GOALPlace: Begin with the End in Mind

Anthony Agnesina, Rongjian Liang, Geraldo Pradipta, Anand Rajaram, Haoxing Ren

Anthony Agnesina, Sr. VLSI Methodology Engineer | ISPD 2025

Agenda

- Motivations
- Cell Density
- James-Stein Empirical Bayes
- Integration with AutoDMP
- Experiments
- Conclusion & Future Work

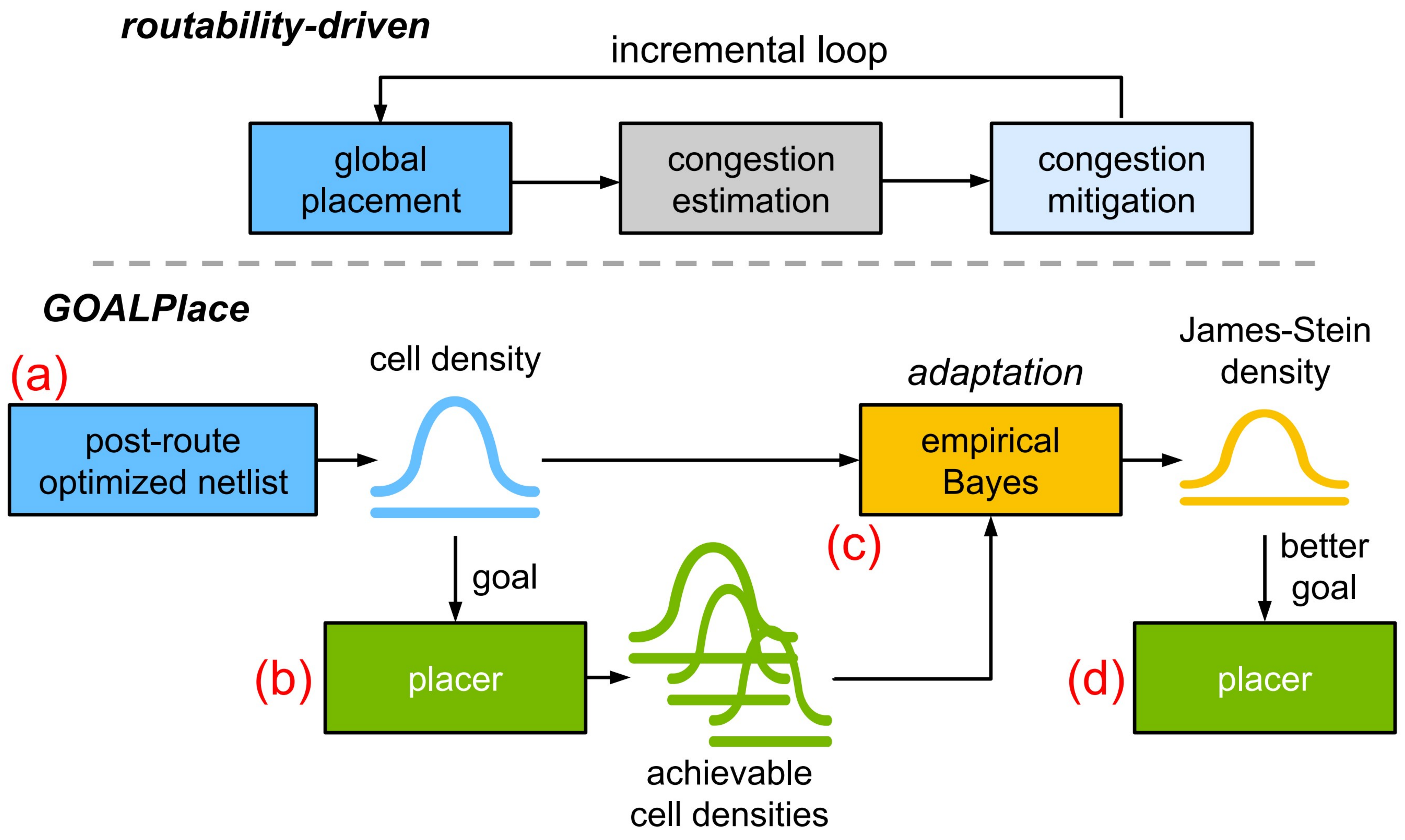
Motivations

A learning-based approach for placement

- Placement significantly impacts routability (and timing)
- Typical routability-driven methods weaknesses
  - Incremental
  - Router-tied

- **Observations**
  - Most visible route/timing opt. layout change is **cell density**
  - Only **post-route** netlist matters and is a blueprint of all optimizations from the flow

