

A Framework for Systematic Evaluation and Exploration of Design Rules

Rani S. Ghaida* and Prof. Puneet Gupta
EE Dept., University of California, Los Angeles
(rani@ee.ucla.edu), (puneet@ee.ucla.edu)

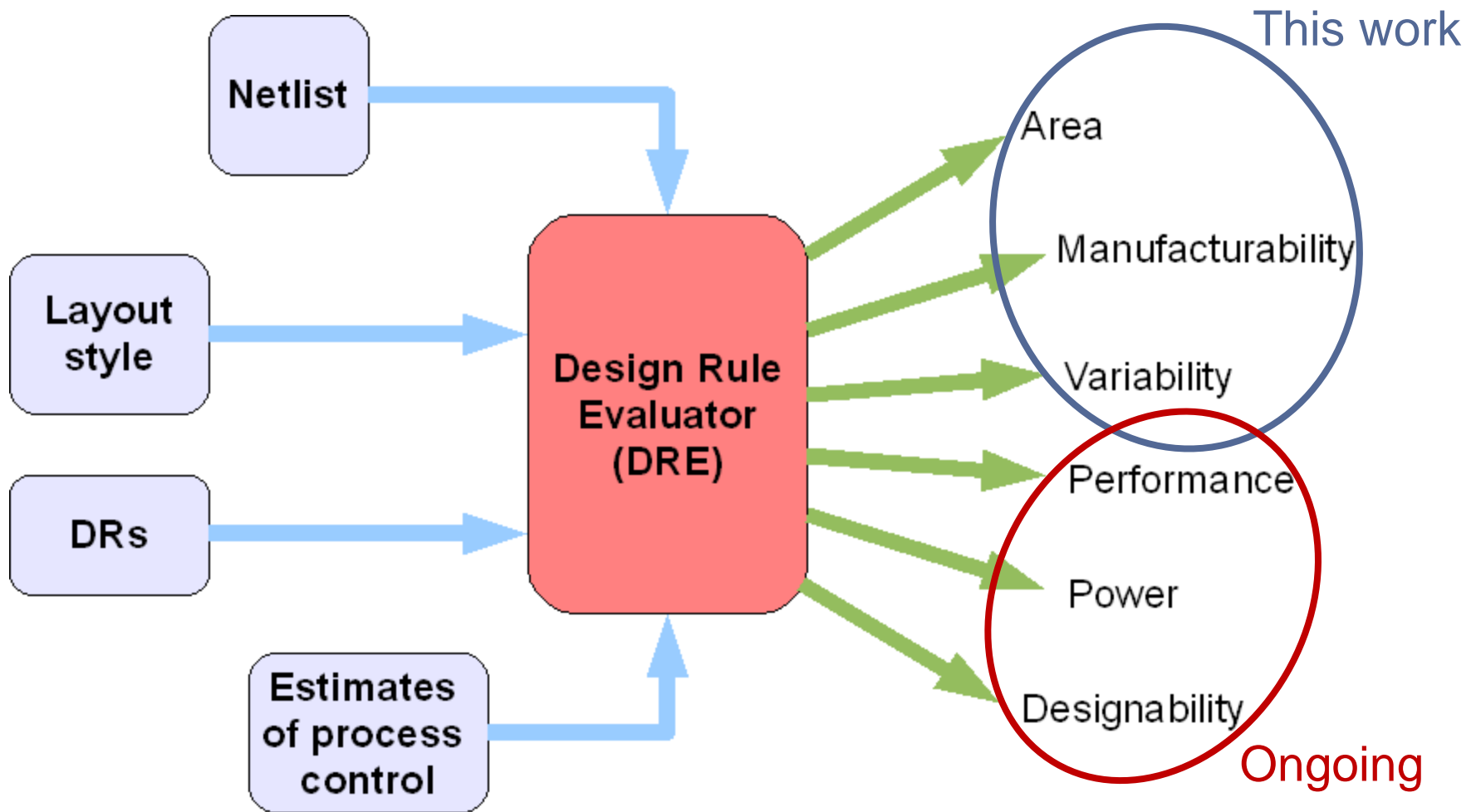
Work partly supported by IMPACT, SRC, and NSF.

NanoCAD Lab

Motivation

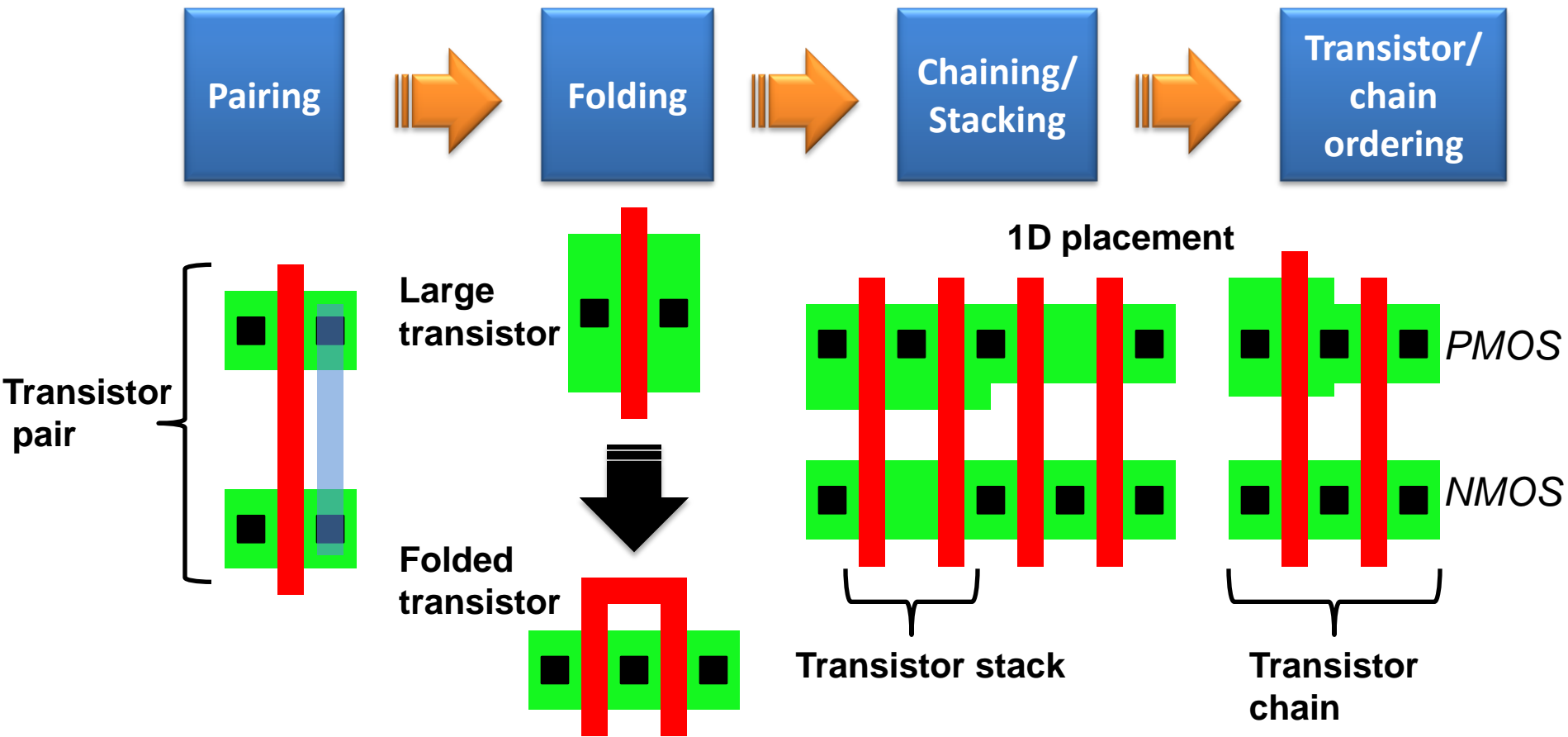
- Industry is faced with many technology options for scaling to every next node
- DRs being the biggest design-relevant quality metric for a technology
 - Evaluating DRs is absolutely necessary to decide on technology choices
- Automated and systematic cell-level DR evaluator
 - focus on *early* evaluation of layout-methodologies and DRs *before* exact process and design technologies are known
 - avoid explicit simulation or excessive reliance on accurate models

