MOSFET Scaling Trends, Challenges, and Potential Solutions Through the End of the Roadmap: A 2005 Perspective

Peter M. Zeitzoff
Howard R. Huff
Outline

- Introduction
  - MOSFET scaling and its impact
  - Front-end approaches and solutions
  - Non-classical CMOS
  - Summary
Introduction

• IC Logic technology: following Moore’s Law by rapidly scaling into deep submicron regime
  – Increased speed and function density
  – Lower power dissipation and cost per function

• But the scaling results in major MOSFET and process integration issues, including
  – Simultaneously maintaining satisfactory $I_{on}$ (drive current) and $I_{leak}$
  – High gate leakage current for very thin gate dielectrics
  – Control of short channel effects for very small transistors
  – Etc.

• Potential solutions & approaches
  – Material and process (front end): high-k gate dielectric, metal gate electrodes, strained Si, …
  – Structural: non-classical CMOS device structures

• This talk gives an updated perspective from the 2003 International Technology Roadmap for Semiconductors (ITRS)
Metrology and Characterization Issues

• Dimensional scaling: meeting metrology requirements for accuracy and precision becomes increasingly challenging
  – Example: electrical and physical measurement of Tox < 1.2 nm
  – Another example: CD measurement
    • Line edge (and width) roughness is increasingly critical

• Potential solutions (high-k, metal gate electrodes, strained Si, non-classical CMOS) raise significant metrology and characterization challenges
### Key Overall Chip Parameters for High-Performance Logic, from 2003 ITRS

<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>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MPU</strong></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>Physical Gate Length (nm)</td>
<td>45</td>
<td>37</td>
<td>32</td>
<td>28</td>
<td>25</td>
<td>22</td>
<td>20</td>
<td>18</td>
<td>14</td>
<td>13</td>
<td>10</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td><strong>On-chip local clock</strong></td>
<td>2,976</td>
<td>4,171</td>
<td>5,204</td>
<td>6,783</td>
<td>9,285</td>
<td>10,972</td>
<td>12,369</td>
<td>15,079</td>
<td>20,065</td>
<td>22,980</td>
<td>33,403</td>
<td>39,683</td>
<td>53,207</td>
</tr>
<tr>
<td><strong>Allowable Maximum Power</strong></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>High-performance with heatsink (W)</td>
<td>149</td>
<td>158</td>
<td>167</td>
<td>180</td>
<td>189</td>
<td>200</td>
<td>210</td>
<td>218</td>
<td>240</td>
<td>251</td>
<td>270</td>
<td>288</td>
<td>300</td>
</tr>
<tr>
<td>Cost-performance (W)</td>
<td>80</td>
<td>84</td>
<td>91</td>
<td>98</td>
<td>104</td>
<td>109</td>
<td>114</td>
<td>120</td>
<td>131</td>
<td>138</td>
<td>148</td>
<td>158</td>
<td>168</td>
</tr>
<tr>
<td><strong>Functions per chip at production (million transistors [Mtransistors])</strong></td>
<td>153</td>
<td>193</td>
<td>243</td>
<td>307</td>
<td>386</td>
<td>487</td>
<td>614</td>
<td>773</td>
<td>1,227</td>
<td>1,546</td>
<td>2,454</td>
<td>3,092</td>
<td>4,908</td>
</tr>
</tbody>
</table>

- Rapid scaling of $L_g$ is driven by need to improve transistor speed
- Clock frequency, functions per chip (density) scale rapidly, but allowable power dissipation rises slowly with scaling
Outline

- Introduction
  - MOSFET scaling and its impact
- Front-end approaches and solutions
- Non-classical CMOS
- Summary
Device Scaling Approach: 2003 ITRS

- Simple models capturing essential MOSFET physics → embedded in a spreadsheet
  - Room T, nominal devices assumed
  - Key parameters include: $L_g$, $T_{ox}$, $V_{dd}$, $V_t$, series parasitic resistance, drive current, leakage current, gate capacitance, subthreshold slope, etc.

- Using spreadsheet, MOSFET parameters are iteratively varied to meet ITRS targets for either
  - Scaling of transistor speed OR
  - Scaling for specific, low levels of leakage current
MOSFET Intrinsic Performance Parameter

- Transistor intrinsic delay, $\tau$
  - $\tau \sim CV_{dd}/(I_{on})$
  - $I_{on}$ units: $\mu$A/µm
  - $C \sim C_L$
    - Gate dominated case: appropriate for local, dense logic
    - $C \sim C_L = C_{gate} \sim C_{ox}L_g + C_{parasitic}$
    - $C_{ox} \sim \varepsilon_{ox}/T_{ox}$
    - $\tau$ is the delay for a load consisting of one transistor’s gate capacitance; shortest logic delay possible

- Transistor intrinsic switching frequency = $1/\tau$
  - Good metric for transistor performance
  - To maximize $1/\tau$, maximize $I_{on}$
Different Applications ➔ Different ITRS Drivers

- **High-performance chips** (MPU, for example)
  - Driver: maximize chip speed ➔ maximize transistor performance
    - Goal of ITRS scaling: $1/\tau$ increases at ~ 17% per year, historical rate
      - Must maximize $I_{on}$
      - Consequently, $I_{leak}$ is relatively high

- **Low-power chips** (mobile applications)
  - Driver: minimize chip power (to maximize battery life) ➔ minimize $I_{leak}$
    - Goal of ITRS scaling: specific, low level of $I_{leak}$
      - Consequently, $1/\tau$ is considerably less than for high-performance logic
1/τ and Isd,leak scaling for High-Performance and Low-Power Logic. Data from 2003 ITRS.

