Synergistic Modeling & Optimization for Physical and Electrical DFM

David Z. Pan

Dept. of Electrical and Computer Engineering

The University of Texas at Austin

dpan@ece.utexas.edu

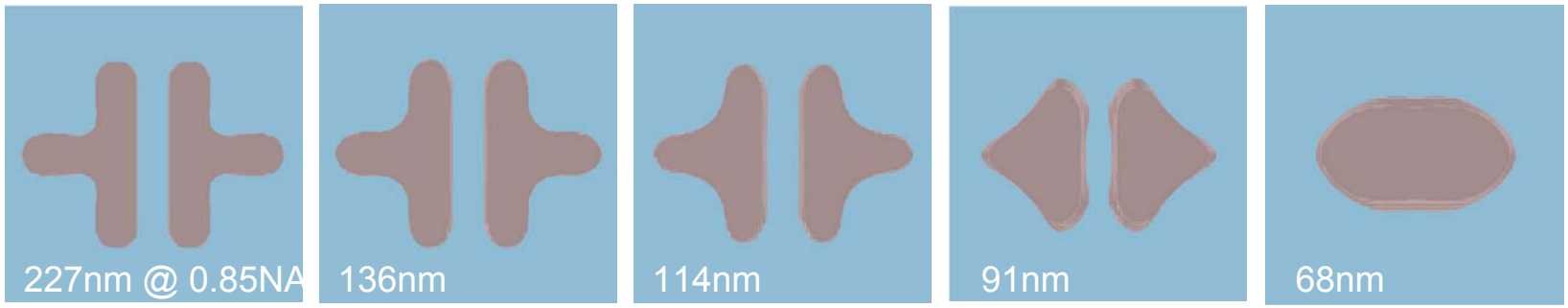
http://www.cerc.utexas.edu/utda

Sponsored by NSF, SRC, Fujitsu, IBM, Intel, Qualcomm, Sun, KLA-Tencor

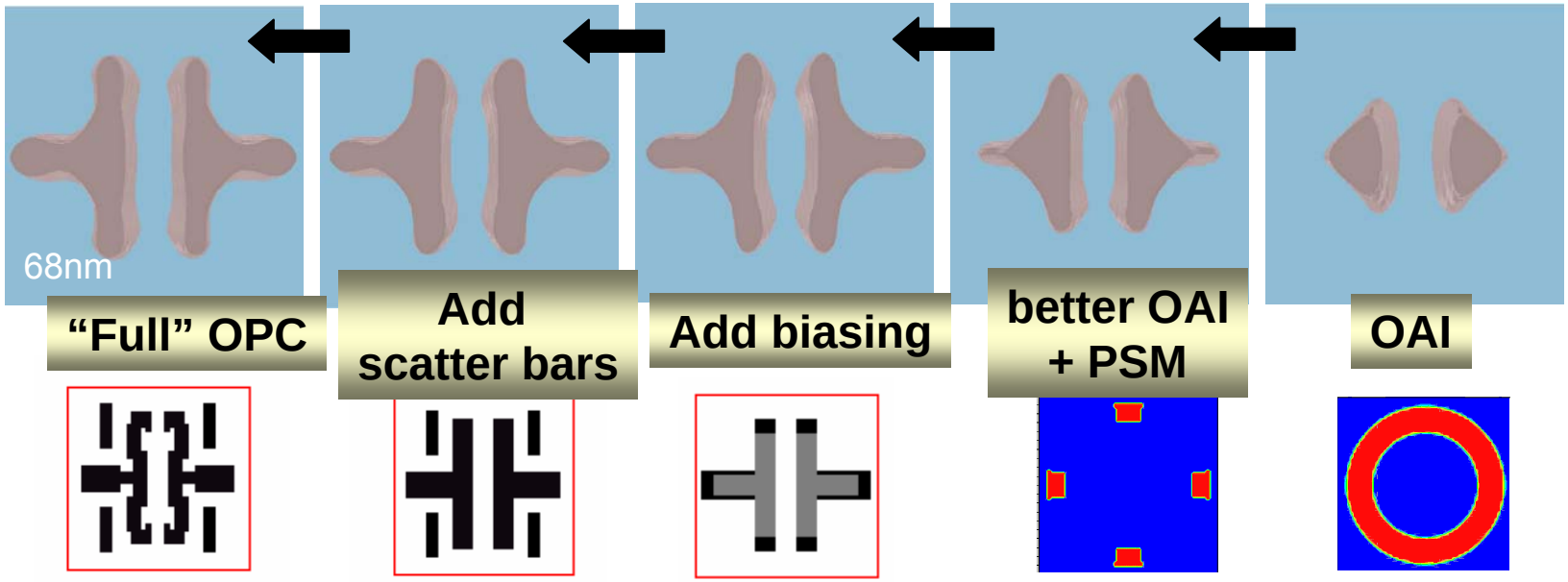
Outline

- Background & Motivation
- DFM Modeling, OPC and Characterizations
  - Design-oriented, PV-aware lithography modeling/OPC
  - Variational characterization & electrical analysis
- DFM Optimizations
  - DFM aware routing
  - Variation-tolerant design
- Conclusions

Litho: WYS != WYG



Resolution Enhancement Techniques (RET)….



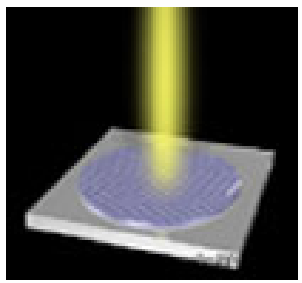
OPC: optical proximity correction; OAI: Off-axis illumination
PSM: phase shift mask

(Source: ASML)

Lithography State of the Art

- Industry stuck with 193nm lithography (in the next 5 years, 45nm, 32nm, likely 22nm)
  - Push the limit of RET and immersion lithography, …
  - NGL still many challenges [SPIE'07]
    - EUVL, E-Beam, nano-imprint…
    - 157nm declared dead

Semiconductor News



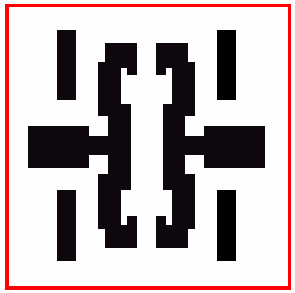
IBM sees immersion at 22-nm, pushing out EUV

IBM plans to extend 193-nm immersion lithography down to the 22-nm node for logic production, thereby possibly pushing out EUV again.

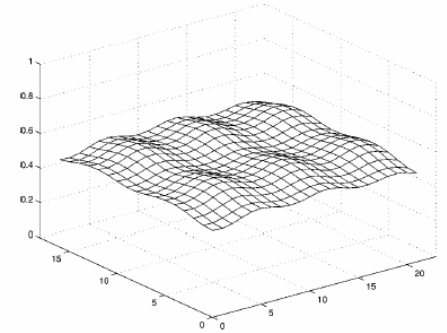
[EE Times 2/23/07]

- Very deep sub-wavelength (VDSW) + very deep sub-micron (VDSM)