Overview of DR Evaluator



- Fast layout estimation ← *Fast topology generation + congestion estimation*

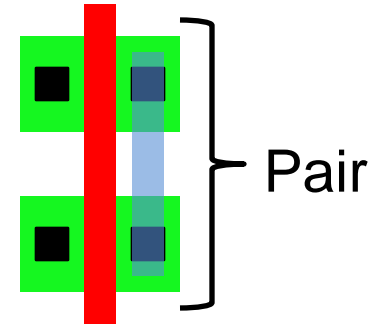
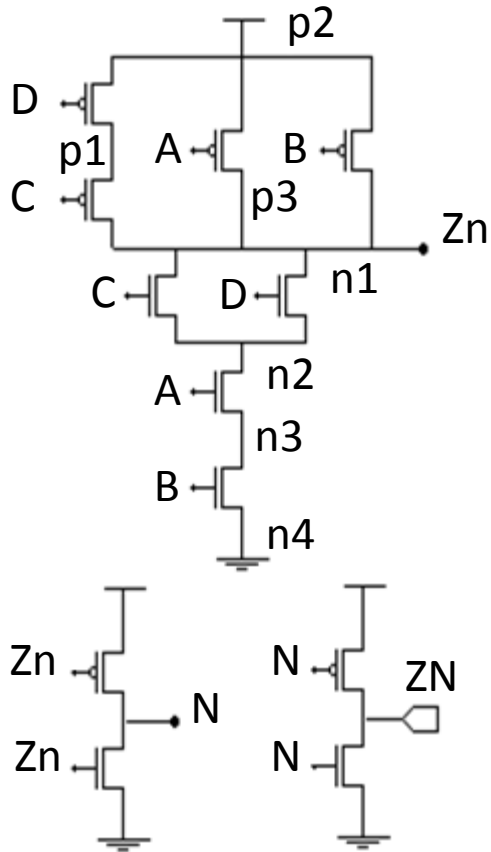
Assumptions and Flow



- CMOS circuits with dual transistors, multiple outputs, any transistor size, 1D transistor placement, i.e. on same row
- Intra-cell routing in poly and M1 layers only (M2 ongoing)

Example – Pairing

OAI211_X4 Netlist

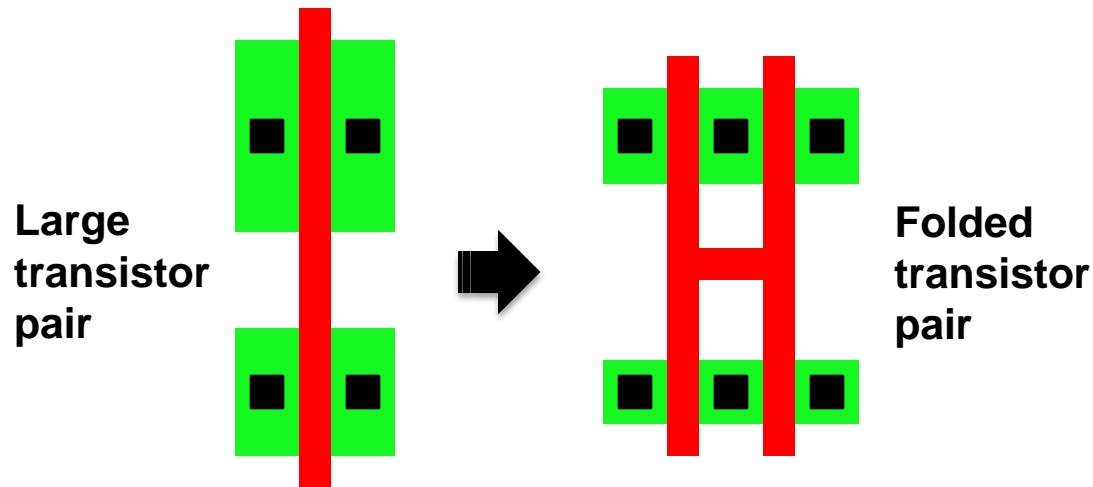


Pairing score table

	TN1	TN2	TN3	TN4	TN5	TN6
TP1	5.9	1.4	0	0	0	0
TP2	0	5	0	0	0	0
TP3	1	1	4.9	0	0	0
TP4	1	1	0	5.6	0	0
TP5	0	0	0	0	6.6	1.1
TP6	0	0	0	0	0.7	6.2

