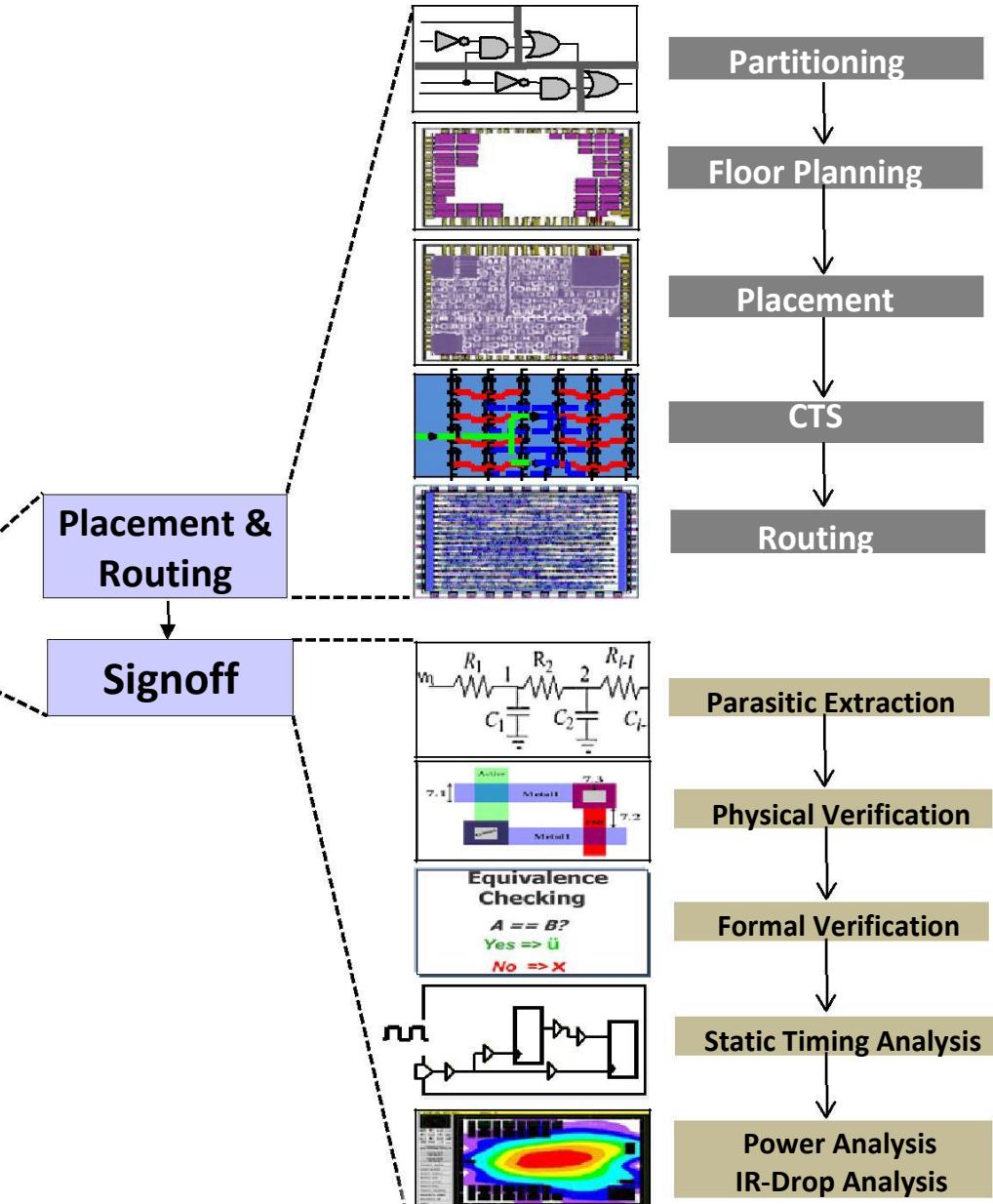
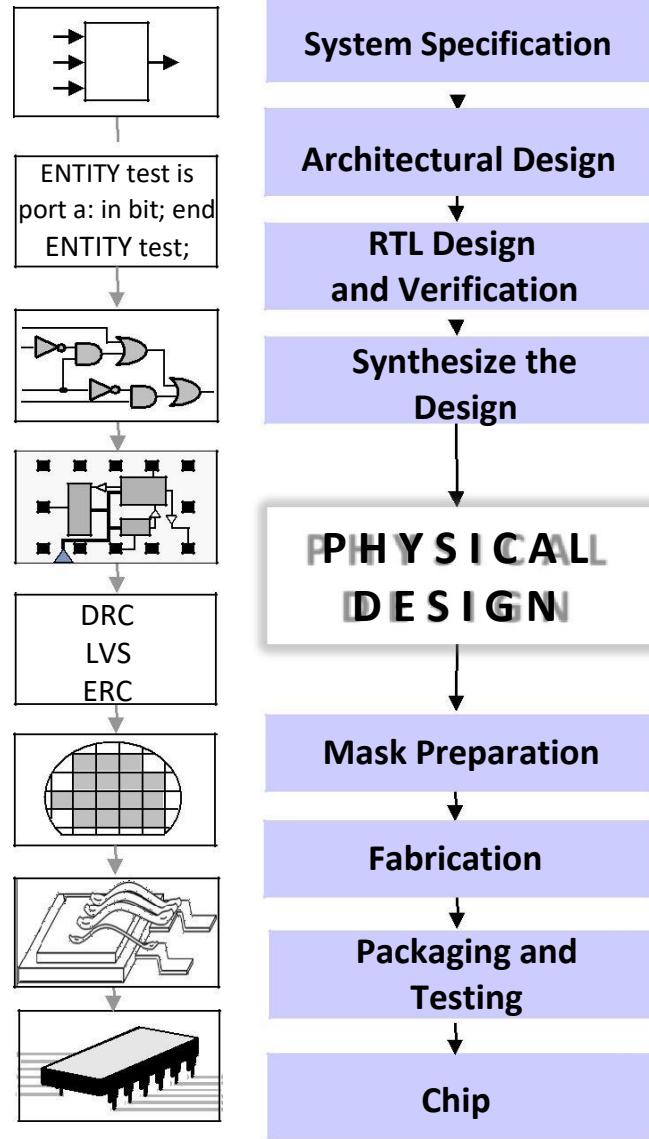


ASIC Physical Design

Outline

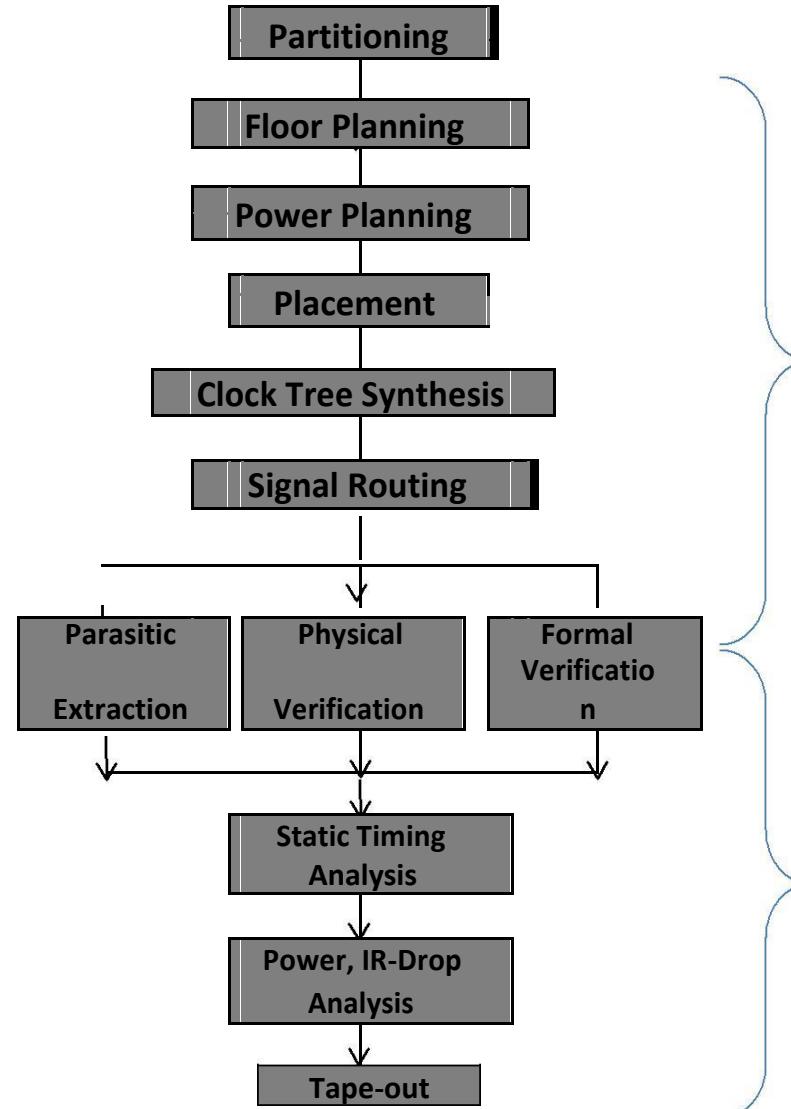
- **ASIC Design Flow**
- **Physical Design**
 - Introduction to Physical Design
 - Physical Design Inputs
 - Physical Design Flow
 - Import Design & Partitioning
 - Floorplanning & Power planning
 - Placement & Placement Optimizations
 - CTS & CTS Optimizations
 - Routing & Routing Optimizations
 - Physical Verification (DRC, LVS, ERC)
 - DFM Checks
 - Formal Verification (LEC)
 - Parasitic Extraction (RC Extraction)
 - Timing Analysis (STA), Power Analysis & IR Drop Analysis
 - Tapeout

ASIC Design Flow



Physical Design Introduction

- Structural Representation to Physical Implementation
i.e., Netlist to GDSII
- Stages
 - Placement and Routing (PnR)
 - Signoff
- Objectives
 - Timing
 - Congestion
 - Area
 - Power
- Possible Issues
 - Timing Violations
 - Congestion Issues
 - Design Rule Violations



PnR

Signoff



Physical Design Inputs

- Netlist (.v or .vhdl)
 - Netlist contains
 - Std. Cell instance – Name & Drive Strength
 - Macros & Memories instances
 - Netlist also consists of
 - Ports of Standard Cells and Macros
 - Interconnection details
- Constraints
 - Types of Constraints
 - Design Rule Constraints
 - Optimization Constraints
 - Design Rules from the Fab.
 - Max. Cap./ Transition/Fanout
 - Clock Uncertainties
 - Optimization Constraints from the designer
 - Timing Constraints/ Exceptions
 - Delay Constraints (Latency, Input Delay, Input Transition, Output Load and Output Transition)
 - Power and Area Constraints
 - System interface
 - Synopsys Design Constraints (SDC)
 - Timing Constraints
 - Clock Definition (Time Period, Duty Cycle)
 - Timing Exceptions (False Paths, Asynchronous Paths)
 - Non-Timing Constraints
 - Operating conditions
 - Wire load models
 - System interface, Design rule constraints (DRVs - Max. Cap./ Transition/Fanout)
 - Area constraints, Multi-voltage and Power optimization constraints
 - Logic assignments

Physical Design Inputs

- Liberty Timing

File (.lib or .db)

- Cell Logical View/ The Timing Library
- Std. Cell lib, Macro lib, IO lib
- Gate Delay = function of input transition time and output capacitance

- Library Exchange

Format (LEF)

- Cell Abstract View/ The Physical Library
- Std. Cell LEF
- Macro LEF
- IO LEF

- LIB contains

- Cell Type and Functionality
- Delay Models (WLD/ NLDM/ CCS)
- Pin/ Cell Timings and design rules
- PVT Conditions
- Power Details (Leakage and Dynamic)

- LEF contains

- Cell Name, Shape, Size, Orientation & Class
- Port/Pin Name, Direction and Layout Geometries
- Obstruction/ Blockages
- Antenna Diff. Area

Physical Design Inputs

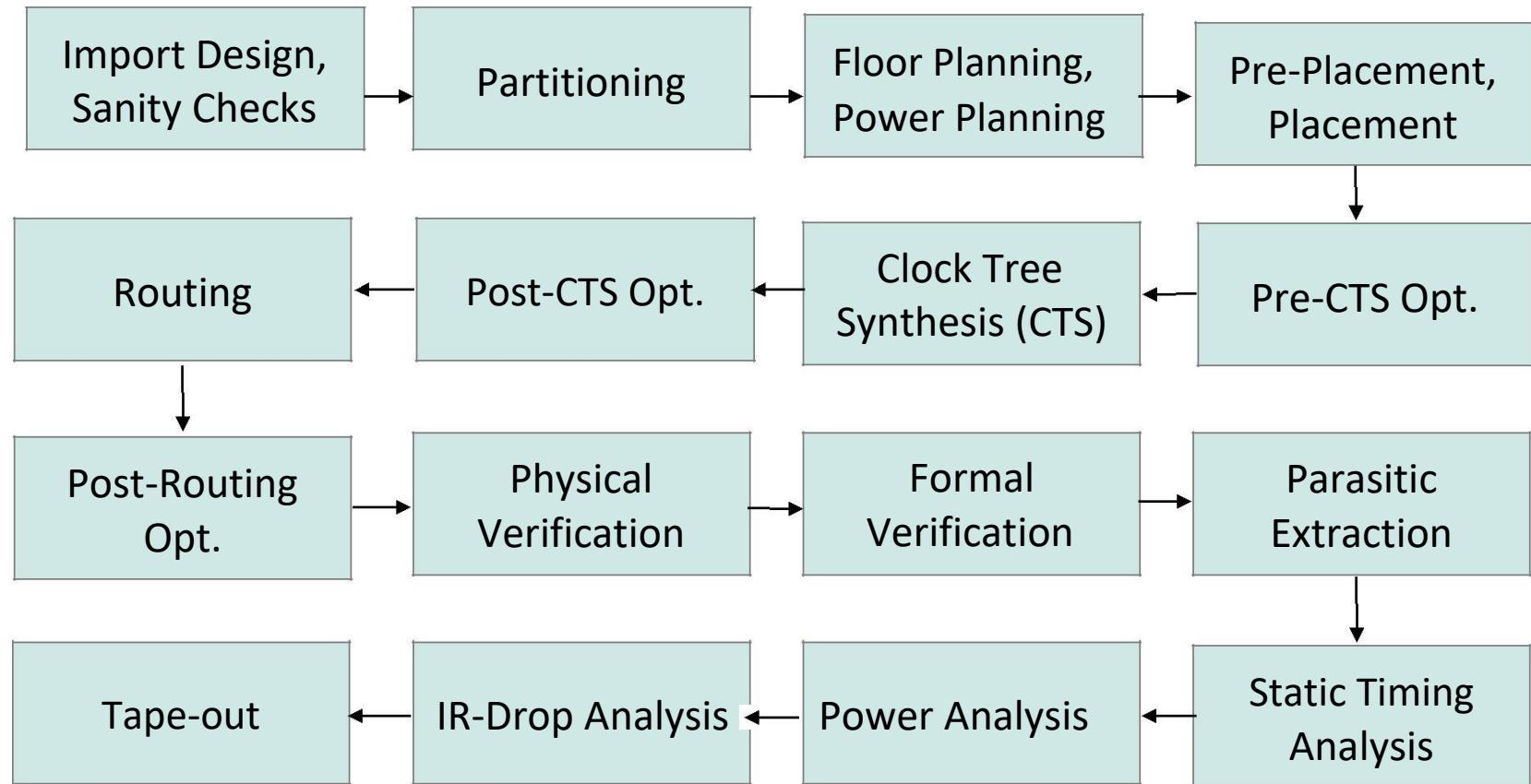
- Technology Related files
 - Technology file
 - Defines Units and Design Rules for Layers and Vias as per the Technology
 - Name and Number conventions of Layers and Vias
 - Physical and Electrical parameters of Layers and Vias
 - E.g.
Direction/Type/Pitch/Width/Offset/
Thickness/Resistance/Capacitance/
Max. Metal Density/Antenna Rule/
Blockages/Design Rules
 - Manufacturing Grid definition
 - Site/Unit Tile definition
 - Technology file has to load before loading other LEF files since it holds the layer information for that particular technology
 - .tech.lef (Cadence Format)
 - .tf - technology file (Synopsys Format)

- Interconnect Parasitic file
 - Used for layer parasitic extraction
 - Contains Layer/ Via capacitance and resistance values in a Lookup Table (LUT) format
 - Also used to generate parasitic formats for the extraction tools (e.g. nxtgrd, captbl)
 - Extraction tool formats are more accurate than interconnect parasitic formats
 - .ict - Interconnect Technology Format (Cadence Format)
 - .itf - Interconnect Technology Format (Synopsys Format)
 - .ptf - Process Technology File (Mentor Graphics Format)
- Map file
 - Useful if there are 2 different naming conventions in Technology file, LEF or Interconnect Parasitic file

Physical Design Inputs

- Power Specification File
 - Power Modes & Power Domains
 - Tie Up supply & Tie Low supply
 - Power Nets & GND Nets
- Optimization Directives
 - Don't use
 - Cells that are not supposed to optimize
 - Size only/ use only
 - Upsizing/ Downsizing only with this list of cells
- Design Exchange Formats
 - List & locations of Components, Vias, Pins, Nets, Special nets
 - Die dimensions, Row definitions, Placement and Bounding Box Data, Routing Grids, Power Grids, Pre-routes
 - .def, .fp are the common formats
- Clock Tree Constraints/ Specification
 - Root Pin Definition
 - Insertion Delay (ID) and Skew Target
 - Maximum Capacitance/ Transition/ Fanout (DRVs)
 - Transition can be classified into Leaf Transition and Buffer Transition
 - No. of Buffer Levels (Tree depth)
 - List of Buffers/ Inverters for CTS
 - List of Through pin, Preserved Pin, Exclude Pin
 - NDRs can be defined in CTS Spec. for the Clock Tree Routing
 - Macro Models
- IO Information File
 - Pin/ Pad locations
 - Edge and order for IO Placement
 - .tdf, .io are common formats

Physical Design Flow



Import Design

- Import Design
 - The following input files information are loaded to the PnR tool
 - Netlist (.v/ .vhdl/ .edif)
 - Physical Libraries (.lef)
 - Timing Libraries (.lib)
 - Technology Files
 - Constraints (.sdc)
 - IO Info. File (optional)
 - Power Spec. File (optional)
 - Optimization Directives (optional)
 - Clock Tree Spec. File (optional at floorplan stage)
 - DEF/ FP (optional if floorplan is not done)
 - Core area is approximately calculated by the tool from the Netlist
 - While Importing, first we have to load the LEF files and then LIB files

Import Design

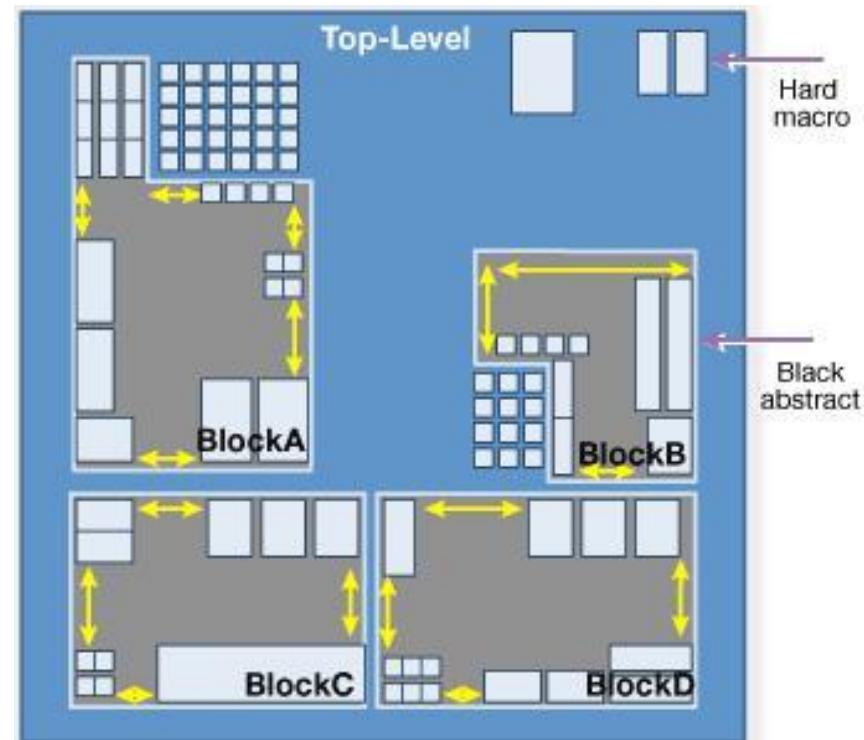
- Sanity Checks
 - Sanity Checks mainly checks the quality of netlist in terms of timing
 - It also consists of checking the issues related to Library files, Timing Constraints, IOs and Optimization Directives
 - Some of the Netlist Sanity Checks:
 - Floating Pins
 - Unconstrained Pins
 - Un-driven i/p Ports
 - Unloaded o/p Ports
 - Pin direction mismatches
 - Multiple drivers etc.
 - Other possible issues include Unconnected/ Wrongly Connected Tie-high/ Tie-low Pins and Power Pins (since Tie-up or Tie-down connectivity always through Tie-Cells)

Partitioning

- Physical Design Netlist
 - All Ports must be defined and should be present
 - No Assignment Statements (1'b0 or 1'b1 statements): Assignment statements causes feed-through (i/p directly to o/p) and can be avoided by adding buffers
 - No Unmapped Cells
 - No Combinational Timing Loops
- Styles of Implementation
 - Flat
 - Small to Medium ASIC
 - Better Area Usage Since no reserve space around each sub-design for power/ground
 - Hierarchical
 - For very large design
 - When sub-systems are design individually
 - Possible only if a design hierarchy exist

Partitioning

- The Hierarchical Partitioning is done prior to Floorplan
- Partition can be done based on
 - Design Hierarchy
 - Timing Criticality
 - Functionality
 - Clock Domain
 - Design Files
 - Block Size
- Partitioning Inputs and Outputs by Registers
- Minimize Cross-Partition-Boundary IO
- For Sub-block designs, the Partitioning is not required
- For Full Chip only we need to design with Partitioning



Floorplanning

- Terminologies and Definitions
 - Utilization
 - Area of the core that is used by placed Standard Cells and Macros expressed in percentage
 - Manufacturing Grid
 - The smallest geometry that semiconductor foundry can process or smallest resolution of your technology process (e.g. 0.005)
 - All drawn geometries during Physical Design must snap to this grid While Masking fab. use this as reference lines
 - Standard Cell Site/ Standard Cell Placement Tile/ Unit Tile
 - The minimum Width and Height a Cell that can occupy in the design
 - The Standard Cell Site will have the same height as Standard Cells, but the width will be as small as your smallest Filler Cell
 - It's one Vertical Routing Track and the Standard Cell
 - Height All Standard Cells must be multiple of Unit Tile
 - Standard Cell Rows
 - Rows are actually the Standard Cell Sites abut side by side and then Standard Cells are placed on these Rows
 - Cells with the equal no. of Track definition will have same height

Floorplanning

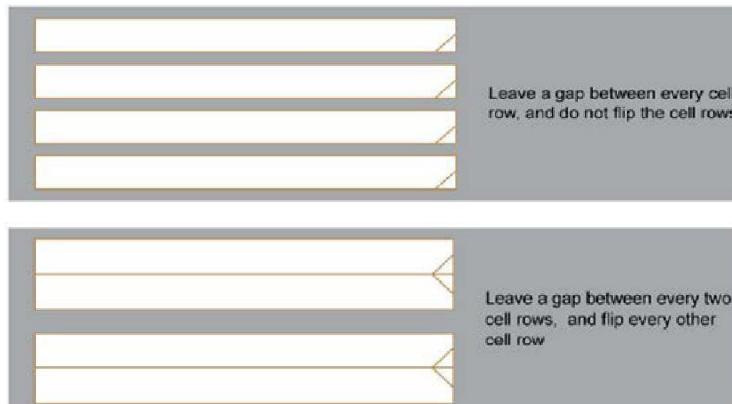
- Terminologies and Definitions
 - Placement Grid
 - Placement Grid is made up of Standard Cell Site
 - Its always a multiple of Manufacturing Grid
 - Placement Grid is made up of the Rows which are composed of Sites
 - Routing Grid and Routing Track
 - Horizontal and Vertical line drawn on the layout area which will guide for making interconnections
 - The Routing Grid is made up of the Routing Tracks
 - Routing Tracks can be Grid-based, Gridless based or Subgrid-based
 - Flight-line/ Fly-line
 - Virtual connection between Macros and Macro or Macros and IOs
 - Macro
 - Any instances other than Standard Cell and is as loaded as black box to the design is Macro
 - Intellectual Property (IP) e.g. RAM, ROM, PLL, Analog Designs etc.
 - Hard Macro: IP with Layout implemented
 - Soft Macro: IP without Layout implemented (HDL)

Floorplanning

- Steps in Floorplan
 - Initialize with Chip & Core Aspect Ratio (AR)
 - Initialize with Core Utilization
 - Initialize Row Configuration & Cell Orientation
 - Provide the Core to Pad/ IO spacing (Core to IO clearance)
 - Pins/ Pads Placement
 - Macro Placement by Fly-line Analysis
 - Macro Placement requirements are also need to consider
 - Blockage Management (Placement/ Routing)

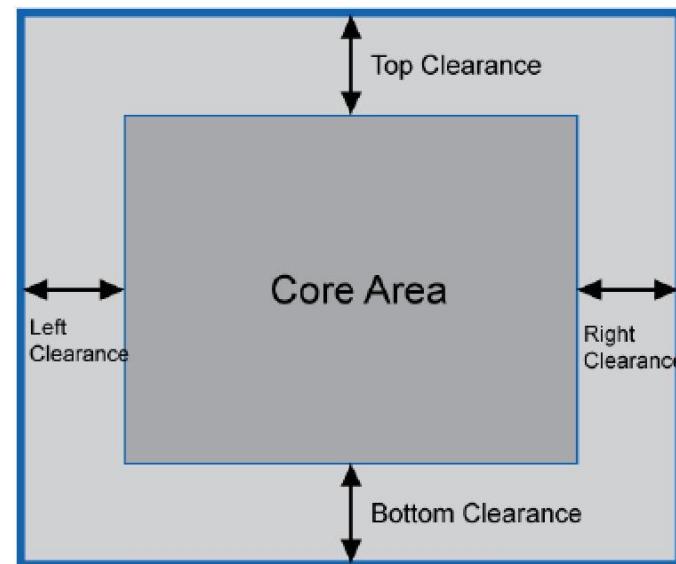
Floorplanning

- Initialization
 - Row Configuration
 - Slanting lines in the side of the cell rows denote the Cell Orientation



Most common because of better space utilization

- Core to Pad/ IO spacing
 - Core to IO clearance
 - Used for Placing IOs and Power Ring



Floorplanning

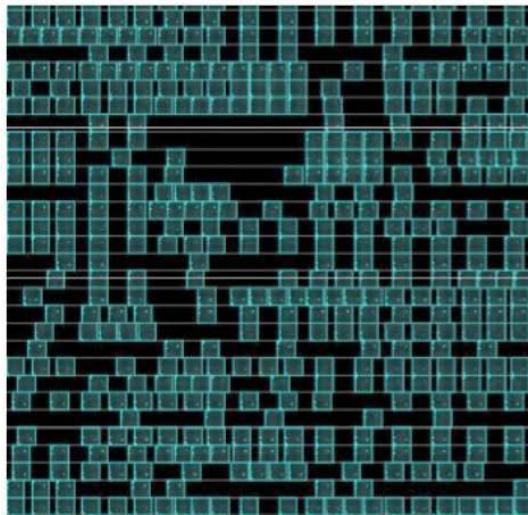
- Initialization

- Utilization = + x 100 %

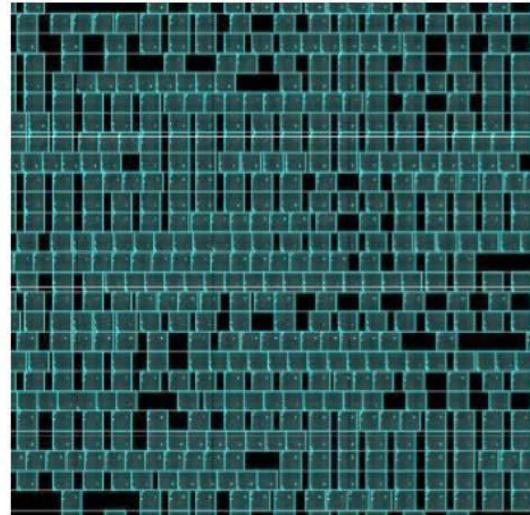
Aspect Ratio ==

or simply Height/Width

- Aspect Ratio decides the shape
- Full chip Aspect Ratio can have a maximum value of 1.25



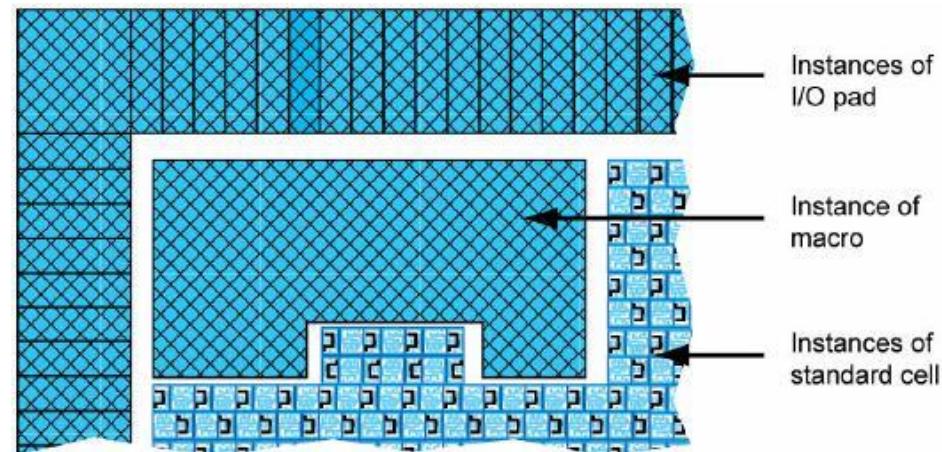
Low standard-cell utilization



High standard-cell utilization

Floorplanning

- IO Placement
 - Chip Level its IO Pads and Block Level its IO Pins
 - Pin is a logical entity and is a property of a Port
 - Port is a physical entity and a Port have only 1 Pin associated with it Netlist will have Pins and Layout will have Ports
 - Unplaced Port is not represented in the Layout
 - Different types of IOs
 - Signal Pads/Pins
 - Core Power Pads/Pins
 - IO Power Pads/Pins
 - Corner Pads (Doesn't hold any logic, provides IO Pad Ring connectivity)
 - Filler Pads (Fill the gaps between IO pads to get the Ring Connectivity)
 - Physical-only pads that are not part of the input Gate level Netlist need to be inserted prior to reading IO constraints

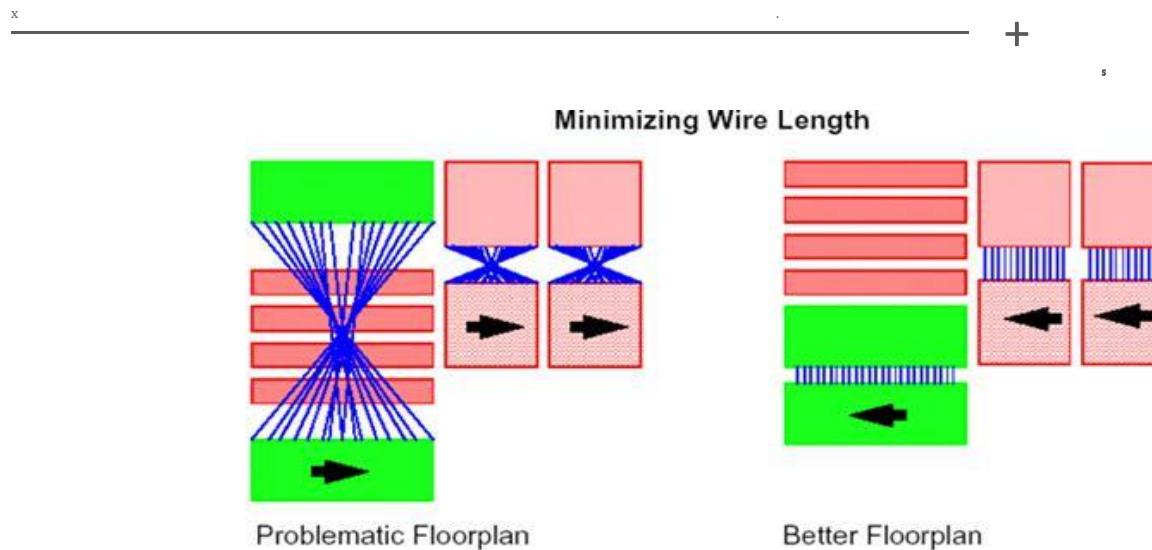


Floorplanning

- **IO Placement**
 - IO Pads enables the design to operate at different voltages with the help of Level Shifters, Pre-Drivers (at Core Voltage) Post-Drivers (at IO Voltage)
 - No of Core Power Pads needed: _____
 . x x
 - There will be 1 Core GND Pad along with every Core Power Pad
 - No. of IO Power Pads needed:

Floorplanning

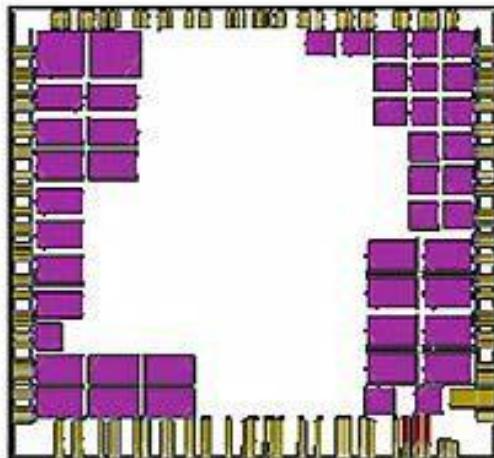
- Macro Placement
 - Fly-line Analysis (For Connectivity information)
 - Macro keep-out (For Uniform Standard Cell Region)
 - Channel Calculation (Critical for Congestion and Timing)
 - Avoid odd shaped area for Standard Cells
 - Funnel shaped Macro Placements are preferred
 - Fix the Macro locations, so that tool wont alter during Optimization
 - Spacing between Macro:



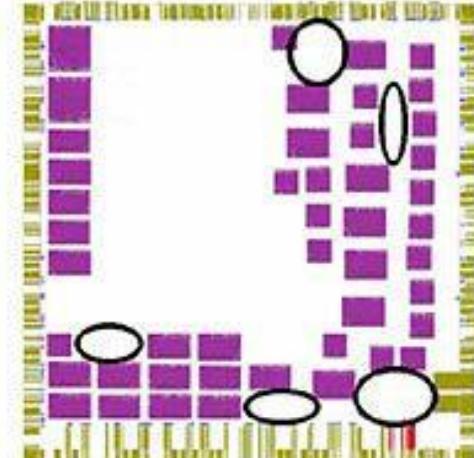
Floorplanning

- Macro Placement Tips
 - Place macros around chip periphery, so that core area will be clustered
 - Consider connections to fixed cells when placing Macros
 - In advanced Technology Nodes Macro Orientation is fixed since the Poly Orientation can't vary, so there will be restrictions in Macro Orientation
 - Reserve enough room around Macros for IO
 - Routing Reduce open fields as much as possible
 - Provide necessary Blockages around the Macro

Homogeneous Standard Cell Area With Aligned Macros

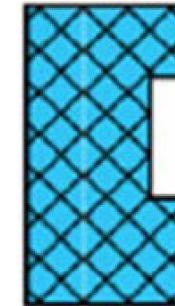


Irregular Macro Placement With Traps for Standard Cells



Floorplanning

- Blockages
 - Placement Blockage & Routing Blockage
 - Both of the Blockages can again be classified as-
 - Hard, Soft and Partial Blockages
 - Hard Blockage
 - Complete Standard Cell Blockage
 - Soft Blockage
 - Non-Buffering Blockage
 - Partial Blockage
 - Partial Standard Cell Blockage and is used to avoid congestion
 - We can Block Standard Cells as per the required percentage value
 - Keep-out/ Halo
 - Halo is similar to Soft Blockage (Terminology in Cadence EDI)
 - Its basically a keep-out Macro margin
 - Halo respects Macro while other Blockages respect location i.e., even if Macro is moved Halo also moves along with it



Rectilinear
Macro
Without
Blockage



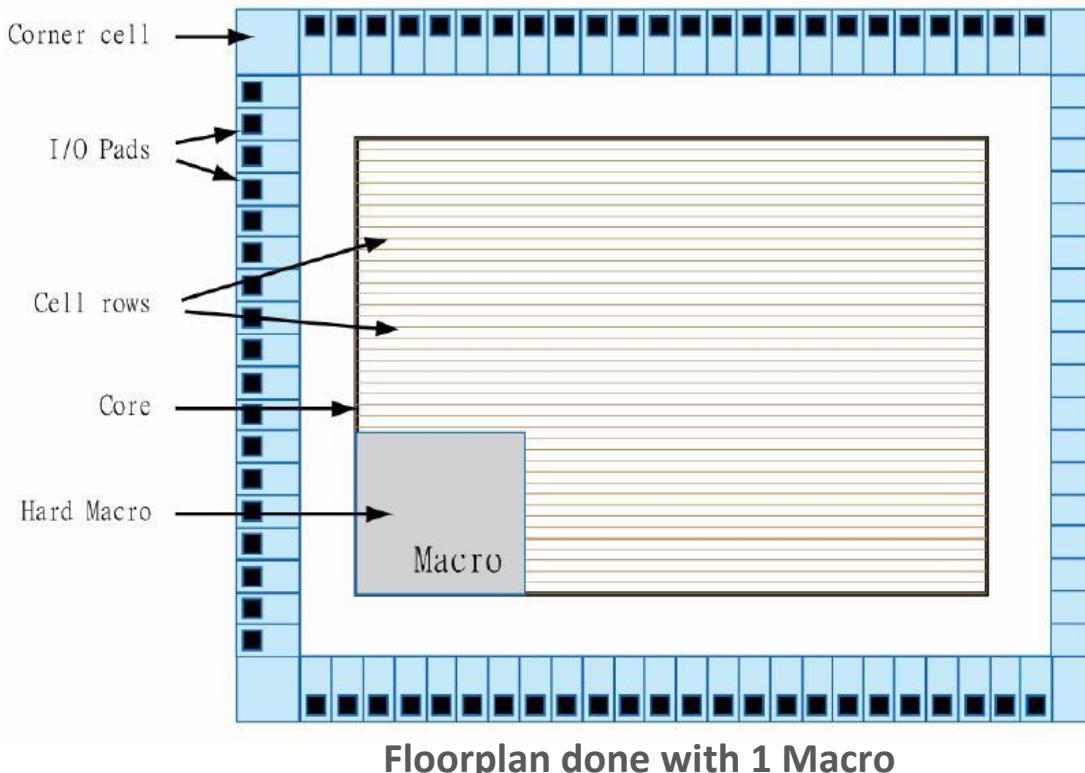
With
Blockage



Halo around Macro

Floorplanning

- Issues arises due to bad Floorplan
 - Congestion near Macro Pins/ Corners due to insufficient Placement Blockage
 - Std. Cell placement in narrow channels led to Congestion
 - Macros of same partition which are placed far apart can cause Timing Violation



Power planning

- Power Plan
 - To connect Power to the Chip by considering issues like EM and IR Drop
 - Power Routing also called Pre-Routing
 - Pre-Routing includes creating Power Ring, Stripes/Mesh/Grid, and Standard Cell Power Rails
 - Power Planning also includes Power Via insertion
 - IO Rings are established through IO Cell abutment and through IO Filler Cells
 - Power Trunks are constructed between Core Power Ring and Power Pads
 - Trunk is a piece of metal that connects IO Pad and Core Ring
 - Technical information required for Power Planning:
 - Total Dynamic Power info. will get from Compiler
 - Technology File will provide Current Density (J_{MAX})
 - LEF will prove the Metal Layer width
 - Technology Library will provide Core Voltage

Power planning

- Levels of Power Distribution
 - Rings
 - V_{DD} and V_{SS} Rings are formed around the Core and Macro
 - Stripes
 - Carries V_{DD} and V_{SS} around the chip
 - Carries V_{DD} and V_{SS} from Rings across the chip
 - Power Stripes are created in the Core Area to tap power from Core Rings to the core area
 - Rails (Special Route)
 - Connect V_{DD} and V_{SS} to the standard cell
 - Standard Cell Rails are created to tap power from Power Stripes to Std. Cell Power/Ground Pins
 - Power Vias
 - Insert all Power Vias between Ring & Grid, Grid & Rail and Vertical Grid & Horizontal Grid
 - Trunks
 - Connects Ring to Power Pad

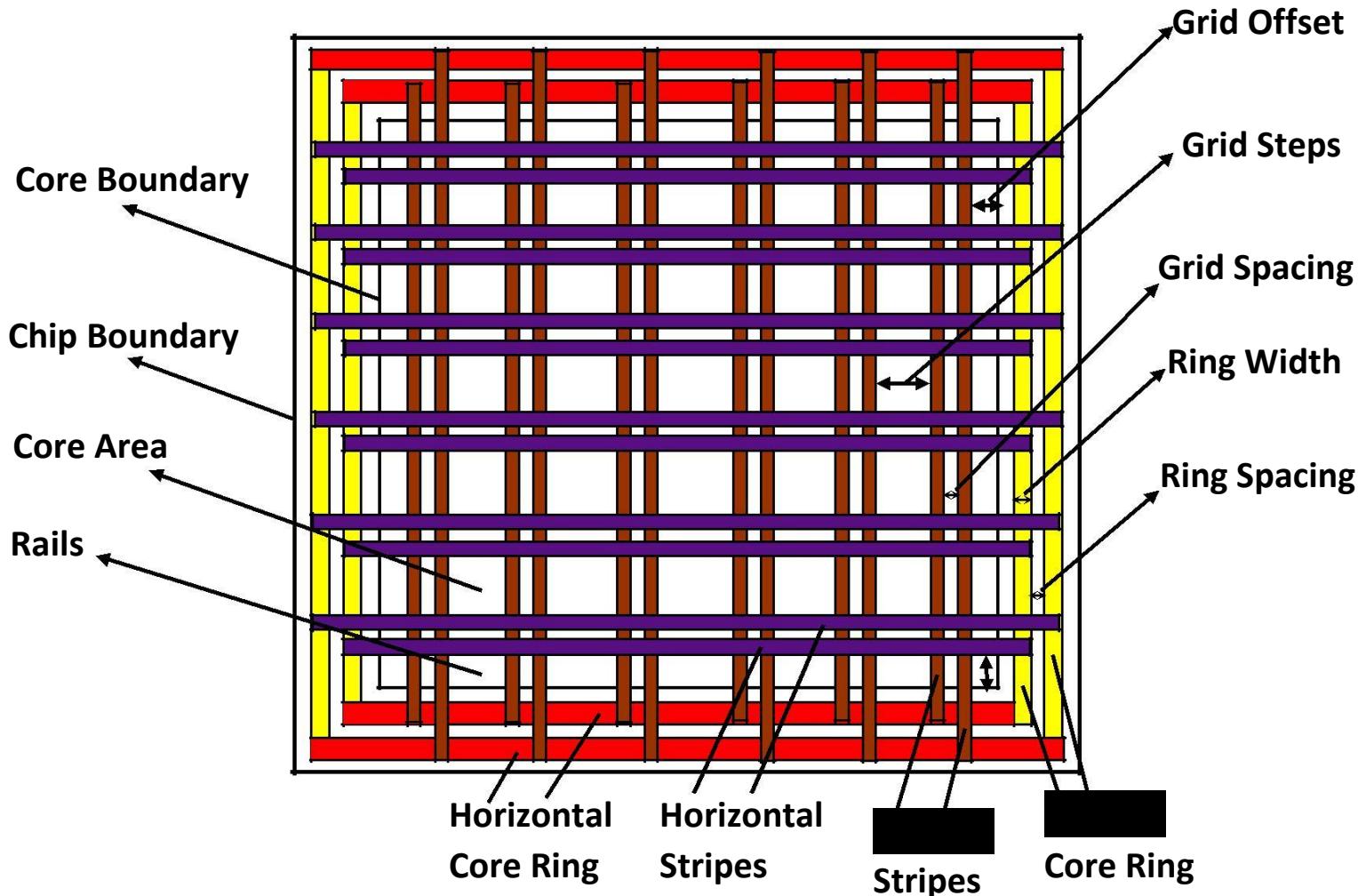
Power planning

- Power Plan: Calculations
 - Total Dynamic Core Current = _____

- Pad to Core Trunk Width = _____
- Core Ring Width = _____
- Power Stripes Spacing =
_____ - (_____) _____ + _____

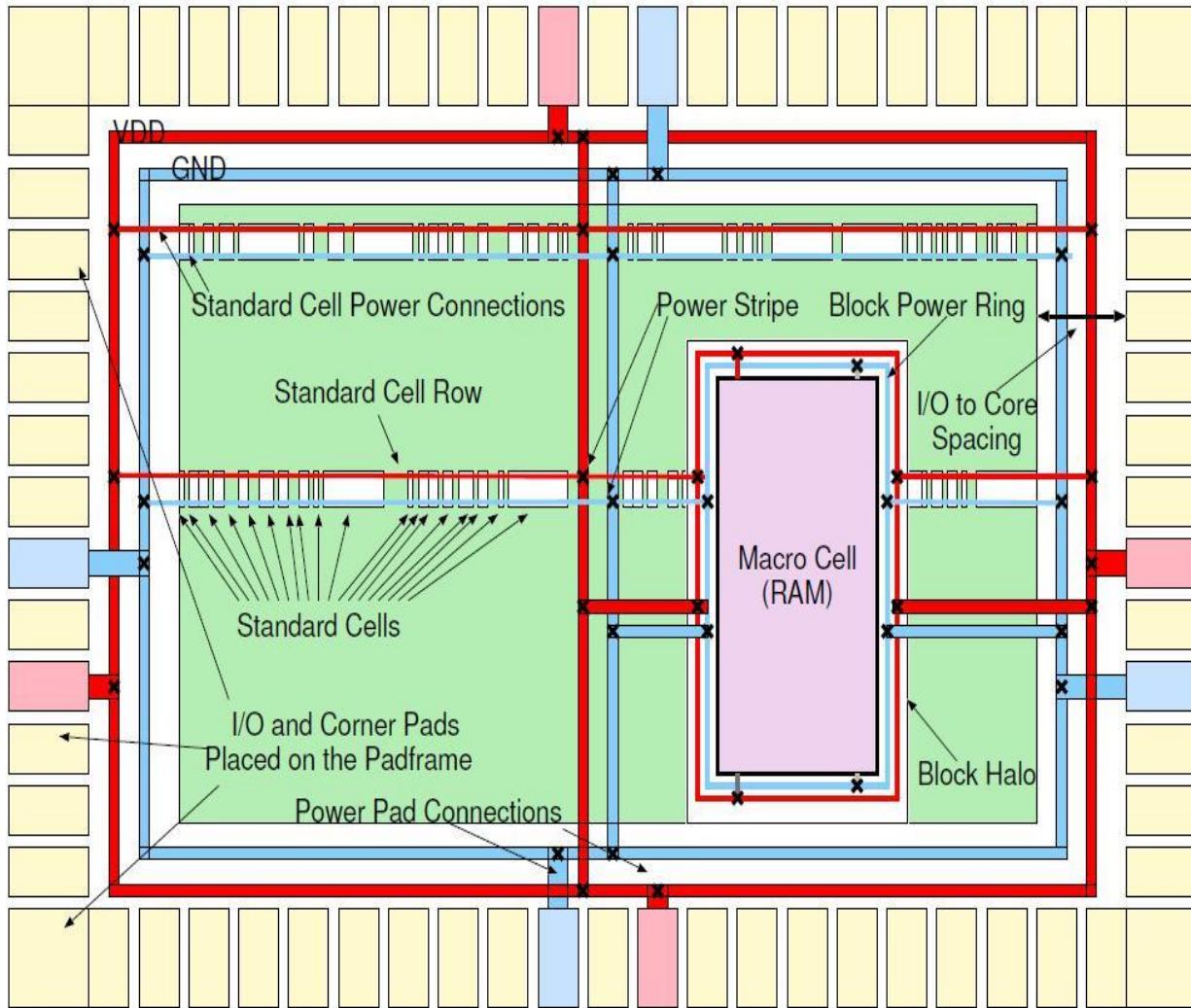
Power planning

- Sub-block Configuration



Power planning

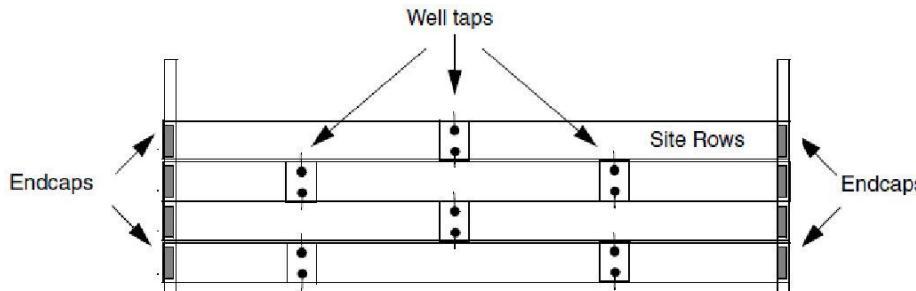
- Full Chip Configuration



- Save Floorplan (.def / .fp)

Pre-Placement

- Physical-Only Cells (Well Taps, End Caps)
 - These library cells do not have signal connectivity and connect only to the power and ground rails
 - End Caps ensure that gaps do not occur between the Well and Implant Layers and also prevents DRC violations by satisfying Well tie-off requirements for core rows
 - Well Taps help to tie Substrate and N-wells to V_{DD} and V_{SS} levels and thus prevent Latch-up
- Special Cells (Spare cells, Decap Cells)
 - Spare Cells for ECO and Decaps for avoiding Instantaneous Voltage Drop (IVD)
 - Place Decaps closer to Power Pads or any larger Drivers



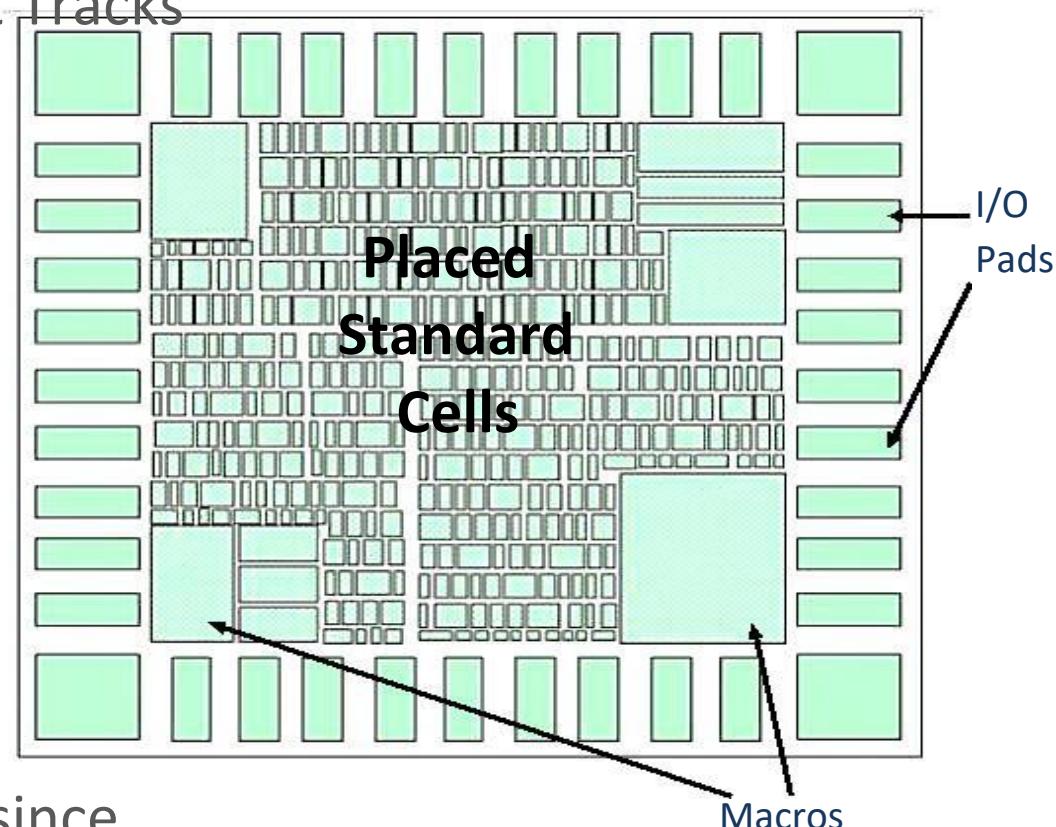
- Cell Padding
 - Cell Padding is done to reserve space for avoiding Routing Congestion
 - Cell Padding adds Hard Constraints to Placement
 - The Constraints are honored by Cell, Legalization, CTS, and Timing Optimization

Pre-Placement Optimization

- Pre-Placement Optimization Goals
 - Routability
 - Performance (Timing)
 - Power (with Cells)
 - Optimizations before Placement
 - Delay models must be removed (if any)
 - Zero-RC (0-RC) Optimization
 - Isolation Cell Insertion
 - Multi Corner Multi Mode (MCMM) settings before Std. Cell Placement
 - Zero-RC Optimization
 - Optimizes the netlist without any delay models, thus provides an optimal starting point for placement
 - Timing during 0-RC Opt and that of during Synthesis has to be matched
 - Else indicate problems in the Technology File, Timing Library, Constraint Files, or overall design
 - Logical restructuring and up/down size are optimizations at the 0-RC stage
- Take care of don't use cells while doing optimization

Placement

- Automated Standard Cell Placement for placing the Standard Cells in Placement Tracks
- Placement Objectives
 - Total wire length
 - Routability
 - Performance
 - Power
 - Heat distribution
- Timing checks only with slow corners at Placement stage
- Only Setup Time check, since buffers are getting added during Clock Tree Synthesis

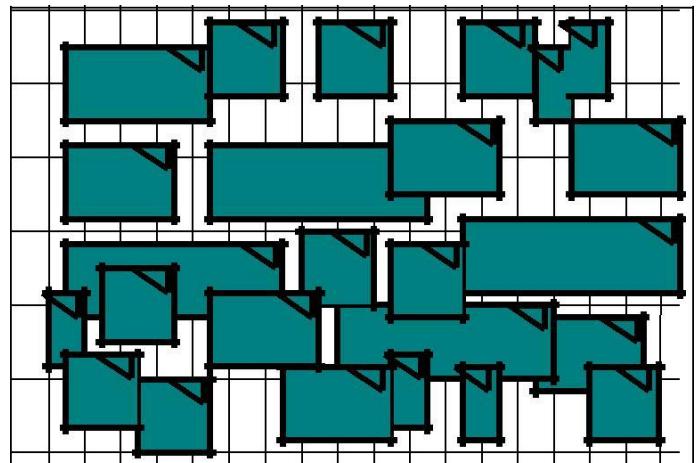


Placement

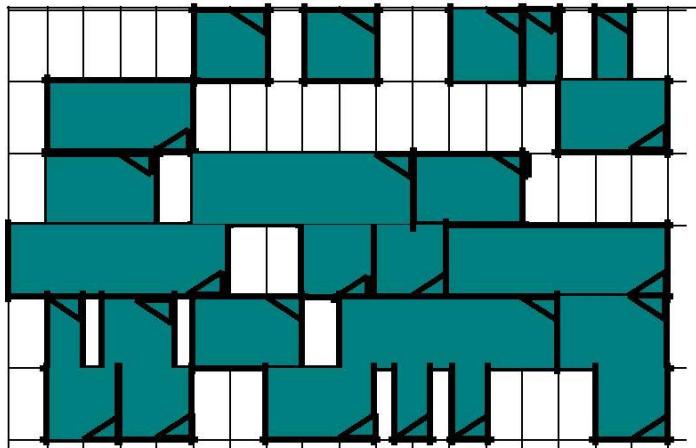
- Placement Methods
 - Timing Driven Placement
 - To Refine placement based on congestion, timing and power
 - To optimize large sets of path delays
 - Net Based
 - Congestion Driven Placement
 - To distance standard cell instances from each other such that more routing tracks are created between them
- Control the delay on signal path by imposing an upper bound delay or weight to net

Placement

- Placement Stages
 - Global Placement
 - Detail Placement
 - Placement Legalization
 - In-Place Optimizations
- Global/ Coarse Placement
 - To get the approximate initial location
 - Cells are not legally placed and there can be overlapping
- Detail/ Legal Placement
 - To avoid cell overlapping
 - Cells have legalized locations
 - Legalize placement will place the cells in their legal position with no overlap



Global/ Coarse Placement



Detail/ Legal Placement

Placement

- Placement Legalization
 - Placed Macros are legally oriented with Standard Cell Rows
- In-Place Optimizations
 - Scan Chain Reordering
- After Placement, report Congestion, Utilization and Timing
- Tie off cell instances provide connectivity between the Tie-high and Tie-low logical inputs pins of the Netlist instances to Power and Ground
- Tie off cells are placed after the placement of Standard Cells
- After placement check the Cell Density
- Global Route (GR)
 - Whole region is divided into an array of rectangular sub-regions each of which may accommodate tens of routing tracks in each dimension called Global Cells
 - Global Route is performed to estimate the inter-connect parasitics and Routing Congestion Map

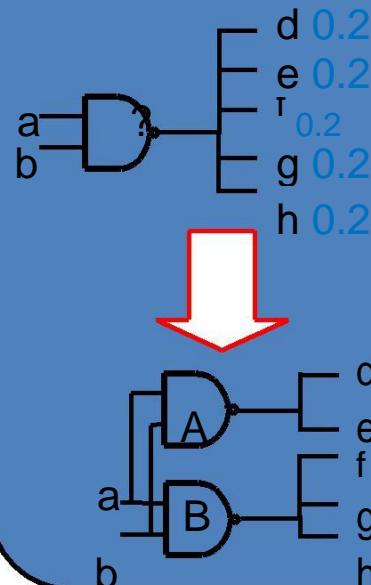
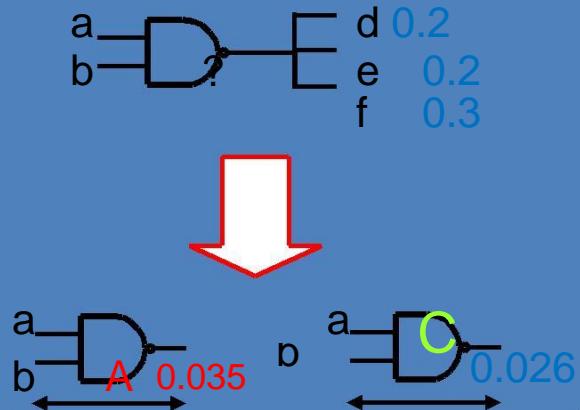
Pre CTS Optimization/ Placement Opt.

