Nano-CMOS Technology

June 1, 2011

Lanzhou Jiaotong University

Hiroshi Iwai,
Tokyo Institute of Technology
Tokyo Institute of Technology
Founded in 1881, Promoted to Univ. 1929
Institute Overview

Established in 1881 → 130th anniversary in 2011

3 undergraduate schools
School of Science, School of Engineering, School of Bioscience and Biotechnology

7 graduate schools
Science and Engineering Science, Science and Engineering Technology, Bioscience and Biotechnology, Interdisciplinary Graduate School of Science and Engineering, Information Science and Engineering, Decision Science and Technology, Innovation Management

Total Number of Students

<table>
<thead>
<tr>
<th></th>
<th>Undergraduate</th>
<th>Graduate</th>
<th>Master's</th>
<th>Doctoral</th>
<th>Teaching Staff</th>
<th>Student/Instructor</th>
<th>Staff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tokyo Inst.</td>
<td>5,000</td>
<td>5,000</td>
<td>3,500</td>
<td>1,500</td>
<td>1,200</td>
<td>8.3</td>
<td>550</td>
</tr>
<tr>
<td>Per Year</td>
<td>1,200</td>
<td>1,800</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Einstein Visit
Total 982
(As of May. 1, 2005)
岩井研メンバー
（2009年11月1日現在）

教授
岩井洋

准教授（共同研究）
筒井一生

客員教授
Simon Min Sze

客員教授
服部健雄

特任教授
名取研二

連携教授
杉井信之

連携教授
西山彰

特任准教授
Parhat Ahmet

助教
角嶋邦之

博士
研究員
Milan Kumar Bera

博士
課程
佐々木雄一朗

D3
下村浩

S3
宋在烈

D3
館喜一

D2
川那子高暢

D3
佐藤創志

D2
富田隆治

D2
Kaimath Rexal Maimat

修士
課程
Abudutelimu
Abudureheman

D1
幸田みゆき

D1
李映勜

D3
藤井孝

D3
小林勇介

D2
新井英朗

D2
中山寛人

D2
船水清永

D2
細田真

D2
又野克哉

M2
タリクス・ハザンダ

M2
Mohammad Shohil Hadi

M1
小柳友常

M1
小澤健児

M1
神田高志

M1
澤田剛伸

M1
茂森直登

M1
向井弘樹

M1
呉研

M1
Dou Chunmeng

学部

研究生

スタッフ

松本昭子

辛川美琴

西澤 正子
• There were many inventions in the 20th century: Airplane, Nuclear Power generation, Computer, Space aircraft, etc
• However, everything has to be controlled by electronics
• Electronics
  Most important invention in the 20th century
• What is Electronics: To use electrons, Electronic Circuits
Electronic Circuits started by the invention of vacuum tube (Triode) in 1906

Thermal electrons from cathode controlled by grid bias

Cathode (heated)  Grid  Anode (Positive bias)

Same mechanism as that of transistor
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
J. E. LILIENFELD

DEVICES FOR CONTROLLED ELECTRIC CURRENT

Filed March 28, 1928
Capacitor structure with notch

**Negative bias**

- Gate Electrode
- Gate Insulator
- Semiconductor
- Electron

No current

**Positive bias**

- Electric field

Current flows
Today’s transistor: MOSFET for CMOS LSI

Surface Potential (Negative direction)

0 bias for gate

Positive bias for gate

Electron flow
However, no one could realize MOSFET operation for more than 30 years.

Because of very bad interface property between the semiconductor and gate insulator

Even Shockley!
Very bad interface property between the semiconductor and gate insulator

Drain Current was several orders of magnitude smaller than expected

Even Shockley!
However, they found amplification phenomenon when investigating Ge surface when putting needles. This is the 1st Transistor:
Not Field Effect Transistor, But Bipolar Transistor (another mechanism)

1947: 1st transistor

Bipolar using Ge

J. Bardeen
W. Bratten, W. Shockley
1958: 1st Integrated Circuit

Connect 2 bipolar transistors in the Same substrate by bonding wire.
1960: First MOSFET by D. Kahng and M. Atalla

Top View

Si
Source

Al Gate

Drain

Si

Si/SiO₂ Interface is extraordinarily good
1970,71: 1st generation of LSIs

**DRAM** Intel 1103

**MPU** Intel 4004
MOS LSI experienced continuous progress for many years.

<table>
<thead>
<tr>
<th>Name of Integrated Circuits</th>
<th>Number of Transistors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960s IC (Integrated Circuits)</td>
<td>~ 10</td>
</tr>
<tr>
<td>1970s LSI (Large Scale Integrated Circuit)</td>
<td>~1,000</td>
</tr>
<tr>
<td>1980s VLSI (Very Large Scale IC)</td>
<td>~10,000</td>
</tr>
<tr>
<td>1990s ULSI (Ultra Large Scale IC)</td>
<td>~1,000,000</td>
</tr>
<tr>
<td>2000s ?LSI (? Large Scale IC)</td>
<td>~1000,000,000</td>
</tr>
</tbody>
</table>
MOSFET: Metal Oxide Semiconductor Field Effect Transistor

Use Gate Field Effect for switching

N-MOS (N-type MOSFET)
CMOS

Complimentary MOS

When NMOS is ON, PMOS is OFF
When PMOS is ON, NMOS is OFF
Needless to say, but....

