Although We’ve Come to the End of the Road(map): The Future of CMOS

Nicole DiLello
6.Insight
April 3, 2007
Boyz II Men

Although we've come
To the end of the road
Still I can't let go
It's unnatural
You belong to me
I belong to you

♥ You = silicon ♥
Outline

- Introduction
- Current technology
- Next generation
- 20 years from now
- Conclusion
Outline

- Introduction
- Current technology
- Next generation
- 20 years from now
- Conclusion
Moore’s Law
nMOSFET

$V > 0$

Gate

Source

Drain

n

n

p-well

Gate oxide
pMOSFET

\[ V < 0 \]
Before continuing...

- \( \text{nMOS} + \text{pMOS} = \text{CMOS} \)
- Standard MOSFET
  - Made entirely from Si
  - Crystalline
  - n/p regions are doped
MOSFET Characteristics

\[ I_{D,\text{linear}} = \frac{W}{L} \mu C_{\text{ox}} (V_{GS} - V_T - \frac{1}{2} V_{DS}) V_{DS} \]

\[ I_{D,\text{Sat}} = \frac{W}{2L} \mu C_{\text{ox}} (V_{GS} - V_T)^2 \]

\[ V_{\text{DS,sat}} = V_{GS} - V_T \]

\[ V_{GS} \uparrow \]

linear

saturation

cutoff

\[ V_{\text{GS}} = V_T \]
Following Moore’s Law...

\[ I_{D,\text{linear}} = \frac{W}{L} \mu C_\text{ox} (V_{GS} - V_T - \frac{1}{2} V_{DS}) V_{DS} \]

\[ I_{D,\text{Sat}} = \frac{W}{2L} \mu C_\text{ox} (V_{GS} - V_T)^2 \]

- Make transistors smaller
- International Technology Roadmap for Semiconductors (ITRS)
  - U.S., Europe, Japan, Taiwan, Korea
  - Chip manufacturers, academia
### ITRS (2005)