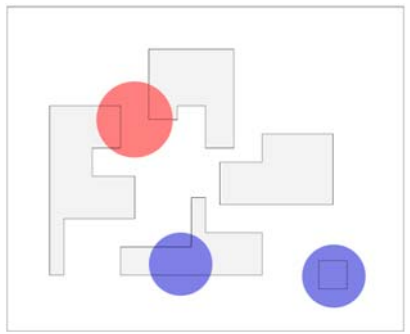
Other Manufacturing Challenges



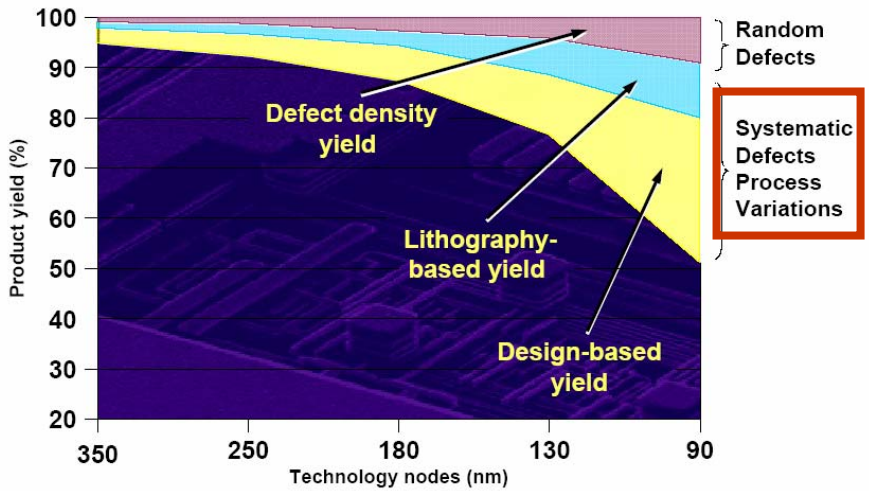
Litho



CMP

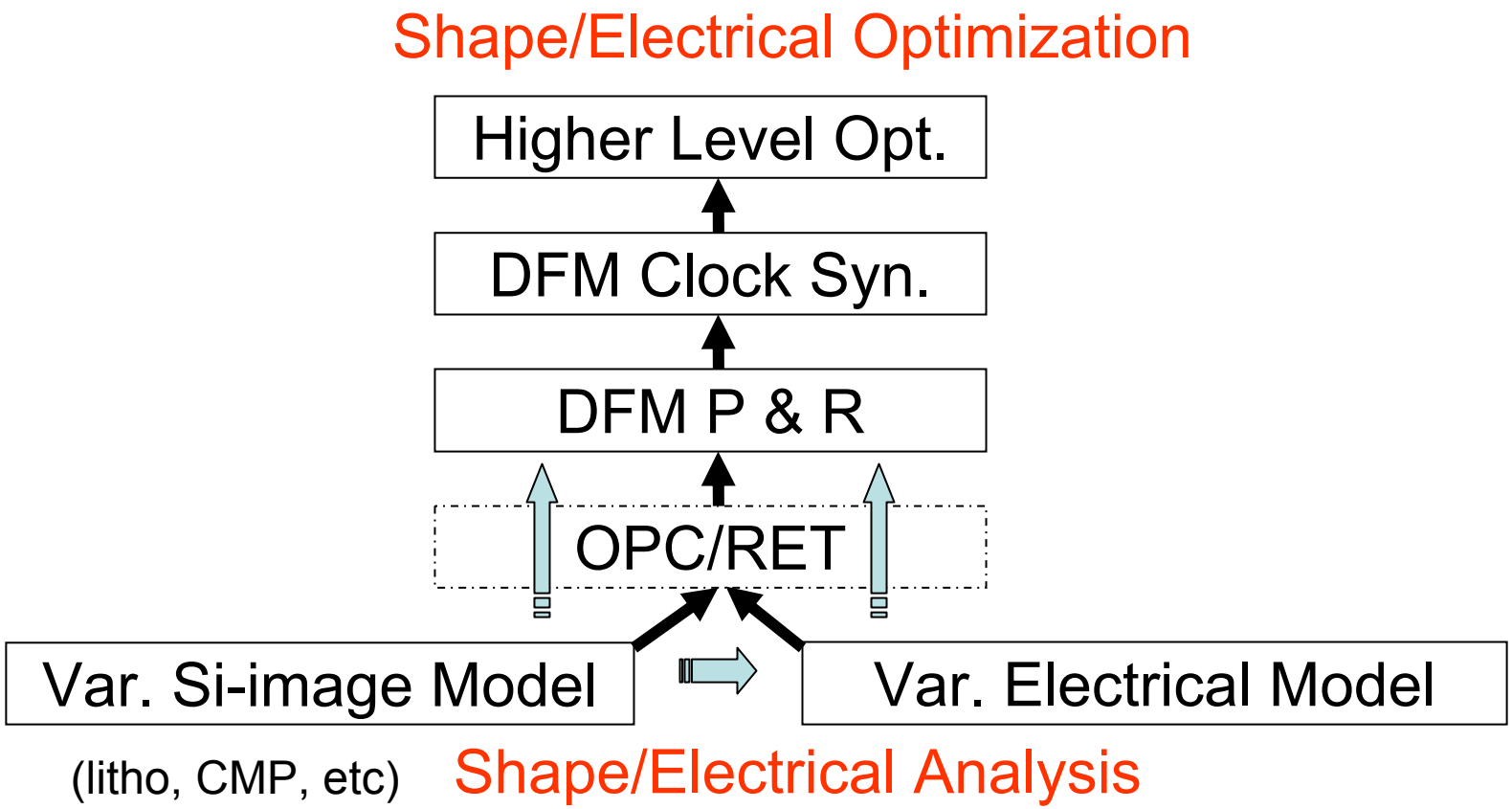


Random defects



Call for Synergistic DFM

- Modeling, extraction and characterization
- Insert proper metrics into the main design flow

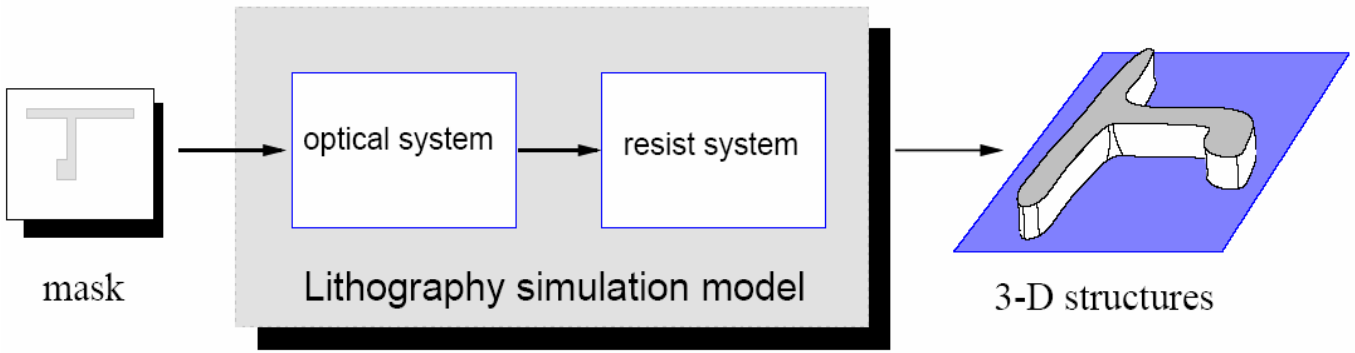


Outline

- Background & Motivation
- DFM Modeling, OPC and Characterizations
  - Design-oriented, PV-aware lithography simulation
  - Variational characterization & electrical analysis
- DFM Optimizations
  - DFM aware routing
  - Variation-tolerant design
- Conclusions

Lithography Model

- To guide lithography aware physical design, *fast* yet *high-fidelity* lithography modeling/metrics are essential
- Process-oriented vs. Design-oriented
- Two key stages in litho-model (in a simplified view)
  - Optical system: will generate aerial image from mask
  - Resist system: photoresist and patterning inside the wafer



Aerial Image Simulation

- Hopkins image equation is generic for optical system
  - U_I: Image complex amplitude
  - F: mask transmission function
  - K: optical system transmission function

$$U_I(x_1, y_1) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(x_0, y_0) K(x_1 - x_0, y_1 - y_0) dx_0 dy_0$$

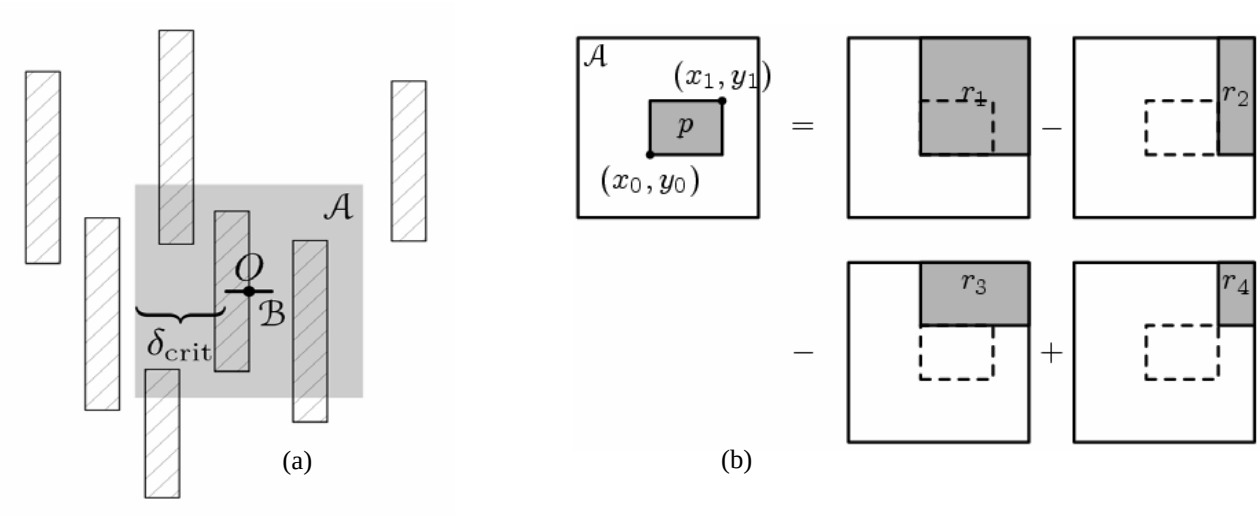
  - Image intensity function for partial coherent system

$$I_I(x_1, y_1) = \iint \iint J_0(x_0 - x_0', y_0 - y_0') F(x_0, y_0) F^*(x_0', y_0')$$
$$\times K(x_1 - x_0, y_1 - y_0) K^*(x_1 - x_0', y_1 - y_0') dx_0 dy_0 dx_0' dy_0'$$

Aerial Image Simulation

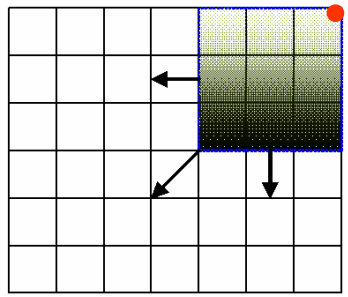
- Directly using Hopkins equation can be extremely slow
- Kernel decomposition into the sum of a small number of fully coherent systems (SOCS) [Cobb 98, Mitra et al 2005]

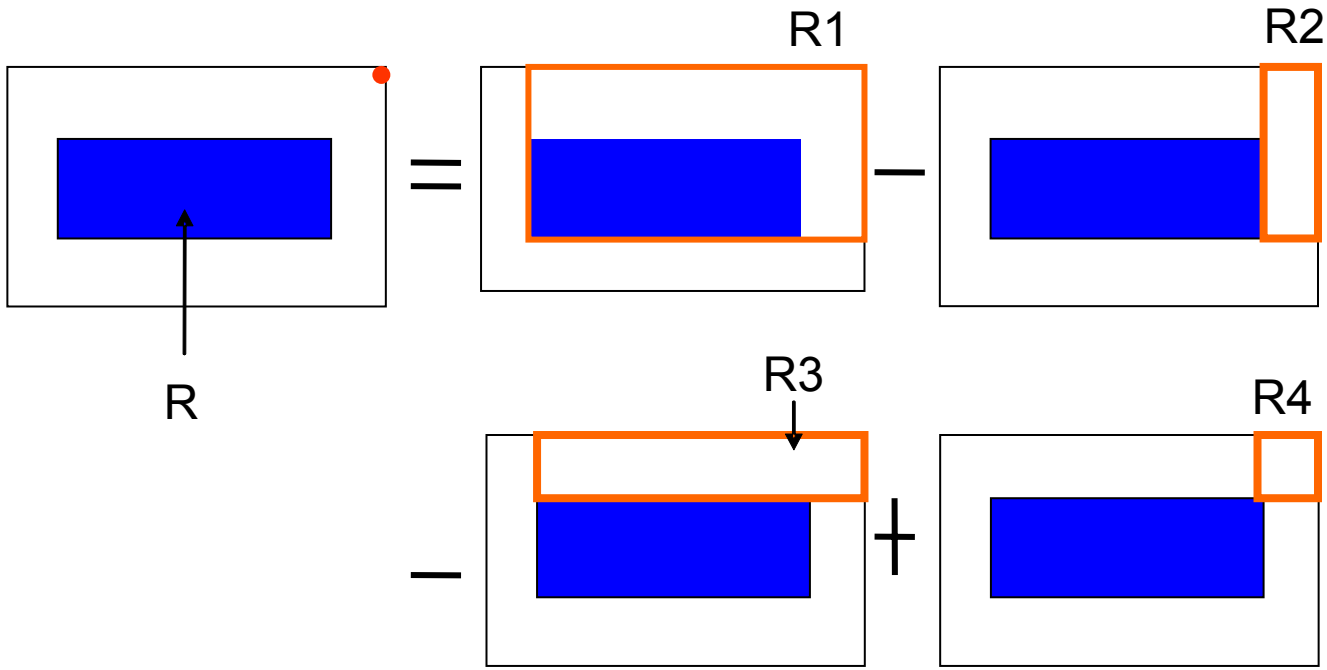
$$I_1(x,y) = \sum_{i=0}^{P-1} \left| \sum_{j \in A_{(x,y)}} (F_j * K_i)(x,y) \right|^2$$



