


# ECE606: Solid State Devices

## Lecture 25

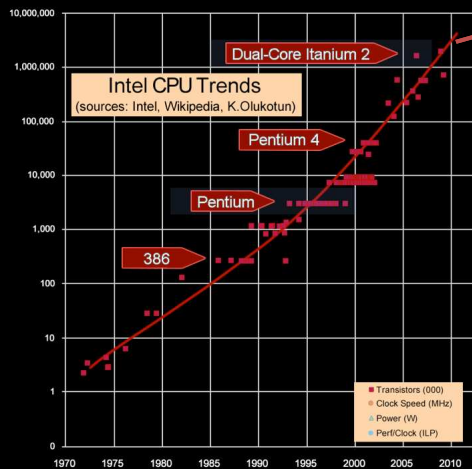
### Modern MOSFETs

Gerhard Klimeck  
[gekco@purdue.edu](mailto:gekco@purdue.edu)



	Equilibrium	DC	Small signal	Large Signal	Circuits
Diode					
Schottky					
BJT/HBT					
<b>MOSFET</b>					

# Moore's Law Forever?

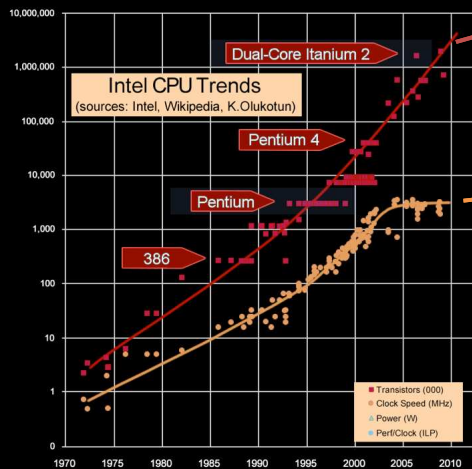


Number of transistors:  
Moore's law is continuing

<http://jai-on-asp.blogspot.com>

2005: free lunch is over, updated 2009

# CPU's are not getting faster!



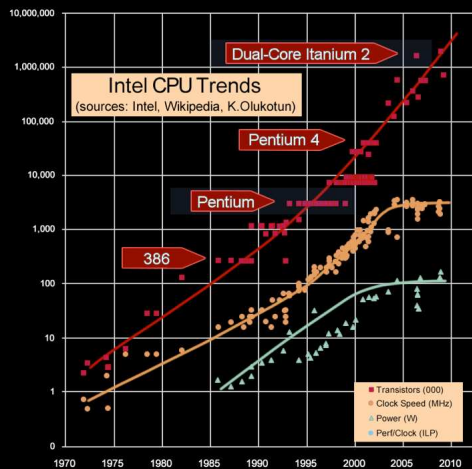
Number of transistors:  
Moore's law is continuing

Clock speed:  
no longer scaling

<http://jai-on-asp.blogspot.com>

2005: free lunch is over, updated 2009

# Power is the Limit!



Number of transistors:  
Moore's law is continuing

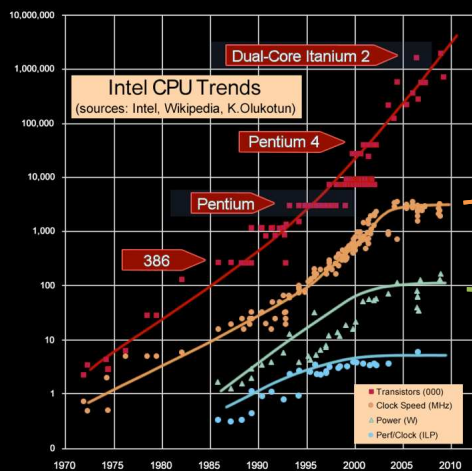
Clock speed:  
no longer scaling

Power:  
today's limitation  
~100W

<http://jai-on-asp.blogspot.com>

2005: free lunch is over, updated 2009

# Limited Performance Improvements



Number of transistors:  
Moore's law is continuing

Clock speed:  
no longer scaling

Power:  
today's limitation  
~100W

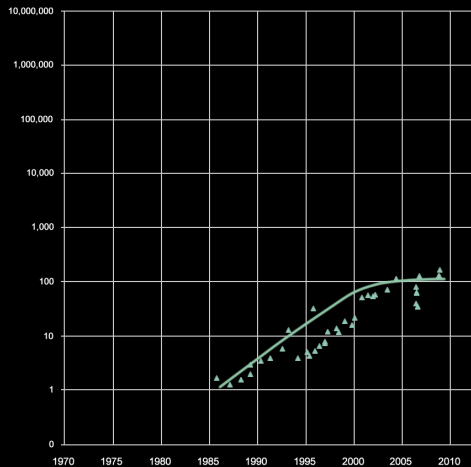
Performance Gain:  
**limited**

Supply voltage ( $V_{dd}$ ) stopped  
scaling at around 2003

<http://jai-on-asp.blogspot.com>

2005: free lunch is over, updated 2009

# What is Special about 100W ?



Power:  
today's limitation  
~100W

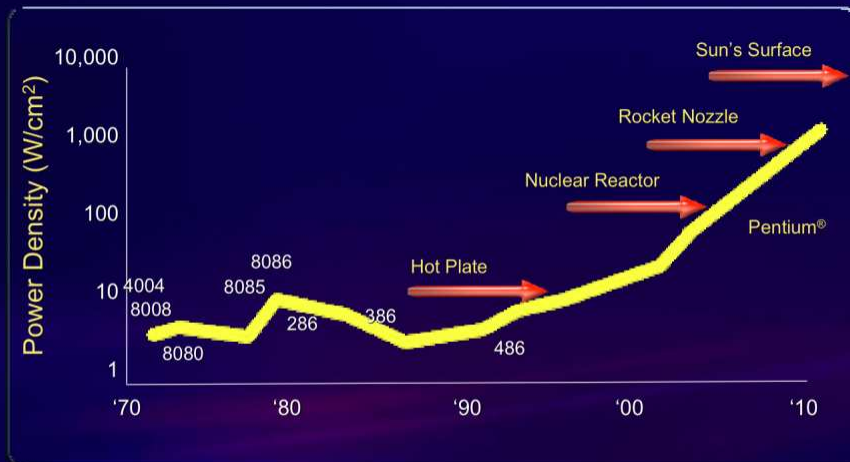
<http://jai-on-asp.blogspot.com>

2005: free lunch is over, updated 2009

# Intel Projection from 2004

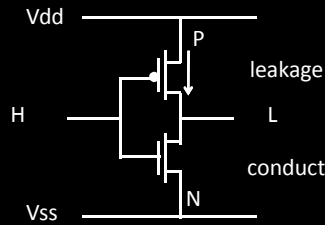
## How do we burn power?

