Introduction to VLSI

ECE 261/461

Prof. Eby Friedman

Lecture # 14

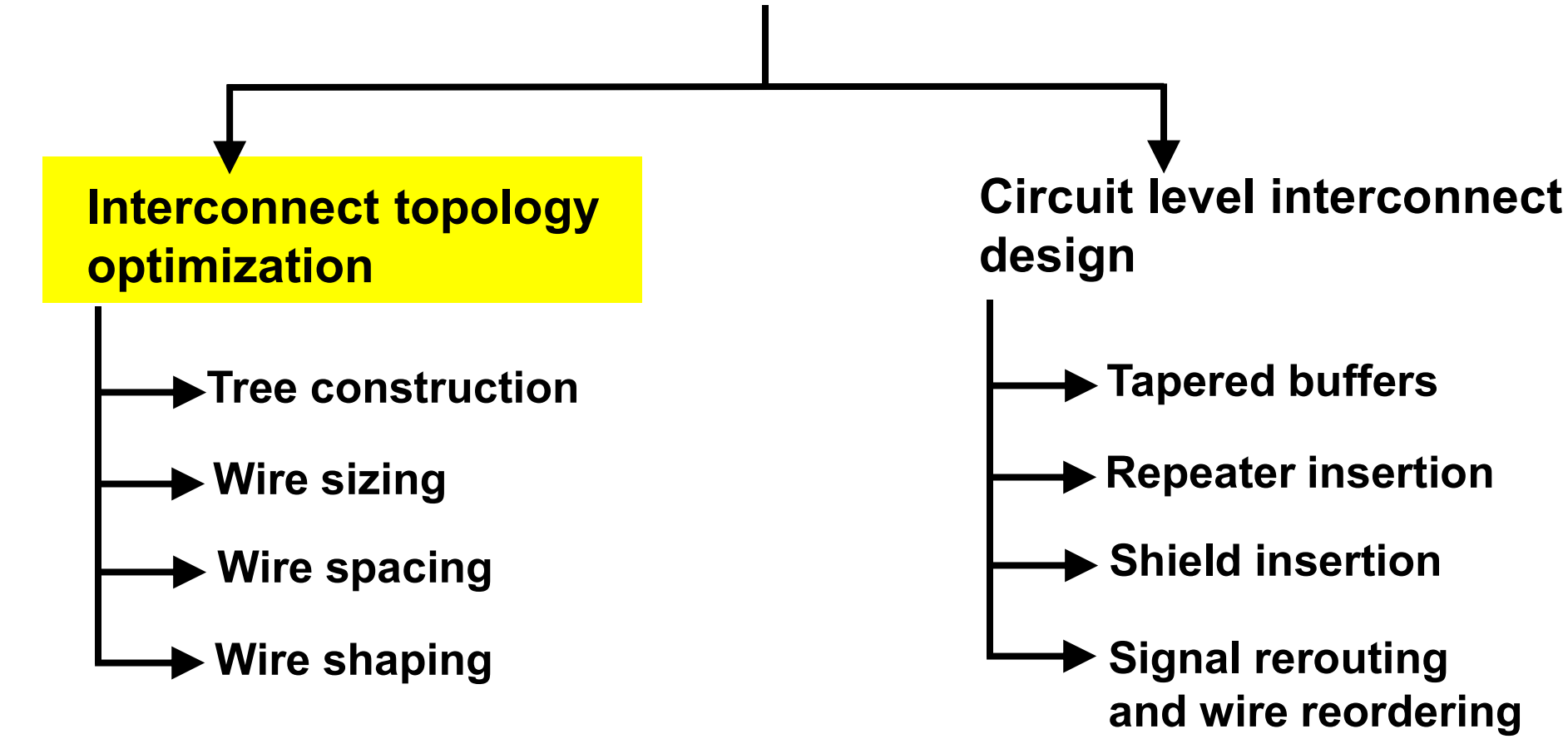
Global Signaling

UNIVERSITY of ROCHESTER

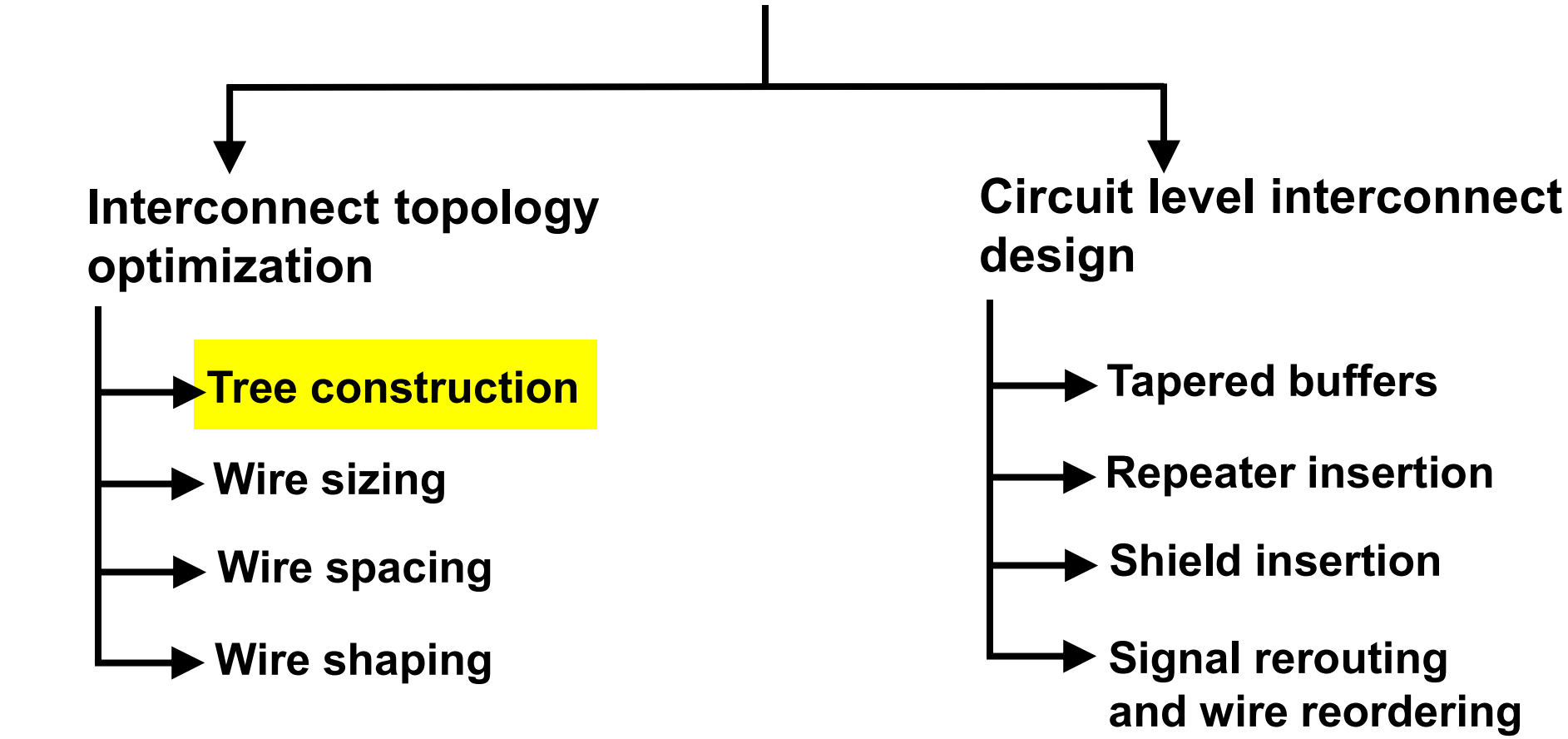
What Are Global Signaling Methodologies?

- Reliable transfer of signals from one block to another block under power, delay, and noise constraints
- Global signaling refers to a set of *interconnect design* methodologies at different abstraction levels
  - To satisfy specific performance requirements
    - Power dissipation
    - Latency
    - Noise
    - Physical area
    - Reliability

Global Signaling Methodologies



Global Signaling Methodologies

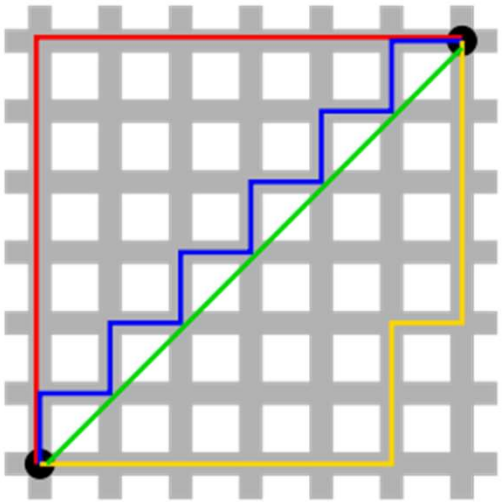


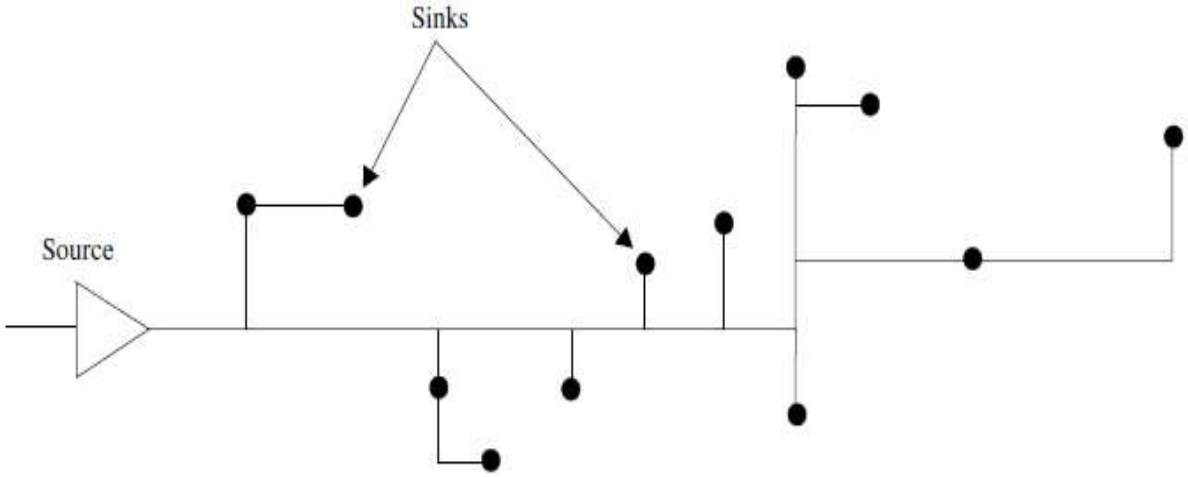
Interconnect Tree Construction

- Tree network is a common structure
- Signals transmitted from root to each leaf
  - Advantage: Existing tree algorithms applied to interconnect optimization
- If circuit is dominated by gates
  - Interconnect modeled as lumped capacitance
  - Minimum rectilinear steiner tree (MRST) used to minimize the *total wire length*
    - Speed and power is minimized
    - Rectilinear: only right angles permitted
- If circuit is dominated by interconnects
  - Interconnect impedance should be considered
  - MRST produces different delays at different sinks
  - Different tree construction techniques maximize slack at each sink

Tree Construction in Interconnect Dominated Circuits

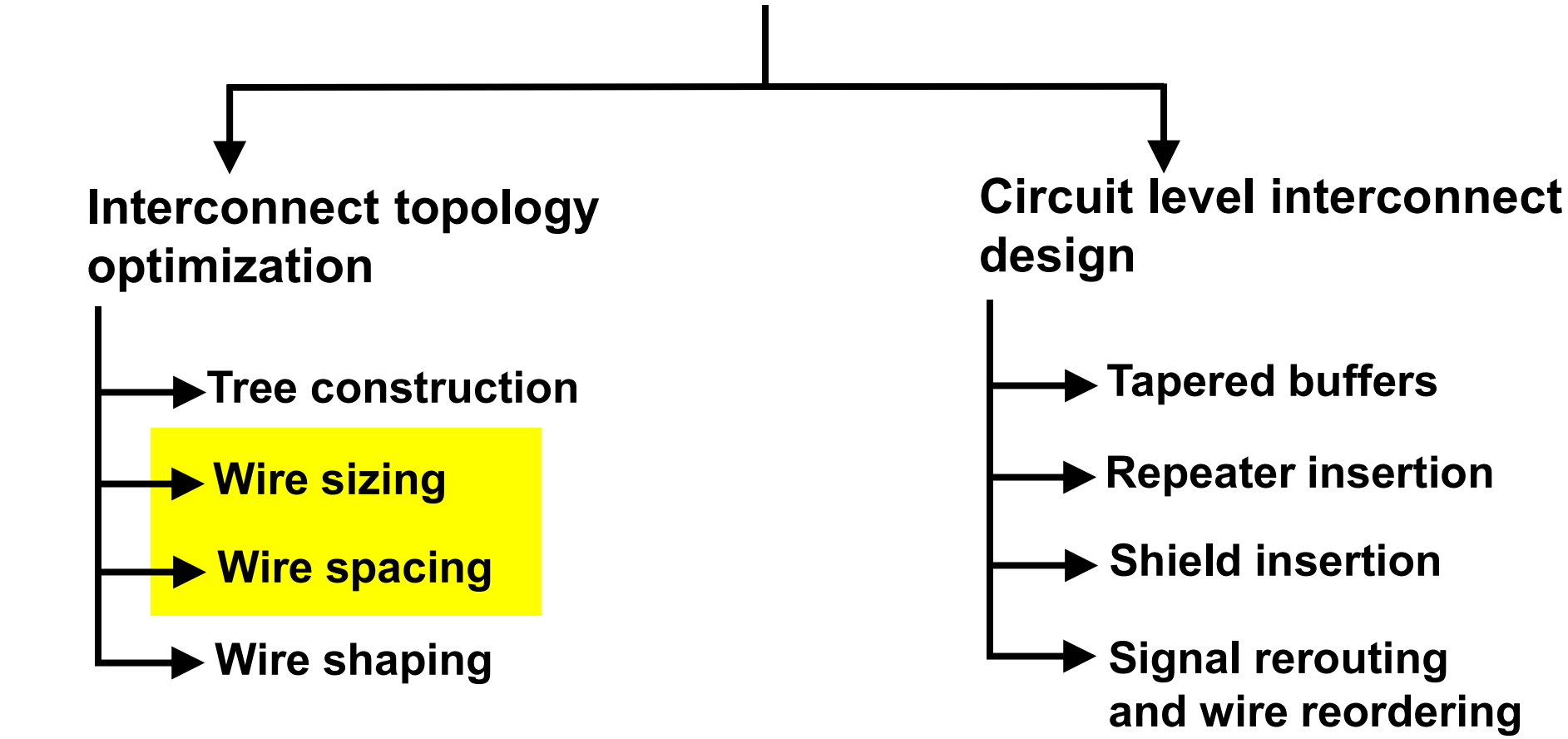
- A-Tree, P-Tree, C-Tree
- A-Tree is a rectilinear tree where the Manhattan distance from each source to sink is minimized
  - Manhattan distance:





  - Subject to this constraint, the total wire length is also minimized

Global Signaling Methodologies

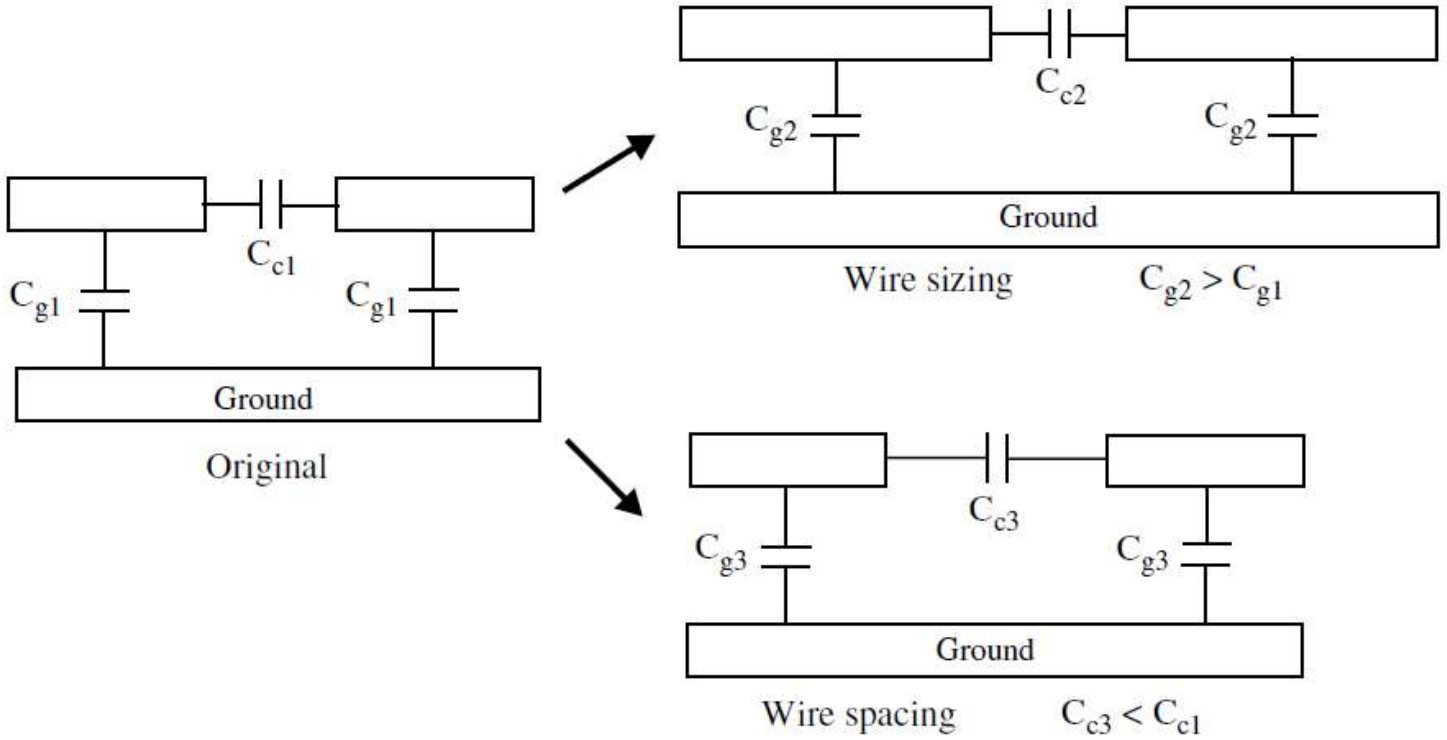


Interconnect Wire Sizing and Spacing

- **Increasing interconnect width and spacing between two wires are common design techniques to reduce coupling noise**

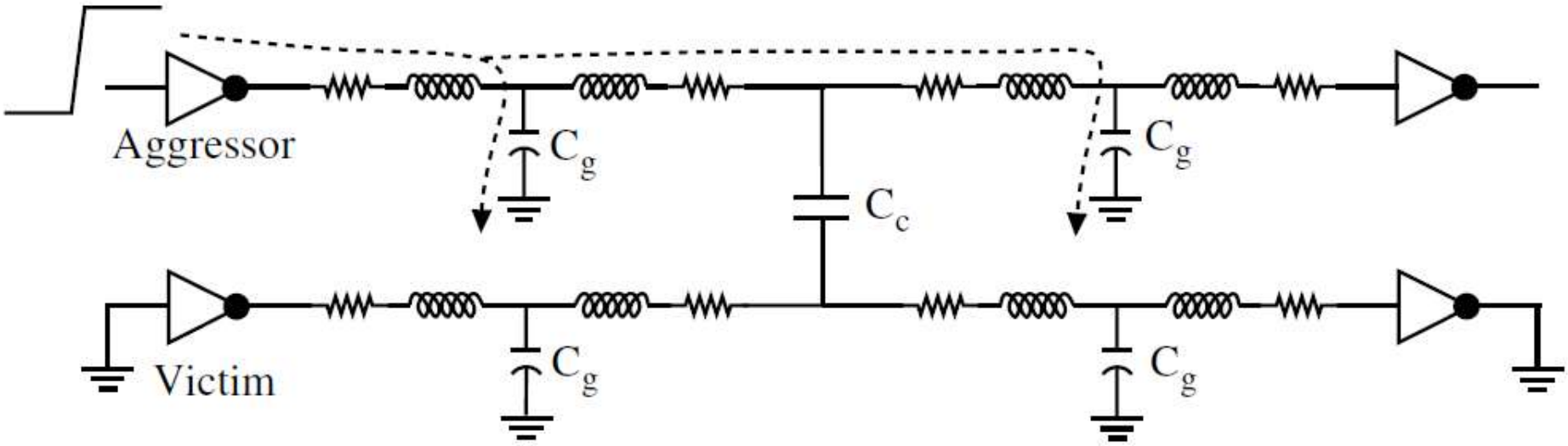
- **Thickness and inter-layer dielectric specified by technology**
- **Wire width and spacing are the *design parameters***
  - **Can be varied to satisfy different design criteria**
  - **Exploit tradeoffs among delay, bandwidth, power, and area**

Wire Sizing and Spacing to Reduce Crosstalk

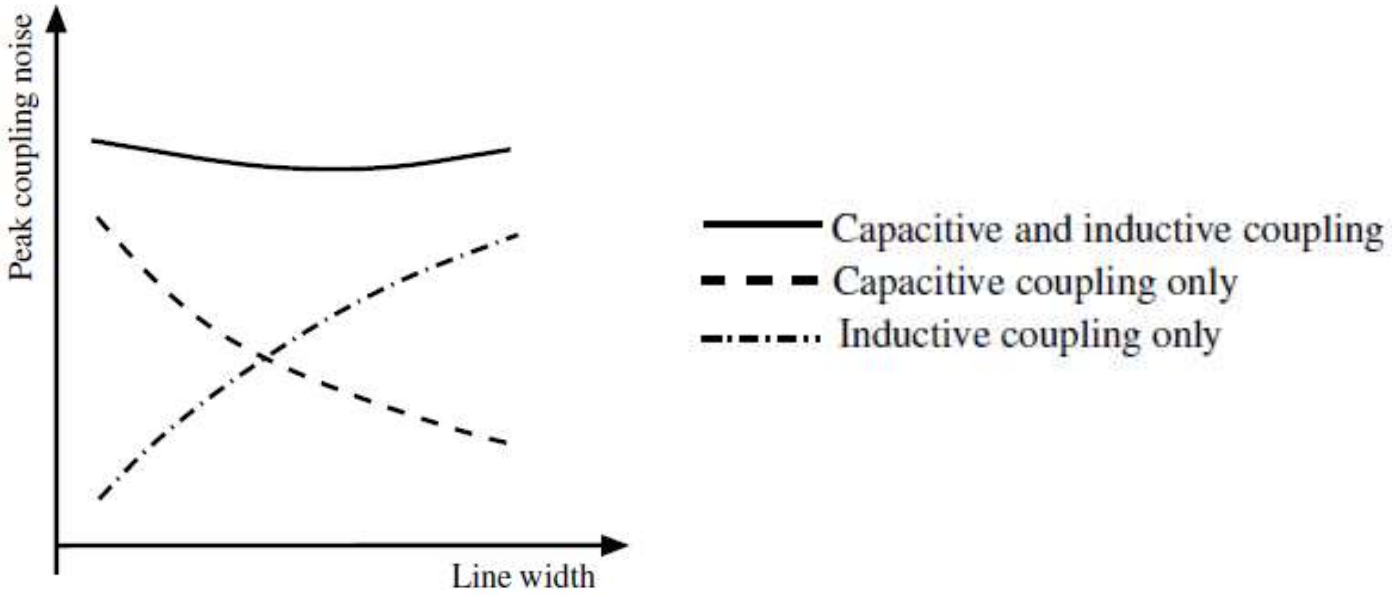


- Increased ground capacitance reduces coupling by behaving as a filter (increasing the width)
  - Disadvantages: increased area and delay
- Increased spacing between conductors reduces coupling capacitance