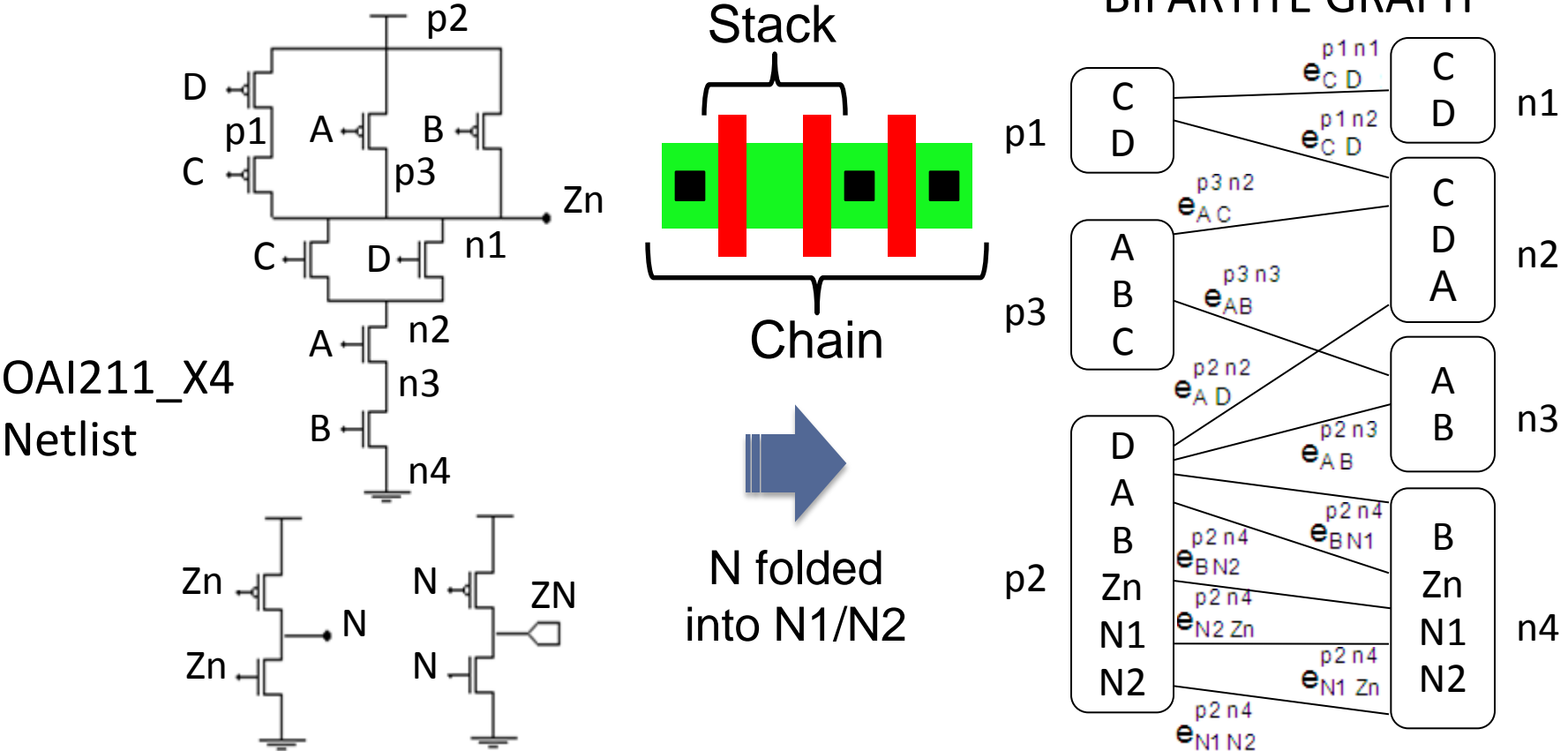
- Matching problem solved optimally with *Hungarian algorithm*

Example – Folding



- Given fixed cell height (as a rule), exhaustive search to find optimal p/n transistor folding sizes.
- Transistor pairs with a larger height are then folded into multiple pairs

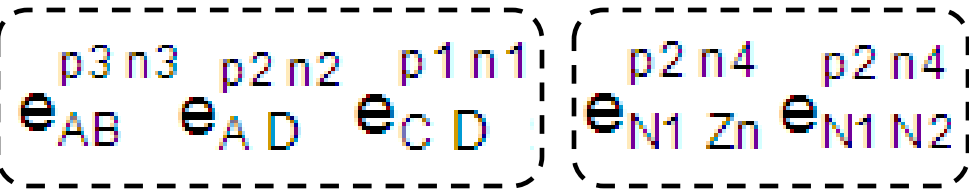
Example – Chaining



- Vertices → nodes in netlist and contain pairs the node connects to
- Edges → possible diffusion sharing between pairs
- If large # of folds, cluster into groups and treat each as single pair
- Chaining: find maximum set of compatible edges

Example – Chaining and Ordering

OPTIMAL ABUTMENT

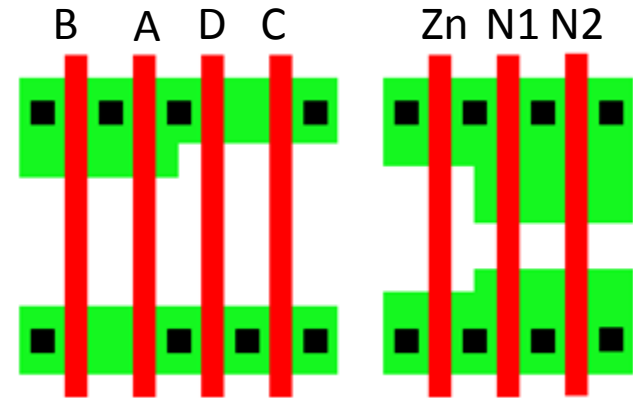


ORDERING

FLIPPING



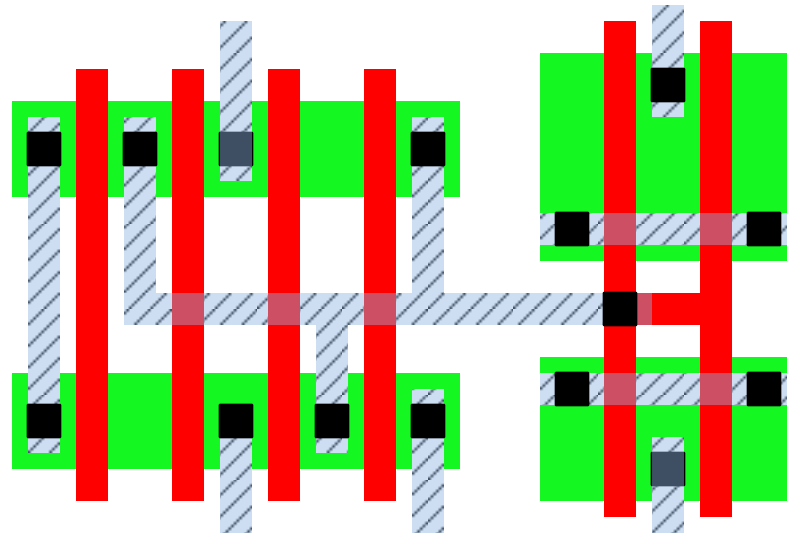
LAYOUT



2 chains/2 stacks

- **Chaining/Stacking:** C. Hwang *et al.*, TCAD 1990
 - Edges are sorted according to upper bound on the number of abutment after selection
 - Construct solutions in depth-first search with tree pruning
 - Only need to examine first several solutions to find optimal in most cases
- **Chain ordering:**
 - Min cut placement of chains with exhaustive search for small # of chains
 - For large # of chains, partition with FM algorithm and run exhaustive search to order partitions and chains within partitions
 - Chains are possibly flipped to minimize WL

