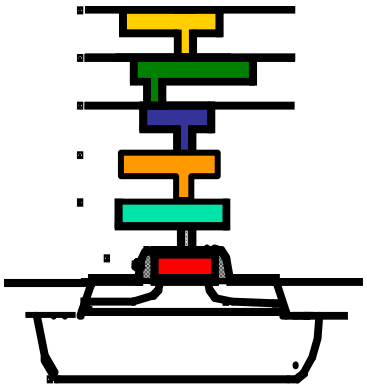
Scaling and Evaluation of Carbon Nanotube Interconnects for VLSI Applications

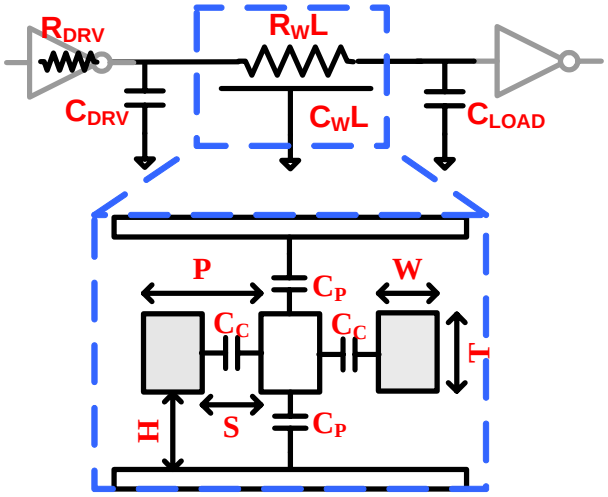
Fred Chen, Ajay Joshi,
Vladimir Stojanović and Anantha Chandrakasan