- Cell Sizing
 - Sized up/ down to meet optimizing for timing and area
 - Up sizing will give timing advantage and Down sizing will give area advantage
- V_T Swapping
 - To optimize for leakage power (HVT, RVT/SVT, LVT)
- Cloning
 - To reduce fanout
- Buffering
 - Long nets are buffered or remove buffers to bring the timing advantage
- Re-Buffering
 - To improve slews, reduce net capacitance and reduce fanout
- Logical Restructuring
 - To optimize timing and area without changing the functionality of the design
 - Breaking complex cells into simpler cells or vice versa
- Pin Swapping

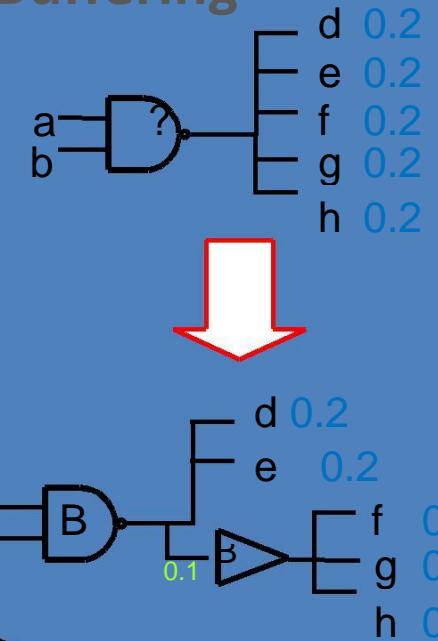
Pre CTS Optimization/ Placement Opt.

- Optimization Techniques

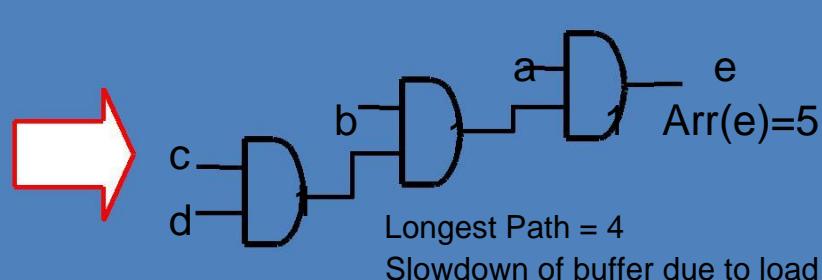
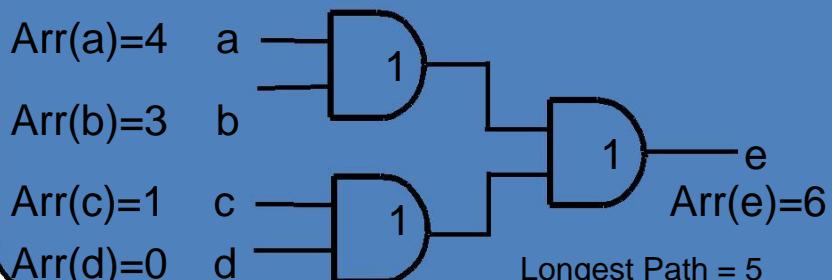
- Resizing- Cloning



- Buffering



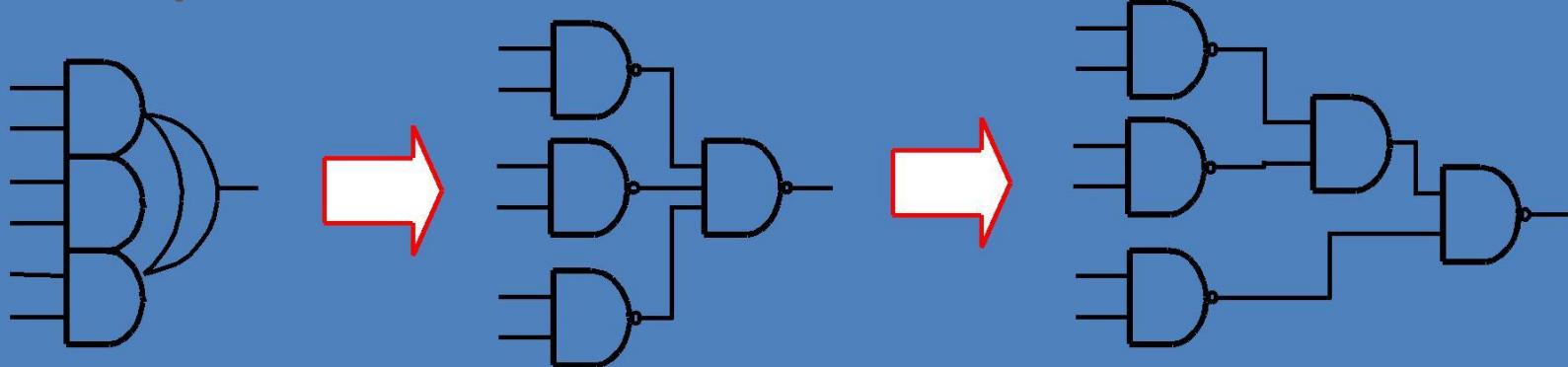
- Redesign Fan-in Tree



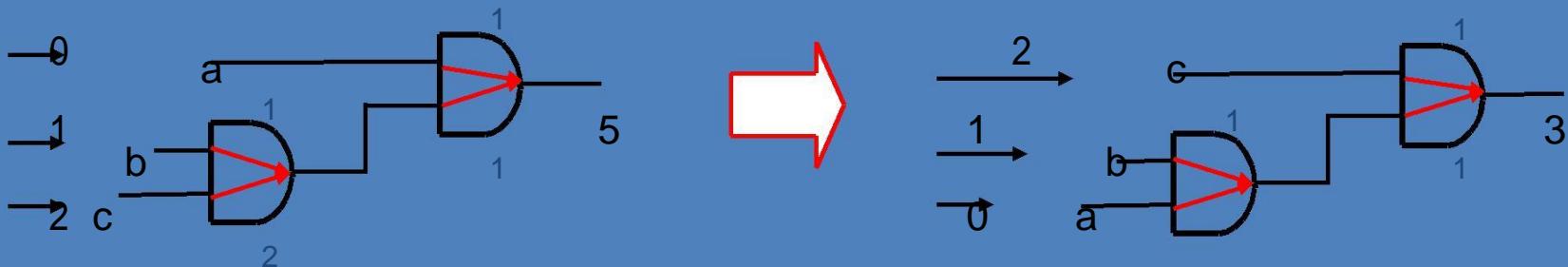
Pre CTS Optimization/ Placement Opt.

- Timing Optimization Techniques

- Decomposition



- Swap Commutative Pins

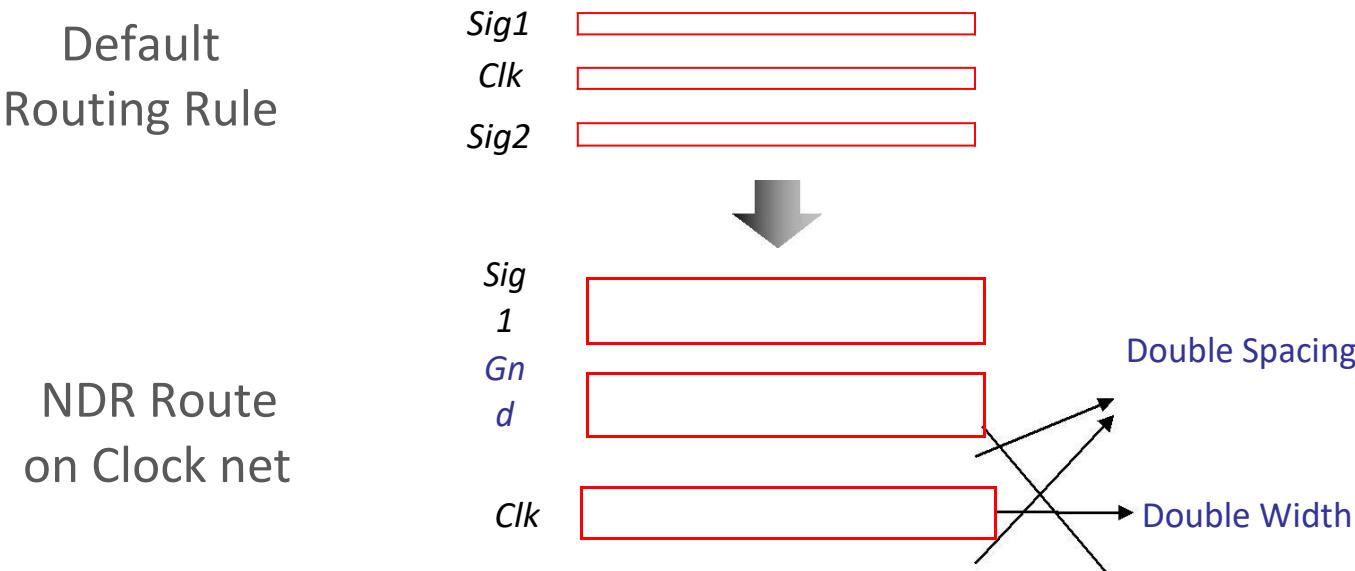


Pre CTS Optimization

- Set the Optimization Directives
 - don't_use, size_only
- Perform High Fanout Nets Synthesize (HFNS)
 - High Fanout Nets are Synthesized before Clock Tree Synthesis
 - HFNS is the Buffering of High Fanout Nets
 - Usually High Fanout Nets may have Fanout of more than 1000
 - Eg., Reset, Clear etc.
- Set CTS Routing Rules
 - Shielding
 - Non Default Rules (NDR)
- Set RC Delay Models

Pre CTS Optimization

- Non-Default Rule (NDR)
 - The user-defined Routing rules apart from the default Routing Rule
 - Often used to “harden” the sensitive nets like Clock Nets
 - NDRs make the Clock Routes less sensitive to CrossTalk or EM effects
 - Double/ Triple Width for avoiding Electromigration
 - Double/ Triple Spacing for avoiding Crosstalk
 - NDRs will improve Insertion Delay



Gnd

► G
r
o
u
n
d

S
h
i
e
l
d
i
n
g

Sig2

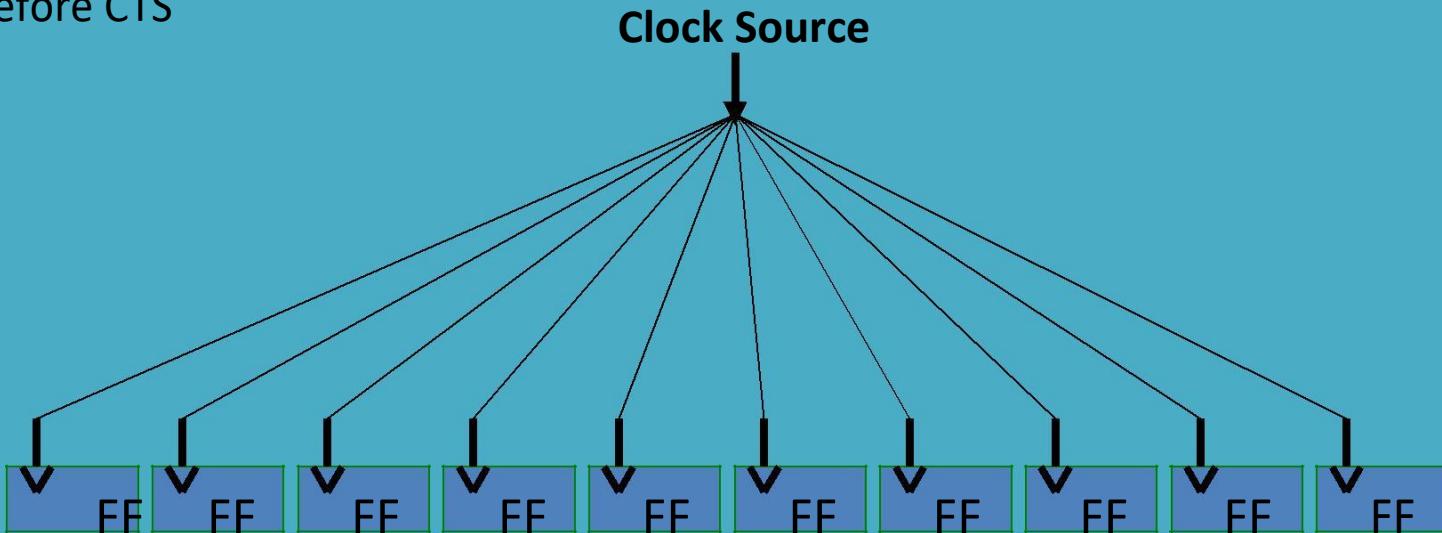


Clock Tree Synthesis (CTS)

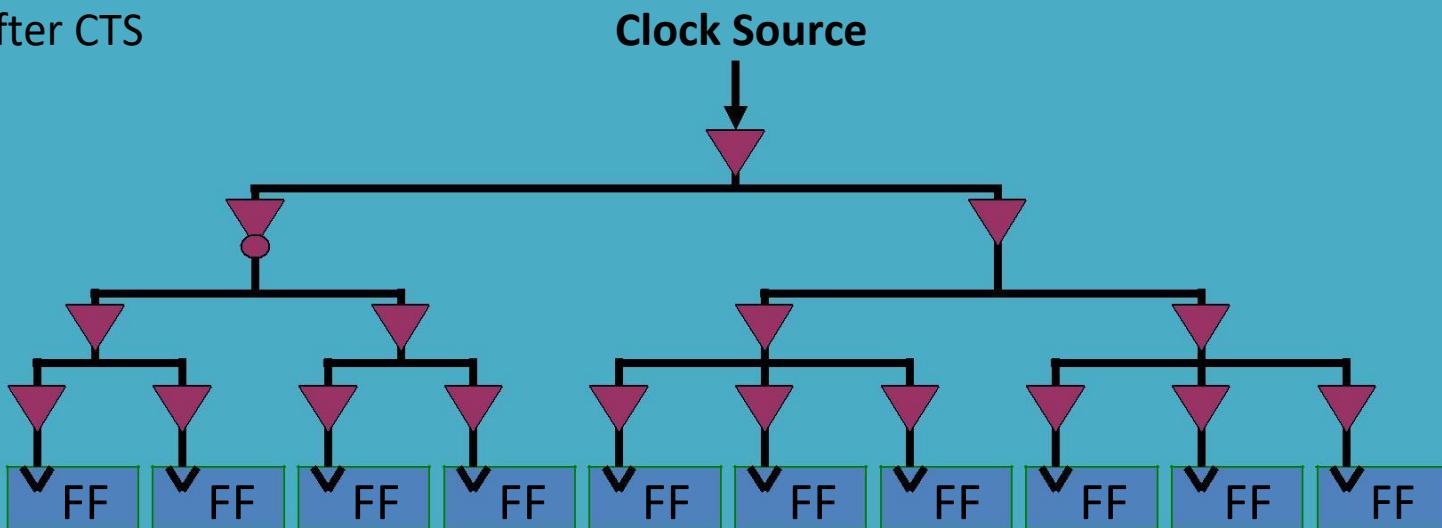
- The Clock Problem
 - Clock skew
 - Long clock insertion delay
 - Skew across clocks
 - Heavy clock net loading
 - Clock is power hungry
 - Clock to signal coupling effect (CrossTalk)
 - Electromigration on clock net
- Clock Tree is a path from the Clock Source (Root) to Clock Sinks (Leaf)
- Clock Tree Synthesis is the process of creating this Clock Path from Clock Source to Clock Sinks
- All Clock pins of flip Flop are considered as Clock Sinks (Leaf); where the Clock Tree Synthesis ends

Clock Tree Synthesis (CTS)

Before CTS

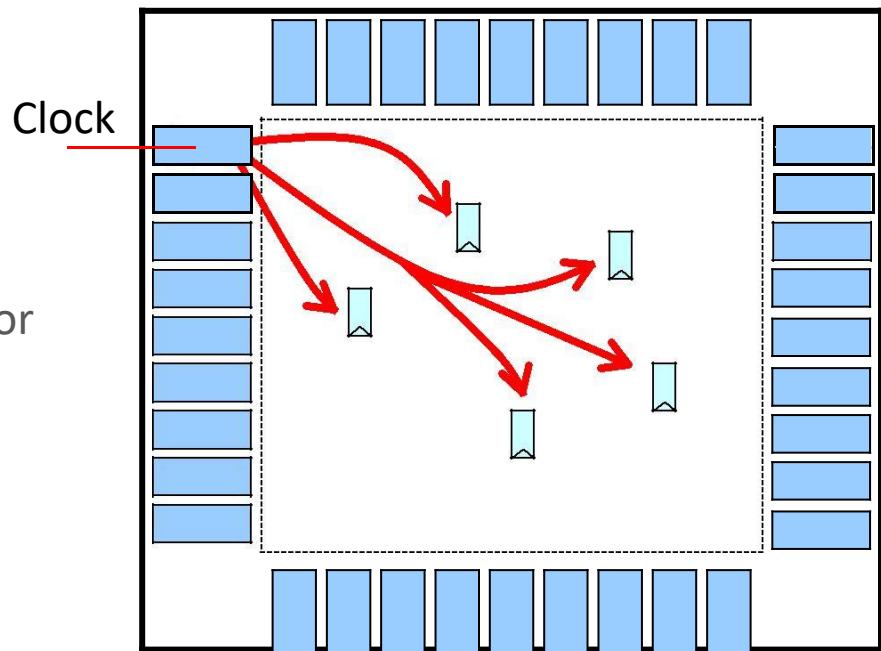


After CTS



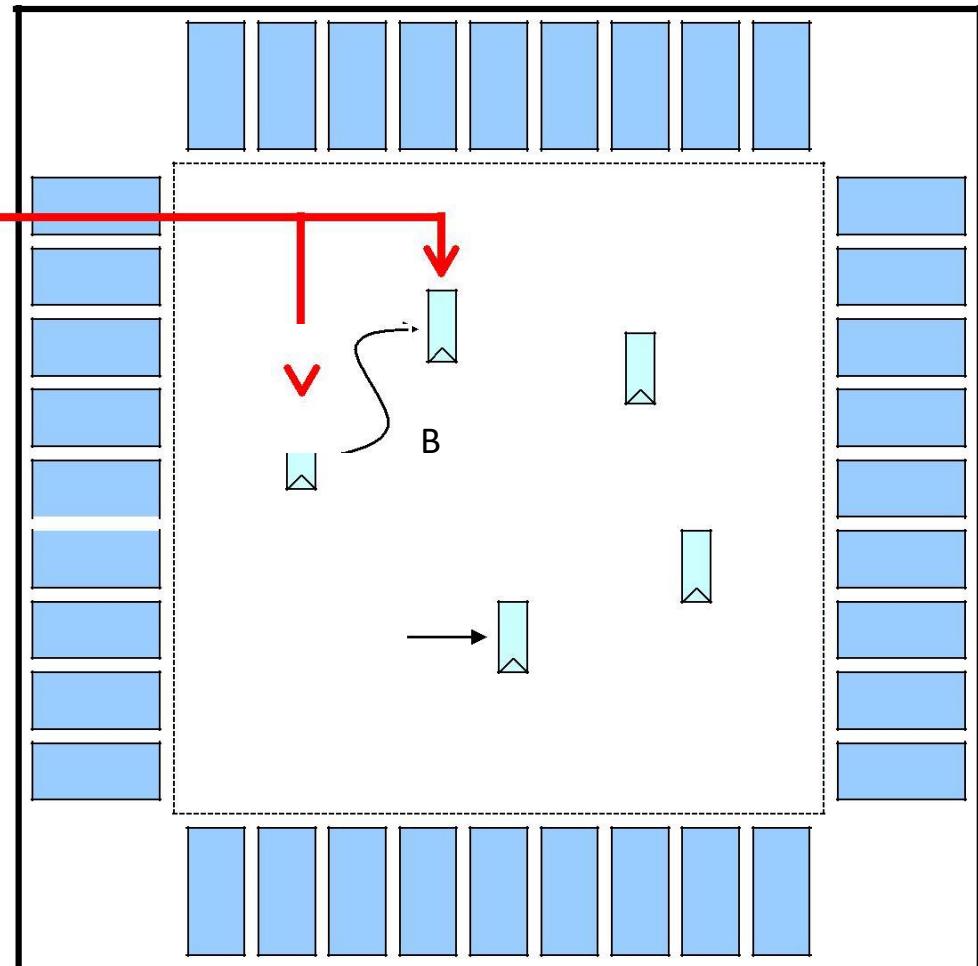
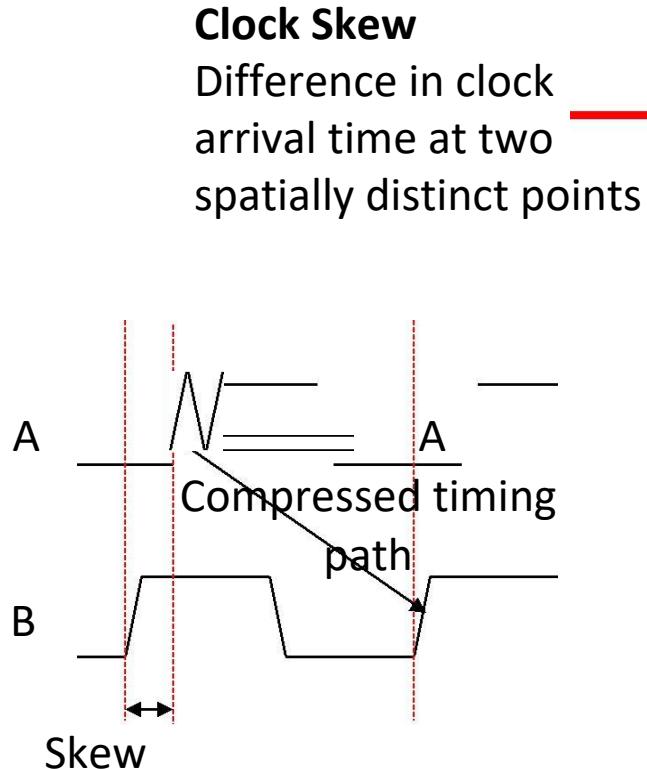
Clock Tree Synthesis (CTS)

- Main concerns for Clock Design
 - Skew
 - Most important concern for clock networks
 - For increased clock frequency, skew may contribute over 10% of the system cycle time
 - Due to variations in trace length, metal width and height, coupling caps
 - It can also be due to variations in local clock load, local power supply, local gate length and threshold, local temperature
 - Power
 - Very important, as clock is a major power consumer
 - It switches at every clock cycle
 - Noise
 - Clock is often a very strong aggressor
 - May need shielding
 - Delay
 - Not really important
 - But Slew Rate is important (sharp transition)



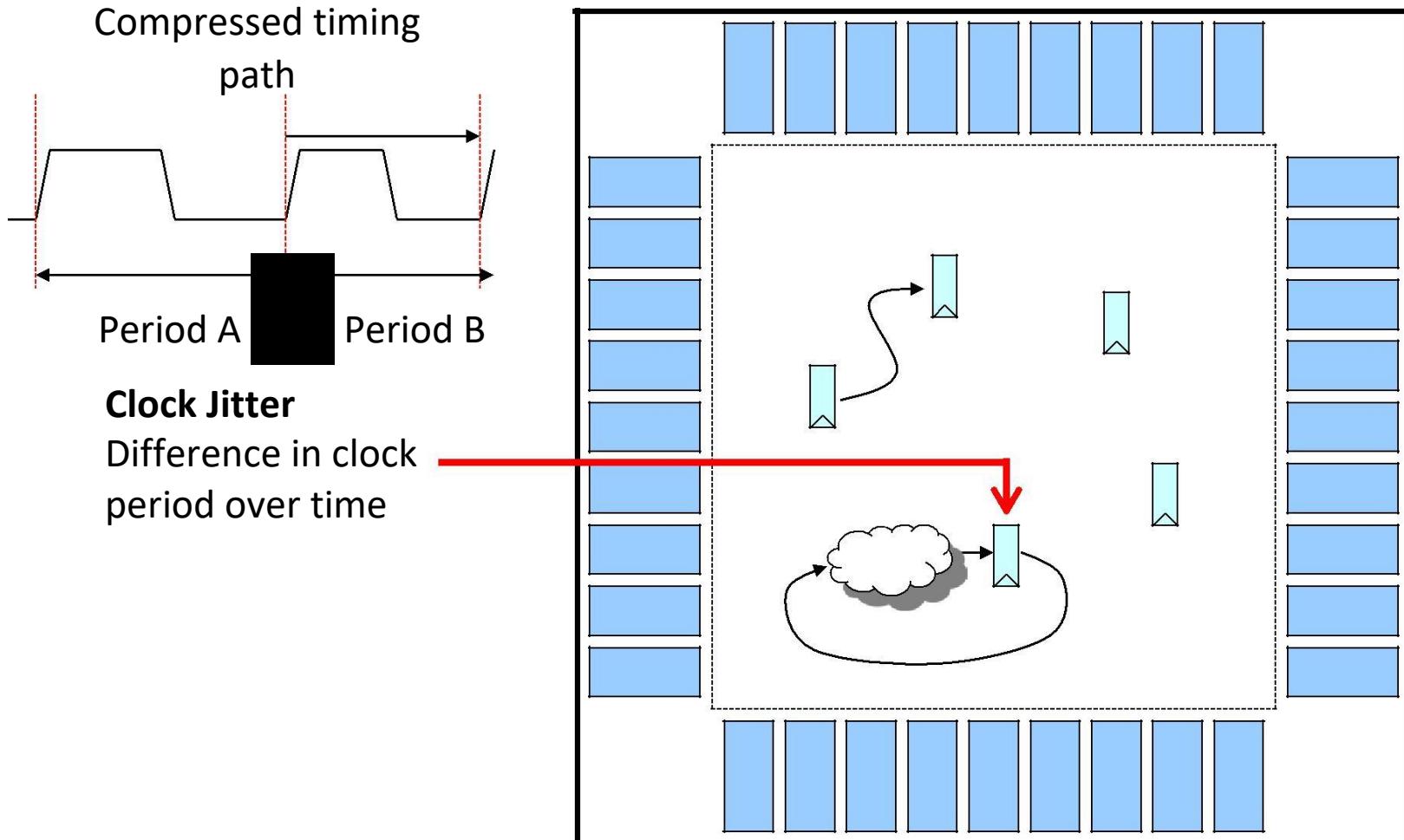
Clock Tree Synthesis (CTS)

- Clock Skew: Spatial Clock Variation



Clock Tree Synthesis (CTS)

- Clock Jitter: Temporal Clock Variation



Clock Tree Synthesis (CTS)

- CTS Pre-requisites
 - Legally Placed and Optimized with acceptable Congestion
 - Timing should be good
 - No Design Rule Violations
 - Power/Ground nets are pre-routed
 - HFNS done
 - Logical/Physical Library should have special Clock Cells
- CTS Objects
 - The timer starts from every Clock Source and traces forward over Combinational Arcs until it reaches the Clock Pin of a flop or another Clock Source
 - All Pins/ Timing Arcs in the forward trace before a valid Leaf are considered to be in the clock network
 - Pin or Combinational Timing Arcs that trace to a non-clock pin are not part of Clock Tree network (e.g. D pin of FF)
 - Sequential elements are traced through if it is a source of the Generated Clock
 - Clock tracing after the propagation of Case Analysis
 - Clock tracing should be Mode aware
 - Inverters are added in Clock Tree for better Duty Cycle
 - Limit the buffer/inverter list to just 3 or 4 buf/inv sizes

Clock Tree Synthesis (CTS)

- CTS Flow
 - Check and fix Macro locations
 - Read CTS SDC: Clock Tree begins at SDC defined clock pin and ends at stop pin of the flop
 - Generate CTS Specification file
 - Max. Skew
 - Max. and Min. Insertion Delay
 - Max. Transition, Capacitance, Fanout
 - No. Buffer levels (Tree depth)
 - Buffer/ Inverter list
 - Clock Tree Routing Metal Layers
 - Clock Tree Leaf Pin, Root Pin, Preserve Pin, Through Pin and Exclude Pin
 - Compile CTS using CTS Spec. file
 - Place Clock Tree Cells
 - Route Clock Tree (Optional and can be done during Signal net routing also)

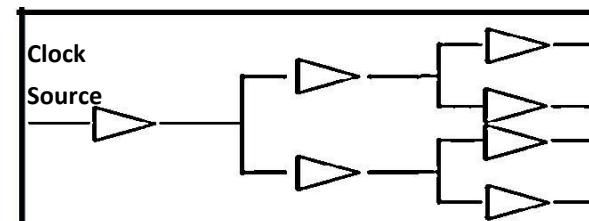
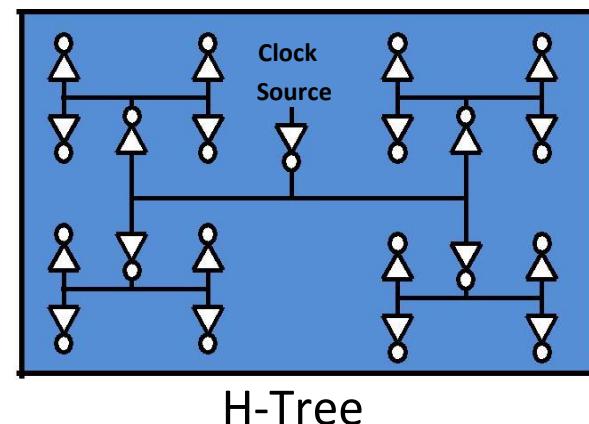
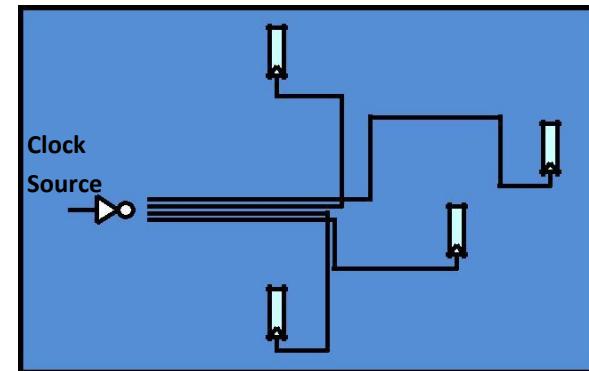
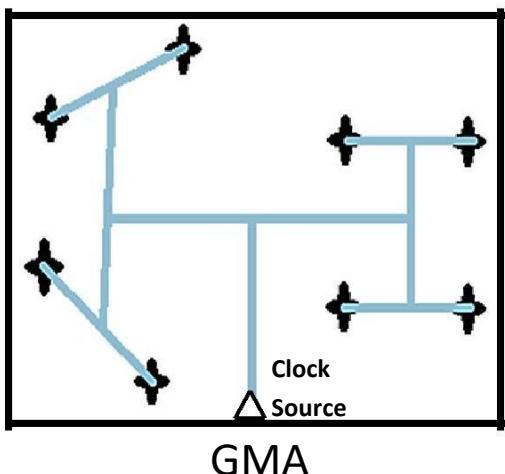
Example of CTS spec file

```
AutoCTSRootPin SH1/I23/Z
ExcludePin + XPU/CAM/C
MaxDelay 5ns MinDelay 0ns

Buffer buf1 buf2 inv1 inv2 dell1
MaxSkew 500ps
MaxDepth 20
LeafPin + FPU/CORE/A rising
END
```

Clock Tree Synthesis (CTS)

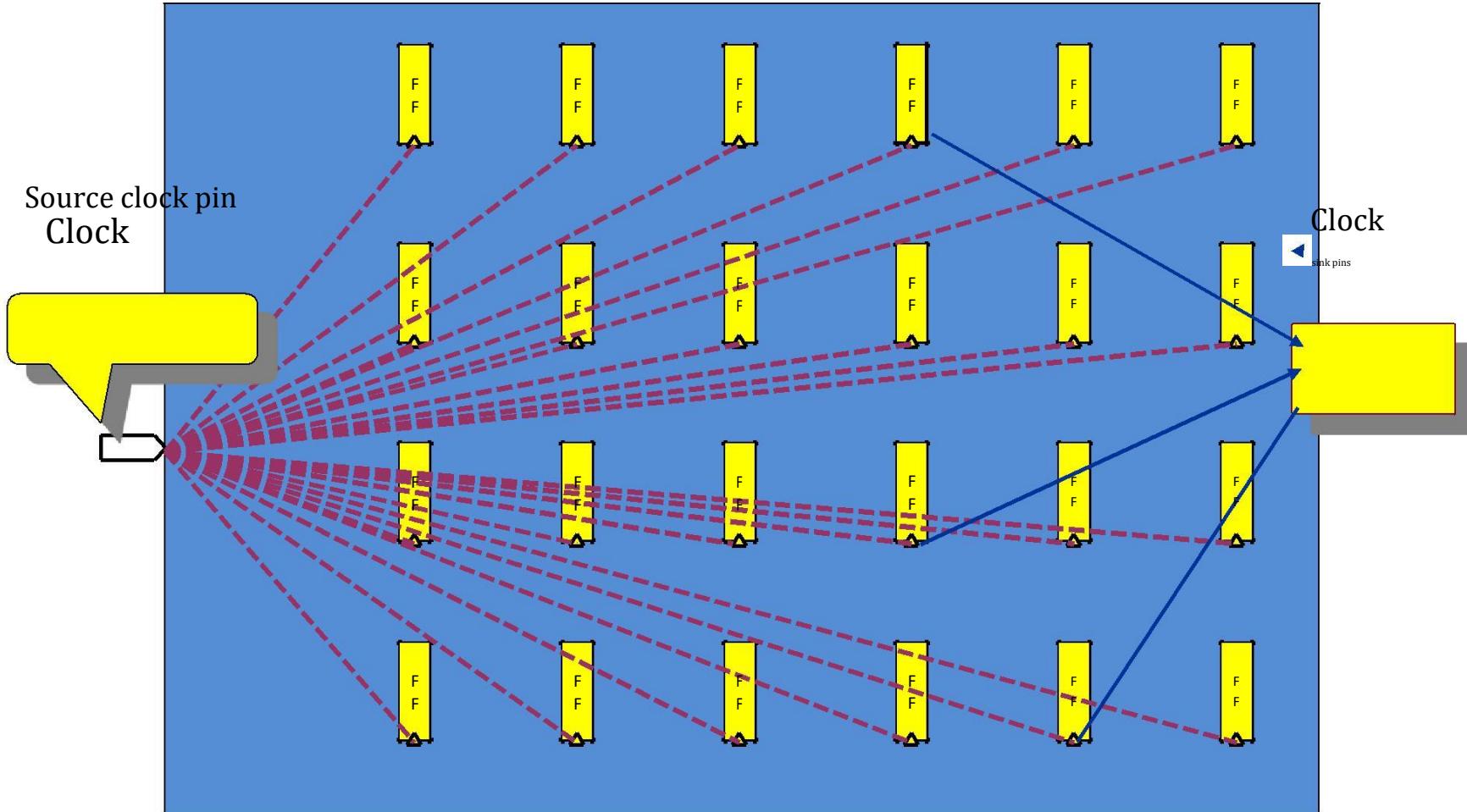
- CTS Algorithms
 - RC Tree Based CTS
 - H Tree based Algorithm
 - X Tree based Algorithm
 - Method of Mean and Median (MMM)
 - Geometric Matching Algorithm (GMA)
 - Pi Configuration



Pi Configuration

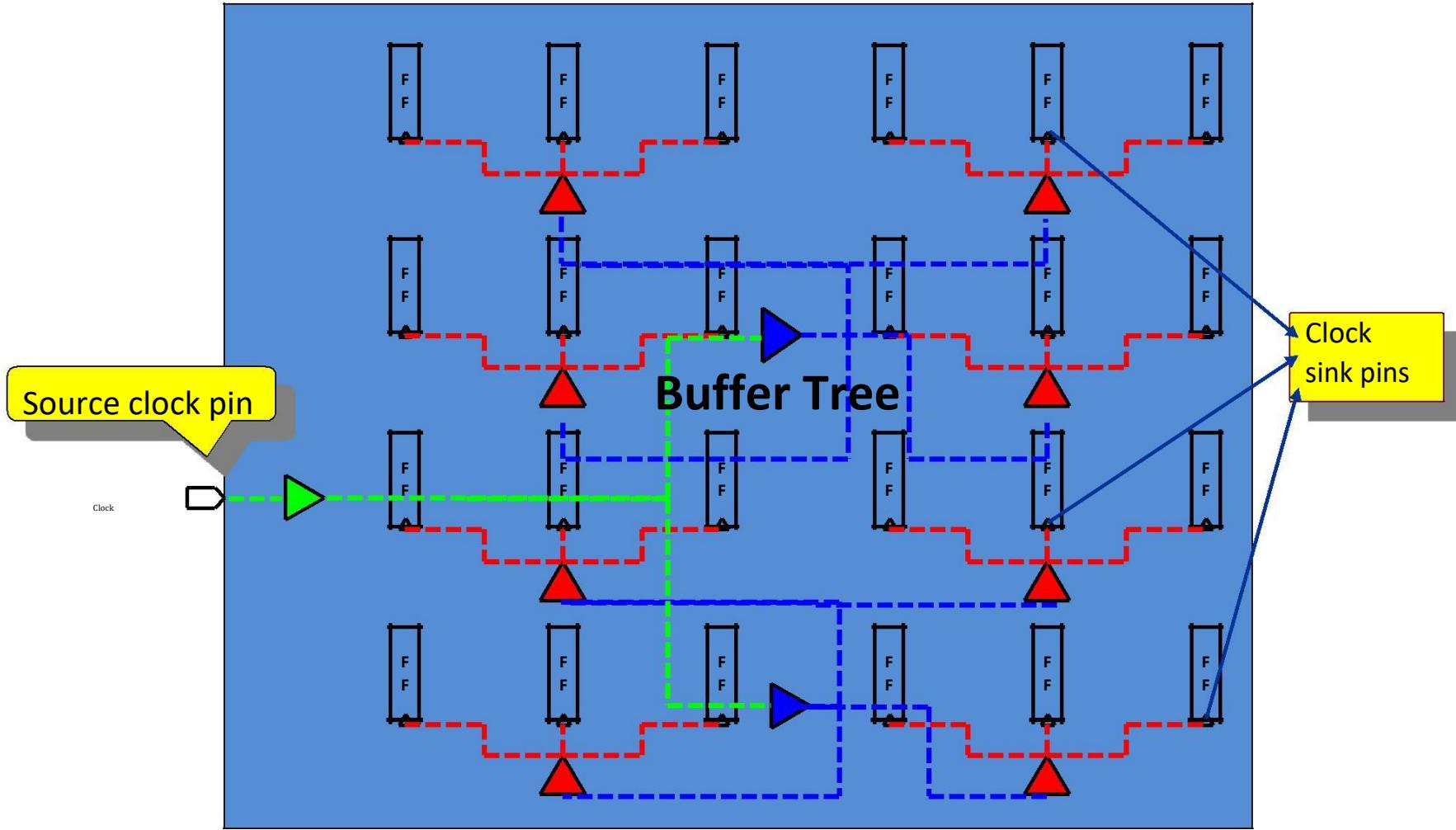
Clock Tree Synthesis (CTS)

- Before CTS all Clock Pins are driven by a single Clock Source



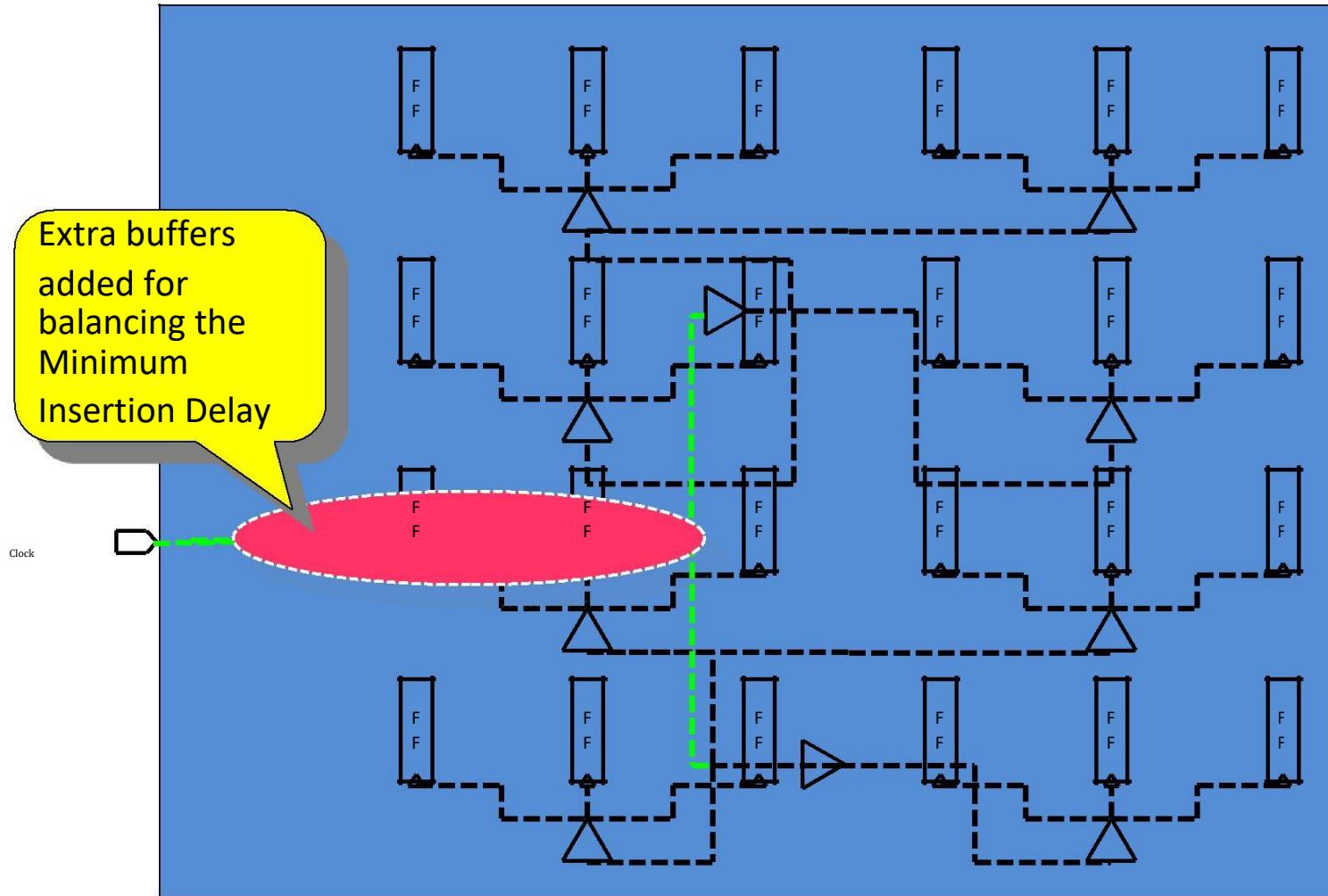
Clock Tree Synthesis (CTS)

- After CTS the buffer tree is built to balance the loads and minimize the skew



Clock Tree Synthesis (CTS)

- After CTS a “delay line” is added to meet the minimum Insertion Delay (ID)



Clock Tree Synthesis (CTS)

- Analyze the Clock Tree
 - Report Timing (both Setup and Hold)
 - If timing not met then check clocks be grouped (balanced together)
 - Report Insertion Delay & Skew and verify that the targets are achieved
 - Report DRV targets (Fanout, Capacitance and Transition)
 - Check the intended Leaf Cell (Clock Sinks) is reached
 - Check the Clock Tree Exceptions are not in the Clock Tree
 - Report the pre-existing cells, such as Clock Gating Cells
 - Do Quality-of-Report (QoR)
 - Check Clock Tree converges either with itself or with another Clock Tree
 - Clock Tree has timing relationship with other Clock Trees for inter Clock Skew balancing
 - Check Design Rule Constraints
 - Check Routing Constraints
 - Report Power and Area

Post CTS Optimization

- Post CTS Optimization
 - Optimization with Useful Skew
 - Optimization with Total Negative Slack (TNS)
 - Fine Grid Spacing
 - Post CTS Optimization Techniques
 - Shielding
 - Sizing
 - Buffer re-location
 - Level adjustment
 - Optimize the design for Hold Time
 - Hold Violations should be fixed first in Best Corner and then in Worst Corner
 - Area Optimizations

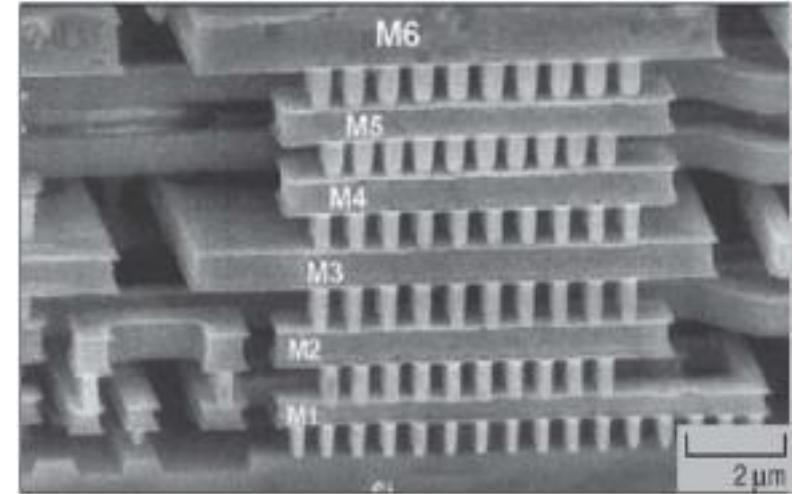
Routing

- Importance of Routing as Technology shrinks

- Device (Gate) delay decreases
 - Interconnect resistance increases
 - Vertical heights of interconnect layers increase, in an attempt to offset increasing interconnect resistance
 - Area component of interconnect capacitance no longer dominates
 - Lateral (sidewall) and fringing components of capacitance start to dominate the total capacitance of the interconnect
 - Interconnect capacitance dominates total Gate loading

- Routing Objectives

- Skew requirements
 - Open/Short circuit clean
 - Routed paths must meet setup and hold timing margin
 - DRVs max. Capacitance/ Transition must be under the limit
 - Metal traces must meet foundry physical DRC requirements
 - Layout geometries should meet Current Density specification



Multi-level Interconnection (MLI)
Technology Layer stacks

Routing

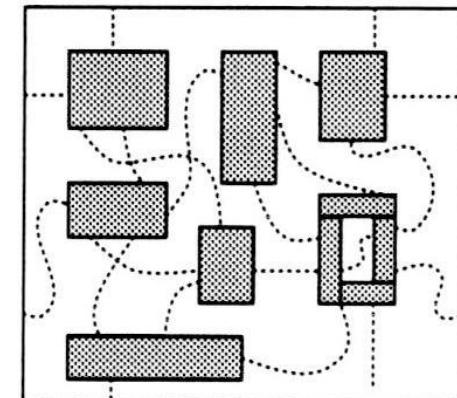
- Routing Stages

- Trial/Global Routing

- Identifying routable path for the nets driving/ driven pins in a shortest distance
 - Does not consider DRC rules, which gives an overall view of routing and congested nets
 - Assign layers to the nets
 - Identify and assign net segments over the specific routable window called Global Route Cell (GRC)
 - Avoid congested areas and also long detours
 - Avoid routing over blockages
 - Avoid routing for pre-route nets such as
 - Rings/Stripes/Rails Uses Steiner Tree and Maze algorithm

- Track Assignment

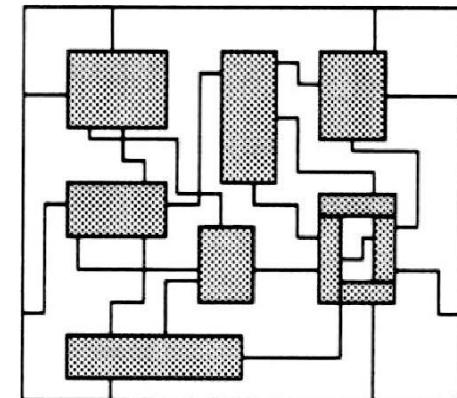
- Takes the Global Routed Layout and assigns each nets to the specific Tracks and layer geometry
 - It does not follow the physical DRC rules
 - It will do the timing aware Track Assignment
 - It helps in Via Minimization



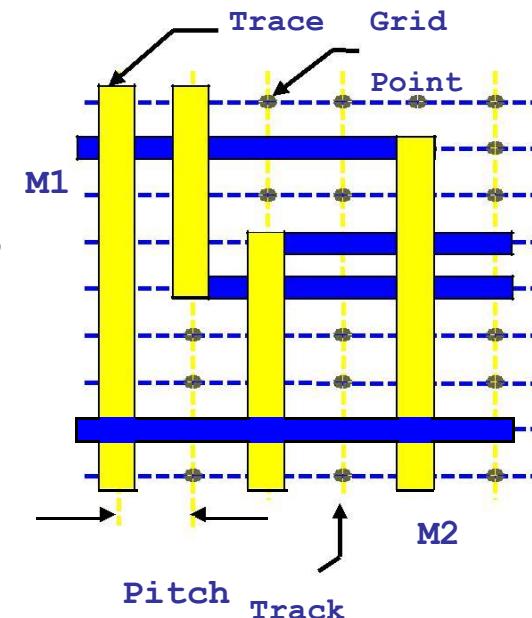
Global Routing

Routing

- Routing Stages
 - Detail/Nano Routing
 - Detailed routing follows up with the track routed net segments and performs the complete DRC aware and timing driven routing
 - It is the final routing for the design built after the CTS and the timing is freeze
 - Filler Cells are adding before Detailed Routing
 - Detail Routing is done after analyze the cause for congestion in the design, add density screen or change flooplan etc.
- Grid Based Routing
 - Metal traces (routes) are built along and centered upon routing tracks on the grid points
 - Various types of grids are Manufacturing Grid, Routing Grid (Pitch) and Placement Grid
 - Grid dimension should be multiple of Manufacturing Grid

