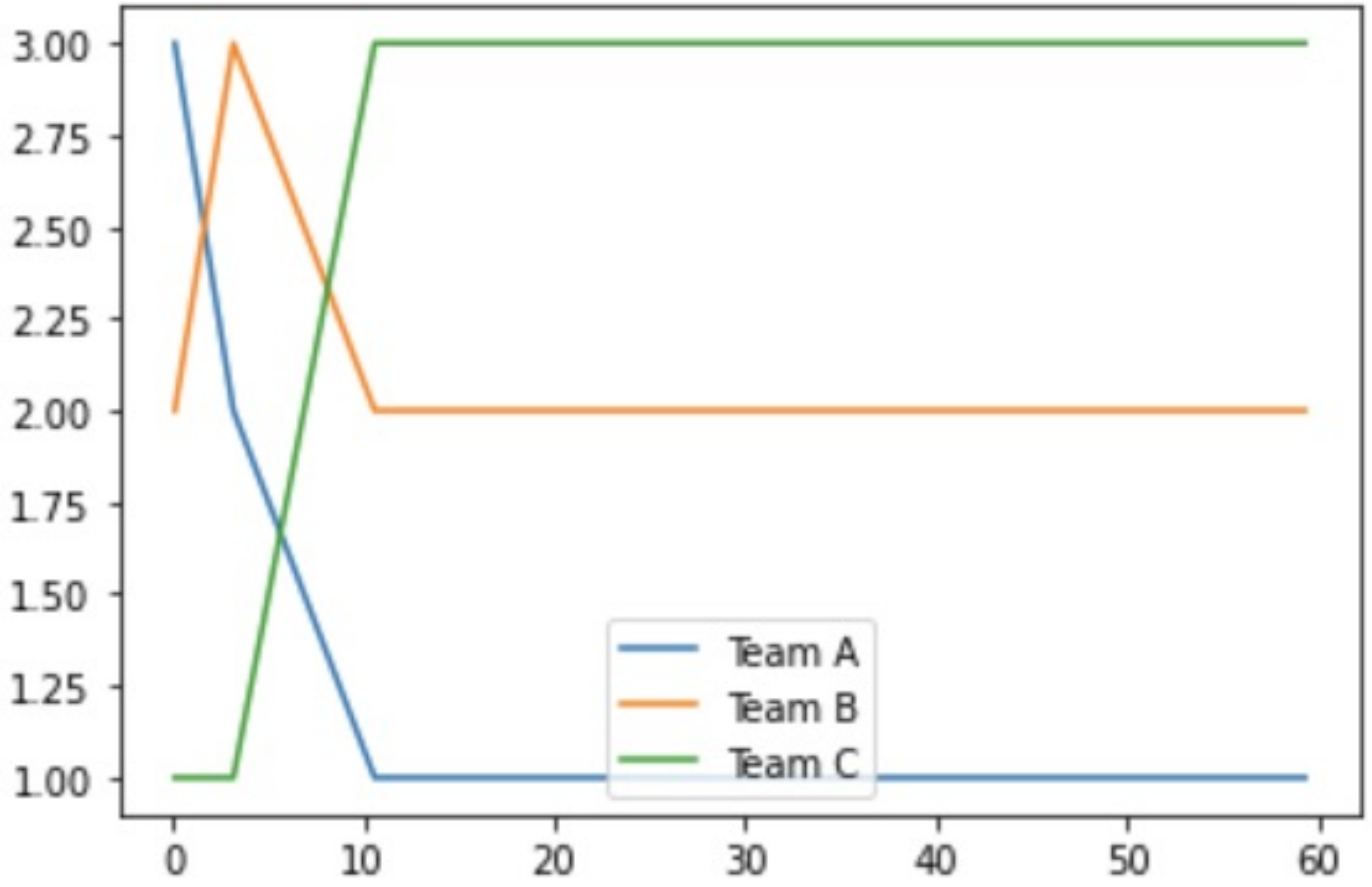


#nets (Million)

