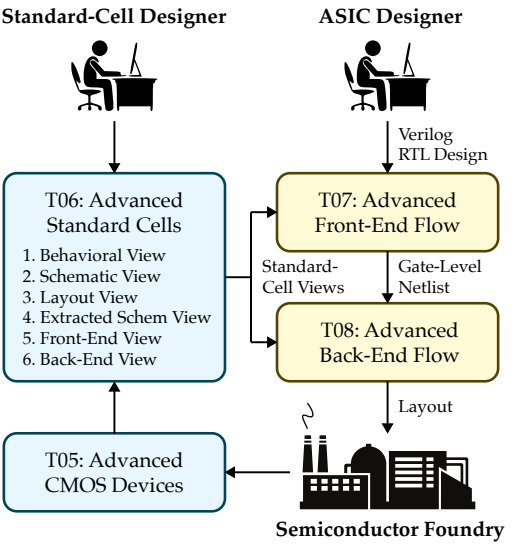
ECE 6745 Complex Digital ASIC Design
Topic 5: Advanced CMOS Devices

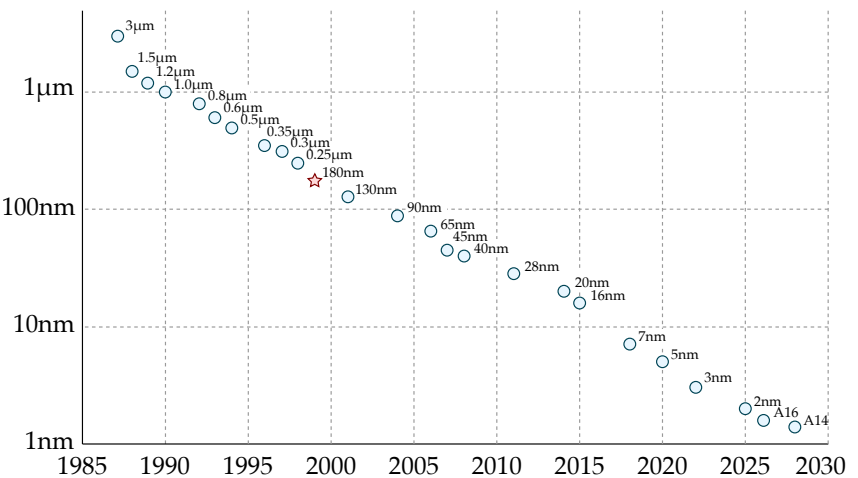
School of Electrical and Computer Engineering
Cornell University

revision: 2026-02-24-12-53



1. Transistors

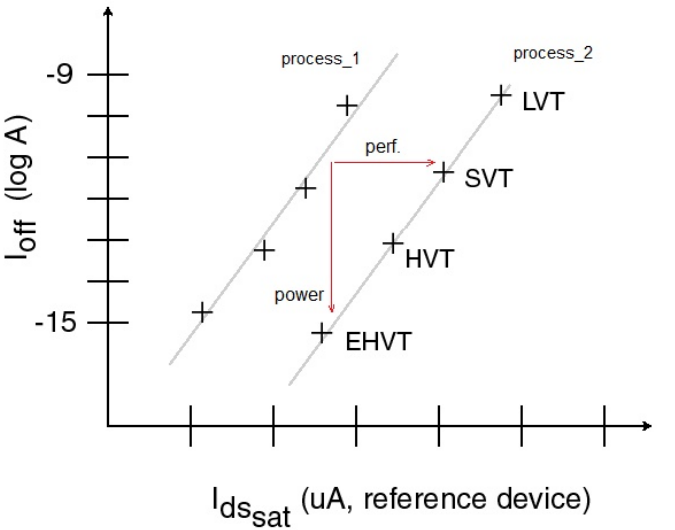
- Technology scaling has slowed, and technology node names after 90 nm no longer correspond to physical transistor dimensions
- Effective transistor length: ≈3 µm in 1980s → ≈10–15 nm in 2020s
- 10,000× increase in transistor density per chip over four decades
- Each technology node still offers significant improvements in performance, power, and/or cost per transistor, but at ever greater fixed foundry costs and non-recurring engineering tapeout costs



TSMC Technology Nodes

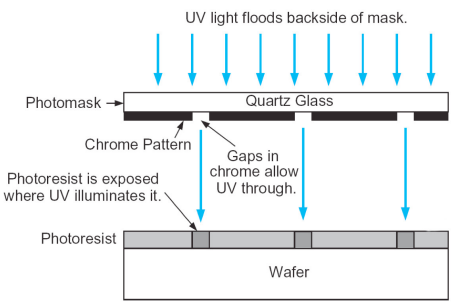
1.1. 130nm: Multiple Threshold Voltages

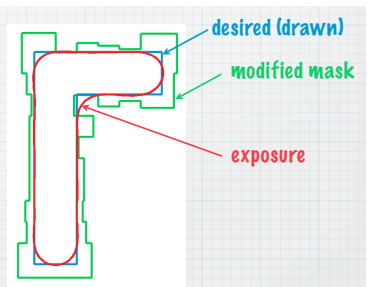
- Advanced technology nodes provide many transistors with *multiple threshold voltages* to trade-off leakage vs drive current

