Transition from Planar MOSFETs to FinFETs and its Impact on Design and Variability

Victor Moroz

Berkeley Seminar, October 28, 2011
Discussion Topics

• Transistor scaling
• FDSOI
• SuVolta
• Stress engineering
• Non-silicon materials
• FinFET
• Summary
Discussion Topics

• Transistor scaling
  ◦ FDSOI
  ◦ SuVolta
  ◦ Stress engineering
  ◦ Non-silicon materials
  ◦ FinFET
  ◦ Summary
TEM Images of High Performance FETs

All TEM images here have the same scale

- Very little change in physical gate length, only ~0.9x per node
- The gate pitch is scaling fast, as 0.7x per node and area scales as 0.5x
- Most of the transistor innovation is in stress engineering and HKMG
TEM Images of High Performance FETs

All TEM images here have the same scale

- Very little change in physical gate length, only ~0.9x per node
- The gate pitch is scaling fast, as 0.7x per node and area scales as 0.5x
- Most of the transistor innovation is in stress engineering and HKMG
ITRS for a long time (from 2003 to 2008) insisted on extrapolating poly gate length to zero

This corresponds to straightforward 0.7x scaling per generation

Meanwhile, the industry kept a much slower, 0.9x scaling pace from 90nm to 30nm
Transistor Size Evolution: Poly Gate

• ITRS for a long time (from 2003 to 2008) insisted on extrapolating poly gate length to zero

• This corresponds to straightforward 0.7x scaling per generation

• Meanwhile, the industry kept a much slower, 0.9x scaling pace from 90nm to 30nm
Transistor Size Evolution: S/D Overlap

- Source/drain overlap has been quickly shrinking
Scaling of Leff (Junction-to-Junction)

90 nm node

- Lgate shrinks very slow
- S/D overlap shrinks fast
- Leff stays almost fixed

32 nm node
Transistor Size Evolution: L Effective

- Effective channel length did not change at all!

- Effective channel length is defined as distance between source and drain junctions

- It determines how far the electrons/holes have to go

- It determines carrier transport and transistor variability
Role of Stress Engineering in CMOS

- Classic scaling is dead
- Stress engineering is huge, bigger than HKMG
- At 32nm node, stress enhances hole mobility by 3.5x
- SiGe plays a key role in PMOS
- Si:C is used in NMOS, but is less efficient

K. Kuhn et al, ECS 2010 (Intel)
Transistor Size Evolution: Extrapolation?

- There is a sweet spot in terms of $L_{\text{eff}}$ that nobody wants to change
- This is not a “brick wall”, but the best $I_{\text{on}}/I_{\text{off}}$ ratio
- Shorter transistors with conventional planar bulk architecture are inferior
Transistor Size Evolution: Extrapolation?

- There is a sweet spot in terms of $L_{eff}$ that nobody wants to change.
- This is not a “brick wall”, but the best $I_{on}/I_{off}$ ratio.
- Shorter transistors with conventional planar bulk architecture are inferior.
- This trend can not continue, as there’s no space left for the contact.
Transistor Size Evolution: Gate Pitch

- Conventional planar MOSFET quickly runs out of space for contacts

200nm space for a contact at 90nm node

30nm space for a contact at 32nm node

No space left for contact

120nm for gate and spacers

80nm for gate and spacers

Technology node, nm

Gate length + 2 spacers

Gate length

Size, nm

© Synopsys 2011
Are Smaller Transistors Possible?

• Yes, here are several examples:
  

• But, their performance is inferior to the current MOSFET with $Leff = 25\text{nm}$
Why MOSFETs Dislike Scaling?

Good current is controlled by the gate (can be turned off)

Bad current is too far from the gate

High halo doping is used to suppress DIBL, but it degrades the on-current and increases BTBT leakage
Typical Drain Junction Leakage Map

- Band-to-band tunneling happens at the tip of drain extension, just outside of the gate.
- This is where high field overlaps with high halo doping.
What Can Be Done?

