Compact Modeling Aided Technology Design and Projection Considering System-Level Performance

Lan Wei, H.-S. Philip Wong
Stanford University
{lanw, hspwong}@stanford.edu
CMOS Device Pitch Scaling

Contacted Gate Pitch (density, cost)

Gate Length (performance)

Technology Node (nm)

1960's

Data from IEDM, VLSI
Projected by ITRS

L_{gate}: Data from IEDM, VLSI
L_{pitch}: Projected by ITRS

Length (nm)

10^1

10^2

10^3
In the old days...

Proportional scaling → Intrinsic on-state performance improvement → Delay, power, area

Device → Library → Circuit
In the future

Non-scalable factors
- Leakage
- $V_{dd}$ ($V_{th}$)
- On current
- Parasitics – R, C

Device + Circuit
Outline

Near future CMOS Technology
• Gate control vs. current drive

Mid future CMOS Technology
• Parasitic engineering

Far future Post-Si Technology
• Carbon nanotube FET (CNFET)
CMOS Scaling Today

- **Present path**
  - Higher velocity (strain, exotic channel materials Ge, III-V)
  - Good electrostatics for gate length scaling
    - Ultra-thin body SOI, FinFET
    - 1D FETs (nanowires, carbon nanotubes)

Which way is more efficient?

\[
\tau = \frac{CV}{I} = \left( \frac{C_{\text{intrinsic}} + C_{\text{par}}}{I} \right) \cdot V
\]
Device modeling and circuit simulation

Ref: A. Khakifirooz et al, T-ED09
S, Natarajari, et al, IEDM 08.
Efficiency depends on applications

32nm HP

10% speed up

HP: Improve $I_{on}$ is efficient

Transport property: strain/new materials

32nm LstP

10% speed up

Improve $DIBL$ is efficient

Electrostatic property: FDSOI/FinFET/GAA
Efficiency also depends on circuit complexity

INV Improve $I_{on}$ is efficient

Transport property: strain/new materials

NAND/NOR: Improve $DIBL$ is efficient

Electrostatic property:
FDSOI/FinFET/GAA

32nm HP Inverter

32nm HP NOR

$10\%$ speed up

$DIBL (V/V)$

$DIBL (V/V)$
Take-home message #1

The most efficient knob to improve device performance at circuit-level depends on

- Application
- Circuit design
The Future of CMOS Scaling

- In the limit of $L_g \to 0$ and ballistic transport
  - Speed is determined by parasitic capacitance and series resistance
  - Realistic performance estimation must include parasitics
    - A 3D problem, some parasitics do not scale

$$\tau = \frac{CV}{I} = C \cdot R = \left( C_{\text{intrinsic}} + C_{\text{par}} \right) \cdot \left( R_{\text{channel}} + R_{\text{series}} \right)$$

- $< 50\%$
- $> 50\%$
- $< 70\%$
- $> 30\%$
Performance Boosting By Parasitics Engineering

\[ \tau = \frac{CV}{I} = C \cdot R = \left( C_{\text{intrinsic}} + C_{\text{par}} \right) \cdot \left( R_{\text{channel}} + R_{\text{series}} \right) \]

- Gate
- Contact, isolation, spacer, etc.
- Pitch
Performance Boosting By Parasitics Engineering

\[
\tau = \frac{CV}{I} = C \cdot R = (C_{\text{intrinsic}} + C_{\text{par}}) \cdot (R_{\text{channel}} + R_{\text{series}})
\]

Reduce parasitics when:
- Keep scaling \(L_{\text{pitch}}\)
- Relax \(L_{\text{gate}}\) scaling

Gate + Contact, isolation, spacer, etc. = Pitch

50% 50% 70% 30%
Capacitance Components

Ref: L. Wei et al, VLSI-TSA 2009  
T. Skotnicki, IEDM 2009, short course
Capacitance Model

- Implemented in MASTAR5
- Used by ITRS Roadmap 2009 Edition
- Bulk, FDSOI, Double-gate FET, FinFET
- http://www.itrs.net/models/htm

L. Wei et al, VLSI-TSA 2009
Selective Device Structure Scaling: Basic Idea

(a): Reduce $L_{\text{cont}}$

(b): Reduce $H_{\text{plug}}$

(c): Reduce $H_{\text{gate}}$

(d): Reduce $L_{\text{gc}}$

Parasitic capacitance ↓
Series resistance ↑

Parasitic capacitance ↑
Series resistance ↓

$\tau = R \cdot C$ ↓
Inverter Delay Improvement – (bulk CMOS and UTBSOI)

- 15% faster
- 45% smaller area

Devices with conventional layout

Normalised FO4 Delay vs Contacted Gate Pitch $L_{pitch}$ (nm)

(a) + (b) + (c) + (d)
Selectively scaled footprints

L. Wei et al, TED 2009.
Circuit-Level Improvement

- **Conventional layout**
  - 0.10mm²
  - 228mW
  - 4.17G

- **Optimal layout**
  - 0.07mm²
  - 205mW
  - 5.06G

- **Optimal Device**
  - Maximum Frequency: 50% faster
  - Power: 10% smaller

- **Normal Device**
  - Maximum Frequency: 120%
  - Power: 100%

Low-k spacer lowers the gate-fringing capacitance, thus speeds up the ring oscillator.