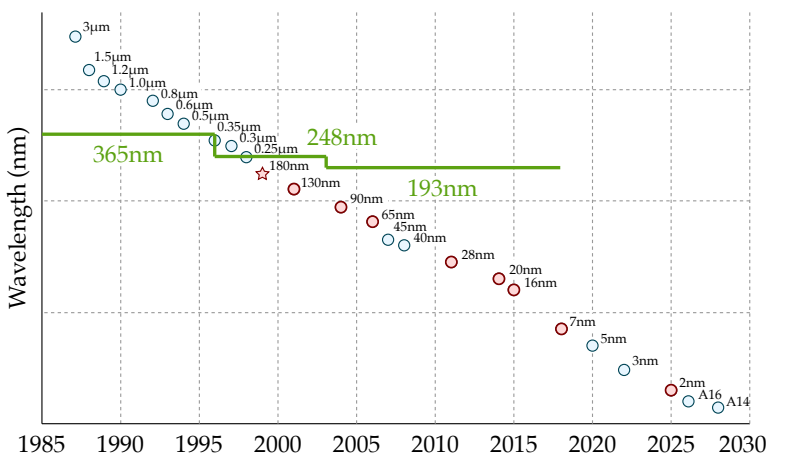


1.2. 90nm: Optical Proximity Correction

- If mask features are smaller than the wavelength of light used to expose the mask, then the resulting exposure will be blurry
- *Optical proximity correction* (OPC) models light diffraction and interference at sub-wavelength scales to pattern smaller features

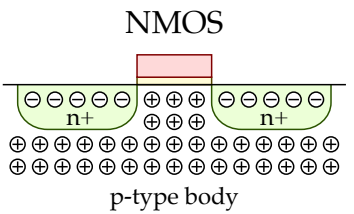


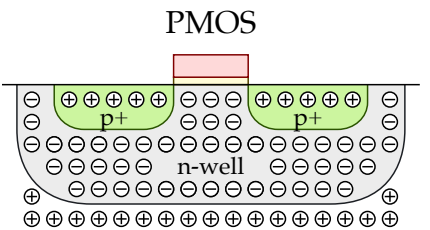


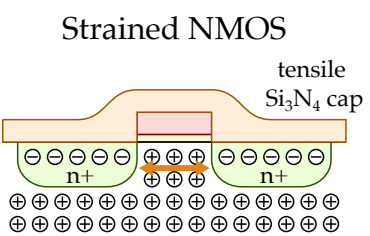


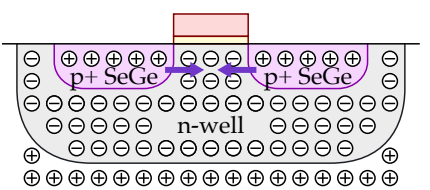
1.3. 65nm: Strained Silicon

- Transistors can be fabricated with *strained silicon* to increase carrier mobility
  - NMOS uses *tensile strain* (stretch channel) → increases electron mobility
  - PMOS uses *compressive strain* (squeeze channel) → increases hole mobility



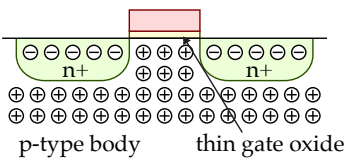


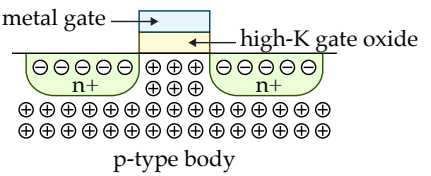




1.4. 28nm: High-K Dielectric and Metal Gate

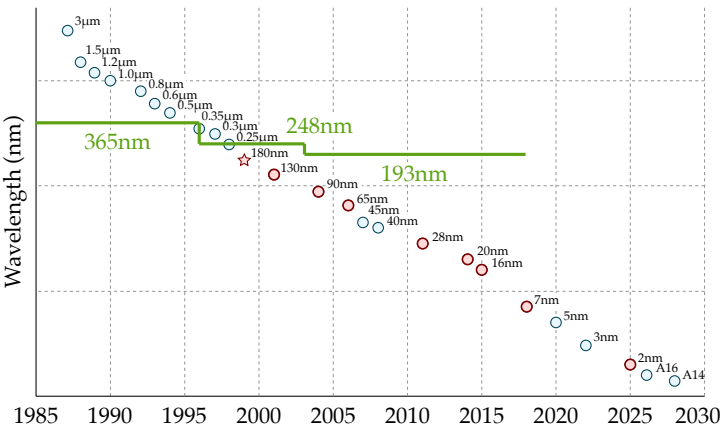
- As gate oxides became extremely thin (< 2 nm), leakage current increased dramatically
- Replacing SiO_2 with a *high-k dielectric* reduces leakage
  - High-k material has larger dielectric constant (k)
  - Allows physically thicker gate oxide
  - Maintains same gate capacitance
- Need to replace polysilicon gate with *metal gate* to interface with high-k dielectrics

