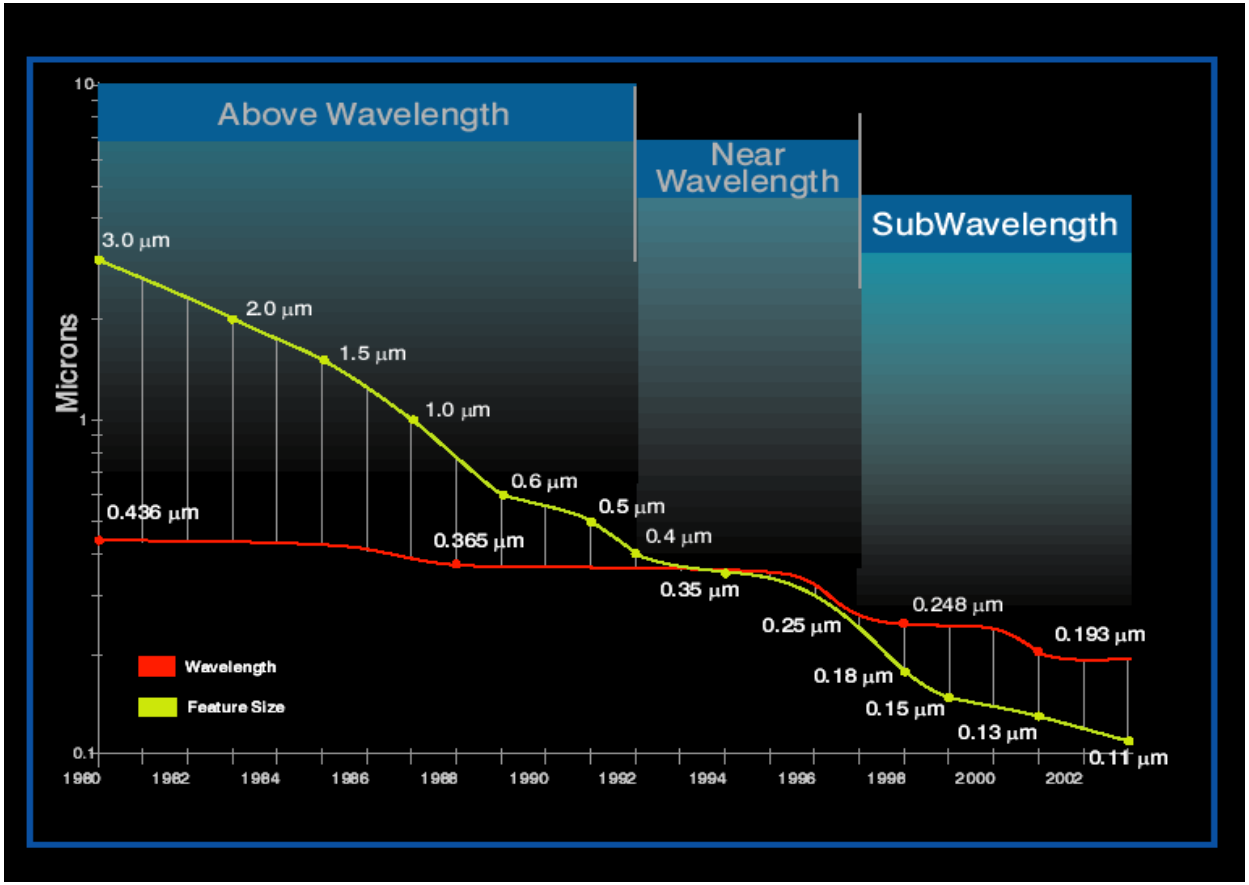
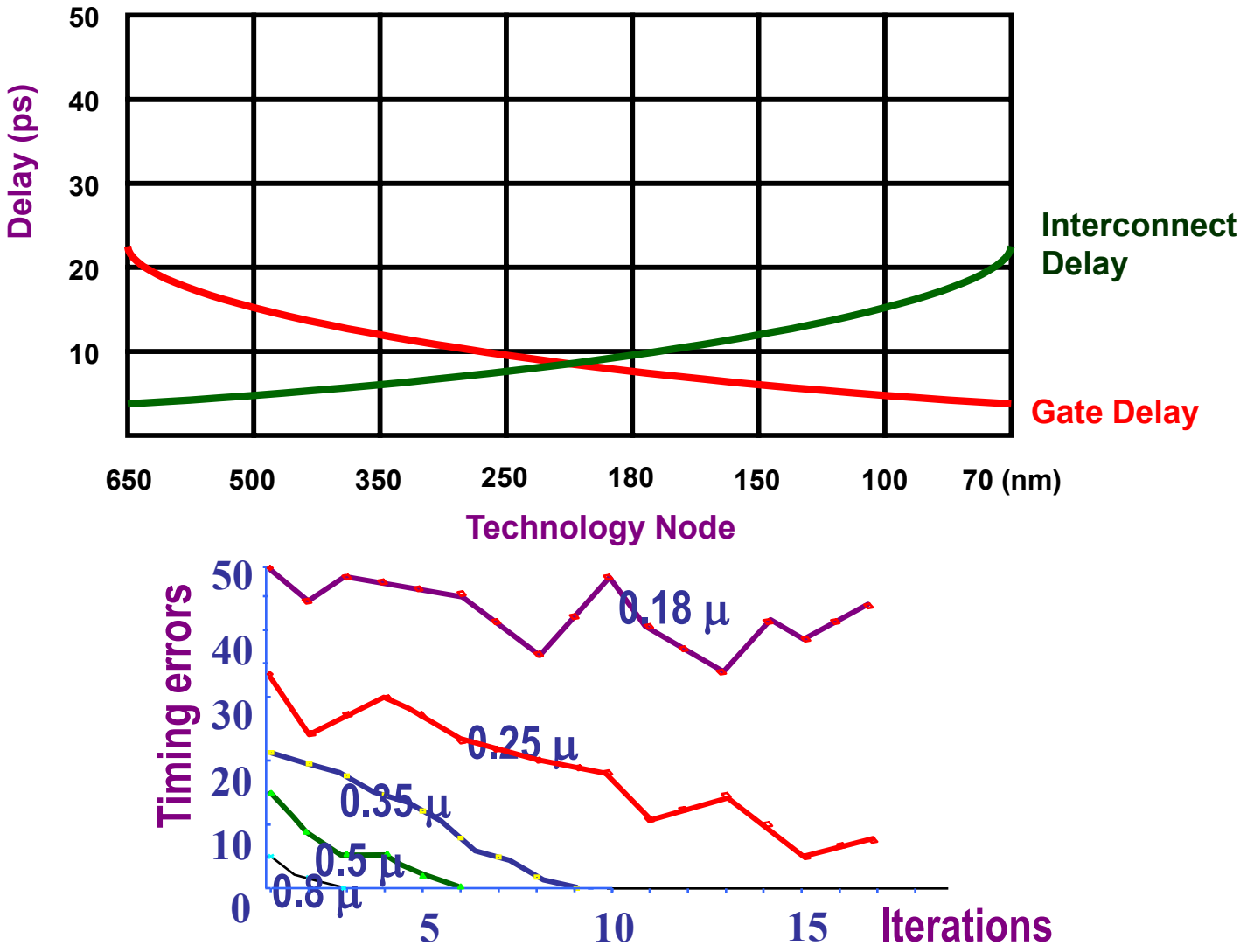
Challenges in Nanometer Physical Design

- Course Contents
  - Interconnect-driven design flow
  - Signal/Power Integrity: crosstalk and IR drop
  - Design for Manufacturability

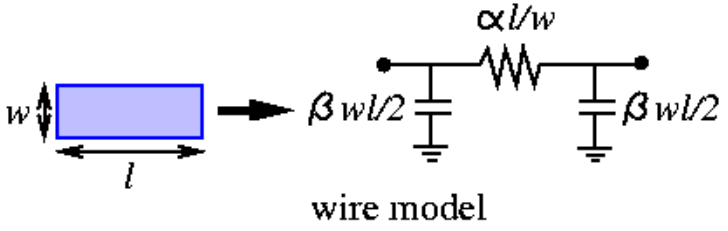


Interconnect Delay Creates the Timing Closure Problem

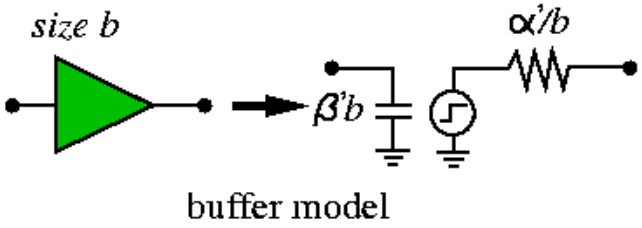


Improving Delay by Buffering

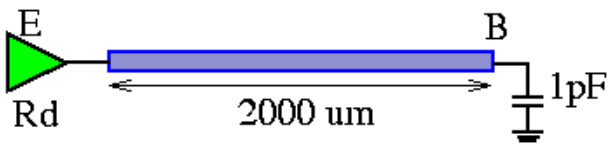
- 0.18 um technology: Wire: α = 0.068 Ω /μm, β = 0.118 *fF*/ μm^2; buffer: α ' = 180 Ω / unit size, β' = 23.4 *fF*/unit size; driver resistance R_d = 180 Ω; unit-sized wire, buffer.

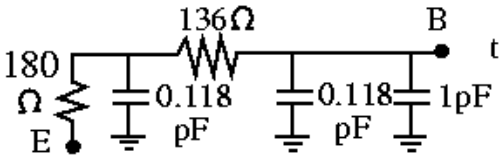


wire model

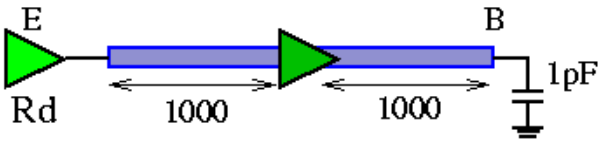


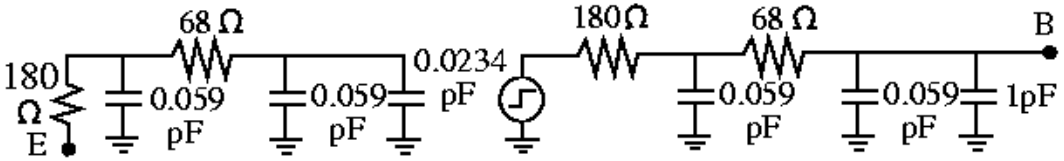
buffer model



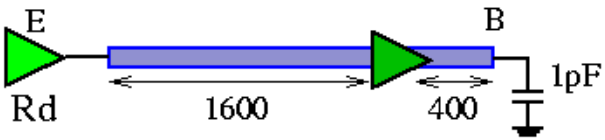


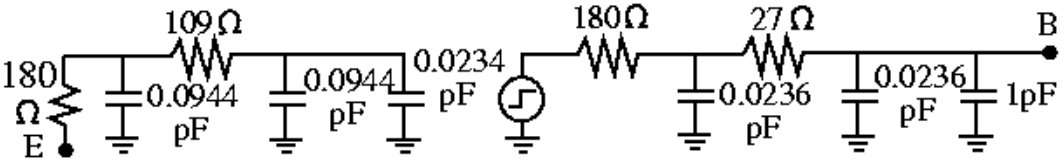
t_{EB} = 180(0.118+0.118+1)+136(0.118+1)
= 375 psec





t_{EB} = 180(0.059+0.059+0.0234)+68(0.059+0.0234)
+ 180(0.059+0.059+1)+68(0.059+1) = 304 psec

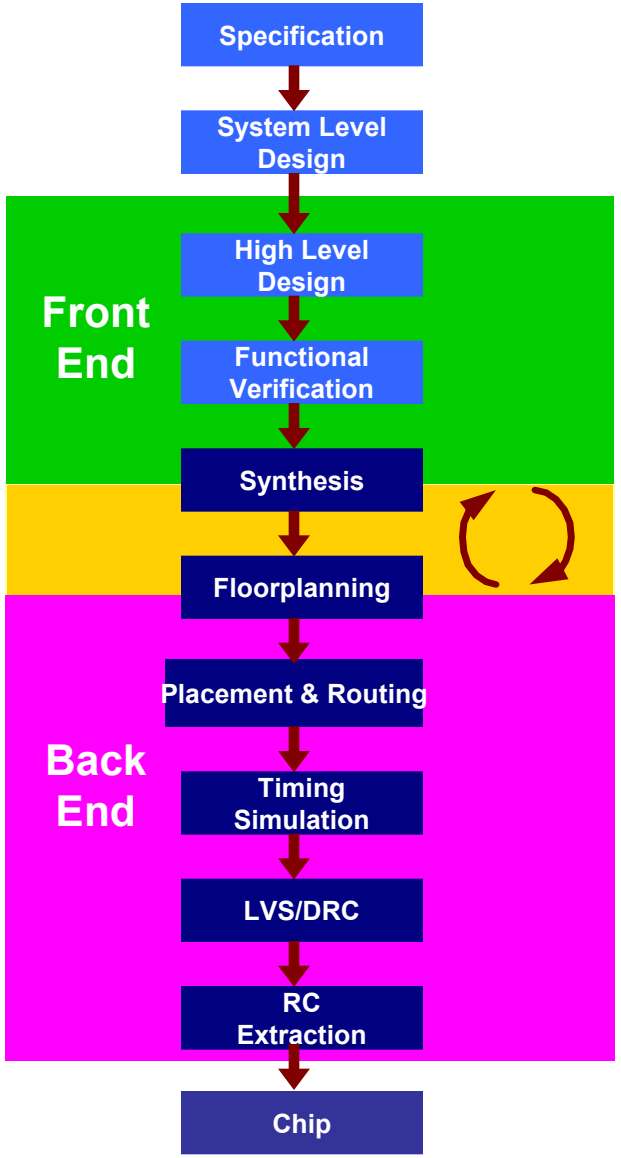




t_{EB} = 180(0.0944+0.0944+0.0234)+109(0.0944+0.0234)
+ 180(0.0236+0.0236+1)+27(0.0236+1) = 267 psec

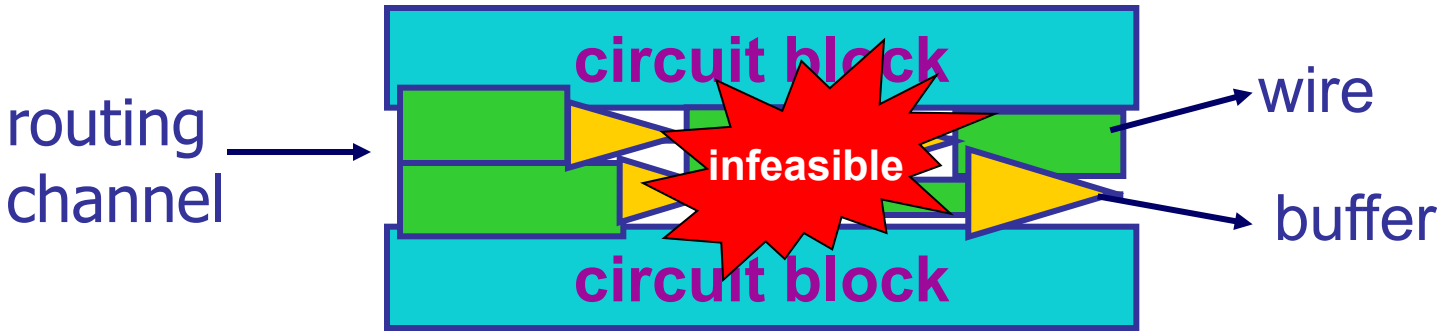
VLSI Design Flow

- Traditional design flow
  - Two-step process
  - Physical design is performed independently after logic design
- In nanometer design
  - Interconnect dominates delay
  - Timing closure
  - Signal integrity
- New design flow
  - Capture real technology behaviors early in the design flow
  - Break the iteration between physical design and logic design

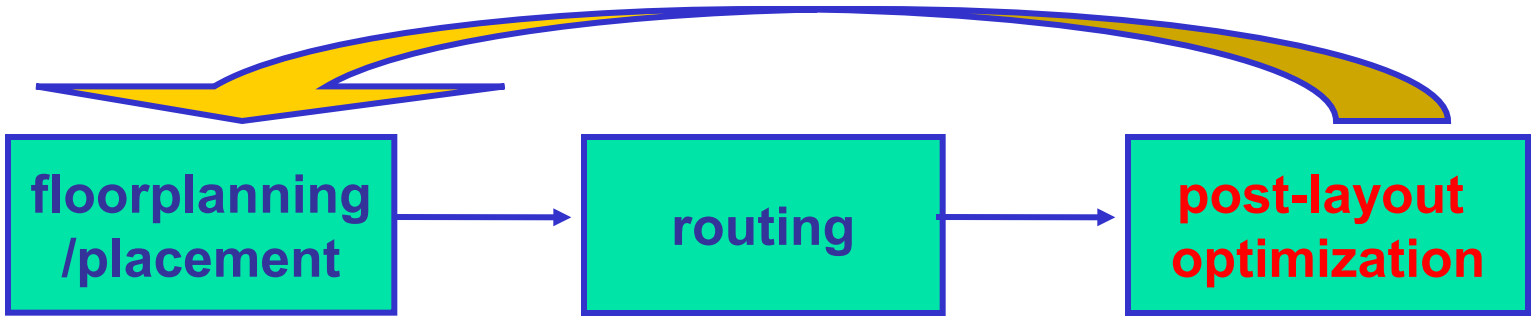


Interconnect-Centric Design for Timing Closure

- Traditional post-layout optimization is not feasible for deep submicron

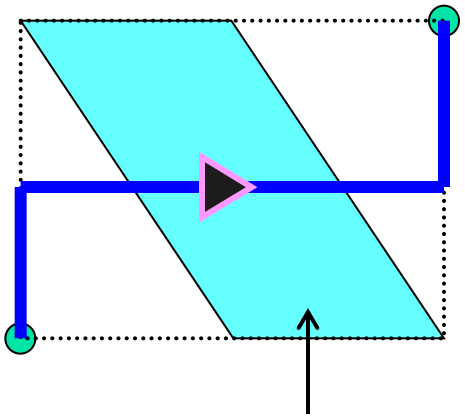


- Shall integrate buffer-block design into floorplanning

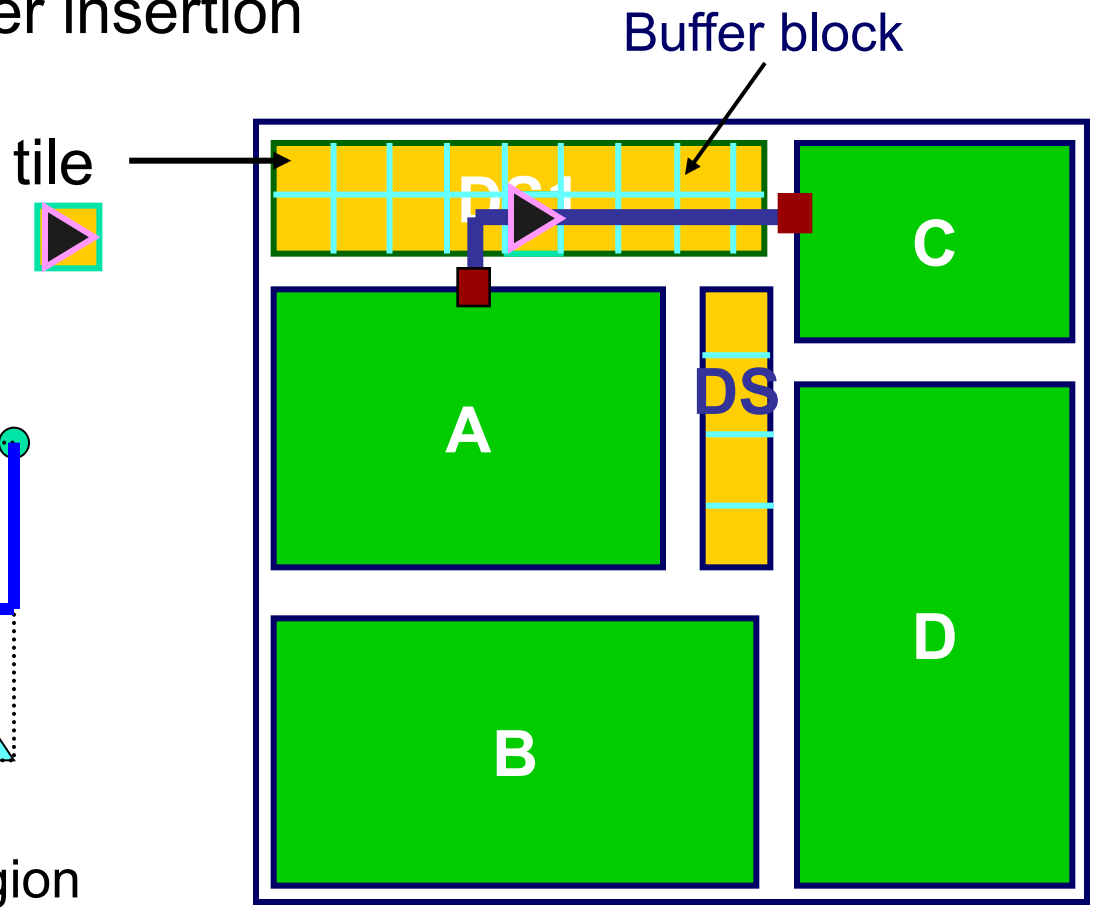


Buffer Block Planning for Timing Closure

- Pick tiles for buffer insertion



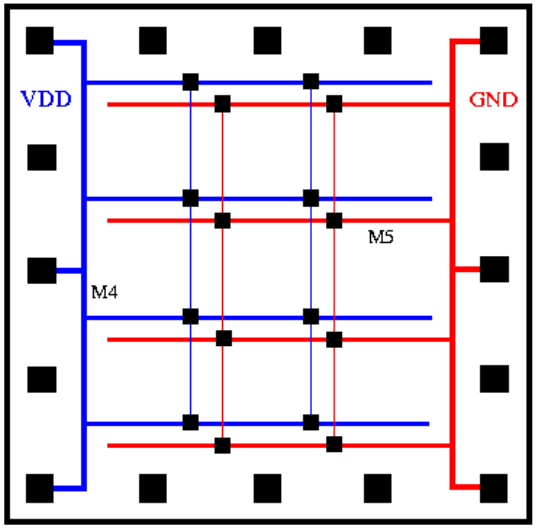
Feasible region



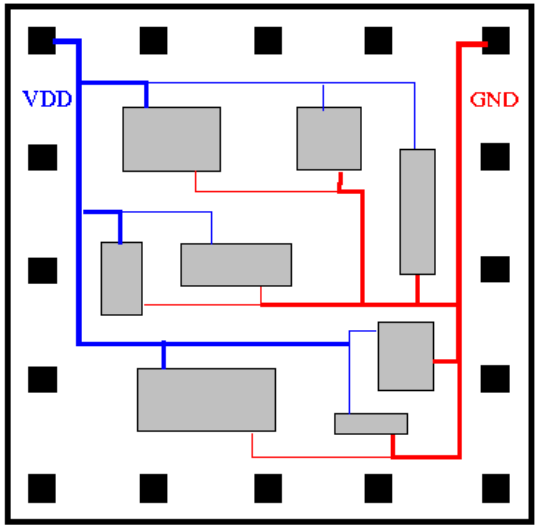
A, B, C, D: circuit blocks
DS: dead space

Power/Ground Routing Revisited

- Are usually laid out entirely on metal layers for smaller parasitics.
- Two steps:
  1. **Construction of interconnection topology:** non-crossing power, ground trees.
  2. **Determination of wire widths:** prevent metal migration, keep voltage drop small, widen wires for more power-consuming modules and higher density current (1.5 mA per μm width for Al). (So area metric?)



