2024 ICCAD CAD Contest Problem C: Scalable Logic Gate Sizing using ML Techniques and GPU Acceleration

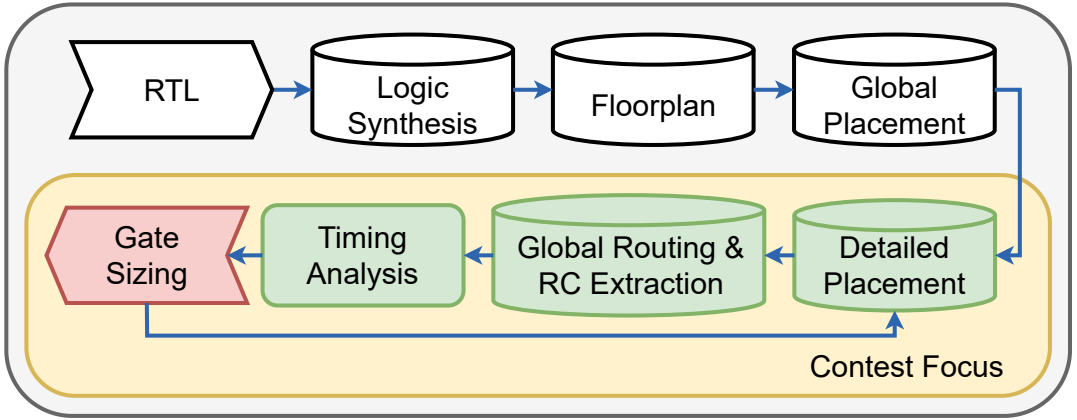
Bing-Yue Wu, Rongjian Liang, Geraldo Pradipta,
Anthony Agnesina, Haoxing Ren, and Vidya A. Chhabria

Outline

- **Problem Description**
- Evaluation Methodology
- Benchmarks
- Results
- Winners

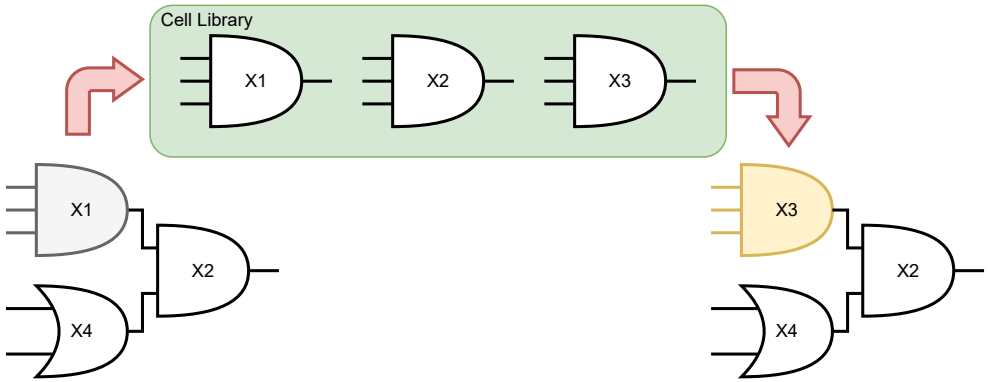
Contest Problem Description

- Timing optimization is important for power, performance, and area (PPA) optimization.
- Modern timing optimization methods are heuristic-based and do not guarantee an optimal solution.
- The NP-hard logic gate sizing problem is an integral part of timing optimization.



Gate Sizing Challenges

- Select a size for every netlist instance from the standard cell library.
- Each size has a different delay, area, and power value.
- **Challenges:**
  - **Non-convexity characteristics of the delay models**
  - **Discrete space of gate sizes**
  - **Large number of near-critical paths**
  - **Tradeoff between power and timing**



Gate Sizing Problem Formulation

- Use AI/GPU-accelerated methods to speed up the gate sizing process.
- **Contest Focus:**

  Objective:

  - $\min \sum_{i \in \mathrm{I}} \mathrm{LeakagePower}(c_i)$

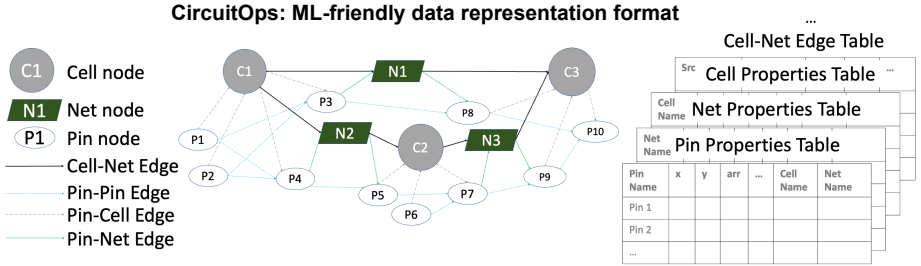
  Subject to:

  - $\mathrm{slack}(\mathrm{pin}_i) \geq 0, \forall i \in \mathrm{I}$
  - $\mathrm{slew}(\mathrm{pin}_i) \leq \mathrm{MaximumPermittedSlew}(\mathrm{pin}_i), \forall i \in \mathrm{I}$
  - $\mathrm{load}(\mathrm{pin}_i) \leq \mathrm{MaximumPermittedLoad}(\mathrm{pin}_i), \forall i \in \mathrm{I}$
  - $c_i \in \mathrm{ChoiceInLibrary}_i, \forall i \in \mathrm{I}$

  where I is a set of all instances in the design

Comparison Against Previous Sizing Contests

- ISPD' 2012 and ISPD' 2013
- Provide benchmarks in standard EDA file format and in CircuitOps format

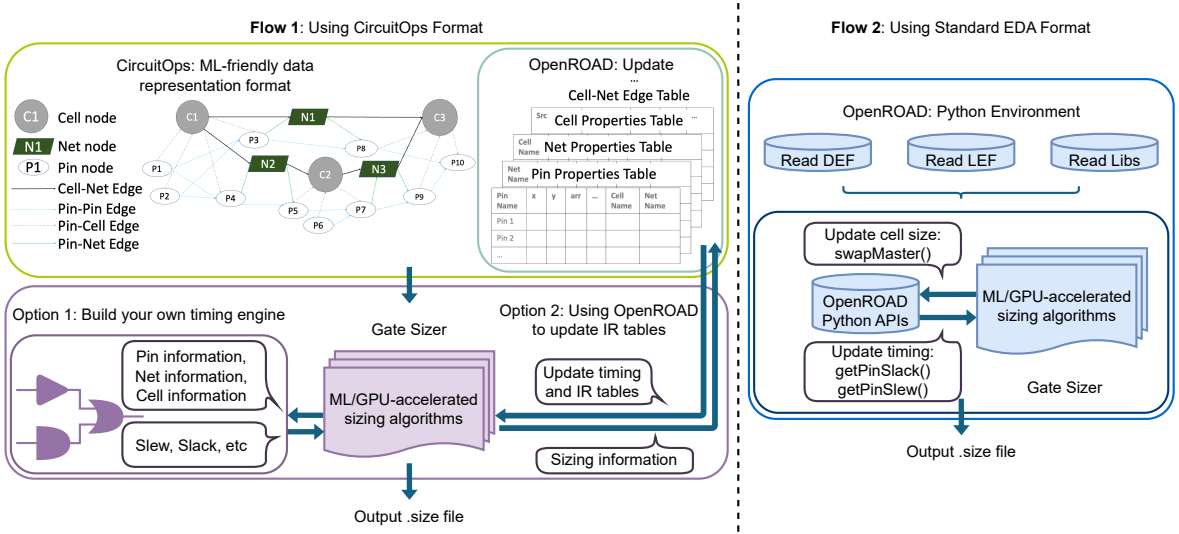


- Consider legalization and global routing in evaluation

Design in the Contest Benchmark Captured using OpenROAD

Contest Flow

- **Flow 1**: Get design information using the CircuitOps format, then perform gate sizing.
- **Flow 2**: Get design information from the standard EDA files, then perform gate sizing.
- Example scripts of OpenROAD Python APIs are provided.
  - Swap the gate, get timing information, etc.



