Supported by Cadence Design Systems, Inc., NSF

Performance-Impact Limited Area Fill Synthesis

Yu Chen, Puneet Gupta, Andrew B. Kahng

(UCLA, UCSD)

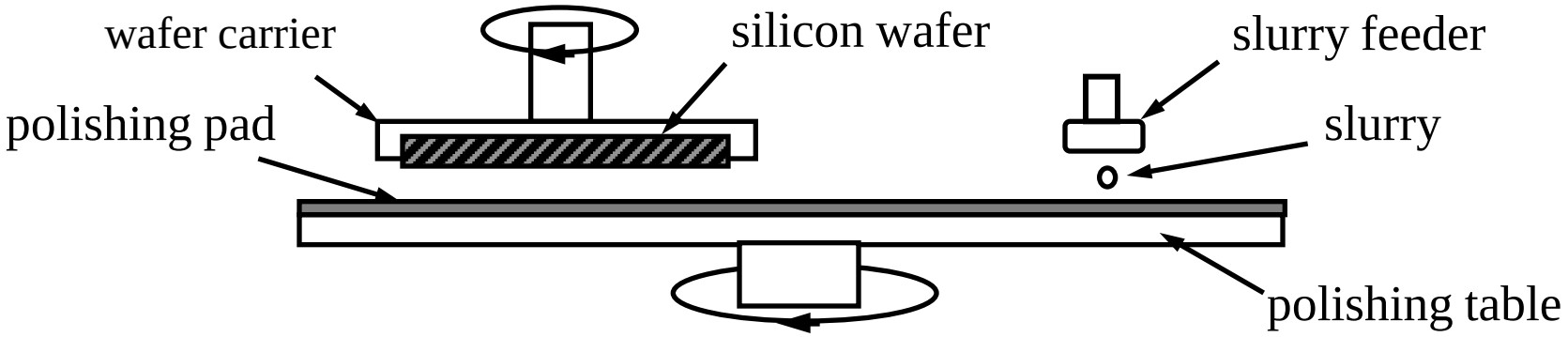
http://vlsicad.ucsd.edu

Outline

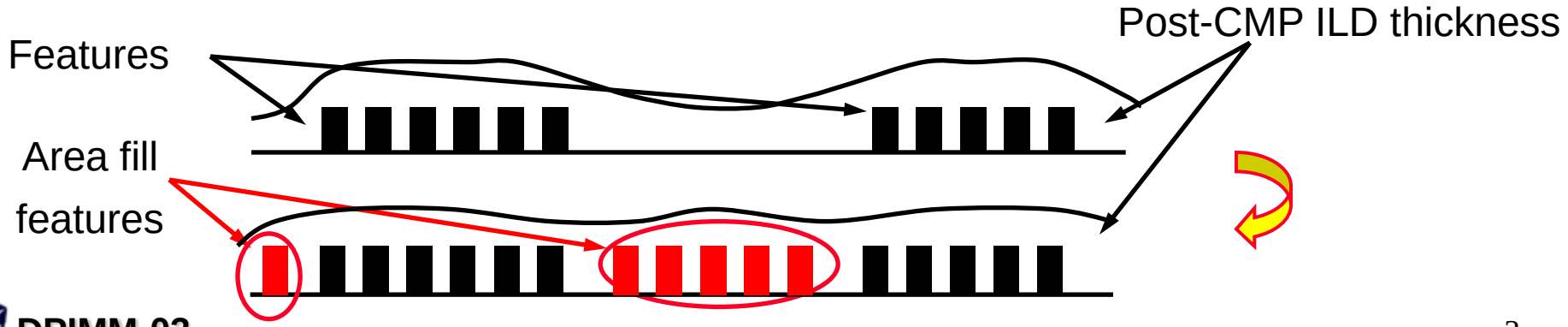
- Chemical Mechanical Planarization and Area Fill
- Performance-Impact Limited (PIL) Fill Problem
- Slack Site Column and Scan-Line Algorithm
- Linear Programming Approaches
- Greedy Method
- Computational Experiences
- Conclusion and Future Works

CMP & Area Fill

- Chemical-Mechanical Planarization (CMP)
  - Polishing pad wear, slurry composition, pad elasticity make this a very difficult process step