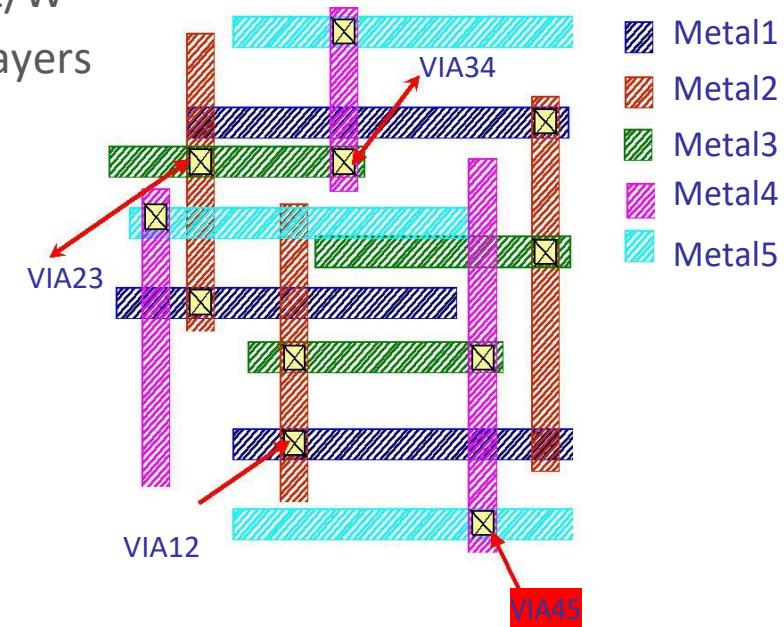


Detailed Routing



Routing

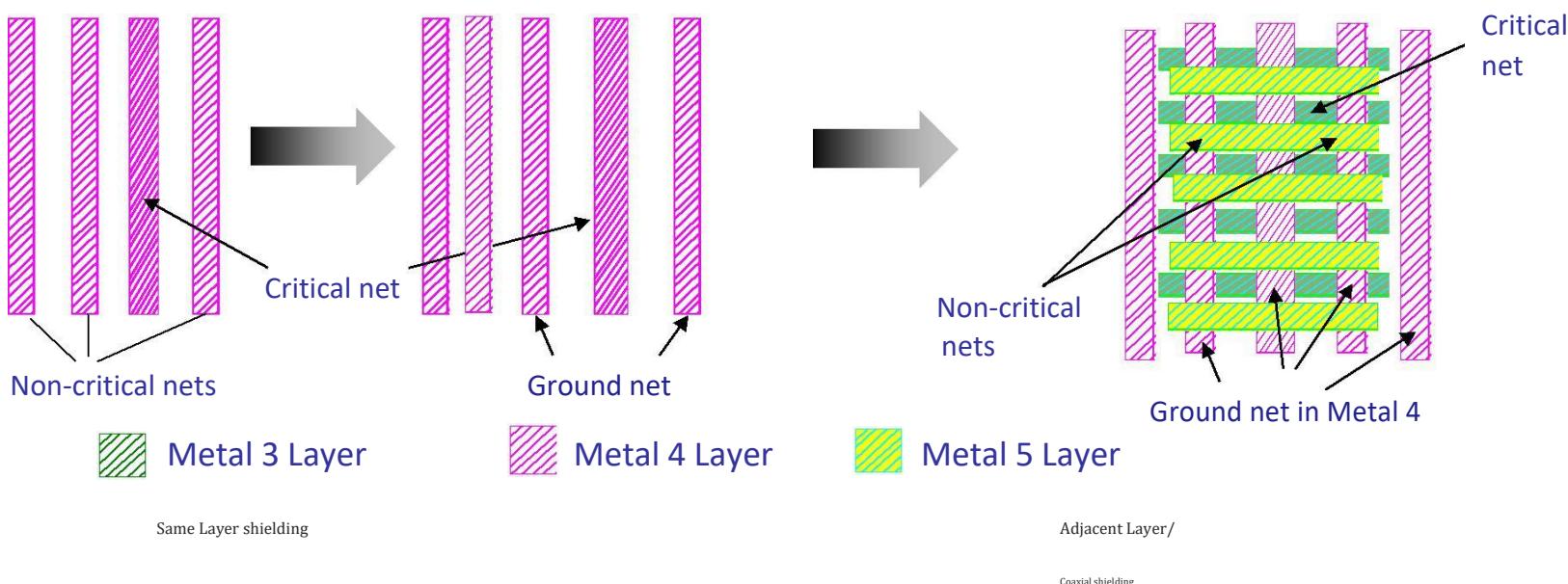
- Routing Preferences
 - Typically Routing only in “Manhattan” N/S E/W directions
 - E.g. layer 1 – N/S Layer 2 – E/W
 - Spacing checks with the adjacent layers
 - Width checks for all layers
 - Via dimension rules
 - Slotting rules
 - A segment cannot cross another segment on the same wiring layer
 - Wire segments can cross wires on other layers
 - Power and Ground have their own layers, mostly the top layers



- Layer Routing directions: Each metal layer has its own preferred routing direction and are defined in a technology rule file
 - M1: Horizontal, M2: Vertical , M3: Horizontal, M4: Vertical and so on
- In some cases, we can avoid following preferred routing direction for smart routing (Non-preferred direction)

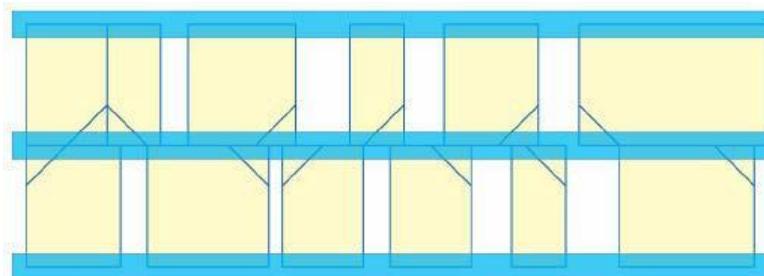
Post Routing Optimization

- Signal Integrity (SI) Optimization by NDRs and Shielding for the sensitive nets
- Types of Shielding for sensitive nets
 - Same layer shielding
 - Adjacent layer/ Coaxial shielding

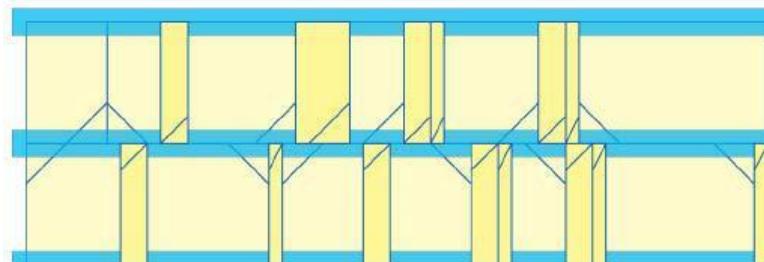


Post Routing Optimization

- Filler Cell insertion
 - Filler Cells can be inserted before or after Detailed Routing
 - If Fillers contain metal routing other than Pre-Routing then Fillers should be inserted before Routing
 - Width of the smallest Filler Cell is the Placement Grid Width
 - Once Fillers are inserted then the placement is fixed and tool can't move Cells for further optimization



Before filler cell is inserted



After filler cells are inserted

Post Routing Optimization

- Metal Fill
 - Filling up the empty metal tracks with metal shapes to meet metal density rules
 - 2 types of Metal Fill
 - Floating Metal Fill: Doesn't completely shield the aggressor nets, so SI will be prominent
 - Grounded Metal Fill: Completely shields the aggressor nets, so less SI
 - impact Grounded Metal Fill is complex as compared to Floating Metal Fill
 - Metal Density Rule helps to avoid Over Etching/ Metal Erosion
- Spare Cells Tie-up/ Tie-down
 - Tie Cells connects the Gate of Cells to VDD/ VSS so reduces ESD
 - Tie-up Cells help in avoid Power Bounce
 - Tie-down Cells help in avoid Ground Bounce
 - Tie Cells are basically MOS in Diode-Connected configuration

Physical Verification (DRC)

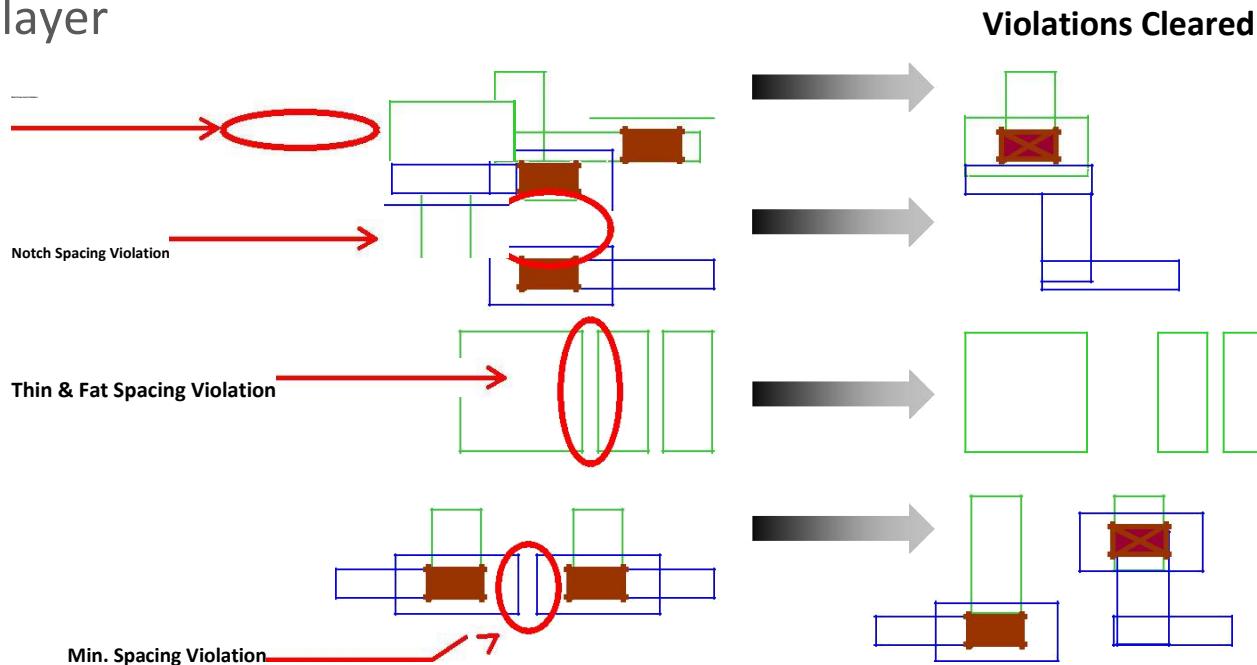
- Design Rule Check (DRC) is the process of checking physical layout data against fabrication-specific rules specified by the foundry to ensure successful fabrication
- Process specific design rules must be followed when drawing layouts to avoid any manufacturing defects during the fabrication of an IC
- Process design rules are the minimum allowable drawing dimensions which affects the X and Y dimensions of layout and not the depth/vertical dimensions
- As Technology Shrinks
 - Number of Design Rules are increasing
 - Complexity of Routing Rules is increasing
 - Increasing the number of objects involved
 - More Design Rules depending on Width, Halo, Parallel Length
- Violating a design rule might result in a non-functional circuit or low Yield

DRC Rule	130nm	90nm	65nm	45nm
Width-based Spacing	1-2	2-3	3-5	7
Min-Area Rule	1 pitch	2 pitch	3 pitch	5 pitch
Cut Number (Via)	N/A	1-2	4-5	5-6
Dense EoL (OPC)	N/A	N/A	M1/M2	All Layers
Min-step (OPC)	N/A	1	5	5



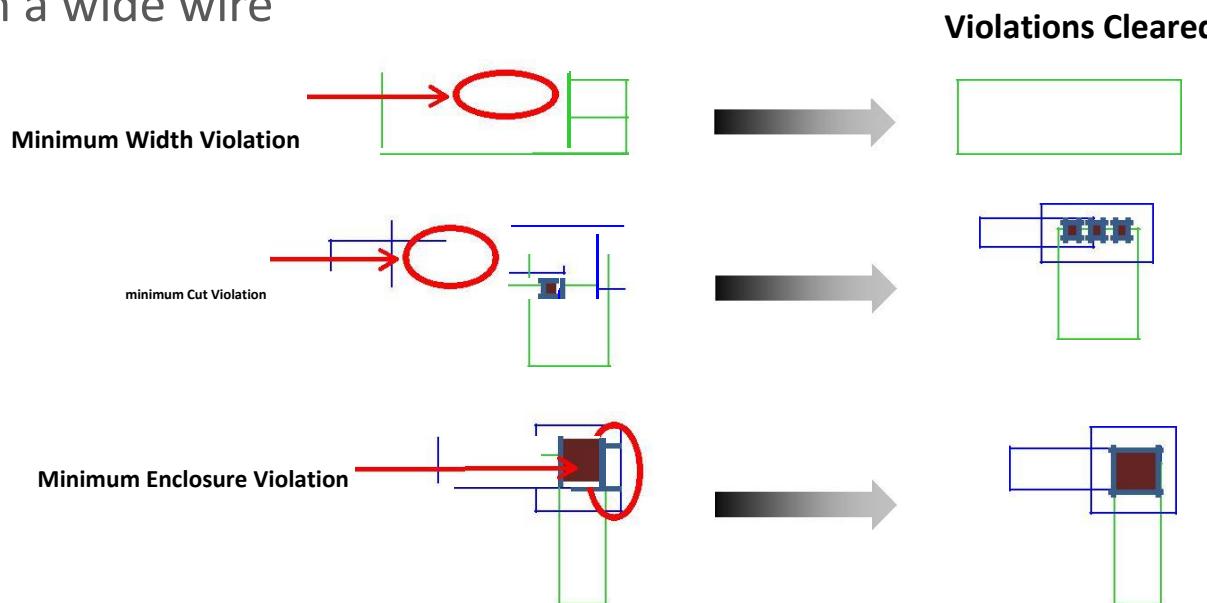
Physical Verification (DRC)

- Design Rule examples
 - Maximum Rules: Manufacturing of large continuous regions can lead to stress cracks. So ‘wide metal’ must be ‘slotted’ (holes)
 - Angles: Usually only multiples of 45 degree are allowed
 - Grid: All corner points must lie on a minimal grid, otherwise an “off grid error” is produced
 - Minimum Spacing: The minimum spacing between objects on a single layer



Physical Verification (DRC)

- Design Rule examples
 - Minimum Width: The min width rule specifies the minimum width of individual shapes on a single layer
 - Minimum Enclosure/ Overlap: Implies that the second layer is fully enclosed by the first one
 - Notch: The rule specifies the minimum spacing rule for objects on the same net, including defining the minimum notch on a single-layer, merged object
 - Minimum Cut: the minimum number of cuts a via must have when it is on a wide wire

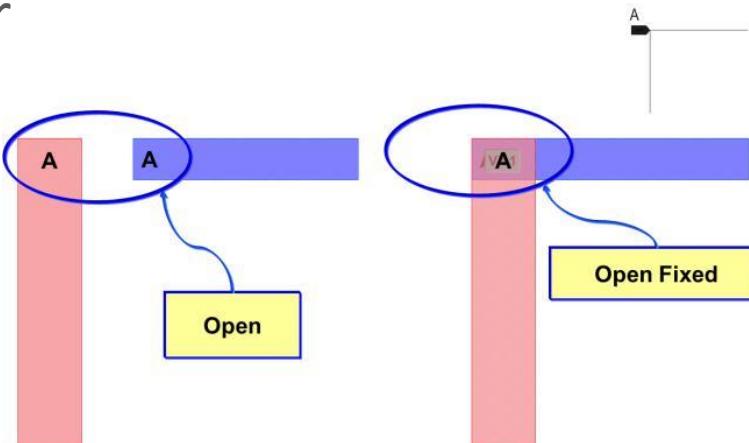


Physical Verification (LVS)

- Layout Versus Schematic (LVS) verifies the connectivity of a Verilog Netlist and Layout Netlist (Extracted Netlist from GDS)
- Tool extracts circuit devices and interconnects from the layout and saved as Layout Netlist (SPICE format)
- As LVS performs comparison between 2 Netlist, it does not compare the functionalities of both the Netlist
- Input Requirements
 - LVS Rule deck
 - Verilog Netlist
 - Physical layout database (GDS)
 - Spice Netlist (Extracted by the tool from GDS)
- LVS checks examples
 - Short Net Error, Open Net Error, Extract errors, Compare errors

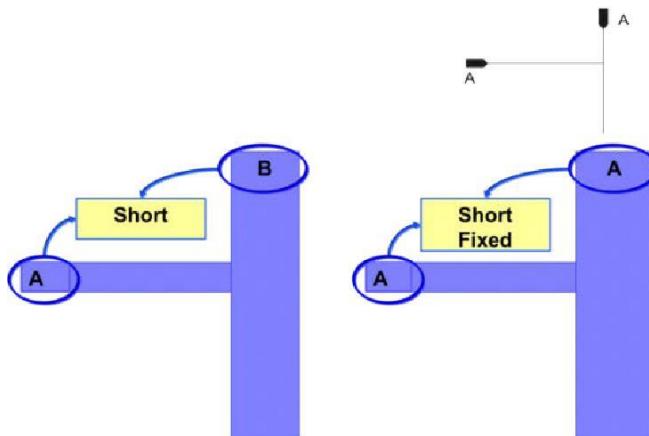
Physical Verification (LVS)

- Open Net Error

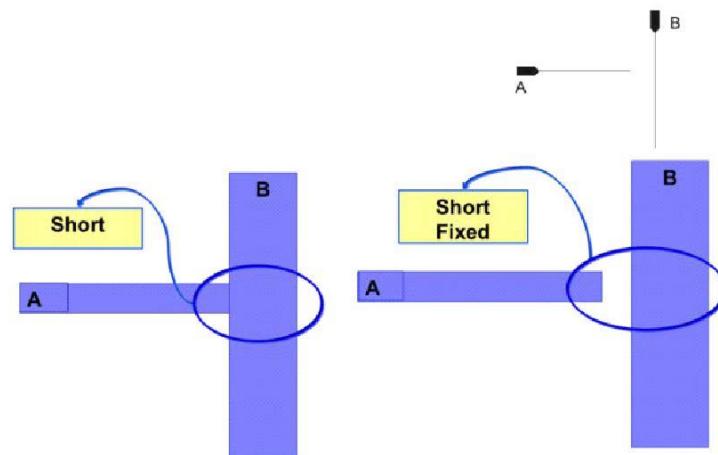


Same net is routed in two different metal layers but not connected

- Short Net Error



Same net with different pin names

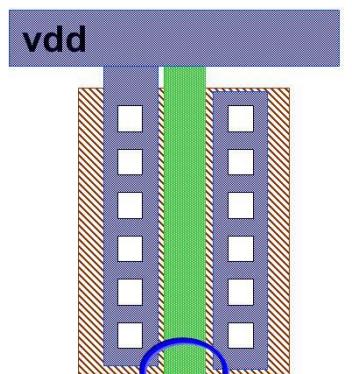


Two different nets shorting together

Physical Verification (LVS)

- Extract Errors
 - Parameter Mismatch
 - Device parameters on schematic and layout are compared
 - Example: Let us consider a transistor here, LVS checks are necessary parameters like width, length, multiplication factor etc.

Layout :



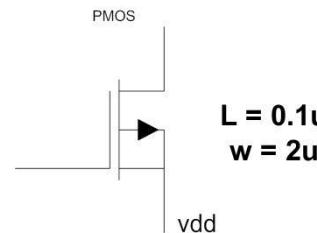
**Parameter
Violation**

$L = 0.13\text{u}$
 $w = 2\text{u}$

**Violation
Fixed**

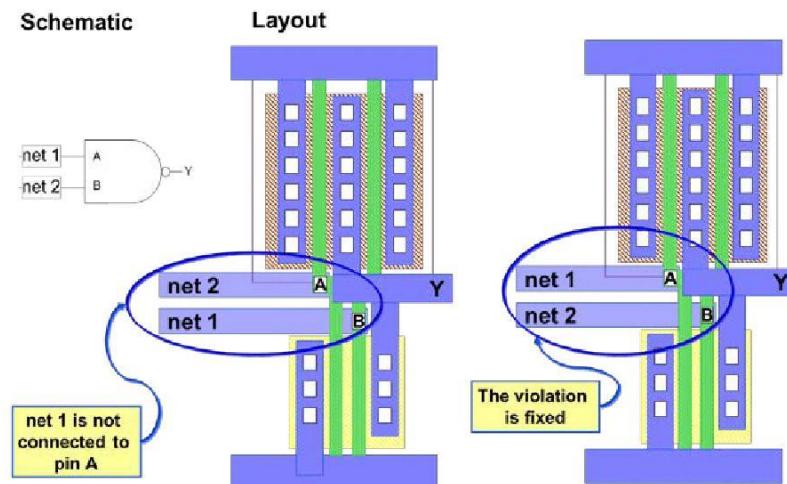
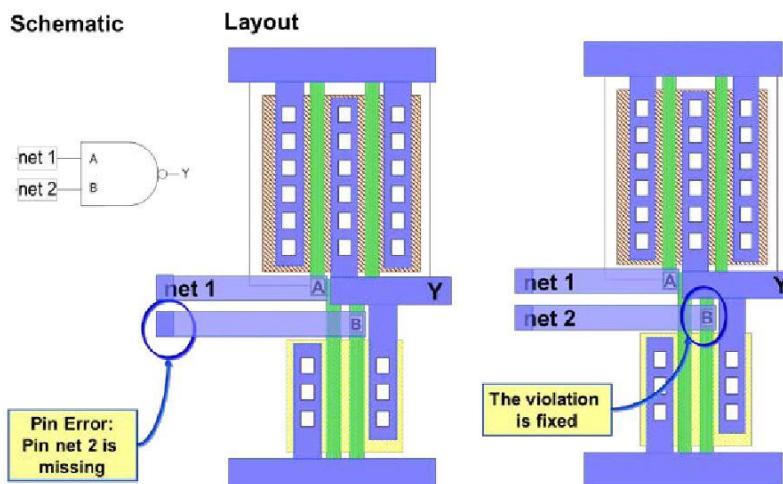
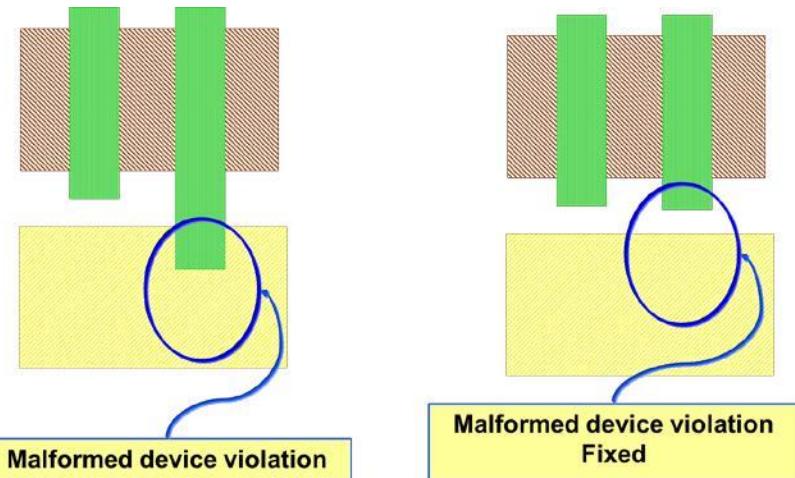
$L = 0.1\text{u}$
 $w = 2\text{u}$

Schematic:



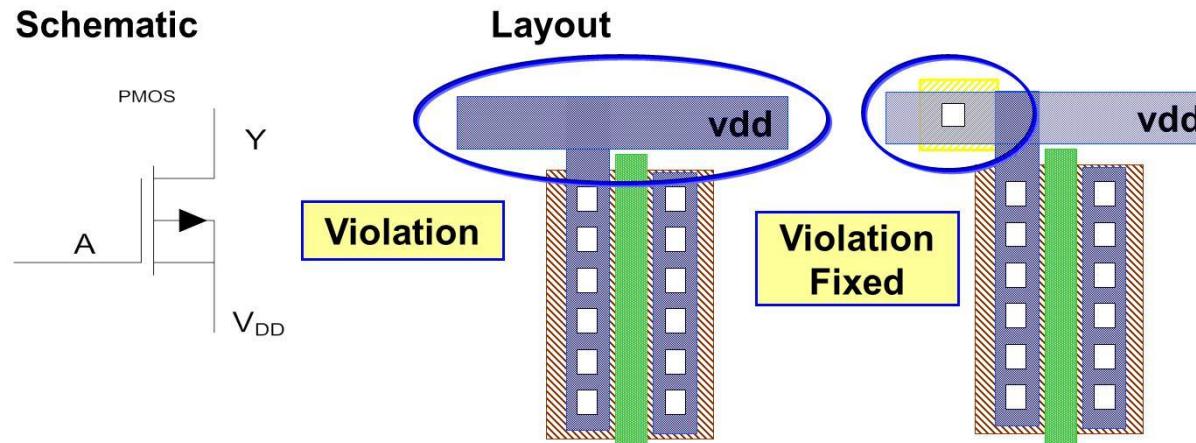
Physical Verification (LVS)

- Compare Errors
 - Malformed Devices
 - Pin Errors
 - Device Mismatch
 - Net Mismatch



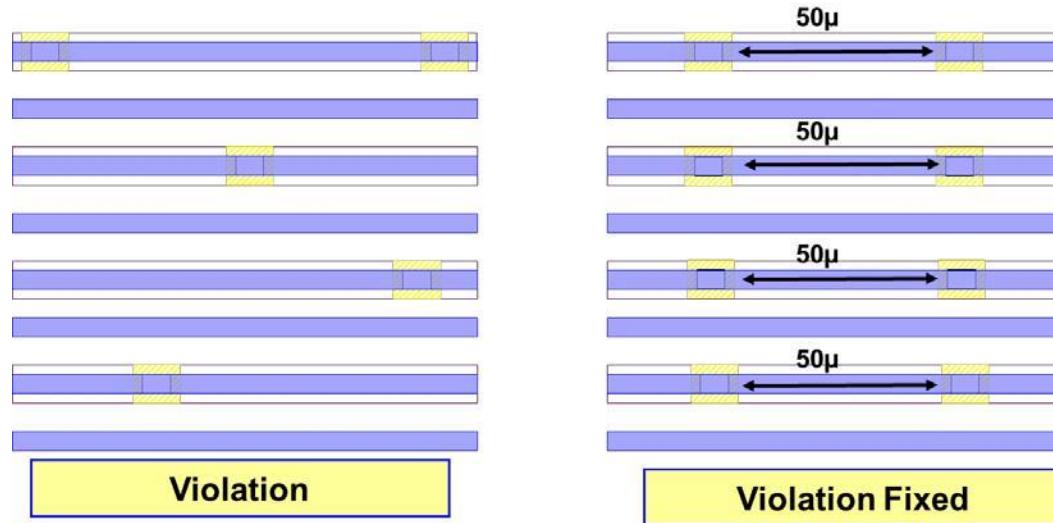
Physical Verification (ERC)

- Electrical Rule Check (ERC) is used to analyze or confirm the electrical connectivity of an IC design
- ERC checks are run to identify the following errors in layout
 - To locate devices connected directly between Power and Ground
 - To locate floating Devices, Substrates and Wells
 - To locate devices which are shorted
 - To locate devices with missing connections
- Well Tap connection error: The Well Taps should bias the Wells as specified in the schematics



Physical Verification (ERC)

- Well Tap Density Error: If there is no enough Taps for a given area then this error is flagged
- Taps need to be placed regularly which biases the Well to prevent Latch-up
 - e.g., In typical 90nm process the Well Tap Density Rule require Well-taps to be placed every 50 microns
- Tools: Mentor Graphics Calibre, Synopsys Hercules, Cadence Assura, Magma Quartz



DFM Checks

- Antenna Check (Gate-Oxide Integrity check)
 - Maximum net length restriction connected to Gate terminal
- Redundant Contacts/ Via
 - Multiple Via improves both Yield and Timing by resistance paralleling
- Metal Filling
 - Narrow Metal Layer separated from other Metal Layers may get high density of etchant than closely spaced wires
 - Over etched filling up empty tracks with metal shapes to meet Metal Density Rules
- Metal Slotting
 - Wide metal lines (Power Nets) expands significantly due to the high temperature during fabrication leads to destruction of the isolation and passivation layer that protect the wafer
 - To avoid it put slots or holes in these metal layers at regular intervals
 - Slotting also prevent the stress damage during wafer dicing and packaging

Formal Verification

- Formal Verification
 - Verify the two representations of circuit design exhibits same behavior
 - Checks the behavior of the Combinational Logics by checking the Compare Points
 - Targets implementation errors and not the design errors
 - Power checks: checks Power Switches/ Retention Cells/ Isolation Cells/ Level Shifters and all power connectivity
 - If any manual editing in the design then LEC has to be done at any point of time
- Formal Verification
 - Complete coverage
 - Effectively exhaustive simulation
 - Cover all possible sequences of inputs
 - Check all corner cases
 - No test vectors are needed
- Informal Verification (Simulation)
 - Incomplete coverage
 - Limited amount of simulation
 - Spot check a limited number of input sequences
 - Many corner cases not checked
 - Designer provides test vectors

Formal Verification

- Types of Formal Verification
 - Gate-level to Gate-level (Logical Equivalence Check after Routing)
 - To ensure that some netlist post-processing did not change the functionality of the circuit
 - RTL to Gate-level (after Synthesis)
 - To verify that the netlist correctly implements the original RTL code
 - RTL to RTL (before Synthesis)
 - To verify that two RTL descriptions are logically identical
- Logical Equivalence Check (LEC) will have two stages
 - Constrains setup stage
 - Logical Equivalence Check stage
- Tool will report equivalent/ non-equivalent/ abort/ not-checked
- Input Requirements
 - Netlists (.v)
 - Library (.lib and .lef)
 - Constraints (.sdc)
- Tools: Mentor Graphics FormalPro, Cadence Conformal, Synopsys Formality, Magma Quartz Formal

Parasitic Extraction

- Parasitic Extraction: Importance
 - Shrinking process geometries
 - New device structures
 - An increasing number of metal layers at each new process node
 - Much more closer nets at each new process node
 - Increasing wire aspect ratio of height to width
 - Increasing operating frequency
- Parasitic Capacitance can be reduced by using higher metals, provide spacing, shielding, Avoid parallel routing
- At higher clock frequencies, RC interconnect modeling is no longer adequate and inductance must be included in interconnect modeling
- Reluctance (Inductance) effect becomes more and more prominent as the resistance (both device and interconnect) decreases and the operating frequency increases

Parasitic Extraction

- Capacitance

$$C = \epsilon_0 W H / d$$

- Transistors

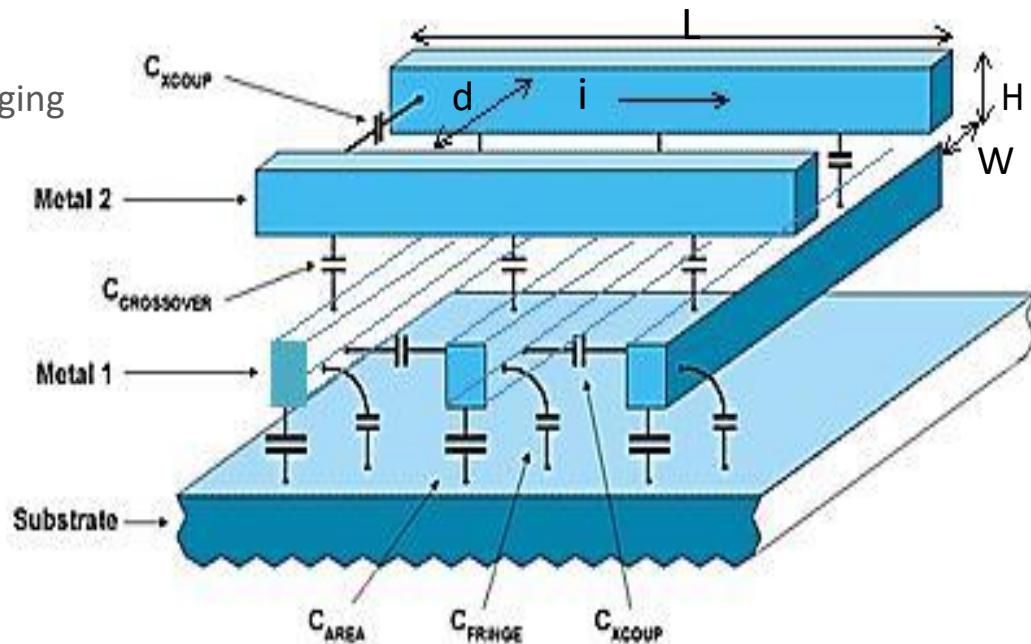
- Depends on area of transistor gate, physical of materials, thickness of insulator, diffusion to substrate

- Poly to Substrate

- Parallel plate and fringing

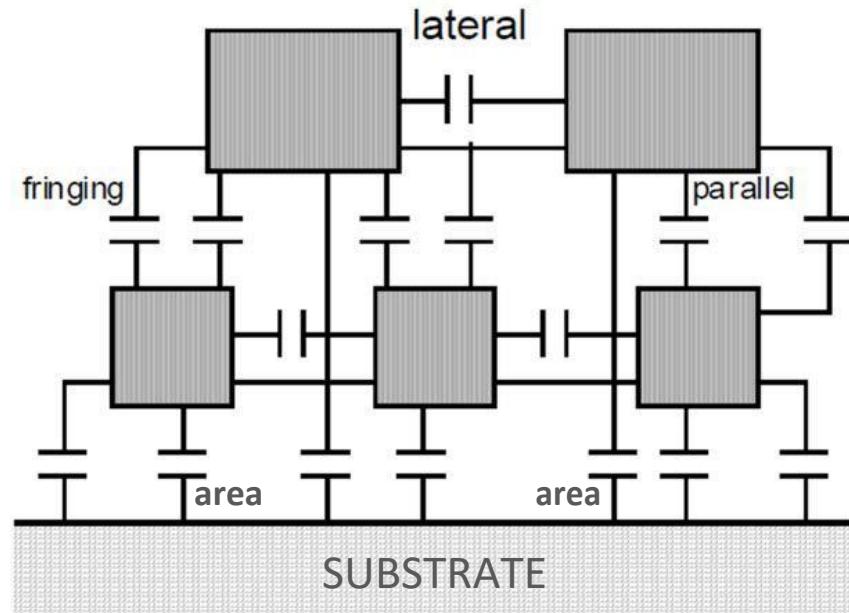
- Capacitance between conductors

- Coupling Capacitance
- Area Capacitance
- Fringing Capacitance
- Crossover Capacitance



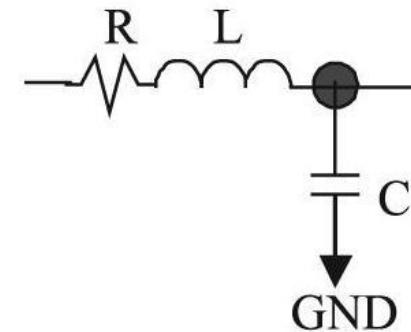
Parasitic Extraction

- Coupling Capacitance/ Lateral Capacitance
 - The capacitance between nets on the same Metal layer
 - Dominant over interlayer capacitances with every new process technology
- Fringing Capacitance
 - Capacitance between nets of different Metal layers and other layers due to Sidewall Capacitance
- Parallel/Crossover Capacitance
 - Capacitance between nets of 2 different Metal layers
- Area Capacitance
 - Capacitance between Metal layers and Substrate
- In modern processes, the width of interconnect wires at lower levels of metal is so small that the Fringing Capacitance of the wire is larger than the Area Capacitance



Parasitic Extraction

- Resistance
 - $R = \rho L/H W$
 - Wire Resistivity
 - Complex 3D geometry around Vias
- Inductance
 - Self Inductance; $V = L \frac{di}{dt}$
 - Mutual Inductance, $M = K \sqrt{L_1 L_2}$
 - At high frequency Skin effect possibility
- Models used for Parasitic Extraction
 - Lumped-C, Lumped-RC, Lumped-RLC
 - Pi segment
 - Pin-to-pin delays are modeled by RC delays



Parasitic Extraction

- Sub-femto Farad accuracy required for extraction of designs at advanced technology nodes
- STA tool uses extraction data at fast corner while calculating hold and slow data while calculating setup to be pessimistic as possible, so that your chip doesn't fail after it comes back from the fab
- Common Extraction Formats: Standard Parasitic Format (SPF), Reduced Standard Parasitic Format (RSPF), Detailed Standard Parasitic Format (DSPF), Standard Parasitic Extraction Format (SPEF)
- Tools: Synopsys Star-RCXT, Cadence QRC, Mentor Graphics Calibre xRC

Timing Analysis

- Static Timing Analysis: Methodical analysis of a digital circuit to determine if the timing constraints imposed are met and to check the design is working properly
- Static Timing Analysis Flow
 - Read the inputs required
 - Setting up Constraints: IO Delay Constraints, DRVs, Timing Exceptions (False/ Multi-Cycle paths), Recovery and Removal, Minimum Pulse Width
 - Construct Timing Graph: Partition Clock Domain, Ideal/ Propagated Clock, Case Analysis
 - Propagation
 - Timing Report: End points with violations/ Paths enumeration
- Input Requirement
 - Routed Netlist (.v)
 - Libraries (.lib only)
 - Constraints (.sdc)
 - Delay Format (.sdf)
 - Parasitic Values (.spef)
- Tools: Synopsys PrimeTime, Cadence ETS, Cadence Tempus

Timing Analysis (SI)

- Signal Integrity (SI)
 - SI refers to the quality of the signal transportation during the circuit operation
 - In deep sub-micron the delays associated with the logic elements far outweighed delays associated with the interconnect
 - SI effects like Crosstalk (both noise and timing), Voltage (IR) Drop, Waveform Integrity and Electromigration have complex interdependencies
 - When the technology shrinks, the effect of coupling capacitance also increases
 - Crosstalk is the undesirable phenomenon, caused by the cross coupling capacitance between metal wires in a chip
 - Signal Integrity comes as an added feature of Timing Signoff tools
 - Crosstalk effects can be analyzed by enabling the SI switch in tools
 - If Crosstalk is enabled then the tool will by default do the timing in On Chip Variation (OCV) mode
 - Tool can read the .spef file which consists of coupling capacitance info.

Power Analysis & IR Drop Analysis

- Power Analysis
 - Static/ Leakage Power Analysis
 - Dynamic Power Analysis
- IR Drop Analysis
 - Static IR Drop Analysis
 - Dynamic IR Drop Analysis
- Tools for Power and IR Drop Analysis
 - Synopsys Prime Power
 - Cadence EPS and Voltus
 - Apache Redhawk
- Tape-out
 - Final GDSII (Graphical Data Stream Information Interchange) or CIF (Caltech Intermediate Format) to Foundry
 - GDS contains Physical Layout information

Thank You

Analysis in ASIC Physical Design

Outline

- **Timing Analysis**
 - Dynamic vs. Static Timing Analysis
 - Static Timing Analysis (STA)
- **Congestion Analysis**
- **Power Analysis**
 - Dynamic Power Analysis
 - Static Power Analysis
- **IR Drop Analysis**
 - Dynamic IR Drop Analysis
 - Static IR Drop Analysis

Timing Analysis

Timing Analysis

Dynamic Timing Analysis (DTA)

Verifies functionality of the design by applying input vectors and checking for correct output vectors

Quality increases with the increase of input test vectors

Increased Test Vectors increase Simulation Time

Can be used for synchronous as well as asynchronous designs

Also best suitable for designs having clocks crossing multiple domains

Computational complexity involved in finding the Input Patterns/Vectors that produces maximum delay at the output

Static Timing Analysis (STA)

Checks Static Delay requirements of the circuit without any input or output vectors, so analysis times are relatively short and STA does not check for logical correctness of the design

Clock related all information has to be fed to the design in the form of constraints and the correctness of the constraints decides the quality

Timing can be analyzed for worst case and best case simultaneously and also all timing paths are considered

Not suitable for asynchronous designs

Not suitable for designs having clocks crossing multiple domains

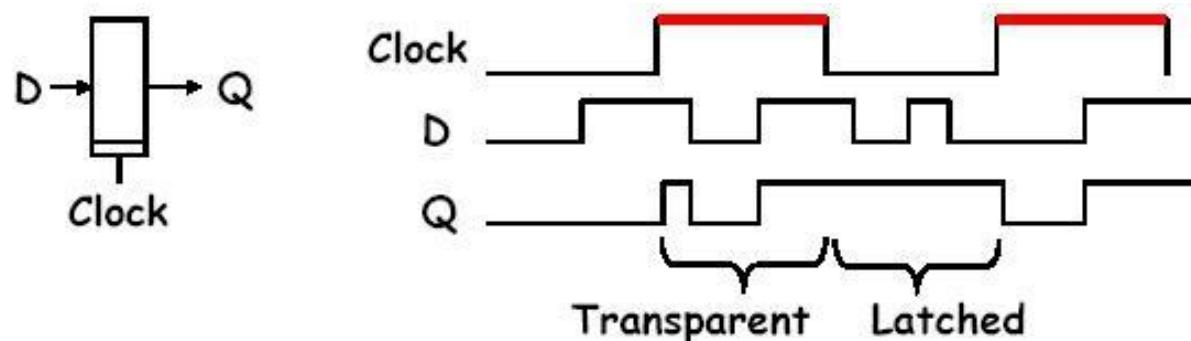
Has more pessimism and thus gives maximum delay of the design and STA and it works with timing models

Static Timing Analysis (STA)

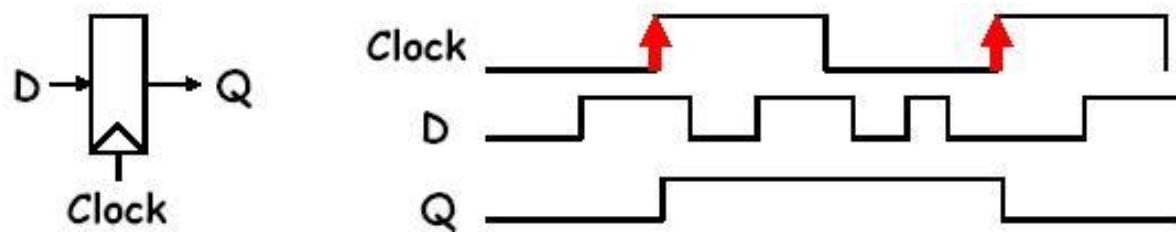
- Static Timing Analysis
 - Effective methodology for verifying the timing characteristics of a design without the use of test vectors
 - Static Timing Analysis can be done only for Register-Transfer-Logic (RTL) designs
 - Functionality of the design must be cleared before the design is subjected to STA
 - STA approach typically takes a fraction of the time it takes to run logic simulation
- STA tool analyzes all paths from each and every start point to each and every end point and compares it against the constraint that exists for that path
- Main steps of STA
 - Break the design into sets of timing paths
 - Calculate the delay of each path
 - Check all path delays to see if the given timing constraints are met

Static Timing Analysis (STA)

- Clocked Storage Elements
 - Transparent Latch, Level Sensitive
 - Data passes through Latch when clock high, latched when clock is low



- D-Type Register or Flip-Flop, Edge-Triggered
 - Data captured on rising edge of clock, held for rest of the cycle



Static Timing Analysis (STA)

- Delays

- Time taken by a signal to propagate through a Cell or Net
- Actual Path Delay is sum of net and Cell Delays along the timing path
- Cell Delay is a function of Input Transition Time (Slew Rate), Total Output Load (Net Cap + Sum of attached pin caps) and Process Parameters (Temperature, Power Level)

- Intrinsic delay

- Internal to the Cell from Input pin to Output pin caused by internal capacitance

- Propagation Delay

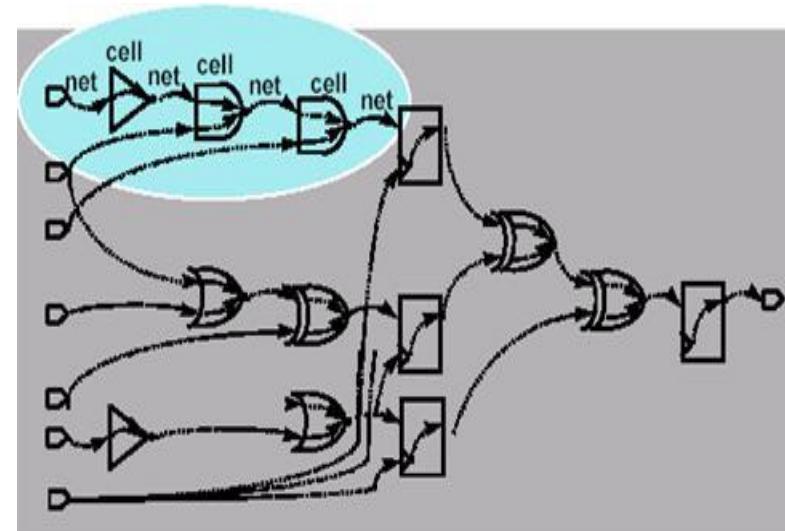
- Delay by a cell for a change of input signal to result a change at output signal as a function of Input Slew and Output load
- Propagation Delay can be Low to High (t_{PLH}) and High to Low (t_{PHL})
- Maximum Propagation Delay (Clock to Q) is considered for Setup check

- Contamination Delay

- Best case delay from valid input to output
- Minimum Propagation Delay (Clock to Q) which is called Contamination Delay is considered for Hold check

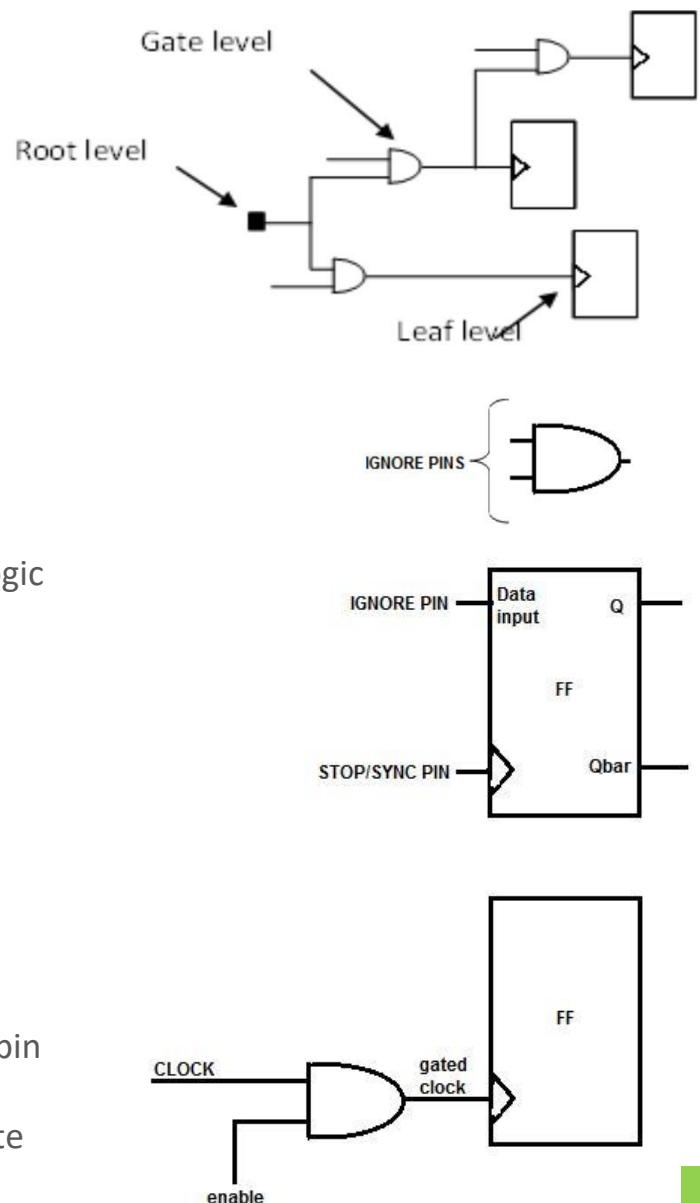
- Net Delay

- Total time for charging/discharging all the parasitic present in the given net



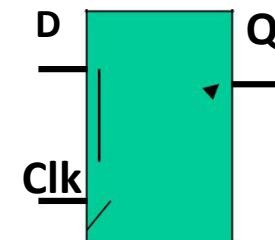
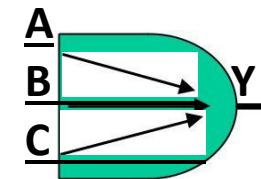
Static Timing Analysis (STA)

- Pins related to Clock Design
 - Start/ Source / Root Pins
 - Source pin of a Clock
 - Stop/ Sink/ Leaf Pins
 - All Clock Pins of Flip Flops
 - Clock wont propagate after this Pin
 - Through pin
 - To make a Clock pin of a flop not a CTS Leaf pin
 - Preserved Pin
 - If we need to preserve a pin w.r.t. location etc.
 - Exclude/ Ignore Pins
 - All non-clock pins (D pin of Flip Flops or combo logic inputs) Not considered for Clock propagation
 - Float Pins (Implicit Stop/ Macro Model)
 - Same as Stop/ Sink Pin but internal Clock Latency of it is considered for Clock Tree
 - Its actually entry pin of the Hard Macro
 - Explicit Sync (Stop) Pin
 - Input of combo logic while considering Clock
 - Tree Important while considering Clock Gating
 - Explicit Exclude (Ignore) Sync Pin
 - Clock Pin of Flop is not considered as Sync/ Stop pin
 - This pin is due to Clock Gating concept
 - In clock gating the signal will be given to AND Gate



Static Timing Analysis (STA)

- Timing Arc
 - Timing Arc is internal to the cell
 - Combinational Cells has Timing Arcs from each Input to each Output of the cell
 - Flip-flops have Timing Arcs from the Clock Input pin to Data Output Q pin (Propagation delay/ Delay Arc) and from Clock Input pin to Data Input D pin (setup, hold checks/ Constraint Arc)
 - Latches have 2 timing arcs:
 - Clock pin to Output Q pin, when D is stable
 - Data D pin to Output Q pin when D changes (Latch is transparent)
- Timing Unate
 - How Output changes for different types of transitions on Input
 - Positive Unate if Output Transition is same as Input Transition
 - Negative Unate if Output Transition opposite to Input Transition
 - Non-Unate if the Output Transition cannot be determined solely from the direction of change of an Input. It also depends upon the state of the other Inputs

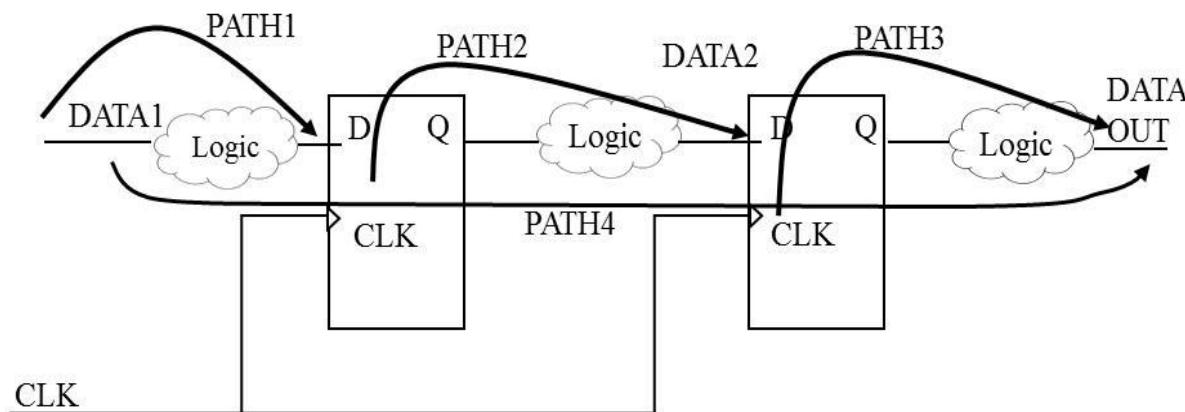


Static Timing Analysis (STA)

- Clock definitions in STA
 - Synchronous Clocks
 - 2 clocks are synchronous w.r.t. each other
 - Timing paths launched by one clock and captured by another
 - Asynchronous Clocks
 - 2 clocks are asynchronous w.r.t. each other
 - If no timing relation, STA can't be applied, so the tool won't check the timing
 - Mutually-Exclusive Clocks
 - Only one clock can be active at the circuit at any given time
 - Generated Clocks
 - Clock generated from a clock source as a multiple of the source clock frequency
 - The frequency can be a multiple or can be divided by the source clock
 - Virtual Clocks
 - Exists but not associated with any pin or port of the design
 - Used as a reference in STA to specify Input Delays and Output Loads relative to a clock (Needed to fix the Input2Reg and Reg2Output Violations)
 - By defining Virtual Clock IO Constraints can be defined relative to this Virtual Clock with no specification of the source port or pin

Static Timing Analysis (STA)

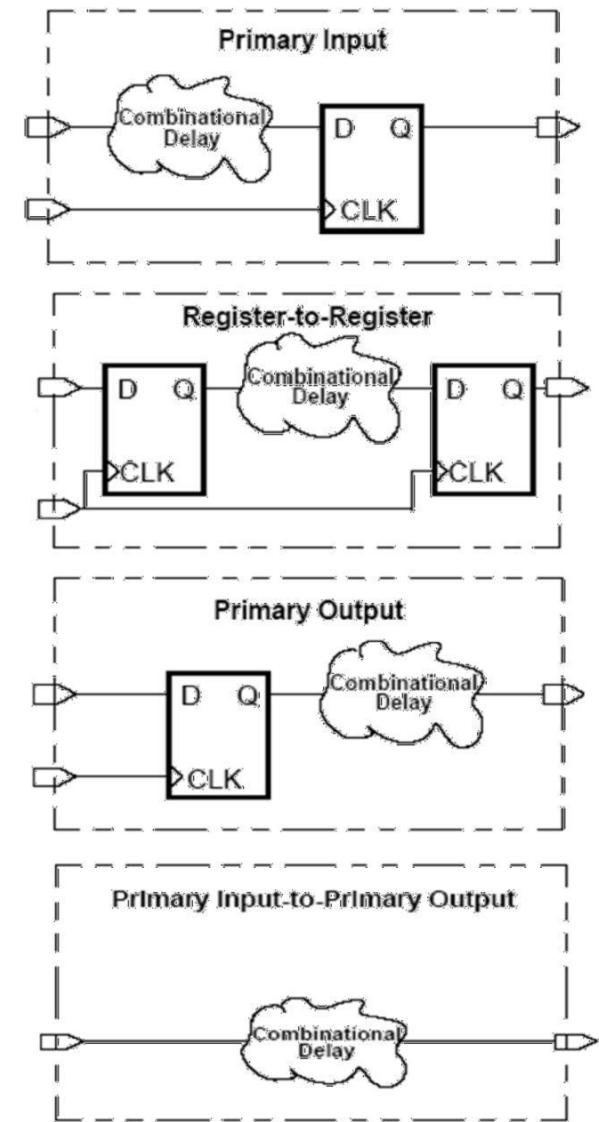
- A Timing Path is a point-to-point path in a design which can propagate data from one flip-flop to another
 - Each path has a start point and an end point
 - Start point: Input ports or Clock pins of flip-flops
 - Endpoints: Output ports or Data input pins of flip-flops