Technology Design Flow

Process Characterization

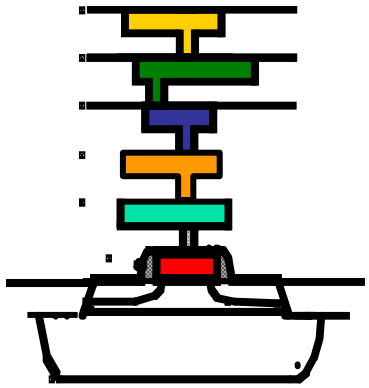


Create Models for Circuits

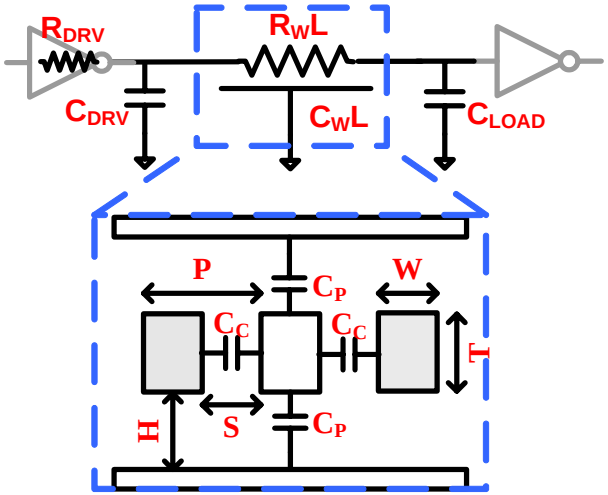


Technology Design Flow

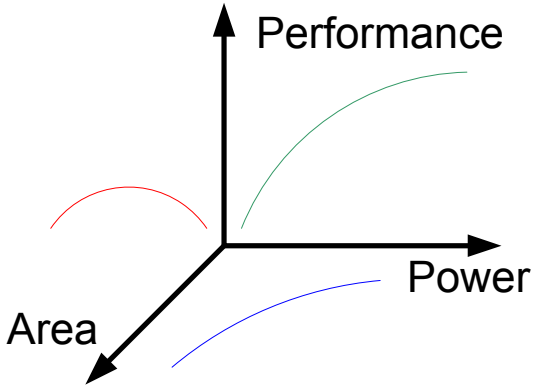
Process Characterization



Create Models for Circuits

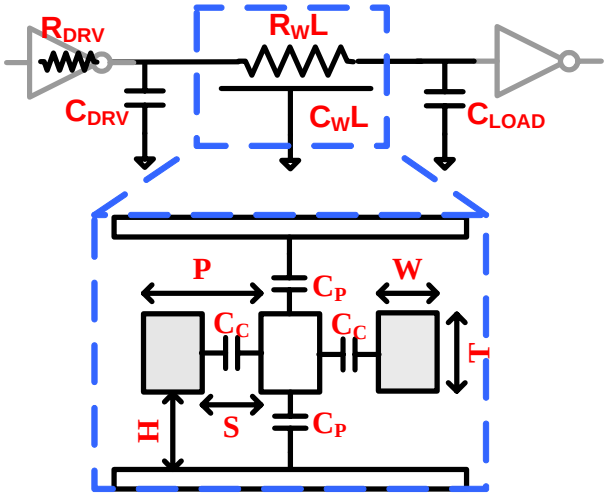


Extract Tradeoff Curves

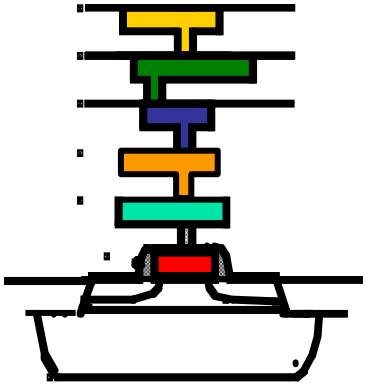


Technology Design Flow

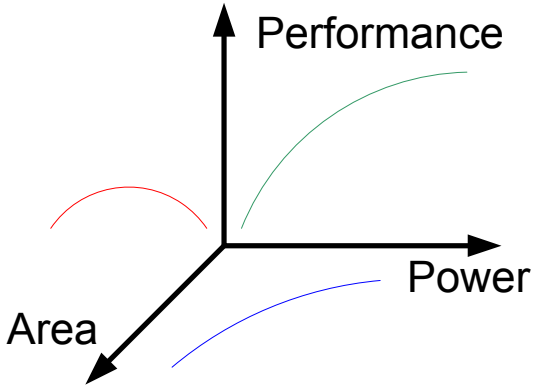
Create Models for Circuits

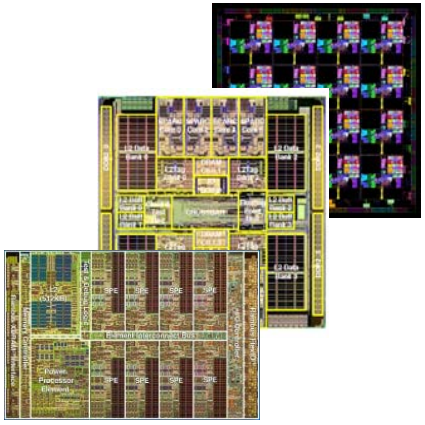


Process Characterization



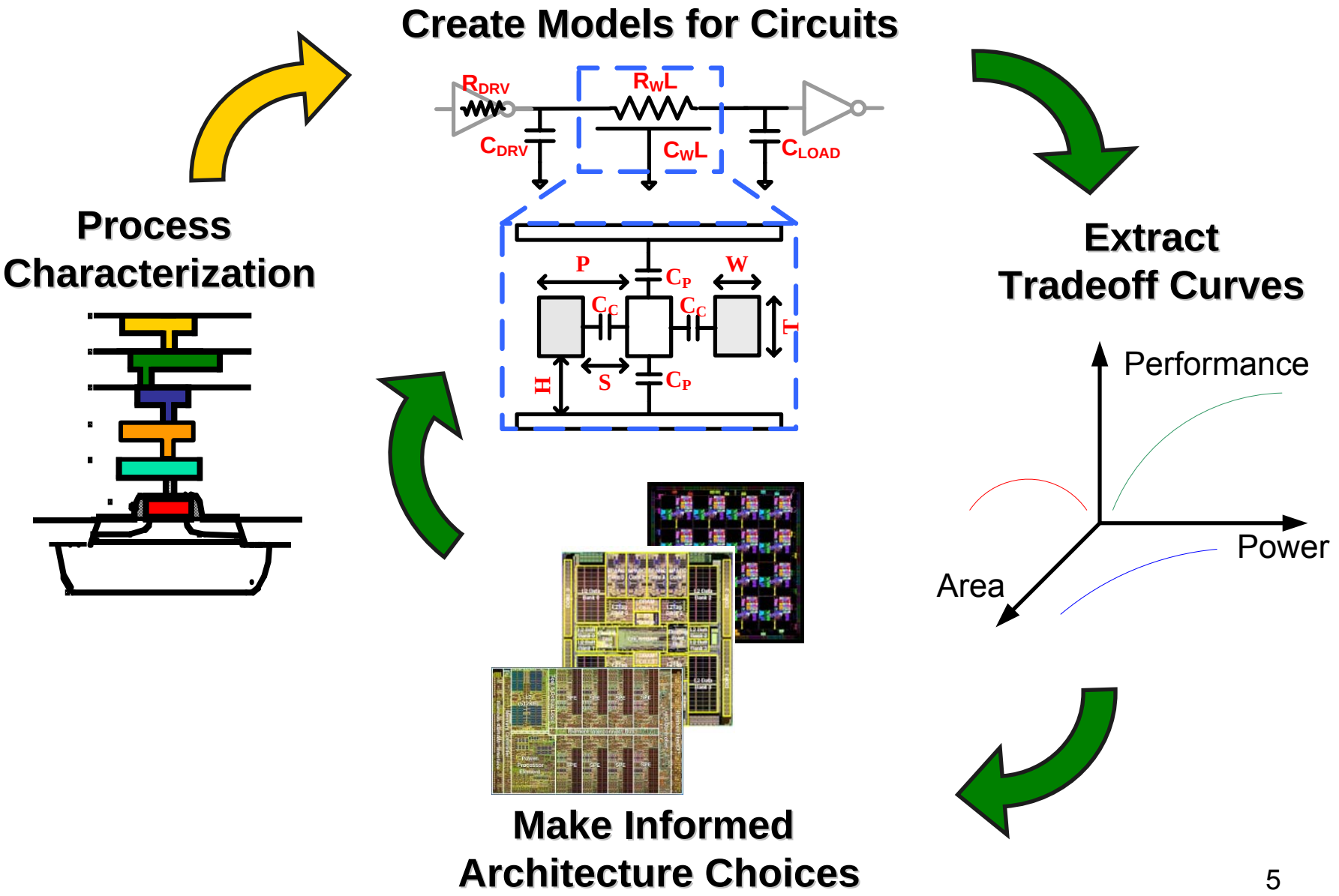
Extract Tradeoff Curves



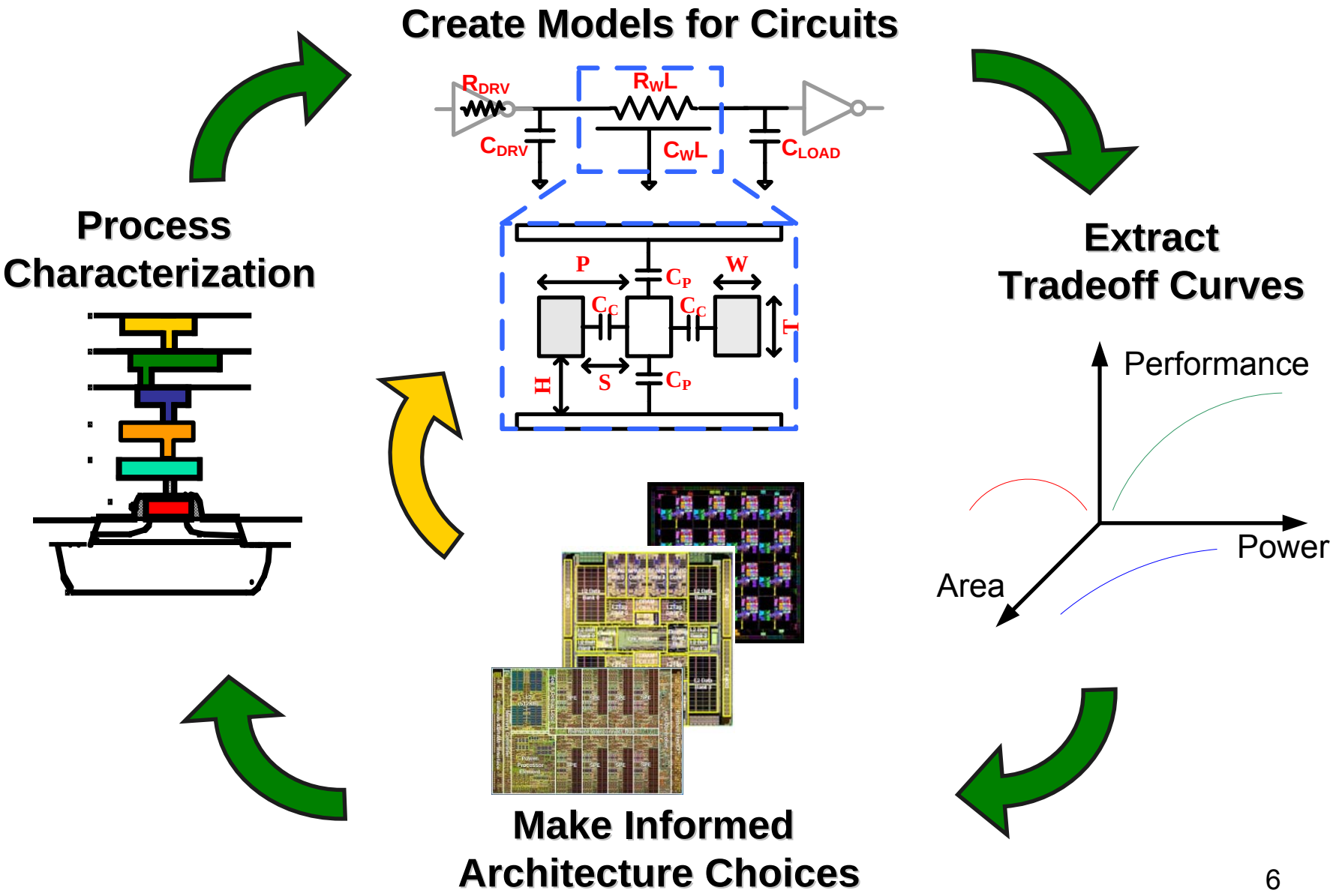


Make Informed Architecture Choices

Technology Design Flow

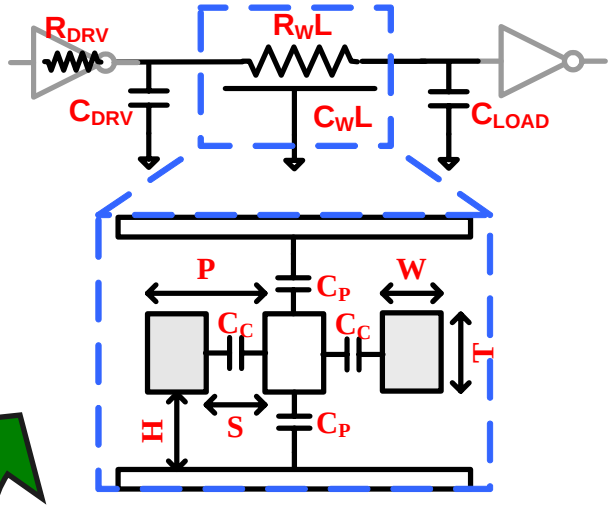


Technology Design Flow

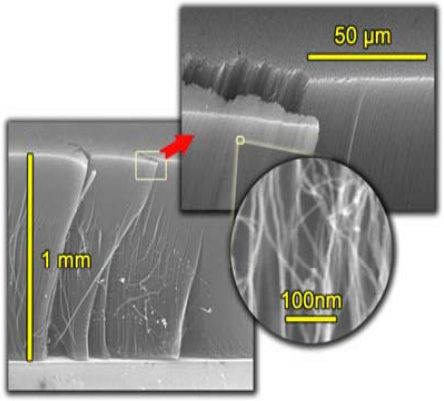


Technology Design Flow

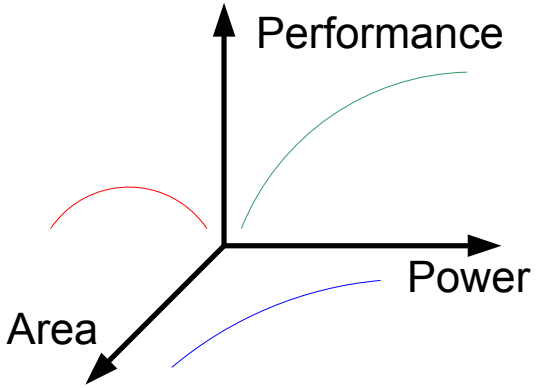
Create Models for Circuits

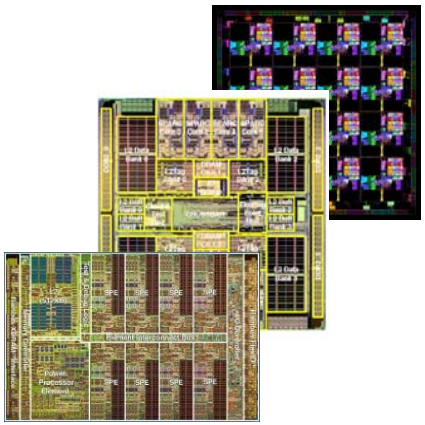


CNT Process Characterization



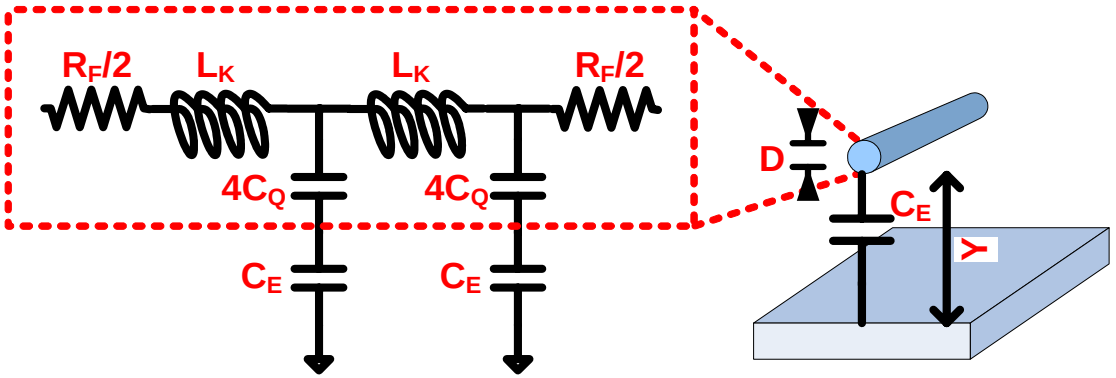
Extract Tradeoff Curves





Make Informed Architecture Choices

Ideal Isolated SWCNT Model



[Burke, *Nanotechnology*]

$$R_F = \begin{cases} h/4e^2 \sim 6.5k\Omega, & L \le L_0 \\ \left(h/4e^2\right)\left(L/L_0\right), & L > L_0 \end{cases}$$

$$L_K = h/2e^2 v_F \sim 16\,nH/\mu m$$

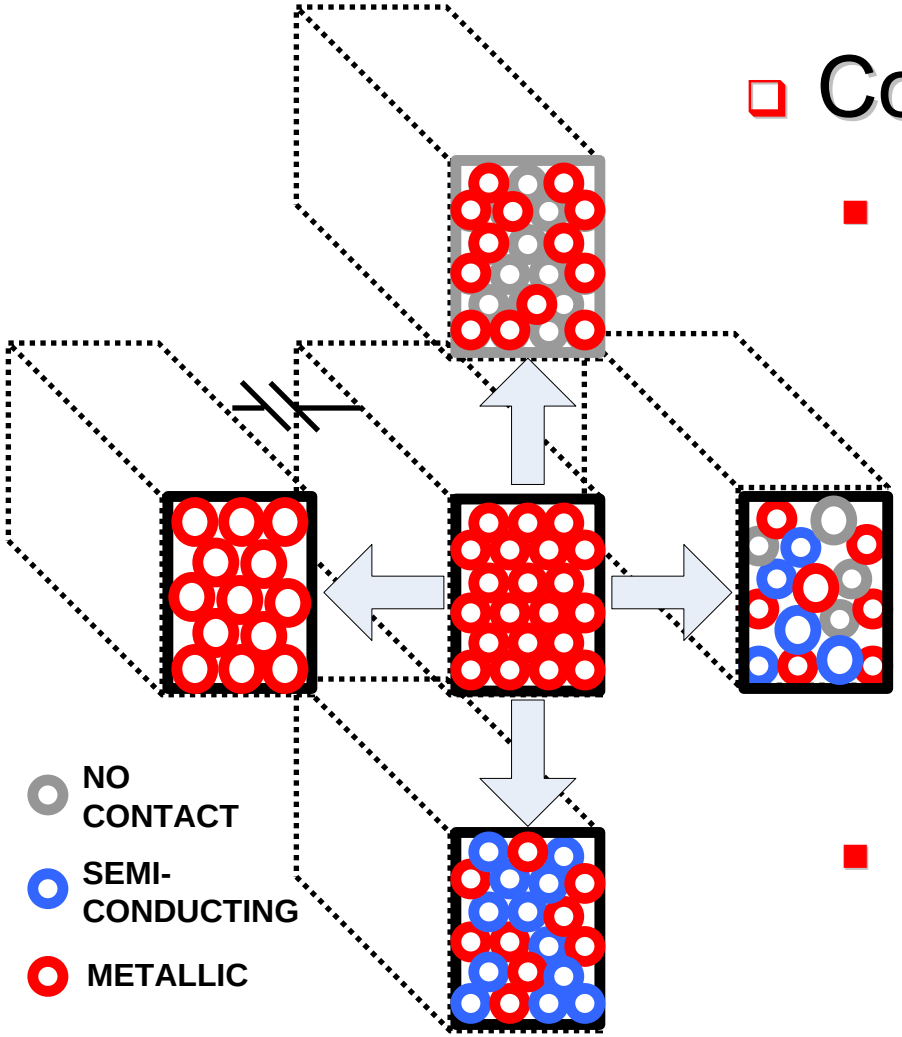
$$C_Q = 2e^2/hv_F \sim 100\,aF/\mu m$$

$$C_E = 2\pi\varepsilon/\ln\left(y/d\right) \sim 34\,aF/\mu m,$$

$$\varepsilon = 2.8\varepsilon_0, y = 97nm, d = 1nm$$

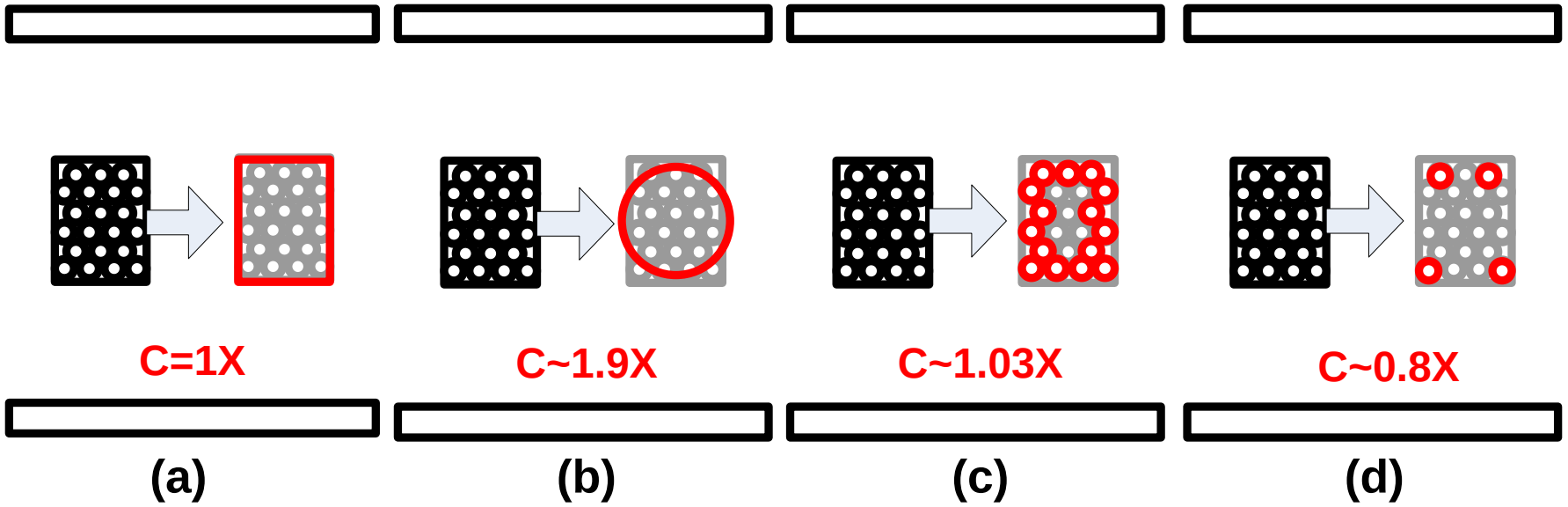
- Each parameter scales with length
  - Resistance – linear function of length beyond mean free path (L_0)
  - Capacitance – dominated by electrostatic cap (C_E)
  - Kinetic Inductance – significant only at very high frequencies (> 7GHz), unobserved thus far*.