Effect of Increased Width



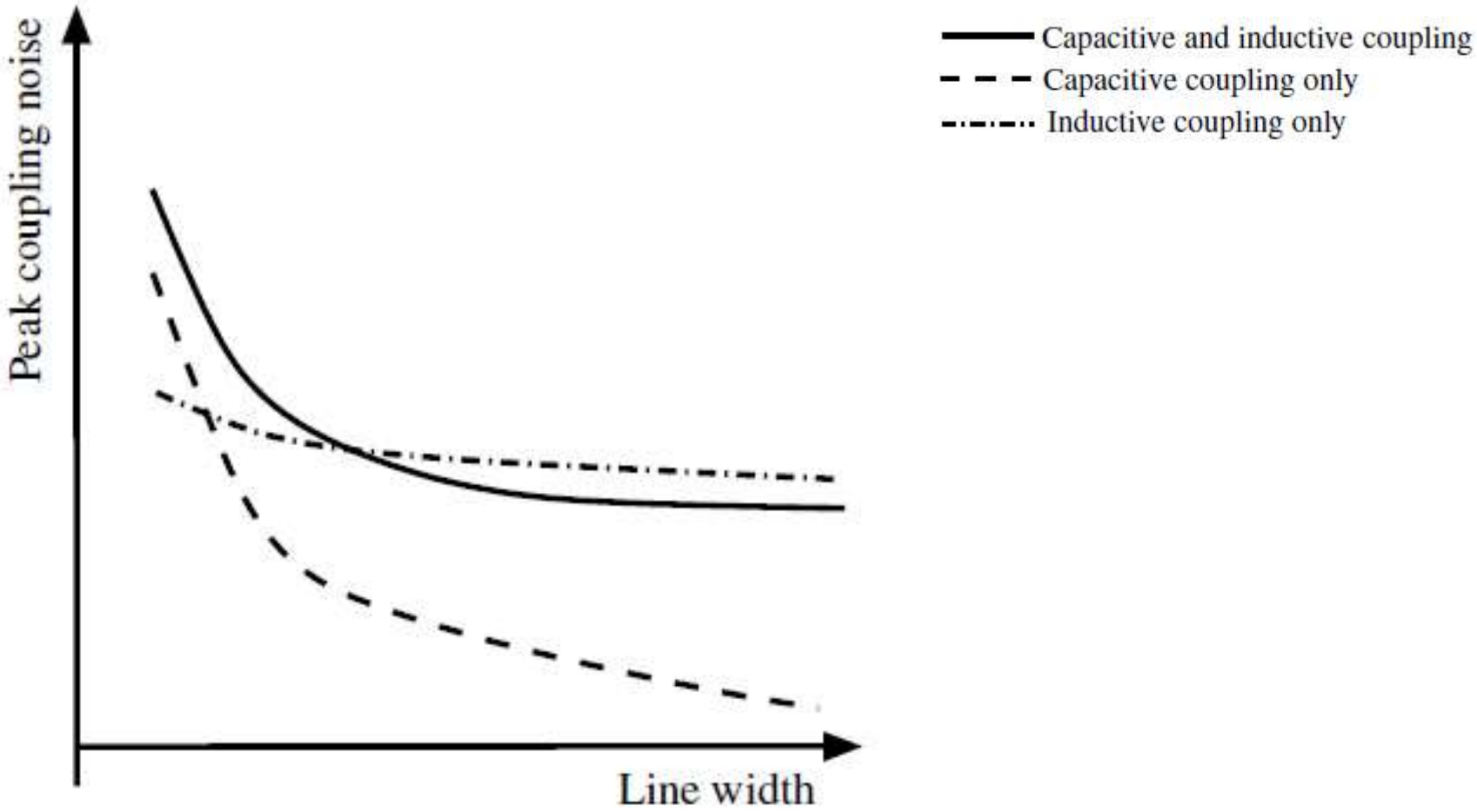
- **What if there is also inductive coupling?**



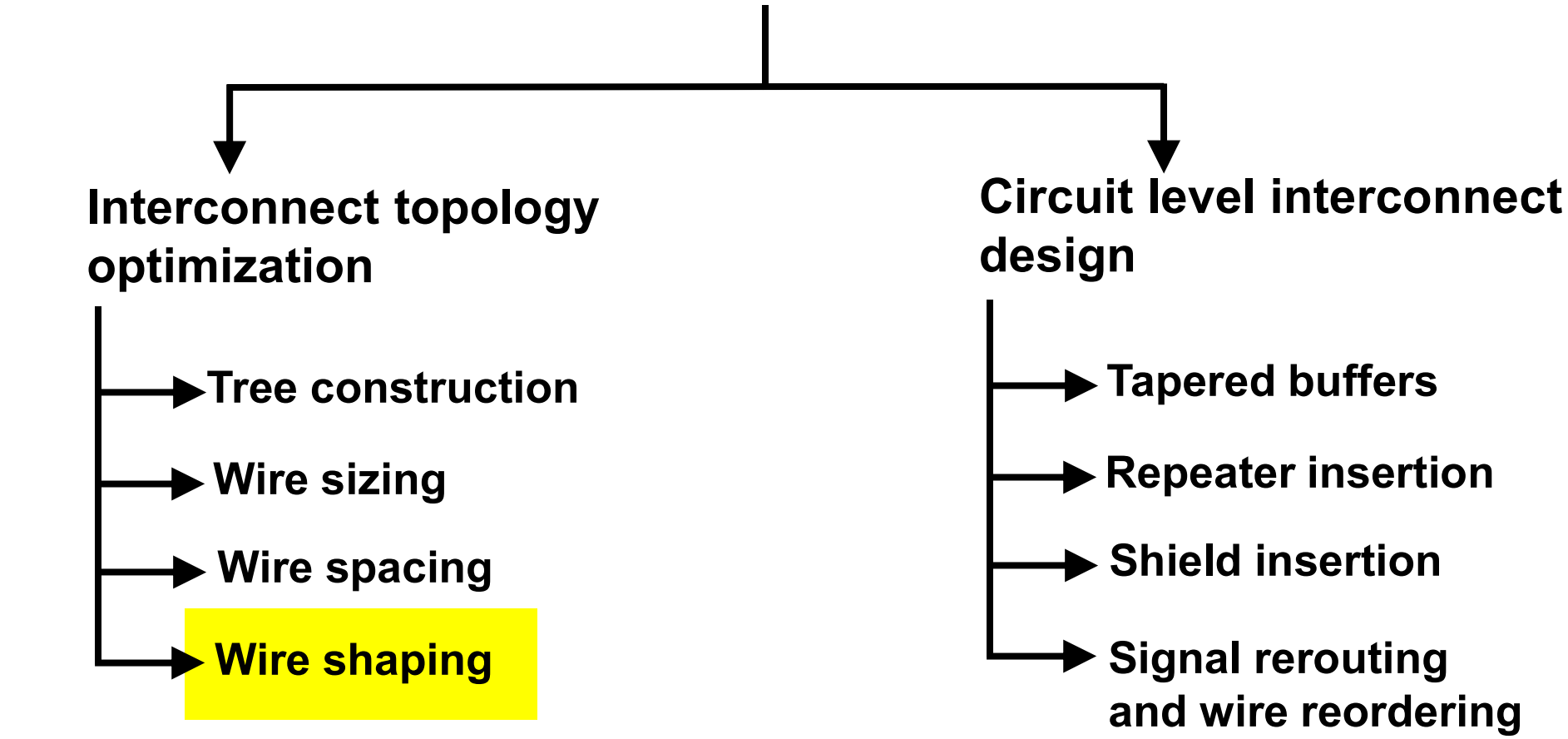
Effect of Increased Spacing

- Capacitive coupling reduced
- Reduction in inductive coupling is not as significant
  - Logarithmic reduction in mutual inductance
  - Long range phenomenon
- Coupling noise is dominated by inductive coupling with wide spacing
  - Further increase in spacing will not significantly lower noise

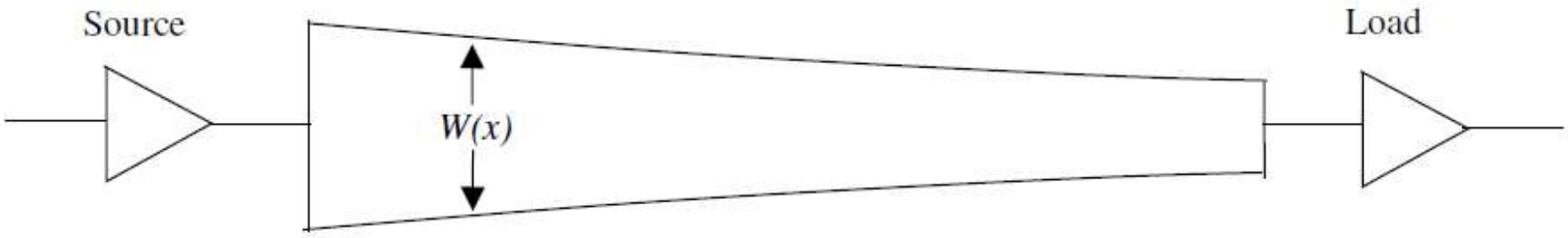
Spacing and Capacitive/Inductive Crosstalk



Global Signaling Methodologies

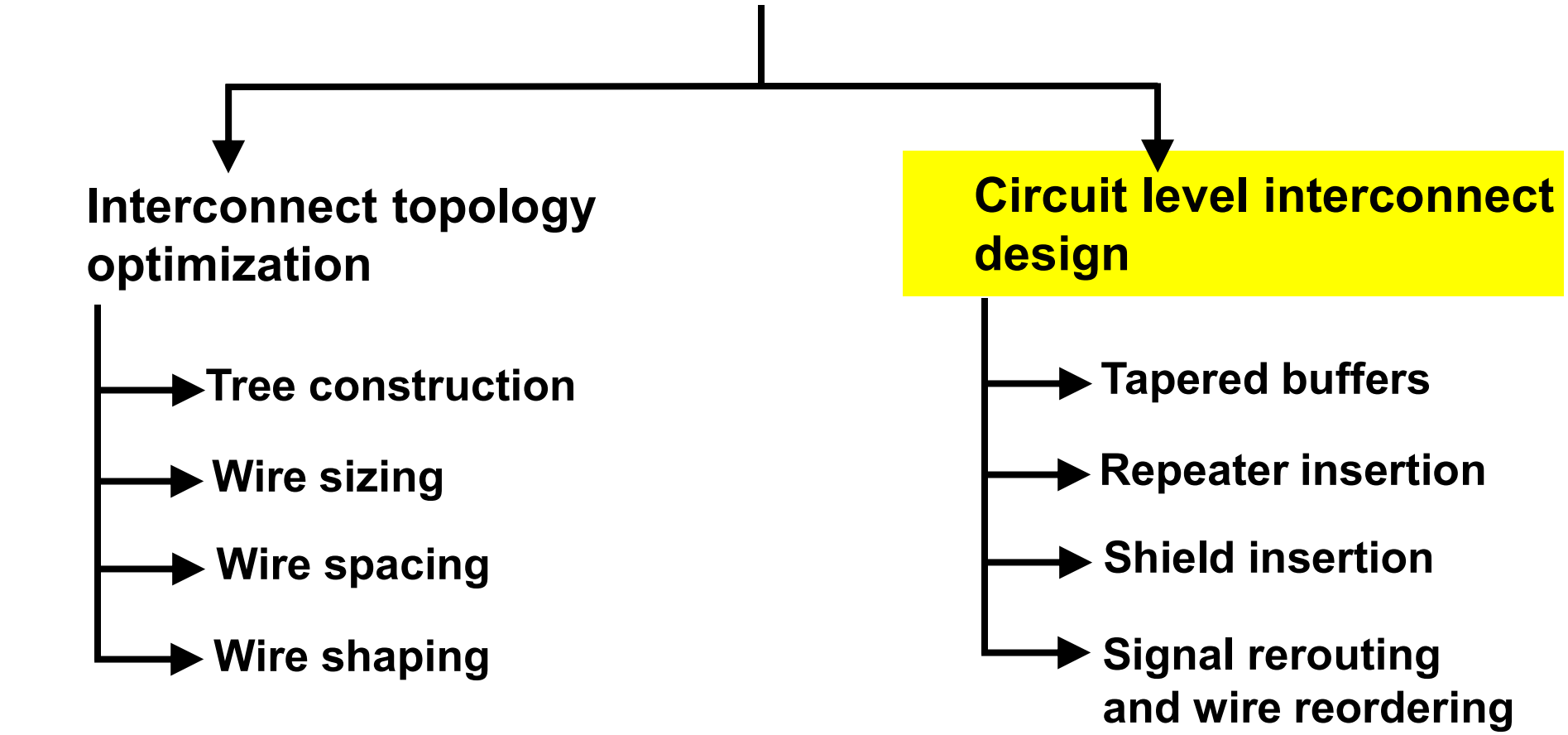


Shape of Interconnect



- **Optimal shape of *RC* interconnect that minimizes Elmore delay - an *exponential taper***
  - **Increased wire width near the source**
  - **Decreased wire width near the load**
  - **Less resistance at near end**
  - **Near end resistance sees more downstream capacitance than far end resistance**
  - **Total *RC* delay is reduced (similar for *RLC* lines)**
  - **Difficult to fabricate !!**

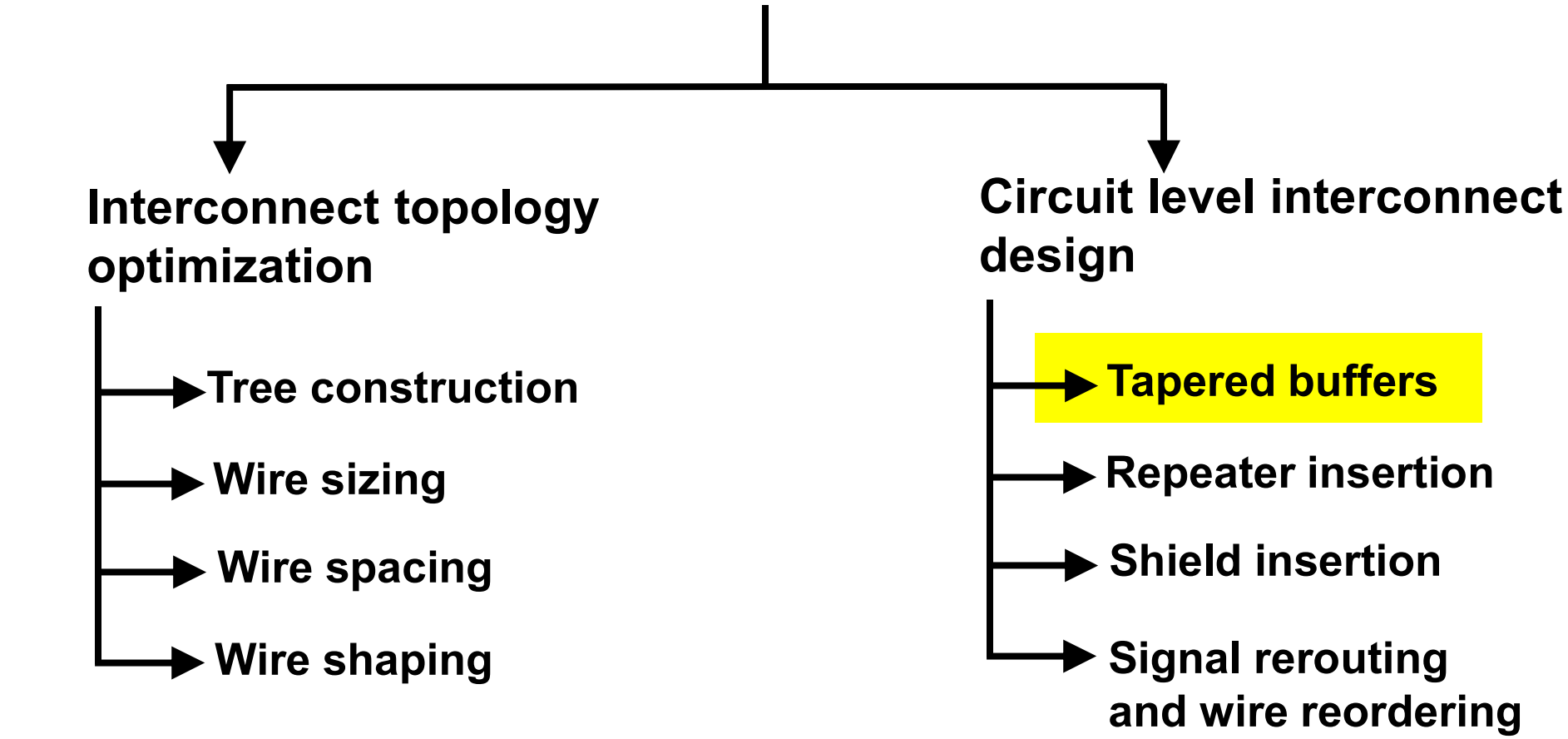
Global Signaling Methodologies



Driving an Interconnect

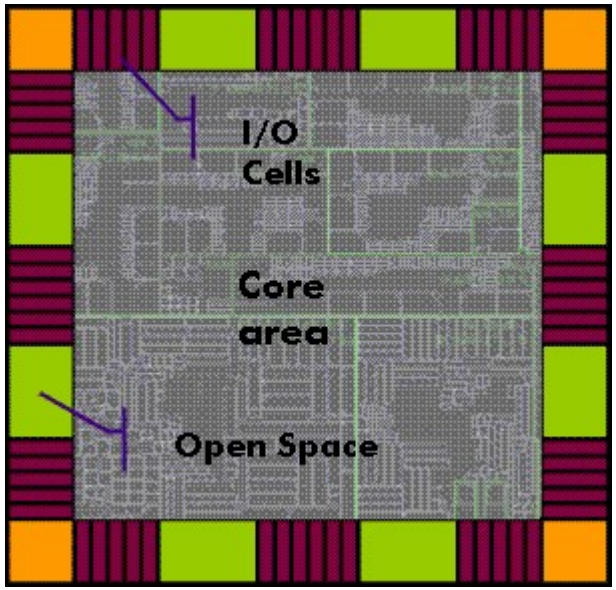
- **Load determines the appropriate circuit level optimization**
  - **Capacitive load:** ***tapered buffer***
  - **Resistive load:** ***repeater insertion***
  - **Inductive load:** ***repeater insertion with fewer repeaters***

Global Signaling Methodologies

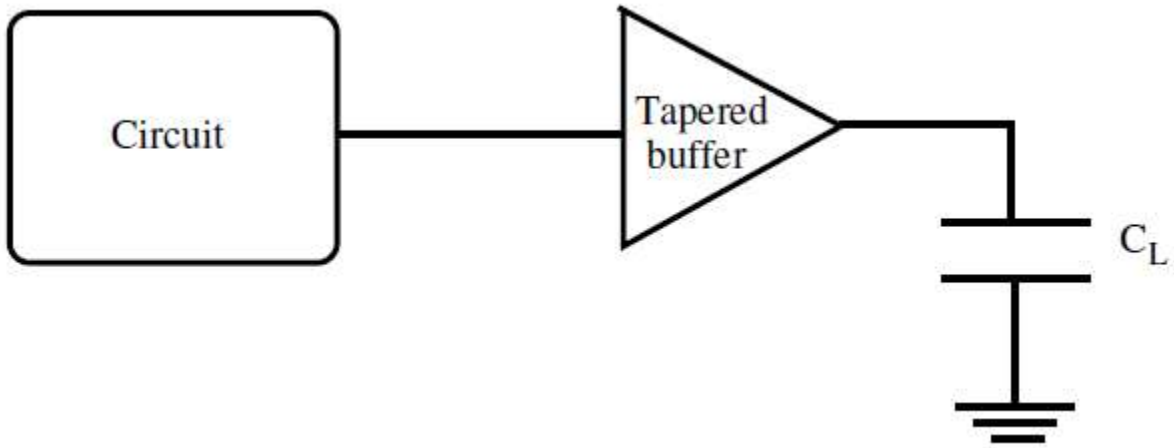


Driving a Large Capacitive Load

- **Tapered buffers to drive large capacitive load**
  - **Large capacitive loads common in ICs**
  - **On-chip**
    - **High fan-out gates**
    - **Long global interconnects**
    - **Signals driving output pads**
  - **Off-chip**
    - **Chip-to-chip communication lines**

