Stretching the Barriers – An analysis of MOSFET Scaling

Presenters (in order) –
Zeinab Mousavi
Stephanie Teich-McGoldrick
Aseem Jain
Jaspreet Wadhwa

Why Small?

- Higher Current
- Lower Gate Capacitance
- Higher Speed
- Cost Efficient

Notes for the first set of slides are at the end of the presentation

Shrinking Methods

1. Constant Voltage Scaling
   - reduce later/vertical sizes by α
   - increase all doping by α

Disadvantages:

- mobility degradation as E → 1(MV/cm)
- device breakdown

Shrinking Methods (cnt.)

2. Constant Field Scaling
   - reduce later/vertical sizes by α
   - increase all doping by α
   - reduce Vth, Vdd by α

Disadvantages:
- decrease in current → $I_{D,scaled} = (I_D/\alpha)$

Advantages:

- ✓ higher speed $T_{delay,scaled}=(T_{delay}/\alpha)$
- ✓ lower capacitance $C_{scaled}=(C/\alpha)$
- ✓ lower power dissipation $P_{scaled}=(P/\alpha)$
- ✓ Increase no. devices/area scales w/ α
- ✓ constant g_m

Scaling Effects

1. Lower V_g → Lower V_{th}

- device does not turn off completely
- weak inversion layer
- increased power consumption when “off”

2. More Devices per Area

- increase interconnect cap
- increase heat production
- Increase delay

Scaling Effects (cnt.)

3. Fabrication Difficulties

- required high density
- precise alignment

4. Mobility Degradation

- high E → scattering → μ reduction
  → I_D reduction → g_m reduction

Scaling Effects (cnt.)

5. V_{th} Variation

- -1(mV/K) → 50 mV variation
- process → 50 mV variation
- V_{th} dependence on L (L variation→V_{th} variation)

6. Velocity Saturation

- lateral E ≥ 1 (V/μm) → V ≈ 10^7 (cm/s)
- premature saturation
- reduction of g_m

Scaling Effects (cnt.)

7. Hot Carrier Effects

- increased lateral E → increase carrier instant.
  kinetic energy & velocity
- ionization of Si atoms → finite $I_{drain\text{-}substrate}$
- carrier injection into gate → current out of gate
- current leakage → higher power consumption

Scaling Effects (cnt.)

8. Carrier Tunneling

- high current drive → reduce t_{ox} thickness
  → carrier tunneling through oxide

hi-k dielectric

- current drive → electrical thickness
- tunneling → physical thickness
- $T_{eq}=(k_{ox}/k)T_{phys}$

Hi-k Consequences:

- thermal stability with Si
- interface issues: scattering
- higher physical thickness → drain fringing fields → lower source-channel potential barrier
- need high potential barrier
  - lower tunneling
  - lower hot carrier emission

- References

Razavi, B. Design of Analog CMOS Integrated Circuits.
New York: McGraw Hill, 2004

Wong, H.-S.P. “Beyond the Conventional Transistor.”
IBM Journal of Research and Development

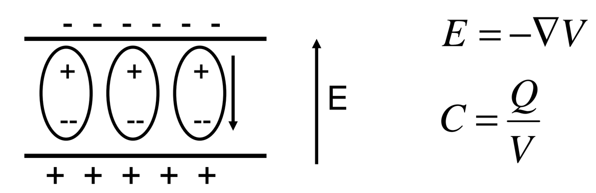
High κ Dielectrics

Stephanie

High-k gate dielectrics /edited by Michel Houssa, Bristol, England; Philadelphia :IOP Publishing, c2004

What’s special about dielectrics?

- Insulators
- Material becomes polarized in an electric field
- Used to reduces the electric field between 2 surfaces of a capacitor
- The voltage across the capacitor decreases



$$E = -\nabla V$$

$$C = \frac{Q}{V}$$

Traditional Dielectric SiO_2

- Advantages
  - Can be grown thermally on Si
    - Control over thickness and uniformity
  - Stable interface with Si
  - Low density of intrinsic interface defects
  - Thermally and chemically stable
    - Manufacturing steps can involve annealing at 1000º C
  - Large band gap
  - Functional down to around 1.5nm

Traditional Dielectric SiO_2

- Problems
  - Theoretical scaling limit ~7 Angstroms
    - Below don't have full band gap
  - Leakage due to tunneling
    - Tunneling proportional to exp(thickness)
    - 1.6 - 1.4 very thinnest that can still work with acceptable leakage
  - Defects limit thickness to around 2nm