• Simple solution:
  – Keep the channel close to the gate
  – Remove current paths that are away from the gate

• This can be achieved in either:
  – FDSOI, or
  – FinFET
Transistor Size Evolution: Future

- At 20nm node, the trend can continue
- At 15nm node, switch to FinFETs or FDSOI is necessary
- FinFETs benefit from S/D underlap, not overlap
Transistor Size Evolution: ITRS 2009

- At 20nm node, the trend will continue
- At 15nm node, switch to FinFETs or FDSOI is necessary
- FinFETs benefit from S/D underlap, not overlap
- ITRS 2009 is in line with this vision (finally!)
Transistor Size Evolution: SRAM

- SRAM occupies large part of the chip
- Yet, it’s size lags 4 generations behind the logic!
- This is due to the:
  - Variability
  - Leakage
Scaling vs. Modeling Approaches

<table>
<thead>
<tr>
<th>Technology node, nm</th>
<th>Lgate, logic</th>
<th>S/D overlap</th>
<th>Leff, logic</th>
<th>Linear trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of lattice atoms along Leff</th>
<th>Number of lattice atoms in channel</th>
<th>Number of electrons/holes per switch</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>&gt;= 640,000</td>
<td>&gt;= 60</td>
</tr>
<tr>
<td>64</td>
<td>&gt;= 160,000</td>
<td>&gt;= 40</td>
</tr>
<tr>
<td>40</td>
<td>&gt;= 32,000</td>
<td>&gt;= 20</td>
</tr>
<tr>
<td>16</td>
<td>&gt;= 1,000</td>
<td>&gt;= 3</td>
</tr>
</tbody>
</table>
Scaling vs. Modeling Approaches

- Lgate, logic
- S/D overlap
- Leff, logic
- Linear trend

<table>
<thead>
<tr>
<th>Ballistics fraction</th>
<th>Adequate model</th>
</tr>
</thead>
<tbody>
<tr>
<td>~10%</td>
<td>Augmented drift-diffusion</td>
</tr>
<tr>
<td>~30%</td>
<td>Augmented drift-diffusion</td>
</tr>
<tr>
<td>&gt;30%</td>
<td>NEGF + scattering</td>
</tr>
<tr>
<td>&gt;50%</td>
<td>DFT</td>
</tr>
</tbody>
</table>
Discussion Topics

- Transistor scaling
- FDSOI
- SuVolta
- Stress engineering
- Non-silicon materials
- FinFET
- Summary
ETSOI aka UTSOI aka FDSOI
Current Is Always Within Gate’s Reach

Current density map

SOI thickness has to be ~5nm +/-1nm across 300mm wafer
20nm FDSOI MOSFET Stress Engineering

- FDSOI has very low stress transfer efficiency for both SiGe S/D and strained gate
- Only ~12% of stress is transferred to the channel
- Compare this to >50% stress transfer efficiency for bulk planar FETs & FinFETs

FDSOI similar to the one reported by IBM at IEDM 2009
FDSOI Summary

- Can be scaled better than planar MOSFETs
- Low off-state leakage – good for LP
- Similar layout style to planar MOSFETs
- Expensive
- No good stress engineering
- Can not compete with FinFET’s performance
Discussion Topics

- Transistor scaling
- FDSOI
- SuVolta
- Stress engineering
- Non-silicon materials
- FinFET
- Summary
• Keep the same masks, same libraries as in standard CMOS
• Less Vt variability, therefore possible to reduce Vdd and power consumption
“SuVolta Claims Half the Power for Mobile SoCs”

- Keep all stress engineering tricks
- Add a couple epi steps
- In a way, it mimics FDSOI
- Can be scaled down to 15nm
Discussion Topics

- Transistor scaling
- FDSOI
- SuVolta
- **Stress engineering**
- Non-silicon materials
- FinFET
- Summary
Stress Engineering

• A must for Si
• Perhaps not necessary for high-\(\mu\) materials
• No good way to do it in SOI
• No SMT for FinFETs and nano-wires
• For gate-last HKMG:
  – CESL is useless
  – Main stress source is strained elevated S/D
Stress Source Evolution

- CESL
- NMOS SMT
- NMOS eSiGe
- Gate-Last
- Gate-Last
- Gate-Last
- NMOS eSi:C

90 nm, 65 nm, 45 nm, 32 nm, 22 nm
{111} Stacking Faults Due to SMT

Intel 32nm NMOSFET

- During SPER with restricted boundary, amorphized Si S/D gets dislocations with {111} stacking faults
- Each stacking fault is a missing {111} plane (i.e. vacancies)
- The missing {111} plane creates tensile stress in the direction perpendicular to the plane
Stress Induced by {111} Stacking Fault