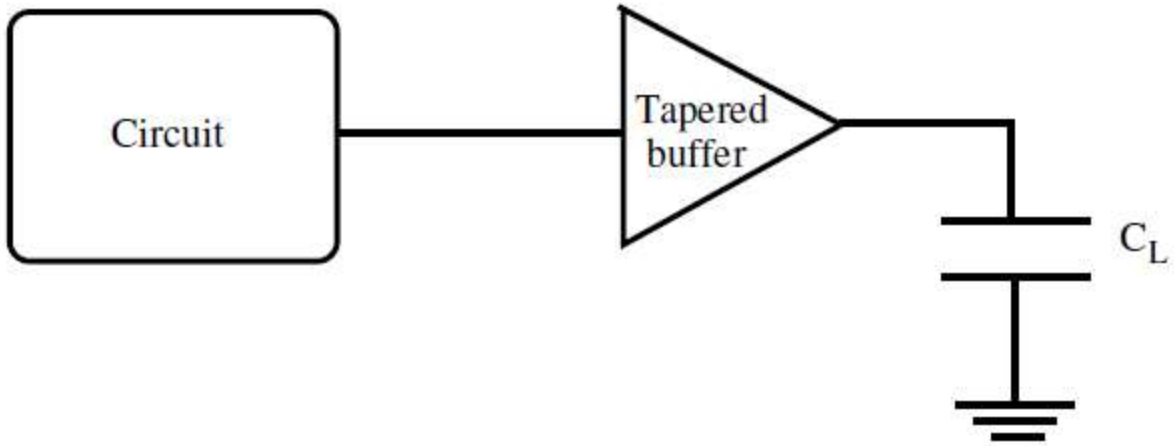


Tapered Buffers



- Buffer circuits quickly source and sink large amounts of current at sufficient speed
- Simply increasing the size of buffer does not work
  - Previous stage experiences the same problem
- *Tapered buffer* structure satisfied this need
- Placed between circuit and large capacitive load

Two Primary Objectives of a Tapered Buffer



- **Isolate the preceding circuit from large load**
- **Amplify the signal along the way**

Tapered Cascaded Buffers

- Capacitive loading (minimal interconnect resistance)

Off-chip, clock loads, data buses

$T_{PD} \propto R_{TR} C_{LOAD}$

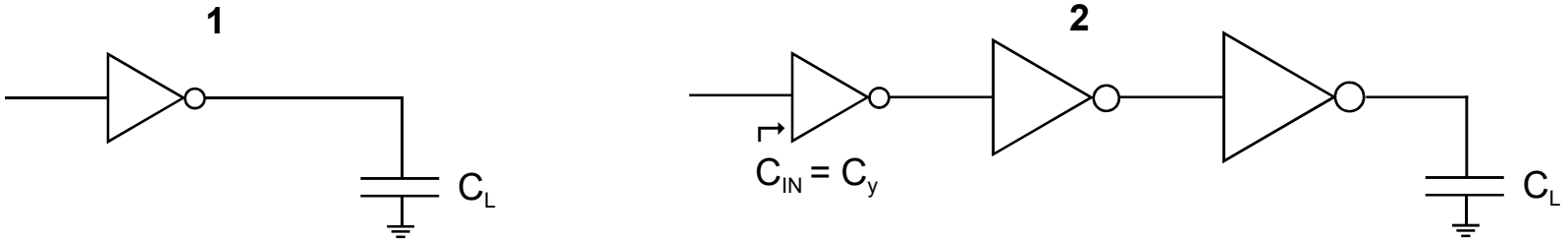
- Choice of R_{TR} is application specific $\Rightarrow$ choosing W

$$I_{DS} = K' \frac{W}{L} \left[(V_{GS} - V_T) \right]^2$$

V_T, K′ - Process dependent

V_S, V_G, V_D - Bias conditions

L ~ Typically chosen as minimum



$t_{PD1} > t_{PD2}$

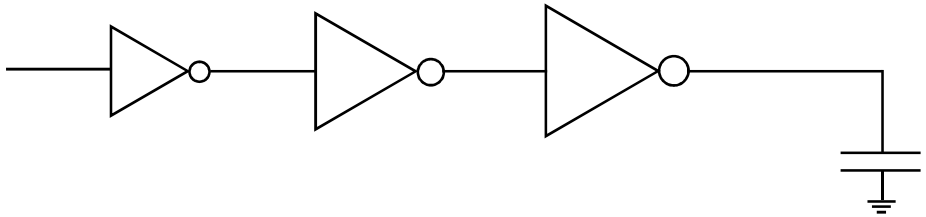
Jaeger, JSSC ‘75

Tapered Cascaded Buffers (continued)

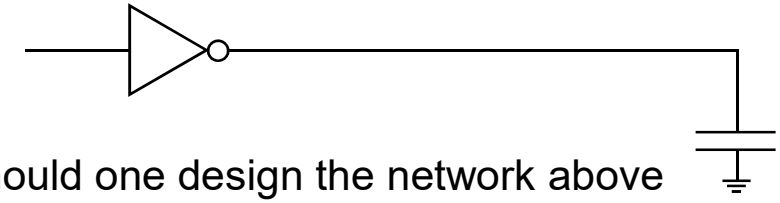
- P_{SC} will be worse since larger I_{SC} during $V_{TN} < V_{IN} < V_{DD} + V_{TP}$ but next stage will be better
- Area will be worse ↓
- P_D will be worse ↓

  $\therefore$ Therefore, there is a nontrivial optimal solution which is application-specific

- It is possible that a buffer driving a large load would be considerably faster and dissipate less power with <u>extra</u> buffers



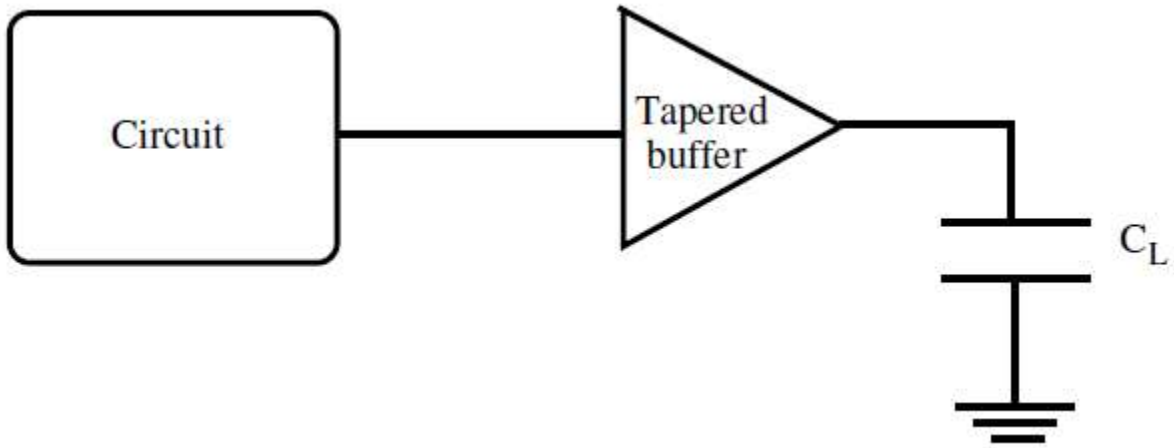
Is faster, uses less power, and requires more area, than (less P_{SC}, not P_{DYN})



- Large P_{SC} of next stage, implied by capacitor

- The question is how should one design the network above
- Application specification defines the maximum delay which for a given C_{LOAD} defines W_{TR}
  - – However, can satisfy specification with less power dissipation if cascade buffers
- Also, what if the load is <u>resistive</u> and capacitive?
  - – Use repeaters

Two Fundamental Conditions For Tapered Buffers

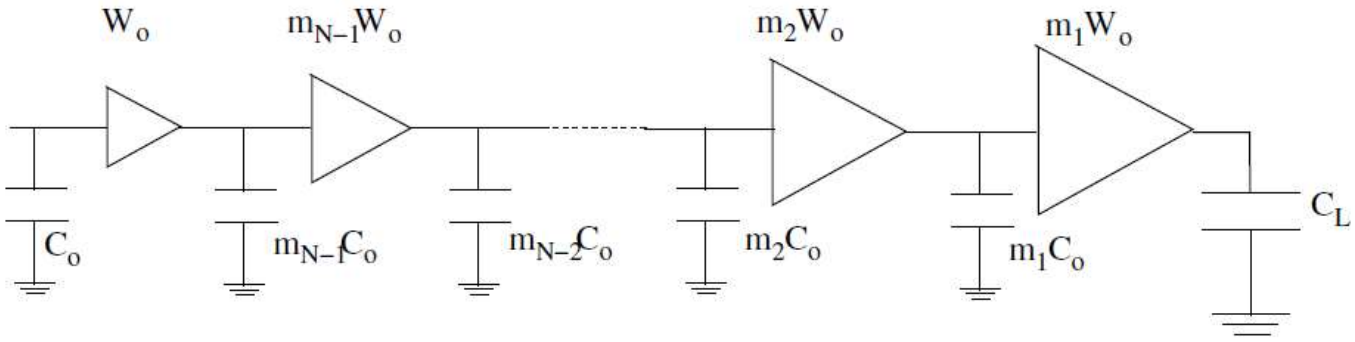


- **Preceding circuit should be able to drive the tapered buffer**
- **Tapered buffer should be able to drive the large capacitive load**

- **First introduced by Lin and Linholm in 1975**

H. C. Lin and L. W. Linholm, "An Optimized Output Stage for MOS Integrated Circuits," *IEEE Journal of Solid-State Circuits*, Vol. SC-10, No. 2, pp. 106–109, April 1975

Tapered Buffer System

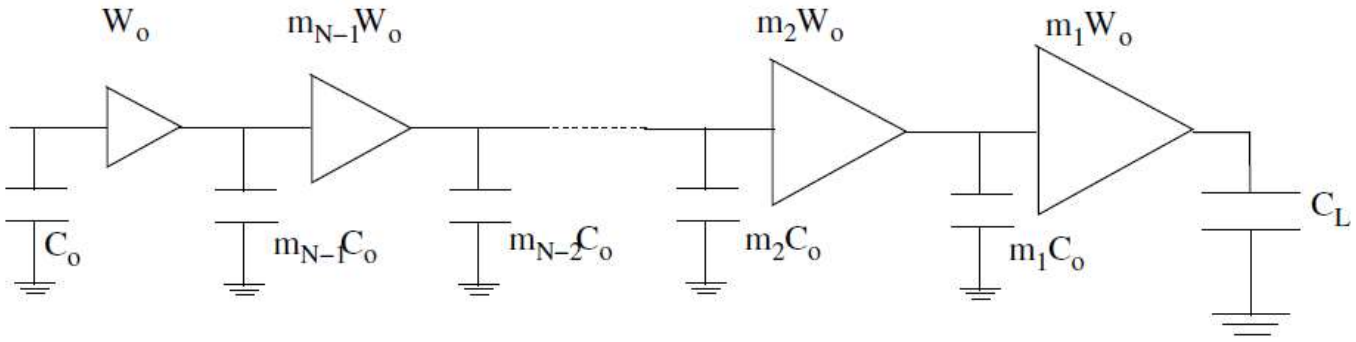


Tapered buffer structure proposed by Lin and Linholm. The factors $m_1, m_2, ..., m_{N-1}$ are selected to equalize the delay of each stage.

- **Consists of series of cascaded tapered inverters**
- **Each transistor channel width is fixed multiple of the previous inverter**
- **Each inverter stage has equal rise, fall, and delay times**
  - **output current drive capability / output capacitance = constant**

$$K = I/C \text{ per stage}$$

Optimum Number of Stages



Tapered buffer structure proposed by Lin and Linholm. The factors $m_1, m_2, ..., m_{N-1}$ are selected to equalize the delay of each stage.

- **Lin and Linholm did not consider optimum number of stages to minimize the entire delay**
- **Jaeger differentiated total delay with respect to N and set the equation to zero to find optimum number of stages N**

$$t_{po} M^{1/N_{opt}} \left(1 - \frac{\ln M}{N_{opt}}\right) = 0 \qquad N_{opt} = \ln M = \ln\left(\frac{C_L}{C_o}\right)$$

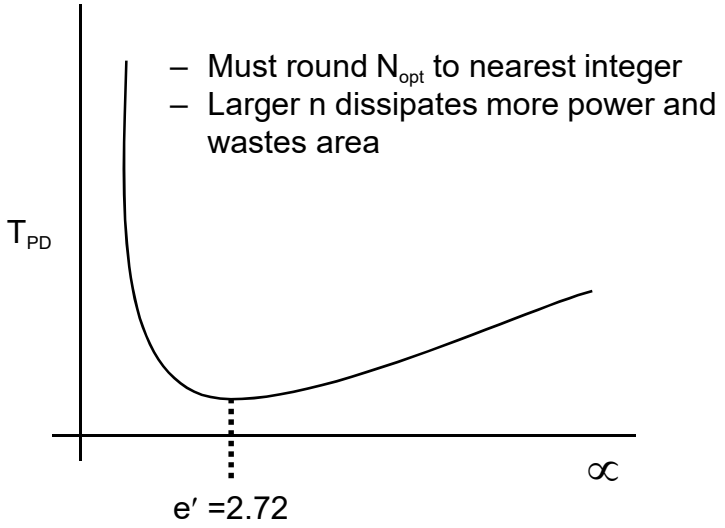
Optimal Sizing of Cascaded Buffers

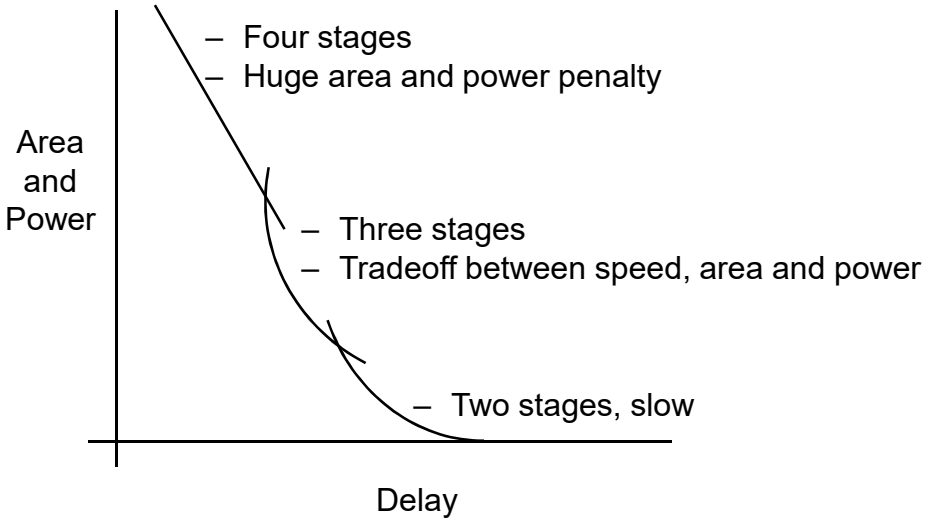
- Don't minimize stage delay, minimize total path delay

$$\propto = \frac{W_i}{W_{i-1}} = e^1 = 2.71$$

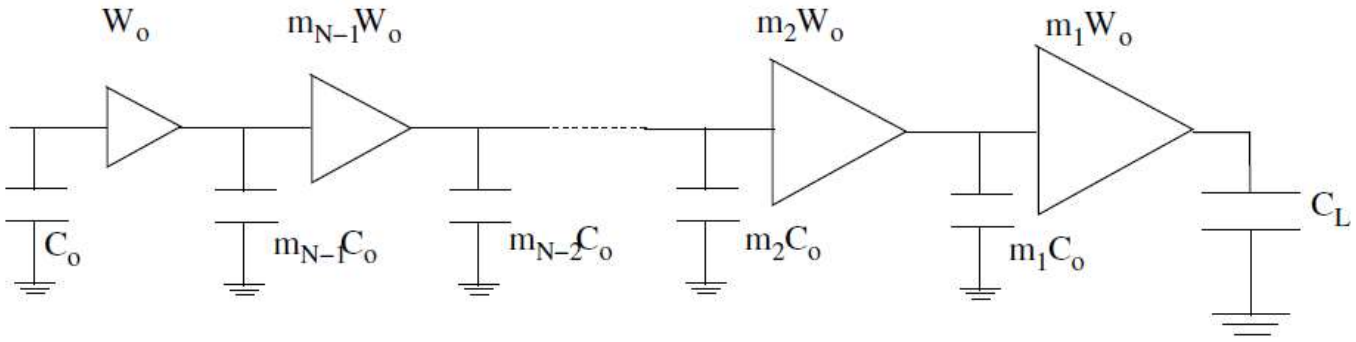
$$A_N = \frac{A_y (M-1)}{e^1 - 1}$$

$$N_{opt} = \ln \frac{C_L}{C_y}$$





Optimum Number of Stages and Tapering Factor



Tapered buffer structure proposed by Lin and Linholm. The factors $m_1, m_2, ..., m_{N-1}$ are selected to equalize the delay of each stage.

- **Tapering factor is implicitly set**
  - **Once the number of stages is determined**

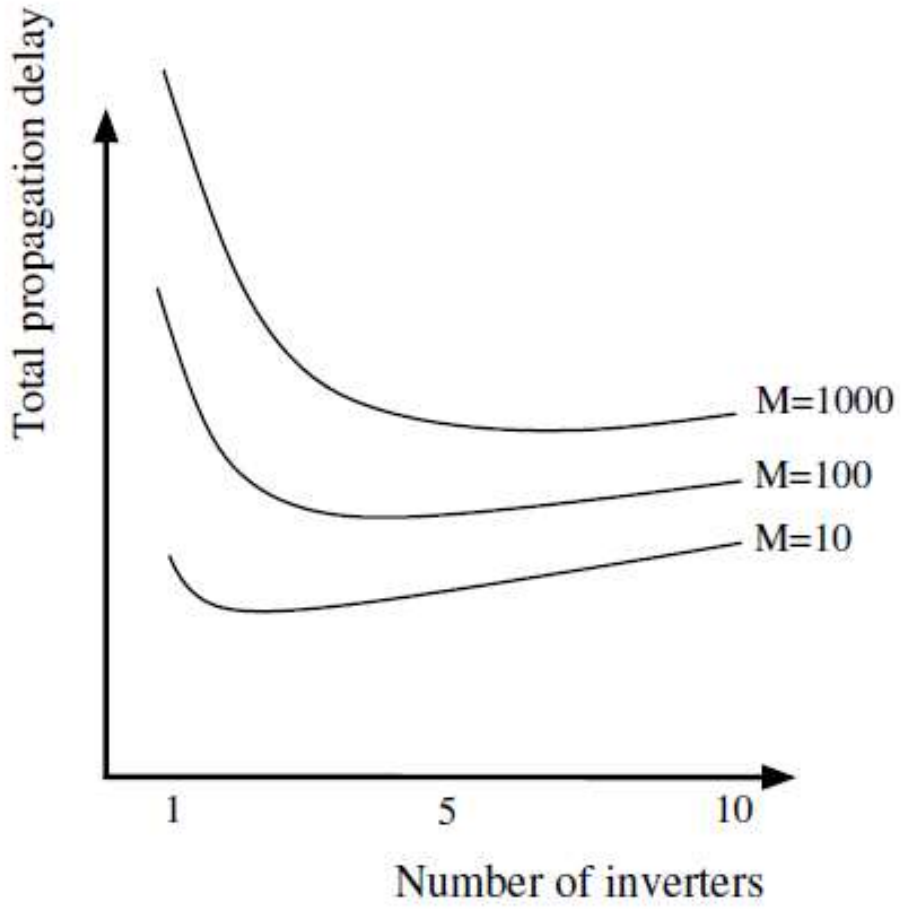
$$N_{opt} = \ln M = \ln(\frac{C_L}{C_o})$$

$$F_{opt} = \frac{m_k}{m_{k+1}} = e \approx 2.72$$

$$F = \frac{m_k}{m_{k+1}} = \frac{m_{k+1}}{m_{k+2}} = M^{1/N} = \text{tapering factor}$$

- **Optimum tapering factor is independent of N and M**

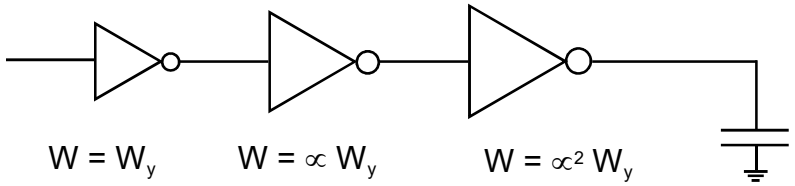
Propagation Delay versus Number of Inverters



- If N is too small
  - Delay is high
- In M is not an integer
- Delay is weak function of N around optimum N

Sizing of CMOS Inverter Stages

- Minimum delay occurs when the delay of each stage is equal
  - *i.e.*, $R_{TR} \downarrow$, $C_g \uparrow$ for each succeeding stage



  - Where W_y is minimum width for that technology

If drain capacitance C_x is non-negligible

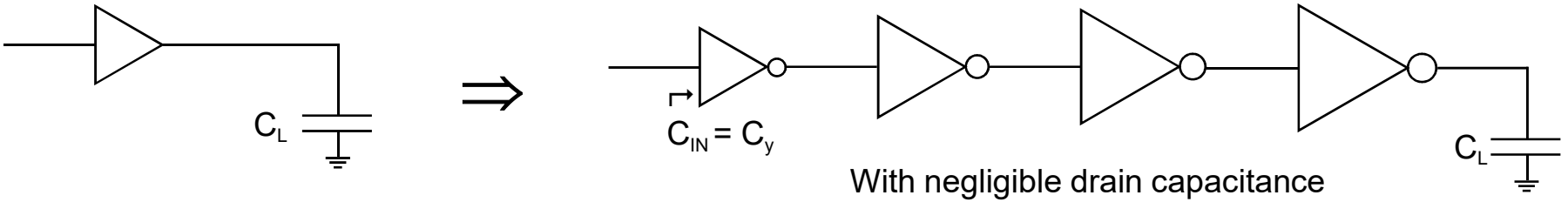
$$N_{opt} = \frac{\ln \frac{C_L}{C_y}}{\ln \propto} \qquad \propto \left[\ln \propto - 1\right] = \frac{C_x}{C_y} \qquad \frac{T_{D_{min}}}{\tau} = \propto \ln\left[\frac{S_n}{S_y}\right]$$

- Minimum delay of a chain of symmetric inverters proportional to logarithm of ratio of size of last inverter to first inverter

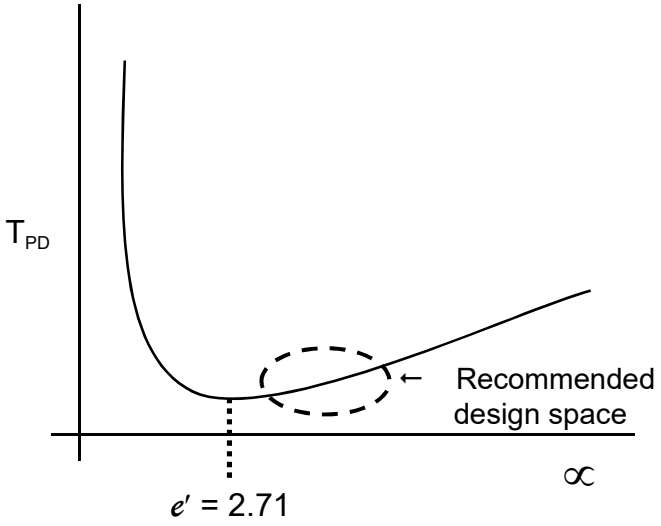
<u>Assumptions</u>

- 100% of output capacitance scales with the following stage
  - No interconnect capacitance
    - Split capacitance model
  - Variable tapering factor
    - Good delay with much less area and power
      - Initially $\propto$ is small (close to 1)
      - Close to exponential $\propto$ in later stages
  - Good choice for extremely high capacitive loads

Tapered Cascaded Buffers



With negligible drain capacitance



Optimal number of stages $\sim N_{opt} = \ln \frac{C_L}{C_y}$

Tapering factor for cascaded buffers $\sim \propto = \frac{S_n}{S_{n-1}} = e^1 \Rightarrow 3.0 \text{ to } 4.0$

With non-negligible drain capacitance

Optimal number of stages $\sim N_{opt} = \frac{\ln \frac{C_L}{C_y}}{\ln \propto}$

Tapering factor for cascaded buffers $\sim \propto \left[\ln \propto - 1\right] = \frac{C_x}{C_y}$

C_x - output drain/source capacitance

Transcendental in $\propto$

What About Power Dissipation?