**CMOS Technology:**

Indispensable for our human society

**All the human activities are controlled by CMOS**

living, production, financing, telecommunication, transportation, medical care, education, entertainment, etc.

**Without CMOS:**

There is no computer in banks, and world economical activities immediately stop.

Cellarer phone dose not exists
Downsizing of the components has been the driving force for circuit evolution

<table>
<thead>
<tr>
<th>Year</th>
<th>Components</th>
<th>Size (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1900</td>
<td>Vacuum Tube</td>
<td>10 cm</td>
</tr>
<tr>
<td>1950</td>
<td>Transistor</td>
<td>cm</td>
</tr>
<tr>
<td>1960</td>
<td>IC</td>
<td>mm</td>
</tr>
<tr>
<td>1970</td>
<td>LSI</td>
<td>10 μm</td>
</tr>
<tr>
<td>2000</td>
<td>ULSI</td>
<td>100 nm</td>
</tr>
</tbody>
</table>

In 100 years, the size reduced by one million times. There have been many devices from stone age. We have never experienced such a tremendous reduction of devices in human history.
Downsizing

1. **Reduce Capacitance**
   - Reduce switching time of MOSFETs
   - Increase clock frequency
     - Increase circuit operation speed

2. **Increase number of Transistors**
   - Parallel processing
     - Increase circuit operation speed

Downsizing contribute to the performance increase in double ways

Thus, downsizing of Si devices is the most important and critical issue.
Scaling Method: by R. Dennard in 1974

W_{dep}: Space Charge Region (or Depletion Region) Width

- W_{dep} has to be suppressed
- Otherwise, large leakage between S and D

Leakage current

Potential in space charge region is high, and thus, electrons in source are attracted to the space charge region.

K = 0.7 for example

\[ X, Y, Z : K, \quad V : K, \quad Na : 1/K \]

By the scaling, W_{dep} is suppressed in proportion, and thus, leakage can be suppressed.

\[ W_{dep} \propto \sqrt{V/Na} : K \]

Good scaled I-V characteristics
## Downscaling merit: Beautiful!

### Geometry & Supply voltage

<table>
<thead>
<tr>
<th>Component</th>
<th>Expression</th>
<th>Scaling K: K=0.7 for example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drive current in saturation</td>
<td>$I_d = V_{sat} W_g C_o (V_g - V_{th})$</td>
<td>$C_o$: gate C per unit area</td>
</tr>
<tr>
<td>$I_d$ per unit $W_g$</td>
<td>$I_d/W_g = 1$</td>
<td>$I_d$ per unit $W_g = I_d/W_g$ = 1</td>
</tr>
<tr>
<td>Gate capacitance</td>
<td>$C_g = \varepsilon_o \varepsilon_{ox} L_g W_g / t_{ox}$</td>
<td>$KK/K = K$</td>
</tr>
<tr>
<td>Switching speed</td>
<td>$\tau$</td>
<td>$\tau = C_g V_{dd}/I_d$</td>
</tr>
<tr>
<td>Clock frequency</td>
<td>$f$</td>
<td>$f = 1/\tau = 1/K$</td>
</tr>
<tr>
<td>Chip area</td>
<td>$A_{chip}$</td>
<td>$\alpha$: Scaling factor</td>
</tr>
<tr>
<td>Integration (# of Tr)</td>
<td>$N$</td>
<td>$N \rightarrow \alpha/K^2 = 1/K^2$, when $\alpha=1$</td>
</tr>
<tr>
<td>Power per chip</td>
<td>$P$</td>
<td>$fNCV^2/2 \rightarrow K^{-1}(\alpha K^{-2})K (K^1)^2 = \alpha = 1$, when $\alpha=1$</td>
</tr>
</tbody>
</table>