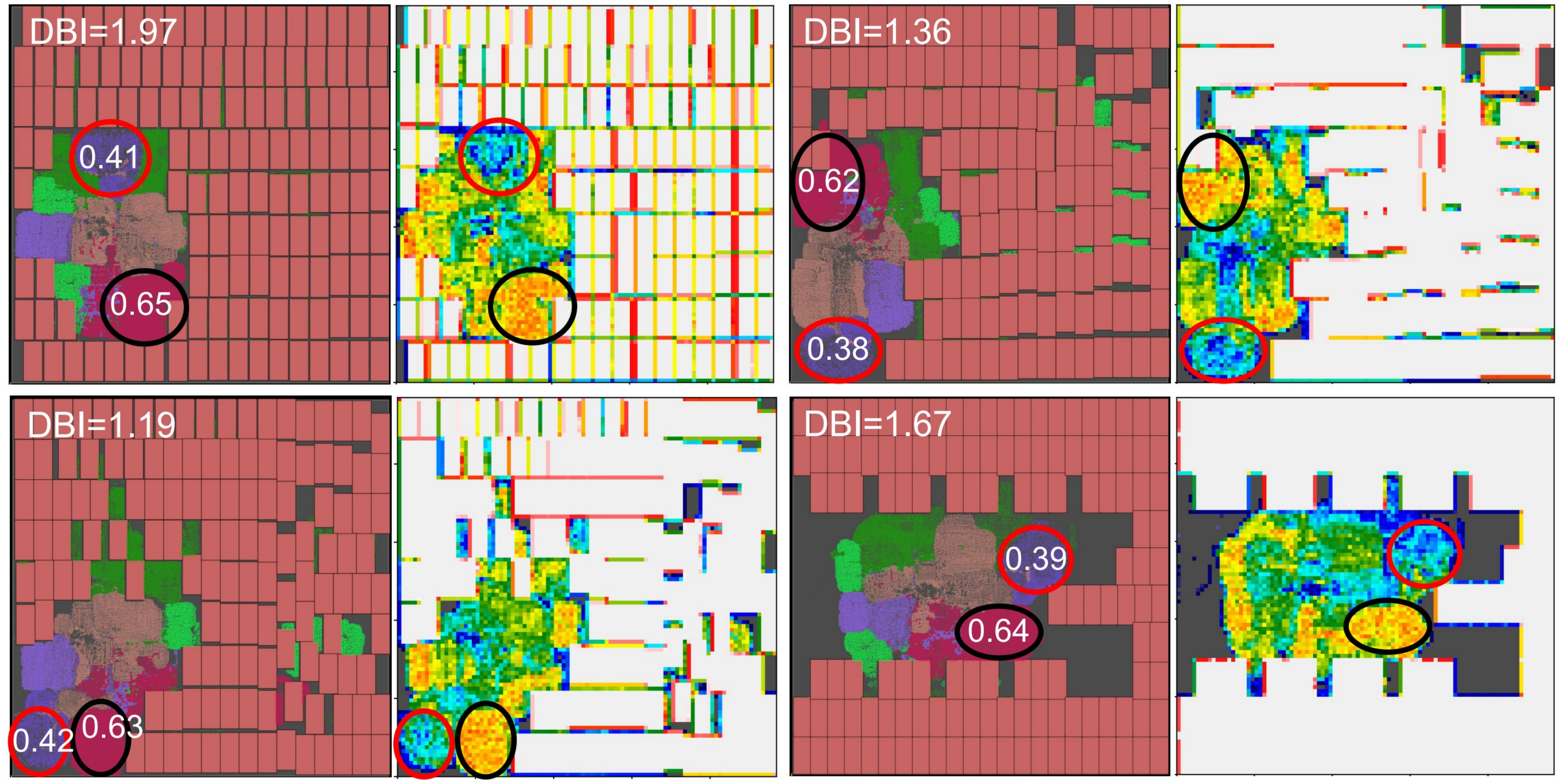
- **Idea**
  - Can we optimize directly at the placement stage for the **right density, i.e., "begin with the end in mind"?**



Motivations

Exposing the importance of density

- Analyzing placement, density, and timing
  - Coarse view from clustering to highlight correlations
    - Based on **hierarchy** and **Leiden** algorithm
      - Respect RTL logical groups and timing boundaries
    - Tune parameters to optimize the Davies–Bouldin index and timing/density correlation
    - Fast implementation based on cuGraph
  - Multiple post-route placements from commercial EDA tool