Runtime performance: Team B > Team C ~= Team A

Ranking vs. #Nets (Main Track)

ranking



#nets (Million)

The larger the number of nets, the better Team A performs

2024 ACM International Symposium on Physical Design

GPU/ML-Enhanced Large Scale Global Routing

HONORABLE MENTION

Team etuoReL

Xingyu Tong, Guohao Chen, Benchao Zhu, Jiawei Li, Jiaming Chang,
Yuzheng Lin, Yuhao Ren, Chang Liu
Fudan University

Iris Hui-Ru Jiang
General Chair
ISPD 2024

Gracieli Posser
Technical Program Chair
ISPD 2024

Rongjian Liang, Anthony Agnesina,
Wen-Hao Liu, Haoxing Ren
NVIDIA

2024 ACM International Symposium on Physical Design

GPU/ML-Enhanced Large Scale Global Routing

HONORABLE MENTION

Team SCAW

Lang Feng, Hongxin Kong, Xupengkai Lu, Xihao Liang, Junxi Feng, Xiaokun Lin, Jixiang Zhu, Junhao Guo, Wenchao Qian, Yujie Wang
Sun Yat-sen University,
Institute of Computing Technology, Chinese Academy of Sciences,
and Advanced Micro Devices, Inc.

Iris Hui-Ru Jiang
General Chair
ISPD 2024

Gracieli Posser
Technical Program Chair
ISPD 2024

Rongjian Liang, Anthony Agnesina,
Wen-Hao Liu, Haoxing Ren
NVIDIA

Team SCAW Members

- Students:


Xupengkai Lu


Xihao Liang

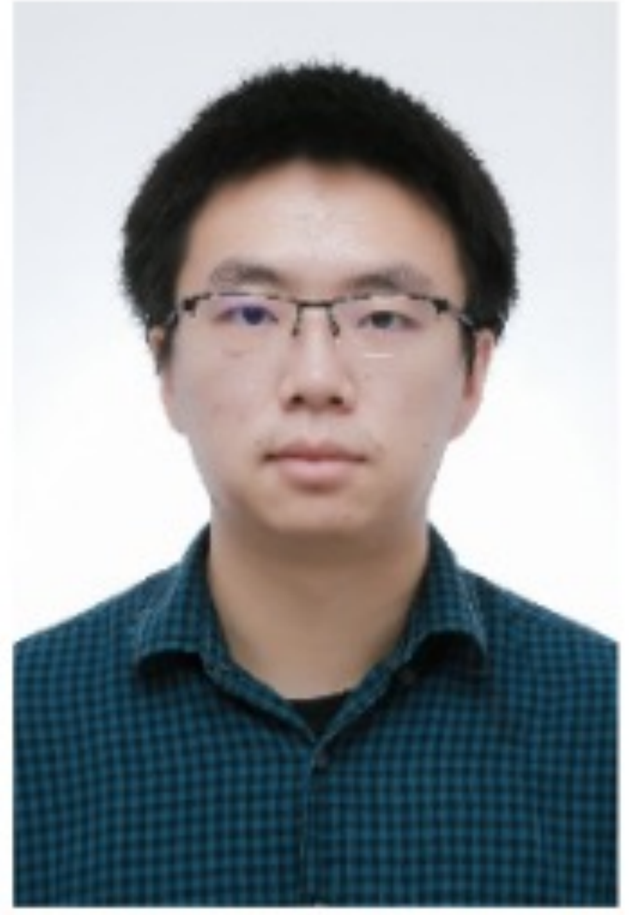

Junxi Feng


Xiaokun Lin


Jixiang Zhu


Junhao Guo

- Advisors:


Lang Feng


Yujie Wang

Team SCAW Methodology

- The opportunities for parallel computing of multiple phases are investigated, and the best configuration combination is selected