---

$\alpha$: Scaling factor

In the past, $\alpha > 1$ for most cases

$K^1(K^1)^2 = \alpha = 1$, when $\alpha=1$
<table>
<thead>
<tr>
<th>k= 0.7 and $\alpha = 1$</th>
<th>k= 0.5 and $\alpha = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single MOFET</strong></td>
<td></td>
</tr>
<tr>
<td>Vdd $\rightarrow$ 0.7</td>
<td>Vdd $\rightarrow$ 0.5</td>
</tr>
<tr>
<td>Lg $\rightarrow$ 0.7</td>
<td>Lg $\rightarrow$ 0.5</td>
</tr>
<tr>
<td>Id $\rightarrow$ 0.7</td>
<td>Id $\rightarrow$ 0.5</td>
</tr>
<tr>
<td>Cg $\rightarrow$ 0.7</td>
<td>Cg $\rightarrow$ 0.5</td>
</tr>
<tr>
<td>$P \ (\text{Power})/\text{Clock}$ $\rightarrow$ $0.7^3 = 0.34$</td>
<td>$P \ (\text{Power})/\text{Clock}$ $\rightarrow$ $0.5^3 = 0.125$</td>
</tr>
<tr>
<td>$\tau \ (\text{Switching time})$ $\rightarrow$ 0.7</td>
<td>$\tau \ (\text{Switching time})$ $\rightarrow$ 0.5</td>
</tr>
<tr>
<td><strong>Chip</strong></td>
<td></td>
</tr>
<tr>
<td>N (# of Tr) $\rightarrow$ $1/0.7^2 = 2$</td>
<td>N (# of Tr) $\rightarrow$ $1/0.5^2 = 4$</td>
</tr>
<tr>
<td>f (Clock) $\rightarrow$ $1/0.7 = 1.4$</td>
<td>f (Clock) $\rightarrow$ $1/0.5 = 2$</td>
</tr>
<tr>
<td>P (Power) $\rightarrow$ 1</td>
<td>P (Power) $\rightarrow$ 1</td>
</tr>
</tbody>
</table>
Actual past downscaling trend until year 2000

Change in 30 years

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Ideal scaling</th>
<th>Real Change</th>
<th>Ideal scaling</th>
<th>Real Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_g$</td>
<td>$K$</td>
<td>$10^{-2}$</td>
<td>$I_d$</td>
<td>$K(10^{-2})$</td>
</tr>
<tr>
<td>$t_{ox}$</td>
<td>$K(10^{-2})$</td>
<td>$10^{-2}$</td>
<td>$I_d/\mu m$</td>
<td>$1$</td>
</tr>
<tr>
<td>$V_{dd}$</td>
<td>$K(10^{-2})$</td>
<td>$10^{-1}$</td>
<td>$N$</td>
<td>$\alpha/K^2(10^5)$</td>
</tr>
<tr>
<td>$A_{chip}$</td>
<td>$\alpha$</td>
<td>$10^1$</td>
<td>$P$</td>
<td>$\alpha(10^1)$</td>
</tr>
</tbody>
</table>

Vd scaling insufficient, $\alpha$ increased → N, Id, f, P increased significantly

Past 30 years scaling

Merit: N, f increase
Demerit: P increase

V_{dd} scaling insufficient

Additional significant increase in $I_d$, f, P

Many people wanted to say about the limit. Past predictions were not correct!!

<table>
<thead>
<tr>
<th>Period</th>
<th>Expected limit(size)</th>
<th>Cause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Late 1970’s</td>
<td>1µm:</td>
<td>SCE</td>
</tr>
<tr>
<td>Early 1980’s</td>
<td>0.5µm:</td>
<td>S/D resistance</td>
</tr>
<tr>
<td>Early 1980’s</td>
<td>0.25µm:</td>
<td>Direct-tunneling of gate SiO_2</td>
</tr>
<tr>
<td>Late 1980’s</td>
<td>0.1µm:</td>
<td>‘0.1µm brick wall’(various)</td>
</tr>
<tr>
<td>2000</td>
<td>50nm:</td>
<td>‘Red brick wall’ (various)</td>
</tr>
<tr>
<td>2000</td>
<td>10nm:</td>
<td>Fundamental?</td>
</tr>
</tbody>
</table>
Historically, many predictions of the limit of downsizing. VLSI text book written 1979 predict that 0.25 micro-meter would be the limit because of direct-tunneling current through the very thin-gate oxide.
Finally, there appears to be a fundamental limit of approximately quarter micron channel length, where certain physical effects such as the tunneling through the gate oxide ..... begin to make the devices of smaller dimension unworkable.
Direct-tunneling effect

Potential Barrier

Wave function

Direct tunneling current

Direct tunneling leakage current start to flow when the thickness is 3 nm.
Direct tunneling leakage was found to be OK! In 1994!

MOSFETs with 1.5 nm gate oxide

- $L_g = 10 \mu m$
- $L_g = 5 \mu m$
- $L_g = 1.0 \mu m$
- $L_g = 0.1 \mu m$
Gate leakage: \( I_g \propto \text{Gate Area} \propto \text{Gate length (Lg)} \)

Drain current: \( I_d \propto 1/\text{Gate length (Lg)} \)

Lg \( \to \) small,
Then, \( I_g \to \) small, \( I_d \to \) large,
Thus, \( I_g/I_d \to \) very small
Never Give Up!