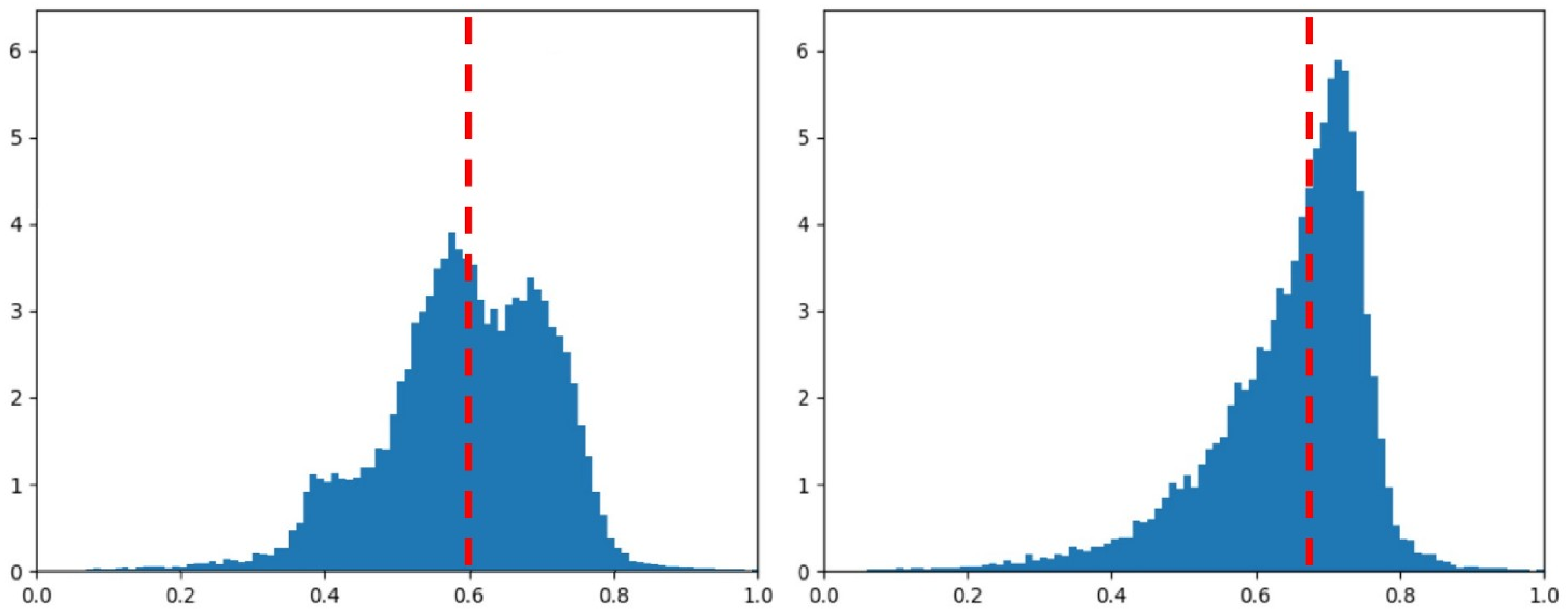
- **Observations**
  - Clustering accurately predicts cells that consistently co-locate across placements
  - **Strong correlations** between lower density and higher timing criticality

Design	#cells	#clusters	ρ_{DT}	$\rho_{DT,\text{tim-crit.}}$	$\sigma_{\text{cluster-dens.}}$	DBI	RT
Ariane	100K	8	0.69	0.84	0.08	1.55	11s
NVDLA	150K	10	0.77	0.88	0.11	1.21	16s
BlackParrot	650K	25	0.61	0.71	0.09	2.05	22s
MemPool Group	2.7M	69	0.72	0.91	0.10	2.40	31s

