Introduction
   Kristin De Meyer, IMEC

1. Technology Scaling and Roadmap
   Hiroshi Iwai, Tokyo Institute of Technology

2. 22nm Device Architecture and Performance Elements
   Kelin Kuhn, Intel Corporation

3. Lithography for the 22nm Technology Node
   Kurt Ronse, Geert Vandenberghe, IMEC

4. BEOL Technology for the 22nm Technology Node
   Jeffrey Gambino, IBM

5. Device/Circuit Interactions at the 22nm Technology Node
   Kaushik Roy, Purdue University
Outline

1. Scaling
2. ITRS Roadmap
3. Voltage Scaling/ Low Power and Leakage
4. SRAM Cell Scaling
5. Roadmap for further future as a personal view
1. Scaling
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.

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

$K=0.7$ for example

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

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!

<table>
<thead>
<tr>
<th>Geometry &amp; Supply voltage</th>
<th>$L_g$, $W_g$</th>
<th>$K$</th>
<th>Scaling $K : K=0.7$ for example</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{ox}$, $V_{dd}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Drive current in saturation</th>
<th>$I_d$</th>
<th>$K$</th>
<th>$I_d = \nu_{sat} W_g C_o (V_g - V_{th})$; $C_o$: gate $C$ per unit area</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_d$ per unit $W_g$</td>
<td>$I_d/\mu m$</td>
<td>1</td>
<td>$I_d$ per unit $W_g = I_d / W_g = 1$</td>
</tr>
</tbody>
</table>

| Gate capacitance | $C_g$ | $K$ | $C_g = \varepsilon_o \varepsilon_{ox} L_g W_g / t_{ox}$; $KK/K = K$ |

| Switching speed | $\tau$ | $K$ | $\tau = C_g V_{dd} / I_d$; $KK/K = K$ |

| Clock frequency | $f$ | $1/K$ | $f = 1/\tau = 1/K$ |

| Chip area | $A_{chip}$ | $\alpha$ | $\alpha$: Scaling factor; $In the past, \alpha>1$ for most cases |

| Integration (# of Tr) | $N$ | $\alpha/K^2$ | $N \rightarrow \alpha/K^2 = 1/K^2$, when $\alpha=1$ |

<p>| Power per chip | $P$ | $\alpha$ | $fNCV^2/2 \rightarrow K^{-1}(\alpha K^{-2})K (K^1)^2 = \alpha = 1$, when $\alpha=1$ |</p>
<table>
<thead>
<tr>
<th>k = 0.7 and $\alpha = 1$</th>
<th>k = 0.7$^2$ = 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>$P$ (Power)/Clock $\rightarrow 0.7^3 = 0.34$</td>
<td>$P$ (Power)/Clock $\rightarrow 0.5^3 = 0.125$</td>
</tr>
<tr>
<td>$\tau$ (Switching time) $\rightarrow$ 0.7</td>
<td>$\tau$ (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>
- The concerns for limits of down-scaling have been announced for every generation.

- However, down-scaling of CMOS is still the ‘royal road’* for high performance and low power.

- Effort for the down-scaling has to be continued by all means.

*Euclid of Alexandria (325BC?-265BC?)
   ‘There is no royal road to Geometry’

Mencius (Meng-zi), China (372BC?-289BC?)
   孟子: 王道, 趙道 (Rule of right or virtue vs. Rule of military)
Actual past downscaling trend until year 2000

Change in 30 years

<table>
<thead>
<tr>
<th>Change in 30 years</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>$V_{dd}$</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></th>
<th>Ideal scaling</th>
<th>Real Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_d/\mu m$</td>
<td>$10^{-1}$</td>
<td>$10^1$</td>
</tr>
<tr>
<td>$V_{dd}$</td>
<td>$K(10^{-2})$</td>
<td>$10^{-1}$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>$10^1$</td>
<td></td>
</tr>
</tbody>
</table>

Past 30 years scaling

Merit: N, f increase
Demerit: P increase

V$_{dd}$ scaling insufficient

Additional significant increase in $I_d$, f, P

V$_{dd}$ scaling insufficient, $\alpha$ increased $\rightarrow$ N, $I_d$, f, P increased significantly

- Now, power and/or heat generation are the limiting factors of the down-scaling

- Supply voltage reduction is becoming difficult, because Vth cannot be decreased any more, as described later.

- Growth rate in clock frequency and chip area becomes smaller.
2. ITRS Roadmap
(for 22 nm CMOS logic)
ITRS Roadmap does change every year!