No one knows future!

There would be a solution!

Think, Think, and Think!

Or, Wait the time!
Some one will think for you
Transistor Scaling Continues

Qi Xinag, ECS 2004, AMD
Downsizing limit?

Electron wave length

Channel length?

10 nm
5 nm gate length CMOS is a Real Nano Device!!

Length of 18 Si atoms

H. Wakabayashi et al, NEC

IEDM, 2003
Electron wave length

10 nm

Tunneling distance

3 nm

Downsizing limit!
Channel length
Gate oxide thickness
Prediction now!

Electron wave length
10 nm

Tunneling distance
3 nm

Atom distance
0.3 nm

MOSFET operation
Lg = 3 nm?

Below this, no one knows future!
However, Gate oxide thickness 2 orders magnitude smaller Close to limitation!!

Lg: Gate length downsizing will continue to another 10-15 years
0.8 nm Gate Oxide Thickness MOSFETs operates!!

0.8 nm: Distance of 3 Si atoms!!

- 1.2nm physical SiO2 in production (90nm logic node)
- 0.8nm physical SiO2 in research transistors

By Robert Chau, IWGI 2003
So, we are now in the limitation of downsizing?

Do you believe this or do not?
There is a solution! To use high-k dielectrics

K: Dielectric Constant

Thin gate SiO$_2$  Thick gate high-k dielectrics

Almost the same electric characteristics

However, very difficult and big challenge!

Remember MOSFET had not been realized without Si/SiO$_2$!
Choice of High-k elements for oxide

Candidates

Unstable at Si interface
- Si + MOₓ M + SiO₂
- Si + MOₓ MSiₓ + SiO₂
- Si + MOₓ M + MSiₓOᵧ

Gas or liquid at 1000 K
- Radio active

He
- He

Li Be Mg Na Ca Sc K Ti V Cr Mn Fc Co Ni Cu Zn Ga Ge As Se Br Kr

Sr Y Zr

Cs Ba Hf

Fr Ra Rf Ha Sg Ns Hs Mt

La Ce Pr Nd Pm Sm Eu Gd Tb Dy Ho Er Tm Yb Lu

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
Band Offsets

Calculated value

Dielectric constant

SiO2; 4
Si3N4: ~ 7
Al2O3: ~ 9
HfO2; ~23
Y2O3; ~10
Gd2O3: ~10
La2O3: ~27

HfO2 was chosen for the 1st generation
La2O3 is more difficult material to treat
### Dielectric Constant vs. Band Offset (Measured)

<table>
<thead>
<tr>
<th>Material</th>
<th>Dielectric Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiO₂</td>
<td>3.9</td>
</tr>
<tr>
<td>AlₓSiᵧO₂z</td>
<td></td>
</tr>
<tr>
<td>(Ba,Sr)TiO₃</td>
<td>200-300</td>
</tr>
<tr>
<td>BeAl₂O₄</td>
<td>8.3-9.43</td>
</tr>
<tr>
<td>CeO₂</td>
<td>16.6-26</td>
</tr>
<tr>
<td>CeHfO₄</td>
<td>10-20</td>
</tr>
<tr>
<td>CoTiO₃/Si₃N₄</td>
<td></td>
</tr>
<tr>
<td>EuAlO₃</td>
<td>22.5</td>
</tr>
<tr>
<td>HfO₂</td>
<td>26-30</td>
</tr>
<tr>
<td>Hf silicate</td>
<td>11</td>
</tr>
<tr>
<td>La₂O₃</td>
<td>20.8</td>
</tr>
<tr>
<td>LaScO₃</td>
<td>30</td>
</tr>
<tr>
<td>La₂SiO₅</td>
<td></td>
</tr>
<tr>
<td>MgAl₂O₄</td>
<td></td>
</tr>
<tr>
<td>NdAlO₃</td>
<td>22.5</td>
</tr>
<tr>
<td>PrAlO₃</td>
<td>25</td>
</tr>
<tr>
<td>Si₃N₄</td>
<td>7</td>
</tr>
<tr>
<td>SmAlO₃</td>
<td>19</td>
</tr>
<tr>
<td>SrTiO₃</td>
<td>150-250</td>
</tr>
<tr>
<td>Ta₂O₅</td>
<td>25-24</td>
</tr>
<tr>
<td>Ta₂O₅-TiO₂</td>
<td></td>
</tr>
<tr>
<td>TiO₂</td>
<td>86-95</td>
</tr>
<tr>
<td>TiO₂/Si₃N₄</td>
<td></td>
</tr>
<tr>
<td>Y₂O₃</td>
<td>8-11.6</td>
</tr>
<tr>
<td>YₓSiᵧO₂z</td>
<td></td>
</tr>
<tr>
<td>ZrO₂</td>
<td>22.2-28</td>
</tr>
<tr>
<td>Zr-Al-O</td>
<td></td>
</tr>
<tr>
<td>Zr silicate</td>
<td></td>
</tr>
<tr>
<td>(Zr,Sn)TiO₄</td>
<td>40-60</td>
</tr>
</tbody>
</table>

\[ \sqrt{\phi_B} \ast k \] : Figure of Merit of High-k

---


T. Hattori, INFOS , 2003
CMOS Technology:
Indispensable for our human society

All the human activities are controlled by CMOS
living, production, financing, telecommunication, transportation, medical care, education, entertainment, etc.

