

## CSCE 613 – Week 6 Fall 2005

### CMOS VLSI Design

# Introduction to CMOS Wire

Adapted/extended by James P. Davis, Ph.D.  
Dept. of Computer Science & Engineering  
University of South Carolina

Jan M. Rabaey  
Anantha Chandrakasan  
Borivoje Nikolic

---

---

---

---

---

---

---

---

---

---

## Topics - Week 6

- ❑ Interconnect between MOS transistor structures – the “wire”, in terms of behavior, function and structure (materials from Rabaey et al.)
  - Coverage from Rabaey et al., Chapter 4, §4.1 - §4.3.2, pp. 136 - 148, 4.3.3, 4.4.1 - 4.4.5, pp. 148 - 169.
  - Examples from Ch4 text – the progression of modeling results for the .25 micron CMOS process.
- ❑ Capacitance revisited, modeling and assumptions for the wire (this is a review, reformulation and extension of notes presented earlier from Weste et al.)
- ❑ Resistance revisited, modeling and assumptions for the wire (this is a review, reformulation and extension of notes presented earlier from Weste et al.). We also consider effects of switching frequency.
- ❑ Inductance on the wire, which we consider at large wire sizes with low resistance, and high switching frequencies.

---

---

---

---

---

---

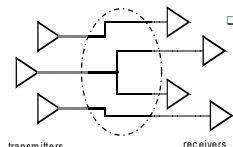
---

---

---

---

## Introduction to The Wire



schematics



physical

- ❑ Scaling of interconnect versus devices
  - ❑ As we scale CMOS devices and their interconnect, the effects of *parasitics* on the wires becomes more pronounced, particularly in scaled circuits running at higher speeds.
  - ❑ Wires scaling effects dominate in areas of speed, power consumption, reliability (noise immunity).
  - ❑ Larger die sizes exacerbate the problem, as line lengths get longer.
- ❑ Effect of “parasitics” on CMOS circuits
  - ❑ Increase propagation delays, and decrease transistor performance.
  - ❑ Affect on energy dissipation & power distribution throughout the circuit.
  - ❑ Create additional sources of noise.
- ❑ Basic wire modeling abstraction
  - ❑ We only use a few dominant parameters.
  - ❑ Each wire in bus network connects “transmitter(s)” to “receiver(s)” via a chain of wire segments w/ length and geometry.

---

---

---

---

---

---

---

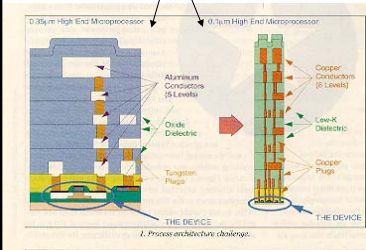
---

---

---

## Interconnect Impact on Chip

Comparing 0.35μ versus 0.1μ CMOS processes, noting their differences in geometry, specifically pertaining to number of layers of interconnect.



### □ Geometry Layering

- As we scale processes, we also add more process layers to the circuit geometry.
- We still have single diffusion and poly layers.
- We have many more layers of metal, sandwiched between layers of insulator.
- It is this more aggressive geometry structuring, coupled with smaller wire feature size versus length, that makes wire modeling a challenge.
- New materials are part of the solution.

---

---

---

---

---

---

---

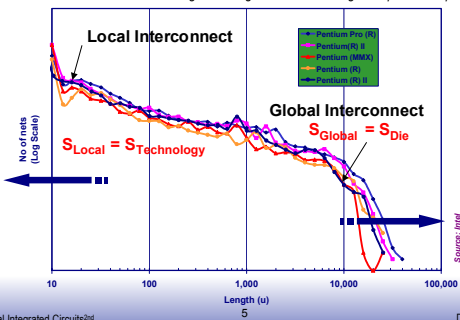
---

---

---

## Nature of Interconnect – Real Data

Here, we are comparing the relative number of nets required for local versus global interconnect for Pentium® class of CMOS devices (with ~30-40M transistors) as a function of actual line length. Technology scaling is facilitating smaller wire size, while increases in die size facilitate longer line lengths—the source of greater parasitic impact.




