UCLA

A Framework for Early and Systematic Evaluation of Design Rules

Rani S. Ghaida and Prof. Puneet Gupta

Motivation

Design Rule (DR)	**Area**	**Speed**	**Power**	**Manufacturability**	**Variability**
Fixed gate-pitch	**Bad**	**?**	**?**	**Good**	**Good**
Diffusion routing	**?**	**?**	**?**	**Bad**	**Bad**
Redundant contacts	**Bad**	**?**	**?**	**Good**	**Good**

- Design rules (DR) are the biggest design-relevant quality metric for a technology.
- Significantly more layout restrictions will be needed to enable future technologies

  → need systematic methods to evaluate the cost (design)/benefit (manufacturing) tradeoffs for DR/technology choices.

Goal

- Automated design rule evaluator in terms of cell area, speed, power, variability, and manufacturability at cell level and chip level.
  - Focus is on *early* co-evaluation/exploration of layout-methodology, technology and DRs → avoid explicit simulation or excessive reliance on accurate models
  - Code, scripts and flows to be made available for open use (Calibre/C++/OpenAccess)

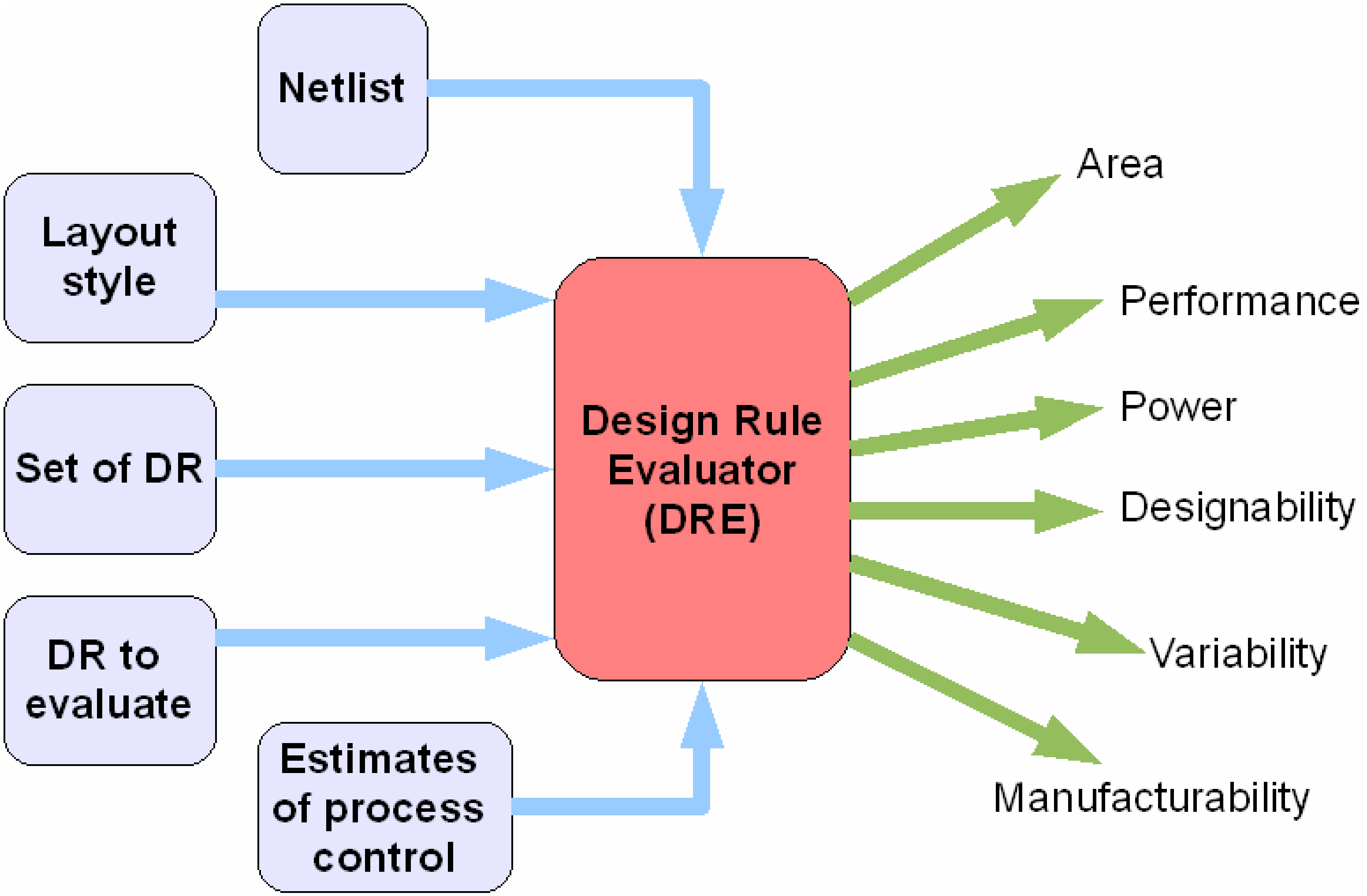
Short term goal

Evaluation at cell-level

Long term goal

Evaluation at chip-level

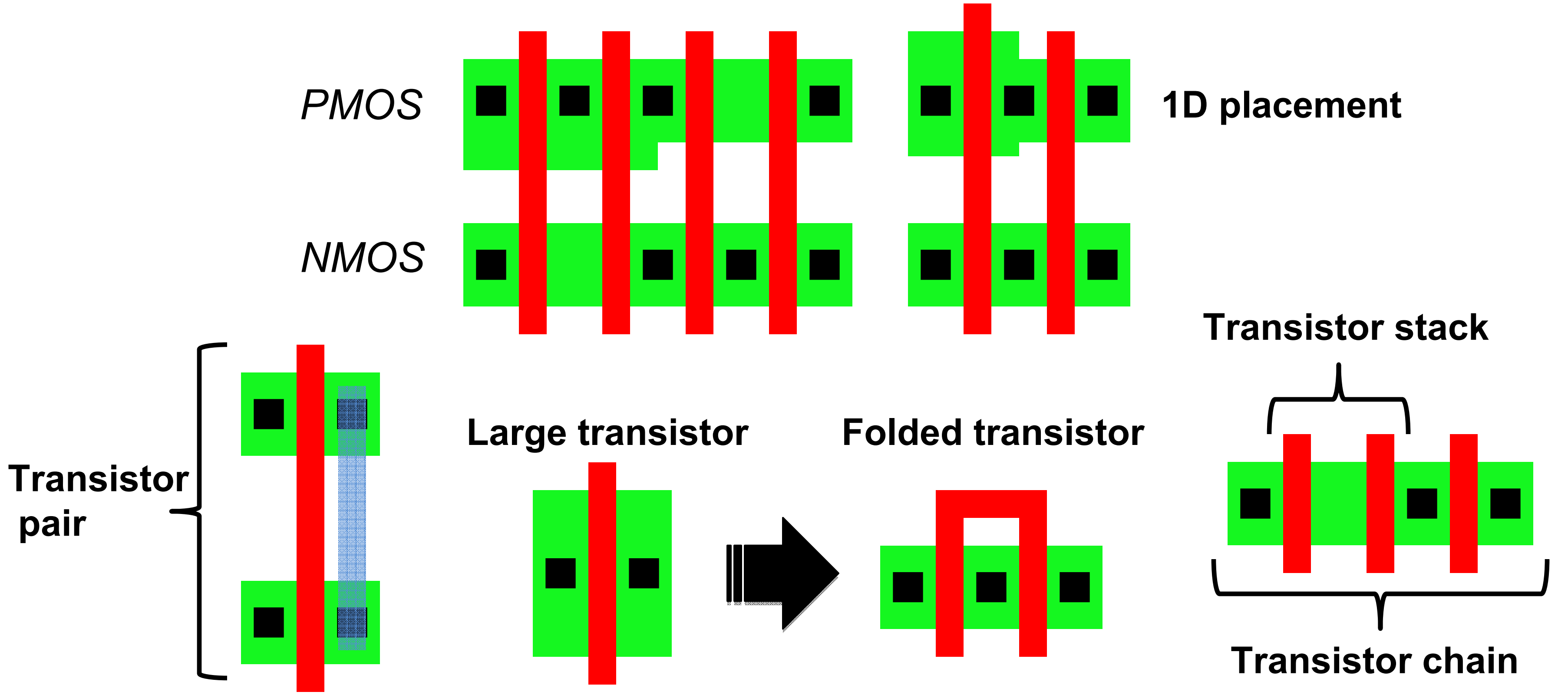
Overview of DR Evaluator



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

Assumptions and Definitions

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



Layout Topology Estimation

Pairing → Folding → Chaining/ Stacking → Transistor/ chain ordering

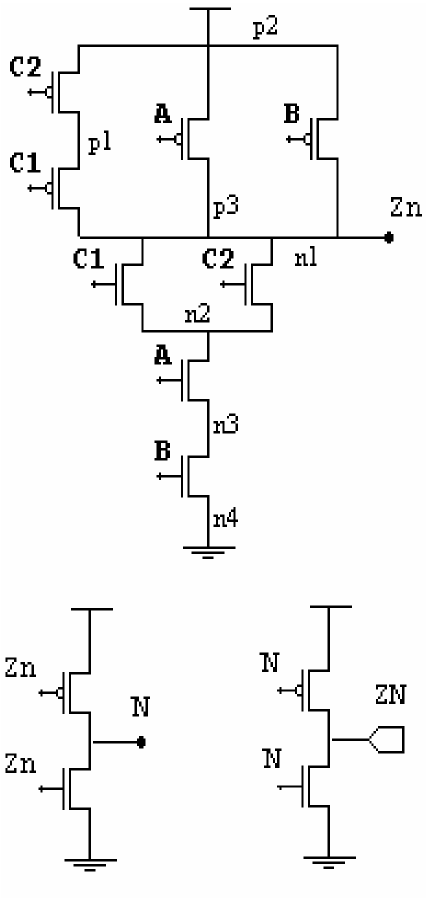
- ***Pairing*** → Matching problem solved with *Hungarian algorithm*.
