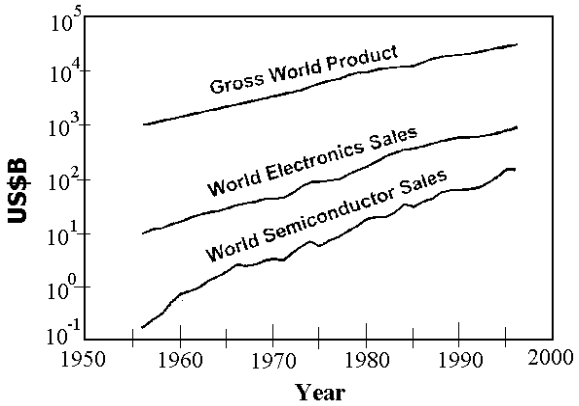
Trends in Integrated Circuits Technology

Prof. Krishna Saraswat

Department of Electrical Engineering
Stanford University
Stanford, CA 94305
saraswat@stanford.edu

Miniaturization => Market growth

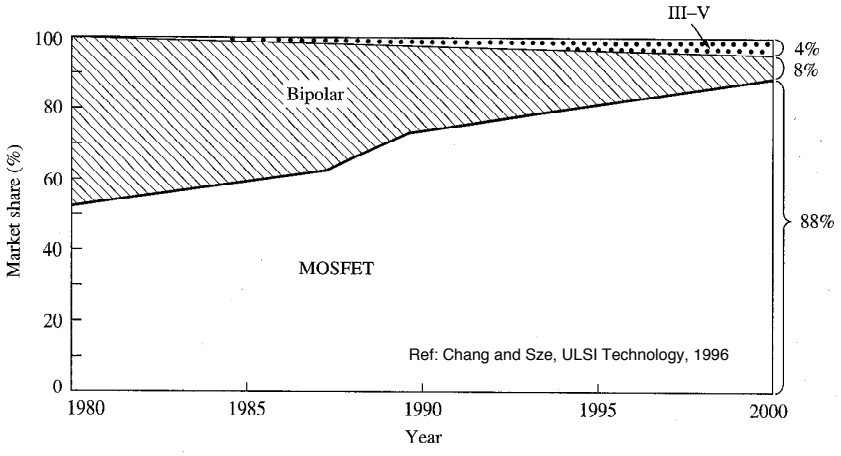


Semiconductors have become increasingly more important part of world economy

Courtesy Prof. Tsu-Jae King

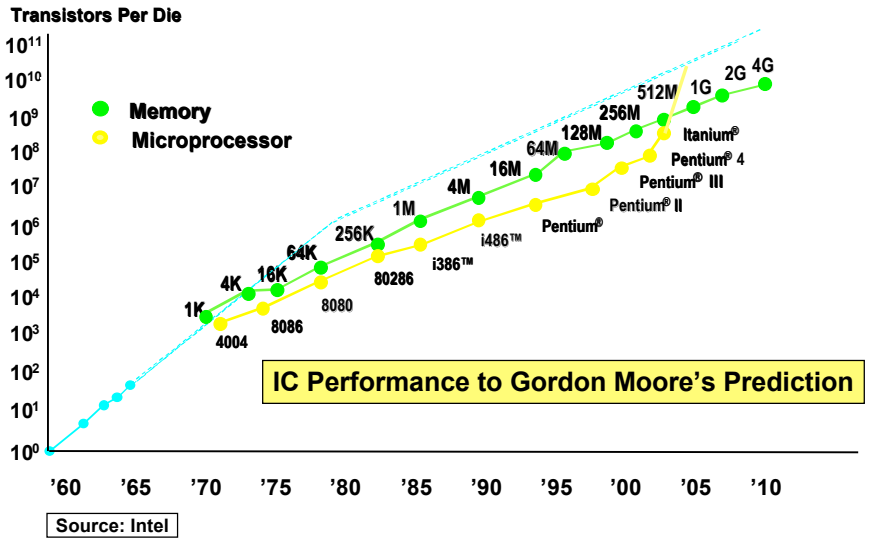
(Sources: VLSI Research Inc.; United Nation yearbook; World Bank Database; IMF)

World IC Market by Technology

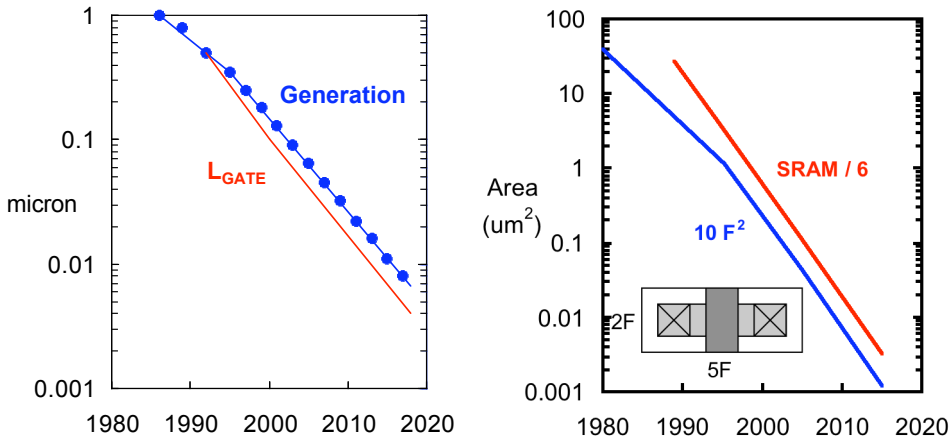


Silicon CMOS has become the pervasive technology

Moore's Law



Feature Size Trend

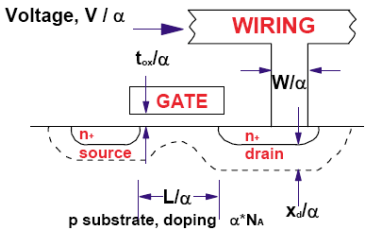


Gate length is not true measure of transistor size

Example: Microprocessor Evolution

Generation:	1.5µ	1.0µ	0.8µ	0.6µ	0.35µ	0.25µ
Intel386™ DX Processor						
Intel486™ DX Processor						
Pentium® Processor						
Pentium® II Processor						

MOS Device Scaling



Constant E Field Scaling

All device parameters are scaled by the same factor α.

- Gate oxide thickness t_{ox} ↓
- Channel length L ↓
- Source/drain junction depth X_j ↓
- Channel doping ↑
- Supply voltage V_D ↓

Why do we scale MOS transistors?