- 2007 Edition
- 2006 Update
- 2005 Edition
- 2004 Update
- 2003 Edition
- 2002 Update
- 2001 Edition
- 2000 Update

http://www.itrs.net/reports.html

The current latest version: ITRS 2007 Edition

ITRS 2008 Update will be published on the web at the end of Dec 2008 or Jan. 2009
However, discussions for ITRS 2008 Update were open at the ‘2008 ITRS Summer Public Conference on 16 July 2008’ held in SF, US

http://www.itrs.net/Links/2008Summer/Presentations.html

The material in this SC is based on 2008 ITRS Summer Public Conference and ITRS 2007 Edition

2008 ITRS Winter Public Conference, was held on Dec.9. in Seoul.  http://www.itrs.net/

The Winter Public Conference data could not be feed-back to this SC material because of print schedule for SC.
What does ‘22 nm’ mean in 22 nm CMOS Logic?

‘XX nm CMOS Technology
Commercial Logic CMOS products

<table>
<thead>
<tr>
<th>Technology name</th>
<th>Starting Year</th>
<th>Physical Gate Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>45 nm</td>
<td>2007</td>
<td>32 nm</td>
</tr>
<tr>
<td>32 nm</td>
<td>2009?</td>
<td>27 nm</td>
</tr>
<tr>
<td>22 nm</td>
<td>2011?~ 2012?</td>
<td>22 nm</td>
</tr>
<tr>
<td>16 nm</td>
<td>2013?~ 2014?</td>
<td>18 nm</td>
</tr>
</tbody>
</table>

ITRS (Likely in 2008 Update)
for High Performance Logic

<table>
<thead>
<tr>
<th>Year</th>
<th>Half Pitch (1st Metal)</th>
<th>Physical Gate Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>68 nm</td>
<td>32 nm</td>
</tr>
<tr>
<td>2008</td>
<td>59 nm</td>
<td>29 nm</td>
</tr>
<tr>
<td>2009</td>
<td>52 nm</td>
<td>27 nm</td>
</tr>
<tr>
<td>2010</td>
<td>45 nm</td>
<td>24 nm</td>
</tr>
<tr>
<td>2011</td>
<td>40 nm</td>
<td>22 nm</td>
</tr>
<tr>
<td>2012</td>
<td>36 nm</td>
<td>20 nm</td>
</tr>
<tr>
<td>2013</td>
<td>32 nm</td>
<td>18 nm</td>
</tr>
<tr>
<td>2014</td>
<td>29 nm</td>
<td>16 nm</td>
</tr>
</tbody>
</table>

Source: 2008 ITRS Summer Public Conf.

‘XX nm’ CMOS Logic Technology:
- In general, there is no common corresponding parameter with ‘XX nm’ in ITRS table, which stands for ‘XX nm’ CMOS.
What does ‘22 nm’ mean in 22 nm CMOS Logic?

- ‘XX nm’ does not correspond to the ‘Half Pitch’ nor ‘Physical Gate Length’ of ITRS.

- ‘XX nm’ is now just a commercial name for CMOS Logic generation of size and its technology.

- Actual parameter values and starting years for commercial products are somewhat different from the above ITRS table, depending on semiconductor companies.

- In 22 and 16 nm technologies, physical gate lengths of high-performance logic device may be close to XX nm.
What does ‘22 nm’ mean in 22 nm CMOS Logic?

8µm → 6µm → 4µm → 3µm → 2µm → 1.2µm → 0.8µm → 0.5µm
- Originally, ‘XX’ means lithography resolution.
- Thus, ‘XX’ was the gate length, and half pitch of lines
- ‘XX’ had shrunk 0.7 in 3 years in average (0.5 in 6 years) those days.
- ‘XX’ value deviated among companies: example: 1.5µm, 1.2µm, 1µm

→ 350nm → 250nm → 180nm → 130nm → 90nm → 65nm → 45nm
- ‘XX’ values were established by NTRS* and ITRS with the term of ‘Technology Node**’ and ‘Cycle***’ using typical ‘half pitch value’.
  *NTRS: National Tech. Roadmap, **Term ‘Technology Node’ is not used now.
  ***Cycle: Period or year for which the half pitch becomes X0.71.
- The gate length of logic CMOS became smaller with one or two generations from the half pitch, and ‘XX’ names ahead of generations have been used for logic CMOS.
- Memory still keeps the half pitch as the value of ‘XX’

→ 32nm → 22nm → 16nm → 11nm → 8nm?? → 5.5nm ??
What does ‘22 nm’ mean in 22 nm CMOS Logic?

Gate length of Logic CMOS became significantly smaller than lithography resolution or half-pitch using special technique such as resist aching (or trimming) method since 350 nm CMOS.

