CoPlace: Coherent Placement Engine with Layout-aware Partitioning for 3D ICs

Bangqi Fu, Lixin Liu, Yang Sun, Wing-Ho Lau, Martin D.F. Wong, Evangeline F.Y. Young

Dept. of Computer Science & Engineering, CUHK

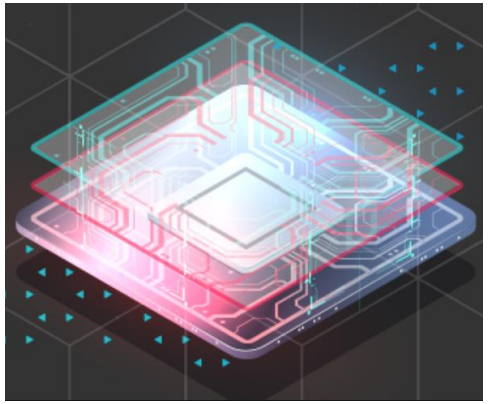
香港中文大學
The Chinese University of Hong Kong

3D-IC Advantages and Challenges

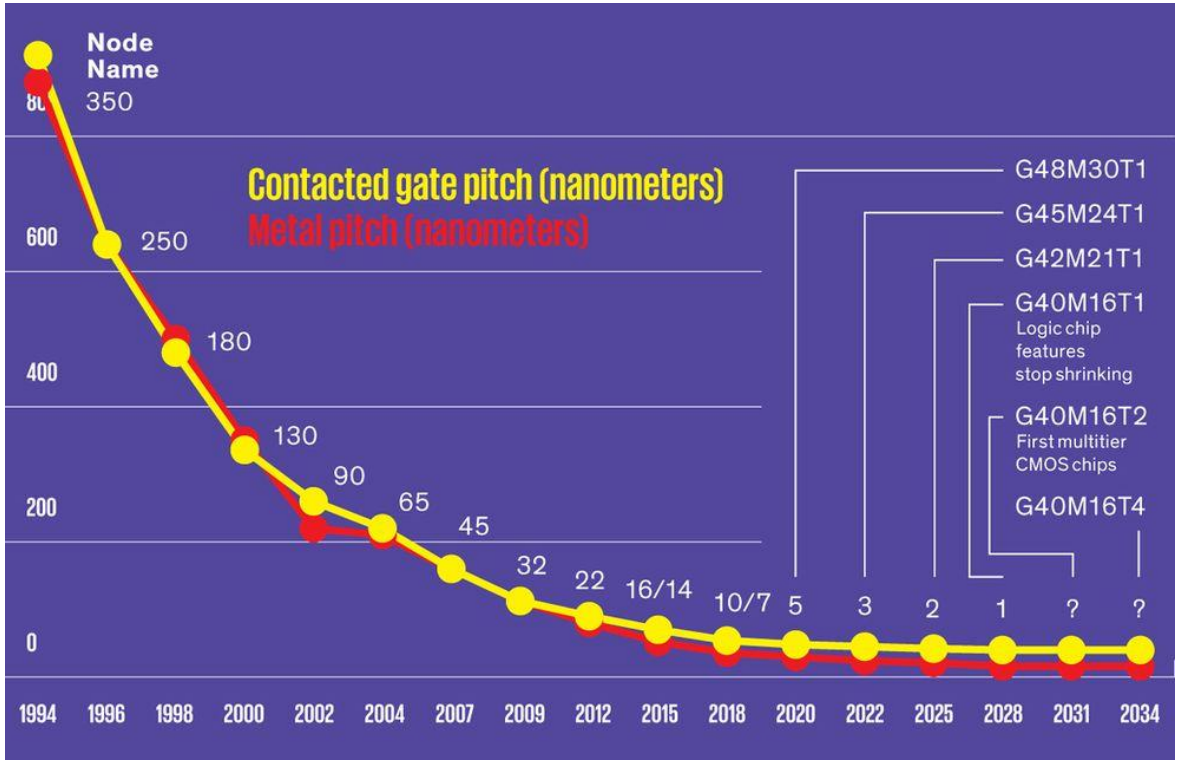
Modern IC size explosively increases

- Technology node meets bottleneck
- 2D-ICs have problems with area, power and cost

3D-IC promises many advantages



- Better timing, area and cost
- Fabrication friendly
- Modularity
- …



[1] https://spectrum.ieee.org/a-better-way-to-measure-progress-in-semiconductors

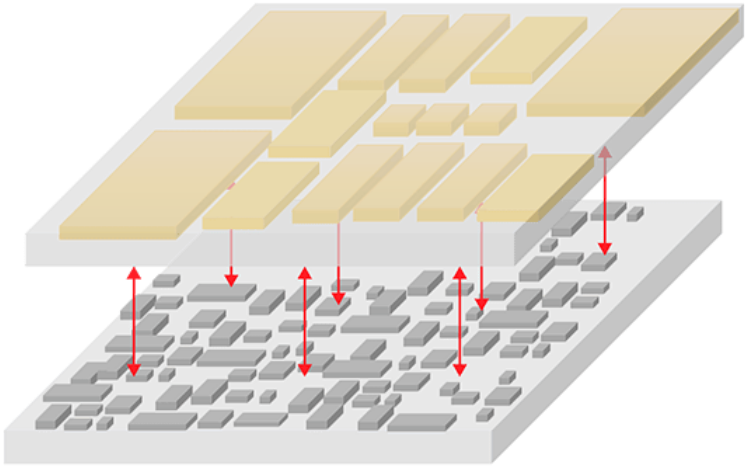
3D-IC Technologies

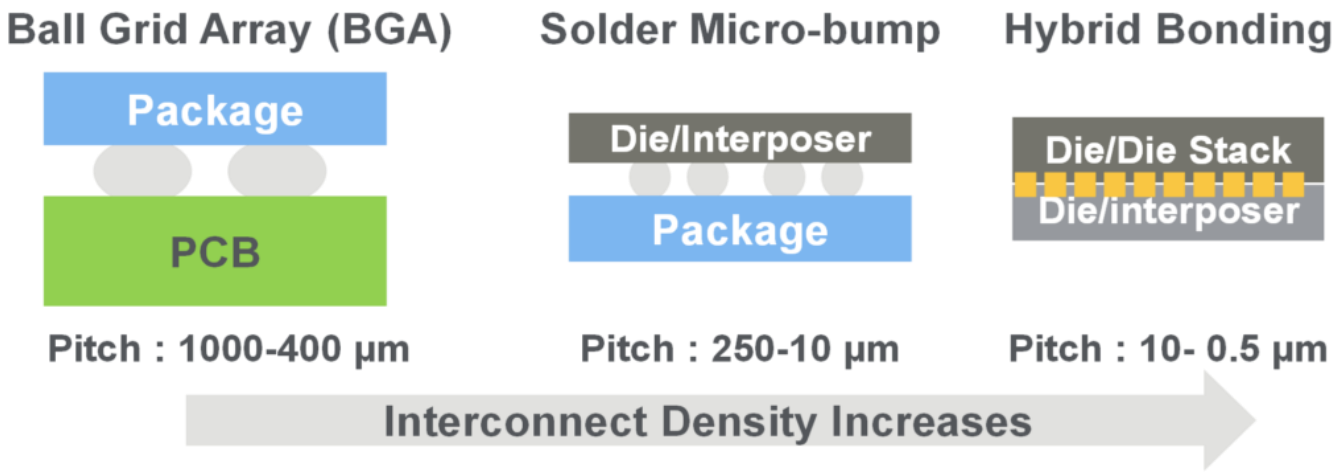
3D-IC varies in technologies

- Die-to-die, multi-die, chiplet, …
- Through-silicon via, monolithic inter-tier via, hybrid bonding terminal