Cell Density

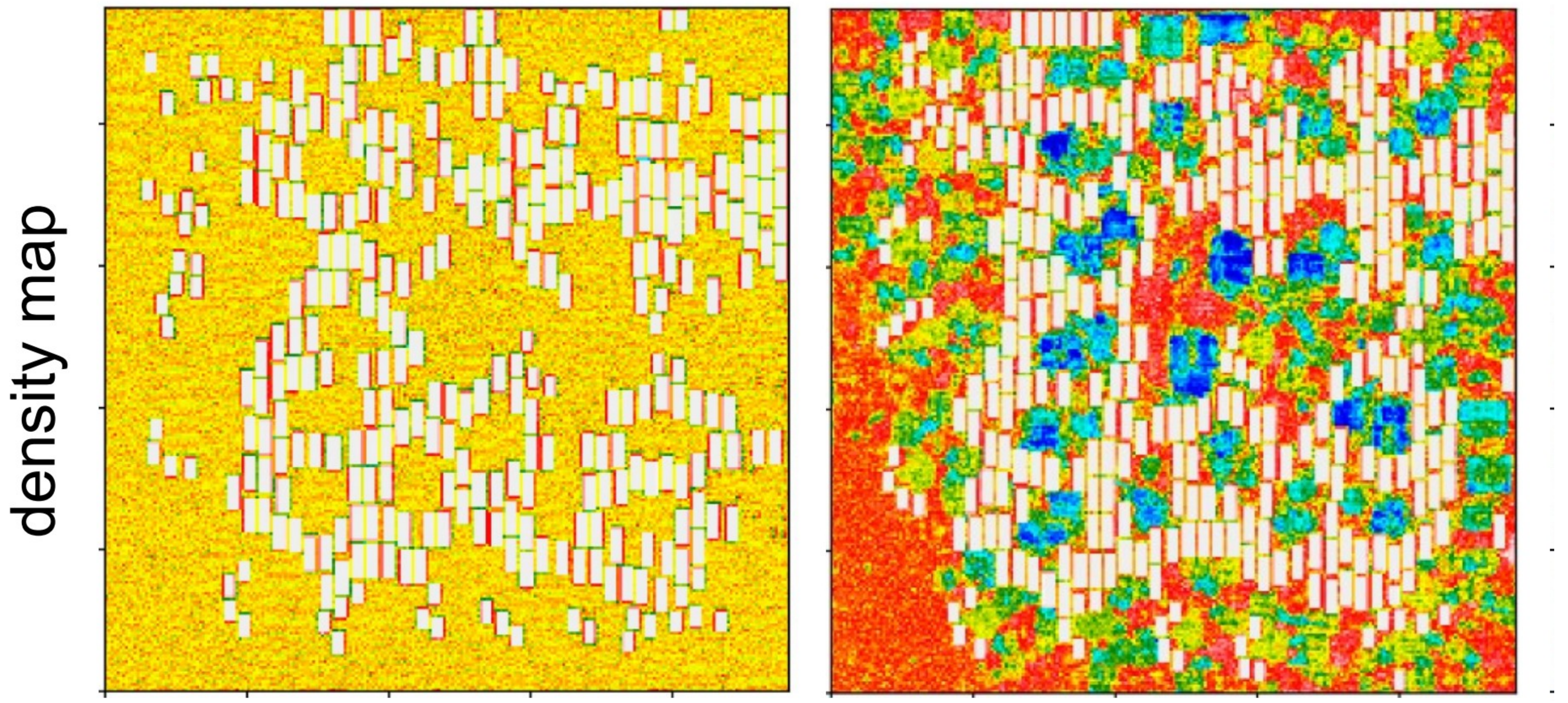
Selection, Computation, and Control

- **Post-route** cell density as our goal
  - Large distribution shift across stages
- Extracting meaningful cell density targets
  - Compute on pre-place netlist using post-route cell locations
  - Remove buffers/non-matching cells to account for later timing-opt steps
  - Density grid w/ bins of size 10x10 standard cell row height
- Density control for DREAMPlace, our global placer
  - Poisson's electric field enforces uniform density targets
  - Cell inflation is best control method
    - Principle is that larger cells lower density around them
    - Inflating each cell their inverse target density is best trade-off
      - Cells in same bin might have different targets
    - Stable convergence without impacting runtime



(a) post-global place (b) post-route optimized

$$\rho_b = \sum_{i \in V} \frac{\mathrm{OA}(i,b)}{A_b} \quad \text{and} \quad \rho_i = \sum_{b \in B} \rho_b \frac{\mathrm{OA}(i,b)}{a_i}$$



uniform density *w/ inflation*

James-Stein Empirical Bayes

Adapting the targets to the placer

- No assumption on the post-route targets
  - E.g., obtained from a commercial EDA tool (our experiments)

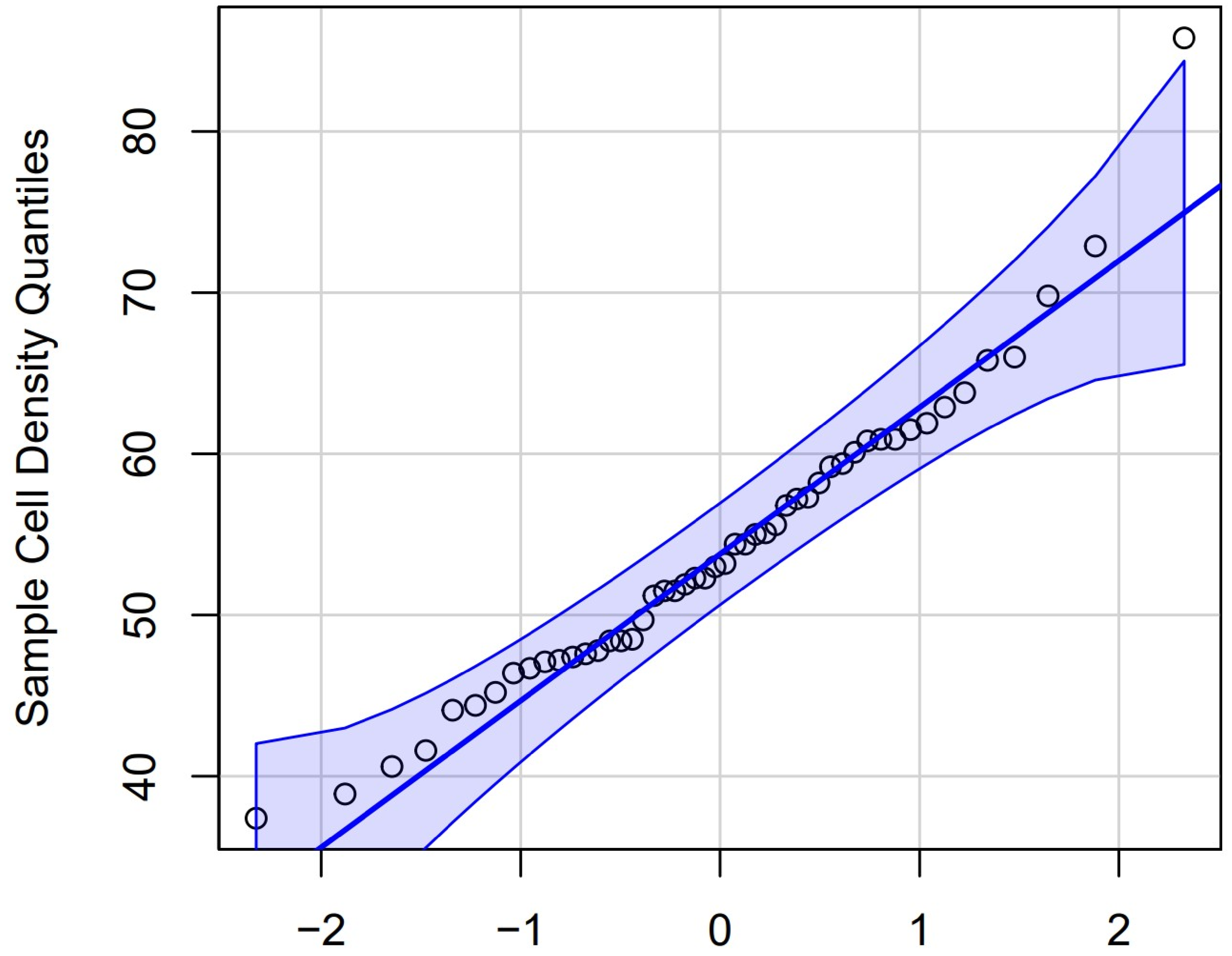
- The post-route target densities might be suboptimal for place
  - Different engines and many steps from place to post-route
  - Different trade-offs at global place (density vs. wirelength mainly)

