Optimal Buffer Sizing w/o Considering Interconnects

- Delay through each stage is $\alpha t_{min}$, where $t_{min}$ is the average delay through any inverter driving an identically sized inverter.

- $\alpha^n = C_L/C_g \Rightarrow n = \ln(C_L/C_g)/\ln \alpha$, where $C_L$ is the capacitive load and $C_g$ the capacitance of the minimum size inverter.

- Total delay $T_{tot} = n\alpha t_{min} = \frac{\alpha}{\ln \alpha} t_{min} \ln \frac{C_L}{C_g}$.

- Optimal stage ratio: $\frac{dT_{tot}}{d\alpha} = 0 \Rightarrow \alpha = e$.

- Optimal delay: $T_{opt} = e t_{min} \ln(C_L/C_g)$.

- Buffer sizes are exponentially tapered ($\alpha = e$).
**Wire Sizing**

- Wire length is determined by layout architecture, but we can choose wire width to minimize delay.
- Wire width can vary with distance from driver to adjust the resistance which drives downstream capacitance.
- Wire with minimum delay has an exponential taper.
- Can approximate optimal tapering with segments of a few widths.
- Recent research claims that buffering is more effective than wire sizing for optimizing delay, and two wire widths are sufficient for area/delay trade-off.
**Optimal Wire-Sizing Function**

- Suppose a wire of length $L$ is partitioned into $n$ equal-length wire segments, each of length $\Delta x = L/n$; unit resistance and capacitance: $\frac{\tilde{R}}{\tilde{C}}$
- The respective resistance and capacitance of $l$-th wire segment can be approximated by $\frac{\tilde{R}}{\tilde{C}} \Delta x/ f(x_i)$ and $\frac{\tilde{C}}{\tilde{C}} \Delta x f(x_i)$, where $f(x_i)$ is the width at position $x_i$.
- Elmore delay: $D_n = R_d \left( C_L + \sum_{i=1}^{n} \tilde{\alpha} f(x_i) \Delta x \right) + \sum_{i=1}^{n} \frac{\tilde{R} \Delta x}{f(x_i)} \left( \sum_{j=i}^{n} \tilde{\alpha} f(x_j) \Delta x + C_L \right)\]
- As $n \to \infty$, $D_n \to D$: $D = R_d \left( C_L + \int_0^L \tilde{\alpha} f(x) dx \right) + \int_0^L \frac{\tilde{R}}{f(x)} \left( \int_x^L \tilde{\alpha} f(t) dt + C_L \right) dx$
- Optimal wire sizing function $f(x) = ae^{-bx}$, where
  
  $a = \frac{\tilde{R}}{bR_d}, \quad b = \sqrt{\frac{R_d C_L}{\tilde{R} \tilde{C}}} - e^{-bL/2} = 0$. 

![Diagram](attachment:image.png)
Wire Sizing & Buffering for Delay Optimization

- Size wires and buffers to optimize delay.
- Trade-off among area, delay, power, clock skew, clock-skew sensitivity, crosstalk, crosstalk sensitivity, etc.
- Popular techniques:
  - Discrete wire/buffer sizing: dynamic programming
  - Continuous wire/buffer sizing: mathematical programming (e.g., Lagrangian relaxation)
Simultaneous Retiming and Clock Skew Scheduling

- Liu, Papaeftymiou, Friedman: Simultaneous retiming and useful clock skew scheduling can further reduce clock period, DAC-99.

- Case 1 (original circuit)
  - Zero clock skew: clock period $\phi = 23\tau$.
  - Schedule $e$ $4\tau$ earlier than $f$: clock period $\phi = 19\tau$. 

![Diagrams for case 1, case 2, case 3]
Simultaneous Retiming and Clock Skew Scheduling (Cont’d)

- Case 2 (fastest retimed circuit with zero skew)
  - Zero clock skew: clock period $\phi = 16\,\tau$.
  - Schedule $f\,1\tau$ earlier than $e$: clock period $\phi = 15\,\tau$.
- Case 3 (fastest retimed circuit with nonzero skew)
  - Zero clock skew: clock period $\phi = 18\,\tau$.
  - Schedule $e\,4\tau$ earlier than $f$: clock period $\phi = 14\,\tau$.

