Section V: Analysis and Verification

Overview

- Signoff timing verification
- Power analysis
- Manufacturability (antennas, OPC, phase-shifting, area fill)
- Inductance modeling

Sign-Off Timing Analysis

Sign-Off Timing Flow

- Extraction
  - Extracted wire RC from the layout must be used in delay calculations
  - Static timing analysis calculates gate and wire delays based on the extracted RC and precharacterized gate attributes
  - The wire and gate delays are summed, then checked against constraints
- Process and physical effects (such as cross talk) must be account for

Setup / Hold

- Setup
  - "Does my signal get there before the clock?"
    - Want to measure under the slowest conditions
    - Bad to miss: lower clock speed / lower yield
- Hold
  - "Is my signal maintained after the clock?"
    - Want to measure under the fastest conditions
    - REALLY bad to miss: chip fails at any frequency

Clocks

- Skew
  - Difference between the clock arrival at different flops
    - Measurable after layout
    - Must be treated as "uncertainty" during synthesis
- Uncertainty
  - PVT variation, characterization inaccuracy, tool inaccuracy

Clocks (#2)

- Jitter
  - Variance in the clock waveform from cycle to cycle
    - Effects only setup. Often lumped with uncertainty
- Latency
  - delay from the clock source to the flop
    - Often set to zero during synthesis
    - Unexpected constraint violations pop up later (clock gating)

Constraints

- Constraints tell the STA tool what to check
  - Most paths are defined by the clocks
  - Other “exceptions” must be added
    - false paths
    - asynchronous clock domains.
- Most chips have separate test modes that run at low frequency
  - Test paths are “false” in functional mode
  - A separate STA run checks the test timing

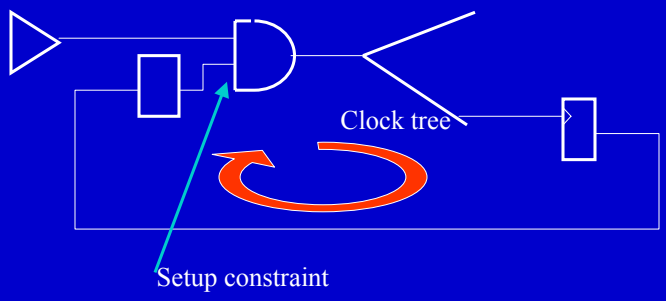
I/O Constraints

- Typically, half of a chip's constraints are for I/O
  - Input signals may be synchronized to an external clock (source synchronous)
  - An output clock may be come along with output data
    - Must be properly aligned
  - I/O data may come on both edges of a clock (ddr)
  - Busses on I/O are often constrained to be balanced

Special constraints

- Clock gating logic also has setup and hold constraints.
  - Can't ignore clock latency
  - Need a latch/flop to avoid glitches on the gating signal
- PLL feedback paths must be carefully balanced

Clock Gating Constraint



Clock gating constraint:
one cycle to get around – including clock tree latency

Timing models

- Most of your timing models come from precharacterized “dot lib” files
- Hard macros may be characterized as
  - lib files
  - Fake netlists with overlaid SDF
  - Synopsys “stamp” models
- You need separate models for slow and fast process conditions.

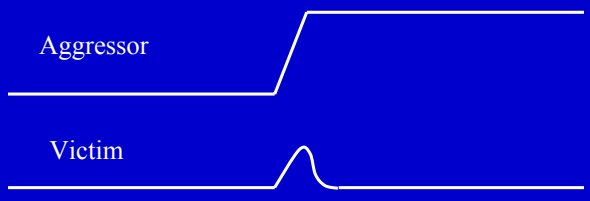
Cross Talk

- Cross talk is caused by capacitive coupling between wires
  - An “aggressor” wire switches
  - A “victim” wire is charged or discharged by the coupling capacitance

Cross Talk (glitches)

- An otherwise quiet victim may look like it has temporarily switched
  - This is bad if:
    - The victim is a clock or asynchronous reset
    - The victim is a signal whose value is being latched at that moment
  - Fixes:
    - Shield/space the wire, use strong drivers, repeater long wires often

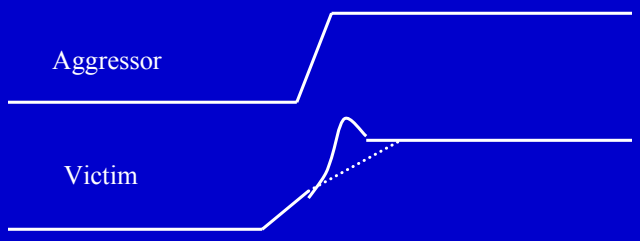
Glitch



Cross Talk (timing pull in)

- A switching victim is aided (sped up) by coupled charge
  - This is bad if:
    - Your path now violates hold time
  - Fixes:
    - Add delay to your path

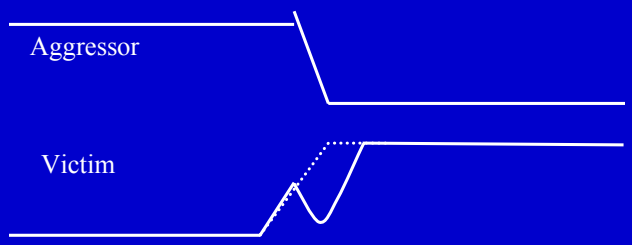
Timing Pull-in



Cross Talk (timing push out)

- A switching victim is hindered (slowed down) by coupled charge
  - This is bad if:
    - Your path now violates setup time
  - Fixes:
    - Space the wires, use strong drivers, add repeaters (to a point)

Timing Push-out



Analyzing cross talk

- Extract all significant coupling capacitance
  - The traditional technique of “coupling to ground” is no longer sufficient
- You can scale the coupling caps to approximate pull in and push out. (a.k.a: “Miller scaling”)
- Or there are tools which will compute accurate pull in and push out from your extracted data. This timing offset can be fed back to STA
- A separate tool may be needed to analyze glitches

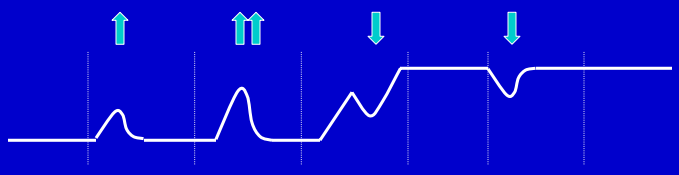
Timing Windows

- Multiple aggressors may affect one victim
  - If the aggressors switch at the same time, their effect is additive
  - An STA tool may produce “timing windows” to tell how the aggressors line up
  - If timing windows are not used, all aggressors are assumed to come at the same time (conservative)

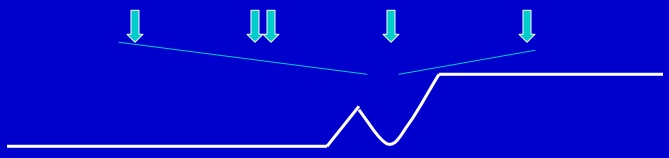
Timing Windows (#2)

- Timing windows are needed for high performance
  - “Stacking” aggressors for push out is too conservative
  - For pull-in and glitches, timing windows are not necessary The conservatism is generally desired anyway

Effect of Timing Windows



Multiple aggressors with timing windows

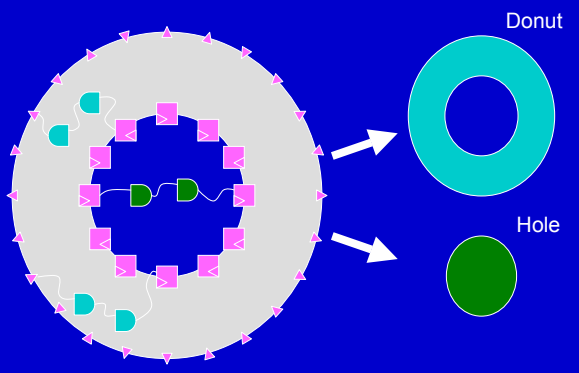


Multiple aggressors without timing windows

Hierarchical Timing: Donuts

- "Donut flow" used for top level timing (and convergence ECO)
  - split block level netlist into "core" and "donut"
  - split block level parasitics " " " "
  - top level netlist + abutment dspf (0 ohm jumpers) connects donut R and C

Donut Timing (#2)



Donut Timing (#3)

- Top level timing run now ~4X – 10Xsmaller
  - top level timing + ECOs for 1.2M objects: 1/2hr
- No block level constraints required

Power Analysis

Power Analysis

- Power Consumption
  - How long with the batteries last?
  - How hot will the chip get?
- Voltage Drop (IR)
  - Impact on timing, considerable at already low supply voltages (e.g. 1.5v)
- Electro-Migration (EM)
  - Reliability

Mars/AstroRail Example

- Prepare libraries
- Prepare files
- Run tools
- Look at log files, reports
- Visually inspect the problem areas

Capacitance Models

- .PARA view of blocks, produced with Star-RCXT
- Or, Apollo LPE (internal extraction engine) if no PARA view (less accurate)
- LPE needs capacitance models in the technology file, (librarian does this)
- Without PARA, or LPE, job cannot run

Other Avant! Views

- Power, PWR views (librarian)
- TIM (static timing view)
- CONN view, used during rail analysis and needed for any non-stdcell that contains part of the power distribution network (padcells, covercells)
- FRAM views (abstracts) with power/ground pins

Net Switching Activity

- "number of switches per unit time"
- Example: if unit is ns, then
  - 4 ns clock = > 2 switches/ 4ns = .5 NS factor
  - Signal switching once every *two* clocks would be 1 / 2*4 = .125 NS factor
- In other words:
  - The clock signal switches 2 times per 4 ns
  - The non-clock " " " " 1 time per 8 ns

Net Switching Activity (#2)

- Hence:
  - Clock net switching activity = .5 switches/ns
  - Data net switching activity = .125 switches/ns
- Most of the time multiple clocks around, very important to cover all the clocks as the switching factor is much higher.