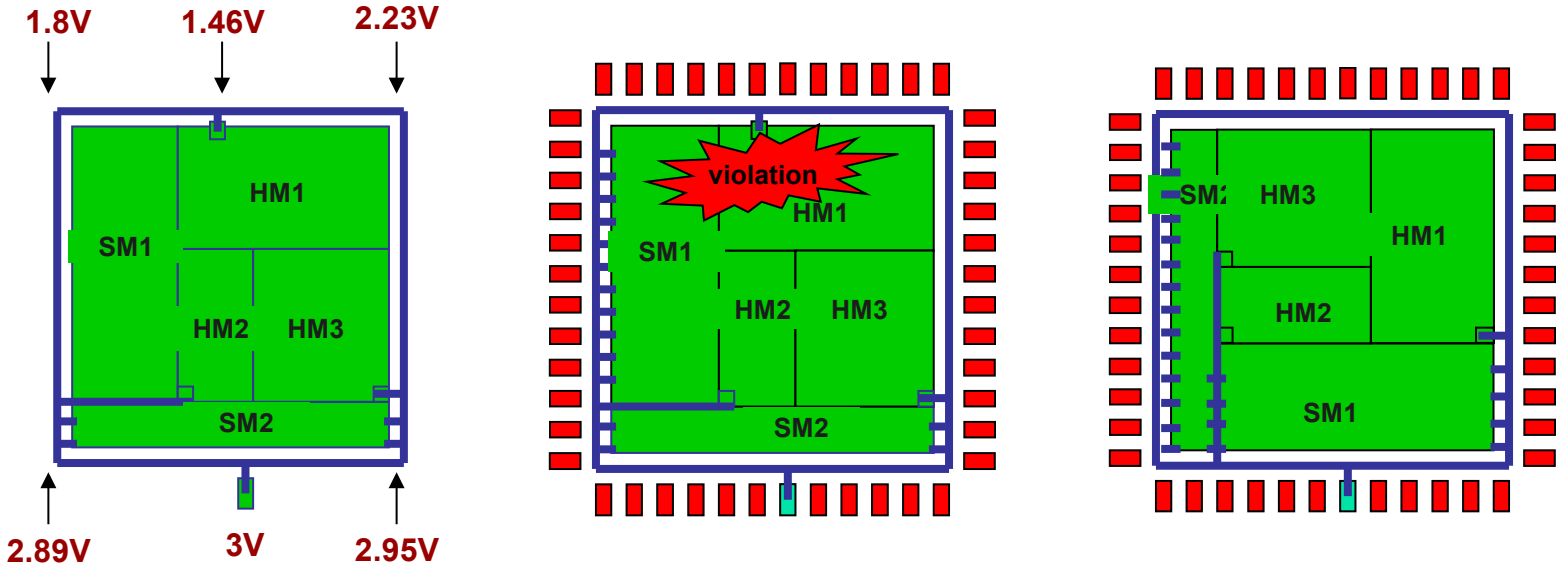
grids



interdigitated trees

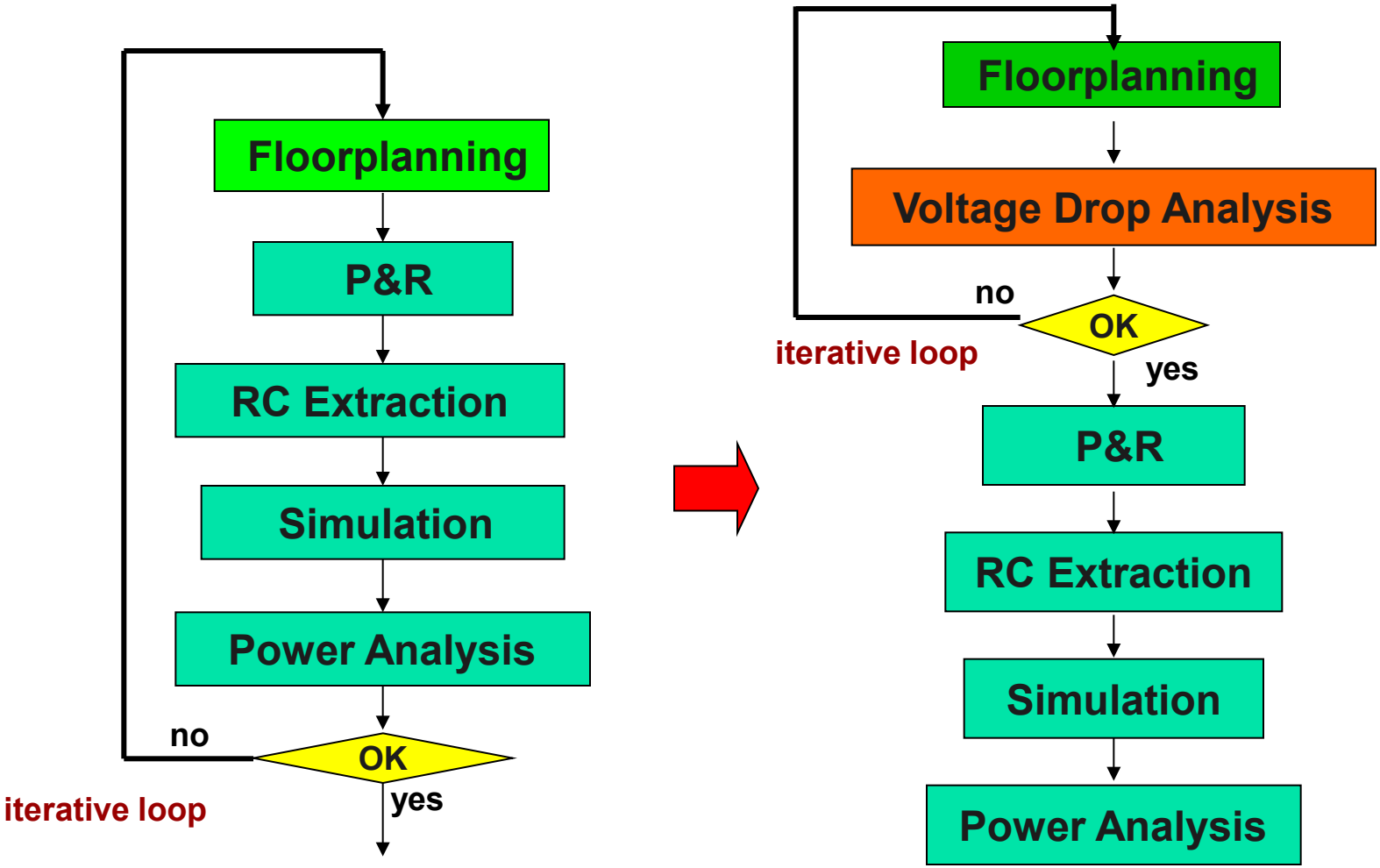
IR (Voltage) Drop

- Power consumption and rail parasitics cause actual supply voltage to be lower than ideal
  - Metal width tends to decrease with length increasing in nanometer design
- Effects of IR drop
  - Reducing voltage supply reduces circuit speed (5% IR drop => 15% delay increase)
  - Reduced noise margin may cause functional failures



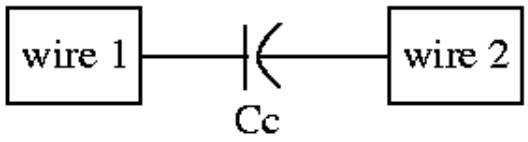
SI-Driven Design Flow for Design Convergence

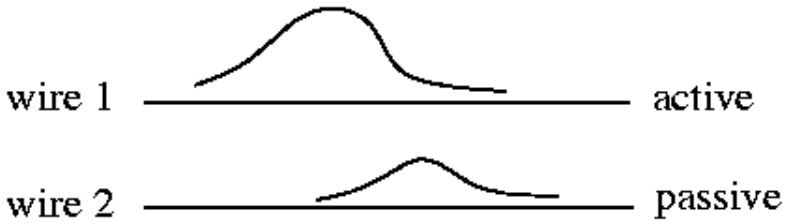
- Typical flow vs. signal integrity-driven design flow



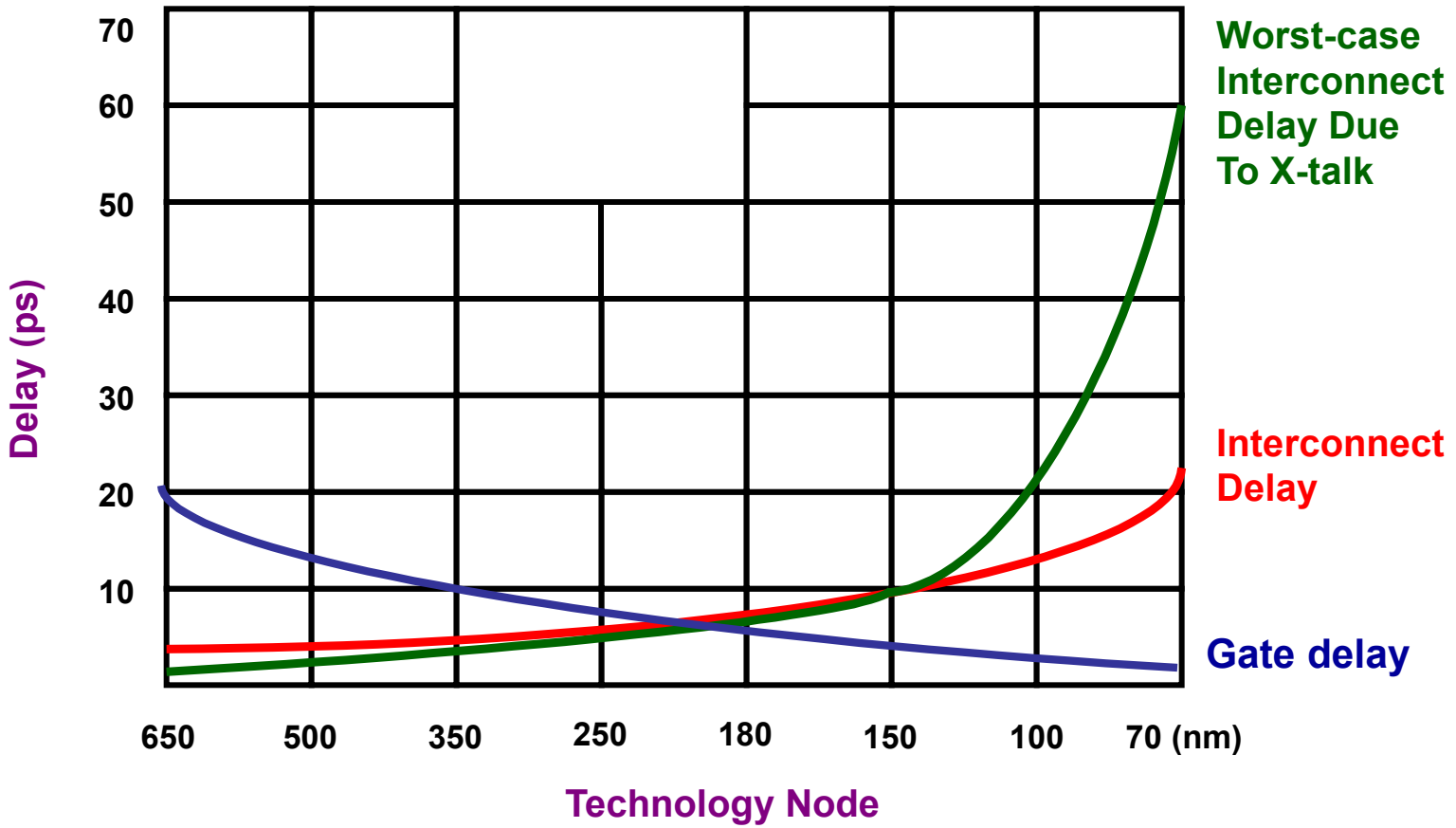
Noise (Crosstalk)

- Wire-to-wire coupling capacitance and inductance introduce **crosstalk**.
- Crosstalk may cause unexpected circuit switching or other undesirable behavior.
- Crosstalk between two wires switching in
  - Different directions: (1) increases signal delays (2) decreases signal integrity.
  - Same directions: (1) decreases signal delays (2) increases signal integrity.
- Capacitive crosstalk proportional to overlapping wire length (switching not the same) and inversely proportional to the distance between wires.





Crosstalk Becomes a First-Order Problem



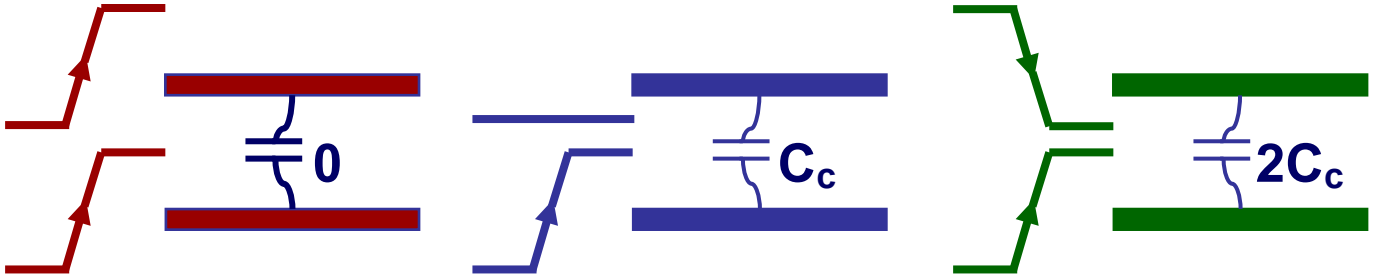
Techniques for Capacitive Crosstalk Reduction

- Capacitive noise is a local effect
- Capacitive crosstalk reduction:
  - Shielding
  - Track permutation: Gao & Liu, ICCAD-93.
  - Buffering: Apert, Devgan, Quay, DAC-98.
  - Wire ordering + Gate/wire sizing: Jiang, Chang, Jou, IEEE TCAD-2K (DAC-99).
  - Wire spacing: Saxena & Liu, DAC-99; Pan & Chang, ICCD-2K.
  - Layer & track assignment: Her, Chang, Chen, Lee, ICCAD-2003.

Crosstalk Modeling Considering Switching Behavior

- When two adjacent wires switch in the same direction, the anti-Miller effect shortens transition.
- When two adjacent wires switch in opposite directions, the Miller effect lengthens transition.
- Crosstalk value should be modeled by the product of switching dissimilarity and coupling capacitance.
- Crosstalk between neighboring wires *i* and *j*:

  crosstalk(i,j) = switching_dissimilarity(i,j) coupling_capacitance(i,j)*



Anti-Miller Effect

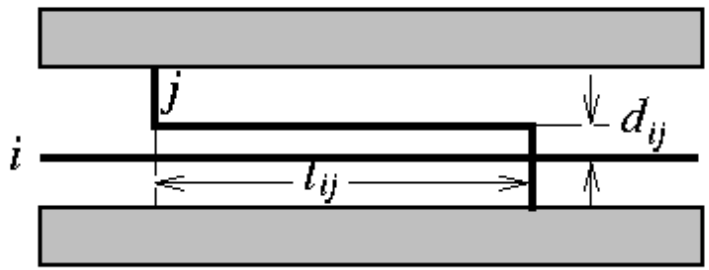
shortens transition

Miller Effect:

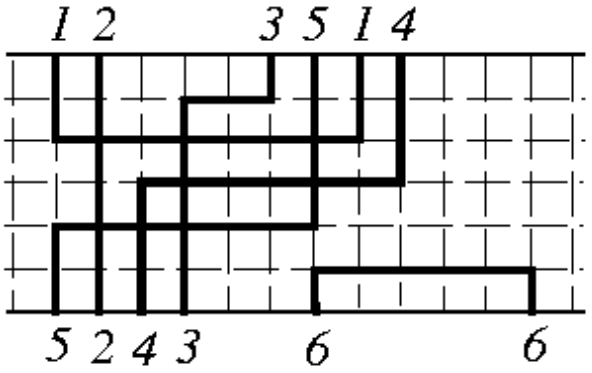
lengthens transition

Simplified Capacitive Crosstalk Computation

- Simplification in gridded channel routing
  - Adjacent tracks or columns only
  - Horizontal or vertical segments overlapping only
  - Linear distance model (k = 1)
- Example: k = 1, $C_{ij} = l_{ij}$ (adjacent tracks/columns only)
  - Net 4: Consider both vertical and horizontal segments $\Rightarrow C_{4,j}$ = 3+3+4+5+2 = 17
  - Net 4: Consider horizontal segments only $\Rightarrow C_{4,j}$= 4+5 = 9



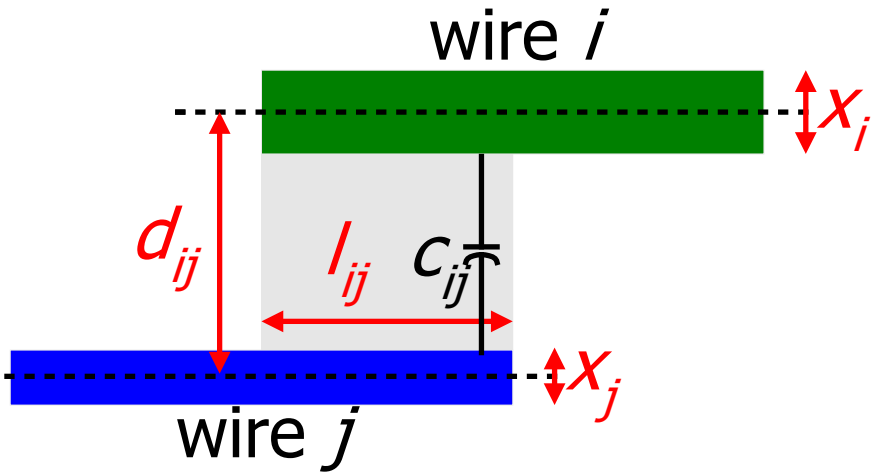
$$C_{ij} = a\frac{l_{ij}}{(d_{ij})^k}$$



$$C_{4,j} = 3 + 3 + 4 + 5 + 2 = 17$$

Coupling Capacitance

- $c_{ij} = \dfrac{\hat{f}_{ij}\, l_{ij}}{d_{ij} - \dfrac{x_i + x_j}{2}} = \dfrac{\hat{f}_{ij}\, l_{ij}}{d_{ij}} \dfrac{1}{1 - \dfrac{x_i + x_j}{2\, d_{ij}}}$



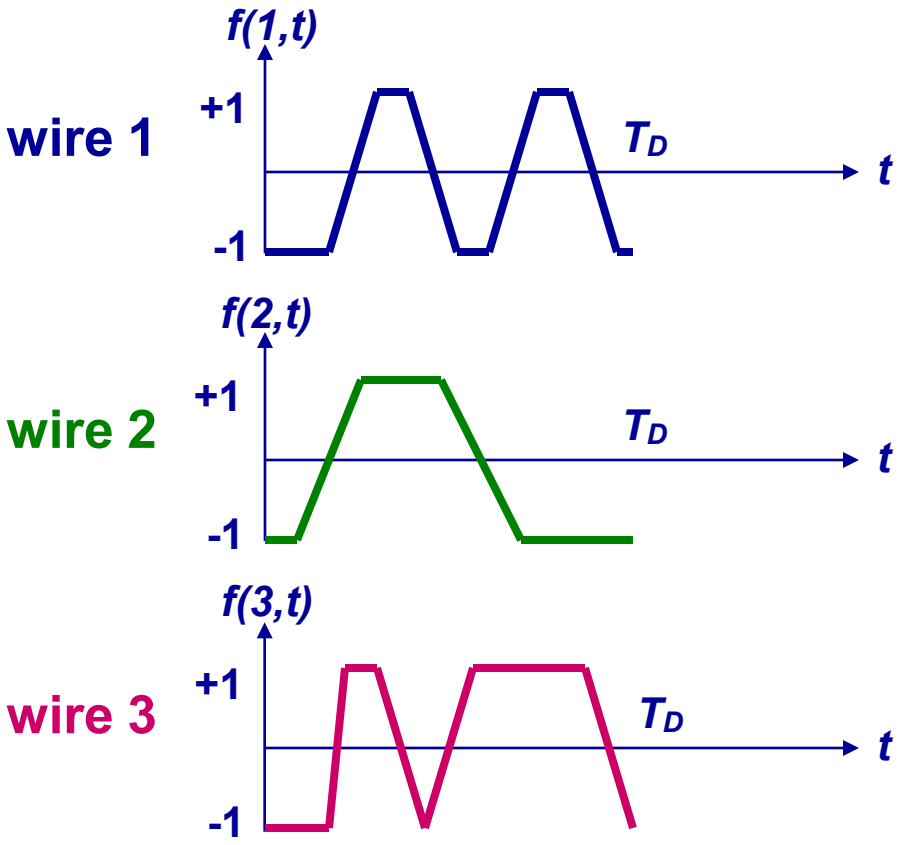
$\hat{f}_{ij}$: **unit-length fringing capacitance**

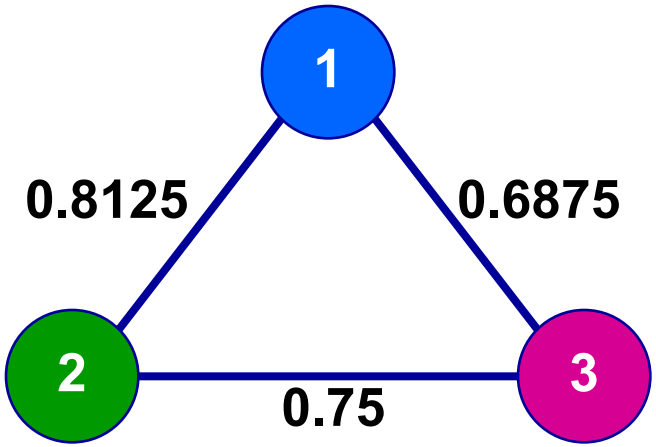
- Include coupling effect into wire capacitance
  - consider crosstalk effect on delay
  - $c_i = \hat{c}_i x_i + f_i + 2\Sigma c_{ij}$

Switching Dissimilarity

Switching dissimilarity between wires *i* and *j*:

