Futur of Nano CMOS Technology

Hiroshi Iwai

Frontier Research Center
Tokyo Institute of Technology
Background for nano-electronics
1900 “Electronics” started.

Device: Vacuum tube
Device feature size: 10 cm
Major Appl.: Amplifier (Radio, TV, Wireless etc.)

→ Technology Revolution

1970 “Micro-Electronics” started.

Device: Si MOS integrated circuits
Device feature size: 10 µm
Major Appl.: Digital (Computer, PC, etc.)

→ Technology Revolution
2000 “Nano-Electronics” started.

Device: Still, Si CMOS integrated circuits

Device feature size: 100 nm

Major Appl.: Digital (µ-processor, cell phone, etc.)

→ Technology Revolution??

Maybe, just evolution or innovation!

But very important so many innovations!
Now, 2013 “Nano-Electronics” continued.

Device: Still, Si CMOS integrated circuits
Device feature size: near 10 nm
Major Appl.: Still Digital (μ-processor, cell phone, etc.)

Still evolution and innovation..

But, so many important emerging applications for smart society.
Questions for future

Future, “Nano-Electronics” still continued?

Device: Still, Si CMOS integrated circuits?
Device feature size: ? nm, what is the limit?

Application: New application?

→ Any Technology Revolution?
What is special or new for Nano-Electronics?

In 1990’s, people expected completely new mechanism or operational principle due the nano size, like quantum mechanical effects.

However, no fancy new operational principle was found.

At least for logic application, there is no success story for “Beyond CMOS devices” to replace Si-CMOS.

Of course, I do not deny the importance of Beyond CMOS technology development. It is becoming very important as CMOS approach its limit.
1. Background for nano-CMOS
First Computer Eniac: made of huge number of vacuum tubes 1946
Big size, huge power, short life time filament

→ dreamed of replacing vacuum tube with solid-state device

Today’s pocket PC made of semiconductor has much higher performance with extremely low power consumption
1960: First MOSFET by D. Kahng and M. Atalla

Top View

Si/\text{SiO}_2\text{ Interface is extraordinarily good}
1970,71: 1st generation of LSIs

DRAM   Intel 1103

MPU    Intel 4004
In 2012

Most Recent SD Card

128GB (Bite)
= 128G X 8bit
= 1T(Tera)bit

1T = 10^{12} = 1 Trillion

World Population: 7 Billion
Brain Cell: 10~100 Billion
Stars in Galaxy: 100 Billion
128 GB = 1Tbit

2.4cm × 3.2cm × 0.21cm

Volume: 1.6 cm³  Weight: 2g

Voltage: 2.7 - 3.6V

Old Vacuum Tube:
5cm × 5cm × 10cm, 100g, 50W

What are volume, weight, power consumption for 1Tbit
Old Vacuum Tube: 5cm X 5cm X 10cm

Volume = (5cm X 10,000) X (5cm X 10,000) X (10cm X 10,000)
= 0.5km X 0.5km X 1km

1Tbit = 10,000 X 10,000 X 10,000 bit

Burj Khalifa
Dubai, UAE
(Year 2010)

828 m

Indian Tower
Mumbai, India
(Year 2016)

700 m

Pingan International Finance Center
Shanghai, China
(Year 2016)

700 m

500 m
Old Vacuum Tube: 50W

Nuclear Power Generator

1MkW = 1BW

1Tbit = $10^{12}$ bit

Power = $0.05kWX10^{12}$ = 50 TW

We need 50,000 Nuclear Power Plant for just one 128 GB memory

In Japan we have only 54 Nuclear Power Generator

Last summer Tokyo Electric Power Company (TEPCO) can supply only 55BW.

We need 1000 TEPCO just one 128 GB memory

Imagine how many memories are used in the world!
So progress of integrated circuits is extremely important for power saving
Brain: Integrated Circuits
Ear, Eye: Sensor
Mouth: RF/Opto device
Stomach: PV device
Hands, Legs: Power device
Near future smart-society has to treat huge data.

Demand to high-performance and low power CMOS become much more stronger.
Semiconductor Device Market will grow 5 times in 12 years, even though, it is very matured market!!

\[ \text{2011} \quad \rightarrow \quad \text{2025} \]

\[ \text{300B USD} \quad \rightarrow \quad \text{1,500B USD} \]

Gartner: By K. Kim, CSTIC 2012
2. Current status of Si-CMOS device technologies
Downsizing

→ Decreasing size

→ Decreasing capacitance

Thus, important for

Decreasing cost, power

Increasing performance
Feature Size / Technology Node

(1970) 10 \(\mu\) m \(\rightarrow\) 8 \(\mu\) m \(\rightarrow\) 6 \(\mu\) m \(\rightarrow\) 4 \(\mu\) m \(\rightarrow\) 3 \(\mu\) m \(\rightarrow\) 2 \(\mu\) m \(\rightarrow\) 1.2 \(\mu\) m

\(\rightarrow\)

0.8 \(\mu\) m \(\rightarrow\) 0.5 \(\mu\) m \(\rightarrow\) 0.35 \(\mu\) m \(\rightarrow\) 0.25 \(\mu\) m \(\rightarrow\) 180 nm \(\rightarrow\) 130 nm

\(\rightarrow\)

90 nm \(\rightarrow\) 65 nm \(\rightarrow\) 45 nm \(\rightarrow\) 32 nm \(\rightarrow\) (28 nm \(\rightarrow\)) 22 nm (2012)

From 1970 to 2013 (Last year)

- 43 years 1 generation
- 18 generations 2.5 years
- Line width: 1/450 Line width: 1/1.43 = 0.70
- Area: 1/200,000 Area: 1/2 = 0.5
Question

What is the problem problem for downsizing?
The problem for downsizing

S and D distance small

→ Ion & Ioff increase

\( I_{\text{off}} \) increase: Transistor cannot be turned-off.

\( I_{\text{off}} \) (Off-leakage current) between S and D

1. Punch-through between S and D
2. Direct-tunneling between S and D
3. Subthreshold current between S and D
1. Punch-through between S and D
Problem for downsizing

Region governed by gate bias
Region governed by drain bias

Gate oxide
Gate metal

Source
Drain
Channel
Substrate

0V < \( V_{\text{dep}} < 1V \)

Depletion Region (DL)
by Drain Bias

DL touch with S Region (DL)

Large \( I_{\text{OFF}} \)

Large \( I_{\text{OFF}} \)
(Electron current)

No \( t_{\text{ox}}, V_{\text{dd}} \) thinning

\( t_{\text{ox}}, V_{\text{dd}} \) thinning

\( t_{\text{ox}} \) and \( V_{\text{dd}} \) have to be decreased for better channel potential control \( \rightarrow I_{\text{OFF}} \) Suppression
1. Punch-through between S and D

There are solutions to suppress the depletion layer

1. Decrease supply voltage → Very difficult

   as explained later

2. Decrease tox to enhance the channel potential controllability by gate bias

3. Gate/channel configuration change to enhance the channel potential controllability by gate bias

   Fin-FET, ET-SOI, etc.
Decrease tox

\[ \rightarrow \text{ Increase the Electric field between Gate & Channel} \]

\[ \rightarrow \text{ Increase the channel potential controllability by gate bias.} \]

\[ \rightarrow \text{ Keep channel potential 0V} \]

\[ \rightarrow \text{ Suppress the depletion layer} \]
$L_{\text{gate}}$ and $t_{\text{ox}}$(EOT) scaling trend

A. Toriumi (Tokyo Univ), IEDM 2006, Short Course
Configuration change for channel and gate structures for better control of channel potential.

Fin-FET, ET-SOI, etc.
Extremely Thin (or Fully-Depleted) SOI

- Make Si layer thin
- Control channel potential also from the bottom

Planar

ET (or FD) SOI
Surrounding gate structure (Multiple gates)

- Make Si layer thin
- Control channel potential also by multiple gates not only from top & bottom but maybe also from side

Planar

Multi gate

Drain bias induced depletion

Si fin or nanowire
Nanowire structures in a wide meaning

Fin
Tri-gate
Tri-gate (Variation)
Ω-gate
All-around
Our work at TIT: Ω-gate Si Nanowire

S. Sato et al., pp.361, ESSDERC2010 (Tokyo Tech.)

- Conventional CMOS process
- High drive current
  (1.32 mA/µm @ $I_{OFF}=117$ nA/µm)
- DIBL of 62mV/V and SS of 70mV/dec for nFET
2. Direct-tunneling between S and D

Wave function of electron penetrates the channel potential barriers by quantum mechanical physics, when the channel length is around 3 nm.
Tunneling distance 3 nm

There is no solutions!

Direct-tunnel current

Energy or Potential for Electron

Source Channel Drain

Tunneling distance

Downsizing limit is @ Lg = 3 nm.
3. Subthreshold current between S and D
Subthreshold leakage current of MOSFET

Subthreshold leakage current:

\[ I_{on} \propto e^{aV_g} \]

Threshold voltage (Vth)

(Vg=0V)

Subthreshold region

(Threshold Voltage)
Subthreshold leakage current

![Graph showing subthreshold leakage current with logarithmic scales for voltage and current.]

- $I_{on}$: Current at $V_g = 0$.
- $I_{off}$: Current at $V_g = 1$.
- $I_d$: Drain current.
- $V_d$, $V_g$: Gate and drain voltages.
- $V_{th}$: Threshold voltage.

**Electron Energy**

Boltzmann statics

Exp ($qV/kT$)

- $L_g \rightarrow 1/2$
- $V_d, V_g \rightarrow 1/2$
- $V_{th} \rightarrow 1/2$

However

$I_{off} \rightarrow 10^3$ in this example

Because of log-linear dependence
Subthreshold leakage current of MOSFET

Subthreshold leakage current of MOSFET is OK at Single Tr. level but not OK for Billions of Trs.