### Switching Circuits & Leakage => Transistors



Intel Developer Forum, Spring 2004 - Pat Gelsinger

# CMOS Inverter



Dynamic / Switching Power:

➤ Charging a capacitor network

$$P \propto f C_L V_{dd}^2$$

➤ Reduce frequency ☹️

➤ Reduce capacitance

=> device size ☺️

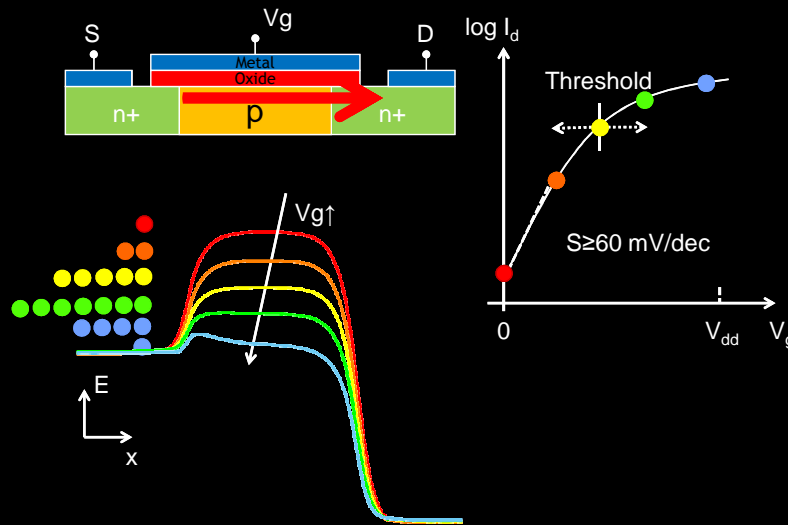
➤ Reduce voltage ☺️

Static Power:

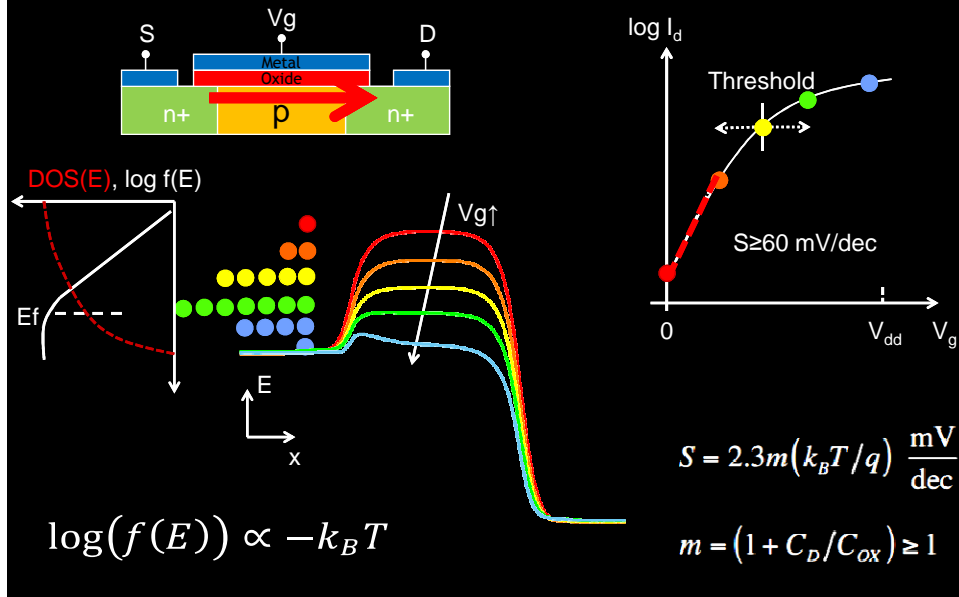
➤ Leakage through transistors

$$P \propto I_{OFF} \propto 1/\exp(V_{dd}) \text{ ☹️}$$

# “Fundamental” Limit



# “Fundamental” Limit



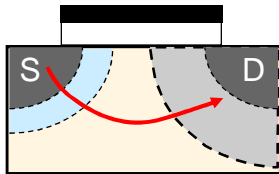
**1. Short channel effect**

2. Control of threshold voltage  $I_D = \frac{\mu C_{ox}}{L_{ch}} (V_G - V_{th}^*)^2$

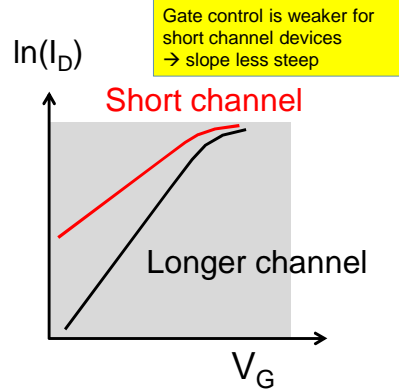
3. Mobility enhancement

4. Conclusion

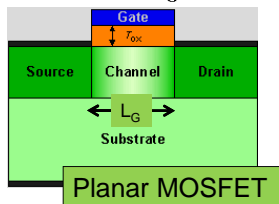
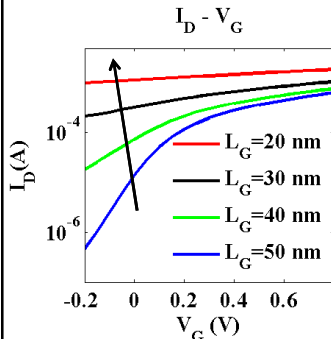
REF: Chapter 19, SDF



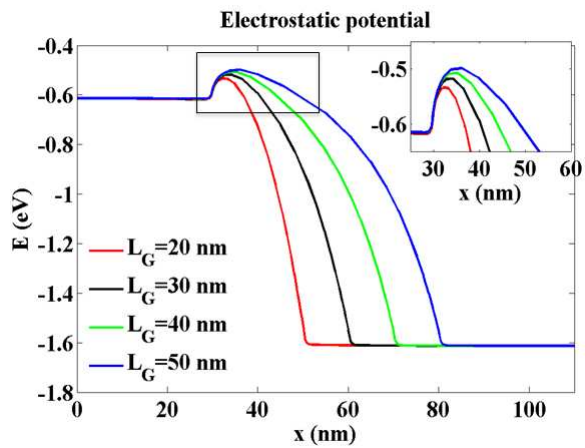
Large Drain bias can lead to "punch-through"



Recall similar problem with bipolar transistor

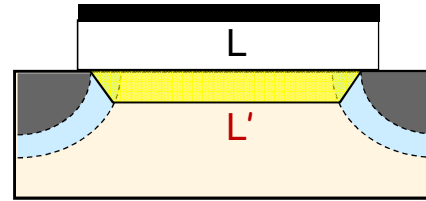
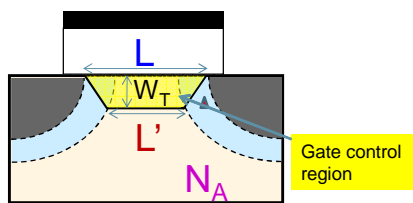
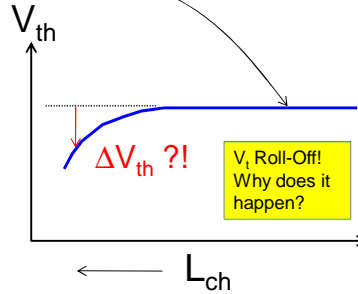
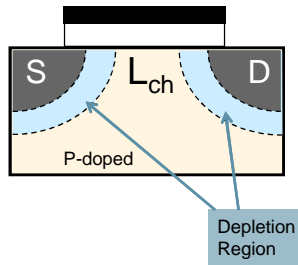


Planar MOSFET



- Short channel effects: **drain induced barrier lowering DIBL**

$$V_{th} = 2\phi_F - \frac{Q_B}{C_{ox}} = 2\phi_F + \frac{qN_A W_T}{C_{ox}}$$



$$V_{th,Short} = 2\phi_F - \frac{Q_{B,Short}}{C_{ox}}$$

$$Q_{B,Short} = \frac{-qN_A \times Z \times W_T \left( \frac{L+L'}{2} \right)}{Z \times L}$$

Area of gate control region

$$= -qN_A W_T \left( \frac{L+L'}{2L} \right)$$

$$V_{th,L} = 2\phi_F - \frac{Q_{B,Long}}{C_{ox}}$$

$$Q_{B,Long} \rightarrow -qN_A W_T \quad (L \cong L')$$

For long channel

$$\Delta V_{th} = -\frac{Q_{B,Long}}{C_{ox}} + \frac{Q_{B,Short}}{C_{ox}}$$

$$= \frac{-qN_A W_T}{C_{ox}} \left[ 1 - \frac{L'+L}{2L} \right]$$



Calculate  $L'$  (geometry exercise)

$$(r_j + W_s)^2 = W_T^2 + \left( r_j + \frac{L - L'}{2} \right)^2$$

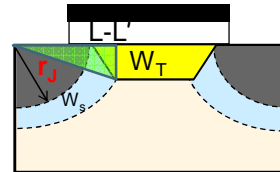
$$L' = L - 2r_j \left( \sqrt{1 + \frac{2W_T}{r_j}} - 1 \right)$$