Fast Lookup Table Technique

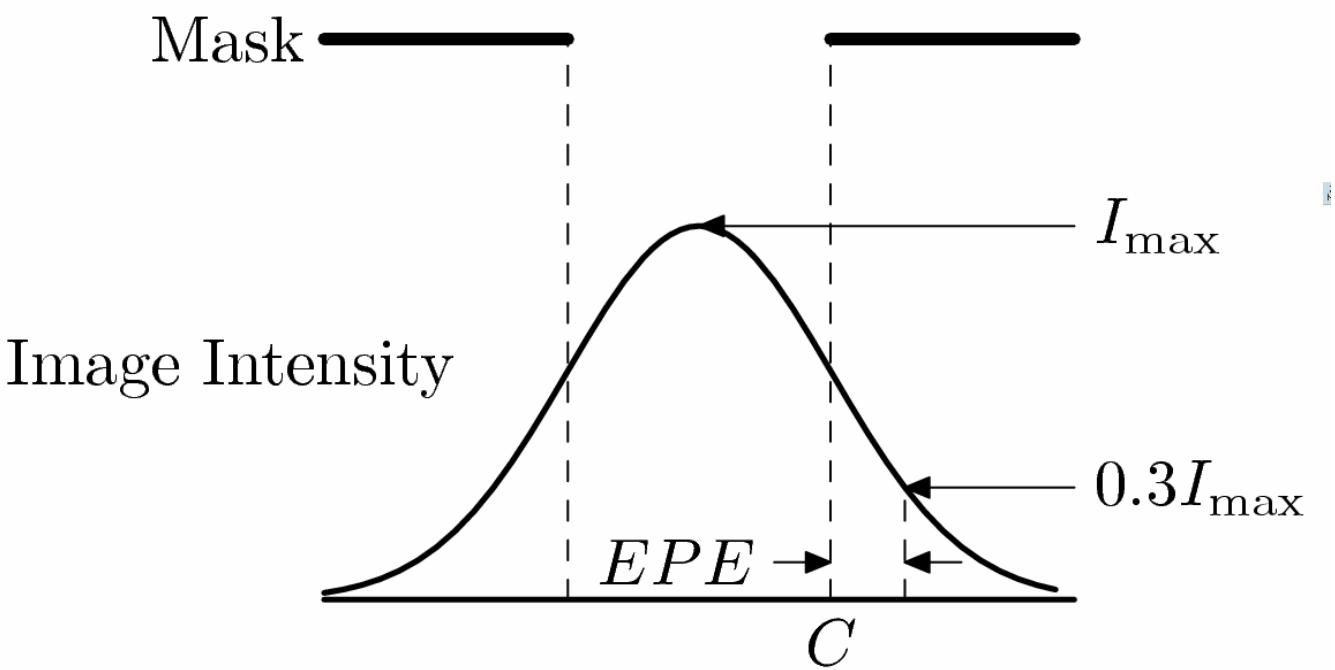
- Store convolution table for rectangles w.r.t the top-right reference point
- Support region is pretty small
- Design-oriented fast simulation by pattern-matching/caching





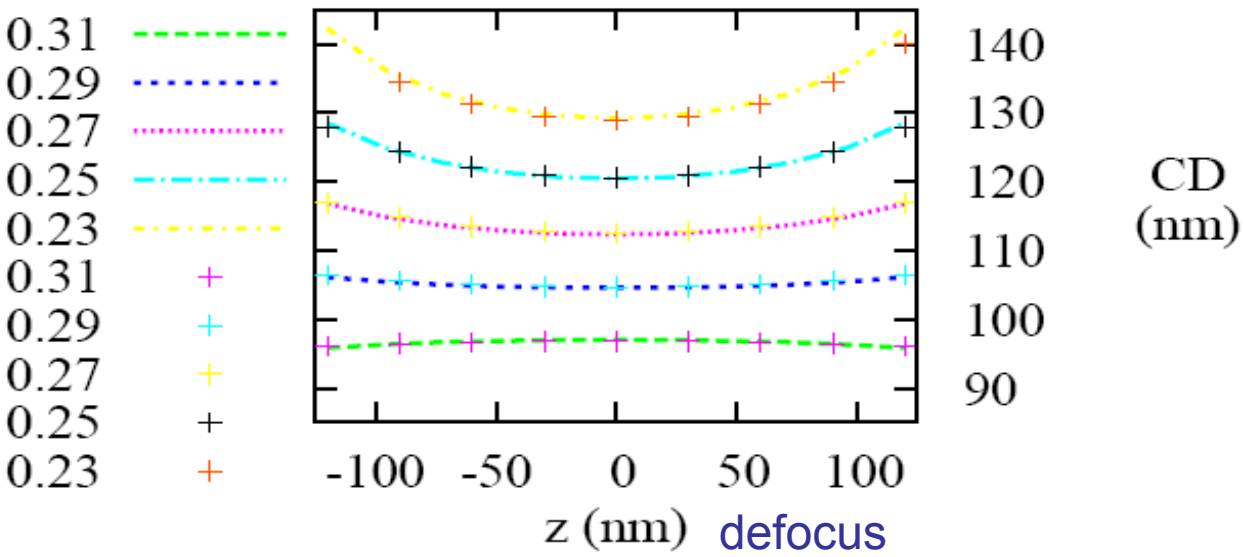
Simple Threshold Resist Model

- From the aerial image => printed image (resist model)
- Example: simple threshold model (e.g., $0.3I_{max}$) to decide where to etch based on image intensity distribution
  - And many more accurate models



Variational Lithography Modeling

- The printed image is subject to process variations
  - Dosage, focus, mask, …
  - Lithography simulation shall be variational aware
  - Extensive process-window sampling TOO SLOW

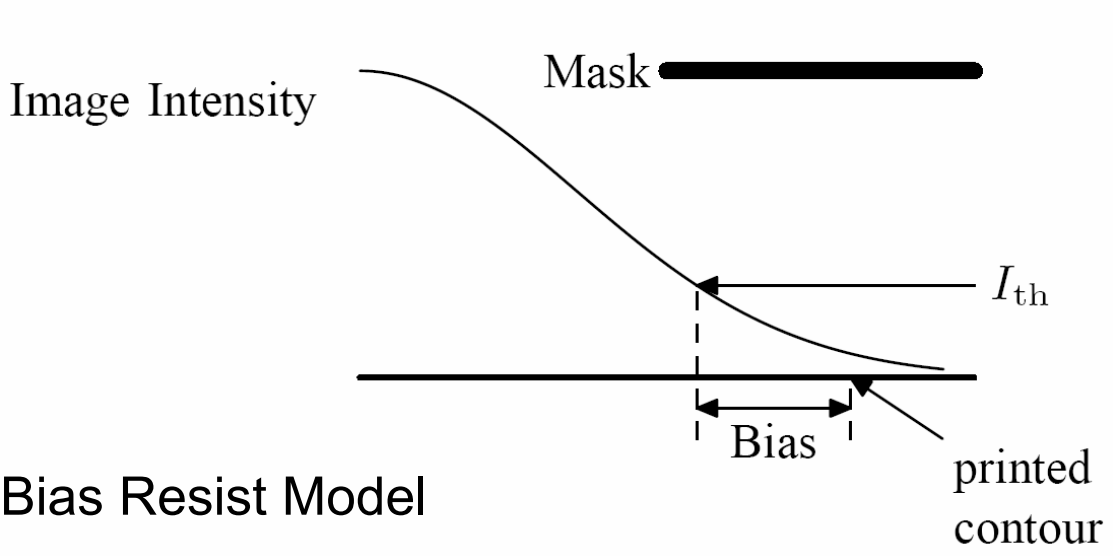


Variational Lithography Model (VLIM) [Yu et al, DAC'06]

- Focus variation: defocus aerial image expansion

$$I(x,y) \cong I_0(x,y) + z^2 I_2(x,y)$$

  - I_0 and I_2 through kernel decomposition and table lookup
- Dosage variation: equivalent threshold variation in the threshold bias resist Model



Threshold Bias Resist Model

Variational EPE

- EPE under process variations
  - analytical function of defocus and dosage

$$E(I_{\mathrm{th}}, z) = E_{\mathrm{iso}} + a_0(1 + a_1 z^2)(I_{\mathrm{th}} - I_{\mathrm{th}_{\mathrm{iso}}})$$

where

$$\begin{cases} a_0 &= \dfrac{E_{\mathrm{target}} - E_{\mathrm{iso}}}{I_0(E_{\mathrm{target}}) - I_{\mathrm{th}_{\mathrm{iso}}}} \\ a_1 &= -\dfrac{I_2(E_{\mathrm{target}}) - I_2(E_{\mathrm{iso}})}{I_0(E_{\mathrm{target}}) - I_{\mathrm{th}_{\mathrm{iso}}}} \end{cases}.$$

Variational EPE Example

- Given certain focus and dosage variations, e.g.

$$z \sim N(\mu_z, \sigma_z^2) \qquad \text{and} \qquad I_{\text{th}} \sim N(\mu_{I_{\text{th}}}, \sigma_{\text{th}}^2)$$

=> Analytical V-EPE metrics, e.g., first moment

$$\begin{aligned}\langle E \rangle &= E_{\text{iso}} + a_0(1 + a_1\sigma_z^2)(\mu_{I_{\text{th}}} - I_{\text{th}_{\text{iso}}}) \\ &= E(0, I_{\text{th}_{\text{iso}}}) + a_0 a_1 \sigma_z^2 (\mu_{I_{\text{th}}} - I_{\text{th}_{\text{iso}}}) \\ &= E_{\text{norm}} + \mu_E\end{aligned}$$

- Concepts are generic to other “raw” distributions of focus/dosage or variational EPE metrics

Applications: PV-OPC and Beyond

- Key Idea: Derive/Extract variational EPE from "raw" variations: focus and dosage, etc.
- Run time only 2-3x compared to nominal process, but explicitly consider process variations
- Can be used to guide true PV aware OPC (PV-OPC) or litho-aware routing

POST-OPC CD MEAN AND EPE VARIANCE COMPARISON.