![Diagram of combinational logic and min/max delay for cases 1, 2, and 3]
Nanometer Electrical Effects

\[ \text{In } \lesssim 0.18 \mu \text{m wire-to-wire cap dominates } (C_W >> C_S) \]

- Interconnect delay
  - Delay modeling
  - Interconnect-driven design flow

- Signal Integrity
  - Crosstalk
    - Capacitive coupling
    - Inductive coupling
  - IR drop

- Manufacturability
  - Antenna effect
  - Metal fill for planarity
  - Optical proximity correction

- Reliability
  - Electromigration
Interconnect Delay Creates the Timing Closure Problem

![Graph showing the relationship between technology node and interconnect delay versus gate delay, as well as timing errors over iterations for different technology nodes.](image_url)
VLSI Design Flow

- **Traditional design flow**
  - Two-step process
  - Physical design is performed independently after logic design

- **In nanometer design**
  - Interconnect dominates delay
  - Timing closure
  - Signal integrity

- **New design flow**
  - Capture real technology behaviors early in the design flow
  - Break the iteration between physical design and logic design
Interconnect-Centric Design for Timing Closure

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

- Shall integrate buffer-block design into floorplanning
Buffer Block Planning for Timing Closure

- Pick tiles for buffer insertion

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

Feasible region

Buffer block
Power/Ground Routing

- Are usually laid out entirely on metal layers for smaller parasitics.
- Two steps:
  1. **Construction of interconnection topology**: non-crossing power, ground trees.
  2. **Determination of wire widths**: keep voltage drop small, widen wires for more power-consuming modules and higher density current

![Diagrams](image_url)
IR (Voltage) Drop

- Power consumption and rail parasitics cause actual supply voltage to be lower than ideal
  - Metal width tends to decrease with length increasing in nanometer design

- Effects of IR drop
  - Reducing voltage supply reduces circuit speed (5% IR drop => 15% delay increase)
  - Reduced noise margin may cause functional failures
Design Flows

- Typical flow vs. suggested flow

![Flowchart Diagram]

- Floorplanning
  - P&R
  - RC Extraction
  - Simulation
  - Power Analysis
  - iterative loop

- Voltage Drop Analysis
  - P&R
  - RC Extraction
  - Simulation
  - Power Analysis

OK

yes

no
Noise (Crosstalk)

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

\[ c_{ij} = \frac{\hat{f}_{ij} l_{ij}}{d_{ij} - \frac{x_i + x_j}{2}} \]

\[ = \frac{\hat{f}_{ij} l_{ij}}{d_{ij}} \left( 1 - \frac{x_i + x_j}{2 d_{ij}} \right) \]

\( \hat{f}_{ij} \): unit-length fringing capacitance

- Include coupling effect into wire capacitance
  - consider crosstalk effect on delay
  - \( c_i = \hat{c}_i x_i + f_i + 2 \sum c_{ij} \)
Crosstalk Becomes a First-Order Problem

![Graph showing the increase in worst-case interconnect delay due to crosstalk as technology nodes decrease. The x-axis represents technology node in nanometers (nm), and the y-axis represents delay in picoseconds (ps). The graph illustrates the degradation of interconnect delay and gate delay as technology nodes shrink, emphasizing the importance of crosstalk in modern circuit design.]
Techniques for Capacitive Crosstalk Reduction

- Capacitive noise is a local effect
- Capacitive crosstalk reduction:
  - Shielding
  - Track permutation: Gao & Liu, ICCAD93.
  - Buffering: Apert, Devgan, Quay, DAC98.
  - Wire ordering + Gate/wire sizing: Jiang, Chang, Jou, IEEE TCAD00.
  - Wire spacing: Saxena & Liu, DAC99; Pan & Chang, ICCD00.
  - Layer & track assignment: Ho, Chang, Chen, Lee, ICCAD03.
Simplified Capacitive Crosstalk Computation

- Simplification in gridded channel routing
  - Adjacent tracks or columns only
  - Horizontal or vertical segments overlapping only
  - Linear distance model ($k = 1$)