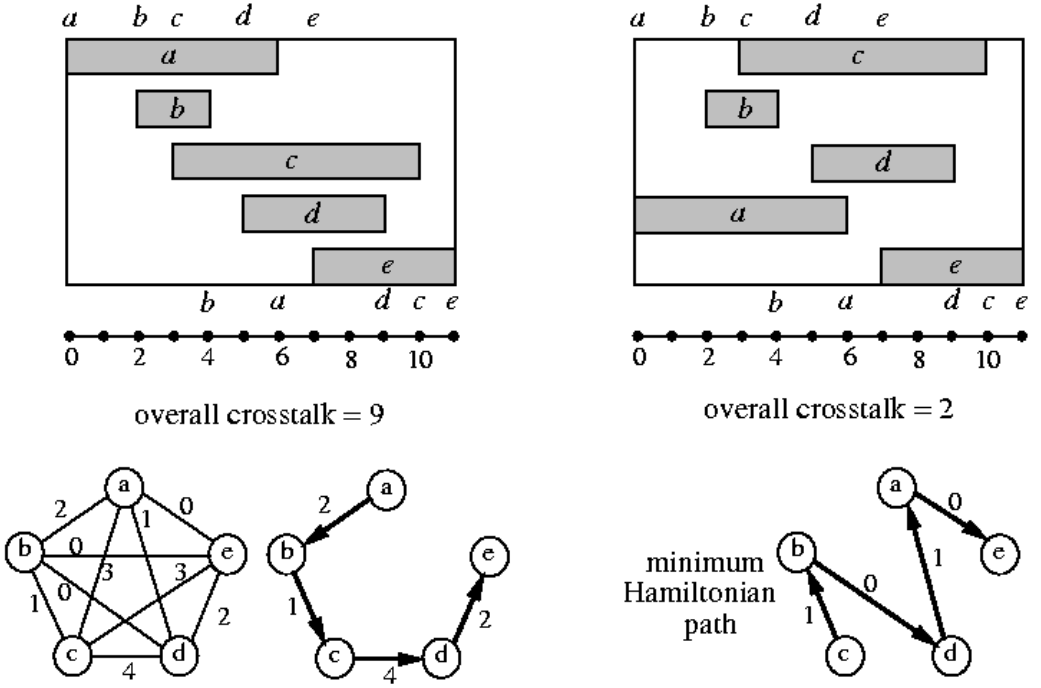
$$switching\_dissimilarity(i,j) = 1 - \frac{\int_0^{T_D} f(i,t) f(j,t) dt}{T_D}$$



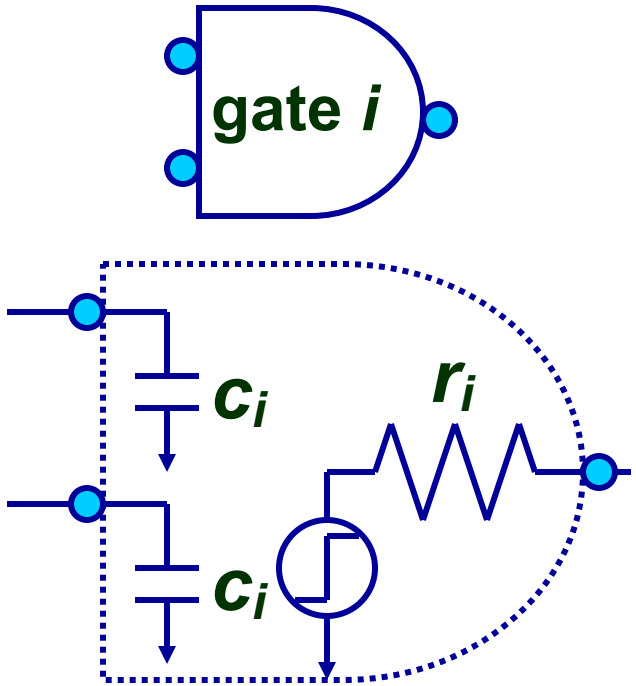


Wire Ordering for Crosstalk Minimization

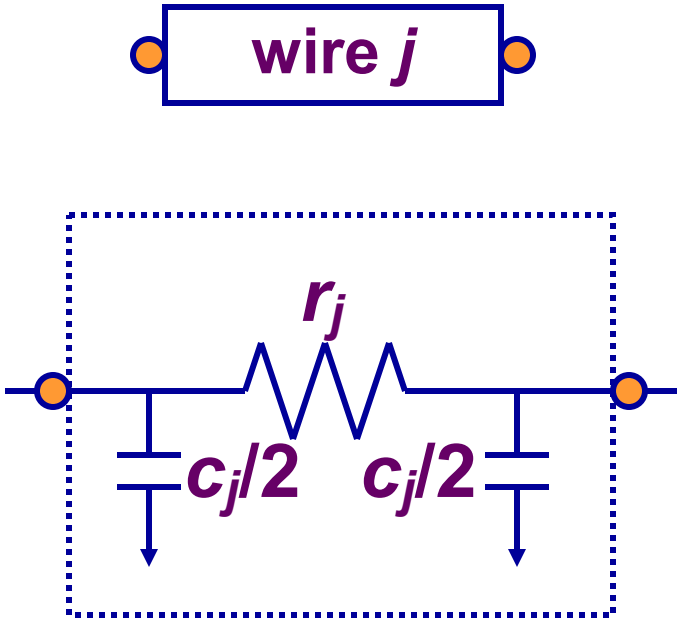
- Goal: Find a wire ordering s.t. the overall crosstalk is minimized.
  - Construct a complete weighted graph $G=(V, E, W)$: $V \leftrightarrow$ wires, $W \leftrightarrow$ coupling length between each pair of wires.
  - The minimum weighted Hamiltonian path induces the minimum crosstalk.



Circuit Model



$$r_i = \hat{r}_i / x_i$$
$$c_i = \hat{c}_i x_i$$



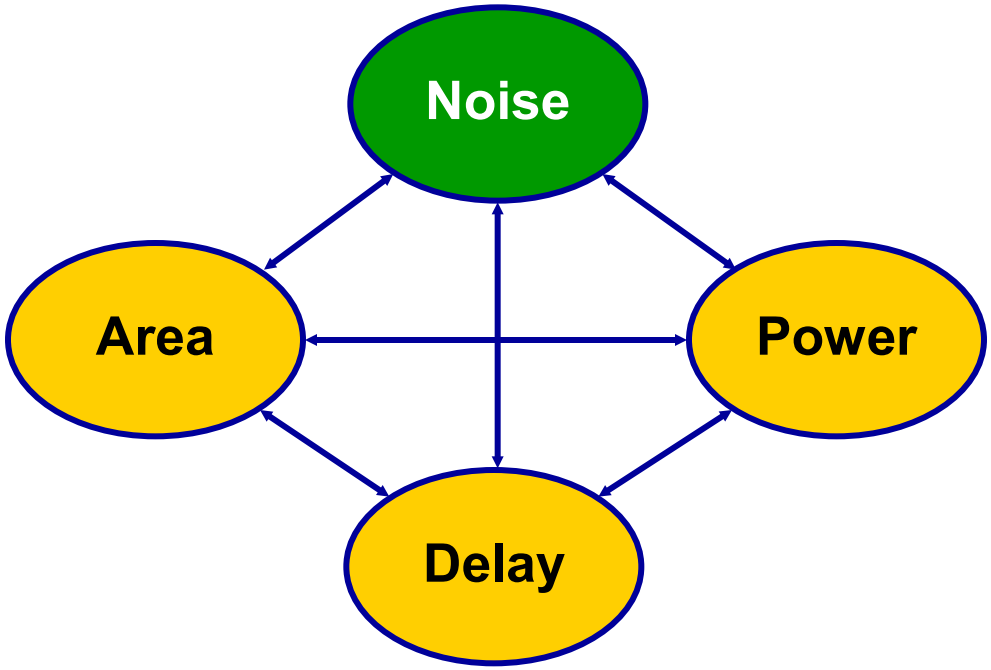
$$r_j = \hat{r}_j / x_j$$
$$c_j = \hat{c}_j x_j + f_j$$

Elmore delay model: $D_i = r_i C_i$

D_i: delay of node i; C_i : downstream capacitance

Crosstalk-Constrained Optimization

- NOISE is a crucial concern in nanometer technology



Goal: simultaneous optimization

Taming Noise and Other Objectives

Switching Dissimilarity Consideration

Wire Ordering

Simultaneous post-layout optimization

by sizing circuit component

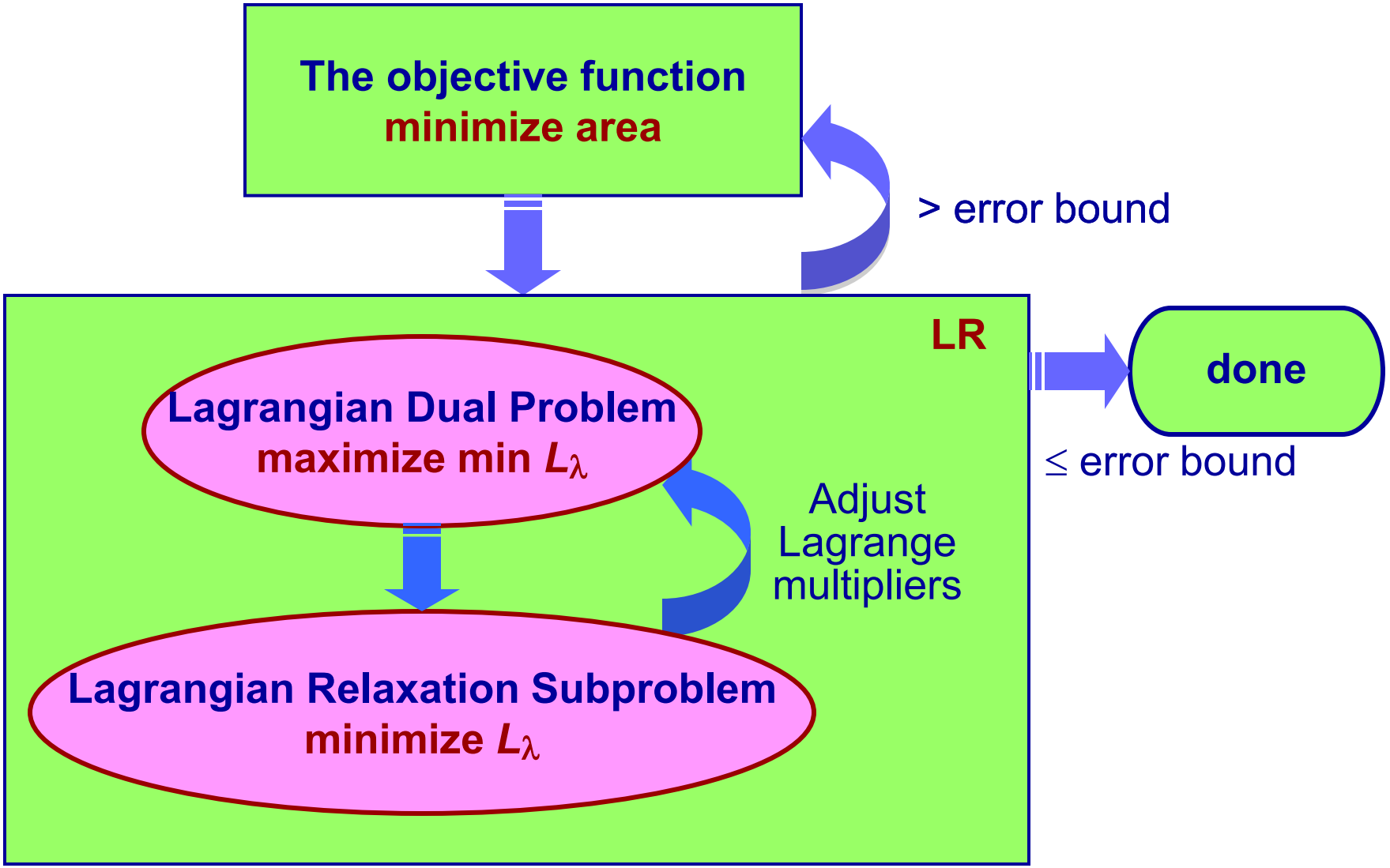
Noise **Area** **Delay** **Power**

Lagrangian Relaxation

Crosstalk-constrained Multi-Objective Optimization

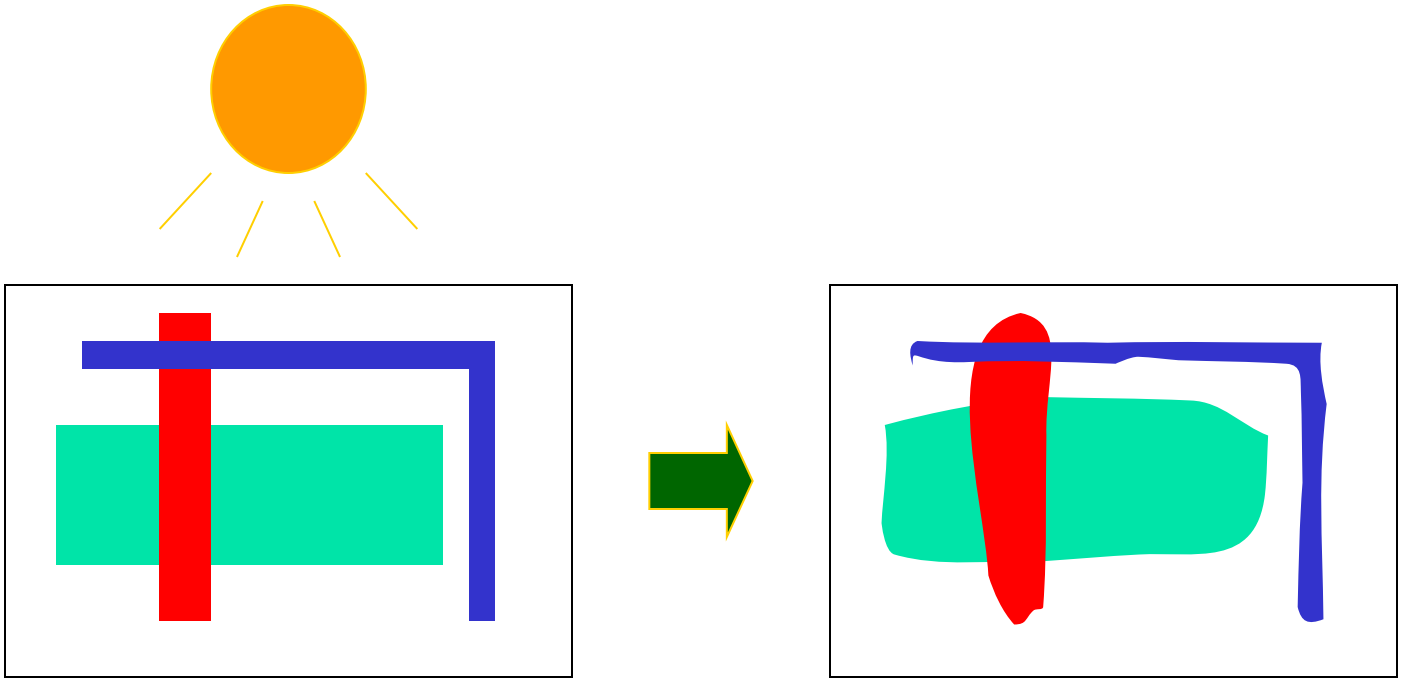
M:		
Minimize	A	Area
Subject to	$D \le D^B$	Delay constraints
	$X \le X^B$	Crosstalk constraints
	$P \le P^B$	Power constraints
	$\mathbf{L} \le \mathbf{x} \le \mathbf{U}$	Sizing constraints

Optimal Gate and Wire Sizing



Design for Manufacturability: Process Variation

- Results from subwavelength lithography
- May create unexpected circuit behavior
- Requires design insensitive to process variation



FEOL Variation

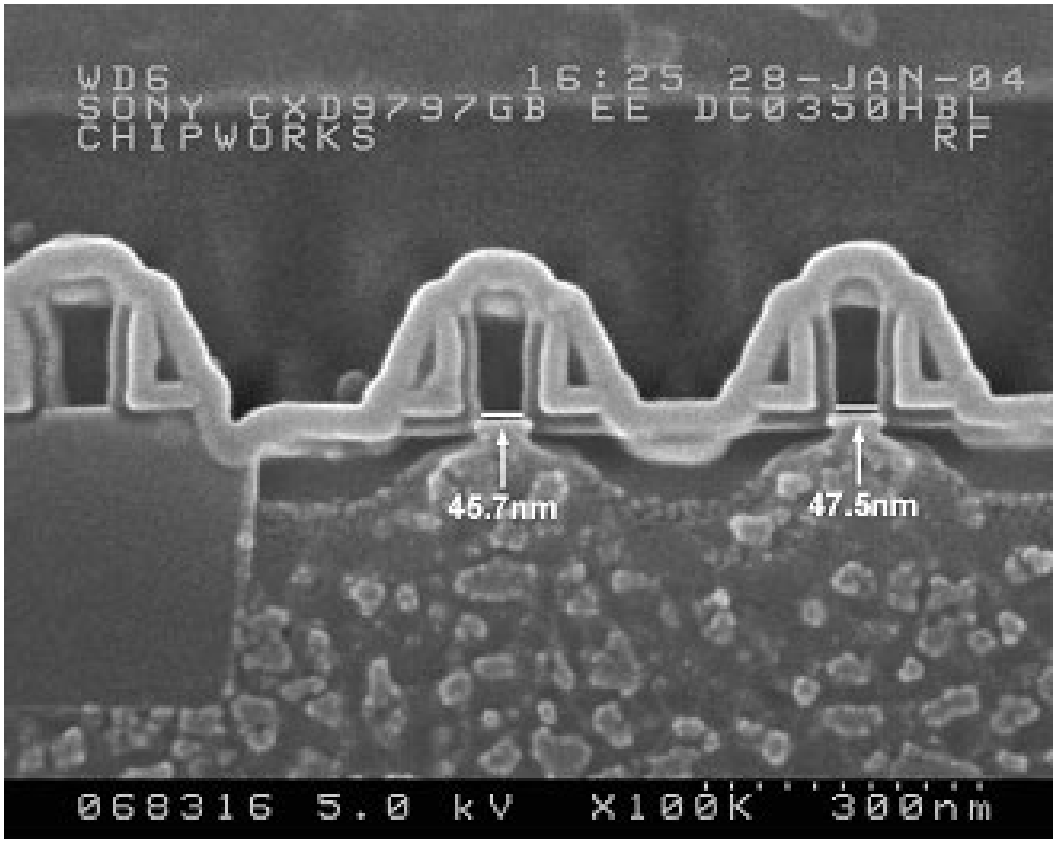
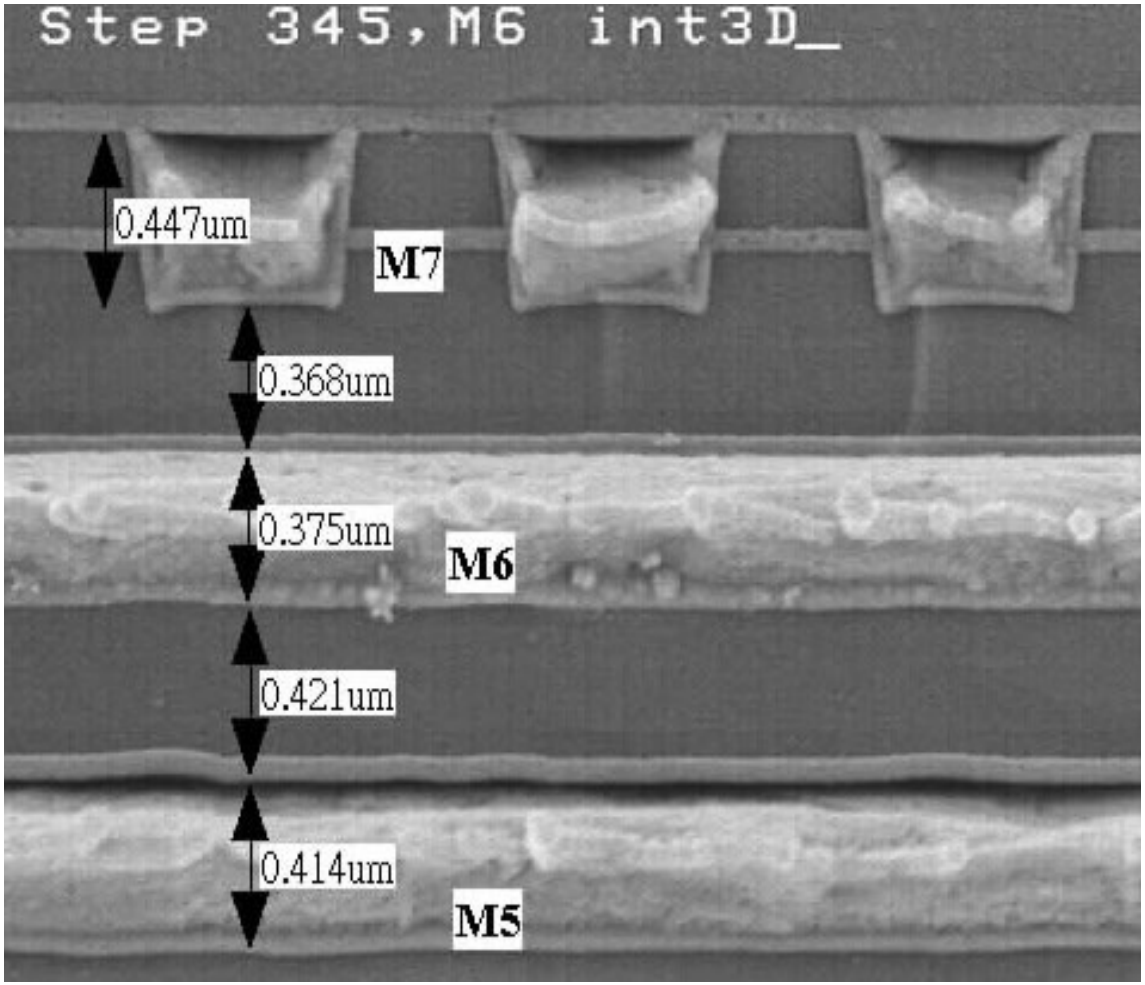


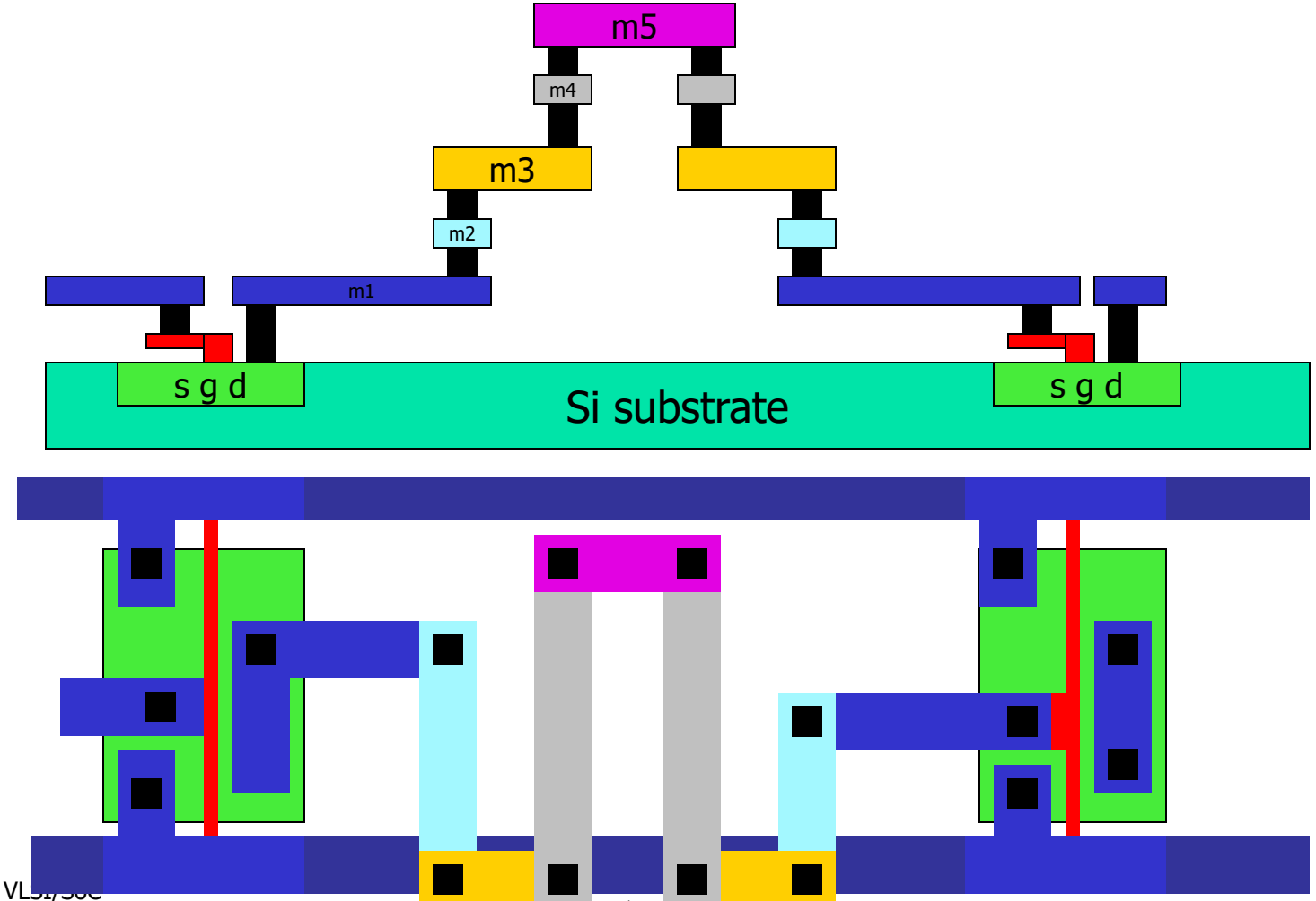
Figure 1 – Deep sub-micron transistors.