Possible solution

- High κ dielectric constant materials (greater than SiO_2)
- $C=A\kappa\varepsilon_0/d$
- Instead of increasing d further increase κ
- For example:
  - κ of SiO_2 = 3.9
  - κ of Al_2O_3 ~ 10.0
  - Need a ~ 2.6 nm thick film of Al_2O_3 to get same capacitance as 1nm thick SiO_2

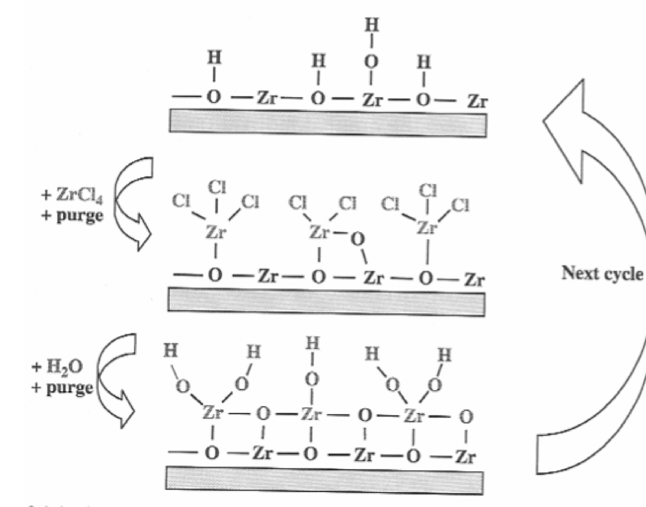
Manufacturing techniques

- Atomic layer deposition
- Chemical vapor deposition
- Pulsed laser deposition

Atomic Layer Deposition

- Thin film deposition method
- Layers grown in sequence
- Process steps
  - 2 or more gaseous chemical reactants added to vacuum chamber
  - Deposit by chemical reaction on the substrate
  - Purge
  - Repeat same or different reaction step to grow layers

Atomic Layer Deposition



High-k gate dielectrics /edited by Michel Houssa, Bristol, England; Philadelphia :IOP Publishing, c2004, pg 18

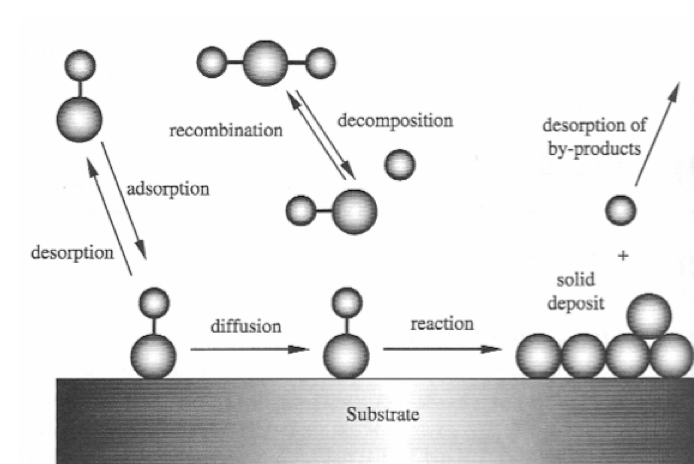
Atomic Layer Deposition

- Advantages
  - After initial layer deposited on substrate all subsequent reaction steps proceed to completion
    - Assuming use enough reactant
  - Allows large area uniformity
  - Film thickness controlled by # rxn steps
  - Can combine multiple chemical reactions
- Disadvantages
  - Step-wise growth
  - Deposition rate is slow ~ 100 - 300 nm/h

Chemical Vapor Deposition

- Gaseous chemicals reactants added to vacuum chamber
- Rxn takes place in the gas phase
- Rxn product and unused reactants deposit onto heated substrate
- 3 possibilities
  - react with the substrate
  - Diffuse on the substrate
  - Desorb from the substrate
- Purge chamber

Chemical Vapor Deposition



High-k gate dielectrics /edited by Michel Houssa, Bristol, England; Philadelphia :IOP Publishing, c2004, pg 66

Chemical Vapor Deposition

- Advantages
  - Already used in industry
  - Can control deposition rates from 1 to 1000 nm/min
  - Can control composition
- Disadvantages
  - Complex chemistry
  - Residual impurities due to by-products of chemicals
  - Extra reactants can nucleate to form particles which deposit on the substrate surface
    - Effects morphology