Modeling CNT Bundles



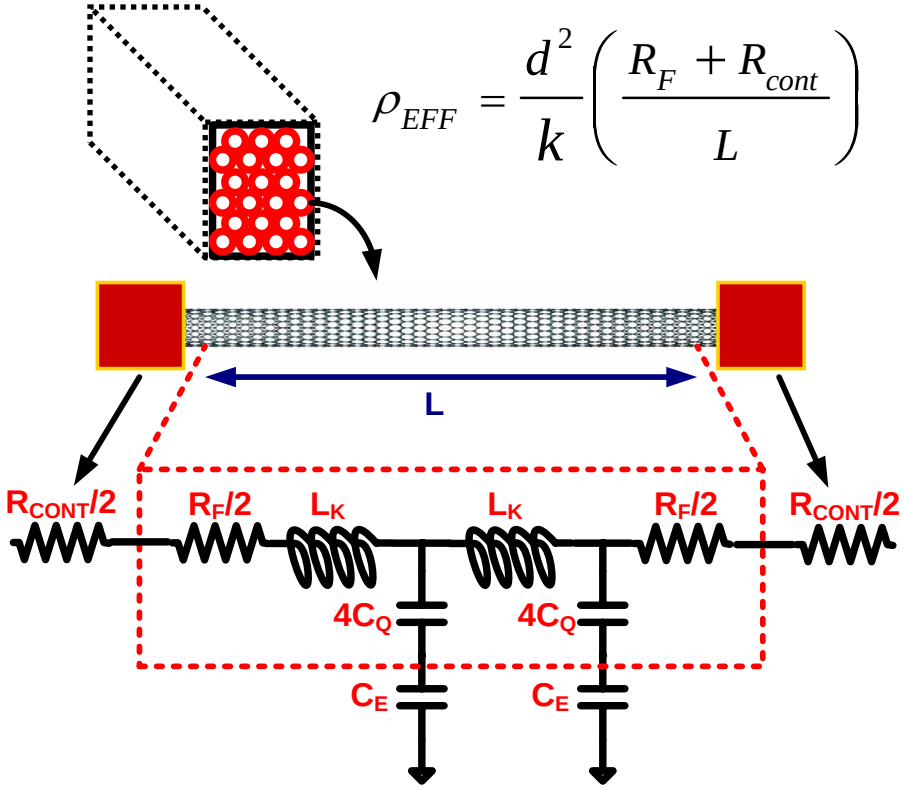
- Consider CNTs a material:
  - 'Resistivity' dependent on:
    - Tube diameter
    - Contact resistance
    - Mean-free path
    - Fraction of contacted metallic CNTs
  - Capacitance dependent on:
    - Bundle dimensions

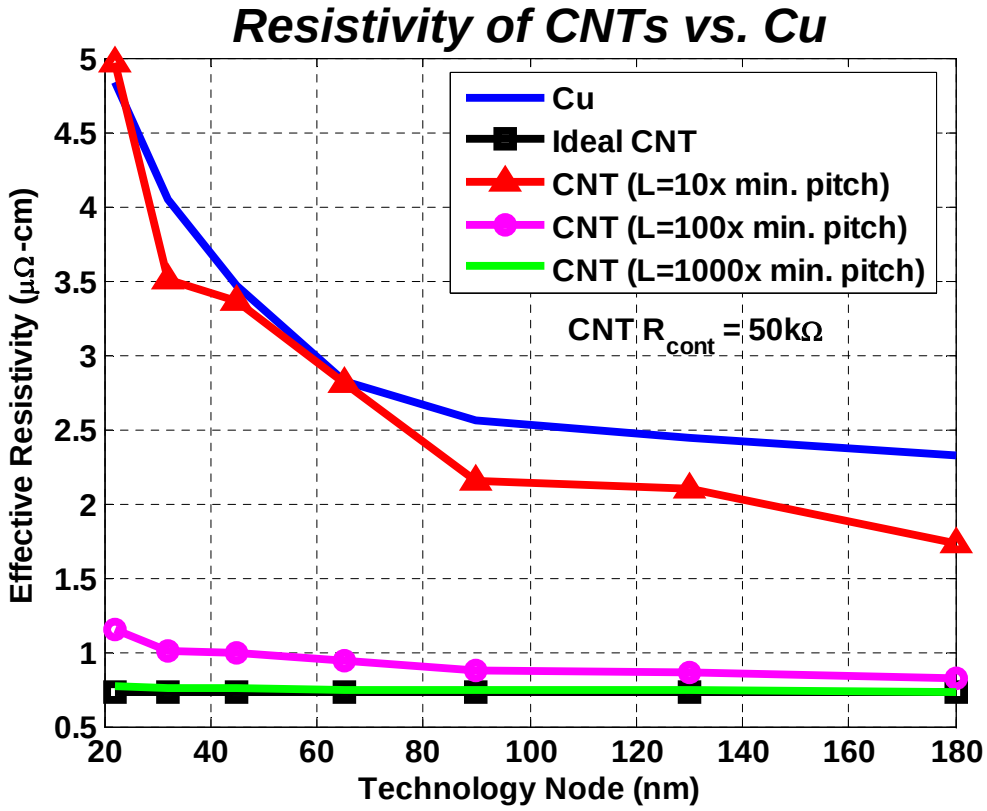
Capacitance of CNT Bundles



- For the same cross-section, electrostatic capacitance for all materials is the same.
  - Prior approaches use same 3D solvers as used for Cu
    [*Raychowdhury, 2004, Srivastava, 2005, Naeemi, 2005, 2007*]
- Assume similar surface roughness between Cu wires and CNT bundles

Effective Resistivity of CNTs

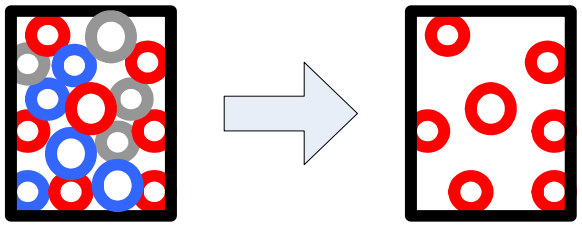




- Resistivity of ideal CNT bundles ~7x lower than Cu at 22nm
- For interconnect lengths > ~1000 gate pitches, contact resistance insignificant

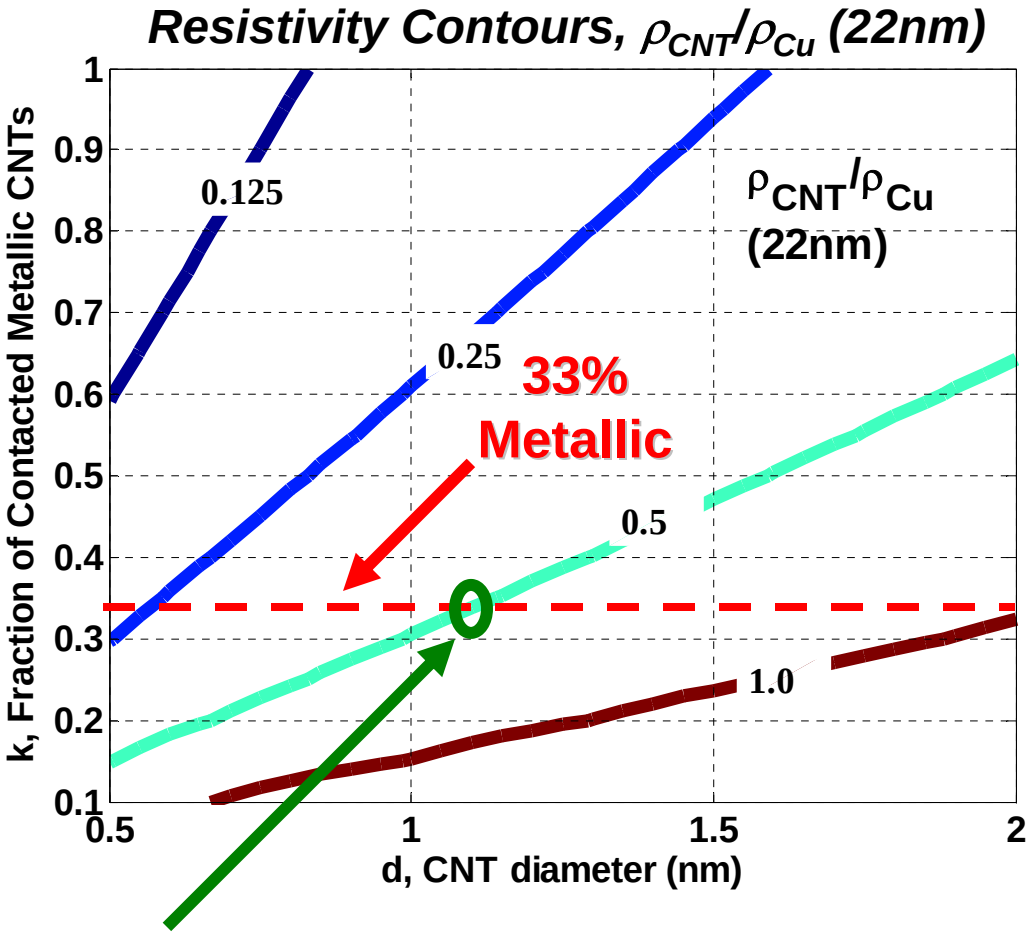
Effective Resistivity: Non-idealities

CONTACTED METALLIC CNT



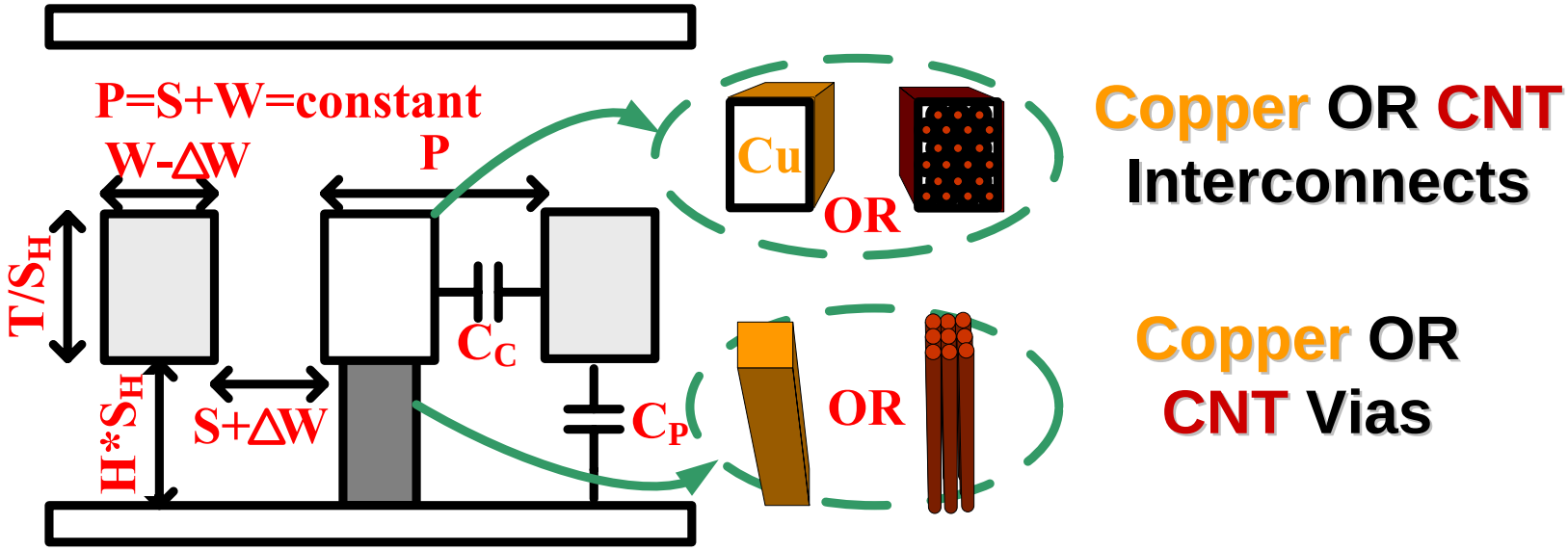
$$\rho_{EFF} = \frac{d^2}{k}\left(\frac{h}{4e^2 L_0} + \frac{R_{cont}}{L}\right),$$

$$L > L_0 \quad where\ L_0 = C_\lambda d$$



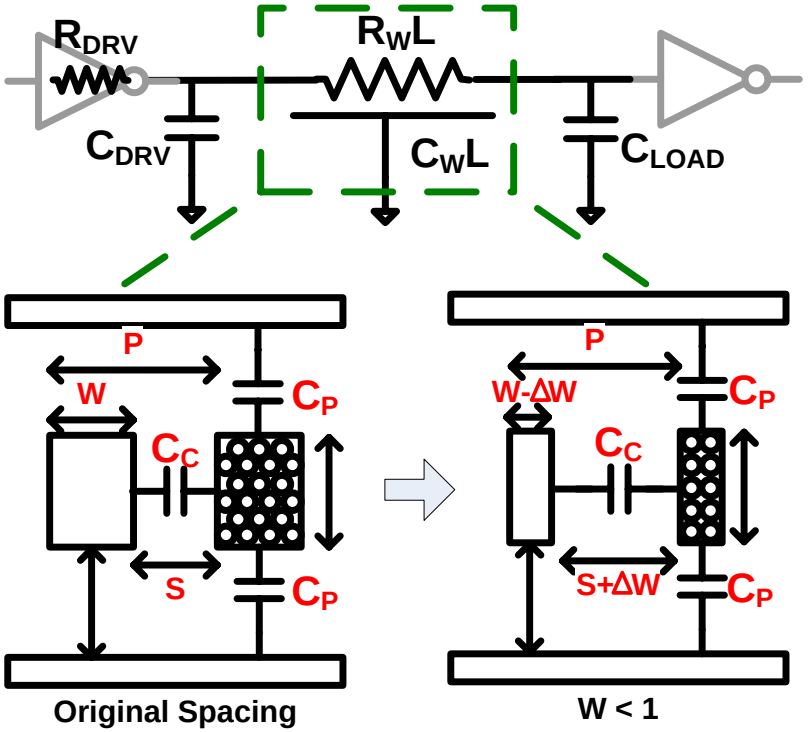
- Current growth limitations can achieve ~2x lower resistivity than copper