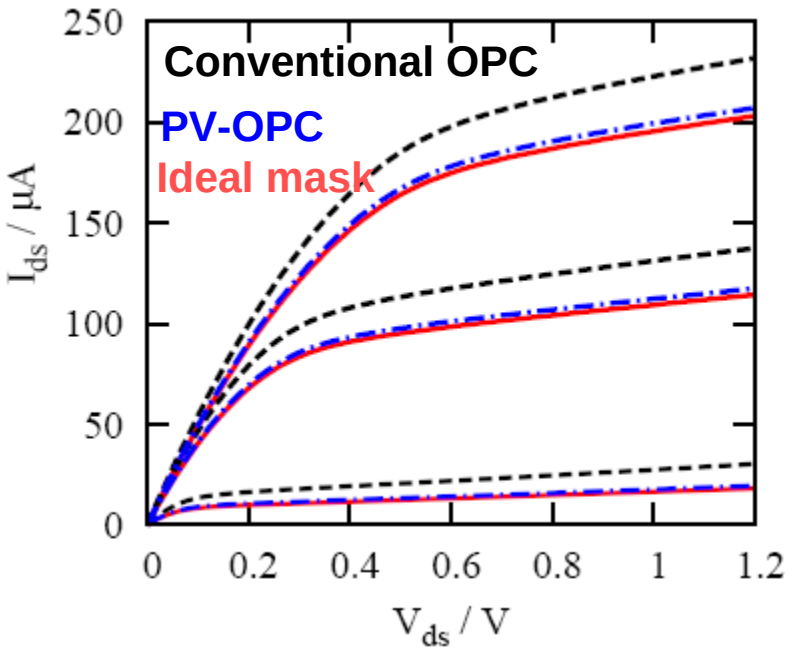
	average CD mean (nm)		average EPE variance (nm)	
	Conventional	PV-OPC	Conventional	PV-OPC
PMOS	61.44	65.22	3.39	3.29
NMOS	61.09	64.35	3.50	3.41

US Patent Pending (supported by SRC)

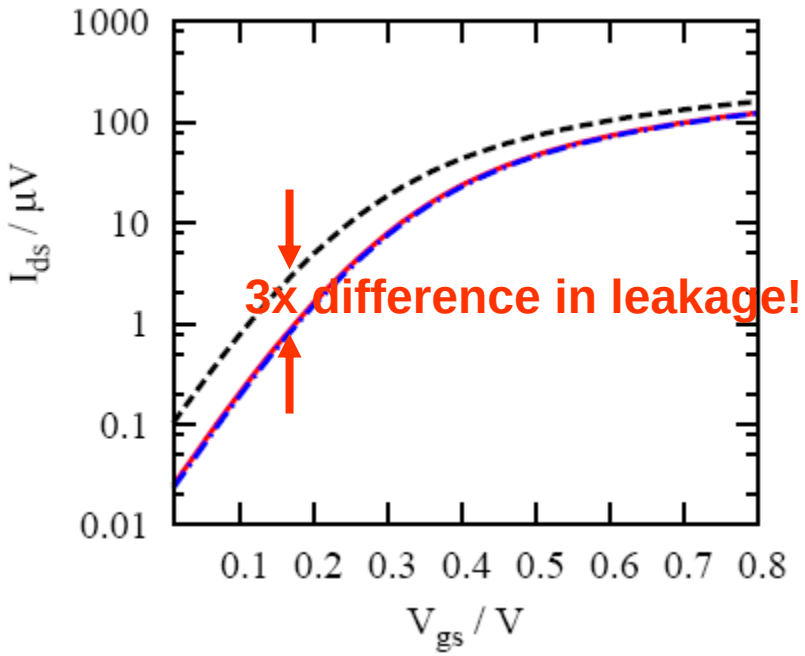
[Yu et al, DAC'06]

Geometry => Electrical

- Electrical characterizations of a 65nm inverter
- PV-OPC is able to meet design intent under process variations



I-V curves with V_{gs} = 0.4, 0.8, 1.2V



NMOS Leakage with V_{ds} = 1.2V

Variational Electrical Analysis

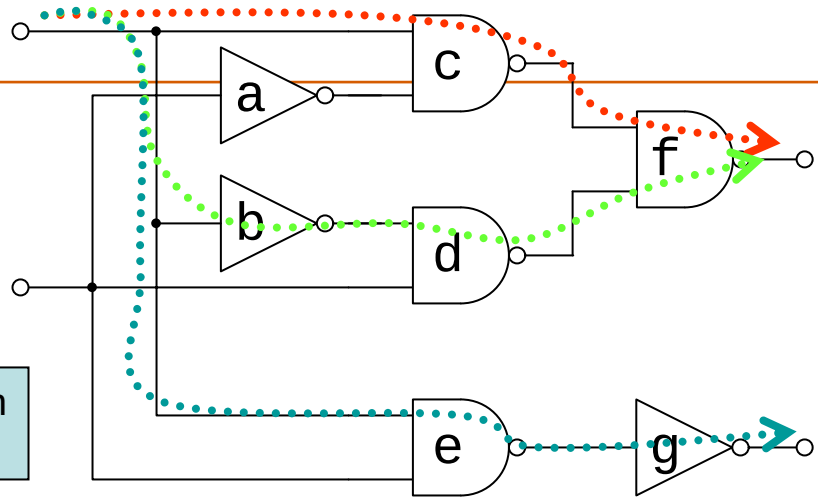
- Layout-dependent non-rectangular gate characterization [Shi+, ICCAD'06]

=> more accurate static/statistical circuit analysis

- A novel sparse-matrix formulation for SSTA [Ramalingam+, ICCAD'06]
  - Model path-based SSTA using sparse matrix multiplication (fast Monte-Carlo in one-short)
  - Handle arbitrary correlations
  - Handle slope propagation
  - Very accurate and fast!
  - Can use block-based SSTA for path-pruning

Path Enumeration

- ½ adder example:



Each column is an input pin

All i/p – o/p, latch-latch paths are represented

a	b	c1	c2	d1	d2	e1	e2	f1	f2	g
		1						1		
	1			1					1	
						1				1
1			1					1		
					1				1	
							1			1

Each row is a complete path

A 1 means input pin is on this path

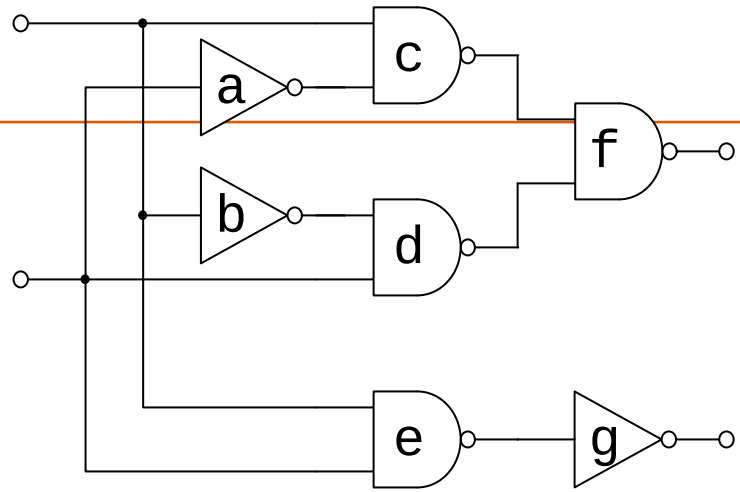
Incidence matrix is quite sparse!

- Path delays can be written in **p** = **A d**.

Delay Modeling

- By linear regression, express gate delay as a polynomial in:
  - Operating environment, e.g., C_{Load}
  - Technology variables, e.g. L, V_T
- Linear regression (≠ Linear models), e.g.:
  - $d = c_0 + c_1 \underbrace{C_{Load} L^2}_{\text{Regression term } z_1} + c_2 \underbrace{L V_T^{1/2}}_{\text{Regression term } z_2} - \ldots$
- Collect all regression terms into one vector **z**
  - $d = \mathbf{c}^T \mathbf{z}$

Gate delay Vector



- Delay for gate k: $d_k = \mathbf{c_k}^T \mathbf{z_k}$
  - d_{c1} denotes the delay from gate c's pin 1 to o/p
- Expanding for all gates/arcs:

$$\begin{pmatrix} d_a \\ d_b \\ d_{c1} \\ d_{c2} \\ d_{d1} \\ d_{d2} \\ \dots \end{pmatrix} = \begin{pmatrix} \mathbf{c_a}^T & & & & & & \\ & \mathbf{c_b}^T & & & & & \\ & & \mathbf{c_{c1}}^T & & & & \\ & & & \mathbf{c_{c2}}^T & & & \\ & & & & \mathbf{c_{d1}}^T & & \\ & & & & & \mathbf{c_{d2}}^T & \\ & & & & & & \end{pmatrix} \begin{pmatrix} \mathbf{z_a} \\ \mathbf{z_b} \\ \mathbf{z_{c1}} \\ \mathbf{z_{c2}} \\ \mathbf{z_{d1}} \\ \mathbf{z_{d2}} \\ \end{pmatrix}$$

- => **d = C z**

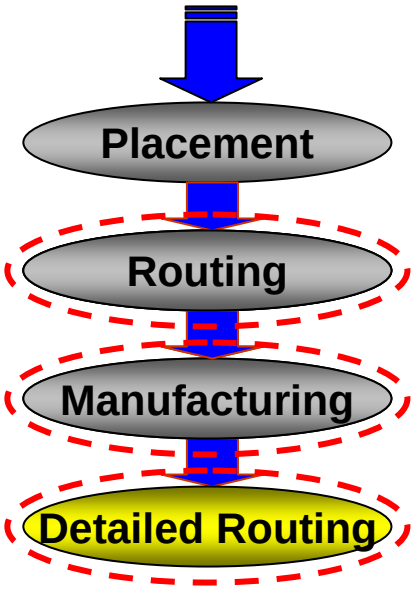
Path Delays using Monte-Carlo

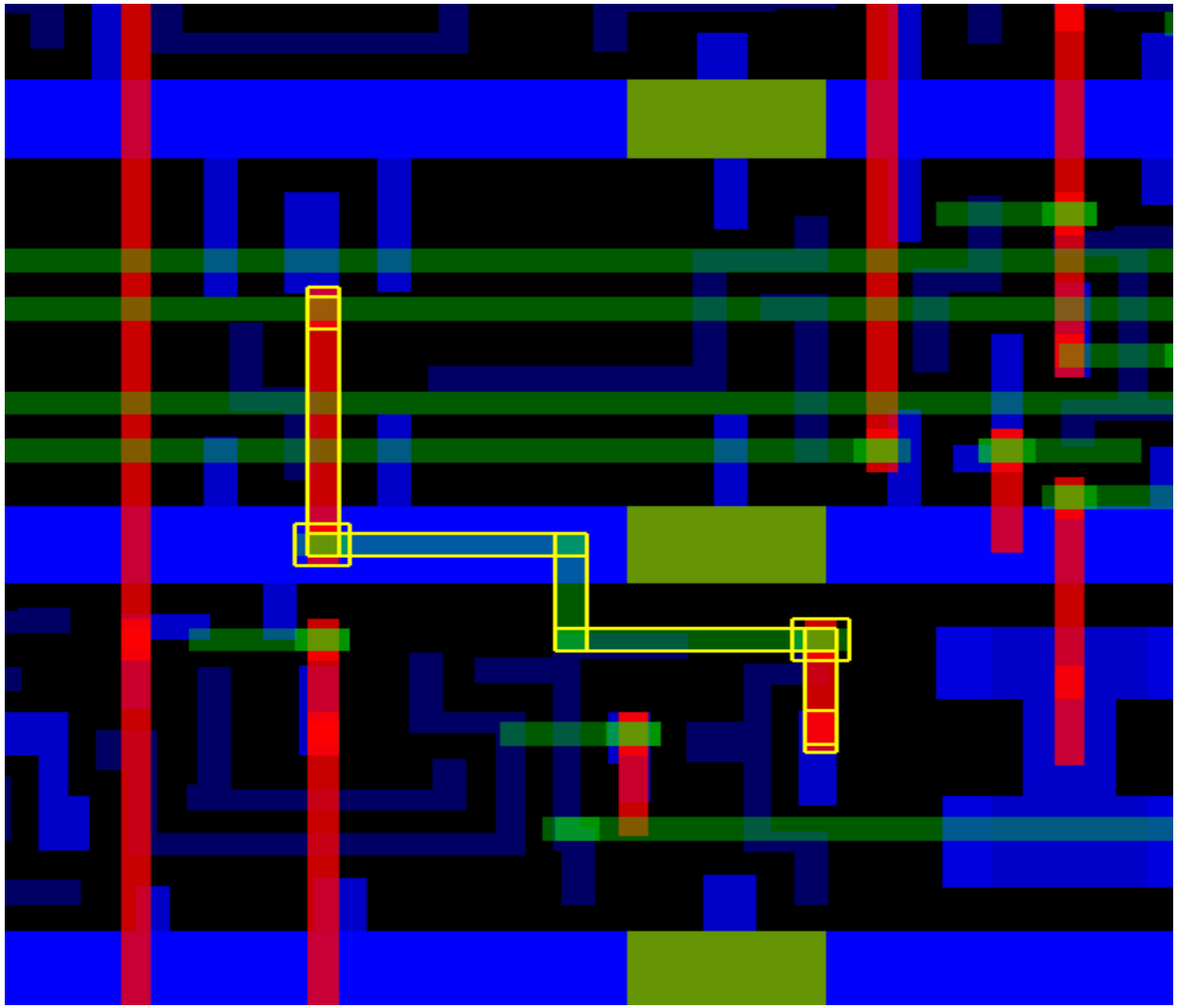
- Equation for all path delays given by:
  - **p = A d = A C z**
- Values of **z** determined by technology settings such as V_T, L, V_{DD} etc.
  - Given an *arbitrary distribution* of **z**, we can generate *n* samples **Z = [z_1 … z_n]**
- Path delays corresponding to these *n* samples:
  - **[p_1 … p_n] = A C [z_1 … z_n]**
- **Fast Monte-Carlo SSTA** in sparse-matrix form:
  - **P = A C Z**
- Handle slope propagation nicely (a new C' matrix)
- Method implemented/used by industry.