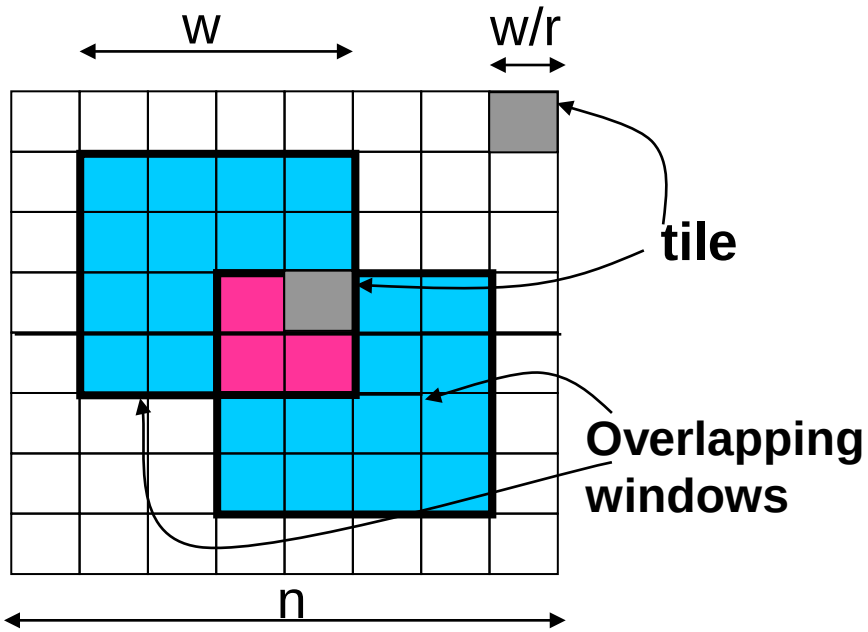


- Area fill feature insertion
  - Decreases local density variation
  - Decreases the ILD thickness variation after CMP



Fixed-Dissection Regime

- To make filling more tractable, monitor only fixed set of $w \times w$ windows
  - offset = w/r (example shown: w = 4, r = 4)
- Partition n x n layout into $nr/w \times nr/w$ fixed dissections
- Each $w \times w$ window is partitioned into r^2 *tiles*



Previous Objectives of Density Control

- Objective for Manufacture = **Min-Var**

  minimize window density variation

  subject to upper bound on window density

- Objective for Design = **Min-Fill**

  minimize total amount of added fill features

  subject to upper bound on window density variation

Performance-Impact Limited Area Fill

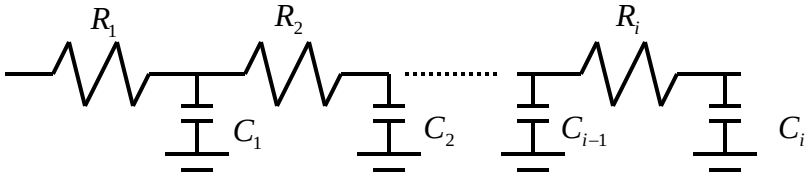
- Why?
  - Fill features insertion to reduce layout density variation
    - ⇨ change coupling capacitance
    - ⇨ change interconnect signal delay and crosstalk
  - No consideration on performance-impact during fill synthesis in current fill methods
- Problem Formulation
  - Two objectives
    - minimizing layout density variation; and
    - minimizing fill features' impact on circuit performance

    ⇨ objective + constraints
  - Minimum Delay with Fill Constraint problem

    Given a fixed-dissection routed layout and the design rule for floating square fill features, insert a prescribed amount of fill in each tile such that the performance impact is minimized

Capacitance and Delay Models

- Interconnect capacitance consists of
  - Overlap Cap. C_a : surface overlap of two conductors
  - Coupling Cap. C_{lt} : two parallel conductors on the same place
  - Fringe Cap. C_{fr} : between two conductors on different places
- Mainly consider fill impact on coupling capacitance
- Elmore Delay Model