Calendar Year

1/τ (GHz)

Isd,leak (µA/µm)

17%/yr ave. increase

1/τ—High Perf

Isd,leak—High Perf

1/τ—Low Power

Isd,leak—Low Power

1/τ and Isd,leak scaling for High-Performance and Low-Power Logic. Data from 2003 ITRS.
Outline

• Introduction
• MOSFET scaling and its impact
  ➢ Front-end material and processing approaches and solutions
• Non-classical CMOS
• Summary
Difficult Transistor Scaling Issues

- Previously discussed scaling results involve determining the required transistor characteristics and performance to meet key scaling targets
  - Assumption: highly scaled MOSFETs with required characteristics can be successfully fabricated

- With scaling, increasing difficulty in meeting transistor requirements without significant technology innovations
  - **High gate leakage**
    - Direct tunneling increases rapidly as $T_{ox}$ is reduced
    - Potential solution: high-k gate dielectric
  - Polysilicon depletion in gate electrode $\rightarrow$ increased effective $T_{ox}$, reduced $I_{on}$
  - Need for enhanced channel mobility
  - Etc.
For Low-Power Logic, Gate Leakage Current Density Limit Versus Simulated Gate Leakage due to Direct Tunneling. Data from 2003 ITRS.

Beyond this point of cross over, oxy-nitride is incapable of meeting the limit ($J_g,\text{limit}$) on gate leakage current density.
High K Gate Dielectric to Reduce Direct Tunneling

- Equivalent Oxide Thickness = EOT = $T_{ox} = T_K \times (3.9/K)$, where 3.9 is relative dielectric constant of SiO$_2$ and K is relative dielectric constant of high K material
  - $C = C_{ox} = \varepsilon_{ox}/T_{ox}$
  - To first order, MOSFET characteristics with high-k are same as for SiO2
- Because $T_K > T_{ox}$, direct tunneling leakage much reduced with high K
  - If energy barrier is high enough
- Candidate materials: LaO$_2$/HfO$_2$/ZrO$_2$(K~15 - 30); Hf, Zr-SiO$_4$ (K~12 - 16); others
  - Major materials, process, integration issues to solve
MOCVD HfO₂ TEM (EOT = 0.95 nm) (HfO₂ on HF-last, N₂O-750°C Pre-Deposition Anneal)

PVD TiN 450 Å
HfO₂ 21 Å
Interfacial layer 12 Å
Silicon substrate

- Effective k for above dielectric stack ≈ 13.5
- k for interfacial layer could be significantly greater than SiO₂ indicating reaction or intermixing of HfO₂ film with interfacial SiO₂

Avinash Agarwal et al., Alternatives to SiO₂ as Gate Dielectrics for Future Si-Based Microelectronics, 2001 MRS Workshop Series (2001)
Difficult Transistor Scaling Issues

• With scaling, increasing difficulty in meeting transistor requirements without significant technology innovations
  – High gate leakage
    • Direct tunneling increases rapidly as $T_{ox}$ is reduced
  – **Polysilicon depletion in gate electrode** $\rightarrow$ **increased effective electrical** $T_{ox}$, **reduced $I_{on}$**
    • Potential solution: metal gate electrodes
  – Need for enhanced channel mobility
  – Etc.
Polysilicon Depletion and Substrate Quantum Effects

- $T_{ox,electric} = T_{ox} + (K_{ox}/K_{si}) \cdot (W_{d,Poly} + X_{C,inv})$
  
  - $K_{ox} = 3.9$
  - $K_{si} = 11.9$

- $T_{ox,electric} = T_{ox} + (0.33) \cdot (W_{d,Poly} + X_{C,inv})$
  - $W_{d,Poly} \sim 1/(\text{poly doping})^{0.5}$
  
  → increase poly doping to reduce $W_{d,Poly}$ with scaling

- But max. poly doping is limited → can’t reduce $W_{d,Poly}$ too much

- Fermi Level pinning with hight-$k$

- Poly depletion and $X_{C,inv}$ become more critical with $T_{ox}$ scaling

- Eventually, poly will reach its limit of effectiveness
Metal Gate Electrodes

- Metal gate electrodes are a potential solution when poly “runs out of steam”: probably implemented at 65 nm tech. generation (2007) or beyond
  - No depletion, very low resistance gate, no boron penetration, compatibility with high-k
- Issues
  - Different work functions needed for PMOS and NMOS ==> 2 different metals may be needed
    - Process complexity, process integration problems, cost
  - Etching of metal electrodes
  - New materials: major challenge
Advanced Gate Stack: Key Metrology and Characterization Challenges

- Transient charge trapping in high-k bulk
  - Characterizing charge trapping
  - Extracting mobility: \( \mu_{\text{eff}} = \frac{(Ld)}{(WQ_{\text{inv}}V_d)} \)
  - Determining \( V_t, V_{FB} \) from C-V
  - Fast pulse measurements help
- Charge in high-k & Interaction between metal gate and high-k: unambiguous determination of \( \phi_m \)
- Gate leakage \( \Rightarrow \) determining EOT from C-V
Example: Fast Transient Electron Trapping with Pulse Measurements on High-k Gate Dielectric

Significant trapping occurs within few $\mu$sec

$V_d = -1$ to $2.5$ V
$PW = 35$ ns, $100$ $\mu$s, $1$ ms

C. Young, SSDM 2004
Difficult Transistor Scaling Issues

• With scaling, increasing difficulty in meeting transistor requirements
  – High gate leakage
    • Direct tunneling increases rapidly as $T_{ox}$ is reduced
  – Polysilicon depletion in gate electrode $\rightarrow$ increased effective $T_{ox}$, reduced $I_{on}$
  – Need for enhanced channel mobility
    • Potential solution: strained Si channels
  – Etc.
Band Engineered MOSFETs: Strained MOSFET Structures

(Courtesy of J. Hoyt - MIT)

- Increased effective mobility, increased $I_{on}$
- Difficult integration issues: manufacturability, thermal stability, simultaneous optimization of both PMOS and NMOS, defects, leakage
- Compatibility with ultra-thin body SOI
- Cost

(J. Welser, J.L. Hoyt, and J.F. Gibbons, IEDM, 1992, pp. 1000-1003.)