Subthreshold Leakage Current

\[ I_{off} \propto \exp(aV_g) \]

\[ V_g = 0V \]

Subthreshold region

\[ I_{on} \propto V_g - V_{th} \]

(Threshold Voltage)

Subthreshold Current

ON

OFF
3. Subthreshold current between S and D

Solution: however very difficult

Keep Vth as high as possible

- Do not decrease supply voltage, Vd
  → However, punchthrough enhanced

- Suppress variability in Vth

Thus, subthreshold current will limit the downsizing, especially for mobile devices
The limit is different depending on application.
How far can we go for production?

**Past** 0.7 times per 2.5 years

10µm → 8µm → 6µm → 4µm → 3µm → 2µm → 1.2µm → 0.8µm → 0.5µm → 0.35µm → 0.25µm → 180nm → 130nm → 90nm → 65nm → 45nm → 32nm

---

**Limit depending on applications**

**Subthreshold punchthrough**

**Direct-tunnel**

**Fundamental limit**

**Intermediate node**

(28nm) → 22nm → 14nm → 11.5 nm → 8nm → 5.5nm? → 4nm? → 2.9 nm? →

- At least 4,5 generations to 8 ~ 5 nm
However, careful about the name of technology!

Recently, Gate length (Lg) is much larger than the Technology name

**22 nm** Technology by Intel

Lg (Gate length) = 30 nm (HP), 34 nm (MP), 34 nm or larger (SP)

IEDM 2012, VLSI 2013

**10 nm** Technology by Leti (FD-SOI)

Lg (Gate length) = 15 nm

ECS 2013
The rate for the shrinkage for the gate length and line pitch will be larger than 0.7 in near future, because of the subthreshold leakage, and also because of the delay in EUV lithography.

As a result, we will have more technology generations until reaching the downsizing limit, and the time to reach the limit will be delayed.
Tri-gate implementation for transistors

C. Auth et al., pp.131, VLSI2012 (Intel)

<table>
<thead>
<tr>
<th></th>
<th>HP</th>
<th>MP</th>
<th>SP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{OX,E}$ (nm)</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>$L_{GATE}$ (nm)</td>
<td>30</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>$I_{OFF}$ (nA/um)</td>
<td>20-100</td>
<td>5-20</td>
<td>1-5</td>
</tr>
</tbody>
</table>

Tri-gate has been implemented since 22nm node, enabling further scaling
How far can we go for production?

Thus, we may go down to “5 nm” technology by choosing whatever gate length we want for the application.
More Moore to More More Moore

Technology node

<table>
<thead>
<tr>
<th>65nm</th>
<th>45nm</th>
<th>32nm</th>
<th>22nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>L_g 35nm</td>
<td>L_g 30nm</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Now

Future

15nm, 11nm, 8nm, 5nm, 3nm

(Fin, Tri, Nanowire)

Si

Planar

28nm

TRi-Gate

Emerging Devices

Si channel

ET: Extremely Thin

Si is still main stream for future !!

M. Bohr, pp.1, IEDM2011 (Intel)
P. Packan, pp.659, IEDM2009 (Intel)
C. Auth et al., pp.131, VLSI 2012 (Intel)
T. B. Hook, pp.115, IEDM2011 (IBM)
S. Bangsaruntip et al., pp.297, IEDM2009 (IBM)
# High-k gate dielectrics

## Hf-based oxides

<table>
<thead>
<tr>
<th>Thickness</th>
<th>EOT Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>45nm</td>
<td>1nm</td>
</tr>
<tr>
<td>32nm</td>
<td>0.95nm</td>
</tr>
<tr>
<td>22nm</td>
<td>0.9nm</td>
</tr>
<tr>
<td>15nm, 11nm, 8nm, 5nm, 3nm,</td>
<td></td>
</tr>
</tbody>
</table>

SiO$_2$ IL (Interfacial Layer) is used at Si interface to realize good mobility.

Technology for direct contact of high-k and Si is necessary.

- EOT = 0.52 nm
- Remote SiO$_2$-IL scavenging HfO$_2$ (IBM)

### Continued research and development

- T.C. Chen, et al., p.8, VLSI 2009, (IBM)
- T. Ando, et al., p.423, IEDM2009, (IBM)

### Hf-based oxides

- TiN
- HfO$_2$
- SiO$_2$
- Si

EOT = 0.9nm
HfO$_2$/SiO$_2$
(IBM)

### EOT = 0.37nm, 0.40nm, 0.48nm

- 0.48 $\rightarrow$ 0.37nm Increase of $I_d$ at 30%

Direct contact with La-silicate (Tokyo Tech.)
How far can we go for production?

Rather than Ioff value, Ion/Ioff ratio is important.

Now, Ion/Ioff ratio is typically $10^6$. However, it degrades significantly with decrease in Vsupply.
\textbf{I}_{\text{ON}} \text{ and } I_{\text{OFF}} \text{ benchmark}

\begin{align*}
\text{NMOS} & \\
& \\
& \\
& \\
& \\
\end{align*}

\begin{align*}
I_{\text{OFF}} \text{ [nA/\mu m]} & \\
& \\
& \\
& \\
& \\
\end{align*}

\begin{align*}
I_{\text{ON}} \text{ [mA/\mu m]} & \\
& \\
& \\
& \\
& \\
\end{align*}

\begin{align*}
\text{PMOS} & \\
& \\
& \\
& \\
& \\
\end{align*}

\begin{align*}
I_{\text{OFF}} \text{ [nA/\mu m]} & \\
& \\
& \\
& \\
& \\
\end{align*}

\begin{align*}
I_{\text{ON}} \text{ [mA/\mu m]} & \\
& \\
& \\
& \\
& \\
\end{align*}

\begin{align*}
\text{a] C. Auth et al., pp.131, VLSI2012 (Intel).} & \quad \text{f] T. Yamashita et al., pp.14, VLSI2011 (IBM).} \\
\text{b] K. Mistry et al., pp.247, IEDM2007 (Intel).} & \quad \text{g] A. Khakifirooz et al., pp.117, VLSI2012 (IBM).} \\
\text{c] H.-J. Cho et al., pp.350, IEDM2011 (Samsung).} & \quad \text{h] G. Bidal et al., pp.240, VLSI2009 (STMicroelectronics).} \\
\text{d] S. Saitoh et al., pp.11, VLSI2012 (Toshiba).} & \quad \text{i] S. Sato et al., pp.361, ESSDERC2010 (Tokyo Tech.)} \\
\text{e] S. Bangsaruntip et al., pp.297, IEDM2009 (IBM).} & \quad \text{j] K. Cheng et al., pp.419, IEDM2012 (IBM).}
\end{align*}
Nanowire/Tri gate MOSFETs have advantage not only suppressing Ioff, but also for increasing Ion over planer MOSFETs

1. Because of higher mobility due to lower vertical electric field

2. Because of higher carrier density at the round corner
Electron Density
\( \times 10^{19} \text{cm}^{-3} \)

Distance from SiNW Surface (nm)

- **Edge portion**
- **Flat portion**

**Inversion areal ratio:** 29%
SiNW Band structure calculation
Cross section of Si NW

First principal calculation,
## Si nanowire FET with 1D Transport

<table>
<thead>
<tr>
<th>Orientation</th>
<th>[001]</th>
<th>[011]</th>
<th>[111]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diameter (nm)</td>
<td>0.86</td>
<td>0.94</td>
<td>0.89</td>
</tr>
</tbody>
</table>

### Wave Number

**G Z G** for [001]

**G Z Z** for [011] and [111]

**Energy (eV)**

-0.5 0.5

(a)  

**Small mass with [011]**

Large number of quantum channels with [001]
Atomic models of a Si quantum dot and Si nanowires

6.6 nm diameter SiQD
(8651 atoms)

10 nm diameter Si(100)NW
(2341 atoms)

20 nm diameter Si(100)NW
(8941 atoms)
RSDFT – suitable for parallel first-principles calculation -

- Real-space Finite Difference
- Sparse Matrix
- FFT free (FFT is inevitable in the conventional plane-wave code)

MPI (Message Passing Interface) library

Kohn-Sham eq. (finite-difference)

\[
\left(-\frac{1}{2}\nabla^2 + v_s[\rho](\mathbf{r}) + \hat{v}^{PP}_{nloc}(\mathbf{r})\right)\phi_n(\mathbf{r}) = \varepsilon_n\phi_n(\mathbf{r})
\]

Higher-order finite difference

\[
\frac{\partial^2}{\partial x^2} \psi_n(x, y, z) \approx \sum_{m=-6}^{6} C_m \psi_n(x + m\Delta x, y, z)
\]

Integration

\[
\int \psi_m(\mathbf{r})\psi_n(\mathbf{r})d\mathbf{r} \approx \sum_{i=1}^{\text{Mesh}} \psi_m(\mathbf{r}_i)\psi_n(\mathbf{r}_i)\Delta x\Delta y\Delta z
\]

3D grid is divided by several regions for parallel computation.

MPI_ISEND, MPI_IRECV

MPI_ALLREDUCE

Higher-order finite difference pseudopotential method
Massively Parallel Computing

with our recently developed code “RSDFT”

J. R. Chelikowsky et al., PRB1994

Highly tuned for massively parallel computers

Based on the finite-difference pseudopotential method (J. R. Chelikowsky et al., PRB1994)

Computations are done on a massively-parallel cluster PACS-CS at University of Tsukuba.

(Theoretical Peak Performance = 5.6GFLOPS/node)

Iwata et al. J. Comp. Phvs.. to be published.