Source: ITRS 2001 Update
Some Problem: Number of most advanced logic CMOS companies is decreasing in generations.
Definition of the Half Pitch

Logic 1st Metal Half Pitch

- DRAM ½ Pitch = DRAM Metal Pitch/2
- MPU/ASIC M1 ½ Pitch = MPU/ASIC M1 Pitch/2

Flash Poly Gate Half Pitch

- FLASH Poly Silicon ½ Pitch = Flash Poly Pitch/2

Typical DRAM/MPU/ASIC Metal Bit Line

8-16 Lines
Typical flash Un-contacted Poly

Source: 2008 ITRS Summer Public Conf.
For example, Typical Half Pitches at ITRS 2007

2007 ITRS Product Technology Trends - Half-Pitch, Gate-Length

Source: 2008 ITRS Summer Public Conf. 2007 - 2022 ITRS Range
Physical gate length in past ITRS was too aggressive. The dissociation from commercial product prediction will be adjusted.

Physical gate length of High-Performance logic will shift by 3-5 yrs.

Source: 2008 ITRS Summer Public Conf.
L of HP and LOP will shift, but LSTP will not.

<table>
<thead>
<tr>
<th>Year</th>
<th>Physical Gate L (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td></td>
</tr>
<tr>
<td>2020</td>
<td></td>
</tr>
<tr>
<td>2021</td>
<td></td>
</tr>
</tbody>
</table>

**HP (High Performance) Logic**
- shifts by 3 – 5 years

**LOP (Low Operation Power) Logic**
- shifts by 1 – 3 years

**LSTP (Low Standby Power) Logic**
- Very minor shift

Source: 2008 ITRS Summer Public Conf.
EOT and Xj shift backward, corresponding to Lg shift

EOT: 0.55 nm → 0.88 nm, Xj: 8 nm → 11 nm @ 22nm CMOS

Likely in 2008 Update

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MPU/ASIC Lg (nm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>25</td>
<td>23</td>
<td>20</td>
<td>18</td>
<td>16</td>
<td>14</td>
<td>13</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6.3</td>
<td>5.6</td>
<td>5</td>
<td>4.5</td>
</tr>
<tr>
<td>2008</td>
<td>32</td>
<td>29</td>
<td>27</td>
<td>24</td>
<td>22</td>
<td>20</td>
<td>18</td>
<td>17</td>
<td>15</td>
<td>14.0</td>
<td>12.8</td>
<td>11.7</td>
<td>10.7</td>
<td>9.7</td>
<td>8.9</td>
<td>8.1</td>
</tr>
<tr>
<td>Shift/Interpolate Formua</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>interp</td>
<td>interp</td>
<td>interp</td>
<td>interp</td>
<td>2009</td>
<td>interp</td>
<td>interp</td>
<td>interp</td>
<td>interp</td>
<td>interp</td>
<td>interp</td>
<td>interp</td>
<td>interp</td>
<td>interp</td>
<td>interp</td>
<td>interp</td>
</tr>
<tr>
<td>EOT w/3E20 poly, bulk MPU (nm)</td>
<td>1.2</td>
<td>0.71</td>
<td>0.54</td>
<td>0.41</td>
<td>0.66</td>
<td>0.64</td>
<td>0.41</td>
<td>0.55</td>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EOT w/3E20 poly, bulk MPU (nm)</td>
<td>1.3</td>
<td>1.2</td>
<td>1.2</td>
<td>1</td>
<td>0.66</td>
<td>0.64</td>
<td>0.41</td>
<td>0.55</td>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EOT w/metal gate, bulk MPU (nm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EOT w/metal gate, bulk MPU (nm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drain Ext. $X_j$ bulk MPU (nm)</td>
<td>12.5</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drain Ext. $X_j$ bulk MPU (nm)</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>9</td>
<td>8.5</td>
<td>7.7</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: 2008/ ITRS Summer Public Conf.

non-steady trend corrected

filled in for metal gate EOT for 2009/10 based on latest conference presentations
Clock frequency does not increase aggressively anymore.

Even decreased!

Advantage in SISC
Era for ‘out of order’

Advantage in RISC
Simple configuration

Source: Mitsuo Saito, Toshiba
Cell Broadband Engine

6GHz capability for SRAM

Source: IBM, Toshiba, Sony
ISSCC2007 and 08

Source: 2007 ITRS Winter Public Conf.
Clock frequency Change in the past ITRS
(Max on chip frequency or ‘Core clock’)

Design Max On-Chip Clock Frequency

- Including 2005 ITRS and Final (Aug'07) 2

2005/06 ITRS “WAS”

• "Gap" Delayed by 3 years in 2005 ITRS

1.29x/year (2x/2.5 yrs)

1.41x/year (2x/2 yrs)

15%/Year

22 nm: 6 GHz?