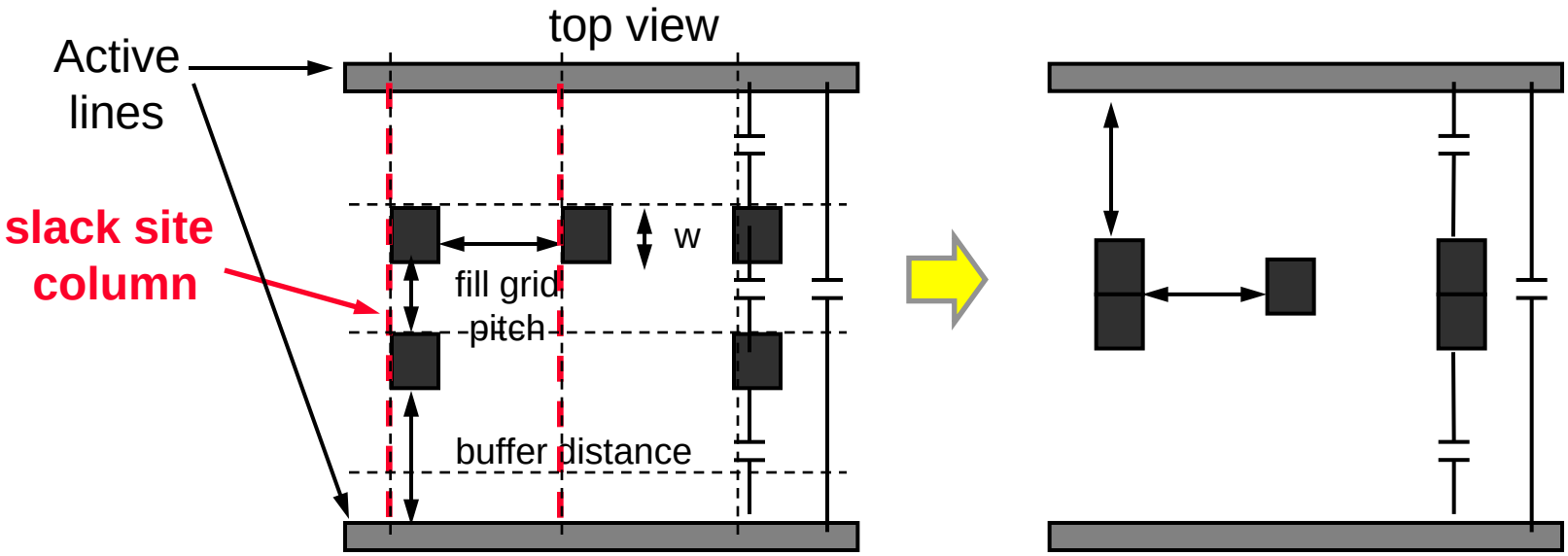


$$\Delta \tau_i = \Delta C_i \cdot \sum_{j=1}^{i} R_j$$

  - First moment of impulse response for a distributed RC interconnect
  - Enjoys additivity property with respect to capacitance along any source-sink path

Slack Site Column

- Grid layout with fill feature size into sites
- Slack Site Column: column of available sites for fill features between active lines

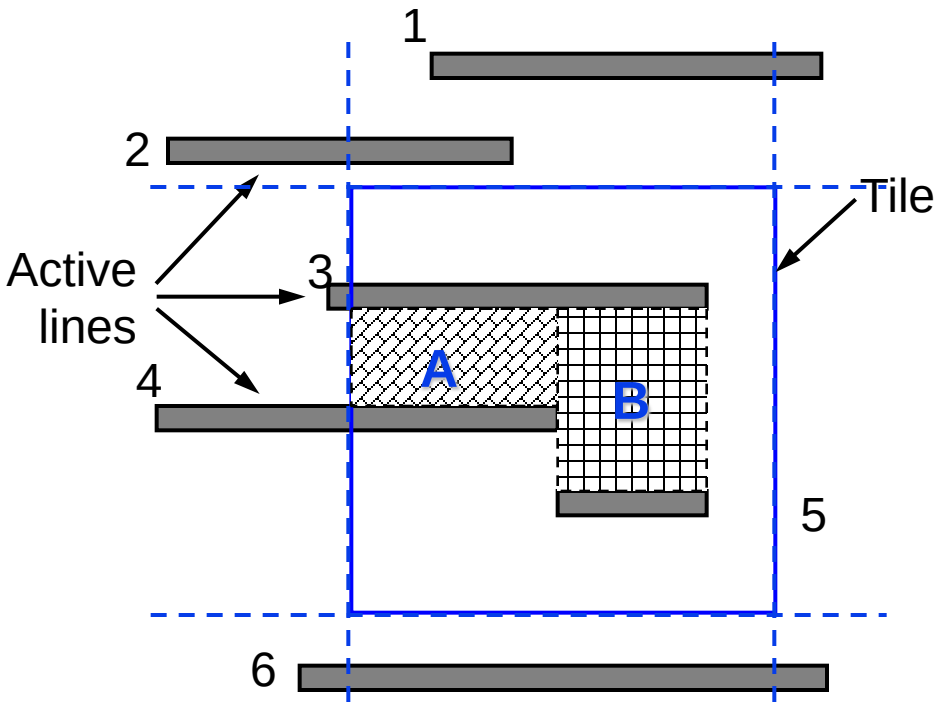


- Per-unit capacitance between two active lines separated by distance d, with m fill features in one column:

$$Cap = \frac{\varepsilon_0 \cdot \varepsilon_r \cdot a}{d - m \cdot w} \approx \frac{\varepsilon_0 \cdot \varepsilon_r \cdot a}{d} + \frac{\varepsilon_0 \cdot \varepsilon_r \cdot a \cdot m \cdot w}{d^2}$$

Slack Site Column Definition I

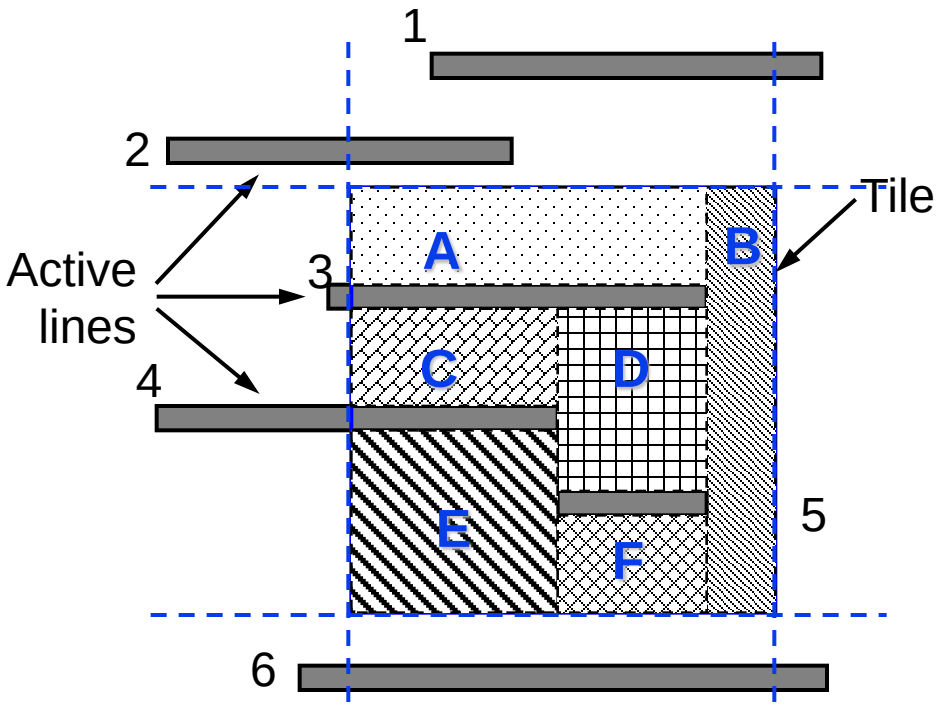
- Slack site columns between active lines within tile
- Drawback:
  - too much slack space cannot be used for fill insertion