---

---

---

---

---

---

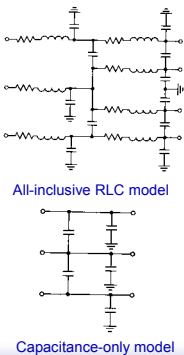
---

---

---

---

## Wire Modeling-1



### □ Parasitic wire models

- The lumped parameter modeling of resistive, capacitive and inductive effects on interconnect are distributed over entire wire length.
- This is necessary when wire length becomes much greater than its line width,  $L_{wire} \gg W_{wire}$ .
- We have effects due to the relationship of the wire to ground, and also between wires at different layers in the circuit geometry.
- We end up with "coupling" effects in the layout that must be managed, that are not present in schematic model.
- These are complex models that, fortunately, we can simplify under certain conditions (see next slide).

---

---

---

---

---

---

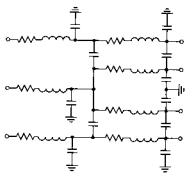
---

---

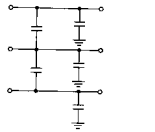
---

---

## Wire Modeling-2



All-inclusive RLC model



Capacitance-only model

- Parasitic wire model simplification
  - Induction can be ignored if the wire resistance is large:  $R_{wire} \gg L_{wire}$  for longer lines.
  - A capacitance only model can be used IF (1) wires are relatively short, (2) wire cross section ( $W_{wire} \times H_{wire}$ ) is large, OR (3) we use a low-resistivity interconnect material.
  - A simplified capacitance model consisting only of wire-to-ground capacitance (ignoring inter-wire effects) can be used IF: (1) inter-wire separation is large, OR (2) wires run together in material for short distance.
  - The specific values for when we use these model assumptions are to be derived by authors in the text.

---

---

---

---

---

---

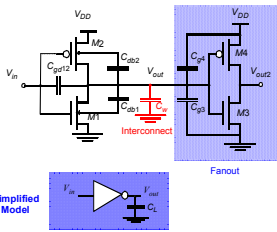
---

---

---

---

## Capacitance of Wire Interconnect



Simplified Model

- Capacitance wire model simplification (redux)
  - In the figure, we see how we incorporate the capacitance wire assumption into the schematic view.
  - We have a "transmitter" (an inverter, with its device capacitances identified) connected to an output line that drives a fanout "receiver" load consisting of the transistors of another inverter.
  - We can model this as the inverter transmitter (or driver) coupled to a load consisting of a capacitance,  $C_L$ , with voltages  $V_{in}$  and  $V_{out}$ .
  - There is a relationship between  $C_L$ ,  $C_{wire}$  and the load values on the fanout that is to be developed.

---

---

---

---

---

---

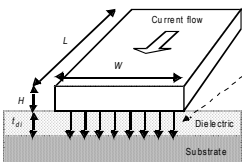
---

---

---

---

## Capacitance - the Parallel Plate Model



$$C_{int} = \frac{\epsilon_{di}}{t_{di}} WL$$

$$S_{C_{wire}} = \frac{S}{S \cdot S_L} = \frac{1}{S_L}$$

- Rectangular wire geometry
  - We can adopt a "parallel-plate" capacitor model IF wire width is much greater than dielectric thickness:  $W_{wire} \gg t_{di}$
  - To minimize the effect of resistance as we scale the wire, we must keep cross section ( $W_{wire} \times H_{wire}$ ) large.
  - Smaller line widths,  $W_{wire}$ , allow more dense packing of circuits onto geometry (at cross-purposes).
  - The ( $W_{wire} / H_{wire}$ ) ratio: as it drops below unity, we must incorporate the "fringing" component of the capacitance model along with the "parallel-plate" component to obtain parameter  $C_{wire}$ .
  - Capacitance is proportional to overlap between conductors, inversely proportional to their separation.

---

---

---

---

---

---

---

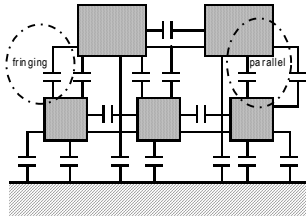
---

---

---



## The Inter-wire Capacitance Model



- The Wire "micro-stripeline" versus the wire interconnect hierarchy:
  - With only 1-2 layers of metal interconnect on top of the poly and diffusion, we can assume the additive capacitance model for parallel plate and fringe relative to ground.
  - Actual CMOS processes have many more layers, densely packed.
  - Each wire is coupled to ground substrate, and to neighboring wires on same and adjacent layers.
  - These "floating" capacitances have parallel plate and fringe aspects relative to time varying voltage levels of these other wire signal paths—a source of "crosstalk" noise.