<table>
<thead>
<tr>
<th>Year of Production</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRAM ½ Pitch (nm)  (contacted)</td>
<td>80</td>
<td>70</td>
<td>65</td>
<td>57</td>
<td>50</td>
<td>45</td>
<td>40</td>
<td>36</td>
<td>32</td>
</tr>
<tr>
<td>MPU/AISC Metal 1 (M2) ½ Pitch (nm) (contacted)</td>
<td>90</td>
<td>78</td>
<td>68</td>
<td>59</td>
<td>52</td>
<td>45</td>
<td>40</td>
<td>36</td>
<td>32</td>
</tr>
<tr>
<td>MPU Physical Gate Length (nm)</td>
<td>32</td>
<td>28</td>
<td>25</td>
<td>22</td>
<td>20</td>
<td>18</td>
<td>16</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td><em>Lg</em> Physical <em>Lg</em> for High-Performance logic (nm)</td>
<td>32</td>
<td>28</td>
<td>25</td>
<td>22</td>
<td>20</td>
<td>18</td>
<td>16</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td><strong>EOT</strong>: Equivalent Oxide Thickness [2]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extraded Planar Bulk (Å)</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>9</td>
<td>7.5</td>
<td>6.5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>UTB FD (Å)</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>DG (Å)</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td><strong>Gate Poly Depletion and Inversion-Layer Thickness</strong> [3]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extraded Planar Bulk (Å)</td>
<td>7.3</td>
<td>7.4</td>
<td>7.4</td>
<td>6.1</td>
<td>5.3</td>
<td>4.7</td>
<td>4.1</td>
<td>3.8</td>
<td>3.5</td>
</tr>
<tr>
<td>UTB FD (Å)</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>DG (Å)</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td><strong>EOT_{eq}</strong>: Electrical Equivalent Oxide Thickness in inversion [4]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extraded Planar Bulk (Å)</td>
<td>19.3</td>
<td>18.4</td>
<td>18.4</td>
<td>11.9</td>
<td>10.3</td>
<td>9.2</td>
<td>7.5</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>UTB FD (Å)</td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>DG (Å)</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td><em>J_{gmax}</em>: Maximum gate leakage current density [5]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extraded Planar Bulk (A/cm²)</td>
<td>1.88E+02</td>
<td>5.36E+02</td>
<td>8.00E+02</td>
<td>9.05E+02</td>
<td>1.10E+03</td>
<td>1.58E+03</td>
<td>2.08E+03</td>
<td>2.34E+03</td>
<td>2.34E+03</td>
</tr>
<tr>
<td>FD/SOI (A/cm²)</td>
<td>7.73E+02</td>
<td>9.93E+02</td>
<td>1.22E+03</td>
<td>1.38E+03</td>
<td>2.07E+03</td>
<td>2.34E+03</td>
<td>2.34E+03</td>
<td>2.34E+03</td>
<td>2.34E+03</td>
</tr>
<tr>
<td>DG (A/cm²)</td>
<td>8.25E+02</td>
<td>7.86E+02</td>
<td>8.46E+02</td>
<td>8.46E+02</td>
<td>8.46E+02</td>
<td>8.46E+02</td>
<td>8.46E+02</td>
<td>8.46E+02</td>
<td>8.46E+02</td>
</tr>
<tr>
<td><em>V_{ddq}</em>: Power Supply Voltage (V) [6]</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td><em>V_{sat}</em>: Saturation Threshold Voltage [7]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extraded Planar Bulk (mV)</td>
<td>195</td>
<td>168</td>
<td>165</td>
<td>160</td>
<td>159</td>
<td>151</td>
<td>146</td>
<td>148</td>
<td>148</td>
</tr>
<tr>
<td>UTB FD (mV)</td>
<td>160</td>
<td>168</td>
<td>167</td>
<td>170</td>
<td>166</td>
<td>166</td>
<td>167</td>
<td>167</td>
<td>167</td>
</tr>
<tr>
<td>DG (mV)</td>
<td>181</td>
<td>184</td>
<td>185</td>
<td>185</td>
<td>185</td>
<td>185</td>
<td>185</td>
<td>185</td>
<td>185</td>
</tr>
<tr>
<td><em>I_{doff}</em>: Source/Drain Subthreshold Off-State Leakage Current [8]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extraded Planar Bulk (uA/μm)</td>
<td>0.06</td>
<td>0.15</td>
<td>0.2</td>
<td>0.2</td>
<td>0.22</td>
<td>0.28</td>
<td>0.32</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>UTB FD (uA/μm)</td>
<td>0.17</td>
<td>0.19</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.29</td>
<td>0.29</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td>DG (uA/μm)</td>
<td>0.1</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
</tr>
</tbody>
</table>
ITRS (2005)

![Red brick wall image]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DRAM % Pitch (µm) (contacted)</td>
<td>28</td>
<td>23</td>
<td>22</td>
<td>22</td>
<td>20</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td>MPUASIC Metal 1 (µm) % Pitch (µm)(contacted)</td>
<td>22</td>
<td>23</td>
<td>22</td>
<td>20</td>
<td>18</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>MPU Physical Gate Length (nm)</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>$L_p$: Physical $L_{min}$ for High Performance logic (nm) [1]</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>$EOT$: Equivalent Oxide Thickness [2]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extended planar bulk (Å)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UTB FD (Å)</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DG (Å)</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Gate Poly: Depletion &amp; Inversion-Layer Thickness [1]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extended planar bulk (Å)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UTB FD (Å)</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DG (Å)</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>$EOT_{tot}$: Electrical Equivalent Oxide Thickness in inversion [4]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extended Planar Bulk (Å)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UTB FD (Å)</td>
<td>9</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DG (Å)</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>$I_{leak}$: Maximum gate leakage current density [5]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extended Planar Bulk (A/cm$^2$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FDSOI (A/cm$^2$)</td>
<td>3.27E+03</td>
<td>3.70E+03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DG (A/cm$^2$)</td>
<td>1.09E+03</td>
<td>1.16E+03</td>
<td>1.22E+03</td>
<td>1.38E+03</td>
<td>1.57E+03</td>
<td>1.83E+03</td>
<td>2.29E+03</td>
</tr>
<tr>
<td>$V_{DD}$: Power Supply Voltage (V) [6]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$V_{sat}$: Saturation Threshold Voltage [7]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extended Planar Bulk (mV)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UTB FD (mV)</td>
<td>164</td>
<td>166</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DG (mV)</td>
<td>190</td>
<td>192</td>
<td>195</td>
<td>200</td>
<td>201</td>
<td>205</td>
<td>208</td>
</tr>
<tr>
<td>$I_{leak}$: Source/Drain Subthreshold Off-State Leakage Current [5]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extended Planar Bulk (µA/µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UTB FD (µA/µm)</td>
<td>0.36</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DG (µA/µm)</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
</tr>
</tbody>
</table>
Outline