Real-Space Density-Functional Theory code (RSDFT)

e.g.) The system over 10,000 atoms → Si_{10701}H_{1996}

(7.6 nm diameter Si dot)

Grid points = 3,402,059
Bands = 22,432

Convergence behavior for Si_{10701}H_{1996}

Computational Time (with 1024 nodes of PACS-CS)

6781 sec. × 60 iteration step = 113 hour
Structure and DOS of Si(100)NWs (D=1nm, 4nm, and 8nm)

D=1 nm
Si21H20 (41 atoms)
KS band gap = 2.60 eV

D = 4 nm
Si341H84 (425 atoms)
KS band gap = 0.81 eV

D=8 nm
Si1361H164 (1525 atoms)
KS band gap = 0.61 eV

KS band gap of bulk (LDA) = 0.53 eV
Band structure of 8-nm-diameter Si nanowire near the CBM

\[ \text{band gap} = 0.608 \text{ eV } (\@\Gamma) \]

Each band is 4-dgenerate.

Effective mass equation

\[ \left[ -\frac{\hbar^2}{2m^*} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) - \frac{\hbar^2}{2m^*} \frac{\partial^2}{\partial z^2} \right] \Phi(r) = (\varepsilon - \varepsilon_{CBM}) \Phi(r) \]

The band structure can be understood that electrons near the CBM in the bulk Si are confined within a cylindrical geometry.
Si nano wire with surface roughness

Si12822H1544 (14,366 atoms)
  • 10nm diameter, 3.3nm height, (100)
    • Grid spacing: 0.45Å (~14Ry)
    • # of grid points: 4,718,592
    • # of bands: 29,024
  • Memory: 1,022GB ~ 2,044GB
SiNW Band compact model
Landauer Formalism for Ballistic FET

\[ I_D = \frac{q}{\pi \hbar} \sum_i \int \left[ f(E, \mu_S) - f(E, \mu_D) \right] T_i(E) dE \]

From \( x_{\text{max}} \) to \( x_{\text{min}} \) \( T_i(E) \approx 1 \)

\[ I_D = G_0 \left( \frac{k_B T}{q} \right) \sum_i g_i \ln \left\{ \frac{1 + \exp\left[ (\mu_S - E_{i0}) / k_B T \right]}{1 + \exp\left[ (\mu_D - E_{i0}) / k_B T \right]} \right\} \]
IV Characteristics of Ballistic SiNW FET

Small temperature dependency
35μA/wire for 4 quantum channels
Model of Carrier Scattering

Linear Potential Approx. : Electric Field $E$

Translation Probability to Drain

$T(\varepsilon) = \frac{F(0) - G(0)}{F(0)}$

Injection from Drain = 0
Résumé of the Compact Model

\[ I = \frac{q}{\pi \hbar} \sum_i g_i \int \left[ f(\varepsilon, \mu_s) - f(\varepsilon, \mu_D) \right] T_i \, d\varepsilon \]

\[ (V_G - V_t) - \alpha \frac{\mu_s - \mu_0}{q} = \frac{|Q_f + Q_b|}{C_G} \]

\[ \mu_s - \mu_D = qV_D \]

\[ C_G = \frac{2\pi \varepsilon_{ox}}{\ln\left( \frac{\sqrt{2r + t_{ox}} + \sqrt{t_{ox}}}{\sqrt{2r + t_{ox}} - \sqrt{t_{ox}}} \right)} \]

\[ C_G = \frac{2\pi \varepsilon_{ox}}{\ln\left( \frac{r + t_{ox}}{r} \right)} \]

Unknowns are \( I_D, (\mu_s-\mu_0), (\mu_D-\mu_0), (Q_f+Q_b) \)
I-V<sub>D</sub> Characteritics (RT)

- Electric current: 20~25 µA
- No saturation at Large <i>V_D</i>
Size and Corner Effects on Electron Mobility of Rectangular Silicon Nanowire MOSFETs
Introduction

Experimental results showed a drastic mobility increase in width < 20 nm

Expectation in facet-driven operation

To analyze deviation from facet-driven operation,
We investigate size and corner effects in the rectangular Si NW MOSFETs.
Simulation methods

Using the parabolic EMA, we self-consistently solved 2D Schrödinger and Poisson equations in equilibrium

- Subband structure, $E_\mu$, and wave function, $\Psi_\mu$,

Using the Kubo-Greenwood formula (low-field limit), we calculated phonon-scattering-limited mobility (R. Kotlyar et al., APL, 2004 / S. Jin et al., JAP, 2007)

$$
\mu_\mu = \frac{q}{n_\mu k_B T} \int_{E_\mu}^\infty \tau_\mu (E) v_\mu^2 (E) \rho_\mu (E) f (E) [1 - f (E)] dE
$$

where

Intrasubband scattering rate

$$
\tau^{ac}_\mu (E) = \frac{\pi \Xi^2 k_B T}{g_v \hbar \rho_{Si} u_j^2} \sum_{\mu'} \delta_{\eta,\eta'} \rho_{\mu'} (E) F_{\mu,\mu'}
$$

Intersubband scattering rate

$$
\tau^j_\mu (E) = \frac{\pi (D_j K_j)^2}{2g_v \rho_{Si} \omega_j} \sum_{\mu} g^{j}_{\eta,\eta'} \rho_{\mu'} (E \pm \hbar \omega_j) F_{\mu,\mu'} \frac{1 - f (E \pm \hbar \omega_j)}{1 - f (E)} \left( N_j + \frac{1}{2} \mp \frac{1}{2} \right)
$$

Form factor

$$
F_{\mu,\mu'} = \iint |\Psi_\mu (x, y)|^2 |\Psi_{\mu'} (x, y)|^2 \, dx \, dy
$$
Materials

Rectangular Si NW MOSFETs with various orientations

To take into account arbitrary direction, effective mass tensor needs to be re-expressed in each coordinate system (M. Bescond et al., Nanotechnol., 2007)

Example effective mass tensor:

Red valley of [100]/(100) Si NW

Constant energy surface

\[
M^{-1} = \begin{pmatrix}
\frac{1}{m_x} & 0 & 0 \\
0 & \frac{1}{m_y} & 0 \\
0 & 0 & \frac{1}{m_z}
\end{pmatrix}
\]
Result 1: Cross-sectional electron density

- Corner and side electron distributions hardly depend on size when \( w > 6 \) nm
Result 2: Size-dependent mobility

Mobility does not drastically change in width larger than 10 nm (we could not explain the drastic mobility increase in the experimental results).

- When \( w < 6 \) nm, mobility is drastically changed by the increase in form factor and change of electron structure.

Now, we discuss the corner effect based on specially resolved mobility analysis.
Spatially resolved mobility analysis

\[ \mu_{\text{local}}(x, y) = \frac{\sum_{\mu} |\Psi_{\mu}(x, y)|^2 n_{\mu,\mu}}{n(x, y)} \]

where

\[ n(x, y) = \sum_{\mu} |\Psi_{\mu}(x, y)|^2 n_{\mu} \]

- Mobility for each subband, \( \mu_{\mu} \), is spatially independent

- Local mobility shows spatially different contribution of each subband
Result 3: Cross-sectional spatially resolved mobility

- We can distinguish orientation and corner effects based on the spatially resolved mobility analysis
- Corner mobility is always lower than the (100)-surface mobility.
Discussion 1: Corner effect

Electron density

Local mobility

Subband occupancy ratio

Velocity & scattering rate

The large ratio of corner electrons belongs to the 2\textsuperscript{nd} Subband, which shows lower velocity and higher scattering rate.

Corner mobility is lower than side mobility.
Discussion 2: Origin of different contribution between subbands

Subband occupancy ratio

The large ratio of corner electrons belongs to the 2nd Subband group.

Probability density

Probability density of the 2nd subband group with high energy is localized at corner because wave function within rapidly varying potential at corner could be represented by high energy plane wave function basis.
Velocity and Form Factor

Trade-off in width dependence

- Velocity increases with occupancy of the ground subband with small effective mass
- Scattering rate increases with increasing form factor
We investigated the size-dependent phonon-scattering-limited mobility of rectangular Si NW MOSFETs

The drastic increase in mobility from experimental results could not be explained by size and corner effects (strain could cause the drastic increase in mobility)

When $w < 6$ nm, the mobility drastically modulated

According to the specially resolved mobility analysis, the corner mobility was lower than the side mobility

Paper with respect to a part of this

Modeling of Quasi-Ballistic Transport in Nanowire MOFSETs

A part of this work has been conducted with group of Prof. Giorgio Baccarani at University of Bologna
Quasi-ballistic transport

As the MOSFET is extremely downscaled, diffusive transport changes to quasi-ballistic transport

K. Natori (JAP, 1994) modeled the ballistic MOSFET operation

M. Lundstrom (EDL, 1997) modeled the quasi-ballistic transport with the $kT$ layer

$$R = \frac{L_{kT}}{\lambda_0 + L_{kT}}$$

R: Backscattering coefficient
$\lambda_0$: Mean free path
$L_{kT}$: $kT$-layer length

(However, the scattering theory with $kT$ layer has not been clearly postulated)

E. Gnani et al. (TED, 2008) and K. Natori (JJAP, 2009) modeled the quasi-ballistic transport with only elastic scattering

K. Natori (JJAP, 2009 / TED, 2011) modeled the quasi-ballistic transport both with elastic and inelastic scatterings

We develop a comprehensive model of the quasi-ballistic transport based on Natori’s model
Natori’s model for quasi-ballistic transport