1. Increase device packing density ~ α^2
2. Improve frequency response (speed) ~ a
3. Power/ckt: ~$1/\alpha^2$, power density constant
4. Improve current drive (transconductance g_m)

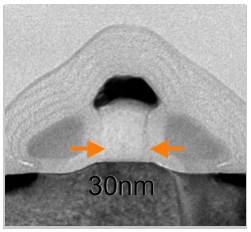
$$g_m = \left.\frac{\partial I_D}{\partial V_G}\right|_{V_D = const}$$

$$\approx \frac{W}{L}\,\mu_n\,\frac{K_{ox}}{t_{ox}}\,V_D \qquad \text{for } V_D < V_{D_{SAT}},\ \textit{linear region}$$

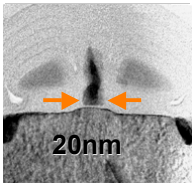
$$\approx \frac{W}{L}\,\mu_n\,\frac{K_{ox}}{t_{ox}}\,(V_G - V_T) \qquad \text{for } V_D > V_{D_{SAT}},\ \textit{saturation region}$$

Intel's Transistor Research down to 10nm

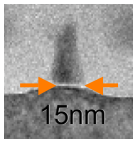
Electronics is Nanotechnology



65nm process
2005 production

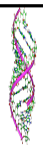


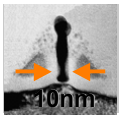
45nm process
2007 production



32nm process
2009 production

DNA is 15 nm wide

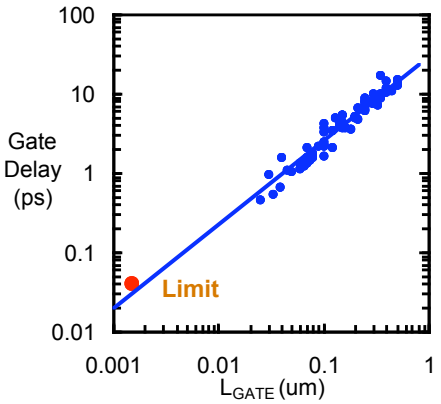




22nm process
2011 production

***Source*: Intel**

Speed increases as a result of scaling

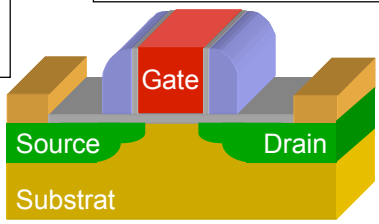


Source: Mark Bohr, Intel

Physical Limits in Scaling Si MOSFET

Gate stack
- Tunneling current ⇒ Increased I_{off}
- Gate depletion ⇒ Increased EOT

Source/Drain
- Contact resistance
- Doping level, abruptness



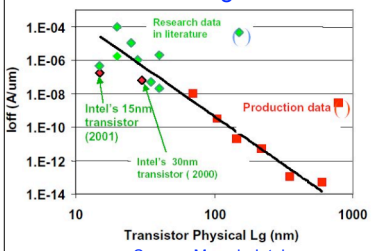
High E-Field
- Mobility degradation
- Reliability

Channel
- Surface scattering - the "universal mobility" tyranny
- DIBL ⇒ drain to source leakage
- Subthreshold slope limited to 60mV/decade (kT/q) ⇒ Increased I_{off}
- V_G - V_T decrease ⇒ reduced I_{ON}

Net result: *Bulk-Si CMOS device performance increase commensurate with size scaling is unlikely beyond the 70 nm node*

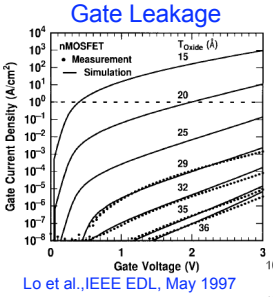
MOSFET Scaling Limit: Leakage

Total Leakage Trend



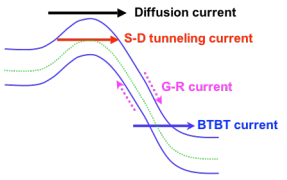
Source: Marcyk, Intel

Gate Leakage

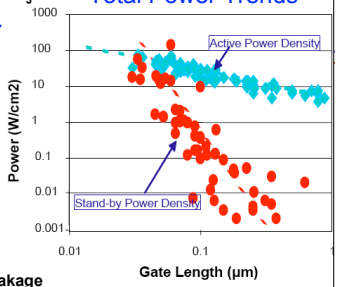


Lo et al.,IEEE EDL, May 1997

S/D Leakage



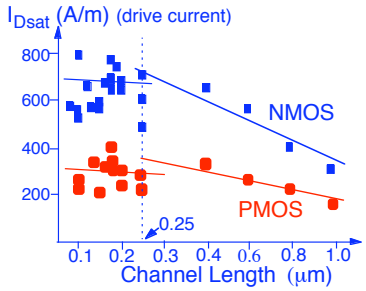
Total Power Trends



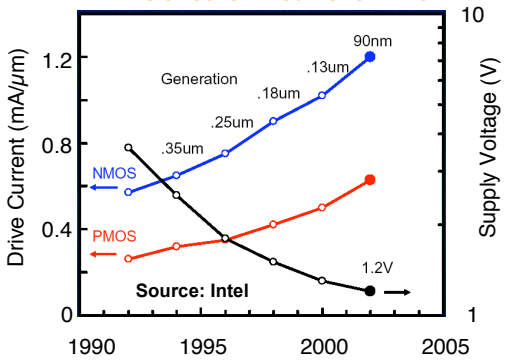
- Ability to control I_{off} will limit gate-length scaling
  - Thermionic emission over barrier
  - QM tunneling through barrier
  - Band-to-band tunneling from body to drain
- To suppress D/S leakage, need to use:
  - Higher body doping to reduce DIBL
    - ⇒ lower mobility, higher junction capacitance, increased junction leakage
  - Thinner gate dielectric to improve gate control ⇒ higher gate leakage
  - Ultra-shallow S/D junctions to reduce DIBL ⇒ higher R_{series}

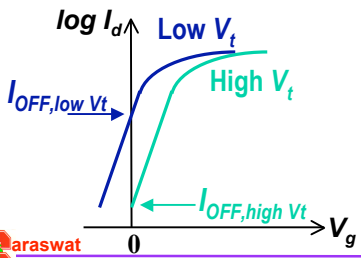
MOSFET Scaling Problem: Saturation of I_{Dsat}

Constant OFF current Limit



Relaxed OFF current Limit

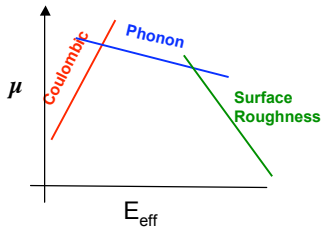




- Data from IBM, TI, Intel, AMD, Motorola and Lucent
- Low OFF current desirable

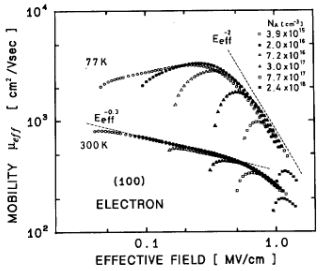
Changhoon Choi, PhD Thesis, Stanford Univ., 2002