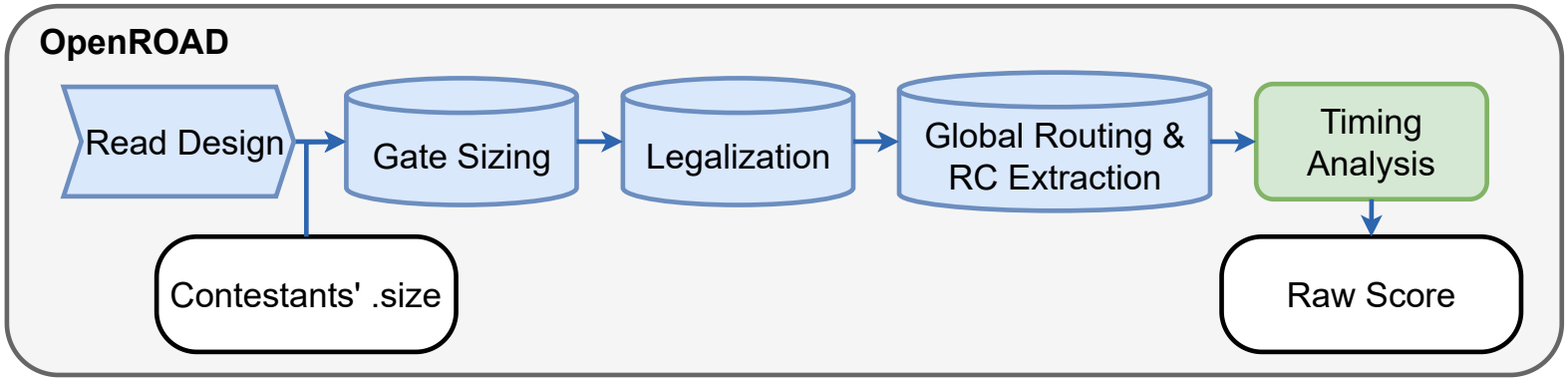
Contributions of This Contest

- **Lower the barrier to entry** for non-EDA experts by turning EDA problems into ML-solvable ones with an **ML-friendly data format**.
- Show an example of “**ML inside**" EDA tools using **OpenROAD Python APIs**.
- Release updated benchmarks in **ASAP7 FinFET technology node** with evaluation scripts to drive logic gate sizing research.
- Encourage the creation of ML/GPU-based gate sizer that account for the impact on the overall physical design flow, including **legalization** and **routability**.

Outline

- Problem Description
- **Evaluation Methodology**
- Benchmarks
- Results
- Winners

Evaluation Flow



Evaluation Score

Convert the constrained optimization problem into a score function using a penalty approach.

$$\text{RawScore} = (\lambda + \text{RunTimeF}) \times [(\text{Leakage}_{\text{tot}} - \text{MinLeakage}_{\text{tot}})$$

$$+\alpha \times |\text{TNS}| \quad \rightarrow \text{slack penalty}$$

$$+ \sum_i \beta_i \times \left(\text{slew}(\text{pin}_i) - \text{MaximumPermittedSlew}(\text{pin}_i)\right) \quad \rightarrow \text{slew penalty}$$

$$+ \sum_i \gamma_i \times \left(\text{load}(\text{pin}_i) - \text{MaximumPermittedLoad}(\text{pin}_i)\right)] \quad \rightarrow \text{load penalty}$$

$$\text{RuntTimeF} = \max\left\{log_2\left(\frac{\text{RunTime}}{\text{MedianRunTime}}\right), \delta\right\}$$

$$\textbf{finalScore} = 100 \times \frac{\textbf{Least RawScore across teams}}{\textbf{RawScore of the team}}$$

Evaluation Platform:
4 NVIDIA A100 GPUs and 8 CPU threads

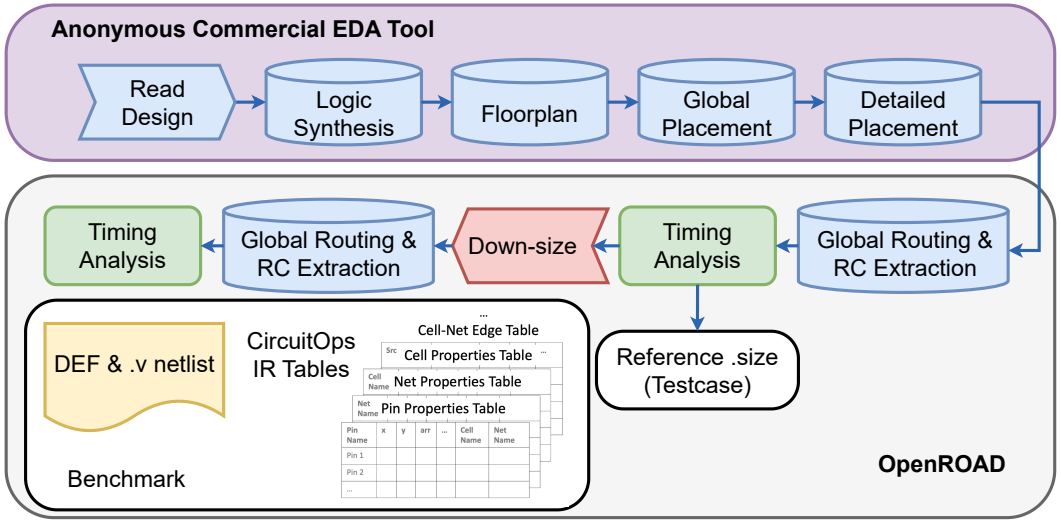
$\alpha, \beta, \gamma, \delta, \lambda$ are empirically assigned after alpha submission