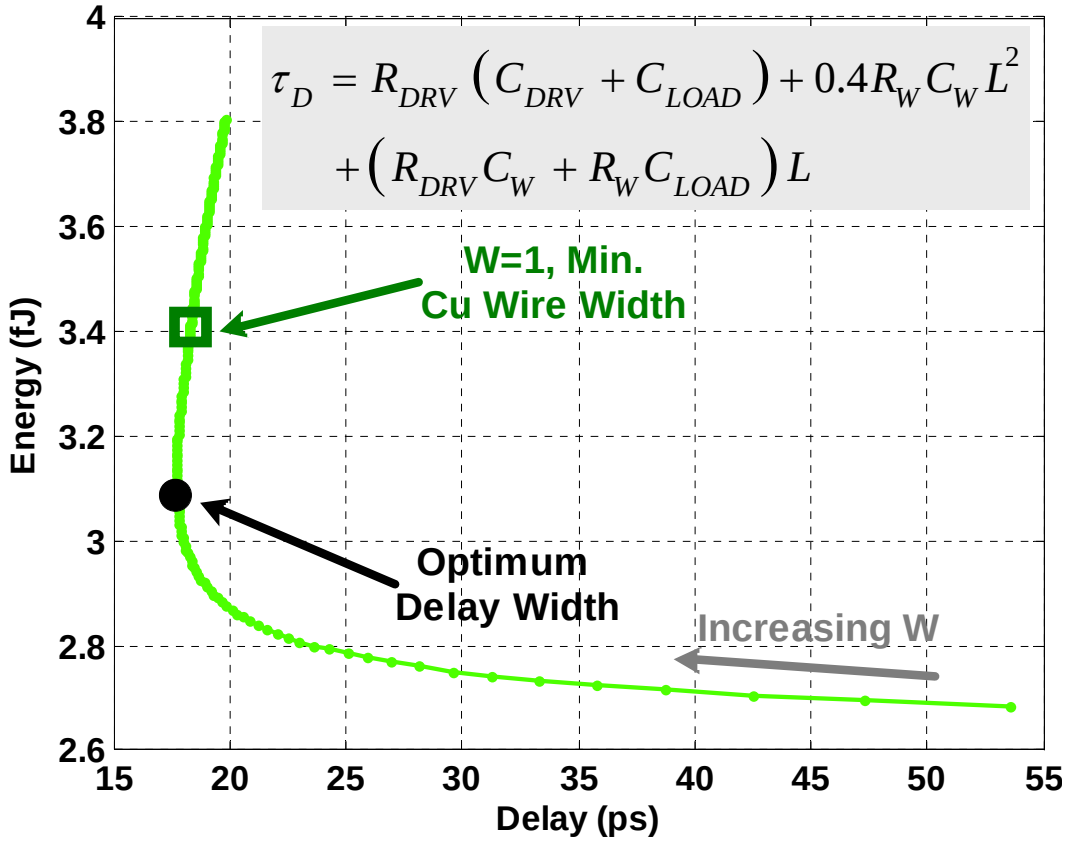
Scaling Interconnect: Design Space



- Consider rescaling the interlayer dielectric (ILD) stackup
  - Scale width: Allow CNT bundles to be grown & assembled at finer widths
  - Increase ILD height (H): CNT vias enable higher aspect ratios, thicker ILD

Scaling Wire Width



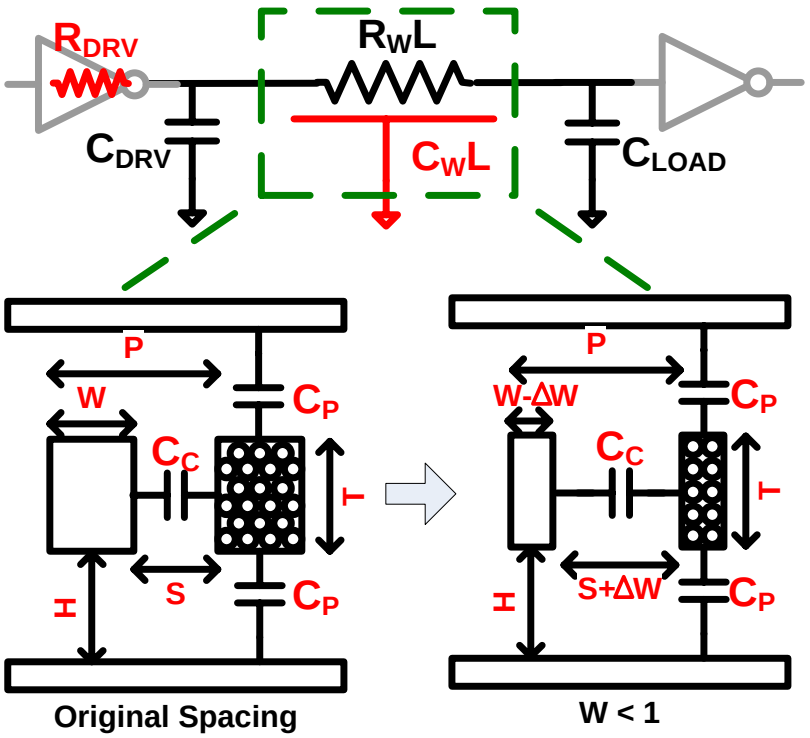


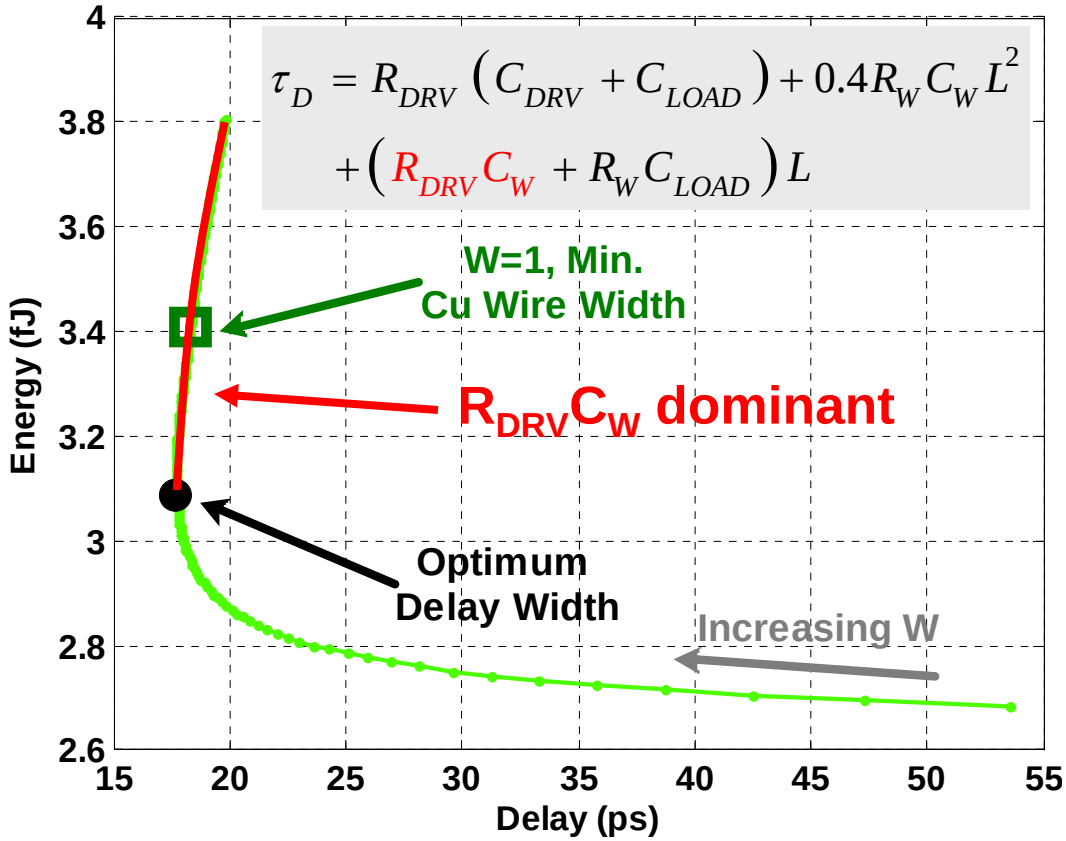
As W ↓ → R_W ↑ → C_W ↓

Cu, 45nm, 8X min. buffer load,
L=250X, FO=4,

- Scale W, but maintain minimum wire pitch (P)
  → constant area

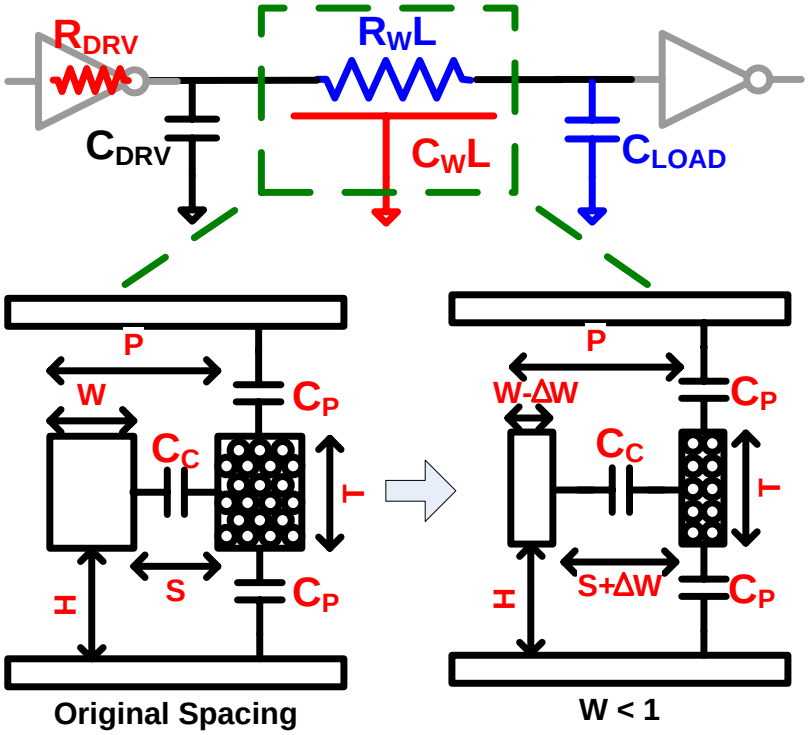
Scaling Wire Width

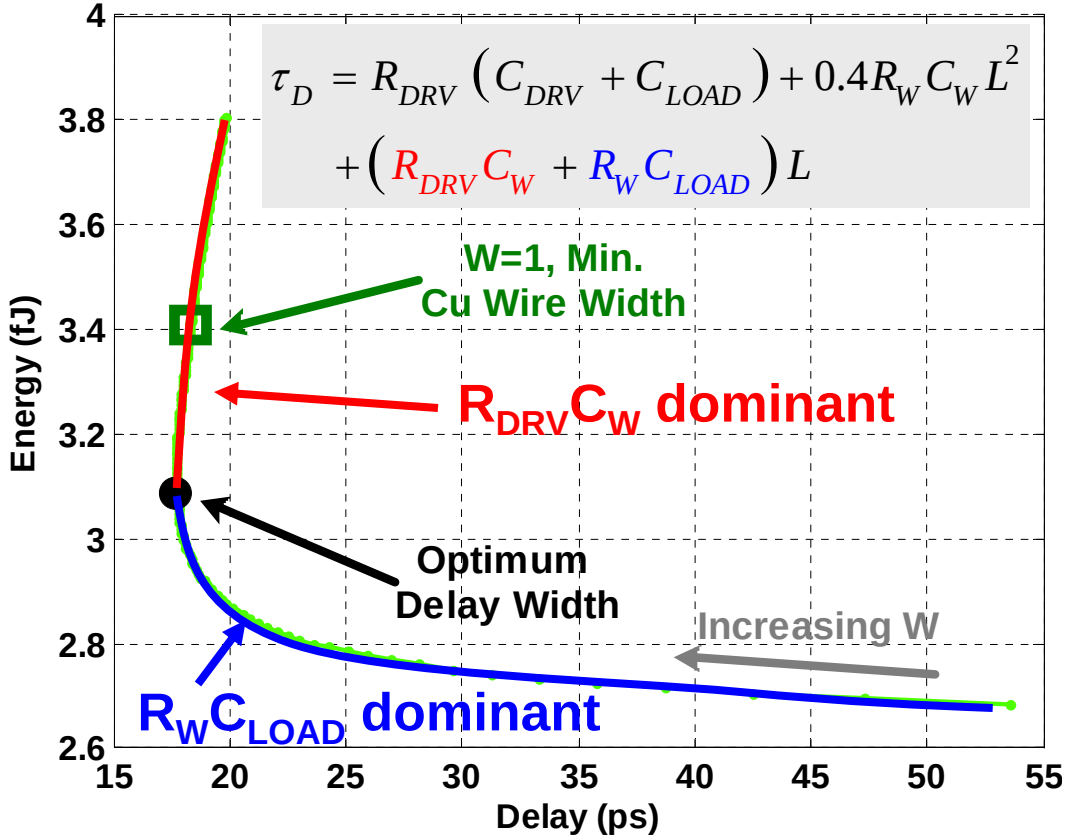




- When delay is dominated by driver resistance, decreasing wire W improves BOTH delay & energy