---

---

---

---

---

---

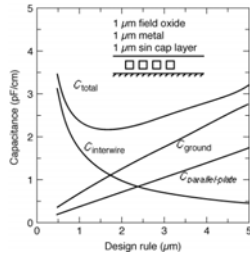
---

---

---

---

## Impact of Inter-wire Capacitance



(from [Bakoglu89])

- The bottom line:
  - Inter-wire capacitances become dominant in multilayer interconnect structures.
  - Effect is more pronounced at higher layers (further away from the substrate).
  - Contribution of  $C_{interwire}$ : for parallel wires routed above a ground plane, with constant dielectric ( $\epsilon_d$ ) and wire thickness ( $H_{wire}$ ), while other dimensions ( $W_{wire}$ ,  $T_{wire}$ ) scaled.
  - As  $(W_{wire} / H_{wire}) < 1.75$ , inter-wire effect dominates (as seen in the plot of Capacitance vs Line width).
  - The figure in the plot "legend" shows the configuration assumption for this conclusion.
  - Compare this result with slide #12 for  $C_{fringe}$  vs  $C_{pp}$ .

---

---

---

---

---

---

---

---

---

---

## Wiring Capacitances (0.25 μm) – for examples

Table rows are capacitor's "top plate".  
Table columns are its "bottom plate".

Use the Field values for C terms when placing wires over thick field oxide (SiO<sub>2</sub>) that isolates different transistors.

$C_{pp}$  values in upper row.  
 $C_{fringe}$  values in lower, shaded row.

Process supports 1 layer Poly and 5 layers Metal.

First 4 metal layers Use same  $W$  and  $\epsilon_d$ .  
But 5<sup>th</sup> layer has  $H_{m5} = 2H_m$  and  $\epsilon_{d5} > \epsilon_d$ .

	Field	Active	Poly	M1	M2	M3	M4
Poly	88						
M1	54						
M2	30	41	57				
M3	40	17	54				
M4	13	15	17	36			
M5	25	27	29	45			
M6	8.9	9.4	10	15	41		
M7	18	19	20	27	49		
M8	6.5	6.8	7	8.9	15	35	
M9	14	15	15	18	27	45	
M10	5.2	5.4	5.4	6.6	9.1	14	38
M11	12	12	12	14	19	27	52

---

---

---

---

---

---

---

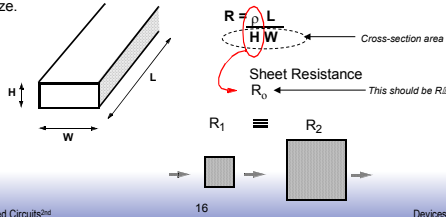
---

---

---

## Wire Resistance Model

- The bottom line:
  - Resistance of a wire  $\propto$  wire length, and  $1/\propto$  to cross section area ( $H \times W$ ).
  - Assume same rectangular geometry as for capacitance model.
  - The "resistivity",  $\rho$ , of the wire material (expressed in Ohm-m) is a constant.
  - The wire thickness,  $H$ , is also a constant (for a given technology), and can be expressed in terms of "sheet resistance" (expressed in "Ohms-per-square").
  - Result is that  $R$  of a square wire segment is independent of its absolute size.




---

---

---

---

---

---

---

---

---

---

## Interconnect Resistance

Material	$\rho$ (Ω-m)
Silver (Ag)	$1.6 \times 10^{-8}$
Copper (Cu)	$1.7 \times 10^{-8}$
Gold (Au)	$2.2 \times 10^{-8}$
Aluminum (Al)	$2.7 \times 10^{-8}$
Tungsten (W)	$5.5 \times 10^{-8}$

- Resistivity  $\rho$ :
  - Aluminum most often used for Metal interconnect material (low cost and process compatibility).
  - Copper has better resistivity constant, however, and is finding its way into more prominent use in newer processes.

Material	Sheet Resistance (Ω/□)
n- or p-well diffusion	1000 - 1500
$n^+$ , $p^+$ diffusion	50 - 150
$n^+$ , $p^+$ diffusion with silicide	3 - 5
$n^+$ , $p^+$ polysilicon	150 - 200
$n^+$ , $p^+$ polysilicon with silicide	4 - 5
Aluminum	0.05 - 0.1

- Sheet resistance  $R_s$ :
  - For the 0.25 micron process, we have avg. values for interconnect materials.
  - This shows use of Diffusion, Poly and Metal lines as wires.
  - Note use of a "Silicide" compound enables processing of Diff and Poly layers at high temp. to increase region's conductivity.