Dividing the channel into two zones: elastic zone and relaxation zone

Scattering matrix

\[
\begin{pmatrix}
    f^+(z_e) \\
    f^-(z_e)
\end{pmatrix}
= \frac{1}{1 - R_1} \begin{pmatrix}
    1 - 2R_1 & R_1 \\
    -R_1 & 1
\end{pmatrix}
\begin{pmatrix}
    f^+(0) \\
    f^-(0)
\end{pmatrix}
\]

\[
R(\varepsilon) = \frac{f^-(0, \varepsilon)}{f^+(0, \varepsilon)} = \frac{R_1(\varepsilon) + [1 - 2R_1(\varepsilon)]}{1 - R_1(\varepsilon)} \frac{f^-(z_e, \varepsilon)}{f^+(z_e, \varepsilon)}
\]

where (from BTE)

\[
R_1(\varepsilon) = \frac{\int_0^{z_e} \frac{1}{\lambda_{ac}(z, \varepsilon)} dz}{1 + \int_0^{z_e} \frac{1}{\lambda_{ac}(z, \varepsilon)} dz}
\]

\[
\frac{f^-(z_e, \varepsilon)}{f^+(z_e, \varepsilon)} \approx \sqrt{\frac{S_{op}}{S_{ac}} - \frac{S_{op}^2}{S_{ac}^2}}
\]

(Endless drain)
Concept of this study

Dividing the device into five zones:
source zone, barrier zone, elastic zone, relaxation zone, drain zone

- Si NW MOSFET is adopted to consider 1D transport
- Adopted elastic scattering is intravalley acoustic phonon (AP) scattering
- Adopted inelastic scattering is g3- and f3-type optical phonon (OP) scattering
  - OP energies are set to 63 meV
One-flux scattering matrix

\[
\begin{pmatrix}
  f^+(z_2) \\
  f^-(z_1)
\end{pmatrix} = \begin{pmatrix} T_1 & R_2 \\ R_1 & T_2 \end{pmatrix} \begin{pmatrix} f^+(z_1) \\
  f^-(z_2)\end{pmatrix}
\]

\[
\begin{pmatrix}
  f^+(z_2) \\
  f^-(z_2)
\end{pmatrix} = \frac{1}{T_2} \begin{pmatrix} T_1 T_2 - R_1 R_2 & R_2 \\ -R_1 & 1 \end{pmatrix} \begin{pmatrix} f^+(z_1) \\
  f^-(z_1)\end{pmatrix}
\]
Expression with one-flux scattering matrices

\[ I = \frac{q}{\pi \hbar} \int [f^+(z_0, \varepsilon) - f^-(z_0, \varepsilon)] d\varepsilon \]

Flux between 0 and \( z_0 \)

\[
\begin{pmatrix}
  f^+(z_0) \\
  f^-(z_0)
\end{pmatrix} = \frac{1}{T_{e_2} T_{b_2}} \begin{pmatrix}
  T_{b_1} T_{b_2} - R_{b_1} R_{b_2} & R_{b_2} \\
  -R_{bl} & 1
\end{pmatrix} \begin{pmatrix}
  T_{e_1} T_{e_2} - R_{e_1} R_{e_2} & R_{e_2} \\
  -R_{el} & 1
\end{pmatrix} \begin{pmatrix}
  f^+ & f^-(0)
\end{pmatrix}
\]

Flux between \( z_0 \) and \( z_d \)

\[
\begin{pmatrix}
  f^+(z_d) \\
  f_D
\end{pmatrix} = \frac{1}{T_{e_2} T_{e_2} T_{d_2}} \begin{pmatrix}
  T_{d_1} T_{d_2} - R_{d_1} R_{d_2} & R_{d_2} \\
  -R_{dl} & 1
\end{pmatrix} \begin{pmatrix}
  T_{e_1} T_{e_2} - R_{e_1} R_{e_2} & R_{e_2} \\
  -R_{el} & 1
\end{pmatrix} \begin{pmatrix}
  f^+ & f^-(z_0)
\end{pmatrix}
\]

Simultaneously solving them, \( f^+(z_0) \) and \( f(z_0) \) are described by \( R_S, T_S, f_S, \) and \( f_D \)

\[
\begin{align*}
  f^+(z_0) &= T_S f_S + R_S f^-(z_0) \\
  f^-(z_0) &= \frac{T_S R}{1 - R_S R} f_S + \frac{T}{1 - R_S R} f_D
\end{align*}
\]

where

\[
\begin{align*}
  R_S &= \frac{R_{b_2} - R_{e_2} (T_{b_1} T_{b_2} - R_{b_1} R_{b_2})}{1 - R_{e_2} R_{b_1}} \\
  T_S &= \frac{T_{e_1} T_{b_1}}{1 - R_{e_2} R_{b_1}} \\
  T &= \frac{T_{e_2} T_{r_2} T_{d_2}}{1 - R_{e_1} R_{r_1} - R_{r_2} R_{d_1} - R_{e_1} R_{d_1} (T_{r_1} T_{r_2} - R_{r_1} R_{r_2})}
\end{align*}
\]

\[
\begin{align*}
  R &= \frac{R_{e_1} (1 - R_{r_2} R_{d_1}) + R_{r_1} (T_{e_1} T_{e_2} - R_{e_1} R_{e_2}) + (T_{e_1} T_{e_2} - R_{e_1} R_{e_2}) (T_{r_1} T_{r_2} - R_{r_1} R_{r_2})}{1 - R_{e_1} R_{r_1} - R_{r_2} R_{d_1} - R_{e_1} R_{d_1} (T_{r_1} T_{r_2} - R_{r_1} R_{r_2})}
\end{align*}
\]

Solving BTE for each zone, we can derive \( R_S \) and \( T_S \)
**R_{b1,2} and T_{b1,2} for barrier zone**

\[
\frac{df^+(z, \varepsilon)}{dz} = \frac{1}{\lambda_{ac}(z, \varepsilon)} f^+(z, \varepsilon) - \frac{1}{\lambda_{ac}(z, \varepsilon)} f^-(z, \varepsilon)
\]

\[
\frac{df^-(z, \varepsilon)}{dz} = \frac{1}{\lambda_{ac}(z, \varepsilon)} f^-(z, \varepsilon) - \frac{1}{\lambda_{ac}(z, \varepsilon)} f^+(z, \varepsilon)
\]

where

\[
\frac{1}{\lambda_{ac}(z, \varepsilon)} \equiv \frac{1}{2v(z, \varepsilon) \tau_{ac}(z, \varepsilon)} = \frac{mS_{ac}}{4\pi \hbar [E_1(z_0) + \varepsilon - E_1(z)]}
\]

\[
R_{b1}(\varepsilon) = R_{b2}(\varepsilon) = \frac{\int_{z_s}^{z_0} \frac{1}{\lambda_{ac}(z, \varepsilon)} \, dz}{1 + \int_{z_s}^{z_0} \frac{1}{\lambda_{ac}(z, \varepsilon)} \, dz}
\]

\[
T_{b1}(\varepsilon) = T_{b2}(\varepsilon) = 1 - R_{b1}(\varepsilon)
\]

If potential is given, we can calculate R and T.
**R_{e1,2} and T_{e1,2} for elastic zone**

BTE

\[
\begin{align*}
\frac{df^+(z,\varepsilon)}{dz} &= \frac{1}{\lambda_{uc}(z,\varepsilon)} f^+(z,\varepsilon) - \frac{1}{\lambda_{uc}(z,\varepsilon)} f^-(z,\varepsilon) \\
\frac{df^-(z,\varepsilon)}{dz} &= \frac{1}{\lambda_{uc}(z,\varepsilon)} f^-(z,\varepsilon) - \frac{1}{\lambda_{uc}(z,\varepsilon)} f^+(z,\varepsilon)
\end{align*}
\]

where

\[
\frac{1}{\lambda_{uc}(z,\varepsilon)} = \frac{1}{2\nu(z,\varepsilon)\tau_{ac}(z,\varepsilon)} = \frac{mS_{ac}}{4\pi[E_i(z_0) + \varepsilon - E_i(z)]}
\]

\[
\begin{align*}
R_{e1}(\varepsilon) &= R_{e2}(\varepsilon) = \frac{\int_{z_0}^{z_f} \frac{1}{\lambda_{uc}(z,\varepsilon)} dz}{1 + \int_{z_0}^{z_f} \frac{1}{\lambda_{uc}(z,\varepsilon)} dz} \\
T_{e1}(\varepsilon) &= T_{e2}(\varepsilon) = 1 - R_{e1}(\varepsilon)
\end{align*}
\]

If potential is given, we can calculate R and T.
Validation of the approximation in elastic zone

Scattering rate under equilibrium
(d = 3 nm, nondegenerate)

- OP scattering can be neglected below the 63 meV
$R_{r1,2}$ and $T_{r1,2}$ for relaxation zone (1)