Pulsed Laser Deposition

- Heated substrate in vacuum chamber
- Laser is outside chamber
- Laser is focused on a source material inside chamber
- Source material is vaporized
- Sublimes onto substrate
- Advantages
  - Deposition can take place in an oxygen environment
  - High deposition rate
  - Can use higher pressures so compounds that are unstable at low pressures can be made into dielectrics

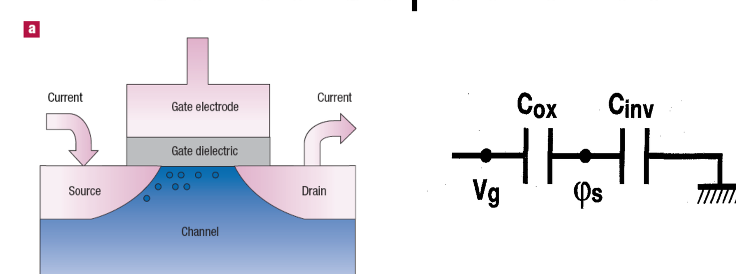
New material properties

- Same defect densities as SiO_2
- Significant reduction of electron tunneling for high κ dielectic materials of thickness of 1nm as compared to SiO_2
- Thermally and chemically stable
- Chemical stability with respect to SI
  - When depositing high κ dielectric oxide onto silicon, a thin low-κ interfacial layer can form
  - $1/C_{total} = 1/C_{low\ \kappa} + 1/C_{high\ \kappa}$
  - Equivalent oxide thickness to give same capacitance as SiO_2 increases when interfacial layer present

New material properties

- 3 ranges for new dielectric materials
- Ultra- high-κ (κ>100)
  - Allows thickest dielectric material
  - Field-induced barrier lowering effect problem
    - Dielectric rectangular rather than a film
    - Channel potential controlled be gate electrode, source and drain & MOSFET hard to turn off
- Mid-range high -κ (10<κ<100)
  - Common materials in CMOS
  - κ not necessarily high enough to suppress gate leakage current sufficiently compared to potential problems
- Moderate high-κ (4<κ<10)

Series Capacitance



• Series Capacitance in MOSFET consists of the Oxide Capacitance C_{ox} and the inversion layer Capacitance C_{inv}

• Inversion layer capacitance has significant impact on performance of scaled MOSFET

Inversion Capacitance

- Reason: The total gate capacitance , which determines the transconductance is reduced by the inversion layer capacitance
- We can depict the Inversion layer Capacitance as follows:

$$C_{inv} = \frac{q\,\partial N_s}{\partial \varphi_s}$$

• Ns is the surface carrier concentration and φ_s is the surface potential

•We find that at lower Ns, Cinv is determined by the finite effective density of states and at higher Ns, Cinv is determined quantum mechanically by the finite inversion thickness layer

Inversion Capacitance

- We can represent the total capacitance as follows:

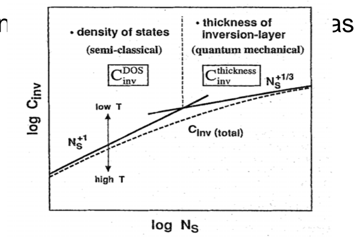
$$C_{tot} = C_{ox} \cdot \frac{1}{1 + \frac{C_{ox}}{C_{inv}}}.$$

- if C_{ox} is comparable to C_{inv}, then C_{tot} becomes significantly lower than C_{ox}
- Since, device scaling rule demands thinner gate oxide, thus resultant larger C_{ox} , the degradation of C_{tot} due to C_{inv} becomes larger with the shrinkage of devices
- IMP: quantitative understanding of C_{inv} not sufficient
- N_s dependence of C_{inv} near room temperature not clarified yet

Inversion Capacitance C_{inv}

- One aspect of C_{inv} is from the finite density of states and other from the inversion layer thickness
- If the density of states is finite, a finite amount of change in the surface potential is always necessary to increase Ns
- Therefore the C_{inv} due to finite den as follows

$$C_{inv}^{DOS} = \frac{q\,\partial N_s}{\partial \varphi_s} = \frac{q^2}{2k_B} N_s^{+1} \cdot T^{-1}.$$