Strained Si Device Structures

 Courtesy of Patricia Mooney (IBM Corp.) From P. M. Mooney et al., presented at the American Physics Society Meeting, Austin, TX, March 3-7, 2003.

modified band structure of Si under biaxial tensile strain ==> enhanced mobility

need relaxed Si$_{1-x}$Ge$_x$ with 0.15<$x<$0.35
Alternate Approach: Uniaxial Process Induced Stress

NMOS: uniaxial tensile stress from stressed SiN film

PMOS: uniaxial compressive stress from sel. SiGe in S/D

Strained Si: Metrology and Characterization Challenges

- Measuring strain distribution with high spatial resolution in deep sub-micron structures
  - Possible approaches
    - X-ray diffraction (XRD)
    - Raman spectroscopy
    - Convergence Beam Electron Diffraction (CBED)
    - Electron Diffraction Contrast (EDC)
Outline

• Introduction
• Scaling and its impact
• Front end approaches and solutions
  ➢ Non-classical CMOS
• Summary
• 65 nm tech. generation (2007, $L_g = 25$nm) and beyond: increased difficulty in meeting all device requirements with classical planar, bulk CMOS (even with material and process solutions: high K, metal electrodes, ...)
  - Control of SCE
  - Impact of quantum effects and statistical variation
  - Impact of high substrate doping
  - Control of series S/D resistance ($R_{\text{series,s/d}}$)
  - Others

• Alternative device structures (non-classical CMOS) may be utilized
  - Ultra thin body, fully depleted: single-gate SOI and multiple-gate transistors
Transistor Structures

Planar Bulk

- Current solution
- Wafer cost / availability
- SCE scaling difficult
- High doping effects and Statistical variation
- Parasitic junction capacitance

Partially Depleted SOI

- Lower junction cap
- F.B. performance boost
- F.B. history effect
- SCE scaling difficult
- Wafer cost/availability

REFERENCES
Schematic cross section of planar bulk, UTB SOI, and DG SOI MOSFET
Transistor Structures

Planar Bulk

- Depletion Region
- Substrate

Partially Depleted SOI

- Buried Oxide (BOX)
- Substrate

Fully Depleted SOI

- BOX
- Substrate

+ Lower junction cap
+ Light doping possible
- SCE scaling difficult
- High $R_{\text{series.s/d}}$ elevated S/D
- Sensitivity to Si thickness (very thin)
- Wafer cost/availability

+ Lower junction cap
+ F.B. performance boost
- F.B. history effect
- SCE scaling difficult
- Wafer cost/availability

+ Current solution
+ Wafer cost / availability
- SCE scaling difficult
- High doping effects and Statistical variation
- Parasitic junction capacitance

REFERENCES
To reduce SCE’s, aggressively reduce Si layer thickness.

E-Field lines

Single-Gate SOI

Courtesy: Prof. J-P Colinge, UC-Davis
Double Gate Transistors

+ Enhanced scalability
+ Lower junction capacitance
+ Light doping possible, with near-midgap metal gate
+ ~2x drive current
- ~2x gate capacitance
- High $R_{series,s/d}$ $\rightarrow$ raised S/D
- Complex process

Summary: more advanced, optimal device structure, but difficult to fabricate, particularly in this SOI configuration

REFERENCES
Field Lines for Single and Double-Gate MOSFETs

E-Field lines

To reduce SCE’s, aggressively reduce Si layer thickness

Double gates electrically shield the channel

Single-Gate SOI

Double-Gate

Courtesy: Prof. J-P Colinge, UC-Davis
Other Double-Gate Transistor Structures (FinFET)

Perspective view of FinFET. Fin is colored yellow.

Gate overlaps fin here

SiO₂

Source

Drain

BOX

Substrate Silicon

Fin

Key advantage: relatively conventional processing, largely compatible with current techniques → current leading approach

Courtesy: T-J. King and C. Hu, UC-Berkeley

Top View of FinFET

Arrow indicates current flow direction

Source

Poly Gate

Drain

SiO₂
Types of Multiple-Gate Devices

1: Single gate
2: Double gate
3: Triple gate
4: Quadruple gate (GAA)
5: Π gate

Buried Oxide

Courtesy: Prof. J-P Colinge, UC-Davis
Metrology and Characterization Challenges for Non-Classical CMOS

- C-V measurements for thin, fully-depleted Si
- Single-gate SOI: measurement of very thin body thickness, 10 nm and less
- Multiple-gate
  - Measurement of fin height and width, high AR
  - Measuring roughness of vertical fin edges
  - Measuring high-k film thickness on vertical fin edges
  - Measuring 2D and 3D doping profiles in thin fins
Outline

• Introduction
• Scaling and its impact
• Front end approaches and solutions
• Non-classical CMOS
➤ Summary
Summary

• Rapid transistor scaling will continue through the end of the Roadmap
  - Transistor performance will improve rapidly, but leakage will be hard to control
  - Many technology innovations will be needed in relatively short time to enable this rapid scaling
    • Front-end potential solutions include high-k gate dielectric, metal gate electrodes, and enhanced mobility through strained silicon
      – High-k needed first for low-power (mobile) chips in ~ 2006
    • Structural potential solutions: non-classical CMOS
      • The technology innovations will raise significant challenges for metrology and characterization
• Non-classical CMOS and front-end solutions being pursued in parallel, and will likely be combined in the ultimate, end-of-Roadmap device
  – \( L_g < 10\text{nm} \) MOSFETs expected by the end of the Roadmap in 2018
BACK-UP
Potential Solutions for Power Dissipation Problems, High-Performance Logic

• Due to high leakage, static power dissipation is a special challenge

• Increasingly common approach: multiple transistor types on a chip\(\rightarrow\)multi-Vt, multi-Tox, etc.
  – Only utilize high-performance, high-leakage transistors in critical paths—lower leakage transistors everywhere else
  – Improves flexibility for SOC

• Electrical or dynamically adjustable $V_t$ devices (future possibility)

• Circuit and architectural techniques: pass gates, power down circuit blocks, etc.

