Future of NanoCMOS after Scaling Limit

September 8, 2008

@Auditorio Marino Troncoso
Pontificia Universidad Javeriana,
Bogota, Columbia

Hiroshi Iwai,
Tokyo Institute of Technology
• 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
4 wives of Lee De Forest

1906 Lucille Sheardown
1907 Nora Blatch
1912 Mary Mayo, singer
1930 Marie Mosquini, silent film actress

Mary

Marie
First Computer Eniac: made of huge number of vacuum tubes. Big size, huge power, short life time filament.

à 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.
History of Semiconductor devices

1947, 1\textsuperscript{st} Point Contact Bipolar Transistor:  
Ge Semiconductor, Bardeen, Brattin  
\rightarrow Nobel Prize

1948, 1\textsuperscript{st} Junction Bipolar Transistor,  
Ge Semiconductor, Schokley  
\rightarrow Nobel Prize

1958, 1\textsuperscript{st} Integrated Circuits,  
Ge Semiconductor, J.Kilby  
\rightarrow Nobel Prize

1959, 1\textsuperscript{st} Planar Integrated Circuits,  
Noice

1960, 1\textsuperscript{st} MOS Transistor, Kahng,  
Si Semiconductor

1963, 1\textsuperscript{st} CMOS Circuits, C.T. Sah and F. Wanlass
History of Electronic Devices

1900
- Vacuum tube
- Diode
- Triode

20
- MOSFET
- MISFET

30
- MOSFET
- Bipolar

50
- Solid-State Circuits
- IC
- LSI

60
- Si-MOSFET
- IC
- LSI

70
- Triode
- Diode
- 1st Electronic circuits

2000
- ULSI
- VLSI

- Low Power
- High speed
- High integration
- Silicon Technology
- High Integration
- Low Power
J. E. LILIENFELD

DEVICES FOR CONTROLLED ELECTRIC CURRENT

Filed March 28, 1928

J. E. LILIENFELD
Capacitor structure with notch

**Negative bias**

- Gate Electrode
- Gate Insulator
- Semiconductor
- Electron

No current

**Positive bias**

- Electric field

Current flows
Use Gate Field Effect for switching

Gate Electrode Poly Si
Gate Insulator SiO₂
Substrate Si

MOSFET: Metal Oxide Semiconductor Field Effect Transistor

Source
Channel
N-MOS (N-type MOSFET)

Si Substrate
Drain
Surface Potential (Negative direction)

0 bias for gate

Positive bias for gate

Electron flow
N-MOS (N-type MOSFET)
Source

Gate

Drain

Current flow

Electron flow

P-MOS (P-type MOSFET)
Source

Gate

Drain

Current flow

Hole 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.

Jack S. Kilby
1960: First MOSFET by D. Kahng and M. Atalla

Top View

Si

Source

Al Gate

Drain

Si

Si/SiO₂ Interface is extraordinarily good

Al

SiO₂

Si
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>~</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>~1,000,000</td>
</tr>
</tbody>
</table>
MOSFET: Metal Oxide Semiconductor Field Effect Transistor

Use Gate Field Effect for switching

Gate Electrode: Poly Si
Gate Insulator: SiO₂
Substrate: Si

Source
Channel
Drain
Si Substrate

N-MOS (N-type MOSFET)
N-MOS (N-type MOSFET) 
Source

Gate

Drain

Electron flow

Current flow

P-MOS (P-type MOSFET) 
Source

Gate

Drain

Hole flow

Current flow
n-MOSFET

OFF

Electrons

ON

High Potential Region
n-MOSFET

- Drain Current (A)
- Positive voltage (V)
- Gate bias

Threshold voltage

OFF

p-MOSFET

- Drain Current (A)
- Negative voltage (V)
- Gate bias

Threshold voltage

OFF
CMOS

Complimentary MOS

Inverter

PMOS

NMOS

When NMOS is ON, PMOS is OFF

When PMOS is ON, NMOS is OFF
Input  Output

PMOS (OFF)

NMOS (ON)
Inverter

Input  Output

[Diagram showing circuit with inverters and connections]
Oscillator
Latch (Memory)

\[ \begin{array}{c}
1 & \rightarrow & 0 & \rightarrow & 1 \\
1 & \rightarrow & 0 & \rightarrow & 1 \\
\end{array} \]
CMOS: Low Power: No DC current from Power supply to the ground

\[ P = \frac{1}{2} CV_D^2 \]

\[ P = \frac{1}{2} fCV_D^2 \]
2 input NAND Circuit

AND
Input 1   1 1 0 0
Input 2   1 0 1 0
Output    1 0 0 0

NAND = NOT \cdot AND
Input 1   1 1 0 0
Input 2   1 0 1 0
Output    0 1 1 1
Needless to say, but....