Take-home message #2

Parasitic engineering is vital for circuit-level performance

- “Selective device structure scaling” can effectively improve device performance at the circuit-level and extend the technology roadmap for a few generations beyond the currently perceived limits.
Carbon Nanotube FET (CNFET)

**TECHNOLOGY**
- Structure innovation
- Material innovation

**DESIGN**
- Optimization
- System-level impact

![Diagram of Carbon Nanotube FET (CNFET)](image)

L. Wei, et al, IEDM 2009
Transport Model

Ballistic transport

Piecewise constant quantum capacitances

Non- iterative model for $\varphi_{ch}$

Highest barrier = max ($\varphi_{ch}$, $\varphi_S$, $\varphi_D$)

Non- iterative model for intrinsic device current
D. Frank et al, IBM J. RES. & DEV. 2006
L. Wei et al, IEDM 2009
For both high performance and low power applications: CNFET is 5x faster.

Device parameters (including $L_{\text{gate}}$) are optimized.
At the optimum performance,

\[
\left( \frac{I_{\text{eff}}}{I_{\text{off}}} \right)_{PDSOI} \approx \left( \frac{I_{\text{eff}}}{I_{\text{off}}} \right)_{\text{CNFET}}
\]

- Conventional way of benchmarking
  - maximizing \( I_{on} \) at fixed \( I_{off} \) and \( V_{dd} \) (not always optimized)
- Optimized design
  - Design criteria depends on application constraints
  - Targeting at the system-level performance

At the optimum performance,

\[
\left( \frac{I_{\text{eff}}}{I_{\text{off}}} \right)_{\text{PDSOI}} \approx \left( \frac{I_{\text{eff}}}{I_{\text{off}}} \right)_{\text{CNFET}}
\]

Chip power=10W, incl. variability

<table>
<thead>
<tr>
<th>L_{\text{gate}}</th>
<th>V_{\text{dd}}</th>
<th>I_{\text{eff}}</th>
<th>I_{\text{eff}}/I_{\text{off}}</th>
<th>C_{\text{load_ave}}</th>
</tr>
</thead>
<tbody>
<tr>
<td>nm</td>
<td>V</td>
<td>mA/\mu m</td>
<td></td>
<td>fF</td>
</tr>
<tr>
<td>PDSOI (optimal)</td>
<td>21nm</td>
<td>0.7V</td>
<td>0.234</td>
<td>456</td>
</tr>
<tr>
<td>CNFET (optimal)</td>
<td>10nm</td>
<td>0.4V</td>
<td>0.62</td>
<td>334</td>
</tr>
</tbody>
</table>

Good short channel performance

Ballistic transport

Shorter/narrower devices and shorter wires
Take-home message #3

System-level optimization is essential for device design and benchmarking.

- Design criteria depends on application constraints
- Targets the system-level performance
Conclusions

- A holistic view is required for device design and engineering
  - Circuit-level performance must be considered
  - Device design is application-orientated
  - Compact modeling facilitates high-speed circuit and system simulation
Conclusions

- A holistic view is required for device design and engineering
  - Circuit-level performance must be considered
  - Device design is application-orientated
  - Compact modeling facilitates high-speed circuit and system simulation

Device design is not a stand-alone task!
Acknowledgement

- **Sponsors and Collaborators**
  - Dr. David. J. Frank, Dr. Leland Chang, Dr. Jie Deng (IBM)
  - Dr. Frédéric Boeuf, Dr. Thomas Skotnicki (STMicroelectronics)
  - Prof. Dimitri Antoniadis (MIT)
  - Dr. Kwok Ng (SRC)

Stanford INMP
(Toshiba, Intel, TI, IBM, AMD, TEL, AMAT, COSAR)
Conclusions

- **A holistic view is required for device design and engineering**
  - Circuit-level performance must be considered
  - Device design is application-orientated
  - Compact modeling facilitates high-speed circuit and system simulation

References:
L. Wei et al, IEDM07 & 09 (Stanford/IBM)
L. Wei et al, VLSI-TSA 09 (Stanford/STMicroelectronics)
L. Wei et at, TED 09 (Stanford/IBM)
L. Wei et al, SSDM 09 (Stanford/STMicroelectronics/MIT)