Outline

- Problem Description
- Evaluation Methodology
- **Benchmarks**
- Results
- Winners

Benchmark Generation

- Eight designs from the TILOS MacroPlacement GitHub and two from OpenCores.
  - Down-size all instances to the smallest size available.
  - Provide a reference .size file for ML label annotation.



"TILOS MacroPlacement." https://github.com/TILOS-AI-Institute/ MacroPlacement, 2024.

Benchmark Statistics

$$\text{RawScore} = (\lambda + \text{RunTimeF}) \times ((\text{Leakage}_{\text{tot}} - \text{MinLeakage}_{\text{tot}}) + \alpha \times |\text{TNS}| + \sum_i \beta_i \times (\text{slew}(\text{pin}_i) - \text{MaximumPermittedSlew}(\text{pin}_i)) + \sum_i \gamma_i \times (\text{load}(\text{pin}_i) - \text{MaximumPermittedLoad}(\text{pin}_i)))$$

$$\text{RuntTimeF} = \max\left\{log_2\left(\frac{\text{RunTime}}{\text{MedianRunTime}}\right), \delta\right\}$$

MacroPlacment GitHib
OpenCores

Design	Gate Count	WNS (ns)	TNS (ns)	Total Slew Violation Difference (ns)	Total Load Capacitance Violation Difference (fF)	Total Leakage (μW)
NV_NVDLA_partition_m	27,553	-0.595	-156.323	258.761	256	1.672
NV_NVDLA_partition_p	79,919	-1.519	-6,306.64	6,125.512	5,292	5.539
ariane136	145,776	-1.298	-10,143.711	14,843.895	15,463	17,539.095
mempool_tile_wrap	187,851	-1.315	-10,458.099	12,069.07	10,779	2,590.189
aes_256	278,465	-0.284	-212.965	942.81	1,300	16.771
hidden1	38,089	-1.069	-1,136.054	4,073.071	6,811	2.762
hidden2	149,396	-1.214	-9,400.457	16,582.889	16,661	17,152.379
hidden3	184,863	-1.563	-2,436.288	19,755.937	33,088	16,513.594
hidden4	260,483	-3.185	-25,334.022	19,138.199	27,548	21.024
hidden5	283,750	-0.324	-370.293	3.483	NA	16.17

Benchmark Statistics

- Sized with provided reference .size file:

Design	Gate Count	WNS (ns)	TNS (ns)	Total Slew Violation Difference (ns)	Total Load Capacitance Violation Difference (fF)	Total Leakage (μW)
NV_NVDLA_partition_m	27,553	-0.207	-10.266	NA	NA	2.693
NV_NVDLA_partition_p	79,919	-0.126	-17.899	0.074	NA	6.635
ariane136	145,776	-0.214	-27.613	23.713	NA	17,545.15
mempool_tile_wrap	187,851	-0.191	-2.889	40.511	41	2,594.179
aes_256	278,465	NA	NA	NA	NA	16.918
hidden1	38,089	-0.217	-20.302	5.394	NA	2.874
hidden2	149,396	-0.233	-162.019	19.842	NA	17,156.382
hidden3	184,863	-0.319	-4.742	666.407	NA	16,514.495
hidden4	260,483	-0.291	-104.095	410.963	NA	21.904
hidden5	283,750	NA	NA	NA	NA	26.831

Outline

- Problem Description
- Evaluation Methodology
- Benchmarks
- **Results**
- Winners

Submission Statistics

- Number of teams submitted **alpha** submission = 19 teams
- Number of teams submitted **beta** submission = 20 teams
- Number of teams submitted **final** submission = 24 teams