- Missing (i.e. vacancy) {111} plane due to SMT for a 30nm deep amorphized Si
- ~350 MPa tensile longitudinal stress is present in the channel
- Each pair of stacking faults increases electron mobility by 5%
It is a Bad Idea to Amorphize FinFETs

Discussion Topics

- Transistor scaling
- FDSOI
- SuVolta
- Stress engineering
  - Non-silicon materials
- FinFET
- Summary
Three Major Issues for Non-Si Channels

Can not scale $L_{eff} < 15 \text{nm}$

Can not scale $W < 4 \text{nm}$

Huge interface traps

Planar MOSFET modeled with Schroedinger eq’n L. Smith et al, MRS 2007

Opportunity Window for Non-Si Channel

- Any new technology has to last at least 2 nodes
- SiGe channel is easier to manufacture
- High Ge content SiGe or pure Ge or GeSn to follow
- III-V materials have a narrow opportunity window
Discussion Topics

- Transistor scaling
- FDSOI
- SuVolta
- Stress engineering
- Non-silicon materials
- **FinFET**
- Summary
FinFET Analysis

- Parasitic R & C
- Conformal doping
- Stress engineering
- Patterning
- Proximity effects
- Leakage
- Collapse
- Variability

{111} facets
Gate 1
Gate 2
Drain

© Synopsys 2011
Thin-Layer Mobility for (100) & (110) Si

Electrons

Holes

Severe mobility degradation happens below ~3 nm

Experimental Data
- e⁻ µ for (100) – K. Uchida, IEDM2002 pp47-50
- h⁺ µ for (100) – K. Uchida, IEDM2002 pp47-50
- e⁻ µ for (110) – T. Hiramoto, IEDM2005 pp729-732
- h⁺ µ for (110) – T. Hiramoto, EDL2005 pp836-838
10nm FinFET: Improves As Fin Narrows

- This means that it really pays off to narrow the fin
- To narrow the fin, two STI depths are required

**Assumptions:**
- $L=10\text{nm}$, $L_{\text{eff}}=16\text{nm}$, EOT=1nm, Fin pitch=3*fin width, $V_{dd}=0.8V$, Undoped bulk fin, $I_{off}=1\text{nA/\mu m}$, Fin height=30nm
Orientation Effects

• Auto-Orientation is a must
• High-order planes?
• Rough side surfaces?
Orientation Effects

- Auto-Orientation is a must
- High-order planes?
- Rough side surfaces?
Planar MOSFETs are not Planar Anyway!

Even in a 90nm planar MOSFET, the channel surface is not planar, but everybody models it as a flat (100)
Variability

• Key for scaling power consumption
• Currently dominated by RDF
• For undoped FinFETs and nano-wires variability will be dominated by CD and LER
Non-Uniform Fin Stress Patterns

-1 GPa

-5 GPa

Zero stress

Huge stress variations, especially at the S/D junctions

To get average fin stress of ~2 GPa, peak stress in the fin exceeds dangerous 5 GPa
Patterning

• Double vias are not necessary any more, so:
  – No corners/jogs
  – All lines on all levels can be straight lines

• Patterning can be done with double and quadruple spacer litho

• LER is an increasing issue

• \{111\} epi and etch facets can help to extend patterning to the end of roadmap
Larger Scale: Time evolution.
Larger Scale: Time evolution.
Larger Scale: Time evolution.
Microscopic Surface Roughness

- {100} planes are the roughest
- {110} planes are smoother
- {111} planes are the smoothest

{111} nano-islands

{100} plane

{110} plane

{111} plane
FinFET Conclusions

- Requires changes in process & design
- Very sensitive to geometry
- Requires spacer lithography
- Insensitive to random dopant fluctuations
- Provides high performance
- Easy for stress engineering
- Enables scaling to 15nm and 10nm
- Will likely become mainstream architecture
STT-RAM: DRAM Replacement?

- Major players consider it for DRAM
- Good potential for “universal memory”
- Can be used for low-power logic
  - Requires different circuits
  - Driven by I, not V
- CoFeB demonstrated working down to 5nm islands
Summary

• Conventional MOSFETs can not be scaled beyond 20nm without performance degradation
• FDSOI is incompatible with stress engineering
• Stress engineering boosts driving current by 2x
• FinFETs will likely become mainstream transistor architecture