Outline

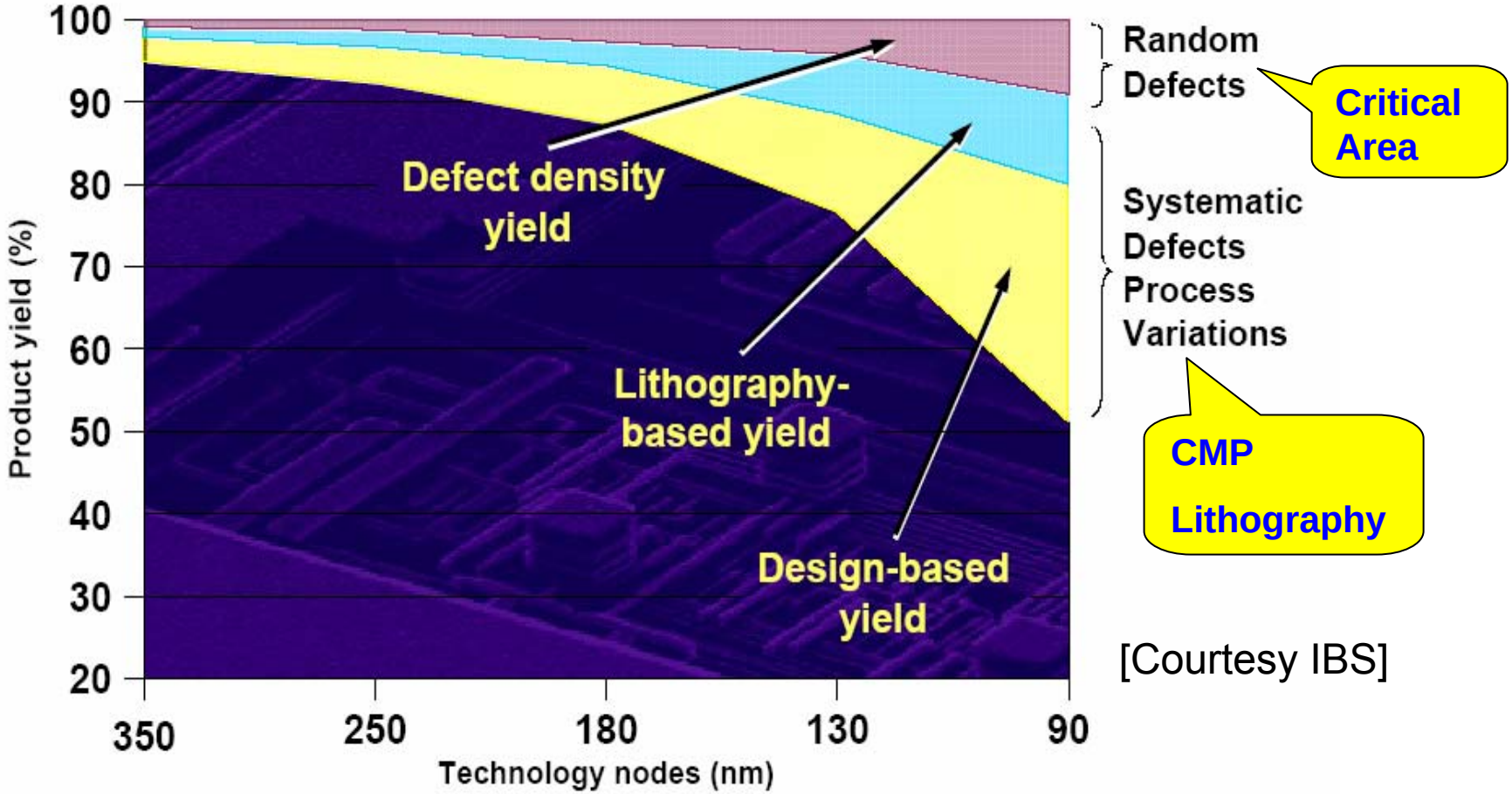
- Background & Motivation
- DFM Modeling, OPC and Characterizations
  - Design-oriented, PV-aware lithography simulation
  - Variational characterization & electrical analysis
- DFM Optimizations
  - DFM aware routing
  - Variation-tolerant design
- Conclusions

Routing in VLSI Physical Design



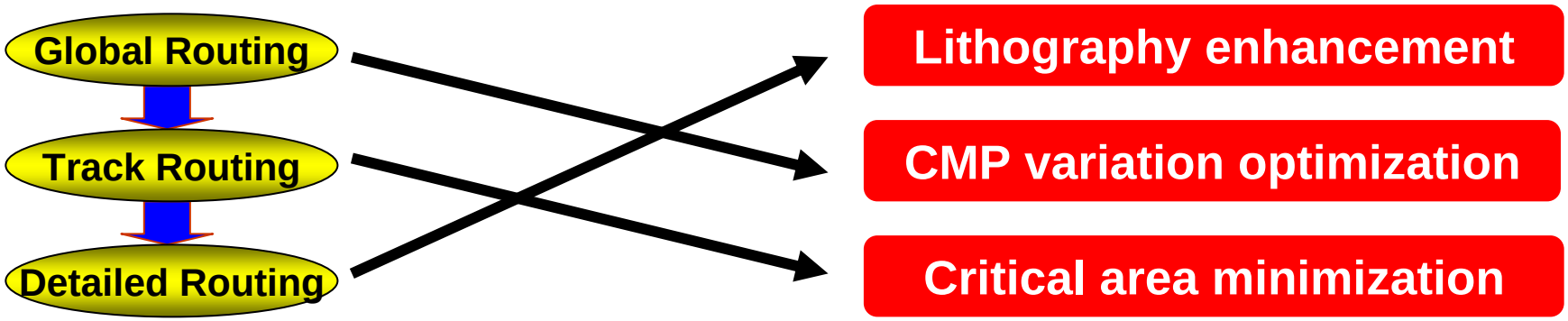


Yield Loss Mechanisms



[Courtesy IBS]

Which Stage to Tackle What?



- CMP variation optimization
  - Minimum effective window (20x20μm²) and global effect
  - Global routing plans approximate routing density
- Critical area optimization
  - Adjacent parallel wires contribute majority of critical area.
  - Track router has good flexibility with wire ordering/spacing/sizing.
- Lithography optimization
  - Effective windows (1-2μm²) are small
  - Detailed router performs localized connection between pins/wires.

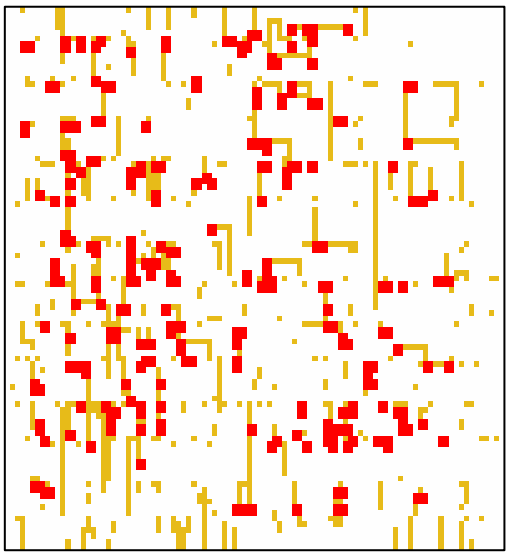
Litho-Aware Routing

- More & more metal layers need RETs
- Rule vs. model based approach
- Rules -
  - Exploding number of rules
  - Very complicated rules
  - Too conservative or not accurate rules
  - No smooth tradeoffs (either follow or break rules)
- How about directly link litho-models with routing?
  - Lithography simulations could be extremely slow !
  - A full-chip OPC could take a week
  - Accuracy vs. Fidelity (Elmore-like)
  - Design-oriented vs. process-oriented

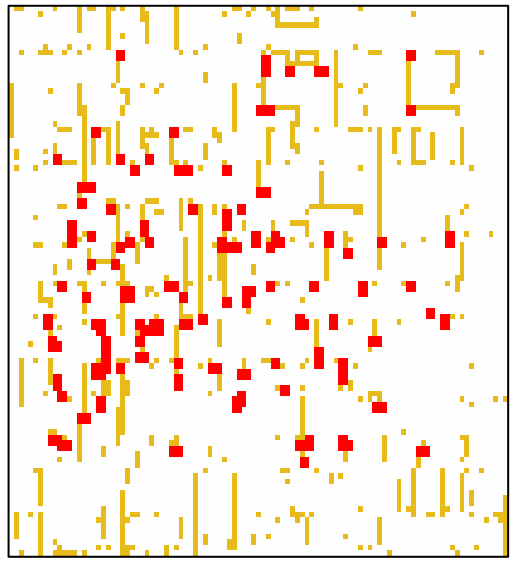
RET-Aware Detailed Routing (RADAR) [Mitra et al, DAC'05]

- First work to truly link lithography modeling with design implementation level
- Introduce the concept of Litho- Hotspot Maps
- Fast lithography simulation to generate LHM
  - Guided by our design-oriented fast litho simulations
- Post-routing optimizations to reduce hotspots
  - Wire spreading
  - Ripup-Reroute (RR) with blockages (protect "good" regions)