Without CMOS:
There is no computer in banks, and world economical activities immediately stop.

Cellarer phone dose not exists
In 100 years, the size reduced by one million times. There have been many devices from stone age. We have never experienced such a tremendous reduction of devices in human history.
Downsizing

1. **Reduce Capacitance**
   - Reduce switching time of MOSFETs
   - Increase clock frequency
     - Increase circuit operation speed

2. **Increase number of Transistors**
   - Parallel processing
     - Increase circuit operation speed

---

Downsizing contribute to the performance increase in double ways

Thus, downsizing of Si devices is the most important and critical issue.
Most recent SD Card
Most Recent SD Card

128GB = 128Gbite
= 128G X 8bit = 1024Gbit
= 1.024T(Tera)bit

1T = 10^{12} = 1Trillion

World Population : 6 Billion
Brain Cell : 10 to 100 Billion
Stars in Galaxy : 100 Billion
Most Recent SD Card
2.4cm X 3.2cm X 0.21cm

Volume : 1.6cm³

Weight: 2g

Voltage: 2.7 - 3.6V

Old Vacuum Tube:
5cm X 5cm X 10cm, 100g, 100W

1Tbit = 10k X 10k X 10k bit

Volume = 0.5km X 0.5km X 1km
= 0.25 km³ = 0.25 X 10¹² cm³

Weight = 0.1 kg X 10¹² = 0.1 X 10⁹ ton = 100 M ton

Power = 0.1 kW X 10¹² = 50 TW

Supply Capability of Tokyo Electric Power Company: 55 BW
Question:

How far we can go with downscaling?
How far can we go?

Past: 0.7 times per 3 years

1973年

8µm → 6µm → 4µm → 3µm → 2µm → 1.2µm → 0.8µm → 0.5µm

→ 0.35µm → 0.25µm → 180nm → 130nm → 90nm → 65nm → 45nm

Now

In 40 years: 15 generations,
Size 1/200, Area 1/40,000

Future

→ 32nm → 22nm → 16nm → 11.5 nm → 8nm → 5.5nm? → 4nm? → 2.9 nm?

- At least 5.6 generations, for 15 ~ 20 years
- Hopefully 8 generations, for 30 years
**HP, LOP, LSTP for Logic CMOS**

![Graph showing HP, LOP, and LSTP CMOS technologies with their respective characteristics.](image)

- **HP CMOS (high Performance)**
  - Highest Ion, Lowest CV/I
  - High leakage
  - Medium Vdd

- **LOP CMOS (Low Operation Power)**
  - Lowest Vdd
  - Medium Ion, *medium CV/I*
  - Medium leakage

- **LSTP CMOS (Low Standby Power)**
  - Lowest leakage
  - Low Ion, *high CV/I*
  - High Vdd

Source: 2007 ITRS Winter Public Conf.
Scaling Method: by R. Dennard in 1974

**Wdep**: Space Charge Region (or Depletion Region) Width

Wdep has to be suppressed. Otherwise, large leakage between S and D.

Potential in space charge region is high, and thus, electrons in source are attracted to the space charge region.

K=0.7 for example

By the scaling, Wdep is suppressed in proportion, and thus, leakage can be suppressed.

**Good scaled I-V characteristics**

\[ W_{dep} \propto \sqrt{V/Na} : K \]

\[ X, Y, Z : K, \quad V : K, \quad Na : 1/K \]
Scaling of high beyond 0.5 nm is important

Power of FET = $CV^2/2 \cdot D^3 (=L^3)$

Problems
→ SCE
→ Variation in Vth
→ Increase in Off-leakage current

Solution

ITRS2007
Down scaling is the most effective way of Power saving.

It has been always discussed about the limit of downscaling, but the down scaling of MOSFETs is still possible for another 10 or 20 years!

3 important technological items for DS.

1. Thinning of high-k beyond 0.5 nm

2. Metal S/D

3. Si-Nanowire FET
High k gate stack
ITRS

Vdd stay high

Year

Vdd (V)

2008 update


2001

2008


2008 update (bulk)

2008 update (UTB)

2008 update (DG)

2007 (bulk)

2007 (UTB)

2007 (DG)

2005 (bulk)

2005 (UTB)

2005 (DG)

2003

2001

1999

ITRS EOT limit = 0.5 nm?