- ***Folding*** → exhaustive search to find optimal p/n transistor folding sizes given fixed cell height. Transistors with larger height are then folded.
- ***Chaining/Stacking***: fast depth-first search algorithm with tree pruning [1] is used to find optimal chaining. Configurable preference is used to find solution with more transistor stacks.
- ***Ordering***: exhaustive search to find best order of chains. Then, chains are possibly flipped to minimize wire length.

[1] C. Hwang *et al.*, "A Fast Transistor-Chaining Algorithm for CMOS Cell Layout," 1990.

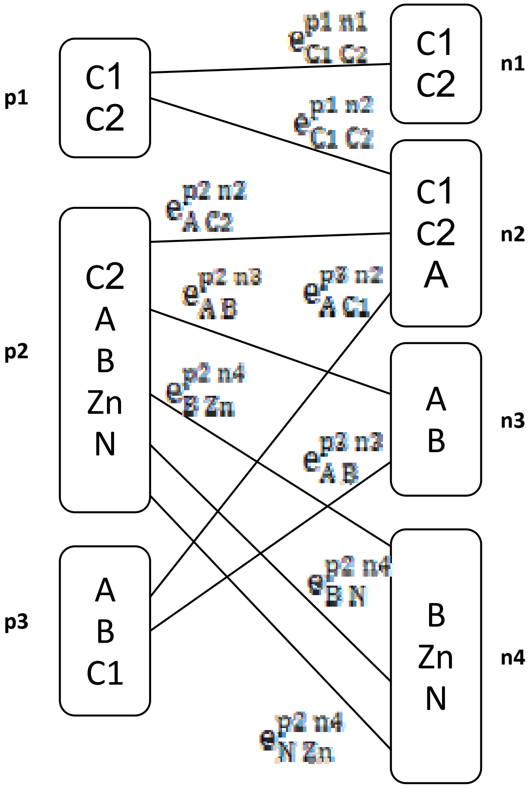
Topology Estimation – Example

Pairing + Folding

OAI211_X4



BIPARTITE GRAPH

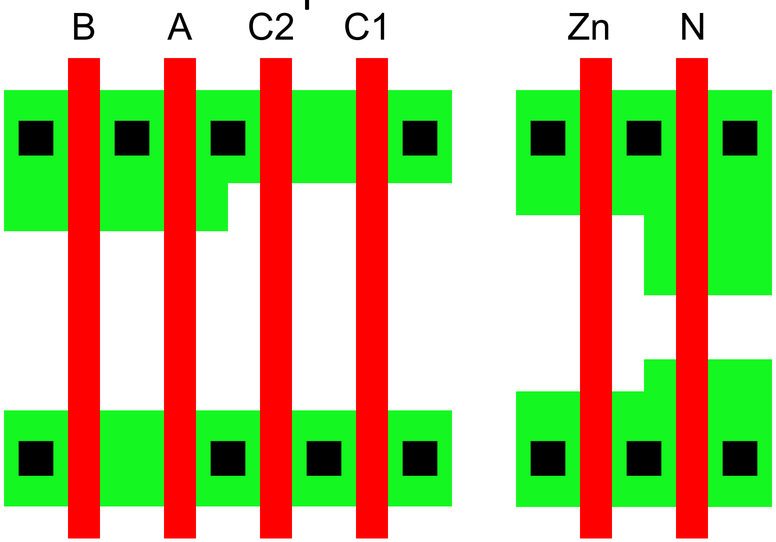


Optimal abutment

$e^{p1\,n1}_{C1\,C2}$, $e^{p2\,n4}_{N\,Zn}$, $e^{p3\,n3}_{A\,B}$, $e^{p2\,n2}_{A\,C2}$.