- Example: $k = 1$, $C_{ij} = l_{ij}$ (adjacent tracks/columns only)
  - Net 4: Consider both vertical and horizontal segments $\Rightarrow C_{4,j} = 3 + 3 + 4 + 5 + 2 = 17$
  - Net 4: Consider horizontal segments only $\Rightarrow C_{4,j} = 4 + 5 = 9$

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

\[ C_{4,j} = 3 + 3 + 4 + 5 + 2 = 17 \]
Wire Ordering for Crosstalk Minimization

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

- When two adjacent wires switch in the same direction, the anti-Miller effect shortens transition.
- When two adjacent wires switch in opposite directions, the Miller effect lengthens transition.
- Crosstalk value should be modeled by the product of switching dissimilarity and coupling capacitance.
- Crosstalk between neighboring wires $i$ and $j$: 
  \[ \text{crosstalk}(i,j) = \text{switching\_dissimilarity}(i,j) \times \text{coupling\_capacitance}(i,j) \]

\[ \text{Anti-Miller Effect} \]
shortens transition

\[ \text{Miller Effect:} \]
lengthens transition
Switching Dissimilarity

- Switching dissimilarity between wires $i$ and $j$:

\[
\text{switching\_dissimilarity}(i,j) = 1 - \frac{\int_{0}^{T_D} f(i,t)f(j,t)\,dt}{T_D}
\]

wire 1

- $f(1,t)$
- $T_D$
- $t$

wire 2

- $f(2,t)$
- $T_D$
- $t$

wire 3

- $f(3,t)$
- $T_D$
- $t$

Graph:

- Node 1
- Node 2
- Node 3

- Edge 1-2: 0.8125
- Edge 1-3: 0.6875
- Edge 2-3: 0.75
• NOISE is a crucial concern in nanometer technology

Goal: simultaneous optimization
Taming Noise and Other Objectives

Switching Dissimilarity Consideration
Wire Ordering

Simultaneous post-layout optimization by sizing circuit component

Lagrangian Relaxation

Noise Area Delay Power
Elmore delay model: $D_i = r_i C_i$

$D_i$: delay of node $i$; $C_i$: downstream capacitance

$r_i = \hat{r}_i / x_i$
$c_i = \hat{c}_i x_i$
Crosstalk-constrained Multi-Objective Optimization

\[ M: \]
Minimize \[ A \] Area
Subject to \[ D \leq D^B \] Delay constraints
\[ X \leq X^B \] Crosstalk constraints
\[ P \leq P^B \] Power constraints
\[ L \leq x \leq U \] Sizing constraints
Lagrangian Relaxation

LP Problem $P$

arrival time variables

LP Problem $PP$

Lagrangian Relaxation

LR Problem $LRS_1$

optimality conditions

LR Problem $LRS_2$
Neighborhood and Dominating Index

- Neighborhood $N(i)$: the set of $i$'s adjacent wires
- Dominating index $I(i)$: $\{ j \mid j > i \text{ and } j \in N(i) \}$

- $N(2) = \{1\}$, $I(2) = \emptyset$
- $N(1) = \{2, 3\}$, $I(1) = \{2, 3\}$
- $N(3) = \{1\}$, $I(3) = \emptyset$
### Problem Formulation $P$

<table>
<thead>
<tr>
<th>Minimize</th>
<th>$\sum_{i=s+1..n+s} \alpha_i x_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject to</td>
<td>$\sum_{i \in \delta} D_i \leq A^B, \forall \delta \in \Delta,$</td>
</tr>
<tr>
<td></td>
<td>$\sum_{i \in W} \sum_{j \in l(i)} c_{ij} \leq X^B,$</td>
</tr>
<tr>
<td></td>
<td>$V^2f \sum_{i=s+1..n+s} c_i \leq P^B$</td>
</tr>
<tr>
<td></td>
<td>$L_i \leq x_i \leq U_i, \forall s+1 \leq i \leq n+s$</td>
</tr>
</tbody>
</table>