- **Adapt** the targets to the placer
  1. Estimate the placer’s cell density distribution from (inflated) runs
  2. Compute a better target density
     - Reconcile the target densities (what we think is good) with the placer densities (what we can achieve)

- **James-Stein Empirical Bayes**
  - Each cell density has a true expectati μ_i “from nature”
  - We observe the target density once, assumed sampled from the (prior) Gaussian distribution (see Quantile plot)

$$z_i \sim \mathcal{N}(\mu_i, \sigma_0^2) \quad (i = 1, 2, \ldots, N)$$

  - Compute an estimator fo μ_i better th $\hat{\mu}_i^{(\mathrm{MLE})} = z_i$



Quantile-Quantile plot of density of cells across placements for most of the cells

James-Stein Empirical Bayes

Adapting the targets to the placer

- **James-Stein Empirical Bayes**
  - Bayes rule gives the best estimator but has unknowns

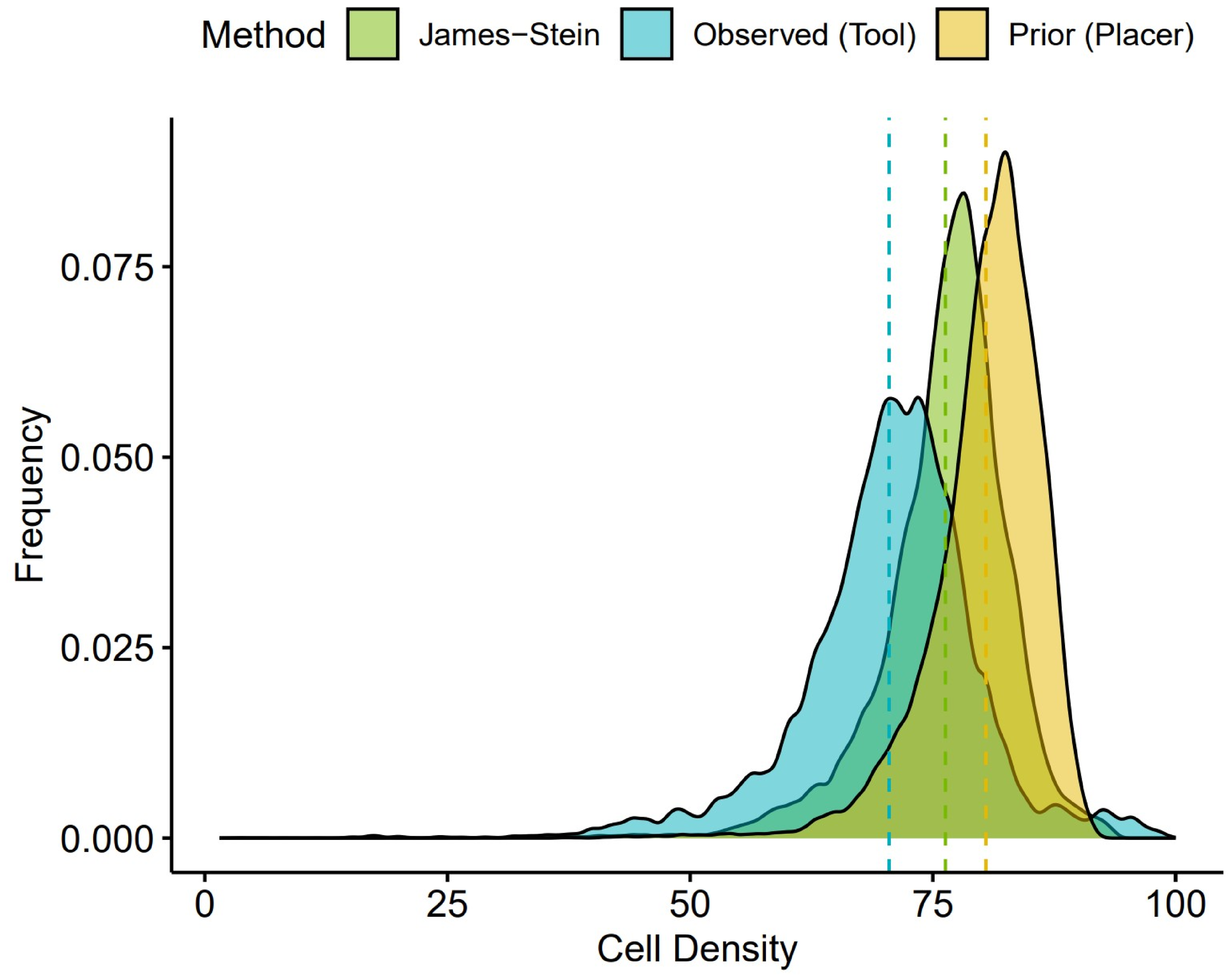
$$\hat{\mu}_i^{(\mathrm{Bayes})} = M_i + B(z_i - M_i)$$