BTE

\[
\begin{align*}
-\frac{df^+(z,\varepsilon)}{dz} &= \frac{1}{\lambda_{ac}(z,\varepsilon)} f^+(z,\varepsilon) - \frac{1}{\lambda_{op}(z,\varepsilon)} f^-(z,\varepsilon) + \frac{1}{\lambda_{op}(z,\varepsilon)} \frac{1 - [(f^+(z,\varepsilon - \hbar\omega) + f^-(z,\varepsilon - \hbar\omega))/2]}{f^+(z,\varepsilon)} f^+(z,\varepsilon) \\
\frac{df^-(z,\varepsilon)}{dz} &= \frac{1}{\lambda_{ac}(z,\varepsilon)} f^-(z,\varepsilon) - \frac{1}{\lambda_{op}(z,\varepsilon)} f^+(z,\varepsilon) + \frac{1}{\lambda_{op}(z,\varepsilon)} \frac{1 - [(f^+(z,\varepsilon - \hbar\omega) + f^-(z,\varepsilon - \hbar\omega))/2]}{f^-(z,\varepsilon)} f^-(z,\varepsilon)
\end{align*}
\]

where

\[
\frac{1}{\lambda_{ac}(z,\varepsilon)} = \frac{1}{2v(z,\varepsilon)\tau_{ac}(z,\varepsilon)} = \frac{mS_{ac}}{4\pi\hbar[E_i(z_0) + \varepsilon - E_i(z)]}
\]

\[
\frac{1}{\lambda_{op}(z,\varepsilon)} = \frac{1}{v(z,\varepsilon)\tau_{op}(z,\varepsilon)} = \frac{mS_{op}}{2\pi\hbar \sqrt{[E_i(z_0) + \varepsilon - E_i(z)][E_i(z_0) + \varepsilon - E_i(z) - \hbar\omega]}} \approx \frac{mS_{op}}{2\pi\hbar [(E_i(z_0) + \varepsilon - E_i(z)]}
\]

\[
f^+(z,\varepsilon - \hbar\omega) \approx f_D(\varepsilon - \hbar\omega)
\]

\[
f^-(z,\varepsilon - \hbar\omega) \approx f_D(\varepsilon - \hbar\omega)
\]

We approximate $U(z_0) + \varepsilon - U(z) >> \hbar\omega$, to analytically solve BTE

To avoid excessive relaxation in degenerate system
R_{r1,2} and T_{r1,2} for relaxation zone (2)

\[
R_r = R_{r2} = \frac{1 - \exp\left(-2\sqrt{2X + X^2} \int_{z_1}^{z_2} \frac{1}{\lambda_{ac}(z, \varepsilon)} \, dz\right)}{\left(\frac{1 + X/2 + X/2}{2}\right) - \left(\frac{1 + X/2 - X/2}{2}\right) \exp\left(-2\sqrt{2X + X^2} \int_{z_1}^{z_2} \frac{1}{\lambda_{ac}(z, \varepsilon)} \, dz\right)}
\]

\[
T_r = T_{r2} = \frac{2\sqrt{2X + X^2} - \exp\left(-2\sqrt{2X + X^2} \int_{z_1}^{z_2} \frac{1}{\lambda_{ac}(z, \varepsilon)} \, dz\right)}{\left(\frac{1 + X/2 + X/2}{2}\right) - \left(\frac{1 + X/2 - X/2}{2}\right) \exp\left(-2\sqrt{2X + X^2} \int_{z_1}^{z_2} \frac{1}{\lambda_{ac}(z, \varepsilon)} \, dz\right)}
\]

where

\[X = \frac{S_{op}}{S_{ac}} [1 - f_D (\varepsilon - \hbar \omega)]\]

If potential is given, we can calculate R and T
$R_{s2,d2}$ and $T_{s1,d1}$ for source and drain zones (1)

Constant potential

AP +

OP (both in- and out-scattering)

BTE

$$\frac{df^+(z,\varepsilon)}{dz} = \frac{1}{\lambda_{ac}(z,\varepsilon)} f^+(z,\varepsilon) - \frac{1}{\lambda_{ac}(z,\varepsilon)} f^-(z,\varepsilon) + \frac{1}{\lambda_{op}(z,\varepsilon)} \left( f^+(z,\varepsilon - \hbar \omega) + f^-(z,\varepsilon - \hbar \omega) \right) / 2$$

$$\frac{df^-(z,\varepsilon)}{dz} = \frac{1}{\lambda_{ac}(z,\varepsilon)} f^{-}(z,\varepsilon) - \frac{1}{\lambda_{ac}(z,\varepsilon)} f^{+}(z,\varepsilon) + \frac{1}{\lambda_{op}(z,\varepsilon)} \left( f^+(z,\varepsilon + \hbar \omega) + f^-(z,\varepsilon - \hbar \omega) \right) / 2$$

Where (for source zone)

$$f^+(z,\varepsilon - \hbar \omega) \approx f_s(\varepsilon - \hbar \omega)$$

$$f^-(z,\varepsilon - \hbar \omega) \approx f_s(\varepsilon + \hbar \omega)$$

To avoid excessive relaxation in the degenerate system

To apply appropriate in-scattering rate

Detailed balance condition
\( R_{s2,d2} \) and \( T_{s1,d1} \) for source and drain zones (2)

Modified BTE

\[
\begin{align*}
\frac{df^+(z,\epsilon)}{dz} &= \left( \frac{1}{\lambda_{ac}^s(\epsilon)} + \frac{1}{\lambda_{out}^s(\epsilon)} \right) f^+(z,\epsilon) - \frac{1}{\lambda_{ac}^s(\epsilon)} f^-(z,\epsilon) - \frac{1}{\lambda_{in}^s(\epsilon)} \\
\frac{df^-(z,\epsilon)}{dz} &= \left( \frac{1}{\lambda_{ac}^s(\epsilon)} + \frac{1}{\lambda_{out}^s(\epsilon)} \right) f^-(z,\epsilon) - \frac{1}{\lambda_{ac}^s(\epsilon)} f^+(z,\epsilon) - \frac{1}{\lambda_{in}^s(\epsilon)}
\end{align*}
\]

\( R_{s2}(\epsilon) = \frac{1 - \exp(-2Y_s(\epsilon)L_s)}{Y_s(\epsilon) + \frac{1}{\lambda_{out}^s(\epsilon)} - \frac{1}{\lambda_{out}^s(\epsilon)} \exp(-2Y_s(\epsilon)L_s)} \)  

\( T_{s1}(\epsilon) = 1 - R_{s2}(\epsilon) \)

\( T_{d2}(\epsilon) = 1 - R_{d1}(\epsilon) \)

With the same process

\( R_{d1}(\epsilon) = \frac{1 - \exp(-2Y_d(\epsilon)L_d)}{Y_d(\epsilon) + \frac{1}{\lambda_{out}^d(\epsilon)} - \frac{1}{\lambda_{out}^d(\epsilon)} \exp(-2Y_d(\epsilon)L_d)} \)

\( T_{d1}(\epsilon) = 1 - R_{d1}(\epsilon) \)

We can calculate \( R \) and \( T \) from length of source and drain zones.
Summary of derived Rs and Ts

**Elastic zone**

\[
R_{e1}(\varepsilon) = R_{e2}(\varepsilon) = \int^{z_e}_{z_i} \frac{1}{\lambda_{ae}(z, \varepsilon)} dz
\]

\[
T_{e1}(\varepsilon) = T_{e2}(\varepsilon) = 1 - R_{e1}(\varepsilon)
\]

where

\[
\frac{1}{\lambda_{ae}(z, \varepsilon)} \equiv \frac{mS_{ac}}{4\pi h[U(z_0) + \varepsilon - U(z)]}
\]

**Barrier zone**

\[
R_{b1}(\varepsilon) = R_{b2}(\varepsilon) = \int^{z_b}_{z_i} \frac{1}{\lambda_{ae}(z, \varepsilon)} dz
\]

\[
T_{b1}(\varepsilon) = T_{b2}(\varepsilon) = 1 - R_{b1}(\varepsilon)
\]

**Relaxation zone**

\[
R_{r1}(\varepsilon) = R_{r2}(\varepsilon) = \frac{1 - \exp \left( -2\sqrt{2X + X^2} \int^{z_r}_{z_i} \frac{1}{\lambda_{ae}(z, \varepsilon)} dz \right)}{\left( \sqrt{1 + \frac{X}{2}} + \sqrt{\frac{X}{2}} \right)^2 - \left( \sqrt{1 + \frac{X}{2} - \frac{X}{2}} \right)^2 \exp \left( -2\sqrt{2X + X^2} \int^{z_r}_{z_i} \frac{1}{\lambda_{ae}(z, \varepsilon)} dz \right)}
\]

\[
T_{r1}(\varepsilon) = T_{r2}(\varepsilon) = \frac{2\sqrt{2X + X^2} - \exp \left( -2\sqrt{2X + X^2} \int^{z_r}_{z_i} \frac{1}{\lambda_{ae}(z, \varepsilon)} dz \right)}{\left( \sqrt{1 + \frac{X}{2}} + \sqrt{\frac{X}{2}} \right)^2 - \left( \sqrt{1 + \frac{X}{2}} - \frac{X}{2} \right)^2 \exp \left( -2\sqrt{2X + X^2} \int^{z_r}_{z_i} \frac{1}{\lambda_{ae}(z, \varepsilon)} dz \right)}
\]

where

\[
X \equiv \frac{S_{op}}{S_{ac}} [1 - f_D(\varepsilon - h\omega)]
\]

**Source zone**

\[
R_{s1}(\varepsilon) = \frac{1 - \exp(-2Y_s(\varepsilon)L_s)}{Y_s(\varepsilon) + \frac{1}{\lambda_{out}(\varepsilon)}}
\]

\[
T_{s1}(\varepsilon) = 1 - R_{s1}(\varepsilon)
\]

where

\[
Y_s(\varepsilon) = \frac{2}{\lambda^d_{in}(\varepsilon)[\lambda^d_{in}(\varepsilon) + (\lambda_{out}(\varepsilon))^2]}
\]

\[
\frac{1}{\lambda^d_{in}(\varepsilon)} = \frac{1}{\lambda_{op}(\varepsilon)[1 - f_s(\varepsilon - h\omega)]}
\]

\[
\frac{1}{\lambda_{op}(\varepsilon)} = \frac{mS_{op}}{2\pi h[[E_i(z_0) + \varepsilon - E_i^1][E_i(z_0) + \varepsilon - E_i^1 - h\omega]]}
\]

**Drain zone**

\[
R_{d1}(\varepsilon) = \frac{1 - \exp(-2Y_d(\varepsilon)L_d)}{Y_d(\varepsilon) + \frac{1}{\lambda_{out}(\varepsilon)}}
\]

\[
T_{d2}(\varepsilon) = 1 - R_{d1}(\varepsilon)
\]

where

\[
Y_d(\varepsilon) = \frac{2}{\lambda^d_{in}(\varepsilon)[\lambda^d_{in}(\varepsilon) + (\lambda_{out}(\varepsilon))^2]}
\]

\[
\frac{1}{\lambda^d_{in}(\varepsilon)} = \frac{1}{\lambda_{op}(\varepsilon)[1 - f_d(\varepsilon - h\omega)]}
\]

\[
\frac{1}{\lambda_{op}(\varepsilon)} = \frac{mS_{op}}{2\pi h[[E_i(z_0) + \varepsilon - E_i^1][E_i(z_0) + \varepsilon - E_i^1 - h\omega]]}
\]
Validation by a numerical simulation

Importing self-consistent potential, Rs and Ts are semi-analytically calculated in our model.