For HP Logic Delay

Saturation

Is 0.5nm real limit?
To use high-k dielectrics

Thin gate SiO$_2$

Almost the same electric characteristics

Thick gate high-k dielectrics

K: Dielectric Constant

However, very difficult and big challenge!

Remember MOSFET had not been realized without Si/SiO$_2$!
### Choice of High-k elements for oxide

**Candidates**

| H   | Li | Be   | Mg   | Na  | K   | Ca  | Sc  | Ti   | V   | Cr  | Mn  | Fe  | Co  | Ni  | Cu  | Zn  | Ga  | Ge  | As  | Se  | Br  | Kr |
|-----|----|------|------|-----|-----|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|     |    |      |      |     |     |     |     |      |     |     |     |     |     |     |     |     |     |     |     |     |     |

**Unstable at Si interface**

- Si + MOₓ M + SiO₂
- Si + MOₓ MSiₓ + SiO₂
- Si + MOₓ M + MSiₓOᵧ

**Gas or liquid at 1000 K**

**Radio active**

- He

**Active**

- B
- C
- N
- O
- F
- Ne

**Al Si P S Cl Ar**

**Candidates**

- Na Al Si P Sc Ar
- K Sc Ti V Cr Mn Fe Co Ni Cu Zn Ga Ge As Se Br Kr
- Sr Y Zr
- Cs Ba Hf
- Fr Ra Rf Ha Sg Np U
- La Ce Pr Nd Pm Sm Eu Gd Tb Dy Ho Er Tm Yb Lu

- Ac Th Pa U Np Pu Am Cm Bk Cf Es Fm Md No Lr

- La₂O₃ based dielectrics are thought to be the next generation materials, which may not need a thicker interfacial layer.

- HfO₂ based dielectrics are selected as the first generation materials, because of their merit in:
  1) band-offset,
  2) dielectric constant,
  3) thermal stability.

---

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

Leakage Current by Tunneling

XPS measurement by Prof. T. Hattori, INFOS 2003
High-k gate insulator MOSFETs for Intel: EOT=1nm

EOT: Equivalent Oxide Thickness
For the past 45 years, SiO₂ and SiON have been used as gate insulators. Today, EOT = 1.0 nm. The EOT limit is 0.7~0.8 nm. One order of magnitude reduction in EOT is possible. EOT can be reduced further beyond 0.5 nm by using direct contact to Si and choosing appropriate materials and processes.
Cluster tool for high-k thin film deposition
Reports on direct contact of high-k/Si


Control of oxygen atoms

Direct HfO₂/Si structure

Our approach

K. Kakushima, et al., ESSDERC2009

W

La₂O₃  k=23

La-silicate  k=8~14

Silicate formation

La₂O₃+Si+nO₂

La₂SiO₅, La₁₀(SiO₄)₆O₃
La₉.₃₃Si₆O₂₆, La₂Si₂O₇

La₂O₃ can easily achieve direct contact of high-k/Si

2011-6-20  Event, Venue information
Gate Leakage vs EOT, (Vg=|1|V)
EOT = 0.48 nm

Transistor with La2O3 gate insulator

Our results
EOT=0.37nm
La2O3

EOT=0.37nm    EOT=0.40nm    EOT=0.48nm

Vth=-0.06V    Vth=-0.05V    Vth=-0.04V

W/L = 50µm /2.5µm

Vd (V)   Vd (V)   Vd (V)

0.48 → 0.37nm Increase of Id at 30%
$\mu_{\text{eff}}$ of W/La$_2$O$_3$ and W/HfO$_2$ nFET on EOT

- W/La$_2$O$_3$ exhibits higher $\mu_{\text{eff}}$ than W/HfO$_2$
- $\mu_{\text{eff}}$ start degrades below EOT=1.4nm
FET characteristics of W/La$_2$O$_3$ on EOT

All characteristics start to degrade or shift below EOT=1.4nm

- Aggressive $N_{fix}$ generation at EOT<1.2nm
- $N_{fix} = 7 \times 10^{12}$ cm$^{-2}$

$W/L_g=50\mu$m/2.5$\mu$m

$N_{fix}$ and Dit

Si sub. metal

All characteristics start to degrade or shift below EOT=1.4nm
Gate Metal Induced Defects Compensation

Suppression of aggressive shift in $V_{fb}$

TEM  EDX

Metal Gate
MgO
La2O3
Si

Gate Metal Induced Defects Compensation

Suppression of aggressive shift in $V_{fb}$
Mobility Improvement with Mg Incorporation

Recovery of $\mu_{\text{eff}}$ mainly at low $E_{\text{eff}}$
Si nanowire FET
Because of off-leakage control,

Planar $\rightarrow$ Fin $\rightarrow$ Nanowire
Nanowire FET

ITRS 2009

Bulk  Fin  Nanowire

Scaling Pathways

w and w/o 3rd gate?