Minimize  $L'$

$$\Delta V_{th} = \frac{-qN_A W_T}{C_{ox}} \left[ 1 - \frac{L' + L}{2L} \right]$$

$$= \frac{-qN_A W_T}{C_{ox}} \frac{r_j}{L} \left( \sqrt{1 + \frac{2W_T}{r_j}} - 1 \right) = \alpha_0 \text{ Minimum acceptable ...}$$

Minimize  $r_j$



$r_j$  junction curvature

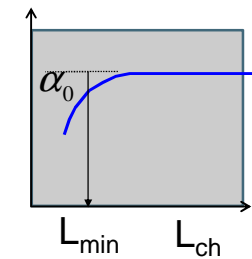


**Shallow junction/geometry of transistors**

laser annealing of junctions, FINFETs

**Reduced substrate doping  $N_A$**   
consider  $W_T$  and junction breakdown

$V_{th}$



$$L_{min} = \frac{qN_A W_T}{C_{ox}} \frac{r_j}{\alpha_0} \left( \sqrt{1 + \frac{2W_T}{r_j}} - 1 \right)$$

**Thinner gate oxides**

Consider tunneling current

**Higher gate dielectric**

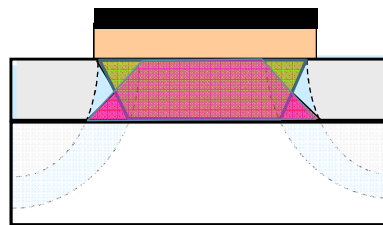
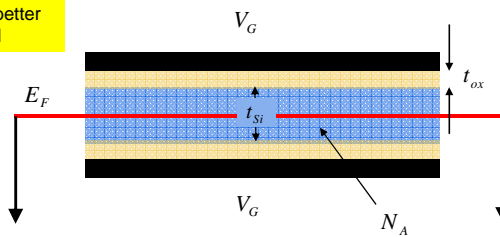
Consider bulk traps

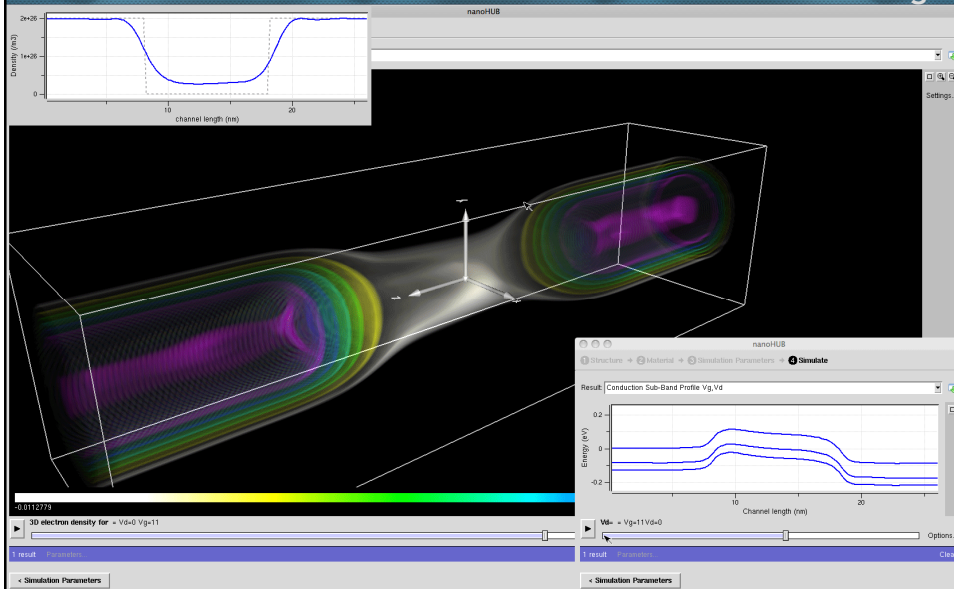
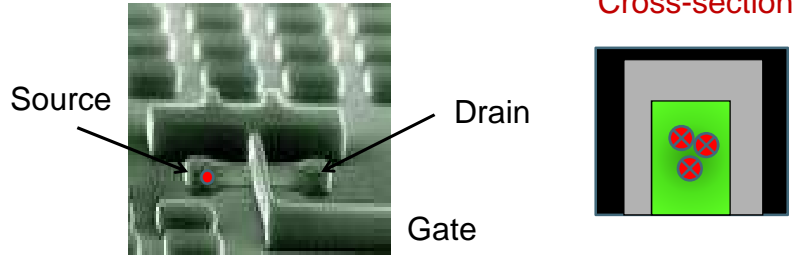


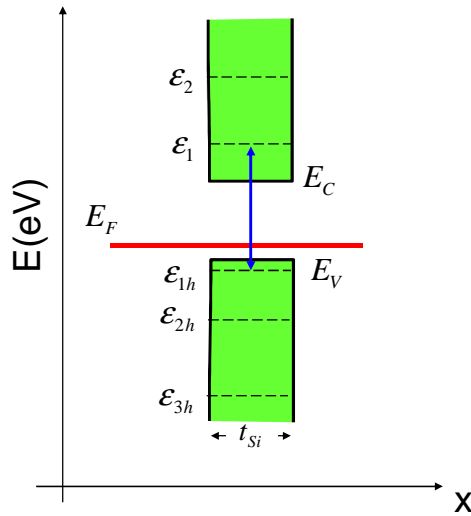
1. Short channel effect
- 2. Control of threshold voltage**
3. Mobility enhancement
4. Conclusion

Two Gates → better channel control

SOI: Silicon On Insulator



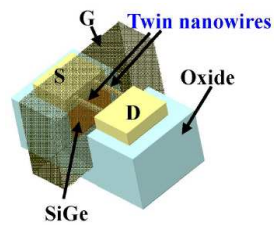




$$\epsilon_n = \frac{\hbar^2 n^2 \pi^2}{2m^* t_{Si}^2}$$

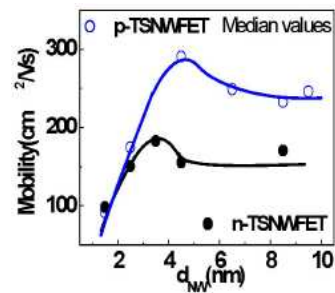
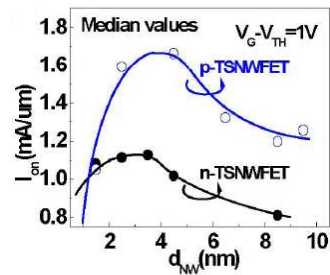
$$E'_G = E_G + \epsilon_1 + \epsilon_{1h}$$

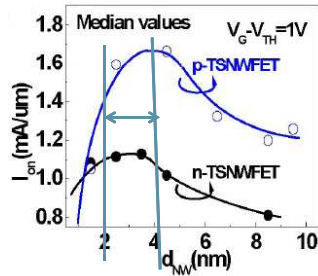
Band-gap widening  
Fluctuation in thickness



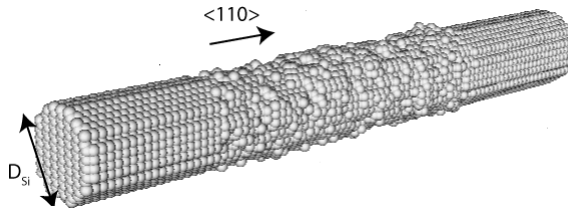
[1] S. D. Suk, IEDM, 2007

- Significant reduction of ON-current/mobility in NW with diameter less than 3 nm.
- What causes this?