  - "Empirical Bayes" estimates the unknowns from data/our placement runs

$$\hat{\mu}_i^{(\mathrm{JS})} = \hat{\mu}_i^{(\mathrm{P})} + \underbrace{\left(1 - \frac{(N-2)\sigma_0^2}{S}\right)}_{\text{shrinking factor}} \left(z_i - \hat{\mu}_i^{(\mathrm{P})}\right)$$

mean density from runs

  - Trade-off between target and placer
  - Parameters estimated by pooling information from all c $S = \sum \left(z_i - \hat{\mu}_i^{(\mathrm{P})}\right)^2$
    - Extremely data-efficient
    - Large-scale inference metho $R^{(\mathrm{JS})} \Big/ R^{(\mathrm{Bayes})} = 1 + \dfrac{2\sigma_0^2}{NA}$



- Enhancement for **timing**
  - James–Stein concentrates on **global savings**
  - Unusual cases can suffer from being shrunk toward the mean
  - **Restrict deviation of timing-critical cells**

$$\hat{\mu}_i^{(\mathrm{JS/D})} = \begin{cases} \max\left(\hat{\mu}_i^{(\mathrm{JS})}, \hat{\mu}_i^{(\mathrm{MLE})} - D_i\sigma_0\right) & \text{for } z_i > \hat{\mu}_i^{(\mathrm{P})} \\ \min\left(\hat{\mu}_i^{(\mathrm{JS})}, \hat{\mu}_i^{(\mathrm{MLE})} + D_i\sigma_0\right) & \text{for } z_i \le \hat{\mu}_i^{(\mathrm{P})} \end{cases}$$

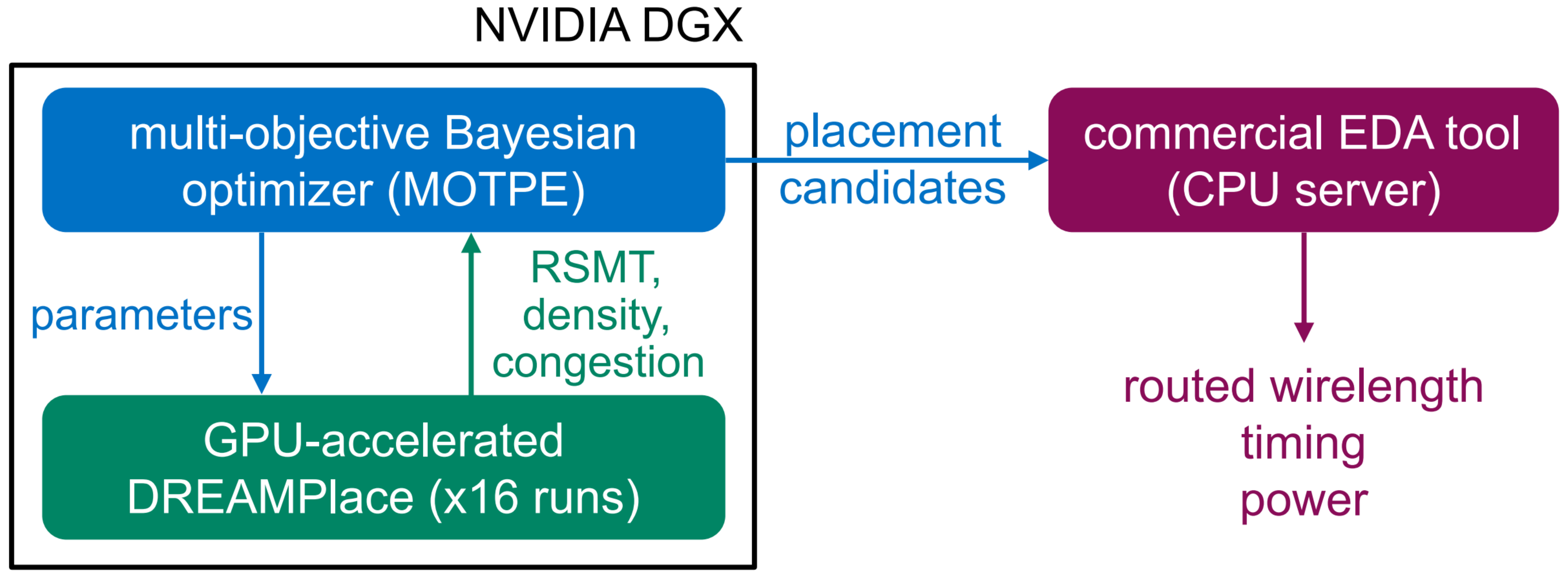
Integration with AutoDMP

Searching for better placements

- AutoDMP is a multi-objective Bayesian optimization methodology for placement
- New objectives (PPA axes)
  - Wirelength, congestion, and ~~uniform density~~ (fixed to 1 for inflation)
  - Add **Hellinger distance** to encourage placements that match the target cell density distribution