Bulk or SOI  Fin

Si Nanowire

First Year of IC Production


Device Structures
Planar bulk
FDSOI
Multiple Gate

2015
22 or 16nm node
Multiple Gate (Fin) FET

2020
11 or 8nm node
Nanowire FET
Si nanowire FET as a strong candidate

1. Compatibility with current CMOS process
2. Good controllability of $I_{OFF}$
3. High drive current

1D ballistic conduction

Multi quantum Channel

High integration of wires

Leakage current

Gate: OFF

source

drain

cut-off
Off Current

- bulk
- FinFET
- SiNWFET
- GeNWFET
- ITRS (Planar)
- ITRS (SOI)
- ITRS (DG)

- Off Current (nA/μm)
- Ion Current (μA/μm)

- Bulk
- DG
- ITRS (SOI)
- ITRS (DG)

- Si Nanowire (dia~3nm)
- Si Nanowire (dia~10nm)
Increase the Number of quantum channels

By Prof. Shiraishi of Tsukuba univ.

Energy band of Bulk Si

Energy band of 3 x 3 Si wire

4 channels can be used
Maximum number of wires per 1 μm

Front gate type MOS
- 165 wires /μm

Surrounded gate type MOS
- 33 wires/μm

6nm pitch
- By nano-imprint method

Metal gate electrode (10nm)

High-k gate insulator (4nm)

Si Nano wire (Diameter 2nm)

30nm pitch:
- EUV lithography

Surrounded gate MOS

Source
Gate
Drain
Device fabrication

Si/Si$_{0.8}$Ge$_{0.2}$ superlattice epitaxy on SOI

Anisotropic etching of these layers

Isotropic etching of SiGe

The NW diameter is controllable down to 5 nm by self limited oxidation.

SiN HM

Gate depositions
HfO$_2$ (3nm)
TiN (10nm)
Poly-Si (200nm)

Gate etching

S/D implantation
Spacer formation
Activation anneal
Salicidation

Standard Back-End of-Line Process

Process Details:
C. Dupre et al.,
3D-stacked Si NWs with Hi-\textit{k}/MG

Wire direction : \textit{<110>}
50 NWs in parallel
3 levels vertically-stacked
Total array of 150 wires
EOT \sim 2.6 \text{ nm}
SiNW Band structure calculation
Cross section of Si NW

First principal calculation,

D=1.96nm [001]  
D=1.94nm [011]  
D=1.93nm [111]
Si nanowire FET with 1D Transport

Orientation  [001]  [011]  [111]
Diameter (nm) 0.86 0.94 0.89

Energy (eV)
0
-1
0
1
G Z G
Wave Number (a)

Orientation  [001]  [011]  [111]
Diameter (nm) 3.00 3.94 1.93

Energy (eV)
0
-1
0
1
G Z G
Wave Number (b)

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)

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

Higher-order finite difference

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

Integration

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

Iwata et al, J. Comp. Phys., to be published

Real-Space Density-Functional Theory code (RSDFT)

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

Highly tuned for massively parallel computers

Compared on a massively-parallel cluster PACS-CS at University of Tsukuba.
(Theoretical Peak Performance = 5.6GFLOPS/node)

e.g.) The system over 10,000 atoms \( \text{Si}_{10701}\text{H}_{1996} \) (7.6 nm diameter Si dot)

Convergence behavior for \( \text{Si}_{10701}\text{H}_{1996} \)

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

Computational Time (with 1024 nodes of PACS-CS)

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

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

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

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

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

- KS band gap = 0.608 eV (@\(\Gamma\))

Each band is 4-dgenerate.

Effective mass equation

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

The band structure can be understood that electrons near the CBM in the bulk Si are Confined within a cylindrical geometry.
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

Si nano wire with surface roughness
SiNW Band compact model
Landauer Formalism for Ballistic FET

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

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 $E$

Transmission Probability to Drain

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

Injection from Drain $= 0$
Résumé of the Compact Model

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

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

\[ \mu_S - \mu_D = qV_D \]

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

Planar Gate

GAA

(Electrostatics requirement)

\[ |Q_f + Q_b| = \frac{q}{\pi} \sum_i g_i \left[ \int_{-\infty}^{\infty} \frac{dk}{1 + \exp \left( \frac{\varepsilon_i(k) - \mu_S}{k_B T} \right)} - \int_{-\infty}^{0} \frac{1}{1 + \exp \left( \frac{\varepsilon_i(k) - \mu_S}{k_B T} \right)} - \frac{1}{1 + \exp \left( \frac{\varepsilon_i(k) - \mu_D}{k_B T} \right)} \right] T_i(\varepsilon_i(k))dk \]