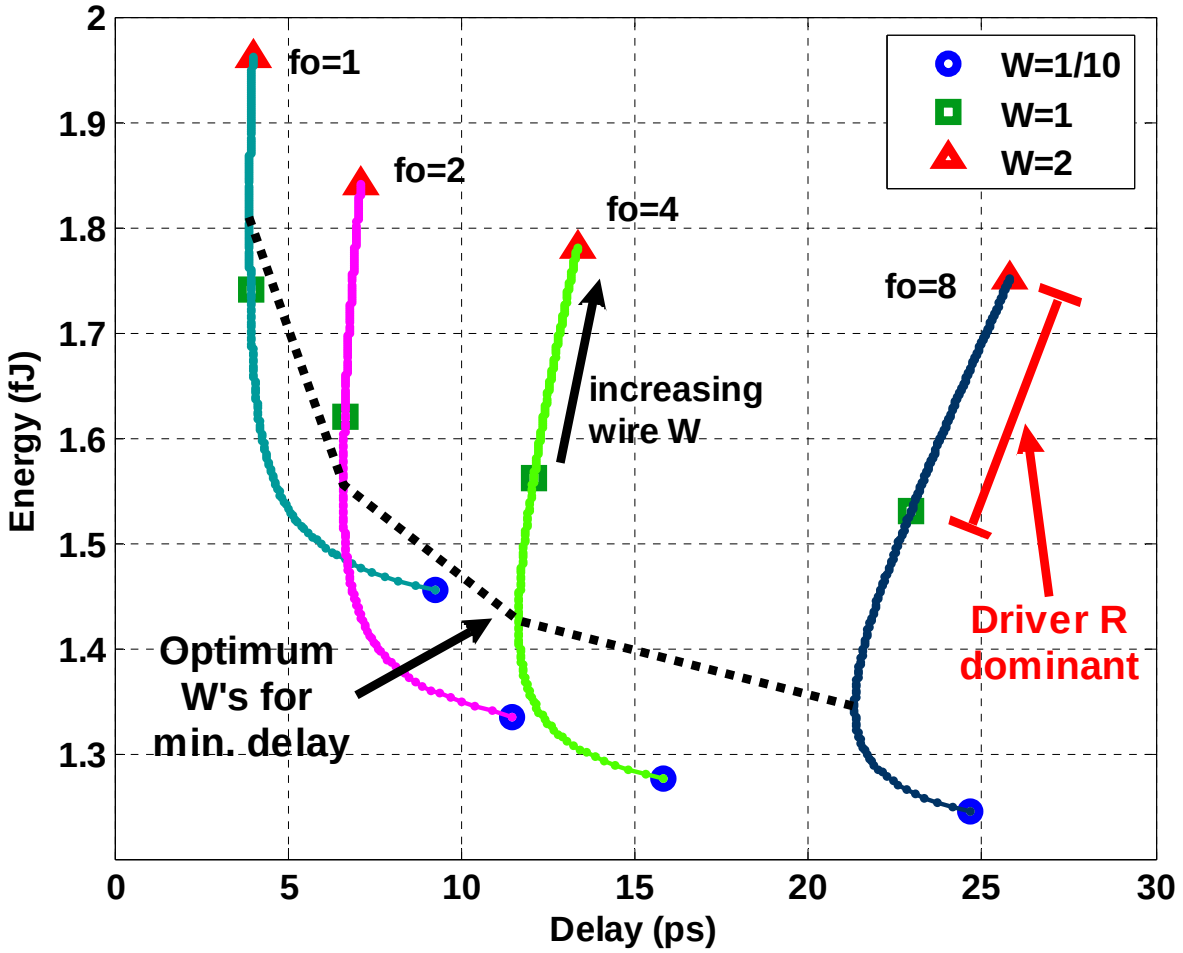
Scaling Wire Width





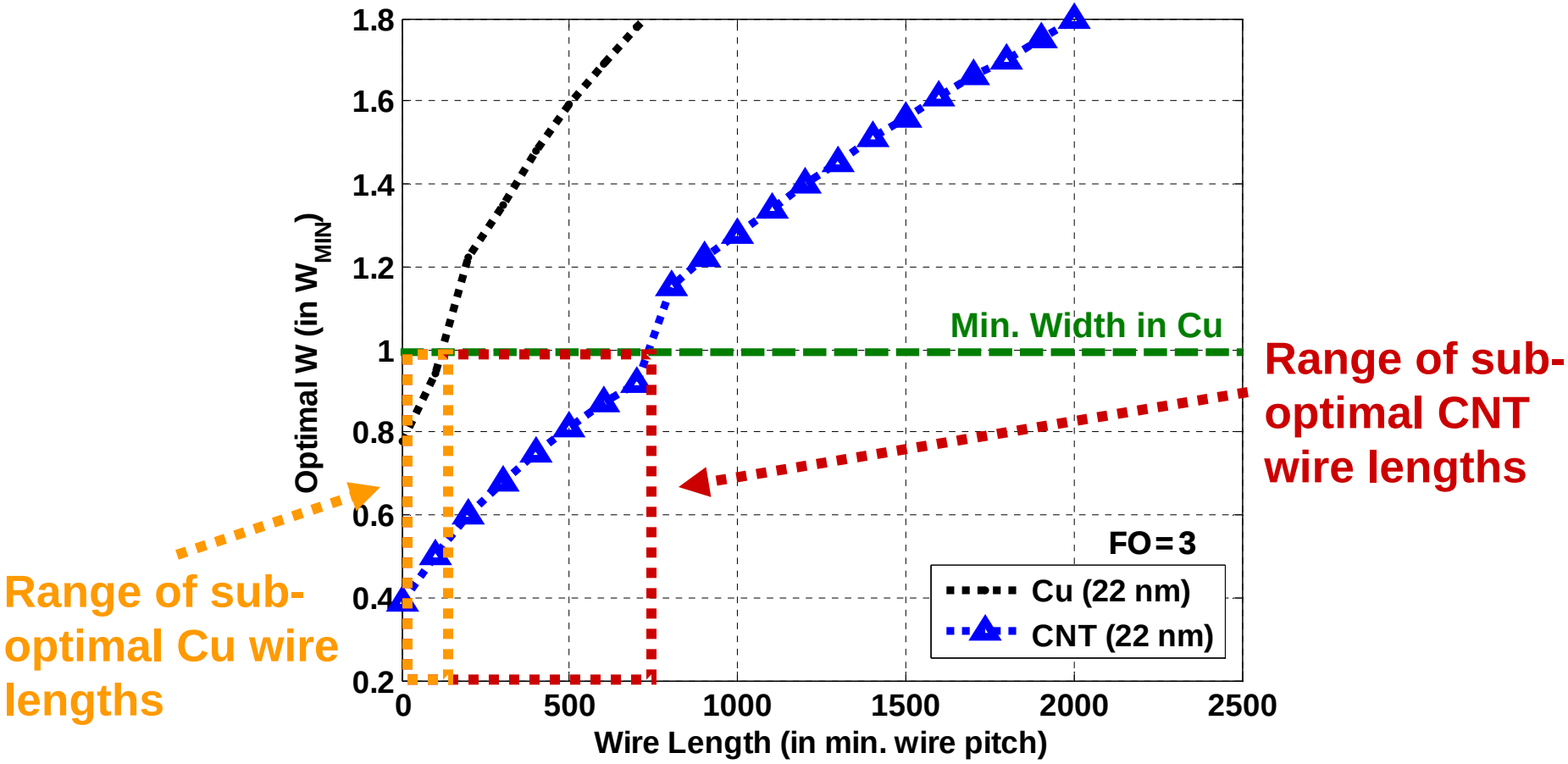
- When delay is dominated by wire resistance, scaling W trades off delay for energy
- Optimum delay width separates the two regions

Optimum Delay Wire Width (CNT, 22nm)



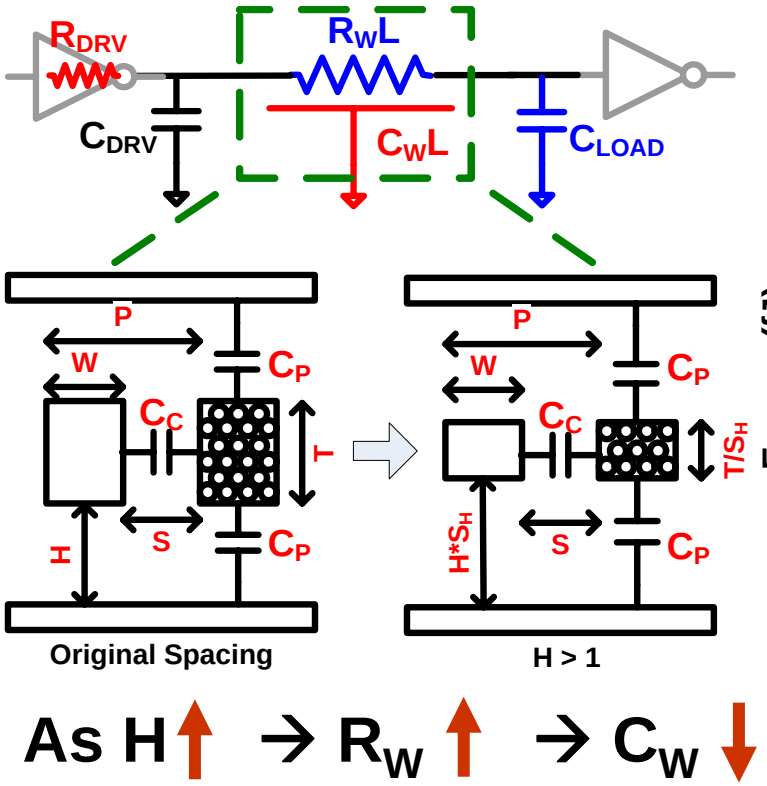
- When optimum delay wire width < 1, wires narrower than the min Cu width result in energy AND delay improvement

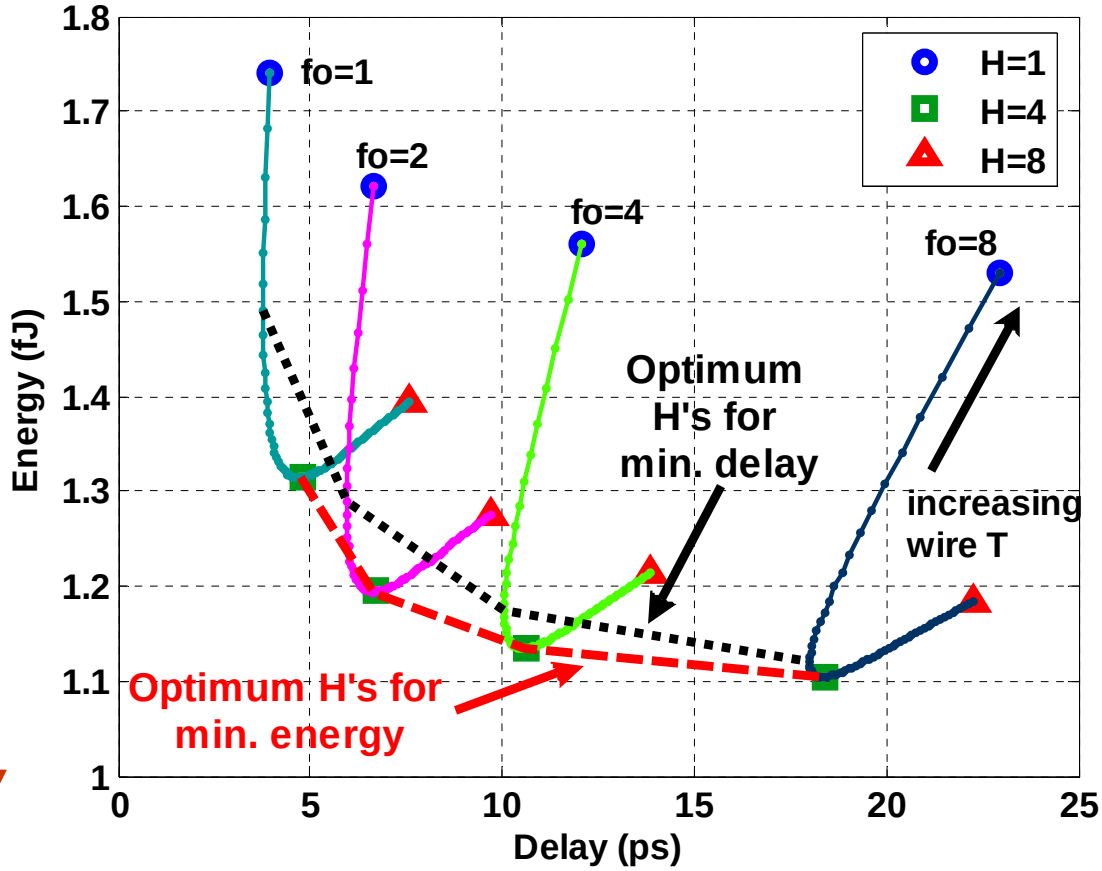
CNTs Require Rescaling



- Range of sub-optimally sized wires is greater if CNTs are used with the same cross-section as copper.