Timing Paths

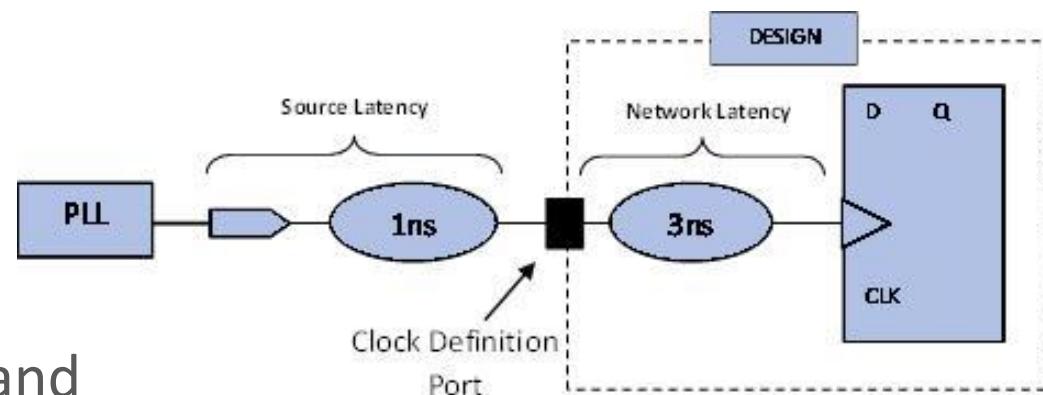
Static Timing Analysis (STA)

- Timing Path Groups
 - Timing paths are grouped into path groups by the clocks controlling their endpoints
 - Input pin/port to Register
 - Delays off-chip + Combinational logic delays up to the first sequential device
 - Register to Register
 - Start at a sequential device
 - CLK-to-Q transition delay + the combinational logic delay + external delay requirements
 - Register to Output pin/port
 - Delay and timing constraint (Setup and Hold) times between sequential devices for synchronous clocks + source and destination clock propagation times
 - Input pin/port to Output pin/port
 - Delays off-chip + combinational logic delays + external delay requirements



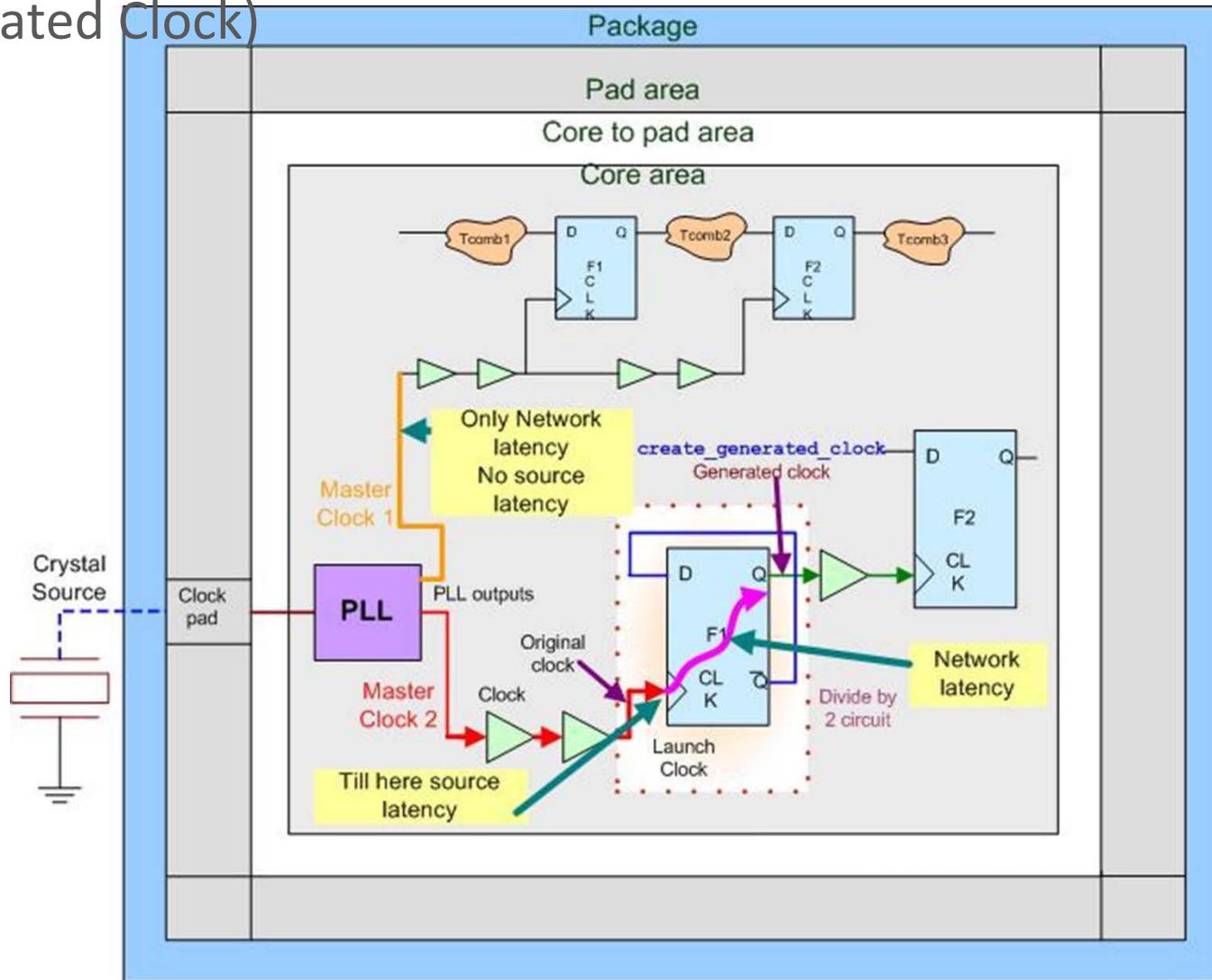
Static Timing Analysis (STA)

- Clock Latency
 - Total time taken by the clock signal to reach the input of the register
 - Source latency is the time between clock sources to clock definition ports
 - Network latency is the time between clock definition ports to clock leaf cells in the design
- Insertion Delay (ID)
 - ID is the clock latency, but after Clock Tree is synthesized
- ID is the physical delay and Clock Latency is the virtual delay
- Latency is a target given to the tool through SDC file or clock tree attribute file and Insertion Delay is the achieved delay value after CTS



Static Timing Analysis (STA)

- Source and Network Latency (Original Clock & Generated Clock)



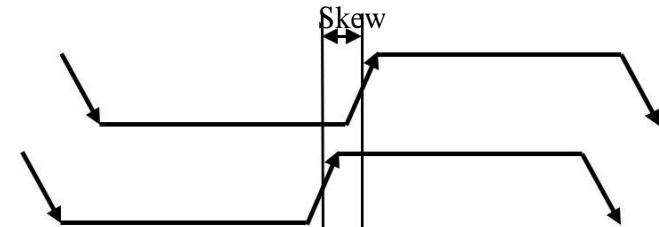
Static Timing Analysis (STA)

- Clock Uncertainty
 - Clock Uncertainty is the time difference between the arrivals of clock signals at registers in one clock domain or between domains
 - Uncertainties include Clock Skew, Clock Jitter and Clock Margin

- Clock Skew

- Clock Skew refers to the absolute time difference in clock signal arrival between two points in the clock network

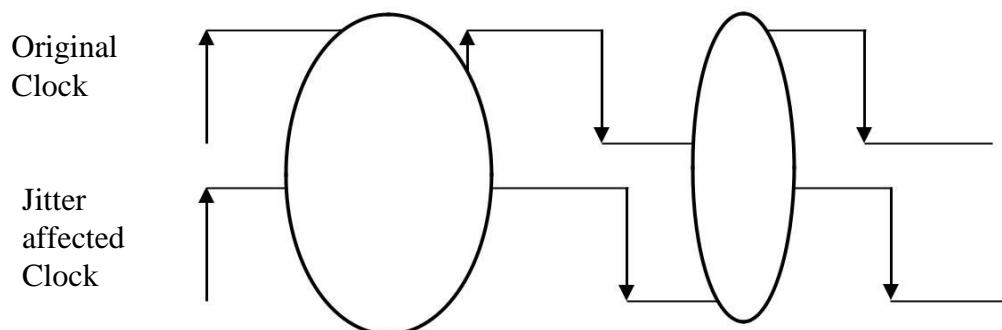
$$T_{\text{LAUNCH_CLOCK}} - T_{\text{CAPTURE_CLOCK}} = T_{\text{SKEW}}$$



- Positive Skew occurs when the Capture Clock is late w.r.t. Launch Clock
 - Negative Skew occurs when the Capture Clock is early w.r.t. Launch Clock
 - Local Skew is the Skew between the clock phase delays of two flip-flops which are the Source and Target flop of a path (Source and Destination flop)
 - Global Skew is the difference between the longest and shortest branch of a Clock Tree (Maximum Insertion Delay – Minimum Insertion Delay)

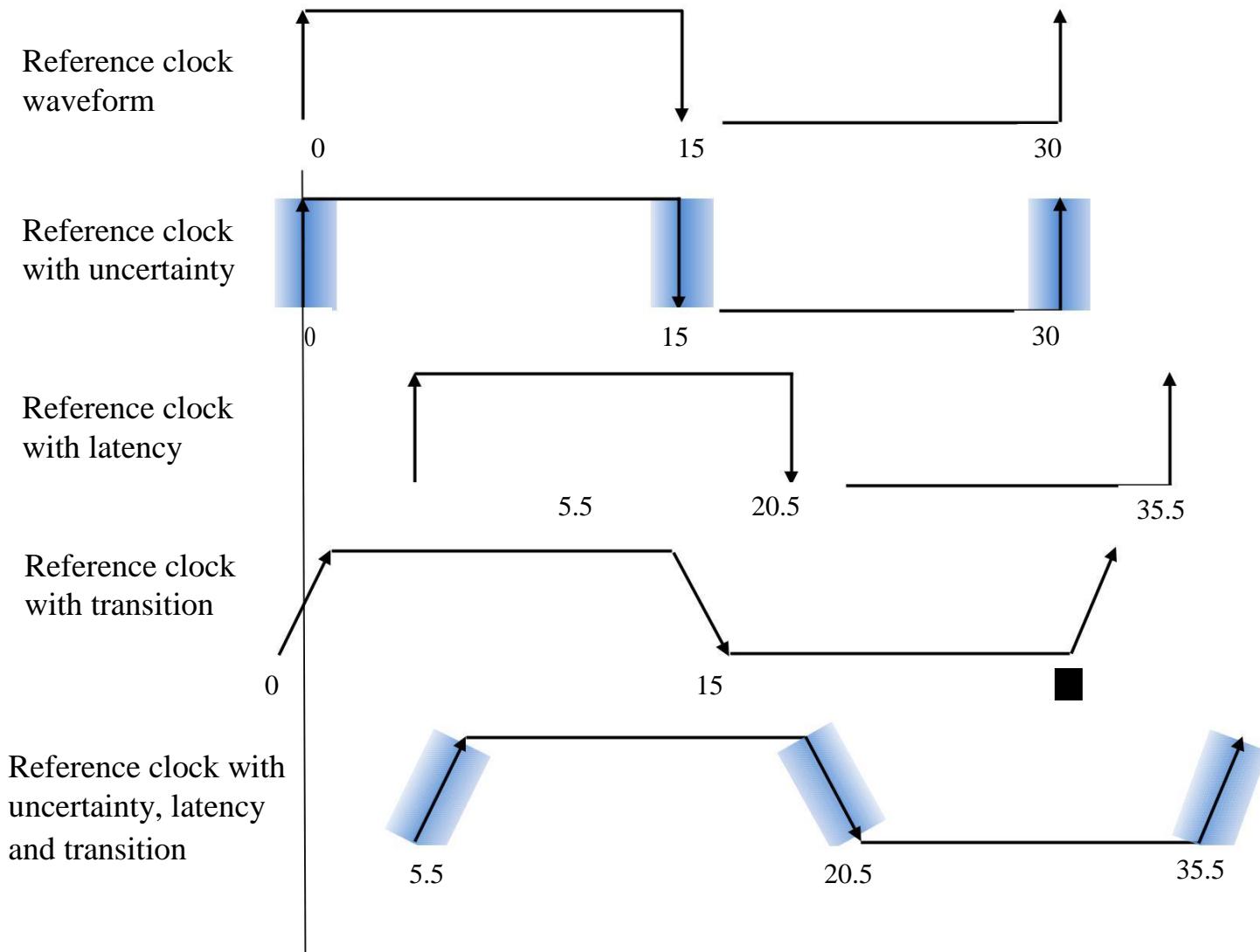
Static Timing Analysis (STA)

- Clock Jitter
 - Jitter is the short-term variations of a signal with respect to its ideal position in time
 - The two major components of Jitter are random Jitter and deterministic Jitter
 - Factors causing Jitter includes imperfections in Clock oscillator, supply voltage variations, Temperature variations, Crosstalk



- Glitch
 - Unexpected switching of any waveform
 - Due to late arrival time of Gate and it is for a short period of time
 - Cause extra delay and also it can cause extra power from false transitions

Static Timing Analysis (STA)

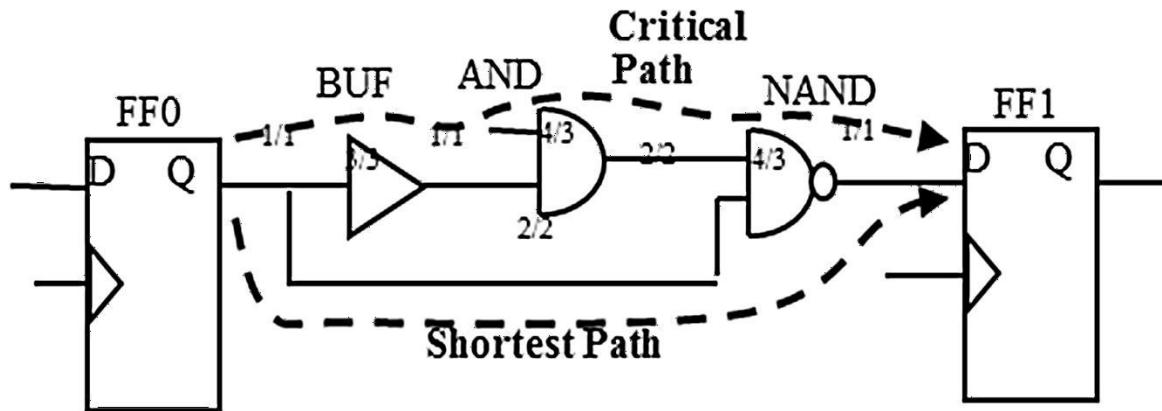


Static Timing Analysis (STA)

- Pulse Width
 - Pulse Width is the time between the active and inactive states of the same signal
 - Minimum high pulse width is the amount of time after the rising edge of a clock, that the clock signal of a clocked device must remain stable
 - Minimum low pulse width is the amount of time after the falling edge of a clock, that the clock signal of a clocked device must remain stable
- Duty Cycle
 - Percentage of clock period having high pulse
 - Typically clock waveforms are of 50% Duty Cycle
- Transition/ Slew
 - Time taken by a signal to change the state (Volts/Second)
 - Rise Slew (t_R) is called Rise Time and Fall Slew (t_F) is called Fall Time
 - Minimum/ Maximum Transition is the Minimum/ Maximum slope allowed at leaf pins
 - Transition affects Power Dissipation, Latency and Pulse width

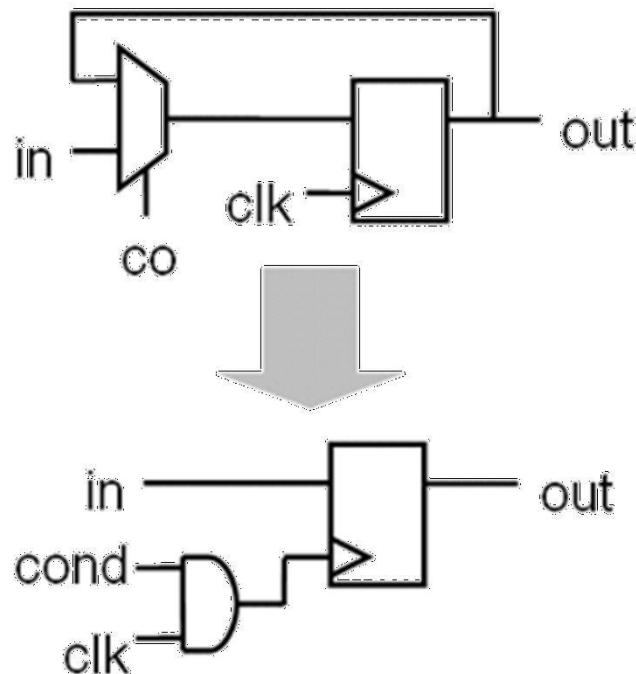
Static Timing Analysis (STA)

- Asynchronous Path
 - A path from an input port to an asynchronous set or clear pin of a sequential element
- Critical Path
 - The path which creates longest delay
 - Also called worst path/ late path/ max. path
 - Timing sensitive functional paths no additional gates are allowed to be added to the path
- Shortest Path
 - One that takes the shortest time; this is also called the best path or early path or a min path



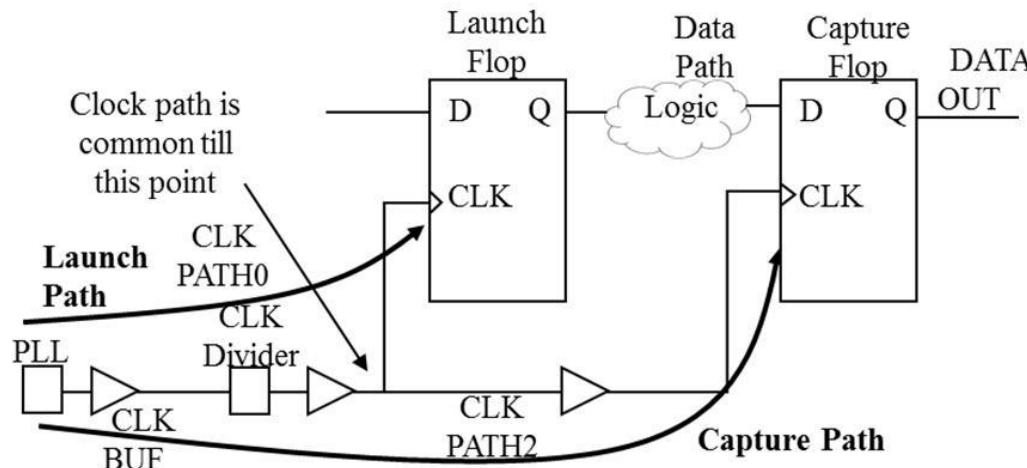
Static Timing Analysis (STA)

- Clock Gating Path
 - Path passed through a “gated element” to achieve additional advantages
 - Clock Gating transformation does not change the state of the flops and register



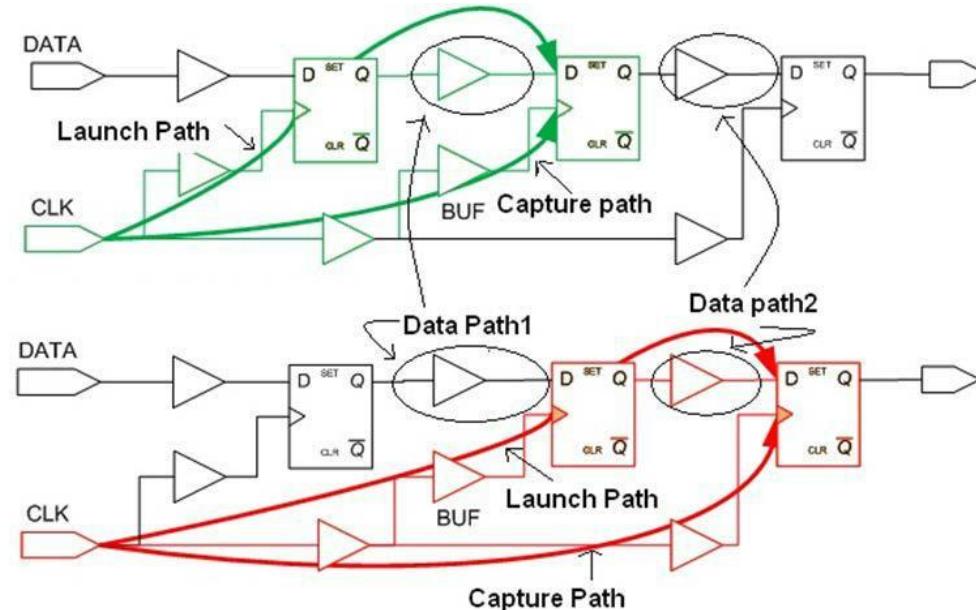
Static Timing Analysis (STA)

- Launch Path
 - Launch path is launch clock path which is responsible for launching the data at launch flip flop
- Capture Path
 - Capture path is capture clock path which is responsible for capturing the data at capture flip flop
- Arrival Time
 - Launch path and data path together constitute arrival time of data at the input of capture flip-flop
- Required Time
 - Capture clock period and its path delay together constitute required time of data at the input of capture register



Static Timing Analysis (STA)

- Common Path Pessimism
 - Same Clock Path may be a Launch Path for one Data Path and can be a Capture Path for another Data Path
 - While doing OCV derating, same path may get both Min./ Max. delay
 - But a path can have either as a Maximum delay or a Minimum delay (or anything in between) but never both delays at the same time
 - STA tools will have techniques to remove artificially introduced pessimism between the Launch Clock Path and the Capture Clock Path



Static Timing Analysis (STA)

- Slack
 - Difference between Required Time (RT) and Arrival Time (AT)
 - Positive Slack at a node implies that the arrival time at that node may be increased without affecting the overall delay of the circuit
 - Negative Slack implies that a path is too slow, and the path must speed up if the whole circuit is to work at the desired speed
- Setup Time

- Setup time is the minimum amount of time the data signal should be held steady before the clock event so that the data are reliably sampled by the clock

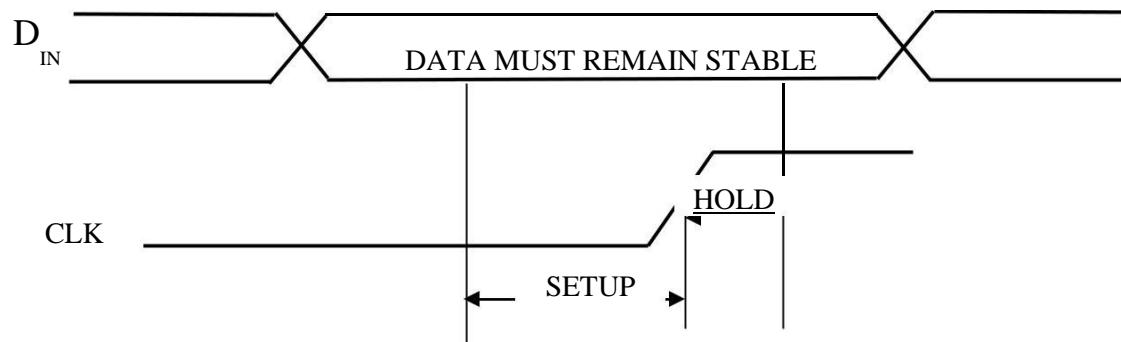
$$T_{\text{LAUNCH_CLOCK}} + T_{\text{CLK-Q_MAX}} + T_{\text{COMB_MAX}} \leq T_{\text{CAPTURE_CLOCK}} - T_{\text{SETUP}}$$

- Hold Time
 - Hold time is the minimum amount of time the data signal should be held steady after the clock event so that the data are reliably sampled

$$T_{\text{LAUNCH CLOCK}} + T_{\text{CLK-Q_MIN}} + T_{\text{COMBO_MIN}} \geq T_{\text{CAPTURE_CLOCK}} + T_{\text{HOLD}}$$

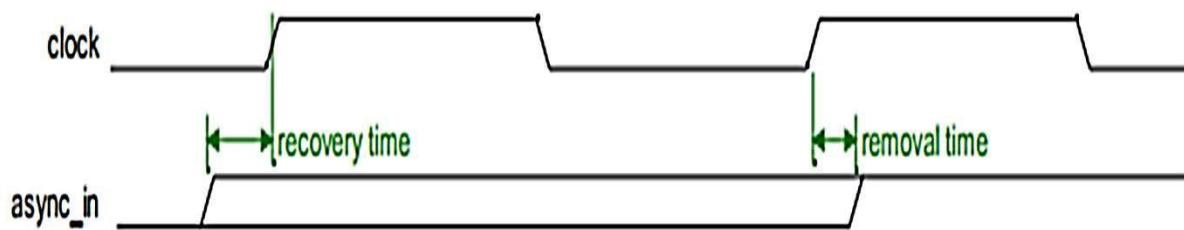
Static Timing Analysis (STA)

- Setup Time and Hold Time Violations
 - If Setup time, T_{SETUP} for a flip-flop and if the data is not stable before T_{SETUP} from the active edge of clock, then there is a Setup Violation at that flip-flop
 - If hold time, T_{HOLD} for a flip flop and if the data is not stable after T_{HOLD} time from the active edge of clock, then there is a hold violation at that flip-flop
 - For a single cycle circuit the signal has to propagate through Data path in one clock cycle



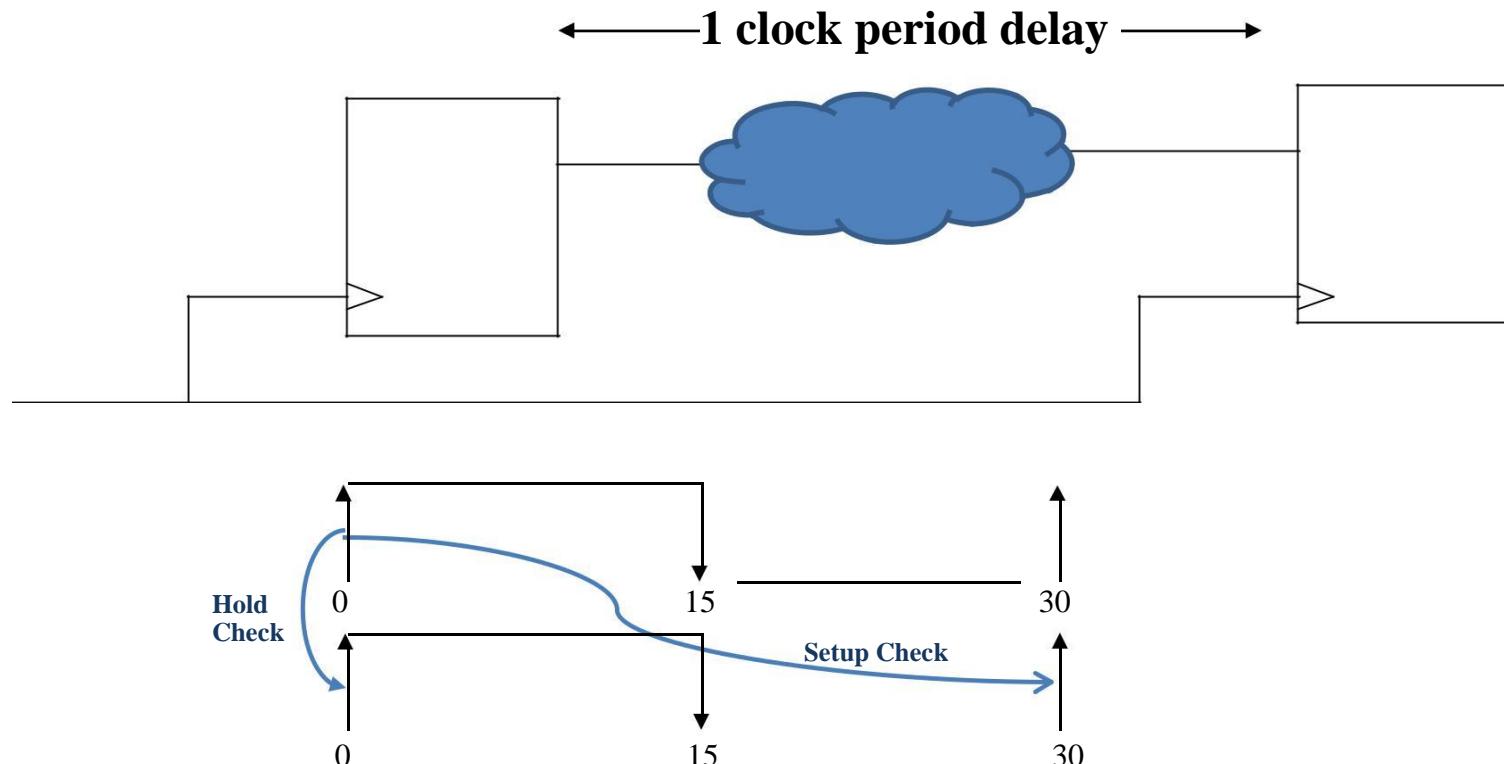
Static Timing Analysis (STA)

- Recovery Time
 - Recovery time is the minimum time that an asynchronous control input pin must be stable after being de-asserted and before the next clock transition (active edge)
- Removal Time
 - Removal time is the minimum time that an asynchronous control input pin must be stable before being de-asserted and before the previous clock transition (active edge)
- Recovery Time and Removal Time Violations
 - This check is to ensure that the asynchronously signal rise/ fall edge is not occurring at the clock edge; it should be some time before or after the clock edge
 - If that violates, then Recovery Time and Removal Time Violations
 - Although a flip-flop is asynchronously SET or CLEAR, the negation from its RESET state is synchronous



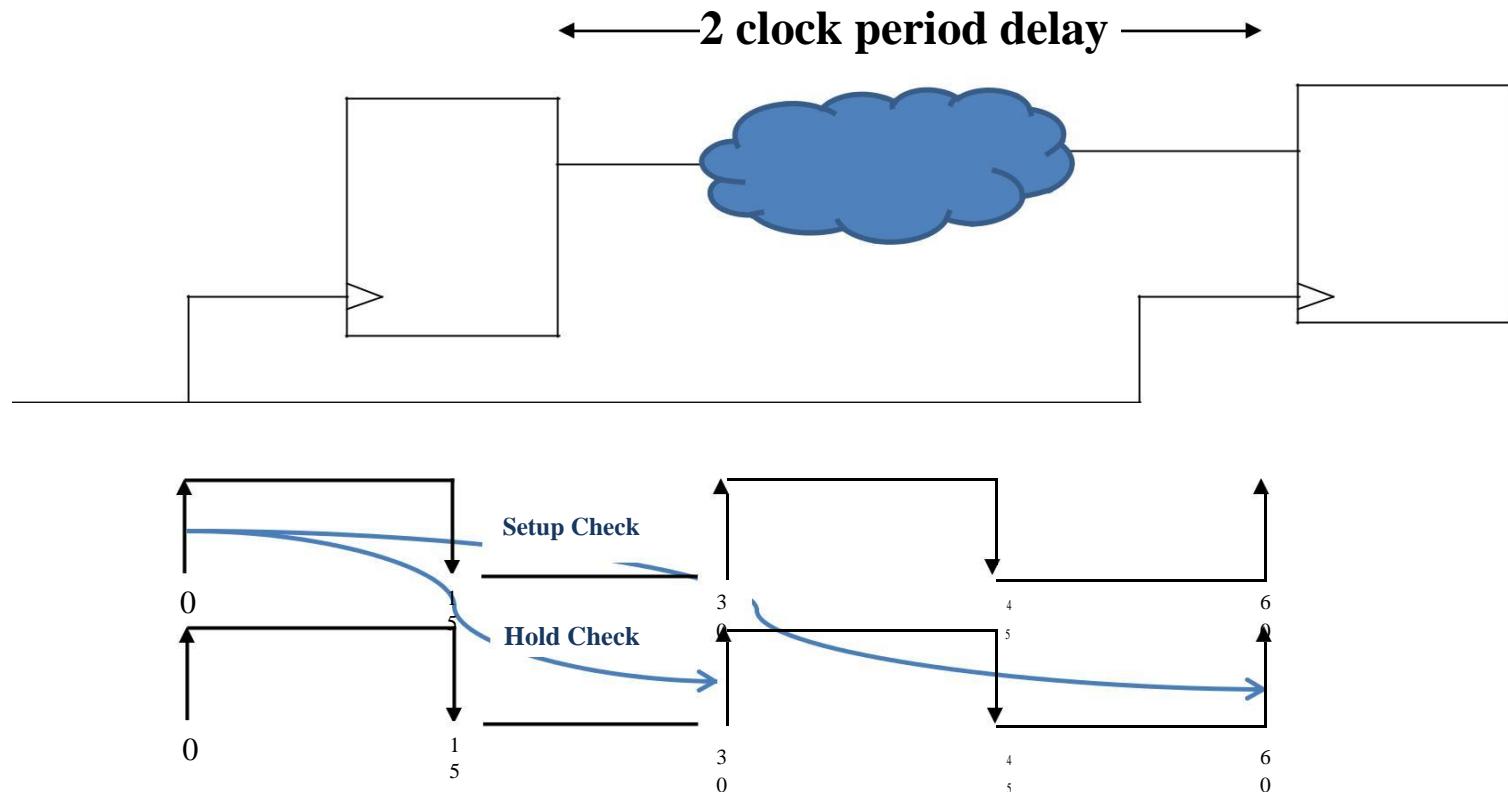
Static Timing Analysis (STA)

- Single Cycle Path
 - Timing path that is designed to take only one clock cycle for the data to propagate from the start point to the endpoint
 - Start point and endpoint are flops clocked by the same clock
 - By default tool will consider all timing paths as single cycle paths



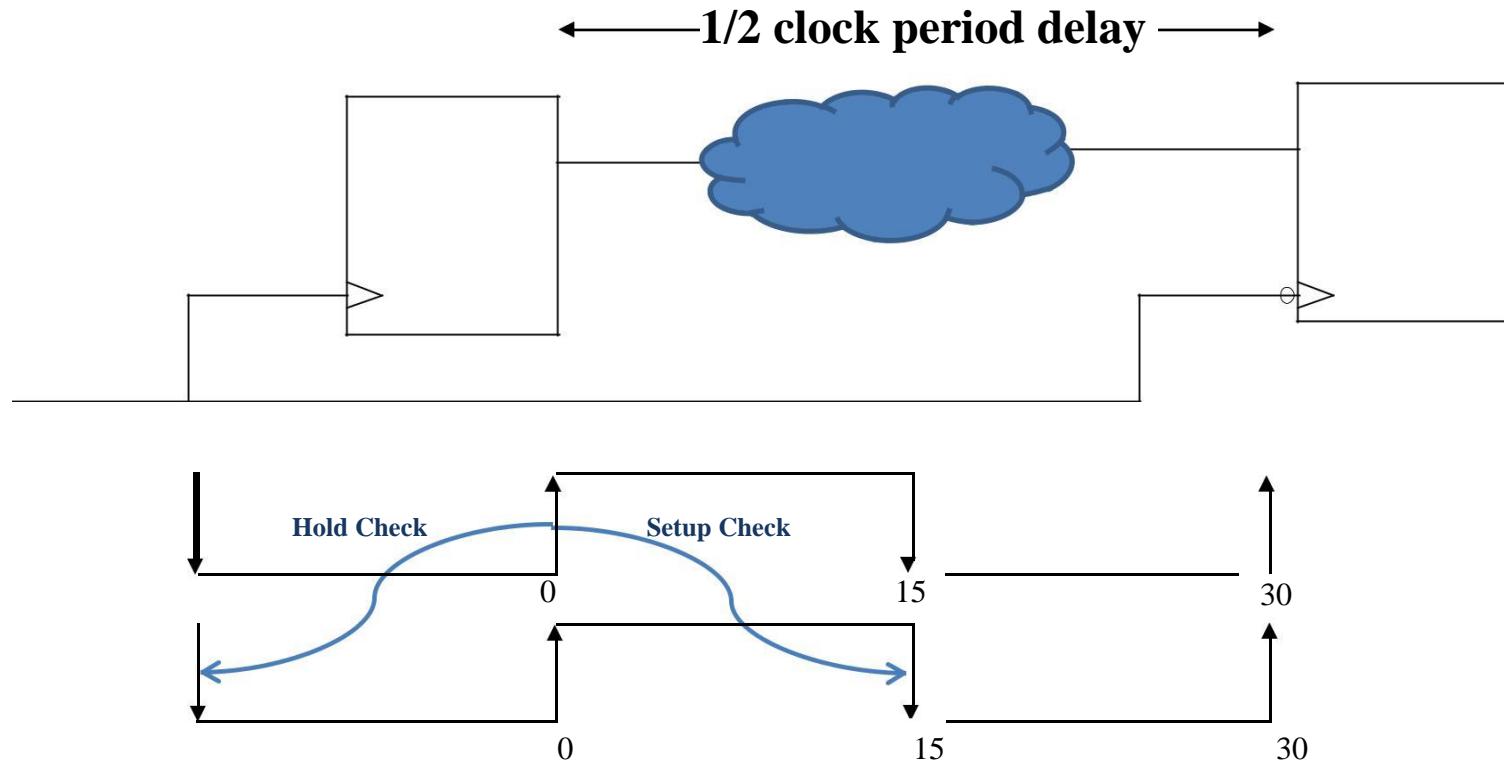
Static Timing Analysis (STA)

- Multi-Cycle Path
 - Timing path that is designed to take more than one clock cycle for the data to propagate from the start point to the endpoint
 - Start point and endpoint are flops clocked by the same clock
 - Need to specify the Launch edge and Capturing edge in SDC



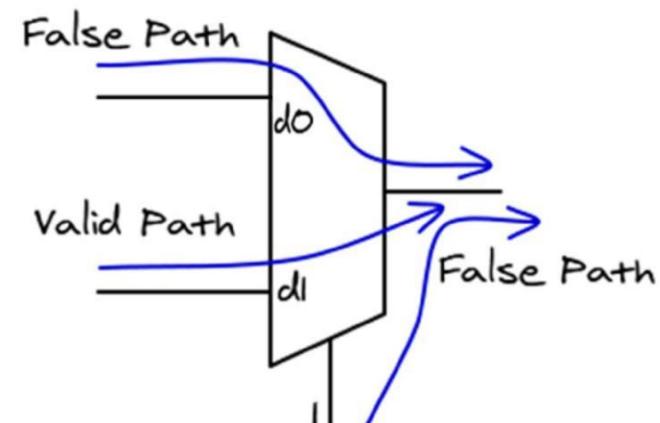
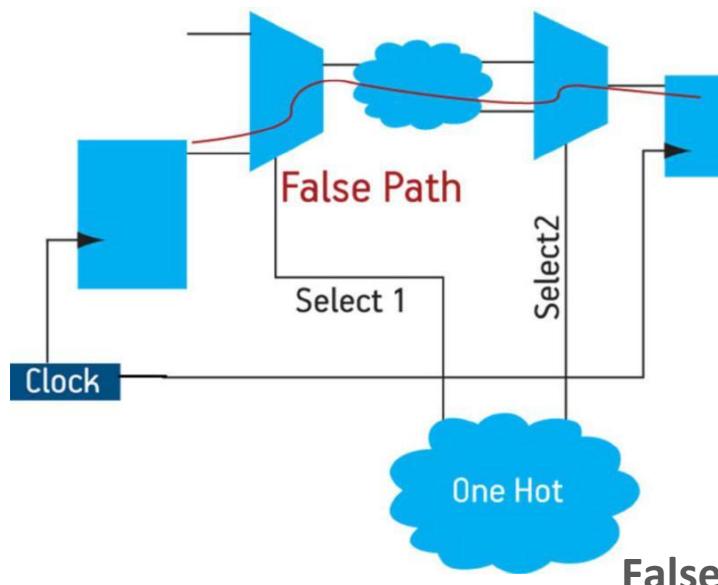
Static Timing Analysis (STA)

- Half Cycle Path
 - Timing path that is designed to take half clock cycle (both of the clock edges) for the data to propagate from the start point to the endpoint
 - Start point and endpoint are flops clocked by the same clock
 - No need to specify the Launch edge and Capturing edge in SDC, since the tool can identify it from the netlist



Static Timing Analysis (STA)

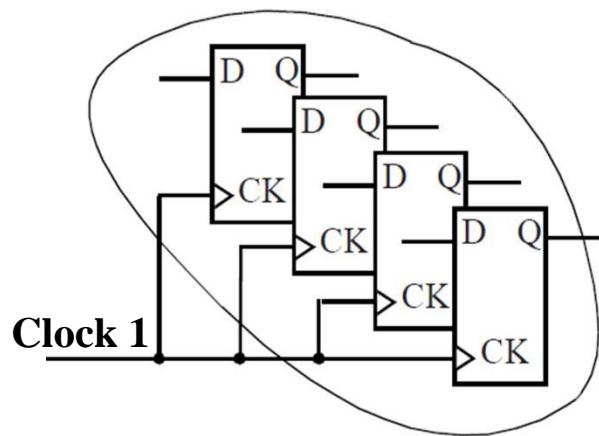
- False Path
 - Physically exist in the design but are Logically/ Functionally inactive/ incorrect path
 - Means no data is transferred from Start Point to End Point
 - The goal in STA is to do timing analysis on all “true” timing paths, so these paths are excluded from timing analysis
 - Similarly timing can be disable for a pin or port or cell where the delay will be computed but won’t report it



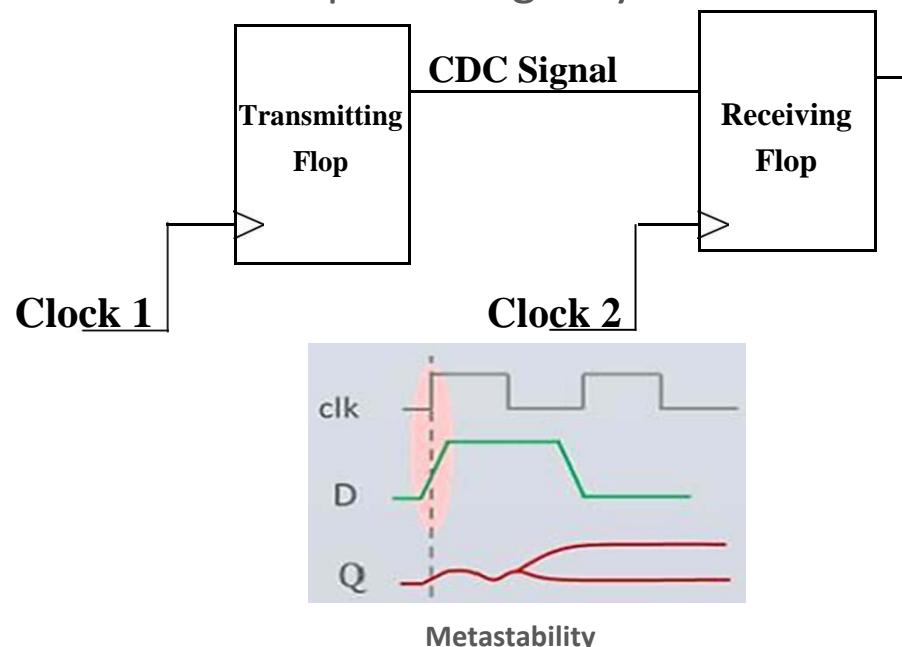
False Path Examples

Static Timing Analysis (STA)

- Clock Domain Crossing (CDC)
 - For designs with Asynchronous Clock Domains, the CDC signal violates the Setup/ Hold window of the receiving clock, resulting in metastability
 - Metastability results in unpredicted values and unpredictable delays
 - Those clocks has to be balanced together else, due to difference in the latency that may lead to timing violations
 - Max. Delay Constraint is used to make CDC paths to get synchronized

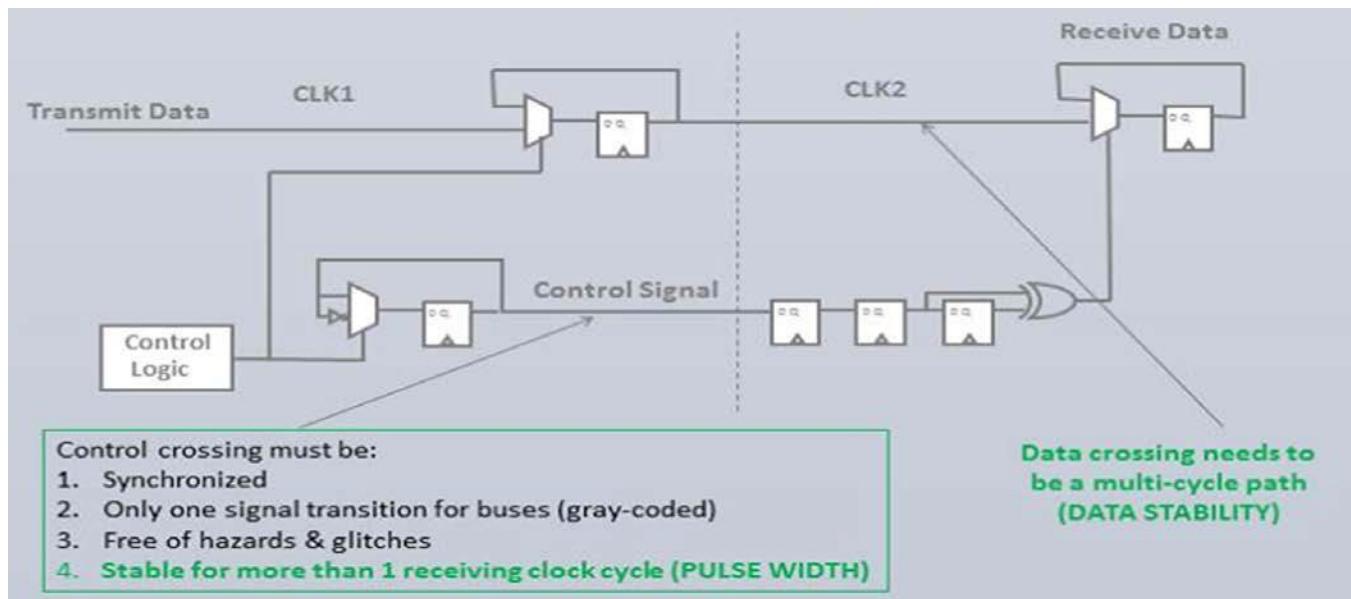


1 Clock Domain



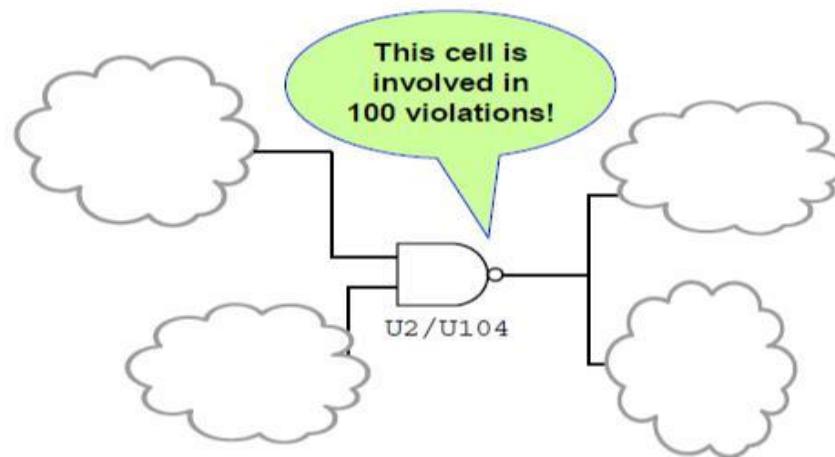
Static Timing Analysis (STA)

- Clock Domain Synchronization Scheme
 - Pulse Width check
 - The control signals is stable for longer than one receive clock period
 - Ensures that data will not be lost due to inadequate width of the control signal
 - Data Stability check
 - The data updated by the transmit domain cannot be captured by the immediately following receive clock edge
 - Ensures that the captured data will not be metastable in the receive domain



Static Timing Analysis (STA)

- Bottleneck Analysis
 - Lists the cells causing the timing violations on multiple paths
 - By identifying and fixing the violation caused by a Bottleneck Cell improved timing can be achieved

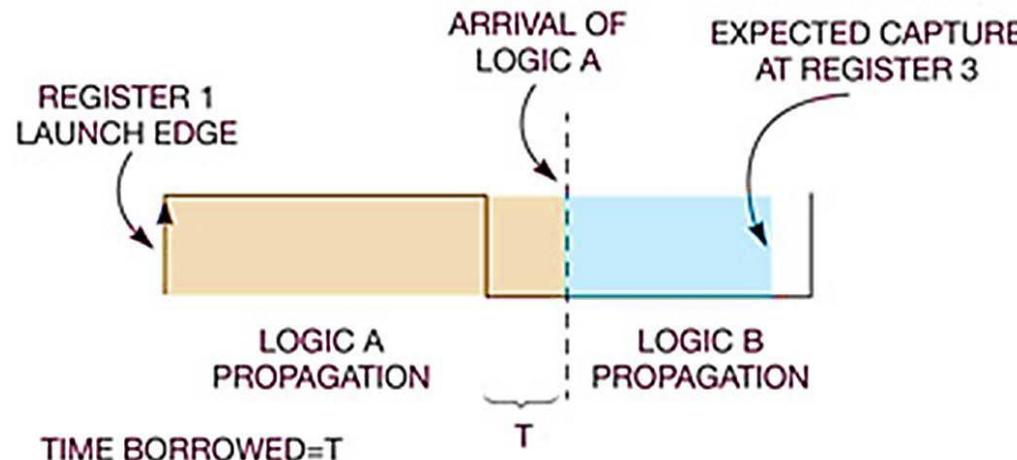


Static Timing Analysis (STA)

- Multi-VT Cells
 - Different threshold voltages are achieved by implanting dopants in different concentration
 - Need Multi-VT Library
 - Sub-threshold leakage varies exponentially with V_T compared to the weaker dependency of delay over V_T
 - If the optimization target is power performance, first use the HVT cells library and then try LVT cells
 - If the optimization target is to meet timing then first use LVT cells and then HVT cells
 - If you swap the capture flop from SVT to LVT or HVT, there will be very minimal setup/hold impact in most flops, it is of zero impact for hold
 - If you swap the launch flop from SVT to LVT or HVT, Setup will be improved and hold will be impacted correspondingly
 - High Voltage Threshold (HVT)
 - Use in non-timing critical paths
 - Use in power critical paths
 - Has low leakage and low speed
 - Low Voltage Threshold (LVT)
 - Use in timing critical paths
 - Use in non-power critical paths
 - Has high leakage and high speed
 - Standard Voltage Threshold/ Regular Voltage Threshold (SVT/ RVT)
 - Medium delay and medium power requirement

Static Timing Analysis (STA)

- Time Borrowing
 - Time Borrowing is basically for Latched based Timing Analysis
 - Edge-triggered flip-flops change states at the clock edges, whereas latches change states as long as the clock pin is enabled
 - In latch based design longer combinational path can be compensated by shorter path delays in the subsequent logic stages
 - The technique of Borrowing Time from the shorter paths of the subsequent logic stages to the longer path is called Time Borrowing or Cycle Stealing

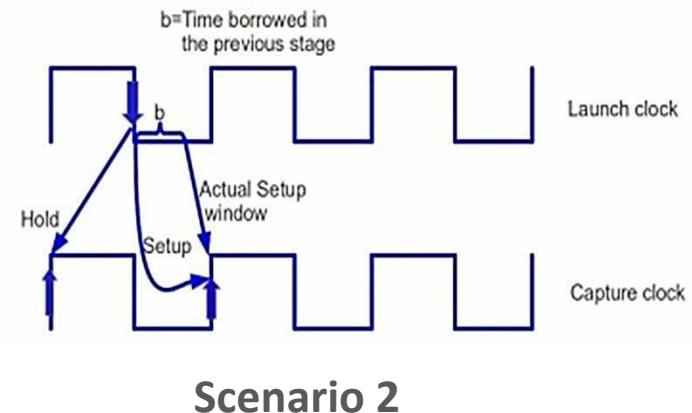
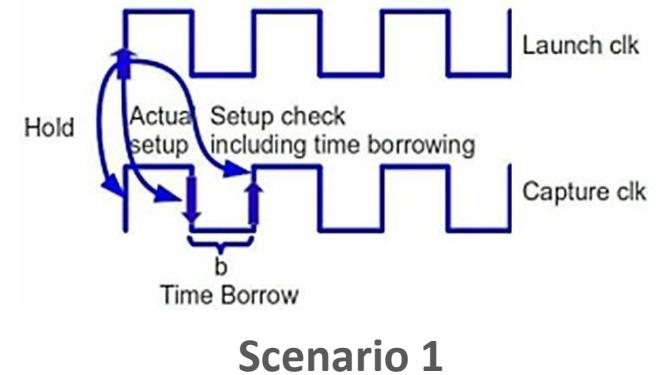
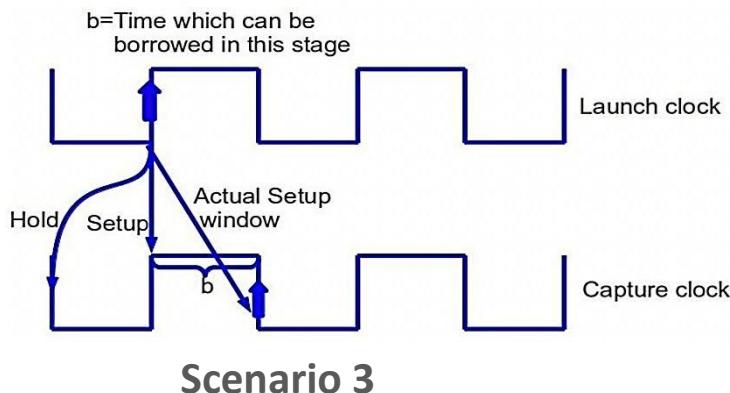


Static Timing Analysis (STA)

- Time Borrowing
 - Time Borrowing typically only affects setup slack calculation since time borrowing slows data arrival times
 - When the clocks of the Launching and Capturing Latches are out of phase, time borrowing is not to happen
 - Timing borrowing can be multistage
 - Maximum Borrow Time:
Clock Pulse Width minus the library Setup Time of the Latch
 - Negative Borrow Time:
Arrival Time minus the clock edge is a negative number, the amount of time borrowing is negative (no borrowing)

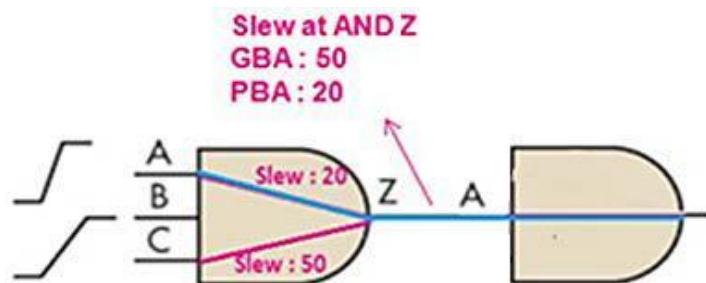
Static Timing Analysis (STA)

- Time Borrowing: Scenarios
 - Scenario 1: When data is launching from a positive edge triggered flip flop and capture is to a negative level sensitive latch
 - Scenario 2: When launch is from a negative level sensitive latch and capture is to a positive edge triggered flip flop
 - Scenario 3: When launch and capture are from positive level sensitive latches



Static Timing Analysis (STA)

- Types of Static Timing Analysis
 - Path Based STA (PBA)
 - First, extract all possible topological paths
 - Next, for each path calculate it's delay and compare it with endpoint (required) value
 - Calculate the Arrival Time (AT) by adding cell delay in timing paths
 - Check all path delays to see if the given Required Arrival Time (RAT) is met
 - Graph Based STA (GBA)
 - Two types of timing data :
 - Arrival times, AT (propagated forward from inputs) Required
 - Arrival Times RAT (propagated from outputs) Slack is calculated on every design element: Slack = RT – AT

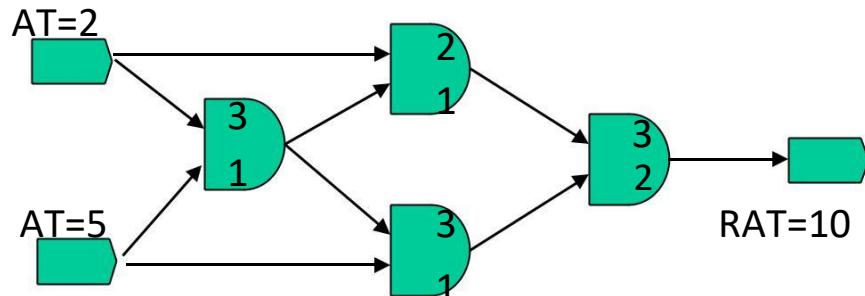


Static Timing Analysis (STA)

Path Based STA (PBA)	Graph Based STA (GBA)
<ul style="list-style-type: none">• Path specific STA• Wont use worst skew• Intensive computation required• Less Pessimistic• More accurate• Timing constraints will be checked at end points of the timing paths• Not favorable for large no. of paths• PBA select either max. path or min. path• Timing information associated with topological paths (collections of design elements)• Traces every possible timing paths• Always done after GBA	<ul style="list-style-type: none">• Parameter based STA• Wont use worst skew• Not so intensive computations• More Pessimistic• Less accurate compare to PBA• Timing constraints will be checked at each node of the timing paths• Not favorable for large no. of Corners• GBA the max. path alone is Selected• Timing information associated with discrete design elements (ports, pins, gates)• Its incremental; breadth based

Static Timing Analysis (STA)

- Block-based STA vs. Path-based STA (example)



Path-based:

$2+2+3 = 7$ (OK)

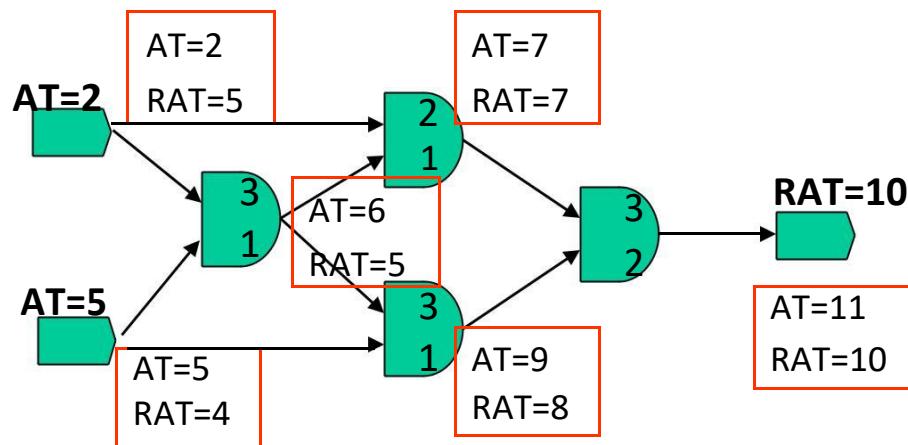
$2+3+1+3 = 9$ (OK)

$2+3+3+2 = 10$ (OK)

$5+1+1+3 = 10$ (OK)

$5+1+3+2 = 11$ (Problem!)