**CMOS Technology:**
Indispensable for our human society

All 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
Exponential Cost Reduction

Cost per Transistor

As the number of transistors goes UP
Cost per transistor goes DOWN

Source: WSSS/Dataman/Intel

SPIE Microlithography 2006, San Jose, CA USA
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 MOSFE
   - Reduce power consumption

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

Thus, downsizing of Si devices is the most important and critical issue.
Keep increase of the number of components. Cost per components decreases!

Gordon Moore

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 SiC</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 and fluctuations in the positions of impurities in the depletion layers begin to make the devices of smaller dimension unworkable.
Direct-tunneling effect

Potential Barrier

Wave function
Direct tunneling leakage was found to be OK! In 1994 MOSFETs with 1.5 nm gate oxide

Direct tunneling leakage current start to flow when the thickness is 3 nm.

Lg = 10 µm  
Lg = 5 µm  
Lg = 1.0 µm  
Lg = 0.1µm
Gate leakage: \( I_g \propto \text{Gate Area} \propto \text{Gate length (Lg)} \)

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

Lg \( \rightarrow \) small,
Then, \( I_g \rightarrow \) small, \( I_d \rightarrow \) large,   Thus, \( I_g/I_d \rightarrow \) very small

Gate leakage current start to flow when the thickness is 3 nm.
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

90nm node
Lg=50nm

65nm node
Lg=35nm

45nm node
Lg=25nm

32nm node
Lg=15nm

22nm node
Lg=10nm

~30% every two years

ITRS 2003 (HP)
Downsizing limit? Channel length?

Electron wave length

10 nm

Gate Oxd

Channel
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

Atom distance: 0.3 nm

Gate oxide thickness

Channel length

Downsizing limit!
Prediction now!

Electron wave length
10 nm

Tunneling distance
3 nm

Atom distance
0.3 nm

MOSFET operation
Lg = 2 ~ 1.5 nm?

Below this, no one knows future!
Maybe, practical limit around 5 nm.

When Gate length Smaller,  
→ Subthrehold Leakage Current Larger

Subtheshould Leakage Current

Subthreshold Current Is OK at Single Tr.
But not OK For Billions of Trs.
We have to reduce the Supply voltage.

Then Vth should be lowered.

\( Vg = 0V \)

**Subthreshold leakage current increase**

- \( 10^{-6}A \)
- \( 10^{-7}A \)
- \( 10^{-8}A \)
- \( 10^{-9}A \)
- \( 10^{-10}A \)

**Vth lowering**
Prediction now!

Electron wave length
10 nm

Tunneling distance
3 nm

Atom distance
0.3 nm

Practical limit for integration
Lg = 5 nm?

MOSFET operation
Lg = 2 ~ 1.5 nm?

Below this, no one knows future!
However, Gate oxide thickness is 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 operate

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 200
So, we are now in the limitation of downsizing?

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

Almost the same electric characteristics

However, very difficult and big challenge!
Remember MOSFET had not been realized without Si/SiO₂!
### Historical Trend of New Material for Gate Stack

<table>
<thead>
<tr>
<th>Year</th>
<th>MOSFET</th>
<th>Gate Stack in production</th>
<th>PMO</th>
<th>NMOS</th>
<th>CMOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>MOSFET</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1970</td>
<td>1st FET IC</td>
<td></td>
<td>SPM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td>LSI</td>
<td>SiO$_2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td></td>
<td>SiO$_x$N$_y$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td>Hf base</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Gate insulator:**
- SiO$_2$
- SiO$_x$N$_y$
- Hf base

**Gate electrode:**
- Al
- N$^+$Poly Si
- Double Poly Si
- Double Metal
- MoSi$_2$ → WSi$_2$ TiSi$_2$ CoSi$_2$ NiSi

**Silicide above Poly Si electrode:**
- Ta$_2$O$_5$
- Al$_2$O$_3$

**R&D for high-k MOSFET:**
- Ni$_3$Si$_4$
- Al$_2$O$_3$, ZrO$_2$

**Stack NO(Si$_3$N$_4$/SiO$_2$):**
- (O)NO, Ni$_3$Si$_4$
- Ta$_2$O$_5$, Al$_2$O$_3$,

**DRAM Capacitor:**
- (O)NO, Ni$_3$Si$_4$
- Ta$_2$O$_5$, Al$_2$O$_3$,

**NV Memory:**
- NO, AO (Al$_2$O$_3$/SiO$_2$)
- (O)NO, Si$_3$N$_4$

**Analog/RF:**
- Ta$_2$O$_5$, 

---

The diagram illustrates the historical trend of new materials for gate stacks in MOSFETs, LSI, and related devices, showing the transition from SiO$_2$ to SiO$_x$N$_y$ and beyond, with developments in gate electrode and silicide materials.
Choice of High-k elements for oxide