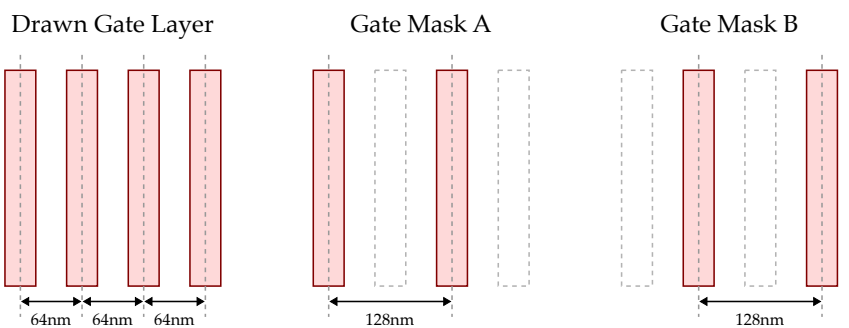




1.5. 20nm: Double Patterning

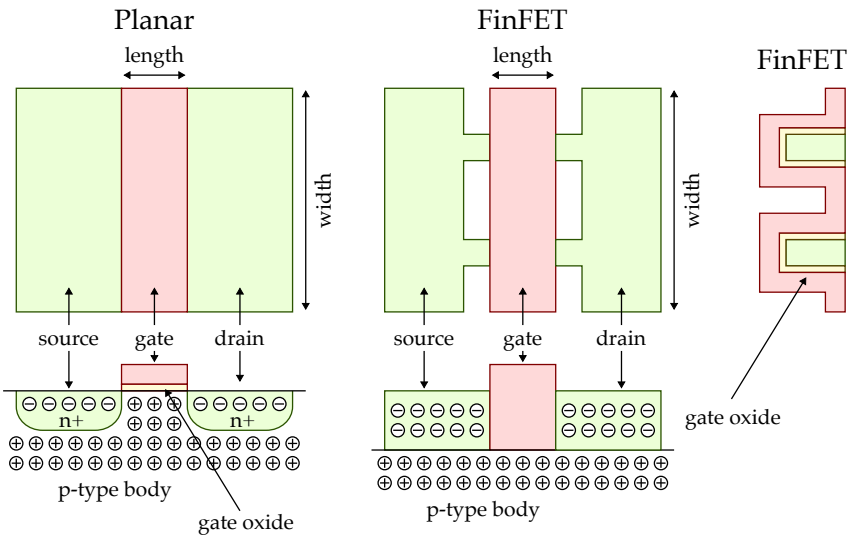
- Feature sizes became smaller than what a single lithography exposure could reliably print
- *Double patterning* splits one layer into two separate masks
  - First mask prints half the features
  - Second mask prints the remaining features

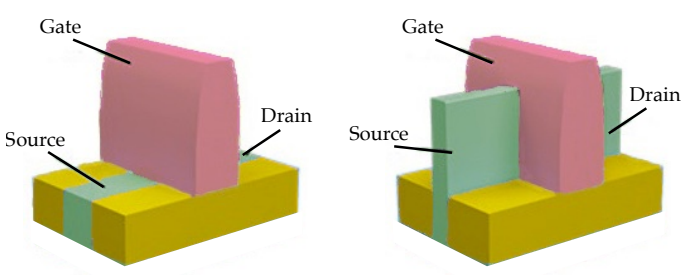




1.6. 16nm: FinFETs

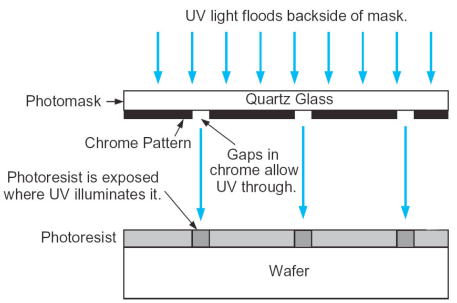
- Planar transistor gate controls channel from the top
- *FinFETs* wrap the gate around the channel
- More elctrostatic control → better switching behavior