Advantages of die-to-die bonded 3D-IC

- High interconnection density
- Flexibility
- Cost-effective



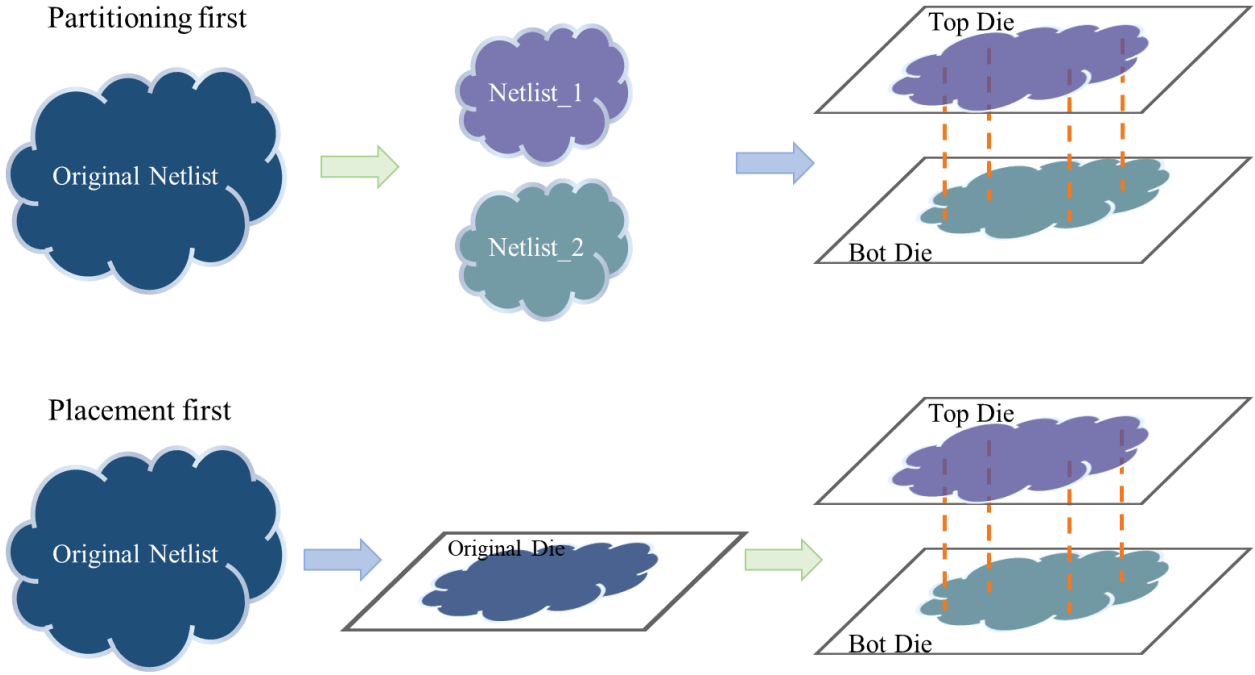


[2] https://www.synopsys.com/blogs/chip-design/upgrading-3d-ic-packaging-for-faster-ai-inference-with-psm.html

3D-IC Technologies

3D-IC design methodologies

- There exists a gap between placement and partitioning
- Chicken and egg problem
  - Partitioning first solution:
    - Lacks physical information
  - Placement first solution
    - Cannot preserve good placement quality

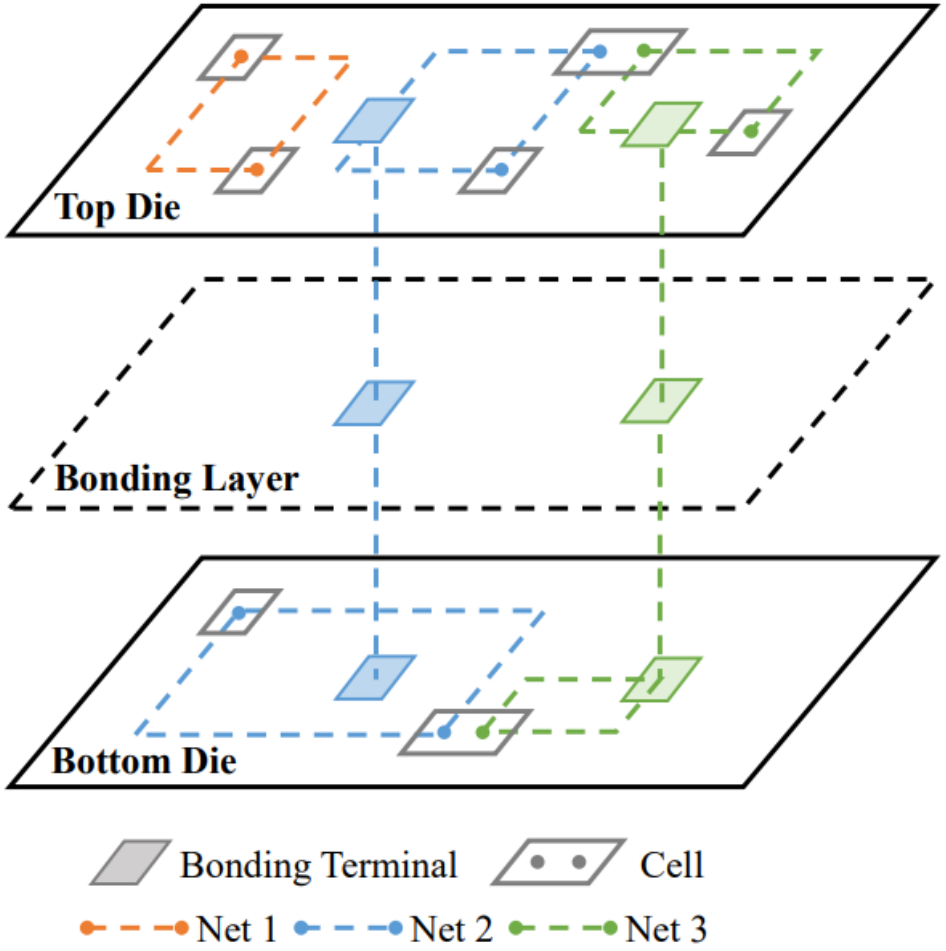


Algorithms that strongly couple the placer and partitioner is needed!