RET-aware Routing on a 65nm Design



Initial routing (after design closure)



40% EPE reduction
(much less litho hot spots)

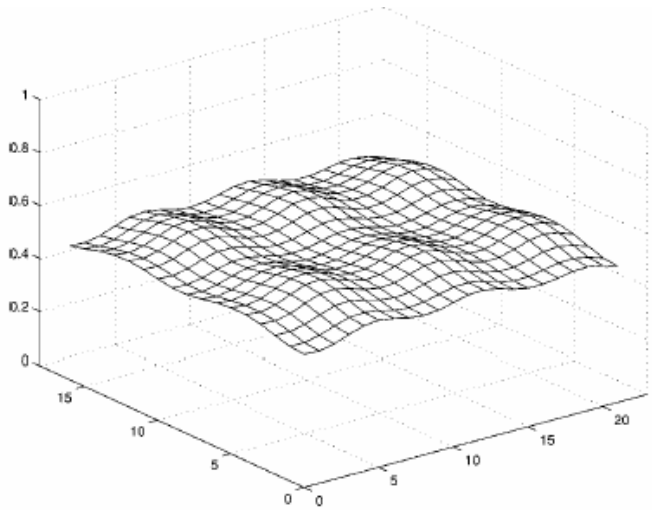
[Mitra et al, DAC'05]

More on Litho-Aware Routing

- Lithography hotspot is a generic concept
- Other metrics may be used: Not just EPE
  - Possibly under process variations, such as variational EPE (V-EPE) metric [Yu+, DAC'06]
  - Other Predictive OPC modeling [SPIE'07]
- We are developing a new correct-by-construction litho-aware router

CMP-Aware Routing

- Lithography interact with CMP
  - CMP => defocus
- CMP-Aware Routing
  - Need scalable yet high-fidelity CMP model
  - Best attacked at the **global** routing stage

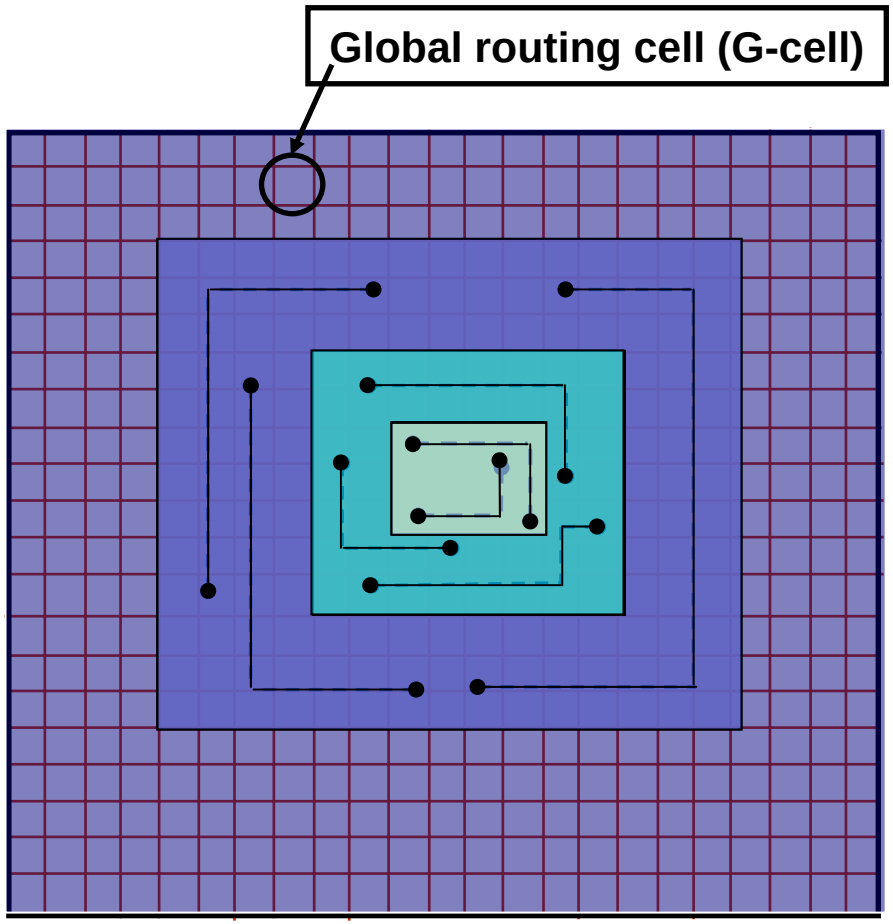


Wire Density Driven Paradigm
[Cho et al, ICCAD'06]

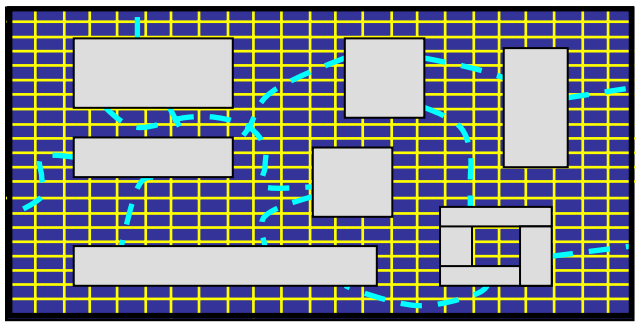
- In general, more uniform wire density
  - Better CMP topography variations
- Wire density will affect timing too
  - Lower wire density => less resistance (higher thickness)
  - Lower wire density => less capacitance
- Wire density: a unified metric for CMP and Timing optimization (one stone, two birds!)
- Wire density vs. congestion driven GR
  - Though correlated, indeed different during global routing
  - Density: wires inside global routing cells
  - Congestion: wires crossing the boundaries of global routing cells

BoxRouter [DAC'06, ICCAD'07]

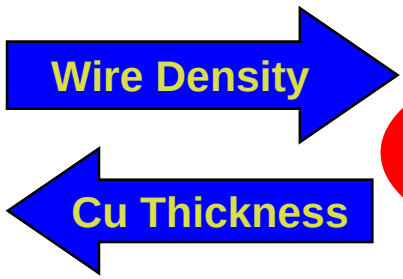
- Incremental Box Expansion
  - From most congested regions
- Progressive ILP
- Adaptive maze routing
- PostRouting
  - Negotiation-based approach
- DAC'06 BPA Candidate
- 2nd place (in 3D) ISPD07 Routing Contest
- US Patent filed by SRC
- Open Source Released

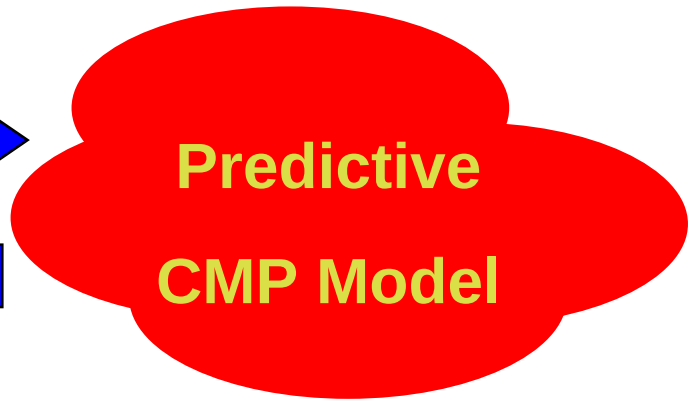


Predictive Copper CMP Model



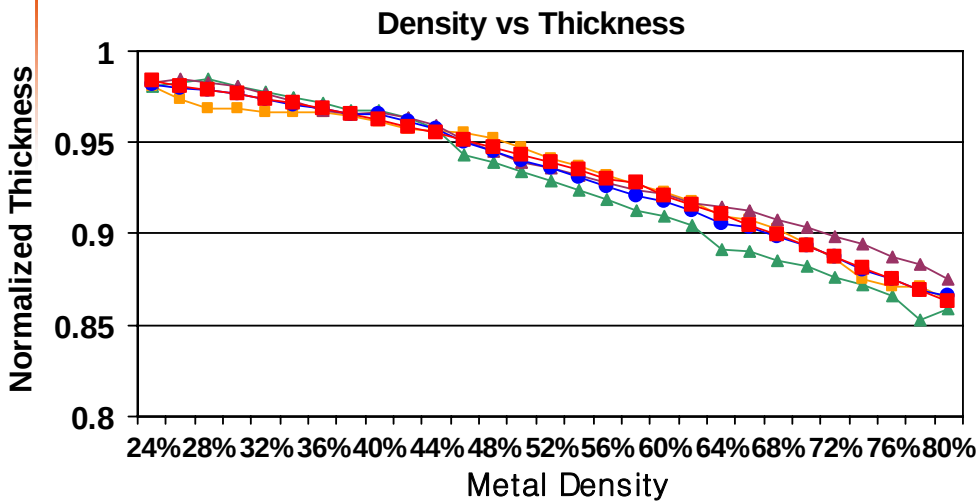
Global Routing

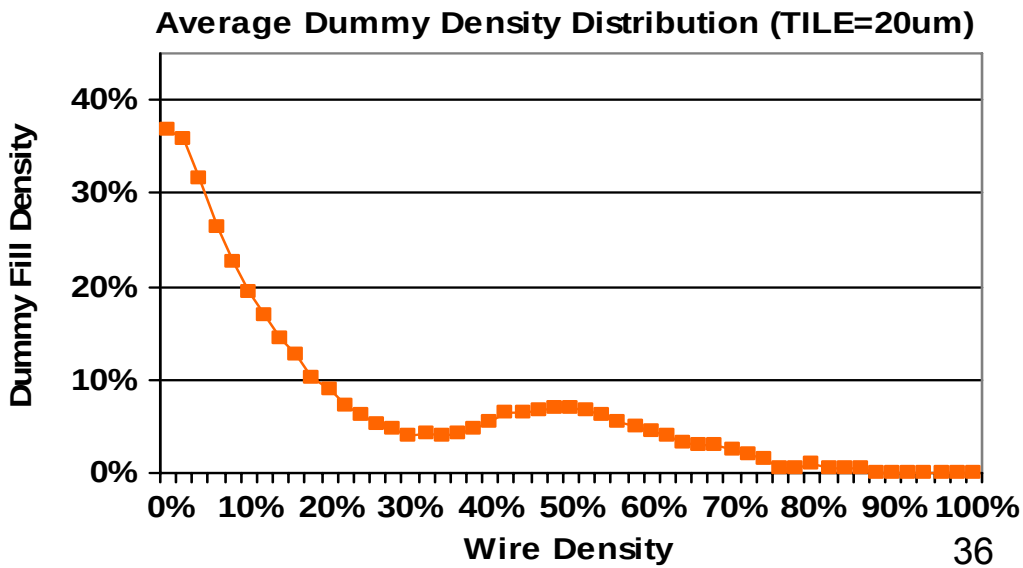




Verified by physics-based simulations

Predictive Copper CMP Model





Calculate Wire Density

↓

Required Dummy Density
(From Lookup Table)