ITRS2001

ITRS2003

ITRS2005

ITRS2007

15%/Year

8%/Year

New Design TWG 2007 ITRS Final “IS” Ave 8% CAGR

Past < Future

2007 Des TWG Actual History of Average On-Chip ~ 21% CAGR

Source: 2008 ITRS Summer Public Conf.
Structure and technology innovation (ITRS 2007)

Source: 2008 ITRS Summer Public Conf.
Technology innovation described in ITRS 2007

Alternative material (Ge, III-V) and structure (Nanowire) in channel region.

Source: 2007 ITRS Winter Public Conf.
Timing of CMOS innovations shifts backward.

**Bulk CMOS has longer life now!**

Correspond to 22nm Logic CMOS

Source: 2008 ITRS Summer Public Conf.
## Wafer size (ITRS 2007)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MPU/ASIC Metal 1 (M1) ½ Pitch (nm)</td>
<td>68</td>
<td>59</td>
<td>52</td>
<td>45</td>
<td>40</td>
<td>36</td>
<td>32</td>
<td>28</td>
<td>25</td>
</tr>
<tr>
<td>MPU Physical Gate Length (nm)</td>
<td>25</td>
<td>23</td>
<td>20</td>
<td>18</td>
<td>16</td>
<td>14</td>
<td>13</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>MPU High-Performance Total Chip Area (mm²)</td>
<td>310</td>
<td>246</td>
<td>195</td>
<td>310</td>
<td>246</td>
<td>195</td>
<td>310</td>
<td>246</td>
<td>195</td>
</tr>
<tr>
<td>MPU High-Performance Active Transistor Area (mm²)</td>
<td>31.7</td>
<td>25.1</td>
<td>20.0</td>
<td>31.7</td>
<td>25.1</td>
<td>20.0</td>
<td>31.7</td>
<td>25.1</td>
<td>20.0</td>
</tr>
</tbody>
</table>

**General Characteristics** *(99% Chip Yield)*

| Maximum Substrate Diameter (mm)—High-volume Production (>20K wafer starts per month)* | 300   | 300   | 300   | 300   | 300   | 450   | 450   | 450   | 450   |

*Source: ITRS 2007*

??

Maybe delay??
Gate CD (Critical Dimension) Control

ITRS 2007

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MPU/ASIC Metal 1 (M1) ( \frac{1}{2} ) Pitch (nm)(contacted)</td>
<td>68</td>
<td>59</td>
<td>52</td>
<td>45</td>
<td>40</td>
<td>36</td>
<td>32</td>
<td>28</td>
<td>25</td>
</tr>
<tr>
<td>MPU Physical Gate Length (nm)</td>
<td>25</td>
<td>23</td>
<td>20</td>
<td>18</td>
<td>16</td>
<td>14</td>
<td>13</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Total maximum allowable etch 3σ (nm), including photoresist trim and gate etch</td>
<td>1.5</td>
<td>1.38</td>
<td>1.2</td>
<td>1.08</td>
<td>0.96</td>
<td>0.84</td>
<td>0.78</td>
<td>0.66</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Source: ITRS 2007

2008 Update

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MPU Physical Gate Length (nm)</td>
<td>32</td>
<td>29</td>
<td>27</td>
<td>24</td>
<td>22</td>
<td>20</td>
<td>18</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>( L_{gate} ) 3σ variation (nm) ( [\mu] )</td>
<td>3.82</td>
<td>3.49</td>
<td>3.18</td>
<td>2.9</td>
<td>2.65</td>
<td>2.42</td>
<td>2.21</td>
<td>2.02</td>
<td>1.84</td>
</tr>
</tbody>
</table>

Source: 2008 ITRS Summer Public Conf.

Gate CD control color changed to ‘white’ through 2011 and to ‘yellow’ for 2012 reflecting the new Lg scaling
# ITRS2008 Low-k Roadmap Update

## Correspond to 22nm Logic

<table>
<thead>
<tr>
<th>Year of Production</th>
<th>Near-term</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2008</td>
</tr>
<tr>
<td>Interlevel metal insulator – effective dielectric constant ($k$)</td>
<td>2.7-3.0</td>
</tr>
<tr>
<td>Interlevel metal insulator – bulk dielectric constant ($k$)</td>
<td>2.3-2.7</td>
</tr>
</tbody>
</table>

**Source:** 2008 ITRS Summer Public Conf.

$k$ value increases by $0.1 \sim 0.3$
Roadmap towards 22nm technology and beyond

- Physical gate length downsizing rate will be less aggressive.

- Corresponding to the above, performance increase would slow down – Clock frequency, etc.