Example of self-consistently calculated potential profile

![Graph of E(z) vs position (nm)](image)

- $d=3\text{nm}$
- $L_s=10\text{nm}$
- $L_g=30\text{nm}$
- $L_g'=10\text{nm}$
- $V_g-V_t=0.3\text{V}$
- $V_d=0.3\text{V}$

We compare the results from our model and the numerical simulation based on the deterministic solution of 1D MSBTE.
Numerical simulation based on the deterministic solution of 1D MSBTE

Deterministic solution of 1D MSBTE is more efficient than 1D MSMC in the low dimensionality of the nanowire

Results are the same as those from 1D MSMC
Validation 1: Backscattering coefficient

- This model is in good agreement with the numerical simulation especially below 63 meV
- Modeling for drain zone is available even in substantially long drain
Validation 2: Carrier distribution function at $z_0$

Saturation region
$L_s = 10 \text{ nm}$
$V_g - V_t = 0.3 \text{ V}$
$V_d = 0.3 \text{ V}$

Long source
$L_s = 100 \text{ nm}$
$V_g - V_t = 0.0 \text{ V}$
$V_d = 0.3 \text{ V}$

- Distribution function at the top of the barrier taken into account
- Modeling of source zone is available even in substantially long source
Validation 3: I-V characteristics

- This model is in quantitatively good agreement with the numerical simulation.
Discussion 1: Potential profile to investigate quasi-ballistic transport

- $U_s$ and $U_d$ are self-consistently calculated with donor impurity density
- $R(\varepsilon)$ can be analytically derived
- To calculate drain current, we need numerical integration over energy
Discussion 2: Ballisticity

- When $L_g = 10$ nm, 20% is backscattered.
- When $L_d > 50$ nm, 20% is backscattered from drain zone.
- Considering realistic 1D drain, we can not obtain ‘ballisticity = 1’.
Discussion 3: Source length dependence

- When $L_s < 10$ nm, elastic source approximation is available

- In a shorter 1D source, the resistivity seems to be smaller
• In $L_g > 50$ nm, Natori’s model is in good agreement with this model

• In $L_g < 10$ nm, elastic approximation is in good agreement with this model
Discussion 5: Comparison with other models (2)

- Elastic approximation could not describe backscattering in long drain.
- In $L_g > 100$ nm, Natori’s model is in good agreement with this model.
Advantages of this model

- Describing distribution function of the carrier injected from realistic source
- Describing transmission coefficient for the carriers injected from drain
- Taking into account finite length of drain and source as well as channel
We have successfully modeled the quasi-ballistic transport in nanowire MOSFETs. The device is divided into five zones. Backscattering coefficient, distribution function, and drain current were derived with one-flux scattering matrix and BTEs. Results from this model were in good agreement with those from a numerical simulation. 1D drain length was important to adjust backscattering coefficient. 1D source length was also important to represent distribution of the carriers injected from realistic source.
Conclusions

By modeling gate capacitance and quasi-ballistic transport, we could interpret device physics

- We successfully modeled the gate capacitance to distinguish different contributions of the quantum effects
- The decrease in $x_{avg}$ caused the positive effect on the total capacitance

- We successfully developed the quasi-ballistic transport model of NW MOSFETs
- $L_s$ and $L_d$ were also important to estimate the drain current

By numerical simulation, we investigated what parameter control performance in ultra-scaled NW MOSFETs

- We showed the effects of the size-dependent properties on performance
- Desirable diameter could be around 5 nm

- We can suppress effective mass fluctuation by adopting square cross section

- Size and corner effects without strain could not cause the drastic mobility increase from experimental results
- In width < 6 nm, mobility drastically modulated
SiNW FET Fabrication

- S/D & Fin Patterning
- Sacrificial Oxidation
- Oxide etch back
- SiN sidewall support formation
- Gate Oxidation & Poly-Si Deposition
- Gate Lithography & RIE Etching
- Gate Sidewall Formation
- Ni SALISIDE Process (Ni 9nm / TiN 10nm)
- Backend

Standard recipe for gate stack formation
(a) SEM image of Si NW FET (Lg = 200nm)
(b) high magnification observation of gate and its sidewall.
Fabricated SiNW FET
Recent results to be presented by ESSDERC 2010 next week in Seville

Wire cross-section: 20 nm X 10 nm

On/Off $>10^6$, 60uA/wire

$L_g = 65$nm, $T_{ox} = 3$nm
Bench Mark

Gate Length (nm)  

$I_{ON} (\mu A / wire)$  

V$_{DD}$: 1.0~1.5 V  

Our Work
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NW Size (nm)</td>
<td>Rect.</td>
<td>Rect.</td>
<td>Rect.</td>
<td>Cir.</td>
<td>Cir.</td>
<td>Elliptical</td>
<td>Elliptical</td>
</tr>
<tr>
<td>Lg (nm)</td>
<td>10x20</td>
<td>10x20</td>
<td>14</td>
<td>10</td>
<td>10</td>
<td>12</td>
<td>13x20</td>
</tr>
<tr>
<td>EOT or Tox (nm)</td>
<td>65</td>
<td>25</td>
<td>100</td>
<td>30</td>
<td>8</td>
<td>65</td>
<td>35</td>
</tr>
<tr>
<td>Vdd (V)</td>
<td>3</td>
<td>1.8</td>
<td>1.8</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1.5</td>
</tr>
<tr>
<td>Ion(uA) per wire</td>
<td>1.0</td>
<td>1.1</td>
<td>1.2</td>
<td>1.0</td>
<td>1.2</td>
<td>1.2</td>
<td>1.0</td>
</tr>
<tr>
<td>Ion(uA/um) by dia.</td>
<td>60.1</td>
<td>102</td>
<td>30.3</td>
<td>26.4</td>
<td>37.4</td>
<td>48.4</td>
<td>43.8</td>
</tr>
<tr>
<td>Ion(uA/um) by cir.</td>
<td>3117</td>
<td>5010</td>
<td>2170</td>
<td>2640</td>
<td>3740</td>
<td>4030</td>
<td>2592</td>
</tr>
<tr>
<td>SS (mV/dec.)</td>
<td>1609</td>
<td>2054</td>
<td>430</td>
<td>841</td>
<td>1191</td>
<td>1283</td>
<td>825</td>
</tr>
<tr>
<td>DIBL (mV/V)</td>
<td>70</td>
<td>79</td>
<td>68</td>
<td>71</td>
<td>75</td>
<td>75</td>
<td>85</td>
</tr>
<tr>
<td>Ion/Ioff</td>
<td>~1E6</td>
<td>&gt;1E6</td>
<td>&gt;1E5</td>
<td>~1E6</td>
<td>&gt;1E7</td>
<td>&gt;1E7</td>
<td>~2E5</td>
</tr>
</tbody>
</table>

Ref[11] by Stmicro Lg=25nm, Tox=1.8nm
This work Lg=65nm, Tox=3nm
This work

Planer FET ▲
S. Kamiyama, IEDM 2009, p. 431
P. Packan, IEDM 2009, p.659

SiナノワイヤFET★
Y. Jiang, VLSI 2008, p.34
H.-S. Wong, VLSI 2009, p.92
S. Bangsaruntip, IEDM 2009, p.297
C. Dupre, IEDM 2008, p. 749
S.D.Suk, IEDM 2005, p.735
G.Bidel, VLSI 2009, p.240

$L_g=500\sim65\text{nm}$

$I_{\text{ON}}/I_{\text{OFF}}$ Bench mark

$\log_{10} I_{\text{OFF}}$ (nA/μm)

$\log_{10} I_{\text{ON}}$ ($\mu$A/μm)

1.0 $\sim$ 1.1 V

1.2 $\sim$ 1.3 V

This work
Chapter 4

H₂ annealing

H₂ annealing
750°C, 2min

Electron mobility $\mu_{\text{elec}}$ [cm²/Vs]

Hole mobility $\mu_{\text{hole}}$ [cm²/Vs]

Change the cross section

- A little improvement of surface roughness
- Increase of coulomb scattering
The p+ region acts as a ‘substrate’ contact.

Increase of interface trap density
3. Problems

We have to decrease Si layer thickness to better control of channel potential by gate bias.