BEOL Variation



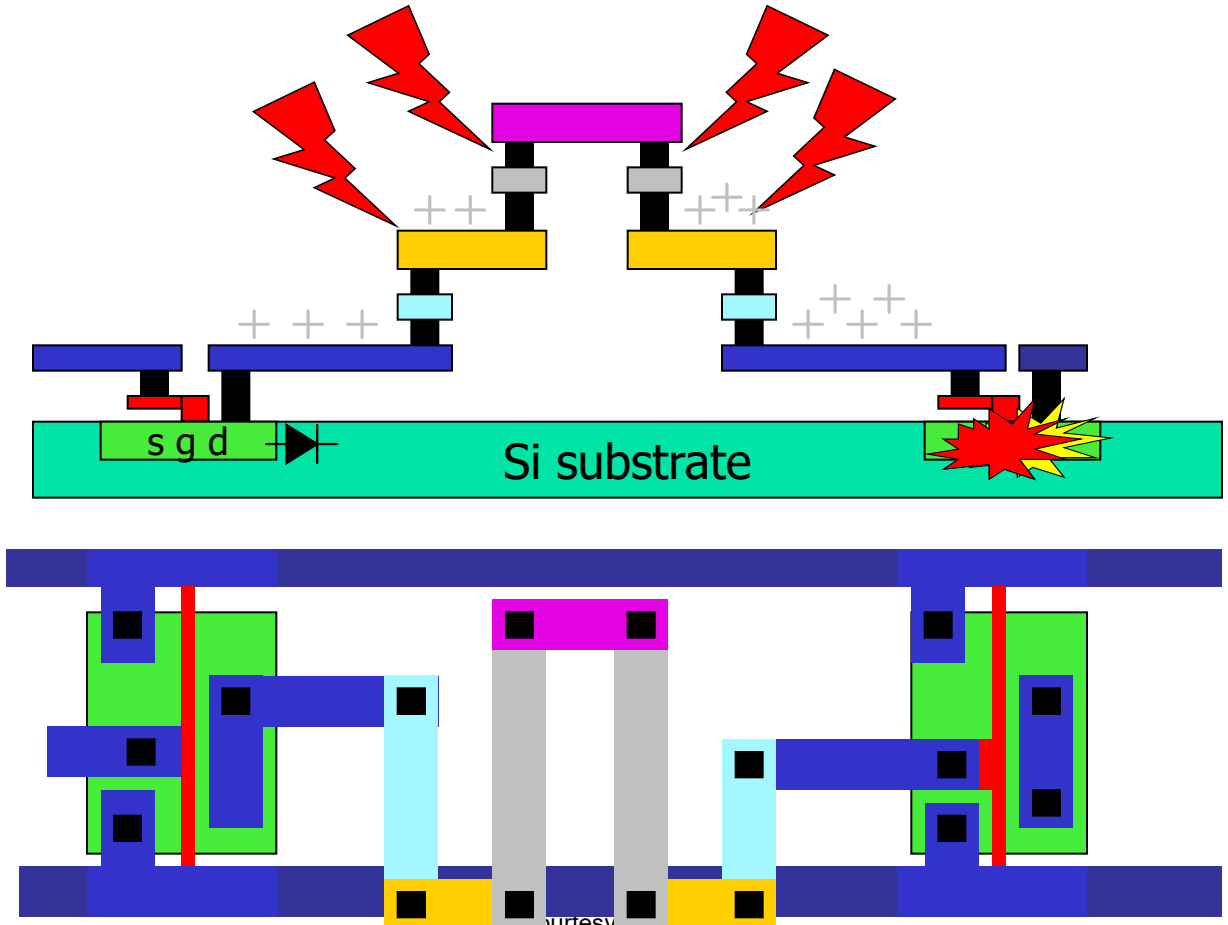
Design for Manufacturability: Antenna Effect

- Notes courtesy of Dr. Patrick Groeneveld
- Layout for a 5-layer interconnect:



Antenna Effect: The Problem

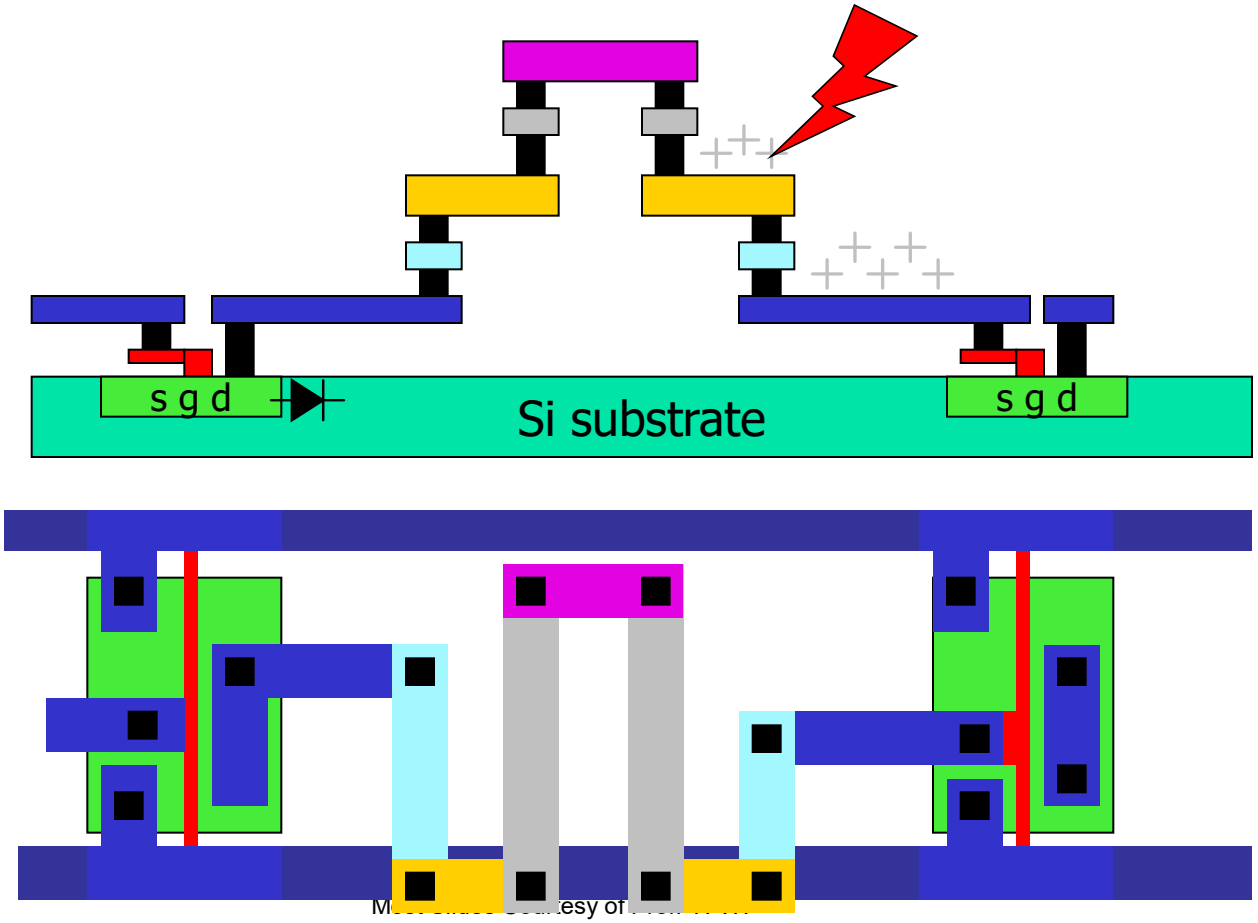
- Unconnected wires act as “antennas” that pick up electrical charge gained from plasma etching and CMP
- The longer the wires, the more the charge.



Chang, T.-C. Wang and KJ Chang

Antenna Effect (1/2)

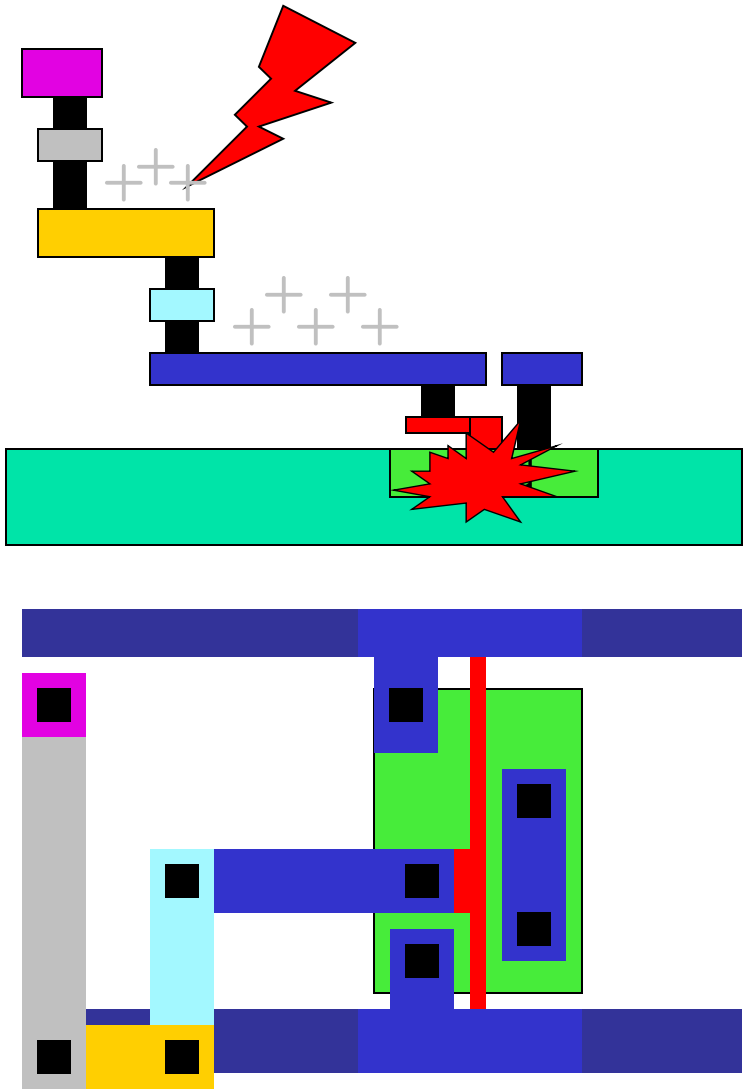
- Wires are always shorted in the highest metal layer.
- 0.18 (0.13) um technology: the maximum length of an "antenna" wire is 500 um (20 um).



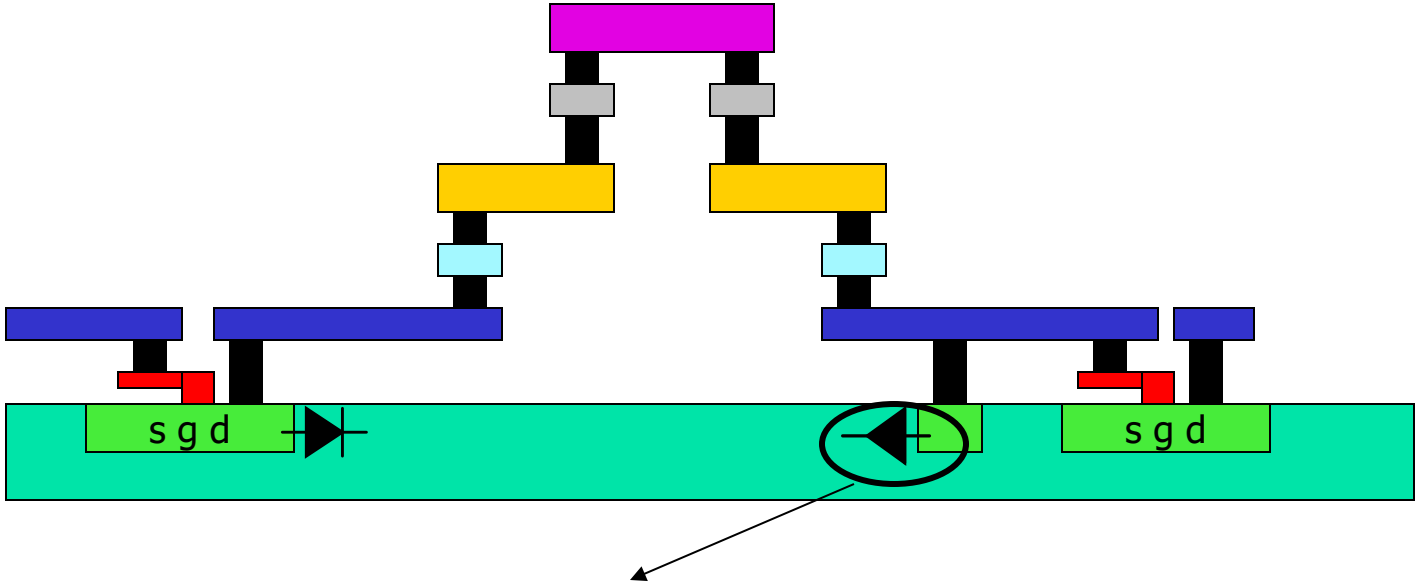
Chang, T.-C. Wang and KJ Chang

Antenna Effect (2/2)

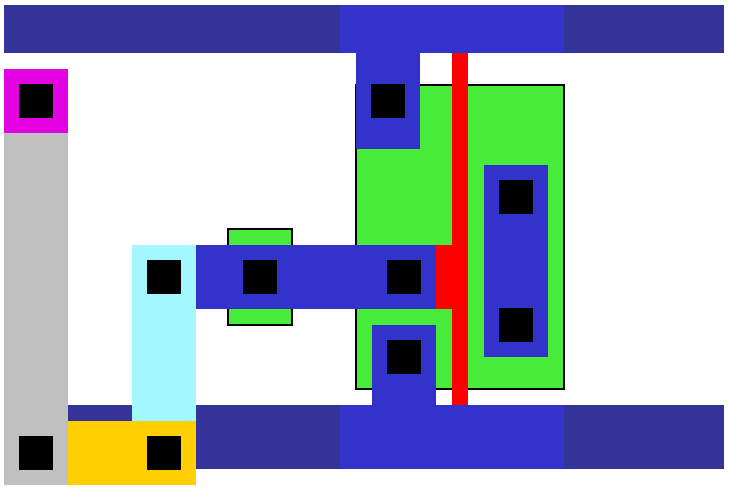
- Depends on the gate size
  - Aggressive down sizing makes the problem worse!
- Depends on length of the part of the wire that is “unshorted” (that is, not connected to a diffusion drain area)
- The calculation of this design rule is surprisingly different per fab.



Fixing Antenna Effect Using Diodes

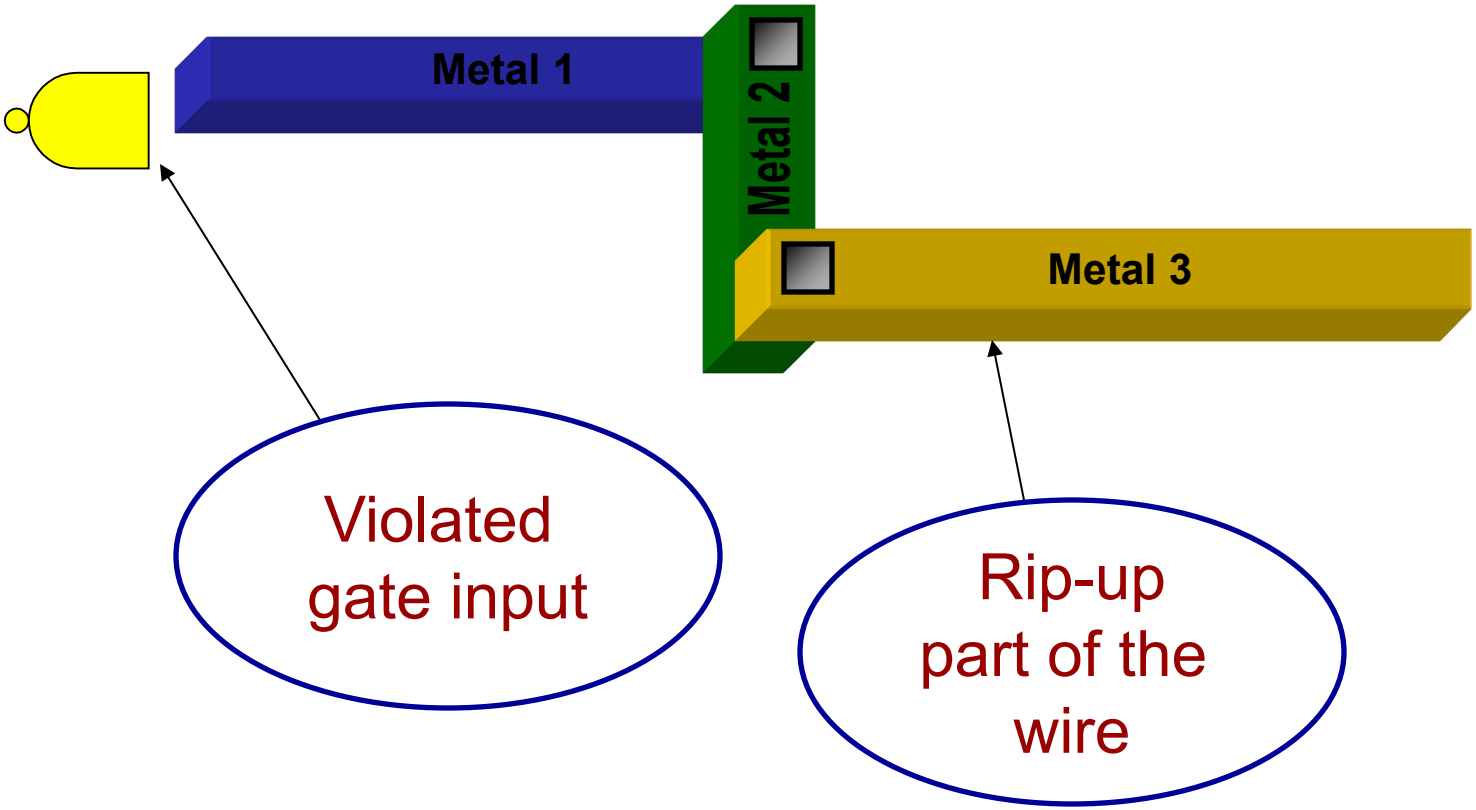


- Insert a diode cell next to each input.
  - Costs significant area
  - Adds capacitance at worst spot.



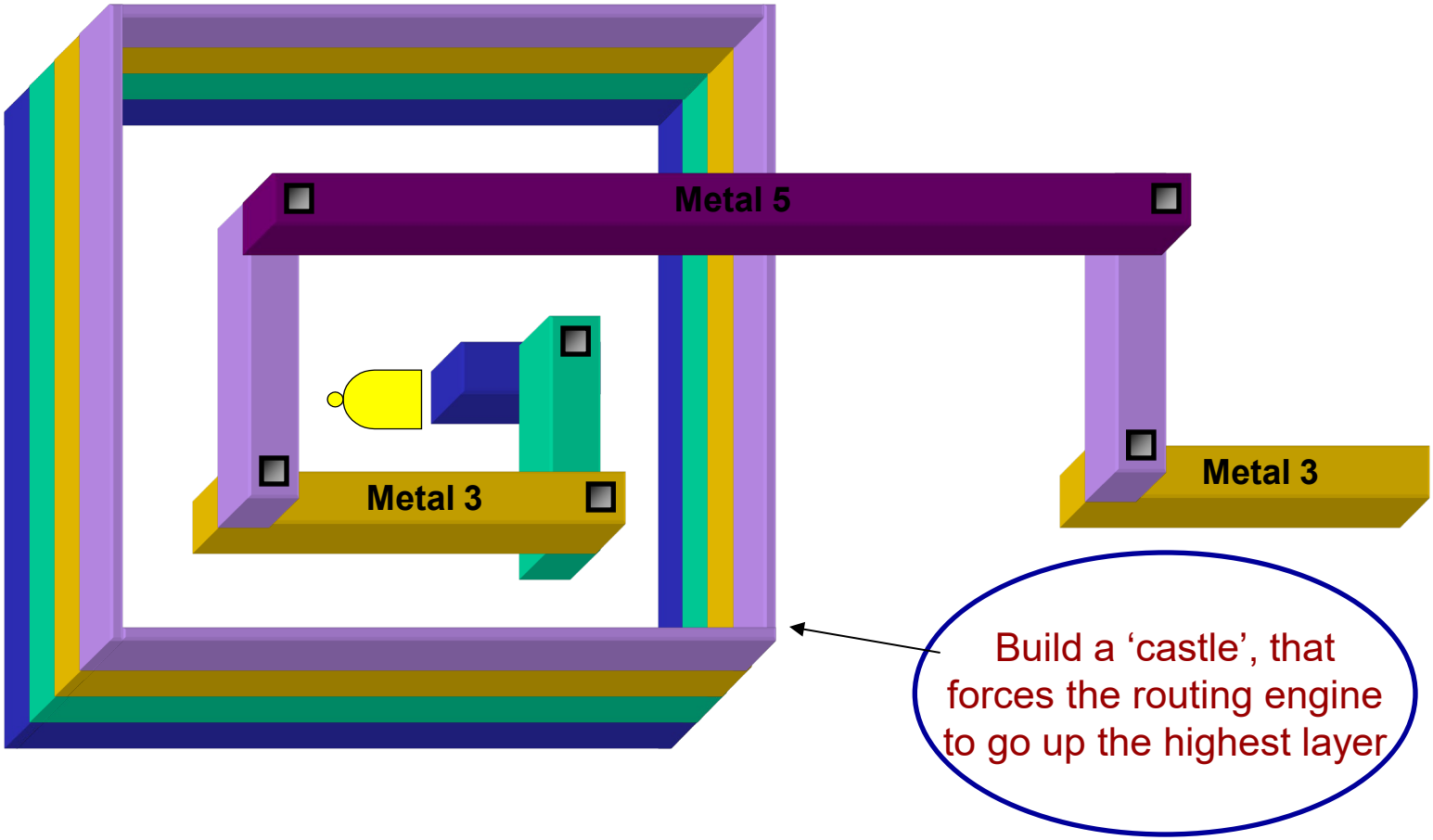
Fixing Antenna Effect through Jumpers

- Adding 'jumpers' after routing
  - The idea: Force a routing pattern that "shoots up" to the highest layer as soon as possible.