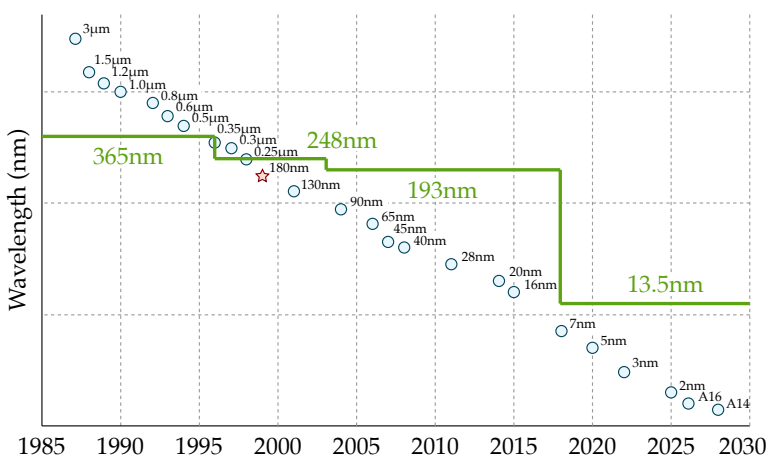




1.7. 7nm: Extreme Ultra-Violet Lithography

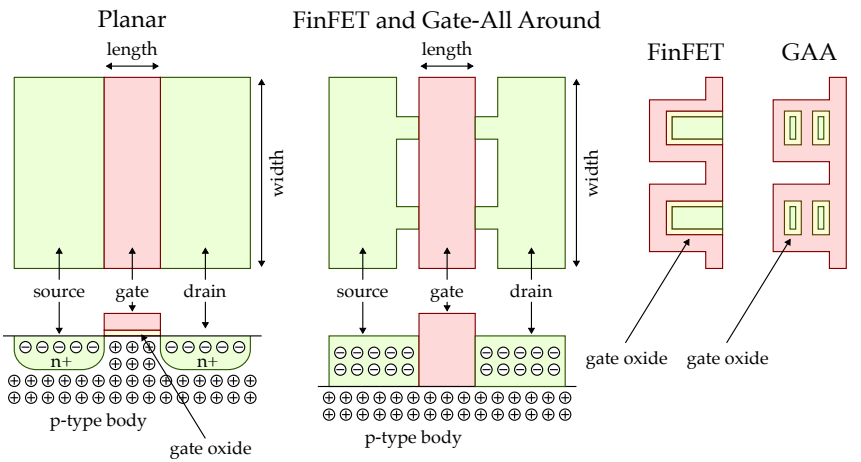
- *Extreme ultra-violet lithography* (EUV) can print dense features
  - EUV light absorbed by air and glass, requires vacuum system with reflective mirrors (no lenses)
  - Light source is extremely complex, power-hungry, initially quite weak
  - Tool cost is very high

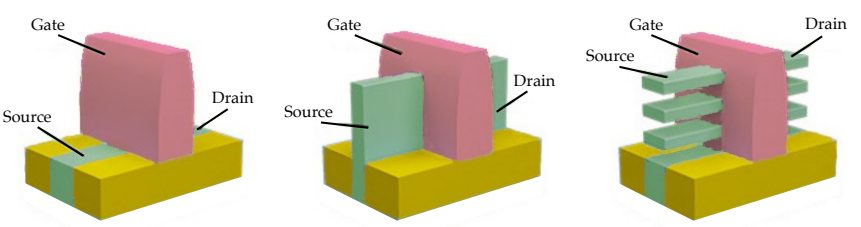




1.8. 2nm: Gate-All-Around Nanosheets

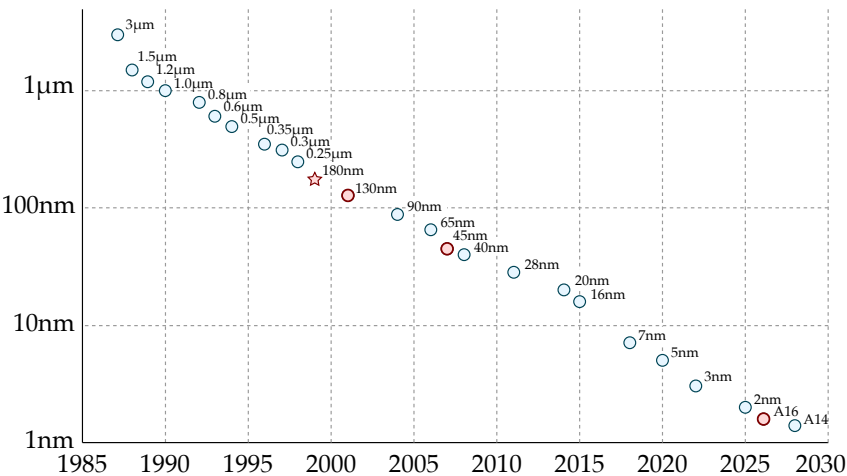
- As channel length continued shrinking, even FinFETs insufficient
- *Gate-all-around* (GAA) surrounds the channel completely
  - Gate controls channel from all sides
  - Maximizes electrostatic control