- Introduction
- Current technology
- Next generation
- 20 years from now
- Conclusion
Strain Engineering - nMOS

\[ I_{D,\text{linear}} = \frac{W}{L} \mu C_{ox} (V_{GS} - V_T - \frac{1}{2} V_{DS}) V_{DS} \]

\[ I_{D,\text{Sat}} = \frac{W}{2L} \mu C_{ox} (V_{GS} - V_T)^2 \]
Strain Engineering - pMOS

\[ \mu_e \text{ increases for tensile strain} \]
\[ \mu_h \text{ increases for compressive strain} \]
Nitride film causes tensile stress in channel $\Rightarrow$ increase $\mu_e$

SiGe causes compressive stress in channel $\Rightarrow$ increase $\mu_h$

Can induce stress for both nMOS and pMOS on the same wafer!
Outline

- Introduction
- Current technology
- Next generation
- 20 years from now
- Conclusion
C\textsubscript{ox} rears its ugly head

\[ I_D = \frac{W}{L} \mu_e C_{ox} (V_{GS} - V_T - \frac{1}{2} V_{DS}) V_{DS} \]

\[ I_D = \frac{W}{2L} \mu_e C_{ox} (V_{GS} - V_T)^2 \]

Problems with both!

\[ C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \]
SiO$_2$: How do I love thee? Let me count the ways…

- Native oxide for Si → easy to grow
- Good quality: resistant to water, other atmospheric elements
- Matches Si lattice well → no dangling bonds
- High breakdown voltage
- $\varepsilon_{\text{SiO}_2} = 3.9 \varepsilon_0$
High-\(\kappa\) (\(k\), if you’re lazy)

- Measured in terms of “equivalent oxide thickness” (EOT)
- Can make thicker layers
  - \(t_{\text{ox}} \approx 1\) nm

\[
t_{\text{high-}\kappa} = \frac{K_{\text{high-}\kappa}}{K_{\text{SiO}_2}} t_{\text{SiO}_2}
\]

If \(\kappa = 16\), can have a thickness of 4 nm that gives roughly the same \(C_{\text{ox}}\) as 1 nm of \(\text{SiO}_2\).
SiO$_x$N$_y$

- Relatively easy to integrate (just add some N$_2$)
- Increases $\kappa$ a bit ($\kappa_{\text{Si}_3\text{N}_4} \sim 7$)
- Intel introduced at 90 nm node (2004)
- Probably limited to a thickness of $\sim 1.3$ nm
- Need a better fix
# Some Other Options

<table>
<thead>
<tr>
<th>Material</th>
<th>Dielectric constant ($\kappa$)</th>
<th>Band gap $E_G$ (eV)</th>
<th>$\Delta E_C$ (eV) to Si</th>
<th>Crystal structure(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiO$_2$</td>
<td>3.9</td>
<td>8.9</td>
<td>3.2</td>
<td>Amorphous</td>
</tr>
<tr>
<td>Si$_3$N$_4$</td>
<td>7</td>
<td>5.1</td>
<td>2</td>
<td>Amorphous</td>
</tr>
<tr>
<td>Al$_2$O$_3$</td>
<td>9</td>
<td>8.7</td>
<td>2.8$^a$</td>
<td>Amorphous</td>
</tr>
<tr>
<td>Y$_2$O$_3$</td>
<td>15</td>
<td>5.6</td>
<td>2.3$^a$</td>
<td>Cubic</td>
</tr>
<tr>
<td>La$_2$O$_3$</td>
<td>30</td>
<td>4.3</td>
<td>2.3$^a$</td>
<td>Hexagonal, cubic</td>
</tr>
<tr>
<td>Ta$_2$O$_5$</td>
<td>26</td>
<td>4.5</td>
<td>1−1.5</td>
<td>Orthorhombic</td>
</tr>
<tr>
<td>TiO$_2$</td>
<td>80</td>
<td>3.5</td>
<td>1.2</td>
<td>Tetrag.$^c$ (rutile, anatase)</td>
</tr>
<tr>
<td>HfO$_2$</td>
<td>25</td>
<td>5.7</td>
<td>1.5$^a$</td>
<td>Mono.$^b$, tetrag.$^c$, cubic</td>
</tr>
<tr>
<td>ZrO$_2$</td>
<td>25</td>
<td>7.8</td>
<td>1.4$^a$</td>
<td>Mono.$^b$, tetrag.$^c$, cubic</td>
</tr>
</tbody>
</table>