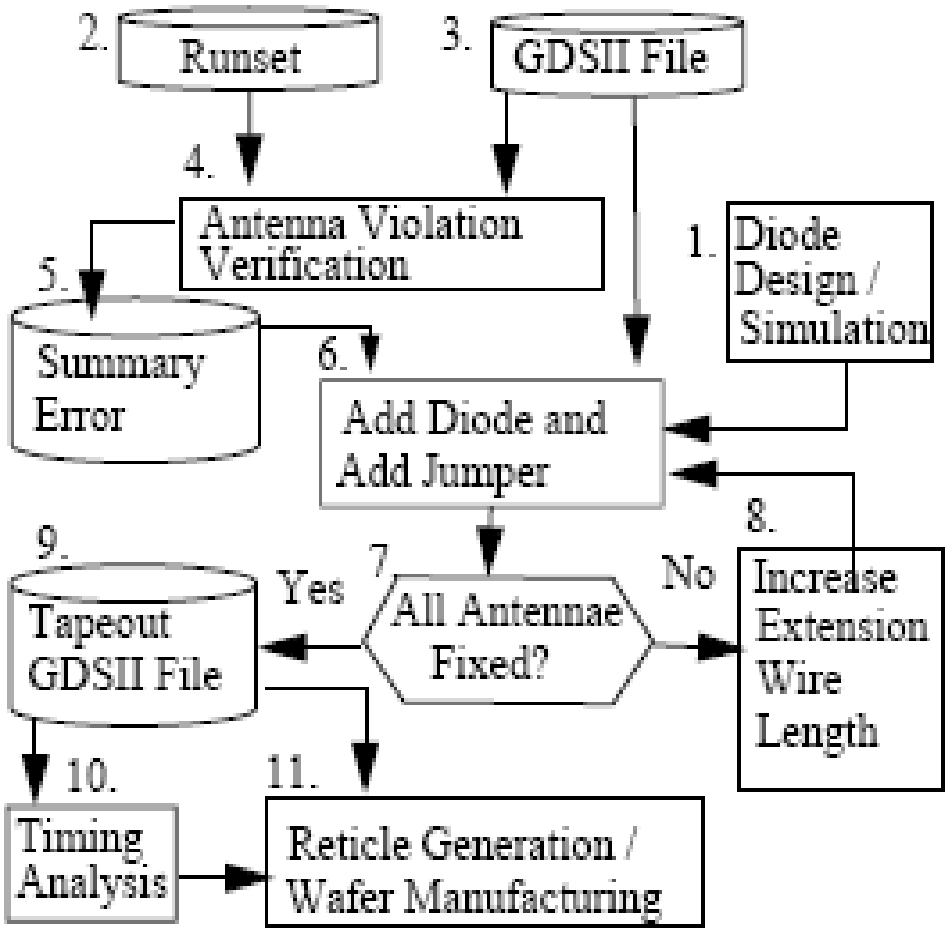
Routing Jumpers

- Adding “jumpers” after routing

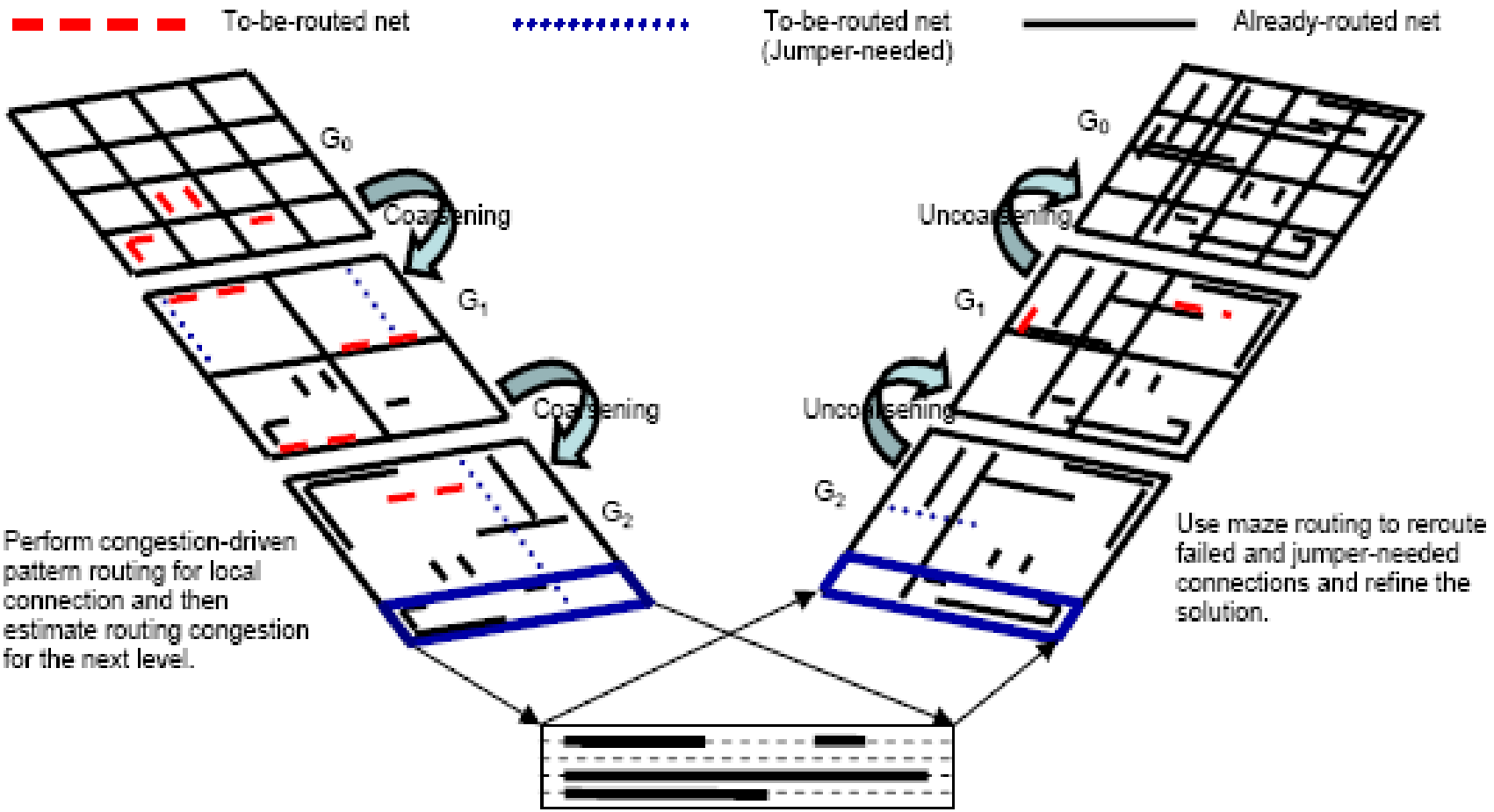


Overall flow for fixing antenna problems

Antenna ratio violation:
wire area/gate area ≤ spec ratio



Source: "Fixing antenna problem by dynamic diode dropping and jumper insertion," ISQED00

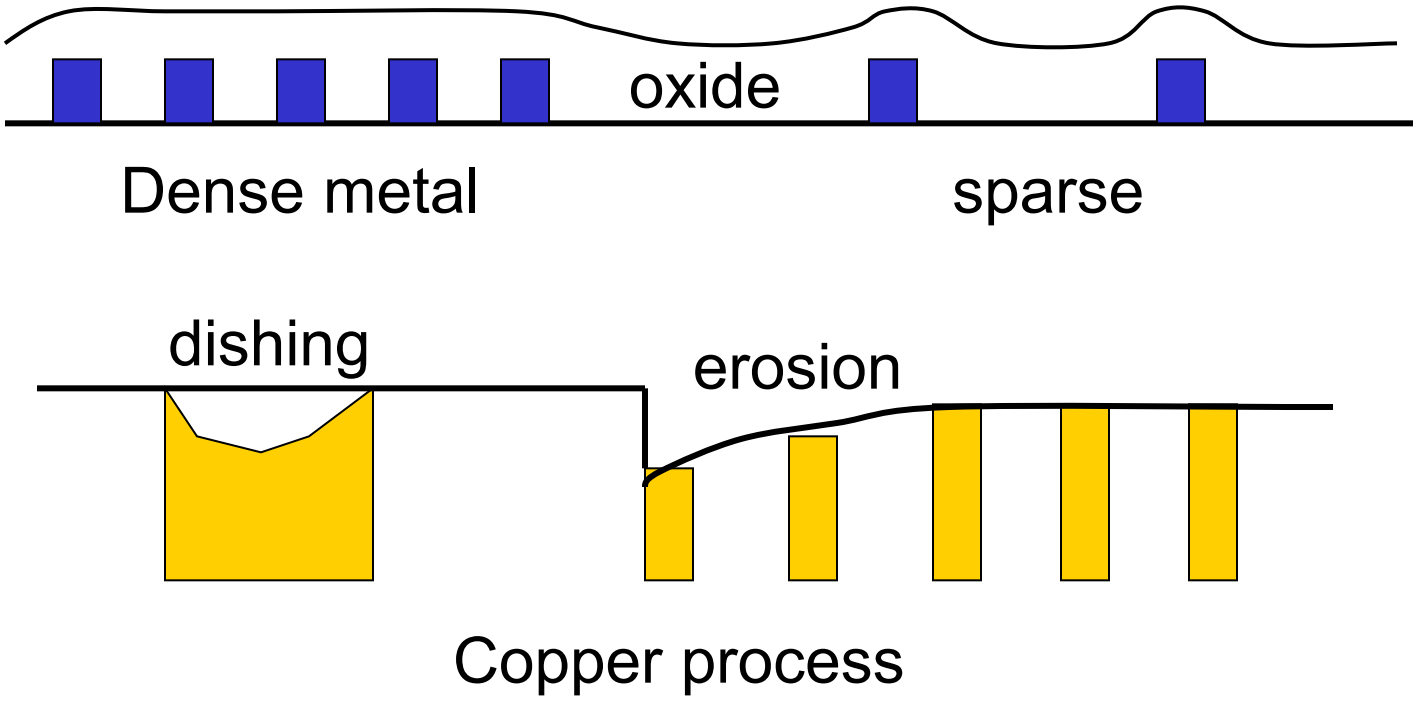


- Break in two those segments of two-pin nets that need jumpers, if the length of them exceeds the minimum allowable gate-strength.
- If they have not exceeded the minimum allowable gate-strength, then try to assign the remaining segments to the highest layer.
- Perform track assignment for long segments, and route short segments by a maze router.
- Perform an antenna check process for every terminal. If nets have antenna violation, rerouted them at the uncoarsening stage.

Source: "Multilevel routing with antenna avoidance," ISPD04

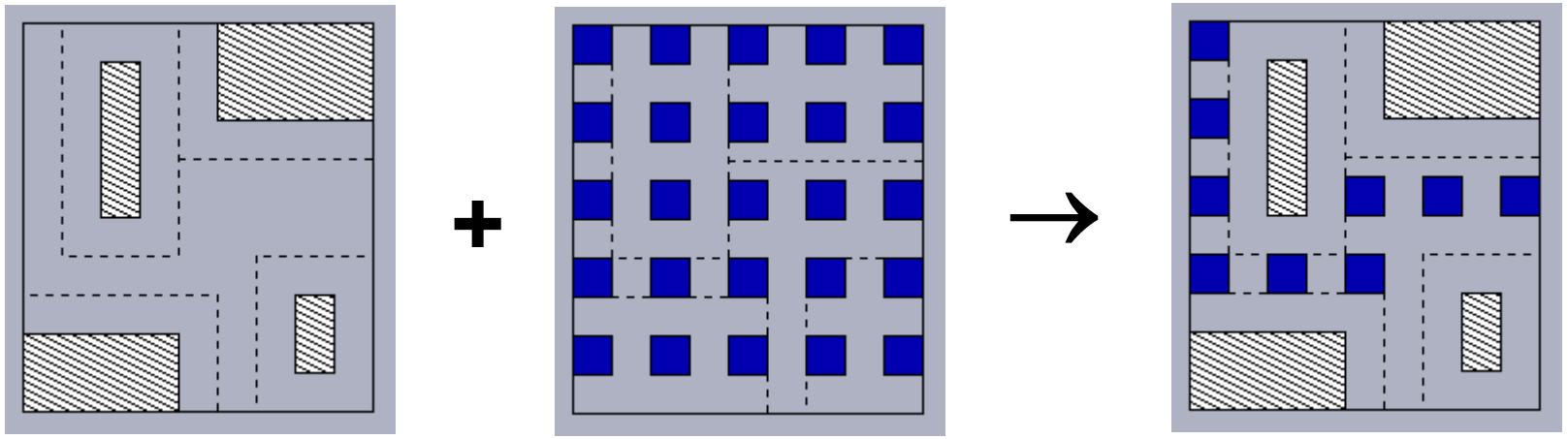
Metal Filling for Manufacturability

- Add dummy features to achieve global planarization for performance and manufacturability
- May change the parasitics
  - Coupling capacitance? Coupling inductance?



Rule-Based Tiling

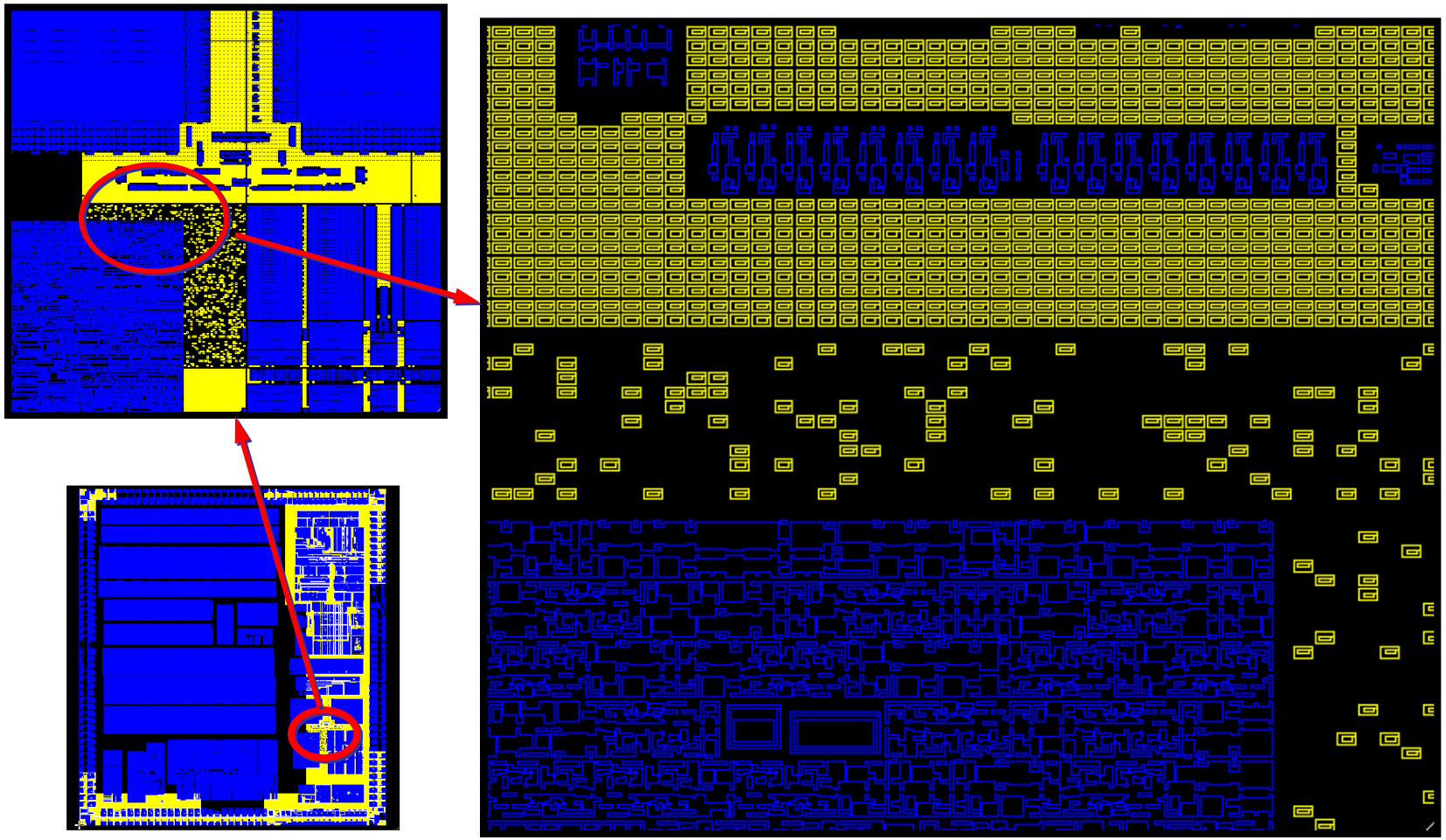
Source: RET trends by W. Grobman, ISPD 2001



- Done with Boolean operations
- Only density of the template is variable
- Not adequate for arbitrary design

Model-Based Tiling

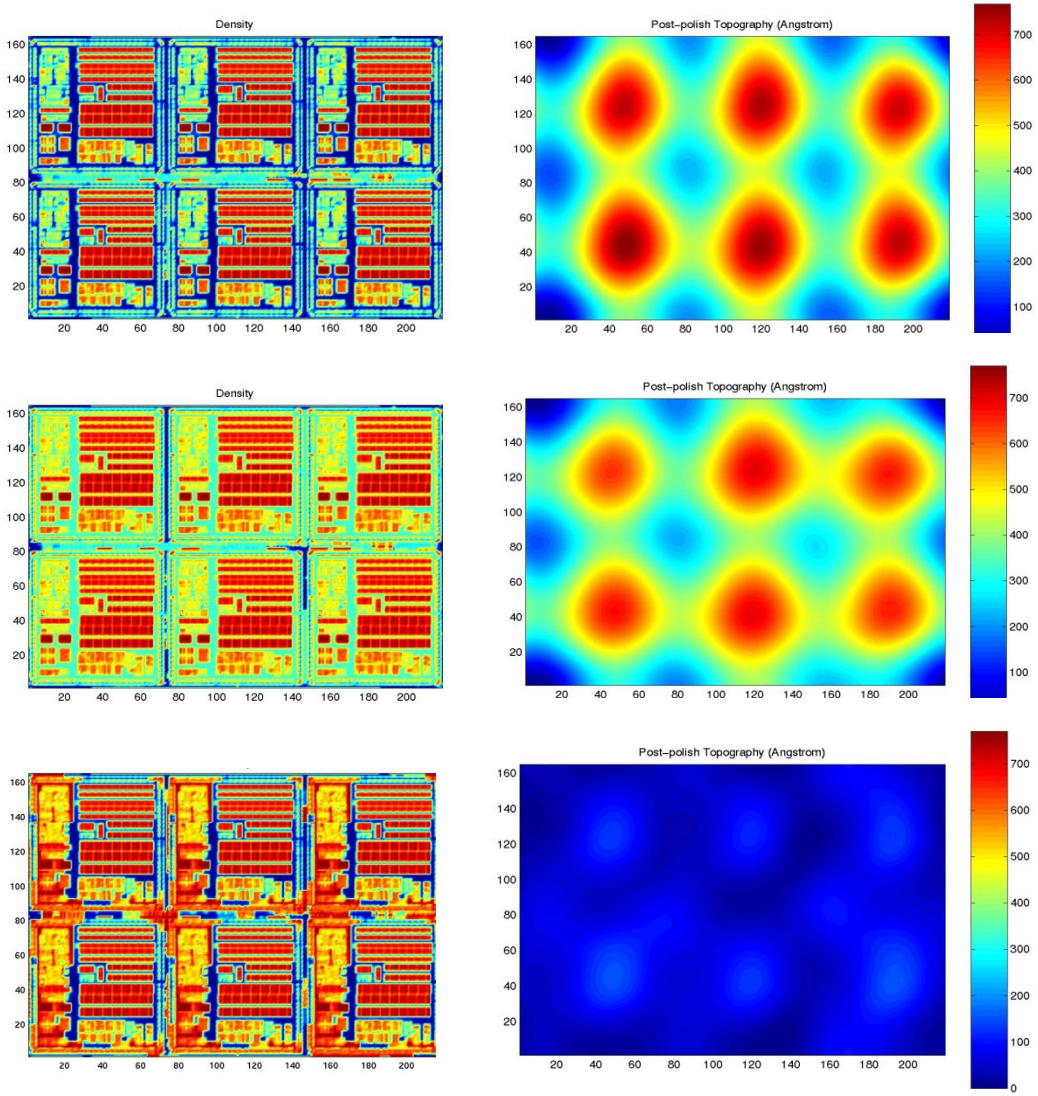
Source: RET trends by W. Grobman, ISPD 2001



- Different amount of tiles at different locations
- Uses Linear Programming and in-house software

Large Manufacturability Enhancement considering Model-Based Tiling

ISPD 2001 - Resources, Verification W. Grobman Mask Making



Un-tiled reticle (768A)
(un-manufacturable)

Conventional
Rule-Based Tiling (702A)
(9% uniformity improvement)

Model-Based Tiling (152A)
(80% uniformity improvement)

More experimental results (1/3)

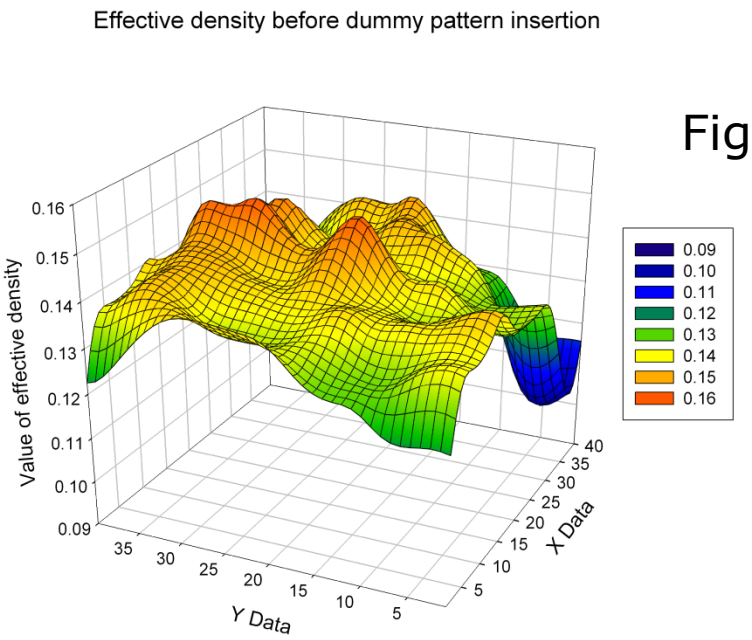


Figure.5 Effective density before dummy pattern insertion

Max(ρ)	Min(ρ)	Sta_Dev(ρ)	Peak to Peak(ρ)	Max local pattern density
0.151912	0.101997	0.00826395	0.049915	0.388521