Problem Formulation

Partition and place the original netlist to minimize the cross-die HPWL

- Partition the original circuit $G = (V, E)$ into 2 sub-circuits
- Maximum utilization constraint U_M
- A cut net must have one hybrid bonding terminal
- HPWL is calculated separately on each die
- Hybrid bonding terminal spacing constraint



Analytical Placement Algorithm

Relaxed Objective Function

$$\min_p HPWL(p) = \min_p \sum_{e \in E} HPWL_e(p) \quad \text{s.t.} \quad D_b \leq D_t, \forall b \in B$$

$$\Rightarrow \quad \min_p \sum_{e \in E} \underbrace{WL_e(p)}_{\text{Wirelength}} + \underbrace{\lambda D(p)}_{\text{Density}}$$

- Net wirelength
  - Weighted average (WA) HPWL

$$WL_e(x) = \frac{\sum_{i \in e} x_i e^{x_i/\gamma}}{\sum_{i \in e} e^{x_i/\gamma}} - \frac{\sum_{i \in e} x_i e^{-x_i/\gamma}}{\sum_{i \in e} e^{-x_i/\gamma}}$$

- Cell density
  - Electrostatic system model

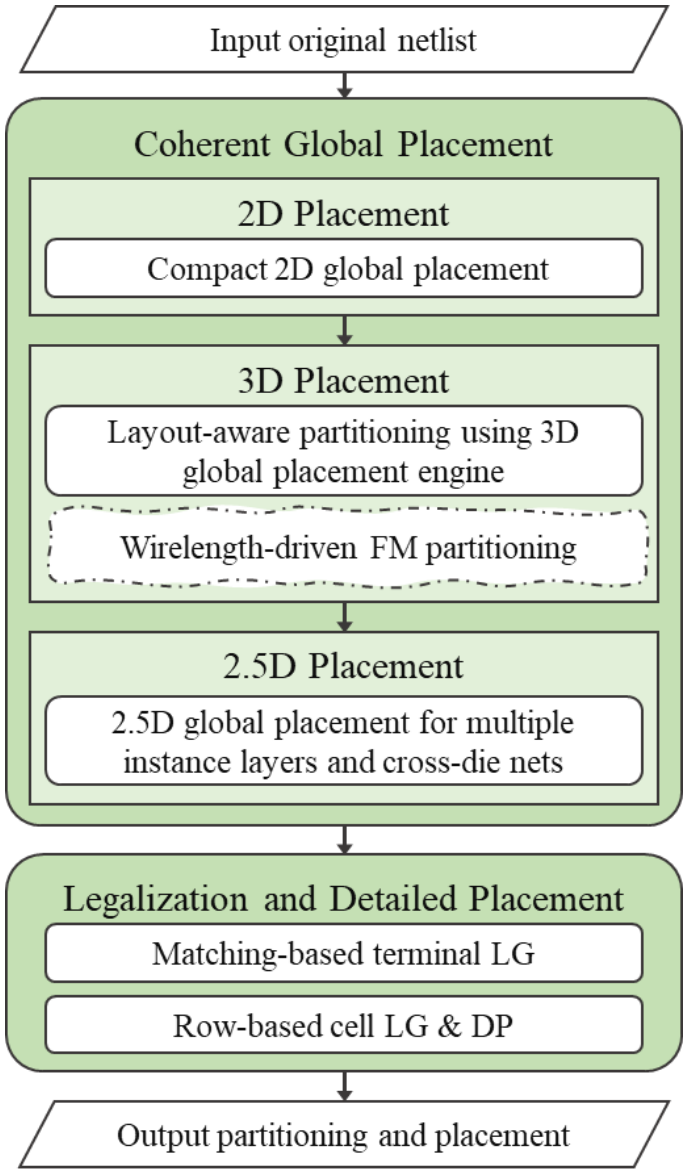
$$\begin{cases} \nabla \cdot \nabla \psi(x,y) = -\rho(x,y), \\ \hat{\mathbf{n}} \cdot \nabla \psi(x,y) = 0, (x,y) \in \partial R, \\ \iint_R \rho(x,y) = \iint_R \psi(x,y) = 0, \end{cases}$$

Proposed Framework

Overall Flow

Coherent placement stages

- Compact 2D global placement
  - Initial solution
- 3D placement on expanded z-dimension
  - Layout-aware partitioning
- 2.5D multi-layer placement
  - Remove overlaps
- Wirelength-driven FM partitioning
  - Refinement
- Legalization and detailed placement



- Coherent placement stages
- Unified placement engine

Proposed Framework

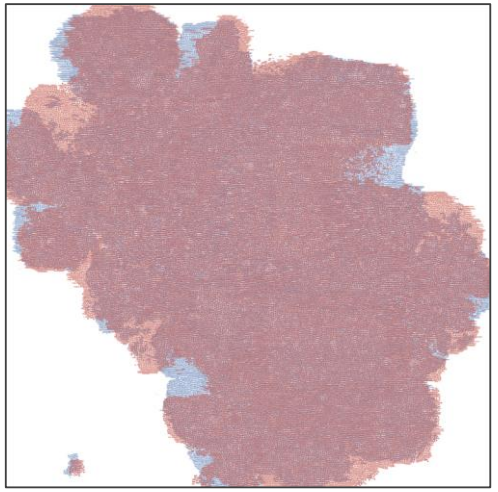
Compact 2D global placement

Serves as an initial solution

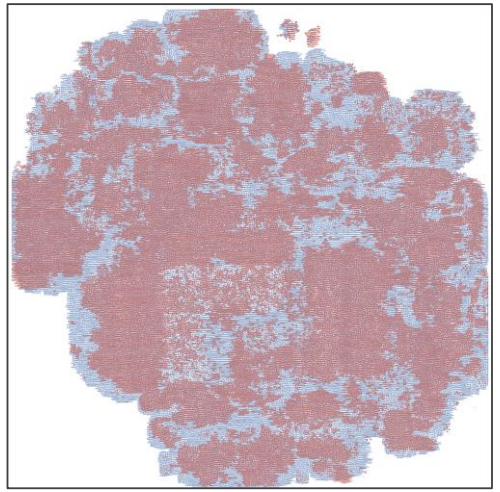
- Less displacement → correct utilization estimation
- Establish enough vertical connection → 2 dies are well-aligned
  → balanced utilization for 2 dies
- $U^0 = U^1 = \frac{A^0 \times A^1}{(A^0 + A^1) \times A^{die}}$

Normalized cell size

- $\bar{s} = \tau \times s^0 + (1 - \tau) \times s^1$
- $\tau = \frac{U^0 \times A^{die}}{A^0}$ is the percentage of cells on die 0