only 2 type-I
slack blocks

Slack Site Column Definition II

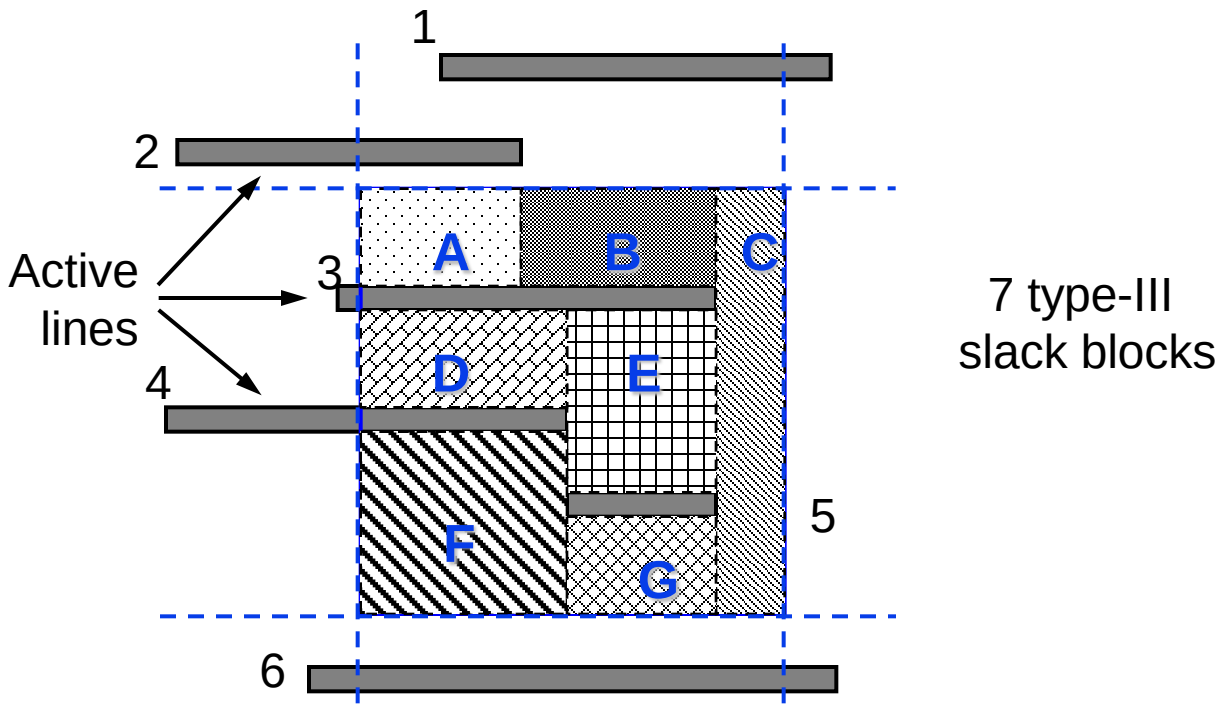
- Slack site columns between active lines within tile and tile boundaries
- Drawback:
  - insert fill features into slack columns without consideration of associated active lines (e.g., block B)
  - inaccurate estimation of delay impact



6 type-II
slack blocks

Slack Site Column Definition III

- Slack site columns between boundaries and active lines in adjacent tiles
- Advantage
  - Possible impact of fill feature at any position can be considered
- Scan-line algorithm: Scan whole layout from bottom (left) boundary for horizontal (vertical) routing direction



ILP Approaches I

- Based on linear approximation for coupling cap. due to fill features
- Objective: minimize weighted incremental delay due to fill features

Minimize: $\sum_{l=1}^{L} W_l \cdot \Delta\tau_l$ W_l : num of downstream sinks of segment l

- Constraints:

$$F = \sum m_k$$ for all overlapping columns

covered slack sites = # fill features

$$Cap_k = \frac{\varepsilon_0 \cdot \varepsilon_r \cdot a \cdot w}{d_k^2} \cdot m_k$$ for each column

Incremental cap. due to m_k features in each column

$$\Delta\tau_l = \sum_{k=1}^{l} Cap_k \cdot (R_l + \sum_{j=1}^{k} r_{l_j})$$ for each segment

Total Elmore delay increment

$$0 \le m_k \le C_k$$ # covered slack sites ≤ column size

ILP Approaches II

- Based on pre-built lookup table for coupling capacitance calculation
  - Fill features have the same shape
  - Potential num of fill features in each column is limited
  - Other parameters are constant for the whole layout