More experimental results (2/3)

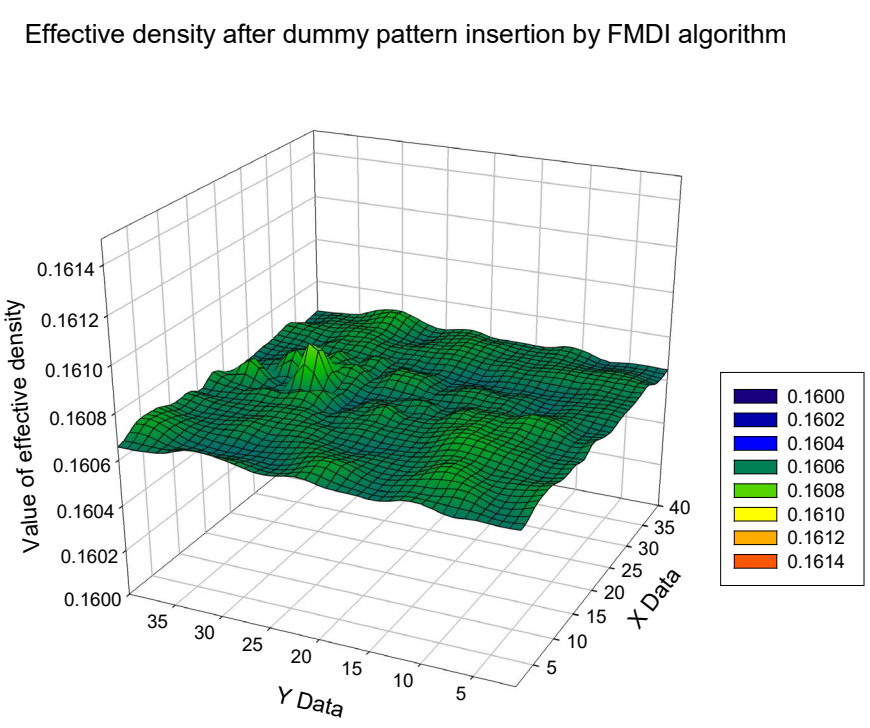


Figure.6 FMDI algorithm result

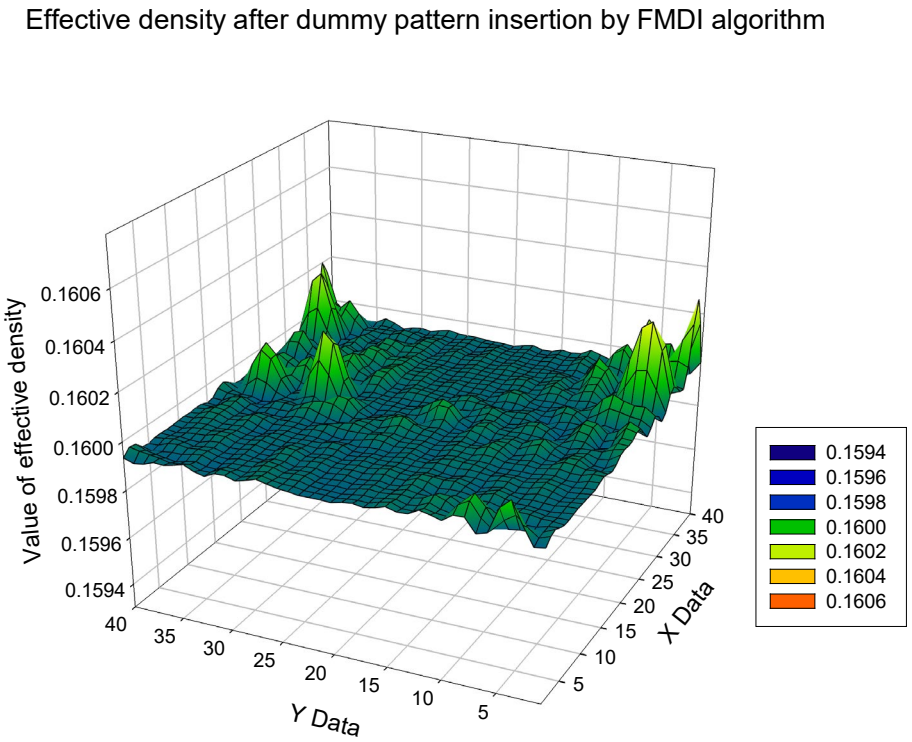
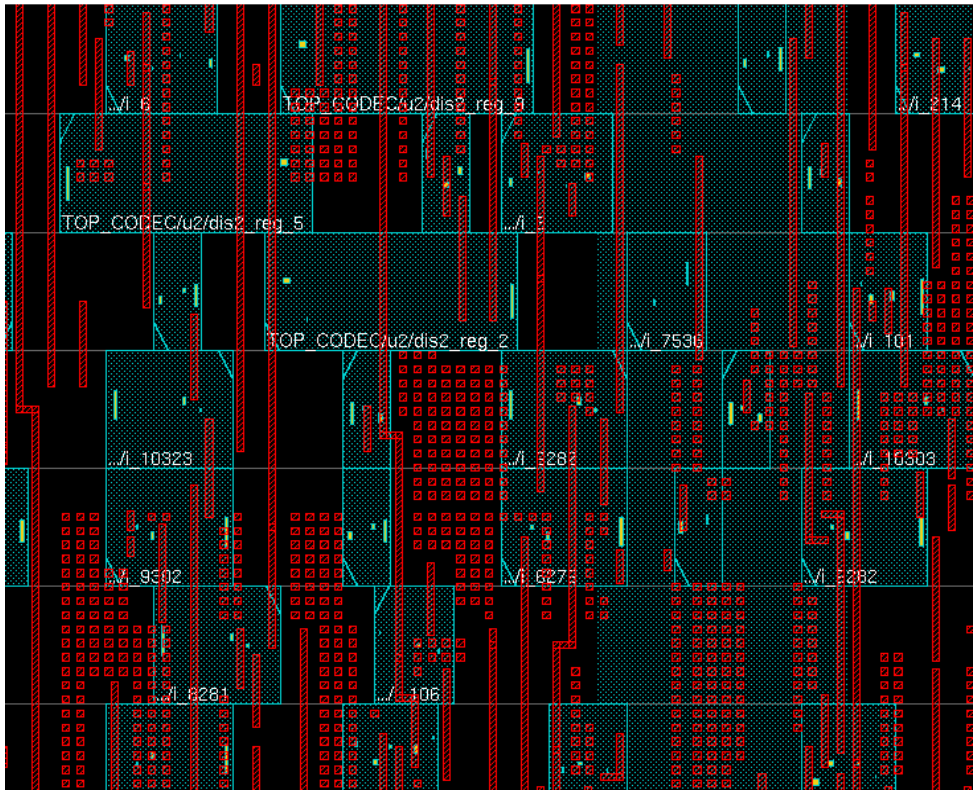


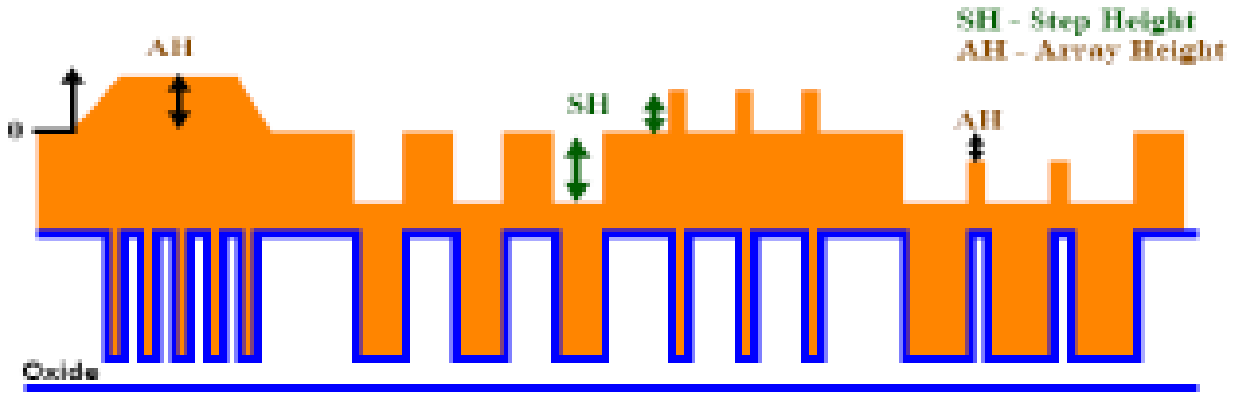
Figure.7 Iterated Smart algorithm result

More experimental results (3/3)

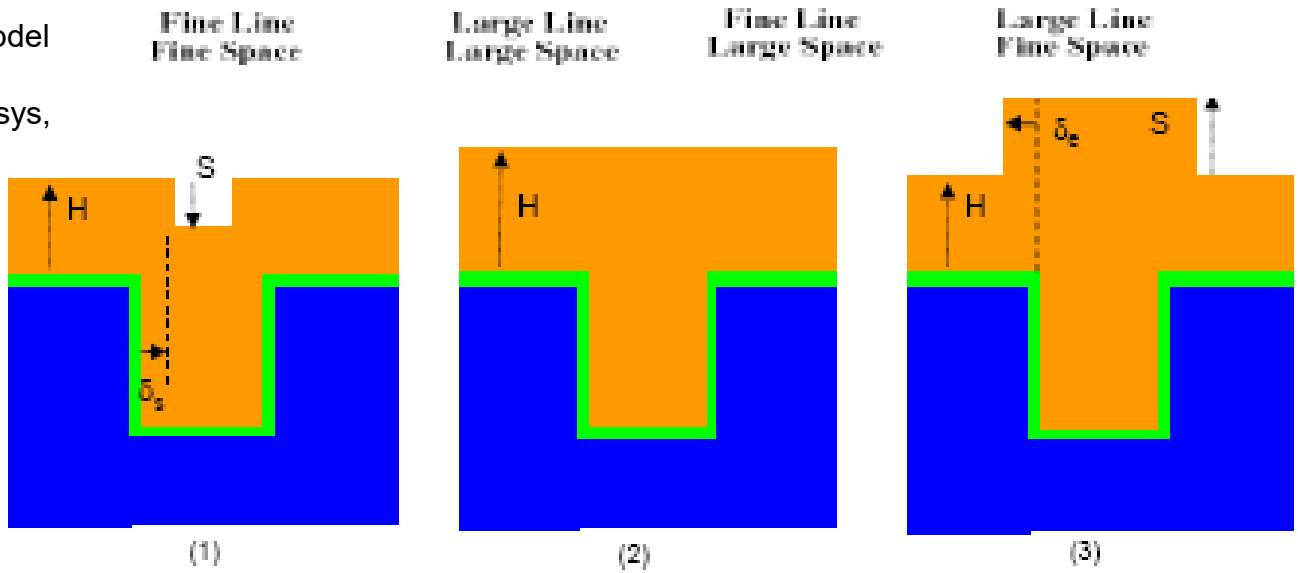


Dummy fill for Cu CMP process

- Cu CMP is different from ILD CMP
- Should consider both deposition (ECP) and polish (CMP) processes



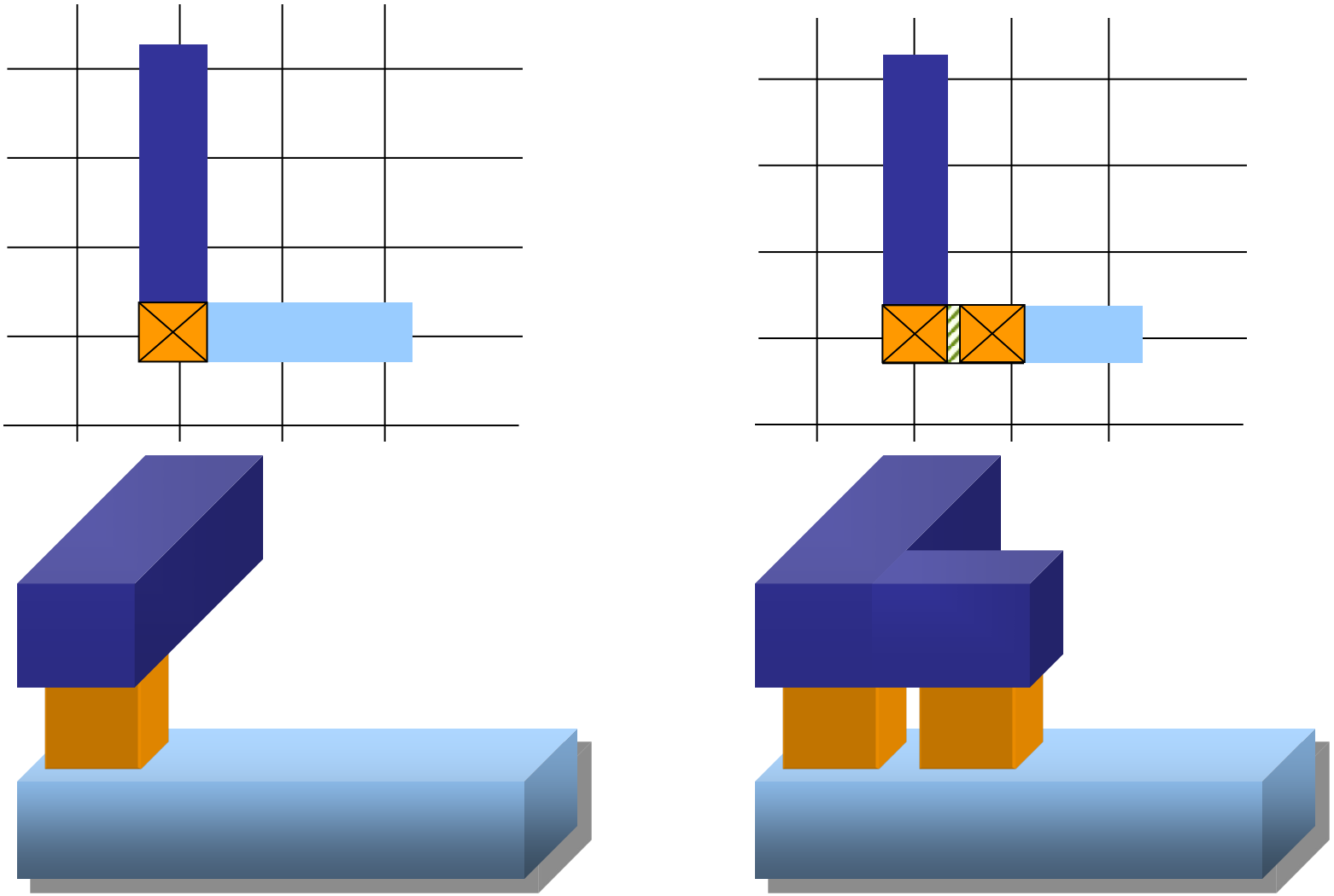
Source: CMP-aware Model based Dummy Fill by Charles Chiang, Synopsys, talk in NTHU, 2006



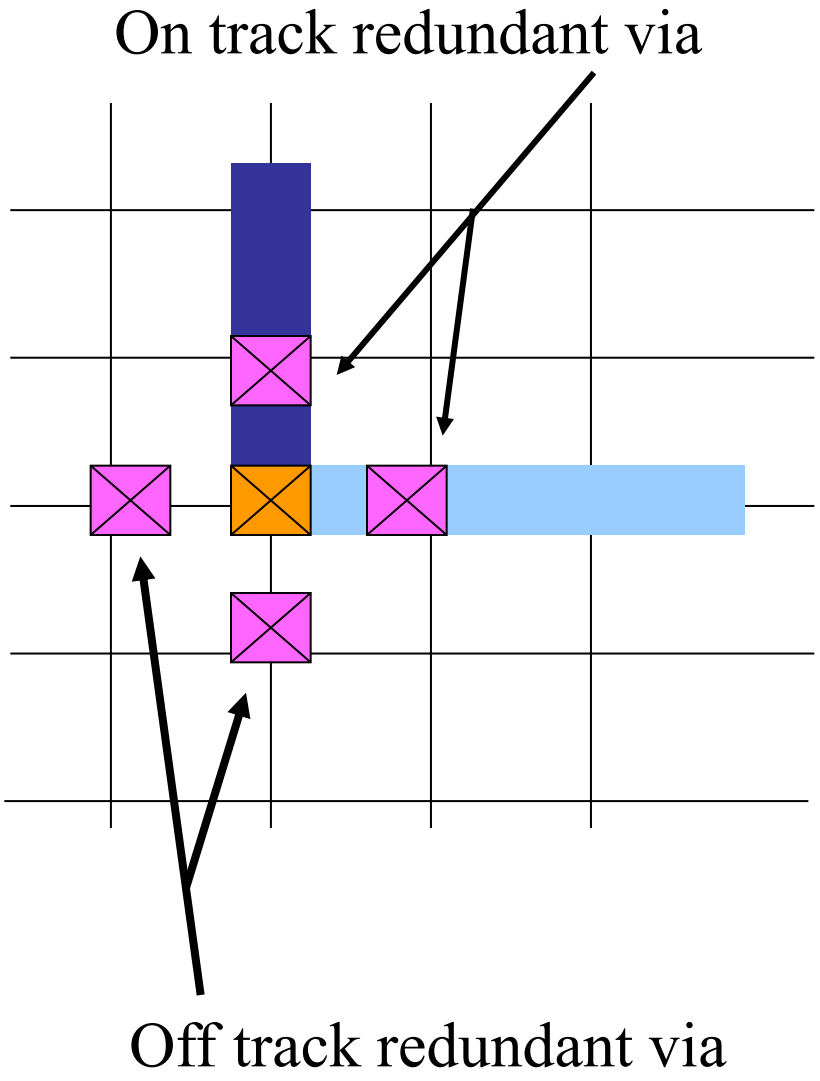
Redundant Via

- Yield loss by via failure becomes critical and requires a good control
- A good solution is to add a redundant via adjacent to a single normal via as a backup
- The extra via enables a single via failure to be tolerated
- Redundant via insertion can be considered in routing stage or post-detailed routing stage

Illustration of Redundant Via



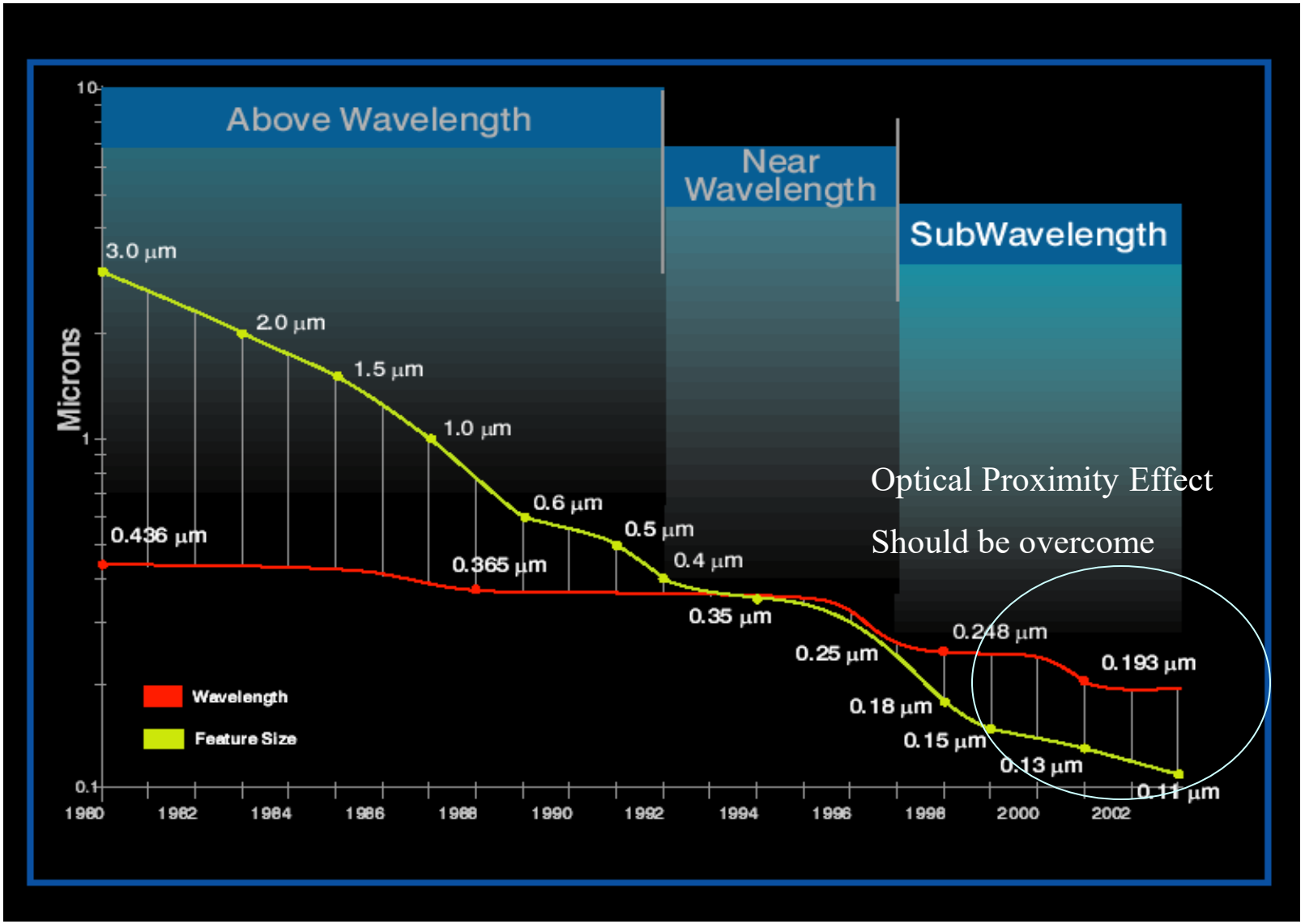
On/Off Track Redundant Via



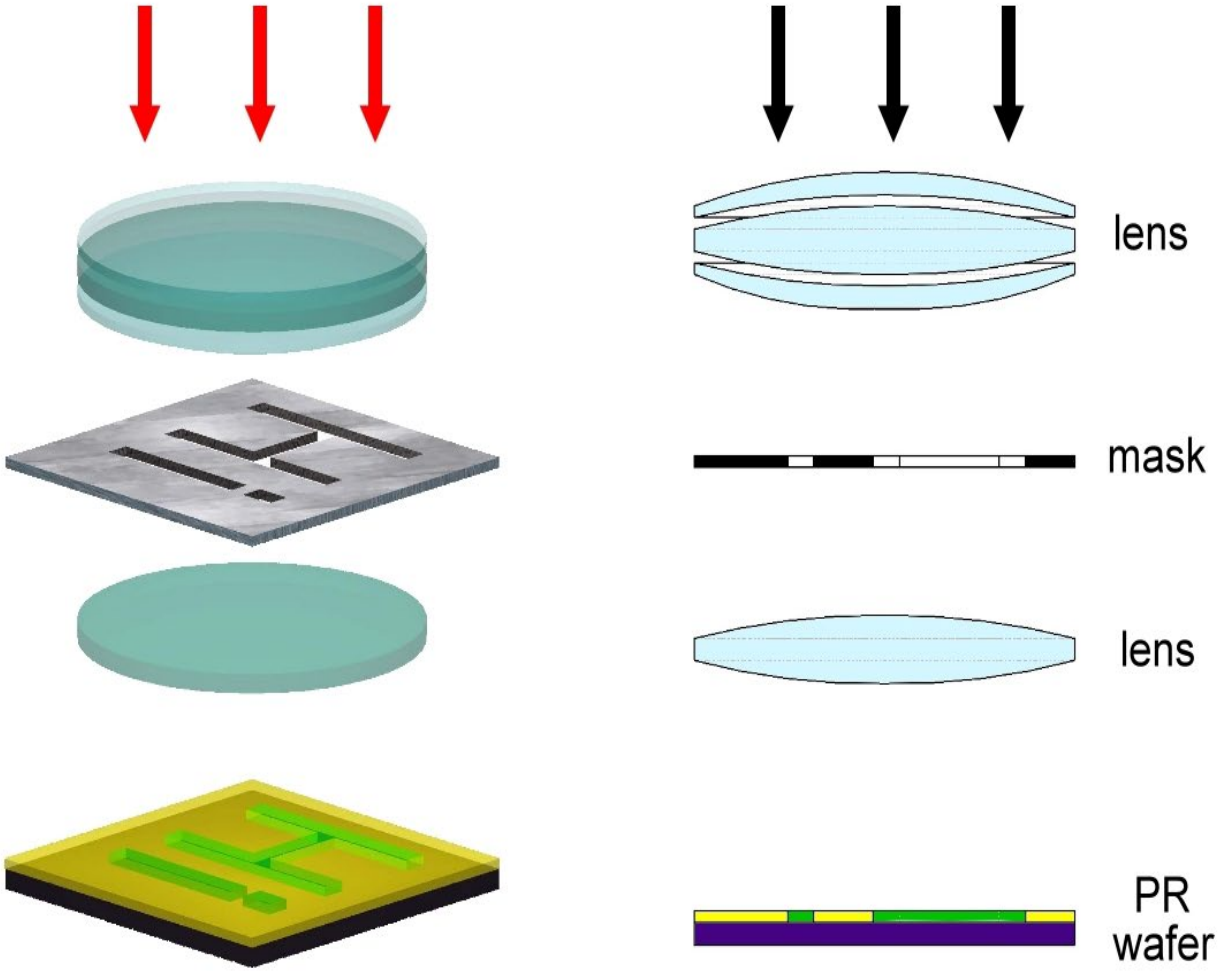
- On track Redundant Vias are more preferable since they
  - take less routing resource
  - have better electrical characteristics