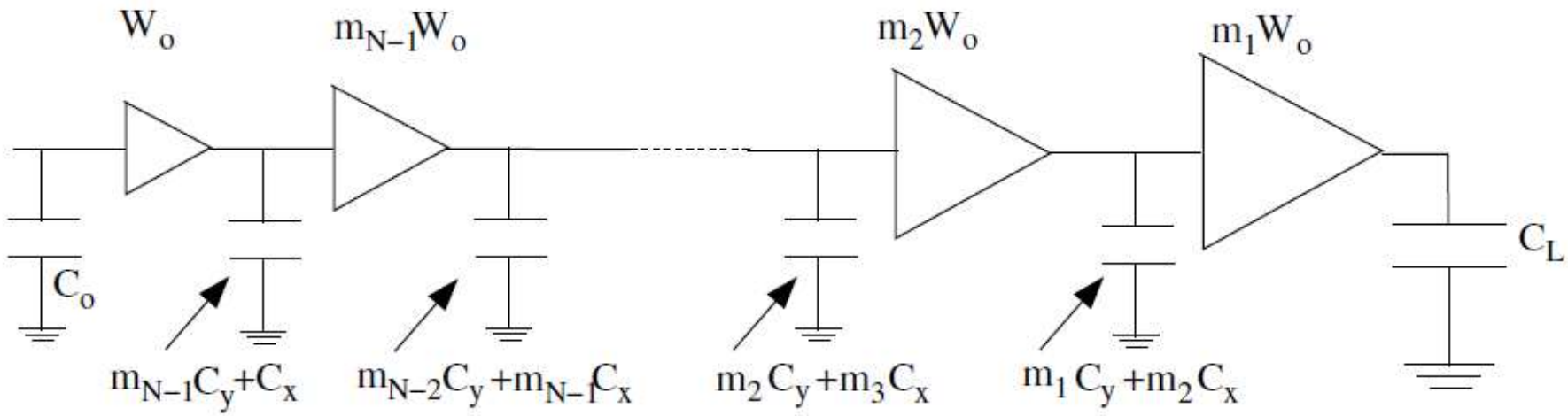
Veendrick model: tapering factor is higher than $e^1 = 2.7$

- **Fewer stages are required**
  - **Smaller area**
  - **Lower power**
  - **Non-optimum delay**

	Jaeger	Veendrick
F	e	11.5
N	5	2
Delay	5.5 ns	6.5 ns
Power	79 mW	16.5 mW

- **Optimizing single parameter can have adverse effect on another parameter**
- **Moderate compromise can be helpful**

Split Capacitor Model



Tapered buffer structure with split-capacitor model

- **More accurate interstage capacitance model**
  - **Input gate capacitance of next inverter stage**
  - **Output diffusion capacitance of previous inverter stage**

Design of Cascaded CMOS Buffers Summary

- This approach describes the optimal choice of $\propto$ for a series of cascaded buffers for minimizing speed
  - However, heavy cost in area as approach 2.71 for minor improvement in speed
  - Therefore, for typical high speed circuits, choose $\propto_C = 3.0$

Assumptions

- 100% of output capacitance scales with following stage
- No drain/source or interconnect capacitance
  - Split capacitance model
- Area/power dissipation tradeoffs
- Variable tapering factor
  - With local C_{INT}, $C_{S/D}$ - could use variable $\propto$
  - C_x - output capacitance, C_y - input capacitance

$$\propto \left[\ln \propto - 1\right] = \frac{C_x}{C_y}$$

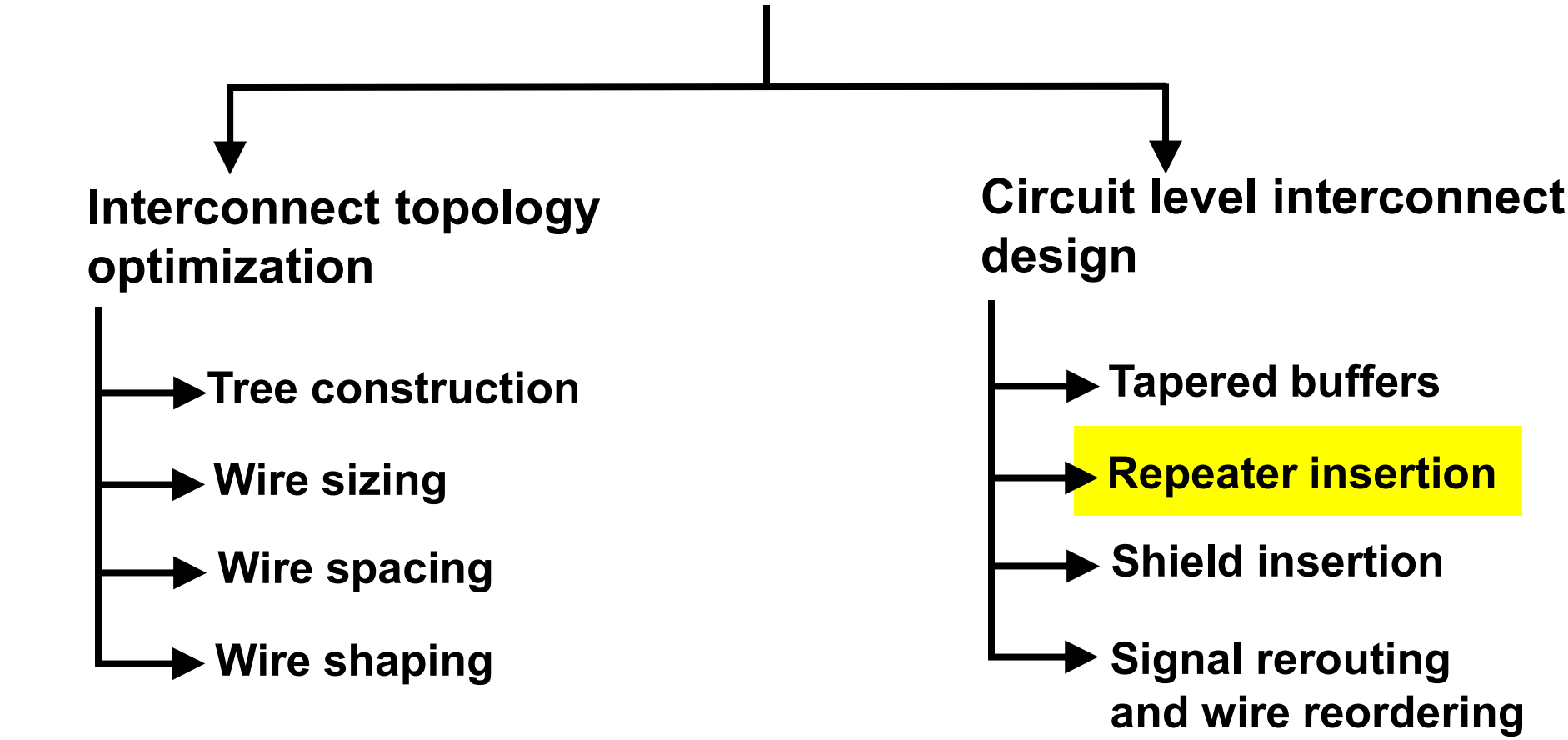
if $C_x = 0$, $\propto = e'$

Transcendal *equation* in $\propto$

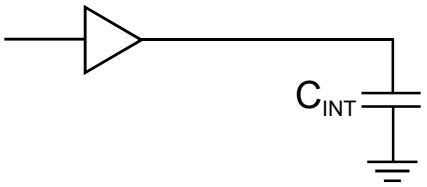
$$N = \frac{\ln \frac{C_L}{C_y}}{\ln \propto}$$

- Primary result of variable tapering is good delay with much less power
- Initial buffers, $\propto$ is small, close to 1
  - Add exponential $\propto$ later
- Good choice for high capacitive loads

Global Signaling Methodologies

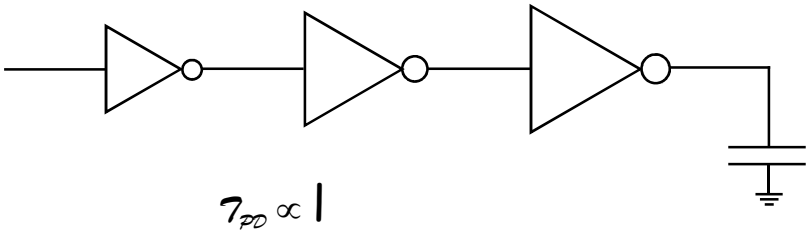


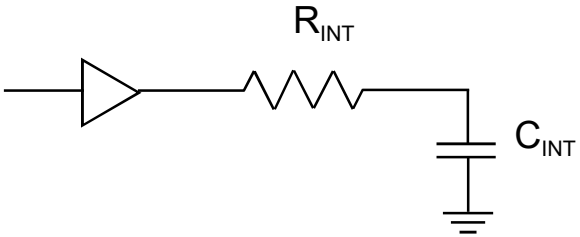
Cascaded Buffers



$\Rightarrow$

Cascaded buffers





$$T_{PD} \propto l_C l_R = l^2$$

- R_{INT} and C_{INT} increase linearly with length

- Assume R_{INT} = 100 Ω – fairly small if increase buffer size

R_{DSMIN} (W = .28 μm) = 2210 Ω
$\Rightarrow R_{DS}$ (W = 5.0 μm) = 118 Ω

Delay Dependence on Interconnect Length

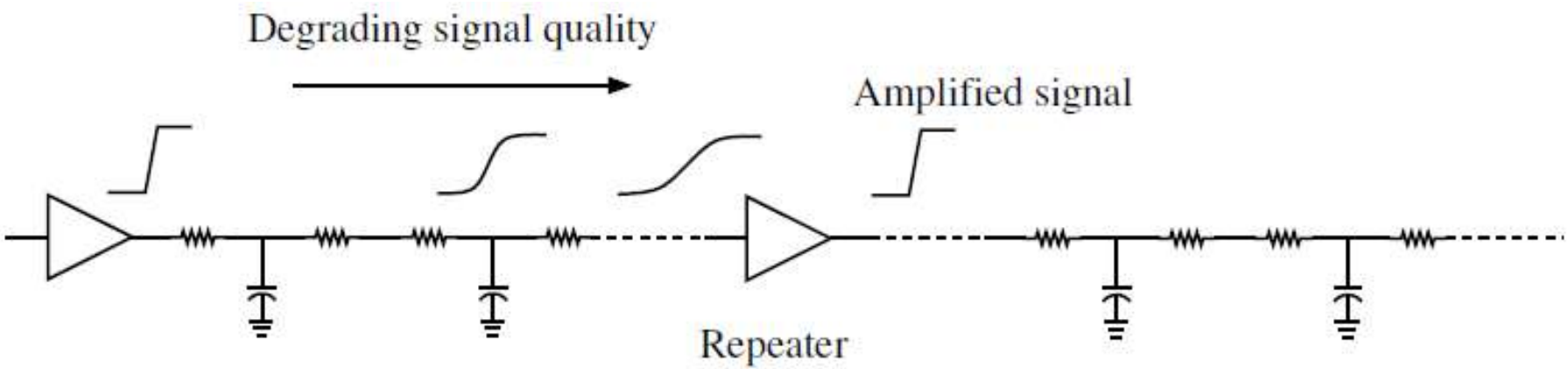
$$T_{d-RC} = R_{int}C_{int}[0.377 + 0.693(R_T C_T + R_T + C_T)]$$

$$R_T = R_{tr}/R_{int} \qquad C_T = C_g/C_{int}$$

$$T_{d-RC} \approx 0.377RCl^2$$

- ***Repeater insertion*** **commonly used to overcome this quadratic dependency**

Repeater Insertion



- **Delay reduced by dividing interconnect into smaller sections**
- **Repeater is inverting or non-inverting buffer placed at specific locations along interconnect**
- **Amplifying nature of inverters is exploited to restore signal**

Repeaters

- Large buffer with R_{DS} = 118 Ω cannot drive fairly small R_{INT} = 100 Ω

  R_{DS} (W = 10 μm) = 60 Ω

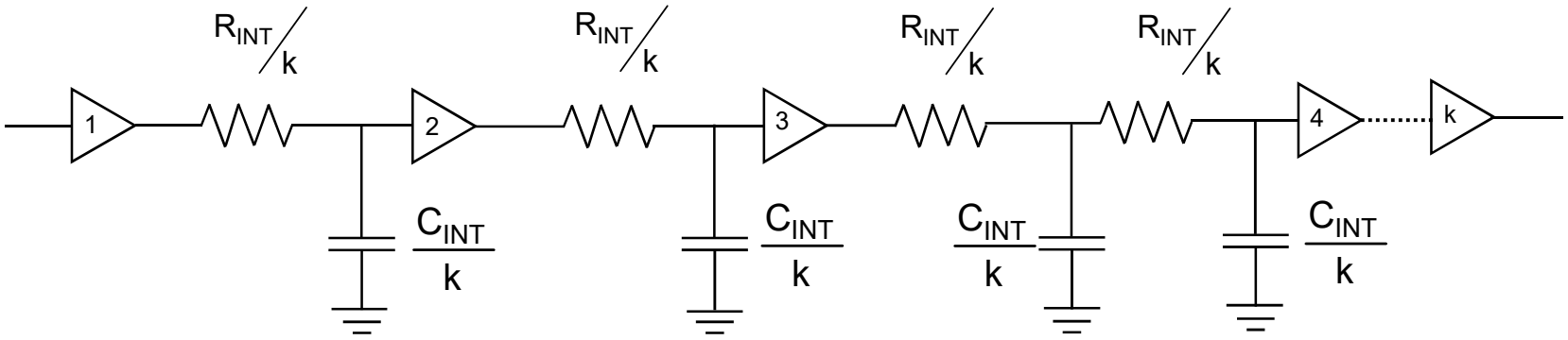
  – Can drive R_{INT} but not efficiently

  – Buffer overdrive problem — What about if R_{INT} = 200 Ω?

- Partition line into multiple sections

  – Make $\tau_{PD} \propto \ell^2$ by inserting buffers or repeaters along the line $\Rightarrow \ell$

  $R_{ON} = R_{DS} >> R_{INT}$ of each section of the line



  – k - number of sections

  – h - size of inverter

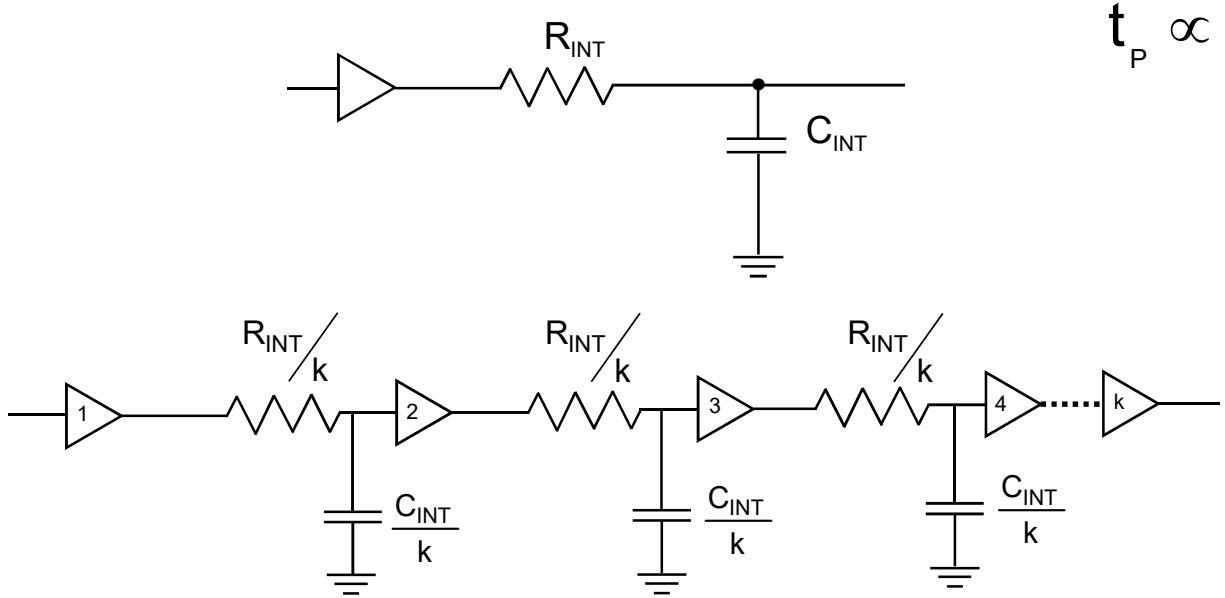
Repeaters

- When interconnect resistance is comparable to or larger than the on-resistance of the driver, $R_{INT} \geq R_{ON}$

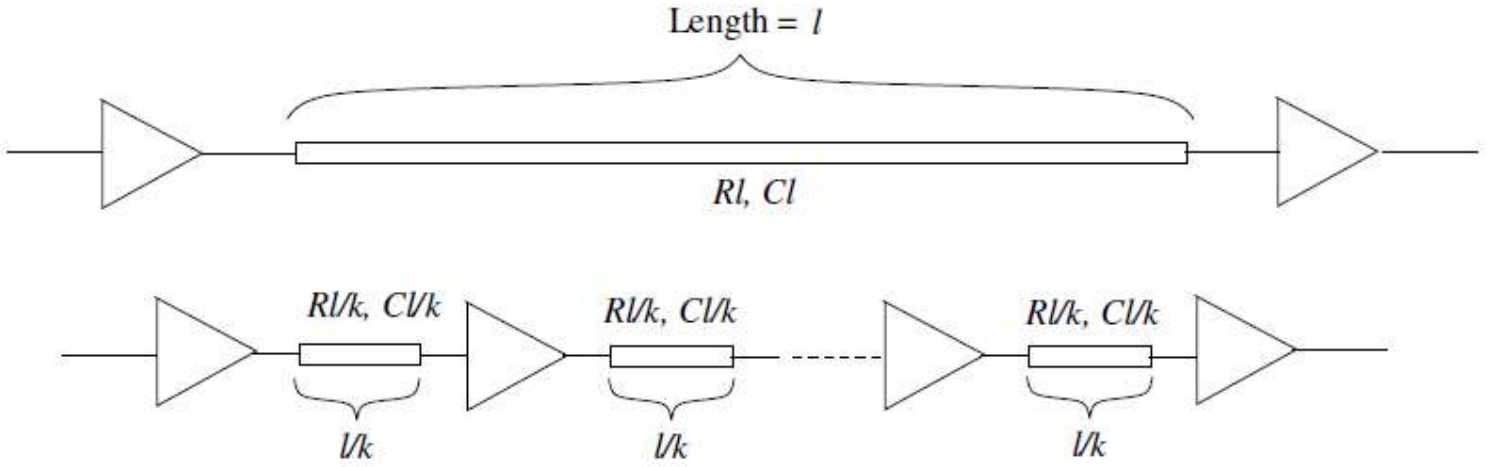
$$t_{PD} \propto \ell^2 \text{ where } \ell \text{ is the interconnect length}$$

- Since both r_{INT} and c_{INT} increase linearly with length
- By inserting localized buffers, such that $R_{ON} >> R_{INT}$
  - t_{PD} is linear with length

$$t_P \propto \ell$$



Reduction in Delay

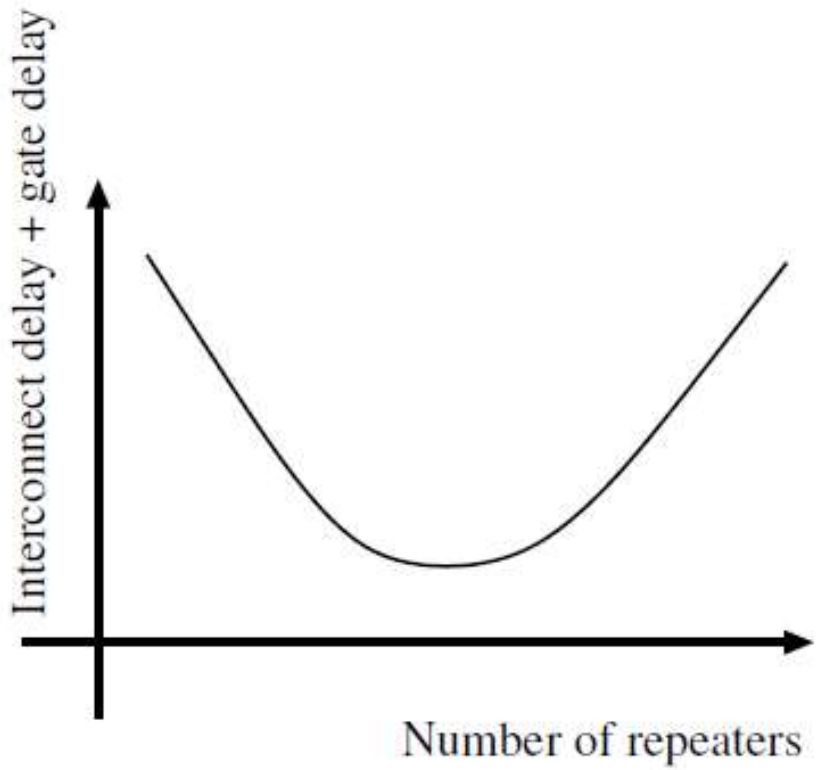


Splitting the interconnect into k segments with repeaters to reduce overall signal propagation delay

$$T_{d-RC} \approx 0.377k\frac{Rl}{k}\frac{Cl}{k} = \frac{0.377RCl^2}{k}$$

- **Original delay reduced by *k***
- **Interconnect delay decreases with increasing *k***
- **Additional repeaters increase gate delay**
- **What is optimum number of repeaters?**