(a) Balanced utilization



(b) Unbalanced utilization

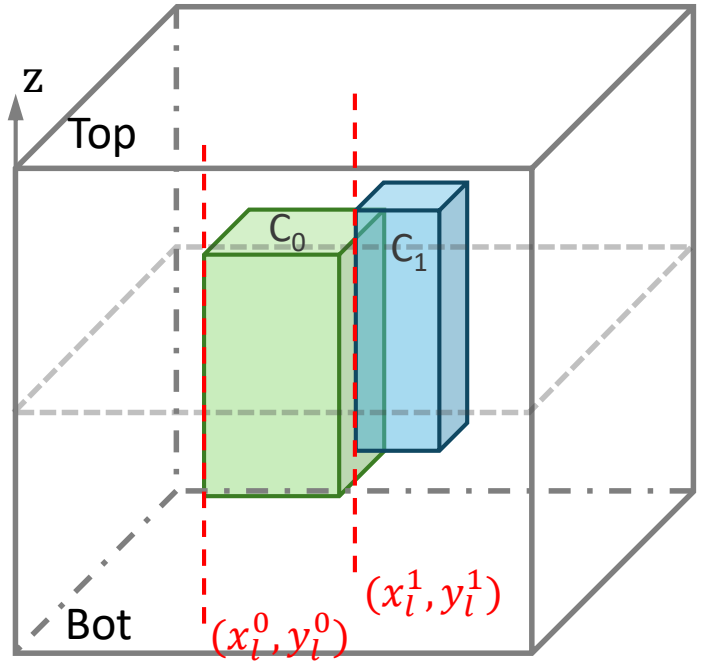
Proposed Framework

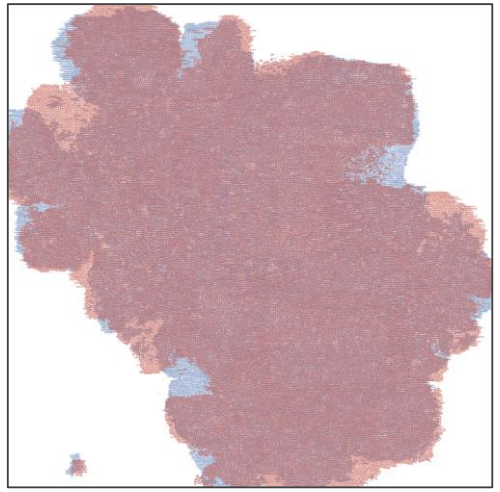
3D Global Placement on z-direction

Serves as layout aware partitioning

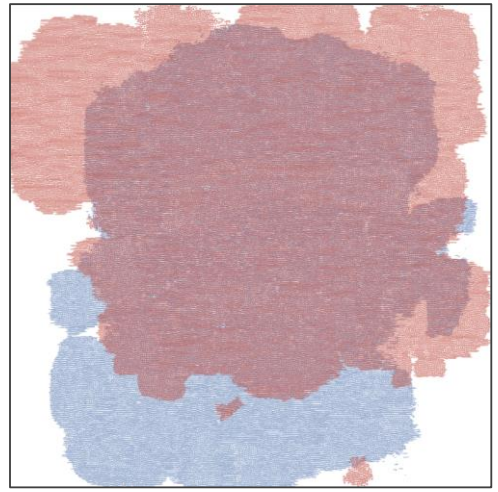
- Preserve initial placement quality → local density is balanced
- Establish enough vertical connection → proper nets are cut

Expand placement space into 3D





(a) Layout aware



(b) Cutsize driven

Proposed Framework

3D placement engine

- 3D spatial density $\rho(x, y, z)$

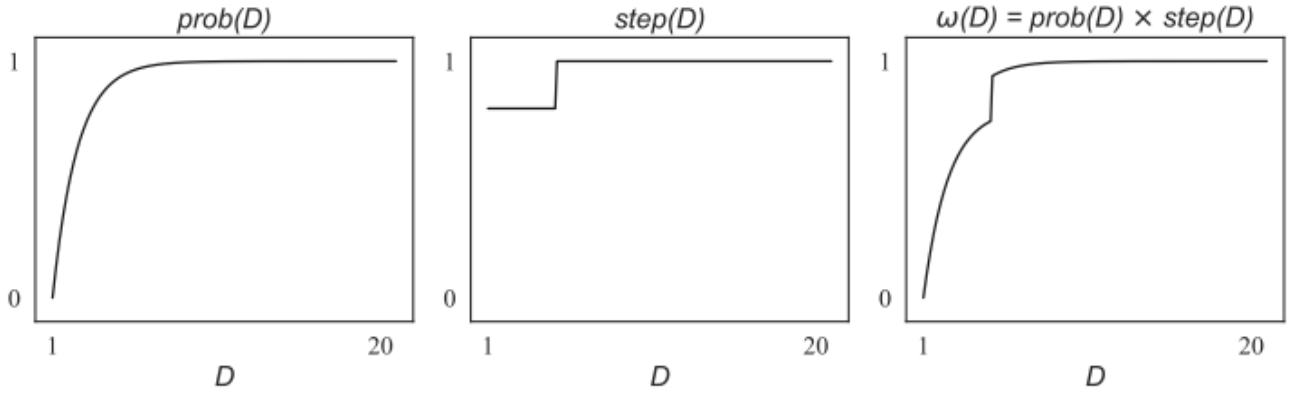
$$\begin{cases} \nabla \cdot \nabla\psi(x,y,z) = -\rho(x,y,z), \\ \hat{\mathbf{n}} \cdot \nabla\psi(x,y,z) = 0, (x,y,z) \in \partial R, \\ \iiint_R \rho(x,y,z) = \iiint_R \psi(x,y,z) = 0 \end{cases}$$

- 3D electrostatic fields distribution

$$\begin{cases} E_x = \sum_{\{j,k,l\}} \frac{a_{j,k,l}\, w_j}{w_j^2 + w_k^2 + w_l^2} \sin(w_{jx}) \cos(w_{ky}) \cos(w_{lz}) \\ E_y = \sum_{\{j,k,l\}} \frac{a_{j,k,l}\, w_k}{w_j^2 + w_k^2 + w_l^2} \cos(w_{jx}) \sin(w_{ky}) \cos(w_{lz}) \\ E_z = \sum_{\{j,k,l\}} \frac{a_{j,k,l}\, w_l}{w_j^2 + w_k^2 + w_l^2} \cos(w_{jx}) \cos(w_{ky}) \sin(w_{lz}) \end{cases}$$

- 3D electrostatic fields distribution

- 3D weighted wirelength $sHPWL$
  - $prob(\mathcal{D}) = 1 - 0.5^{\mathcal{D}-1}$, the probability that a $\mathcal{D}$-degree net is cut
  - $step(\mathcal{D}) = \alpha$ if $\mathcal{D} < T$; 1 otherwise
  - $sHPWL = prob(\mathcal{D}) \times step(\mathcal{D}) \times HPWL_{x,y}$
- Small nets are more preferred to cut



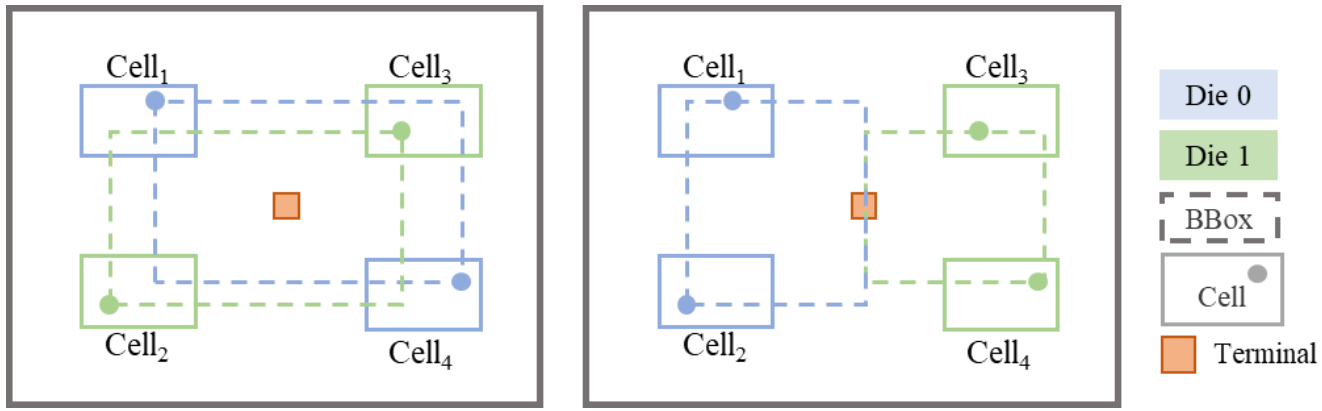
- $\min \sum_{e \in E} sHPWL_e(z) + \lambda D(z)$