<table>
<thead>
<tr>
<th>Candidates</th>
<th>Unstable at Si interface</th>
<th>Gas or liquid at 1000 K</th>
<th>Radio active</th>
<th>HfO₂ based dielectrics are selected as the first generation materials, because of their merit in 1) band-offset, 2) dielectric constant, 3) thermal stability</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>Si + MOₓ M + SiO₂</td>
<td></td>
<td>He</td>
<td></td>
</tr>
<tr>
<td>Li B Be</td>
<td>Si + MOₓ MSiₓ + SiO₂</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Na Mg Al Si P S Cl Ar</td>
<td>Si + MOₓ M + MSiₓOᵧ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K Sc Ti V Cr Mn Fc Co Ni Cu Zn Ge As Se Br Kr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rh Sr Y Zr</td>
<td>Na Al Si P S Cl Ar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cs Ba Hf</td>
<td>Si + MOₓ M + MSiₓOᵧ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>La Ce Pr Nd Sm Eu Gd Tb Dy Ho Er Tm Yb Lu</td>
<td></td>
<td></td>
<td></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>Ac Th Pa U Np Pu Am Cm Bk Cf Es Fm Md No Lr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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] vs. Dielectric Constant $\varepsilon(0)$

XPS measurement by Prof. T. Hattori, INFOS 2003
Intel’s announcement, January 26, 2007, and IEDM Dec 2007

Hafnium-based high-k material by ALD: EOT= 1nm
Specific gate metals ( Intel’s trade secret)
  Different Metals for NMOS and PMOS
Use of 193nm dry lithography

From 65 nm to 45 nm Tech.
  Tr density: 2 times increase
  Tr switching power: 30% reduction
  Tr switching speed: 20% improvement
  S-D leakage power: 5 times reduction
  Gate oxide leakage: 10 times reduction

45nm processors (Core™2 family processors "Penryn") running Windows* Vista*, Linux* etc.

45nm production in the second half of 2007
High-k gate insulator MOSFETs for Intel: EOT=1nm

EOT: Equivalent Oxide Thickness
History and future of Transistor
Shrinking, Shrinking, and Shrinking!
and then, Shrinking, Shrinking, and Shrinking!

\[ C, V \propto L \quad \text{C: Capacitance \ V: Voltage} \]

Switching speed \( CV/I \) \( \rightarrow \) Decrease

Power consumption \( CV^2/2 \) \( \rightarrow \) Decrease

Integration density: \( 1/L^2 \) \( \rightarrow \) Increase

1970 \quad 2007

Gate length \quad 10,000 \text{ nm} \quad 25 \text{ nm}

Gate Oxd Thickness \quad 100 \text{ nm} \quad 1 \text{ nm}
For the past 45 years, SiO2 and SiON have been used as gate insulators. Currently, EOT=1.0nm. The EOT limit is 0.7~0.8nm. By introducing High-k materials and choosing appropriate materials and processes, EOT can be reduced further beyond 0.5nm with direct contact to Si.
### Choice of High-k elements for oxide

**Candidates**

<table>
<thead>
<tr>
<th>Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
</tr>
<tr>
<td>Li</td>
</tr>
<tr>
<td>Be</td>
</tr>
<tr>
<td>Na</td>
</tr>
<tr>
<td>K</td>
</tr>
<tr>
<td>Rh</td>
</tr>
<tr>
<td>Cs</td>
</tr>
<tr>
<td>Fr</td>
</tr>
<tr>
<td>La</td>
</tr>
<tr>
<td>Ac</td>
</tr>
</tbody>
</table>

**Unstable at Si interface**

- Si + MO<sub>x</sub> M + SiO<sub>2</sub>
- Si + MO<sub>x</sub> MSi<sub>x</sub> + SiO<sub>2</sub>
- Si + MO<sub>x</sub> M + MSi<sub>x</sub>O<sub>y</sub>

**La<sub>2</sub>O<sub>3</sub> based dielectrics are thought to be the next generation materials, which may not need a thicker interfacial layer.**