The Tap File

- Specifies the boundary conditions, I.e. the location and layer of the ideal voltage sources
- IR Drop and EM simulations use the information and it must be correct, for example for BGA style pads, the centers of the bumps should be used

The Tap File (#2)

- The layer of metal and location for the bond-wire attachment point should be used for linear padrings
- For a PnR block, the tap file should contain the locations of the power mesh tips at the edge of the block

Power Consumption Summary

- Switching power
- Short Circuit power
- Internal power
- Leakage power

Switching Power

- Power consumed to charge or discharge the capacitive net, which depends on:
  - Amount of C on the net (from PARA view fo the block, or from LPE based on TLU models)
  - Voltage Supply Value, V (defined by user)
  - Switching Activity, f (defined by user)
  - = ½ * C * V^2 * f

Short Circuit Power

- Power consumed by current that flows between power net and ground with both pull-up / pull-down logic are temporarily on during switching. Depends on:
  - Transition time of the input signal (calculated by running STA)
    - Long nets driven by small drivers will cause long transition times

Internal Power / Leakage Power

- Internal Power consumed when inputs of a device change but the output does not. Depends on the internal structure of the device
- Leakage power consumed by the current that leaks through the semiconductor junctions. Depends on the physical characteristics of the junctions.

Voltage Drop

- Distribution network not ideal, Resistances R, all over the place
- Current * R = > Voltage drop
- If drop too large, noise margins suffer, sensitivity to cross-talk increases
- Reduces drive strengths, effecting the timing

Voltage Drop (#2)

- For large padring style die, center of chip is the problematic issues (assuming pads are symmetrical, etc)
- For BGA, life is easy, Max drop a function of distance between pads over the core
- Ground rails will also have voltage drop: Ground Bounce. Should be similar if distribution networks look the same, but BE CAREFUL and check it
- Typical targets 5% each for VDD Drop / VSS Bounce.

Electro-Migration

- Movement of current in a wire is carried by electrons..too much and metal is damaged. Damage rate proportional to wire cross section and temperature.
- Two kinds of current flow:
  - Uni-directional (power supplies)
  - Bi-directional (signals)
- Bi-directional flow has healing effect since atoms are push back and forth, therefore problem not as bad.

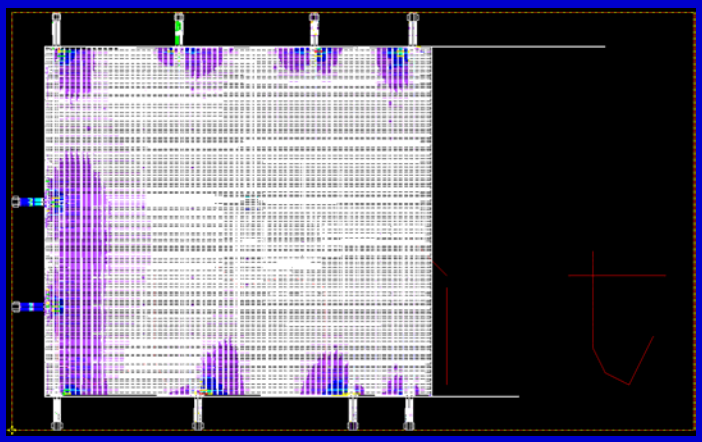
Electro-Migration (#2)

- Trends making this problem worse, (.13 copper should help)
- Dielectric materials between metal layers getting worse for thermal conductivity. Metal self-heading (joule heating) has a harder time being dispersed
- Signal EM becoming more of an issue (I.e. BUFX20 in .13 require two vias?)

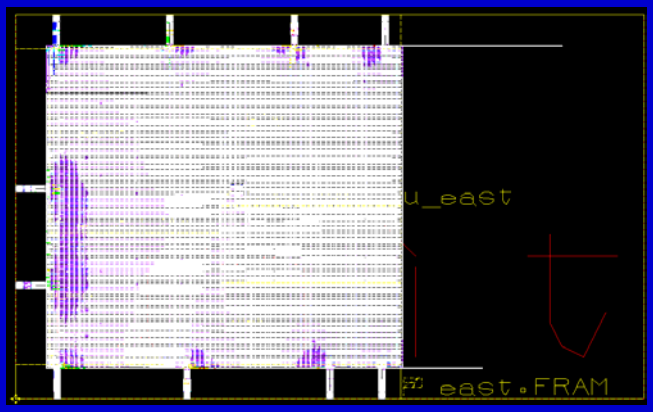
Power: Correct By Construction

- Best EM and IR drop solution is robust up-front distribution:
  - Each power pad has a fixed current capacity
  - Number of power pads initially determined, but as soon as netlist matures tools should be run to calculate preliminary power consumption
  - Padring designs must have good power/ground padcell - to - mesh (e.g. ring) interface.
  - Too few attachment points create bottlenecks

Power Bottleneck



Power w/ Improved Ring



Manufacturability

Antennas

Charging in Semiconductor Processing

- Many process steps use plasmas, charged particles
- Charge collects on conducting poly, metal surfaces
- Capacitive coupling: large electrical fields over gate oxides
- Stresses cause damage or complete breakdown
- Induced threshold (Vt) shifts affect device matching (e.g., for analog) and timing predictability

Antenna Ratio Limits

- Standard solution: limit *antenna ratio*
- Antenna ratio = $(A_{poly} + A_{M1} + \ldots) / A_{gate\text{-}ox}$
- E.g., antenna ratio < 400
- A_{Mx} = metal(x) area that is electrically connected to node without using metal (x+1), **and** not connected to an active area

Antenna Solutions

- General solution: bridging (break antenna by moving route to higher layer)
  - Introduces extra vias, congestion
  - As antenna ratios get small (and, gate areas decrease rapidly) more bridges are needed
- Antennas also solved by protection diodes
  - Also a costly solution: leakage power, area penalties, *timing penalties*
- Note: Antenna ratio limits of ~400 kill router performance (completion rate, runtime) !

Observations re Antenna Ratios

- Observation: There should be TWO antenna ratios
  - One for thick-ox (400), one for thin-ox (~2000)
  - Only the more constraining one has been traditionally put into design rules
  - Thick-ox mostly applies to I/O drivers; thin-ox applies to core → can (safely?) relax the router's antenna rules by factor of 5x in core
- Observation: There should be TWO types of antenna ratios
  - One is a *per-layer* rule, one is a *cumulative* rule
  - Sometimes per-layer rules are put into the design rules, but they are not as physically meaningful as cumulative rules

OPC and PSM

Subwavelength Optical Lithography — Technology Limits

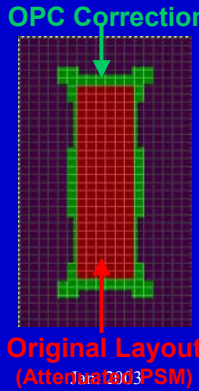
- Implications of Moore's Law for feature sizes
- Steppers not available; WYSIWYG (layout = mask = wafer) fails after .35μm generation
- Optical lithography
  - circuit patterns optically projected onto wafer
  - feature size limited by diffraction effects
  - Rayleigh limits
    - resolution R proportional to λ / NA
    - depth of focus DOF proportional to λ / NA^2
- Available knobs
  - amplitude (aperture): OPC
  - phase: PSM

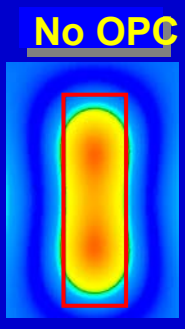
Next-Generation Lithography and the Subwavelength Gap

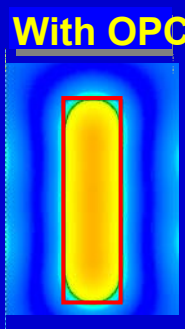
- EUV
- X-rays
- E-beams
- All require significant R&D, major infrastructure changes
- > 30 years of infrastructure and experience supporting optical lithography

Optical Proximity Correction (OPC)

- Layout modifications improve process control
  - improve yield (process latitude)
  - improve device performance







Optical Proximity Correction (OPC)

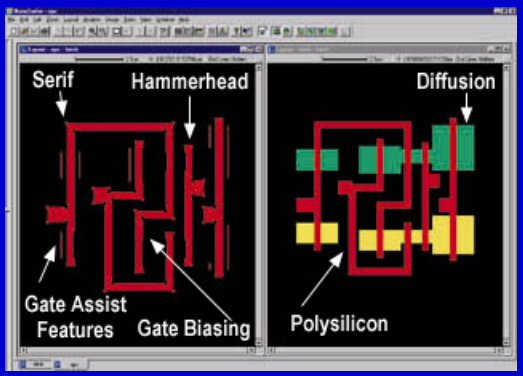
- Cosmetic corrections; complicates mask manufacturing and dramatically increases cost
- Post-design verification is essential

Rule-based OPC
- apply corrections based on a set of predetermined rules
- fast design time, lower mask complexity
- suitable for less aggressive designs

Model-based OPC
- use process simulation to determine corrections on-line
- longer design time, increased mask complexity
- suitable for aggressive designs

OPC Features

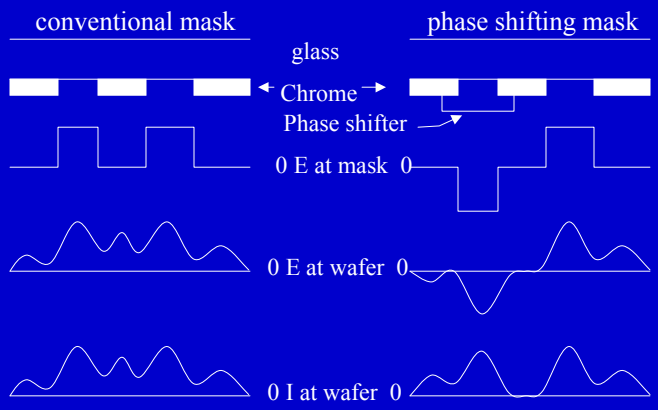
- Serifs - for corner rounding
- Hammerheads - for line-end shortening
- Gate assists (subresolution scattering bars) - for CD control
- Gate biasing - for CD control
- Affects custom, hierarchical and reuse-based layout methodologies