Delay versus Number of Repeaters



- Optimum number of repeaters that minimize overall delay
- Two parameters
  - Number of repeaters k
  - Uniform size h

Optimum Number and Size of Repeaters

- R_o and C_o are output resistance and input capacitance of minimum size inverter
- Inverter with size *h*, R_o/h
- Delay per stage

$$T_{d-RC} = R_{int}C_{int}[0.377 + 0.693(R_TC_T + R_T + C_T)]$$

$$R_T = R_{tr}/R_{int} \qquad C_T = C_g/C_{int}$$

- Total delay = *k* x delay of single stage

$$T_{d-RC} = k(0.377\frac{R_{int}C_{int}}{k^2} + 0.693\frac{R_0C_{int}}{kh} + 0.693\frac{hR_{int}C_0}{k} + 0.693R_0C_0)$$

Repeaters - Basic Design Expressions

R_o ~ on-resistance of buffer
C_g ~ gate capacitance of buffer

$$T_{50\%} = k\left[0.7\,R_o\left(\frac{C_{INT}}{k} + C_g\right) + \frac{R_{INT}}{k}\left(0.4\,\frac{C_{INT}}{k} + 0.7\,Cg\right)\right] \qquad (1)$$

$$\frac{dT}{dk} = 0 \qquad 0.4\,\frac{R_{INT}\,C_{INT}}{k^2} = 0.7\,R_o\,C_g$$

- Segment delay (between repeaters) should be equal to repeater delay
  - To achieve shortest delay

$$k = \sqrt{\frac{0.4\,R_{INT}\,C_{INT}}{0.7\,R_o\,C_g}} \quad \text{for } k \geq 2 \text{ and an integer}$$

$$T_{50\%} = 0.7\,R_o\,C_{INT} + 1.1\sqrt{R_o\,C_g\,R_{INT}\,C_{INT}} + 0.7\,R_{INT}\,C_g \qquad (2)$$

$$1.1\sqrt{R_o\,C_g\,R_{INT}\,C_{INT}} < 0.7\,R_o\,C_g + 0.4\,R_{INT}\,C_{INT}$$

↖ From buffer w/o repeater equation
$T_{50\%} = .4\,R_{INT}\,C_{INT} + 0.7\,(R_{tr}\,C_{INT} + R_{tr}\,C_L + R_{INT}\,C_L)$

- Repeaters improve the delay of resistive lines since $k \geq 2$, $R_{INT}\,C_{INT} > 7\,R_o\,C_g$

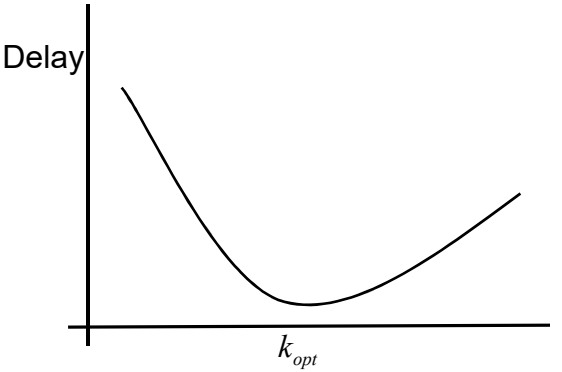
Repeaters - Basic Design Expressions

- Interconnect dominates for few stages
- Repeater delay dominates with too many stages

$C_T = \frac{C_L}{C_{INT}}$ $R_T = \frac{R_{tr}}{R_{INT}}$ $R_{tr} = \frac{R_o}{h}$ - Minimizes size inverter - Increases gain $C_L = h\, C_o$ - Minimize size gate capacitance

$$T_{PD}(RC) = 0.74\, R_{TR}\, C_L \left(R_T + C_T + R_T\, C_T + 0.5\right)$$

$$= 0.37\, R_{INT}\, C_{INT} + 0.74\left(R_{INT}\, C_L + R_{TR}\, C_{INT} + R_{INT}\, C_L\right)$$

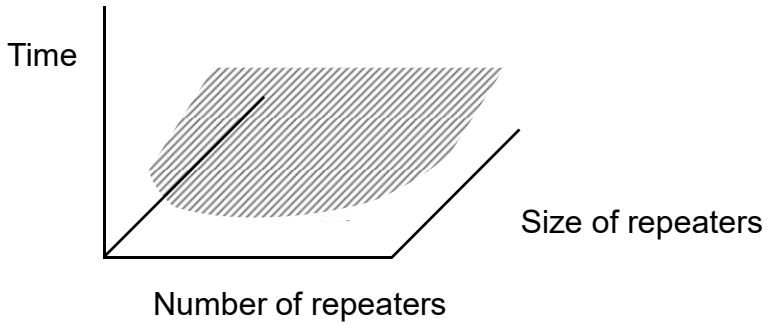


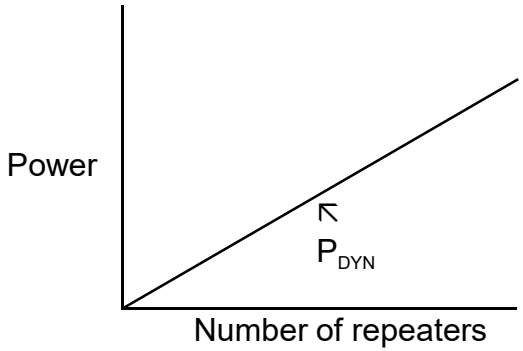
$$\frac{\partial T_{PD_{total}}(h,k)}{\partial h} = 0 \Rightarrow h_{opt}(RC) = \sqrt{\frac{R_o\, C_{INT}}{R_{INT}\, C_o}} \qquad \frac{\partial T_{PD_{total}}(h,k)}{\partial k} = 0 \Rightarrow k_{opt}(RC) = \sqrt{\frac{R_{INT}\, C_{INT}}{2\, R_o\, C_o}}$$

For $k \geq 2$ and an integer

- Segment delay (between repeaters) should be equal to the repeater delay
  - To achieve the shortest delay

For uniformly sized repeaters





Optimum Number and Size of Repeaters

- **Total delay = *k* x delay of single stage**

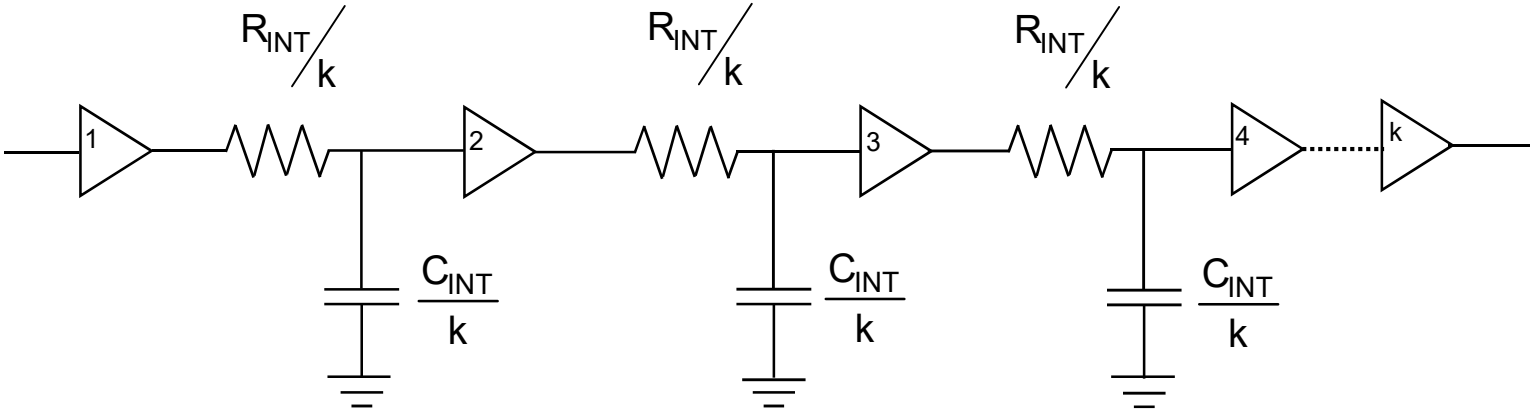
$$T_{d-RC} = k(0.377\frac{R_{int}C_{int}}{k^2} + 0.693\frac{R_0C_{int}}{kh} + 0.693\frac{hR_{int}C_0}{k} + 0.693R_0C_0)$$

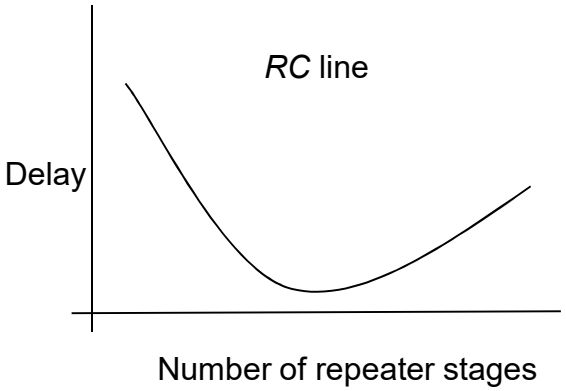
- **Take partial derivatives to find optimum *h* and *k***

$$k_{opt-RC} = \sqrt{\frac{R_{int}C_{int}}{2.3R_0C_0}} \qquad h_{opt-RC} = \sqrt{\frac{R_0C_{int}}{R_{int}C_0}}$$

- **Optimum number of stages determined by ratio of interconnect delay to gate delay**
  - **Higher ratio → k should be increased since gate delay is less significant**
- **Optimum size chosen to balance output resistance of repeater (R_o/h) and interconnect resistance (R_{int}/k)**

Repeater Design Expressions - *RC* Line





$$C_T = \frac{C_L}{C_{INT}} \quad R_T = \frac{R_{TR}}{R_{INT}}$$

$$T_{PD}\,(RC) = 0.37\,R_{INT}\,C_{INT} + 0.74\left(R_{INT}\,C_L + R_{TR}\,C_{INT} + R_{TR}\,C_L\right)$$

$$h_{opt}\,(RC) = \sqrt{\frac{R_o\,C_{INT}}{R_{INT}\,C_o}} \qquad k_{opt}\,(RC) = \sqrt{\frac{R_{INT}\,C_{INT}}{2\,R_o\,C_o}}$$

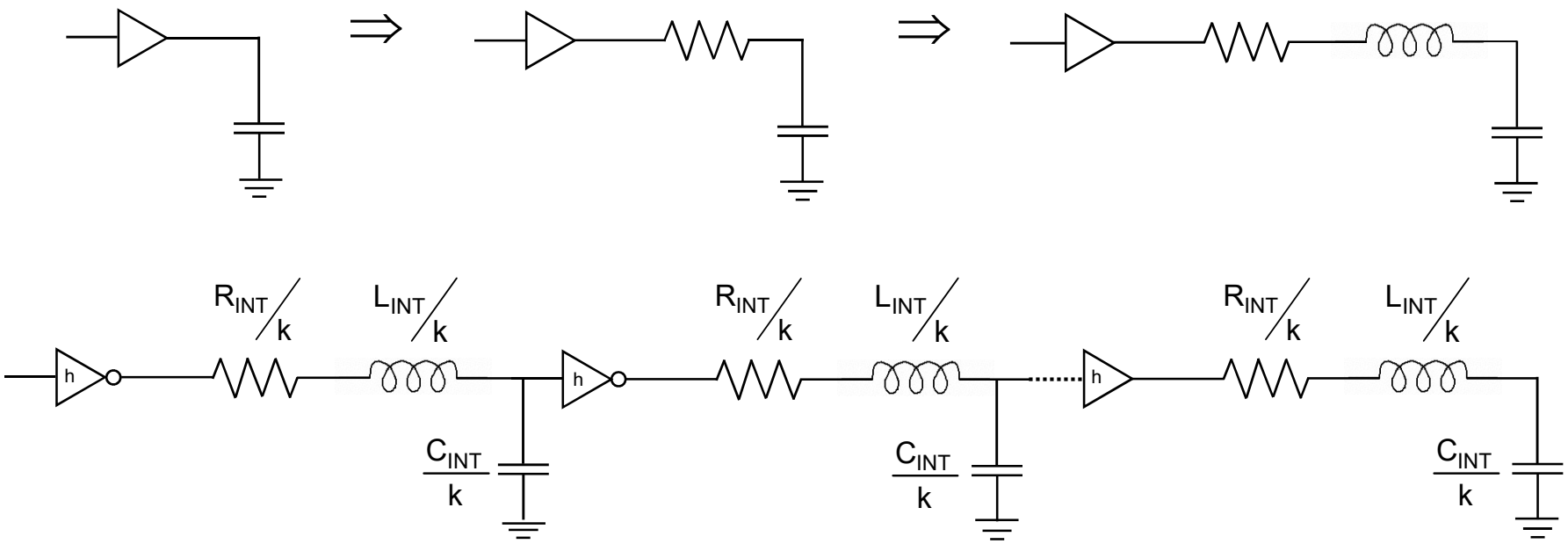
Driving an Interconnect

- **Load determines the appropriate circuit level optimization**
  - **Capacitive load: Use *tapered buffer***
  - **Resistive load: Use *repeater insertion***
  - **Inductive load: Use *repeater insertion with fewer repeaters***

Repeater Insertion in *RLC* Interconnect

Repeater Insertion: Lines vs. Trees

- Resulting solution less optimal than if tree is optimized as a tree
  - If tree optimized branch by branch



$$R_{tr} = \frac{R_o}{h}$$ ~ Minimum size inverter ~ Increased gain

$$C_L = h\,C_o$$

Minimum size gate capacitance

Repeater Insertion in RLC Interconnect (continued)

$$h_{opt}\,(RLC) = \sqrt{\frac{R_o\,C_{INT}}{R_{INT}\,C_o}} \bullet h'\,(T_{L/R})$$

$$k_{opt}\,(RLC) = \sqrt{\frac{R_{INT}\,C_{INT}}{2\,R_o\,C_o}} \bullet k'\,(T_{L/R})$$

$$\text{where } T_{L/R} = \sqrt{\frac{L_{INT}/R_{INT}}{R_o\,C_o}}$$

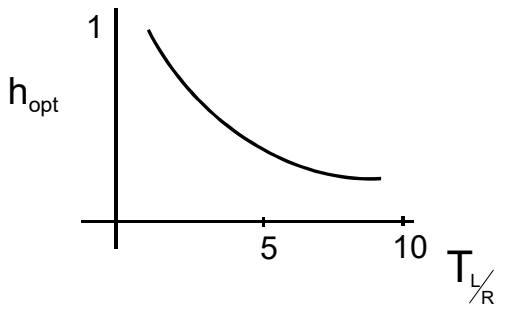
- General case is analytically intractable
  - Use curve fitting

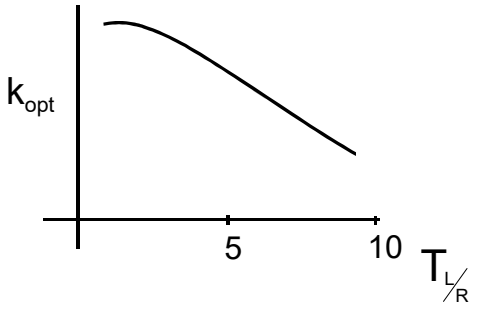
$$h_{opt}\,(RLC) = \sqrt{\frac{R_o\,C_{INT}}{R_{INT}\,C_o}}\;\frac{1}{\left[1+0.16\,(T_{L/R})^3\right]^{0.24}}$$

$$k_{opt}\,(RLC) = \sqrt{\frac{R_{INT}\,C_{INT}}{2\,R_o\,C_o}}\;\frac{1}{\left[1+0.18\,(T_{L/R})^3\right]^{0.3}}$$

0.3

Error < 0.5% of numerical solutions

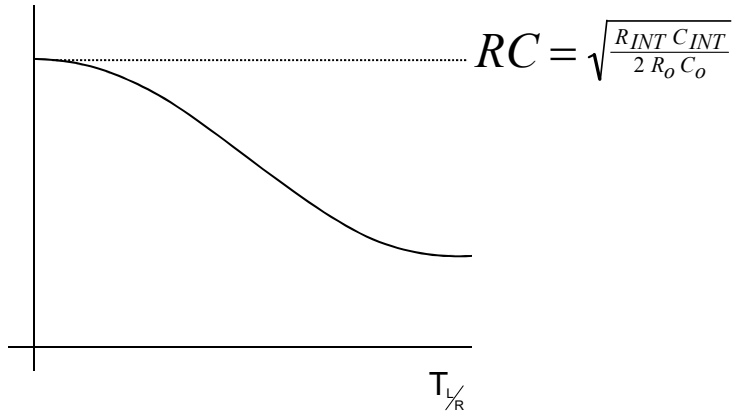




Repeater Insertion in *RLC* Interconnect

h_{opt} (*RLC*) = h_{opt} (*RC*)
k_{opt} (*RLC*) = k_{opt} (*RC*)

} For special case of an *RC* impedance
$L_t \Rightarrow 0 \left(T_{L/R} \Rightarrow 0\right)$



- As $T_{L/R}$ increases, (inductance increases), number of sections k_{opt} decreases

$$T_{PD}(RC) \propto l^2 \qquad T_{PD}(LC) \propto l$$

$$l < T_{PD}\ (RLC) < l^2$$

- Inserting repeaters in *RC* lines to gain performance primarily due to this ℓ^2 relationship
- No repeaters should be inserted in a lossless line ($R_{INT} = 0$),
  - − Would only increase the delay
    - ∴ as L ↑, k_{opt} ↓
- Ignore effects of inductance in repeater insertion process
  - Delay ↑ , area ↑, power ↑

Driving an Inductive Line

- **Repeater insertion methodologies should be reconsidered**
  - **For those cases where inductance cannot be neglected**
- **Delay of *RLC* line**

$$T_{D-RLC} = \sqrt{LC}(e^{2.9(\alpha_{asym}l)^{1.35}}l + 0.74\alpha_{asym}l^2)$$

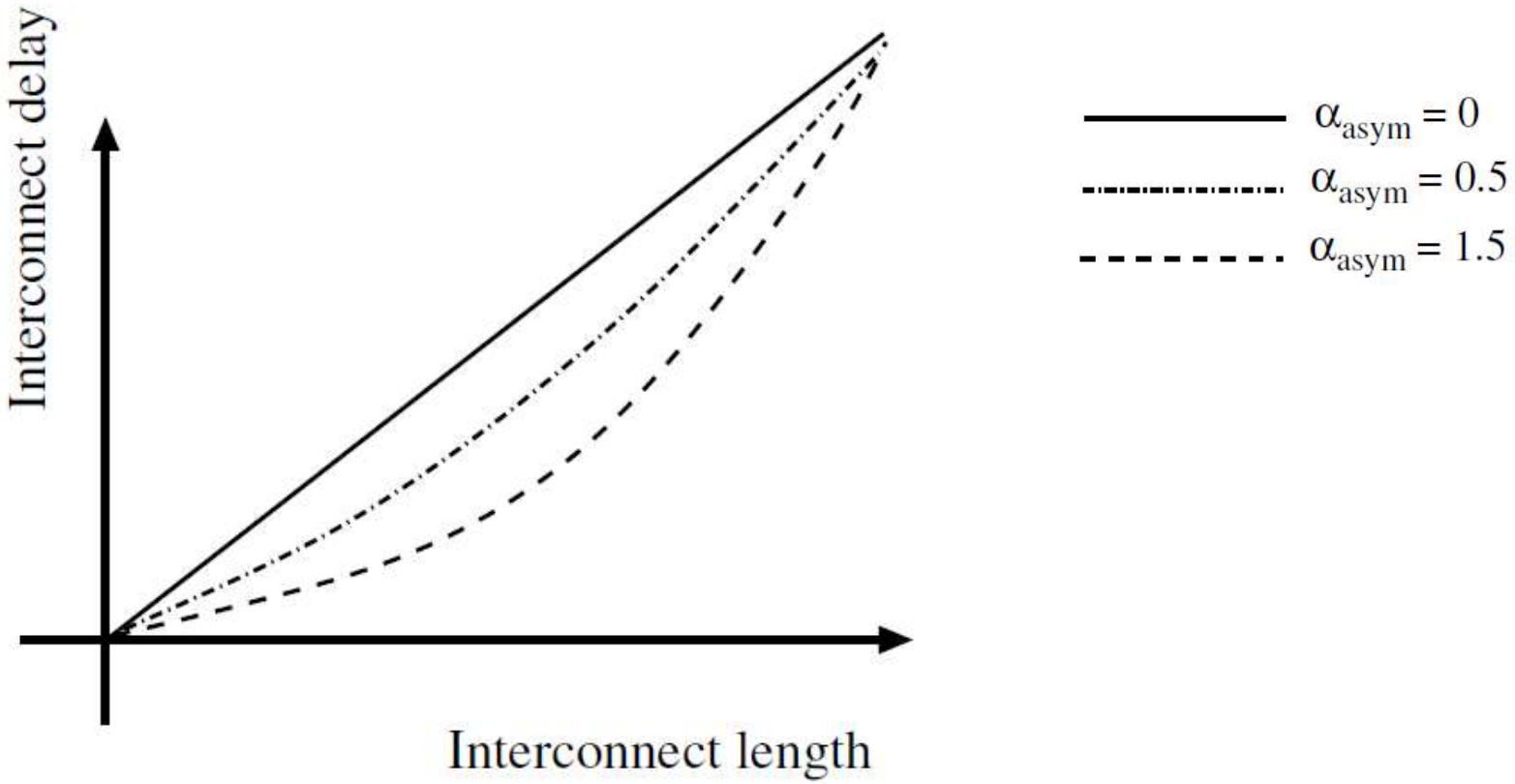
$$\alpha_{asym} = \frac{R}{2}\sqrt{\frac{C}{L}}$$

- **Important conclusion: Quadratic dependence of delay on line length for *RC* lines**
  - **Approaches linear dependence for *RLC* lines**
    - **L → 0 (resistive line)**
      - **Quadratic dependence**
    - **R → 0 (lossless line)**
      - **Linear dependence**

Delay Dependence on Line Length

- **For an inductive line, this dependence**

$$\alpha_{asym} = \frac{R}{2}\sqrt{\frac{C}{L}}$$



Optimum Number and Size of Repeaters

- **Optimum number and size of repeaters for *RC* lines multiplied by *error factor* to find optimum number and size of *RLC* lines**