Effects of Scaling Bulk MOSFET on Mobility



$$\frac{1}{\mu_{eff}} = \frac{1}{\mu_C} + \frac{1}{\mu_{ph}} + \frac{1}{\mu_{sr}}$$

$$E_{eff} = \frac{q}{\varepsilon_{Si}}(N_{dep} + \eta \cdot N_{Channel})$$

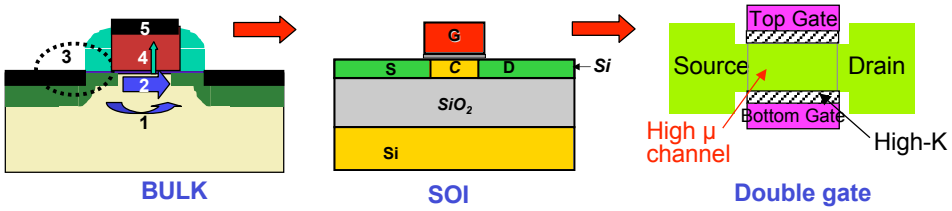


S. Takagi *et al.*, *IEEE TED*, **41** (1994) 2357.

N_{dep}= depletion charge density
$N_{Channel}$ = charge induced in the channel

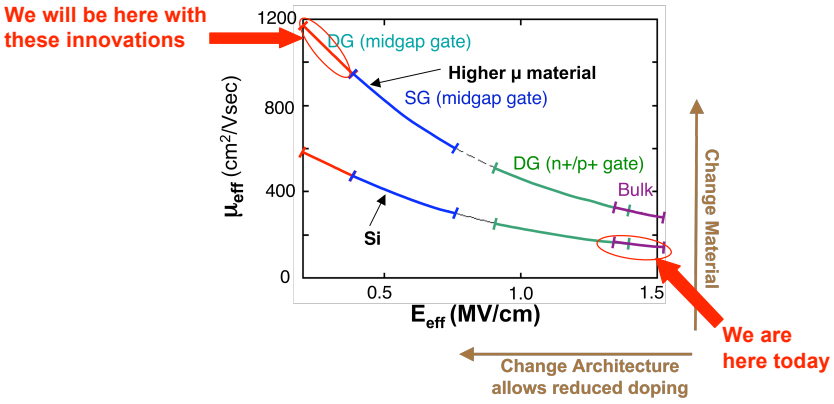
- Increases in substrate doping ⇒ N_{dep} ⇑
- Gate oxide thickness decrease ⇒ $N_{Channel}$ ⇑
- E_{eff} increases with scaling ⇒ μ ⇓
- Reduced gate oxide thickness increases remote charge scattering ⇒ μ ⇓
- High k dielectrics have higher coulombic scattering due to surface states and phonon scattering ⇒ μ ⇓

New Structures and Materials for Nanoscale MOSFETs



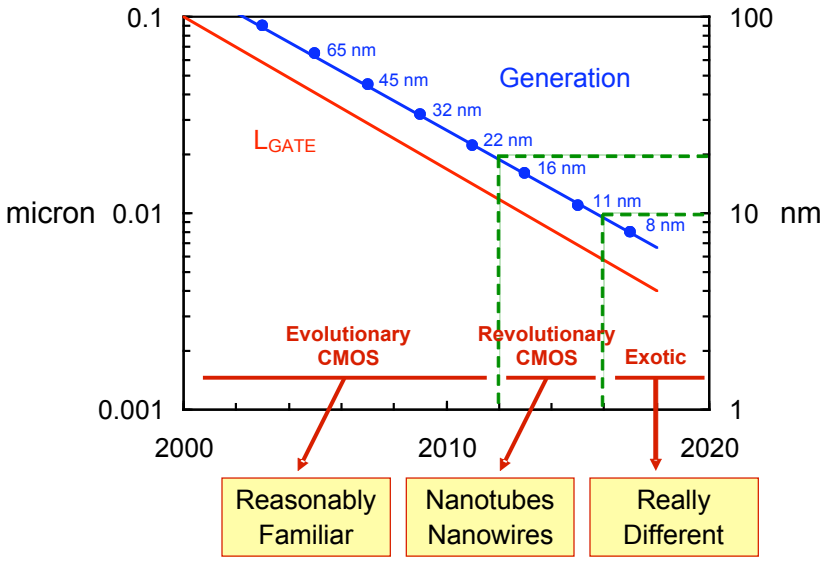
1. Electrostatics - Double Gate
   - Retain gate control over channel
   - Minimize OFF-state drain-source leakage
2. Transport - High Mobility Channel
   - High mobility/injection velocity
   - High drive current for low intrinsic delay
3. Parasitics - Schottky S/D
   - Reduced extrinsic resistance
4. Gate leakage - High-K dielectrics
   - Reduced power consumption
5. Gate depletion - Metal gate

Combining New Device Structures with New Materials



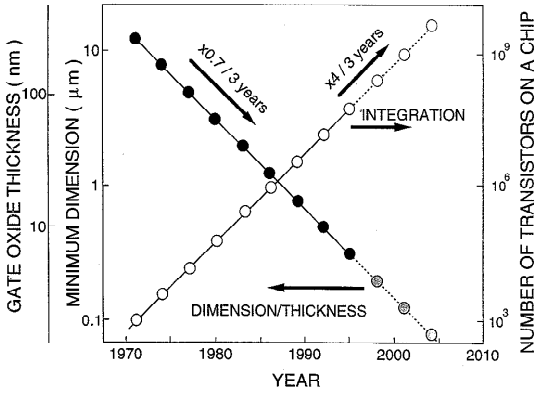
- With better injection and transport we may be able to improve MOSFET I_{ON}
- With better electrostatics we may be able to minimize I_{off}

Nanotechnology Eras



Source: Mark Bohr, Intel

Scaling of MOS Gate Dielectric



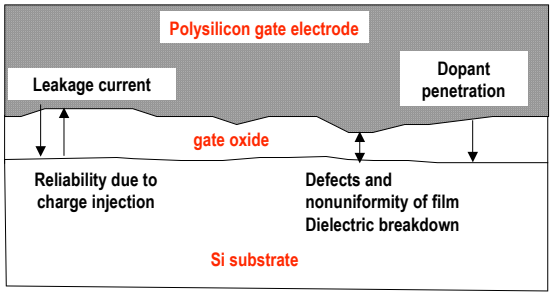
$$I_D \propto g_m \propto \frac{K}{thickness}$$

(Ref: S. Asai, Microelectronics Engg., Sept. 1996)

Gate SiO_2 thickness is approaching < 10 Å to improve device performance

- **How far can we push MOS gate dielectric thickness?**
- **How will we grow such a thin layer uniformly?**
- **How long will such a thin dielectric live under electrical stress?**
- **How can we improve the endurance of the dielectric?**

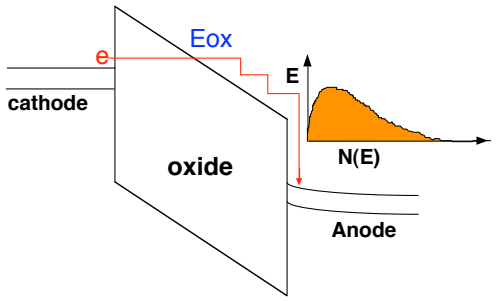
Problems in Scaling of Gate Oxide



- Below 20 Å problems with SiO_2
  - Gate leakage => circuit instability, power dissipation
  - Degradation and breakdown
  - Dopant penetration through gate oxide
  - Defects

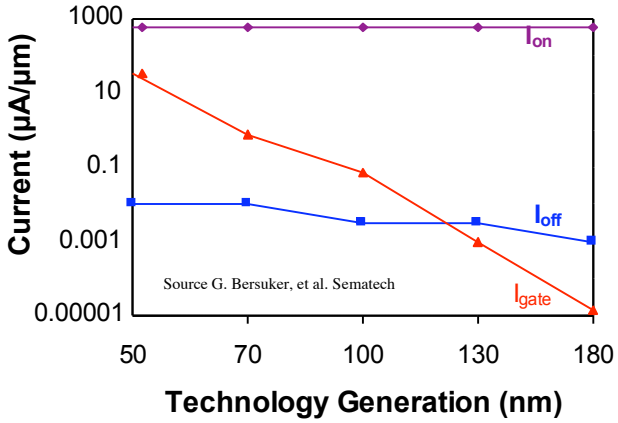
Dielectric Degradation

- Degradation during device operation due to high E field causing current injection
- Degradation during fabrication due to charging in plasma processing



What are the mechanisms for damage and breakdown?
How can we engineer the gate dielectric to minimize the damage?