• Since, location of inversion of inversion layer away from the Si/SiO_2 interface, finite inversion layer increases the oxide thickness layer

$$C_{inv}^{thickness} = \frac{\varepsilon_{\mathrm{Si}}}{Z_{inv}}$$

Inversion Capacitance

$$C_{inv}^{thickness} = \frac{\varepsilon_{\mathrm{Si}}}{Z_{inv}}$$

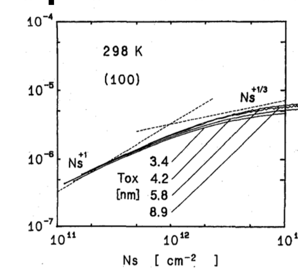
$$= \left(\frac{4\varepsilon_{\mathrm{Si}}^2 q^2 m_3}{9\hbar^2}\right)^{1/3} \left(N_{depl} + \frac{11}{32} N_s\right)^{1/3}$$

$$\approx \left(\frac{11\varepsilon_{\mathrm{Si}}^2 q^2}{72\hbar^2}\right)^{1/3} \cdot m_3^{1/3} \cdot N_s^{1/3}$$

- C_{inv} thickness is independent of temperature and is in proportion to the one-third power of N_s
- Thus, in order to obtain total Cinv,

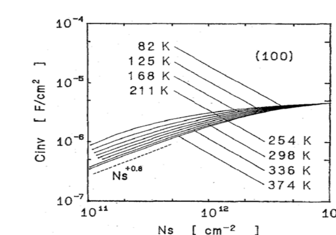
$$C_{inv}^{-1} = (C_{inv}^{DOS})^{-1} + (C_{inv}^{thickness})^{-1}$$

N_s dependence of C_{inv}



- The almost identical curves indicate that Cinv at a same value of Ns should be independent of the oxide thickness
- Mild scattering of value attributed to variation in measured value of Tox
- Cinv increases with an increase in Ns
- Influence of Cinv upon the gate capacitance becomes more severe for lower Ns
- With increasing Ns, the Ns dependence of Cinv becomes weaker

Temperature dependence of C_{inv}



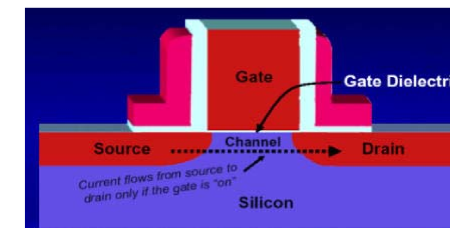
- As the temperature increases, the Cinv in lower Ns regions becomes smaller and the Ns dependency approaches close to linear dependence
- At higher temperatures, the Cinv is almost inversely proportional to T

Leakage Currents

- When the gate oxide of the MOSFET breaks down, a leakage path is created between channel and gate
- When gate oxide is ultra thin, the first breakdown is a soft breakdown
- As long as the power supply can cope with the additional load due to the leakage current, an inversion layer will still form in the channel and the transistor will function properly
- Even when one or more transistors show a soft breakdown, the circuit may not fail if the circuit can tolerate the reduced performance levels
- Even hard breakdown has shown not to cause failure if the leakage current is not too high

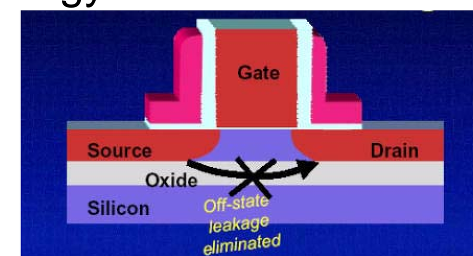
Current CMOS Design

- CMOS Design basics have not changed
- Performance and scalability issues are resolved by tweaking the design
- Different techniques are being used to extend the lifetime of the Planar CMOS



Silicon on Insulator

- Primarily used by AMD and IBM
- Addition of insulator reduces current leakage and increases switching speed
- Technology has been around since the 1940s

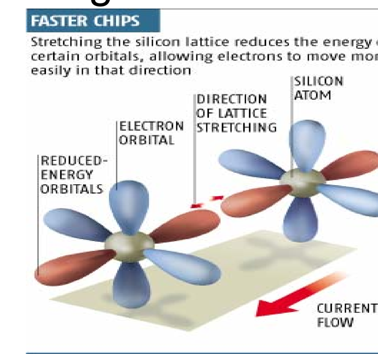


Strained Silicon

- The crystal structure of Si lattice is one of the limiting factors for the flow of charge carriers (proposed in early 90s)
- If the lattice structure is “deformed” appropriately, we can increase the performance of transistors