Results Overview

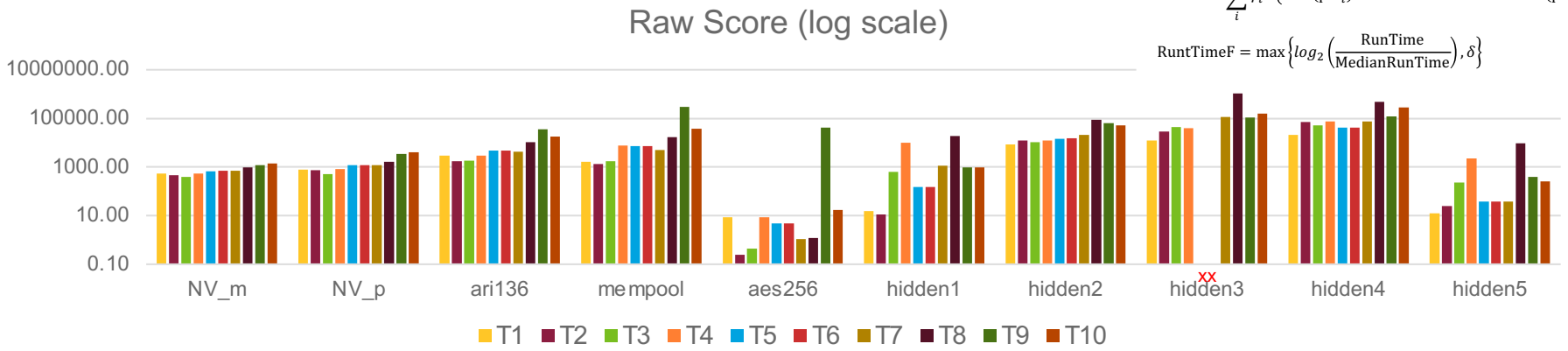
- 16 out of the 24 teams have valid results on at least one case

Team	NV_m	NV_p	ari136	mempool	aes256	hidden1	hidden2	hidden3	hidden4	hidden5
T1	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
T2	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
T3	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
T4	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
T5	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Failed	Pass	Pass
T6	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Failed	Pass	Pass
T7	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
T8	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
T9	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
T10	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
T11	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
T12	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
T13	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
T14	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
T15	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Failed	Pass
T16	Pass	Failed	Failed	Failed	Failed	Pass	Failed	Failed	Failed	Failed

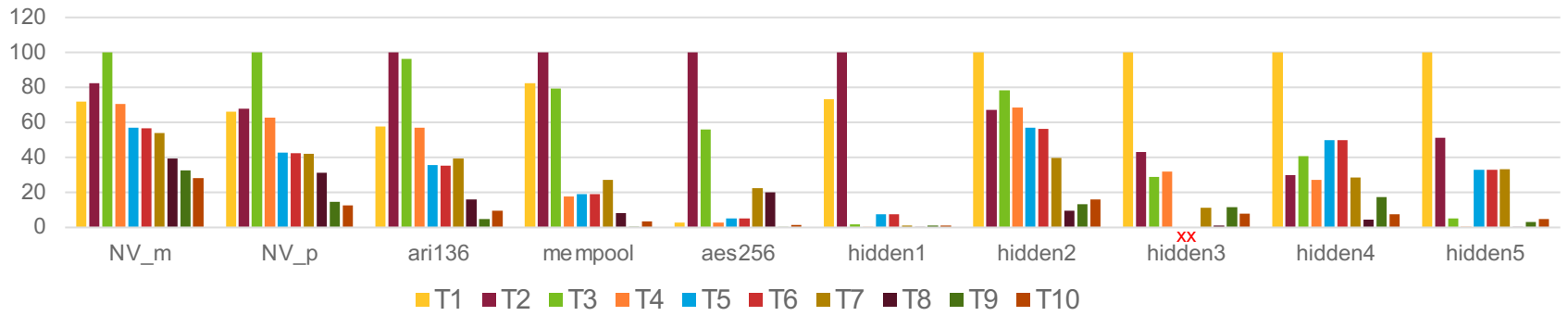
Result Overview: Score Distribution

$$\text{RawScore} = (\lambda + \text{RunTimeF}) \times ((\text{Leakage}_{\text{tot}} - \text{MinLeakage}_{\text{tot}}) + \alpha \times |\text{TNS}| + \sum_i \beta_i \times (\text{slew}(\text{pin}_i) - \text{MaximumPermittedSlew}(\text{pin}_i)) + \sum_i \gamma_i \times (\text{load}(\text{pin}_i) - \text{MaximumPermittedLoad}(\text{pin}_i)))$$