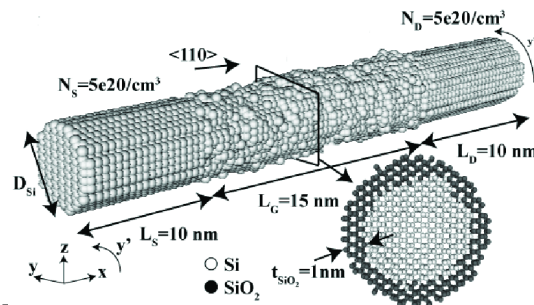




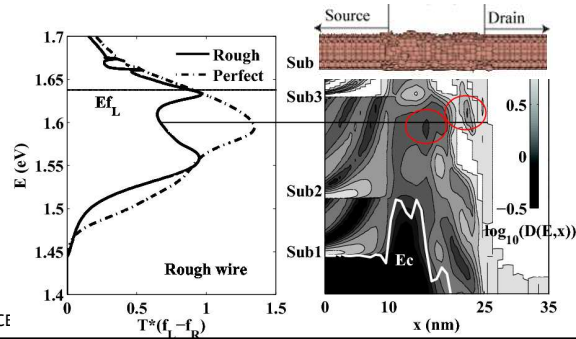
- Bandstructure: tight-binding/full-band
  - 3D-atomistic interface roughness
  - SiO<sub>2</sub> included in transport
  - Low/High-drain bias
- ➔ Comparison with experiments



- Realistic/atomistic rough interface between Si/SiO<sub>2</sub>
  - » Adapt experimentally generated statistical function: How do we know the generated interface roughness is correct?
  - » Many statistical samples needed: computational cost high
  - » Electrons may penetrate into the SiO<sub>2</sub> region. How do we count it?
  - » Computational cost increases as we take into account SiO<sub>2</sub> in tight-binding approximation.

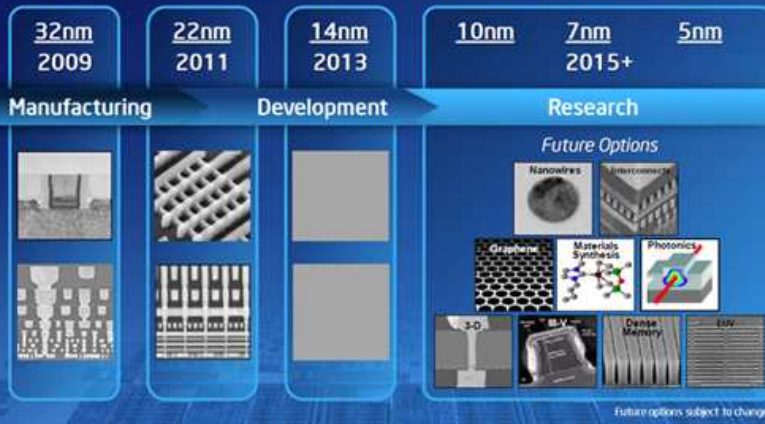


- To correctly include scattering mechanisms (interface roughness/phonon scattering)  $\langle 110 \rangle$  NW
  - » Beyond the effective mass approximation, full band tight-binding simulation is needed: computational cost is higher
    - ✓ Effective mass cannot understand non-parabolicity/anisotropy in the bandstructure of  $\langle 110 \rangle$  oriented NW
    - ✓ At high drain/gate bias drain side of the channel, higher subbands are mixing and influence the scattering



## Intel Roadmap

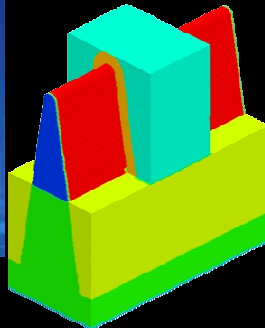
### Innovation Enabled Technology Pipeline Our Visibility Continues to Go Out ~10 Years



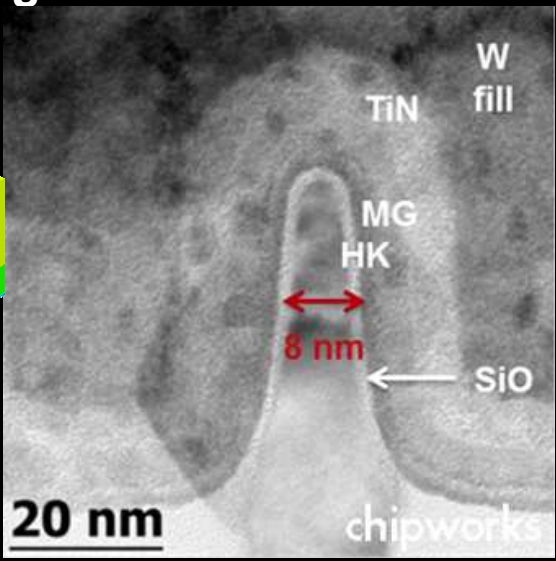


**22nm 2011**

**Today: non-planar 3D devices  
Better gate control!**



**Intel 22nm finFET**



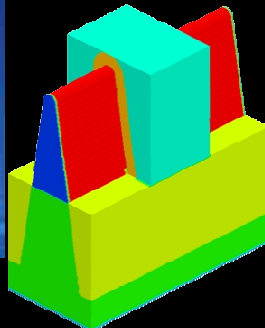
W fill  
TiN  
MG  
HK  
8 nm  
SiO

**20 nm** chipworks

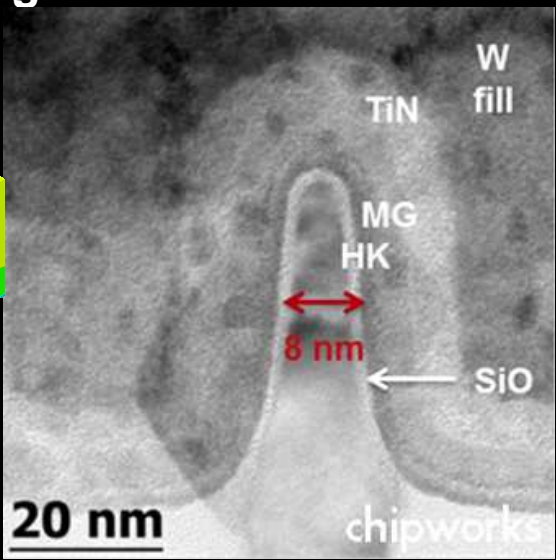
[http://www.goldstandardsimulations.com/index.php/news/blog\\_search/simulation-analysis-of-the-intel-22nm-finfet/](http://www.goldstandardsimulations.com/index.php/news/blog_search/simulation-analysis-of-the-intel-22nm-finfet/)  
<http://www.chipworks.com/media/wpmu/uploads/blogs.dir/2/files/2012/08/Intel22nmPMOSfin.jpg>

**22nm 2011**

**Today: non-planar 3D devices  
Better gate control!**



**22nm = 176 atoms**  
**8nm = 64 atoms**



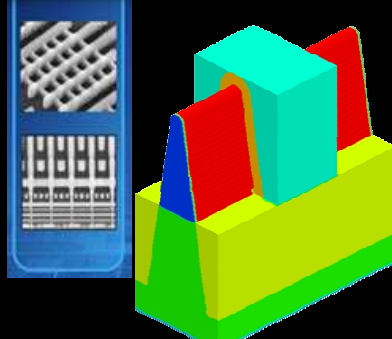
W fill  
TiN  
MG  
HK  
8 nm  
SiO

**20 nm** chipworks

[http://www.goldstandardsimulations.com/index.php/news/blog\\_search/simulation-analysis-of-the-intel-22nm-finfet/](http://www.goldstandardsimulations.com/index.php/news/blog_search/simulation-analysis-of-the-intel-22nm-finfet/)  
<http://www.chipworks.com/media/wpmu/uploads/blogs.dir/2/files/2012/08/Intel22nmPMOSfin.jpg>

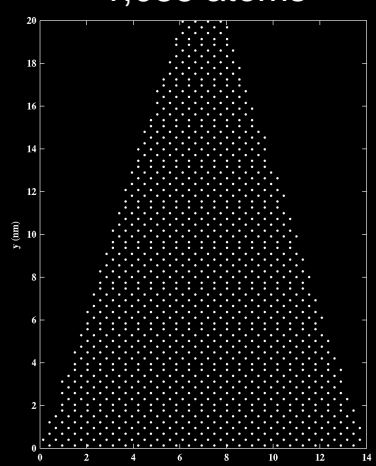
## Today: non-planar 3D devices Better gate control!