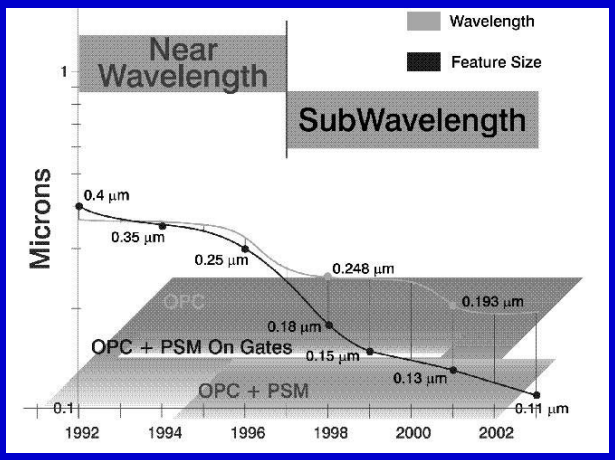
OPC Issues

- WYSIWYG broken → (mask) verification bottleneck
- Pass functional intent down to OPC insertion
  - OPC insertion is for predictable circuit performance, function
  - Make corrections to win $$$, reduce perf variation → **cost-driven RET**
- Pass limits of manufacturing up to layout
  - don't make corrections that can't be manufactured or verified
  - Mask Error Enhancement Factor, etc.
- Layout needs models of OPC insertion process
  - geometry effects on cost of required OPC to yield function
  - costs of breaking hierarchy (beyond known verification, characterization costs)

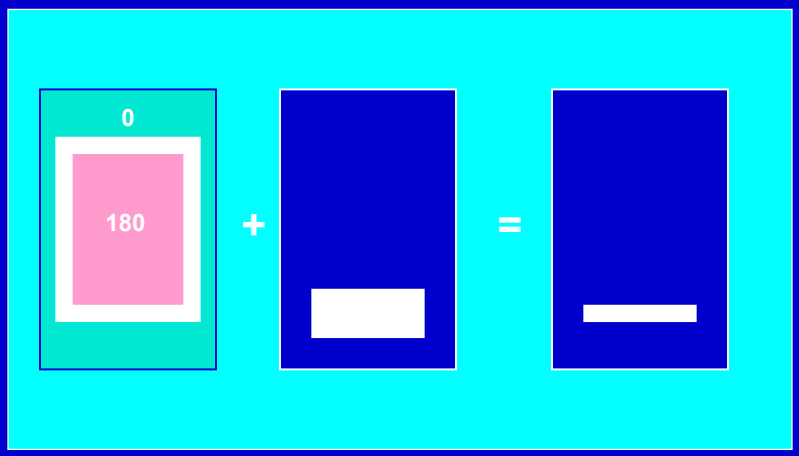
Phase Shifting Masks

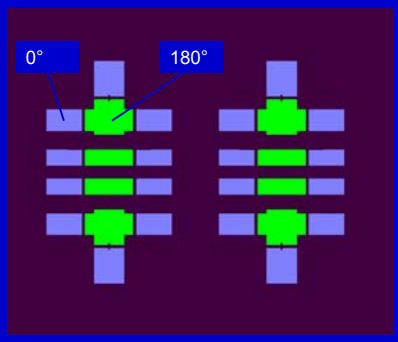


Applicability of OPC and PSM

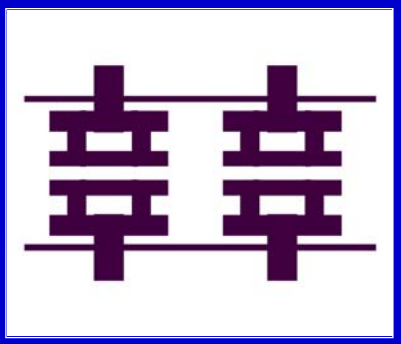


Double-Exposure Bright-Field PSM



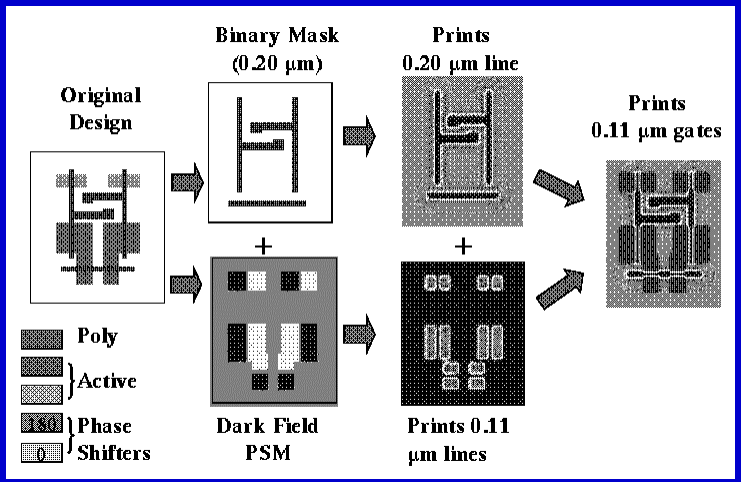


1. Alternate PSM Mask



2. Trim Mask (COG)

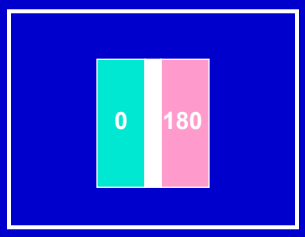
Gate Shrinking and CD Control



Phase Assignment Problem

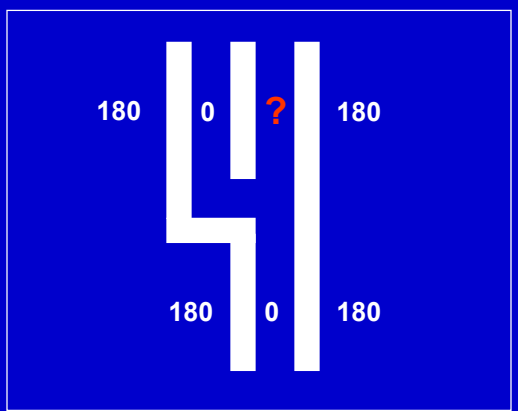
- Assign 0, 180 phase regions such that critical features with width (separation) < B are induced by adjacent phase regions with opposite phases

Bright Field AltPSM



Key: Global 2-Colorability

- If there is an odd cycle of "phase implications"
  → layout cannot be manufactured
  - layout verification becomes a global, not local, issue



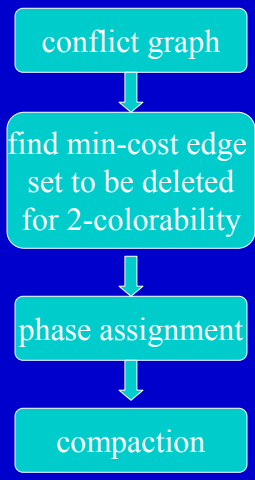
EDA Implications of PSM

- Must partition responsibility for phase-assignability
  - Good layout practices (local geometry)
    - (open) problem: is there a set of "design rules" that guarantees phase-assignability of layout ? (no T's, no doglegs, even fingers...)
  - **Automatic phase conflict resolution / bipartization (global colorability)**
  - Enable reuse of layout (free composability)
    - problem: how can we guarantee reusability of phase-assigned layouts, such that no odd cycles can occur when the layouts are composed together in a larger layout ?

Basic (Compaction-Oriented) Layout Approach

- Analyze input layout
- Find min-cost set of perturbations needed to eliminate all "odd cycles"
- Induce constraints for output layout
  - i.e., PSM-induced (shape, spacing) constraints
- Compact to get phase-assignable layout
- Key: Minimize the set of new constraints, i.e., break all odd cycles in conflict graph by deleting a minimum number of edges.

Phase-Assignable Layout Flow



PSM Flow Issues

- PSM must be "transparent" to ASIC auto-P&R
  - "free composability" is basis of cell-based methodology!
  - focus on poly layer → focus on placer, not router
- Iteration between placer and a separate tool is unacceptable
  - interface to auto-P&R tools is bulky (e.g., 100s of MB for DEF), slow
  - no known convergent method for post-P&R phase-assignability checks to drive P&R to guaranteed correct solution (very difficult!)
- → P&R tool MUST deliver guaranteed phase-assignable poly layer

Types of Composability

- Same-row composability
  - any cell can be placed immediately adjacent (in the same row) to any other cell
- Adj-row composability
  - any cell can be placed in an adjacent cell row to any other cell, with the two cells having intersecting x-spans
- Four cases of cell libraries (G = guaranteed; NG = not guaranteed)
  - Case 1: adj-G, same-G
    - constrained cell layout; transparent to placer
  - Case 2: adj-G, same-NG
  - Case 3: adj-NG, same-G
  - Case 4: adj-NG, same-NG
    - unconstrained cell layout; least transparent to placer

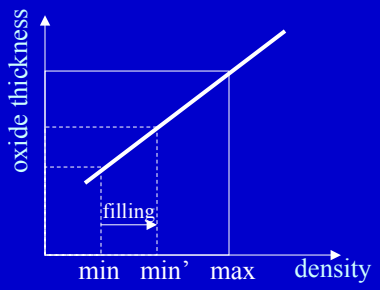
Density Control for CMP

Density Control for CMP

- Chemical-mechanical polishing (CMP)
  - applied to interlayer dielectrics (ILD) and inlaid metals
  - polishing pad wear, slurry composition, pad elasticity make this a very difficult process step
- Cause of CMP variability
  - pad deforms over metal feature
  - greater ILD thickness over dense regions of layout
  - "dishing" in sparse regions of layout
  - huge part of chip variability budget used up (e.g., 4000Å ILD variation across-die)

Min-Variation Objective

- Relationship between oxide thickness and local feature density

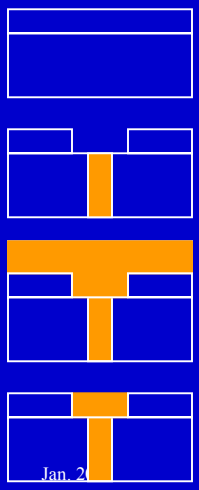


- Minimizing variation in window density over layout preferable to satisfying lower and upper density bounds

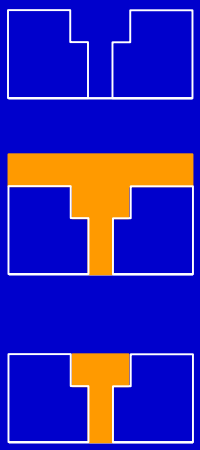
Damascene, Dual-Damascene

- Named after ancient technique for inlaying metal in ceramic or wood for decoration

