Technology Scaling and Roadmap for 22nm CMOS and beyond

May 1, 2009

@University College of Dublin

Hiroshi Iwai

Tokyo Institute of Technology
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, T_{ox}, V_{dd} )</th>
<th>( K )</th>
<th>Scaling ( K ): ( K=0.7 ) for example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drive current in saturation</td>
<td>( I_d )</td>
<td>( K )</td>
<td>( I_d = v_{sat} W_g C_o (V_g - V_{th}) ) ( C_o ): gate C per unit area</td>
</tr>
<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>
<tr>
<td>Gate capacitance</td>
<td>( C_g )</td>
<td>( K )</td>
<td>( C_g = \epsilon_o \epsilon_{ox} L_g W_g / t_{ox} ) ( \epsilon_{ox} ): gate C per unit area</td>
</tr>
<tr>
<td>Switching speed</td>
<td>( \tau )</td>
<td>( K )</td>
<td>( \tau = C_g V_{dd} / I_d )</td>
</tr>
<tr>
<td>Clock frequency</td>
<td>( f )</td>
<td>1/( K )</td>
<td>( f = 1/\tau = 1/K )</td>
</tr>
<tr>
<td>Chip area</td>
<td>( A_{chip} )</td>
<td>( \alpha )</td>
<td>( \alpha ): Scaling factor ( \rightarrow ) In the past, ( \alpha &gt; 1 ) for most cases</td>
</tr>
<tr>
<td>Integration (# of Tr)</td>
<td>( N )</td>
<td>( \alpha/K^2 )</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>( \alpha )</td>
<td>( fNCV^2/2 ) ( \rightarrow ) ( K^{-1}(\alpha K^{-2})K (K^1)^2 = \alpha ) = 1, when ( \alpha = 1 )</td>
</tr>
<tr>
<td></td>
<td>( k = 0.7 ) and ( \alpha = 1 )</td>
<td>( k = 0.7^2 = 0.5 ) and ( \alpha = 1 )</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------------------------------------------------</td>
<td>-------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Single MOFET</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( V_{dd} )</td>
<td>( 0.7 )</td>
<td>( 0.5 )</td>
<td></td>
</tr>
<tr>
<td>( L_g )</td>
<td>( 0.7 )</td>
<td>( 0.5 )</td>
<td></td>
</tr>
<tr>
<td>( I_d )</td>
<td>( 0.7 )</td>
<td>( 0.5 )</td>
<td></td>
</tr>
<tr>
<td>( C_g )</td>
<td>( 0.7 )</td>
<td>( 0.5 )</td>
<td></td>
</tr>
<tr>
<td>( P ) (Power)/Clock</td>
<td>( 0.7^3 = 0.34 )</td>
<td>( 0.5^3 = 0.125 )</td>
<td></td>
</tr>
<tr>
<td>( \tau ) (Switching time)</td>
<td>( 0.7 )</td>
<td>( 0.5 )</td>
<td></td>
</tr>
<tr>
<td><strong>Chip</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( N ) (# of Tr)</td>
<td>( 1/0.7^2 = 2 )</td>
<td>( 1/0.5^2 = 4 )</td>
<td></td>
</tr>
<tr>
<td>( f ) (Clock)</td>
<td>( 1/0.7 = 1.4 )</td>
<td>( 1/0.5 = 2 )</td>
<td></td>
</tr>
<tr>
<td>( P ) (Power)</td>
<td>( 1 )</td>
<td>( 1 )</td>
<td></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, 貫道 (Rule of right vs. Rule of military)
# Actual past downscaling trend until year 2000

## Change in 30 years

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

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

## Past 30 years scaling
- **Merit:** \( N, f \) increase
- **Demerit:** \( P \) increase

## Vdd scaling insufficient
- Additional significant increase in \( I_d, f, P \)

### Equation

\[ V_d \text{ scaling insufficient, } \alpha \text{ increased } \rightarrow \text{ } N, I_d, f, P \text{ increased significantly} \]

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

- Supply voltage reduction is becoming difficult, because $V_{th}$ 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)
What is a roadmap? What is ITRS?