Use of Polycide increases conductivity by reducing sheet resistance.

---

---

---

---

---

---

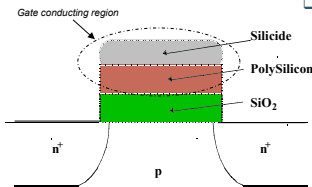
---

---

---

---

## Interconnect - the Polycide Gate



- Design Objective:
  - Want to be able to use Poly and Diffusion materials as better conductors of current.
  - Reduce resistance of conducting materials, Poly and Diffusion layers.
  - Add a silicon compound layer on the gate Poly, source and drain Diffusion regions.
  - Increases conductivity by reducing  $R_s$ .

Silicides:  $WSi_2$ ,  $TiSi_2$ ,  $PtSi_2$  and  $TaSi$

Conductivity: 8-10 times better than Poly alone.

---

---

---

---

---

---

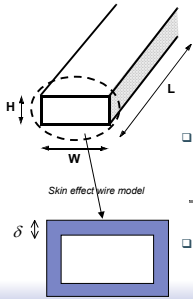
---

---

---

---

## Resistance – Skin Effect



- **Resistance at very-high switching frequencies, in wider wires**
  - R is no longer linear, but is frequency dependent (GHz range, e.g., Pentium® clock frequency)
  - Current flows nearer to surface of conducting wire, drops off beyond depth  $\delta$ . Equ. 4.6
  - Skin depth  $\delta$ : depth at which current falls off to a value of  $e^{-1}$  of its nominal value.
    - Al at 1 GHz frequency,  $\delta = 2.6$  microns.
- **Skin effect approximation**
  - Assume uniform current flow in "outer shell" of rectangular wire with thickness of  $\delta$ . Wire cross-section ( $H \times W$ )  $\approx 2(W \times H) \times \delta$
  - Calculating  $r$  (i.e., R per unit length): Equ. 4.7
- **Implications**
  - Find the frequency  $f_s$  where  $\delta = \frac{1}{2} \max(W, H)$ . If  $f > f_s$ , then we'll have increased resistance. Eq 4.8
  - Clocking: wide wires & high freq  $\Rightarrow$  "skin"  $r$ .

---

---

---

---

---

---

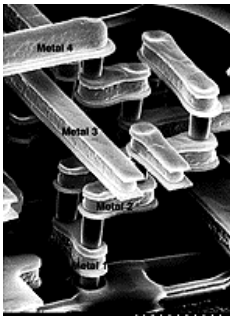
---

---

---

---

## Minimizing Resistance



- **Employ selective Technology Scaling**
  - Scale L and W, leave H and  $t_{\text{eff}}$  alone (as with capacitance)?
- **Use Better Interconnect Materials**
  - Reduces average wire-length
  - e.g. Copper, Silicides
- **More Interconnect Layers**
  - Reduces average wire-length.
  - But add more wiring layers, which increases capacitance effects (design tradeoff here).
  - However, to minimize *contact resistance*, we want to keep signal wires on a single layer, and avoid excessive contacts and vias.

---

---

---

---

---

---

---

---

---

---

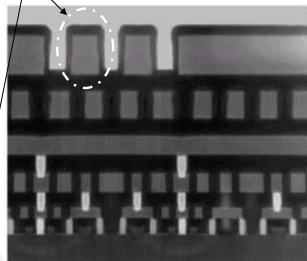
## Example: Intel 0.25 micron Process

As we get to higher layers of metalization, we see W and H get much larger. This starts to change the model for resistance, and introduces inductance, in designs operating at higher switching frequencies (GHz range).