**22nm**  
2011



22nm = 176 atoms  
8nm = 64 atoms

1,085 atoms




[http://www.goldstandardsimulations.com/index.php/news/blog\\_search/simulation-analysis-of-the-intel-22nm-finfet/](http://www.goldstandardsimulations.com/index.php/news/blog_search/simulation-analysis-of-the-intel-22nm-finfet/)  
<http://www.chipworks.com/media/wpmu/uploads/blogs.dir/2/files/2012/08/Intel22nmPMOSfin.jpg>


## Roadmap of finite atoms!

**Atomistic Modeling**  
=>  
**NEMO**


**32nm**  
2009




**22nm**  
2011



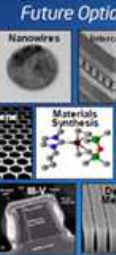
**14nm**




**10nm**



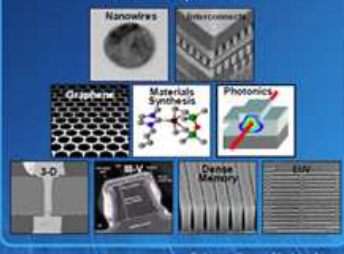
**7nm**



**5nm**



**Future Options**



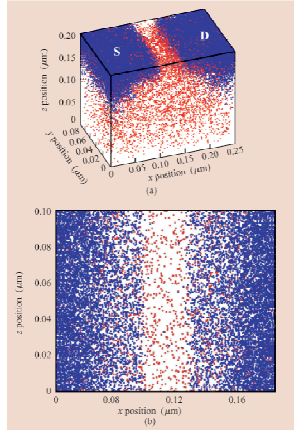
Future options subject to change

nm Node	22	14	10	7	5
Node atoms	176	122	80	56	40
Critical atoms	64	44 <sup>(?)</sup>	29 <sup>(?)</sup>	20 <sup>(?)</sup>	14 <sup>(?)</sup>
Electrons	160-190	64-80	30-38	18-23	11-15



$$V_{th} = 2\phi_F - \frac{Q_B}{C_{ox}} = 2\phi_F - \frac{qN_D W_T}{C_{ox}}$$

$$\sigma_{V_T} = 3.19 \times 10^{-8} \left( \frac{t_{ox} N_A^{0.4}}{\sqrt{L_{eff} W_{eff}}} [V] \right),$$



Variation of V<sub>T</sub> in short channel devices

Stronger effect of dopant number fluctuations on V<sub>T</sub>

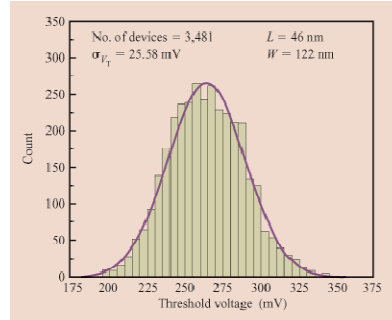


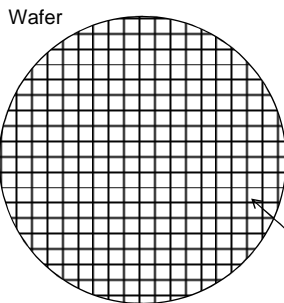
Figure 2

Threshold voltage histogram for FETs in the 90-nm-technology node.

IBM Journal of Res. And Tech. 2003.

$$V_{th} = 2\phi_F - \frac{Q_B}{C_{ox}} = 2\phi_F - \frac{qN_A W_T}{C_{ox}}$$

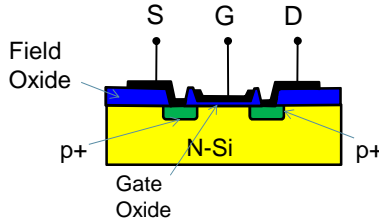
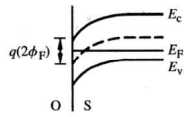
$$\sigma_{V_T} = 3.19 \times 10^{-8} \left( \frac{t_{ox} N_A^{0.4}}{\sqrt{L_{eff} W_{eff}}} [V] \right),$$



If every transistor has different V<sub>th</sub> and therefore different current, circuit design becomes difficult

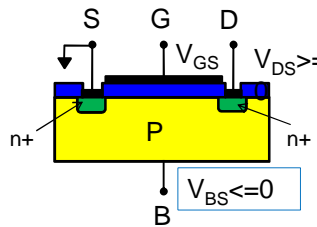
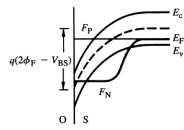
$$I_D = \frac{\mu C_{ox}}{L_{ch}} (V_G - V_{th}^*)^2$$

$$V_{th} = \psi_s - \frac{Q_B}{C_{ox}} = \psi_s + B\sqrt{\psi_s}$$



$$\psi_s = 2\phi_F$$

Control channel inversion voltage through back gate bias

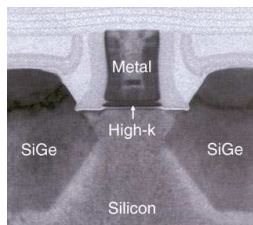


$$\psi_s = 2\phi_F - V_{BS}$$

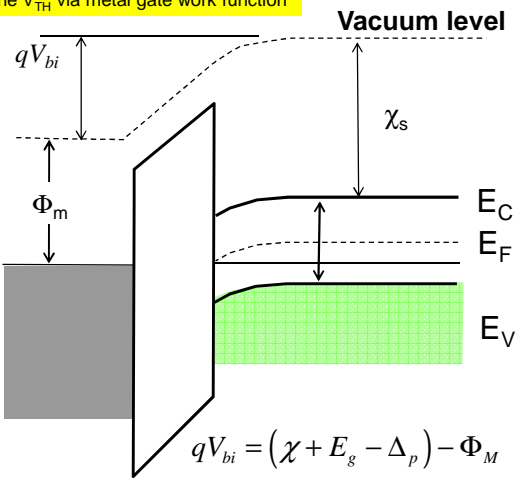
Tune  $V_T$  through back gate bias  $V_{BS}$



High-k/metal gate MOSFET



Tune  $V_{Th}$  via metal gate work function



$$Q_i = C_{ox}(V_G - V_{th})$$

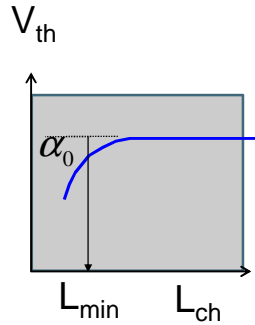
$$V_{th} = -V_{FB} + \psi_s - \frac{Q_B}{C_{ox}}$$

$$qV_{bi} = (\chi + E_g - \Delta_p) - \Phi_M = qV_{FB}$$



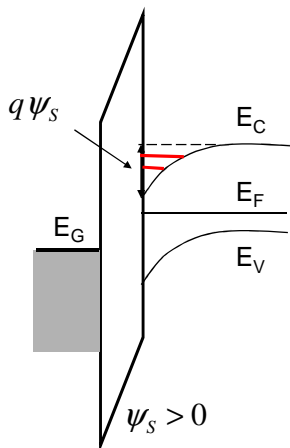
**Shallow junction/geometry of transistors**  
laser annealing of junctions, FINFETs