2024 ACM International Symposium on Physical Design

GPU/ML-Enhanced Large Scale Global Routing

THIRD PLACE

Team metaRoute

Zhijie Cai, Min Wei, Yilu Chen, Zhaoyi Wu, Hongzhi Ding, Peng Zou, Zhifeng Lin, Jianli Chen
Fudan University, Fuzhou University and Shanghai LEDA technology

Iris Hui-Ru Jiang
General Chair
ISPD 2024

Gracieli Posser
Technical Program Chair
ISPD 2024

Rongjian Liang, Anthony Agnesina, Wen-Hao Liu, Haoxing Ren
NVIDIA

Team metaRoute members

- Zhijie Cai
  - Graduate Student, Fudan University
- Min Wei
  - PhD Student, Fudan University
- Yilu Chen
  - PhD Student, Fuzhou University
- Zhaoyi Wu
  - Graduate Student, Fuzhou University

Advisors

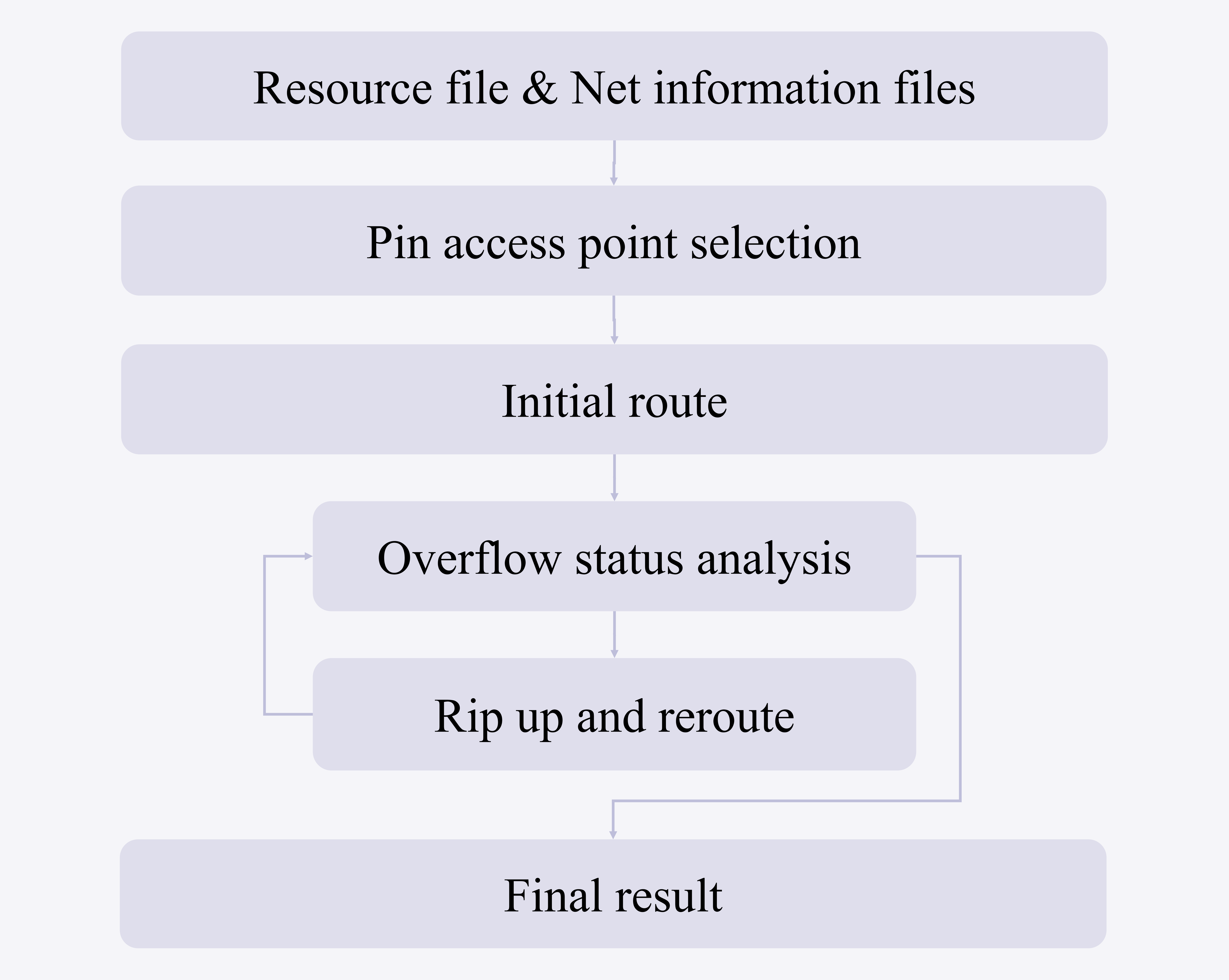
- Hongzhi Ding
  - Shanghai LEDA technology
- Peng Zou
  - Shanghai LEDA technology
- Zhifeng Lin
  - Professor, Fuzhou University
- Jianli Chen
  - Professor, Fudan University