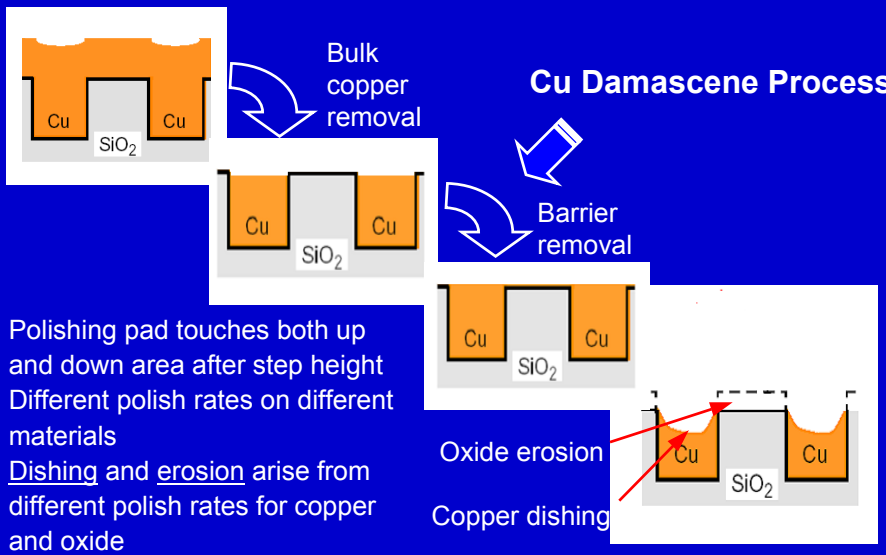
• Single Damascene



• Dual Damascene

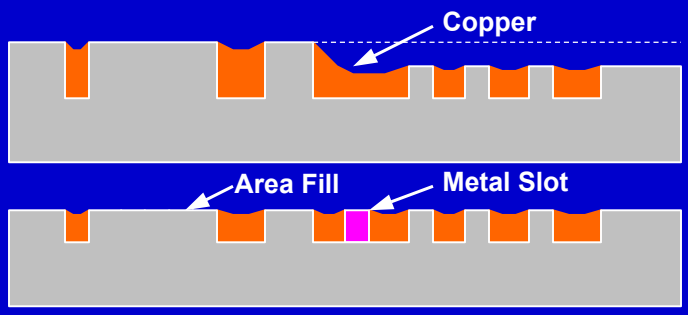


Dual-Damascene Cu Process



- Polishing pad touches both up and down area after step height
- Different polish rates on different materials
- Dishing and erosion arise from different polish rates for copper and oxide

Cu CMP Area Fill / Metal Slot



- ***Dishing*** can thin the wire or pad, causing higher-resistance wires or lower-reliability bond pads
- ***Erosion*** can result in a sub-planar dip on the wafer surface, causing short-circuits between adjacent wires on next layer
- *Oxide erosion* and *copper dishing* can be controlled by *area filling* and *metal slotting*

Density Control for CMP

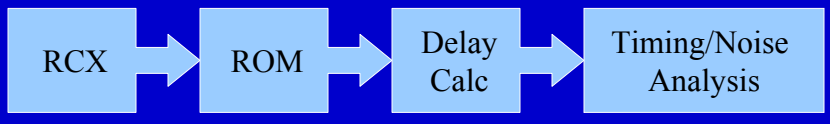
- Layout density control
  - density rules minimize yield impact
  - uniform density achieved by post-processing, insertion of dummy features
- Performance verification (PV) flow implications
  - accurate estimation of filling is needed in PD, PV tools (else broken performance analysis flow)
  - filling geometries affect capacitance extraction by > 50%
  - is a multilayer problem (coupling to critical nets, contacting restrictions, active layers, other interlayer dependencies)

Density Rules

- Modern foundry rules specify layout density bounds to minimize impact of CMP on yield
- Density rules control local feature density for $w \times w$ windows
  - e.g., on each metal layer every 200um × 200um window must be between 35% and 70% filled
- Filling = insertion of "dummy" features to improve layout density
  - typically via layout post-processing in PV / TCAD tools
    - boolean operations on layout data
  - affects vital design characteristics (e.g., RC extraction)
  - accurate knowledge of filling is required during physical design and verification

Need Density-Aware Layout

- Performance verification flow:



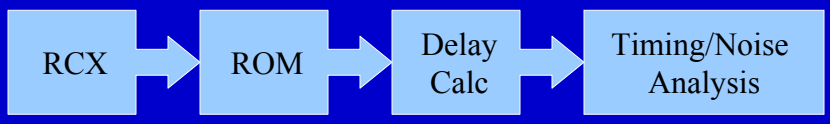
- Filling/slotting geometries affect RC extraction

VICTIM LAYER TOTAL CAPACITANCE (10^{-15} F)			
Same layer-i neighbors?	Fill layers i-1, i+1?	$\varepsilon = 3.9$	$\varepsilon = 2.7$
N	N	2.43 (1.0)	1.68 (1.0)
N	Y	3.73 (1.54)	2.58 (1.54)
Y	N	4.47 (1.84)	3.09 (1.84)
Y	Y	5.29 (2.18)	3.66 (2.18)

Up to 1% error in extracted capacitance
Reliability also affected (e.g. slotting of power stripes)

Need Density-Aware Layout

- Performance verification flow:



- Can be considered as ``single-layer'' problem

Middle Victim Conductor Total Capacitance (10^{-15} F)			
Fill layer offset	Fill geometry	$\varepsilon = 3.9$	$\varepsilon = 2.7$
N	10×1	3.776 (1.0)	2.614 (1.0)
N	1×1	3.750 (0.99)	2.596 (0.99)
Y	10×1	3.777 (1.00)	2.615 (1.00)
Y	1×1	3.745 (0.99)	2.593 (0.99)

- Caveat: contacting, active+gate layers, other layer interactions

Limitations of Current Density Control Techniques

- Current techniques for density control have three key weaknesses:
  - (1) only the average ***overall*** feature density is constrained, while local variation in feature density is ignored
  - (2) density analysis does not find ***true*** extremal window densities - instead, it finds extremal window densities only over fixed set of window positions
  - (3) fill insertion into layout does not minimize the maximum variation in window density
- In part, due to PV tool heritage: Boolean operations, inability to touch layout, etc.

Layout Density Control Flow

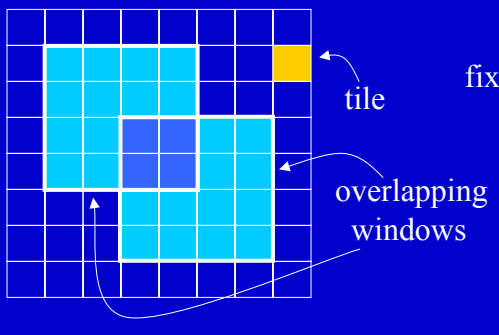
Density Analysis
- find total feature area in each window
- find maximum/minimum total feature area over all w × w windows

- find slack (available area for filling) in each window

Fill synthesis
- compute amounts, locations of dummy fill
- generate fill geometries

Fixed *r*-Dissection Regime

- Feature area density bounds enforced only for ***fixed*** set of ***w*** × ***w*** windows
- Layout partitioned by r^2 distinct fixed dissections
- Each ***w*** × ***w*** window is partitioned in r^2 ***tiles***



fixed *r*-dissection with *r* = 4

Filling Problem

- **Given** design rule-correct layout of k disjoint rectilinear features in $n \times n$ region
- **Find** design rule-correct **filled** layout
  - no fill geometry is added within distance B of any layout feature
  - no fill is added into any window that has density $\geq U$
  - minimum window density in the filled layout is maximized (or has density $\geq$ lower bound L)

Filling Problem in Fixed-Dissection Regime

- **Given**
  - fixed r-dissection of layout
  - feature ***area***[T] in each tile T
  - ***slack***[T] = area available for filling in T
  - maximum window density U
- **Find** total fill area p[T] to add in each T s.t. any $w \times w$ window W has density $\leq U$ and $\min_W \sum_{T \in W}$ (area[T] + p[T]) is maximized

Fixed-Dissection LP Formulation

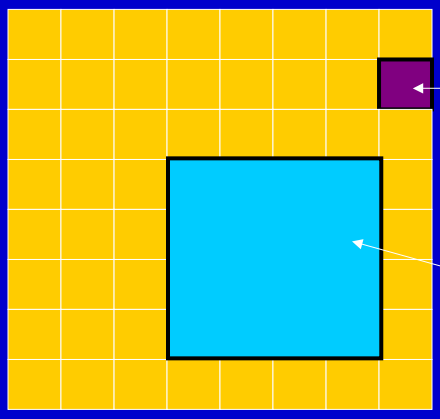
- Maximize M (= lower bound on window density)
- Subject to:
  - For any tile T: $0 \le p[T] \le \textit{pattern} \times slack[T]$
  - For any window W:

$$\sum_{T \in W} p[T] + area[T] \le U \times w^2$$

$$M \le \sum_{T \in W} (p[T] + area[T])$$

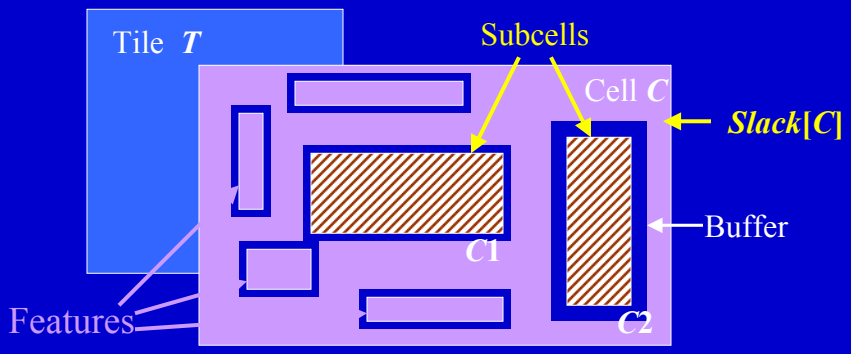
(***pattern*** = max achievable pattern area density)