Routing and Congestion Estimation



- M1 wiring for S/D-to-S/D and gate connections that cannot be on poly using single-trunk Steiner tree
 - horizontal trunk in center of cell
- Estimate congestion of vertical/horizontal tracks based on wire length and blockage by wires in orthogonal direction

$$C = \text{Occupied Track-Length} / \text{Available TL}$$

$$= (\text{WL} + \text{Blocked TL}) / \text{Available TL}$$

$$\text{Blocked TL} = \text{Blockage}_{\text{Actual}} + \text{Blockage}_{\text{Orthog}}$$

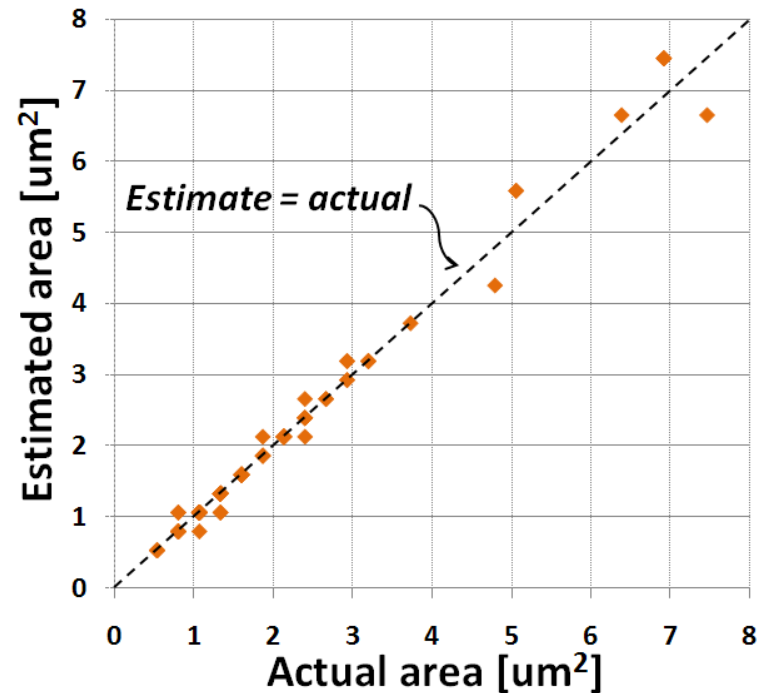
M1 Area Estimation

- If $C > C_{th}$ → increase cell-area to accommodate M1 wiring
- C_{th} captures routing efficiency, I/O pin accessing, and congestion:

$$C_{th} = \alpha + \left| \frac{U_x - U_y}{U_x + U_y} \right| \times \beta - \gamma$$

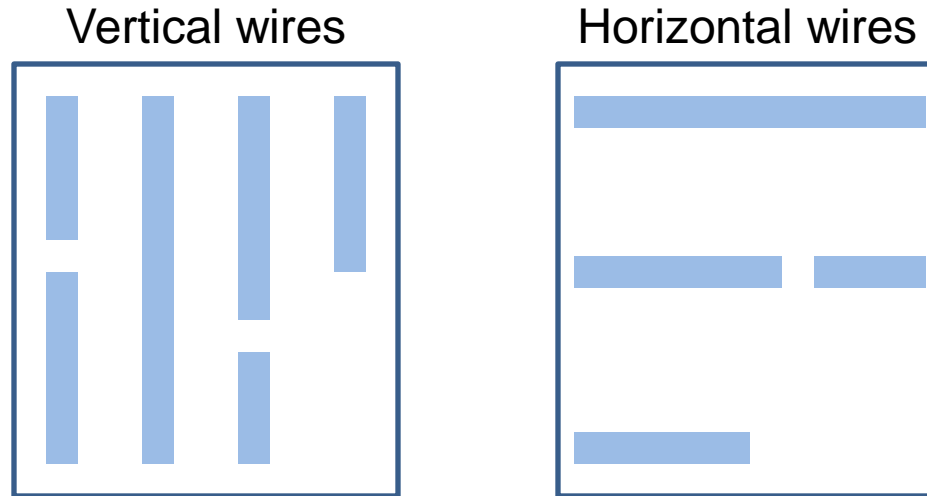
- α and β determined empirically from actual cells from previous generation or trial routes
- γ is for I/O pin-access requirement and is specified by user
- U is utilization w/o considering blockage from orthogonal wiring
- When $U_x \approx 0$, $\left| \frac{U_x - U_y}{U_x + U_y} \right| \rightarrow 1 \Rightarrow C_{th}$ larger
- When $U_x \approx U_y$, $\left| \frac{U_x - U_y}{U_x + U_y} \right| \rightarrow 0 \Rightarrow C_{th}$ smaller

Validation of Area Estimation and Runtime



- Layout of Nangate cell-library (104 cells) were estimated
 - Area estimated with 2.4% error on average
 - Runtime of evaluation procedure is 20 minutes real time on a 2GHz clock speed and 2MB cache processor
- Easily parallelizable with no overhead since cell topology generation is independent

Evaluation of Manufacturability



- Probability of survival (POS) from:
 - Overlay of Poly/M1/Active to contacts and poly-to-active (normal distribution)
 - Contact-hole failures
 - Particle defects: place wires on equally spaced tracks, use a compact model for CAA for M1/poly/contact shorts/opens and gate-to-contact shorts. Example for M1/poly wires:

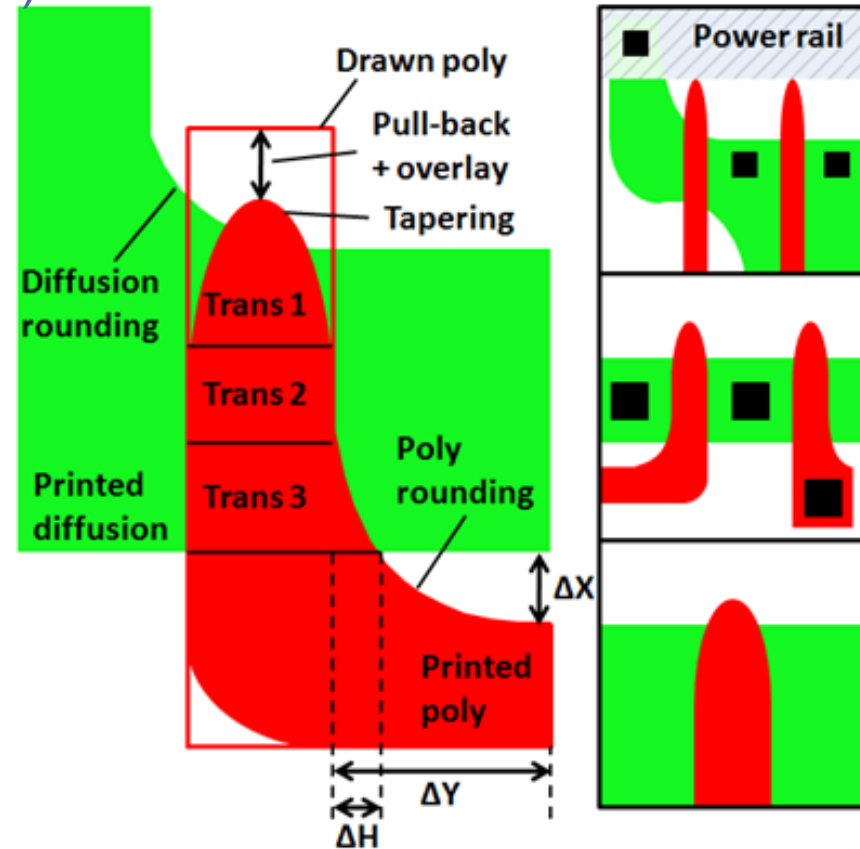
Evaluation of Variability

- We consider the sources:
 - diffusion and poly imperfections under average overlay error and line-end pullback (corner-rounding, line-end tapering)
 - CD variability (using distribution)

- Variability index is the total change in drive current

$$\Delta\left(\frac{W}{L}\right) = \frac{\sum_{all\ gates} \left| \Delta\left(\frac{W}{L}\right)_i \right|}{\left(\frac{W_{tot}}{L}\right)_{ideal}}$$

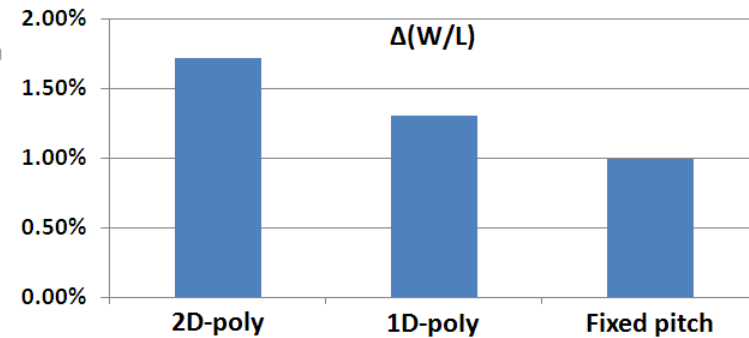
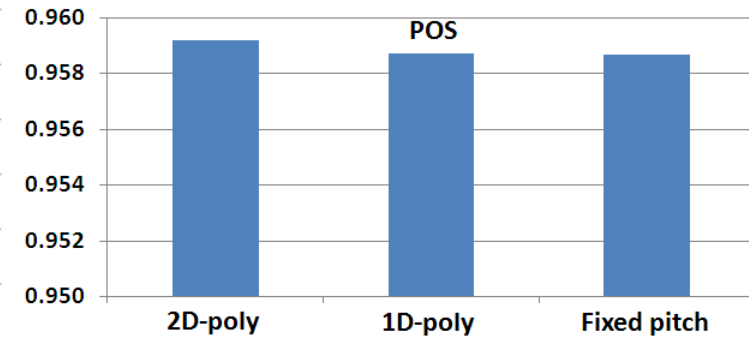
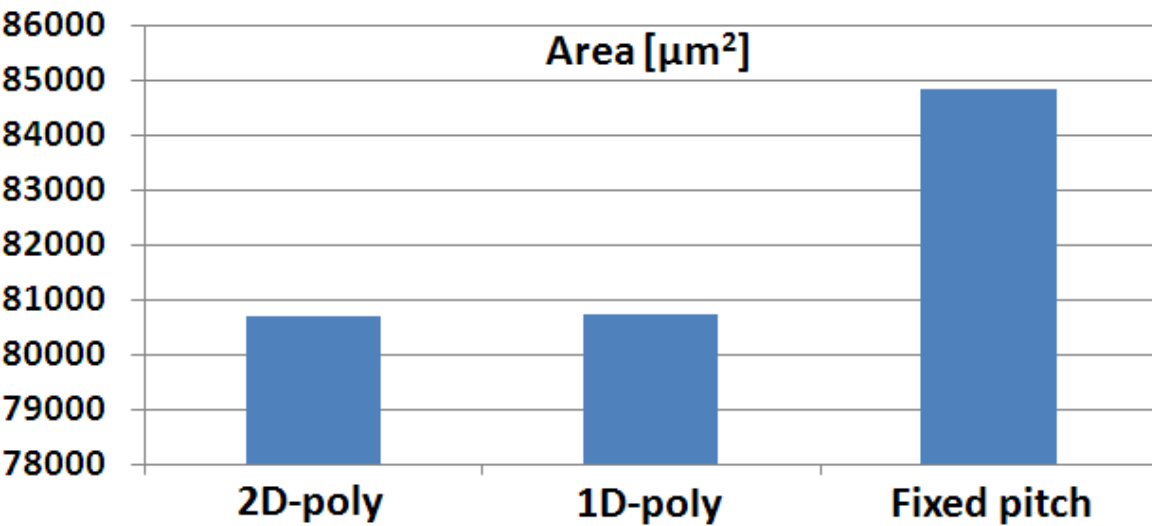
where i is the source of variability