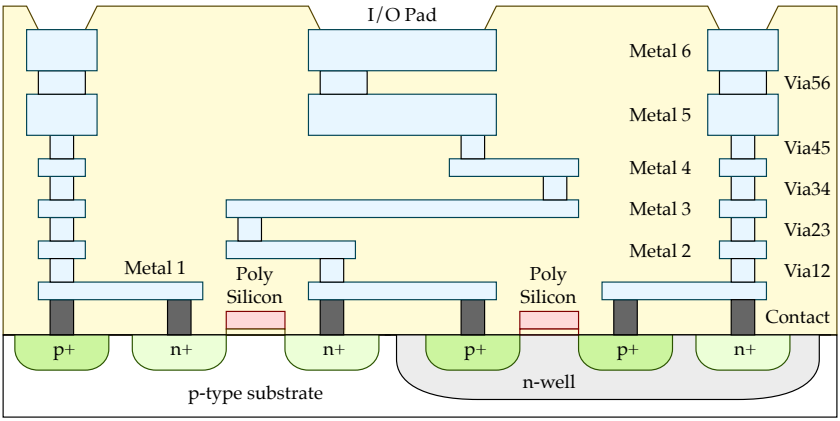
2. Interconnect

- Interconnect gradually become as important or even more important than transistors in determining overall chip performance and power
- Must consider both the resistance, capacitance, and possibly even inductance of interconnect
- Metal 1 pitch is now one of the key challenges in technology scaling



TSMC Technology Nodes

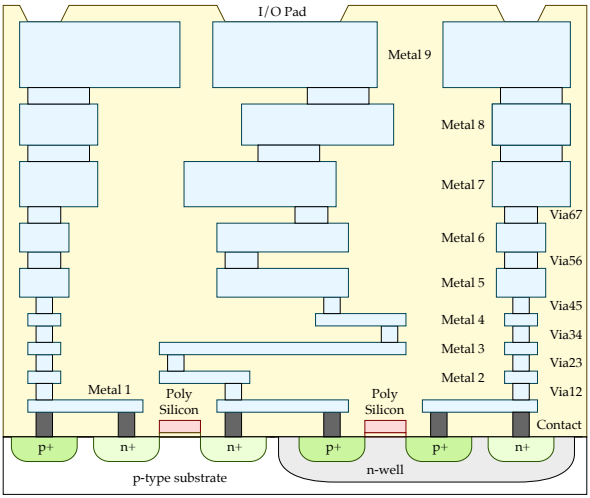
2.1. 130nm: Low-K Dielectric and Copper



Dielectric	Relative Permittivity (ϵ_r)
Silicon Dioxide (SiO_2)	3.9
Silicon Flouride (SiOF)	3.1
SiLK™ Polymer	2.6

Metal	Resistivity ($\mu\Omega$·cm)
Silver (Ag)	1.6
Copper (Cu)	1.7
Gold (Au)	2.2
Aluminum (Al)	2.8
Tungsten (W)	5.3
Molybdenum (Mo)	5.3
Titanum (Ti)	43.0