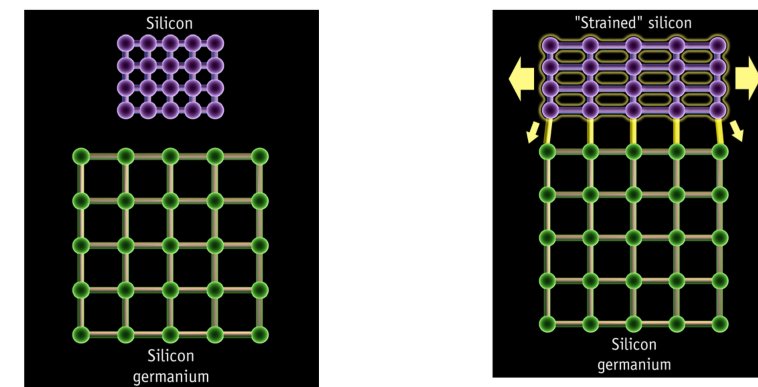
Strained Silicon – Orbital Model

- As the Si lattice is “strained” (stretched or compressed), the flow rate of charge carriers changes



Creating Strained Silicon

- Silicon can be strained with the aid of materials such SiGe



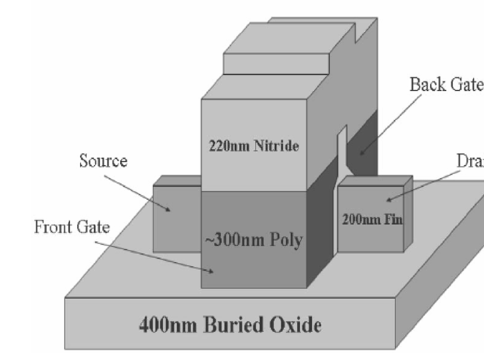
Strained Si Demo

- Movie time.

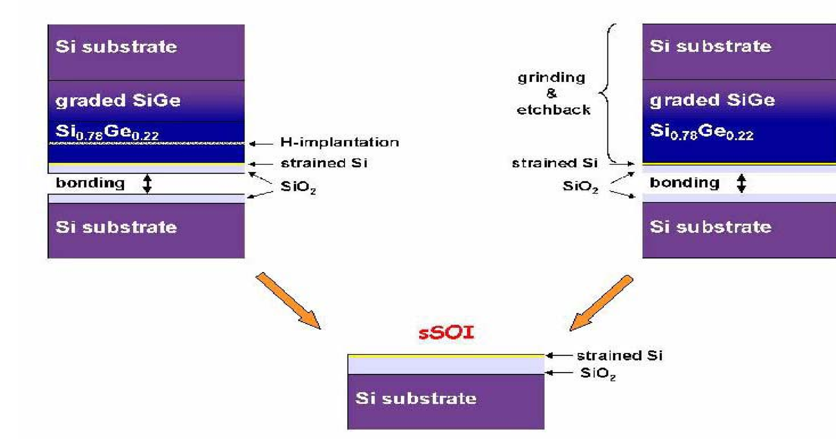
Other Techniques

- A combination of techniques has saved the CMOS transistor from going extinct
- Planar transistors are expected to be out of date at the 45 – 32 nm manufacturing level
- A possible replacement for planar transistors is the FinFET (3D Transistors)
- Intel and AMD have proposed the use of Tri-gate transistors

FinFET



sSOI



Q/A

- Questions?

Notes Corresponding to 1st set of slides

Slide 1

Why Small?

mosFET act like resistor : $\downarrow$ resistance $\Rightarrow$ $\uparrow$ current

$\downarrow$ gate size $\Rightarrow$ $\downarrow C_{gate}$

$\downarrow$ switching time $\Rightarrow$ $\uparrow$ speed

$\uparrow$ # of FET on chip $\Rightarrow$ $$$

Slide 2

Shrinking Methods

I) Constant Voltage Scaling

- shrink all dimensions by same factor
- keep V_{DD} constant

$\downarrow t_{ox}$, V_{DD} cnst $\Rightarrow \vec{E}\uparrow \Rightarrow \mu\uparrow \Rightarrow$ speed $\uparrow$

But as $\vec{E} \rightarrow 1\,\text{mV/cm} \Rightarrow \mu\downarrow$ & breakdown