↓

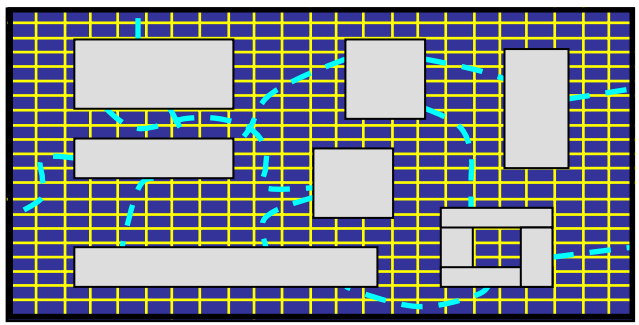
Metal Density =
Wire Density + Dummy Density

↓

Thickness =
$A*(1- \text{Metal Density}^2/B)$

[Cho et al, ICCAD'06]

Global Routing Flow With Predictive CMP Model



Global Routing

BoxRouter

Wire Density

Cu Thickness

Dummy Fill Density From Lookup Table

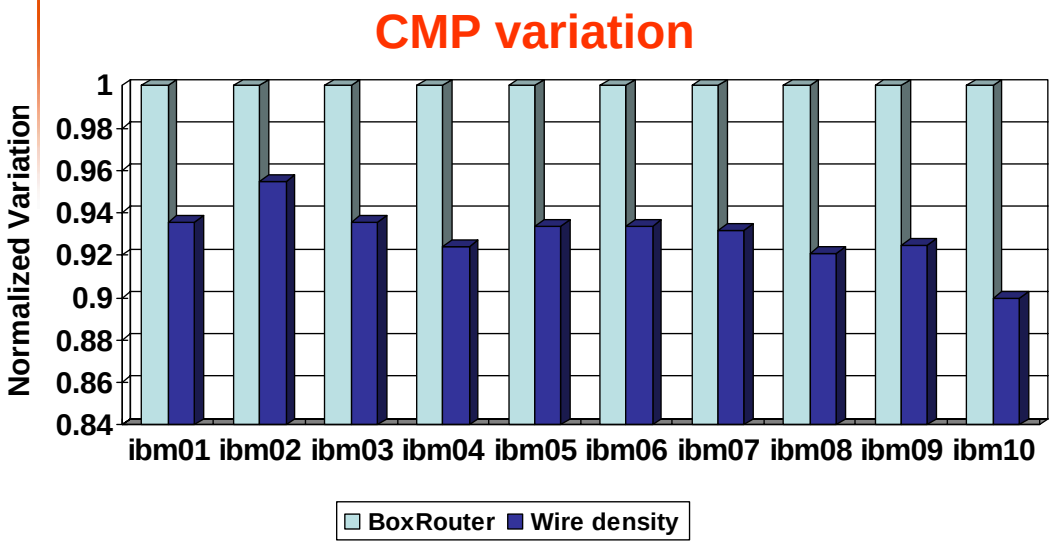
Metal Density = Wire Density+ Dummy Fill Density

Cu Thickness

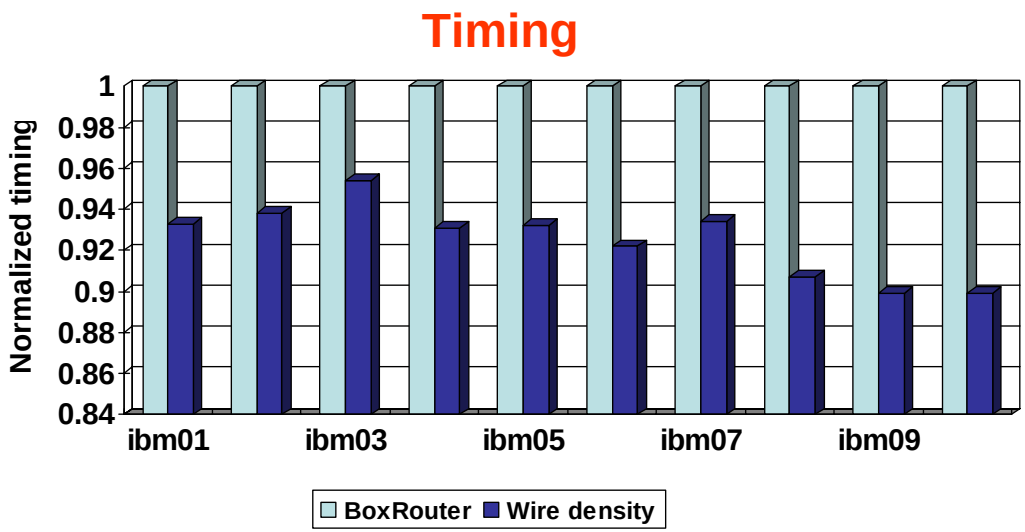
Predictive CMP Model for Global Routing

- Predictive CMP Model guides global router
  - More uniform wire distribution
  - Consider metal blockages (power/ground rail, IPs)

Experimental Results



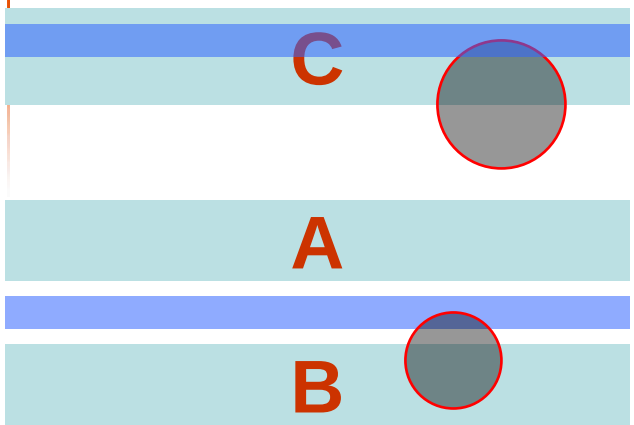
- On average 7.5% reduction
- Up to 10.1% reduction



- On average 7% reduction
- Up to 10% reduction

[Cho et al, ICCAD'06]

Critical Area For Yield



Critical area due to open defect
- Reduce defect size
- Increase wire width

Critical area due to short defect
- Reduce defect size
- Increase wire spacing

- Random defect can cause open/short defect
  - Wire planning for critical area reduction
- Defect size distribution
  - Chance of getting larger defect decreases rapidly [TCAD85]
- Concurrent optimization for open/short defect
  - Larger wire width for open, but larger spacing for short defect
  - Limited chip area

Track Routing for Yield (TROY) [Cho et al, DAC'07]

- TROY is the first yield-driven track router
  - Wire ordering to minimize overlapped wirelength between neighbors
    - Preference-aware Minimum Hamiltonian Path
  - Wire sizing/spacing to minimize critical areas
    - Second-order conic programming (SOCP)
    - global optimal in nearly linear time
- Result is very promising
  - 18% reduction in yield loss due to random defects
  - TROY is very scalable

Math 101 for Critical Area

$$F(x) = kx^{-r} \qquad \text{for } x_{min} \le x < \infty$$

Defect size distribution *r* =3

$$A_i^o(x) = \begin{cases} 0 & \text{for } 0 \le x < w_i \\ L_i(x - w_i) & \text{for } w_i \le x < 2w_i + S_{min} \\ L_i(w_i + S_{min}) & \text{for } 2w_i + S_{min} \le x < \infty \end{cases}$$

Critical are due to open defect

$$A_{ij}^s(x) = \begin{cases} 0 & \text{for } 0 \le x < s_{ij} \\ l_{ij}(x - s_{ij}) & \text{for } s_{ij} \le x < 2s_{ij} + W_{min} \\ l_{ij}(s_{ij} + W_{min}) & \text{for } 2s_{ij} + W_{min} \le x < \infty \end{cases}$$

Critical are due to short defect

$$POF_i^o = \int_{x_{min}}^{\infty} F(x)\frac{A_i^o(x)}{A_{chip}}\,dx = \frac{kL_i}{2A_{chip}}\left(\frac{w_i + S_{min}}{2w_i^2 + S_{min}w_i}\right)$$

$$POF_{ij}^s = \int_{x_{min}}^{\infty} F(x)\frac{A_{ij}^s(x)}{A_{chip}}\,dx = \frac{kl_{ij}}{2A_{chip}}\left(\frac{s_{ij} + W_{min}}{2s_{ij}^2 + W_{min}s_{ij}}\right)$$

Probability of failure (*POF*) for open/short defects

$$POF_i^o \approx \frac{kL_i}{2A_{chip}}\left(a\frac{S_{min}}{w_i} - b\right) \quad \left(1 \le \frac{w_i}{S_{min}} \le 20\right)$$

$$POF_{ij}^s \approx \frac{kl_{ij}}{2A_{chip}}\left(a\frac{W_{min}}{s_{ij}} - b\right) \quad \left(1 \le \frac{s_{ij}}{W_{min}} \le 20\right)$$

Approximated *POF*

- *POFs* for open/shorts are convex functions
  - Global optimal can be found, but *POF* is complicated.
  - Approximated POF is also convex, but easy enough to enable SOCP.

TROY Brute-forth Formulation

$$\begin{aligned}
\textbf{min}: \quad & \alpha \textstyle\sum_i (POF_i^o + POF_i^{o*}) + (1-\alpha) \sum_{i,j>i} POF_{ij}^s \\
\textbf{s.t}: \quad & |p_i - M_i| \le d_i && \forall i \\
& S_{min} \le s_{ij} \le p_i - p_j - \frac{(w_i+w_j)}{2} + (1-o_{ij})N && \forall i,j \\
& S_{min} \le s_{ij} \le p_j - p_i - \frac{(w_i+w_j)}{2} + o_{ij}N && \forall i,j \\
& o_{ij} \in \{0,1\} && \forall i,j \\
& B_k + \tfrac{w_i}{2} \le p_i \le T_k - \tfrac{w_i}{2} && \forall i \in P_k \\
& W_{min} \le w_i \le W_{max} && \forall i
\end{aligned}$$

***POF* for short**

***POF* for open**

Min/max width and spacing constraints, and desired locations

- Integer nonlinear programming
  - Extremely hard to solve
  - Integer variable for the wire order o_{ij} (above/below relationship)

TROY Strategy

$$
\begin{aligned}
\textbf{min}: \quad & \alpha \textstyle\sum_i (POF_i^o + POF_i^{o*}) + (1-\alpha) \sum_{i,j>i} POF_{ij}^s \\
\textbf{s.t}: \quad & |p_i - M_i| \le d_i && \forall i \\
& S_{min} \le s_{ij} \le p_i - p_j - \tfrac{(w_i+w_j)}{2} + (1+o_{ij})N && \forall i,j \\
& S_{min} \le s_{ij} \le p_j - p_i - \tfrac{(w_i+w_j)}{2} + o_{ij}N && \forall i,j \\
& o_{ij} \in \{0,1\} && \forall i,j \\
& B_k + \tfrac{w_i}{2} \le p_i \le T_k - \tfrac{w_i}{2} && \forall i \in P_k \\
& W_{min} \le w_i \le W_{max} && \forall i
\end{aligned}
$$