$$H(D_P, D_T) = \frac{1}{\sqrt{2}} \left\| \sqrt{D_P} - \sqrt{D_T} \right\|_2$$

- New parameter to shift the target density distribution to explore the trade-off between congestion, density, and wirelength



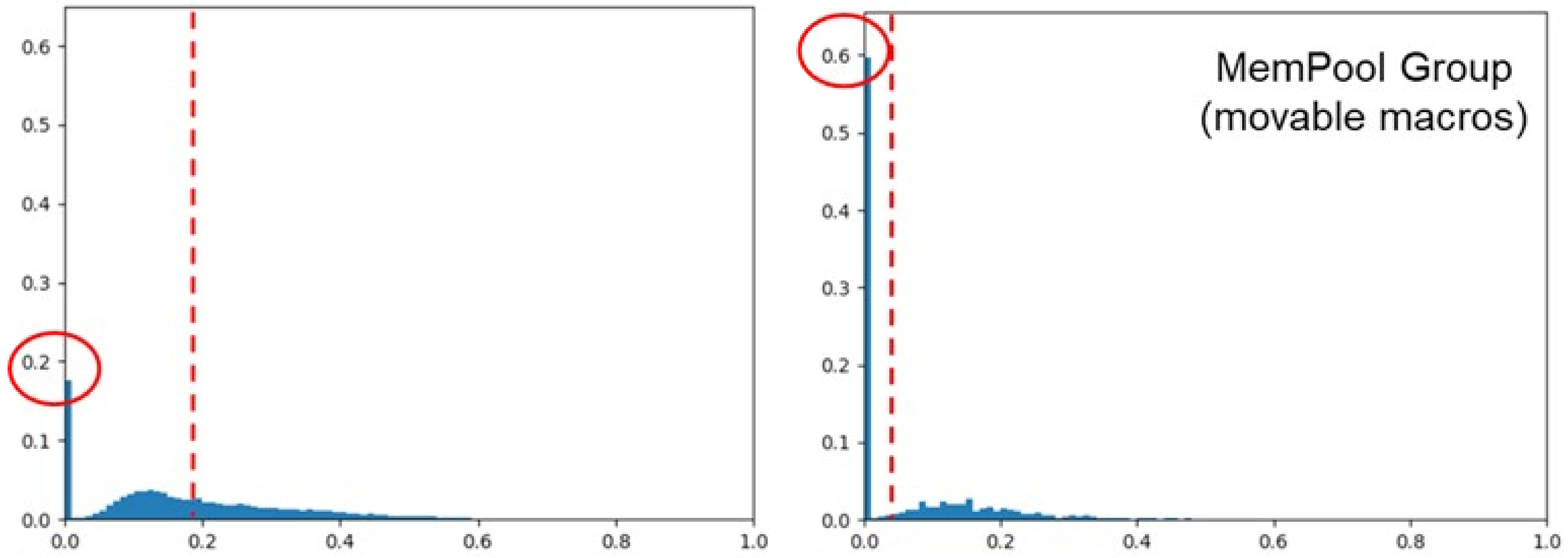
Parameter	Search Range	$\widehat{c_V}$ (%) RSMT	$\widehat{c_V}$ (%) Cong.	Divg. Rate
*horiz. initial position	[0.2, 0.8] (%)	2.2	0.9	0.0
*vert. initial position	[0.2, 0.8] (%)	2.0	1.1	0.0
*horiz. macro halo	technology dep.	1.8	1.3	0.0
*vert. macro halo	technology dep.	1.7	1.2	0.0
~~target density d_{target}~~	~~$[a_{util} - 0.2, a_{util}]$ (%)~~	~~-~~	~~-~~	~~-~~
density weight	$[1e^{-6}, 1.0]$	3.1	1.7	0.0
smooth HPWL model	{LSE, WA}	0.7	1.1	0.0
smooth HPWL initial γ_0	[0.10, 0.50]	5.1	1.9	0.0
GD initial LR lr_0	$[1e^{-4}, 1e^{-2}]$	1.4	1.0	0.0
GD LR decay	[0.99, 1.0]	6.7	2.3	53.2
GD optimizer	[Adam, Nesterov]	1.2	0.8	54.2
# horiz. global bins	{256, 512, 1024, 2048}	1.3	0.9	0.0
# vert. global bins	{256, 512, 1024, 2048}	3.1	1.3	21.1
λ update lower coeff. L	[0.90, 0.99]	4.2	1.9	0.0
λ update upper coeff. U	[1.01, 1.15]	27.0	7.5	1.8
λ update Δ HPWL$_{REF}$	$[1.5e^5, 5.5e^5]$	2.3	1.2	0.0
target density shift	[−0.2, 0.2]			