Fixed-Dissection LP Formulation



Hierarchical Density Control

- Hierarchical filling = master cell filling



Hierarchical LP Formulation

- For any cell instance *C* of master cell **C** and tile T, $\gamma[C,T]$ is portion of slack[**C**] in intersection of *C* with T:

$$\gamma[C,T] = \text{slack}(C \cap T)/\text{slack}[\mathbf{C}]$$

- New variable d[**C**] per each master cell **C**:

d[**C**] = ***filling per master cell*** **C**

- New constraints:
  - for total amount of filling added into tile T:

$$p[T] = \sum_{C \cap T} d[\mathbf{C}] \cdot \gamma[C,T]$$

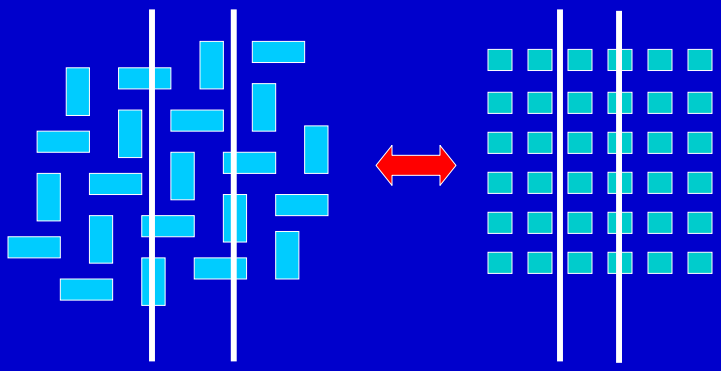
  - for amount of filling added into each master cell C:

$$0 \le d[\mathbf{C}] \le \textit{pattern} \times \text{slack}[\mathbf{C}]$$

Synthesis of Filling Patterns

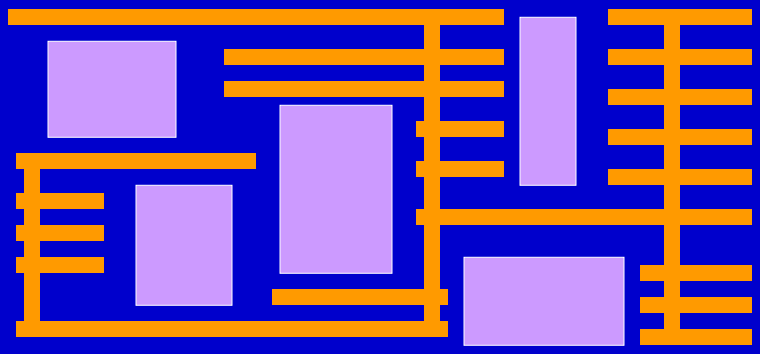
- Given area of filling pattern p[i,j], insert filling pattern into tile T[i,j] ***uniformly*** over available area
- Desirable properties of filling pattern
  - uniform coupling to long conductors
  - either grounded or floating

Basket-Weave Fill Pattern



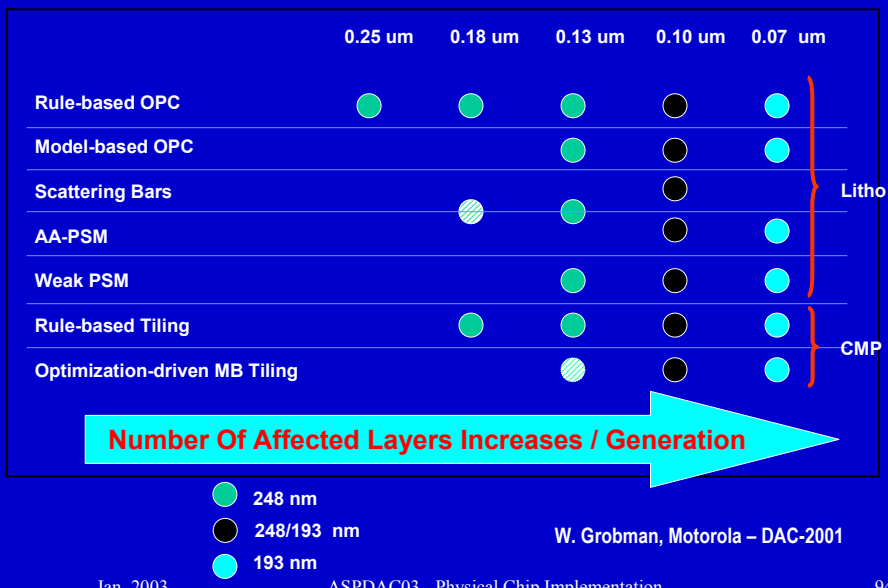
Each vertical/horizontal crossover line has same overlap capacitance to fill

Grounded Fill Pattern



Fill with horizontal stripes,
then span with vertical lines

Reticle Enhancement Roadmap



W. Grobman, Motorola – DAC-2001

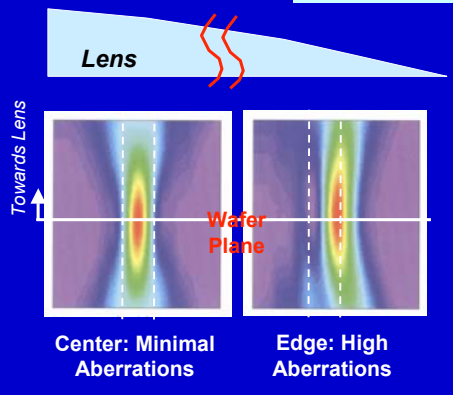
Optical Lithography Becomes Harder

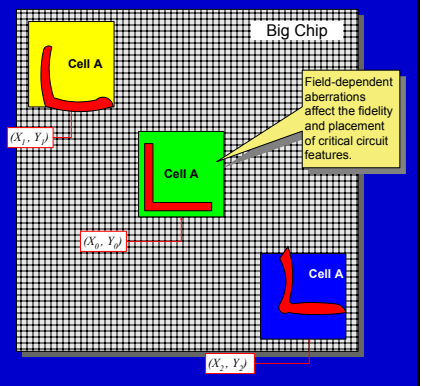
- Process window and yield enhancement
  - Forbidden width-spacing combinations (defocus window sensitivities)
  - Complex "local DRCs"
- Lithography equipment choices (e.g., off-axis illumination)
  - Forbidden configurations (wrong-way critical-width doglegs, or diagonal features)
- OPC subresolution assist features (scattering bars)
  - Notch rules, critical-feature rules on local metal

Many Other Optical Litho Issues

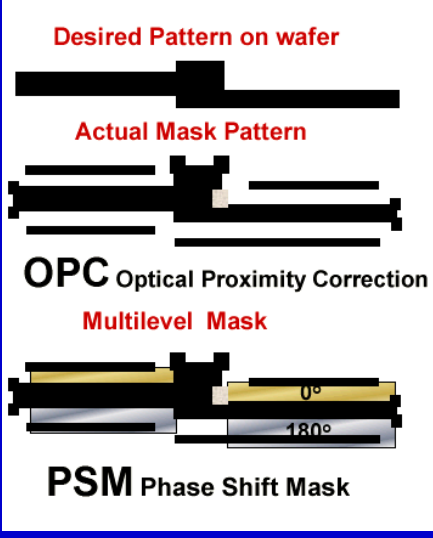
- Example: Field-dependent aberrations cause placement errors and distortions

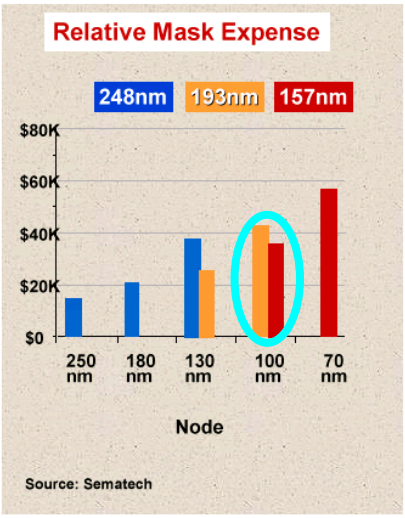
$$\mathrm{CELL\_A}(X_1, Y_1) \neq \mathrm{CELL\_A}(X_0, Y_0) \neq \mathrm{CELL\_A}(X_2, Y_2)$$





Mask NRE Cost (1999 Sematech)



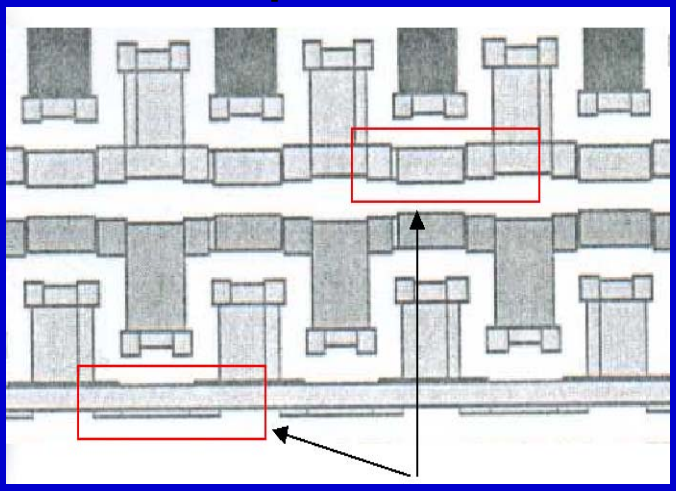


"$1M mask set" at 100nm, but average only 500 wafers per set

Mask Data and $1M Mask NRE

- Too many data formats
  - Most tools have unique data format
  - Raster to variable shaped-beam conversion is inefficient
  - Real-time manufacturing tool switch, multiple qualified tools → duplicate fractures to avoid delays if tool switch required
- Data volume
  - OPC increases figure count acceleration
  - MEBES format is flat
  - ALTA machines (mask writers) slow down with > 1GB data
  - Data volume strains distributed manufacturing resources
- Refracturing mask data
  - 90% of mask data files manipulated or refractured: process bias sizing (iso-dense, loading effects, linearity, …), mask write optimization, multiple tool formats, …