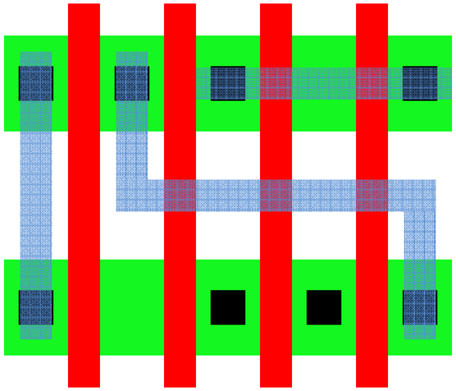
Chaining, stacking, ordering

LAYOUT



2 chains
1 stack in P
1 stack in N

M1-Congestion Estimation



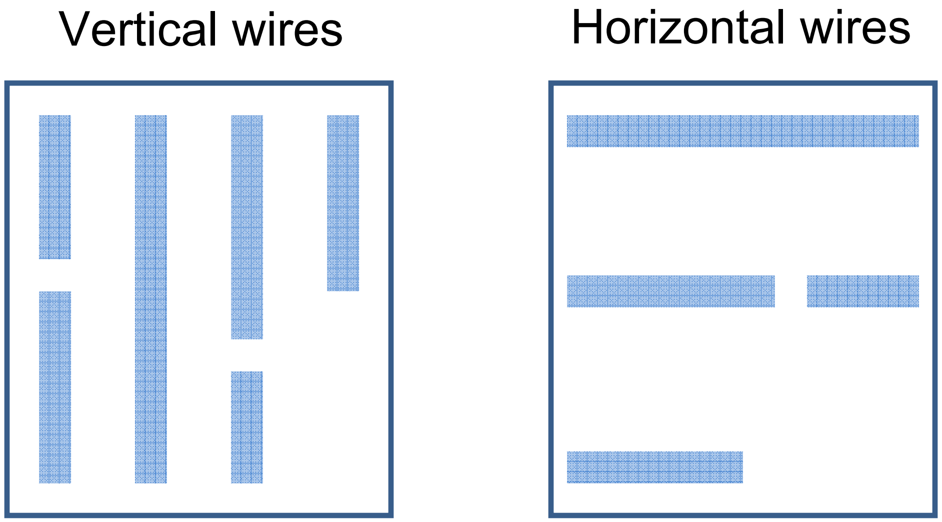
- M1 wiring for S/D-to-S/D connections and, also, gate connections in case of poly-congestion or restricted poly-routing
- Estimate congestions of vertical/horizontal tracks based on wire length and blockage by wires in orthogonal direction
- If $C > C_{th} \rightarrow$ increase cell-area to accommodate M1 wiring
- C_{th} captures routing efficiency, I/O pin accessing, and congestion:

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

- α and β determined empirically from actual cells, γ user-defined

Evaluation of Manufacturability

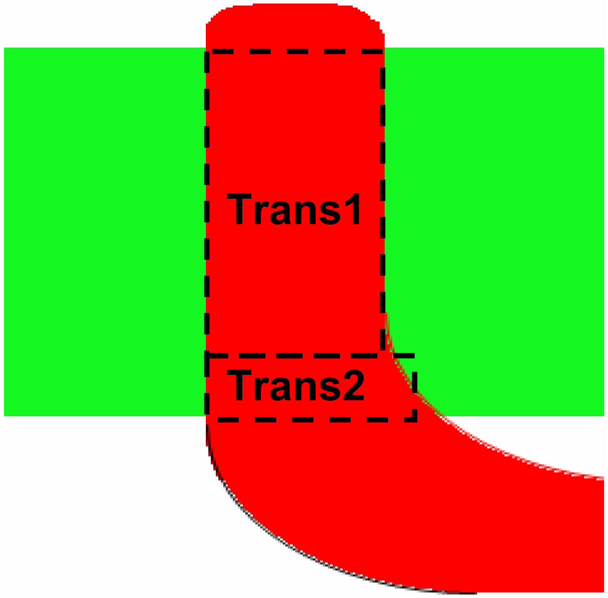
- Probability of survival (POS) from:
  - Overlay of Poly/M1/Active to contacts and poly-to-active ($3\sigma = 13$nm for 45nm node)
  - Contact-holes (0.00004 ppm failure rate).
  - Particle defects (2793 faults/m^2): place wires on equally spaced tracks, run CAA for M1/poly/contact shorts/opens and gate-to-contact shorts. Example for M1/poly wires:


Vertical wires

Horizontal wires

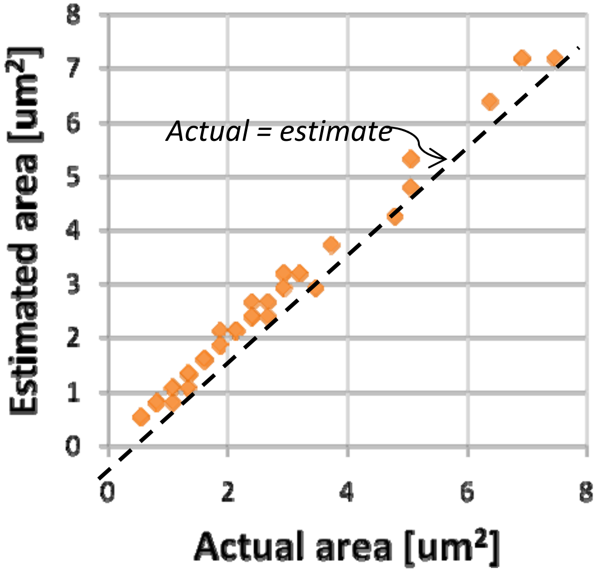
Evaluation of Variability

- $\Delta\frac{W}{L}$ from line-width variation (CDU), poly-rounding, diffusion-rounding, and line-end tapering (including overlay and pull-back).
- E.g. for poly-rounding, model corner-rounding and then approximate to multiple transistor slices in parallel:



Cell-Area Verification

- Layout of Nangate cell-library (107 cells) were estimated
  - Area estimated with 5.5% error on average
  - 65 out of 107 with exact estimate (discrete cell-width)



- Testing Setup:
  - Evaluate major debatable DRs and layout styles
  - FreePDK 45nm-DRs of Nangate cell-lib
  - MIPS benchmark circuit (17091 instances with 59 cell types) normalized to 10x10mm chip-area

Poly-Routing Study

Power conn.	Poly-routing			Limited poly-routing			No poly-routing		
	Area	POS	Δ(W/L)	Area	POS	Δ(W/L)	Area	POS	Δ(W/L)
Diff	33017	0.953	5.17%	33017	0.953	5.17%	33017	0.952	5.17%
M1	30520	0.942	3.28%	30950	0.942	3.28%	31471	0.942	3.28%

- ***M1 vs diff power connections***:
  - Area reduced by 7.6% and variability improved significantly (less diffusion rounding)
  - POS reduced slightly because power-contact redundancy implemented at no cost
- ***Power connections with diffusion***:
  - No area overhead for poly-routing restrictions (from low M1-congestion).
  - CDU improvement of 1D-poly not modeled
  - Variability with poly routing affected only for cells containing max size transistors which are not used in this design
  - POS reduced slightly when no-poly routing because additional contacts are necessary
- ***Power connections with M1***:
  - Area increased by 3.1% when no poly-routing (from high M1-congestion)
  - POS same. M1 more fault-tolerant than poly (good), but more contacts needed (bad)

Fixed Gate-Pitch Study

	Non-fixed pitch (no diff)			Fixed gate-pitch (no diff)			Non-fixed pitch (w/ diff)			Fixed gate-pitch (w/ diff)		
	Area	POS	Δ(W/L)	Area	POS	Δ(W/L)	Area	POS	Δ(W/L)	Area	POS	Δ(W/L)
MIPS	30520	0.942	3.28%	32006	0.942	3.28%	33017	0.953	5.17%	37946	0.956	3.65%