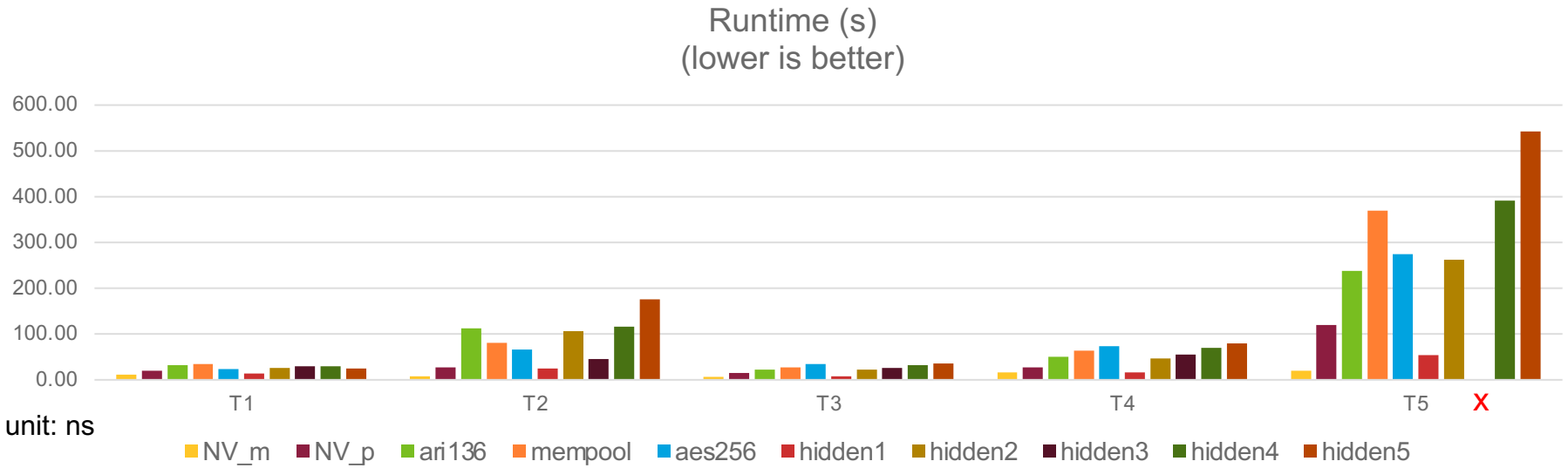
$$\text{RuntTimeF} = \max\left\{log_2\left(\frac{\text{RunTime}}{\text{MedianRunTime}}\right), \delta\right\}$$



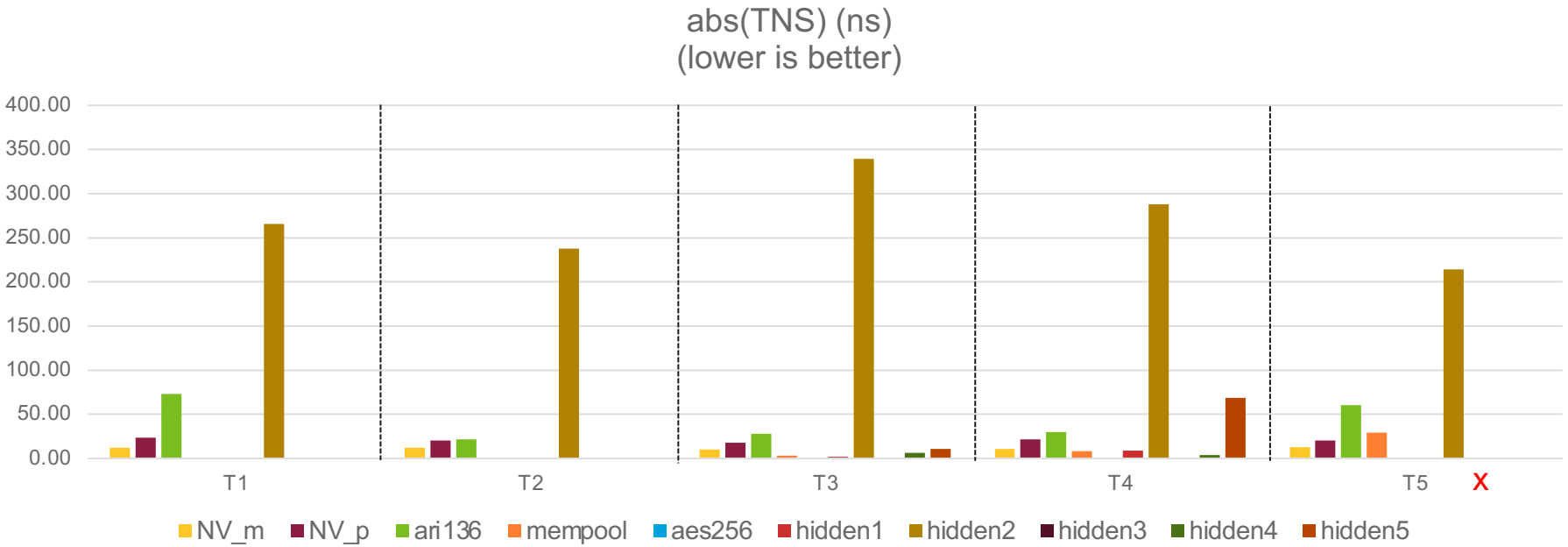
$$\textbf{finalScore} = 100 \times \frac{\textbf{Least RawScore across teams}}{\textbf{RawScore of the team}}$$