$5+1+2 = 8$ (OK)



Block-based:

Critical path is determined as collection of gates with the same, negative slack:

$$\text{Slack} = \text{RT} - \text{AT}$$

In our case, we see one Critical path with slack = -1

Static Timing Analysis (STA)

- STA Summary Report

timeDesign Summary

Setup mode	all	reg2reg	in2reg	reg2out	in2out	clkgate
WNS (ns):	-7.815	-5.368	-7.815	-0.582	-7.110	N/A
TNS (ns):	-2113.3	-1239.7	-1969.2	-1.269	-38.582	N/A
Violating Paths:	757	708	375	8	6	N/A
All Paths:	1811	1344	819	18	6	N/A
Real	757	708	375	8	6	N/A
DRVs	135 (135)	136 (136)	136 (136)	135 (135)	136 (136)	135 (135)
Nr nets(terms)	370 (14467)	388 (14485)	388 (14485)	370 (14467)	388 (14485)	370 (14467)
Worst Vio	-3.518	-7.767	-7.767	-3.518	-7.767	-3.518
max_cap	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
max_tran	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
max_fanout	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

Density: 78.864%

Routing Overflow: 0.00% H and 0.23% V

Congestion Analysis

Congestion Analysis

- As the Technology advances, millions of transistors can be packed onto the surface of a chip
- Thus the increased circuit density introduces additional Congestion
- Intuitively speaking, Congestion in a layout means too many nets are routed in local regions
 - This causes detoured nets and un-routable nets in Detailed Routing
- Congestion Analysis
 - Routing Congestion Analysis
 - Congestion in general referred to Routing Congestion
 - Routing congestion is the difference between supplied and available tracks A track is nothing but a routing resource which fills the entire Core
 - Placement Congestion Analysis
 - Placement Congestion is due to overlap of Standard Cells, it is called Overlapping rather than called as Congestion
 - Overlapping issue can be fixed by aligning cells to the Placement Grid by Legalization

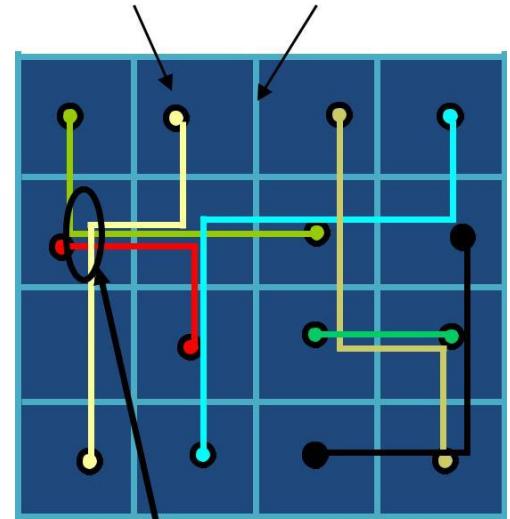
Congestion Analysis

- In recent years, several congestion estimation and removal methods have been proposed
- They fall into two categories: Congestion estimation and removal during global routing stage, and Congestion estimation and removal during Placement stage
- To estimate Congestion, tool does Initial/ Global Routing
- Congestion reports are generated after each Routing stages which shows the difference between supplied and demanded Tracks or G-cells
 - Overflow = Routing Demand - Routing Supply (0% otherwise)
 - Usually starts the initial Target Utilization with 65% to 70%
 - 7/3 in a 2D congestion map : There are 7 routes that are passing through a particular edge of a Global Route Cell (GRC), but there are only 3 routing tracks available. There is an overflow of 4.

Congestion Analysis

- Causes for Routing Congestion
 - Missing Placement Blockages
 - Inefficient floorplan
 - Improper macro placement and macro channels
(Placing macros in the middle of floorplan etc.)
 - Floorplan the macros without giving routing space for interconnection between macros
 - High Cell Density (High local utilization)
 - If your design had more number of AOI/OAI cells you will see this congestion issue
 - Placement of standard cells near macros
 - High pin density on one edge of block
 - Too many buffers added for optimization
 - No proper logic optimization
 - Very Robust Power network
 - High via density due to dense power mesh
 - Crisscross IO pin alignment is also a problem
 - Module splitting

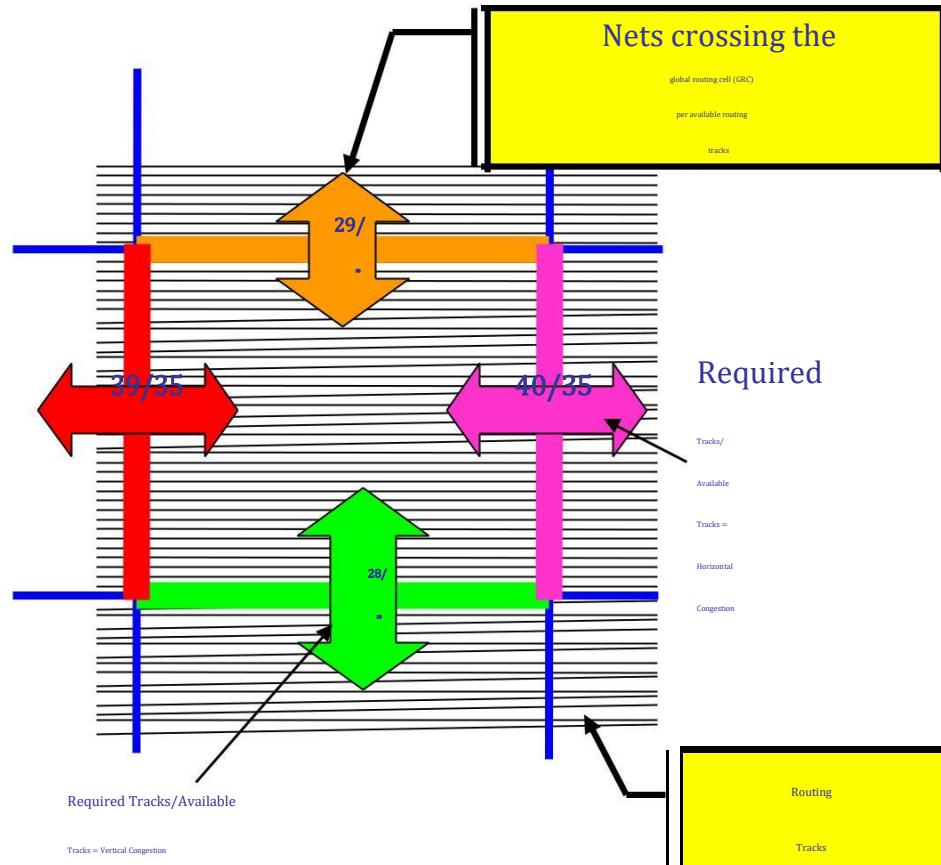
Global Bin Global Bin Edge



Routing demand = 3
Assume routing supply is 1, overflow = 3 - 1 = 2

Congestion Analysis

- Congestion Fixes
 - Add placement blockages in channels and around macro corners
 - Review the macro placement
 - Reduce local cell density using density screens
 - Reordering scan chain to reduce congestion
 - Congestion driven placement with high effort
 - Continue the iterations until good congestion results
 - Density screen is applied to limit the density of standard cells in an area to reduce congestion due to high pin density

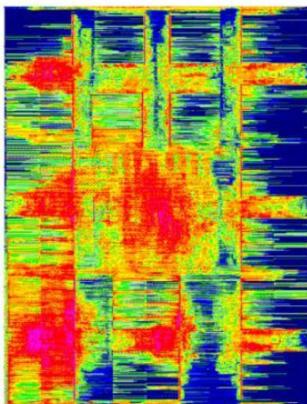


Congestion Analysis

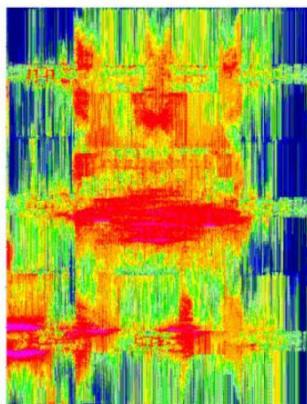
- Routing congestion, results when too many routes need to go through an area with insufficient “routing tracks” to accommodate them

Example Congestion Map

Horizontal Congestion

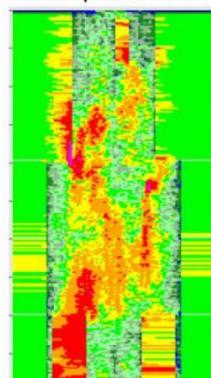


Vertical Congestion

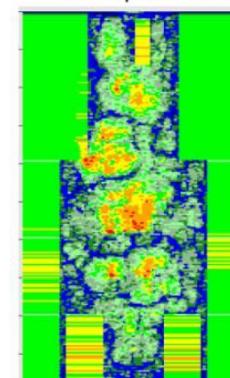


Minimizing Wire Generally Improves Congestion

Bad placement



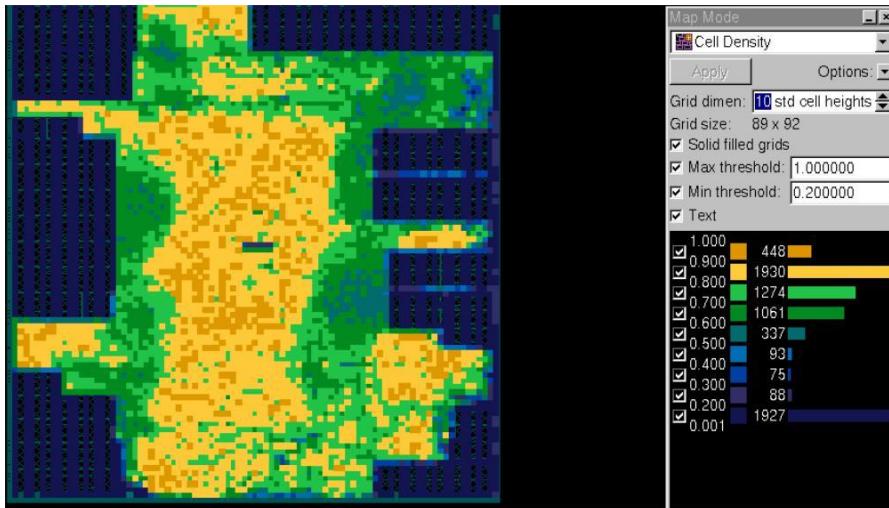
Good placement



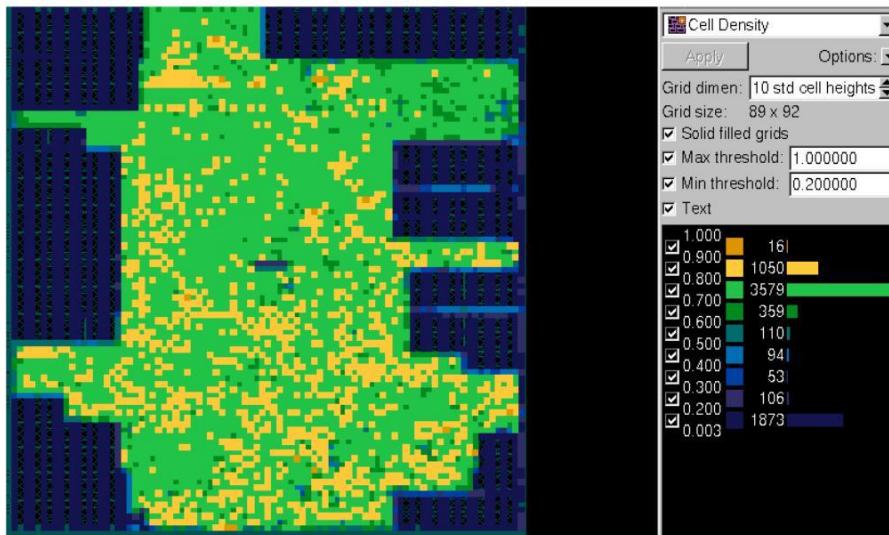
- Two major categories: Global Congestion and Local Congestion
 - Global Congestion: This occurs when there are a lot of chip-level or inter-block wires that need to cross an area
 - Local Congestion: This occurs when the floorplan has macros and other routing blockages that are too close together to get enough routes through to connect to the macros

Congestion Analysis

- Congestion Profiles



Before Fixing



After Fixing

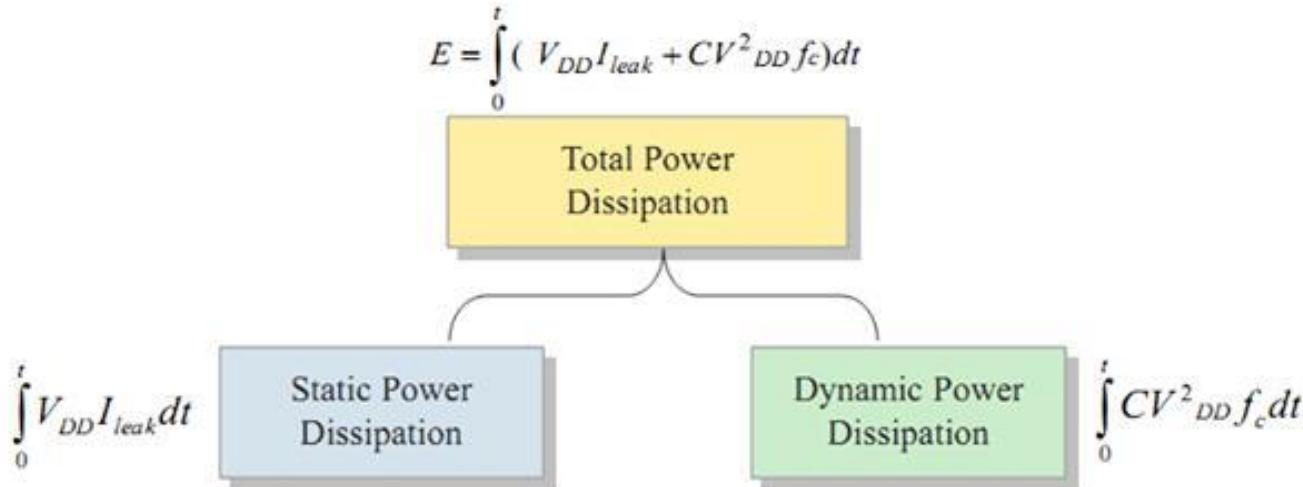
Power Analysis

Power Analysis

- Power Analysis
 - Power Density of the Integrated Circuit increase exponentially with every Technology generation
$$P_{\text{TOTAL}} = P_{\text{DYNAMIC}} + P_{\text{LEAKAGE}}$$
$$P_{\text{DYNAMIC}} = P_{\text{SWITCHING}} + P_{\text{SHORT_CIRCUIT}}$$
 - Leakage Power (Static Power): Leakage at almost all junctions due to various effects
 - Reverse Biased Diode Leakage
 - Gate Induced Drain Leakage
 - Gate Oxide Tunnelling
 - Sub-threshold Leakage
 - Switching Power: When signal change their state, energy is drawn from the power supply to charge up the load capacitance from 0 to VDD
 - Short Circuit Power (Crowbar Power/ EM Rush Through Power): Finite non-zero rise and fall times of transistors which causes a direct current path between the Power and Ground`

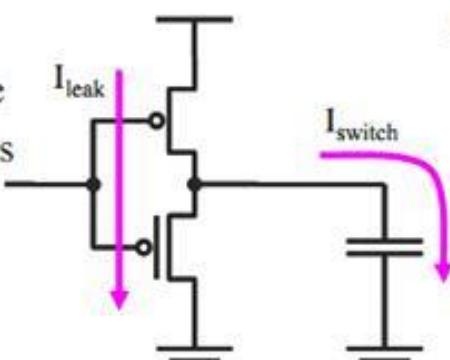
Power Analysis

- Power Analysis
 - Static/ Leakage Power Analysis
 - Dynamic Power Analysis



Minimize I_{leak} by:

- Lower operating voltage
- Fewer leaking transistors



Minimize I_{switch} by:

- Lower operating voltage
- Less switching capacitance
- Less switching activity

- With Shrinking technology Static leakage increases which results in more focus in Reducing leakage power for advanced technologies

Power Analysis

- Static Power/ Leakage Power
 - It is the power consumed when the device is powered up but no signals are changing value (when the transistors are not switching)
 - In CMOS devices, static power consumption is due to leakage
 - Sub-threshold leakage occurs when a CMOS gate is not turned completely OFF

$$I_{SUB} = \mu C_{ox} V_{th}^2 \frac{W}{L} \cdot e^{\frac{V_{GS}-V_T}{nV_{th}}}$$

where

μ - Carrier mobility

C_{ox} - Gate capacitance

V_T - Threshold voltage

V_{GS} - Gate-Source voltage

W and L - Dimensions of the transistor

V_{TH} - Thermal voltage, $kT/q = 25.9\text{mV}$ at room temperature

n - function of device fabrication process (ranges 1.0 -2.5)

Power Analysis

- Static Power Dissipation
 - Leakage Power, is consumed when the transistors are not switching
 - Dependent on the voltage, temperature and state of the transistors
 - $\text{Leakage Power} = V * I_{\text{leak}}$
- Types of Static Leakages
 - Reverse biased diode leakage from the diffusion layers and the substrate
 - Gate Induced Drain Leakage
 - Gate Oxide Tunnelling
 - Sub-threshold Leakage caused by reduced threshold voltages which prevents the Gate from completely turning OFF
- Static Power Reduction Techniques
 - Using Multi V_T cell in the design and optimizing for leakage by replacing high V_T cell for non timing critical paths
 - Power Gating
 - Power Shut-off groups of logic which are not used
 - Voltage Scaling
 - Multi V_{DD} and Voltage Island
 - Multi-threshold CMOS (Back Biasing)

Power Analysis

- Dynamic/ Switching Power
 - Dynamic power is the power consumed when the device is active, when signals are changing values (by switching logic states)
 - Primary source of dynamic power consumption is switching power

$$P_{DYN} = A C V^2 F$$

where,

A is activity factor, i.e., the fraction of the circuit that is switching

C is Load capacitance

V is supply voltage

F is clock frequency

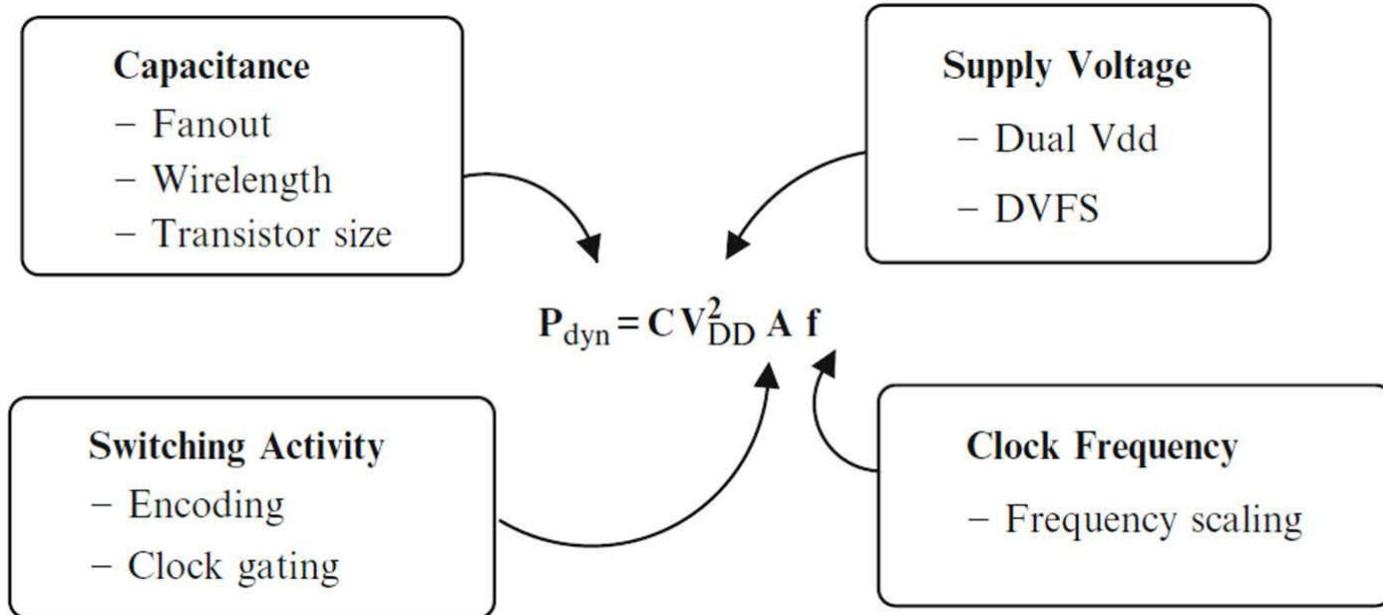
- Dynamic Power Calculation depends on
 - Switching frequency
 - Transition
 - Output load
 - Cell internal power

Power Analysis

- Dynamic Power Dissipation
 - Dynamic power is dissipated any time the voltage on a net changes due to some stimulus
- Types of Dynamic Power
 - Net Switching Power = $(C_{int} * V^2 * f)$
 - Internal Power = $(C_{int} * V^2 * f) + (V * I_{sc})$
 - Short Circuit : = $(V * I_{sc})$ During switching both PMOS and NMOS becomes on which results in a short circuit current
 - Internal Capacitance Loading Power = $(C_{int} * V^2 * f)$ is the power consumed while charging/discharging internal nets

Power Analysis

- Dynamic Power Reduction Techniques

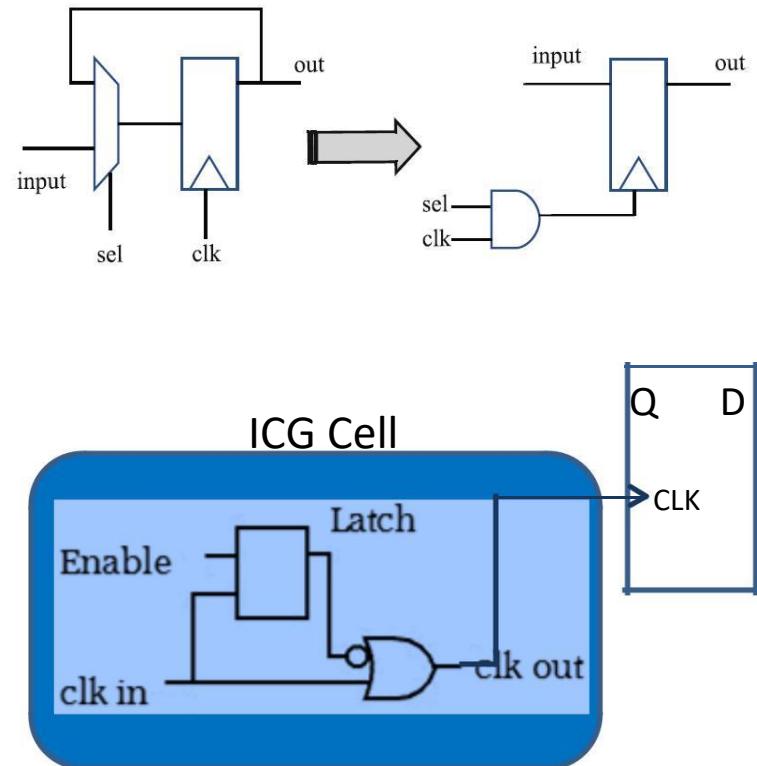


Power Analysis

- Dynamic Power Reduction Techniques

- Clock Gating

- Architectural Technique to reduce Dynamic Power along the Clock Path
 - Clock gates should be placed at the Root of the Clock
 - Results in small delay, more area and makes the design complex
 - Clock Gating logic is generally in the form of "Integrated clock gating" (ICG)
 - Sequential clock gating is the process of extracting/propagating the enable conditions to the upstream/downstream sequential elements, so that additional registers can be clock gated
 - As the granularity on which you gate the clock of a synchronous circuit approaches zero, the power consumption of that circuit approaches that of an asynchronous circuit: the circuit only generates logic transitions when it is actively computing



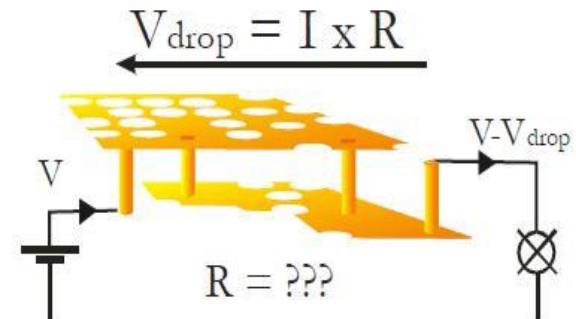
IR Drop Analysis

IR Drop Analysis

- IR Drop
 - The voltage that gets to the internal circuitry is less than that applied to the chip, since every metal layer offers resistance to the flow of current
 - When a current, I passes through a conductor with resistor R , it exhibits a voltage drop V which is equal to the resistance times the current,

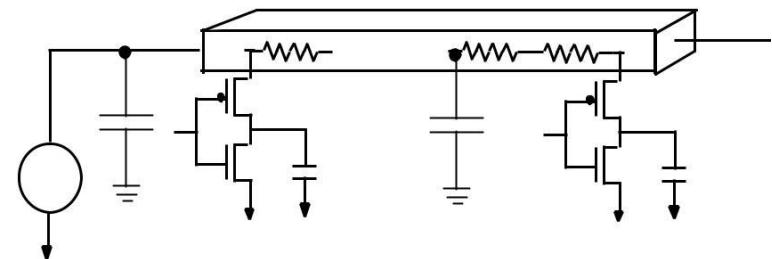
Ohm's law, $V=IR$

- IR Drop is defined as the average of the peak currents in the power network multiplied by the effective resistance from the power supply pads to the center of the chip
- IR Drop is a reduction in voltage that occurs on both Power and Ground networks
- IR Drop Analysis ensures that Power Delivery Network (PDN) is robust, and that your system will function to specification
- IR Drop is determined by the current flow and the supply voltage
- As distance between supply voltage and the component increases the IR Drop also increases



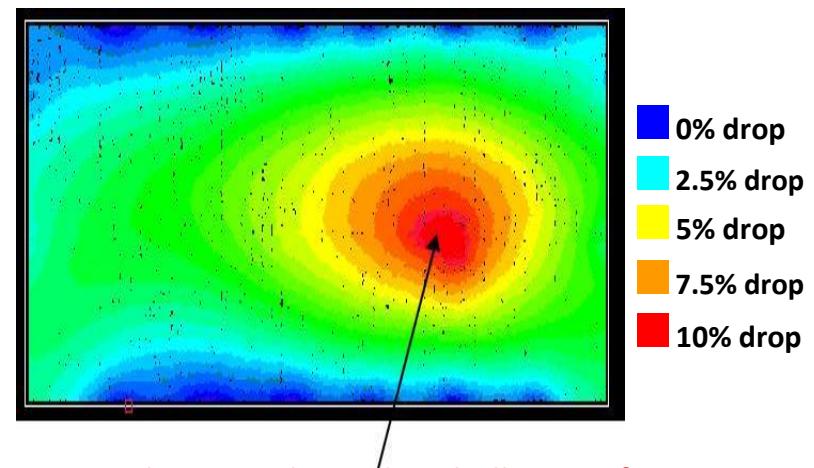
IR Drop Analysis

- IR Drop Analysis
 - IR Drop Analysis will compute the actual I_{DD} and I_{SS} currents, because these values are time-dependent
 - IR Drop Analysis will compute Global IR drop which is important and more accurate, but cannot be compute separately (parallel) for smaller blocks, which may led to bigger run time
 - Local IR Drop
 - IR Drop become a local phenomenon when a number of gates in close proximity switches at once
 - Local IR Drop can also be caused by a higher resistance to a specific portion of the Grid
 - Global IR Drop
 - IR Drop is a global phenomenon when activity in one region of a chip causes an IR Drop in other regions
 - In a well-meshed power grid with equally distributed currents, the power grid typically has a set of equipotential IR Drop surfaces that form concentric circles centered in the middle of the chip
 - So the center of the chip usually has the largest IR Drop or the lowest supply voltage
 - Peak IR Drop is much larger than the Average IR Drop
 - Peak IR Drop happens in the worst-case switch patterns of the gates



IR Drop Analysis

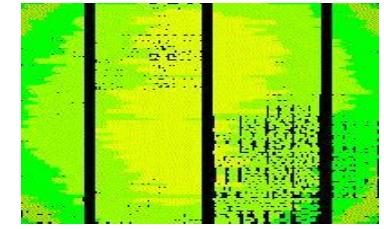
- Types of IR Drop
 - Static IR Drop
 - Static IR drop is average voltage drop for the design
 - The average current depends totally on the time period
 - Static IR drop was good for signoff analysis in older technology nodes where sufficient natural decoupling capacitance from the power network and non-switching logic were available
 - Localized switching is only considered
 - Only be a few % of the supply voltage
 - Can be reduced by lowering the resistance of Supply and Signal Paths
 - Static IR Drop methodology
 - Extract power grid to obtain R
 - Select stimulus
 - Compute time averaged power consumption for a typical operation to obtain I
 - Compute: $V = IR$
 - Non time-varying



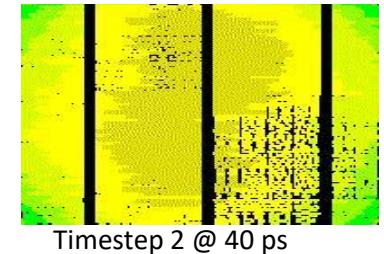
- Typical static voltage drop bulls-eye of an appropriately constructed power grid
- But 10% static voltage drop is very high

IR Drop Analysis

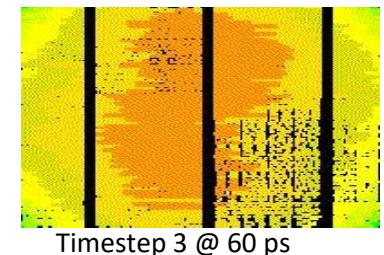
- Types of IR Drop
 - Dynamic IR Drop
 - When large amounts of circuitry switch simultaneously, causing peak current demand
 - Dynamic IR drop is mainly due to Instantaneous Voltage Drop (IVD) and it can be controlled by inserting Decap Cells in the Power network
 - Dynamic IR drop depends on switching activity and switching time of the logic and is less dependent on the a clock period
 - Instantaneous current demand could be highly localized and could be brief within a single clock cycle (a few hundred ps)
 - Vector dependent, so VCD-based analysis is required
 - Dynamic IR Drop methodology
 - Extract power grid to obtain on-chip R and C
 - Include RLC model of the package and bond wires
 - Select stimulus
 - Compute time varying power for specific operation to obtain $I(t)$
 - Compute $V(t) = I(t)*R + C*dv/dt*R + L*di/dt$



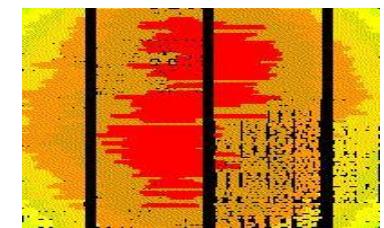
Timestep 1 @ 20 ps



Timestep 2 @ 40 ps



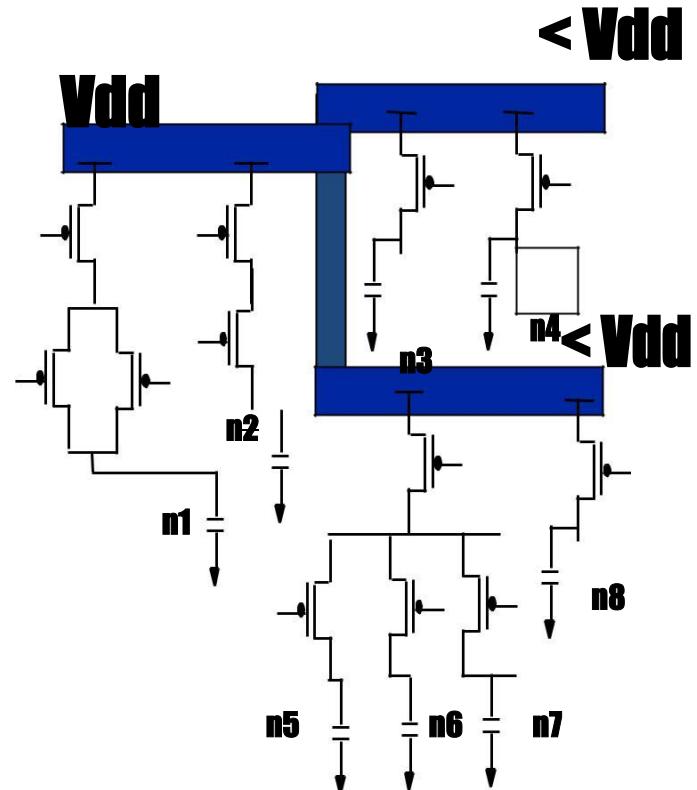
Timestep 3 @ 60 ps



Timestep 4 @ 80 ps

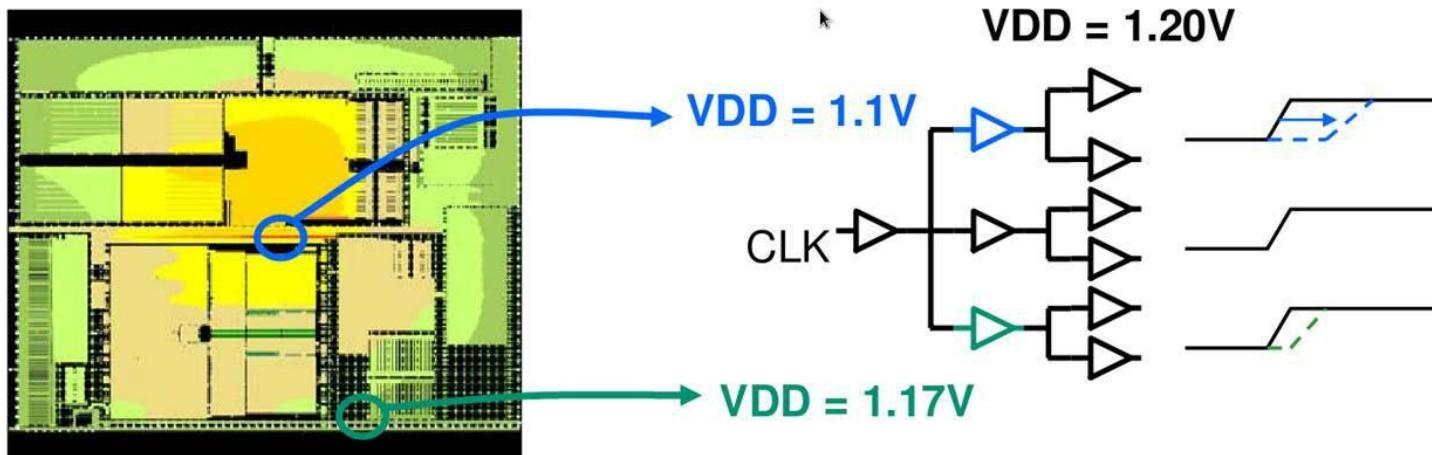
IR Drop Analysis

- IR Drop: Reasons
 - Improper placement of Power/Ground Pads
 - Wrong Block placement
 - Bad global power routing
 - Insufficient Core Ring, Power Strap width
 - Lesser no of Power Straps
 - Missing Vias
 - Insufficient number of Power Pads
- IR Drop: Robustness Checks
 - Open circuits
 - Missing or insufficient Vias
 - Current Density violations
 - Insufficient Power Rail design



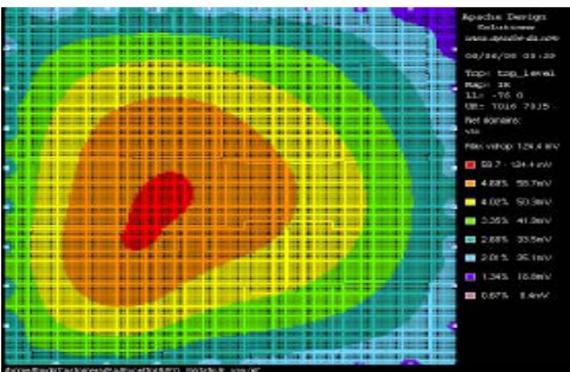
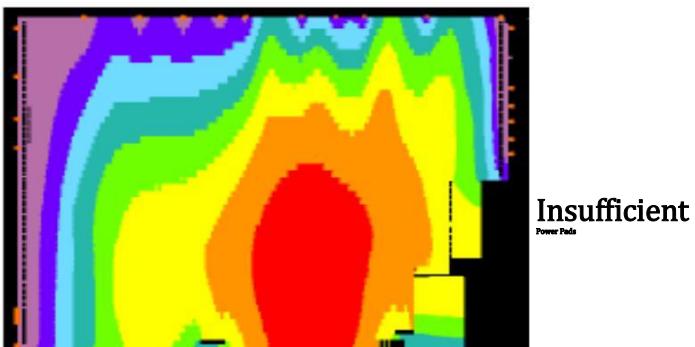
IR Drop Analysis

- IR Drop: Impacts
 - IR Drop Analysis confirms that the worst case voltage drop (which is considered for the worst corner for timing) on a chip meets IR Drop targets
 - Impacts in Timing
 - If this Voltage Drop is too severe, the circuit will not get enough voltage, resulting in the malfunction or timing failure
 - If IR Drop increases Clock Skew then it will result in Hold Time Violations
 - If IR Drop increases Signal Skew then it will result in Setup Time Violations

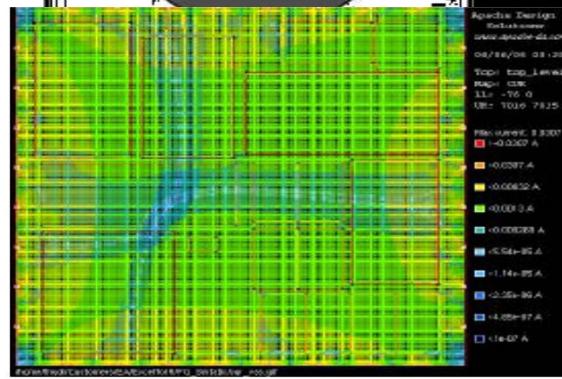
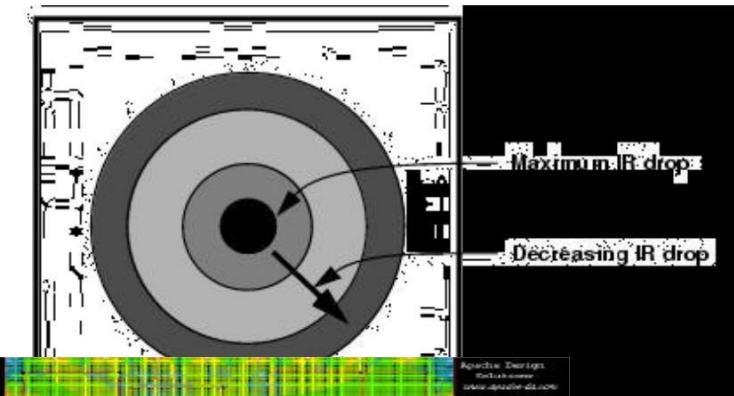


IR Drop Analysis

- IR Drop Plot
 - Power grid has a set of equipotential surfaces that form concentric circles centered in the middle of a block



IR Drop Plot – Before fixing



IR Drop Plot – After fixing

Courtesy: eetimes.com

IR Drop Analysis

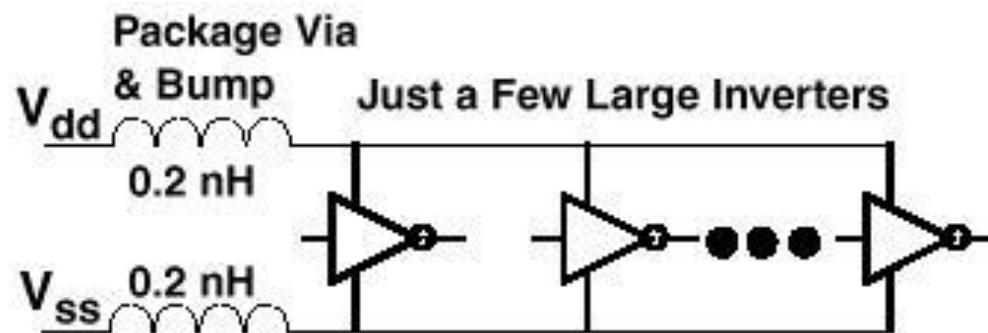
- IR Drop: Remedies
 - Stagger the firing of buffers (bad idea: increases skew)
 - Use different power grid tap points for clock buffers (but it makes routing more complicated for automated tools)
 - Use smaller buffers (but it degrades edge rates/increases delay)
 - Rearrange blocks
 - More V_{DD} pins
 - Connect bottom portion of grid to top portion
 - Distributing supplies symmetrically on the chip
 - Lowering the resistance of Supply and Signal Paths by making supply wires thicker in dimensions than signal wires, $R = \rho \cdot L / A$
 - Decap insertion can solve Dynamic IR drop, at later stage of the design
 - Amount of decap depends on:
 - Acceptable ripple on V_{DD}-V_{SS} (typically 10% noise budget)
 - Switching activity of logic circuits (usually need 10X switched cap)
 - Current provided by power grid (di/dt)
 - Required frequency response (high frequency operation)

IR Drop Analysis

- $I\frac{d}{dt}$ Effects
 - In addition to IR drop, power system inductance is also an issue
 - Inductance may be due to power pin, power bump or power grid
 - Overall voltage drop is:

$$V_{\text{drop}} = IR + L\frac{di}{dt}$$

- As a solution to this effect, distribute decoupling capacitors (decaps) liberally throughout design



Physical Design Essentials

Outline

- **Issues in ASIC Physical Design**
 - Design Parasitics, Latch-up, Electro-Static Discharge, Electromigration, Antenna Effect, Cross Talk, Soft Errors, Self-Heating
- **Cells in ASIC Physical Design**
 - Standard Cells, ICG Cells, Well taps, End caps, Filler Cells, Decap Cells, ESD Clamps, Spare Cells, Tie Cells, Delay Cells, Metrology Cells
- **IO Design**
- **Delay Models**
 - Interconnect Delay Models
 - Cell Delay Models
- **Engineering Change Order (ECO)**
- **Types of Standard Cell Libraries**

Issues in ASIC Physical Design

ASIC Design Parasitics

- ## Parasitic Resistance

 - If resistance increases delay also get increases (Delay= R.C)
 - As technology shrinks interconnects also shrinks and thus wire resistance will get increase
 - To avoid this situation we will increase the height of interconnects
- ## Parasitic Capacitance

 - As technology shrinks height of nets getting increase, so sidewall capacitance is increasing
 - As technology shrinks the dielectric become thinner, the capacitance will get increases
 - To reduce the capacitance, minimize the surface area which can be in common
 - So we keep the adjacent metal layers vertical and horizontal in designs
- ## Parasitic Inductance

 - Mutual inductance affects: High frequency bus
 - Self-inductance affects: Clock nets
 - To limit inductance, we provide current return paths for high frequency signals
 - Separation and Shielding are the possible remedies
 - The rule of thumb has been that when the length of the signal path was long enough to become some percentage of a wavelength that the line itself starts to become a concern for signal integrity
 - Prominent above 500MHz & below 130nm for long wire nets & Power/Clock lines

Latch-up

- What is Latch-up?
 - Phenomenon occur with CMOS/ BiCMOS circuits
 - Generation of a low-impedance path between the VDD supply and the Ground
- Reason for Latch-up
 - Due to regenerative feedback between the parasitic PNP and the NPN Transistors
- Impact in the design
 - PN Junctions can produce Parasitic Thyristor
 - Forms by PNP/ NPN structures
 - Considerable input current is necessary to activate
 - Thyristor formed from parasitic transistors is triggered and generates short-circuit between VDD & GND
 - Results in self destruction/ system failure due to the direct connection between VDD & GND

Latch-up

- NPN Transistor

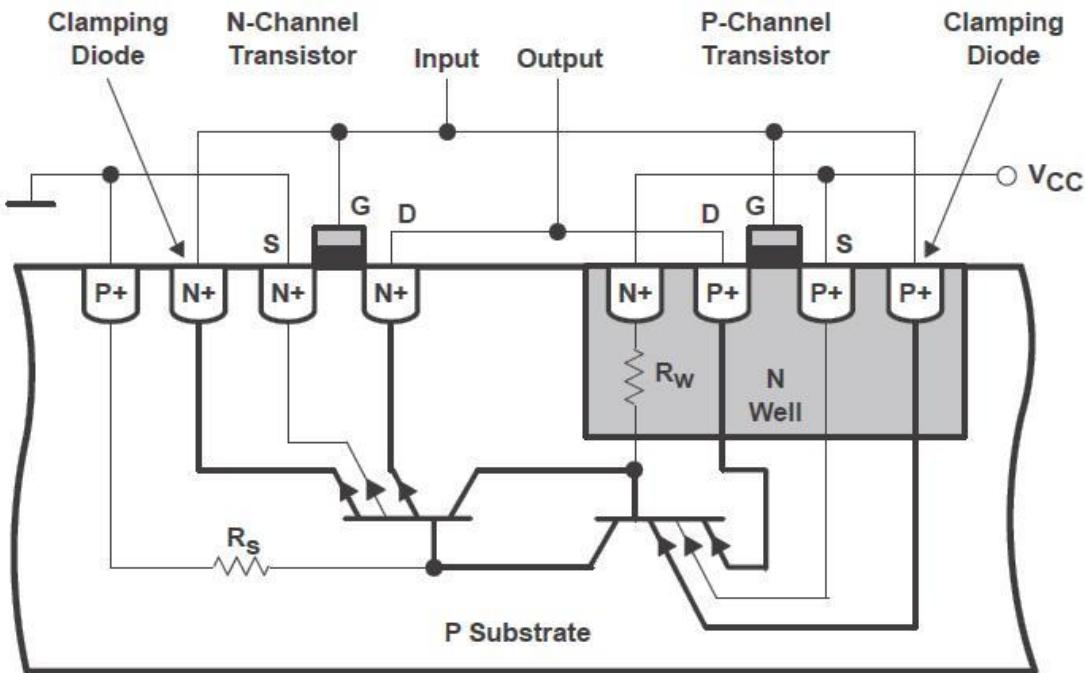
- Emitter – drain /source of the N-channel MOSFET
- Base – P Substrate
- Collector – N Well in which the complementary P- channel MOSFET is located

- PNP Transistor

- Emitter – drain /source of the P-channel MOSFET
- Base – N Well in which the complementary P-channel MOSFET is located
- Collector – P Substrate

- Thyristor/SCR/PNPN diode

- Anode – drain /source of the P- channel MOSFET
- Cathode – drain /source of the N-channel MOSFET
- Gate – P Substrate



Latch-up

— Remedies for Latch-up

- Latch-up resistant CMOS process

Reduces the gain of parasitic transistors (use of Si starting material with a thin epitaxial layer on top of a highly doped substrate)

Increase the holding voltage above VDD supply

Increase the dopant concentration of substrate & well (but will lead to higher V_T)

Retrograde well structure (Highly doped area at bottom and lightly doped at top)

- Layout techniques

Sufficient space between NMOS & PMOS

This reduces the current gain of the parasitic transistors

limited success because can be increased only to a certain limit

Reduce R_S and R_W by keeping Substrate & Well contacts as close as possible

Place substrate contacts as close as possible to the source connection of transistors connected to the supply rails (VSS n-devices, VDD p-devices)

This reduces the value of $R_{SUBSTRATE}$ and R_{WELL}

A very conservative rule would place one substrate contact for every supply (VSS or VDD) connection

In Std. Cells based designs a common Well Tap is taking out as per the need

Guard Rings

Gain of transistors is reduced (in analog designs)

ESD

\emdash Electrostatic Discharge (ESD)

- When two non-conducting materials rub together, then are separated, opposite electrostatic charges remain on both which attempt to equalize each other
- A transient discharge of static charge that arises from either human handling or a machine contact

- Reasons for Electrostatic Discharge

- Thin & vulnerable Gate Oxide of the CMOS makes ESD protection essential for CMOS
- Can be due to inductive or capacitive coupling
- ESD can occur during the removal of extra metal by rubbing in metallization process
- ESD occurs so rapidly that normal GND wires exhibits too much inductance to drain the charge before it can do damage

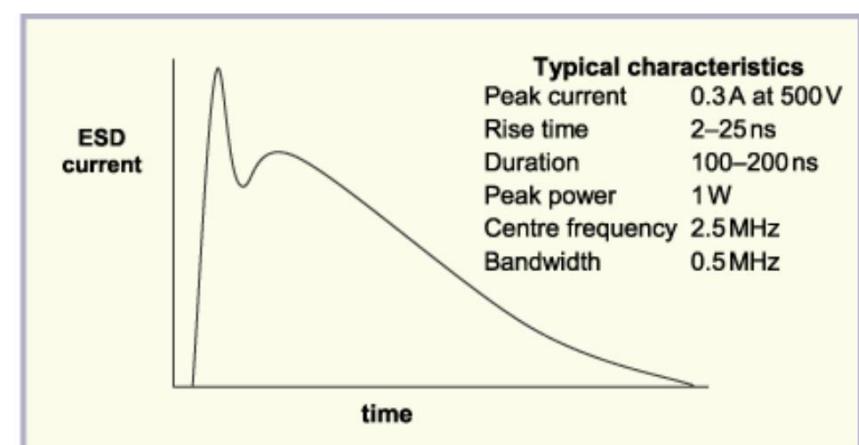
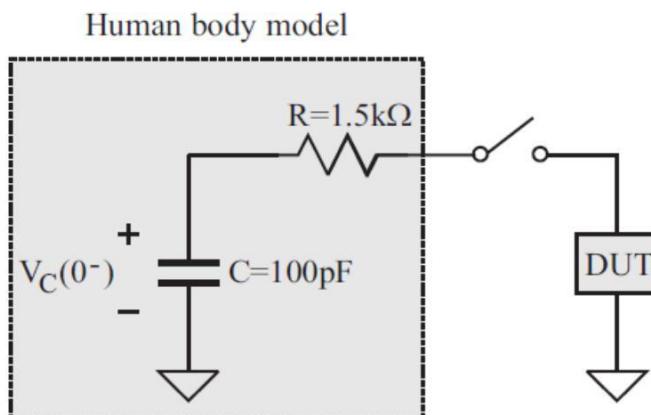
\emdash Impact on the design