**HfO<sub>2</sub> 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
Gate Leakage vs EOT, (Vg=|1|V)

- Al₂O₃
- HfAlO(N)
- HfO₂
- HfSiO(N)
- HfTaO
- La₂O₃
- Nd₂O₃
- Pr₂O₃
- PrSiO
- PrTiO
- SiON/SiN
- Sm₂O₃
- SrTiO₃
- Ta₂O₅
- TiO₂
- ZrO₂(N)
- ZrSiO
- ZrAlO(N)
EOT = 0.48 nm  
Our results

Transistor with La2O3 gate insulator
EOT=0.37nm
La2O3

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

W/L = 50µm /2.5µm

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

0.48 → 0.37nm Increase of Id at 30%
New material research will give us many future possibilities and the most important for Nano-CMOS! Not only for high-k!

- New material for Metal gate electrode
- New material for High-k gate dielectric
- New channel material
- New material For Metal S/D
New materials are important for Not only nano-CMOS logic MOSFETs, But also for new memories!

Limit of high-density memories, such as flash, DRAM will be solved by new materials.

Flash: floating gate $\rightarrow$ gate insulator charge trap like SONOS, MNOS

DRAM $\rightarrow$ New high-k insulator

New memory $\rightarrow$ PRAM, RRAM

Flash

DRAM

PRAM, RRAM
Due to design productivity, yield, and power

1970s: 10 years after single MOSFETs,

PMOS 1kbit
DRAM
Toshiba(1974)

NMOS 1k bit
SRAM  Toshiba
(1974)
1k SRAM
2 inch wafer
1974

64k DRAM
3 inch wafer
1979

64k DRAM
4 inch wafer
1980

1k SRAM
2 inch wafer
1974
1970’s

Toshiba Corporation
300 mm Fab TSMC

Now

Toshiba Oita Works

300 mm Super clean room in Tsukuba, Selete
In a future
No person is necessary!
6 µm NMOS LSI in 1974

Layers
1. Si substrate
2. Field oxide
3. Gate oxide
4. Poly Si
5. S/D
6. Interlayer
7. Aluminum
8. Passivation

Materials
1. Si
2. SiO₂
3. BPSG
4. Al
5. PSG

Atoms
1. Si
2. O
3. P
4. B
5. Al
(H, N, Cl)

Materials
1. Si
2. SiO₂
3. BPSG
4. Al
5. PSG

Atoms
1. Si
2. O
3. P
4. B
5. Al
(H, N, Cl)
New materials

Just examples!
Many other candidates

Y. Nishi, Si Nano Workshop, 2006,
(S. Sze, Based on invited talk at Stanford Univ., Aug. 1999)
Now: After 45 Years from the 1st single MOSFET

32 Gb and 16Gb NAND, SAMSUNG
Samsung’s NAND flash trend

<table>
<thead>
<tr>
<th>Capacity</th>
<th>Node</th>
<th>1st Fabrication</th>
</tr>
</thead>
<tbody>
<tr>
<td>512Mbit</td>
<td>120nm</td>
<td>2000</td>
</tr>
<tr>
<td>1Gbit</td>
<td>100nm</td>
<td>2001</td>
</tr>
<tr>
<td>2Gbit</td>
<td>90nm</td>
<td>2002</td>
</tr>
<tr>
<td>4Gbit</td>
<td>70nm</td>
<td>2003</td>
</tr>
<tr>
<td>8Gbit</td>
<td>60nm</td>
<td>2004</td>
</tr>
<tr>
<td>16Gbit</td>
<td>50nm</td>
<td>2005</td>
</tr>
<tr>
<td>32Gbit</td>
<td>40nm</td>
<td></td>
</tr>
<tr>
<td>256Gbit</td>
<td>20nm</td>
<td></td>
</tr>
</tbody>
</table>

Even Tbit would be possible in future!
Already 32 Gbit:
  larger than that of world population
  comparable for the numbers of neurons
  in human brain

Samsung announced 256 Gbit will be produced in 2010.
Only 4 years from now.
256Gbit: larger than those of # of stars in galaxies
Example: Immersion lithography, plasma doping, laser annealing etc.