Gate Oxide Scaling Issues: Leakage



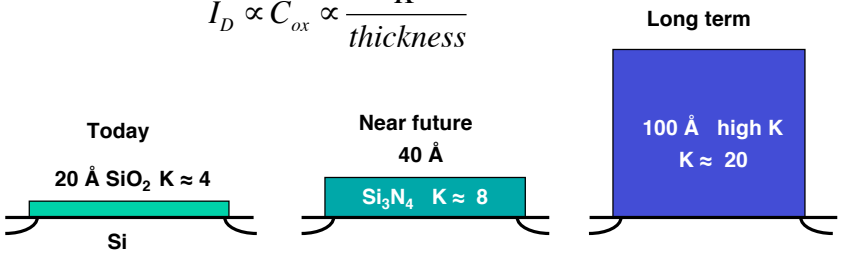
- I_{on} is not increasing with scaling
- I_{gate} ⇑, power dissipation ⇑
- Circuit instability

High-k MOS Gate Dielectrics

$I_{channel} \propto$ charge x source injection velocity
$\propto$ (gate oxide cap x gate overdrive) v_{inj}
$\propto C_{ox} (V_{GS} - V_T)\ E_{source}\ \mu_{inj}$

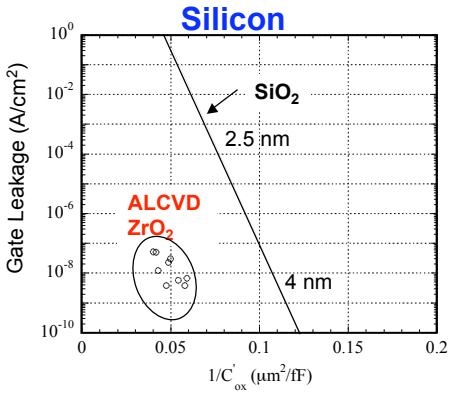
Historically C_{ox} has been increased by decreasing gate oxide thickness. It can also be increased by using a higher K dielectric

$$I_D \propto C_{ox} \propto \frac{K}{thickness}$$

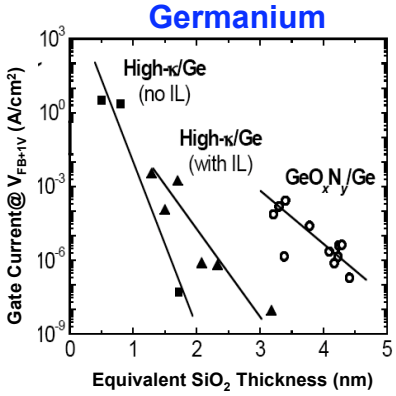


Higher thickness -> reduced gate leakage

Capacitance and Leakage for High-k Gate Dielectric Films Grown Using ALCVD

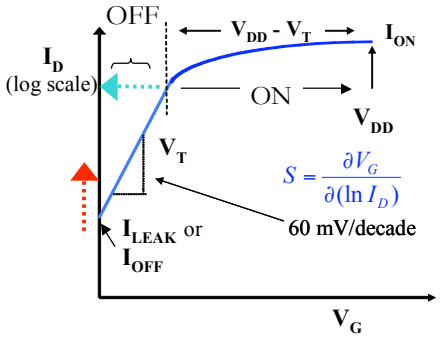


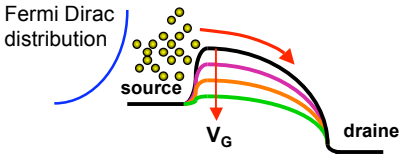
Perkins, Saraswat and McIntyre, Stanford Univ. 2002



Chui, Kim, Saraswat and McIntyre, Stanford Univ. 2004

Subthreshold Behavior





$$Q_I \propto e^{q\phi_s / kT}$$

$$\phi_s = f(V_G) = \frac{1}{\xi} V_G$$

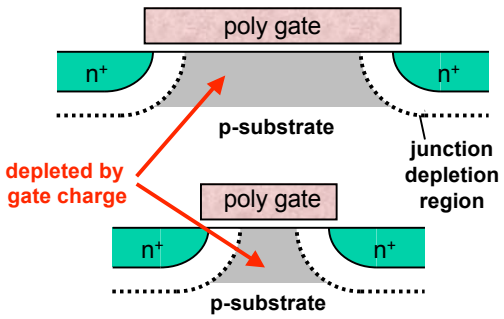
$$I_D \propto Q_I \propto e^{q V_G / \xi kT}$$

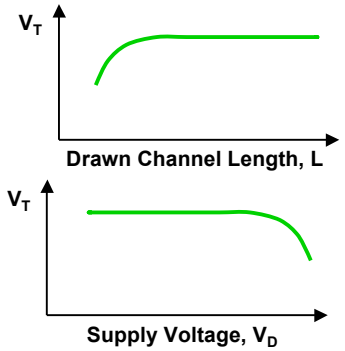
Where $\xi = 1 + \frac{1}{2C'_{ox}} \sqrt{\frac{q\varepsilon_s N_a}{|\phi_p|}}$

- Diffusion of carriers over the barriers to the channel.
- Fermi-dirac distribution of carriers: $e^{-E/kT}$
- Gate reduces the barrier to current flow.

Effect of Reducing Channel Length

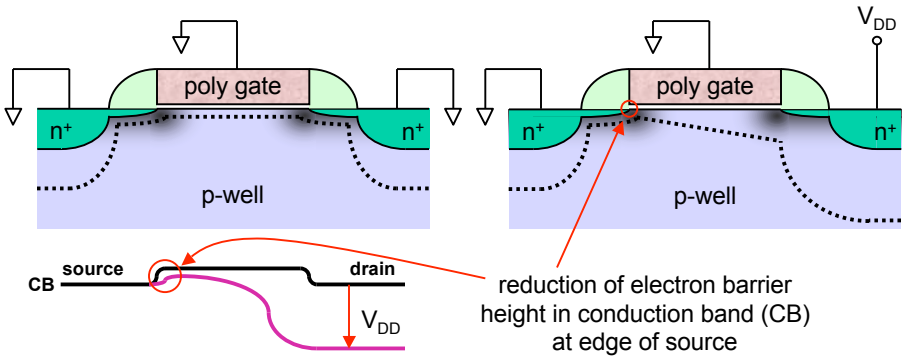
- In devices with long channel lengths, the gate is completely responsible for depleting the semiconductor (Q_B). In very short channel devices, part of the depletion is accomplished by the drain and source bias
- Since less gate voltage is required to deplete Q_B, $V_T\downarrow$ as $L\downarrow$. Similarly, as $V_D\uparrow$, more Q_B is depleted by the drain bias, and hence $V_T\downarrow$. These effects are particularly pronounced in lightly doped substrates.