Roadmap: Prediction of future technologies

ITRS: International Technology Roadmap for Semiconductors made by SIA (Semiconductor Industry Association) with Japan, Europe, Korea and Taiwan.
1992 - 1997: NTRS (National Technology Roadmap)
1998 - : ITRS (International Technology Roadmap)
ITRS Roadmap does change every year!

2006 Update    2002 Update    
2004 Update    2000 Update    

http://www.itrs.net/reports.html
Subthreshold Leakage (A/µm)

Operation Frequency (a.u.)

Source: 2007 ITRS Winter Public Conf.
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>
</tr>
</thead>
<tbody>
<tr>
<td>45 nm</td>
<td>2007</td>
</tr>
<tr>
<td>32 nm</td>
<td>2009?</td>
</tr>
<tr>
<td>22 nm</td>
<td>2011?~</td>
</tr>
<tr>
<td></td>
<td>2012?</td>
</tr>
<tr>
<td>16 nm</td>
<td>2013?~</td>
</tr>
<tr>
<td></td>
<td>2014?</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?

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

- 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: 1.5 µm, 1.2 µm, 1 µm

→ 350 nm → 250 nm → 180 nm → 130 nm → 90 nm → 65 nm → 45 nm

- ‘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’

→ 32 nm → 22 nm → 16 nm → 11 nm → 8 nm?? → 5.5 nm ??
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.

Source: 2008 ITRS Summer Public Conf.
Definition of the Half Pitch

Logic 1\textsuperscript{st} Metal Half Pitch

- DRAM $\frac{1}{2}$ Pitch
  = DRAM Metal Pitch/2
- MPU/ASIC M1 $\frac{1}{2}$ Pitch
  = MPU/ASIC M1 Pitch/2

Metal Pitch

Typical DRAM/MPU/ASIC Metal Bit Line

Flash Poly Gate Half Pitch

- FLASH Poly Silicon $\frac{1}{2}$ Pitch
  = Flash Poly Pitch/2

Typical flash
Un-contacted Poly

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

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.
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

Correspond to 22nm

Source: 2008/ ITRS Summer Public Conf.

<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>2007 MPU/ASIC Lg (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>
<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 MPU/ASIC Lg (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>
<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>2005</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>
<td>interp</td>
<td>interp</td>
<td>interp</td>
<td>interp</td>
</tr>
</tbody>
</table>

EOT w/3E20 poly, bulk MPU (nm)

<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>1.2</td>
<td>0.71</td>
<td>0.54</td>
<td>0.41</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>
</tbody>
</table>

EOT w/3E20 poly, bulk MPU (nm)

<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>1.3</td>
<td>1.2</td>
<td>1.2</td>
<td>1</td>
<td>0.68</td>
<td>0.64</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

EOT w/metal gate, bulk MPU (nm)

<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>0.9</td>
<td>0.75</td>
<td>0.65</td>
<td>0.55</td>
<td>0.50</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>
</tbody>
</table>

EOT w/metal gate, bulk MPU (nm)

<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>1.0</td>
<td>0.95</td>
<td>0.88</td>
<td>0.75</td>
<td>0.66</td>
<td>0.60</td>
<td>0.60</td>
<td>0.53</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Drain Ext. Xj bulk MPU (nm)

<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>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>
<td></td>
</tr>
</tbody>
</table>

Drain Ext. Xj bulk MPU (nm)

<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>11</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>

non-steady trend corrected

filled in for metal gate EOT for 2009/10 based on latest conference presentations
Advantage in RISC
Simple configuration

Advantage in SISC
Era for ‘out of order’

Clock frequency does not increase aggressively anymore.
Even decreased!

Multi Core
Clock
Performance

Source: Mitsuo Saito, Toshiba
source: 2007 ITRS Winter Public Conf.

MPU “GHz” by “Cores” ITRS2007

Cell Broadband Engine

6GHz capability for SRAM

Source: IBM, Toshiba, Sony ISSCC2008 and 08
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

Gamers “Clock-Doubling?”