PROCESS CONTROL: THE INVESTMENT THAT YIELDS

Ref: KLA Tencor
Lithography Challenge

Minimum feature size is scaling faster than lithography wavelength.
Light path in an EUV exposure tool

Reflective Mask

Collector Optics

EUV Light Source

Condenser / Illuminator

Reflective Projection / Reduction Optics

Resist

Wafer
Short-channel effect at downsizing

Source

Gate

Drain

Leakage Current

Space Charge Region
Scaling Method: by R. Dennard

$K=0.7$

for example

$X, Y, Z : K$, $V : K$, $Na : 1/K$

$D \propto \sqrt{V/Na}$ 

$I : K$

<table>
<thead>
<tr>
<th>Downscaling merit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Geometry &amp; Supply voltage</strong></td>
</tr>
</tbody>
</table>

| **Drive current in saturation** | \( I_d \) | \( K \) | \( I_d = \nu_{sat} W_g C_o (V_g - V_{th}) \), \( C_o \): gate C per unit area |
| **\( I_d \) per unit \( W_g \)** | \( I_d / \mu m \) | 1 | \( I_d \) per unit \( W_g = I_d / W_g = 1 \) |

| **Gate capacitance** | \( C_g \) | \( K \) | \( C_g = \varepsilon_o \varepsilon_{ox} L_g W_g / t_{ox} \) |
| **Switching speed** | \( \tau \) | \( K \) | \( \tau = C_g V_d / I_d \) |
| **Clock frequency** | \( f \) | 1/\( K \) | \( f = 1 / \tau = 1 / K \) |
| **Chip area** | \( A_{chip} \) | \( \alpha \) | Scaling \( \alpha \) |
| **Integration (# of Tr)** | \( N \) | \( \alpha / K^2 \) | \( N \rightarrow \alpha / K^2 \) |
| **Power per chip** | \( P \) | \( \alpha \) | \( f N C V^2 / 2 \rightarrow K^{-1} (\alpha K^{-2}) K (K^1)^2 = \alpha 1 \text{ when } \alpha = 1 \) |
What will be real Downscaling?

Is $K$ the same for all the parameters?

$L_g, W_g, t_{ox}, V_d \rightarrow K? \quad A_{chip} \rightarrow \alpha?$

$I_d \rightarrow K? \quad I_d/\mu m \rightarrow 1? \quad f \rightarrow 1/K?$

$C_g \rightarrow K? \quad \tau \rightarrow K? \quad N \rightarrow \alpha/K^2? \quad P \rightarrow \alpha?$
### Past downscaling trend

#### Change in 30 years

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Ideal scaling</th>
<th>Real Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_g$</td>
<td>$K$</td>
<td>$10^{-2}$</td>
</tr>
<tr>
<td>$t_{ox}$</td>
<td>$K(10^{-2})$</td>
<td>$10^{-2}$</td>
</tr>
<tr>
<td>$V_d$</td>
<td>$K(10^{-2})$</td>
<td>$10^{-1}$</td>
</tr>
<tr>
<td>$A_{chip}$</td>
<td>$\alpha$</td>
<td>$10^1$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Ideal scaling</th>
<th>Real Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_d/\mu m$</td>
<td>$K(10^{-2})$</td>
<td>$10^{-1}$</td>
</tr>
<tr>
<td>$N$</td>
<td>$\alpha/K^2(10^5)$</td>
<td>$10^4$</td>
</tr>
<tr>
<td>$P$</td>
<td>$\alpha(10^1)$</td>
<td>$10^5$</td>
</tr>
</tbody>
</table>

#### Past 30 years scaling

**Merit:** $N$, $f$ increase  
**Demerit:** $P$ increase

$V_d$ scaling insufficient  
Additional significant increase in $I_d$, $f$, $P$
Microprocessors Trend expected in 2001


<table>
<thead>
<tr>
<th>Year</th>
<th>Lg (nm)</th>
<th>Tox (nm)</th>
<th>f (GHz)</th>
<th>P (mW)</th>
<th>N (k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1972</td>
<td>10,000</td>
<td>1200</td>
<td>0.00075</td>
<td>a few 100</td>
<td>2.25k</td>
</tr>
<tr>
<td>2002</td>
<td>sub-70</td>
<td>1.4</td>
<td>2.53</td>
<td>several 10</td>
<td>50 M</td>
</tr>
<tr>
<td>2008</td>
<td>sub-25</td>
<td>0.7</td>
<td>30</td>
<td>10k</td>
<td>1.8B</td>
</tr>
</tbody>
</table>

Heat Generation

- **2002**: 10W/cm² Hot Plate
- **2006**: 100W/cm² Surface of Nuclear Reactor
- **2010**: 1000W/cm² Rocket Nozzle
- **2016**: 10000W/cm² Sun Surface

**Power Increase**

- Heat generation increase
- Clock Frequency increase
- Tr. Number increase

**Solution:**

- Low supply Voltage

Power = CV^2f

Source: Intel

Paul Packan, Intel Corporation, IEDM Short Course 2007
Problem is the huge production cost and investment.
Examples of foundry facility: UMC

**Fab 12A (Tainan, Taiwan)**
- US$3 billion investment
- Production since 2001
- 38K wafers/month by E/06
- 90, 65nm in production

**Fab 12i (Singapore)**
- US$3.6 billion investment
- Production since 2004
- 25K wafers/month by E/06
- 130, 90nm in production
- Ready for 65nm pilot
Volume production with larger wafer is a solution.