Significant decrease in conduction or Ion.
**Short-channel effect**

T. Skotnicki, IEDM 2009 Short Course (STMicroelectronics)

<table>
<thead>
<tr>
<th>Channel</th>
<th>Expressions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulk</td>
<td>$EI = 1 \times \left(1 + \frac{X_j^2}{L_{el}^2} \right) \times \frac{T_{ox}}{L_{el}} \times \frac{T_{dep}}{L_{el}}$</td>
</tr>
<tr>
<td>PD SOI</td>
<td>$EI = 1 \times \left(1 + \frac{X_j^2}{L_{el}^2} \right) \times \frac{T_{ox}}{L_{el}} \times \frac{T_{dep}}{L_{el}}$</td>
</tr>
<tr>
<td>Thick BOX FD SOI</td>
<td>$EI = 1 \times \left(1 + \frac{T_{si}^2}{L_{el}^2} \right) \times \frac{T_{ox}}{T_{si} + \lambda T_{box}} \times \frac{L_{el}}{L_{el}}$</td>
</tr>
<tr>
<td>UTBB (FD SOI or SON)</td>
<td>$EI = 1 \times \left(1 + \frac{T_{si}^2}{L_{el}^2} \right) \times \frac{T_{ox}}{T_{si} + \lambda T_{box}} \times \frac{L_{el}}{L_{el}}$</td>
</tr>
<tr>
<td>FinFET (independent gates)</td>
<td>$EI = 1 \times \left(1 + \frac{T_{si}^2}{L_{el}^2} \right) \times \frac{T_{ox}}{T_{si} / 2} \times \frac{L_{el}}{L_{el}}$</td>
</tr>
<tr>
<td>FinFET (common gates)</td>
<td>$EI = 1 \times \left(1 + \frac{T_{si}^2 / 4}{L_{el}^2} \right) \times \frac{T_{ox}}{T_{si} / 2} \times \frac{L_{el}}{L_{el}}$</td>
</tr>
</tbody>
</table>

SCE = $0.64 \frac{\varepsilon_s}{\varepsilon_{ox}} \times EI \times \Phi_d$

DIBL = $0.8 \frac{\varepsilon_s}{\varepsilon_{ox}} \times EI \times V_{ds}$

- if max ch. doping is $10^{20}$ cm$^{-3}$ => Tep~12nm then:
  - Tbox = 145nm
  - $\lambda T_{box} = 18$nm
- if max ch. doping is $10^{19}$ cm$^{-3}$ => Tep~12nm then:
  - Tbox = 10nm
  - $\lambda T_{box} = 3$nm
- if $T_{si} \ll 3$nm then:
  - Tbox = 10nm
  - $\lambda T_{box} = 0.46$nm

LeL,min (DIBL=100mV/V)

= 34nm

≈ 30nm

≈ 10nm

≈ 6nm

≈ 2.7nm
**Drain-induced barrier lowering**

**T. Skotnicki, IEDM 2010 Short Course (STMicroelectronics)**

\[ \text{DIBL} = 0.80 \frac{\varepsilon_{Si}}{\varepsilon_{ox}} \left( 1 + \frac{X_j^2}{L_{el}^2} \right) \frac{T_{ox}}{L_{el}} \frac{T_{dep}}{L_{el}} V_{DS} \]

- **BULK**
  - \( T_{dep} = \frac{3}{4} L_{el} \)
  - 140 mV/V

- **III-V**
  - \( T_{ox} + 2A \)
  - \( \varepsilon_{Si} + 15\% \)
  - 210 mV/V

- **FDSOI**
  - \( T_{dep} = T_{si}/2 \)
  - \( X_j = T_{si} \geq 6 \text{nm} \)
  - 110 mV/V

- **ETSOI**
  - \( T_{dep} = T_{si} + \lambda T_{box} \)
  - \( X_j = T_{si} \geq 6 \text{nm} \)
  - 80 mV/V

- **FinFET**
  - \( T_{dep} = T_{si}/2 \)
  - \( X_j = T_{si}/2 \)
  - \( T_{si} \geq 10 \text{nm} \)
  - 70 mV/V

120
Sub-threshold Slope

\[ S = \frac{kT}{q} \ln \left( 1 + \frac{C_{\text{dep}}}{C_{\text{ox}}} \left( 1 + \frac{\varepsilon_{\text{Si}} T_{\text{ox}}}{\varepsilon_{\text{ox}} L_{\text{el}} L_{\text{el}}} \left( 1 + \frac{3 T_{\text{dep}}}{4 L_{\text{el}}} \right)^{1+2 V_{\text{ds}}/\Phi_{\text{d}}} \right) \right) \]

- BULK
  - \( T_{\text{dep}} = \frac{3}{4} L_{\text{el}} \)
  - \( X_j = \frac{3}{4} L_{\text{el}} \)
  - Slope: 95 mV/dec

- III-V
  - \( T_{\text{ox}} = 2A \)
  - \( \varepsilon_{\text{Si}} = 15\% \)
  - Slope: 110 mV/dec

- FDSOI
  - \( T_{\text{dep}} = \frac{T_{\text{si}}}{2} \)
  - \( X_j = \frac{T_{\text{si}}}{2} \)
  - Slope: 85 mV/dec

- ETSOI
  - \( T_{\text{dep}} = T_{\text{si}} + \lambda T_{\text{box}} \)
  - \( X_j = T_{\text{si}} \)
  - Slope: 75 mV/dec

- FinFET
  - \( T_{\text{dep}} = \frac{T_{\text{si}}}{2} \)
  - \( X_j = \frac{T_{\text{si}}}{2} \)
  - \( \lambda = 1\% \)
  - Slope: 65 mV/dec

- UTBB
  - \( \Phi_d = 1\% \)
  - Slope: 75 mV/dec

- RBB => Pstat

T. Skotnicki, IEDM 2010 Short Course (STMicroelectronics)
3. Problems

    Mobility degradation.
Problems in Multi-gate

S. Bangsaruntip et al., pp.297, IEDM2009 (IBM),
K. Tachi et al., pp.313, IEDM2009 (CEA-LETI)

Decreasing the diameter of NW

Need to decrease diameter for SCH

Significant $\mu$ degradation at diameter $< 10$ nm

Decreasing the diameter of NW

Improved short-channel control

Severe mobility degradation
Problems in SOI

Mobility is also decreased with decreasing the Si thickness of SOI transistor similar to the NW transistor.
Problem for nanowire

When wire diameter becomes less than 10 nm, sudden drop of $I_d$

1. **Mobility degradation**
   - Extremely small distance between the electron and all around Si surface.
   - Strong scattering of electrons by interaction with all around Si surface.

2. **Electron density decrease**
   - Decrease of DOS in extremely narrow wire.

If diameter cannot be scaled, SCE cannot be suppressed.

Then, again aggressive EOT scaling of high-k is necessary.
3. Problems

Carrier density degradation.
Number of quantum channels

By Prof. Shiraishi of Tsukuba univ.

Energy band of Bulk Si

Energy band of 3 x 3 Si wire
Diameter dependence

By Profs. Oshiyama and Iwata, U. of Tokyo
Wire cross section dependence.
What cross section gives best solution for SCE suppression and drive current?
Significant increase in $V_{th}$ variability with decreasing Fin width
- Smaller wire/fin width is necessary for SCE suppression
- But mobility and $I_{ON}$ severely degrade with wire/fin width reduction
- Therefore even in multi-gate structures, EOT scaling should be accelerated to provide SCE immunity
High-k beyond 0.5 nm
Gate oxide should be thicker than mono atomic layer

0.8 nm gate oxide thickness MOSFETs operate

0.8 nm → Distance of 3 Si atoms → 2 mono layers

R.Chau, et al., (Intel) IWGI 2003
Gate Leakage Power Density becomes significantly large with $L_g$ reduction, and thus, with $t_{ox}$ thinning!!
To use high-k dielectrics

**K**: Dielectric Constant

Thin SiO$_2$

K=4

Almost the same electric characteristics

Thick high-k dielectrics

K=20

5 times thicker

Small leakage Current

However, very difficult and big challenge!
Equivalent Oxide Thickness (EOT): gate dielectrics itself, $C_{ox}$

Capacitance Equivalent Thickness (CET): entire gate stack,

Metal gate can eliminate the poly-Si depletion.

Inversion CET = $T_{inv} \approx EOT + 0.4\text{nm}$ with metal gate electrode

$C_{metal}$ is finite because of quantum effect. In other words electron is not a point charge located at the interface but distributed charge.
### Choice of High-k elements for oxide

<table>
<thead>
<tr>
<th>Candidates</th>
<th>Unstable at Si interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>Gas or liquid at 1000 K</td>
</tr>
<tr>
<td>Li Be</td>
<td>Radio active He</td>
</tr>
<tr>
<td>Na Mg</td>
<td>B C N O F Ne</td>
</tr>
<tr>
<td>Ca Sc</td>
<td>Al Si P S Cl Ar</td>
</tr>
<tr>
<td>K</td>
<td>1) band-offset, 2) dielectric constant, 3) thermal stability</td>
</tr>
<tr>
<td>Rh Sr Y Zr</td>
<td>La₂O₃ based dielectrics are thought to be the next generation materials, which may not need a thicker interfacial layer</td>
</tr>
<tr>
<td>Cs Ba Hf</td>
<td>HfO₂ based dielectrics are selected as the first generation materials, because of their merit in 1) band-offset, 2) dielectric constant, 3) thermal stability</td>
</tr>
</tbody>
</table>

- **Choice of High-k elements for oxide**
  - HfO₂ based dielectrics are selected as the first generation materials, because of their merit in 1) band-offset, 2) dielectric constant, 3) thermal stability.
  - La₂O₃ based dielectrics are thought to be the next generation materials, which may not need a thicker interfacial layer.

- **R. Hauser, IEDM Short Course, 1999**
Conduction band offset vs. Dielectric Constant

Leakage Current by Tunneling

Band Discontinuity [eV]

Dielectric Constant \( \varepsilon(0) \)

XPS measurement by Prof. T. Hattori, INFOS 2003
Direct high-\(k\)/Si by silicate reaction

Direct contact can be achieved with La\(_2\)O\(_3\) by forming silicate at interface
Control of oxygen partial pressure is the key for processing.