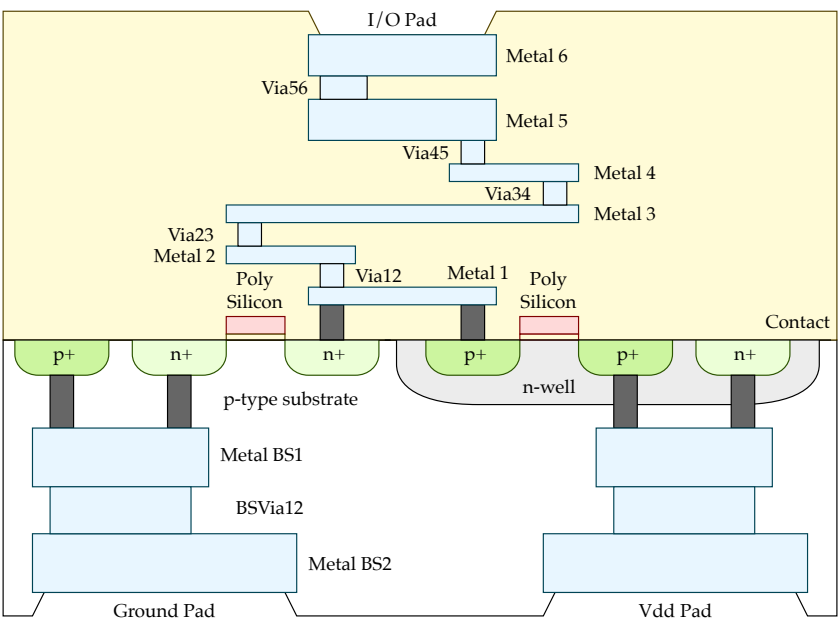
2.2. 45nm: Tall Metal Stacks



Layer	T (nm)	W (nm)	S (nm)	P (nm)	Relative W
M9	7µm	17.5µm	13µm	30.5µm	218×
M8	720	400	410	810	5×
M7	504	280	280	560	3.5×
M6	324	180	180	360	2.25×
M5	252	140	140	280	1.75×
M4	216	120	120	240	1.5×
M3	144	80	80	160	1×
M2	144	80	80	160	1×
M1	144	80	80	160	1×

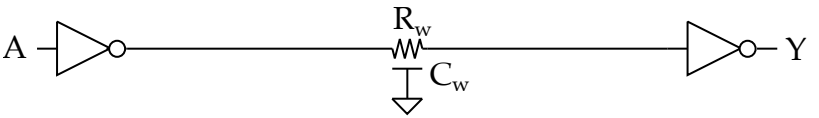
2.3. A16: Back-Side Power Delivery

- Traditional chips route signal and power on the same front-side metal layers
  - Power grid competes with signal routing
  - Increases congestion and IR drop
- *Back-side power delivery* moves power routing to the back of the wafer
  - Signals remain on the front side
  - Power delivered through vertical vias
- Separates power and signal networks

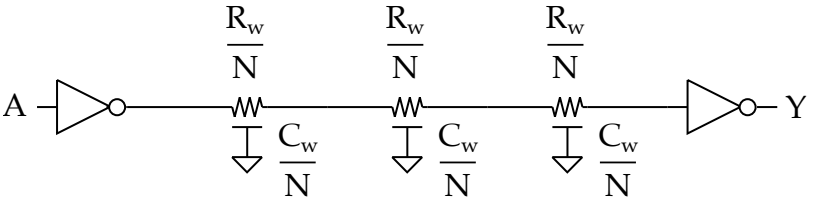


2.4. RC Model

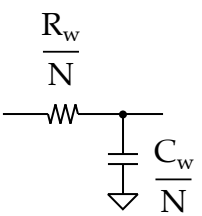
- Resistance increases with length
- Capacitance increases with length



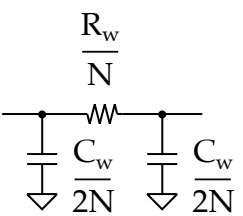
- Think of wire divided into many small segments each with an equal portion of the total wire resistance and capacitance
- The more segments the more accurate the RC model



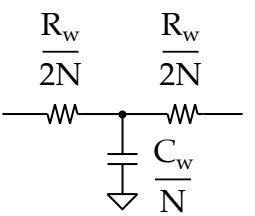
- Various RC models are possible for each segment



L-Model

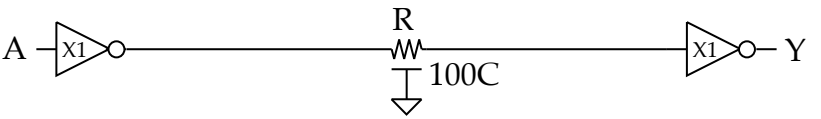


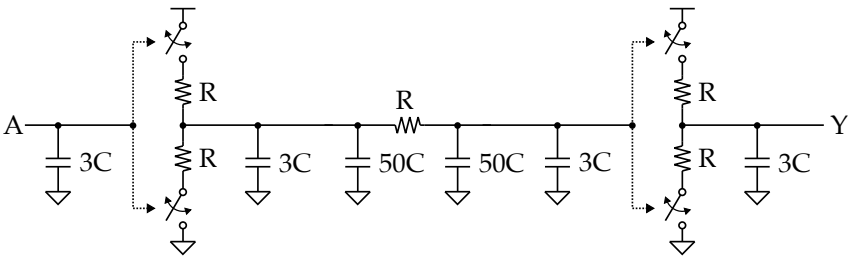
π-Model

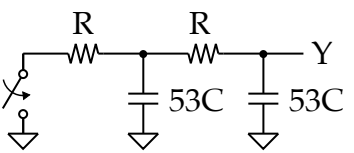


T-Model

- Consider an INVX1 standard cell driving a long wire and INVX1



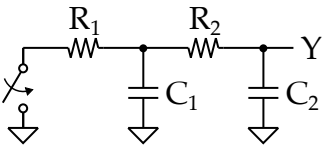




How to we estimate the propagation delay
of such an RC model?