<table>
<thead>
<tr>
<th></th>
<th>12&quot;-Fab</th>
<th>8&quot;-Fab</th>
<th>12&quot;/8&quot; Usage ratio</th>
<th>12&quot;/8&quot; Wafer size ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td>1,100 KWH/Wafer</td>
<td>660 KWH/Wafer</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>6.1 M3/Wafer</td>
<td>4.7 M3/Wafer</td>
<td>1.3</td>
<td>2.25</td>
</tr>
<tr>
<td>Waste Water</td>
<td>3.8 M3/Wafer</td>
<td>2.9 M3/Wafer</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>Waste Air</td>
<td>20,000 CMH/Wafer</td>
<td>13,000 CMH/Wafer</td>
<td>1.5</td>
<td></td>
</tr>
</tbody>
</table>

TSMC
When do we start planning for next wafer size transition?

We are here

When does this happen?

200mm/1990  
(125/150mm - 1981)  
9 yrs + 2 yrs delay*

300mm/2001  
9 yrs? + 2 yrs delay?

450mm/2012  
?  
9 yrs + ? yrs delay

675mm/2021  
?
Crystal pulling furnace becomes too huge
Si crystal height cannot be very long because of its weigh

**Furnace**
- Height: 12 m
- Weight: 36 ton
- Hot zone: 40 inch
- Cusp-type super conductive magnet

**Crystal**
- Diameter: 400 mm
- Weight: over 400 kg
- Body length: over 1 m
How about the integration of such small-geometry MOSFETs in a chip?

1) Integration of huge number of the ultra-small MOSFETs would consume too huge power and thus, creates too huge heat?

2) Integration of such ultra-small MOSFETs causes too huge variations in the transistor characteristics, which could make the circuit design impossible?

3) There are too many number of transistors in a chip for the circuit designers to manipulate? (design crisis),

4) There would be no merit of transistor downsizing in performance and power, because of RC (resistance capacitance product) of interconnect cannot be reduced aggressively any more?

5) Who will pay the huge development and production costs for the integration of such ultra-small MOSFETs? Note that the prices for the recent process equipments and the lithography mask became extremely high.
These concerns have been argued in the past 15 years at every new generation of the products, like the wolf boy.

Fortunately, the wolf has not come, and the concerns have not come true.

It is expected that we can go with several more generations for the integration.

There will be still a room for squeezing the technologies to obtain the merit of the scaling-down for integration.
The continuous progress of CMOS technologies for
- high-performance
- low power
is very important because of the 3 reasons:

1) Rapid progress of aging population and falling birth rate

1) Global warming

1) Semiconductor industry and world economy
1) Rapid progress of aging population and falling birth rate:

Replacement of some of the human jobs by intelligent machines – such as human type robot for elderly-care, for example.

For, the daily family use, much higher intelligence and much lower power consumption than those of today are required.
Robot in 21c cannot be made without integrated circuits

Robot (21C)

Karakuri (Windup Mechanical) dolls (18C) in Japan
2) Recent Significant Global Warming

![Graph showing recent sea level rise and glacier thickness change](image)
We need to reduce CO2 generation!

Low power technology is an urgent request.

**Carbon Dioxide Variations**

The Industrial Revolution has caused a dramatic rise in CO2.

- Ice Age Cycles

![Graph showing historical CO2 concentrations](graph.png)
3) Semiconductor industry, and world economy

If there is no more downsizing such as 
45 → 32 nm Logic, 8 Gbit → 16 Gbit Memory

- LSIs will not be sold well, and semiconductor companies will face a disaster.

- Equipment and martial companies as well.

- There is no more R & D for semiconductors and many people will loose their jobs.