5 metal layers  
Ti/Al - Cu/Ti/TiN  
Polysilicon dielectric

LAYER	PITCH	THICK	A.R.
Isolation	0.67	0.40	-
Polysilicon	0.64	0.25	-
Metal 1	0.64	0.48	1.5
Metal 2	0.93	0.90	1.9
Metal 3	0.93	0.90	1.9
Metal 4	1.60	1.33	1.7
Metal 5	2.56	1.90	1.5
	$\mu\text{m}$	$\mu\text{m}$	

Layer pitch, thickness and aspect ratio




---

---

---

---

---

---

---

---

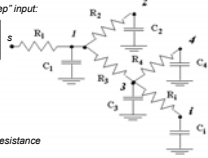
---

---



## The Lumped RC Model

Applying a "step" input:



Shared path resistance

$$R_{jk} = \sum R_j \Rightarrow (R_j \in [\text{path}(s \rightarrow i) \cap \text{path}(s \rightarrow k)])$$

$$\tau_{Di} = \sum_{k=1}^N C_k R_{ik}$$

### Model axioms

- Lump total wire resistance of each wire segment into single R between nodes in network.
- Lump total capacitance into single node capacitor to GND.
- Single differential equ., single RC time constant for solution.

### Model topology

- RC tree: (1) single input node, (2) C<sub>i</sub> between node i and GND, (3) no resistive loops.
- Unique resistive path from source node s to any node i.

### Elmore delay calculation

- Path resistance, R<sub>ij</sub>.
- Shared path resistance: s to k, i.
- Elmore delay at node i: Equ 4.13

---

---

---

---

---

---

---

---

---

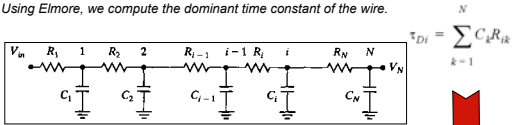
---

## Lumped RC Model - Elmore Delay

Consider the simple, non-branched RC ladder network: a good approximation of an R-C wire. Using 4.13, we reach a simplification where the shared-path resistance is replaced by the path resistance. This is summed with capacitance at each node over the wire segments.

A wire of length L is partitioned into N identical segments, each with length L/N. Capacitance = cL/N, Resistance = rL/N (these are per-unit-length for each segment).

Using Elmore, we compute the dominant time constant of the wire.



$$\tau_{Di} = \sum_{k=1}^N C_k R_{ik}$$

This reformulation of Equ. 4.13 into 4.14 (p. 154) allows the RC tree to be represented (approximately) by the RC ladder network.

Time constant  $\tau_i$  can be derived for any node i in the network:  
 $\tau_{Di} = C_i R_i + C_j (R_i + R_j) + \dots + C_N (R_i + R_j + \dots + R_N)$

$$\tau_{DN} = \sum_{i=1}^N C_i \sum_{j=1}^i R_j = \sum_{i=1}^N C_i R_{ii}$$

---

---

---

---

---

---

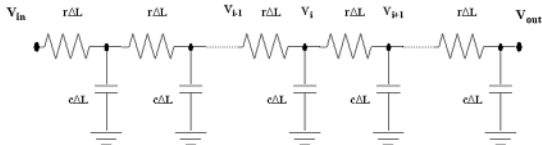
---

---

---

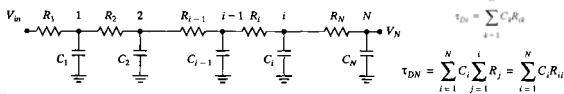
---

## The Elmore Delay Model - RC Chain



Consider the simple, non-branched RC ladder network: a good approximation of an R-C wire. Using 4.13, we reach a simplification where the shared-path resistance is replaced by the path resistance.

A wire of length L is partitioned into N identical segments, each with length L/N. Capacitance = cL/N, Resistance = rL/N (these are per-unit-length for each segment). Using Elmore, we compute the dominant time constant of the wire.



$$\tau_{DN} = \sum_{i=1}^N C_i R_{ii}$$

$$\tau_{DN} = \sum_{i=1}^N C_i \sum_{j=1}^i R_j = \sum_{i=1}^N C_i R_{ii}$$

---

---

---

---

---

---

---

---

---

---

## Lumped RC Model – Dominant Time Constant

Assume: Wire modeled by N equal-length segments

$$\tau_{DN} = \left(\frac{L}{N}\right)^2 (rc + 2rc + \dots + Nrc) = (rcL^2) \frac{N(N+1)}{2N^2} = RC \frac{N+1}{2N}$$

We can compute the dominant time constant  $\tau = RC$  of the wire by summing over N segments, extracting the Length components L/N for both r and c (per-unit-length) multiplied by the series expansion. Reformulating terms, and inserting  $R = rL$  and  $C = cL$  as total lumped R and C of the wire.