1.29x/year (2x/2.5yrs)

1.41x/year (2x/2yrs)

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

Past < Future

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

22 nm: 6 GHz?

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)/(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>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

Correspond to 22nm

Maybe delay??
Gate CD (Critical Dimension) Control

ITRS 2007

Correspond to 22nm Logic

<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)(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 [AAL]</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

Correspond to 22nm Logic

<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>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interlevel metal insulator – effective dielectric constant (κ)</td>
<td>2.7-3.0</td>
<td>2.5-2.8</td>
<td>2.5-2.8</td>
<td>2.5-2.8</td>
<td>2.1-2.4</td>
<td>2.1-2.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year of Production</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interlevel metal insulator – effective dielectric constant (κ)</td>
<td>2.9-3.3</td>
<td>2.6-2.9</td>
<td>2.6-2.9</td>
<td>2.6-2.9</td>
<td>2.4-2.8</td>
<td>2.4-2.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year of Production</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interlevel metal insulator – bulk dielectric constant (κ)</td>
<td>2.3-2.7</td>
<td>2.1-2.4</td>
<td>2.1-2.4</td>
<td>2.1-2.4</td>
<td>1.8-2.1</td>
<td>1.8-2.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year of Production</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interlevel metal insulator – bulk dielectric constant (κ)</td>
<td>2.5-2.8</td>
<td>2.3-2.6</td>
<td>2.3-2.6</td>
<td>2.3-2.6</td>
<td>2.1-2.4</td>
<td>2.1-2.4</td>
</tr>
</tbody>
</table>

Source: 2008 ITRS Summer Public Conf.

k value increases by 0.1 ~ 0.3
Historical Transition of ITRS Low-k Roadmap

ITRS1999
ITRS2001
ITRS2003
ITRS2005
ITRS2007,8

Source: 2008 ITRS Summer Public Conf.
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

$\Delta V_{th} > V_{th}$

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

Subthreshold leakage current of MOSFET

$V_g = 0V$

Subthreshold leakage current

$I_{on}$

$I_{off}$

$V_{th}$ (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

\[(300 - 100)\text{mV}/(60\text{mV}/\text{dec}) = 3.3 \text{ dec}\]

**Subthreshold slope (SS)**

\[SS = (\ln 10)(kT/q)(C_{ox} + C_D + C_{it})/C_{ox} > \sim 60 \text{ mV/decade at RT}\]

**SS value:**

Constant and does not become small with down-scaling

---

**Log scale Id plot**

- **Ion**
- **Ioff**

- **Vdd**
  - Vdd=0.5V
  - Vdd=1.5V

- **Vth**
  - Vth = 300mV
  - Vth = 100mV

- **Vg (V)**
  - Vg = 0V
ITRS for HP logic

S-D leakage current

**Id-sd-leak has to be stay less than 1μA/μm**

Saturation current (μA/μm)

**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/loff ratio

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

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!

---

**Diagram:**
- High-k
- Gate oxd
- BO (Buried oxd)
- Enhanced by high-k
- Enhanced from backside
- Enhanced from both side
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(C₁) + EOT(C₃) > 0.5nm
Small effect to decrease EOT(C₂) beyond 0.5nm?
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)

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


EOT scaling below 0.5nm
Still useful for larger drain current
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.
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
   $SS = (\ln 10)\frac{(kT/q)(C_{ox}+C_D+C_{it})}{C_{ox}}$

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!
Enhanced D-Electric-field

Bulk  DG

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

$L_g=16\text{nm}$, $t_{ox}(\text{EOT})=0.5\text{nm}$,
$\text{Dopant@Channel}=8.1\times10^{18}\text{cm}^{-2}$

$\frac{\partial V(x,y)}{\partial V_d} \bigg|_{V_d=1\text{V}}$

$\Lambda$: Penetration Depth of DIBL

$\text{DIBL: Drain Induced Barrier Lowering}$

Comparison of Bulk and DG

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

$W_{\text{fin}} = 10.7\text{ nm}$ $W_{\text{fin}} = 30\text{ nm}$ $W_{\text{fin}} = 40\text{ nm}$