- Introduction of innovative structures – UTB SOI and DG delayed, and bulk CMOS has longer life than predicted by previous ITRS roadmaps.
3. Voltage Scaling
/ Low Power and Leakage
Difficulty in Down-scaling of Supply Voltage: Vdd

Because, $V_{th}$ cannot be down-scaled anymore, $V_{dd}$ down-scaling is difficult.

$V_{dd} - V_{th}$ determines the performance (High $I_d$) and cannot be too small.

$\Delta V_{th}$: $V_{th}$ variation

$> \Delta V_{th}$

Margin for $V_{th}$ variation is necessary
Subthreshold leakage current of MOSFET

Subthreshold leakage current

\[ I_{th} \]

\[ V_g = 0 \text{V} \]

Subthreshold region

\[ (Threshold \ Voltage) \]

Subthreshold Current

Is OK at Single Tr. level

But not OK
For Billions of Trs.
Vth cannot be decreased anymore

significant Ioff increase

Vth: 300mV → 100mV
Ioff increases with 3.3 decades
(300 – 100)mV/(60mV/dec) = 3.3 dec

Subthreshold slope (SS)
= (Ln10)(kT/q)(Cox+C_D+C_it)/Cox
> ~ 60 mV/decade at RT

SS value:
Constant and does not become small with down-scaling
**ITRS for HP logic**

**S-D leakage current**

Id-s leak has to be stay less than $1\mu\text{A}/\mu\text{m}$

**Saturation current**

Id-sat growth will be modest in 2008 update

Source: ITRS and 2008 ITRS Summer Public Conf.

2008 Values are from ITRS Public Conf. and still under discussion
ITRS for HP logic

Ion/Ioff ratio

Year


2008 up (bulk)
2008 up (UTB)
2008 up (DG)
2007 (bulk)
2007 (UTB)
2007 (DG)
2005 (bulk)
2005 (UTB)
2005 (DG)
2003
2001
1999

2001
1999

Others

2003-2008

Source: ITRS and 2008 ITRS Summer Public Conf.

2008 Values are from ITRS Public Conf. and still under discussion
ITRS for HP logic

Vdd will stay higher in 2008 update

Vth-sat will be around 0.1V

Source: ITRS and 2008 ITRS Summer Public Conf.

2008 Values are from ITRS Public Conf. and still under discussion
ITRS for HP logic

Vth-sat / Vdd

2008 Values are from ITRS Public Conf. and still under discussion

Source: ITRS and 2008 ITRS Summer Public Conf.
HP, LOP, LSTP for Logic CMOS

Subthreshold Leakage ($A/\mu m$)

Operation Frequency (a.u.)

Source: 2007 ITRS Winter Public Conf.
Comparison of Isd-leak and Id-sat for HP, LOP, LSTP

### Isd_leak for HP

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Isd_leak</td>
<td>1.0E-3</td>
<td>1.0E-2</td>
<td>1.0E-1</td>
<td>1.0E+0</td>
<td>1.0E+1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Isd_leak for LOP

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Isd_leak</td>
<td>500</td>
<td>1000</td>
<td>1500</td>
<td>2000</td>
<td>2500</td>
<td>3000</td>
<td></td>
</tr>
</tbody>
</table>

### Isd_leak for LSTP

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Isd_leak</td>
<td>200</td>
<td>400</td>
<td>600</td>
<td>800</td>
<td>1000</td>
<td>1200</td>
<td>1400</td>
</tr>
</tbody>
</table>

### Id_sat for HP

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Id_sat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Id_sat for LOP

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Id_sat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Id_sat for LSTP

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Id_sat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: ITRS and 2008 ITRS Summer Public Conf.

2008 Values are from ITRS Public Conf. and still under discussion
Comparison of Ion/loff for HP, LOP, LSTP

Source: ITRS and 2008 ITRS Summer Public Conf.

2008 Values are from ITRS Public Conf. and still under discussion
Comparison of Vdd and Vth-sat for HP, LOP, LSTP

Vdd for HP

Vdd for LOP

Vdd for LSTP

Vth for HP

Vth for LOP

Vth for LSTP

2008 Values are from ITRS Public Conf. and still under discussion

Source: ITRS and 2008 ITRS Summer Public Conf.
Comparison of Vth-sat/Vdd for HP, LOP, LSTP

Source: ITRS and 2008 ITRS Summer Public Conf.
2008 Values are from ITRS Public Conf. and still under discussion
SS (Subthreshold Slope) becomes worse in the following cases

1. Improper down-scaling
   Ex. When $T_{ox}$, $W_{dep}$, or $V_{dd}$ is not scaled

2. High impurity doping in channel or substrate
   High impurity Conc.
   $\rightarrow C_D$ increase
   $\rightarrow$ SS increase
   $SS = (\ln 10)(kT/q)(C_{ox}+C_D+C_{it})/C_{ox}$

3. Enhanced Drain-Electric-field penetration through oxide
   Ex. High-k, SOI, Multi-gate (Double gate: DG)
   DG and SOI often show better SS, but be careful!
Improper down-scaling

Could we squeeze technologies for ultimate CMOS scaling?