Experimental Setup

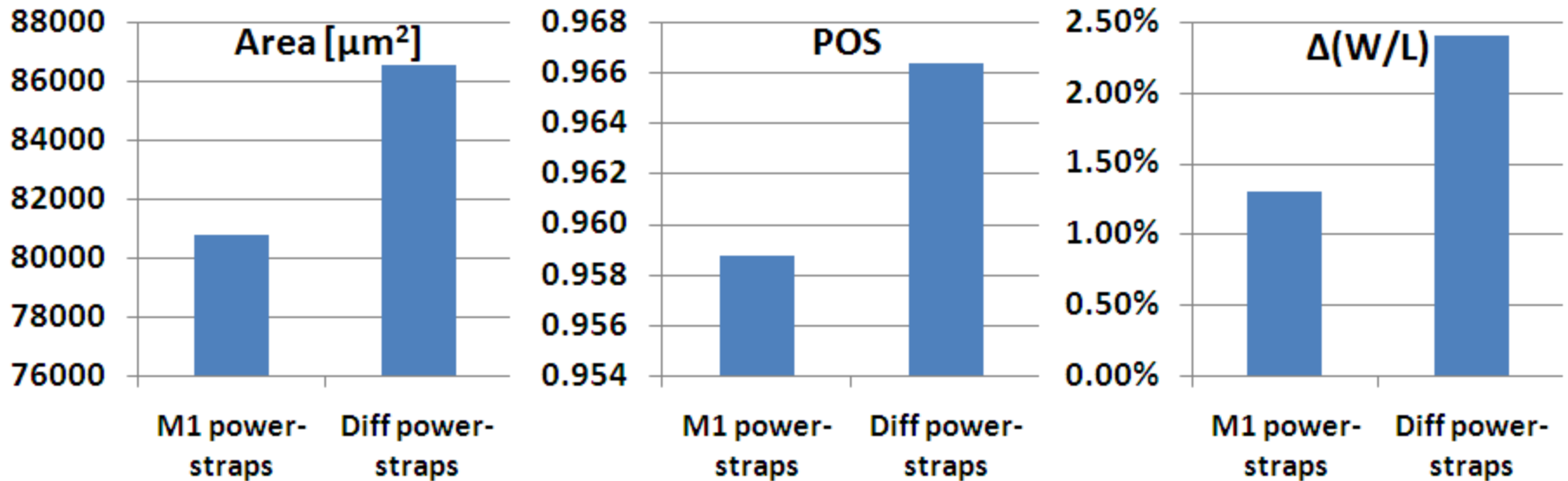
- 45nm FreePDK and 65nm process from a commercial vendor
- Benchmark designs varying from 4K to 43K cells synthesized using Nangate 45nm Open Cell Lib (scaled for 65nm testing)
- POS values normalized to a 10x10mm chip-area
- Baseline experiment with:
 - 1D-poly (non-fixed pitch)
 - M1 power-straps
 - 9-track cell-height

Evaluation of Poly-Restrictions



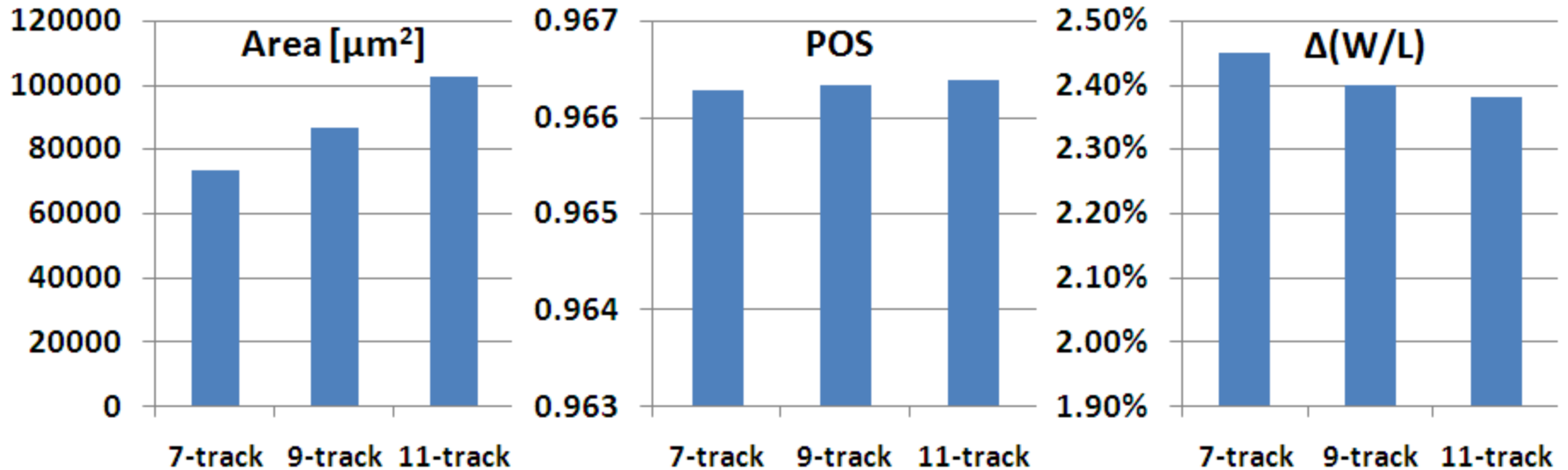
- 2D vs. 1D-poly
 - Almost identical cell-area due to pairing and small overhead for gate-alignment according to FreePDK DRs
 - 32% less variability with 1D-poly
- Fixed vs. non-fixed pitch 1D-poly
 - 23% less variability
 - 5% area overhead because min = contacted gate pitch