• Improved heat removal, electro-thermal modeling and design
<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>Strained Si–HP</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>High-k (Low Power)</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>Elevated S/D</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>High-k (HP)</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>Metal Gate (HP, dual gate)</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>Metal Gate (Low Power, dual gate)</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>Ultra-thin Body (UTB) SOI, single gate (HP)</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>Metal gate (near midgap for UTBSOI)</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>Strained Si (Low Power)</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>Multiple Gate (HP)</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>Ultra-thin Body (UTB) SOI, single gate (Low power)</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>Multiple Gate (Low Power)</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>Quasi-ballistic transport (HP)</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>Quasi-ballistic transport (LOP)</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>
</tbody>
</table>

This timeline is from PIDS evaluation for the 2003 ITRS.
International Technology Roadmap for Semiconductors (ITRS)

- Industry-wide, fully international effort to map IC technology generations for the next 15 years
  - For each technology generation
    - Projects targets for technology characteristics and requirements
    - Assesses key needs and gaps
    - Lists potential solutions
  - Provides common reference for semiconductor industry: device manufacturers, equipment and materials vendors, researchers
    - Useful for planning
    - Focus: stimulating needed R&D, not intended to restrict research
    - Enabling factor in continuing to follow Moore’s Law
  - Much of this talk is based on the 2003 ITRS (formally presented in Dec., 2003)
## Typical Technology Requirements Table: High-Performance Logic. Data from 2003 ITRS.

<table>
<thead>
<tr>
<th>Year in Production</th>
<th>Units</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Lgates (High Performance)</td>
<td>nm</td>
<td>45</td>
<td>37</td>
<td>32</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>EOT (Equivalent Oxide Thickness)</td>
<td>Å</td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Gate Poly Depletion &amp; Inversion-Layer Thickness</td>
<td>Å</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Inversion Gate Dielectric Thickness Value</td>
<td>Å</td>
<td>21</td>
<td>20</td>
<td>18</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Maximum Gate Leakage Limit</td>
<td>A/cm²</td>
<td>2.2E+02</td>
<td>4.5E+02</td>
<td>5.2E+02</td>
<td>6.0E+02</td>
<td>9.3E+02</td>
</tr>
<tr>
<td>Power Supply Voltage</td>
<td>V</td>
<td>1.2</td>
<td>1.2</td>
<td>1.1</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Saturation Threshold Voltage</td>
<td>V</td>
<td>0.21</td>
<td>0.20</td>
<td>0.20</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>Source/Drain Subthreshold Off-State Leakage Drain Current</td>
<td>uA/um</td>
<td>0.03</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Effective NMOS Current Drive</td>
<td>uA/um</td>
<td>980</td>
<td>1110</td>
<td>1090</td>
<td>1170</td>
<td></td>
</tr>
<tr>
<td>Sub-threshold Slope Adjustment Factor (Full Depletion/Dual-Gate Effects)(0-1)</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobility Enhancement Factor</td>
<td></td>
<td>1</td>
<td>1.3</td>
<td>1.3</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>Effective Saturation Carrier Velocity Enhancement Factor</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effective Parasitic Rsd</td>
<td>ohm-um</td>
<td>180</td>
<td>180</td>
<td>180</td>
<td>171</td>
<td></td>
</tr>
<tr>
<td>Ideal NMOS Device Gate Capacitance</td>
<td>F/um</td>
<td>7.4E-16</td>
<td>6.4E-16</td>
<td>6.1E-16</td>
<td>5.7E-16</td>
<td>6.6E-16</td>
</tr>
<tr>
<td>Parasitic Fringe/Overlap Capacitance</td>
<td>F/um</td>
<td>2.4E-16</td>
<td>2.4E-16</td>
<td>2.4E-16</td>
<td>2.3E-16</td>
<td>2.2E-16</td>
</tr>
<tr>
<td>NMOS Device Time Constant</td>
<td>ps</td>
<td>1.20</td>
<td>0.95</td>
<td>0.86</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>Relative Performance Improvement (compared to 2003)</td>
<td></td>
<td>1.00</td>
<td>1.26</td>
<td>1.39</td>
<td>1.60</td>
<td></td>
</tr>
<tr>
<td>Nominal Gate Delay (NAND Gate)</td>
<td>ps</td>
<td>30.24</td>
<td>23.94</td>
<td>21.72</td>
<td>18.92</td>
<td></td>
</tr>
<tr>
<td>NMOS Device Static Power Dissipation due to Drain &amp; Gate Leakage</td>
<td>Watts/um</td>
<td>3.96E-07</td>
<td>6.60E-07</td>
<td>6.05E-07</td>
<td>6.05E-07</td>
<td>8.43E-07</td>
</tr>
<tr>
<td>NMOS Device Power Delay Product</td>
<td>Joules/um</td>
<td>1.41E-15</td>
<td>1.27E-15</td>
<td>1.03E-15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Design Max On-Chip Clock Frequency vs. A&P Max Off-Chip (Chip-to-Board) Frequency

**Year**


**Max. Freq. (Ghz)**

1.0 10.0 100.0

**Design**

2001 ITRS

**2003 ITRS**

1.17x/year (2x/4.5yrs)

1.29x/year (2x/2.5yrs)

"Gap" Delayed by 3 years

1.41x/year (2x/2yrs)

**IS: Design/Architecture:** reduction of maximum # of inverter delays to flat at 12 beginning 2007

**WAS:** (2001 ITRS: flat at 16 after 2006)

**MPU Clock Frequency Historical Trend:**

Gate Scaling, Transistor Design contributed ~17-19%/year

Architectural Design innovation contributed additional ~13-21%/year

**Actual Scaling Acceleration or Equivalent Scaling Innovation Needed to maintain historical trend**

Goal: Increase Speed by 2x Speed/2-2.5 years

**Courtesy:** Alan Allan, Intel
ITRS Projections of $V_{dd}$ and $V_t$ Scaling. Data from 2003 ITRS.
Key MOSFET Scaling Results, 2003 ITRS: Performance and Leakage

• High-performance logic
  - **Average 17%/yr improvement** in $1/\tau$ is attained
  - $I_{sd,\text{leak}}$ is high, particularly for 2007 and beyond ➔ chip static power dissipation scaling is an issue

• Low-power logic
  - **Very low** $I_{sd,\text{leak}}$ target is met
  - $I_{\text{gate,leak}}$ is also very low: difficult to meet this ➔ need for high-k gate dielectric
    - $1/\tau$ is considerably lower than for high-performance, but close to 17%/yr improvement in $1/\tau$ is still attained

• ITRS MOSFET targets are chosen to drive the technology scaling ➔ pretty aggressive
Frequency scaling: Transistor Intrinsic, Fanout-3 NAND Gate, Chip Clock for High-Performance Logic. Data from 2003 ITRS.