2.5. Elmore's Delay

- Consider the following general RC circuit



- Requires complicated second-order model

$$V_Y(t) = V_{DD}\frac{\tau_1 e^{-t/\tau_1} - \tau_2 e^{-t/\tau_2}}{\tau_1 - \tau_2}$$

$$R^* = \frac{R_2}{R_1} \qquad C^* = \frac{C_2}{C_1}$$

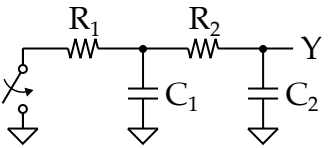
$$\tau_{1,2} = \frac{R_1C_1 + (R_1 + R_2)C_2}{2}\left(1 \pm \sqrt{1 - \frac{4R^*C^*}{[1 + (1 + R^*)C^*]^2}}\right)$$

- We do not want to determine the exact waveform for V_Y
- So we can use Elmore's delay to estimate the propagation delay
- Approximate this second-order model with a first-order model that preserves the propagation delay but not the full waveform shape

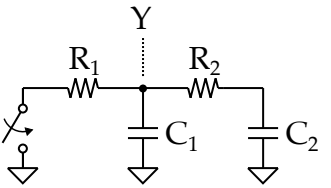
$$V_Y = V_{DD}\, e^{-t/\tau}$$

$$\tau = R_1C_1 + (R_1 + R_2)C_2$$

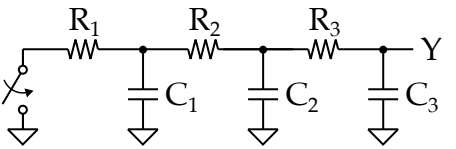
- Each capacitor contributes delay proportional to the resistance between it and the input
- Only useful for estimating propagation delay
- Best when one τ is much larger than the other τ
- Even if $\tau_1 = \tau_2$, error is $< 15\%$