Evaluation of Power-Strap Styles



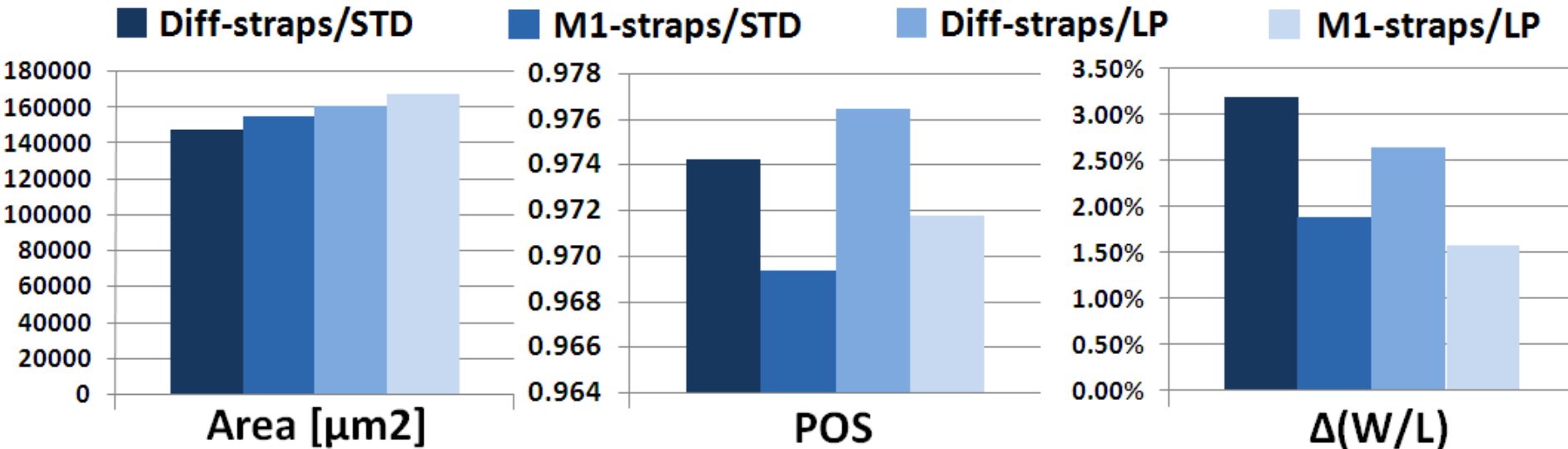
- 7% area overhead with diff straps (not true for small cell-height)
 - Specific to FreePDK, opposite results for 65nm commercial process
- 84% larger variability with diff-power straps
 - diffusion-rounding is dominant in tested cells
- Manufacturability benefits of diffusion power straps:
 - Gate-to-contact shorts are reduced
 - Contact redundancy for power connections on power rail (no cost)

Evaluation of Cell-Height



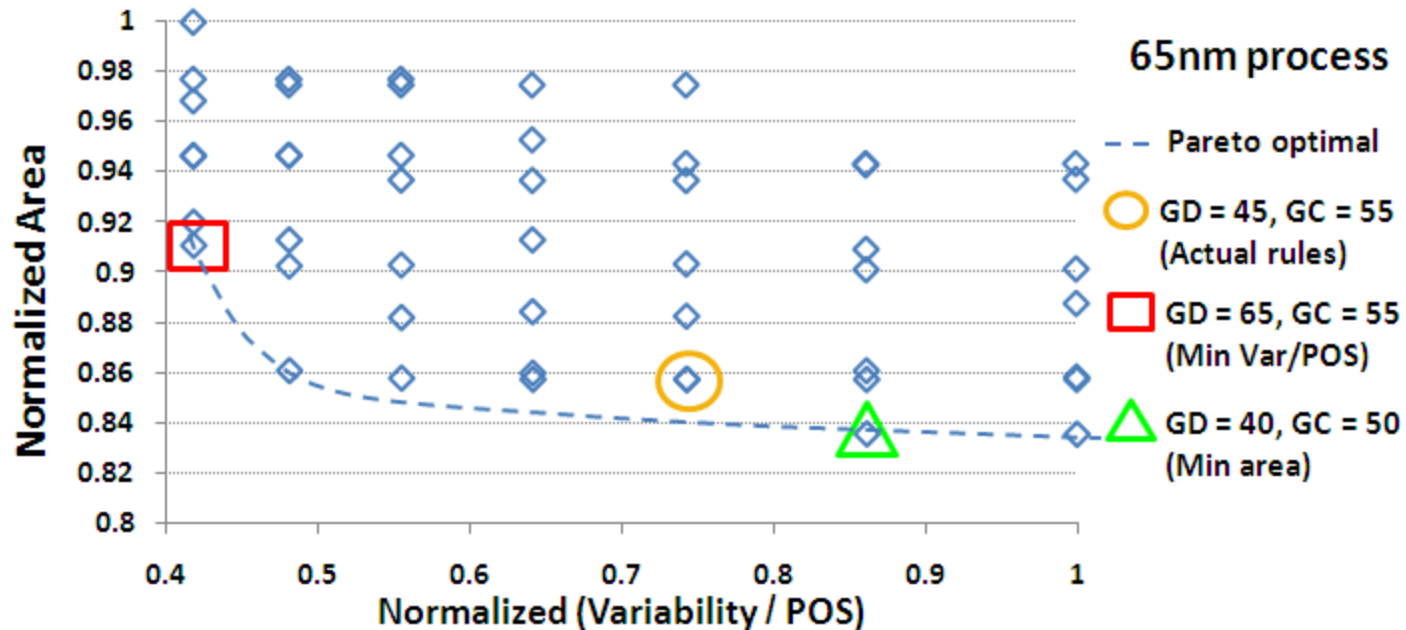
- Minor effect of cell-height decision on variability
 - poly rounding and line-end tapering affected by cell-height decision are second-order sources of variability
- In general, smaller cell-height \rightarrow smaller area,
 - Not true for (large) cells in high-performance designs

Comparison of DRs from Different Processes



- Compare DRs of std and LP 65nm process from same vendor with diffusion/M1 power-strap style and 1D-poly patterning
 - LP better in terms of variability and manufacturability, but std process is more area-efficient (7.9% less area)

Exploration of Gate-Spacing Rules

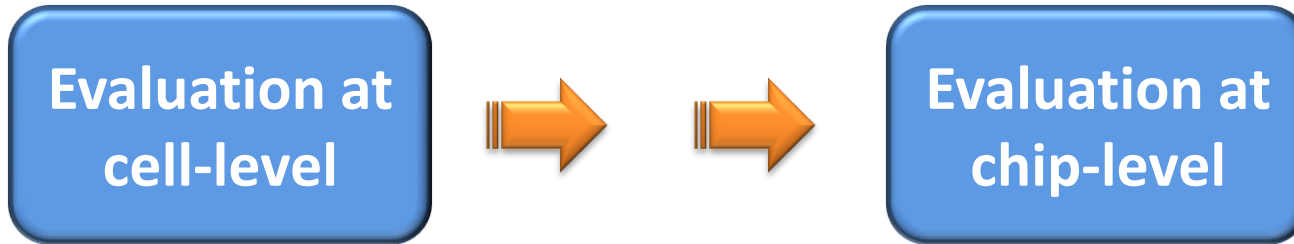


- Consider gate-to-diffusion (GD) and gate-to-contact (GC) rules in 65nm process and use diff power-straps and 1D-poly styles
 - Solution corresponding to process GD/GC actual values falls very near the Pareto optimal frontier
 - example shows the fidelity of our evaluation metrics and approach

Summary

- Flexible framework for:
 - Early co-evaluation of technologies, DRs, and cell library architectures *before* exact process and design technologies are known
 - Compare DRs from different processes
 - Can be used in DR optimization loops to narrow down on reasonable DR choices
- C++ source code available for download at <http://nanocad.ee.ucla.edu/Main/Projects>

Future Work



- Address DR effects on other layout characteristics including performance, power, and some notion of designability
- Introduce a 2D printability model (not based on field simulation)
- Extrapolate DR evaluation to the chip level and include intermediate and global metal/via layers
- Study interactions and tradeoffs of variability and area

Thank you!