- ESD can also burn-out device/ interconnect if thermally initiated
- PMOS is stronger than NMOS in ESD protection, because snap back holding voltage is lower for NMOS

ESD

— Human Body Model (HBM)

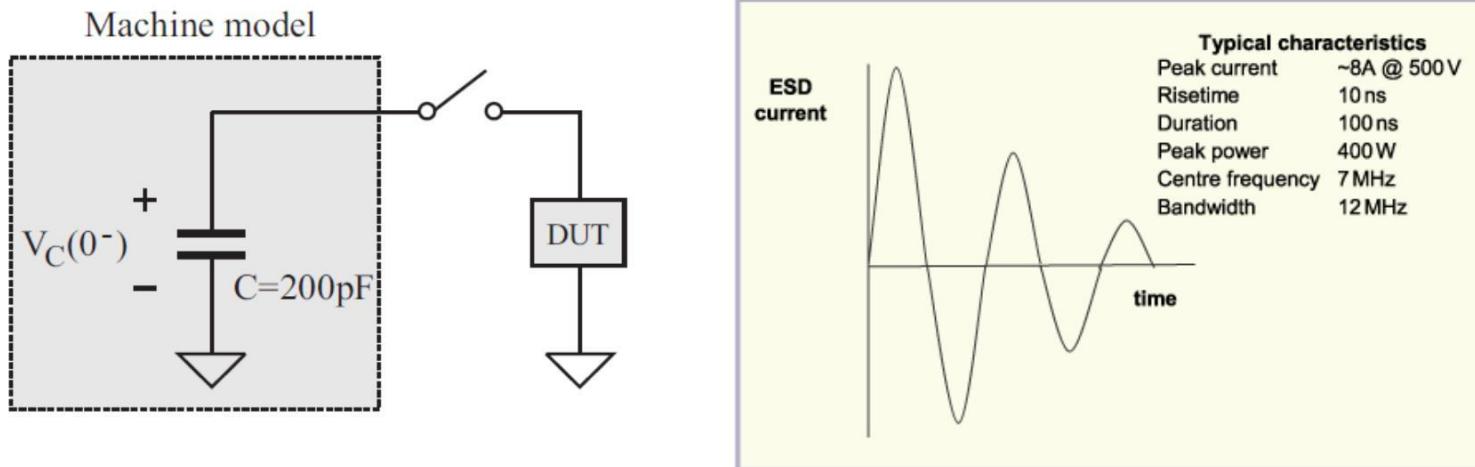
- The actual capacitance of the human body is between 150 pF and 500 pF & the internal resistance of the human body ranges from a few kilo-ohms to a few hundred
- Peak current $\approx 1.3A$, rise time $\approx 10\text{-}30\text{ns}$



ESD

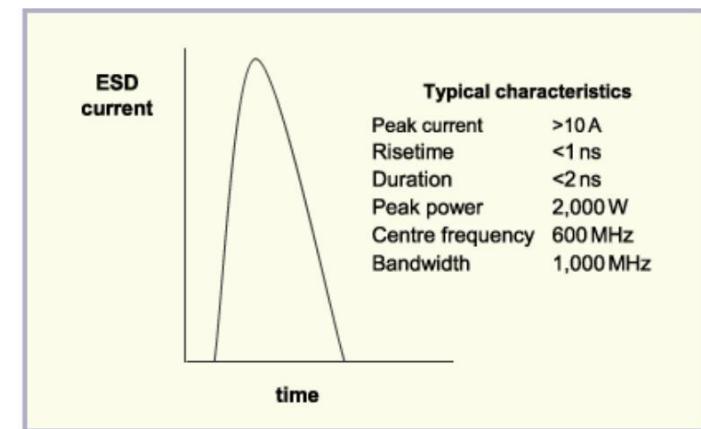
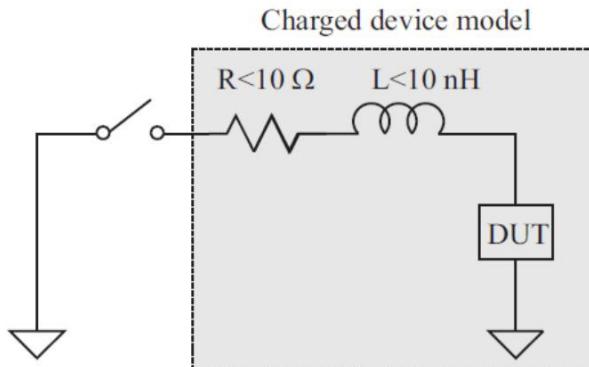
— Machine Model (MM)

- MM models the ESD of manufacturing / testing equipment
- Peak current $\approx 3.7A$, rise time $\approx 15-30\text{ns}$, bandwidth $\approx 12 \text{ MHz}$
- ESD stress caused by charged machines is severe because of zero body resistance
- MM ESD withstand voltage is typically one tenth of HBM
- Most ESD protection circuits can only protect HBM and MM



ESD

- Charged Device Model (CDM)
 - CDM models the ESD of charged integrated circuits
 - As more and more circuits and functions getting integrated causes large Die size which provides large body capacitance which in turn stores charges for CDM in the body of IC
 - Inductance in the model is mainly due to the inductance of bond wires
 - Gate oxide breakdown is the signature failure of CDM stress, in contrast to the thermal failure signature of HBM and MM stress
 - CDM stress is the most difficult ESD stress to protect against since fastest transient and has the max. peak current
 - Peak current $\approx 10A$, rise time $\approx 1ns$



ESD

‐ ESD Protection

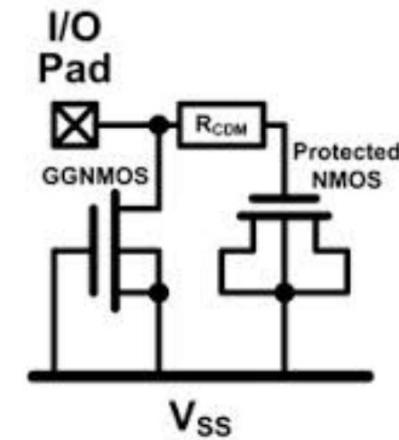
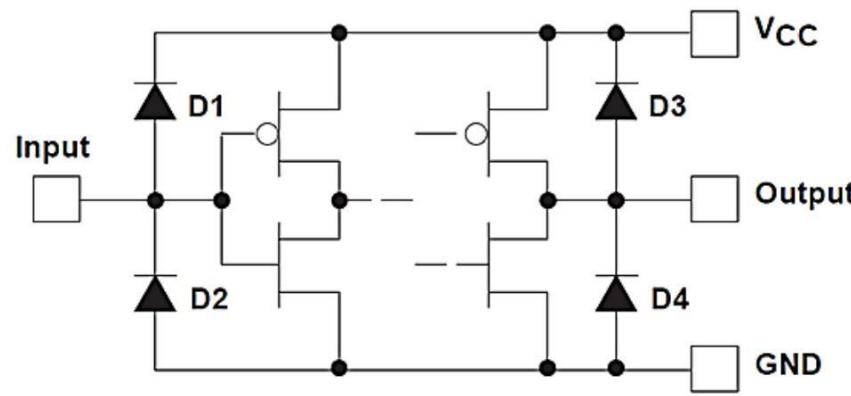
- The integration of Clamping Diodes

Limits the dangerous voltages and conduct excess currents into regions of the circuit that are safe

- The Protection Diodes

‐ Oriented to be blocking in normal operation

‐ Situated between the connection to the component to be protected and the supply voltage lines safe regions consist primarily of the supply-voltage connections

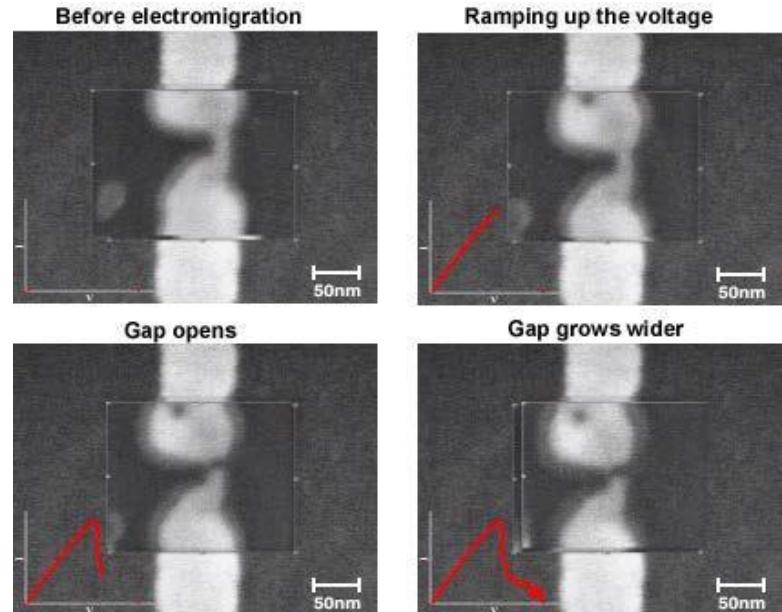


Electromigration

- Electromigration (EM)
 - A failure mechanism caused by high energy electrons impacting the atoms in a material and causing them to shift position
 - Enhanced and directional mobility of atoms under the influence of an electric field

- Reason for Electromigration

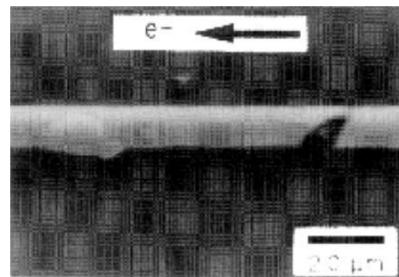
- Forms a positive feedback path where EM will cause an atom to move down a wire, slightly narrowing the wire width at that location and increasing the current density
- This increased current density then further increases electromigration, causing more atoms to be displaced Transport of material caused by the gradual movement of ions in a conductor due to the momentum transfer between conducting electrons & diffusing metal atoms
- It is most problematic in areas of high current density
- Significant as size decreases & is most significant for unidirectional (DC) current



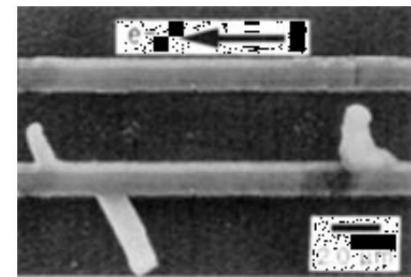
Electromigration

Impact in the design

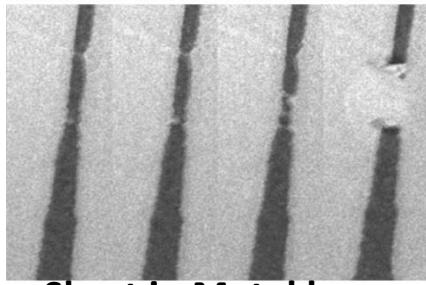
- Excessive EM leads to open (voids) & short circuits (Hillocks) and thus decreases the reliability of the chip
- Approaching life time of device faster
- Increased power consumption
- Higher on-chip temperatures
- High Voltage operation
- High frequency switching



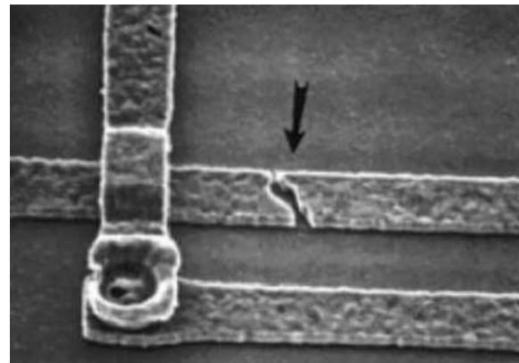
Voids/ Open



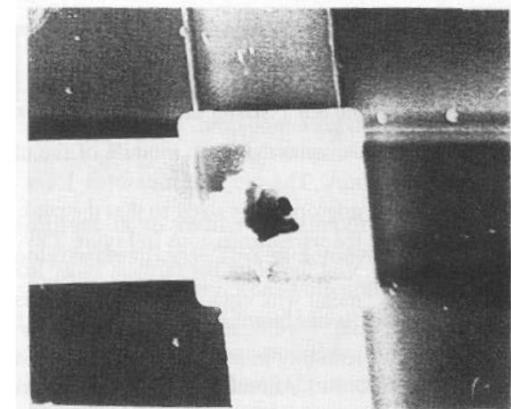
Hillocks/ Short



Short in Metal layer



Open in Metal layer



Open in Via

Electromigration

— EM Remedies and Precautions

- Wire widening to reduce current density
- Good power management techniques

Bigger Power Grids for power nets

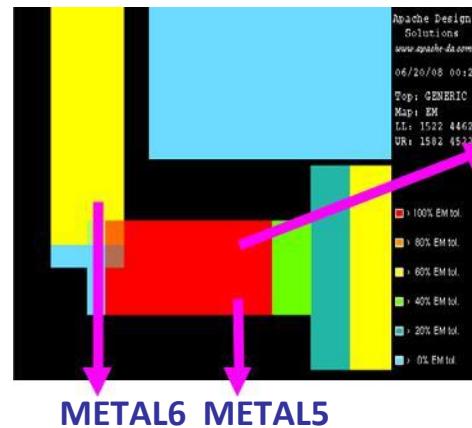
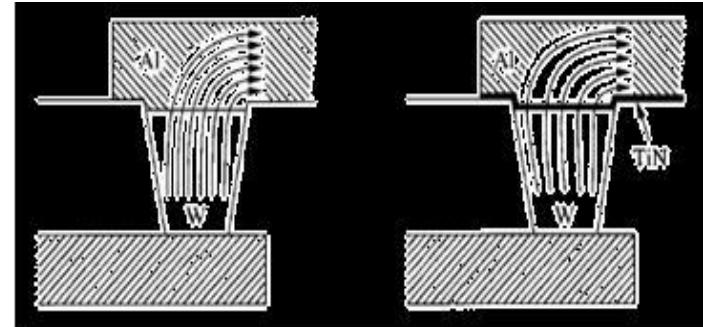
(putting power grids on thicker layers)

Wire-widening for signal nets

Better Power Grid planning

Double sizing for Power Greedy nets

- Providing Redundant Vias
- Designing the circuit to run at lower voltage levels
- EM resistance can be increased by alloying with Copper
- Controlling temperature by using a thermal-aware IC design methodology
- DFM techniques that reduce variability
- Besides, need to be aware of “dishing” effect (CMP)



EM limit for M6 is high compared to M5 hence no violation in M6
Increase M5 width

Electromigration

— Types of EM checks

- Related to Currents
 - 1. Average EM checks
 - 2. RMS EM checks
 - Peak EM checks
- Related to Nets
 - 1. Signal EM checks
 - 2. Power EM checks
- Limits for all these EM checks will be specified in technology file as a function of minimum life of the device, depending on the application
- All the three Current related EM checks need to be satisfied for Signal EM unless otherwise specified
- For Power nets, satisfying Average EM numbers would suffice

• EM failure mechanisms

- Timing Failure: Narrowing of the wire will increase wire resistance, which may cause a timing failure if a signal can no longer propagate within the clock period
- Functional Failure: Electromigration will continue until the wire completely breaks, allowing no further current flow and resulting in functional failure

Electromigration

- EM Rule Types
 - Metal Layer based (This was the only rule used in older technologies)
 - Metal length or width dependent EM Rules
 - Length and width of upper and bottom Metal and also depends on Via width
 - Complex rules with polynomials
- Black's Equation

$$\text{Mean Time To Failure (MTTF)}, t_{50} = C J^{-n} e^{(E_a/kT)}$$

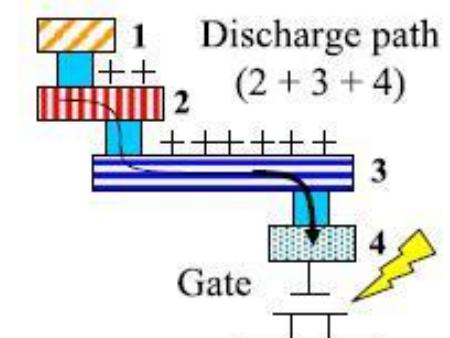
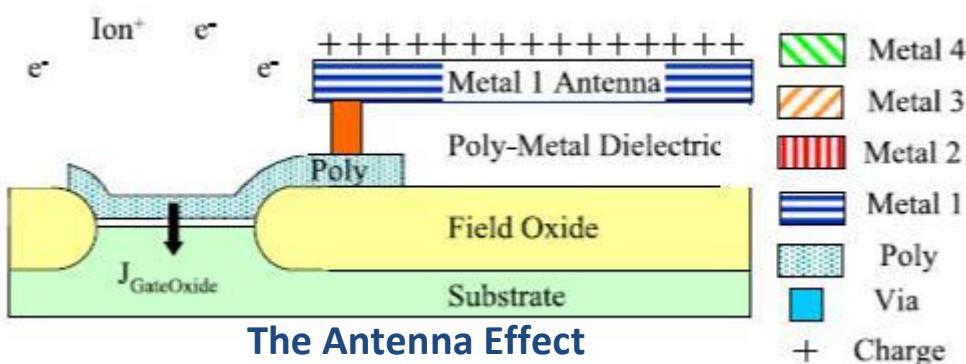
- t_{50} = the median lifetime of the population of metal lines subjected to EM
- C = a constant based on metal line properties (depends on cross sectional area)
- J = the current density ($J_{dc} < 1 - 2 \text{ mA / mm}^2$)
- n = integer constant from 1 to 7; many experts believe that n = 2
- T = temperature in degree Kelvin
- k = the Boltzmann constant
- Ea (Activation Energy) = 0.5 - 0.7 eV for pure Al

Antenna Effect

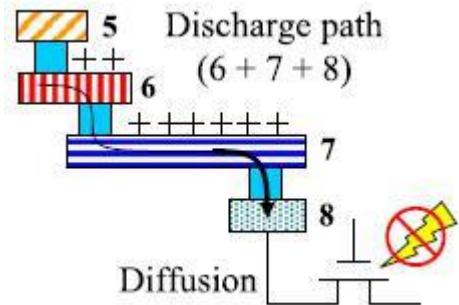
- Antenna Effect
 - A phenomenon of charge accumulation in metal segments that are connected to an isolated Gate (Poly) during the metallization process
 - This phenomenon occurs during process, so also known Process Antenna Effect (PAE)
 - It occurs when conducting net act as antenna, amplifying the charge effect
 - The conductive layers are receiving the charge, so termed as Antenna Effect
- Reason for Antenna Effect
 - Glow discharge during Plasma etching results in electric charging, which when occurred in conductive layer leads to Antenna effect thus termed Plasma-Induced/ Process-Induced damage (PID)
 - Charging occurs when conductor layers not covered by a shielding layer of oxide are directly exposed to Plasma
 - During process like soldering the chip is protected with some shielding
 - But during fabrication there is no such protection & will lead to Antenna effect
 - For Aluminium based process PAE is prominent at Etching stage and for Copper based process PAE is prominent at Chemical-Mechanical Polishing (CMP) stage
 - If the area of a higher metal layer connected to the Gate through lower metal layer/ layers, then the charge of higher metal layer got added to the lower metal layer which can also cause PAE called Accumulative Antenna Effect

Antenna Effect

- Impact in the design
 - If the area of the layer connected directly to the Gate the static charges are discharged through the Gate, the discharge can damage the oxide that insulates the gate and cause the chip to fail
 - Fowler-Nordheim (F-N) tunneling current will discharge through the thin oxide and cause damage to it



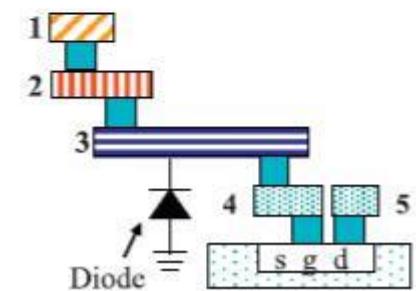
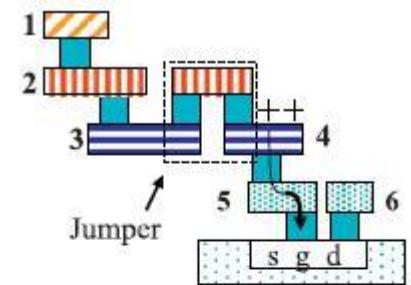
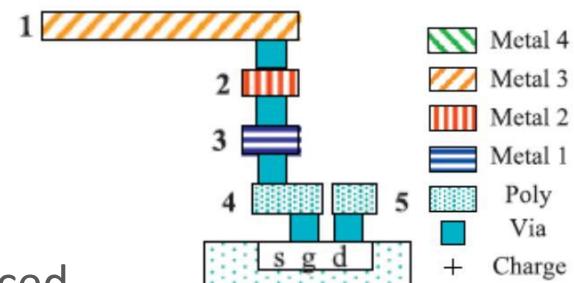
Charge accumulation & discharging on Poly



Charge accumulation & discharging on Diffusion

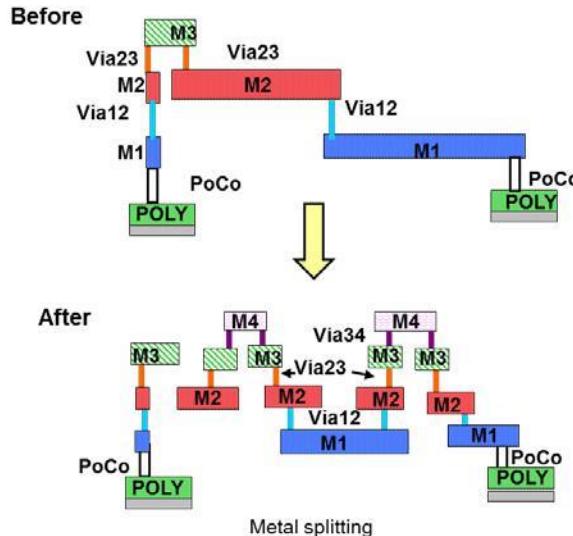
Antenna Effect

- Remedies for PAE
 - Assigning higher metal layers for routing
 - Higher metal layers will not be connected directly to the Gate Connect various metals through Via connections
 - Inserting Jumpers
 - If PAE is in lower layers then PAE can be reduced by connecting it to higher layers through Jumpers
 - Jumpers will reduce the peripheral metal length, which is attached to the Gate
 - Connecting Antenna diode
 - If it is in higher layers, Jumper wont be a solution, hence need diodes
 - As soon as extra charge is induced onto metal/ poly the diode diverts the extra charges to the substrate
 - But for buffer insertion higher metal layers has to come to lower metal layer (M1 or M2) to connect to pins of buffer and go back and also there may not be enough place for buffer insertion
 - After routing only we go for antenna check, so Buffer insertion may lead to congestion and DRC violations

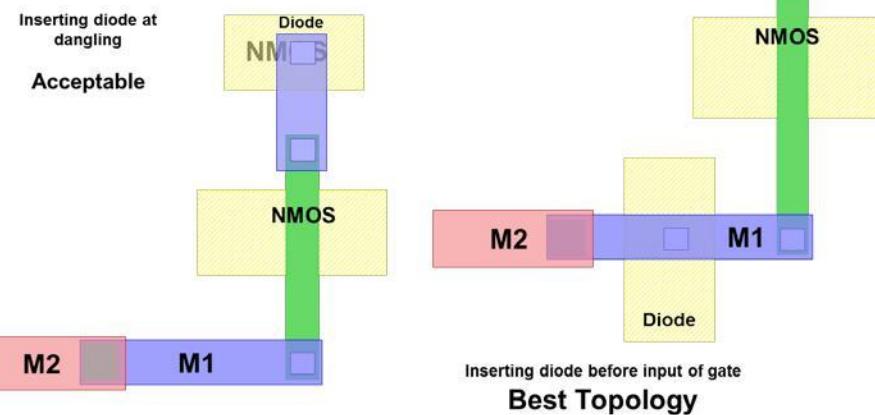


Antenna Effect

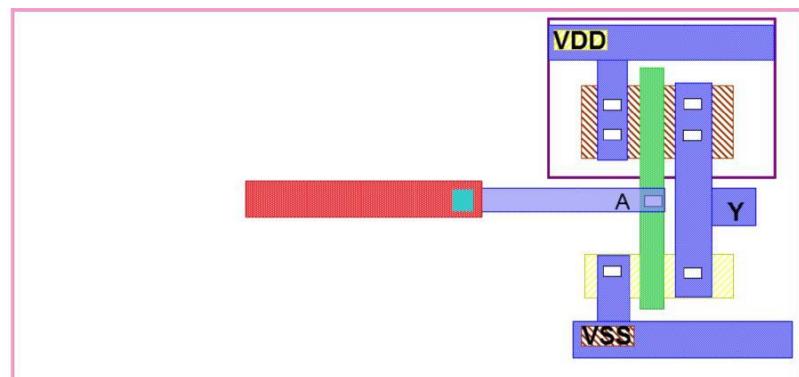
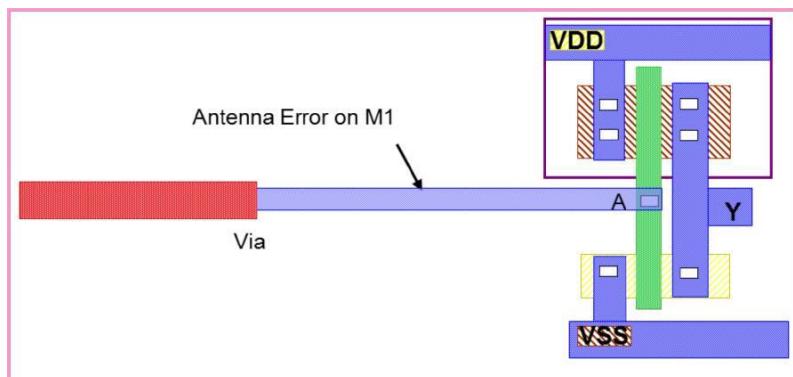
- Remedies for PAE



Metal Splitting: Connecting to higher Metal Layers



Diode placed at the unconnected end of gate
(optional due to resistance of poly)



Move the Via to reduce area of Metal 1

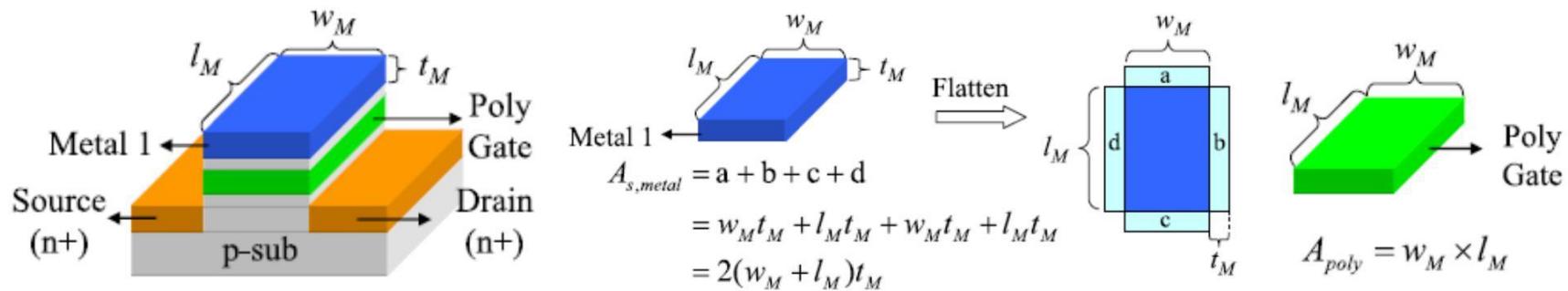
Antenna Effect

— Antenna Ratio (AR)

- A design rule to prevent charge accumulation during Metal/ Poly-Si layer etching which limits the area of metal segment connected to the Gate oxide
- Foundries set a maximum allowable AR for the chips they fabricate
- The AR is defined as the ratio of plasma-exposed area $A_{s,metal}$ to the gate oxide area A_{poly} as formulated,

$$AR = \frac{\text{plasma-exposed area}}{\text{gate oxide area}} = \frac{A_{s,metal}}{A_{poly}} \leq k_{th} ; k_{th} \text{ is the threshold of AR}$$

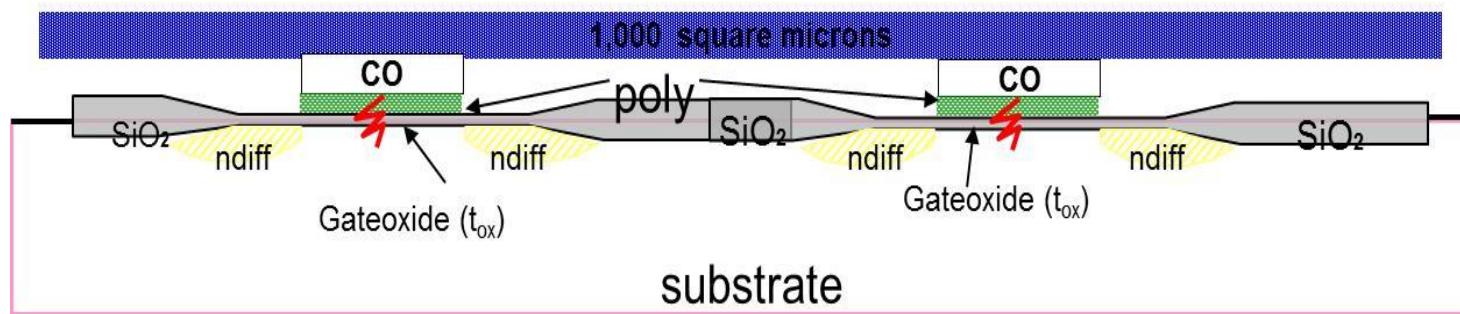
- This rule can be applied to any metal segment connected to the Gate



Antenna Effect

- Antenna Effect possibilities example

- Assume a foundry setting a maximum allowable antenna ratio of 500
- If a net has two input gates that each have an area of 1 square micron, any metal layers that connect to the gates and have an area larger than 1,000 square microns have process antenna violations because they would cause the antenna ratio to be higher than 500



- Dominant as technology shrinks
 - When oxide thickness reduces
 - More metallic structures are added to the chip

Antenna Effect

- Antenna (ANT) Rules

- The Antenna Ratio

- For Aluminium at Etching stage (metal deposition)

The top of the metal is protected by a resist during this step, so the antenna rules for this process should be based on the metal sidewall area

- For Copper at Chemical-Mechanical Polishing (CMP) stage

Charge accumulation occurs during CMP

In this process, the sides of the metal are protected, so the antenna rules need to be based on the metal's top surface area

- Metal used in the process depends on Technology

- From 28nm onwards Aluminium is replacing Copper

Antenna Effect

- PAE as a side effect of the manufacturing process
 - Plasma etchers/ ion implanters induce charge into various structures connected to Gate Oxide
 - This induced charges destroy the Oxide layer - a permanent damage
 - Conductor layer pattern etching processes
 - Amount of accumulated charge is proportional to perimeter length
 - Ashing processes
 - Amount of accumulated charge is proportional to area
 - Ashing processes remove remaining photo resist layers after etching processes of a conductor layer
 - In the late stage of the processes, the area of a conductor layer pattern is directly exposed to plasma
 - Contact etching processes
 - The amount of accumulated charge is proportional to the total area of the contacts
 - Contact etching processes dig holes between two conductor layers
 - In the late stage of the processes, the area of all the contacts on the lower conductor layer pattern is directly exposed to plasma

Crosstalk

— What is Crosstalk?

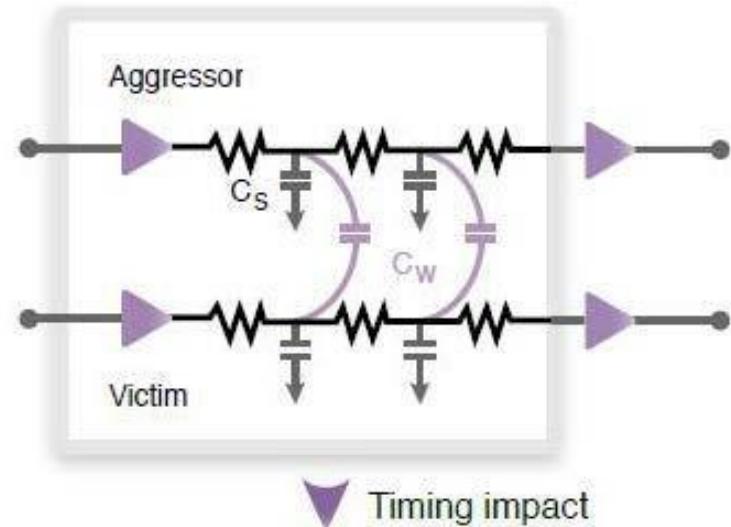
- Refers to a signal affecting another signal being transmitted in vicinity caused by capacitive/ inductive coupling
- Crosstalk is the unwanted coupling of energy between two or more adjacent lines which can change the required signal and is also termed as Xtalk
- Occurs on long adjacent wires
- Can be interpreted as the coupling of energy from 1 line to another via:

Mutual Capacitance,

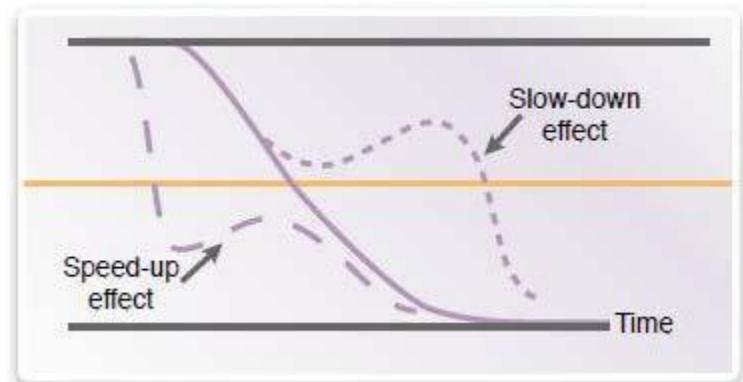
C_m (due to Electric Field)

Mutual Inductance,

L_m (due to Magnetic Field)

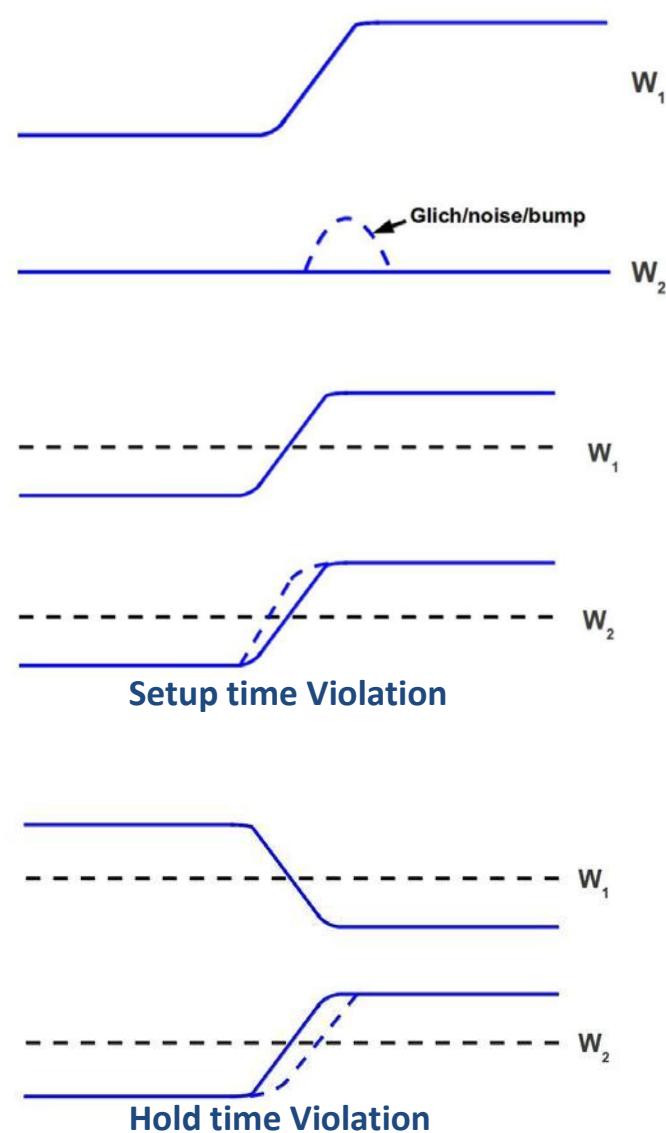


▼ Timing impact



Crosstalk

- Impact of Crosstalk in the design
 - Functional Failures
 - Noise induced glitches
 - If the Glitch duration is that of clock period duration, an extra clock cycle effect
 - Timing violations
 - If aggressor switches in opposite direction to the victim : Setup time Violation
 - If aggressor switches in same direction to the victim : Hold time Violation
 - If the victim line is not terminated at both ends in its characteristic impedance the induced spurious signals can reflect at the ends of the line and travel in the opposite direction down the line
 - Thus a reflected near-end crosstalk can end up appearing at the far end and vice versa

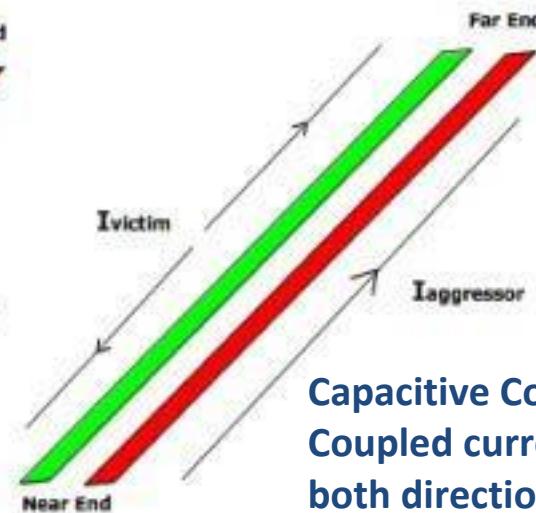
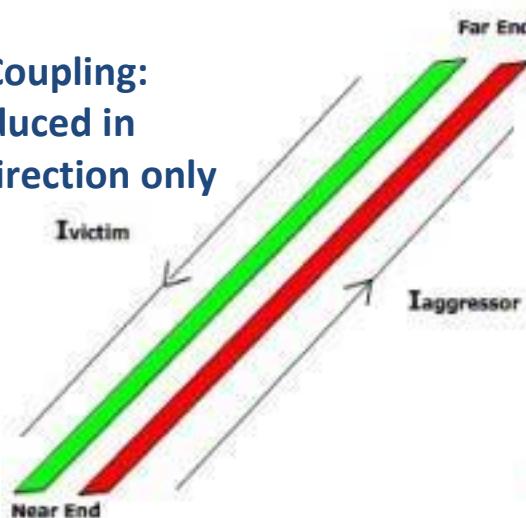


Crosstalk

- Types of Crosstalk

- Energy that is coupled from the actual signal line, the aggressor, onto a quiet passive victim line so that the transferred energy "travels back" to the start of the victim line. This is known as the backward or near-end crosstalk
- Energy that is coupled from the active signal line, the aggressor, onto a quiet passive victim line so that the transferred energy "travels forward" to the end of the victim line. This known as forward or far-end crosstalk

Inductive Coupling:
Current induced in
opposite direction only



Capacitive Coupling:
Coupled current flows in
both directions

Crosstalk

— Remedies to avoid Xtalk

- Its a 3 dimensional problem, so height, width and length matters
- Noise/Bump violations can be fixed by changing the spacing between critical nets
- Shield the clock nets (critical nets) from other nets by ground lines
- Net Re-ordering

Avoid routing the critical nets parallelly

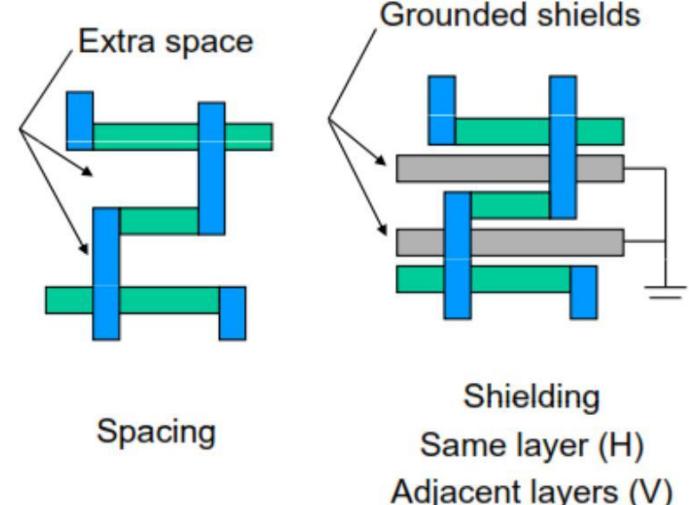
for long distances

- Modify the clock net (critical nets) minimum width from normal value to a larger one

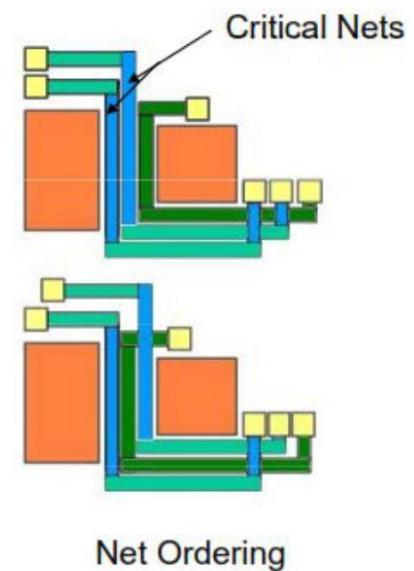
This makes the router to skip a grid near clock net to prevent spacing violation

This technique not only reduces crosstalk, but will also have a lower resistance due to larger line width & less side wall capacitance

- Can be fixed either by upsizing (increasing the drive strength) of the victim, or by downsizing (decreasing the drive strength) of the aggressor



Shielding
Same layer (H)
Adjacent layers (V)

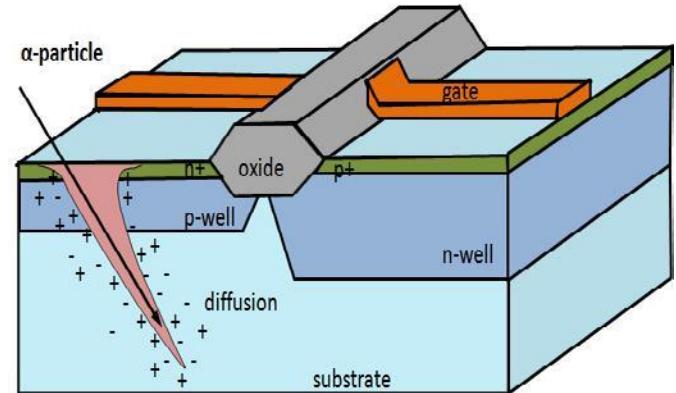


Net Ordering

Soft Errors

- Soft Error (Random Particle Error)

- Soft error is the phenomenon of an erroneous change in the logical value of a transistor, and can be caused by several effects, including fluctuations in signal voltage, noise in the power supply, inductive coupling effects etc., but, majority of soft errors are caused by cosmic particle strike on the chip
- With technology scaling, even low-energy particles can cause Soft Errors
- Soft errors are radiation induced faults which happen due to a particle hit, either by an alpha particle from impurities in packaging material or a neutron from cosmic rays
- When particles strike the silicon substrate they create hole-electron pairs which are then collected by PN-Junctions via drift and diffusion mechanisms
- This collected charge creates a transient current pulse and if it is large enough, it can flip the value stored in the state saving element (bit cell, latch etc.)
- These upsets are called Single Event Upsets (SEU)



Soft Errors

- Impact in the design
 - Soft error can result in incorrect results, segmentation faults, application or system crash, or even the system entering an infinite loop
 - When particle strike happens in combinational circuit, the result is a glitch which can then propagate to a latch where it could be clocked in and incorrect data can be latched
- Precautions to avoid Soft Errors
 - Radiation Hardening: Technique to reduce the Soft Error rate in digital circuits
 - Radiation hardening is often accomplished by increasing the size of transistors who share a Drain/ Source region at the node

Self-Heating

- If current flows through a wire, then due to the resistance of the wire heat will generate
- Oxide surrounding wires is a thermal insulator, so heat tends to build up in wires
- Hotter wires are more resistive & become slower
- Wire self-heating is only a negligible effect in the supply lines on bulk-CMOS ICs
- Self-heating Design Rule/ Self-heating Limit AC current densities for reliability
 - Typical limit: $J_{RMS} < 1.5 \text{ MA/cm}^2$ (for Aluminum nets)
 - It limits the unavoidable degradation of Electromigration lifetime due to temperature increase in the current carrying or in any nearby interconnect

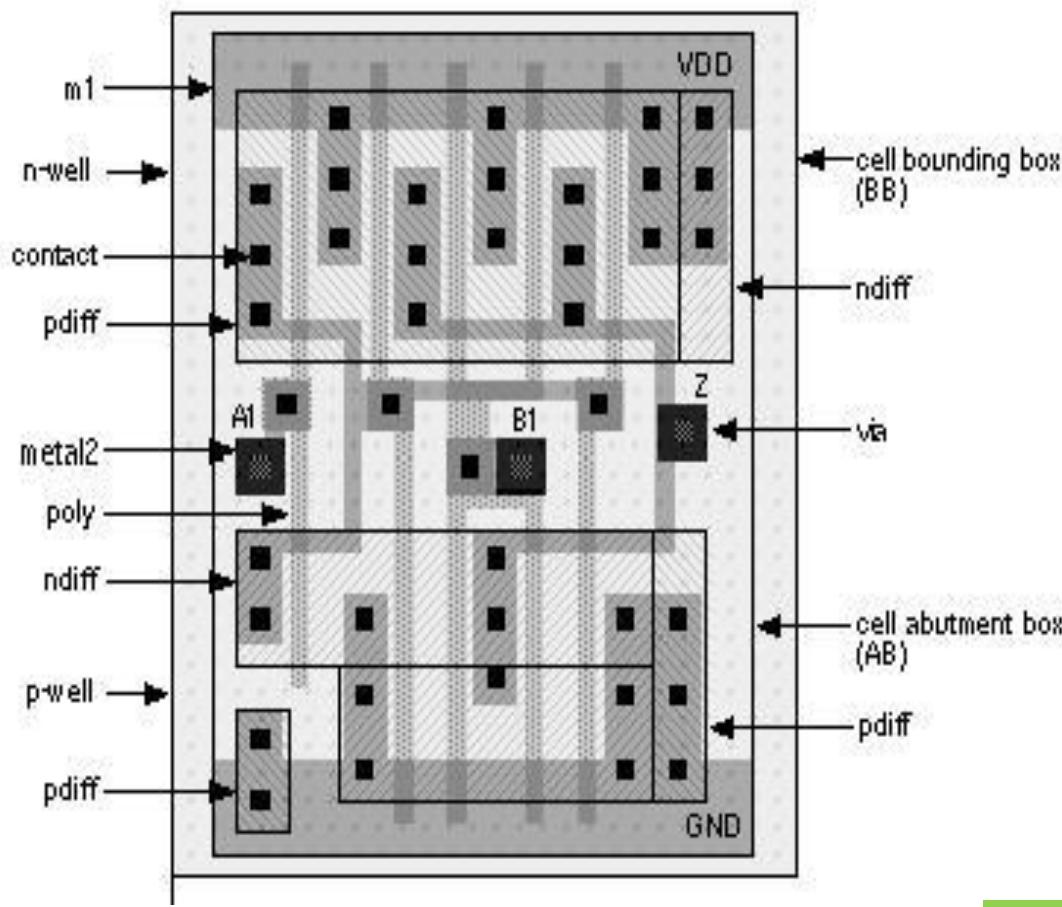
Cells in ASIC Physical Design

Cells in ASIC Physical Design

- Special Cell Requirements in IC Design is to minimize the possible CMOS issues
- More no. of transistors than are necessary for basic functioning. e.g.,
 - To limit the Overshoots and Undershoots
 - To protect the components from destruction
 - To isolates 2 components by PN Junction
- Common Special Cells used in CMOS IC Design:
 - Standard Cells
 - ICG Cells
 - Well taps (Tap Cells)
 - End caps
 - Filler Cells
 - Decap Cells
 - ESD Clamps
 - Spare Cells
 - Tie Cells
 - Delay Cells
 - Metrology Cells

Standard Cells

- A Standard Cell is a group of transistor and its interconnect structures that provides a Boolean logic function (e.g., AND, OR, XOR, XNOR, Inverters) or a storage function (Flip-flop or Latch)
- Std. Cell methodology has helped designers to scale ASICs from comparatively simple single-function ICs, to complex multi-million gate SoCs
- Cell-based methodology makes designer to focus on the implementation (physical) aspects



A Standard Cell Layout

Standard Cells

\emdash The cell's Boolean logic function is called its logical view: functional behavior is captured in the form of a truth table or Boolean algebra equation (for combinational logic), or a state transition table (for sequential logic)

\emdash AOIs (AND-OR-INVERTER) provide a way at the gate level to use less transistors than separate ANDs and a NORs

\emdash ASIC design logic builds upon a standard logic cell library, therefore, do not optimize transistors only logic gates

\emdash Types of Standard Cells

- Buffers (Inverting and Non-inverting)
- Combinational (AND, OR, NAND, NOR, AOI, OAI, OA, AO, MUX)
- Arithmetic (XOR, full-adder, half-adder), Sequential (latches, clock-gates, D-type flip/flops with any optional combination of scan input, set and reset)
- Miscellaneous (ICG Cells, Well Taps, Tie Cells, End Caps, Decaps, Filler Cells, Spare Cells, Delay Cells, Antenna Diode, ESD diodes)

ICG Cells

- Integrated Clock Gating Cells (ICG Cells)
 - During idle modes, the clocks can be gated-off to save dynamic power dissipation on flip-flops
 - Proper circuit is essential to achieve a gated clock state to prevent false glitches on clock path
 - Use a combination of AND and a Latch to avoid any glitches on the clocks. A glitch can propagate a false edge on to the design

Insertion of ICG

- Manual insertion of ICG

The clock gating can be implemented through logic circuits and ICG's

Most of Clock Gating Cells from vendor libraries have a RTL code

- Automated Insertion of ICG –

Some power aware tools insert the ICG's

through automated software algorithms

\emdash Types of Clock Gating Cells

- Latch Based Clock Gating Buffer for Neg-edge

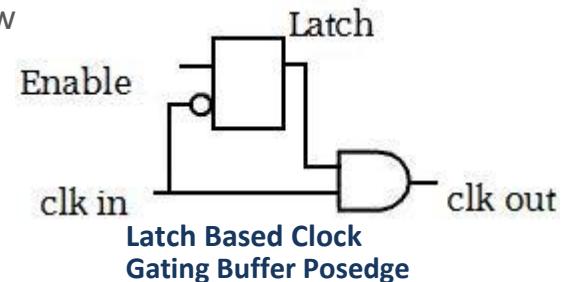
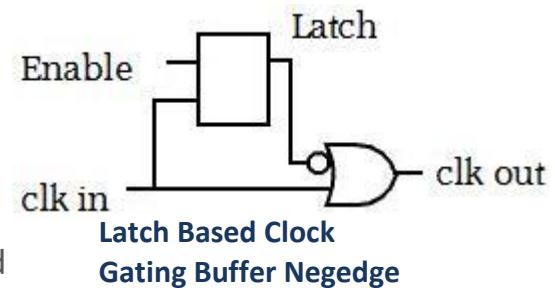
The circuit employs a latch and OR gate with one input inverted

The output clock is always clock gated low when Enable is low

- Latch Based Clock Gating Buffer for Pos-edge

The circuit employs a latch with inverted clock input and a AND gate

The output clock is always clock gated HIGH when Enable is low

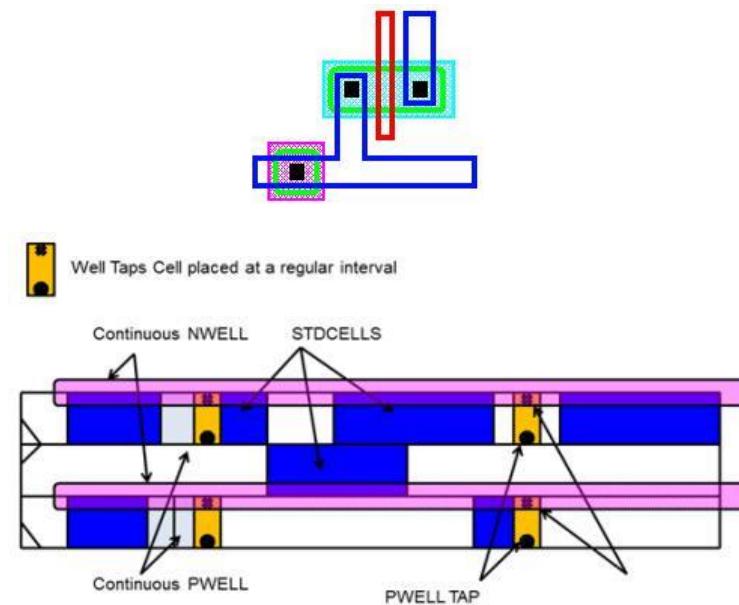


ICG module IO's

- 3 input ports – clock, clock enable and test
- 1 output port – clock for gated clock

Well Taps

- Physical only cell which helps to tie MOS Substrate and N-Well to VDD and GND levels, and thus avoid latch-up possibilities
- Switching circuits dump current into Well/ Substrate and if there is a high resistance between Well/ Substrate and the VDD/ GND grids the Substrate can be at different potential than VDD/ GND which causes latch-up
- Well Tap Cells reduce resistance between VDD/ GND to wells of the Substrate
- Tap Cells are usually placed on the Power Rails of the Standard Cells
- Standard Cells do not have internal tap to N-well (P substrate process) to reduce design complexity of Standard Cells
- These library cells do not have any signal connectivity
- Hence Tap to Wells is done by external cells called "Tap cells" which are sprinkled all over Core Area at regular distance as decided by the foundry
- More Taps reduces resistance, but will also increases core area, so we need a trade-off which will be provided by the foundry
- Place well taps at regular intervals throughout the design with the specified distances and snaps them to legal positions