Runtime of the Top 5 Teams



TNS of the Top 5 Teams

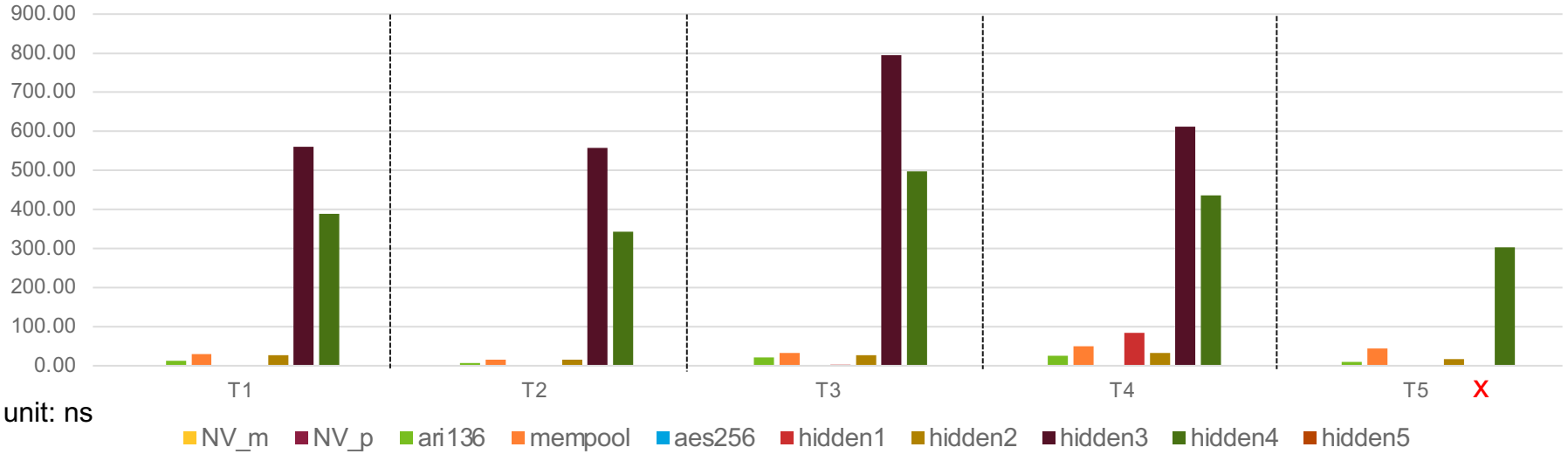


Slew Violation of the Top 5 Teams

Total Slew Violation Difference

$$\sum_i \beta_i \times \left(\text{slew}(\text{pin}_i) - \text{MaximumPermittedSlew}(\text{pin}_i)\right) \qquad \beta_i = \begin{cases} 0 & \text{if } \text{slew}(\text{pin}_i) \le \text{MaximumPermittedSlew}(\text{pin}_i), \\ 20 & \text{otherwise,} \end{cases}$$