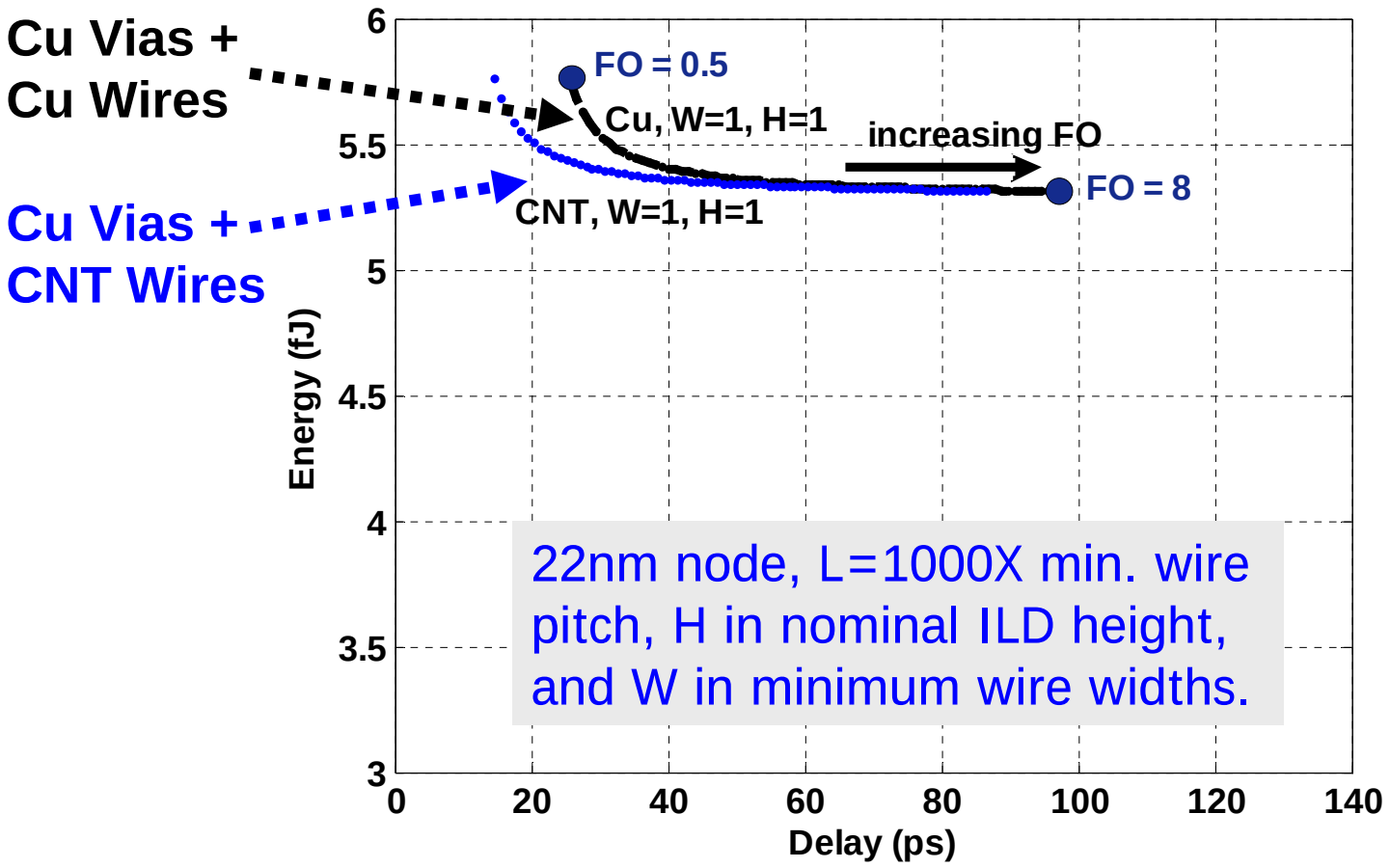
Scaling ILD Height





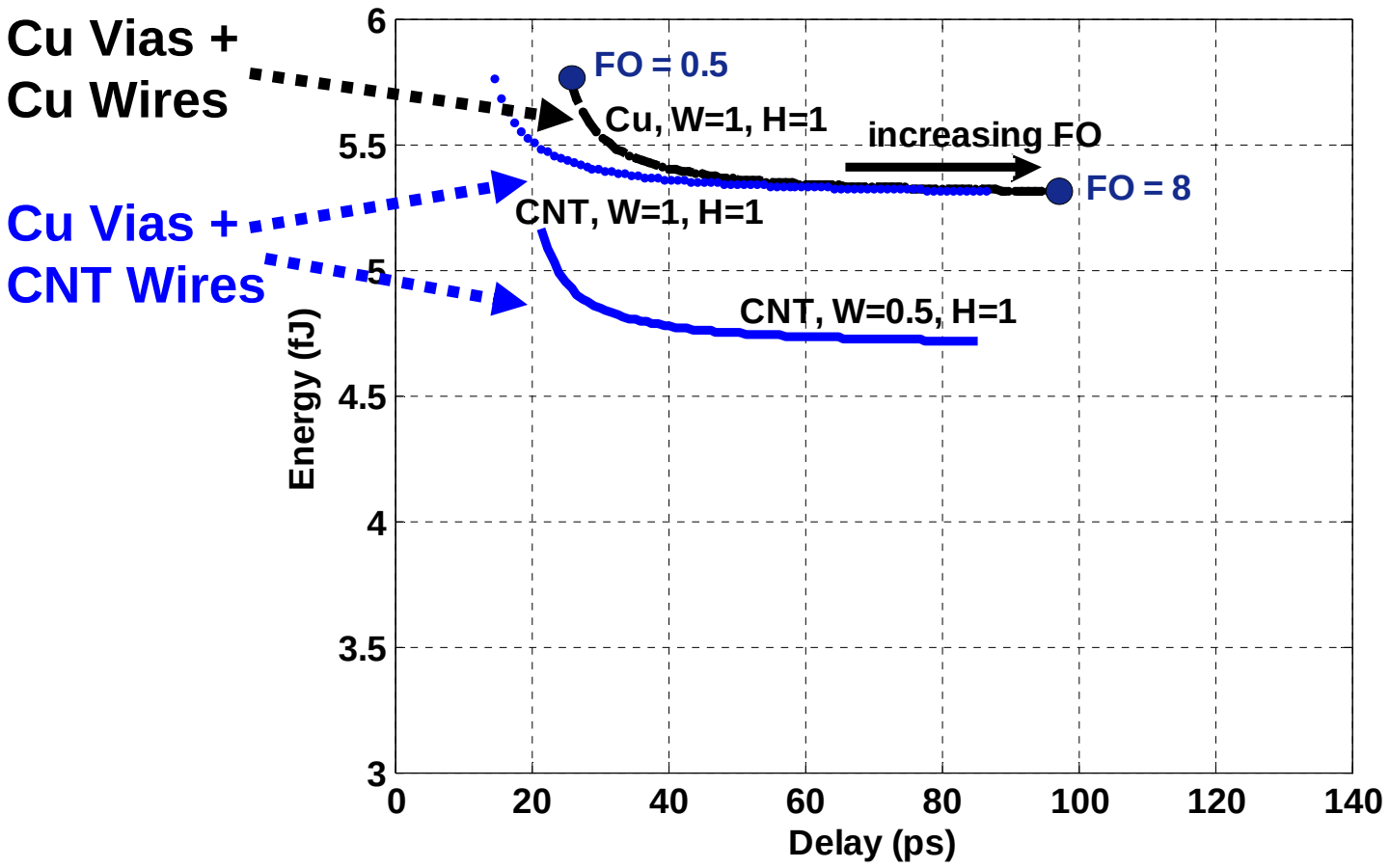
- Scale H, reverse scale T to maintain constant wire bandwidth (for comparison)
- Min. delay (~H=2) and a min. energy point (~H=4) exist

Energy vs. Delay: No Scaling



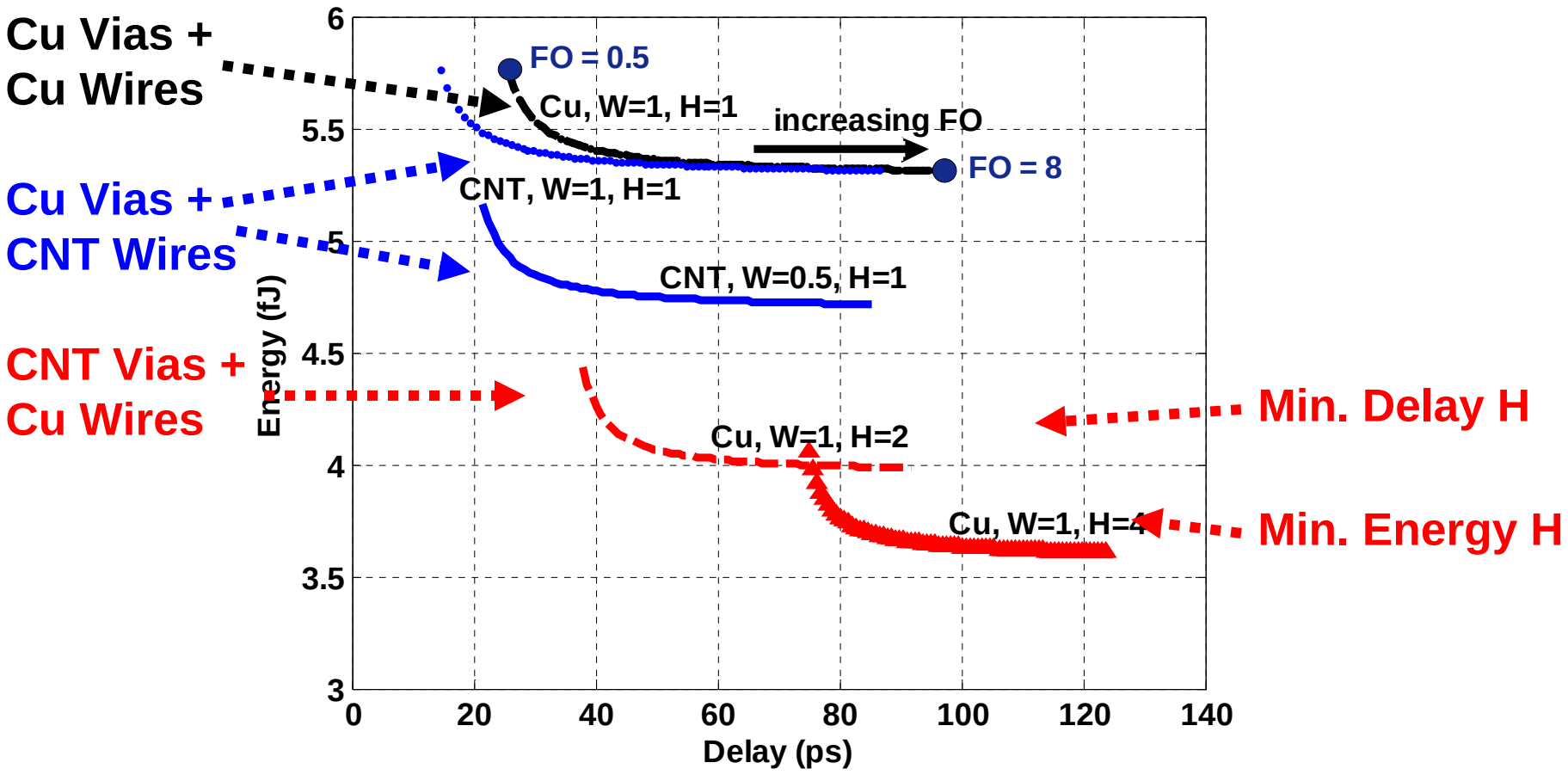
- Directly replacing Cu with CNTs (same cross-section) only yields delay improvement at lower fanouts