Saturation of EOT thinning is a serious roadblock to proper down-scaling.

Is 0.5nm real limit?

Delay Saturation
EOT<0.5nm with Gain in Drive Current is Possible

La$_2$O$_3$ gate insulator

(a) EOT=0.37nm** $V_{th}=-0.04$V
(b) EOT=0.43nm** $V_{th}=-0.03$V
(c) EOT=0.48nm $V_{th}=-0.02$V

W/L=2.5/50μm
PMA 300°C (30min)

Drain current (mA)
0 1 2 3

Drain voltage (V)
0 0.2 0.4 0.6 0.8 1

Drain voltage (V)
0 0.2 0.4 0.6 0.8 1

Drain voltage (V)
0 0.2 0.4 0.6 0.8 1

EOT scaling below 0.5nm

Still useful for larger drain current


* Because Lg is very large (2.5μm), gate leakage is large in case (a). The gate leakage component was subtracted from measured data for case (a). However, if we make small gate length, the gate leakage current should become sufficiently small to be ignored compared with Id as we verified with SiO$_2$ gate before (Momose et al.,IEDM 1994). The gate leakage could be suppressed by modifying material and process in future.

** Estimated by Id value
Thus, in future, maybe continuous development of new techniques could make more proper down-scaling possible.

It is difficult to say, but EOT and Vdd may become smaller than expected today.
Enhanced D-Electric-field

Same parameter condition for both
(2006 ITRS Bulk parameters are used for both Bulk and DG)

\[ \text{Lg} = 16 \text{nm}, \text{tox(EOT)} = 0.5 \text{nm}, \]
\[ \text{Dopant@Channel} = 8.1 \times 10^{18} \text{cm}^{-2} \]

Source: ECS Fall Meeting, Oct 2008, Honolulu,
Y. Kobayashi, A. B. Sachid, K. Tsutsui, K. Kakushima,

**DIBL**: Drain Induced Barrier Lowering

\[ \frac{\partial V(x,y)}{\partial V_d} \quad V_d = 1 \text{V} \]

\[ \Lambda : \text{Penetration Depth of DIBL} \]

\[ \Lambda = 7.6 \text{nm} \quad \Lambda = 17.1 \text{nm} \quad \Lambda = 13.2 \text{nm} \]

Comparison of Bulk and DG

- **DIBL at drain edge**
- **SS**
- **Bulk**

Fin Width (nm)

\[ W_{fin} = 10.7 \text{ nm} \quad W_{fin} = 30 \text{ nm} \quad W_{fin} = 40 \text{ nm} \]
Enhanced D-Electric-field

Comparison of High-k and SiO₂ MOSFETs

Penetration of lateral field from Drain through high-k causes significant short channel effects

$V_{dd}$ will stay higher than predicted by previous ITRS roadmaps.

Solution towards Low $V_{dd}$

Effort to reduce $I_{sd\text{-}leak}$ and increase $I_{d\text{-}sat}$ is important

- **Scaling**: Proper down-scaling
  - Introduction of Next generation high-k, S/D etc.
  - CD* variation control by lithography and etching techniques
    * **CD**: **Critical dimension**

- **Structure**: Bulk $\rightarrow$ UTB-SOI $\rightarrow$ DG $\rightarrow$ Nanowire

- **Variation**: Proper scaling by new tech. – High-k, litho. Etc.
  - $V_{th}$ adjustment by $V_{sub}$ control

- **Circuit techniques**: Dynamic and local Multi-$V_{dd}$, etc.
Random Variability Reduction Scenario in ITRS 2007

Assumption: Random dopant fluctuation is Main source of Random Variability. Line width roughness of Lg and Wg is not considered in this.

Source: 2007 ITRS Winter Public Conf.
Another concern for Low $V_{dd}$ besides $I_{sd\text{-}leak}$ increase

$\rightarrow$ Huge power loss for voltage conversion to such low $V_{dd}$
4. SRAM cell scaling
Intel’s **SRAM** test chip trend

Source: B. Krzanich, S. Natrajan, Intel Developer’s Forum 2007
http://download.intel.com/pressroom/kits/events/idffall_2007/Briefing
Silicon&TechManufacturing.pdf