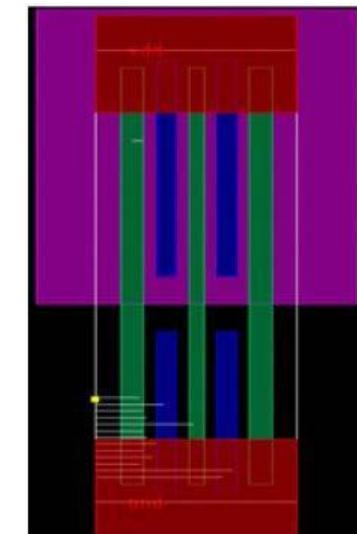


End Caps

- \emdash End-cap cells are preplaced physical-only cells required to meet certain design rules and placed at the ends of the site rows by satisfying well tie-off requirements for the core rows
- \emdash These library cells do not have any signal connectivity
- \emdash They connect only to the power and ground rails once power rails are created in the design
- \emdash They also ensure that gaps do not occur between the well and implant layers i.e. well proximity effect
- \emdash This prevents DRC violations by satisfying well tie-off requirements for the core rows
- \emdash Each end of the core row, left and right, can have only one end cap cell specified
- \emdash However, you can specify a list of different end caps for inserting horizontal end cap lines, which terminate the top and bottom boundaries of objects such as macros
- \emdash End caps have a fixed attribute and cannot be moved by optimization steps
- \emdash A core row can be fragmented (contains gaps), since rows do not intersect objects such as power domains. For this, the tool places end cap cells on both ends of the un-fragmented segment

Filler Cells

- Physical only cells which provide N-Well continuity and avoid N-Well spacing DRC
- Filler cells are inserting for density rules, to meet Core Utilization targets and to avoid sagging of layer
- Filler cells are inserting at the last stage of Placement and Routing
- Some of the small cells also don't have the Bulk/Substrate connection because of their small size (thin cells)



Filler Cell Layout

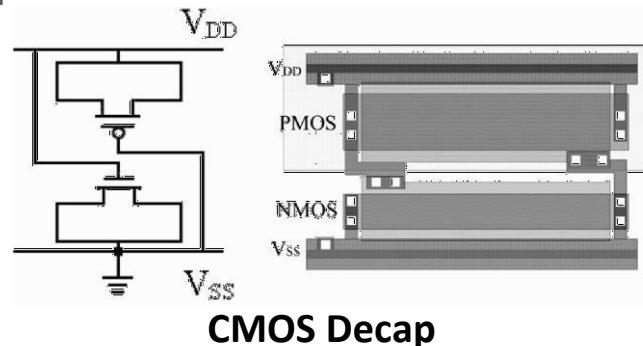
— In those cases, the abutment of cells through inserting Filler Cells can connect those Substrates of small cells to VDD/ GND nets

— i.e. those thin cells can use the bulk connection of the other cells

— Filler cells are used to make up the Poly density (if that filler cell is having any poly structure inside), but certainly not for metal density

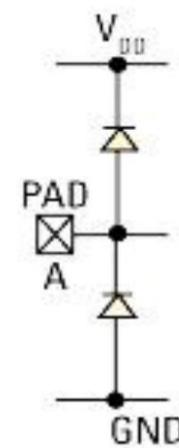
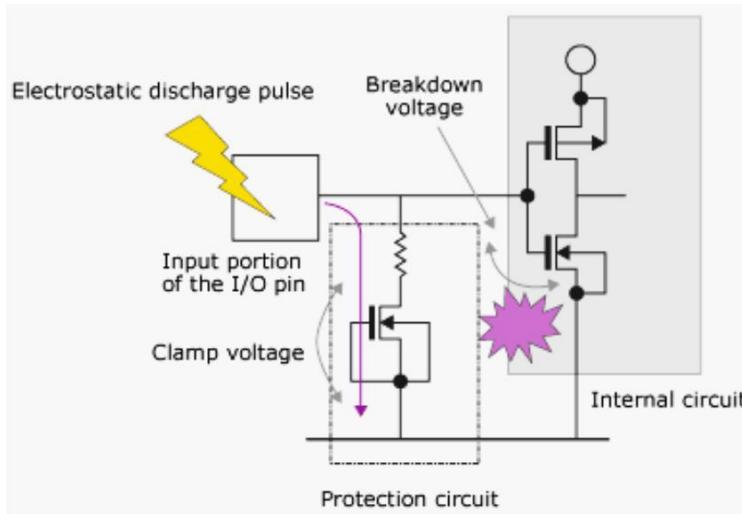
Decap Cells

- Decaps are on-chip decoupling capacitors (Extrinsic Capacitances) that are attached to the power mesh to decrease noise effects (dynamic I.R. Drop)
- Supply voltage variations caused by Instantaneous Voltage Drop (IVD) lead to problems related to spurious transitions and delay variations
- Decap cells are typically poly gate transistors where source and drain are connected to the ground rail, and the gate is connected to the power rail
- Decap helps to smoothen out the Glitches and Ground bounce
- 3% to 8% of the core physical area is required for Decaps referred as decap density
- It is important to place only the necessary amount of decaps since they normally come with a quite serious downside as they are leaky devices
- Another drawback, which many designers ignore, is the interaction of the decap cells with the package RLC network
- Since the die is essentially a capacitor with very small R and L, and the package is a huge RL network, the more decap cells placed the more chance of tuning the circuit into its resonance frequency. That would be trouble, since both VDD and GND will be oscillating
- NMOS Decaps are superior to PMOS decaps because of the high frequency operation and large R_{EFF} and C_{EFF} for the same area



ESD Clamps

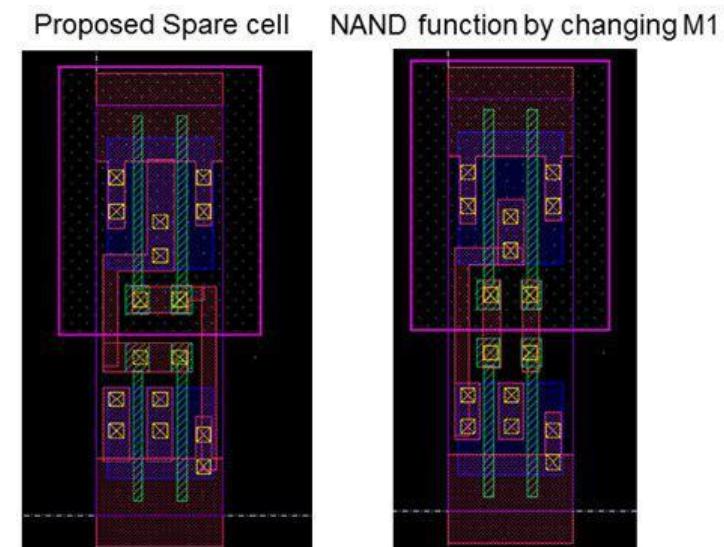
- ESD Clamp/ ESD Diode is the primary protection device that protects against ESD surges at the I/O pad by clamping the voltage and allowing the high ESD current to be discharged safely to the ground terminal
- The main function of ESD Clamp is to protect the Gate oxide
- Snap back device (Diode implementation between the grounds) provides Snapback voltage (ESD Voltage) to get grounded thus the ESD current won't be getting in to Gate
- The design of ESD Clamp must ensure that Electrical Overstress (EOS) events do not cause failure
- The ESD Clamp is essential for HBM, MM, and CDM



Spare Cells

- Pre-placed inactive (with inputs tied off) gates in the empty areas of a design (or even in the crowded areas) before tape-out (Mostly NAND Gates)
- ECO Cells/ Spare Cells are collection of Gates coming in different sizes for doing small functional ECO and connect them with minimal mask changes called a metal-only ECO
- Provides new functions on a design which exhibits post-production problems
- No change is made to the diffusion layer, M1 and a contact layer only need to change

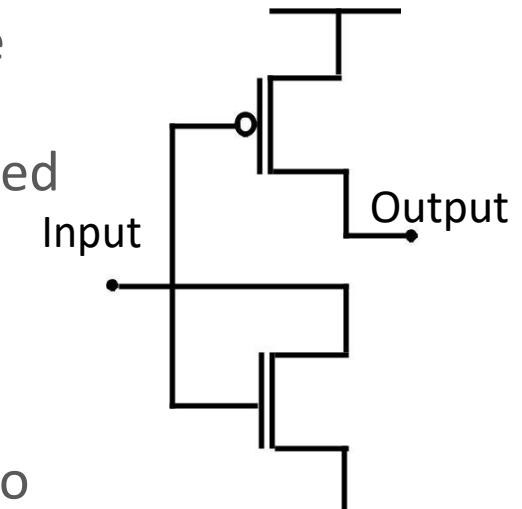
- Disadvantages:
 - They are connected to VSS and VDD and despite having their inputs tied off, they are still drawing Static Current
 - The designer may not have the right cell in the right place at the time of the ECO



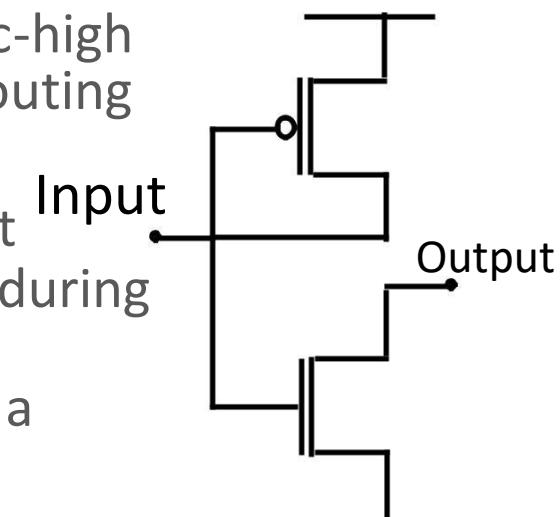
Tie Cells

- Tie-high and Tie-Low cells are used to connect the Gate of the transistor to either Power or Ground
- In deep sub micron process, if the Gate is connected to Power/ Ground, the transistor might be turned ON/ OFF due to Power or Ground Bounce
- The suggestion from foundry is to use Tie Cells for the purpose
- The cells which require VDD, comes and connect to Tie High (so Tie High is a Power Supply Cell), while the cells which wants VSS connects itself to Tie-Low
- Without Tie Cells, unused inputs are tied to logic-high or logic-low, and these connections are made by routing the input pin right to the Power/ Ground grid
- With Tie Cells, unused inputs in the original netlist are tied to logic-high or logic-low, and somewhere during the physical design process, Tie Cells are inserted

\emdash The unused inputs are then connected to a Tie-high or Tie-low Cell



Tie-up Cell



Tie-down Cell

Delay Cells

- Delay cells
 - Are buffer cells with slower transition time
 - Can drive high currents
 - Are helpful in reducing Slew Rate (0-1 or 1-0 Transition Time)
 - Are of wider channel
 - Have delay starting from 20ps to few Nano seconds
 - Will have constant delay
- Delay cell insertion is the conventional way to fix hold time violation tends to penalized in area percentage increment
- Lesser number of delay cells are required for hold time fixing as compared to buffers but it will have area much greater than normal buffers
- Increasing gate width reduces gate capacitance hence reduces delay, but results in higher leakage
- It has inverter in input and a inverter in output and in between these two inverters it has a combination of a inverter and pass transistors. Pair of inverter and pass transistor provide at large delay
- Depending on the delay of the cell, pair of inverter and pass transistor can be repeated multiple times

Metrology Cells

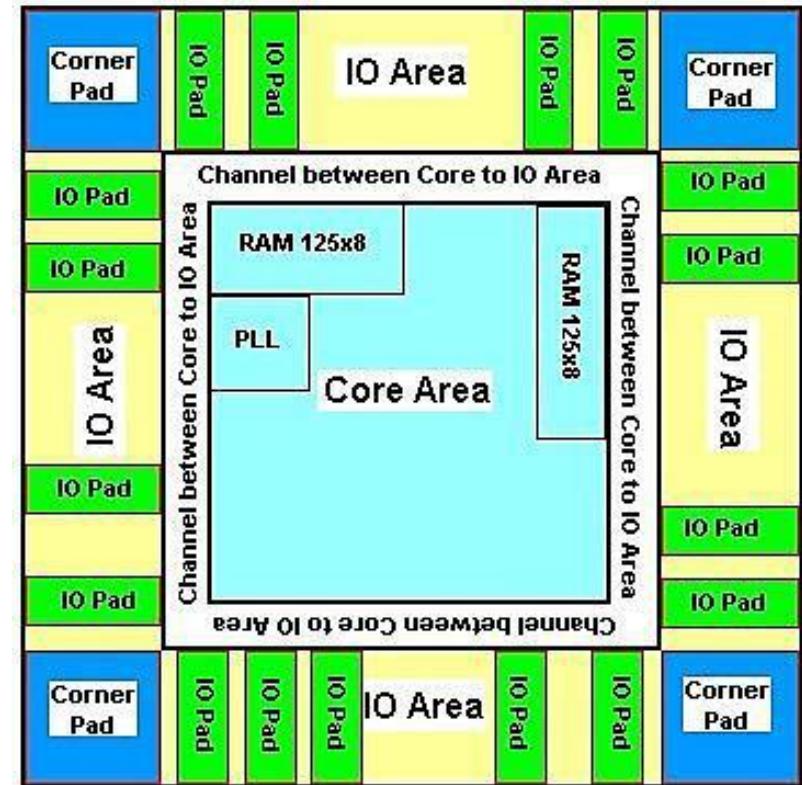
- To enable the reliable re-productivity of micro-scale devices used in high volume and low cost
- To measure and monitor the process parameters during manufacturing
- The effect of process variations during fabrication time can be identified and measured

IO Design

IO Pads

\emdash Input Output Pads

- Input/ Output circuits (I/O Pads) are intermediate structures connecting internal signals from the core of the integrated circuit to the external pins of the chip package
- Typically I/O pads are organized into a rectangular Pad Frame
- The input/output pads are spaced with a Pad Pitch
- Pads will have pins on all metal layers used in design for easy access while routing the design
- Number of layers depends on technology
- Multiple Power Pads are often used to reduce the power
- Pads consists of some logic cells like level shifters and buffers which will control the voltages of input and output signals and to increase/ decrease drive strength



IO Pads

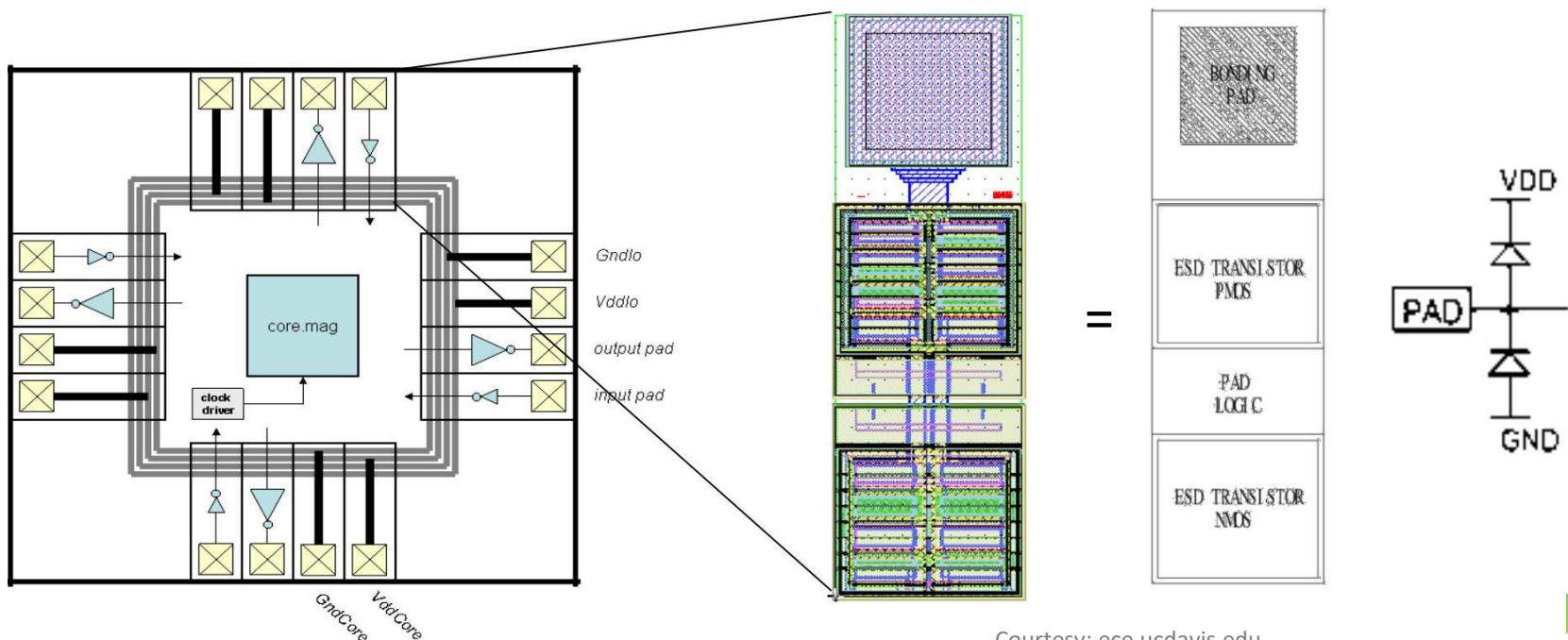
- Structure of Pads

- Bonding Pad

Area to which the bond wire is soldered

The wire goes from the bonding pad to a chip pin

- ESD (Electrostatic Discharge) protection circuitry consisting of a pair of big PMOS, NMOS in a reverse biased diode structure
- Driving and Logic Circuitry for which the area of is designated



Courtesy: ece.ucdavis.edu

IO Pad Design

- Implementation Guidelines

- Isolate sensitive asynchronous inputs such as Clock or Bidirectional Pins from other switching pads with Power/Ground Pads
- Group Bidirectional Pads together such that all are in the input/ output mode
- Avoid continuous placing of simultaneous switching pads
- 2 extra pins = 1 extra pad on 2 sides and 4 extra pins = 1 extra pad on each side
- Power supply pads must be evenly distributed
- The number of Power Pads required are calculated based on the IO Signal Pads power requirement and Core Power requirement (IR drop limit)
- No. of IO Power Pads required in a design,

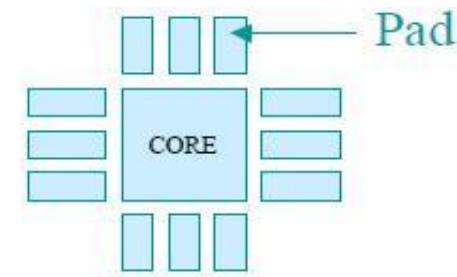
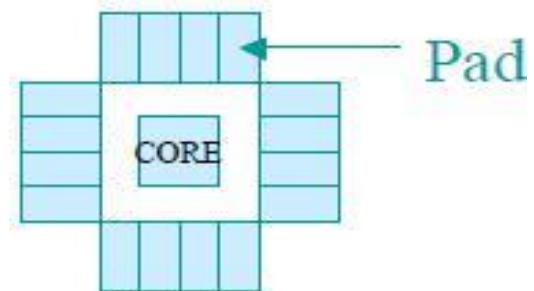
Thumb Rule: One Pair of Power Pads for every 4 or 6 Signal Pads

- No. of Core Power Pads required in a design,

$$\text{Pads per side} = \frac{\text{TotalCorePower}}{(\#side * V_{worst} * \text{MaxAllowableCurrentofPad})}$$

IO Pad Design

- Pad Limited design
 - The area of Pad limits the size of Die
 - No. of IO pads are more or larger in size (technology dependent)
 - Pad limited designs pose several challenges for design implementation and to the backend designers, if Die area is a constraint
 - The Solution would be to use Flip Chip or Staggered IO placement techniques
- Core Limited Design
 - The area of Core limits the size of Die
 - No. of IO Pads are lesser
 - In these designs Inline IOs will be used
 - It can be either due to large no. of Macros the design or due to larger logic
- Types of Pads according to Logic directions
 - Input Pad
 - Output Pad
 - Bidirectional Pad



Types of IO Pad

— Types of Pads according to Logic Styles

- Signal Pads
- Power Pads (Core Power and IO Power)
- Corner Pads

Corner pads contains only connections

in all metal layers defined in technology

These pad used only for IO Ring continuity

and chip metal density on corners and to maintain yield

- Filler Pads

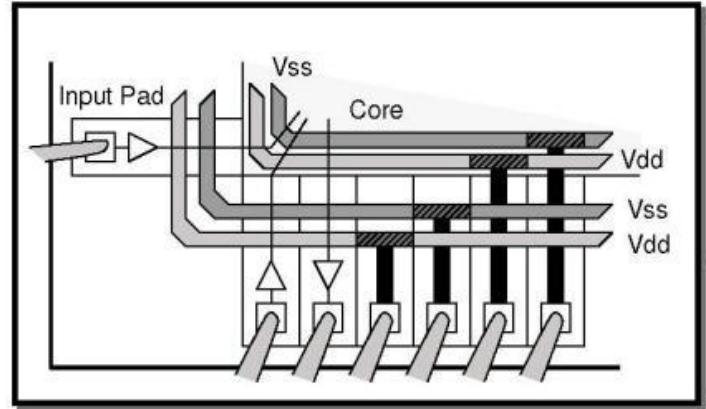
IO Filler Cells contains only the geometrical information of the Power Rings

in all metal layers

Continuity of Power Rings which is responsible for uniform distribution

of power

Electrostatic Discharge protection



Types of IO Pad

- According to the Pad locations
 - Peripheral IO Pads
 - Area IO Pads
- Types of Pads according to Implementation Styles
 - Inline
 - Staggered

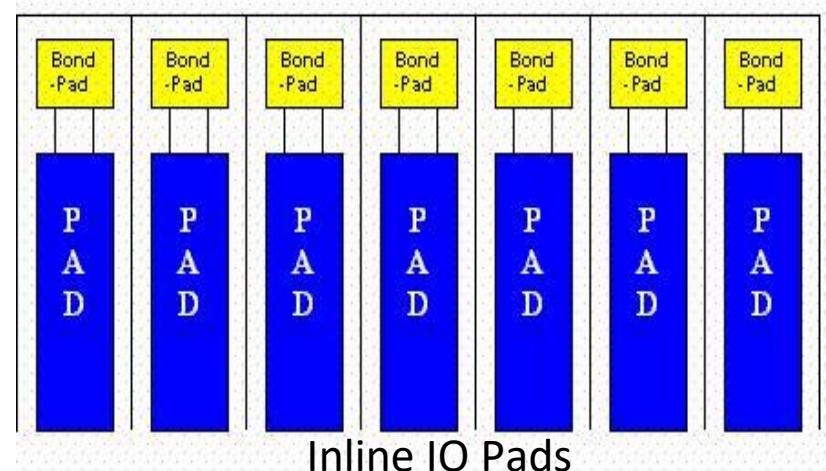
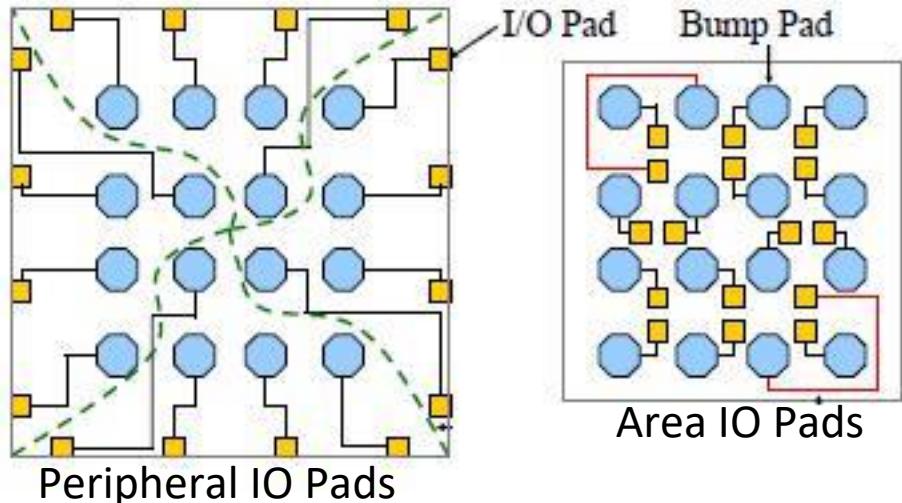
CUP (Circuit-Under-Pad)

Non-CUP (Circuit-Under-Pad)

- Flip Chip

Inline IO Pads

- Pads are placed next to each other, with the corresponding bond pads lined up against each other having a small gap in between
- Minimum Pitch is determined by foundry/vendor and is technology dependent

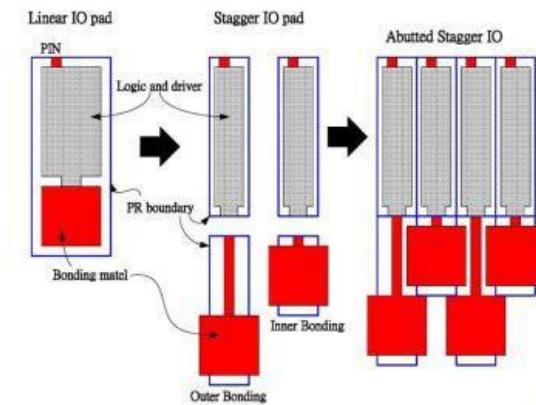
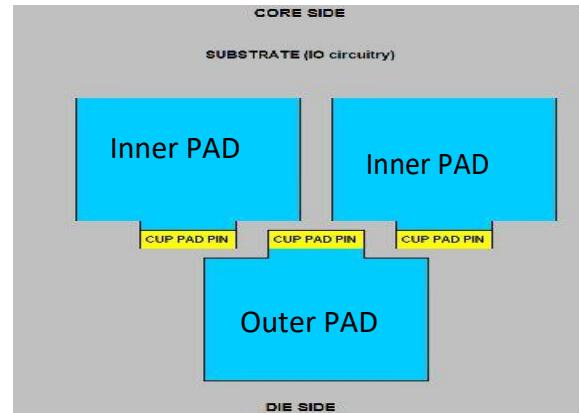


Types of IO Pad

- Staggered IO Pads
 - CUP (Circuit-Under-Pad)
 - Bonding Pad over the IO body itself
 - Bonding Pad have to connected to the PAD Pin of IO
 - Pad pin is located close to the center of the IO body for easier routing, signal integrity, and space saving
- Pros:**
- Reduce the die size since the Bonding Pad does not take any extra space in addition to the IO body itself
- Cons:**
- Advantages include more no. of IO's, Optimal area utilization, Lower cost

- Non-CUP (Circuit-Under-Pad)

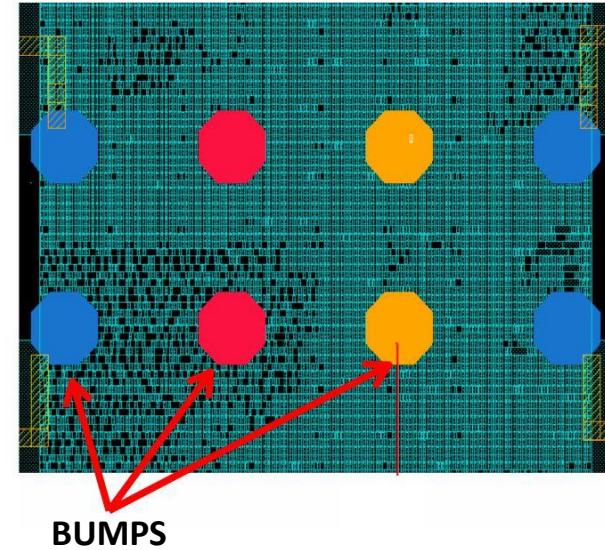
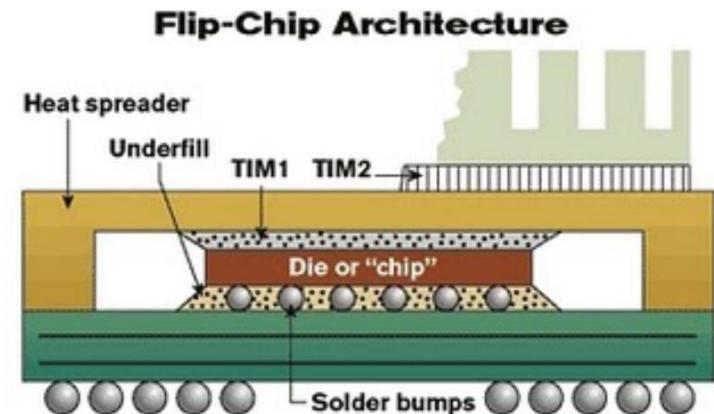
- Pros:**
- Useful technique if design is “Pad Limited”
 - Place an inner and outer Bond Pad alternately
 - A larger number of pads can be accommodated
 - Disadvantage is that the overall height of the pad structure increases significantly



Types of IO Pad

— Flip Chip IO Bumps

- It is simply a direct connection of a flipped electrical component onto a substrate, carrier, or circuit board by means of conductive Bumps instead of the conventional Wire-bond
- In Flip Chip, IO Bumps and driver cells may be placed in the peripheral or core area
- Note, the large octagonal area IO Bumps overlaying placed cells in the core area
 - No chip area benefit for small chips – full Bump array redistribution is very difficult
- In advanced technology nodes a separate Re-distribution layer (RDL) is make use of for the Bump connections



Delay Models

Delay Models

- Delay Calculation
 - The delay calculation is needed because of complex Input Capacitance, Voltage Drop, Voltage Islands, High Impedance nets etc.
 - Delay calculation parameter data are stored as Lookup-Table format
- Delay Models
 - Interconnect Delay Models
 - Lumped RCL Delay Models
 - Wire Load Delay (WLD) Model
 - Elmore Delay Model
 - Arnoldi Delay Model
 - Cell Delay Models
 - Non-Linear Delay Model (NLDM)
 - Scalable Polynomial Delay Model (SPDM)
 - Effective Current Source Model (ECSM)
 - Composite Current Source (CCS) Delay Model

Interconnect Delay Models

Wireload	Elmore	Arnoldi
<ul style="list-style-type: none"> Delays are estimated based on the number of fanout of the cell driving the net Values of unit resistance R and unit capacitance C are given in technology file Fanout vs net length is tabulated in WLMs Once the net length is known delay can be calculated 	<ul style="list-style-type: none"> Delays are estimated based on first moment of impulse response used where speed of calculation is important but the delay through the wire itself cannot be ignored Less accurate, but if the nets are very small, Elmore can provide sufficient accuracy with less run time Inherently cannot handle inductance effect but can be extended to include inductance Need higher order moments Useful for interconnect optimization 	<ul style="list-style-type: none"> More accurate Need more run time Used in cases where the driver resistance is much less than the impedance of the network to ground, especially when a very strong driver is connected to a very resistive network

Cell Delay Models

Non-Linear Delay Model	Effective Current Source Model	Scalable Polynomial Delay Model	Composite Current Source Delay Model
<ul style="list-style-type: none"> Modeled as a linear voltage ramp in series with a resistor Less accurate Less run time Intermediate values are interpolated Assumes load is purely capacitive Variation may range anywhere from 5-10% Linear k-factors required for handling of IR-drop, Delay Transition time are functions of Input slew and Output load 	<ul style="list-style-type: none"> Models a unique dataset for each Voltage-Temperature combination Improved accuracy Easy to characterize Increased characterization Smooth non-linear interpolation/extrapolation Can't use for memory or complex cell characterization Models IR Drop non-linearly Data characterized for three voltage corners 	<ul style="list-style-type: none"> Models the delay and slew values as a function of voltage and temperature SPDM is a polynomial abstract Less accurate More runtime Extra characterization setup required Increased characterization time Extrapolation is unreliable SPDM requires elaborate curve fitting techniques for an accurate curve fit 	<ul style="list-style-type: none"> Modeled as current waveform from a time varying current source More accurate More run time Extra setup required for characterization CCS libraries are huge in size Addresses the effects of deep submicron processes

ECO

ECO

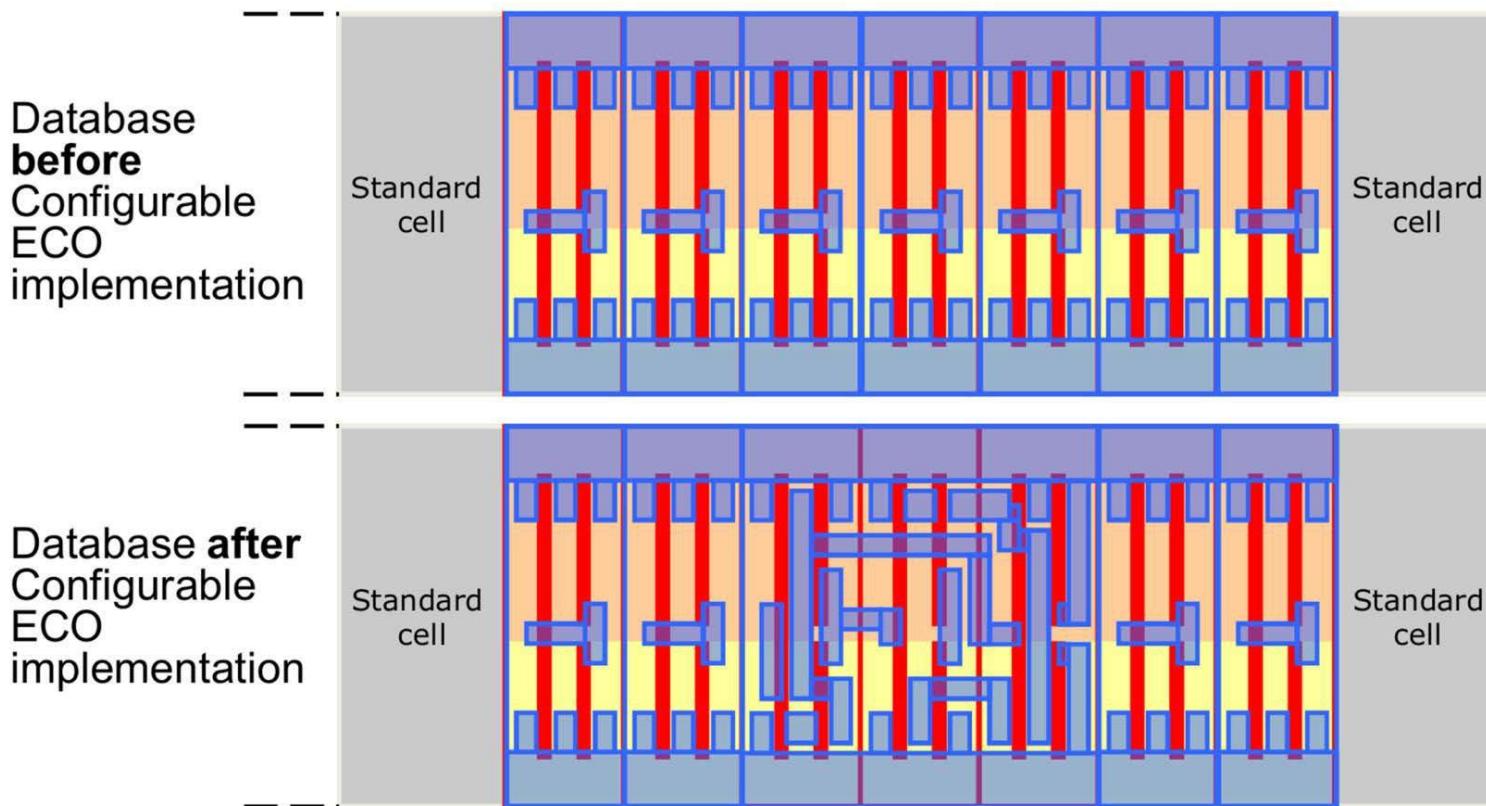
- Engineering Change Order (ECO)
 - Technique to add/ remove the logic with minimum modifications in the design
 - To deliver the product to market as fast as possible with minimum Risk-to-Correctness and Schedule
 - For fixing post Synthesis/ Route/ Silicon issues
 - Fixing both timing and functionality issues
 - Spare Cells placed in the design are used for ECO
 - A Logic Gate/ Flip-flop can be realized using these Spare Cells
 - Different flavors of Gate, of required drive strength can be realized anywhere in the design
 - Only Metal/ Contact changes are needed after fixing the defective design

Types of ECO

- Post Synthesis ECO (ECO after Synthesis)
 - ECO with Synthesized Netlist
- Post Route ECO (ECO after P&R)
 - During minute change in the design after full Tape-out is over
 - Uses Spare cells and metal layers only
 - **Metal Layer ECO**
 - During minute change in RTL after Active/Base Layer Tape-out is over
 - Metal Layer changes only
 - Cleaning-up routing for Signal Integrity (SI)
 - **Active Layer ECO (Base Layer ECO)**
 - During minute change in RTL just after Routing is over
 - Uses Spare Cells
 - NAND Gate (Universal Logic Gate) based Spare Cell can be used to realize the new ECO logic
- Post Silicon ECO (ECO after Fabrication)
 - To recover from minute manufacturing issues
 - Uses Spare cells and metal layers

Types of ECO

- Metal Layer ECO (example)



Types of Standard Cell Libraries

Types of Standard Cell Libraries

- Standard Cell Library Types
 - According to the Density
 - According to the Threshold Voltage (V_{TH})
- Classification according to the Density
 - Ultra High Density (UHD) - 7 Track or 8 Track
 - High Density (HD) - 9 Track
 - High Performance (HP) - 12 Track
- Classification according to the Threshold Voltage (V_{TH})
 - Low VT (LVT) - Fast because of low Gate Delay, but high leakage
 - Standard VT (SVT) or Regular VT (RV)
 - High VT (HVT) - Low leakage, but slow because of high Gate Delay
- Metal 2 pitch is used to calculate the Number of Tracks in different Density Libraries

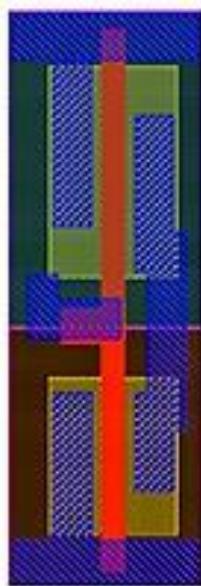
— Sub-threshold Leakage varies exponentially with V_{TH}
compared to the weaker dependency of Delay over V_{TH}

— HVT Cells are used in Non-critical paths to reduce Leakage Power while
SVT Cells are used in Critical paths to meet Timing

Types of Standard Cell Libraries

High Density

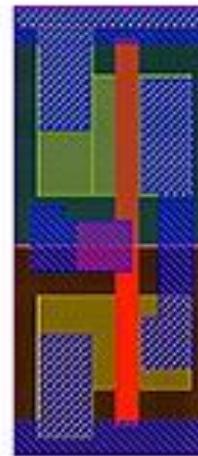
9 or 10-tracks high cells



Balanced transistor size
for high density and good
performance, low power

Ultra-High Density

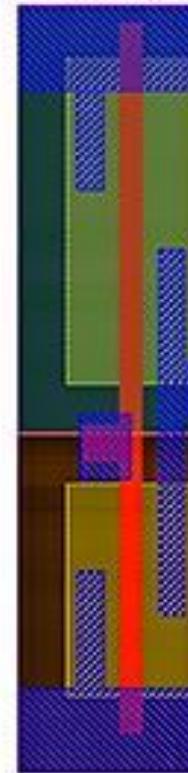
7 or 8-tracks high cells



Small transistors for high
density and low power

High Performance

12-tracks high cells



Large transistors for optimal
speed, but also low power
features

The Discontinuity

SmartPlay Overview

*"To be a leading service provider of **End to End Solutions** enabled by **Innovative Business Models** that provide **Value, Quality and Execution excellence** to our **Customers**"*



Semiconductor
Digital



Analog



Wireless Software



System Design

World-wide Sales

Common Support Functions (HR/Staffing/Ops/Finance)

Common Infrastructure

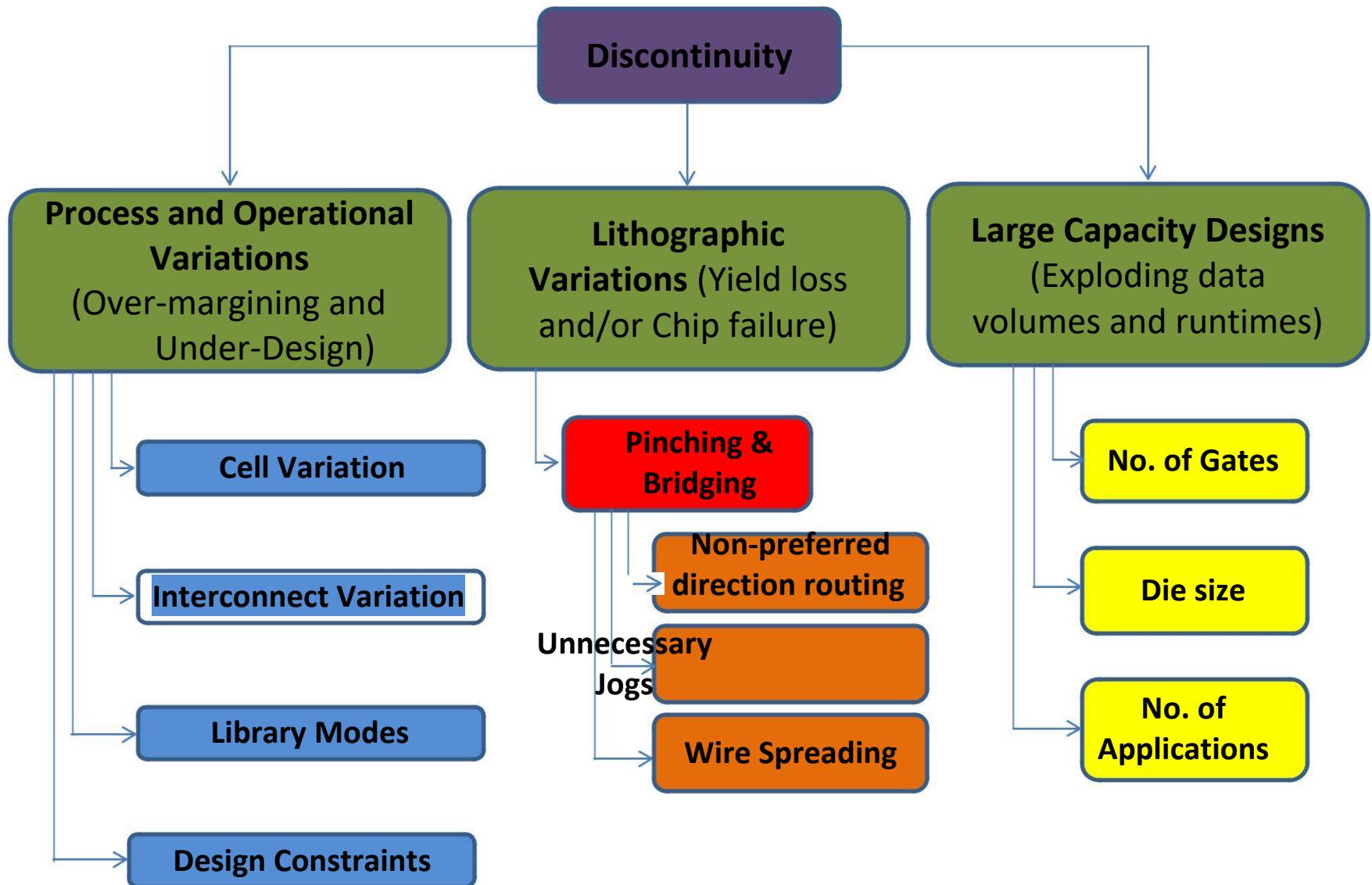
Outline

- **The Discontinuity and its classification**
- - Issues, Need & Rules
 - Resolution Enhancement Techniques
 - Optical Proximity Correction and Scattering Bars
 - Multiple Patterning
 - Phase Shift Masking and Off-Axis Illumination
- **MC/MM/OCV**
 - Corner Analysis
 - PVT/RC Corners
 - Temperature Inversion & Cross Corner Analysis
 - Modes of Analysis
 - Multi-corners/ Multi-modes of Analysis
 - OCV & OCV Enhancements

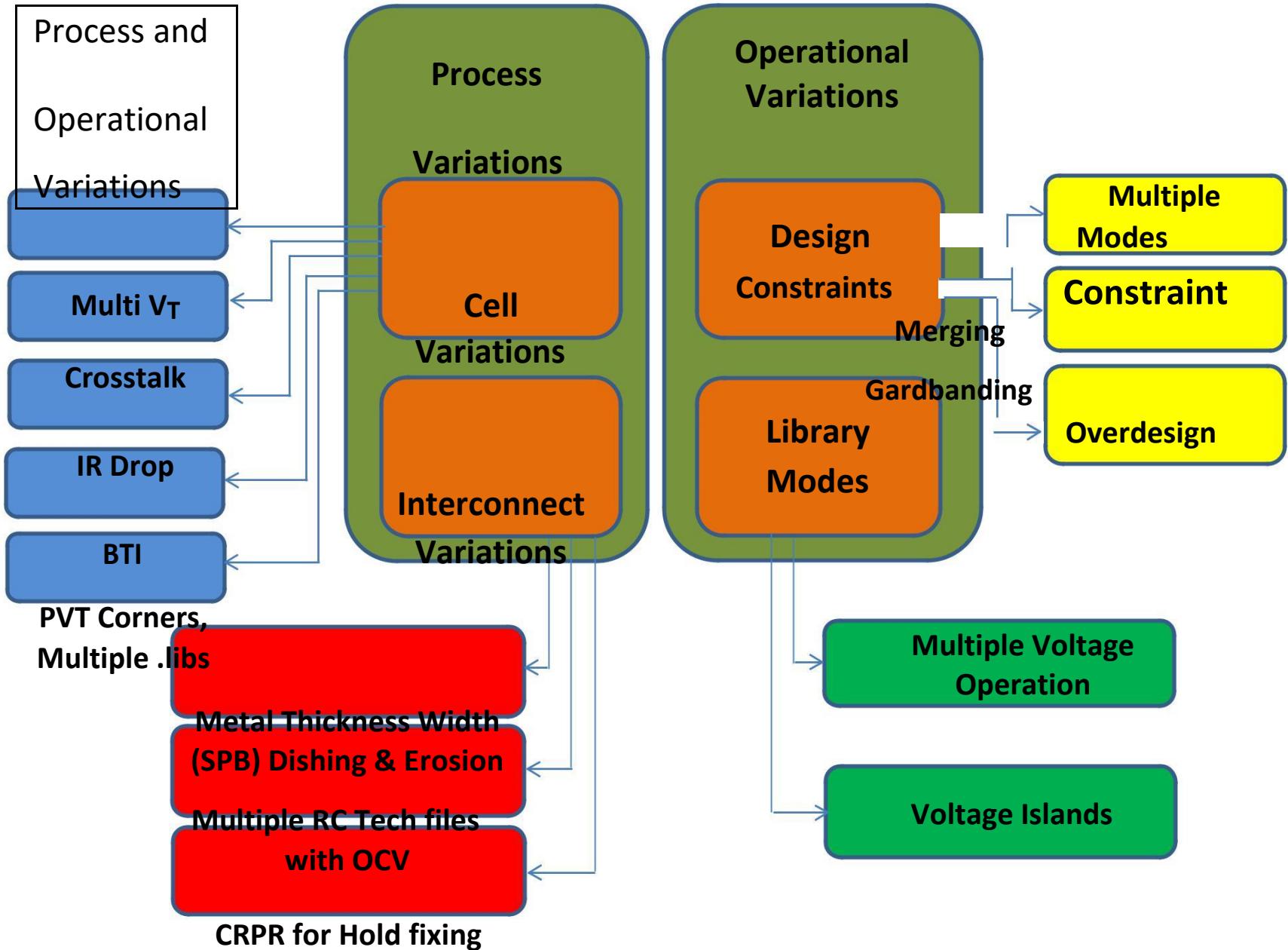
The Discontinuity

- Discontinuity
 - With each new Technology node, previously manageable challenges in physical implementation emerge as extremely disruptive discontinuities
 - At 180nm, timing closure was a disruptive challenge, which led to new physical synthesis technology
 - At 130nm, Signal Integrity (SI) closure was the main discontinuity
 - The new generation of challenges started at 65nm, are in full force at 45nm
 - The challenges will get worse as ICs venture into more advanced Technology nodes like 22/14nm
 - Designers are working at these Technologies to fully understand the new discontinuities
 - Special design enhancements are introduced under the title Design-for-Manufacturability (DFM) and Design-for-Yield (DFY) to overcome these Discontinuities

Discontinuity: Classification



Discontinuity: Classification



- Design for Manufacturability (DFM)/ Design for Yield (DFY)

- Techniques to ensure the design can manufacture successfully with high yield
- To ensures survival of the design, during the complex fabrication process
- Lithography, etch, Chemical Mechanical Polishing (CMP), and mask systematic manufacturing variations surpass random variations as the prime limiters to catastrophic and parametric yield loss

- Yield

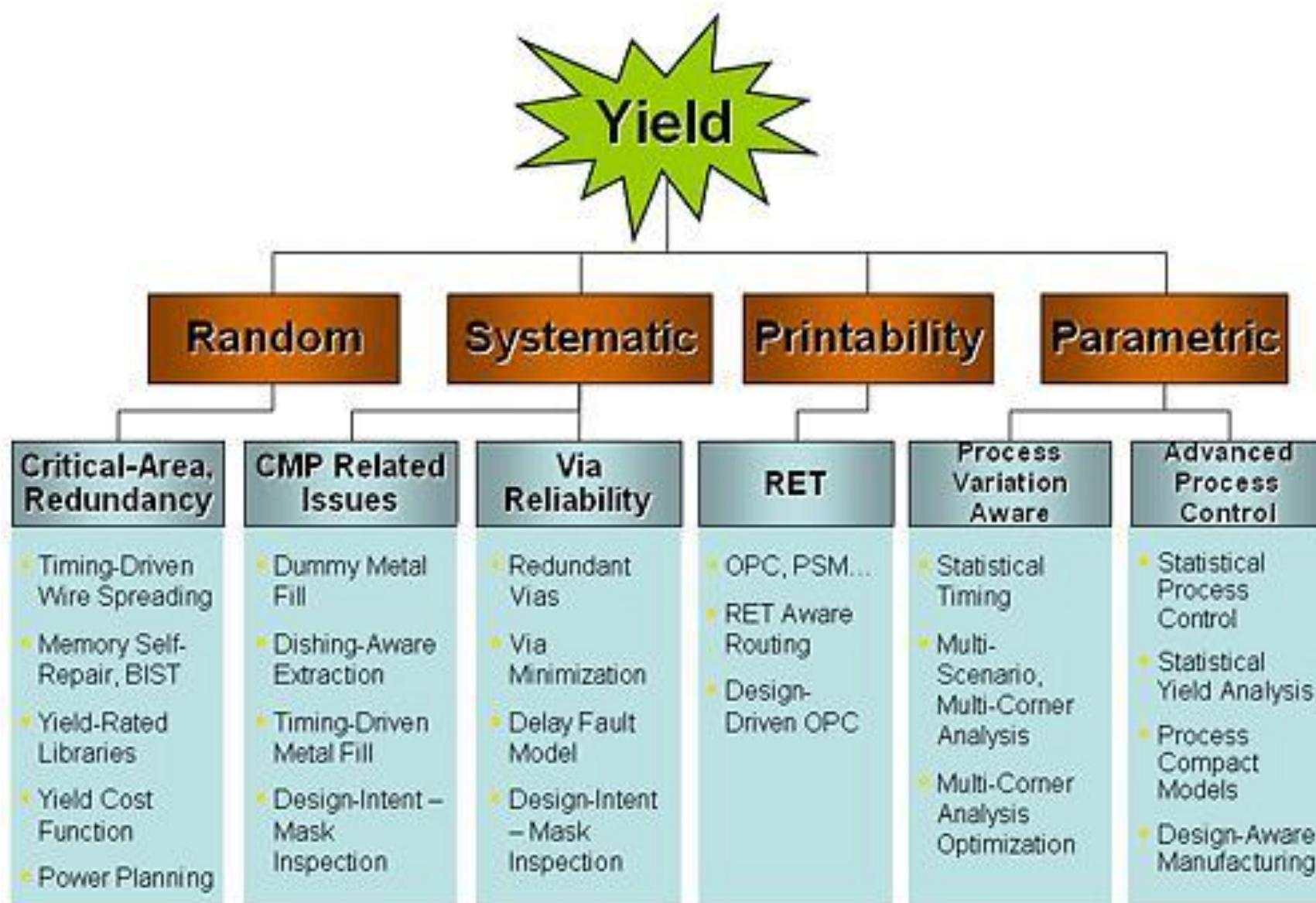
- Percentage of manufactured products that meet all performance and functionality Specifications
- The number of die that work as a percentage of the total number of die on the silicon wafer

$$\text{Yield} = \text{Good Chips} / \text{Total Chips}$$

$$\text{Measured Yield} = \text{Good Parts} + \text{Test Escapes} - \text{False Rejects} / \text{All Parts}$$

- Memory fails more than logic, so repairable memory can improve Yield
- DFY predicts chip yield at two points of the manufacturing flow wafer probe and during final test of the packaged chip and identifies what defects result in yield loss

Yield Classifications



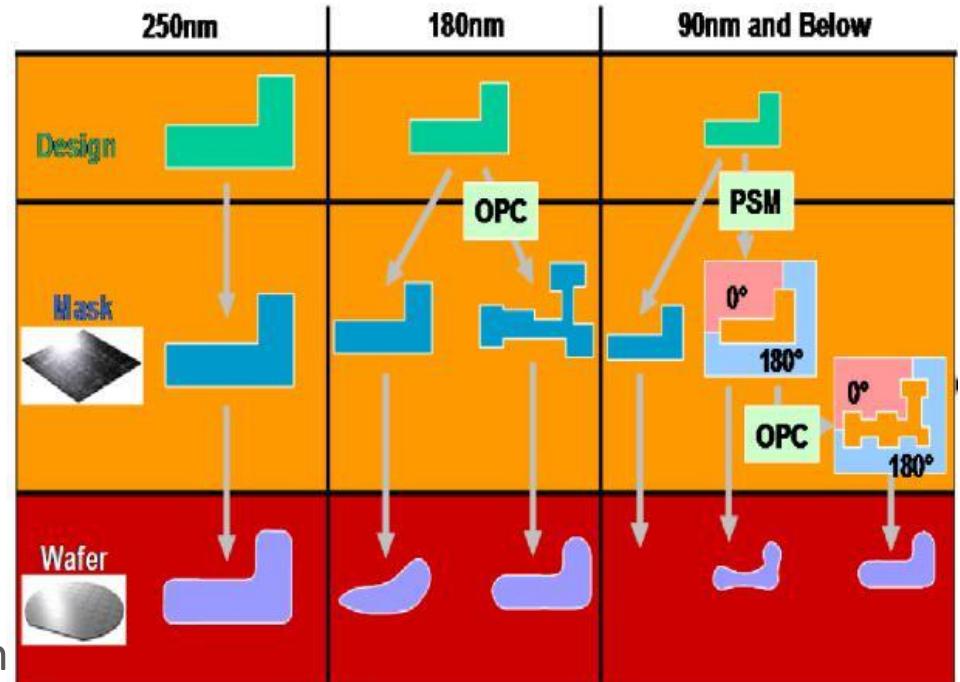
Why ?