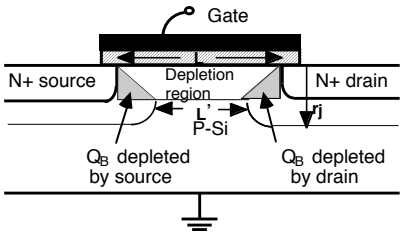


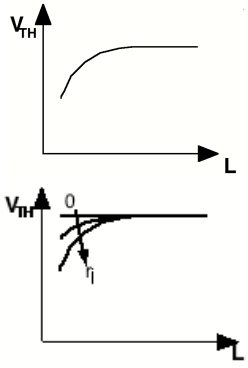
Drain-Induced Barrier Lowering (DIBL)

- Exacerbates subthreshold leakage in short-channel devices
- Soft punchthrough induced by drain-to-substrate depletion region
  - $|V_T| \downarrow$ as $V_D \uparrow$ [drain-induced short channel effects (SCE)]
  - $V_D \uparrow \rightarrow$ drain-to-substrate depletion region grows with more reverse bias
  - Lateral electric fields in drain-induced depletion region lowers source-to-channel barrier, allowing more carriers to diffuse from source to channel



Why do we need to scale junction depth?



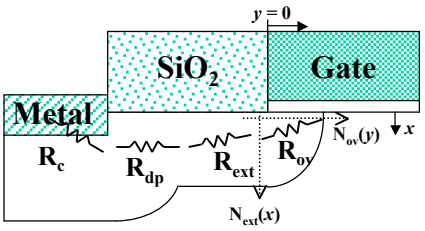


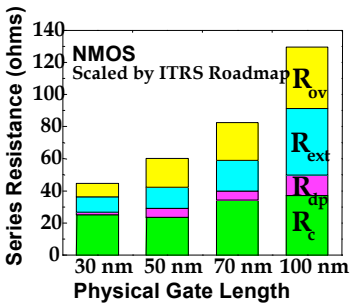
$$V_T = V_{FB} - 2\cdot\phi_F - \frac{Q_B}{C_{ox}}\cdot\left[1-\left(\sqrt{1+\frac{2\cdot W}{r_j}}-1\right)\cdot\frac{r_j}{L}\right]$$

- Roll-off in threshold voltage as the channel length is reduced
- V_T roll-off is reduced as junction depth(r_j) is decreased
- Sheet resistance increases as junction depth is reduced

L. Yau, *Solid-State Electronics*, vol. 17, pp. 1059, 1974

Source/Drain Resistance





Source: Jasonn Woo, UCLA

Problem in junction scaling:

- Sheet resistance of a junction is a strong function of doping density
- Maximum doping density is limited by solid solubility and it does not scale
- Silicidation can minimize the impact of junction sheet resistance (R_s,R_d)
- Contact resistance R_c is one of the dominant components for future technology

Contact Resistance

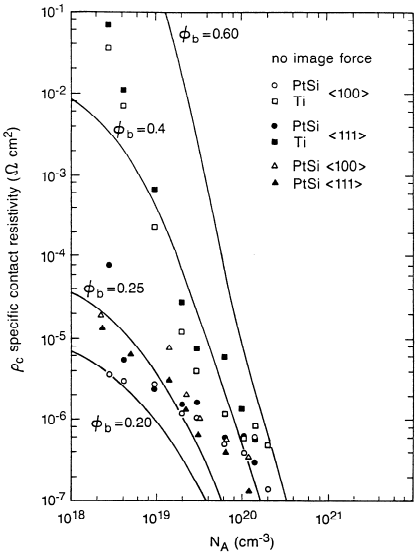
Specific contact resistivity

$$\rho_c = \rho_{co} \exp\left(\frac{2\phi_B}{q\hbar}\sqrt{\frac{\varepsilon_s m^*}{N}}\right) \quad ohm - cm^2$$

Doping density

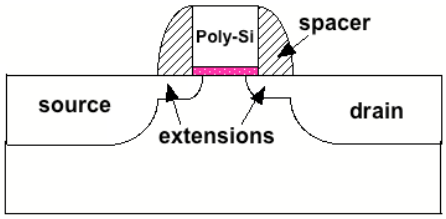
PROBLEMS

- **Contact resistance is a strong function of doping density at the metal/silicon interface**
- **Solid sulubility of dopants does not scale !**

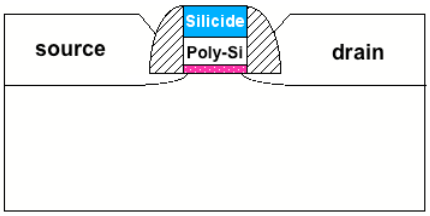


(Ref: S. Swirhun, PhD Thesis, Stanford Univ. 1987)

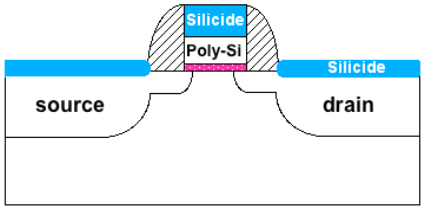
Solutions to Shallow Junction Resistance Problem



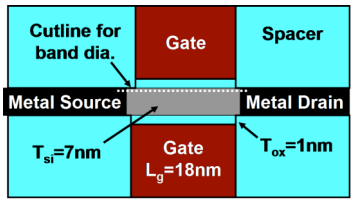
Extension implants



Elevated source/ drain



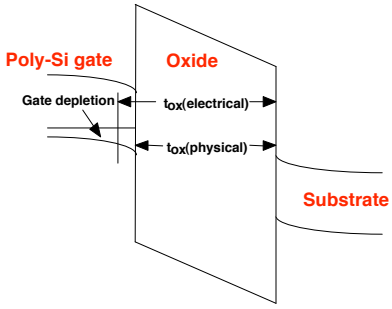
Silicidation



Schottky Source/Drain

Problems with Poly-Si Gate.

This occurs because of high E - field due to a combination of higher supply voltage and thinner gate oxide.

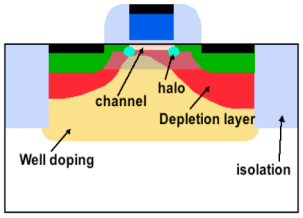


- Effect of depletion is to increase effective t_{ox} and thus reduce C_{ox}
- A reduced C_{ox} implies reduction in g_m and thus I_D(on)
- Ionized impurities in the gate electrode cause "remote charge scattering"
  ⇒ Reduced mobility

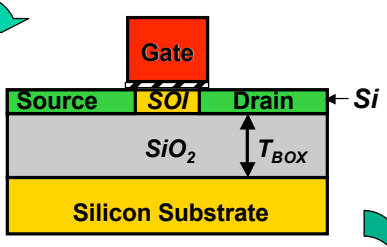
Need metal gate electrode with proper workfunction

Evolution of MOSFET Structures

BULK



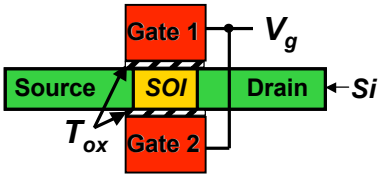
Ultra-Thin Body Single Gate SOI



Advantages of Ultra-Thin Body SOI

- Depleted channel ⇒ no conduction path is far from the gate
- Short channel effects controlled by geometry
- Steeper subthreshold slope
- Lower or no channel doping
- Higher mobility
- Reduced dopant fluctuation

