Scaling and Evaluation of Carbon Nanotube Interconnects for VLSI Applications

Fred Chen, Ajay Joshi, Vladimir Stojanović and Anantha Chandrakasan
Technology Design Flow

Create Models for Circuits

Process Characterization
Technology Design Flow

Create Models for Circuits

Extract Tradeoff Curves

Process Characterization

Performance

Area

Power
Technology Design Flow

Create Models for Circuits

Process Characterization

Make Informed Architecture Choices

Extract Tradeoff Curves
Technology Design Flow

Create Models for Circuits

Extract Tradeoff Curves

Make Informed Architecture Choices

Process Characterization
Technology Design Flow

Create Models for Circuits

Process Characterization

Make Informed Architecture Choices

Extract Tradeoff Curves

- Performance
- Area
- Power
Technology Design Flow

Create Models for Circuits

Extract Tradeoff Curves

Make Informed Architecture Choices

CNT Process Characterization

Performance

Power

Area
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*.

\[
R_F = \begin{cases} 
\frac{h}{4e^2} \sim 6.5k\Omega, & L \leq L_0 \\
(\frac{h}{4e^2})(\frac{L}{L_0}), & L > L_0 
\end{cases}
\]

\[
L_K = \frac{h}{2e^2v_F} \sim 16 \text{nH/}\mu\text{m}
\]

\[
C_Q = \frac{2e^2}{h v_F} \sim 100 \text{aF/}\mu\text{m}
\]

\[
C_E = 2\pi\varepsilon/\ln\left(\frac{y}{d}\right) \sim 34 \text{aF/}\mu\text{m},
\]

\[
\varepsilon = 2.8\varepsilon_0, \ y = 97\text{nm}, \ d = 1\text{nm}
\]
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
  

- 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

\[ \rho_{\text{EFF}} = \frac{d^2}{k} \left( \frac{R_F + R_{\text{cont}}}{L} \right) \]
**Effective Resistivity: Non-idealities**

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

\( L > L_0 \) where \( L_0 = C_\lambda d \)

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

**Resistivity Contours, \( \rho_{\text{CNT}}/\rho_{\text{Cu}} \) (22nm)**

- 33% Metallic
- \( \rho_{\text{CNT}}/\rho_{\text{Cu}} \) (22nm)
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 $W \downarrow \rightarrow R_W \uparrow \rightarrow C_W \downarrow$

- Scale $W$, but maintain minimum wire pitch ($P$)
  $\rightarrow$ constant area

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

Cu, 45nm, 8X min. buffer load, $L=250X$, FO=4,
Scaling Wire Width

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

\[ \tau_D = R_{DRV} \left( C_{DRV} + C_{LOAD} \right) + 0.4R_WC_WL^2 + \left( R_{DRV}C_W + R_WC_{LOAD} \right)L \]
Scaling Wire Width

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

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

Range of sub-optimal Cu wire lengths

Range of sub-optimal CNT wire lengths

Min. Width in Cu
Scaling ILD Height

As $H \uparrow \Rightarrow R_W \uparrow \Rightarrow C_W \downarrow$

- Scale $H$, reverse scale $T$ to maintain constant wire bandwidth (for comparison)
- Min. delay ($\sim H=2$) and a min. energy point ($\sim H=4$) exist
Directly replacing Cu with CNTs (same cross-section) only yields delay improvement at lower fanouts.

- FO = 0.5
- FO = 8

22nm node, L=1000X min. wire pitch, H in nominal ILD height, and W in minimum wire widths.
Energy vs. Delay: Scaling Width

- Scaling wire width down improves energy and delay

- Cu Vias + Cu Wires
- Cu Vias + CNT Wires
- CNT, W=1, H=1
- CNT, W=0.5, H=1
- FO = 0.5
- FO = 8
Scaling wire height up using CNT vias, but Cu wires improves energy with small penalty in delay.
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…

\[ L_{SEG} = \frac{L_{TOTAL}}{N} \]

N = # of repeaters
Core-to-Core Communication

As the cores grow: \( BW_{wire} \downarrow \) but \( N_{wire} \uparrow \)

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…
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

- 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

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

\[ + \left( R_{DRV} C_W + R_W C_{LOAD} \right) L \]

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

\[ C_W = 2 \left( C_P + C_C \right) \]
EDP for Optimal Delay Repeated Wires

Cu Vias + Cu Wires

Cu Vias + CNT Wires

CNT Vias + Cu Wires

CNT Vias + CNT Wires