\emdash Need for

- Current Lithographic techniques (193nm Laser) cannot print deep-submicron technology patterns without distortion
- Higher design complexity and shrinking device geometries
- More devices per unit area on a chip (device density)

\emdash Importance of

- Impact of variations, if not addressed in the design, will cause manufacturing issues, such as poor yields, long yield ramp-up times and poor reliability
- The chips may completely miss the market window or may hit the market window but not economically viable
- The chips may still function, but not at the required/expected speed
- The chips appear to be reliable after volume production, but may suffer catastrophic failures in the field earlier than their expected life-cycle



Solutions

— DFM: Recommendations

— Wire Spreading

The wire distribution spreads wires that are on the same metal layer as well as across different metal layers

The benefits gained from lower routing density are in improved manufacturing yield, reduced crosstalk noise, crosstalk delay and random particle defects

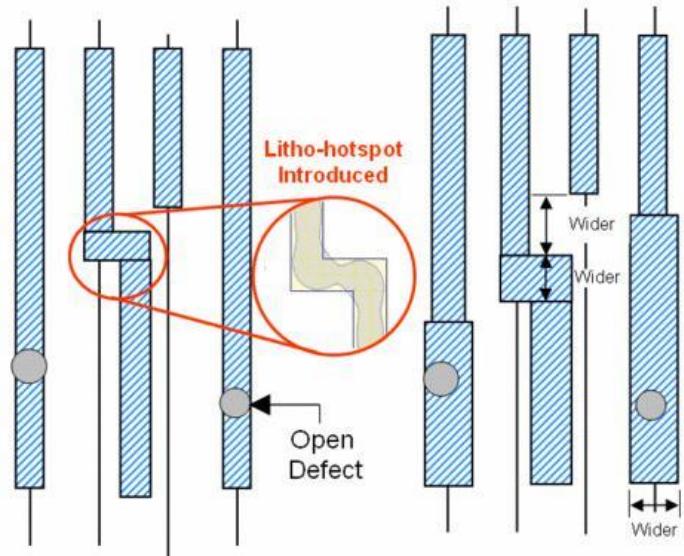
— Metal Fill

Dummy metal fill

Timing aware metal fill

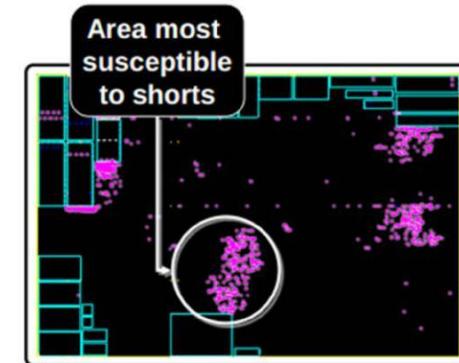
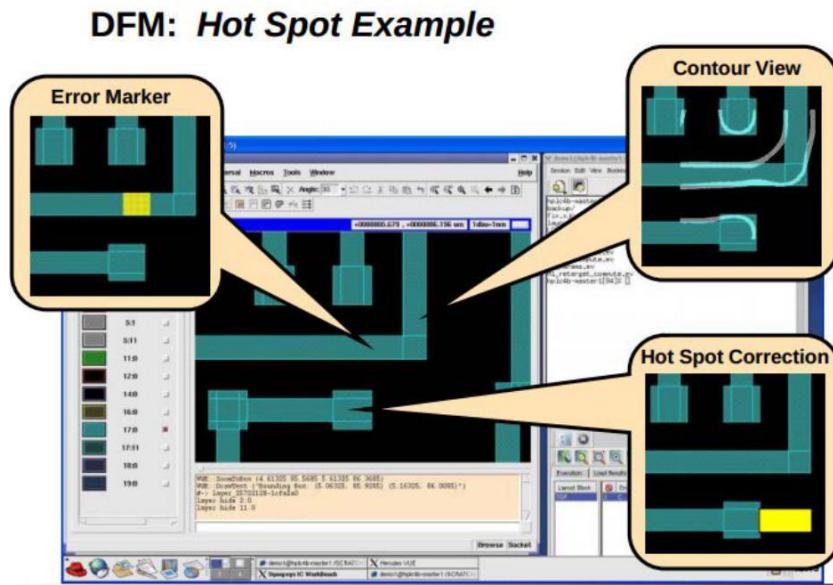
Unbalanced metal density across a chip may cause yield loss, so fill the empty spaces in the design with metal wires to meet the metal density rules required by most fabrication processes

Improved surface planarity helps decrease manufacturing variations that contribute to timing variability



Solutions

- DFM: Recommendations
 - Hot Spots and Critical Area Analysis (CAA)
- Hot Spot/ Critical Area is the region at the center of a random defect which will cause circuit failure (yield loss)
- By analyzing the critical areas, defect-limited yield can be estimated based on the probability of the failures of vias and point defects on routing
- The larger the defect size, the larger the Critical Area
- Critical area reduction improves yield



Solutions

\emdash DFM: Recommendations

- Chemical-Mechanical Polishing (CMP) is a technique for surface smoothing and material removal process to get globally planar wafer surface
- Simultaneous polishing of copper, dielectric and barrier
- Combination of chemical and mechanical interactions

The chemical effect by pH regulators, oxidizers or stabilizers

The mechanical action by submicron sized abrasive particles contained in the slurry flow

between the polishing pad and the wafer surface

- Dishing
 - Difference between the height of the copper in the trench and the height of the dielectric surrounding the copper trench
 - Copper dishing is higher for wider copper line or the spacing
 - It can thin the wire or pad, causing higher-resistance wires or lower reliability bond pads
- Erosion
 - Difference between the dielectric thickness before CMP and after
 - CMP Dielectric erosion is higher for higher density
 - Erosion can result in a sub-planar dip on the wafer surface, causing short-circuits between adjacent wires on next layer
- On-Chip Variation (OCV) from the interconnect thickness variation due to CMP becomes relatively larger and needs to be taken into consideration in the post-layout RC extraction and timing flow
- Solution to CMP is CMP hotspot detection and fixing

Solutions

- DFM: Recommendations

- CMP aware-design

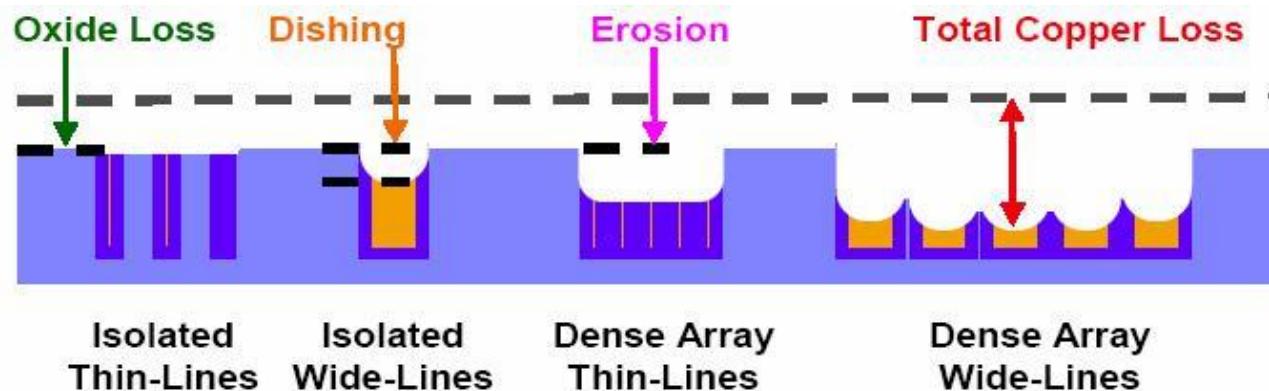
Various degrees of Copper Dishing and Dielectric Erosion occur at different densities and metal line widths

In advanced nodes minimal material removal with atomically flat and clean surface finish has to be achieved

CMP is influenced by line width and pattern density

The dishing and erosion increase slowly as a function of increasing density and go into saturation when the density is more than 0.7

Oxide erosion and copper dishing can be controlled by area filling and metal slotting



Solutions

— DFM: Recommendations

— Redundant Via

Redundant Vias use two, or more, Vias to connect the upper and lower routing layers together

Replacing single Vias with redundant (or double) Vias on signal nets improves reliability and reduce yield loss, due to via failures

Critical Area Analysis (CAA) identifies the requirement of Redundant Vias

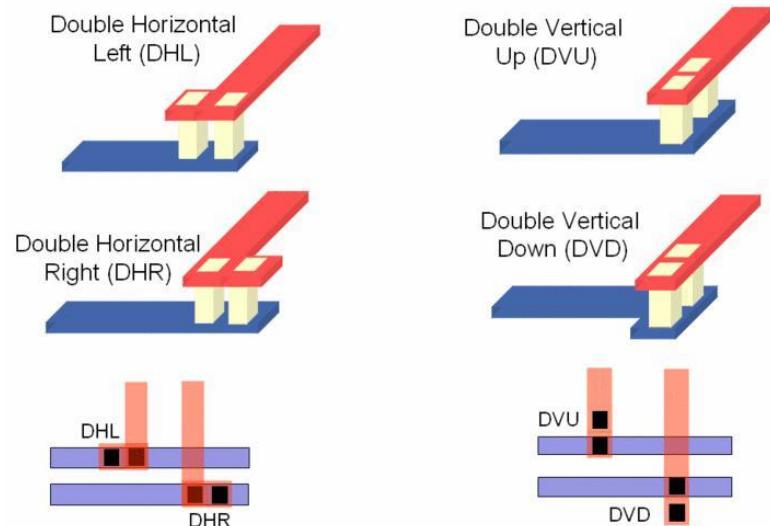
— Resolution Enhancement Techniques (RET)

RET are methods used to modify photo-masks to compensate for limitations in the lithographic processes used to manufacture the chips

Have significantly increased the cost and complexity of sub-micron nanometer photomasks

The photomask layout is no longer an exact replica of the design layout

As a result, reliably verifying RET synthesis accuracy, structural integrity, and conformance to mask fabrication rules are crucial for the manufacture of nanometer regime VLSI designs



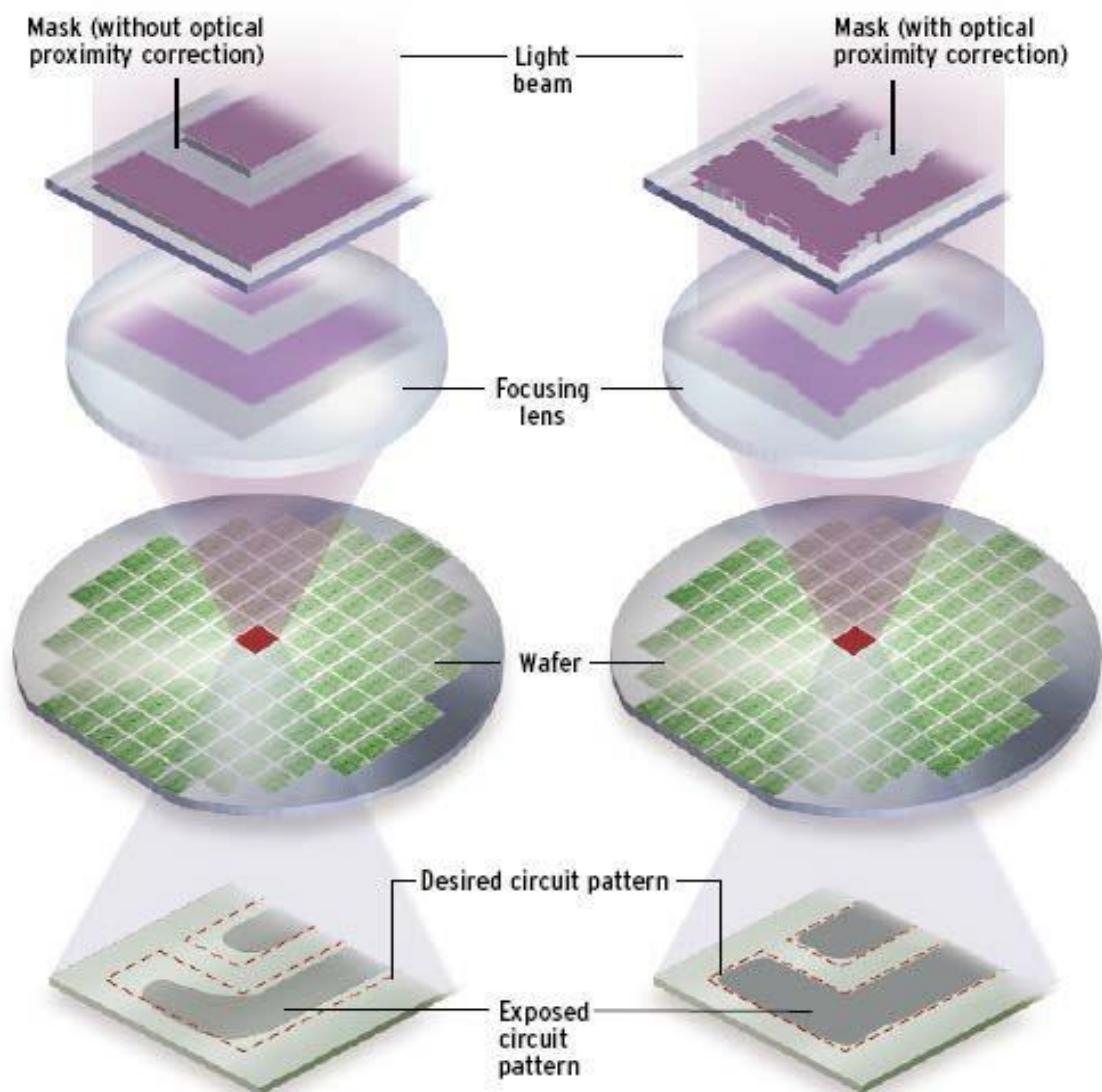
Solutions

- DFM: Recommendations
 - Litho Process Check (LPC)
 - Problem: Some DRC clean layouts do not print on silicon
 - Solution: Must-have litho hotspot detection and fixing of design
 - Layout Dependent Effects
 - Well Proximity Effect (WPE)
 - Poly Spacing Effect (PSE)
 - Length of Diffusion (LOD)
 - OD to OD Spacing Effect (OSE)
 - Layout Patterning Check (LPC)
 - OD/Poly Density

Resolution Enhancement Techniques

— Types of RET

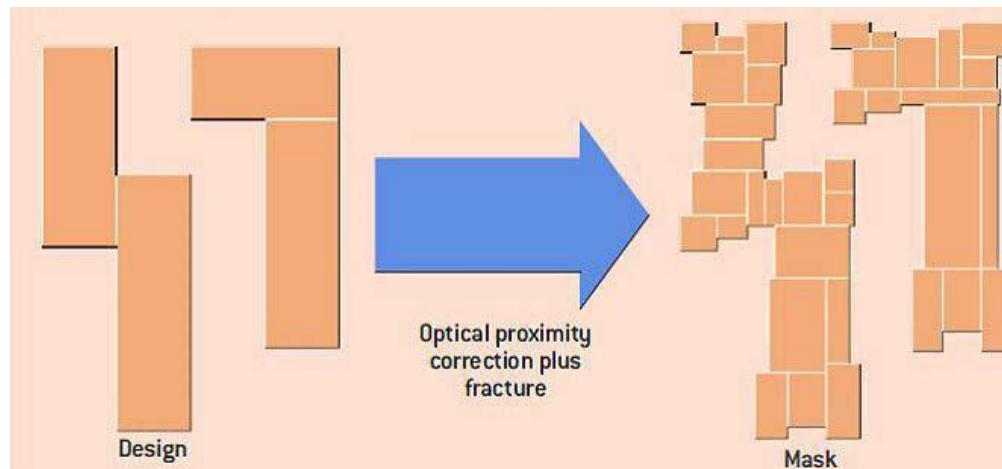
- Optical Proximity Correction (OPC)
- Scattering Bars (SB)
- Double Patterning (DP) or Multiple Patterning
- Phase Shift Masking (PSM)
- Off-axis Illumination (OAI)



Optical Proximity Correction

- Optical Proximity Correction (OPC)

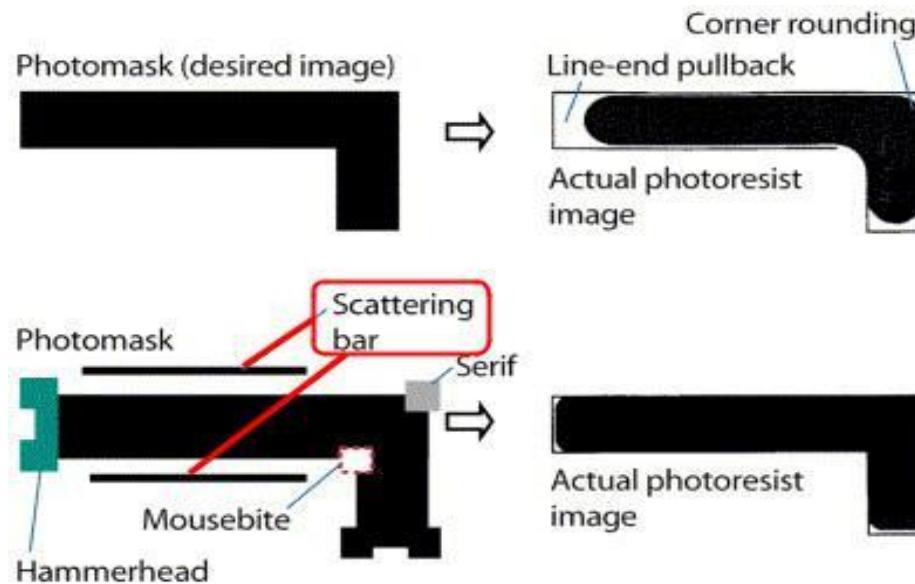
- OPC is a Photo-lithography Enhancement technique commonly used to compensate the mask pattern for image errors due to diffraction or process effects (by reducing the value of the k₁ factor in CD equation)
- OPC is an effective way to deal with geometry distortion from design to chip; however, it does come at a price
- First, there is the cost of the EDA tools you need to implement the OPC corrections
- Second, you have an exponential increase in volume of the data representing the chip's layout, along with a huge increase in the time it takes to process this data and prepare it for photo-mask generation



Scattering Bars

\emdash Scattering Bars (SB)

- Sub resolution assist features that improves the depth of focus of isolated features
- Scattering Bars are added only for the most outer line of the dense pattern



Multiple Patterning

— Multiple Patterning

- Involves decomposing the design across multiple masks to allow the printing of tighter pitches
- 38-nm features with 193-nm light water immersion lithography is the limitation with the current lithographic process
- Multiple Patterning is a technique used in the lithographic process that can create the features less than 38nm at advanced process nodes
- Multiple patterning basically changing the value of K_1 in the Critical Dimension equation
- Double Patterning



❖ ❖ ❖ ❖ ❖

Double patterning counters the effects of diffraction in optical lithography

❖ ❖ ❖ ❖ ❖

Diffraction effects makes it difficult to produce accurately defined deep sub-micron patterns using existing lighting sources and conventional masks

❖ ❖ ❖ ❖ ❖

Diffraction effects makes sharp corners and edges become blur, and some small features on the mask won't appear on the wafer at all

❖ ❖ ❖ ❖ ❖

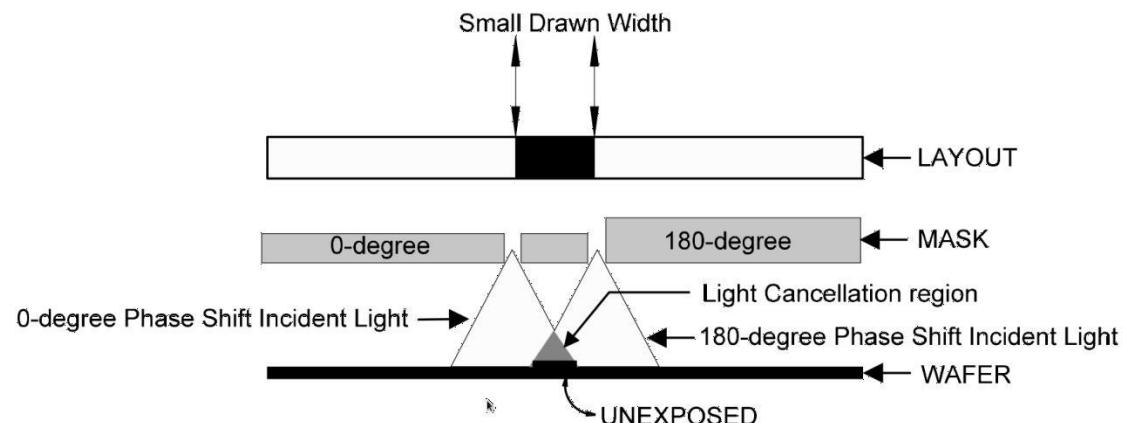
Double patterning is expensive because it uses two masks to define a layer that was defined with one at previous process nodes

Phase Shift Masking

- Phase Shift Masking (PSM) (not considered in PD)

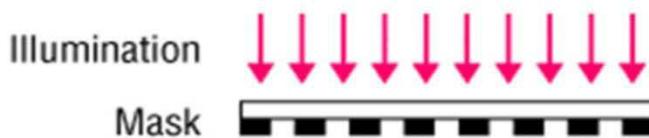
- Phase-shift masks are photo-masks that take advantage of the interference generated by phase differences to improve image resolution in photolithography
- Controlling the phase enables constructive or destructive interference at desired locations in the image plane, thus sharpening or dulling the contrast as desired
- These are photo-masks with structures that manipulate not only the amplitude of the transmitted waves but also their phase
- Etching quartz from certain areas of the mask (alt-PSM) or replacing Chrome with phase shifting Molybdenum Silicide layer (attenuated embedded PSM) to improve CD control and increase resolution
- There exist alternating and attenuated phase shift masks
- Types of masks

❖ Conventional (binary) mask, Alternating phase-shift mask, Attenuated phase-shift mask

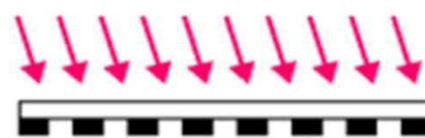


Off-Axis Illumination

- Off-Axis Illumination (OAI) (not considered in PD)
 - Off-axis illumination is one of the practical techniques to enhance resolution of a given optical system with bigger advantage of improvements in depth of focus
 - The specific illumination geometry is designed to enhance the contrast in the wafer plane of the photo-mask features whose dimensions are most Critical
 - With OAI, resolution of a given system can be improved without going for shorter wavelength or higher numerical aperture (NA)
 - This technique basically has no on-axis illumination component as oppose to partial coherence
 - The shape and size of the source plays an important role when different conditions of mask features such as density and orientation are considered
 - To obtain the highest resolution, illumination of the photo-mask is not performed by a disc-shaped source
 - The angular distribution of the illumination beam may have a complex structure, such as an annulus, a set of off-axis circles, or even a continuously varying profile



On-axis illumination



Off-axis illumination

MC/MM/OCV

Corner

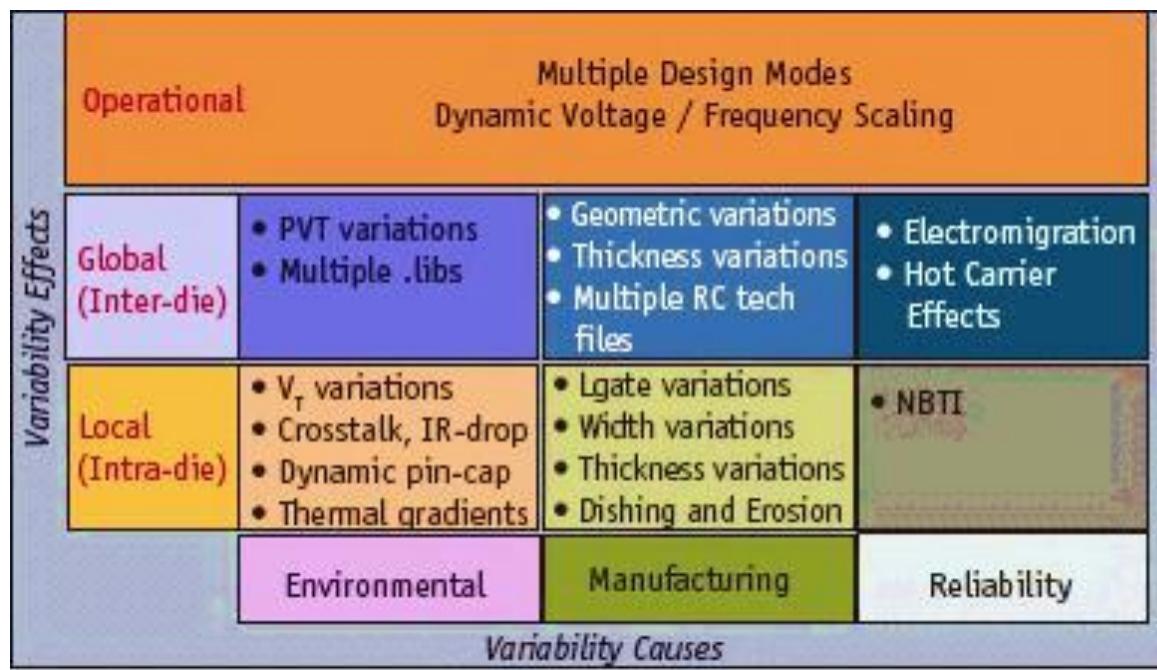
- Corner
 - Characterizes the physical environment for Timing Analysis
 - An extreme point in the PVT/ RC space where cell and net delays have extreme values
 - A particular one cell library and RC-model specified for STA run
 - Corners are meant to capture variations in the manufacturing Process, along with expected variations in the Voltage and Temperature of the environment in which the chip will operate
 - Corners are independent on functional settings
 - As technology shrinks, variations increases since smaller geometries have had a higher variability
 - As a result the number of Corners and Derates also grows

Parameters	Resistance	Capacitance		Remark
		Surface	Coupling	
Temperature ↑	↑	---	---	
Width ↓	↑	↓	---	
Thickness ↓	↑	---	↓	
Space ↓	No Effect	No effect	↑	Space between same metal plate

Corner

\emdash Corner

- It is important to find minimum number of Corners, because run-time and Turn Around Time increases with increased number of Corners
- E.g. run only slow metal at SS for Maximum Frequency
- Also each Corner need its own OCV timing margins
- The more Corners are used, the more pessimistic the timing signoff



Corner

\emdash Corner

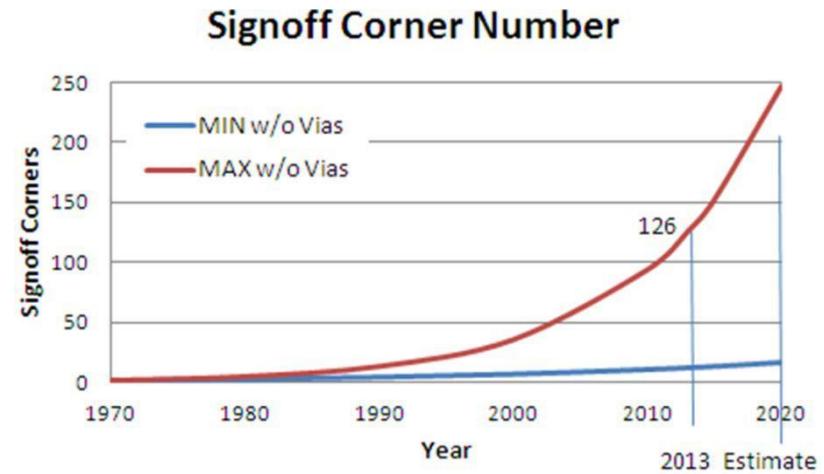
- At each global Corner the Die experiences
 - External Voltage (like Minimum, Maximum, Typical)
 - Temperature (like Minimum, Typical, Maximum)
 - Process Shifts in (independent)
 - Transistors (Slow: SS, Typical: TT, Fast: FF or mixed SF & FS)
 - Interconnects (4 RC-extremes and RC-typical and Via Minimum, Maximum, Typical - Capacitance/ Resistance)
- Vias are independent and not practically correlated with RC-wire models
- Possible Vias models: V_{RCBES} , V_{CBES} , V_{RCWOR} , V_{CWOR} , V_{RCTY}
- Total number of Corners =

$$\{P: SS \& FF \& TT\} \times \{V: Min. \& Max. \& Typ.\} \times \{T: Min. \& Max. \& Typ.\} \times \{RC: R_{C_{BEST}}, C_{C_{BEST}}, R_{C_{WORST}}, C_{C_{WORST}}, R_{C_{TYP}}\}$$

- E.g $3 \times 3 \times 3 \times 5 = 135$ PVT/RC Corners
- By considering Aging Degradation two more corners will come in to picture Beginning-Of-Life (BOL) and End-Of-Life (EOL)

Corner

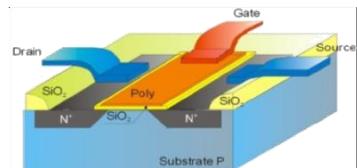
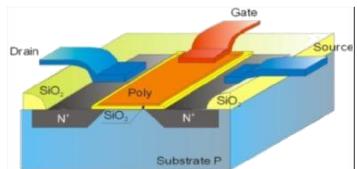
- Corner
 - Even more Corners are needed for advanced nodes due to:
 - Temperature Inversion
 - Non-Linearity in Voltage
 - Designs with multi voltage domains
 - Additional voltages for over- and under-drive design modes
 - DPT (Double Patterning Technology) may add new corners
 - Via Capacitance Corners (additional to resistance corners) due to using wide Vias
 - Using FinFET and 3D structures may also contribute to Corner numbers and may decrease model accuracy
 - Using so many PVT/RC/Via corners will be not acceptable from the design time and costs considerations
 - Additionally, the number of Signoff Scenarios is a product of Corners and Modes (functional, test, etc.) and becomes too big to be handled by the tools



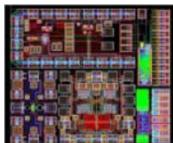
Need for Corner Analysis

Global Variations

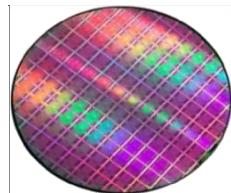
Device to Device



Die to Die



Wafer to Wafer



Lot to Lot



Fab. to Fab.

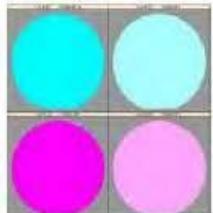


Intra-die

Inter-die

Parametric Variations in the Wafer

Global



Linear



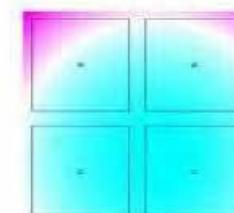
Radial



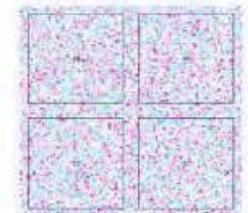
Wafer



Across Reticle

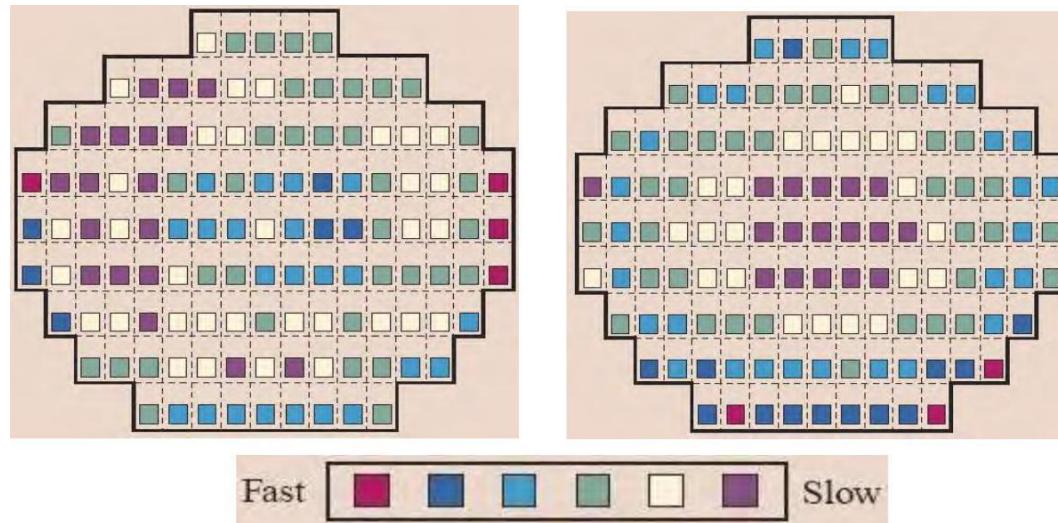


Local

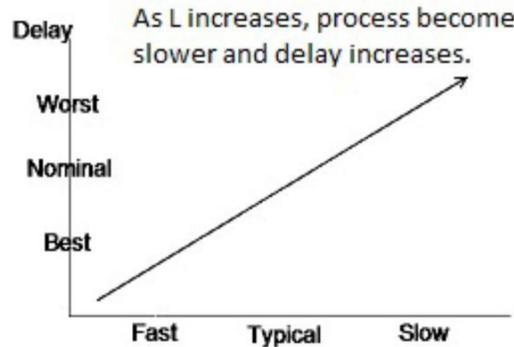


Need for Corner Analysis

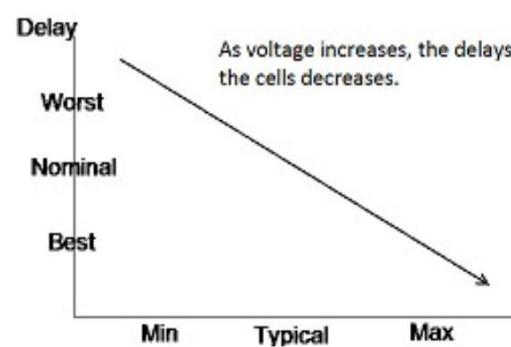
Impact in a Wafer



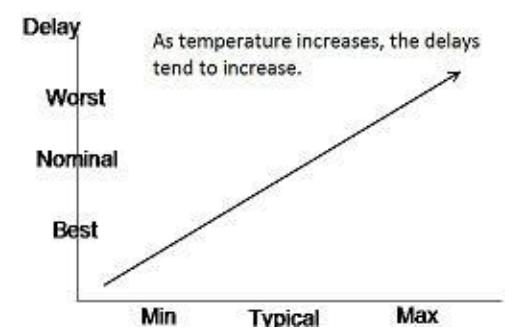
Ideal PVT Plots w.r.t Delay



Process Variations



Voltage Variations



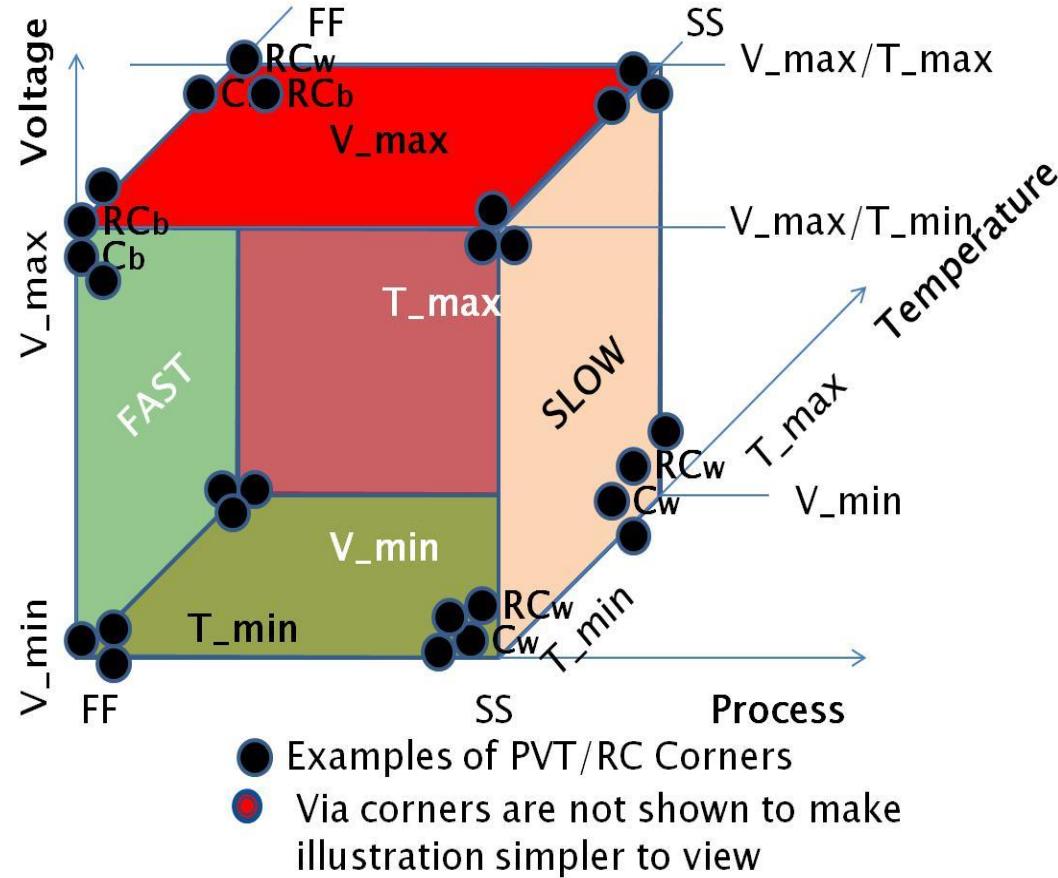
Temperature Variations

PVT Variations

Process	Voltage	Temperature
<ul style="list-style-type: none">Variations makes difference in propagation delaySmaller transistors are faster and have less propagation delayVariations in the process parameters results impurity concentration densities, oxide thicknesses and diffusion depths, sheet resistance and transistor parameters such as threshold voltage or (W/L) variationsWhile manufacturing any die, the dies that are present at the center are accurate in their process values, but the ones lying on the periphery tend to deviate from this process value	<ul style="list-style-type: none">A higher voltage increases current and decreases propagation delayThe voltage drop is due to nonzero resistance in the supply wiresThe self-inductance of a supply line also contributes to a voltage dropThe saturation current of a cell depends on the power supply and thus the propagation delayThroughout the chip also the power supply may vary and hence the propagation delay also variesThe voltage might not be a constant over a period of time	<ul style="list-style-type: none">A higher temperature will decrease the threshold voltage, which results in higher current and lesser delayMore the temperature, lesser the electron and hole mobility and increased propagation delayHigher temperature results increased electron scattering, which in turn reduces current and thus the delay increasesWhen a chip is operating, the temperature can vary throughout the chip due to the power dissipation due to switching, short-circuit and leakage power consumption

Corner Analysis

- PVT/RC Corners



Corner Analysis

- RC Corners

- C_{BEST}

It has minimum capacitance. So also known as C_{MIN} corner

Interconnect Resistance is larger than the Typical corner

This corner results in smallest delay for paths with short nets and can be used for min-path-analysis

- C_{WORST}

Refers to corners which results maximum Capacitance. So also known as C_{MAX} corner.

Interconnect resistance is smaller than at typical corner

This corner results in largest delay for paths with shorts nets and can be used for max-path-analysis

- RC-BEST

Refers to the corners which minimize interconnect RC product. So also known as RC-MIN corner

Typically corresponds to smaller etch which increases the trace width. This results in

smallest resistance but corresponds to larger than typical capacitance

Corner has smallest path delay for paths with long interconnects and can be used for min-

path-analysis

- RC-

Refers to the corners which maximize interconnect RC product. So also known as RC-MAX corner

Typically corresponds to larger etch which reduces the trace width. This results in largest resistance but corresponds to smaller than typical capacitance

Corner has largest path delay for paths with long interconnects and can be used for max-

path-analysis

- Typical

This refers to nominal value of interconnect Resistance and Capacitance

Temperature Inversion

- Temperature Inversion Dependence
 - A problem first described by Vassilios Gerousis of Infineon Technologies in 2003
 - Current, $I = K \cdot \mu \cdot (V_{GS} - V_{TH})^2$; where mobility (μ) and Threshold Voltage (V_{TH}) are functions of Temperature
$$\mu = \mu_0 (T/T_0)^{\alpha_\mu} \quad V_T(T) = V_{T0} + \alpha_{V_T} (T - T_0)$$
 - At high voltage μ determines the Drain current where as at lower voltages V_{TH} determines the drain current
 - So at higher voltages device delay increase with temperature but at lower voltages, device delay decreases with temperature
 - At advanced Technology Nodes though the Threshold Voltage has not reduced much, but the Gate Overdrive Voltage has reduced due to the reduction of supply voltages
 - Therefore, Temperature Inversion Effects are more observed in Technology Nodes below 40nm

Cross Corner Analysis

- Cross Corners
 - The consequence of Temperature Inversion is that the actual worst case for delay can occur at a temperature different from the highest temperature
 - E.g., as high- V_T , low-leakage cells get colder they do not speed up in the way that circuits built around faster low- V_T transistors do
 - The reason being that unlike the older technologies where Process, voltage, temperature (PVT) conditions are chosen with highest temperature to be the worst conditions for synthesis and P&R timing closure which is not true now
 - As a result the worst corner is not always easy to predict thus we need Cross Corners to identify the worst corner
 - The designers have to take into account the libraries corresponding to the lowest temperature PVT due to the temperature inversion effects
- The Two Corner Analysis
 - Late (setup) analysis at weak, minimum voltage, high temperature conditions
 - Early (hold) analysis at strong, maximum voltage, low temperature conditions

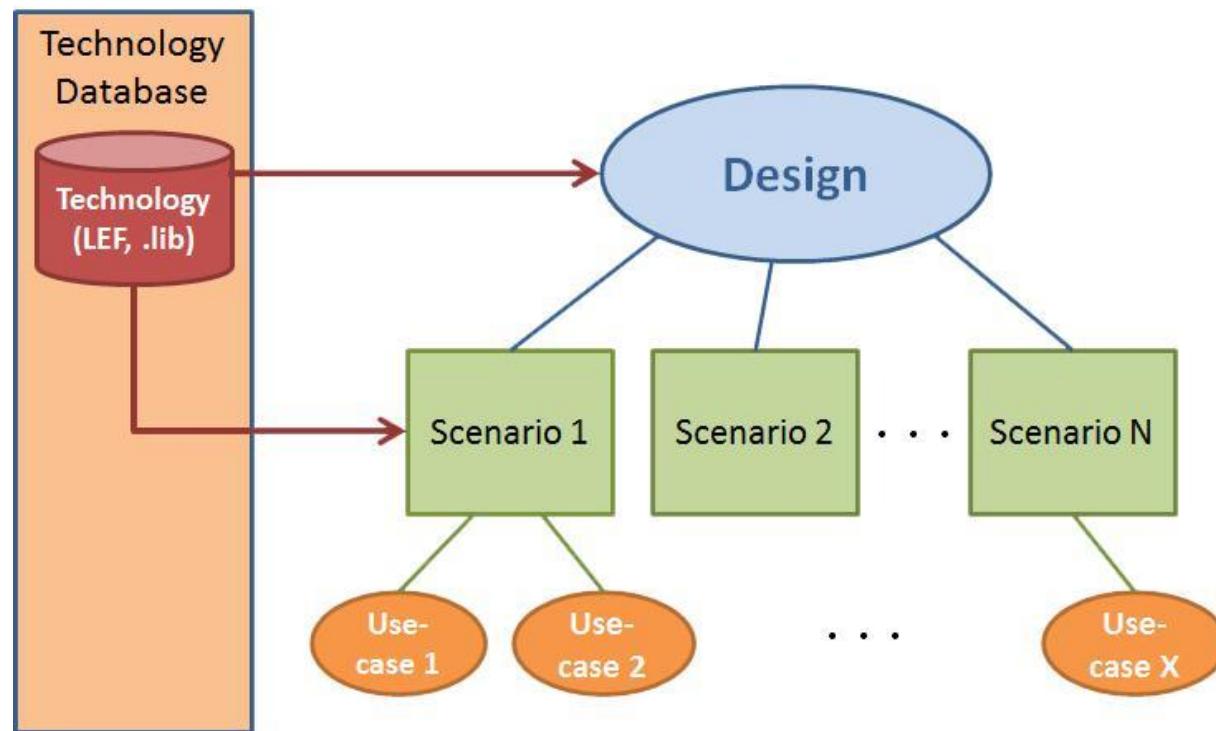
Modes of Analysis

- Modes
 - A Mode is defined as an operational setting of the chip
 - Mode is linked to a unique set of timing constraints
 - Mode can be associated with a set of corners to include only real combinations
 - Mode data is found in .sdc
- Common Operational Modes
 - High-speed clocks mode
 - Slow clocks mode
 - Sleep mode
 - Debug mode
 - Scan capture mode
 - Scan shift mode
 - LBIST mode
 - JTAG mode
 - MBIST mode

MC/MM Analysis

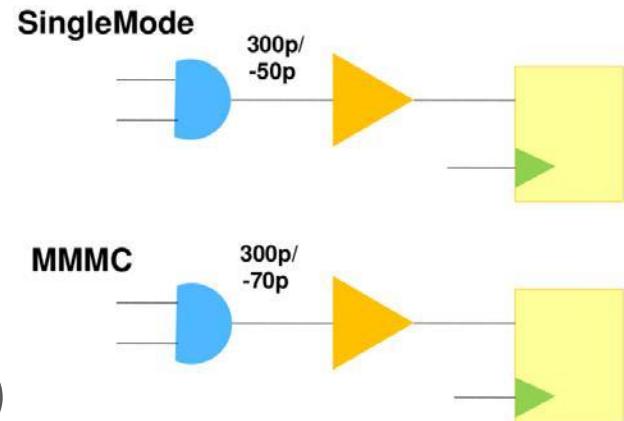
\emdash Scenarios

- A severely limited Corner/Mode views that combines the worst-case parameters to run multiple extraction/timing analysis
- Mode or Corner or a combination of both analyzed and optimized
- E.g. Functional Mode - Slow Corner (func_setup_ss_0.9v_125c)
- E.g. Logic BIST Mode - Fast Corner (lbist_hold_ff_1.1v_m40c)



MC/MM Analysis

- Multi Corner (MC)/ Multi Mode (MM) Analysis (Multi-Scenario)
 - A technique intended to provide high confidence results for timing and other metrics without performing exhaustive simulation of all possible IC conditions
 - MCMM needed because of multiple dominant corners
 - MCMM eliminates the situation where a Hold fix in one mode can break the Setup in the other Modes
 - MCMM helps to avoid switching between different Corners/Modes to fix Setup/Hold violation
 - Avoids over fixing/ under fixing a Hold violation in a particular Corner
 - Reduces Hold buffer count
 - Reduce number of manual timing ECOs
 - Faster design closure
 - Helps in reducing the pessimistic margins and so is also called as Design-for-Variability (DFV)
 - Performed as concurrent analysis & optimization
 - Multi-corner analysis to examine the effects of process and environmental variations as well as changes caused by shifts into different operating modes
 - MCMM is the terminology by Synopsys & MMMC is the terminology by Cadence

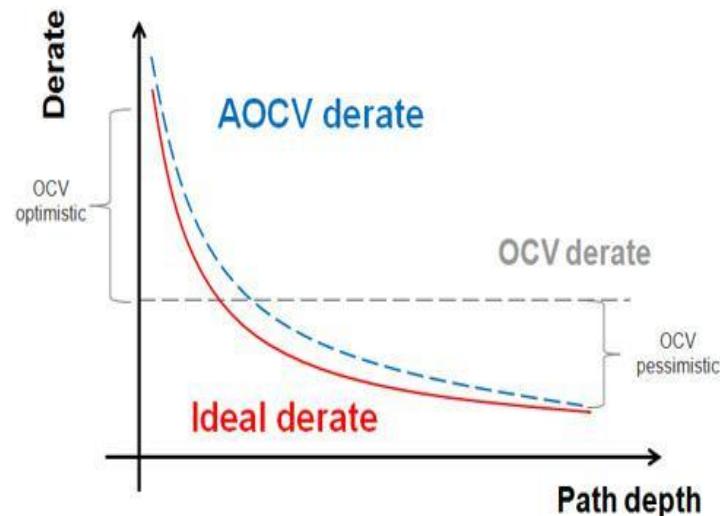


OCV

- On-Chip Variation (OCV)
 - On-chip variation (OCV) is a recognition of the intrinsic variability of semiconductor processes and their impact on factors such as logic timing
 - The number of contributors to timing variability has increased and led to significant variations not just between wafers but across individual wafers and increasingly intra-die
 - ICs from one batch of wafers being ‘slow’ or ‘fast’ relative to nominal estimates
 - Initially, timing analysis accounting for OCV was handled by telling the STA tool to apply a global margin (derate) across the entire chip using a percentage or delay estimate that the designer or the foundry considered safe
 - Timing variation was primarily a consequence of subtle shifts in manufacturing conditions that would lead to ICs from one batch of wafers being ‘slow’ or ‘fast’
 - OCV provides a single derating factor for all instances, so the results can be grossly optimistic or pessimistic
 - So OCV may lead to performance degradation while closing the timing
 - OCV handles global variations with Corners (best case, nominal, and worst-case combinations)
 - The biggest challenge in OCV variations is handling the local uncorrelated variables

OCV Derating

- Derating
 - Derating is a way to model slow and fast signals in On-Chip-Variation (OCV)
 - It is an extra pessimism added in Static Timing Analysis, in order to account for the On -Chip Variation effects
 - 10% derate in simple terms means, over designing the timing by 10%
 - So that chip will work at the desired frequency, even if there is a variation effect across the die
 - Scaling factors can be set independently for data paths, clock paths, cell delays, net delays, and cell timing checks
 - Early and late derates applied to launch paths and capture paths depending upon Setup/Hold Analysis
 - Maximum and minimum derating means to multiply the original timing library delay values by the derate value
 - Derating decreases as process matures
 - E.g. For 65nm designs at earlier days 15% derates added but now a days only 5% derates need to be added



OCV Timing Checks

- Scaling factors can be set independently for data paths, clock paths, cell delays, net delays and cell timing checks
- Early and late derates applied to Launch Paths and Capture Paths depending upon Setup/Hold Analysis
- Setup Check with OCV
 - Maximum possible data arrival is determined by taking the maximum delays along the clock path to the start-point register and the maximum delays along the slowest data path from the start-point register to the endpoint register
 - The earliest possible clock arrival at the end-point register is determined by taking the minimum delays along the clock path to the end-point register
- Hold Check with OCV
 - For hold check, we use min delays for the clock path to the start-point register, min delays through the shortest data path, and max delays for the clock path to the end-point register

OCV Enhancements

- Advanced OCV (AOCV)
 - Uses context-specific derating instead of a single global derate
 - value Reduce design margins and lead to fewer timing violations
 - Determines derate values as a function of logic depth and relative cell or net location
 - As a function of cell depth it gives less pessimistic margins to the path
 - Corrects pessimism and optimism in timing derate by accurately modeling variance
 - Sometimes referred to as Location-based OCV or Stage based OCV
 - Stage based OCV is a systematic correction to liberty timing models for on chip variation based on the logic depth of a path
 - Logic depth and location based approach deals based approach with systematic effects
 - Advanced OCV computes the length of the diagonal of the bounding box that contains the cells being analyzed to select an appropriate derate value from the table constructed by test-chip results
 - Global variations cancel out over long distances
 - For data path derate is a measure of statistical delay/ Corner delay For clock path derate is a measure of slew

OCV Enhancements

- Advanced OCV (AOCV)
 - AOCV table generation is independent of the methodology
 - AOCV table can be easily adapted to tools and is companion to .lib
 - AOCV tables have derate values for each cell for different depths (path length)
 - AOCV Derates are defined by analyzing the ratio of delay at the global corner with local variance to a fixed corner
 - AOCV defines 8 derate values for each cell at each depth

Setup	Hold
Late data rise	Early data rise
Late data fill	Early data fall
Early clock rise	Late clock rise
Early clock fall	Late clock fall

OCV Enhancements

\emdash Statistical OCV (SSTA modeling)

- Statistical OCV (SOCV) is a simplified approach to SSTA that uses a single local variable as Derate
- It is also referred as Parametric OCV (POCV)
- It takes elements of SSTA and implementing them in a way that is less compute-intensive
- It solves the major limitations of AOCV, including variation dependency on slew and load and the assumption that the same cell, or load, is in the path
- It combines delay variations in Cells, Wires and Vias
- It promises near SSTA accuracy for a small additional cost of runtime and memory compared to AOCV
- It can include signoff-accurate signal integrity (SI) analysis
- Handles DPT and some other dynamic effects in a conservative static way
- It ignores correlations and number of timing paths
- SOCV is much more accurate than AOCV, especially for graph-based analysis
- SOCV can be validated with SPICE Monte Carlo Analysis

CRPR/ CPPR

- Common Path Pessimism (CPP)
 - Applying different derating for the Launch and Capture Clock is overly pessimistic
 - The Clock Tree will be at only one PVT condition, either as a maximum path or as a minimum path (or anything in between) but never both at the same time
 - CPP is the delay difference along the common portion of the Clock Tree due to different deratings for Launch and Capture Clock Paths
 - Pessimism caused by different derating factors applied on the common part of the Clock Tree is called Common Path Pessimism (CPP)/ Clock Re-convergence Pessimism (CRP) which should be removed during the analysis

CRP or CPP = (maximum clock delay or skew) - (minimum clock delay or skew)

- Common Path Pessimism Removal (CPPR) or Clock Re-convergence Pessimism Removal (CRPR)

- Both CPPR and CRPR are removal of artificially introduced pessimism between the Launch Clock Path and the Capture Clock Path in timing analysis
 - CPPR - terminology by Cadence
 - CRPR - terminology by Synopsys

Thank You