$^a$Calculated by Robertson (See Ref. 153).

$^b$Mono.$^b$ = monoclinic.

$^c$Tetrag.$^c$ = tetragonal.
Intel’s Announcement

- Jan. 27, 2007, *New York Times*: 45 nm process (in production later this year) will use Hf-based dielectric (HfO$_2$? Si-based alloy? Shhhhh…) and metal gate
- Leakage current is down, drive current is up, power consumption is down
- IBM: Hey, we did it too!
Poly-Si Gates

- Currently: gates made from poly-Si
- Doped at $>10^{20}$ cm$^{-3}$
- Pro: same material for nMOS and pMOS
- Cons: Can only dope so high, poly depletion effects
Poly depletion

Deplete 3 – 4 Å in the gate \(\rightarrow\) adds 3 – 4 Å to dielectric (significant when dielectric is 12 Å) \(\rightarrow\) reduces \(C_{ox}\)

\[
C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}
\]
Metal Gate

- No poly depletion
- Lower series resistance
- Different for nMOS and pMOS

- Incorporating SiGe, Hf-based dielectric, and metal gates show that the industry is willing to (slowly) incorporate new materials
Outline

- Introduction
- Current technology
- Next generation
- 20 years from now
- Conclusion
Look at bigger picture

- What if we change the mode of transportation entirely?
  - Tunneling FET (TFET)?
  - Carbon nanotubes?
  - Computation bubbles?
  - PHOTONICS!

- Before we go crazy-nuts, should probably look into a hybrid system
Integrated electronic/photonic system

- Photonics are good for transmitting data, high frequency applications
- Electronics are good for processing data, especially in a small area
- Let’s use both!
Analog-to-digital Converter

Use of a mode-locked laser → low sampling jitter!

Parallel processing on different wavelengths

Germanium photodetectors
But still Si-based

- Everything on-chip $\rightarrow$ faster!
- Integrate laser
- Si modulator
- Ge photodetectors
  - SiGe already in CMOS process
- $\text{SiO}_2$ and $\text{SiN}_x$ waveguides
  - Already in CMOS process

**Key:** Optical sampling drastically reduces the timing jitter
Germanium photodiode

Diagram showing layers of oxide, N+ polysilicon, intrinsic Ge, and SiO$_2$. 

- N+ polysilicon
- Intrinsic Ge
- SiO$_2$

Text elements:
- > 1 $\mu$m Intrinsic Ge Layer
- P+ Ge seed
- P+ (100) Si Substrate

Scanning electron microscope image showing the cross-section of the photodiode structure.
Another system: Multi-core Processor

- Working within an existing CMOS process
- Materials constraints
- Process constraints
- Pros: 1,000 cores, much more energy efficient, faster
Helpful Classes

- Devices: 6.012, 6.720J, 6.728, 6.730, 6.731
- Processing: 6.152J, 6.774, 6.781
- Optics/Photonics: 6.013, 6.630, 6.631
- Circuits: 6.002, 6.301
- Computer Architecture: 6.823
Groups at MIT

- Strain engineering – Hoyt, Fitzgerald
- High-κ materials – Antoniadis
- Electronic Photonic Integrated Circuits (EPIC) – Kaertner, Hoyt, Ram, H.I. Smith, Ippen
- Multi-core processor – Stojanovic, Asanovic, Hoyt, Ram, Kaertner, H.I. Smith, Schmidt
Conclusions

- “No exponential is forever, but we can delay ‘forever.’”
- New materials: whoo!
- Electronic photonic architectures: whoo!