K. Kakushima, et al., VLSI2010, p.69
Direct high-k/Si by silicate reaction

Direct contact can be achieved with La$_2$O$_3$ by forming silicate at interface. Control of oxygen partial pressure is the key for processing.

K. Kakushima, et al., VLSI2010, p.69
SiO$_x$-IL growth at HfO$_2$/Si Interface

Phase separator

\[
\text{HfO}_2 + \text{Si} + \text{O}_2 \rightarrow \text{HfO}_2 + \text{Si} + 2\text{O}^* \rightarrow \text{HfO}_2 + \text{SiO}_2
\]

H. Shimizu, JJAP, 44, pp. 6131

Oxygen supplied from W gate electrode

D.J. Lichtenwalner, Trans. ECS 11, 319

SiO$_x$-IL is formed after annealing

Oxygen control is required for optimizing the reaction
La-Silicate Reaction at $\text{La}_2\text{O}_3$/Si

Direct contact high-$k$/Si is possible

$\text{La}_2\text{O}_3 + \text{Si} + n\text{O}_2 \rightarrow \text{La}_2\text{SiO}_5, \text{La}_2\text{Si}_2\text{O}_7, \text{La}_{9.33}\text{Si}_6\text{O}_{26}, \text{La}_{10}(\text{SiO}_4)_6\text{O}_3$, etc.

$\text{La}_2\text{O}_3$ can achieve direct contact of high-$k$/Si
Cluster tool for HKMG Stack

- **EB Deposition for HK**
- **Flash Lamp**
- **Sputter for MG**
- **Robot**
- **ALD**
- **RTA**
- **Entrance**
Cluster Chambers for HKMG Gate Stack

EB Deposition: HK
Flash Lamp Anneal
ALD: HK
Sputter: MG
Robot
RTA
Entrance
$L = 0.5 \sim 100 \mu \text{m} (8 \text{kinds})$

$W = 10, 20, 50, 100 \mu \text{m} (4 \text{kinds})$

30 different Trs

26 chips

1 cm × 1 cm

p-Si

n+ SD

n+ SiO₂
Gate Leakage vs EOT, (Vg=|1|V)

- Materials: Al₂O₃, HfAlO(N), HfO₂, HfSiO(N), HfTaO, La₂O₃, Nd₂O₃, Pr₂O₃, PrSiO, PrTiO, SiON/SiN, Sm₂O₃, SrTiO₃, Ta₂O₅, TiO₂, ZrO₂(N), ZrSiO, ZrAlO(N)

- Parameters: EOT (nm), Current density (A/cm²)
However, high-temperature anneal is necessary for the good interfacial property.

A fairly nice La-silicate/Si interface can be obtained with high temperature annealing. (800°C)
ALD of La$_2$O$_3$


ALD is indispensable from the manufacturing viewpoint

Precursor (ligand)

La(iPrCp)$_3$

La(FAMD)$_3$

La gas feed

Ar purge

H$_2$O feed

Ar purge

1 cycle

ALD is indispensable from the manufacturing viewpoint

precise control of film thickness and good uniformity
Physical mechanisms for small Dit

① silicate-reaction-formed fresh interface

② stress relaxation at interface by glass type structure of La silicate.

Fresh interface with silicate reaction

FGA800°C is necessary to reduce the interfacial stress


As depo
La₂O₃(3.5 nm)
TiN/W(60 nm)

Annealed for 2 s
TiN/W(6 nm)

Annealing temperature (°C)

EOT (nm)

Annealed for 2 s
La₂O₃(3.5 nm)

W(60 nm)

TiN/W(6 nm)

EOT=0.55nm

Cg (μF/cm²)

Experiment
Theory

TaN/(45nm)/W(3nm)
900°C, 30min

EOT=0.55nm

Vg (V)
Flat-band voltage (V)

TaN(45nm)/W(3nm) 900°C, 30min

Q_{fix} = 1 \times 10^{11} \text{ cm}^{-2}

Fixed Charge density: 1 \times 10^{11} \text{ cm}^{-2}
Our Work at TIT: High-k

Our result at TIT

EOT=0.40nm

Drain Current (mA) vs. Drain Voltage (V)
- L/W = 5/20µm
- Vg = 1.0V
- Vg = 0.8V
- Vg = 0.6V
- Vg = 0.4V
- Vg = 0.2V
- Vg = 0 V

Electron Mobility [cm²/Vsec] vs. E_{eff} [MV/cm]
- EOT = 0.40nm
- L/W = 5/20µm
- T = 300K
- N_{sub} = 3 \times 10^{16} cm^{-3}
Benchmark of La-silicate dielectrics

**Gate Leakage current**

- **ITRS requirement**
- **Our data: La-silicate gate oxide**

- **Effective Mobility**
  - **at 1 MV/cm**
  - **Solid circle: Our data**
  - **Open square: Hf-based oxides**


*T. Ando, et al., (IBM) IEDM 2009, p.423*
Issues in high-k/metal gate stack

- Oxygen concentration control for prevention of EOT increase and oxygen vacancy formation in high-k
- Flat metal/high-k interface for better mobility
- Suppression of metal diffusion
- Suppression of oxygen vacancy formation
- Small interfacial state density at high-k/Si
- Control of interface reaction and Si diffusion to high-k
- Suppression of gate leakage current
- Endurance for high temperature process
- Oxygen diffusion control for prevention of EOT increase and oxygen vacancy formation in high-k
- Workfunction engineering for $V_{th}$ control
- Interface dipole control for $V_{th}$ tuning
- Suppression of FLP
- Remove contamination introduced by CVD
- Thinning or removal of SiO$_2$-IL for small EOT
- Reliability: PBTI, NBTI, TDDB
- Oxygen concentration control for prevention of EOT increase and oxygen vacancy formation in high-k
- Control of interface reaction and Si diffusion to high-k
<table>
<thead>
<tr>
<th>Gate stack</th>
<th>EOT</th>
<th>Mobility</th>
<th>( V_{th} )</th>
<th>SS</th>
<th>DIBL</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TiN/Cap/HfO₂</td>
<td>0.52nm</td>
<td>110cm²/Vs</td>
<td>~0.4V</td>
<td>90mV/dec</td>
<td>147mV/V</td>
<td>IBM VLSI2011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(at 1x10¹³cm⁻²)</td>
<td>(L₉=24nm)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TiN/Cap/HfO₂</td>
<td>0.55nm</td>
<td>140cm²/Vs</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>IBM VLSI2009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(at 1MV/cm)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>TiN/Cap/HfO₂</td>
<td>0.45nm</td>
<td>115cm²/Vs</td>
<td>0.3V</td>
<td>—</td>
<td>—</td>
<td>IMEC MEE2011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(at 1x10¹³cm⁻²)</td>
<td>(L₉=10um)</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Metal/HfO₂</td>
<td>0.59nm</td>
<td>130cm²/Vs</td>
<td>0.45V</td>
<td>75mV/dec</td>
<td>—</td>
<td>Sematech VLSI2009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(at 1MV/cm)</td>
<td>(L₉=1um)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metal/Hf-based</td>
<td>0.65nm</td>
<td>—</td>
<td>0.3~0.4V</td>
<td>90mV/dec</td>
<td>100mV/V</td>
<td>Samsung VLSI2011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(L₉=30nm)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metal/Hf-based</td>
<td>0.95nm</td>
<td>—</td>
<td>~0.3V</td>
<td>100mV/dec</td>
<td>~200mV/V</td>
<td>Intel IEDM2009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(L₉=30nm)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W/La-silicate</td>
<td>0.62nm</td>
<td>155cm²/Vs</td>
<td>~0.08V</td>
<td>~70mV/dec</td>
<td>—</td>
<td>Tokyo Tech. T-ED2012</td>
</tr>
<tr>
<td>158</td>
<td></td>
<td>(at 1MV/cm)</td>
<td>(L₉=10um)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusions 1

Downsizing of MOSFETs is still important for high-speed low-power operation of logic LSIs.

Ioff will limit the downsizing.

Punchthrough component of Ioff will be suppressed by thinning tox and adopting new configuration such as FinFET or nanowire FET.

Direct tunneling will limit the downsizing @Lg = 3 nm.

Even before that, subthreshold leakage would limit the downsizing @Lg > 3nm, depending on the application.
Conclusions 2

In the application, Ion/Ioff ratio is important.

The ratio is typically $10^6$ for the present devices, however, it degrades significantly with decreasing the supply voltage.

Si nanowire FET has advantage not only on Ioff over planer FET, but also on Ion, because of better mobility and higher channel carrier density.

In order to suppress Ioff with decrease in Lg, the diameter of nanowire, width of fin, or thickness of Si film of SOI need to be shirked.

However, with decreasing the above diameter, width or thickness less than several nano-meter, very significant decrease in Ion occurs, because of the degradation on mobility and carrier concentration.
Conclusions 3

If the diameter, width or thickness cannot be decreased, we need to decrease the EOT of high-k aggressively in order to suppress $I_{off}$.

High-k EOT reduction trend is very slow – 0.05 nm for each generation --, for the moment.

The limit of EOT scaling is expected to be around 0.4 nm or so, considering the additional capacitances of channel and metal gate.

By changing the high-k material from HfO$_2$ to La-silicate, we can obtain the good operation of MOSFET with EOT = 0.4 nm,

- Metal silicide Shottoky S/D will become important in order to suppress the S/D encroachment to the channel by dopant diffusion.
Conclusions 3

Downsizing of MOFET is becoming more and more important for low power high performance application in the future smart society, and will be accomplished in another 10 to 15 years, although the rate of the downsizing will become slow.

Thus, many challenging technology development will be necessary for another 10 to 15 years.

Thank you very much for your attention.