Experiments

Settings and Validation of Assumptions

- Settings
  - Global place pushed to post-route using a commercial EDA tool (Cadence Innovus)
    - Fixed and unfixed macros for global place
    - Using macro and/or standard cell placement for downstream steps
  - Tested on industrial and academic P&R flows/benchmarks/technologies
  - AutoDMP runs on NVIDIA DGX with four A100 GPUs for 1,000 placements

- Validation of working cell inflation
  - Effective control of cell density in DREAMPlace



(a) Per-bin range without cell inflation

(b) Per-bin range with cell inflation

Design	No Infl. (uniform)	Infl. Tool (z)		Infl. JS ($\hat{\mu}^{(JS)}$)	
		cell	cluster	cell	cluster
Ariane	< 0	0.65	0.90	0.72	0.95
NVDLA	< 0	0.61	0.85	0.89	0.97
BlackParrot	< 0	0.58	0.83	0.71	0.92
MemPool Group	< 0	0.76	0.92	0.85	0.99

Experiments

PPA Evaluation

- AutoDMP candidates are of **much better quality** for the EDA tool
- James-Stein targets co-optimize wirelength & congestion
- The timing enhancement provides large **WNS/TNS gains**
- GOALPlace **works across designs/settings**

Post-place average quality reported using the EDA tool's early global router

Method	WL (m)	HPWL (m)	Cong. H/V (%)	Density Hotspot	Updated Paretos
NVDLA-NG45 *(150K cells, 128 macros, 1.11 GHz,* ***movable*** *macros)*					
AutoDMP	9.25	6.69	0.09/0.39	44.5	0/5
+ pin infl.	9.18	6.65	0.08/0.41	46.1	0/5
Ours	**9.09**	6.58	0.07/0.31	44.6	**2/5**
+ JS	**9.08**	**6.54**	**0.05/0.28**	**43.2**	**3/5**
BlackParrot-NG45 *(650K cells, 220 macros, 769 MHz,* ***movable*** *macros)*					
AutoDMP	25.21	19.63	0.08/0.24	57.3	1/5
+ pin infl.	24.92	19.41	0.08/0.26	62.1	0/5
Ours	24.26	**18.94**	0.06/0.19	55.0	1/5
+ JS	**24.20**	19.12	**0.06/0.16**	**53.5**	**3/5**

Post-route PPA on large MemPool Group design

Method	WL (m)	Power (mW)	WNS all (ns)	TNS all (ns)	WNS r2r (ns)	TNS r2r (ns)	Cong. H/V Pre-Route (%)	# DRCs	Search RT	Tool RT
MemPool Group-NG45 *(2.7M cells, 320 macros, 333 MHz,* ***movable*** *macros)*										
Tool	115.5	4124	-0.398	-2668	-0.206	-1548	3.52/1.76	2132	-	32h
AutoDMP (*Macro*)	111.1	4091	-0.330	-2913	-0.226	-1955	3.32/1.66	2651	3h30	28h
Ours (*Macro*)	**110.2**	**4066**	-0.318	-2899	-0.178	-1892	**2.44/1.54**	**210**	*	24h
+ JS (*Macro*)	**109.8**	**4071**	-0.410	-3688	-0.207	-2711	**2.22/1.57**	**144**	2×3h	26h
+ JS Timing (*Macro*)	**110.4**	**4064**	**-0.303**	**-2434**	**-0.130**	**-1339**	**2.31/1.59**	**142**	2×3h	24h
AutoDMP (*Macro/Std-Cell*)	112.0	4087	-0.463	-4708	-0.196	-3438	2.87/1.87	8189	3h30	33h
Ours (*Macro/Std-Cell*)	111.9	4075	-0.355	-4072	-0.169	-3117	**2.00/1.59**	**59**	*	27h
+ JS (*Macro/Std-Cell*)	**110.7**	**4056**	-0.398	-4164	-0.198	-3397	**2.12/1.53**	**76**	2×3h	27h
+ JS Timing (*Macro/Std-Cell*)	**110.7**	**4051**	**-0.327**	**-3586**	**-0.116**	**-1390**	**2.24/1.64**	**60**	2×3h	25h

Post-route average/best PPA

Method	WL	Best WL	Max-Cong.	Best Max-Cong.
ML Accelerator *(400K cells, 65% util., 48* ***movable*** *macros)*				
Tool	-	2.11	-	0.58
AutoDMP	2.27	2.14	0.63	0.61
Ours	2.15	2.11	0.56	0.54
+ JS	2.13	**2.09**	**0.55**	**0.51**

Conclusion & Future Work

- Great potential to **learn from end-of-flow** results
  - Placement density as a prime example

- Our method's **benefits**
  - Inflation is **easily integrable** into a global placer
  - James-Stein estimator is very **data-efficient**
  - Large **PPA benefits** compared to baseline AutoDMP

- Future work
  - Predict target cell densities directly
  - Enabling a **feedback loop** with the EDA tool

NVIDIA