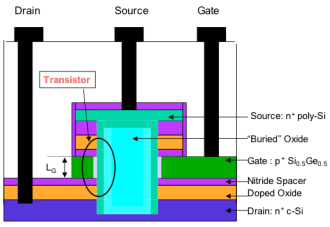
Ultra-Thin Body Double Gate SOI

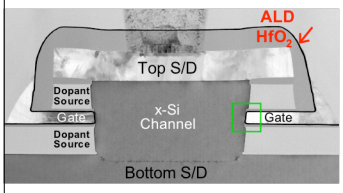


Ref: Philip Wong, IEDM Short Course, 1999

Non Planar MOSFETs

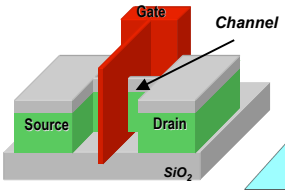
Vertical FET

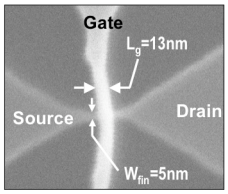




Stanford, AT&T

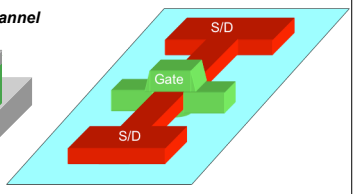
Double Gate FinFET

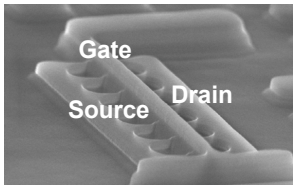




UC Berkeley

Tri Gate FET



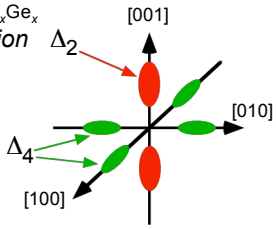


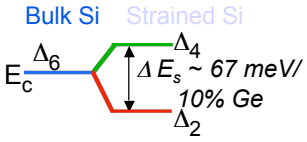
Intel

Transport: Effects of Biaxial Tensile Strain on Si Energy Bands

Hoyt, 2002

Strained Si grown on Relaxed $Si_{1-x}Ge_x$





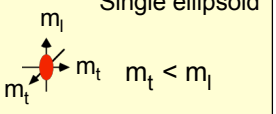
Conduction Band

Additional splitting: ➔
Band repopulation

- reduced intervalley scattering
- smaller in-plane effective transport mass

$$\mu = q\frac{\tau}{m_c^*}$$

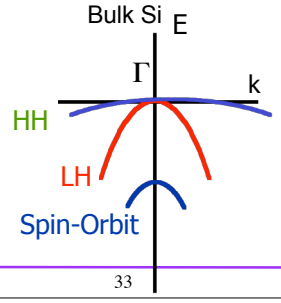
Single ellipsoid

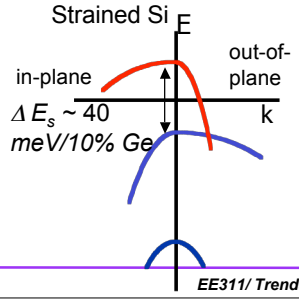


Valley Band

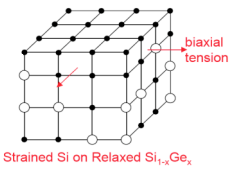
HH/LH degeneracy lifted at Γ

- reduced interband scattering
- smaller in-plane transport mass due to band deformation

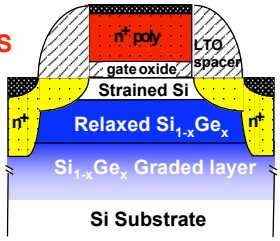




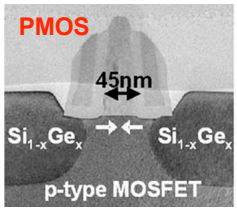
Mobility Enhancements in Strained-Si MOSFETs

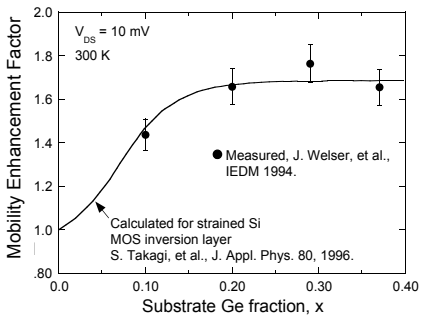


NMOS

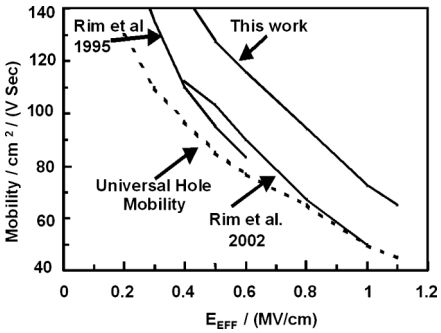


PMOS





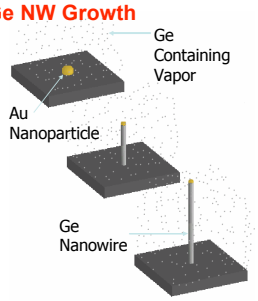
Gibbons Group, Stanford



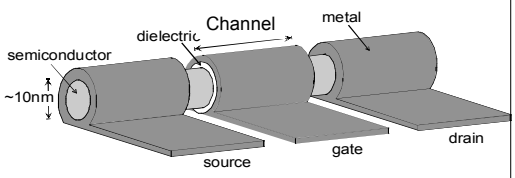
Intel

Nanowire and Nanotube FETs

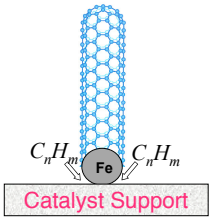
Ge NW Growth



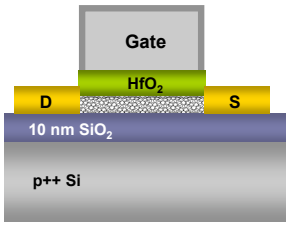
ALD HfO_2 Coated of Ge NW FET



Carbob Nanotube Growth

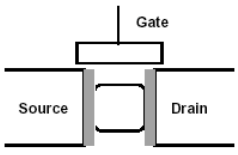


Carbob Nanotube MOSFET

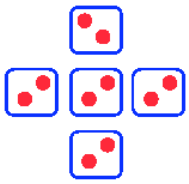


Key Challenge: Controlled growth

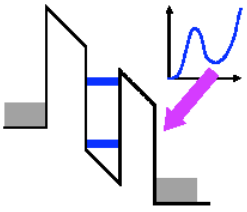
Seemingly Useful Devices



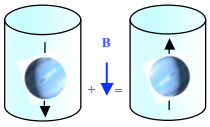
Single Electron Transistors (SET)
Limited Current Drive
Cryogenic operation



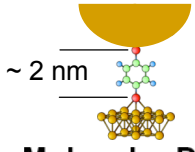
Quantum Dot
Limited Fan-Out
Critical dimension control



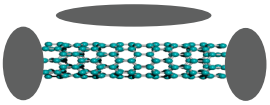
Resonant Tunneling Diode
Challenging fabrication and process integration