Team metaRoute Methodology

Overview of our flow

- Pin access point selection with alignment preference
- Initial route by 2D pattern route and layer assignment
- Analysis current overflow situation to determine next step
- Rip up the problematic nets and reroute them in a parallel friendly mode

Algorithm Flow



2024 ACM International Symposium on Physical Design

GPU/ML-Enhanced Large Scale Global Routing

SECOND PLACE

Team RL-Route

Shiju Lin, Liang Xiao, Jinwei Liu, Evangeline F.Y. Young
The Chinese University of Hong Kong

Iris Hui-Ru Jiang
General Chair
ISPD 2024

Gracieli Posser
Technical Program Chair
ISPD 2024

Rongjian Liang, Anthony Agnesina,
Wen-Hao Liu, Haoxing Ren
NVIDIA

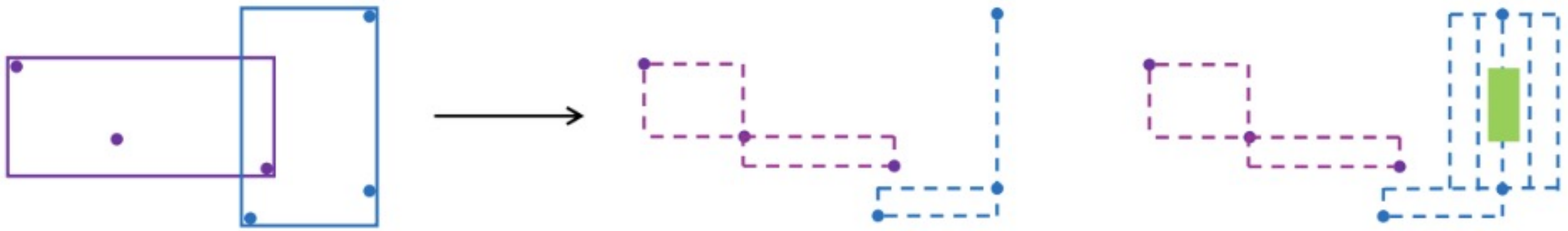
Team RL-Route Methodology

Fine-Grained Massive Net-Level Parallelism

- Net-level parallelism
  - Parallel routing of a batch of “non-overlapping” nets
- Overlap checking
  - **Previous** method based on bounding boxes (BB) (**pessimistic**)
  - **New** method based on fine-grained routing patterns (**accurate**)

Fine-Grained Massive Net-Level Parallelism

- Use CUGR2 as an example to show its impact in each stage
  - 3 stages: L-shape routing → augmented graph routing → sparse graph maze routing



Previous method (BB)
overlap (pessimistic)

RSMT-based L-shape graph
no overlap (accurate)

Augmented routing graph
(to escape the congested green area)
no overlap (accurate)