**Reduced substrate doping NA**  
consider WT and junction breakdown



$$L_{min} = \frac{qN_A W_T r_J}{C_{ox} \alpha_0} \left( \sqrt{1 + \frac{2W_T}{r_J}} - 1 \right)$$

**Thinner gate oxides**  
Consider tunneling current  
**Higher gate dielectric**  
Consider bulk traps



$$\frac{d^2\psi}{dx^2} = \frac{-q}{\epsilon} \left[ p(x) - n(x) |\psi(x)|^2 + N_D^+ - N_A^- \right]$$

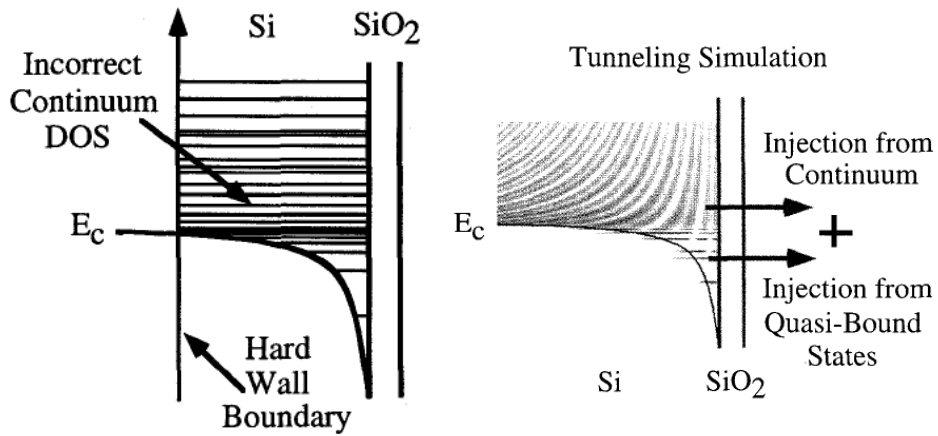
↓  
wavefunction, not potential !

Wave function should be accounted for

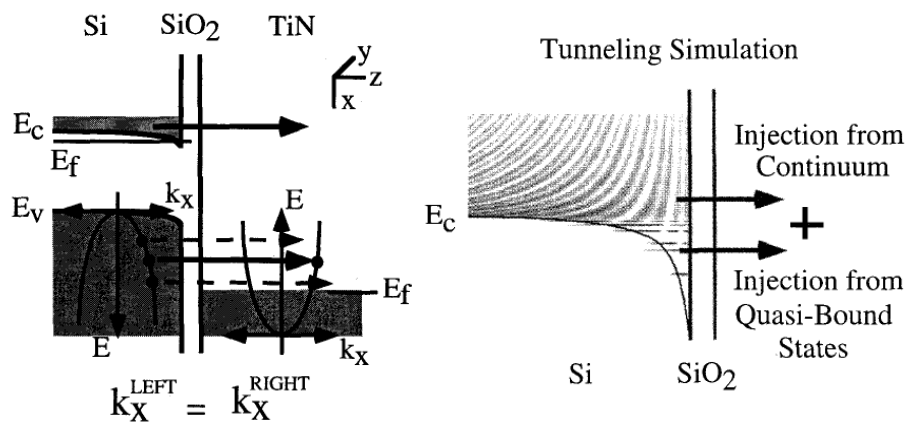
Bandgap widening near the interface must also be accounted for.

Assumption of nondegeneracy may not always be valid

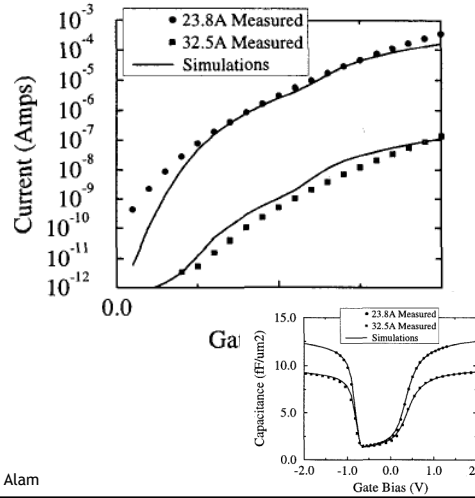
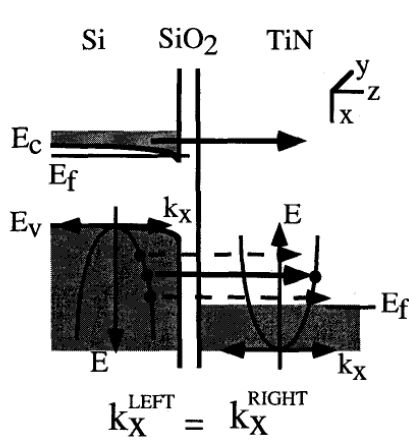
R. Bowen, Chenjing Fernando, Gerhard Klimeck, Amitava Chatterjee, Daniel Blanks, Roger Lake, J. Hu, Joseph Davis, M. Kularni, Sunil Hattangady, I.C. Chen,  
**"Physical Oxide Extraction and Versification using Quantum Mechanical Simulation"**  
 Proceedings of IEDM 1997, IEEE, 869 (1997); doi : 10.1109/IEDM.1997.650518, Cited by 42



R. Bowen, Chenjing Fernando, Gerhard Klimeck, Amitava Chatterjee, Daniel Blanks, Roger Lake, J. Hu, Joseph Davis, M. Kularni, Sunil Hattangady, I.C. Chen,  
**"Physical Oxide Extraction and Versification using Quantum Mechanical Simulation"**  
 Proceedings of IEDM 1997, IEEE, 869 (1997); doi : 10.1109/IEDM.1997.650518, Cited by 42

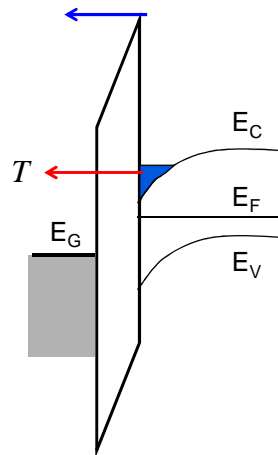
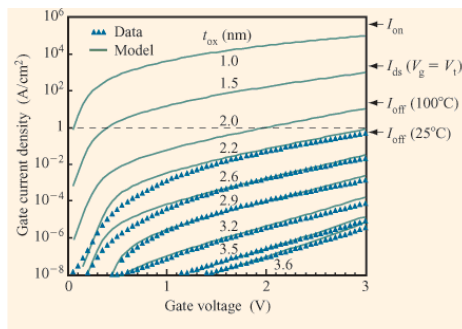


R. Bowen, Chenjing Fernando, Gerhard Klimeck, Amitava Chatterjee, Daniel Blanks, Roger Lake, J. Hu, Joseph Davis, M. Kularni, Sunil Hattangady, I.C. Chen,  
**"Physical Oxide Extraction and Versification using Quantum Mechanical Simulation"**  
 Proceedings of IEDM 1997, IEEE, 869 (1997); doi : 10.1109/IEDM.1997.650518, Cited by 42



Don't make oxides too thin → tunneling!