Energy vs. Delay: Scaling Width



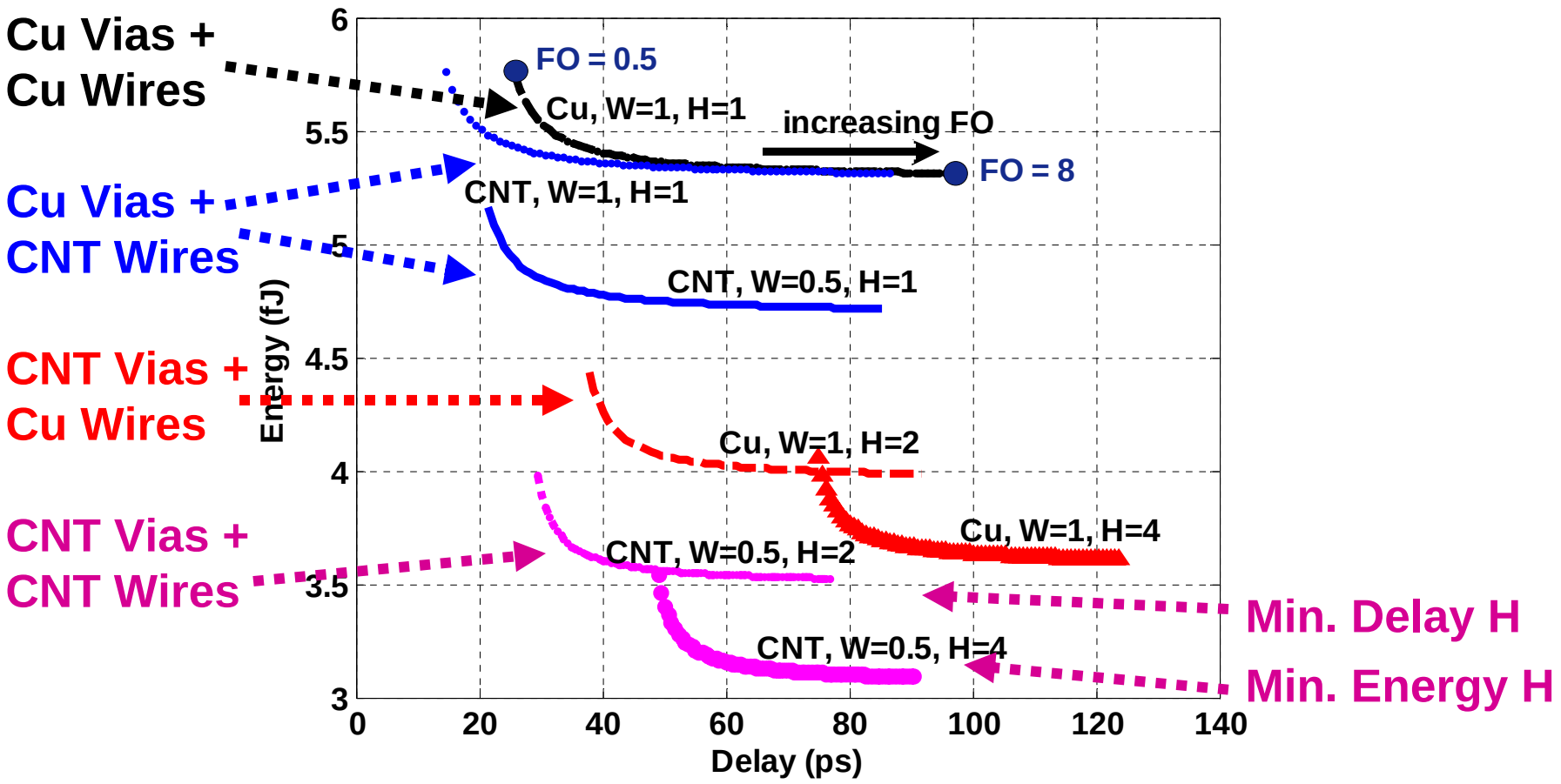
- Scaling wire width down improves energy and delay

Energy vs. Delay: Scaling ILD Height



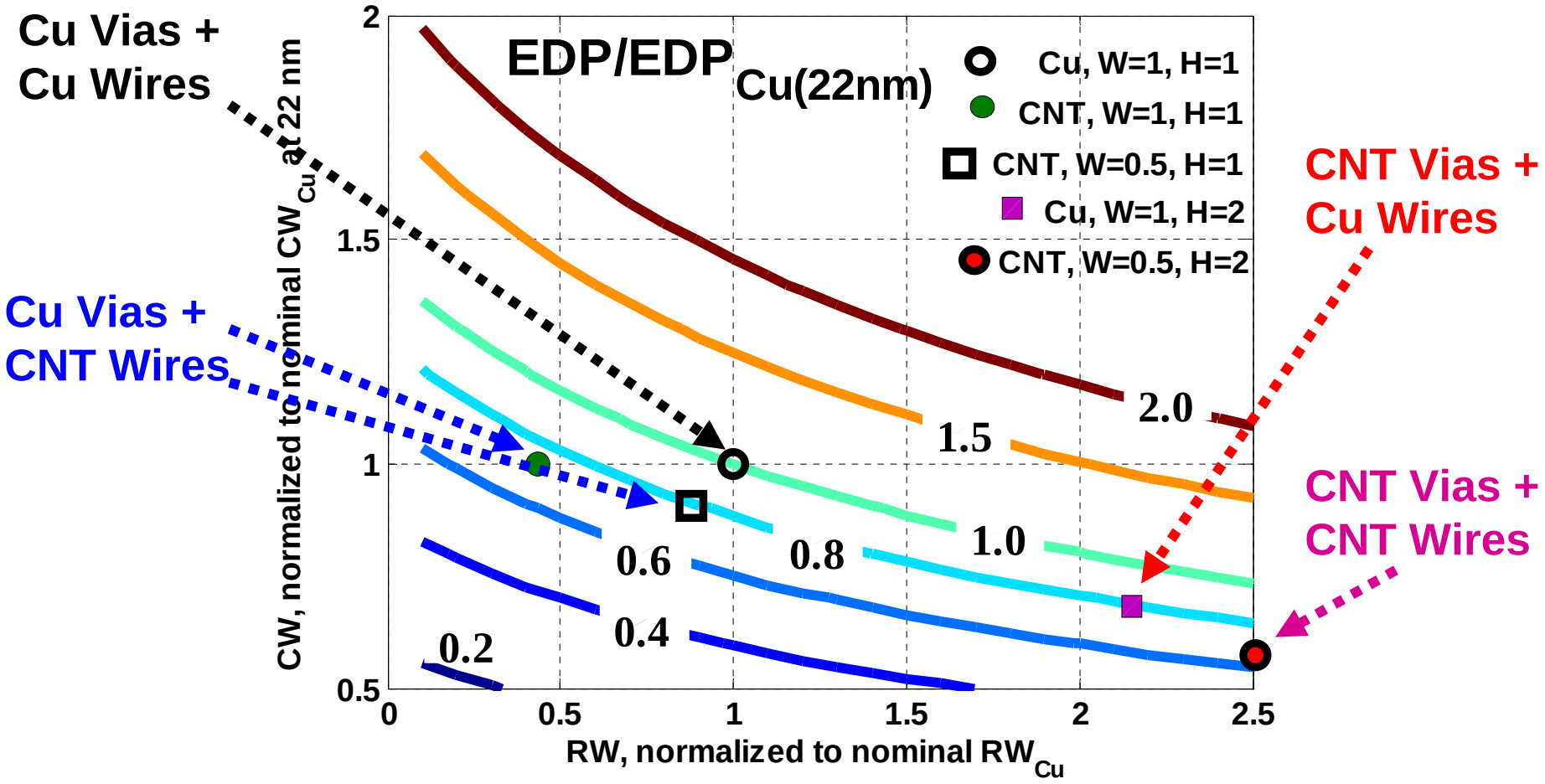
- Scaling wire height up using CNT vias, but Cu wires improves energy with small penalty in delay.

Energy vs. Delay: Scaling W & H



- Scaling both width and height result in almost 33% energy savings for the same delay

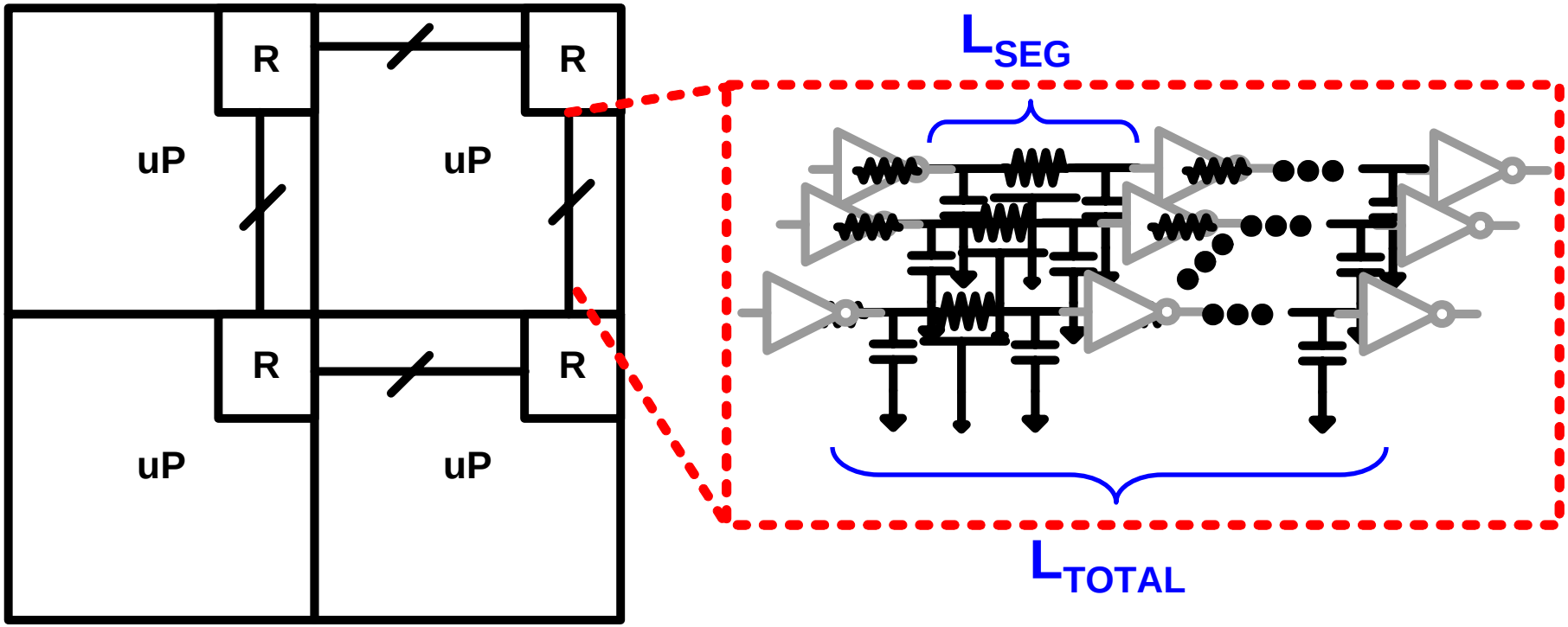
Energy Delay Product (EDP)



- Each configuration maps to a certain R_W and C_W
- Can map other configurations/materials to compare

System Evaluation: On-Chip Network

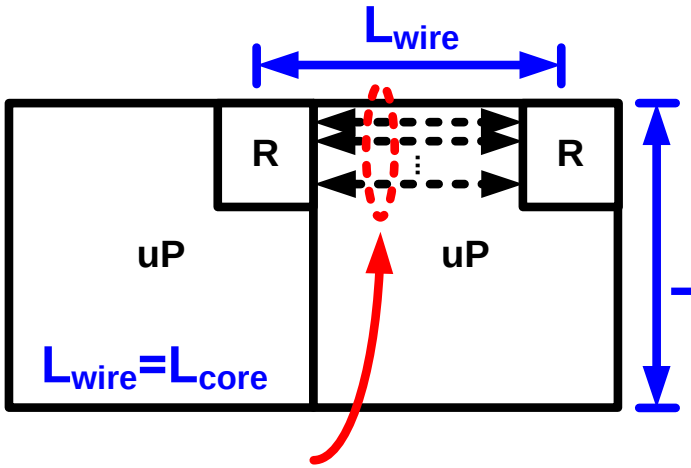
- Abstract the wire + driver model results to system level
- Extend to repeated interconnects…