→ World economy crisis!
History and future of Transistor
Shrinking, Shrinking, and Shrinking!
and then, Shrinking, Shrinking, and Shrinking!

\[ C, V \propto L \]
\[ C: \text{Capacitance}; V: \text{Voltage} \]

Switching speed \( \frac{CV}{I} \rightarrow \text{Decrease} \)
Power consumption \( \frac{CV^2}{2} \rightarrow \text{Decrease} \)
Integration density: \( \frac{1}{L^2} \rightarrow \text{Increase} \)

<table>
<thead>
<tr>
<th>Year</th>
<th>Gate length</th>
<th>Gate Oxid Thickness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td>10,000 nm</td>
<td>100 nm</td>
</tr>
<tr>
<td>2007</td>
<td>25 nm</td>
<td>1 nm</td>
</tr>
</tbody>
</table>
CMOS downsizing is critically important

However now, many people expect that we will reach limit in 2020.

Totally, new paradigm after reaching the downsizing limit.

What will be?
After 2020

There is no decrease in gate length around at 10 ~ 5 nm.

4 reasons.
After 2020

4 reasons for no downsizing anymore
or No decrease in gate length

1. No increase of On-current (Drain current) because of already semi-ballistic conduction.
   Ballistic $\leftrightarrow$ No scattering of carriers in channel
   Thus, all the carrier from the source reach drain

2. Increase of Off-current (Subthreshold current)

3. No decrease of Gate capacitance by parasitic components

4. Increase in production cost.
After 2020

What will be the world with no gate length reduction?
More Moore and More than Moore

Moore’s Law & More

More than Moore: Diversification

- Analog/RF
- Passives
- HV Power
- Sensors
- Actuators
- Biochips

Combining SoC and SiP: Higher Value Systems

Information Processing

Combining SoC and SiP: Higher Value Systems

Non-digital content
System-in-package (SiP)

Interacting with people and environment

Baseline CMOS: CPU, Memory, Logic

Beyond CMOS

More Moore: Miniaturization


Question what is the other side of the cloud?
Metallic

Semiconducting
$E_G \approx 0.8eV/d(\text{nm})$

Diameter (d): $\sim 0.5$-$5 \text{ nm}$, Length: $\sim 10 \text{ nm}$-$10 \text{ cm}$

**a** CNT SB-MOSFET

**b** CNT MOSFET

*Charles M. Lieber, IEDM Short Course, 2008*
The current density is normalized by $2d$, where $d$ is the nanotube diameter.

**CNTFETs outperform Si MOSFETs**

*Charles M. Lieber, IEDM Short Course, 2008*
IBM’s Carbon Nanotube IC

March 2006

5-stage Ring Oscillator built using single walled carbon Nanotubes (SWNT)

S.E Thompson, IEDM Short Course, 2008
The vision:
Graphene FET with top gate and local interconnects

Top Gate

Source

Au

Graphene

SiO₂

Back Gate

Silicon

Drain
Graphene examples

Source: A. Geim
Graphene Field Effect Devices

- conductivity reduced by top gate oxide
- Model of Das Sarma:
  - Minimum conductivity and shift of Dirac Point caused by interface charges
  - without oxide: $\sim 2 \times 10^{12} \text{ cm}^{-2}$
  - with oxide: $\sim 6 \times 10^{12} \text{ cm}^{-2}$
- Charges are suspected to cause scattering
Mobility: a first approximation

Approximation of $\mu$ (Drude model):

$$\mu = \frac{\sigma}{n^* q}$$

with

$$\sigma = \frac{J}{E_{ds}} = \frac{I}{\text{width} \times \text{length}} / V_{ds}$$

$$n = \frac{\varepsilon \times E_{eff}}{q}$$

- Mobility decreases (> 1/10) for graphene "sandwiched" in SiO$_2$
  → dominant scattering mechanism seems to be substrate determined
- Yet higher values than silicon universal mobility and especially higher than Ultra Thin Body SOI MOSFETs

Lemme et al., tbp: SSE

lemme@amo.de  www.amo.de  ECS Meeting, Washington DC, Oct. 12th, 2007
3D Chip Stacking LSI