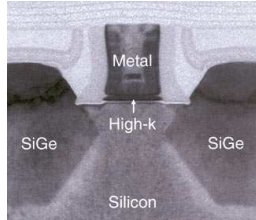
$$J_T = \left[ Q_i(V_G) - \frac{n_i^2}{N_A} e^{-qV_G/\beta} \right] v_{th} \langle T(E) \rangle$$



$$L_{\min} = \frac{qN_A W_T r_J}{\frac{\kappa_{ox} \epsilon_0}{x_0} \alpha} \left( \sqrt{1 + \frac{2W_T}{r_J}} - 1 \right)$$

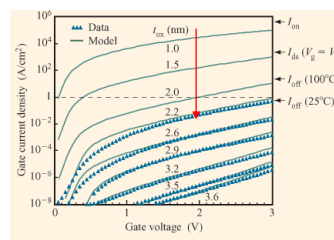
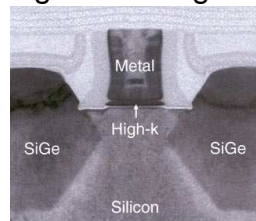
High k oxides allow for smaller  $L_{\min}$  but lot of defects

### High-k/metal gate MOSFET



- Shallow junction and geometry of transistors laser annealing of junctions, FINFET
- Substrate doping NA consider  $W_T$  and junction breakdown
- Thinner gate oxides consider tunneling current
- **Higher gate dielectric** consider bulk traps

### High-k/metal gate MOSFET



$$L_c = \frac{qN_A W_T r_J}{\frac{\kappa_{ox} \epsilon_0}{x_0} \alpha} \left( \sqrt{1 + \frac{2W_T}{r_J}} - 1 \right)$$

$$I_D = \frac{\mu C_{ox}}{L_{ch}} (V_G - V_{th}^*)^2$$

Thicker oxide ( $x_0$ ) for same capacitance ...

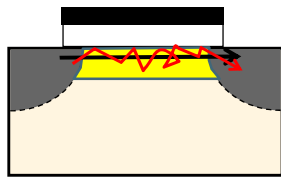
... ensures the drive-current is not reduced  
, but tunneling current is suppressed.



1. Short channel effect
2. Control of threshold voltage
3. **Mobility enhancement**
4. Conclusion

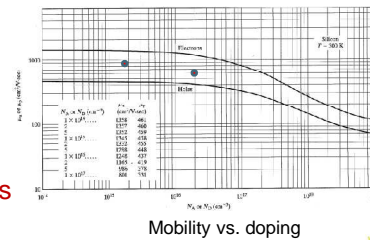
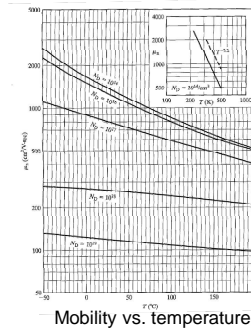
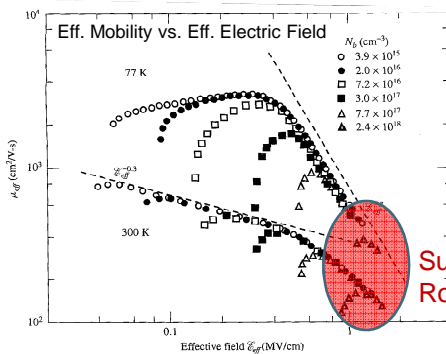
$$I_D = \frac{\mu C_{ox}}{L_{ch}} (V_G - V_{th}^*)^2$$

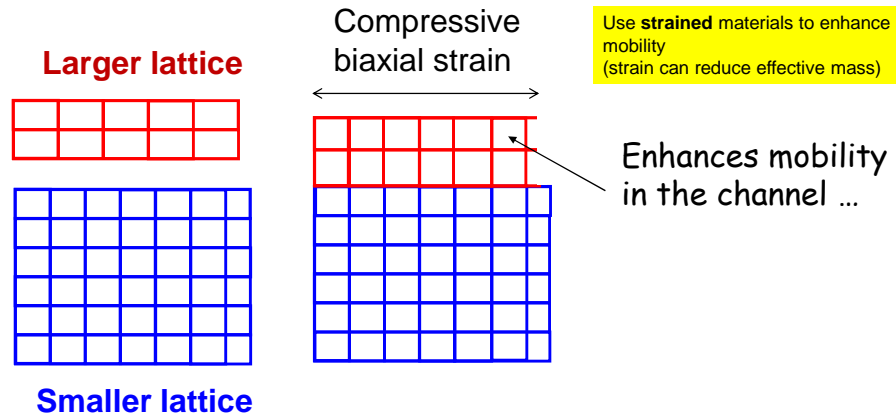
Few words about universal Mobility ...



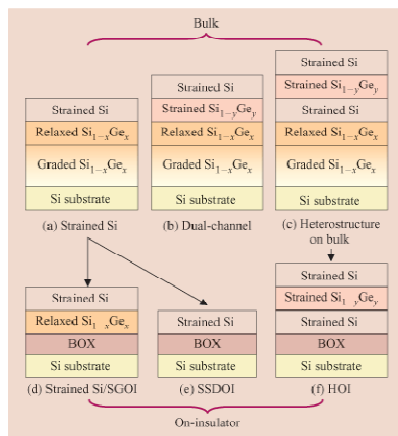
Gate E-field → electrons pulled to surface (surface scattering)

$$\mu_{eff} = \frac{\mu_0}{1 + \theta(V_G - V_{th})}$$

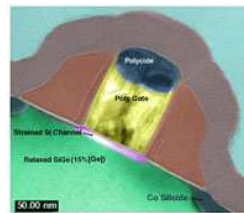
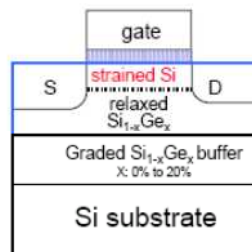




Examples of strained Si structures

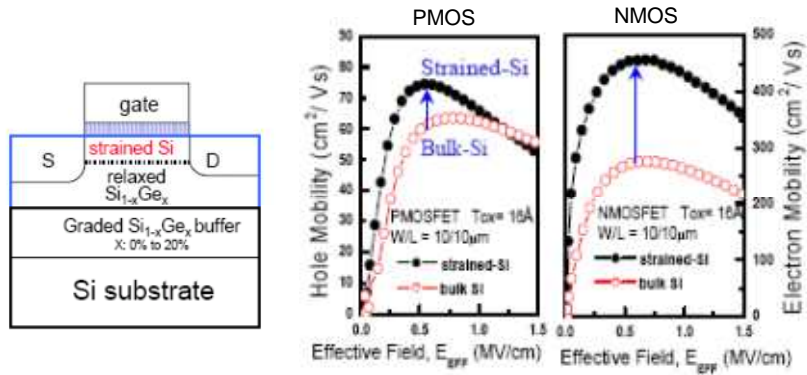


Experimental Device using strained Si

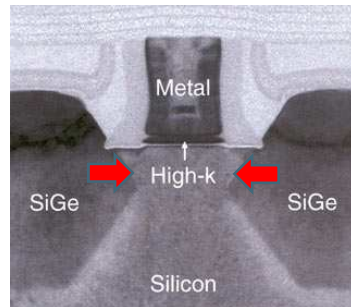




Experimental Device using strained Si

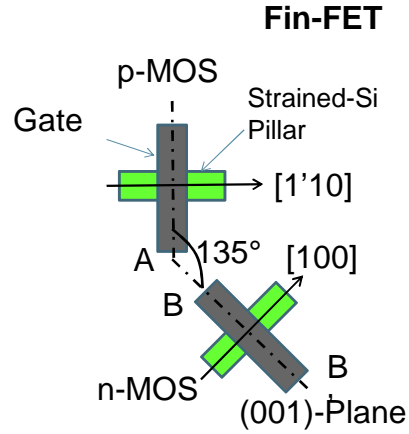
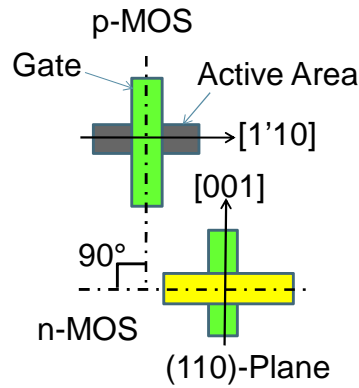


Adapted from Chang et. al, IEDM 2005.