Integer nonlinear programming

Solved by finding minimum Hamiltonian path

Wire ordering

$$
\begin{aligned}
\textbf{min}: \quad & \alpha \textstyle\sum_i \{\delta_i + (1-\tfrac{b}{a})d_i\} + (1-\alpha) \sum_{i,j} \gamma_{ij} \\
\textbf{s.t}: \quad & |p_i - M_i| \le d_i && \forall i \\
& S_{min} \le s_{ij} = p_i - p_j - \tfrac{w_i+w_j}{2} && \forall o_{ij}=1, \forall j \in n_i \\
& l_{ij} W_{min} \le s_{ij} \gamma_{ij} && \forall i, \forall j \in n_i \\
& L_i S_{min} \le w_i \delta_i && \forall i \\
& B_k + \tfrac{w_i}{2} \le p_i \le T_k - \tfrac{w_i}{2} && \forall i \in P_k \\
& W_{min} \le w_i \le W_{max} && \forall i
\end{aligned}
$$

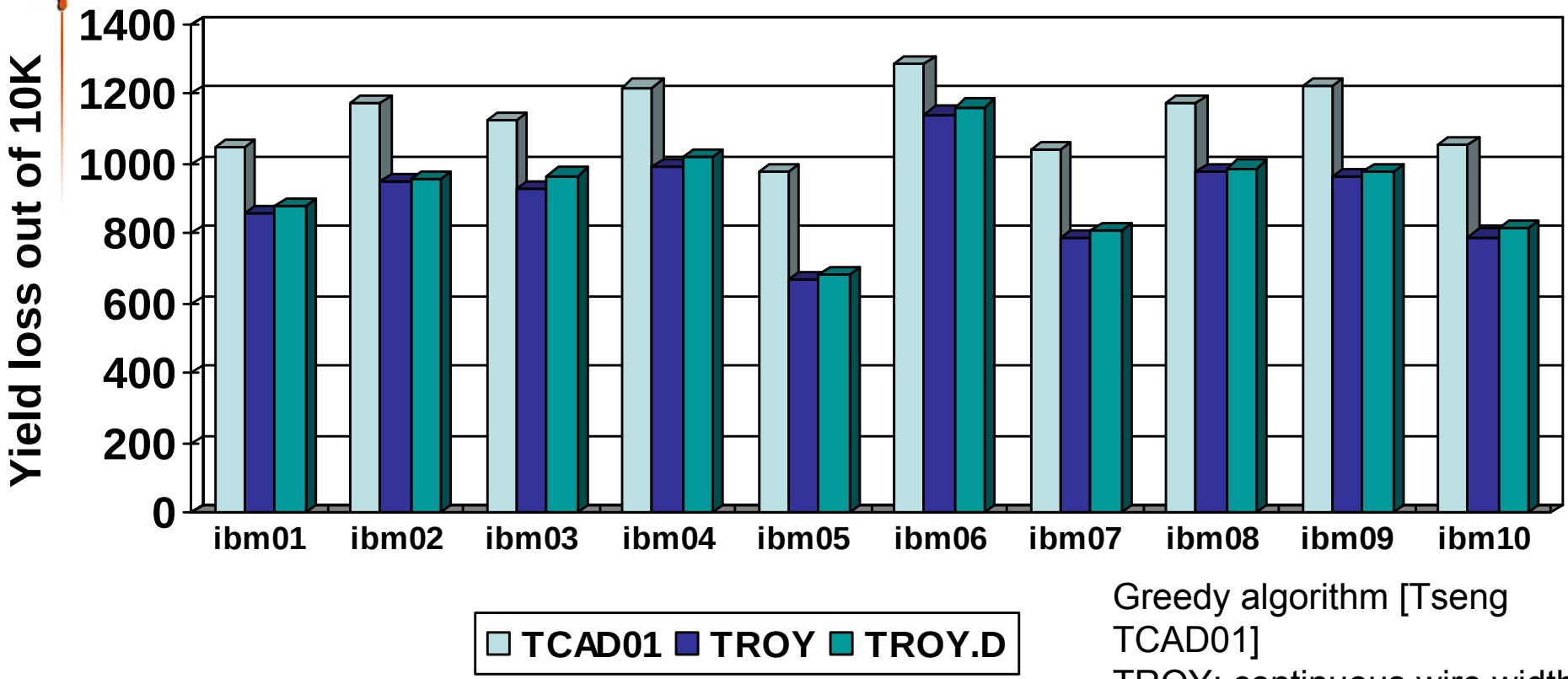
Second order conic programming (SOCP)

Wire sizing/spacing

- SOCP: Global optimal solution in nearly linear time

TROY Results

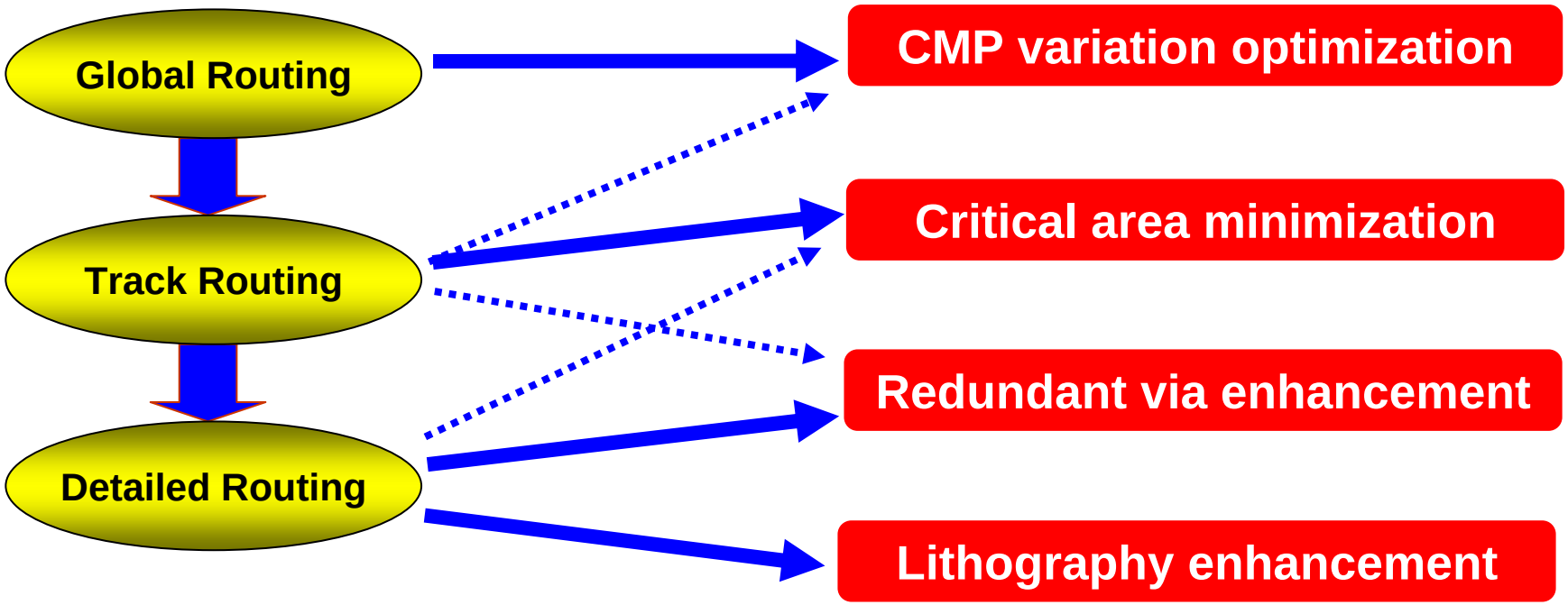
[Cho et al, DAC'07]



Greedy algorithm [Tseng TCAD01]
TROY: continuous wire width
TROY.D: discrete wire width

- Monte-Carlo simulation with 10K defects
- On average 18 % reduction in yield loss, up to 30%
- Discretization only loses 2%

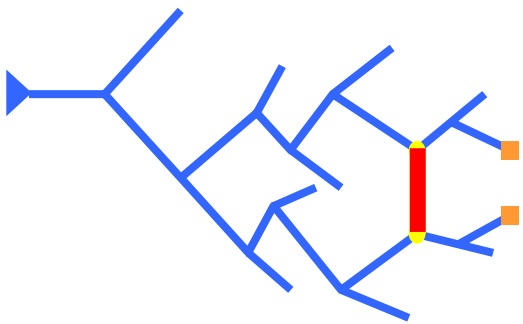
DFM-Aware Routing Wrap-up



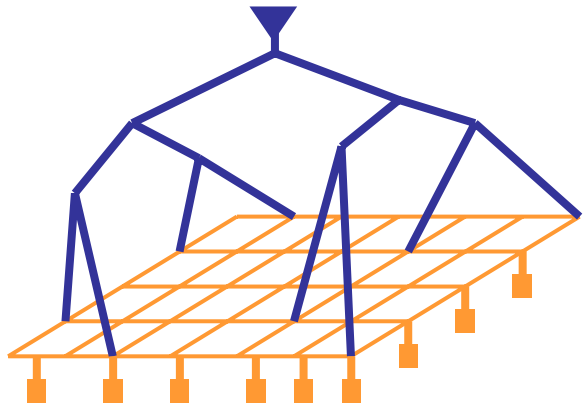
- The reality is of course, MORE COMPLICATED
- Need to consider the interactions between CMP, Litho, RV, CAA, etc.

Variation-Tolerant Clock Synthesis

- Variation reduction => variation tolerance
- Temperature aware clock opt. [ICCAD'05]
- Clock tree link insertion [Rajaram+, ISPD'05, ISQED'06, ISPD'06, ISQED'07]
- Meshworks: clock mesh planning/synthesis [ASPDAC 2008, BPA nominee]



CTS -> add links



Mesh -> remove links?

Conclusions

- 193nm lithography will still be the dominant chip manufacturing workhorse, for next 5+ years
  - Even EUV still has DFM problems
- Synergistic modeling & optimization needed in a unified framework => Design + Mfg Closure
  - DFM in context of DSM
- Current works just scratch the surface
- Need much closer collaborations than ever
  - Between academia and industry (e.g., SRC, IMPACT)
  - Between different camps: design, CAD, process, and system!

Acknowledgment

- Sponsorship by NSF, SRC (core + custom funding from AMD/Cadence/Freescale), IBM, Fujitsu, Qualcomm, Sun, Intel and KLA-Tencor
- PhD students at UTDA: Minsik Cho, Joydeep Mitra, Anand Rajaram, Anand Ramalingam, Sean X. Shi, Peng Yu
- Collaborations/discussions with Dr. Chris Mack, Dr. Ruchir Puri, Dr. Hua Xiang, Dr. Warren Grobman, Dr. Vassilios Gerousis, Dr. Riko Radojcic, et al.