Questions?

Backup Slides

Runtime Improvement in Chaining

- Problem: Runtime for cells with trans folded into large # of folds
- Special case for inverters/buffers
 - Detect them based on connectivity info in netlist
 - Force chaining of fingers of folded trans
 - Give preference to sharing output signal so that power signal is at the edge allowing its sharing with other transistors
 - Runtime improvement at no overhead
- If trans is folded into large # of fingers (user-specified)
 - Group fingers into multiple groups (user-specified)
 - Treat each group as a single pair during chaining
 - Unfold finger groups after chaining is complete
 - Runtime improvement at negligible overhead

Runtime Improvement in Chain-Ordering

- Problem:
 - Runtime of chain-ordering with exhaustive search in cells with large # of chains
- Let user specify limit on # of chains for exhaustive search (Lim)
- For cells with # of chains $>$ Lim:
 - Partition chains into groups with # of chains \leq Lim using FM algorithm to minimize connection cuts between partitions
 - Exhaustive search to order partitions
 - Exhaustive search to order chains within each partition separately while taking the order of partition into account

M1-Congestion Estimation

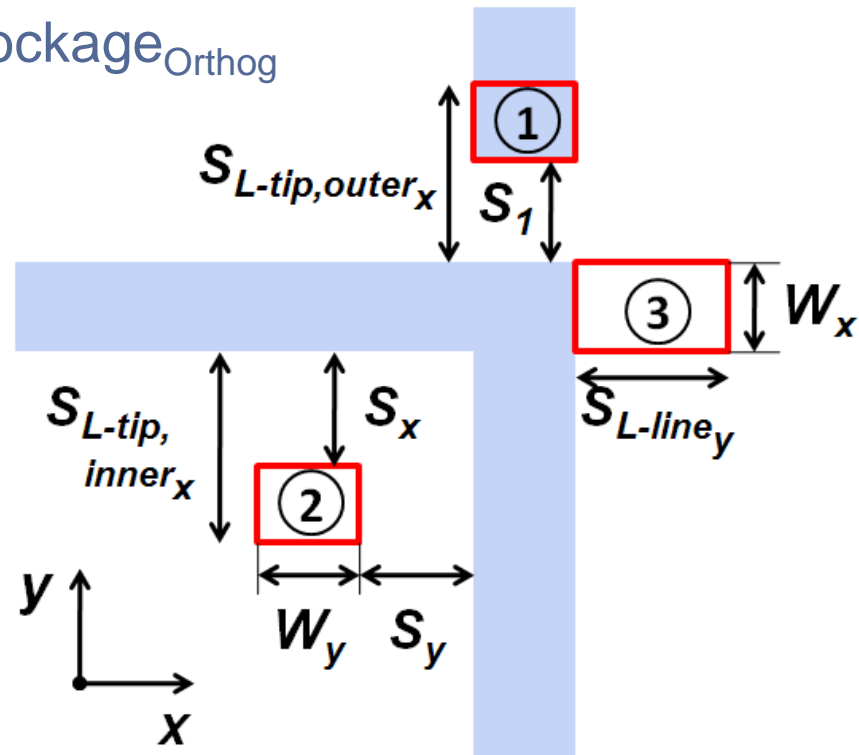
- Estimate congestion of vertical/horizontal tracks based on wire length and blockage by wires in orthogonal direction

$$C = \text{Occupied Track-Length} / \text{Available TL}$$

$$= (\text{WL} + \text{Blocked TL}) / \text{Available TL}$$

$$\text{Blocked TL} = \text{Blockage}_{\text{Actual}} + \text{Blockage}_{\text{Orthog}}$$

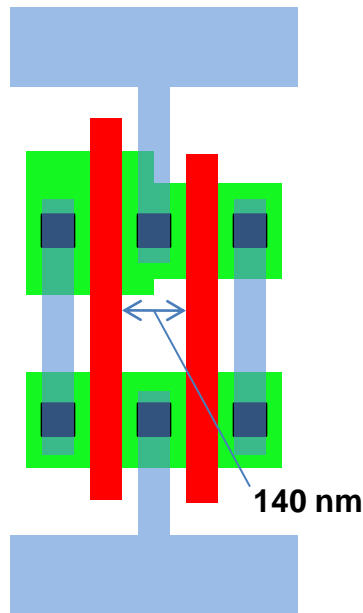
- Example of L-bend with tip facing outer corner:
 - (1) and (2) are actually blocked
 - (3) effectively increases wirelength in orthogonal direction



Outcomes Depend on Set of DRs

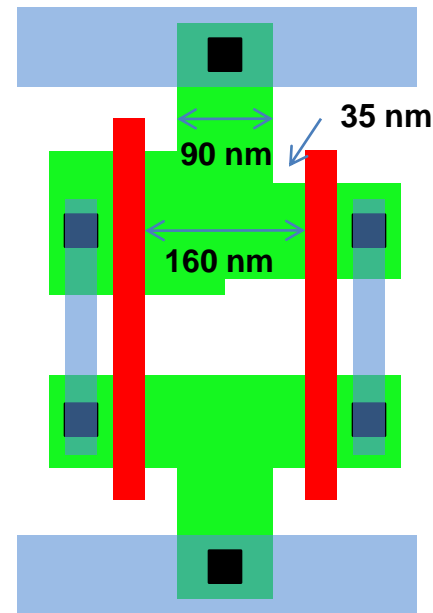
- Special characteristics of FreePDK DRs, e.g.:
 - Too small gate-to-contact spacing => huge effect on POS results for case of redundant contacts
 - Large area overhead for fixed gate-pitch with diffusion power connection

Power connection with M1



Contacted pitch = min pitch

Power connection with diffusion



Have to increase cell-width to next allowable pitch

UCLA_DRE Supported Rules

- Layout styles
 - Poly patterning (2D/limited/1D/fixed-pitch)
 - Power-straps (Metal/Active)
- DRs
 - Poly: LEE, LEG, gate-to-CA, gate-to-active, gate-pitch, etc...
 - Active: spacing, min width, extension beyond gate, etc...
 - CA: width, spacing, poly/active enclosure, CA to active, etc...
 - M1: width, 2D-spacing rules, overhang rules
 - M2/Via1 (ongoing): width, 2D-spacing rules
 - Mx in chip-level evaluation (ongoing)
 - Well (ongoing): active-to-well edge