Slide 3

II) Constant Field Scaling

$$I_{D,scaled} = \frac{1}{2}\mu_n(\alpha C_{ox})\left(\frac{W/\alpha}{L/\alpha}\right)\left(\frac{V_{GS}}{\alpha} - \frac{V_{TH}}{\alpha}\right)^2$$

$$I_{D,scaled} = (I_D/\alpha)$$

Slide 4

- $\uparrow$ speed $\Leftrightarrow$ $\downarrow$ delay time

$$T_{delay,scaled} = \left(\frac{C}{I}\right)V_{dd} = \left(\frac{1}{\alpha}\right)T_{delay}$$

- $$C_{channel\,scaled} = \frac{W}{\alpha}\frac{L}{\alpha}(\alpha C_{ox}) = \frac{1}{\alpha}WLC_{ox}$$

  similarly, all other caps scale down by $1/\alpha$

- Digital Circuits: power dissipation $\downarrow$

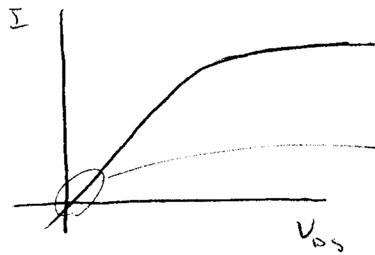
$$P = fCV_{DD}^2 \Rightarrow P_{scaled} = P/\alpha^3$$

- $$g_{m,scaled} = \mu(\alpha C_{ox})\left(\frac{W/\alpha}{L/\alpha}\right)\left(\frac{V_G - V_{TH}}{\alpha}\right) = g_m$$

Slide 5

In practice, voltage does NOT scale down by same factor cause it isn't as ideal as it seems.

L: $1\mu \rightarrow 0.25\mu$, but $V_{DD}: 5 \rightarrow 2.5$ ($V_{TH}: 0.8 \rightarrow 0.4$)



$\rightarrow V_{GS} \simeq V_{TH} \Rightarrow$ subthreshold current significant as device size ↓.

Slide 6

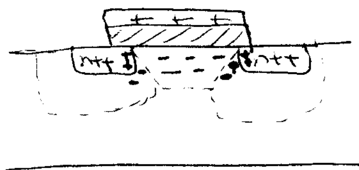
Fabrication Difficulty: control of density & location of dopant in channel, drain, and source becomes difficult as size ↓.

Mobility Degradation: hi $\vec{E} \Rightarrow$ more carriers confined to narrower region below oxide-Si interface $\Rightarrow$ scattering $\Rightarrow \mu \downarrow$

Since in practice haven't used ideal constant field scaling $\Rightarrow$ have higher $\vec{E}$ than conventional MOSFET $\Rightarrow$ small devices have μ degradation

Slide 7

V_{TH} Variation:



part of carrier charge in channel is imaged by S & D areas (S = source, D = drain)

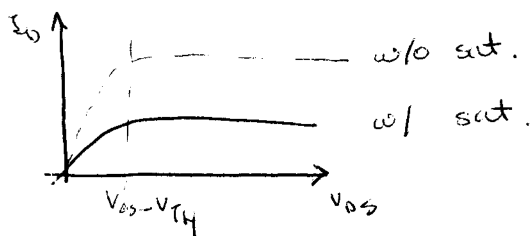
$\Rightarrow$ requires lower V_g to create inversion

But can't control L (length) accura[te]

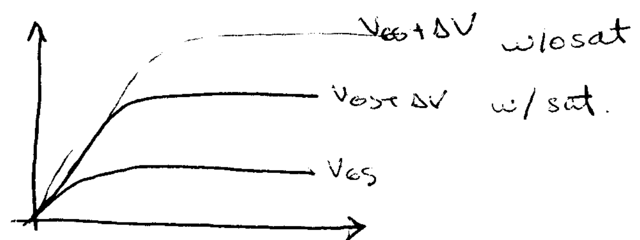
$\Rightarrow V_{TH}$ variation.

Velocity Saturation:

$L < 1\mu m \Rightarrow \vec{V}$ sat



— premature saturation —



— reduction of g_m —

Slide 8

Hot Carrier Effects: lateral $\vec{E}$ ↑ ⇒ ↑ instantaneous velocity & kinetic energy of carriers ⇒ hit Si atoms as accelerate to drain ⇒ ionization of drain ⇒ { ē → absorbed by drain; hole → absorbed by substa

⇒ drain-substrate current

Slide 9

- As V_{DD} ↓, need t_{ox} ↓ (→ electrical thickness) in order to have same charge density in channel

- hi k material: { ↑ physical thickness; ↓ electrical thickness, t_{eq}

Slide 10

- hi temp doping for drain, source, & poly gate