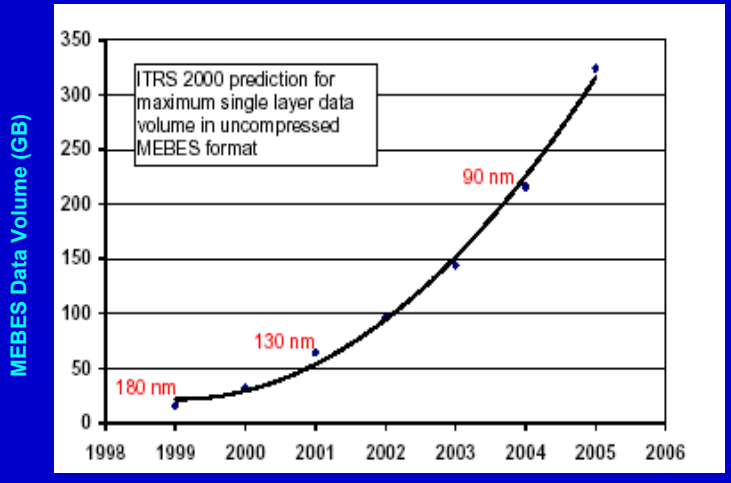
Context-Dependent Fracturing



Same pattern, different fracture

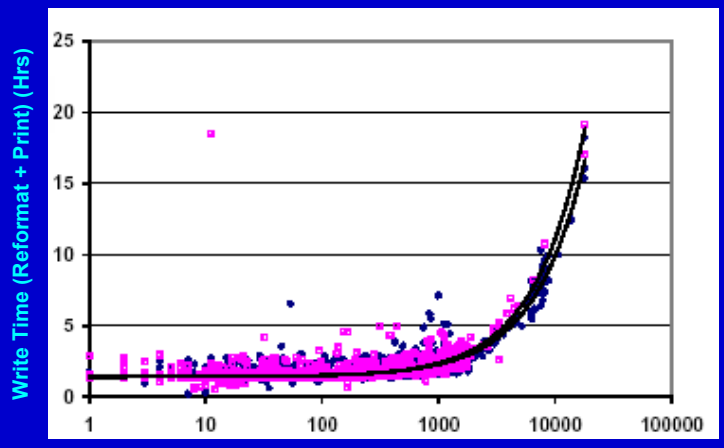
P. Buck, Dupont Photomasks – ISMT Mask-EDA Workshop July 2001

ITRS Maximum Single Layer File Size



P. Buck, Dupont Photomasks – ISMT Mask-EDA Workshop July 2001

ALTA-3500 Mask Write Time

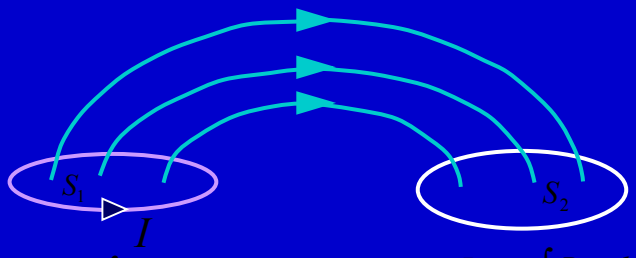


P. Buck, Dupont Photomasks – ISMT Mask-EDA Workshop July 2001

Inductance

What is Inductance?

- Inductance is the flux induced by current variation



$$\Phi_{11} = \int_{S_1} B_1 \bullet ds_1 \qquad \Phi_{12} = \int_{S_2} B_1 \bullet ds_2$$

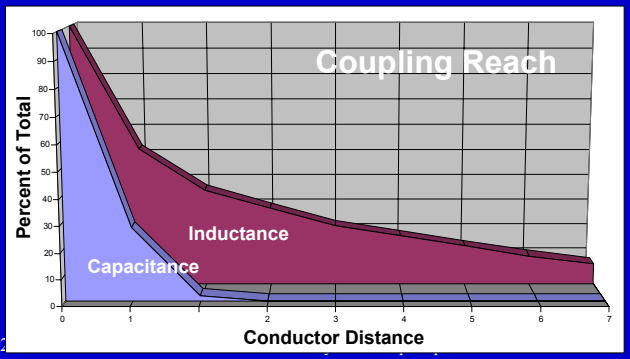
Self Inductance $= \dfrac{\Phi_{11}}{I}$ **Mutual Inductance** $= \dfrac{\Phi_{12}}{I}$

Why Inductance is Important

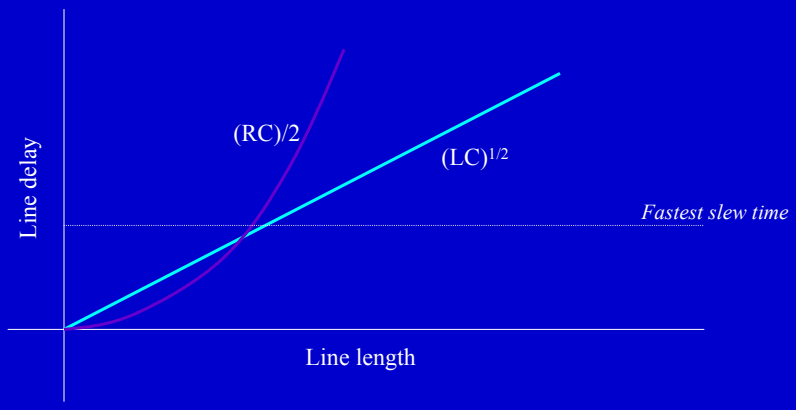
- If $\omega L \approx R$ where $\omega = 2\pi f = 2\pi\left(\dfrac{1}{\pi t_r}\right)$
  - Use of Copper, R is reduced
  - Faster clock speeds
  - Use of thick, low resistance global lines
  - Chips are getting larger ⇒ long lines ⇒ large current loops
  - Note that signal *rise time* determines frequency of interest, not clock frequency

Edge Rate and Coupling Reach

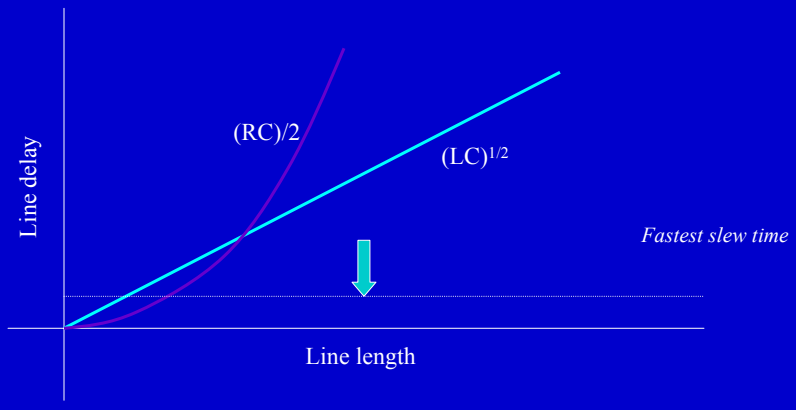
- Fast edge rate ⇒reactance ωL approaches resistance
  - A resistive wire model may not be sufficiently accurate
  - Mutual inductance and resistance become a noise, delay, and modeling concern



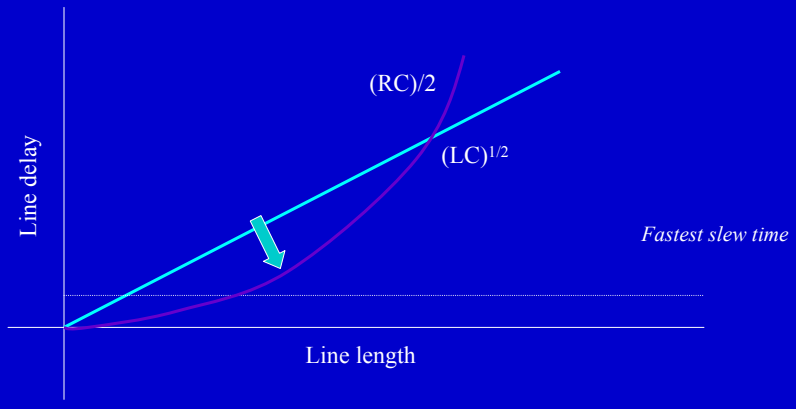
Why Inductance is Important



Why Inductance is Important



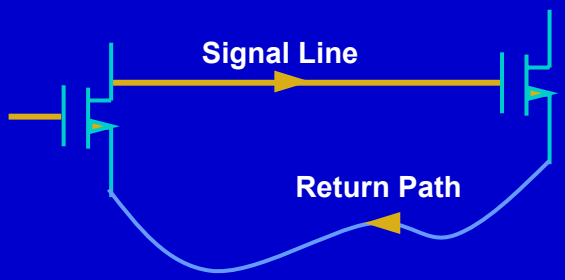
Why Inductance is Important



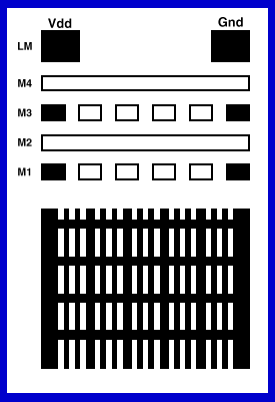
This analysis may require an "effective" inductance that considers simultaneous switching.

On-Chip Inductance

- Inductance is a loop quantity (Flux Linkage to a loop)
- Knowledge of return path is required



Typical On-Chip Power Distribution



Traditional grid structure

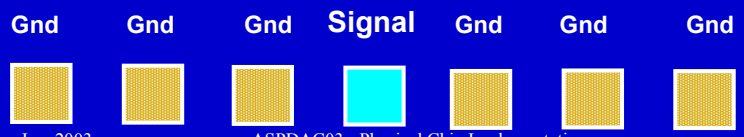
DEC 21164

No ground plane!