<table>
<thead>
<tr>
<th>Process name</th>
<th>Lithography</th>
<th>1st production</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1264</td>
<td>65nm</td>
<td>2005</td>
</tr>
<tr>
<td>P1266</td>
<td>45nm</td>
<td>2007</td>
</tr>
<tr>
<td>P1268</td>
<td>32nm</td>
<td>2009</td>
</tr>
<tr>
<td>P1270</td>
<td>22nm</td>
<td>2011</td>
</tr>
</tbody>
</table>

Only schedule has been published

---

**SRAM down-scaling trend has been kept until 32nm and probably so to 22nm**

![Graph showing cell area vs. year for different process technologies](image)

**Technology**
- **90 nm Process**
  - Cell size: 1.0 \(\mu\text{m}^2\)cell
  - Capacity: 50 Mbit
  - Chip area: 109 mm\(^2\)
  - Functional Si: February ‘02

**65 nm Process**
- Cell size: 0.57 \(\mu\text{m}^2\)cell
- Capacity: 70 Mbit
- Chip area: 110 mm\(^2\)
- Functional Si: April ‘04

**45 nm Process**
- Cell size: 0.346 \(\mu\text{m}^2\)cell
- Capacity: 153 Mbit
- Chip area: 119 mm\(^2\)
- Functional Si: January ‘06

**32 nm Process**
- Cell size: 0.182 \(\mu\text{m}^2\)cell
- Capacity: 291 Mbit
- Chip area: 118 mm\(^2\)
- Functional Si: September ‘07
22 nm technology 6T SRAM Cell: Size = 0.1µm

Announced on Aug 18, 2008

Consortium: IBM (NYSE), AMD, Freescale, STMicroelectronics, Toshiba and the College of Nanoscale Science and Engineering (CNSE)

0.1µm cell size is almost on the down-scaling trend

New technologies introduced
- High-NA immersion lithography
- High-K metal gate stacks
- 25 nm gate lengths
- Thin composite oxide-nitride spacers
- Advanced activation techniques
- Extremely thin silicide
- Damascene copper contacts

Static noise margin of 220 mV at 0.9 V

Source: IEDM2008 Pre-conference Publicity
http://www.btbmarketing.com/iedm/
Cell size reduction trends

1/2 or 2/3 per cycle?

Cell area ($\mu m^2$)

- 0.57 $\mu m^2$ (Intel)
- 0.35 $\mu m^2$
- 0.24 $\mu m^2$
- 0.18 $\mu m^2$
- 0.15 $\mu m^2$
- 0.1 $\mu m^2$

65nm  Apr.2004
45nm  Jan.2006
32nm  Sep.2007
32nm  Dec.2007
32nm  Dec.2007
22nm  Aug.2008
\[ \sigma V_{\text{Tran}} = \left( \frac{4q^3 \varepsilon_{s,p} \phi_B}{2} \right) \cdot \frac{T_{ox}}{\varepsilon_{ox}} \cdot \left( \frac{4\sqrt{N}}{\sqrt{W_{\text{eff}} \cdot L_{\text{eff}}}} \right) \]

\[ = \frac{1}{\sqrt{2}} \left( \frac{c_2}{\sqrt{W_{\text{eff}} \cdot L_{\text{eff}}}} \right) \]

**NMOS Mismatch Coefficient (C_2)**

**improvement with technology scaling**

[Graph showing normalized NMOS C_2 with technology scaling]

Source: K.J.Kuhn
IEDM 2007
Mismatch improvement by layout (Intel)

“tall” design
90nm : 1.0 µm²

“wide” design
65nm : 0.57 µm²

“wide” design
(Square endcaps)
45nm 0.346 µm²

Source: K. J. Kuhn
IEDM2007 Tech. Dig. pp.471
Double patterning for square endcap

- Pattern gate lines/spaces
- Pattern cut mask
- Final gate pattern
- Intel 45nm SRAM cell

Source: M. Bohr, ICSICT2008

Cell evolution is similar

- TSMC 45nm
- TSMC 32nm
- IBM Gr. 32nm
- IBM Alliance 32nm
- IBM Alliance 22nm

TSMC 45nm IEDM 2007
TSMC 32nm IEDM 2007
IBM Alliance 32nm IEDM 2004
IBM Alliance 22nm IEDM 2008

Cell evolution is similar
Most Difficult part of SRAM down-scaling is Vdd down-scaling

Density of on-chip cache SRAM memory is high and thus, Vth cannot be down-scaled too much because of large Isd-leak

Also, under low Vdd, read- and write margin degrades, data retention degrade.