Encapsulate Resin

Si Die

Si Interposer

20 μm pitch

50 μm

10 μm

bonding with Cu/Sn alloy

Cu₃Sn

Upper TV

Cu Bump

Resin

SiN

SiO₂

Lower TV

ASET
<table>
<thead>
<tr>
<th>Device</th>
<th>FET</th>
<th>RSFQ 1D structures</th>
<th>Resonant Tunneling Devices</th>
<th>SET</th>
<th>Molecular</th>
<th>QCA</th>
<th>Spin transistor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Size</td>
<td>100 nm</td>
<td>0.3 μm</td>
<td>100 nm</td>
<td>100 nm</td>
<td>40 nm</td>
<td>Not known</td>
<td>60 nm</td>
</tr>
<tr>
<td>Density (cm$^{-2}$)</td>
<td>3E9</td>
<td>1E6</td>
<td>3E9</td>
<td>3E9</td>
<td>6E10</td>
<td>1E12</td>
<td>3E10</td>
</tr>
<tr>
<td>Switch Speed</td>
<td>700 GHz</td>
<td>1.2 THz</td>
<td>Not known</td>
<td>1 THz</td>
<td>1 GHz</td>
<td>Not known</td>
<td>30 MHz</td>
</tr>
<tr>
<td>Circuit Speed</td>
<td>30 GHz</td>
<td>250–800 GHz</td>
<td>30 GHz</td>
<td>30 GHz</td>
<td>1 GHz</td>
<td>&lt;1 MHz</td>
<td>1 MHz</td>
</tr>
<tr>
<td>Switching Energy, J</td>
<td>2×10$^{-18}$</td>
<td>&gt;1.4×10$^{-17}$</td>
<td>2×10$^{-18}$</td>
<td>&gt;2×10$^{-18}$</td>
<td>&gt;1.5×10$^{-17}$</td>
<td>1.3×10$^{-16}$</td>
<td>&gt;1×10$^{-18}$</td>
</tr>
<tr>
<td>Binary Throughput, GBit/ns/cm$^2$</td>
<td>86</td>
<td>0.4</td>
<td>86</td>
<td>86</td>
<td>10</td>
<td>N/A</td>
<td>0.06</td>
</tr>
</tbody>
</table>

We HAVE IDENTIFIED NO VIABLE EMERGING LOGIC TECHNOLOGIES for Information Processing beyond CMOS
We could keep the Moore’s law after 2020
Without downswing the gate length

What is Moore’s law.
Keep increase of the number of components. Cost per components decreases!

We could keep the Moore’s law after 2020
Without downswing the gate length

What is Moore’s law.
→ to increase the number (#) of Tr. In a chip

Now, # of Tr. in a chip is limited by power.
→ key issue is to reduce the power.
→ to reduce the supply voltage is still effective

To develop devices with sufficiently high drain current under low supply voltage is important.
FinFET to Nanowire

Channel conductance is well controlled by Gate even at L=5nm

F.-L. Yang, VLSI2004
Selection of MOSFET structure for high conduction: Nano-wire or Nano-tube FETs is promising

3 methods to realize High-conduction at Low voltage
M1. Use 1D ballistic conduction
M2. Increase number of quantum channel
M3. Increase the number of wire or tube per area
   3D integration of wire and tubes

For suppression of Ioff, the Nanowire/tube is also good.
1D conduction per one quantum channel:

\[ G = \frac{2e^2}{h} = 77.8 \ \mu\text{S/wire or tube} \]
regardless of gate length and channel material

That is 77.8 mA/wire at 1V supply

This an extremely high value

However, already 20mA/wire was obtained experimentaly by Samsung
Increase the Number of quantum channels

By Prof. Shiraishi of Tsukuba univ.

Energy band of Bulk Si

300 K
- $E_g = 1.12$ eV
- $E_l = 2.0$ eV
- $E_x = 1.2$ eV
- $E_v = 0.044$ eV
- $E_T = 3.4$ eV
- $E_{TR} = 4.2$ eV

4 channels can be used

Energy band of 3 x 3 Si wire
Maximum number of wires per 1 µm

Front gate type MOS
- 165 wires /µm

Surrounded gate type MOS
- 33 wires/µm

- Metal gate electrode (10nm)
- High-k gate insulator (4nm)
- Si Nano wire (Diameter 2nm)

6nm pitch: By nano-imprint method
6nm pitch: EUV lithography

30nm pitch: EUV lithography
Increase the number of wires towards vertical dimension
Our new roadmap

Extended CMOS: More Moore + CMOS logic

Beyond the horizon

Natural direction of downsizing

Si Channel

Diameter = 2nm

Nanowire

Diameter = 10nm

Graphene

Si Fin, Tri-gate

Si Nano wire

III-V & Ge Nano wire

Tube

CNT

Tube, Ribbon

Our new roadmap

High conduction

By 1D conduction

ITRS

ITRS Beyond CMOS

More Moore

Extended CMOS

More Moore ??
Miniaturization of Interconnects on PCB (Printed Circuit Board)
Brain

Ultra small volume
Small number of neuron cells
Extremely low power
Real time image processing
(Artificial) Intelligence
3D flight control

Sensor

Infrared
Humidity
CO₂

Mosquito

System and Algorism becomes more important!

But do not know how?

Dragonfly is further high performance
Thank you for your attention!