$L_{SEG} = L_{TOTAL}/N$

N = # of repeaters

Core-to-Core Communication

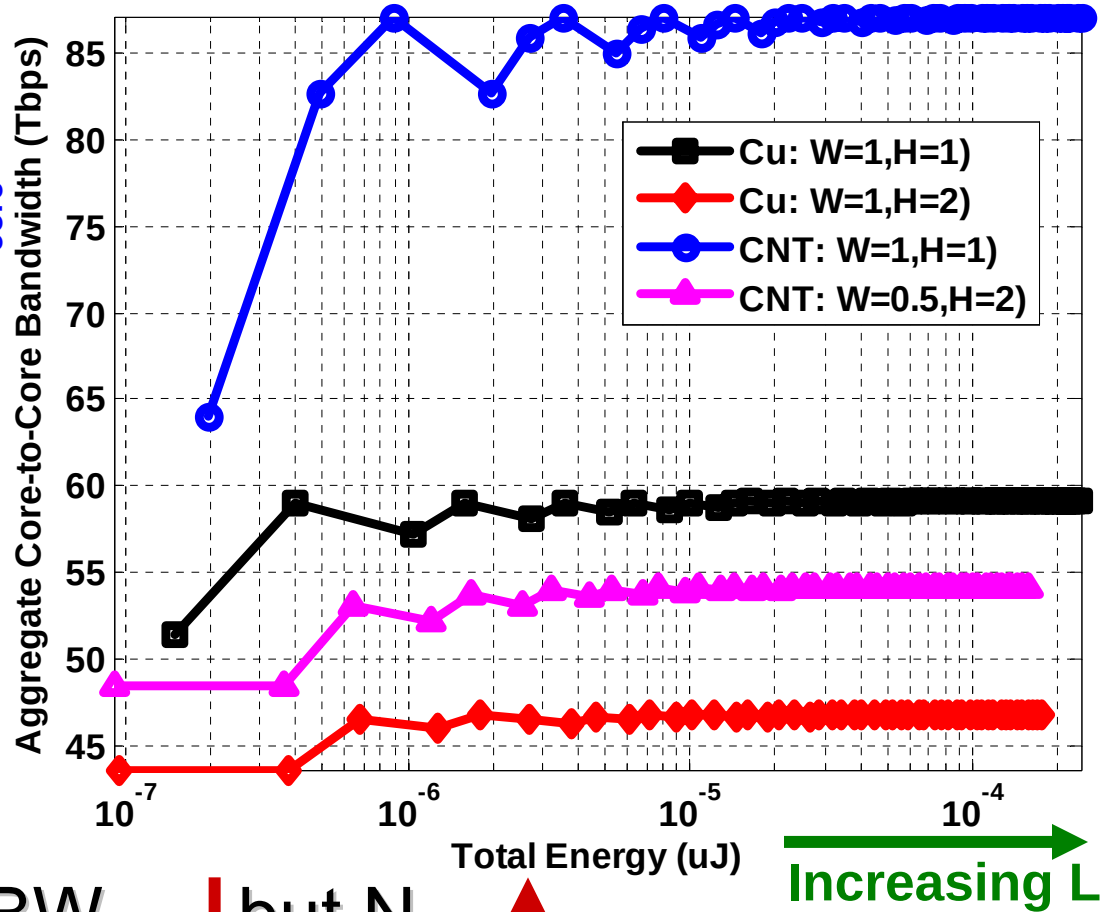


$N_{wire(max)} = L_{wire}/P_{wire\ (min)}$

$BW_{wire} \sim 1/\ _{D,wire}$

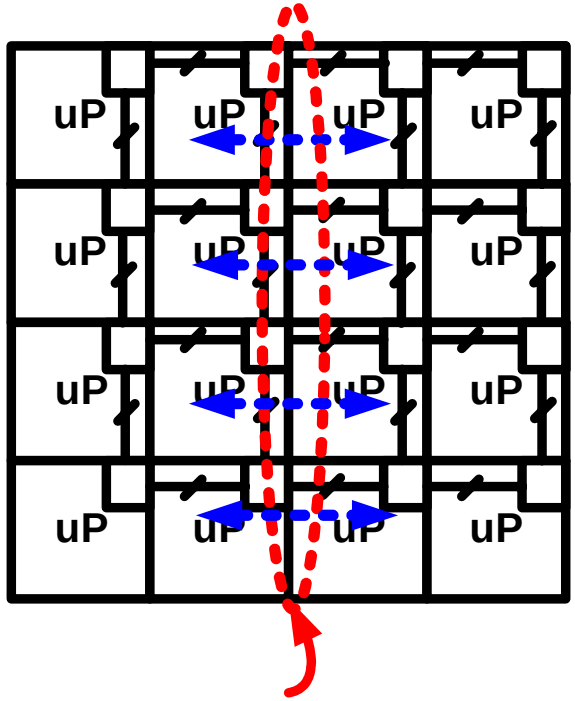
$BW_{agg} = N_{wire}BW_{wire}$

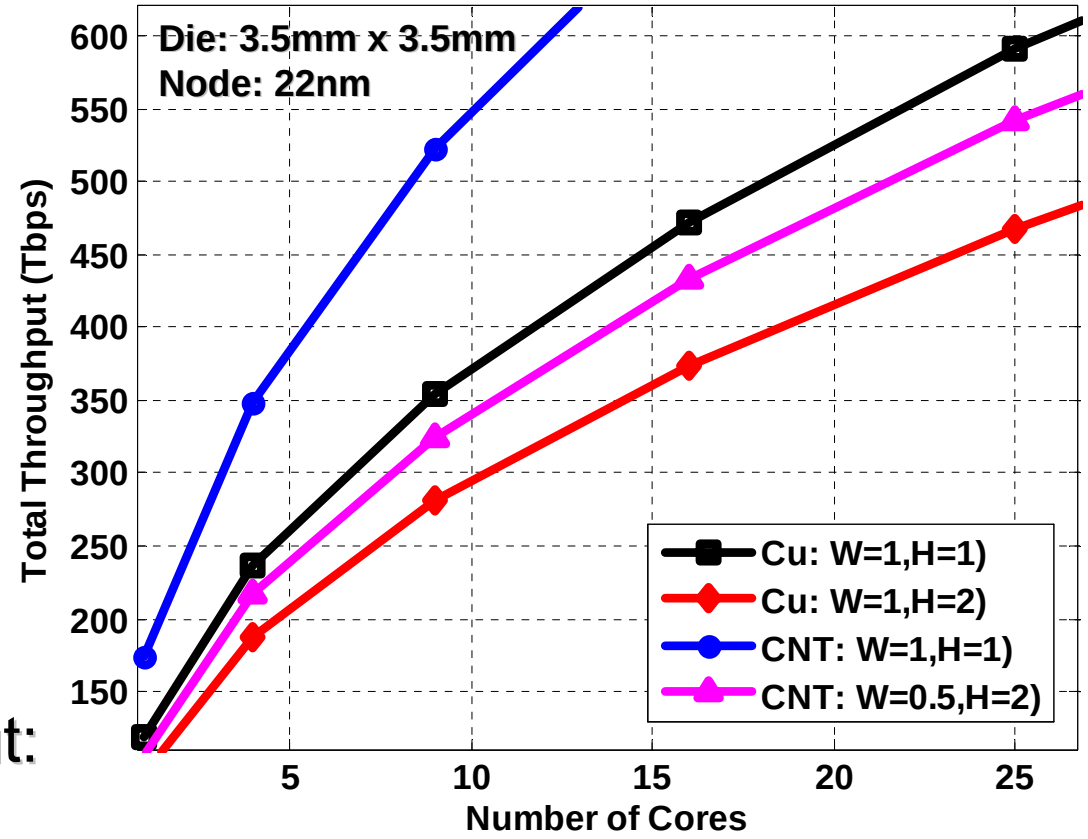
$E_{total} = N_{wire}E_{wire}$



- As the cores grow: BW_{wire} ↓ but N_{wire} ↑
- Bandwidth comes at different energy cost for each configuration

How Many Cores? Infinite?



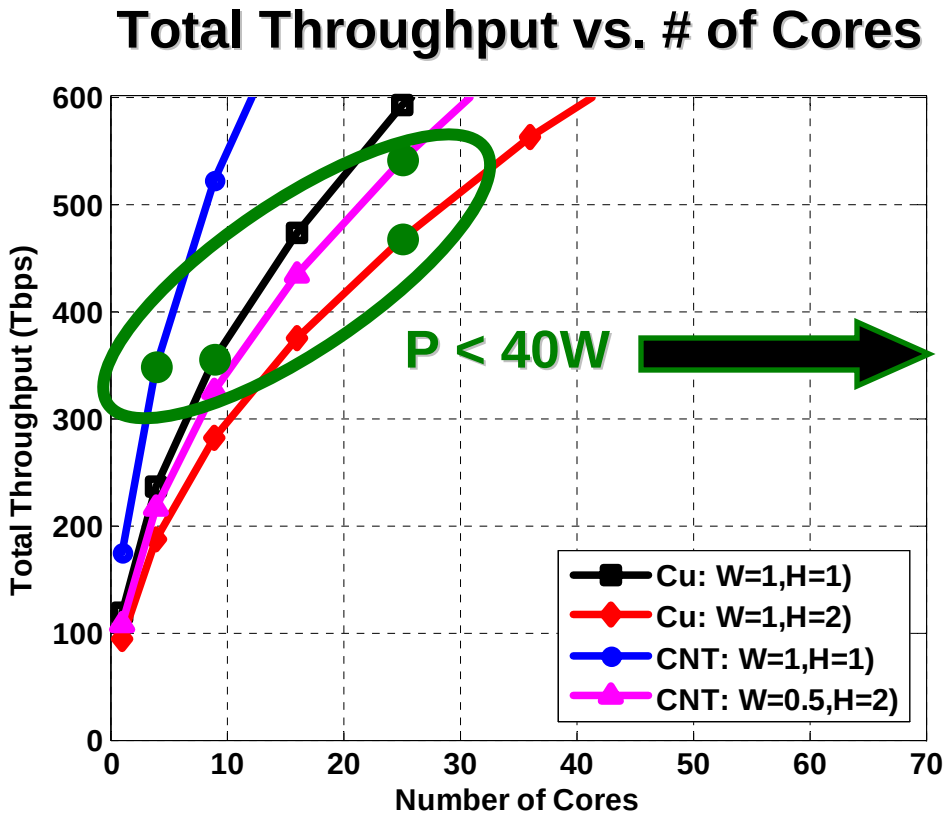


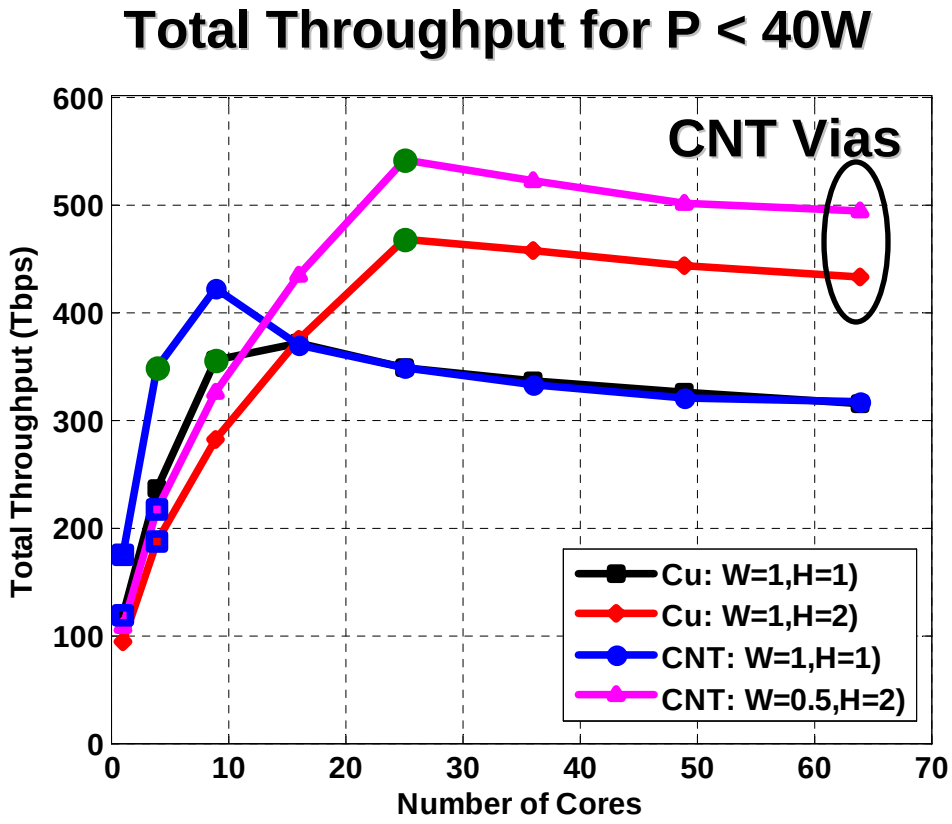
- Total system throughput:

$$T_{sys} = 2\sqrt{N_{core}} \cdot BW_{agg}$$

- Total system throughput monotonically increases, but…

Max. Throughput: Power Constrained





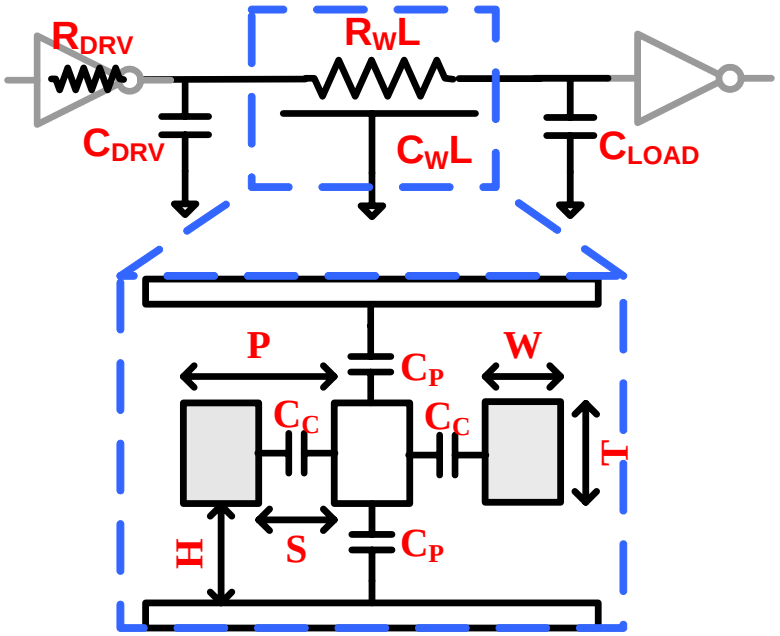
- Throughput saturates with increasing number of cores
- Lower capacitance configurations produce greater power constrained total throughput

Conclusions & Future Work

- Both CNT vias and CNT wires can improve energy/delay of the system (~50% increase in total system throughput)
- To really take advantage of CNTs, ILD stack up needs to be rescaled to proportion R_W & C_W for application needs
- Models need to be verified with measurement data
- Many integration/manufacturing challenges remain

Backup

Interconnect Performance Metric



$$\tau_D = R_{DRV}\left(C_{DRV} + C_{LOAD}\right) + 0.4R_W C_W L^2 + \left(R_{DRV}C_W + R_W C_{LOAD}\right)L$$

$$E_{TOT} = 0.5 \cdot \left(C_{DRV} + C_{LOAD} + C_W L\right) \cdot V_{dd}^2$$

$$C_W = 2\left(C_P + C_C\right)$$

- Compare interconnect materials (Cu & CNT) in the context of a CMOS system
- Use energy and delay of an inverter driven interconnect to evaluate various configurations

EDP for Optimal Delay Repeated Wires