Spintronics
Need high spin injection and long spin coherence time



Molecular Device
Limited thermal stability
New architectures needed

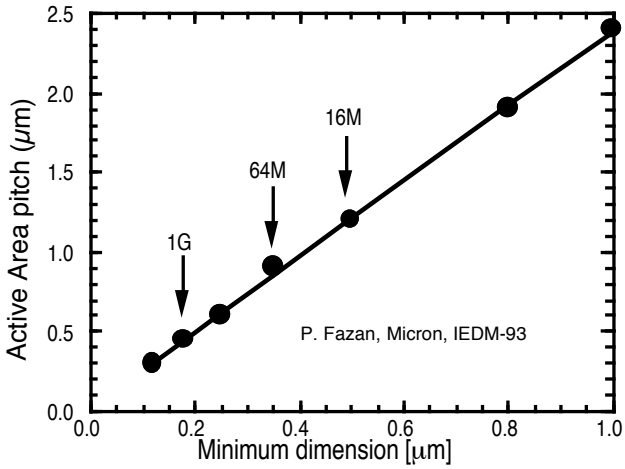


Carbon Nanotubes
Controlled growth

- **In general this device scaling methodology does not take into account many other chip performance and reliability issues, e.g., interconnects, contacts, isolation, etc.**

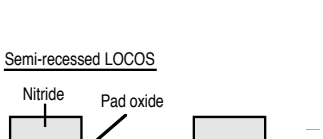
- **These factors are now becoming an obstacle in the evolution of integrated circuits.**

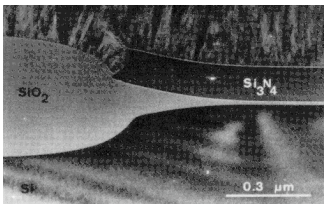
Device Isolation pitch as a function of minimum dimension



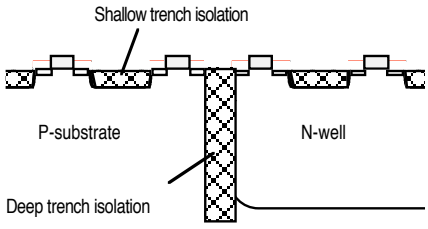
With decreasing feature size the requirement on allowed isolation area becomes stringent.

Scaling of Device Isolation



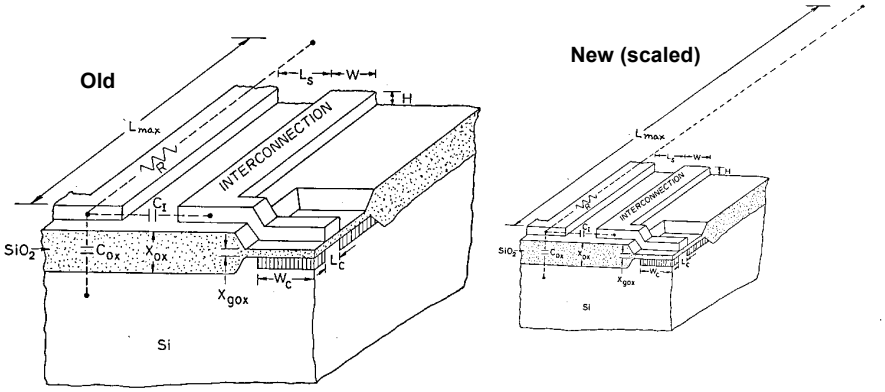


LOCOS based isolation technologies have serious problems in loss of area due to bird's beak.



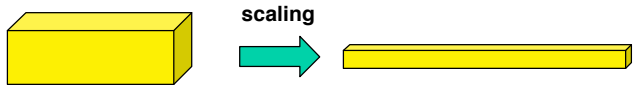
Trench isolation can minimize area loss

Scaling of interconnections

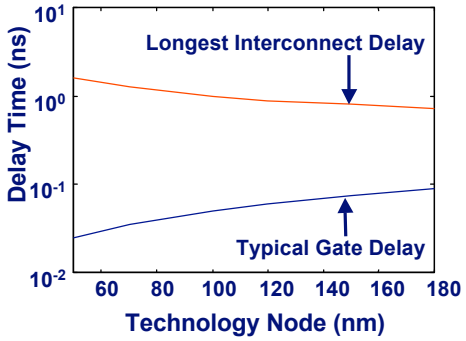


- **Bigger chip => longer interconnects**
- **Scaling to smaller dimensions => reduced cross section**
- **Larger R, L and C**

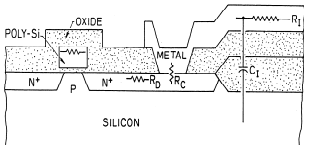
Interconnect Delay Is Increasing



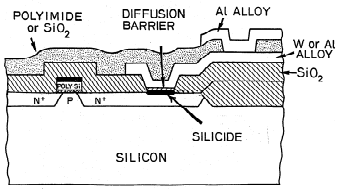
- Chip size is continually increasing due to increasing complexity
  - Increase in R, L and C
- Device performance is improving but interconnect delay is increasing
- Need better materials
  - Metal with lower resistivity
  - Dielectrics with lower K
  - Other solutions, e.g., 3D, optical interconnects



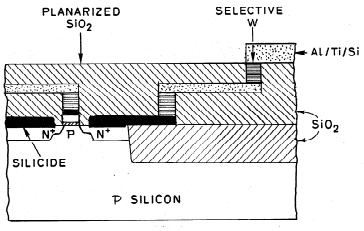
Advances in Backend Technology



1970's Poly-Si gate
Aluminum

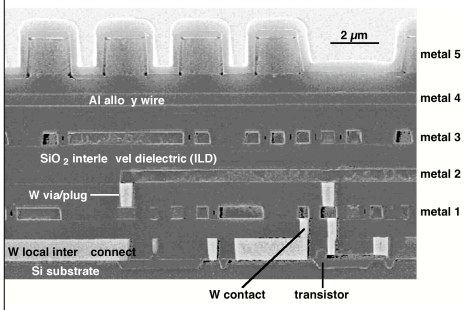


1980's Aluminum alloys
Silicide contacts
Polycide gates
Local planarization

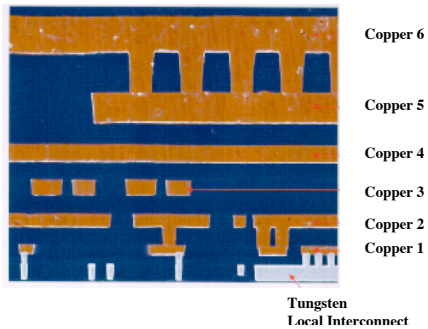


1990's Layerd aluminum/titanium
Salicides
CVD tungsten plugs
Shallow trench isolation
Global planarization

Current Interconnect Technologies

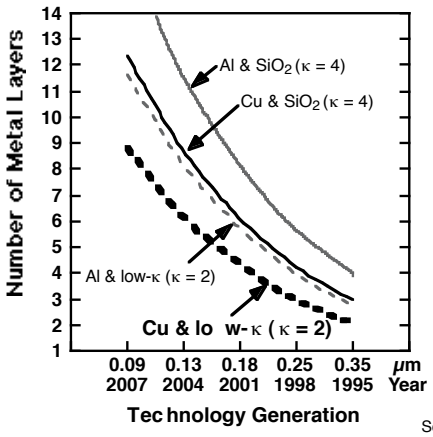


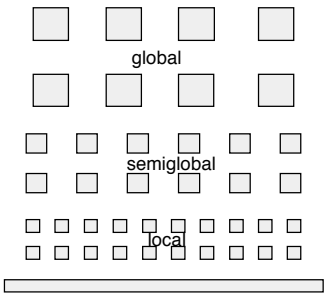
Current Al technology
(Courtesy of Motorola)



Current Cu technology
(Courtesy of IBM)

Why Cu and Low-k Dielectrics?