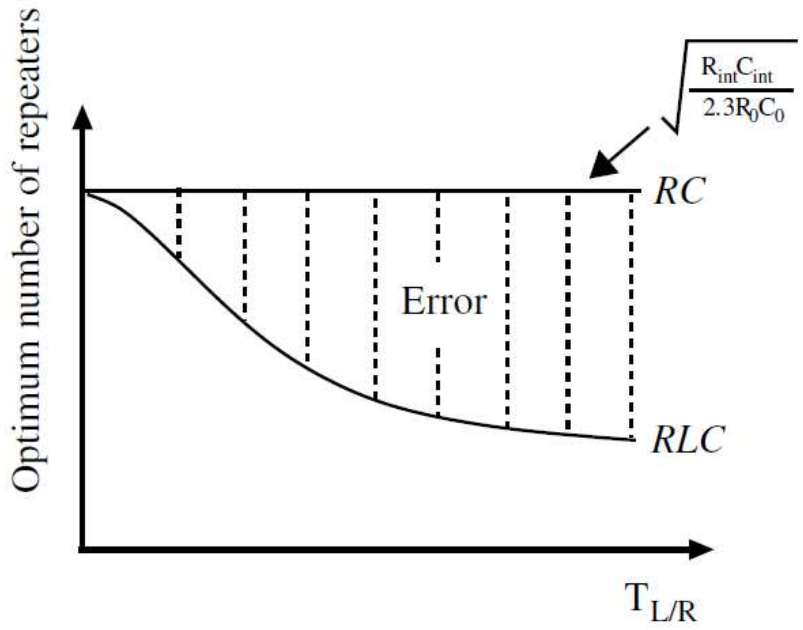
$$k_{opt-RLC} = \sqrt{\frac{R_{int}C_{int}}{2.3R_0C_0}} \times k' \qquad k' = \frac{1}{[1 + 0.18(T_{L/R})^3]^{0.3}}$$

$$h_{opt-RLC} = \sqrt{\frac{R_0C_{int}}{R_{int}C_0}} \times h' \qquad h' = \frac{1}{[1 + 0.16(T_{L/R})^3]^{0.24}}$$

$$T_{L/R} = \sqrt{\frac{L_{int}/R_{int}}{R_0C_0}} = \frac{1}{2\alpha_{asym}}\sqrt{\frac{RC}{R_0C_0}}$$

Y. I. Ismail and E. G. Friedman, "Effects of Inductance on the Propagation Delay and Repeater Insertion in VLSI Circuits," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 8, No. 2, pp. 195–206, April 2000

Effect of Inductance on Optimum Number of Repeaters

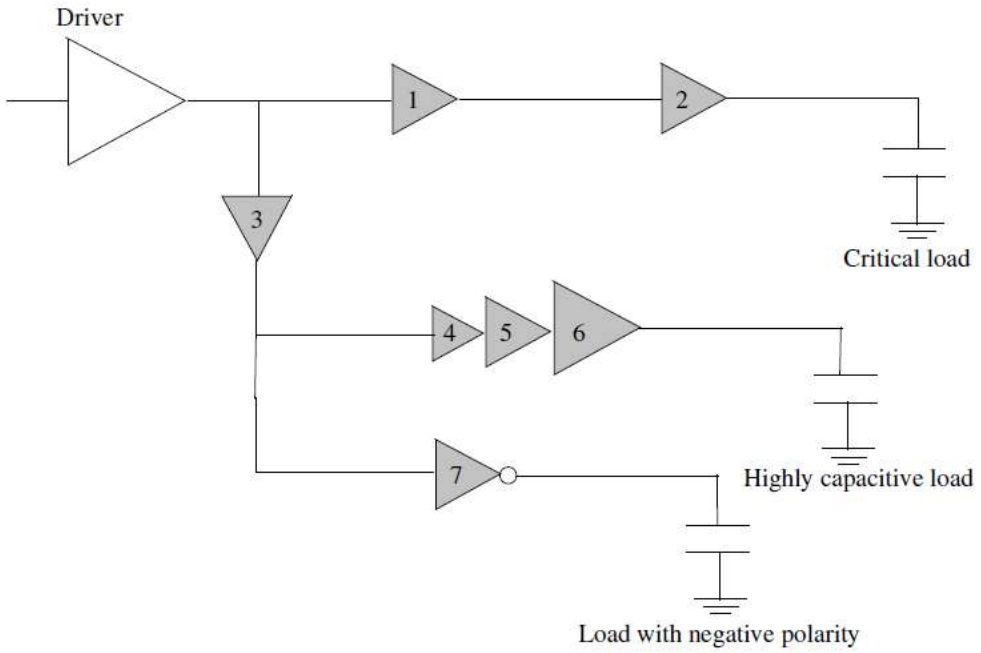


Optimum number of repeaters as a function of $T_{L/R}$ for both RC and RLC models. The error in using an RC model increases as $T_{L/R}$ increases

- **Important conclusions**
  - **Higher error as circuit exhibits more inductive behavior**
  - **Optimum number decreases as inductive effects increase**
    - **Linear dependence of delay on line length**
    - **Additional gate delay of repeaters and unnecessary power dissipation**

Repeater Insertion in Tree Structured Interconnects

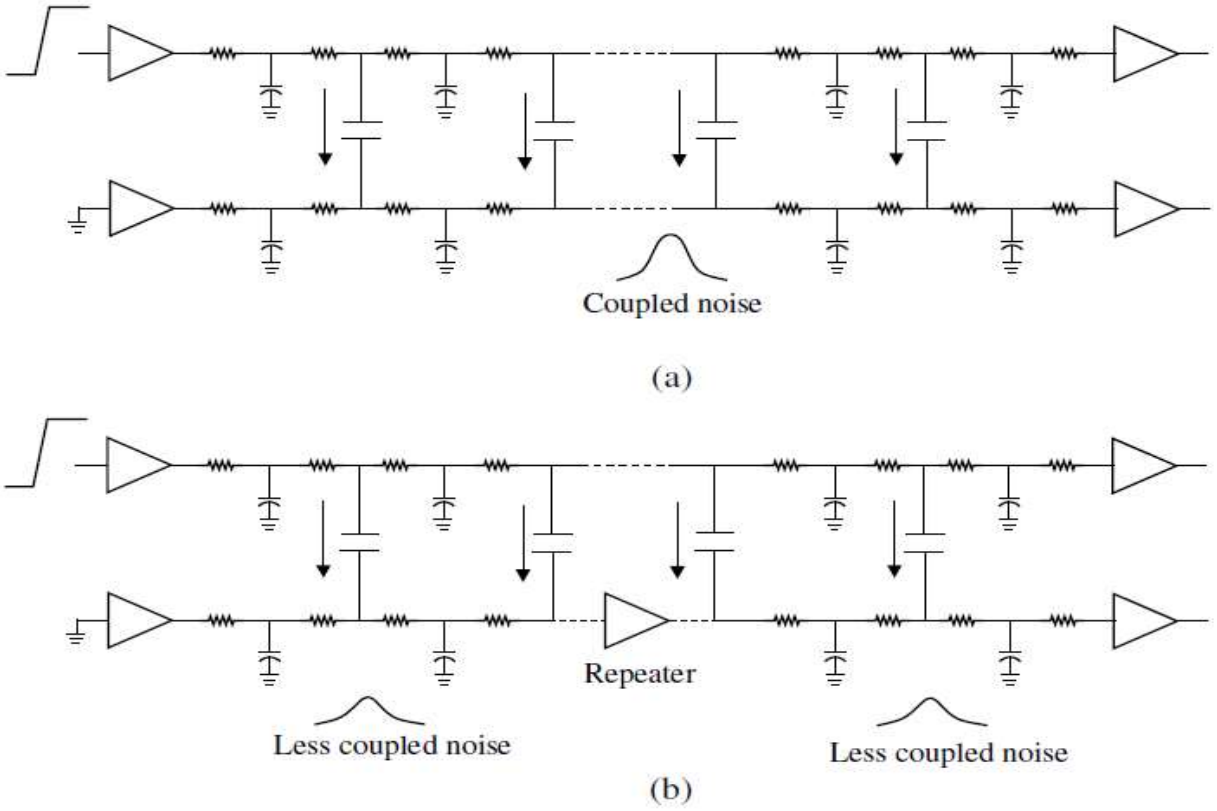
- **Buffered tree is important application of repeater insertion**



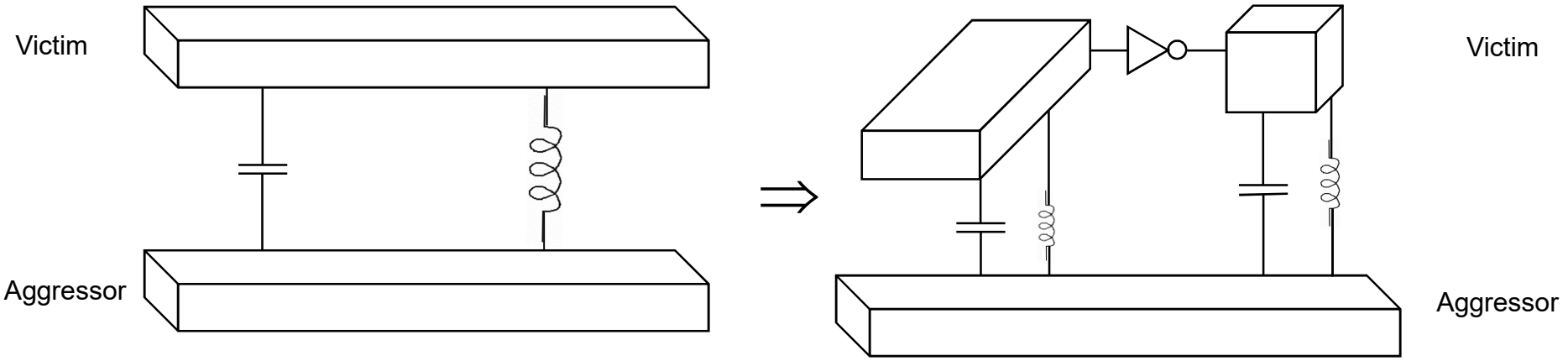
- **Four situations**
  - **Split long interconnect to satisfy delay constraints (1 and 2)**
  - **Isolate large capacitances from critical path (3)**
  - **Drive large capacitances (4, 5, and 6)**
  - **Reversing signal polarity (7)**

Repeater Insertion to Reduce Coupling Noise

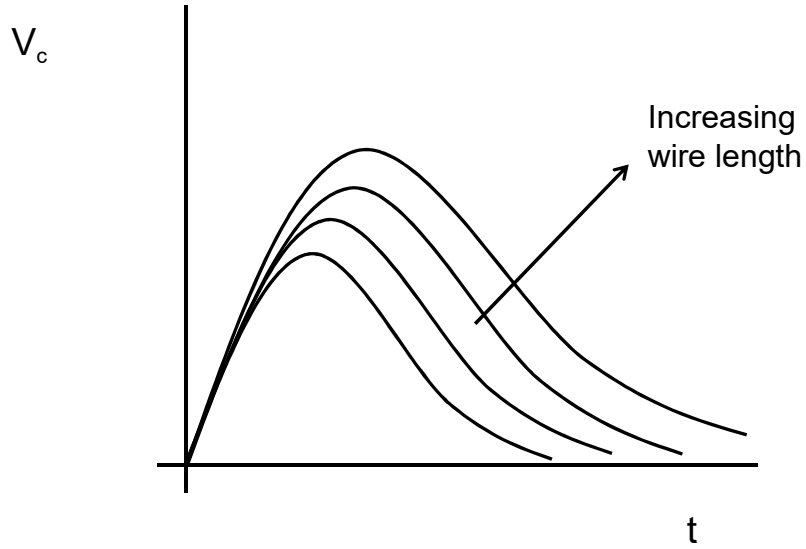
- Repeater insertion not only reduces the delay but also lowers capacitive coupling between interconnects
- Coupling noise proportional to length of two parallel interconnects
- Parallel portion reduced by inserting repeaters



Repeater Insertion - Minimize Coupling Noise

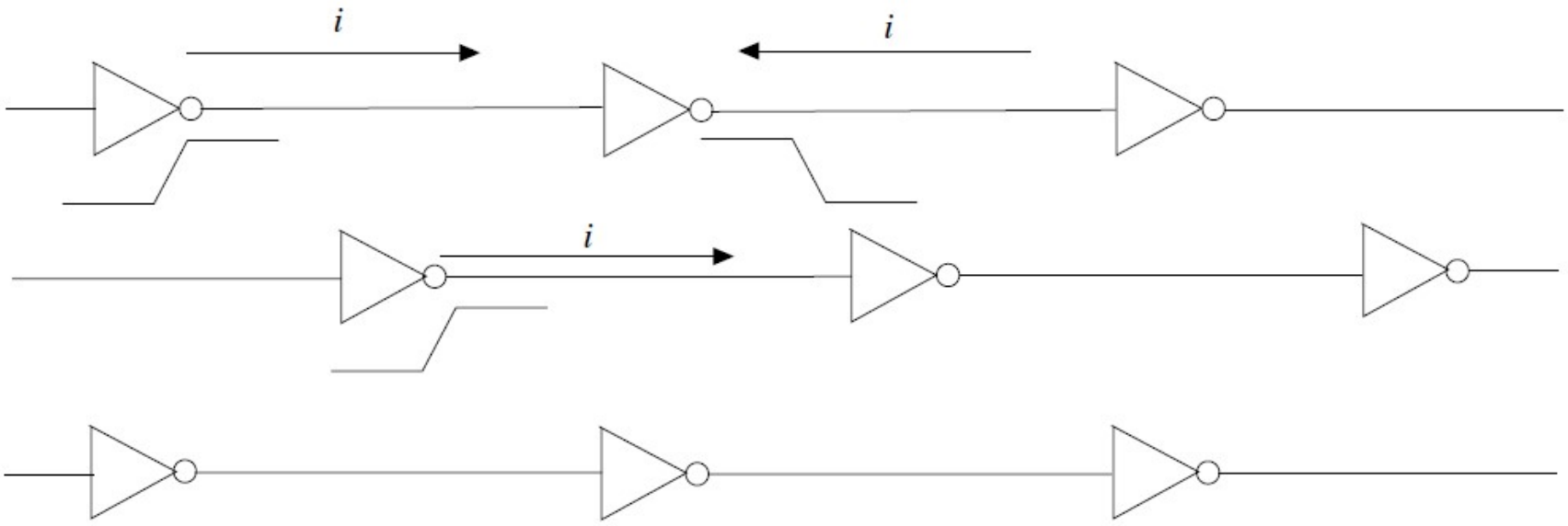


- Much less noise coupling before amplification
  - Signal restoration

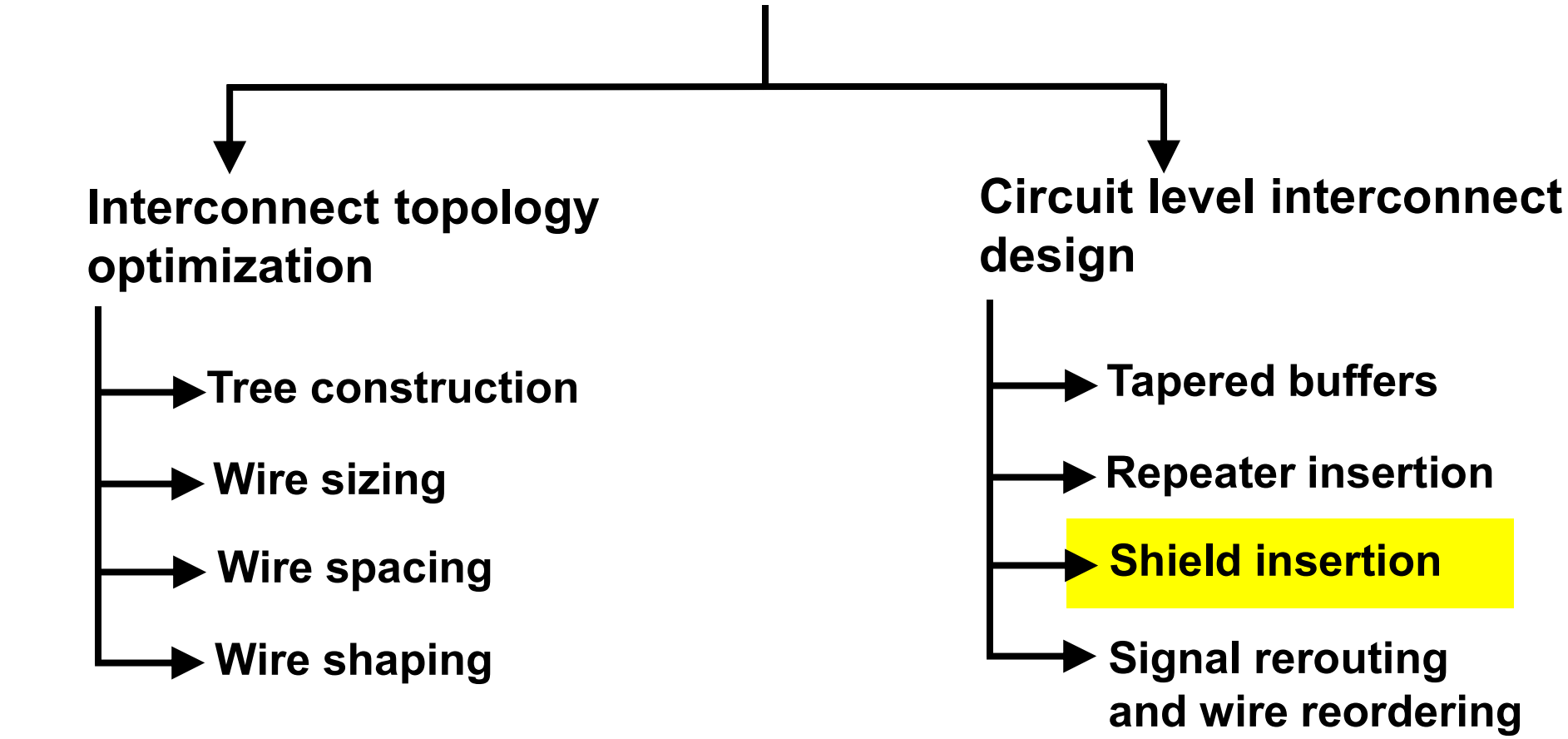


Repeater Staggering

- Reduces worst case delay and crosstalk noise
- Repeaters in adjacent wires are interleaved
- Repeater placed between two adjacent repeaters
- Potential worst case capacitive coupling only for half the line length
- Signals switch in the same direction
  - For other half, coupling is best case
- Delay uncertainty and worst case delay are reduced

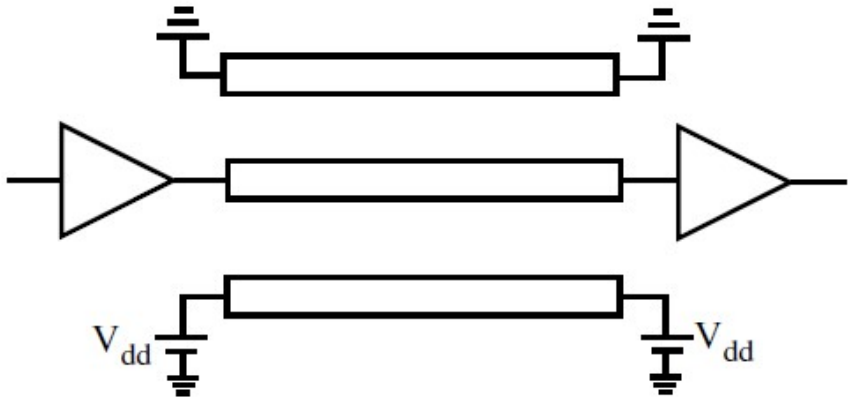


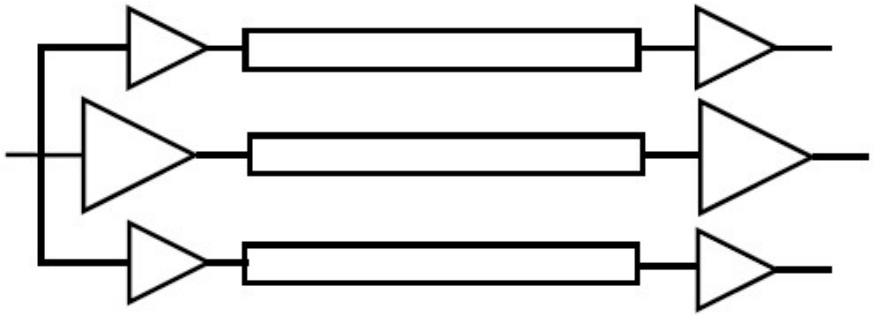
Global Signaling Methodologies



Shield Insertion

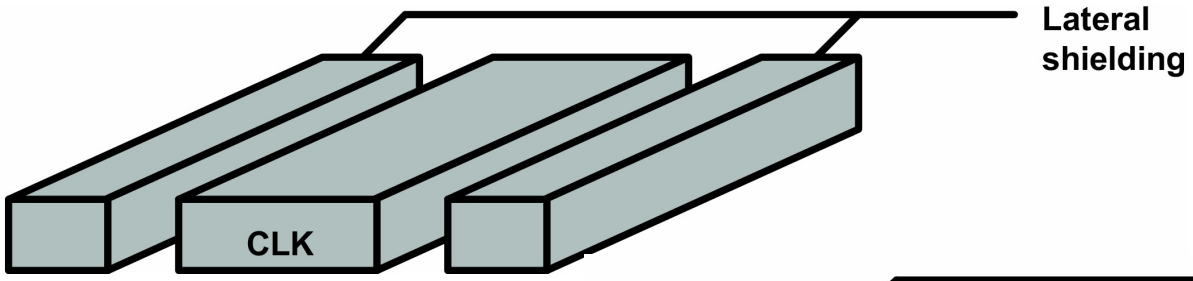
- Widely used technique where power or ground line placed between aggressor and victim to reduce coupling noise
  - Passive shielding
- Active shielding
  - Exploits miller effect
  - Shield line switches in the same direction as the signal line
    - Reducing effective coupling capacitance
- Both techniques consume additional area