Proposed Framework

Wirelength-driven FM Partitioning

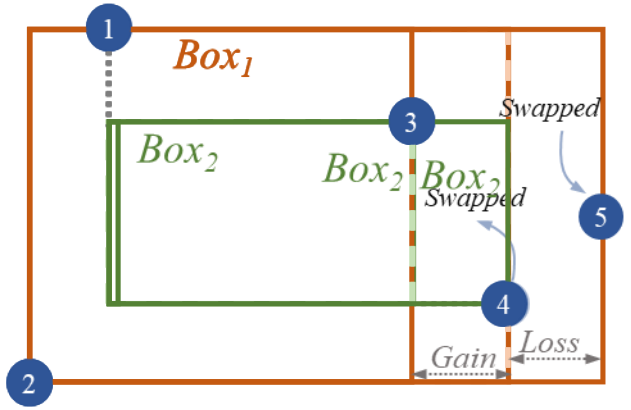
Refine the partitioning to remove HPWL overhead

- Hybrid gain function
  - Reduction in *wirelength*
  - Increasement in *cutsize*
  - Impact on area balance: ΔR^c



Wirelength-driven FM Partitioning

- Swap the cell with the largest gain
- Update the status efficiently with 2 bbox data structure



Proposed Framework

2.5D multi-layer placement

Remove overlaps after partitioning

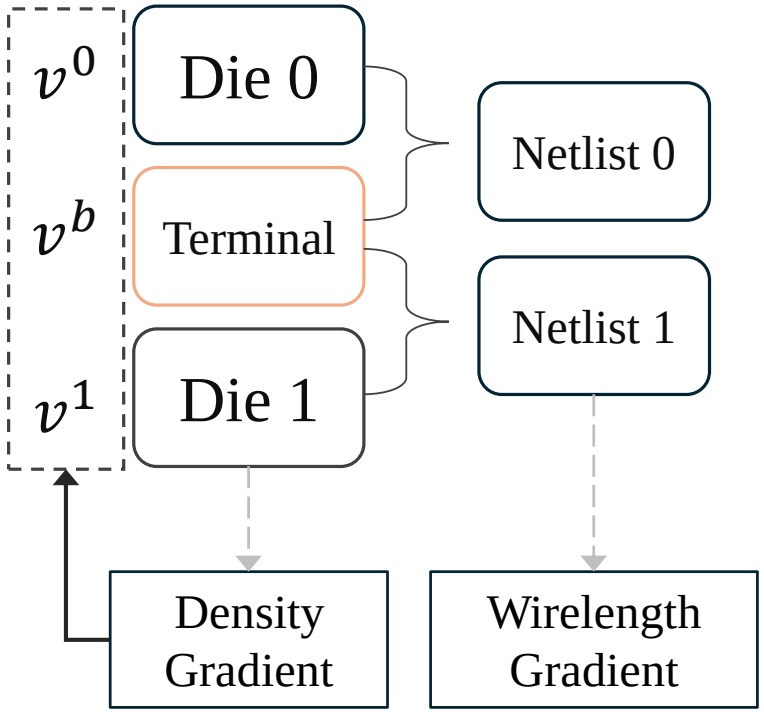
- 2 cell layers
- Instantiate bonding terminal layer with spacing

Non-overlapping constraint ← Spacing constraint

- Place 3 layers simultaneously

Minimize the precise cross-die HPWL

$$\min_{v} \sum_{c\in\{0,1\}} HPWL\left(v^c \cup v^b\right) + \langle \lambda, D \rangle$$



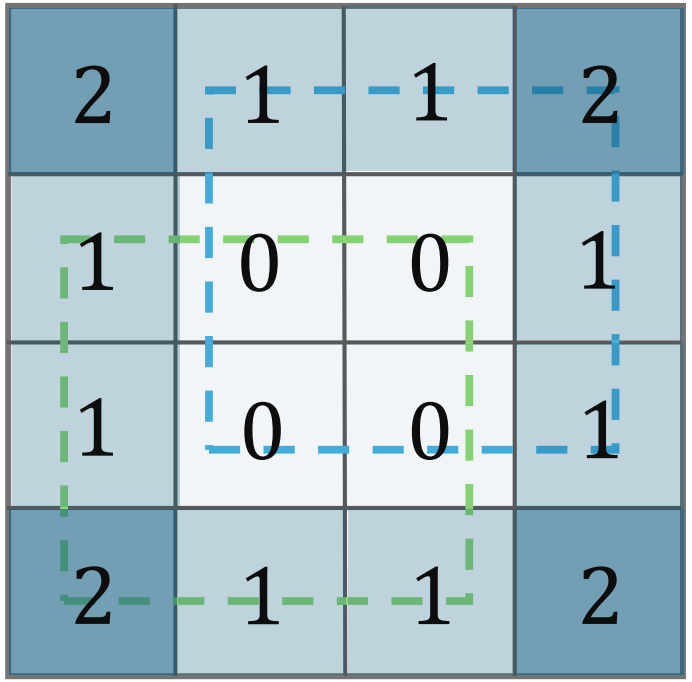
Proposed Framework

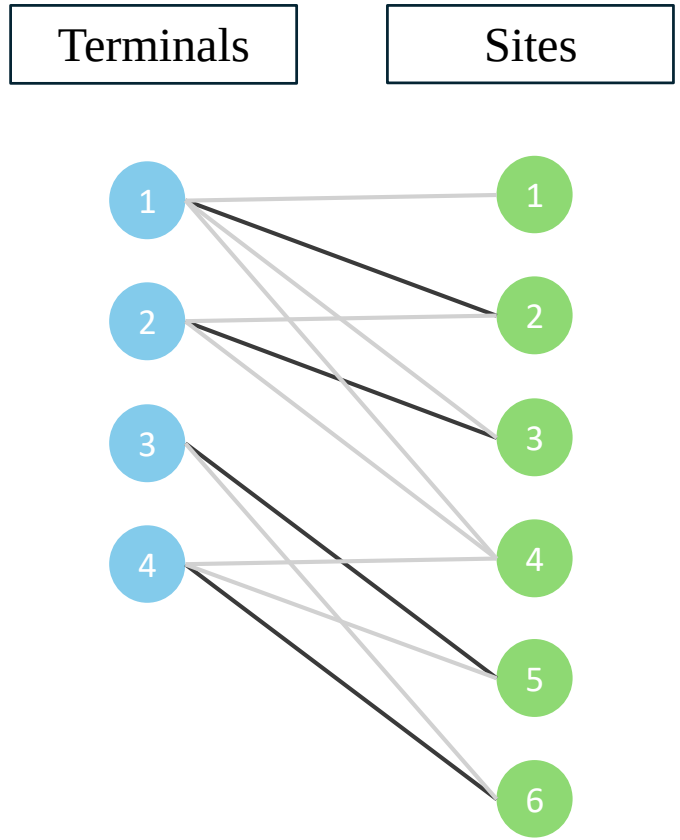
Row-based legalization and detailed placement