(Carrier distribution in Subbands)

\[ T(\varepsilon) = \frac{\sqrt{2D_0} qE}{\left( \sqrt{B_0 + D_0} + \sqrt{D_0} \right) qE + \sqrt{2mD_0 B_0} \ln \left( \frac{qEx_0 + \varepsilon}{\varepsilon} \right)} \]

Unknowns are \( I_D, (\mu_S - \mu_0), (\mu_D - \mu_0), (Q_f + Q_b) \)
Electric current: 20~25 μA
No saturation at Large \( V_D \)
SiNW FET Fabrication
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

SiNW

SiN support

30nm

Poly-Si

SiN

Nanowire

500nm
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
<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>10x20</td>
<td>10x20</td>
<td>14</td>
<td>10</td>
<td>10</td>
<td>12</td>
<td>13x20</td>
<td></td>
</tr>
<tr>
<td>Lg (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>EOT or Tox (nm)</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>Vdd (V)</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) per wire</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 dia.</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>Ion(uA/um) by cir.</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>SS (mV/dec.)</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>DIBL (mV/V)</td>
<td>62</td>
<td>56</td>
<td>15</td>
<td>13</td>
<td>22</td>
<td>40-82</td>
<td>65</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

**$I_{ON}/I_{OFF}$ Bench mark**

![Graph](image)

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

  - Voltage: $1.0 \sim 1.1 \text{V}$

- **Si Nano-wire 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

  - Voltage: $1.2 \sim 1.3 \text{V}$

- **$L_g=500 \sim 65 \text{nm}$**
Electron Density
($x10^{19}$ cm$^{-3}$)

Edge portion

Flat portion

Distance from SiNW Surface (nm)
Primitive estimation!

- Compact model
- Small EOT for high-k (33)
- P-MOS improvement (26)
- Low S/D resistance (20)
- # of wires /1µm (15)
- SiNW (12nm × 19nm)
- Assumption
  \[ I_{ON} \propto L_g^{-0.5} \cdot T_{ox}^{-1} \]

**Nanowire**

- **pMOS**
  - Improved with low S/D resistance

**ITRS**

**Small EOT for high-k**

**Compact model**

**SiNW** (12nm × 19nm)
Transistor Innovations Enable Technology Cadence

- 2003: 90 nm - Invented SiGe Strained Silicon
- 2005: 65 nm - 2nd Gen SiGe Strained Silicon
- 2007: 45 nm - Invented Gate-Last High-k Metal Gate
- 2009: 32 nm - 2nd Gen, Gate-Last High-k Metal Gate
- 2011: 22 nm - First to Implement Tri-Gate

Strained Silicon
High k Metal gate
Tri-Gate
22 nm 3-D Tri-Gate Transistor
Energy-Efficient Performance Built on Moore’s Law

- Lower Transistor Leakage
- Higher Transistor Performance (Switching Speed)

Source: Intel

Active Power per Transistor (normalized)

- 65nm Planar
- 45nm Planar
- 32nm Planar
- 22nm Tri-Gate

22 nm Tri-Gate transistors increase the benefit from a new technology generation

> 50% reduction
22nm Silicon Technology Breakthrough Benefits Broad Range of Intel Architecture Devices

New 22nm 3-D transistors deliver unprecedented performance improvement and power reduction for Intel’s product portfolio

- This benefits smallest handhelds to powerful cloud-based servers
- 37% performance increase at low voltage vs. 32nm planar transistors*
- Consumes only half the power at the same performance level as 2-D transistors on 32nm planar chips*
Our roadmap for R & D

Source: H. Iwai, IWJT 2008

Current Issues

**Si Nanowire**
- Control of wire surface property
- Source Drain contact
- Optimization of wire diameter
- Compact I-V model

**III-V & Ge Nanowire**
- High-k gate insulator
- Wire formation technique

**CNT:**
- Growth and integration of CNT
- Width and Chirality control
- Chirality determines conduction types: metal or semiconductor

**Graphene:**
- Graphene formation technique
- Suppression of off-current
  - Very small bandgap or no bandgap (semi-metal)
- Control of ribbon edge structure which affects bandgap

---

Extended CMOS: More Moore + CMOS logic

PJT (2007~2012)

Si Channel

Si Fin, Tri-gate

Si Nano wire

III-V and Ge Nano wire

Nanowire

Problem: High-k gate oxides, etching of III-V wire

Mechanical Stress, Roughness, Surface control

Selection

Graphene

Graphene formation technique

Control of wire surface property

Compact I-V model

Wire formation technique

Suppression of off-current

Very small bandgap or no bandgap (semi-metal)

Control of ribbon edge structure which affects bandgap
Thank you for your attention!