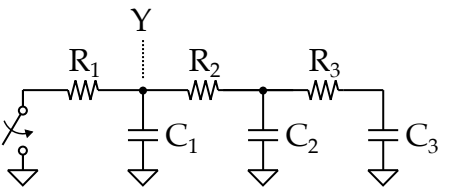
$$t_{pd} = R_1C_1 + (R_1 + R_2)C_2$$



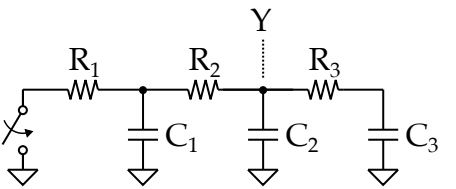
$$t_{pd} = R_1C_1 + R_1C_2$$



$$t_{pd} = R_1C_1 + (R_1 + R_2)C_2 + (R_1 + R_2 + R_3)C_3$$

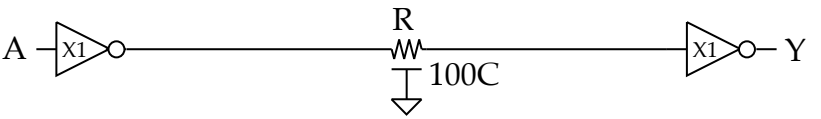


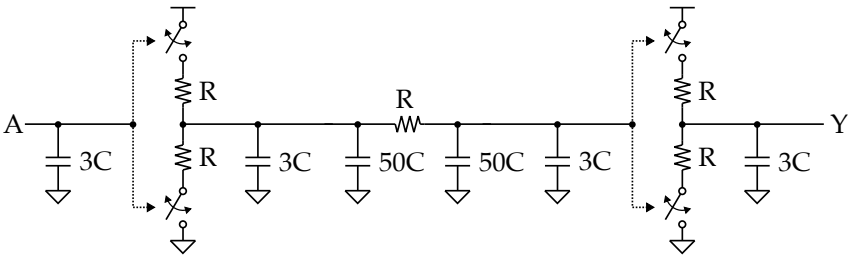
$$t_{pd} = R_1C_1 + R_1C_2 + R_1C_3$$

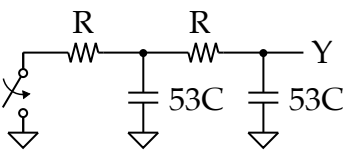


$$t_{pd} = R_1C_1 + (R_1 + R_2)C_2 + (R_1 + R_2)C_3$$

- Consider an INVX1 standard cell driving a long wire and INVX1

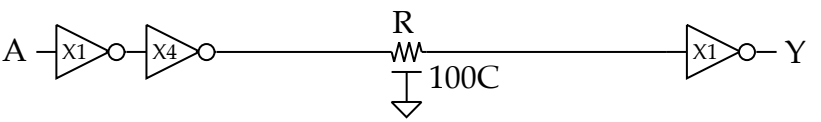




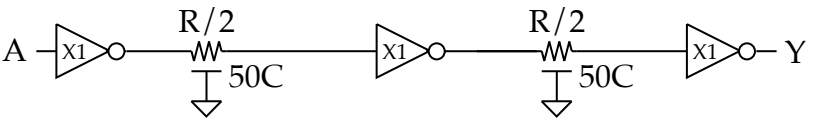


2.6. Larger Drivers and/or Repeaters

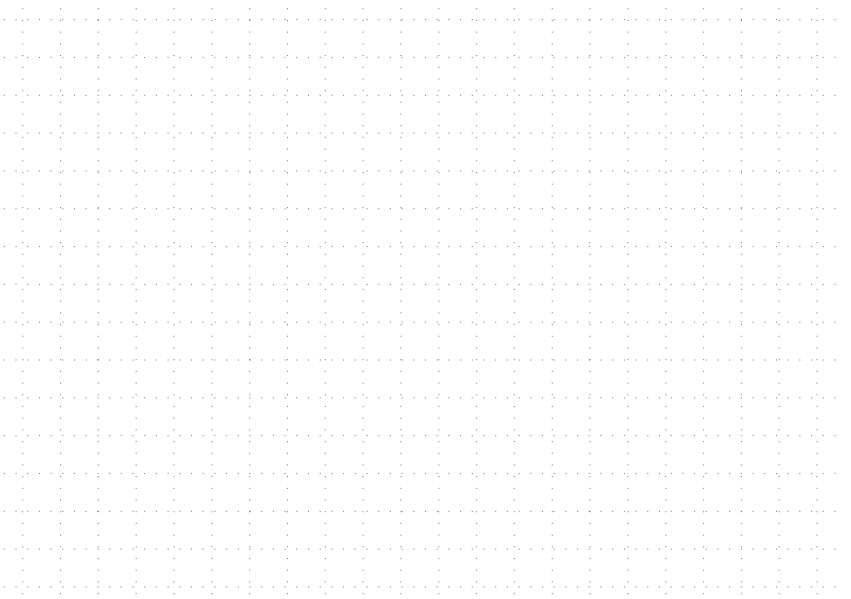
- We can decrease the propagation delay by using a larger driver
- For a fair comparison must normalize the overall input capacitance



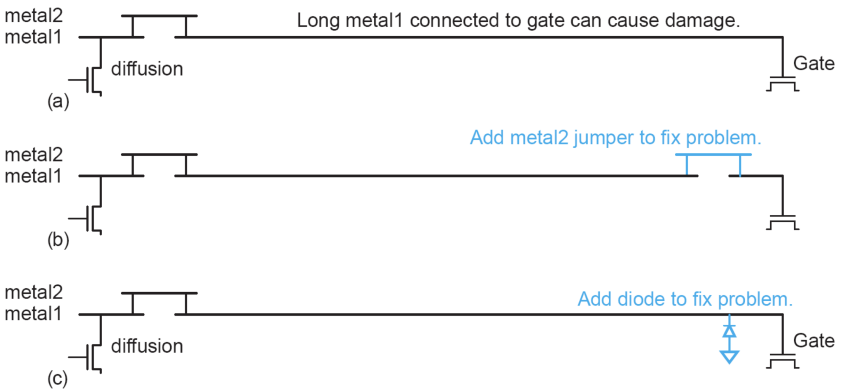
- We can decrease the propagation delay by using multiple repeaters



- There will be an optimal number of repeaters and optimal repeater sizing which balances the wire delay with repeater delay



3. Device-Related EDA Issues



- Antenna Rules
  - Etching metal wires can charge up the voltage on that wire
  - This can destroy fragile thin gate oxides
  - Fix by adding short jumpers to higher metal layer or antenna diodes
- Layer Density and Orientation Rules
  - Low polysilicon density can result in transistor gates under etched
  - Low metal density may cause dishing of copper when polishing
  - Fix by requiring specific density of poly and metal layers
- Layer Orientation Rules
  - Polysilicon in multiple directions can cause process variability
  - Polysilicon with small jogs can complicate optical proximity correction
  - Fix by requiring all polysilicon to be oriented in direction on grid

- Yield Enhancements
  - Space out wires to reduce risk of short circuits and reduce capacitance. Use non-minimum-width wires to reduce risk of open circuits and to reduce resistance.
  - Use at least two vias for every connection to avoid open circuits if one via is malformed and to reduce electromigration wearout.
  - Surround contacts and vias by landing pads with more than the minimum overlap to reduce resistance variation and open circuits caused by misaligned contacts.
  - Use wider-than-minimum transistors; minimum-width transistors are subject to greater variability and tend not to perform as well.
  - Avoid non-rectangular shapes such as 45-degree angles and circles. For specialized circuits such as RAMs that strongly benefit from 45-degree angles, verify masks after optical proximity correction analysis.