Uniaxial strain from source and drain

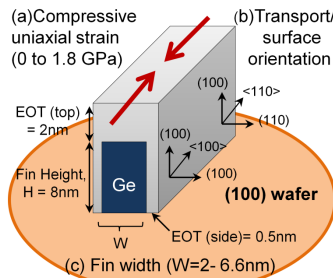
Try different crystal directions to maximize mobility



Takagi, TED 52, p.367, 2005

**Objective:**

- Identifying performance boosters in Ge FinFETs for hole transport.

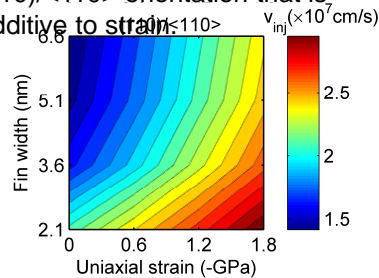


**Approach**

- Self-consistent Top of the barrier model used (NEMO5).
- Computed  $v_{inj}$  for different (a) Compressive strain, (b) Orientation and (c) Fin width scaling at a constant inversion charge ( $1e13/cm^2$ ).

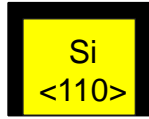
**Results**

- Similar enhancement in  $v_{inj}$  compared to mobility52 →  $v_{inj}$  a good metric for performance.
- (110)/<110> shows ~2.5X enhancement over (100)/<100>** for all strain/Fin width cases.
- Fin width scaling** identified as a performance booster for (110)/<110> orientation that is additive to strain.

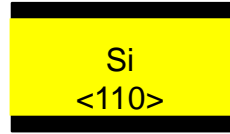




Gate all around



Tri-Gate



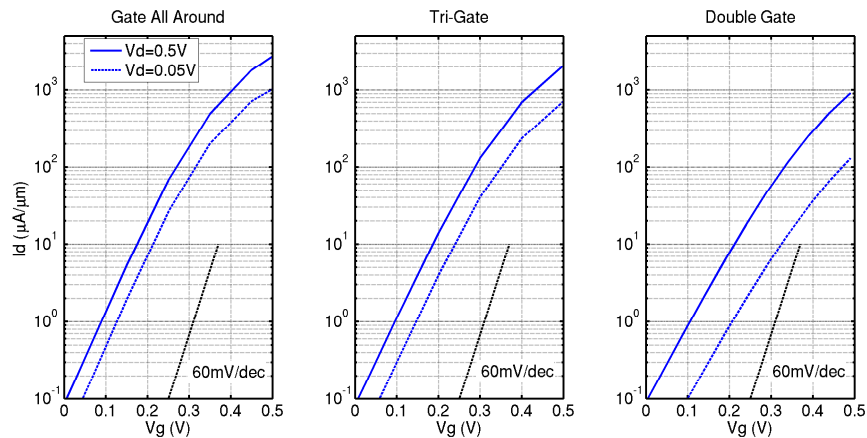
Double Gate

Si <110> 2GPa uniaxial tensile strain 3x3 nm<sup>2</sup>

$m^* = 0.175 m_0$  from TB bandstructure

EOT = 0.47 nm,  $V_{DD} = 0.5$  V

Simulations done by Klimeck group with NEMO5

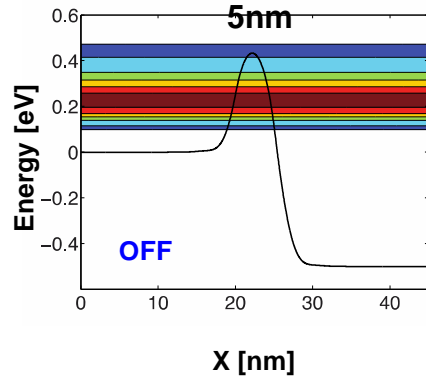


	$I_{ON}$ ( $\mu A/\mu m$ )	SS(mV/dec)	DIBL(mV/V)
Double Gate	~900	101.5	200
Tri-Gate	~2000 (norm. with height)	91.3	120
Gate-All-Around	~3000 (norm. with height)	84.1	90



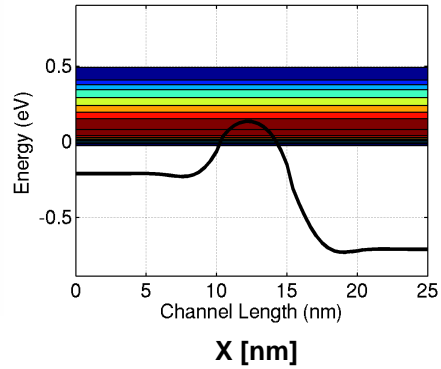
## results: SD tunneling in 5 nm MOSFETs

Double Gate  $L =$



Tunneling Current =  
96%

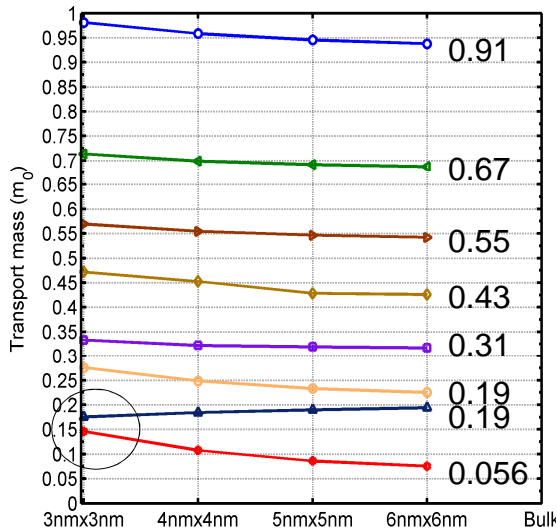
GAA = 5nm



Tunneling Current =  
45%



## results: eff mass engineering



- Si <100>
- Si <112>
- ▲ Si <110> 2 GPa tensile
- In<sub>0.75</sub>Ga<sub>0.25</sub>As <100>
- ◆ Si <111>
- ▲ Si <110> 2GPa compressive
- ▲ Si <112> 2GPa compressive
- Si <100> 2GPa compressive

### “Effective masses”

- DOS eff. mass  $D(E)$
- DOM eff. mass  $M(E)$
- conductivity eff. mass
- confinement eff mass

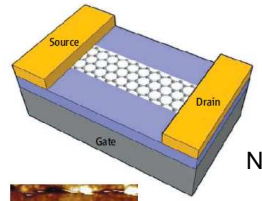
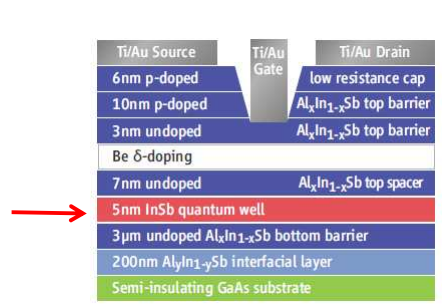


Speed of Charges in Different Materials ( $\text{cm}^2/\text{V}\cdot\text{s}$ )					
Charges	Si	GaAs	$\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$	InAs	InSb
Electrons*	300	7000	10,000	15,000	30,000
Holes*	450	400	200	460	1250

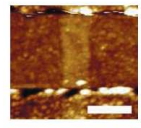
  

Ge
*
1900

\*Electron carrier mobilities measured in transistor channels with electron concentration of  $1 \times 10^{12} \text{ cm}^{-2}$ . Hole mobilities in bulk.



Nature, 2009



New kid. Transistors made from graphene nanoribbons could be blinding fast. But can they perform on an industrial scale?

Ge in PNP transistors, bandgap too small, but now coming back for PMOS

- 1) Short channel effect is a serious concern for MOSFET scaling.
- 2) Many novel solutions at the material, device, circuit level have been proposed to reduce short channel effect.
- 3) The success of these efforts are now reflected in effective MOSFET channel lengths of 30 nm.