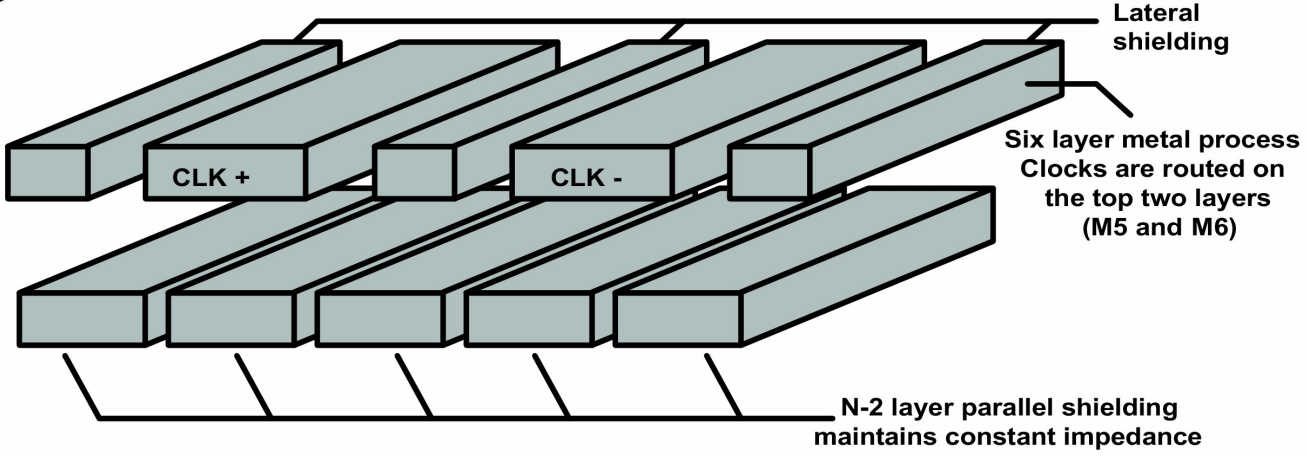




Passive Shielding

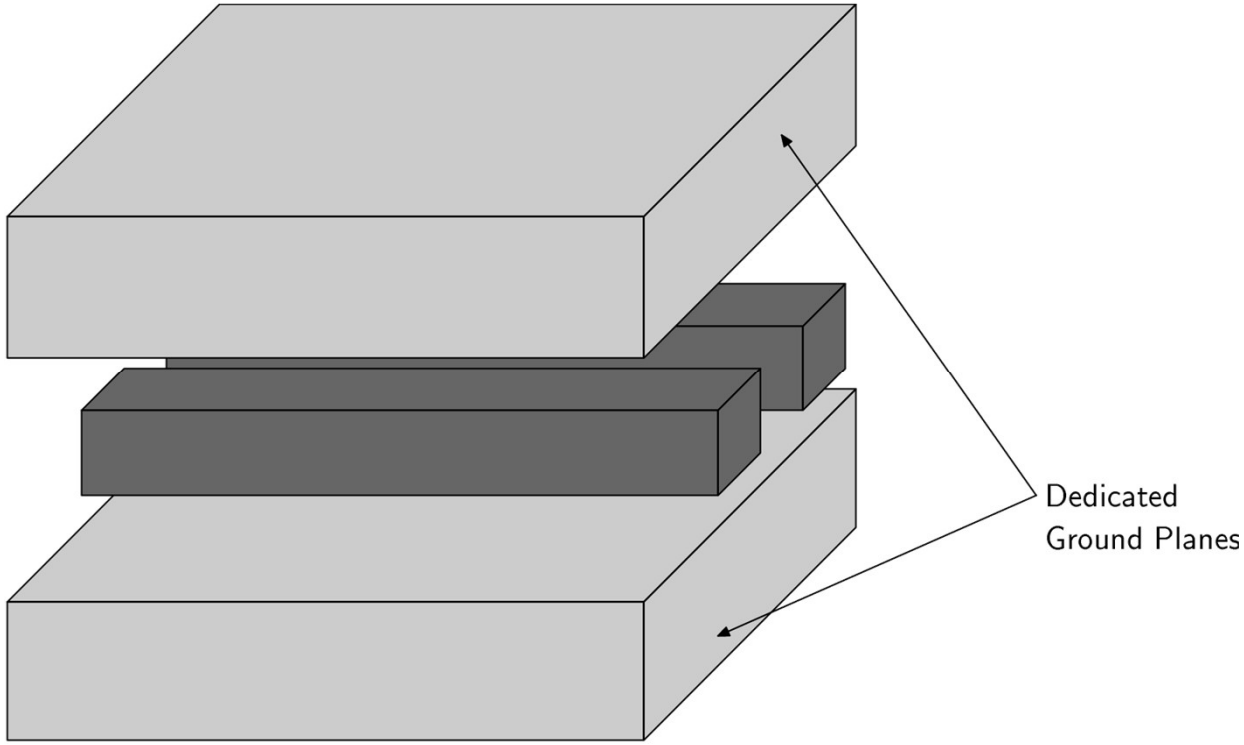
- Signal line isolated from the switching neighbor lines
  - Reduced capacitive coupling noise
- Inductive coupling also reduced due to closer current return path
- Delay uncertainty reduced since effective capacitance is almost fixed
- Clock signals are typically shielded on both sides
- Additional parallel shielding on lower metal layer is possible





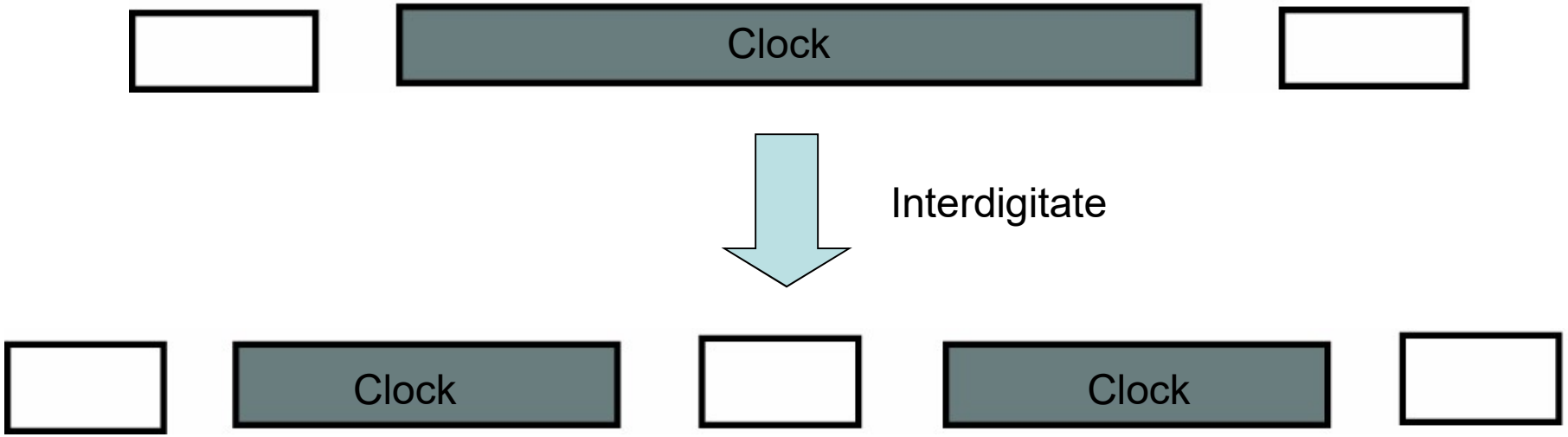
Ground Plane Shielding

- Entire metal plane is dedicated for shielding
- Not practical in modern resource limited ICs
- 600 MHz Alpha processor
  - Clock signal is shielded



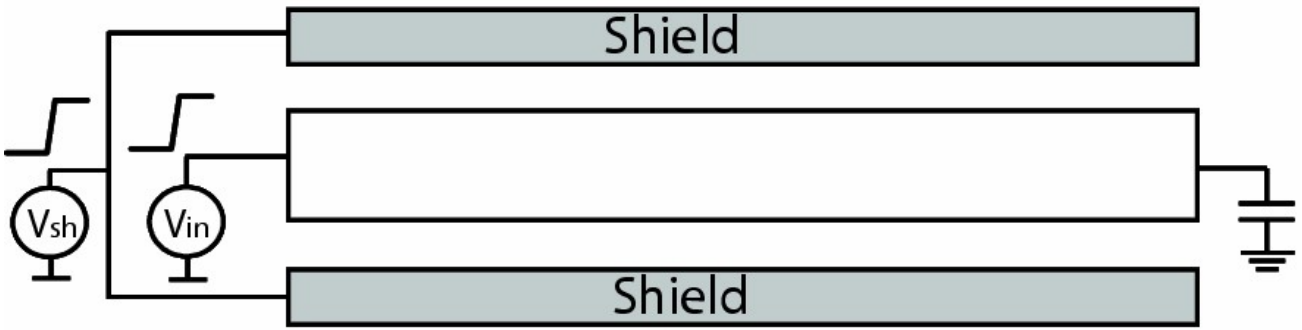
Interdigitated Shielding

- **Forms smaller current loop**
- **Smaller inductance and ringing behavior**
- **Tradeoff between inductance and capacitance/area**



Active Shielding

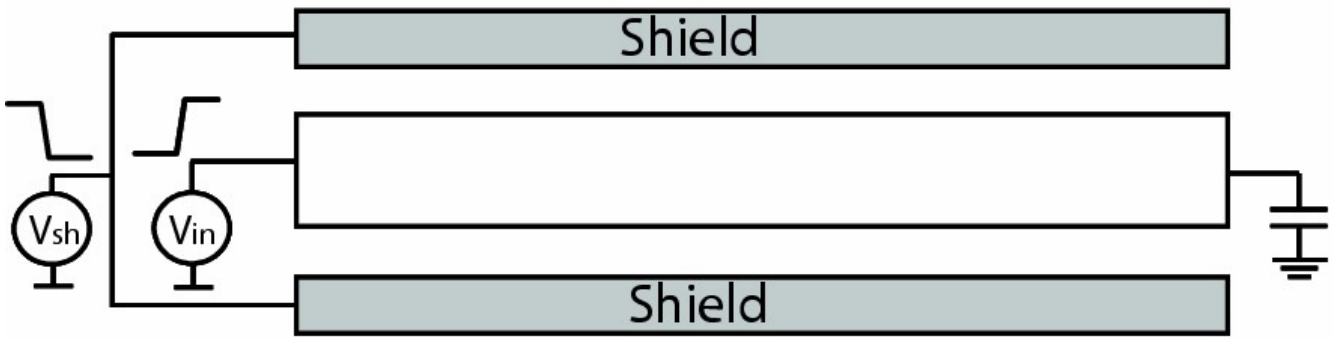
- 16% improvement in performance has been demonstrated
- Consumes more power due to additional switching activity of shield line
- Power due to coupling capacitance is reduced



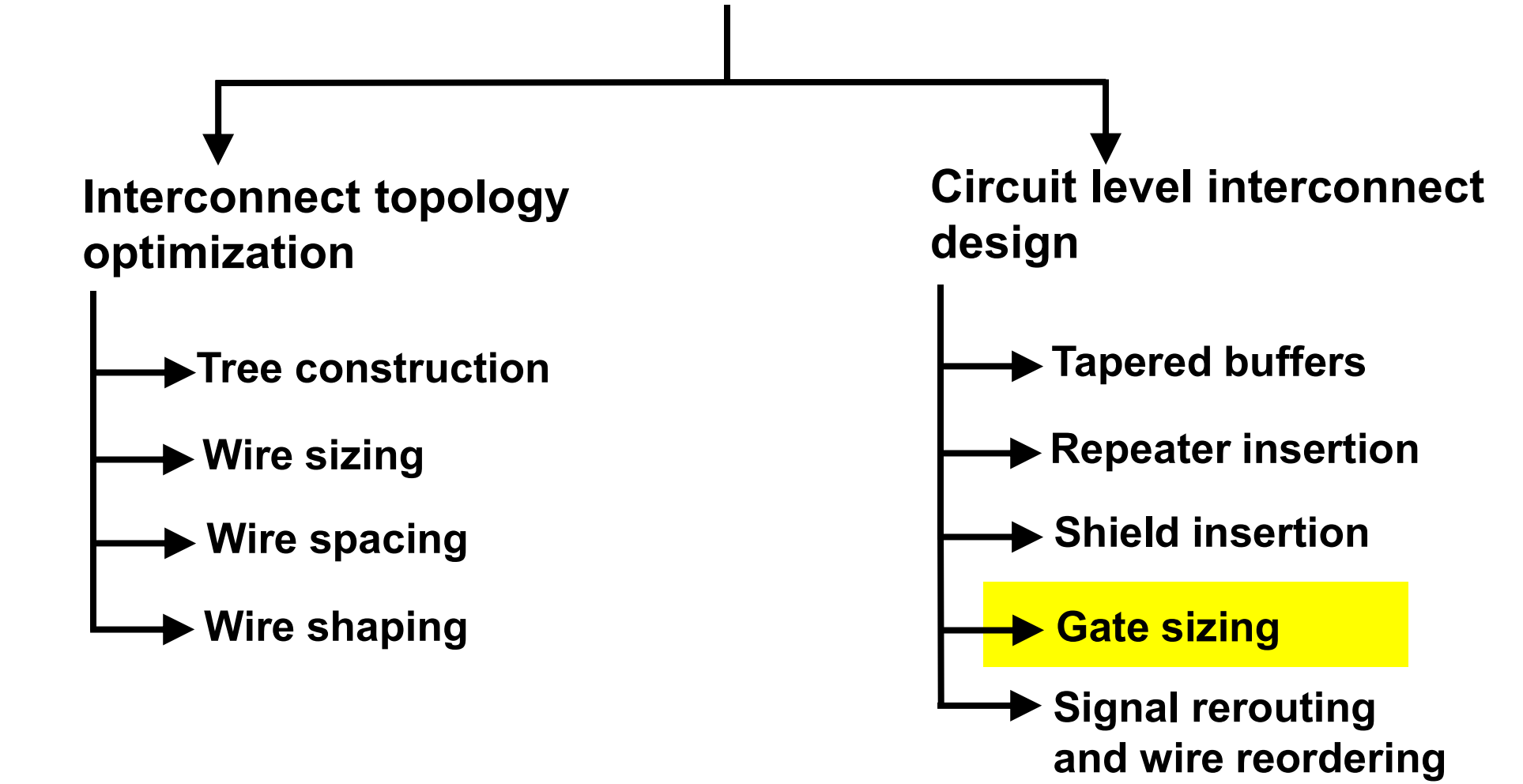
- In-phase switching
  - Active shielding for *RC* lines

Active Shielding for *RLC* Lines

- **Out-of-phase switching active shielding for *RLC* lines**
  - **Exploit effective capacitance to suppress inductive effects**
  - **Higher damping factor**
  - **Less ringing**

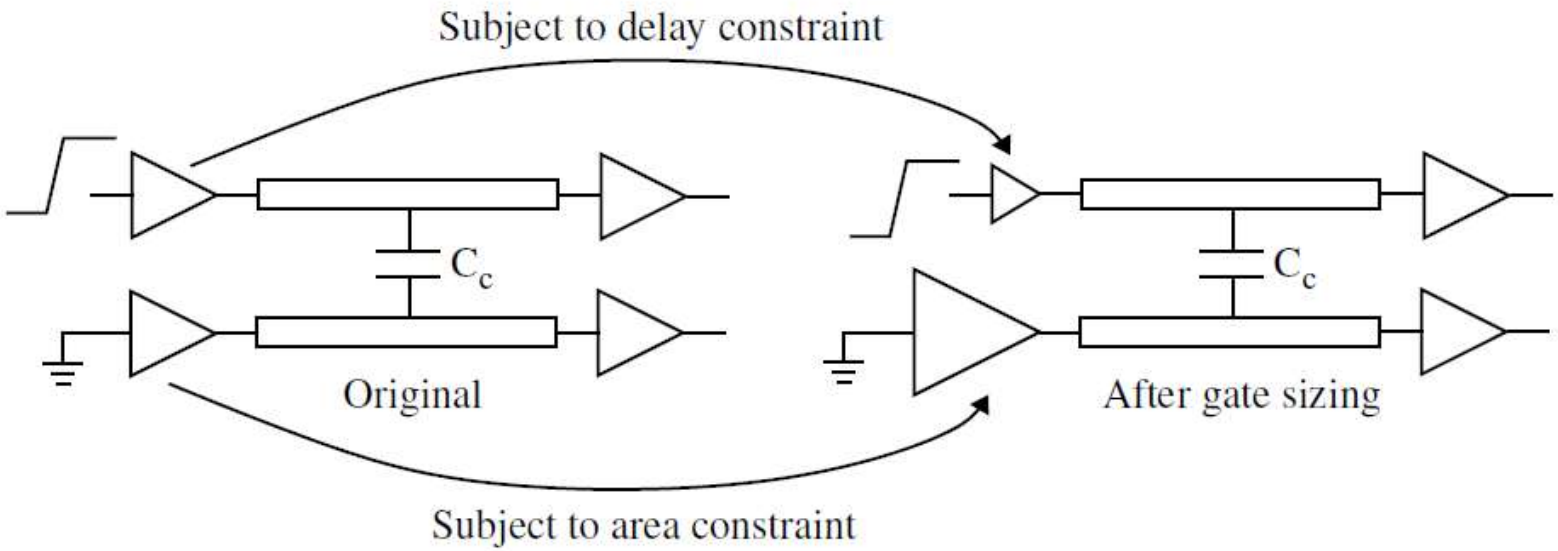


Global Signaling Methodologies



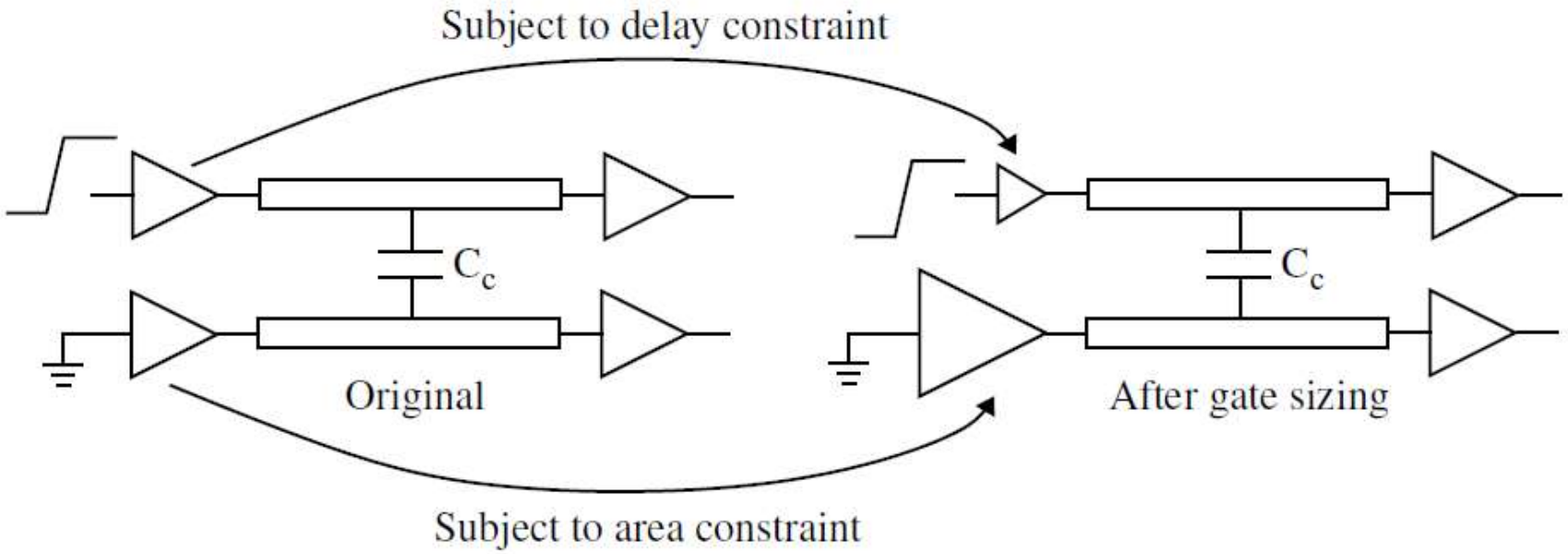
Gate Sizing

- Commonly used technique to exploit tradeoff between speed and power
  - Gates along critical path sized up to satisfy delay
  - Remaining gates sized smaller to reduce power dissipation
- Size of driver and victim also affects coupling noise and noise induced delay variation



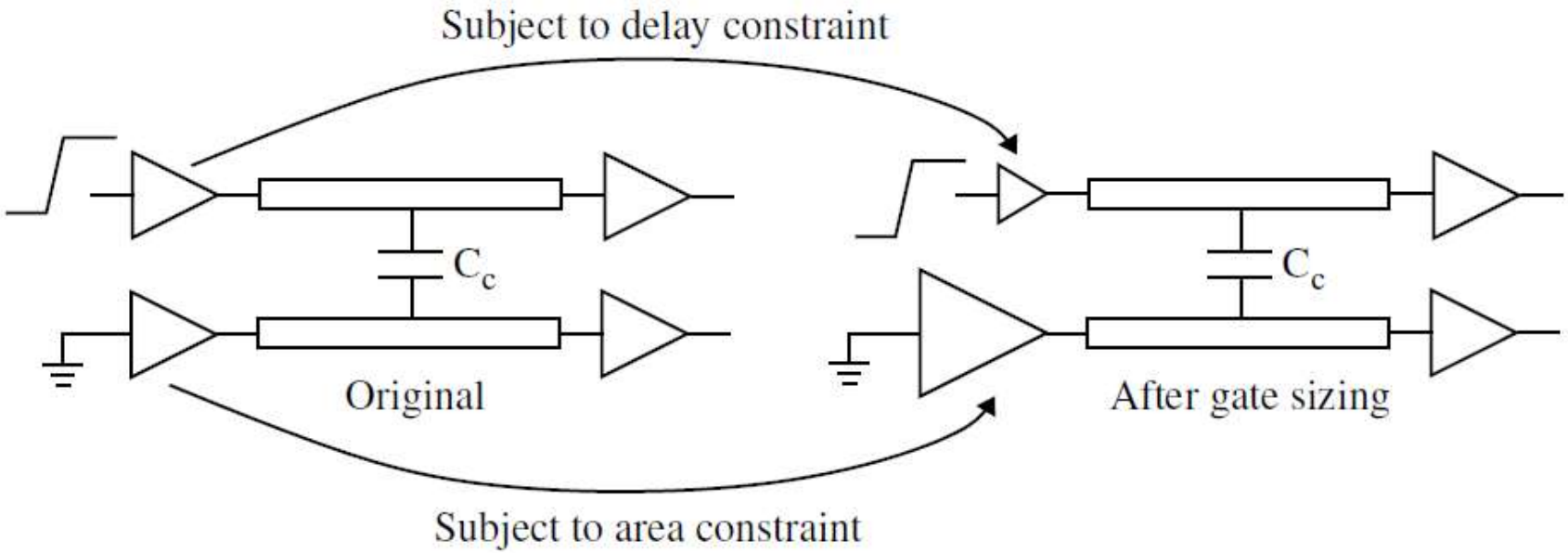
Downsizing the Driver of the Aggressor

- Reduces capacitive coupling noise since driver is weaker
- Slows down signal path
- Tradeoff between delay and coupling noise
- Inductive coupling also reduced since less current is injected