Chip clock: ITRS projection

Chip Clock: assumption is that only improvement here is from transistor speed increase
Hierarchy of IC Requirements and Choices

Overall Circuit Requirements and Choices
- Chip Power
- Chip Speed
- Functional Density
- Chip Cost
- Architecture
- Etc.

Overall MOSFET Requirements and Choices
- \( V_{dd} \)
- MOSFET Leakage
- MOSFET Drive current
- Parasitic series resistance
- Transistor size
- \( V_t \) control
- Reliability
- Etc.

MOSFET Design Choices
- \( T_{ox}, L_g, x_j, R_s \)
- Channel engineering
- Oxynitride or High K gate dielec.
- Classical Planar Bulk or Non-classical CMOS Structures
- Etc.

Process Integration Choices
- Thermal processing
- Overall process flow
- Process modules
- Material properties
- Boron penetration
- Etc.
Key Overall Chip Parameters for High-Performance Logic, from 2001 ITRS

<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>
</tr>
</thead>
<tbody>
<tr>
<td>DRAM Half Pitch nm</td>
<td>130</td>
<td>115</td>
<td>100</td>
<td>90</td>
<td>80</td>
<td>70</td>
<td>65</td>
<td>45</td>
<td>35</td>
<td>22</td>
</tr>
<tr>
<td>Physical Gate Length, Lg nm</td>
<td>65</td>
<td>53</td>
<td>45</td>
<td>37</td>
<td>32</td>
<td>28</td>
<td>25</td>
<td>22</td>
<td>18</td>
<td>9</td>
</tr>
<tr>
<td>Nominal Power Supply Voltage (Vdd) V</td>
<td>1.2</td>
<td>1.1</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9</td>
<td>0.9</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Maximum on-chip local clock frequency GHz</td>
<td>1.7</td>
<td>2.3</td>
<td>3.1</td>
<td>4.0</td>
<td>5.2</td>
<td>5.6</td>
<td>6.7</td>
<td>11.5</td>
<td>19.4</td>
<td>28.8</td>
</tr>
<tr>
<td>Allowable maximum power dissipation, with heatsink W</td>
<td>130</td>
<td>140</td>
<td>150</td>
<td>160</td>
<td>170</td>
<td>180</td>
<td>190</td>
<td>218</td>
<td>215</td>
<td>288</td>
</tr>
<tr>
<td>Number of transistors per chip Millions of transistors</td>
<td>276</td>
<td>348</td>
<td>439</td>
<td>553</td>
<td>697</td>
<td>878</td>
<td>1106</td>
<td>2212</td>
<td>4424</td>
<td>8848</td>
</tr>
</tbody>
</table>

- The DRAM half pitch and Lg are drivers of IC technology scaling, including lithography.
- Technology generations (in red) defined by DRAM half pitch.
  - This is a dense feature: drives functional density and Litho. and Etch.
  - Reduction factor of 0.7X ~ 1/√2 between generations (130nm in 2001, 90nm in 2004, 65nm in 2007, etc.)
  - Three years between generations.
  - Gate length (Lg) ≤ 0.5 X DRAM half pitch.
  - These are isolated features.
  - Rapid scaling of Lg is driven by need to improve transistor speed.
V_{dd} and V_{t} Device Scaling Issues

- We need to scale V_{dd} down rapidly with the technology generations
  - To keep dynamic power dissipation (∼V_{dd}^2) within acceptable bounds
  - For reliability, control of short channel effects (SCE), general device scaling
- \frac{1}{I_{sd, leak}} \text{ exp. dependent on } V_{t}
- I_{on} \text{ strongly dependent on gate overdrive, } (V_{dd} - V_{t})
- Also, V_{dd} ≥ 2 V_{t} for circuit functionality

Scaling requires key tradeoffs between I_{on} and I_{sd, leak}, V_{dd} and V_{t}
- Tradeoff choices driven by application needs
Historical Data and 2001 ITRS Projections for Chip Clock Frequency, High-Performance Logic

Historical Trend:
Transistor scaling has contributed ~ 17-19%/year
Architectural Design innovation contributed additional ~ 21-13%/year

Actual Scaling Acceleration, Or Equivalent Innovation Needed to maintain historical trend

Courtesy: Alan Allan, Intel. Data from 2001 ITRS.
Potential Problem with Static Power Dissipation Scaling: High-Performance Logic

Assumption, to make a point re $P_{static}$: all transistors are high performance, low $V_t$ type
Relative Chip Power Dissipation, High Performance

Assumptions:

- Only one type of (high \(I_{on}\), high \(I_{leak}\)) transistor, and no power reduction techniques.
- Simple scaling
- This is an unrealistic scenario, only meant to clearly illustrate static power dissipation issues

**Chip Static Power Dissipation**

**Chip Dynamic Power Dissipation**

**Allowable Total Chip Power Dissipation**
ITRS Projected Scaling of Power Dissipation per Device

- **Static Power Dissipation/device** (W/device)
  - 2001: 1.0E-13
  - 2006: 1.0E-12
  - 2011: 1.0E-11
  - 2016: 1.0E-10

- **Power-delay product, High-Perf.**
- **Power-delay product, Low Power**
- **Static Pdiss, Low Power**
Impact of Key MOSFET Parameters on Chip Power Dissipation

- \( P_{\text{total}} = P_{\text{dynamic}} + P_{\text{static}} \)
  - \( P_{\text{dynamic}} = C_{\text{active}} V_{\text{dd}}^2 f_{\text{clock}} \)
    - With scaling, \( C_{\text{active}} \) and \( f_{\text{clock}} \) increase rapidly
    - To keep \( P_{\text{dynamic}} \) within tolerable limits, reduce \( V_{\text{dd}} \) with scaling
  - \( P_{\text{static}} = N_{\text{off}} W L \text{leak} V_{\text{dd}} \)
    - With scaling, \( N_{\text{off}} \) increases rapidly, but \( V_{\text{dd}} \) and \( W \) scale down
    - To keep \( P_{\text{static}} \) within tolerable limits, constrain increase of \( \text{leak} \) with scaling
Solutions for Power Dissipation Problems, High-Performance Logic