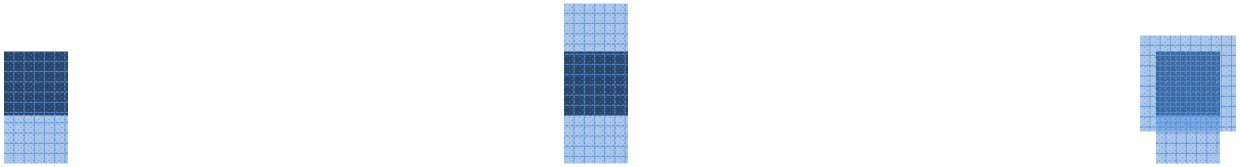
- Case of NO poly-routing in this study
- ***No diff-routing***: 4.9% area overhead, variability unchanged (not including CDU improvement), POS unchanged when fixed gate-pitch.
- ***With diff-routing***: ~15% area overhead, variability improved significantly (not including CDU improvement), POS improved slightly for fixed pitch.
- Better to have M1 power-connections if fixed gate-pitch is adopted.

Contact Redundancy Study

	No redundant contacts			Redundant diffusion contacts			Redundant diffusion and poly contacts		
	Area	**POS**	**Δ(W/L)**	**Area**	**POS**	**Δ(W/L)**	**Area**	**POS**	**Δ(W/L)**
MIPS	30520	0.942	3.28%	36419	0.940	3.28%	38714	0.950	3.28%

- Poly routing allowed but no diff-routing in this study.
- ***Redundancy of diff-contacts***:
  - 19.3% area overhead, POS reduced slightly (better for contact-holes failure but worse for gate-to-contact shorts).
- ***Redundancy of all contacts***:
  - assume poly-contacts can always fit
  - 26.8% area overhead, POS improved slightly.
- *Remark*: contact redundancy not always good for manufacturability as it increases wire length and probability of bridging.

Study of M1-Contact Overhang Rules



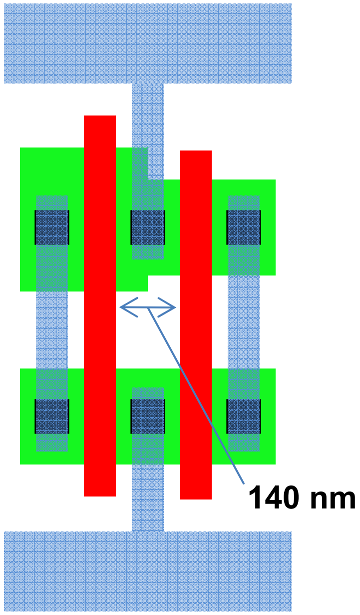
	No M1 contact overhang			2-sided M1 contact overhang (35nm)			All sides M1 contact overhang (15nm)		
	Area	**POS**	**Δ(W/L)**	**Area**	**POS**	**Δ(W/L)**	**Area**	**POS**	**Δ(W/L)**
MIPS	29984	0.942	3.28%	30520	0.942	3.28%	32006	0.942	3.28%

- Poly routing allowed but no diff-routing in this study.
- ***No M1-contact overhang***: 1.7% less area, same POS (less M1 short/opens but larger probability of failure from overlay error and line-end pull-back).
- ***All sides overhang***: 4.9% area overhead even with smaller overhang rule (15nm instead of 35).

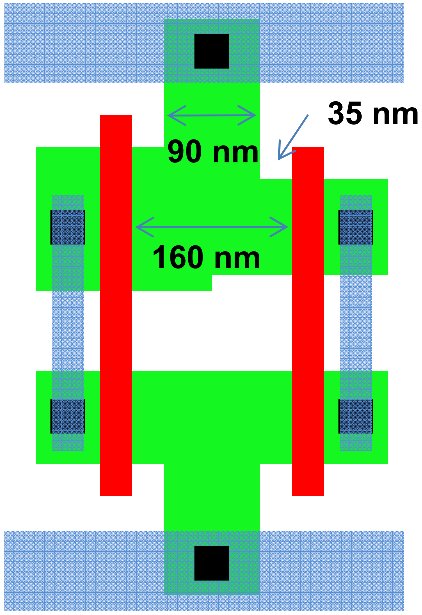
Outcomes Depend on Set of DRs

- Large overhead for fixed gate-pitch in case of diffusion routing
- Special characteristics of FreePDK DRs:
  - E.g. too small gate-to-contact spacing => huge effect on POS results for case of redundant contacts

Power connection with M1



Power connection with diffusion



Have to increase cell-width because of DRs

Summary & Future Work

- Flexible framework for co-evaluation/exploration of:
  - Technologies
  - DRs given a technology
  - Cell library architecture given DRs and technology
- Future work to extrapolate this evaluation to chip-level using buffer insertion and gate sizing as guidelines.