$\alpha_i$: unit size of component $i$

$\delta$: path; $\Delta$: path set; $A^B$: delay bound

$c_{ij}$: coupling capacitance between $i$ and $j$; $X^B$: crosstalk bound

$V$: supply voltage; $f$: working frequency; $P^B$: power bound

$L_i$: lower bound of $x_i$; $U_i$: upper bound of $x_i$

$|\Delta|$ may grow exponentially in the circuit size
Problem Formulation $PP$

Minimize \[ \sum_{i=s+1..n+s} \alpha_i x_i \]
Subject to
\[
\begin{align*}
a_j &\leq A_0, \quad j \in \text{input}(n+s+1) \\
ad + D_i &\leq a_i, \quad \forall \ s+1 \leq i \leq n+s, \quad j \in \text{input}(i) \\
D_i &\leq a_i, \quad \forall \ 1 \leq i \leq s,
\end{align*}
\]
\[
\sum_{i \in W} \sum_{j \in I(i)} c_{ij}(x_i + x_j) \leq X_0,
\]
\[
\sum_{i=s+1..n+s} c_i \leq P_0,
\]
\[
L_i \leq x_i \leq U_i, \quad \forall \ s+1 \leq i \leq n+s
\]

$a_i$: arrival time of $i$; $A_0 = A^B$;

$X_0$: transformed crosstalk bound

$P_0$: transformed power bound

All are positive coefficient polynomials
From LP to Lagrangian Relaxation

\[
\begin{align*}
\text{min} & \quad cx \\
\text{st} & \quad Ax \leq b \\
& \quad x \in X
\end{align*}
\]

\[
\begin{align*}
\text{min} & \quad L(\lambda) = cx + \lambda(Ax - b) \\
\text{st} & \quad x \in X
\end{align*}
\]

LP formulation  \quad Posynomial forms  \quad Lagrange multipliers \( \lambda \)
Lagrangian Relaxation

- Introduce Lagrange multipliers $\lambda$, $\beta$, $\gamma$ to relax constraints to the objective function $L$

\[
L_{\lambda,\beta,\gamma}(x,a) = \sum_{i=s+1}^{n+s} \alpha_i x_i + \sum_{i \in \text{input}(m)} \lambda_{jm}(a_j - A_0) + \sum_{i=s+1}^{n+s} \sum_{j \in \text{input}(i)} \lambda_{ji}(a_j + D_i - a_i) + \sum_{i=1}^{s} \lambda_0(i) (D_i - a_i) + \beta(\sum_{i=s+1}^{n+s} c_i - P_0) + \gamma(\sum_{i \in W} \sum_{j \in \text{input}(i)} c_{ij}(x_i + x_j) - X_0)
\]

- Lagrangian Relaxation Subproblem $LRS1$

Minimize $L_{\lambda,\beta,\gamma}(x,a)$

Subject to $L_i \leq x_i \leq U_i$, $\forall$ $s+1 \leq i \leq n+s$

Subject to only sizing constraints!!
Optimality Conditions

• Theorem:
  The optimality conditions on Lagrange multipliers:
  \[ \sum_{k \in \text{output}(i)} \lambda_{ik} = \sum_{j \in \text{input}(i)} \lambda_{ji}, \text{ for } 1 \leq i \leq n+s \]

  \[ \frac{\partial L_{\lambda, \beta, \gamma}(x^*, a^*)}{\partial a_i} = 0 \]

  \[ \sum \text{ multipliers on incoming edges} \]
  \[ = \sum \text{ multipliers on outgoing edges} \]

\[ 1 + 3 + 2 = 6 = 4 + 2 \]
## Lagrangian Relaxation Subproblem \textit{LRS2}

Minimize \[ L_{\mu,\beta,\gamma}(x) \]

Subject to \[ L_i \leq x_i \leq U_i, \quad \forall \ s+1 \leq i \leq n+s, \]

where \( \mu=(\mu_1, \ldots, \mu_m) \), \( \mu_i = \sum_{j \in \text{input}(i)} \lambda_{ji} \), for \( 1 \leq i \leq m \)

and \( L_{\mu,\beta,\gamma}(x) = \sum_{i=s+1\ldots n+s} \alpha_i x_i + \beta(\sum_{i=s+1\ldots n+s} c_i - P_0) \)
\[ + \gamma(\sum_{i\in W} \sum_{j \in I(i)} \hat{c}_{ij}(x_i+x_j) - X_0) \]
\[ + \sum_{i=1\ldots n+s} \mu_i D_i \]

Apply the optimality conditions and rewrite \( L_{\lambda,\beta,\gamma}(x,a) \).