Return Path Hard to Determine

- Frequency dependent return path
  - At Low Frequency, $(R >> \omega L)$, current tries to
    - minimize impedance
    - minimize resistance
    - use as many returns as possible (parallel resistances)

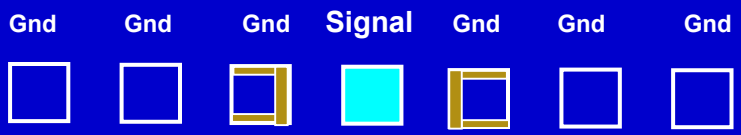
$$(R + j\omega L)$$



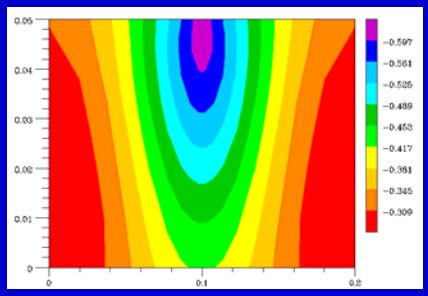
Return Path Hard to Determine

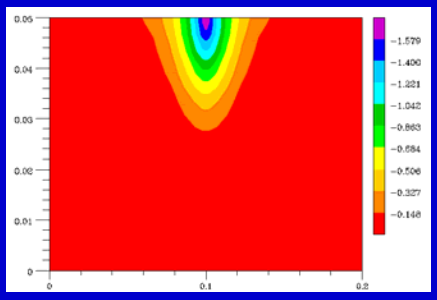
- Frequency dependent return path
  - At High Frequency, $(R << \omega L)$, current tries to
    - minimize impedance
    - minimize inductance
    - use smallest possible loop (closest return path)
  - At frequencies where R becomes frequency-dependent, L dominates.
  - At frequencies where L dominates, current returns have "collapsed."
  - Power and ground lines are always available as low-impedance current returns at high frequencies.

$$(R + j\omega L)$$



Loop Impedance: Substrate Effects





100 MHz 2 GHz

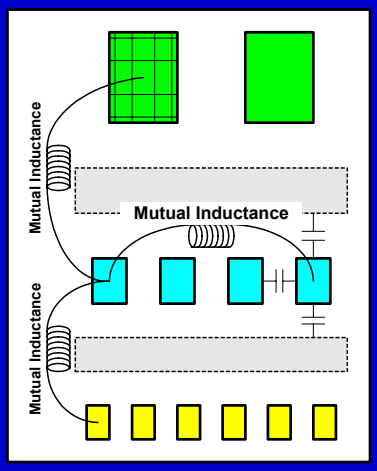
- Must provide localized return paths in metal to reduce return resistance and loop impedance

Inductance vs. Capacitance

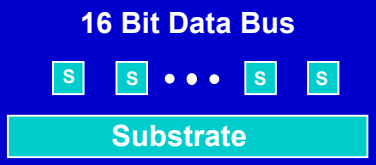
- Capacitance
  - Locality problem easy… electric field lines "suck up" to nearest neighbor conductors
  - Local calculation hard… hence all the effort in "accuracy"
- Inductance
  - Locality problem hard… magnetic field lines are not local; current returns can be complex
  - Local calculation easy… no strong geometry dependence… analytic formulae work very well.

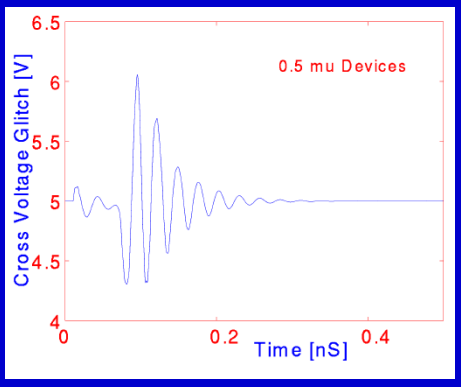
Modeling of Multi-layer, Multi-Conductor Systems

- Capacitance to *nearest* neighbor is sufficient
- Inductance to *distant* neighbors is necessary
- Inductance to wires on *other layers* is also necessary
  - Switching activity often unknown!



Signal Integrity Problems Due to Inductance





- 15 Data lines switching simultaneously
- Vcc = 5V creates a one volt glitch

EDA Challenges From Inductance

- Holistic SI/noise management
  - Inductance effects primarily a noise concern
  - Conventional RC coupling checks can significantly underestimate noise
  - Back-end extraction/verification for LRC(f) is costly
  - Noise estimates may be insufficient if taken without regard to device, loading, and switching conditions
- 3D RLC accurate/fast extraction is essential to accurately model DSM effects
- Global busses, clocks, & critical nets must be shielded to minimize coupling effects
- Tools for 3D analyses of power supply integrity
  - Signal and supply noise are correlated

Inductance Methodology Goals

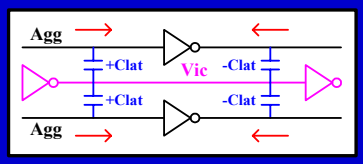
- System to be modeled is very complex
  - Significant design and simulation time
  - Easy for the inexperienced to get it wrong
- Want methodology to facilitate RC modeling but manage away the inductance:
  - Various configurations of supply routing
  - Multiple wiring lengths, loads, and pitches
    - Various distributions of fixed capacitance
    - Various combinations of width and spacing
  - For a range of logic families (static and dynamic)
  - Under worst-case conditions (typically fastest edges)

Intuition For Interconnect Design

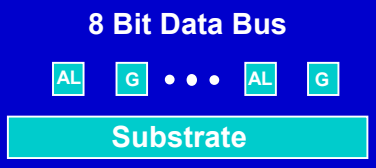
- Seesaw effect between inductance and capacitance
- Minimize *variations* in L and C rather than absolutes
  - Techniques used to minimize variation in capacitive coupling may also benefit inductive coupling
- Employ techniques which reduce net current injection for a given system

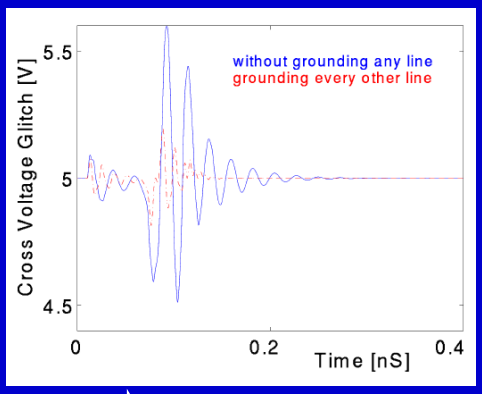
Useful Techniques

- Introduce shield wires (returns) between signals
  - Of benefit even for SR ratios beyond 2:1
  - Required anyway for supply distribution
- Introduce reference planes between layers
  - Provides excellent supply performance (low IR, I^2R)
- Staggered repeater locations
- Signaling mechanisms
  - Complementary logic
  - Predecoded signals



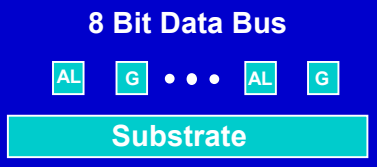
Reduce Coupling by Using Grounded Lines

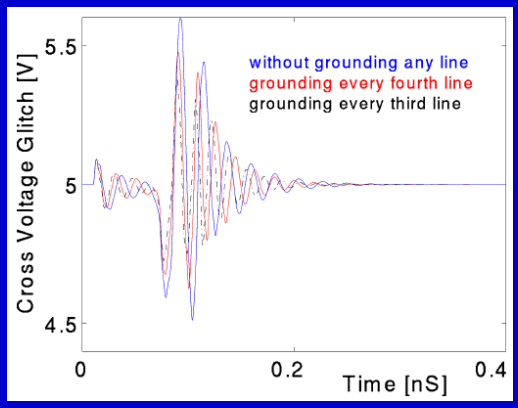




- Grounding every other line ⇨ 3x glitch reduction, 2x routing tracks used

Reduce Area Penalty By Grounding Fewer Lines

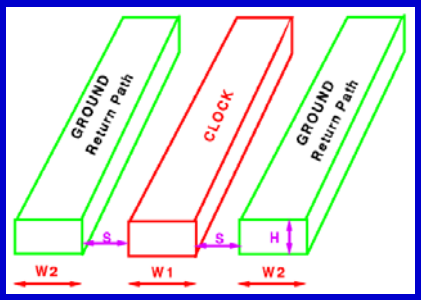


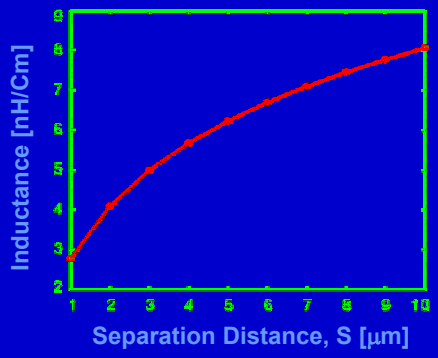


Grounding every third line ⇨ Factor of 1.5 Reduction
Grounding every fourth line ⇨ Factor of 1.2 Reduction!

Reduce Self Inductance By Shielding

Sandwiching signal line between ground lines

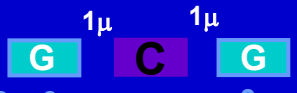




Keep return path as close as possible

Reduce Self Inductance by Optimizing Signal Line Width

Regular Structure



"Optimized" Structure

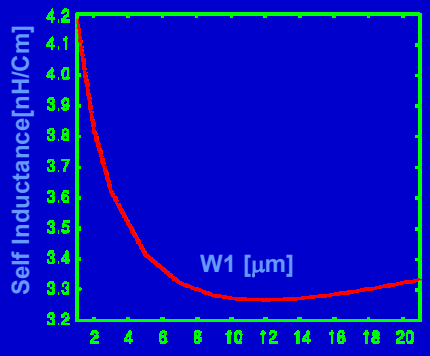
For small W1

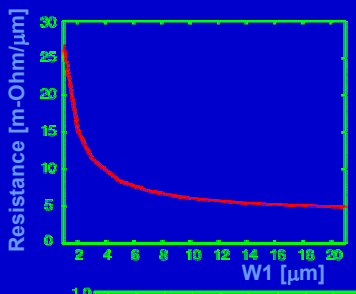
$L \approx \ln(1 + C1/W1) + C2$ As W1 ↑ , L ↓

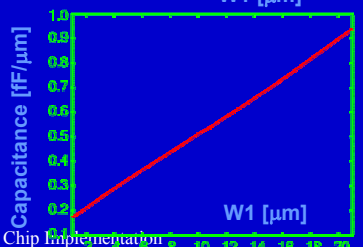
For large W1

$L \approx C3\,W1 + C4$ As W1 ↑ , L ↑

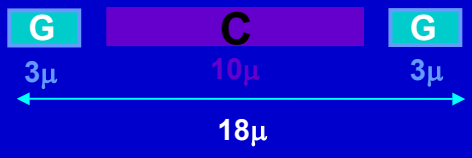
Reduce Self Inductance by Optimizing Signal Line Width