Relationship Between Lithography and Feature Size

Source: RET trends by W. Grobman, ISPD 2001



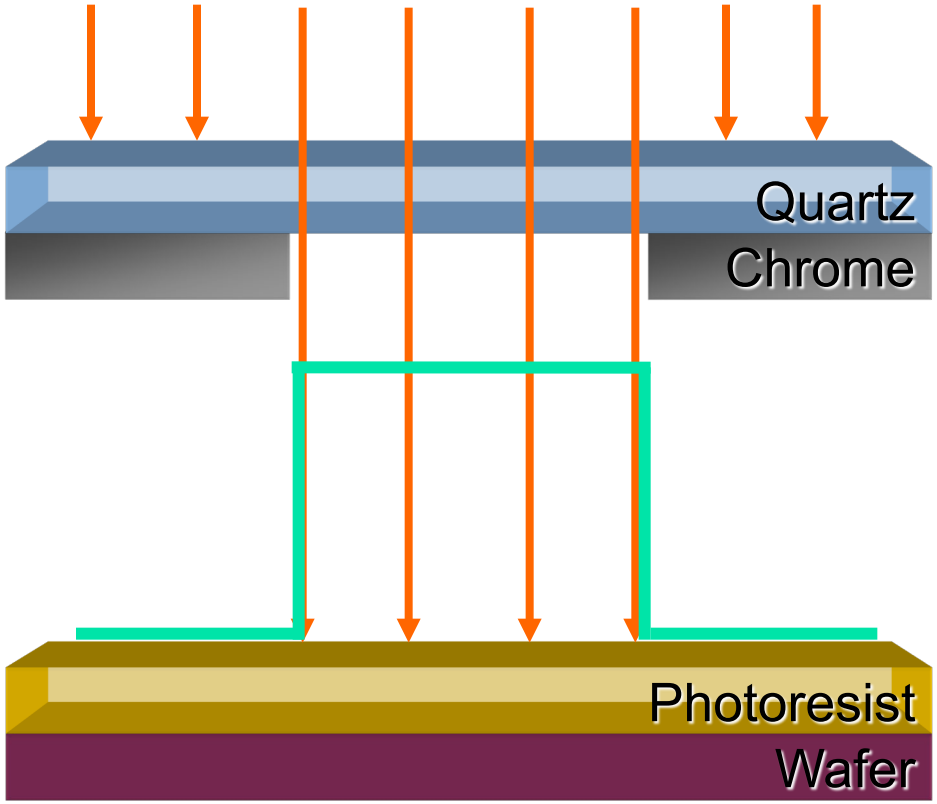
Lithography System



Optical Proximity Effect (1/3)

- Ideal lithography:

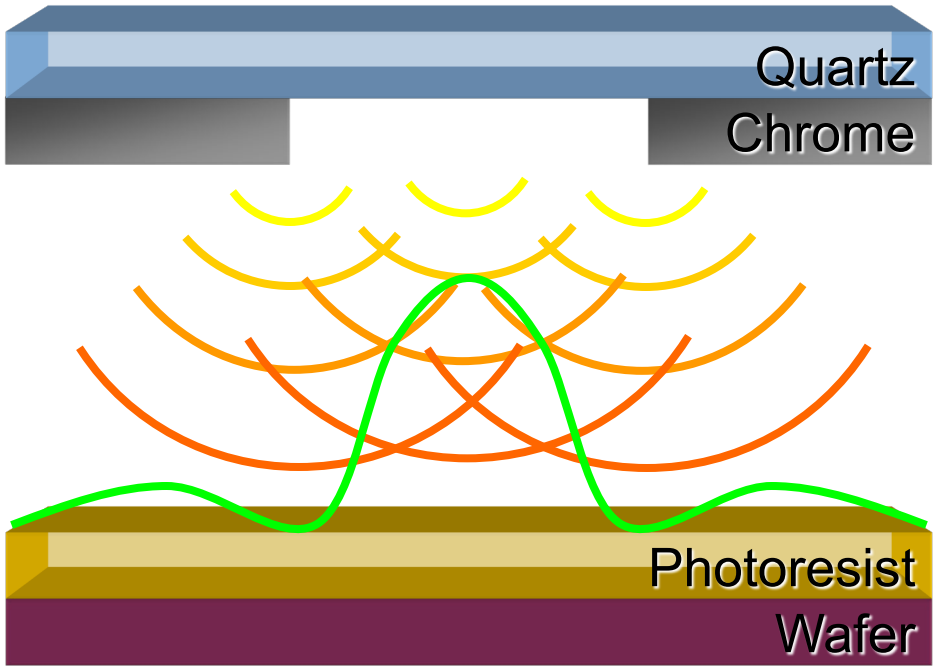
  light passes through features by a straight path



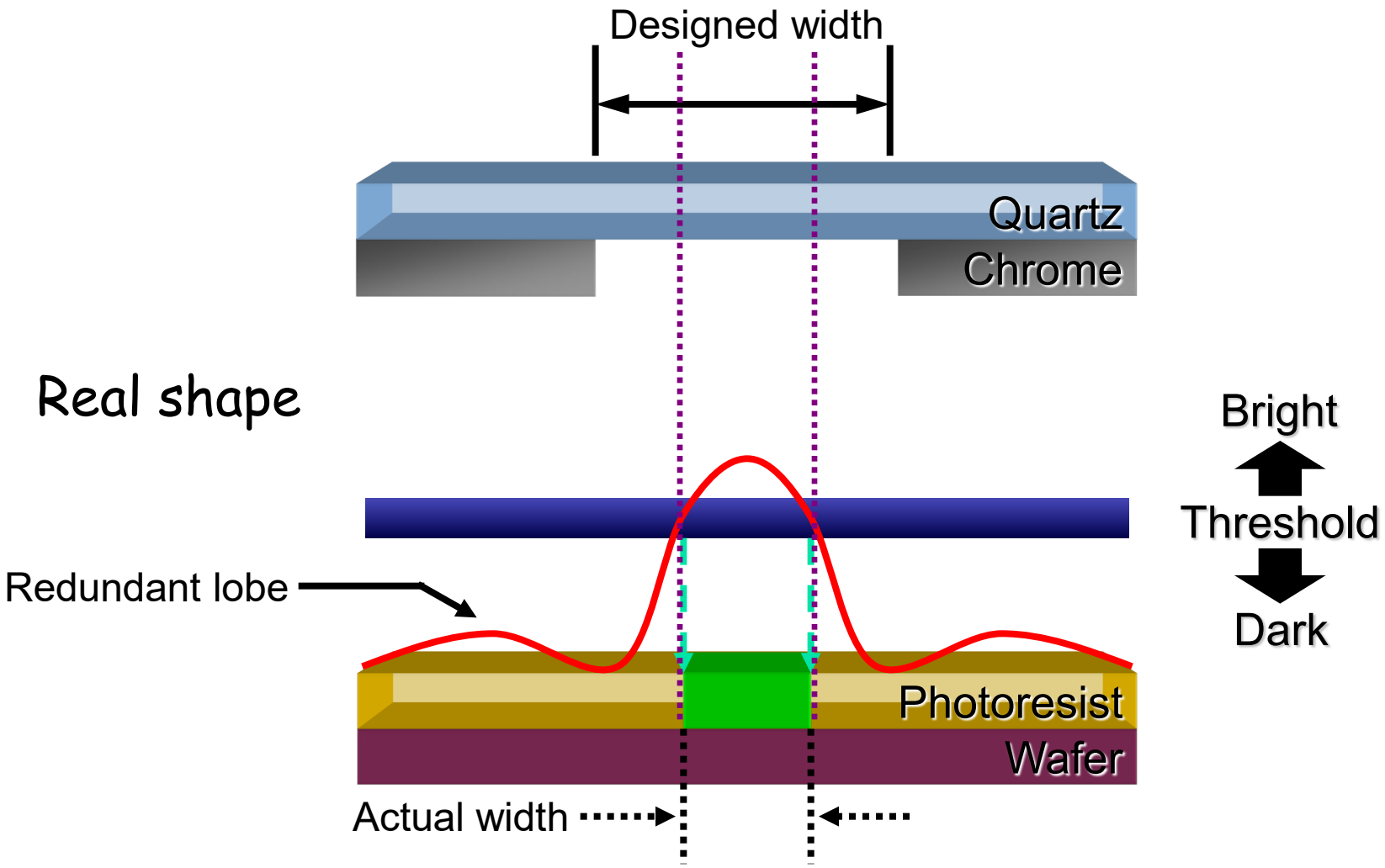
Optical Proximity Effect (2/3)

- Real lithography:

  light behaves like waves when feature size is close to wavelength

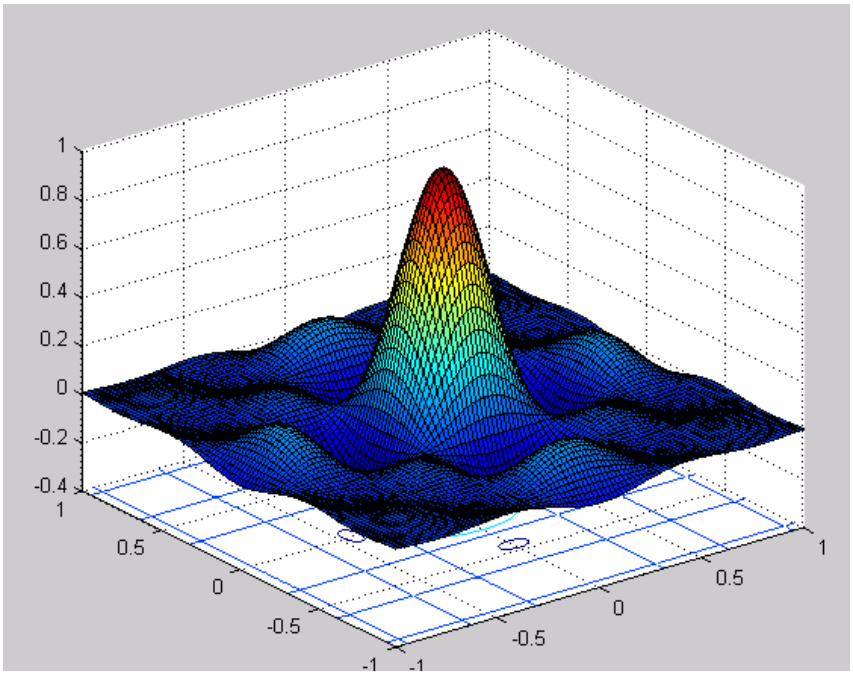


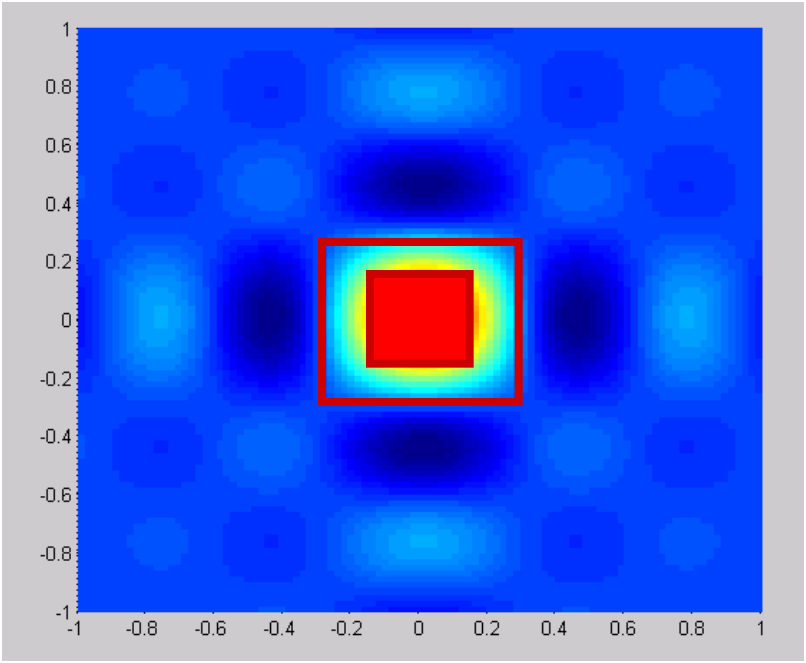
Optical Proximity Effect (3/3)



Simulation results of Diffraction (1/2)

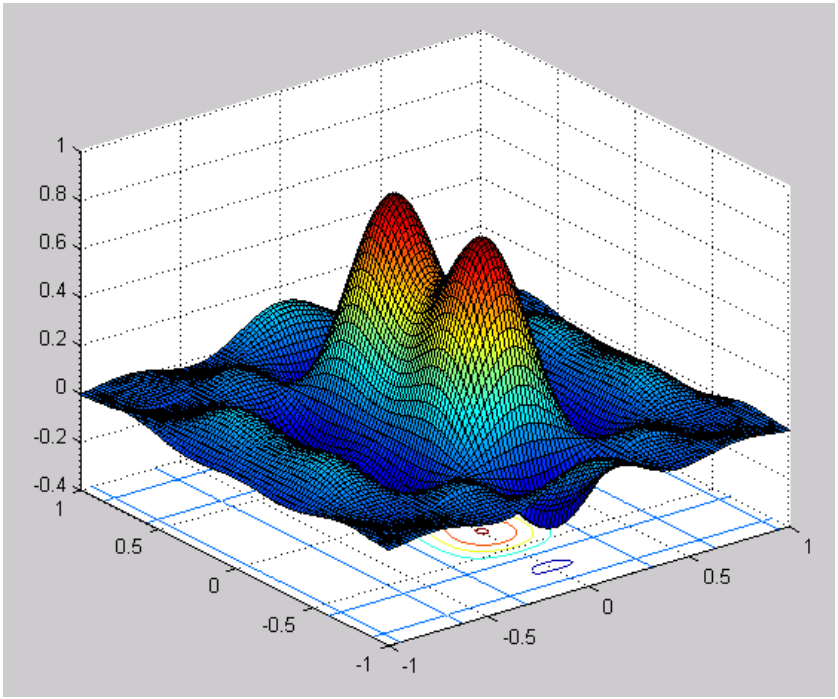
- Diffraction pattern of single aperture

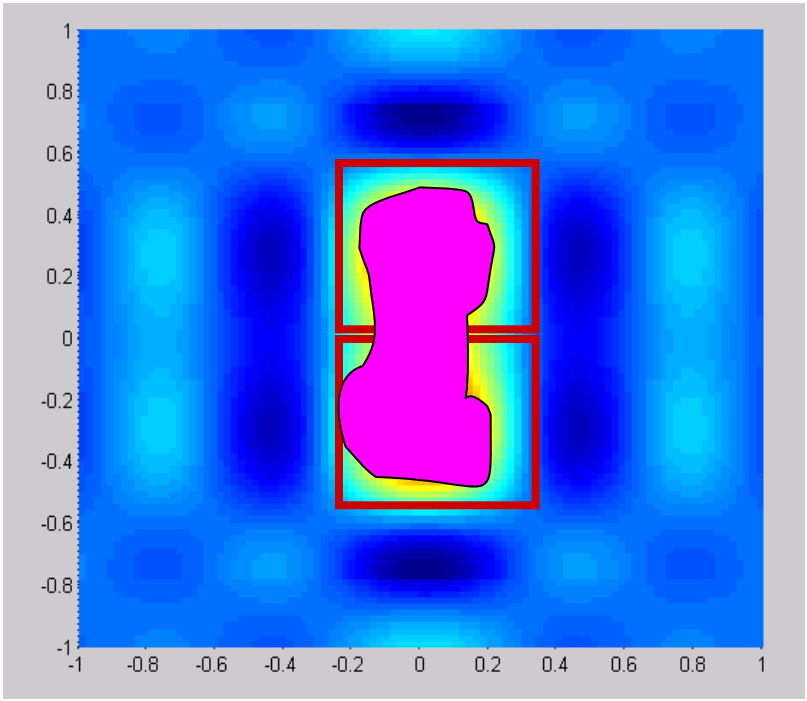




Simulation results of Diffraction (2/2)

- Bridged diffraction pattern

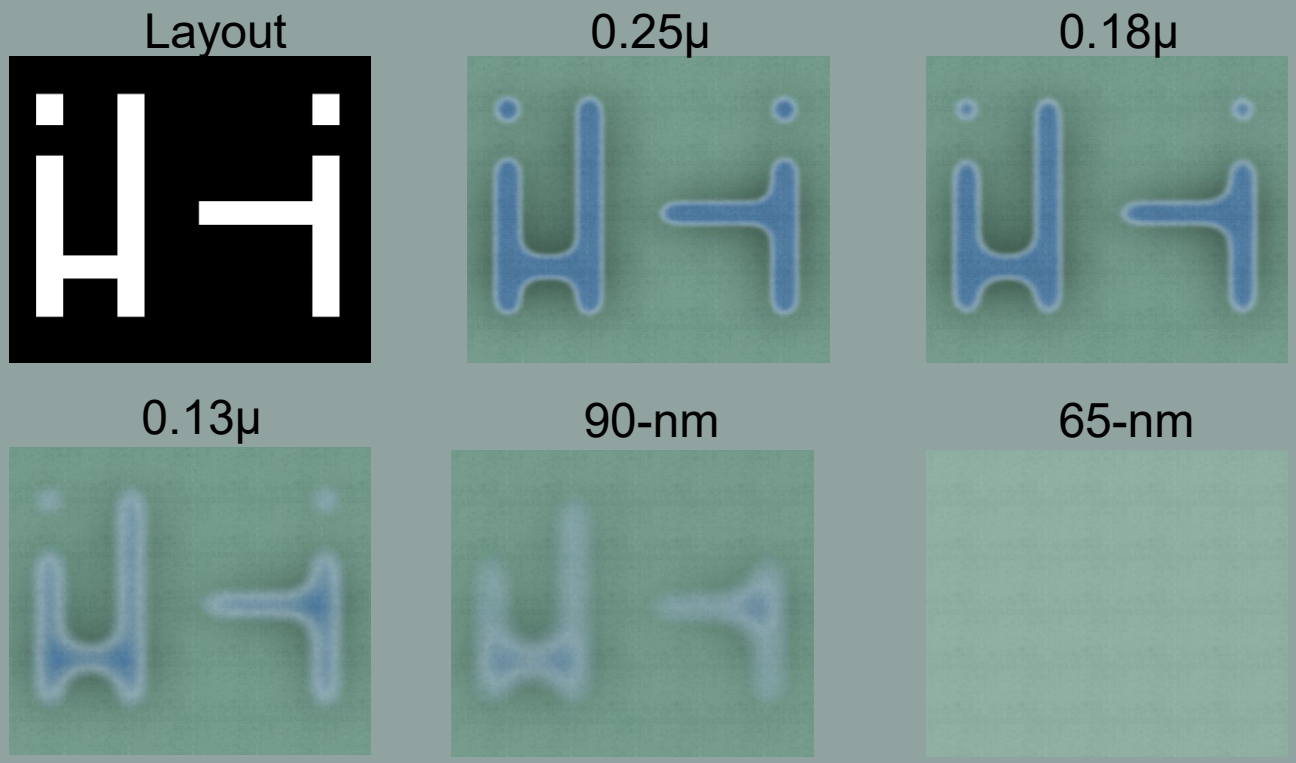




Side lobe effect

Why RET

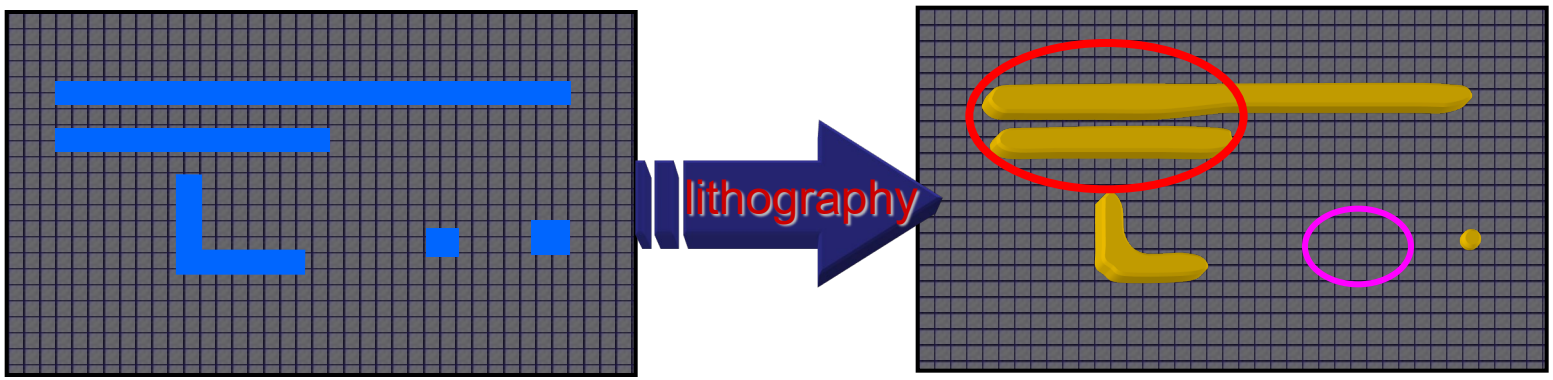
- With standard mask and illumination, features simply will not print