(lower is better)



unit: ns

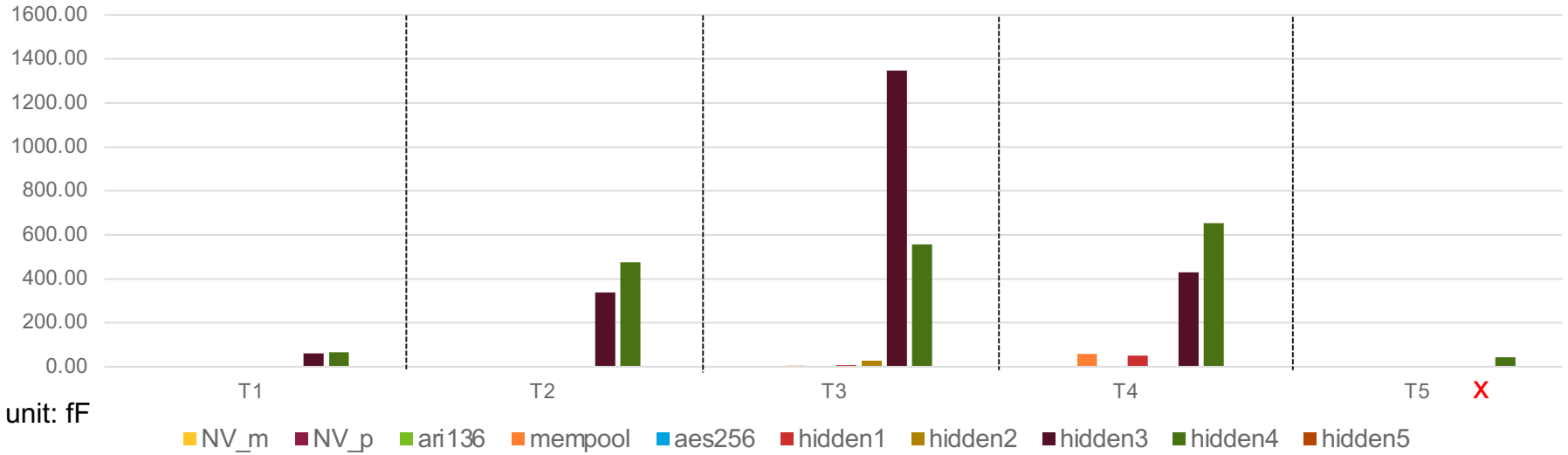
Load Violation of the Top 5 Teams

Total Load Violation Difference

$$\sum_i \gamma_i \times \left(\text{load}(\text{pin}_i) - \text{MaximumPermittedLoad}(\text{pin}_i)\right)$$

$$\gamma_i = \begin{cases} 0 & \text{if } \text{load}(\text{pin}_i) \le \text{MaximumPermittedLoad}(\text{pin}_i), \\ 20 & \text{otherwise,} \end{cases}$$

(lower is better)



Outline

- Problem Description
- Evaluation Methodology
- Benchmarks
- Results
- **Winners**

Honorable Mention – Congratulations!

Problem C		Contributing authors	School
cadc0013	Advisors	Ting-Chi Wang	National Tsing Hua University
	Students	Wen-Xuan Chen	National Tsing Hua University

Problem C		Contributing authors	School
cadc1031	Advisors	Jie-Hong Roland Jiang	National Taiwan University
	Students	Hsin-Ying Tsai	National Taiwan University
		Shao-Jui Wu	National Taiwan University
		Mu-Yao Chung	National Taiwan University
		Tian-Fu Chen	National Taiwan University
		Jiun-Hao Chen	National Taiwan University
		Yu-Hung Pan	National Taiwan University

Third Place – Congratulations!

Problem C		Contributing authors	School
cadc1021	Advisors	Peng Cao	Southeast University
	Students	Yuhan Dong	Southeast University
		Zeyuan Deng	Southeast University
		Yusen Qin	Southeast University
		Xu Cheng	Southeast University
		Junming Jiao	Southeast University
		Zhanhua Zhang	Southeast University

Second Place – Congratulations!

Problem C		Contributing authors	School
cadc1026	Advisors	Evangeline F.Y. Young	The Chinese University of Hong Kong
	Students	Qijing Wang	The Chinese University of Hong Kong
		Tianji Liu	The Chinese University of Hong Kong
		Bangqi Fu	The Chinese University of Hong Kong
		Zhenxuan Xie	The Chinese University of Hong Kong

First Place – Congratulations!

Problem C		Contributing authors	School
cadc1015	Advisors	Yibo Lin	Peking University
	Students	Yufan Du	Peking University
		Zizheng Guo	Peking University

2024 ICCAD CAD Contest Problem C: Scalable Logic Gate Sizing using ML Techniques and GPU Acceleration

Bing-Yue Wu, Rongjian Liang, Geraldo Pradipta,
Anthony Agnesina, Haoxing Ren, and Vidya A. Chhabria