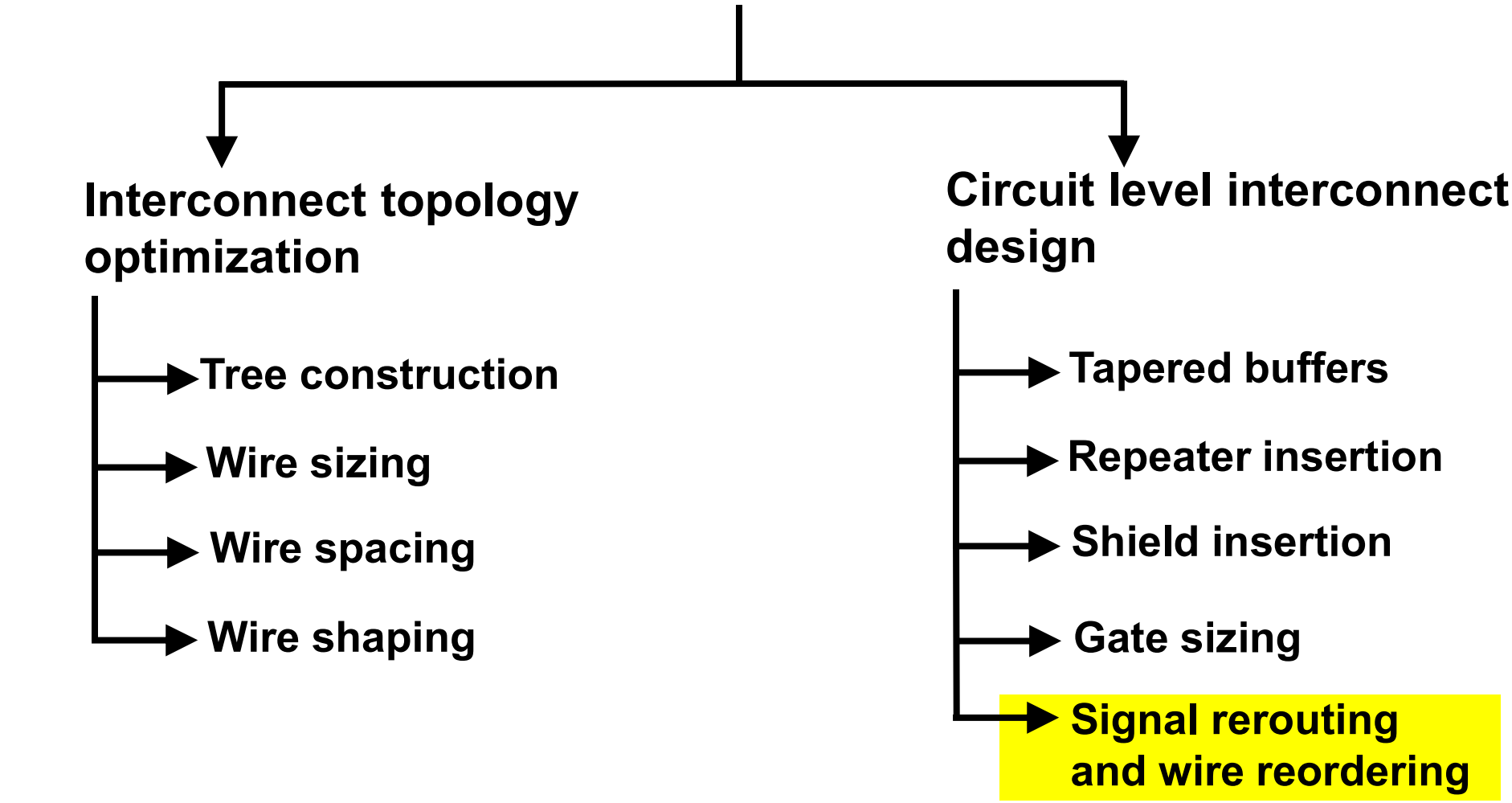


Increasing the Size of Victim Driver

- **Reduces both inductive and capacitive coupling since victim is more effectively connected to ground or Vdd**
- **Increases overall area**
- **Tradeoff between area and coupling noise**



Global Signaling Methodologies

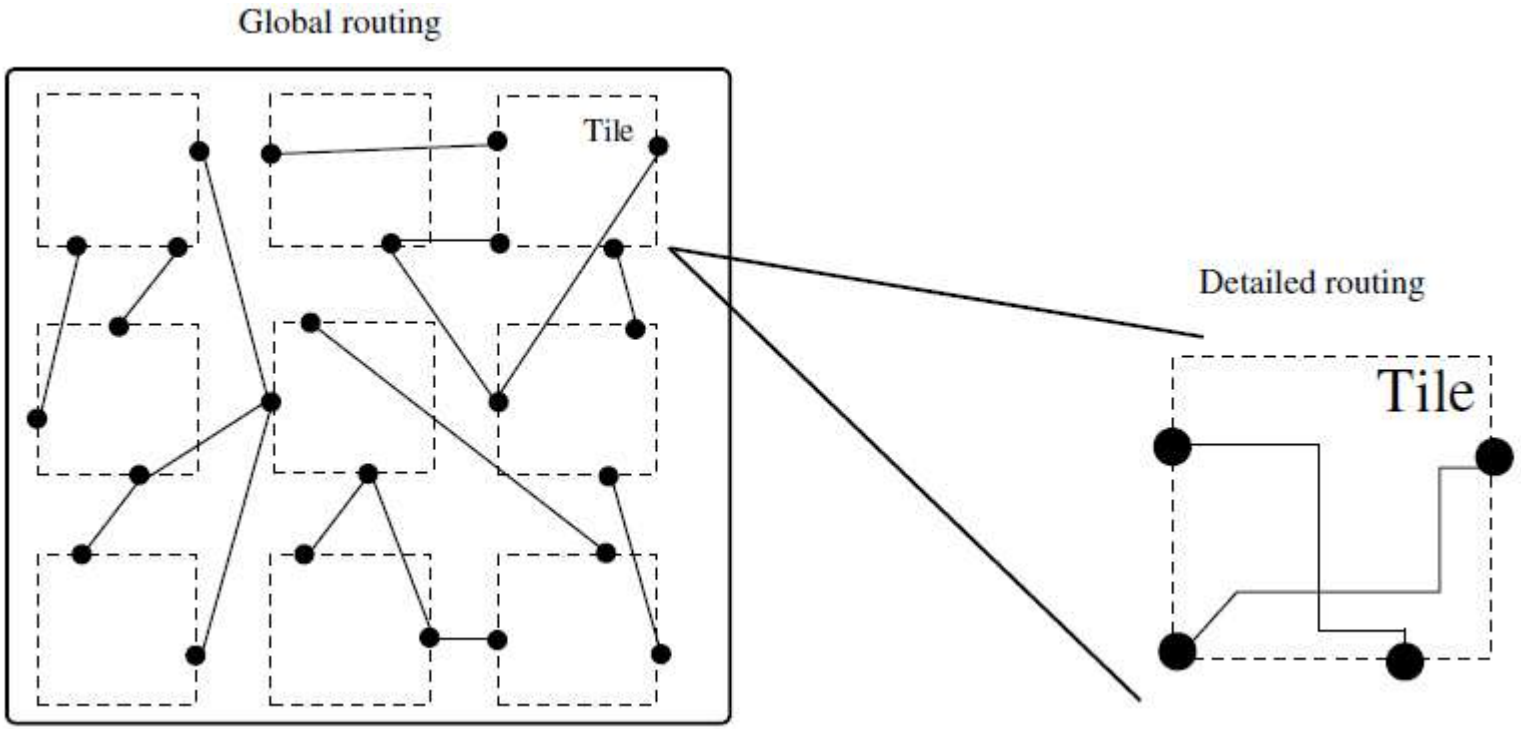


Signal Rerouting

- Interconnect routing has been a focus for a long time
  - Area, power, delay, and noise are all affected
- Target: Given connectivity information, minimize the wire length to achieve highest performance while reducing area and power
- Reconsidered to include crosstalk as a design constraint
  - Spacing and length of overlap between aggressor and victim
- Two step routing process
  - Global routing
  - Detailed routing

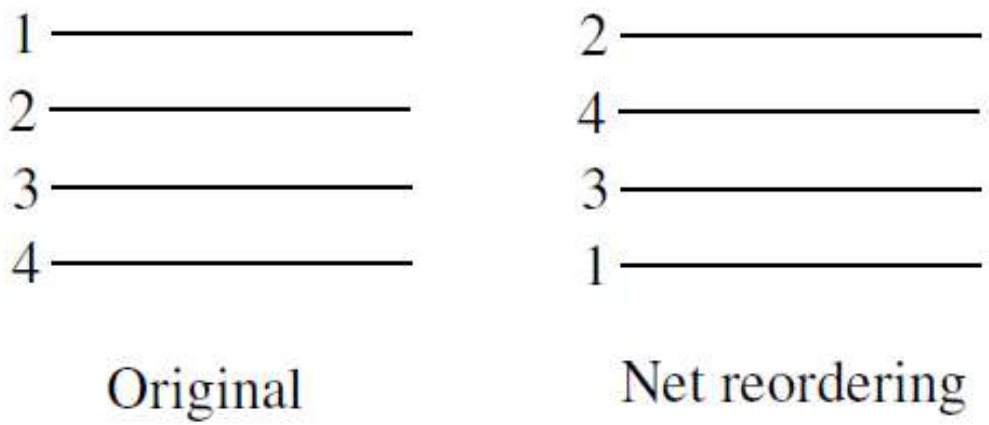
Global and Detailed Routing

- Overall area divided into tiles during global routing
- Path through tiles are determined for each net
- Routing of nets within the tiles achieved during detailed routing



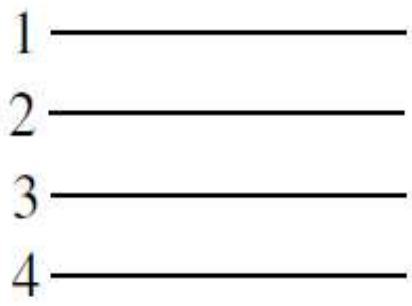
Net Reordering

- Order nets to ensure that sensitive nets are not placed adjacent to each other
- Assume
  - 1 and 2 are mutually sensitive
  - 2 and 3 are mutually sensitive
- Less efficient in reducing long range inductive coupling

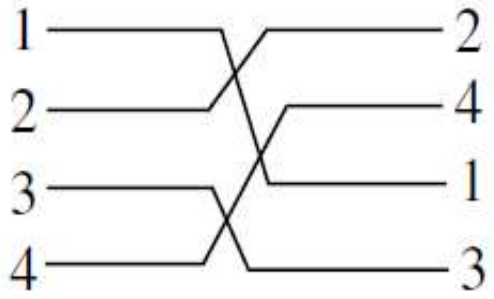


Wire Swizzling

- Wires split into several segments
- Wire sequences in each segment changed to ensure that capacitive coupling averages out for each wire
- Number of permutations required to realize all possible adjacencies is *k/2*
  - For group of *k* wires
- For k = 4, need two permutations
  - 1234 and 2413



Original

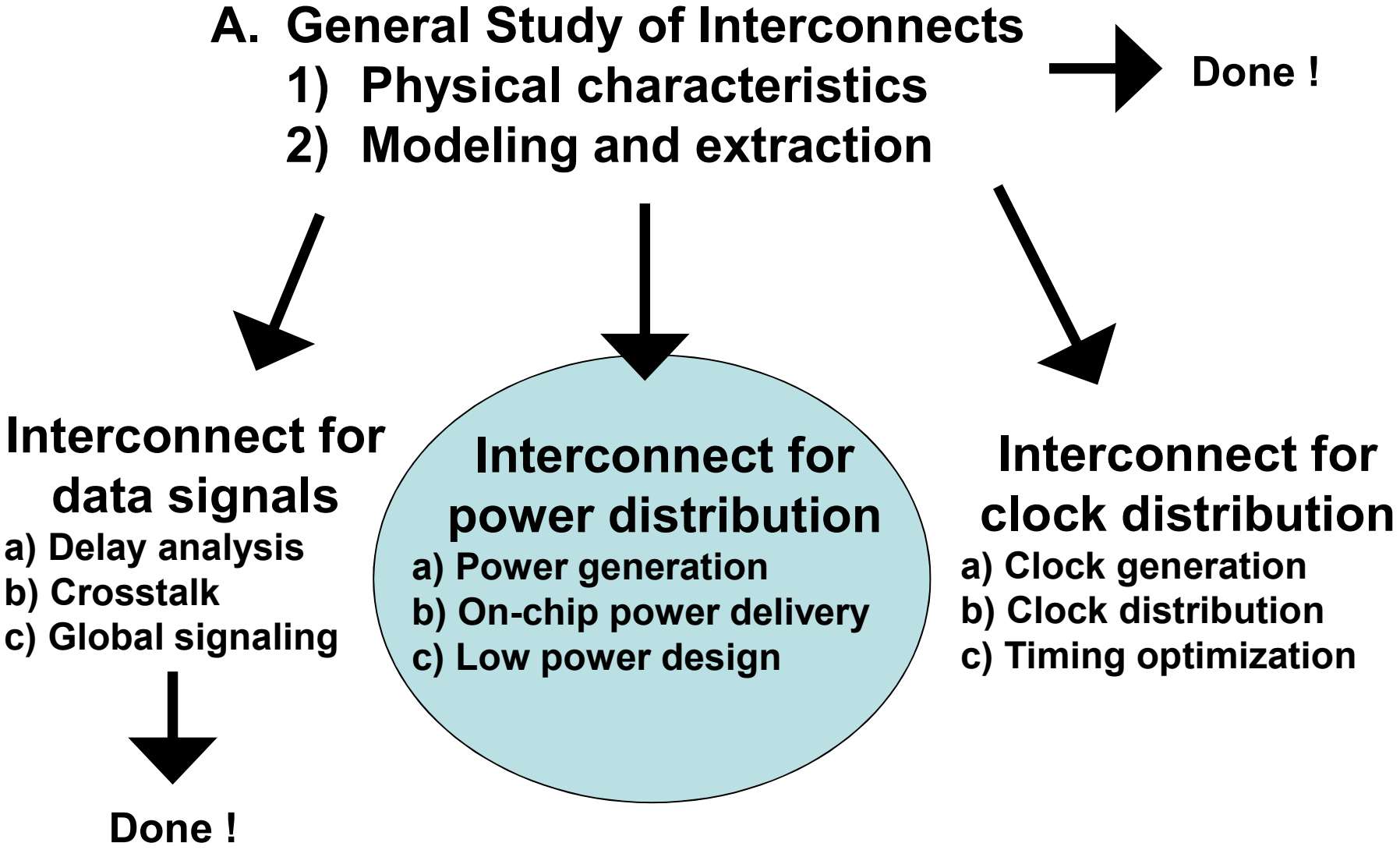


Wire swizzling

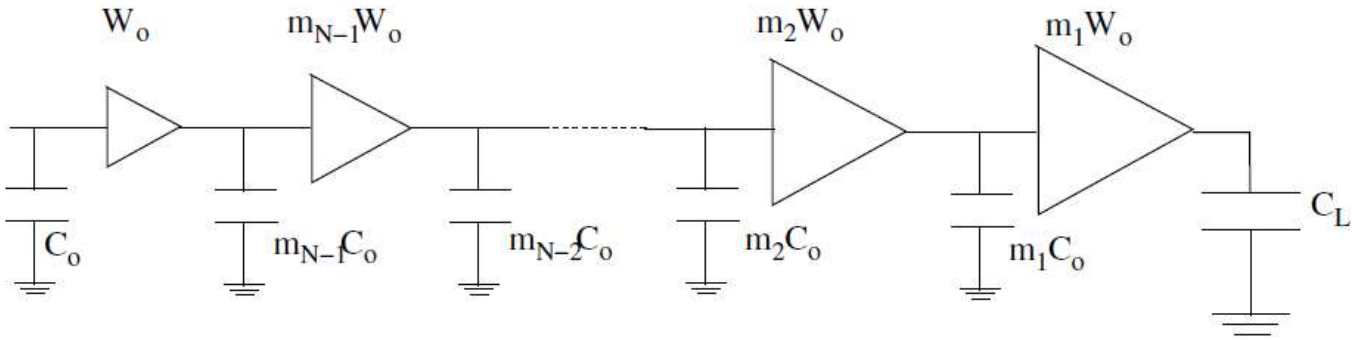
Power/Speed/Noise/Area Tradeoffs

		Delay	Noise		Power	Area
			Capacitive coupling	Inductive coupling		
Wire sizing	Increasing the width of aggressor	Increase	Decrease	Decrease	Increase	Increase
	Decreasing the width of victim	Constant	May increase or decrease	Decrease	Constant	Decrease
Wire spacing		Decrease	Decrease	Slow decrease	Decrease	Increase
Tapered buffer		Decrease	Increase	Increase	Increase	Increase
Repeater insertion		Decrease	Decrease	Increase	Increase	Increase
Passive shield insertion		May increase or decrease	Decrease	Decrease	May increase or decrease	Increase
Active shield insertion		Decrease	Decrease	Increase	Increase	Increase
Gate sizing	Decreasing the size of aggressor driver	Increase	Decrease	Decrease	Decrease	Decrease
	Increasing the size of victim driver	Constant	Decrease	Decrease	Constant	Increase

Interconnect Centric Design



Overall Propagation Delay



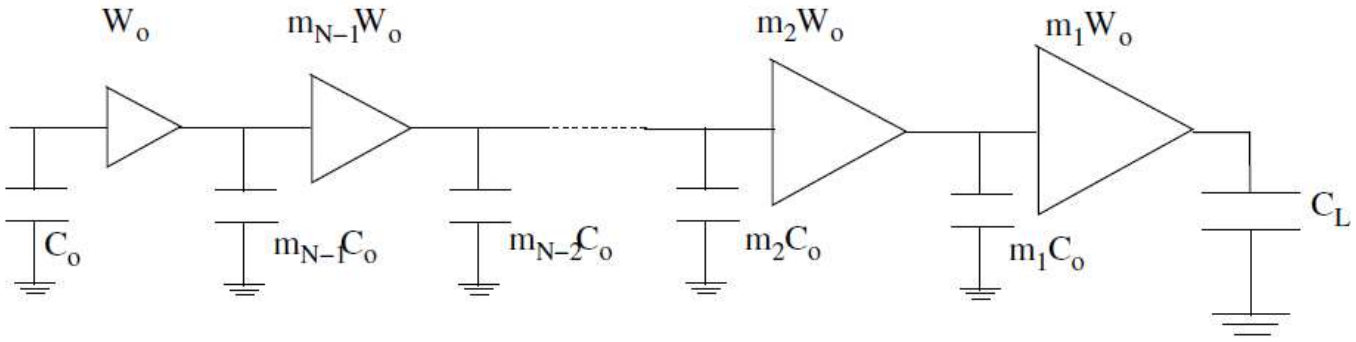
Tapered buffer structure proposed by Lin and Linholm. The factors $m_1, m_2, ..., m_{N-1}$ are selected to equalize the delay of each stage.

- **Assume C_L/C_o = M (load capacitance to interstage capacitance)**
- **Assume delay due to C_o is t_{po}**

$$t_p = t_{po}\left(\frac{M}{m_1} + \frac{m_1}{m_2} + \frac{m_2}{m_3} + ... + \frac{m_{N-2}}{m_{N-1}} + m_{N-1}\right)$$

- **Delay is assumed to be a linear function of the interstage and load capacitances**

Find Minimum Propagation Delay



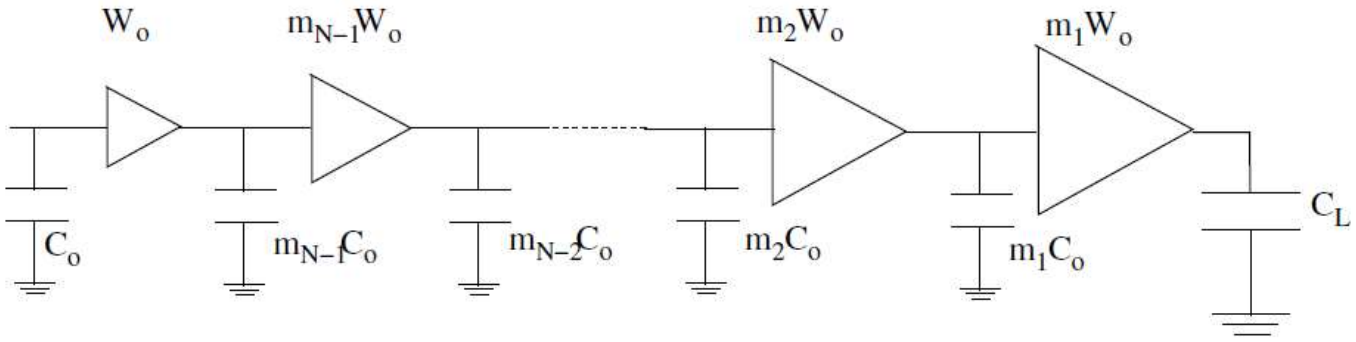
Tapered buffer structure proposed by Lin and Linholm. The factors $m_1, m_2, ..., m_{N-1}$ are selected to equalize the delay of each stage.

$$t_p = t_{po}\left(\frac{M}{m_1} + \frac{m_1}{m_2} + \frac{m_2}{m_3} + ... + \frac{m_{N-2}}{m_{N-1}} + m_{N-1}\right)$$

- **What is the optimum scaling factor? m_1, m_2,…., m_{N-1}**
- **Differentiate total delay with respect to each factor**
- **Set each equation to zero**

$$m_k = M^{(N-k)/N}$$

Find Minimum Propagation Delay



Tapered buffer structure proposed by Lin and Linholm. The factors $m_1, m_2, ..., m_{N-1}$ are selected to equalize the delay of each stage.

$$m_k = M^{(N-k)/N}$$

- **Tapering factor (F) = ratio of the width of any two consecutive inverters**

$$F = \frac{m_k}{m_{k+1}} = \frac{m_{k+1}}{m_{k+2}} = M^{1/N} = \text{tapering factor}$$

- **Total propagation delay will be** $t_p = t_{po} N M^{1/N}$

$$t_p = t_{po}(\log_F M)F$$