2024 ACM International Symposium on Physical Design

GPU/ML-Enhanced Large Scale Global Routing

FIRST PLACE

Team Hippo

Chunyuan Zhao, Yibo Lin
Peking University

Iris Hui-Ru Jiang
General Chair
ISPD 2024

Gracieli Posser
Technical Program Chair
ISPD 2024

Rongjian Liang, Anthony Agnesina,
Wen-Hao Liu, Haoxing Ren
NVIDIA

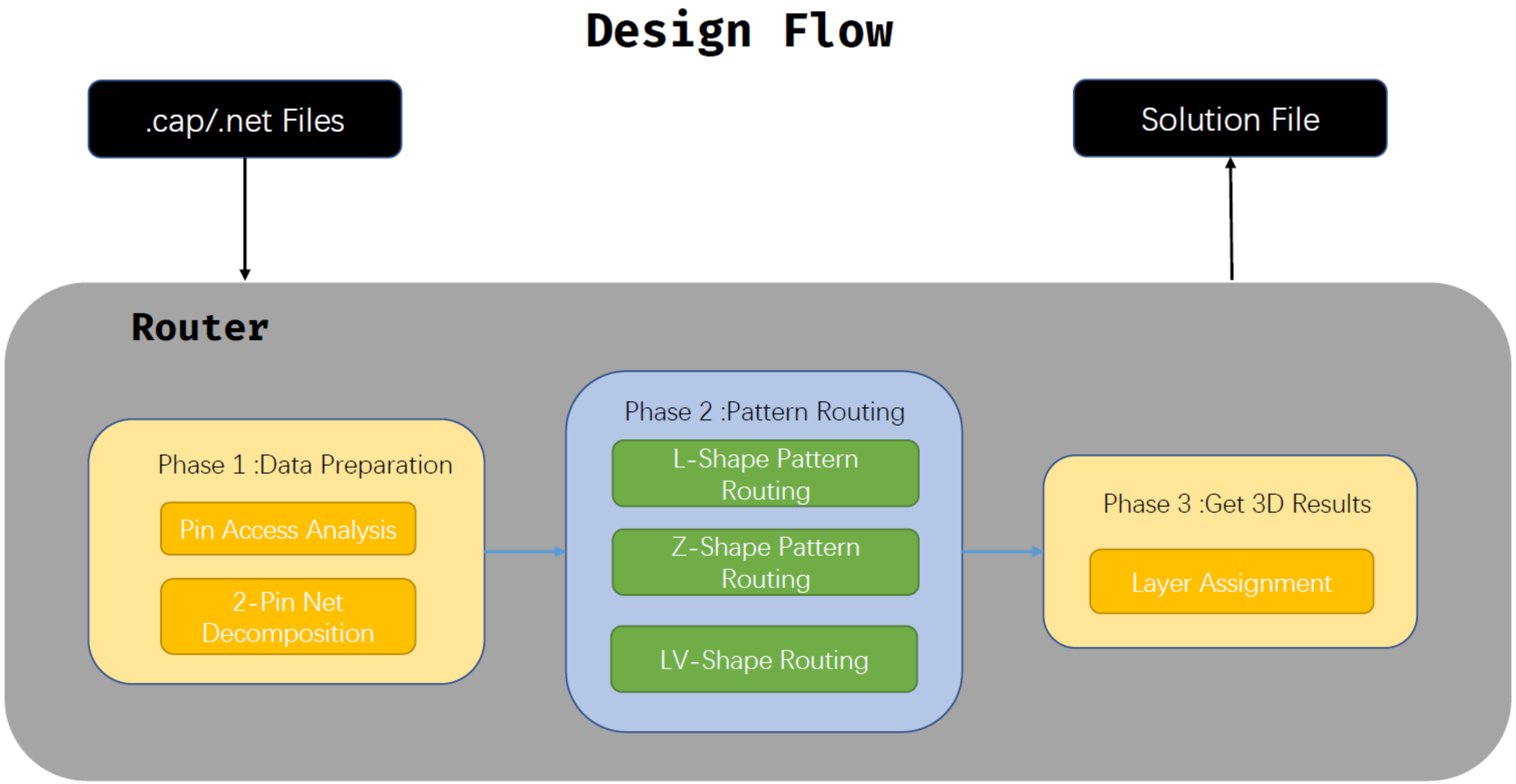
Team Hippo

Team Member

Chunyuan Zhao
4th Year Undergraduate Student
Peking University

Advisor : Prof.Yibo Lin
Peking University

Team Hippo Methodology



2024 ACM International Symposium on Physical Design

GPU/ML-Enhanced
Large Scale Global Routing

HONORABLE MENTION - SPECIAL HONOR TRACK

Team etuoReL

Xingyu Tong, Guohao Chen, Benchao Zhu, Jiawei Li, Jiaming Chang, Yuzheng Lin, Yuhao Ren, Chang Liu
Fudan University

Iris Hui-Ru Jiang
General Chair
ISPD 2024

Gracieli Posser
Technical Program Chair
ISPD 2024

Rongjian Liang, Anthony Agnesina, Wen-Hao Liu, Haoxing Ren
NVIDIA

2024 ACM International Symposium on Physical Design

GPU/ML-Enhanced
Large Scale Global Routing

HONORABLE MENTION
- SPECIAL HONOR TRACK

Team SCAW

Lang Feng, Hongxin Kong, Xupengkai Lu, Xihao Liang, Junxi Feng,
Xiaokun Lin, Jixiang Zhu, Junhao Guo, Wenchao Qian, Yujie Wang
Sun Yat-sen University,
Institute of Computing Technology, Chinese Academy of Sciences,
and Advanced Micro Devices, Inc.

Iris Hui-Ru Jiang General Chair ISPD 2024	Gracieli Posser Technical Program Chair ISPD 2024	Rongjian Liang, Anthony Agnesina, Wen-Hao Liu, Haoxing Ren NVIDIA

2024 ACM International Symposium on Physical Design