\( L_{\mu,\beta,\gamma}(x) \) is independent of \( a \).
Optimal Resizing

• Theorem:
  Let \( x = (x_{s+1}, \ldots, x_{n+s}) \) be a solution.
  The optimal resizing of component \( i \):
  \[
  x_i^* = \min(U_i, \max(L_i, \text{opt}_i)),
  \]
  where \( \text{opt}_i = \sqrt{\mu_i r_i \left(C_j' + \sum_{j \in N(i)} \hat{c}_{ij} x_j\right)}
  \]
  \[
  \sqrt{\alpha_i + (\beta + R_i)\hat{c}_i + \gamma \sum_{j \in N(i)} \hat{c}_{ij}}
  \]
  \[
  \frac{\partial L_{\mu,\beta,\gamma}(x)}{\partial x_i} = 0.
  \]
  Update \( x_i \) using this theorem.
**Lagrangian Dual Problem LDP**

<table>
<thead>
<tr>
<th>Maximize</th>
<th>$D(\lambda, \beta, \gamma)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject to</td>
<td>$\lambda$ satisfies the optimality conditions, where $D(\lambda, \beta, \gamma) = \min L_{\lambda, \beta, \gamma}(x, a)$</td>
</tr>
</tbody>
</table>

LDP converges to the global optimal.
Optimal Gate and Wire Sizing

The objective function
minimize area

Lagrangian Dual Problem
maximize min $L_\lambda$

Lagrangian Relaxation Subproblem
minimize $L_\lambda$

> error bound

Adjust Lagrange multipliers

$\leq$ error bound

done

LR
Multiplier Adjustment

Step 1: \( \lambda_{ji}^{\text{new}} = \lambda_{ji}^{\text{old}} + \theta_k Z_i, \beta^{\text{new}} = \beta^{\text{old}} + \theta_k (\sum_{i=s+1}^{n+s} c_i - P_0), \)

\( \gamma^{\text{new}} = r^{\text{old}} + \theta_k (\sum_{i\in W} \sum_{j\in I(i)} \hat{\gamma}_{ij} (x_i + x_j) - X_0) \)

where \( j \in \text{input}(i), \lim_{k \to \infty} \theta_k \to 0, \sum_{k=1}^{\infty} \theta_k \to \infty, \)

\( Z_i = a_j - A_0 \text{ if } i \in T \)

\( = a_j + D_i - a_i \text{ if } i \in G \cup W \)

\( = D_i - a_i \text{ if } i \in R \)

Step 2: Project \( \lambda \) to nearest feasible solution
Antenna Effect

- During the chip manufacturing process, metal is initially deposited to cover the whole chip.
- Then the unneeded portion of metal is removed by etching, typically in plasma (charged particles).
- The exposed metal collects charge from plasma and form voltage potential.
- If the voltage potential across the gate oxide becomes large enough, the current can damage the gate oxide.
Antenna Effect

- Unconnected wires act as “antennas” that pick up electrical charge.
- The longer the wires, the more the charge.
Fixing Antenna Effect Using Diodes

- Insert a diode cell next to each input.
  - Costs significant area
  - Adds capacitance at worst spot.
Fixing Antenna Effect through Jumpers

- Adding ‘jumpers’ after routing
  - The idea: Force a routing pattern that “shoots up” to the highest layer as soon as possible.
Routing Jumpers

- Adding “jumpers” after routing

Build a ‘castle’, that forces the routing engine to go up the highest layer.
Metal Fill

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

![Diagram of metal fill and parasitics](image-url)
Optical Proximity Correction

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

- A statistical effect, resulting in a gradual increase of the wire resistance, followed by failure.
- The effect depends on
  - The current density -> wide wires
  - Exponentially on temperature.