- Legalize cells

Matching-based legalization

- Legalize terminals
- Limit the solution to region of interest
- Solve the bipartite matching
- 1% improvement over row-based method

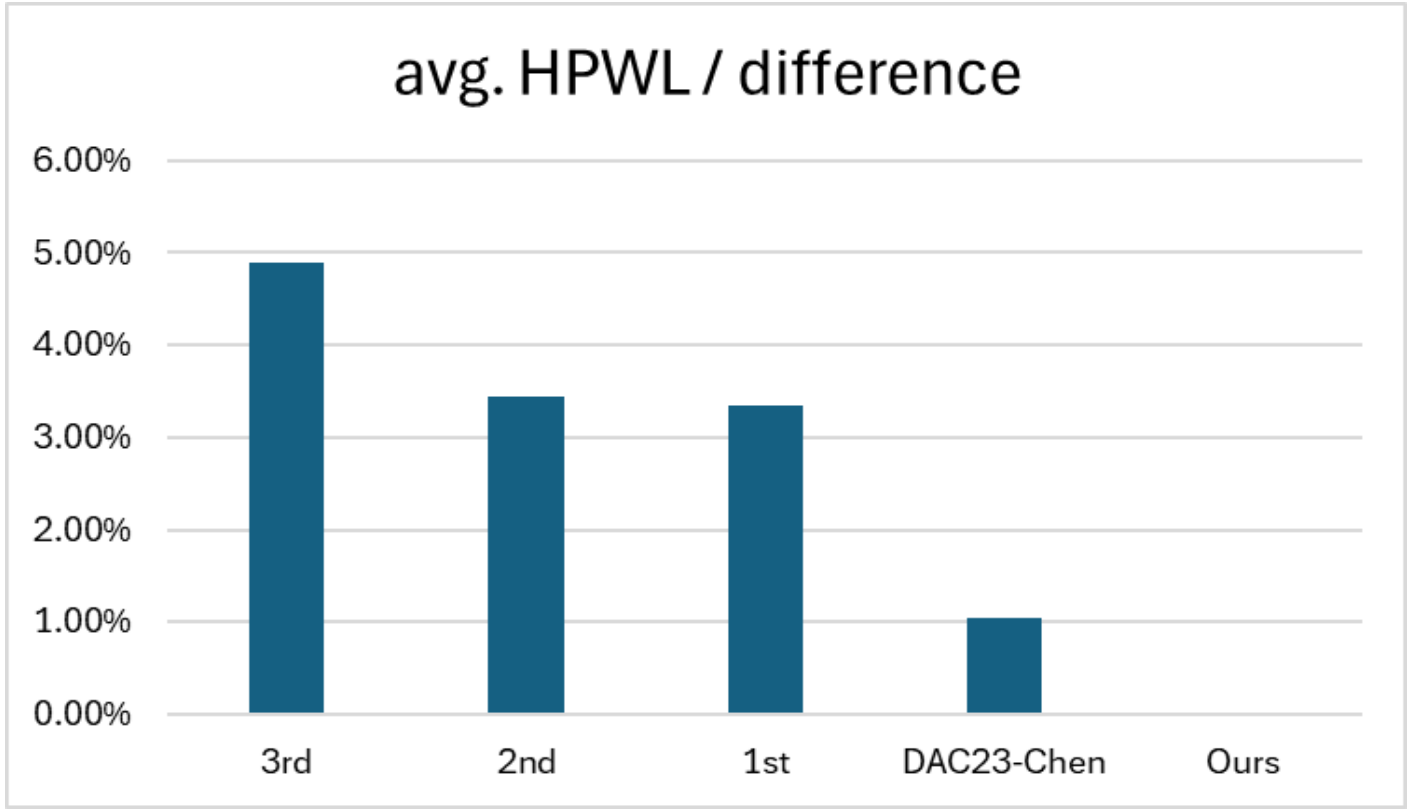




Experimental Results

Experimental Setup

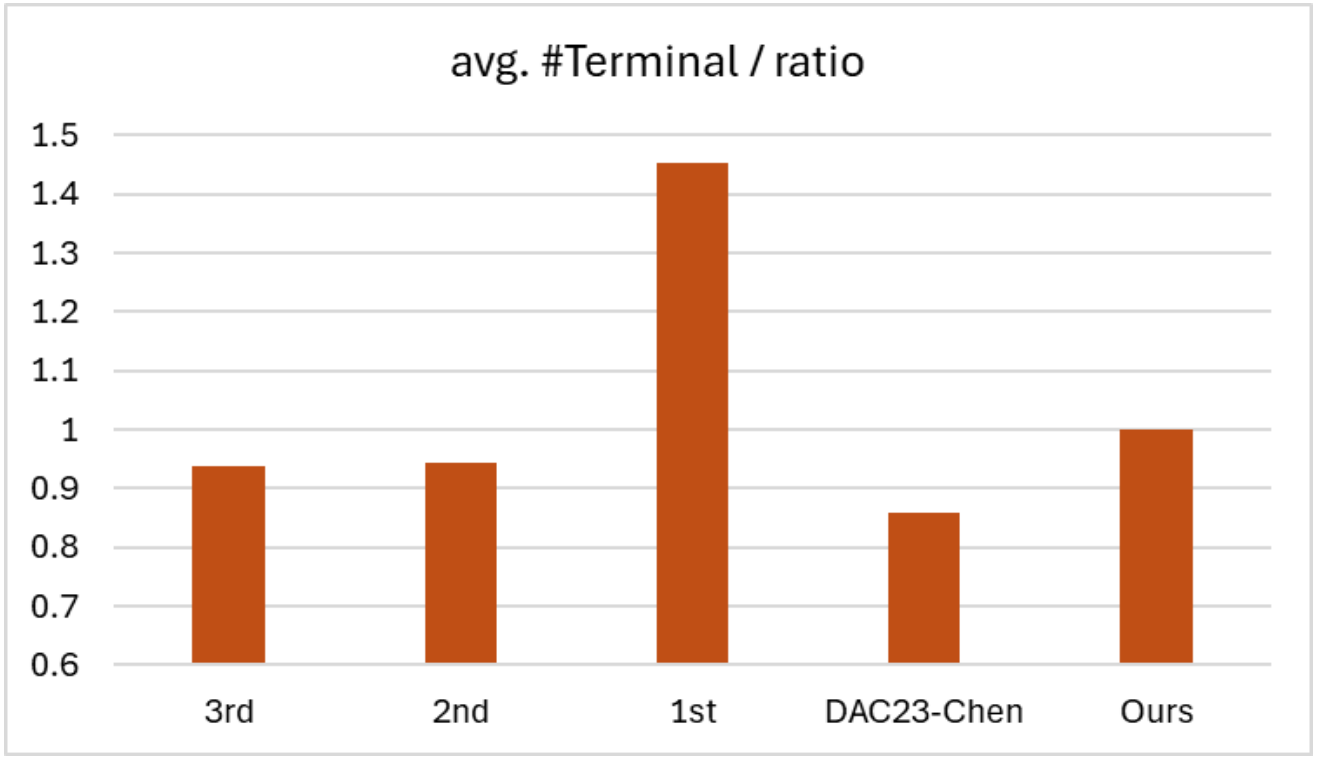
- Tested on ICCAD 2022 Contest benchmark
- Overall of 3% improvement compared to the 1st place team
- 1% improvement compared to the SOTA placer