$\Lambda = 7.6\text{ nm}$ $\Lambda = 17.1\text{ nm}$ $\Lambda = 13.2\text{ nm}$

$\text{SS}$

$\text{DIBL@D Edge (mV/V)}$

Sub-threshold Swing (mV/dec)
Comparison of High-k and SiO$_2$ MOSFETs

Enhanced D-Electric-field

- $L_g = 40$ nm, $V_d = 0.1$ V, EOT = 2 nm
- $k = 390$ too large high-k SiO$_2$
- $K = 3.9$

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 $I_{sd-leak}$ and increase $I_{d-sat}$ 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.
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/dfall_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

<table>
<thead>
<tr>
<th>Technology</th>
<th>Cell size</th>
<th>Capacity</th>
<th>Chip area</th>
<th>Functional Si</th>
</tr>
</thead>
<tbody>
<tr>
<td>90 nm Process</td>
<td>1.0 ( \mu \text{m}^2 \text{cell} )</td>
<td>50 Mbit</td>
<td>109 mm(^2)</td>
<td>February ‘02</td>
</tr>
<tr>
<td>65 nm Process</td>
<td>0.57 ( \mu \text{m}^2 \text{cell} )</td>
<td>70 Mbit</td>
<td>110 mm(^2)</td>
<td>April ‘04</td>
</tr>
<tr>
<td>45 nm Process</td>
<td>0.346 ( \mu \text{m}^2 \text{cell} )</td>
<td>153 Mbit</td>
<td>119 mm(^2)</td>
<td>January ‘06</td>
</tr>
<tr>
<td>32 nm Process</td>
<td>0.182 ( \mu \text{m}^2 \text{cell} )</td>
<td>291 Mbit</td>
<td>118 mm(^2)</td>
<td>September ‘07</td>
</tr>
</tbody>
</table>
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?

Intel
- 65nm Apr. 2004
- 45nm Jan. 2006
- 32nm Sep. 2007

TSMC
- 45nm Dec. 2007
- 32nm Dec. 2007

IBM Alliance (Consortium)
- 32nm Dec. 2007

Press release
- 22nm Aug. 2008
\[ \sigma V_{Tran} = \left( \frac{4 \sqrt{4q^3 \varepsilon_s \phi_B}}{2} \right) \cdot \frac{T_{ox}}{\varepsilon_{ox}} \cdot \left( \frac{\sqrt{N}}{\sqrt{W_{eff} \cdot L_{eff}}} \right) \]

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

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

“tall” design
90nm : 1.0 $\mu m^2$

“wide” design
65nm : 0.57 $\mu m^2$

“wide” design
(Square endcaps)
45nm 0.346 $\mu m^2$

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

a) Pattern gate lines/spaces
b) Pattern cut mask
c) Final gate pattern
d) Intel 45nm SRAM cell

Source: M. Bohr, ICSICT2008

Cell evolution is similar

TSMC 45nm
TSMC 32nm
IBM Alliance 32nm
IBM Alliance 22nm

Cell evolution is similar

TSMC 45nm
TSMC 32nm
IBM Gr. 32nm

Source: M. Bohr, ICSICT2008
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
Voltage/Frequency Partitioning
- DDR Vcc
- Core Vcc
- Uncore Vcc

Dynamic Power Management

Nehalem (Intel) 2, 4 or 8 Cores

Chip

Core

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.
Si nanowire FET with Semi-1D Ballistic Transport

Merit of Si-nanowire

Reduction in $I_{off}$ ($I_{sd-leak}$)
- Good control of $I_{sd-leak}$ by surrounding gate

Increase in $I_{on}$ ($I_{d-sat}$)

Trade off
- Carrier scattering probability
  - Small $\rightarrow$ Large
  - Large $\rightarrow$ Small
- # of quantum channel
  - Small $\rightarrow$ Large
  - Large $\rightarrow$ Small

High Conduction (1D)
- $G_0 = 77.8 \mu S/wire$

Multiple quantum channel (QC) used for conduction

High-density lateral and vertical integration


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
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

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?
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!