Thus, Vdd down-scaling is more severe in SRAM than logic part of the circuits
Intel® Xeon® 7400 Series (Dunnington)

45 nm high-k6 cores
16MB shared L3 cache

Source: Intel Developer Forum 2008

Cache occupies huge area

→ Cell size of SRAM should be minimized
→ Isd-leak should be minimized
   → Vth are often designed to be higher than Min. logic Vth
   → Lg are often designed to be larger than Min. logic Lg
Future Directions For Improving Vmin

- Application
  - Improvement in voltage and temperature tolerance
- Package
  - Separated array / logic voltage to minimize logic noise effect on SRAM
- Design
  - Higher array VDD and improved on-chip supply robustness
  - Increased redundancy
  - Improved timings
  - Cells per BL hierarchical BL structure
  - Write/Read assist and sense-amp design
- Cell and Process
  - Improved bit cell optimization
- NFET/PFET centering and Beta/Gamma control
- Minimize device fluctuation by limiting device-geometry scaling larger cell
- Lpoly, Weff, LER
  - Leakage / defect mechanisms

Source: Harold Pilo IEDM2006 Short Course
Nehalem (Intel) 2, 4 or 8 Cores

Voltage/Frequency Partitioning
- DDR Vcc
- Core Vcc
- Uncore Vcc

Dynamic Power Management

8T SRAM Cell
- 32kB L1 I-cache
- 32kB L1 D-cache
- 256kB L2 -cache

6T SRAM Cell
- 8 MB L3 cache

Source: Intel Developer Forum 2008
6T and 8T Cell

6T Cell
- Cell size is small
- For high density use

8T Cell
- Add separate read function
- Cell size increase 30%
- For low voltage use

Source: Morita et al, Symp. on VLSI Circ. 2007
5. Roadmap for further future as a Personal View
There will be still 4~6 cycles (or technology generations) left until we reach 11 ~ 5.5 nm technologies, at which we will reach down-scaling limit, in some year between 2020-30 (H. Iwai, IWJT2008).

Even after reaching the down-scaling limit, we could still continue R & D, seeking sufficiently higher Id-sat under low Vdd.

Two candidates have emerged for R & D
1. Nanowire/tube MOSFETs
2. Alternative channel MOSFETs (III-V, Ge)

Other Beyond CMOS devices are still in the cloud.

Source: 2008 ITRS Summer Public Conf.

*5.5nm? was added by Iwai
FinFET to Nanowire

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

Ion/loff=230000

N-FET
(Lg =10 nm)
Swing =75 mV/dec
DIBL =80 mV/V
I_eff =10 nA/μm

F.-L.Yang, VLSI2004
Si nanowire FET with Semi-1D Ballistic Transport

Merit of Si-nanowire

- Reduction in Ioff (Isd-leak)
  - Good control of Isd-leak by surrounding gate

Trade off

- Carrier scattering probability
  - Small → Large
  - # of quantum channel
  - Small → Large

- High Conduction (1D)
  - Go=77.8µS/wire

Increase in Ion (Id-sat)

- Multiple quantum channel (QC) used for conduction
- High-density lateral and vertical integration


Comparison with planar bulk MOSFET

Does Nanowire FET have really higher drive current per unit area?
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
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)
- Beyond the horizon
- Natural direction of downsizing
- Mechanical Stress, Roughness, Surface control
- Problem: High-k gate oxides, etching of III-V wire
- Selection
- Tube, Ribbon

ITRS
- ITRS Beyond CMOS
- High conduction
- By 1D conduction
- Extended CMOS
- More Moore ??

2007 2010 2015 2020 2025 2030 2035

Si Fin, Tri-gate
Si Nano wire
Ill-Vand Ge Nano wire
Tube
Ribon
Graphene
CNT
Nanowire
Si Channel
Research
Development
Production
Long term roadmap for development

Source: H. Iwai, IPFA 2006

Miniaturization of Interconnects on PCB (Printed Circuit Board)

Some time in 2020 - 2030

After 2050?

We do not know how?

We do know system and algorithms are important!

But do not know how it can be by us for use of bio?

Long term roadmap for development

We do not know how?

Some time in 2020 - 2030

After 2050?

Sensor
Infrared
Humidity
CO₂

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

Mosquito

Dragonfly brain has even further higher performance
Acknowledgement
I would like to express deep appreciation to the following people for the useful advice and support for material preparation. Special thanks to ITRS committee for the permission to refer roadmap and Public conference.

**ITRS Committee:** Hidemi Ishiuchi (Toshiba), Paolo Gargini (Intel)

**Toshiba Corporation:** Mitsuo Saito, Yukihiro Urakawa, Tomoaki Yabe

**Tsukuba University:** Kenji Shiraishi, Kenji Natori

**Intel Corporation:** Mark Bohr

**IBM Alliance:** B.S. Haran et al,

**Tokyo Institute of Technology:** Kuniyuki Kaukshima, Parhat Ahmet, Takamasa Kawanago, Yeonghun Lee
Thank you for your attention!