- Need Resolution Enhancement Techniques

Resolution enhancement techniques

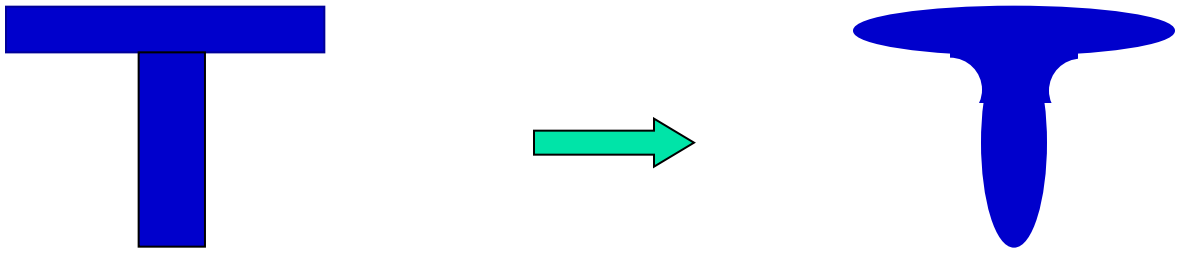
- The age of WYSIWYG came to a dramatic and painful end
- Specialized illumination patterns
- Optical proximity corrections (OPC)
- Phase-shift masks (PSM)
- Subresolution assist features (SRAF)



Original design pattern

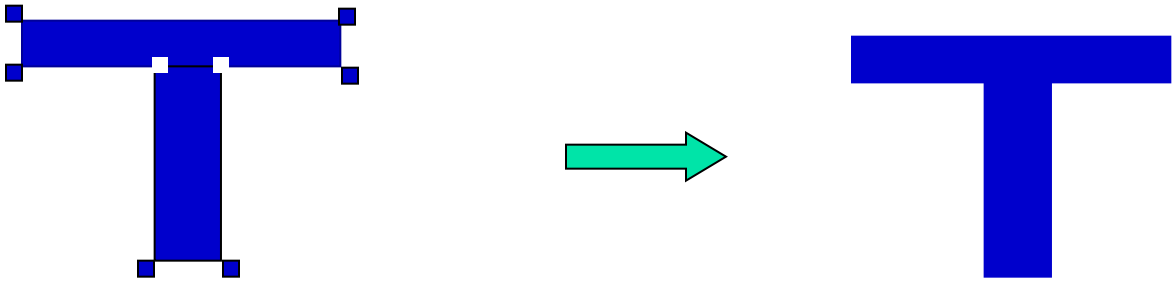
Optical Proximity Correction

- Nanometer process technologies require advanced optical processing, e.g., optical proximity correction (OPC).
- OPC increases the number of vertices in the layout and thus the size of the layout databases.
  - Make the mask cost much more expensive.



Drawn layout

Printed wafer

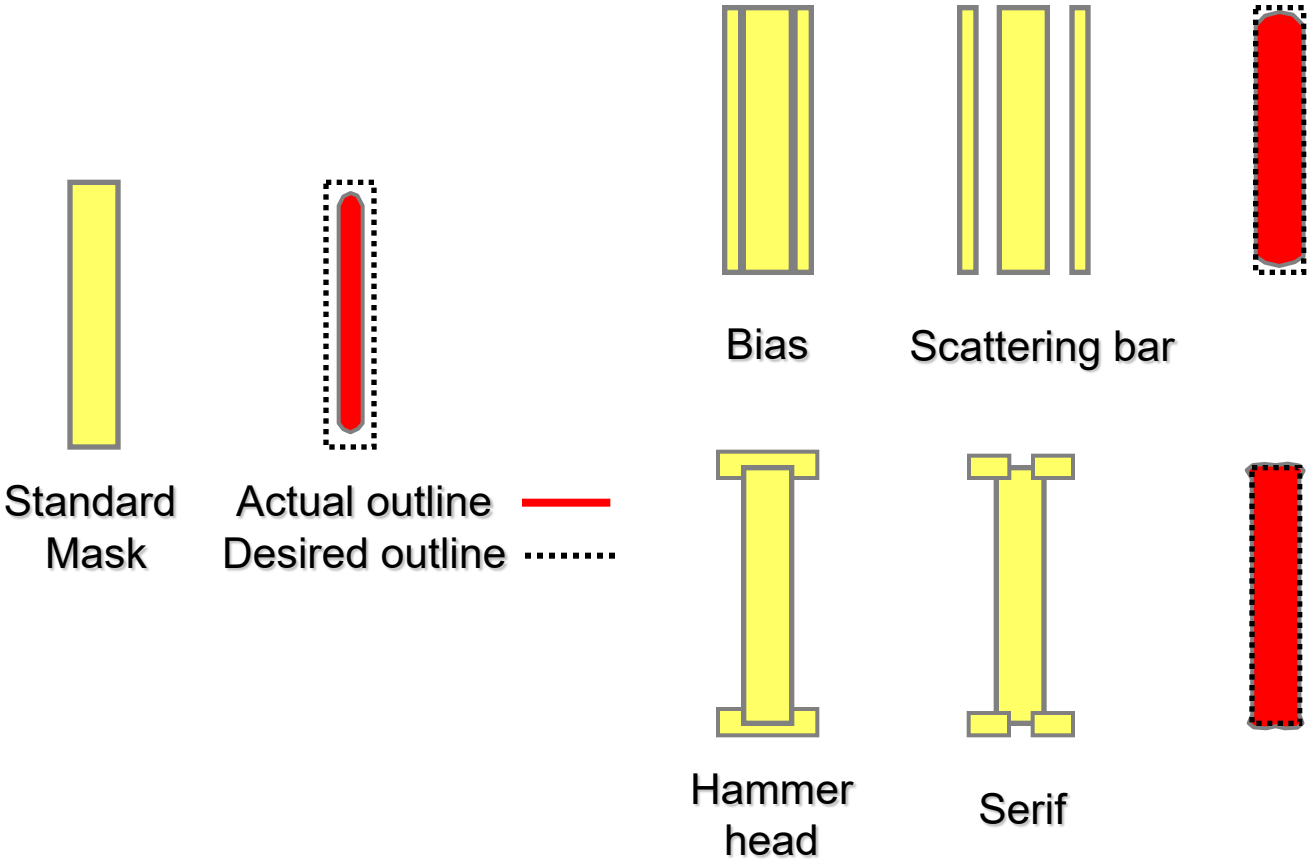


Proximity corrected layout

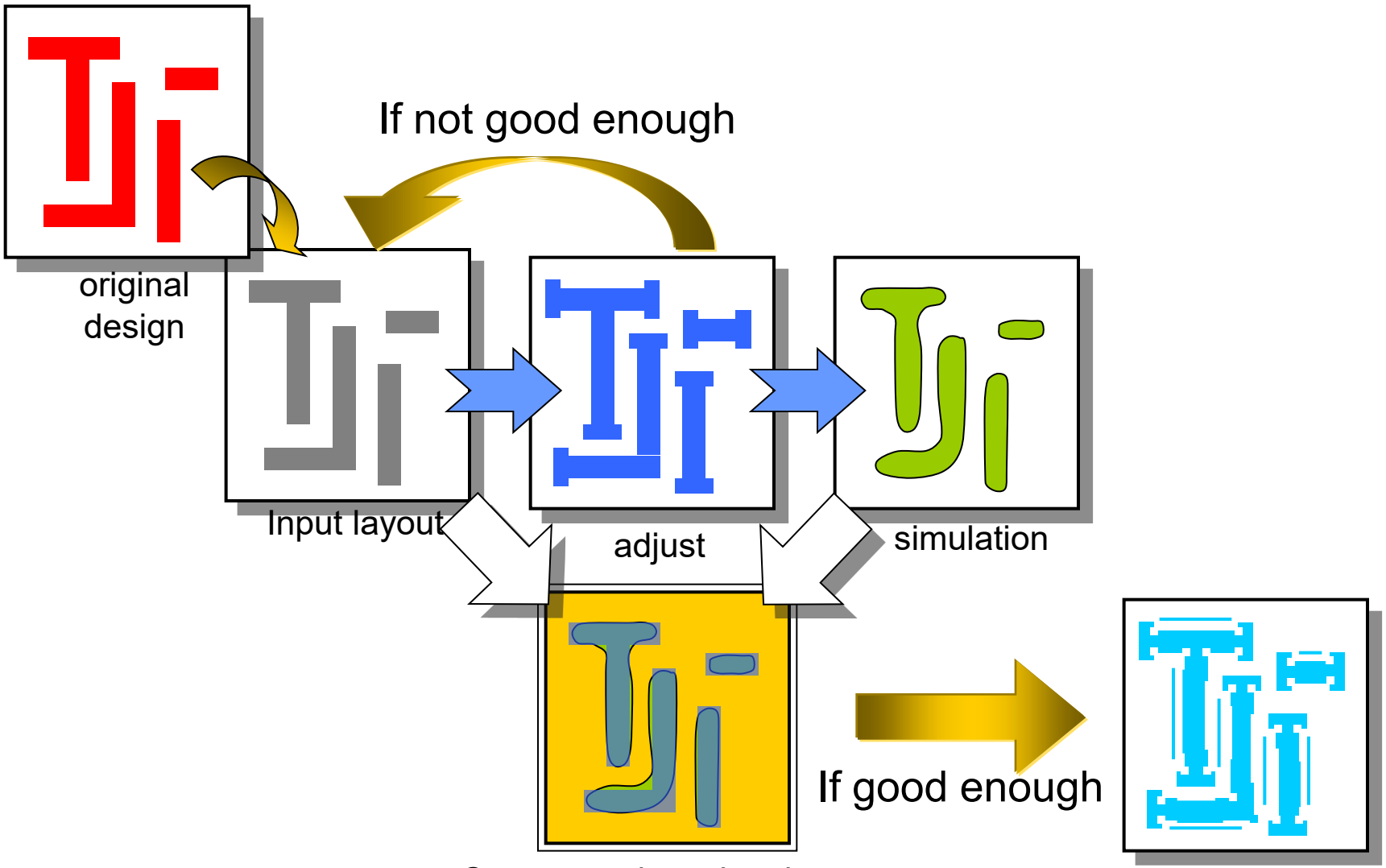
Printed wafer

Rule-based OPC

- Basic treatments of rule-based OPC



Model-based OPC

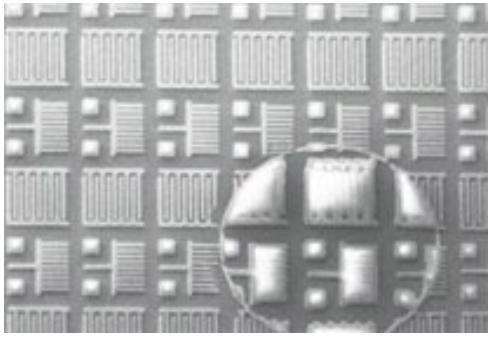


Future Lithographic Technologies

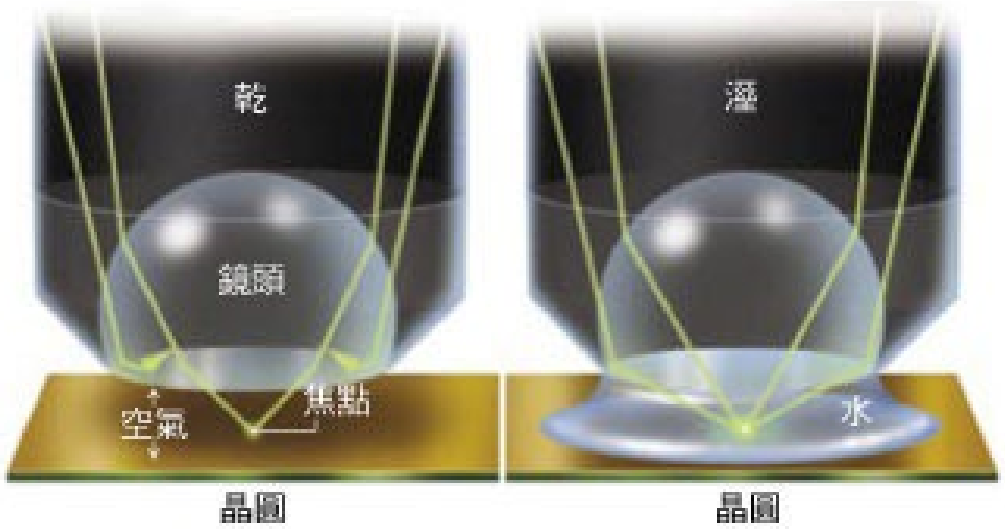
- Brief history of optical lithography
  - Increasing the numerical aperture (NA)
  - Imaging at smaller κ-factors with increased use of RETs and tighter control over process variations
  - Evolutionary reduction of the exposure wavelength
- Evolutionary path
  - Immersion lithography
  - Extreme UV lithography (EUVL, also called soft-X ray)
  - Particle beam lithography
  - Multiple electron beam direct write (EBDW)
  - Nanoimprint lithography (NIL)
  - Multiple exposure/patterning (to 22nm)
  - Direct Self Assembly (DSA)

Immersion Lithography (1/2)

- This will increase NA, then improve DoF



應用浸潤式微影技術，可以在晶圓上投射出更微小的電路，但是水裡的微氣泡會使電路的完整性受到破壞。



在鏡頭與晶圓之間的空隙中填滿水，可提高晶片製造微影技術工具的解析度。雷射光為了要投射出最小的電路線寬，會以極斜的角度穿透鏡頭，如果遇到空隙中的空 氣，很容易就會被反射回去（左圖）。反之，光線若以相同的角度遇到水就會發生偏折，因此可抵達焦點（右圖）。浸潤式微影技術也能改善焦深，也就是成像清晰的狀況下，從鏡頭到影像之間的距離。

Source: 科學人, 2005.8

Immersion Lithography (2/2)

浸潤式微影技術可把水導入微影機台,填滿鏡頭與半導體晶圓上光阻劑之間的空隙,改善晶片上電晶體之間的解析度以及焦深.當鏡頭下的平台挪動晶圓時,機器會把已成像部位上的水吸走.

Source: 科學人, 2005.8

Multiple Exposure/Multiple Patterning

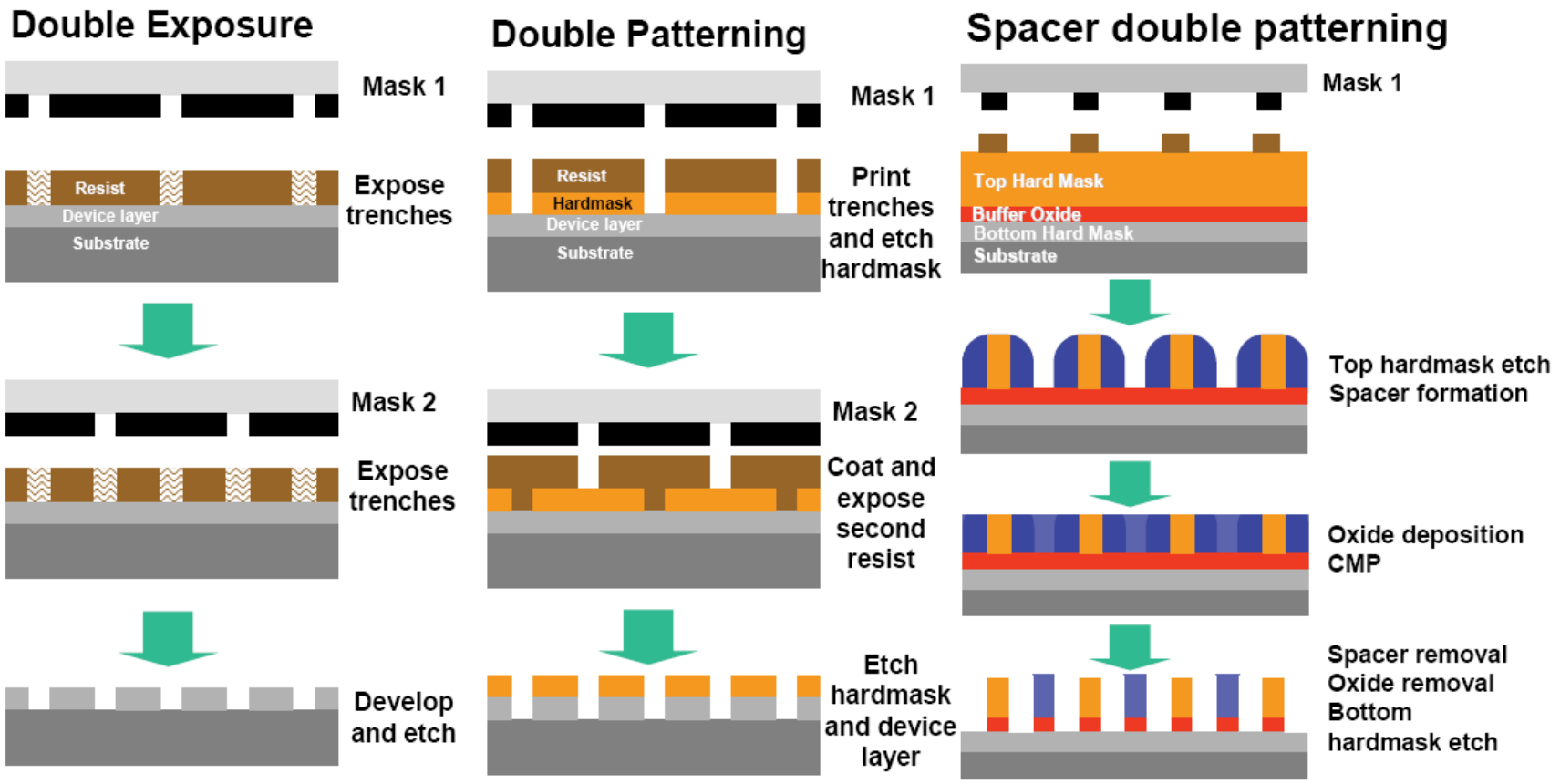


Figure LITH2 Schematic Process Flows for Double Exposure, Double Patterning, and Spacer Double Patterning

Ref. ITRS Roadmap, Lithography, 2007 edition http://www.itrs.net/

Multiple Exposure/Multiple Patterning

- Double exposure and double patterning are both capable of further extending 193nm in 32 and 22nm nodes
  - Successful mask decomposition is essential
- Double exposure
  - Double dipole lithography (DDL) enables semiconductor manufacturers to create smaller chip features by splitting dense circuit patterns into horizontal and vertical masks, then exposing them sequentially
  - Two consecutive mask exposures while the wafer stays on the chuck, allowing only one etch step
- Double patterning technology (DPT) / LELE
  - After the first mask exposure, the wafer is sent to the develop and etch processes, then come back for a second mask exposure (two independent patterning steps)
  - DPT can reduce the effective κ factor to be smaller than 0.25

Patterning solutions for advanced technology nodes

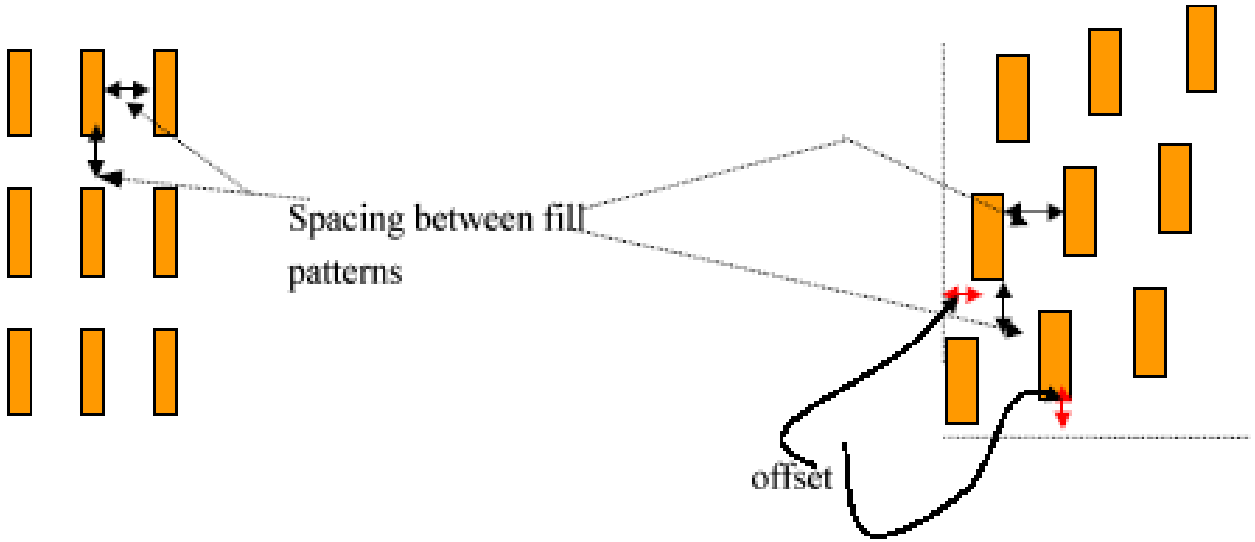
Technology node	Primary solution	First alternative
65-nm	193-nm dry, single exposure	193-nm, double exposure
45-nm (HKMG)	193-nm immersion (193i)	193-nm dry, double patterning
32-nm 28-nm (TSMC)	193i, double patterning	193i, single exposure, high-index fluids
22-nm	193i, double patterning	
16-nm (HKMG FinFET) 14nm	193i, double/triple patterning	
10-nm	193i, multiple patterning (MP)	
7-nm	193i, MP/EUV	EUV
5-nm/3-nm	EUV, double patterning	

Summary

- We have discussed some emerging second order effects in nanometer era
  - Interconnect delay domination
  - Signal integrity: crosstalk and voltage drop
  - DFM: antenna effect, metal filling, RET, and more

2001 MOE IC/CAD Contest Problem 9

- Area filling
  - To control the manufacturing variation due to CMP



2005 MOE IC/CAD Contest Problem 6

- ESD (electrostatic discharge) current path analysis
  - To prevent ESD failure, it is important to verify at design stage whether ESD signal path exists for any two pads.
  - One way to build the protection is to protect all pairs of pins.
  - It may be unnecessary for some protections which protect the pairs of pins that have no current paths.
  - An alternative is to identify those pairs of pins that have current paths between and then to protect those pairs only.