- Calculate coupling capacitance based on LUT

Binary: $$m_{k_n} = \begin{cases} 1 & m_k = n \\ 0 & otherwise \end{cases}$$

$$\sum_{n=1}^{C_k} m_{k_n} = 1 \qquad \text{for each column}$$

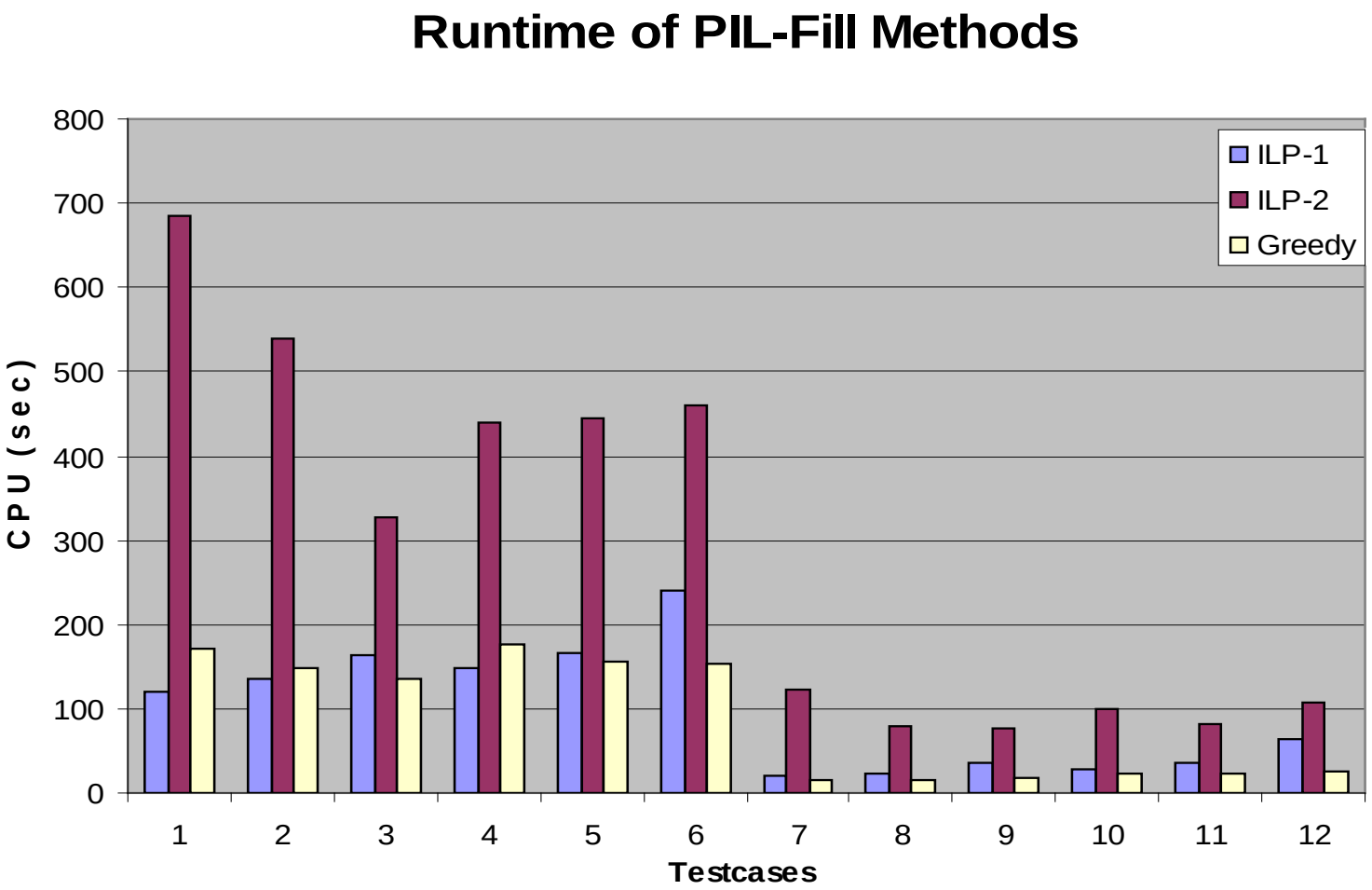
$$Cap_k = \sum_{n=1}^{C_k} f(n, d_k) \cdot m_{k_n} \qquad \text{for each column}$$

integer: $$0 \le m_k \le C_k$$

Greedy Method for PIL Fill

- Run standard fill synthesis to get *#FillFeatures* for each tile
- For each net N_i Do
  - Find intersection with each tile
  - Calculate entry resistances of net N_i in its intersecting tiles
- Get slack columns by scan-line algorithm
- For each tile Do
  - For each slack column Do
    - calculate induced coupling capacitance of slack column
  - Sort all slack columns according to its corresponding delay
  - While *#FillFeatures* > 0 Do
    - select slack column with minimum corresponding delay
    - insert C_k features into the slack column

Computational Experience



- PIL-Fill methods can reduce total delay increase by up to 90%
- LUT based ILP has best performance but longest runtime

Conclusions and Future Research

- Contributions:
  - Approximation for performance impact due to area fill insertion
  - First formulation for Performance-Impact Limited Fill problem
  - Integer Linear Programming based approaches
  - Greedy method

- Future Works
  - PIL Fill with given capacitance budgets for each net
  - Other PIL Fill formulations
    - Min density variation while obeying upper bound on timing impact

Thank You!