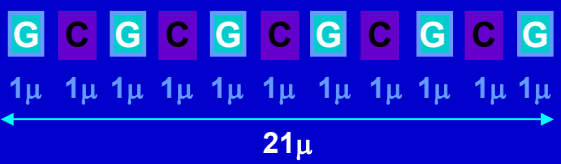


Interdigitate Signal Lines



t_r =100ps

$\frac{\omega L}{R}$ = 1.07

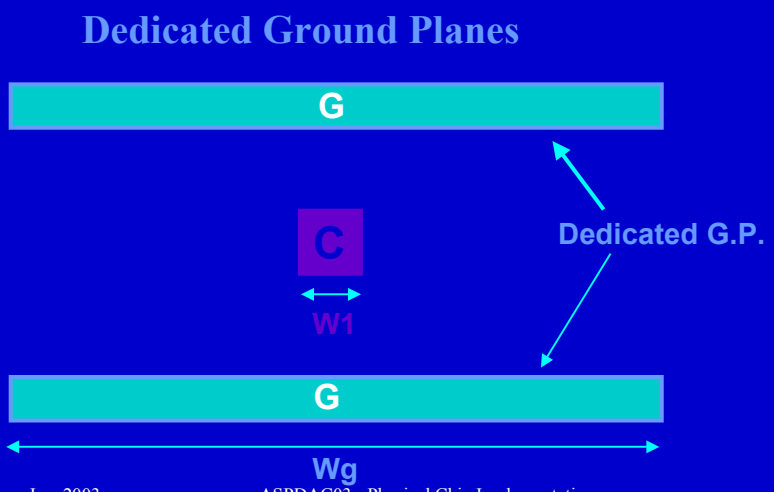


$\frac{\omega L}{R}$ = 0.20

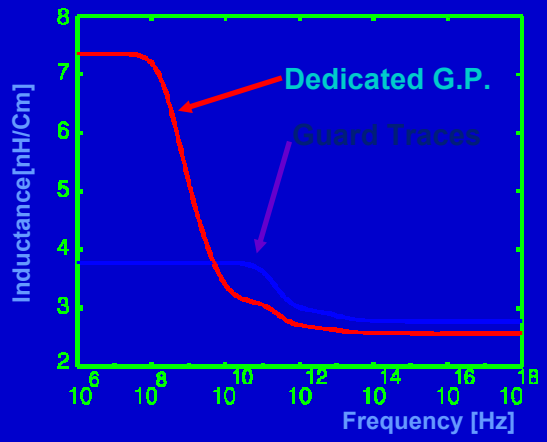
L2 = L1 / 4 C2 = 1.6 C1 R2= 1.3 R1

$\frac{\omega L}{R}$ decreased by a factor of 5.4 ⇒ RC analysis

Dedicated Ground Planes



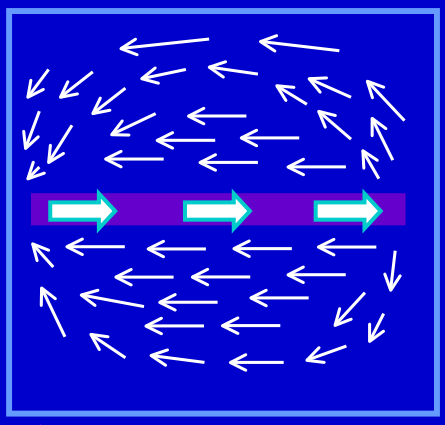
Guard Traces vs. Dedicated Ground Planes



- Below 5 GHz, guard traces are better

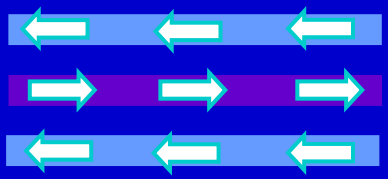
LF Current Distribution

Current Flow in Dedicated G.P.



Top View

Current Flow in Guard Traces



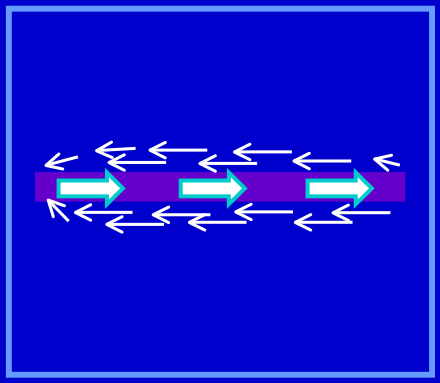
Smaller Current Loops

➡ Smaller Inductance

➡ Current spreads through outer, bigger current loops

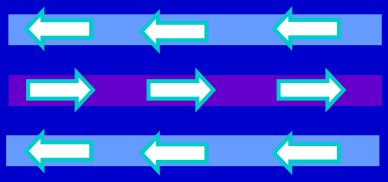
HF Current Distribution

Current Flow in Dedicated G.P.



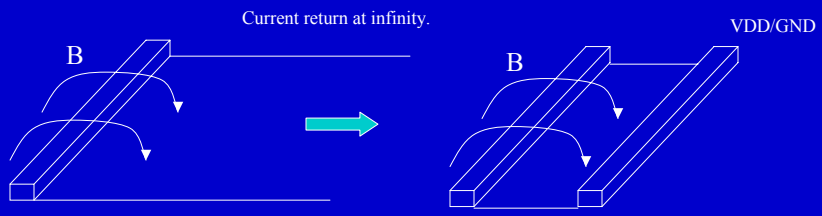
Top View

Current Flow in Guard Traces



Current concentrates underneath signal line → Much smaller current loops, and smaller inductance

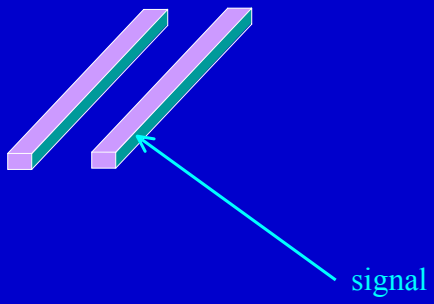
Extraction: Partial Inductance and Return-Limited Inductance



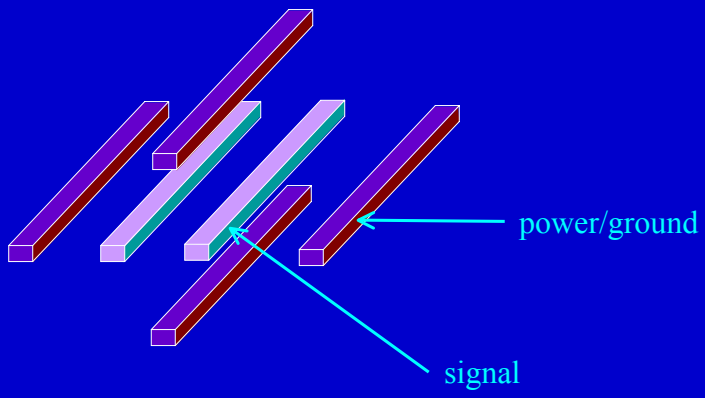
Return-Limited Inductance Extraction

- Need to determine which mutual inductances to discard and wish to use the power-ground network as an "always-available" current return.
- To do this, we:
  - Use the power-ground distribution to divide the interconnect into disjoint interaction regions. Mutual inductances between interaction regions are discarded.
  - Power-ground wires within the interaction region act as a "distributed ground plane".
- A set of geometry-based matrix decomposition rules guide the interaction region definition (halo rules).

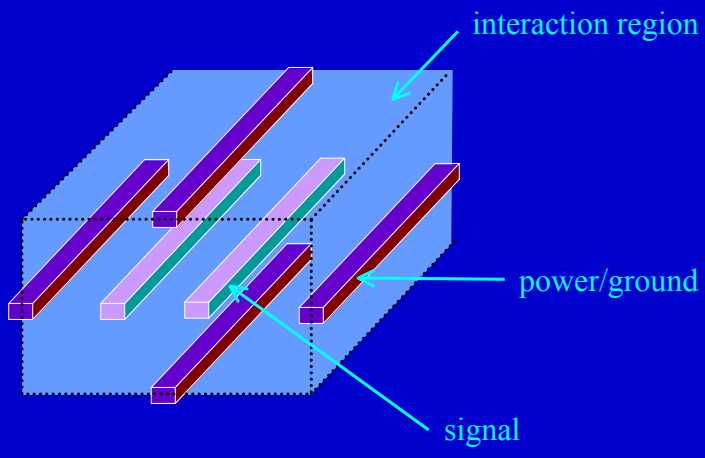
Interaction Region



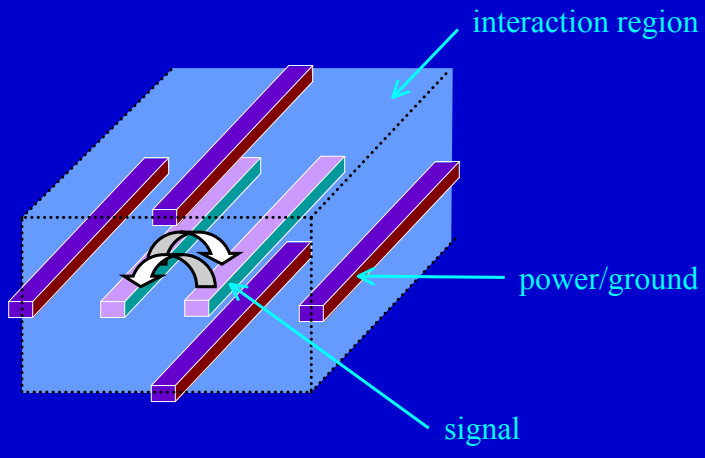
Interaction Region



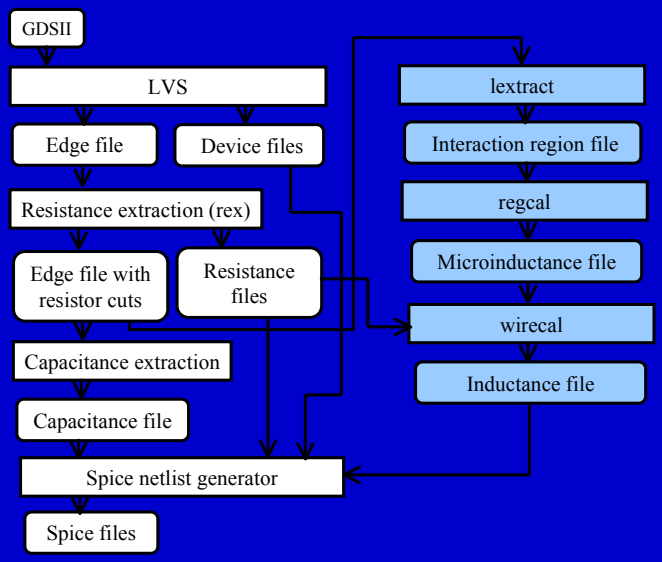
Interaction Region



Interaction Region



Assura RLCX Architecture



Sylvester/Shepard, 2001