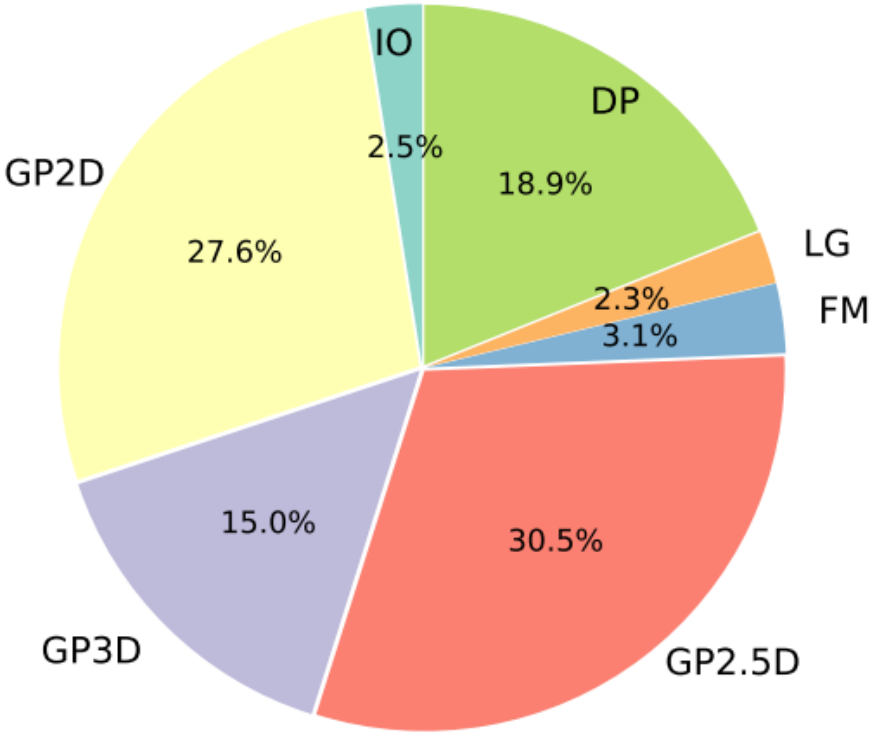
[3] Y.-J. Chen, eta, and Y.-W. Chang, “Late Breaking Results: Analytical Placement for 3D ICs with Multiple Manufacturing Technologies,” in Proc. DAC, 2023