- Increasingly common approach: multiple transistor types on a chip → multi-\(V_t\), multi-\(T_{ox}\)
  - Only utilize high-performance, high-leakage transistors in critical paths—lower leakage transistors everywhere else
  - Improves flexibility for SOC
- Electrical or dynamically adjustable \(V_t\) devices (future possibility)
- Circuit and architectural techniques: pass gates, power down circuit blocks, etc.
Summary: MOSFET Scaling

• MOSFET scaling is the “raw material” for designers to improve chip performance, control power dissipation
  - MOSFET scaling projected to scale at historical ~17% per year in “raw” speed improvement for high-performance logic
  - Design and architectural innovation has contributed about as much, but is expected to slow down in the future: continued MOSFET speed improvement is critically important

• MOSFET scaling goals are critically important
  - High-performance logic emphasizes speed at the expense of high leakage and static power dissipation
  - Low-power logic emphasizes low leakage at the expense of speed

• Static power dissipation is a growing problem for high-performance logic, and there are numerous approaches to dealing with it
High-K Issues

- Process integration
  - Thermal stability of high-k material
    - Retain high-k performance with planar CMOS flow (S/D anneal, etc.,) challenge
  - Chemical, electrical compatibility with polysilicon
    - Boron penetration
    - PMOS $V_t$
    - Metal electrode may be required
  - Interface with Si substrate and gate electrode
    - Deposition / post process anneals $\Rightarrow$ thin SiO$_2$-like layer

- Interface properties: $D_{it}$, $Q_f$, $\mu = \mu$(interfacial “SiO$_2$”)

- Charges and charge trapping in high-k: $V_t$ control and instability

- Mobility degradation

- Leakage, reliability

- New material: major challenge
**Polysilicon Limitations**

- **Polysilicon depletion**
  - Increases effective electrical $T_{ox}$ reduces inversion charge & $I_{on}$
  - More of a problem as $T_{ox}$ is scaled Poly doping must increase with scaling
- **PMOSFETs**: B penetration through very thin oxides
  - Oxy-nitrides & reduction of DT effective now
- **Compatibility with high-k**
- **Gate resistance of very thin gates (even with silicide)**
Transistor Structures

**Planar Bulk**
- Current solution
- Wafer cost / availability
- SCE scaling difficult
- High doping effects and Statistical variation
- Parasitic junction capacitance

**Partially Depleted SOI**
- Lower junction cap
- F.B. performance boost
- F.B. history effect
- SCE scaling difficult
- Wafer cost/availability

**Fully Depleted SOI**
- Lower junction cap
- Light doping possible
- SCE scaling difficult
- High $R_{\text{series,s/d}}$ elevated S/D
- Sensitivity to Si thickness (very thin)
- Wafer cost/availability

**REFERENCES**
Non-Classical CMOS Summary

- Below $L_g = 25\text{nm}$ or so, planar bulk CMOS may not scale effectively
  - Ultra-thin body, single-gate SOI and (eventually) multiple-gate, ultra-thin body MOSFETs are more optimal from a device point of view than planar bulk CMOS. Key issues:
    - Effectiveness of planar bulk CMOS scaling in this regime
      - Working but suboptimal 8nm devices reported in literature
    - Finding effective solutions to difficult processing issues for SOI and multiple-gate
      - Ultimate MOSFET ($L_g < 10\text{nm}$) likely to be multiple-gate with high-k, metal gate electrodes, strained Si, etc.
    - Such devices will require metal electrodes with near-midgap work functions
      - Tuning of work function of single metal gate material may be feasible
High-k Gate Dielectric Candidates and Key Issues

- **Modest k (<10)**
  - Al₂O₃
    - Negative charge, complicated defect structure

- **Medium k (10-25)**
  - Group IV Oxides - ZrO₂, HfO₂
    - Low crystallization temperature
  - Group III Oxides - Y₂O₃, La₂O₃,…
    - Charge
  - Silicates - (Zr, Hf, La, Y, ..) SiO₄
    - Lower k if too dilute
  - Aluminates - (Zr, Hf, La, Y, ..)•Al₂O₃
    - Charge issue, complicated defect structure

- **High k (≥ 25)**
  - Ta₂O₅ , TiO₂
    - Low-barrier height

C. M. Osburn and H.R. Huff, Spring ECS abst. # 366
MOCVD HfO\textsubscript{2} CV Curve (EOT = 0.95 nm)

Corrected Freq Data

- Area = 5.0E-5 cm\textsuperscript{2}
- Frequency = 100 and 250 kHz
- Gate Electrode: PVD TiN
- HfO\textsubscript{2} @ 485°C
- Interface: N2O at 750°C

Gate Leakage
- 4.3 A/cm\textsuperscript{2} @ 1V
- 10 A/cm\textsuperscript{2} @ V\textsubscript{fb}+1

CVC Model
- EOT = 0.95 nm
- V\textsubscript{fb} = -0.239 V
- N\textsubscript{surf} = 2.11E15

Avinash Agarwal et al., (Alternatives to SiO\textsubscript{2} as Gate Dielectrics for Future Si-Based Microelectronics, 2001 MRS Workshop Series (2001)
Difficult Transistor Scaling Issues

- With scaling, increasing difficulty in meeting transistor requirements

  - High gate leakage
    - Direct tunneling increases rapidly as $T_{ox}$ is reduced
    - Potential solution: high-k gate dielectric
  - Poly depletion in gate electrode $\rightarrow$ increased effective $T_{ox}$, reduced $I_{on}$
    - Potential solution: metal gate electrode
  - Need for enhanced channel mobility
    - Potential solution: strained Si channels
  - Etc.
Other Structures of Interest

Vertical FET (one type of double-gate MOSFET)

Silicon on Nothing (SON): localized buried oxide (BOX)

REF: Mark Bohr, ECS Meeting PV 2001-2, Spring, 2001

REF: S. Monfray et al., '01 IEDM, p. 645.
Reproduced by permission of The Electrochemical Society, Inc.
Assumptions for All Logic Types