GPU/ML-Enhanced Large Scale Global Routing

THIRD PLACE - SPECIAL HONOR TRACK

Team Hippo

Chunyuan Zhao, Yibo Lin
Peking University

Iris Hui-Ru Jiang
General Chair
ISPD 2024

Gracieli Posser
Technical Program Chair
ISPD 2024

Rongjian Liang, Anthony Agnesina,
Wen-Hao Liu, Haoxing Ren
NVIDIA

2024 ACM International Symposium on Physical Design

GPU/ML-Enhanced Large Scale Global Routing

SECOND PLACE - SPECIAL HONOR TRACK

Team metaRoute

Zhijie Cai, Min Wei, Yilu Chen, Zhaoyi Wu, Hongzhi Ding, Peng Zou, Zhifeng Lin, Jianli Chen
Fudan University, Fuzhou University and Shanghai LEDA technology

Iris Hui-Ru Jiang
General Chair
ISPD 2024

Gracieli Posser
Technical Program Chair
ISPD 2024

Rongjian Liang, Anthony Agnesina, Wen-Hao Liu, Haoxing Ren
NVIDIA

2024 ACM International Symposium on Physical Design

GPU/ML-Enhanced
Large Scale Global Routing

FIRST PLACE
- SPECIAL HONOR TRACK

Team RL-Route

Shiju Lin, Liang Xiao, Jinwei Liu, Evangeline F.Y. Young
The Chinese University of Hong Kong

Iris Hui-Ru Jiang
General Chair
ISPD 2024

Gracieli Posser
Technical Program Chair
ISPD 2024

Rongjian Liang, Anthony Agnesina,
Wen-Hao Liu, Haoxing Ren
NVIDIA

Acknowledgment

- David Chinnery - Siemens Digital
- Gracieli Posser – Cadence
- Andrew Kahng – University of California at San Diego, OpenROAD
- Matt Liberty – OpenROAD
- Evangeline F.Y. Young – The Chinese University of Hong Kong
- Fangzhou Wang - The Chinese University of Hong Kong
- Yangqing Zhang – NVIDIA
- Haoyu Yang – NVIDIA
- Mingjie Liu - NVIDIA

NVIDIA