For large values of N – the (N+1)/2N goes to 1/2:

$$\tau_{DN} = \frac{RC}{2} = \frac{rcL^2}{2}$$

As N grows large, the lumped RC model approaches that of the distributed rc line model, with conclusions:

- (1) Delay of a wire is quadratic function of wire length L. (So, doubling wire length quadruples line delay.)
- (2) Delay predicted using lumped RC model is approximately half as accurate (in terms of delay calculation) as the distributed rc line model.
- (3) The Elmore delay calculation only considers the dominant time constant, ignoring the others that would be associated with other RC terms, so the delay associated with the dominant RC is crude, depending on wire length.

---

---

---

---

---

---

---

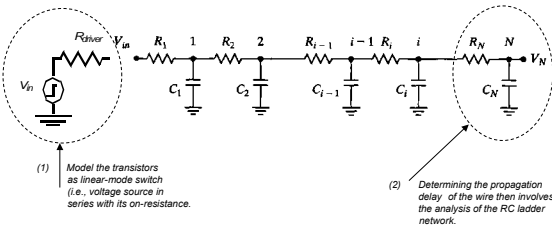
---

---

---

## Lumped RC Model – Plugging It In

We can now relate the wire model to the transistor model as follows:




---

---

---

---

---

---

---

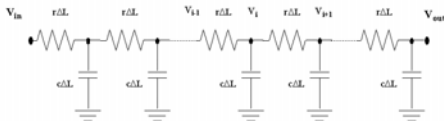
---

---

---

## Distributed rc Model

Lumped RC model is "pessimistic" in its assumption, so we increase the accuracy by distributing the resistance and capacitance throughout wire segments (r and c defined per-unit-length).



A wire of length L is partitioned into N identical segments, each with length L/N. Capacitance = cL/N, Resistance = rL/N (these are per-unit-length for each segment).

Using Elmore, we compute the dominant time constant of the wire.

The voltage at a given node can be solved using partial differential equations (Equ. 4.17 -> Equ. 4.18).

---

---

---

---

---

---

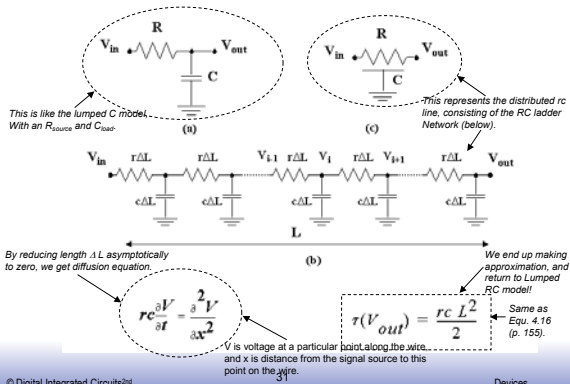
---

---

---

---

## The Distributed RC-line




---

---

---

---

---

---

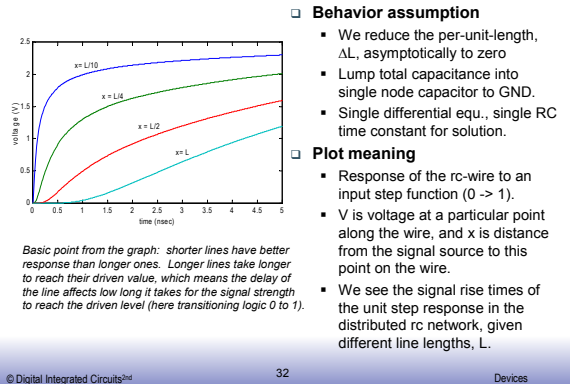
---

---

---

---

## Step-response of RC Wire Model




---

---

---

---

---

---

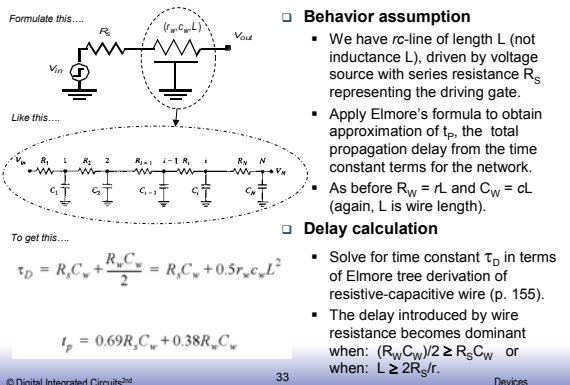
---

---

---

---

## Reduction in Distributed rc Model




---

---

---

---

---

---

---

---

---

---