- All modeling is done for nominal devices, room T
- Models are simplified (spreadsheet-based), assume basic transistor functioning doesn’t change
  - No dynamic Vt
  - S=85 mV/decade
  - $EOT_{electrical} = EOT + 0.8 \text{ nm}/0.4\text{ nm} \Rightarrow 0.8 \text{ nm} \text{ for poly gate, } 0.4 \text{ nm} \text{ for metal gate (in 2007 or beyond)}$
  - $\log(I_{sd,leak}) \sim -Vt/S$
    - Gate leakage and junction leakage are related to $I_{sd,leak}$
  - $I_d, sat \sim g_{m,eff} (Vdd-Vt)$
  - $C_{ideal} = \epsilon_{ox}/(EOT_{electrical}); C_{gate} = C_{ideal} + C_{parasitic}$
  - $\tau = (C_{gate} Vdd)/(I_d, sat) = \text{intrinsic transistor delay}$
  - Parasitic Rs,d is included (20-30% of $Vdd/I_d, sat = R_{on}$)
  - PMOS is like NMOS, except PMOS $I_d, sat$ is 40-50% of NMOS $I_d, sat$
  - S/D junction capacitance is ignored in calculating $\tau$
Simplified Cross-Section of High K Gate Dielectric Stack

- Spacer
- High-k Gate Dielectric Stack
- Gate electrode
- Upper interfacial region
- Bulk high-k film
- Lower interfacial region

Source, Substrate, Drain
Process - Structure - Property Relation

- Crystallinity / polycrystallinity
  - Phase structure
    - Epitaxial alignment to substrate
  - Stoichiometry
  - Bond coordination
  - Morphology
  - Interfacial microroughness
- Retention of amorphicity by doping
- Mixed oxide phase separation
- Spatial inhomogeneity / periodicity in energy gap(s)
Why High-K (Dielectric Constant)?

• Direct tunneling current depends on film physical thickness and barrier height ($\phi$)
  - $I_{DT} \propto [\exp - \sqrt{[2m^*q \phi / (h/2\pi)^2]}] [T_{phys}]$

• Transistor drive current depends on film electrical thickness
  - $I_{DSAT} = (w/2l) (3.9K_0A) (T_{EOT,INV})^{-1} \mu (V_G-V_T)^2; V_G \Rightarrow V_{DD}$
  - $T_{EOT} = T_{phys} \times (k_{SiO_2}/k_{high k})$
    - $k_{SiO_2} = 3.92; k_{high k} \approx 15 - 25$

• Increasing $k$ increases $I_{dsat}$ without increasing $I_{DT}$
  - Transistor performance improves or thickness may be increased (with increased $k$) to reduce gate leakage (direct tunneling) current without loss of transistor performance

• High-k gate dielectric proposed to obviate IC power concern while still achieving required gate electrode capacitive coupling with silicon

• High-k introduces new set of design constraints
Emerging Technology Parametrization

Log size in m
Log switching time in sec
Log cost in $/gate²
Log(energy in Joules)

QUANTUM
RSFQ
OPTICAL
SI CMOS
NEUROMORPHIC
NEMS
PLASTIC
MOLECULAR
Simplified Cross Section of a Typical PMOSFET and NMOSFET

(Not to scale)
Difficult Transistor Scaling Issues

- With scaling, increasing difficulty in meeting transistor requirements
  - High gate leakage
    - Direct tunneling increases rapidly as $T_{ox}$ is reduced
  - Poly depletion in gate electrode → increased effective $T_{ox}$, reduced $I_{on}$
  - Scaling S/D extension and deep S/D
    - High $R_{series,s/d}$ → reduced $I_{on}$
  - Etc.
S/D Extension Issues

Schematic, not to scale

Deep S/D

Silicide

S/D Extension

Lateral Abruptness

ρ_{s,ext}

X_{j,ext}

Deep S/D

Silicide

FMOS

NMOS
S/D Extension Issues

- Increasingly abrupt, shallow, heavily doped profiles required for successively scaled technologies
  - Needed for optimal devices, esp. to control short channel effects (SCE)
  - Difficult $\rho_s - x_{j,ext}$ tradeoffs, esp. for PMOS (B) $\Rightarrow$ difficult to control $R_{S/D,series}$
65 nm node and beyond: may require novel doping and annealing techniques

(Courtesy: L. Larson, D. Sing, R. Tichy of International SEMATECH)
S/D Extension Potential Solutions

• Shorter range
  – Ultra-low energy implants (< 1 KeV, B)
  – Rapid Thermal Processing (RTP) and spike anneal: reduces DT & TED
  – Increase dose as much as possible ==> reduced $R_{\text{series,s/d}}$

• Beyond 90 nm technology generation
  – Laser thermal annealing
  – Doped, selective epi
  – Co-implant
  – Others
Deep S/D & Silicide Issues

Schematic, not to scale

PMOS

Silicide

Deep S/D

S/D Extension

NMOS

Silicide, $\rho_{s,silicide}$

Deep S/D

$R_{\text{contact}}$

$X_j,\text{deep}$

$X$, silicon consumption

• Difficult tradeoffs to keep $R_{\text{series,s/d}}$ scaling down as required
  
  $-\rho_{s,silicide}$ must be minimized $\rightarrow X$ must be maximized
  
  But $X$ must be kept $\leq X_j,\text{deep}/2$ to avoid excessive junction leakage
  
  Also, $R_{\text{contact}}$ must be reduced with scaling
Deep S/D & Silicide Potential Solutions

- Through 90 or 65 nm generation, tradeoffs to get acceptable $R_{\text{series,s/d}}$ are possible
  - Change silicide to get better $\rho_{\text{s,silicide}}$ - $X$ tradeoff: $\text{TiSi}_2 \rightarrow \text{CoSi}_2 \rightarrow \text{NiSi}$

- Potential long-range solutions
  - Elevated S/D: doped, selective epi
  - Reduced $R_{\text{contact}}$
    - Selective CVD silicide tailors Schottky energy barrier
    - Selective deposited metal
Elevated S/D with Selective (Epitaxial) Silicon and Post Implant

- Implant extension
- Form spacer
- Epi deposition
- Silicide formation

Courtesy: Eric Graetz, Infineon
Device Metrics

- **Power**
  - \( P_{\text{dynamic}} = f_{\text{clock}} C_{\text{load}} V_{DD}^2 \) and \( P_{\text{static}} = N_{tr} W I_{\text{leak}} V_{DD} \)