Experimental Results

#Terminals

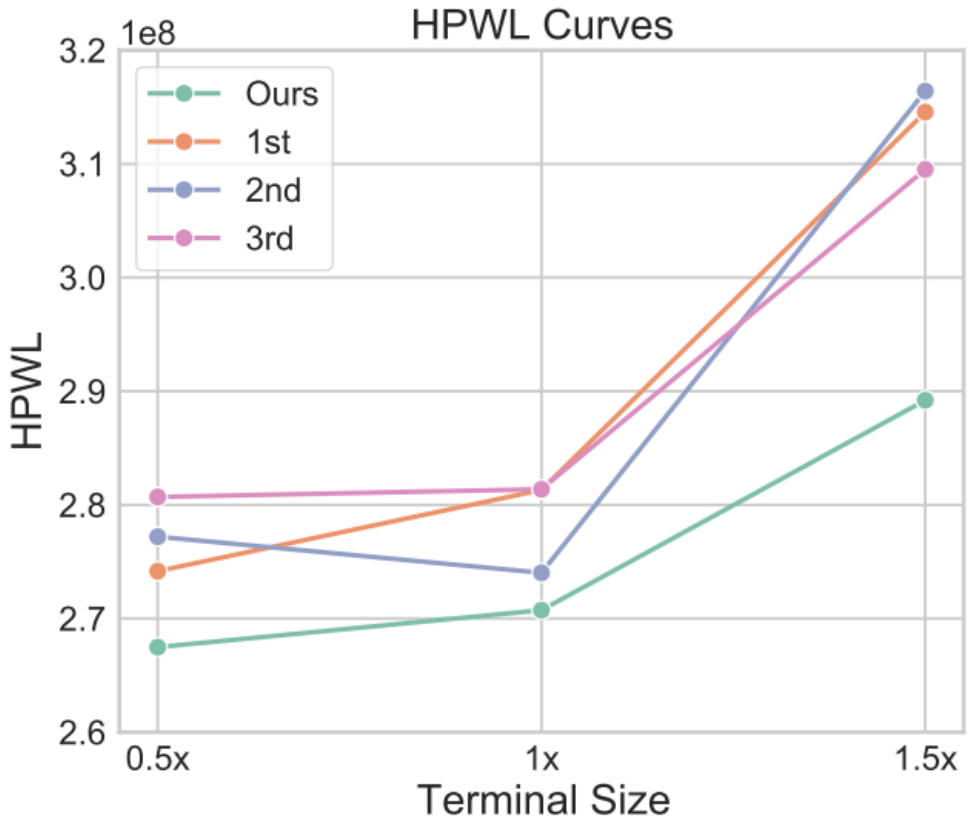


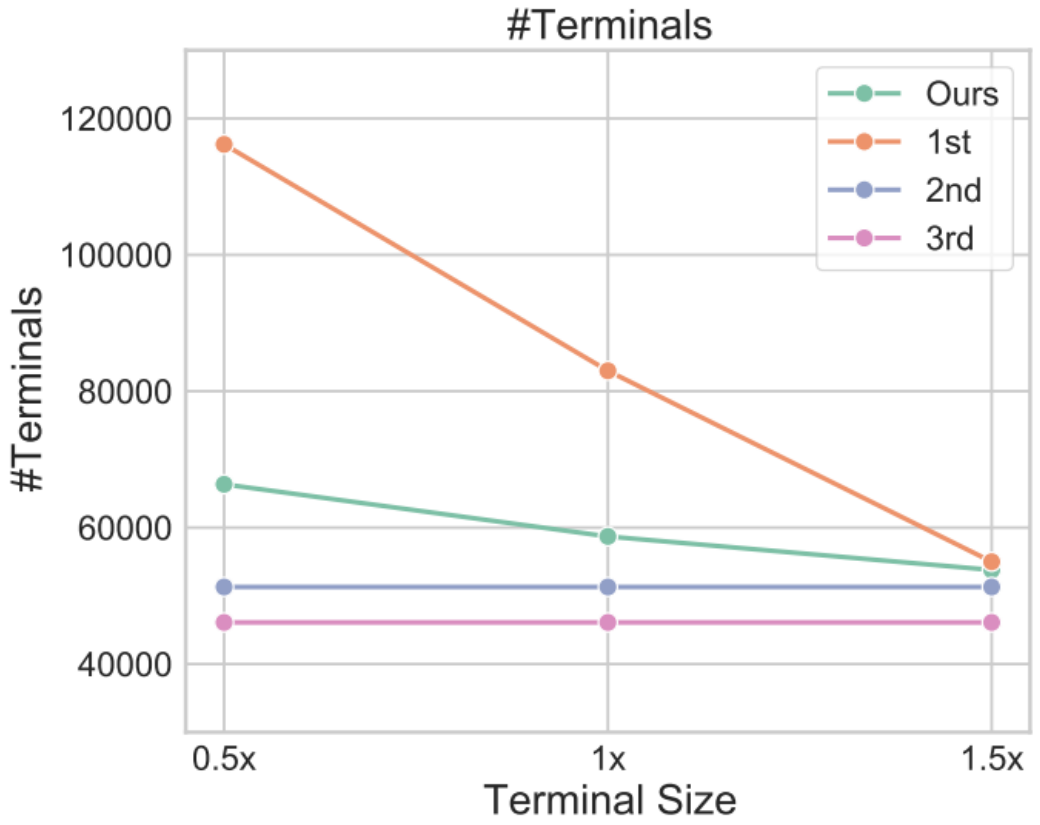
Runtime breakdown



Experimental Results

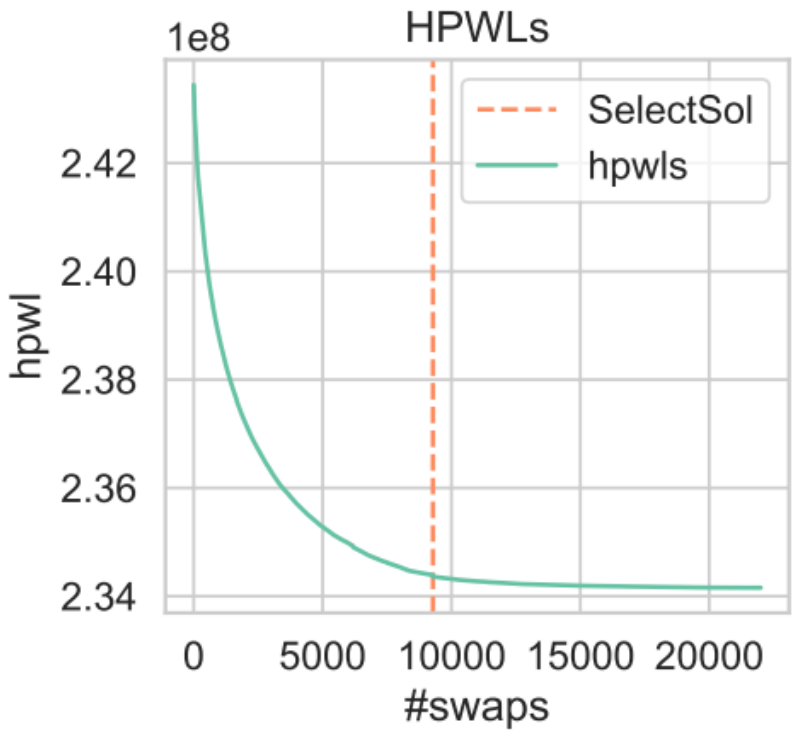
varying terminal size

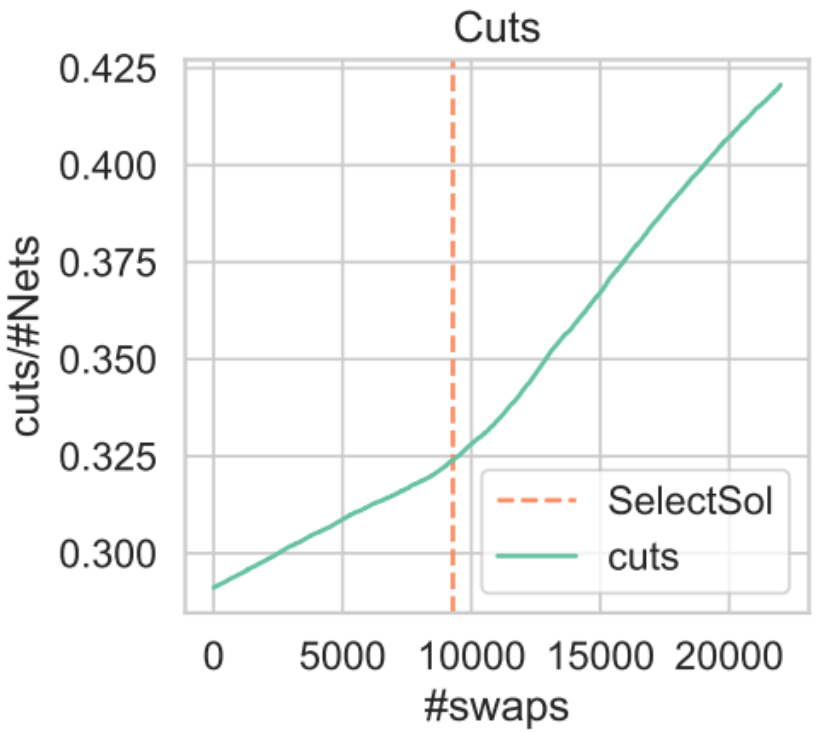


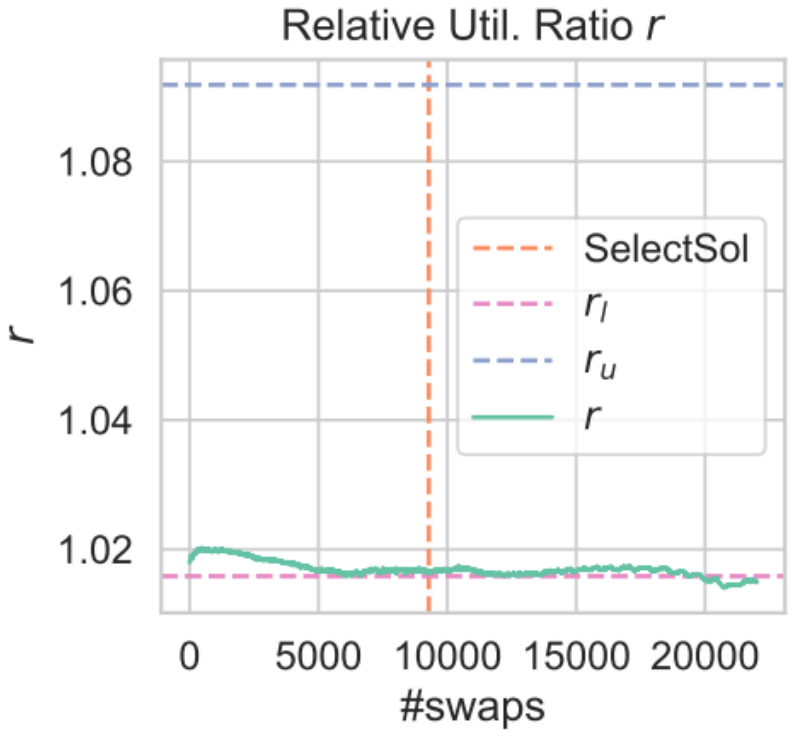


Experimental Results

History curves of wirelength-driven FM







Conclusion

- Coherent framework
  - Preserves placement quality
- Unified placement engine
  - Easy development
- Efficient algorithms
  - Low runtime overhead
- Flexibility
  - Works for other objectives

Thank you