- **Intrinsic transistor gate delay (speed)**
  - \( \tau = C_{\text{load}} V_{DD} / I_{\text{DSAT}} \)
  - Maximum saturated drain current \((I_{DSAT})\): ideal, long-channel device
    - \( I_{DSAT} = (W/2L_{\text{phys}}) (3.9K_o A) (T_{EOT,INV})^{-1} \mu_{\text{eff}} (V_G-V_T)^2 \)
      - \( W \) and \( L_{\text{phys}} \) device width and physical gate length
      - \( T_{EOT,INV} \) = equivalent oxide thickness in inversion
      - \( \mu_{\text{eff}} \) = mobility, generally determined for a long-channel device \((g_m)\)
      - \( V_G-V_T \) = gate overdrive, where \( V_G \) is supply voltage \((V_{DD})\) applied to gate \((V_G \Rightarrow V_{DD})\) and \( V_T \) is threshold voltage
    - \( C_{\text{load}} \approx (3.9K_o A) (T_{EOT,INV})^{-1} = \varepsilon_{\text{ox}} / T_{EOT,INV} \)
  - **Transconductance**
    - \( g_m = (W/L_{\text{phys}}) (3.9K_o A) (T_{EOT,INV})^{-1} \mu_{\text{eff}} V_{DD} \)
    - \( T_{EOT} = (k_{\text{high}} k / k_{\text{SiO2}}) T_{\text{phys}} \)
    - \( S = \text{Sub-threshold swing} \Rightarrow \text{Inverse slope of log } I_D \text{ versus } V_G \)
High Resolution TEM Showing 30 nm Channel Length

Polysilicon Gate

3.5 nm SiO₂

13 layers of Si atoms consumed to create 3.5 nm SiO₂

Electron mean free path

Inversion charge

30 nm Channel Length

78 columns of Si atoms

Source

Drain

Donor atom

Acceptor atom

Courtesy of Yoshi Nishi / Dick Chapman

Mar 25, 2003 2003 International Conference on Characterization and Metrology for ULSI Technology
Representative Theoretical and Universal Mobility Curve

\[ \mu_{pk,\text{univ}} \]
\[ \mu_{pk} \]
\[ \mu_{hi,\text{univ}} \]
\[ \mu_{hi} \]

\[ E_T \]
\[ E_{pk} \]
\[ E_{hi} \]

Universal curve

Representative curve: high-k
Mobility Considerations

- **Theoretical**
  - Low electric field
    - Unscreened (by inversion layer free carriers) ionized dopant scattering centers in silicon
  - High electric field
    - Acoustic phonons
    - Surface microroughness
      - $H \times L$ (where $H$ is height of surface undulation and $L$ is undulation correlation length)
    - Remote scattering due to high-k phonons
- **Experimental adders (not presently theoretically modeled)**
  - Interfacial and high-k bulk traps
  - N, Al and other elemental scatterers
  - Crystalline inclusions in amorphous high k gate dielectric
  - Remote scattering due to gate electrode
- **Universal curve only considers high electric field contributions (extends to low electric field)**
Electron Transport in εMOS™

Unstrained

perpendicular $\Delta_2$ valleys

[001] → [010] → [100]

in-plane $\Delta_4$ valleys

Tensile strain splits conduction band degeneracy

Biaxial Tension

[001] → [010] → [100]

• Reduced intervalley scattering
• Light in-plane effective mass

Courtesy of Matt Currie
AmberWave Systems Corp.
Hole Transport in εMOS™

Unstrained

E

in-plane

out-of-plane

k

Heavy Hole

Light Hole

Split-Off

Biaxial Tension

in-plane

out-of-plane

Tensile strain splits valence band degeneracy

• Reduced intervalley scattering
• Light in-plane effective mass

Courtesy of Matt Currie
AmberWave Systems Corp.
Transistor Structures

Planar Bulk
- Lower junction cap
- F.B. performance boost
- F.B. history effect
- SCE scaling difficult
- Wafer cost/availability

Partially Depleted SOI
- Lower junction cap
- F.B. performance boost
- F.B. history effect
- SCE scaling difficult
- Wafer cost/availability

Fully Depleted SOI
- Lower junction cap
- SCE scaling difficult
- High $R_{series,s/d}$ raised S/D
- Sensitivity to Si thickness (very thin)
- Wafer cost/availability

References:
1. P. Zeitzoff, J. Hutchby and H. Huff, to be pub. in Internat. Jour. Of High Speed Electronics and Systems
2. Mark Bohr, ECS Meeting PV 2001-2, Spring, 2001
Schematic cross section of planar bulk, UTB SOI, and DG SOI MOSFET

Inversion Layer

Depletion Region

Bulk MOSFET

Ultra-Thin Body SOI

Si Substrate

Ultra-thin silicon film

Double-Gate SOI MOSFET

Si Substrate

Ultra-thin silicon film
Schematic cross section of planar bulk, UTB SOI, and DG SOI MOSFET

- **Bulk MOSFET**
  - Inversion Layer
  - Depletion Region

- **Ultra-Thin Body SOI**
  - Si Substrate
  - Ultra-thin silicon film

- **Double-Gate SOI MOSFET**
  - Ultra-thin silicon film

Schematic cross section of planar bulk, UTB SOI and DG SOI MOSFET

- **Bulk MOSFET**
  - Inversion Layer
  - Depletion Region

- **Ultra-Thin Body SOI**
  - Si Substrate
  - Ultra-thin silicon film

- **Double-Gate SOI MOSFET**
  - G1
  - G2
  - BOX
  - Si Substrate
  - Ultra-thin silicon film
Electric field lines from the drain encroach on the channel region. Any increase of drain voltage decreases the threshold voltage (the “NPN” potential barrier between source and drain is lowered).
E-Field lines

Ground-plane SOI MOSFETs
Electrostatic Scaling - Channel Leakage ($I_{off}$)

Gate Leakage

Source Drain

Substrate

Gate

Drain

Gate Leakage

Channel Leakage

Thermionic Emission

QM Tunneling

BTB Tunneling

$E_{CB}$

$E_{VB}$

Source $L_{gate}$ Drain

Sum = $I_{off}$

Channel Leakage

Jim Hutchby