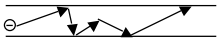
Source: Y.Nishi

Reduced resistivity and dielectric constant results in reduction in number of metal layers as more wires can by placed in lower levels of metal layers.

Cu Resistivity: Effect of Line Width Scaling

- Effect of Cu diffusion Barrier
  - Barriers have higher resistivity
  - Barriers can't be scaled below a minimum thickness

- Effect of Electron Scattering
  - Reduced mobility as dimensions decrease



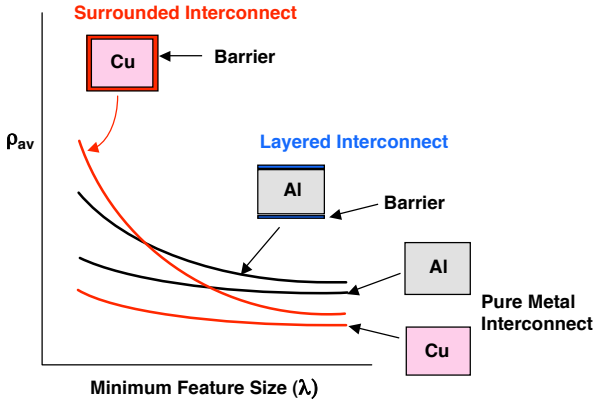
- Effect of Higher Frequencies
  - Carriers confined to outer skin increasing resistivity

Problem is worse than anticipated in the ITRS 1999 roadmap

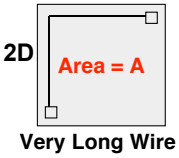
Problems in Scaling of Interconnections

AS λ DECREASES

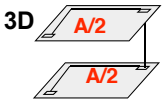
- **Resistivity increases as grain size decreases**
- **Resistivity increases as main conductor size decreases but not the surroundingbarrier size**



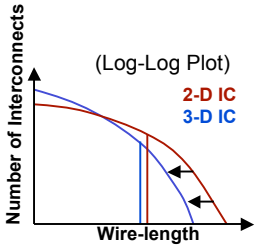
3-D Integration: Motivation

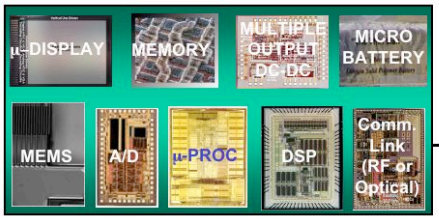


Very Long Wire

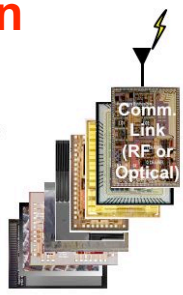


Shorter Wire





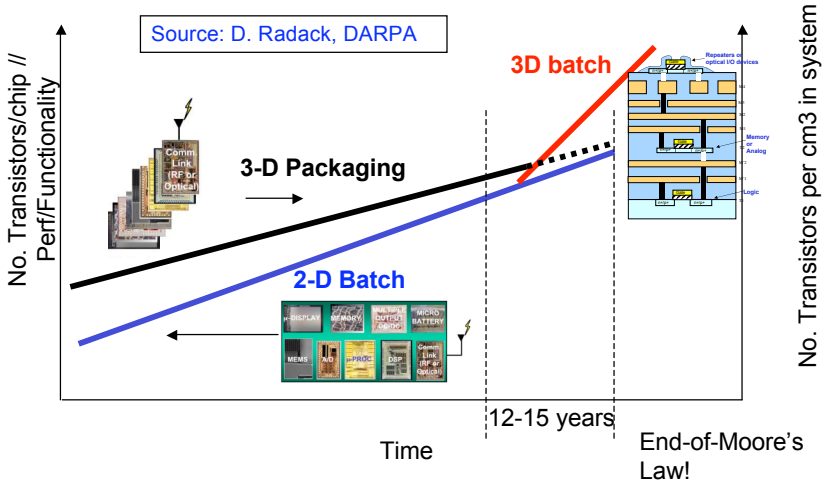
2-D System



3-D System

- Integration of heterogeneous technologies possible, e.g., memory & logic, optical I/O
- Reduce Chip footprint
- Replace long horizontal wires by short vertical wires
- Interconnect length ⇓ and therefore R, L, C ⇓
  - Power reduction
  - Delay reduction

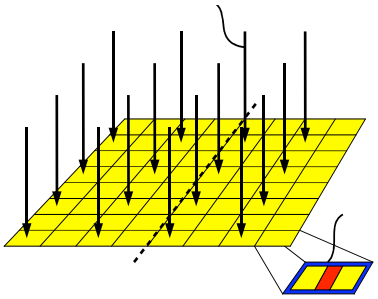
3-D Motivation: Integration Density



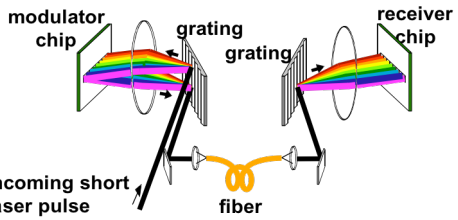
The Best Integrators of Electronic Devices Will Own the Heart of Every System – We have <15 Years to Figure it out

Can Optical Interconnects help?

On-Chip Optical Interconnects



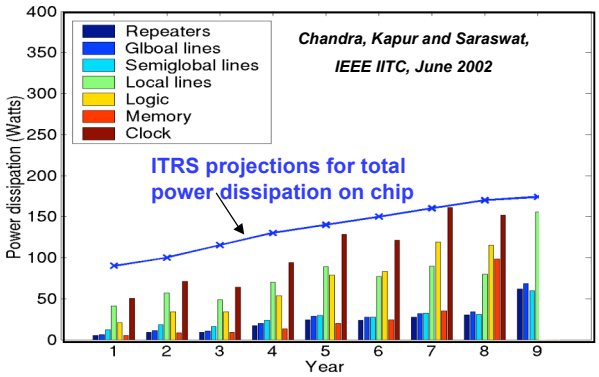
Chip-to-chip Optical Interconnects



Can potentially address many problems of Cu/low-k wires

- **On-Chip Links**
  - Reduce delay
- **Clocking and Synchronization**
  - Reduce jitter and skew
- **High Bandwidth off-chip Links**
- **Reduce power**

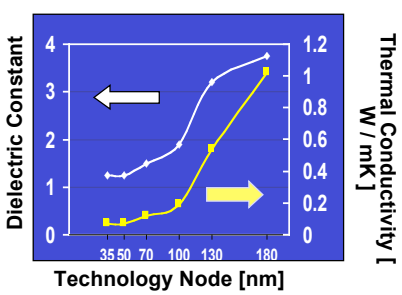
Result: scaling of power components

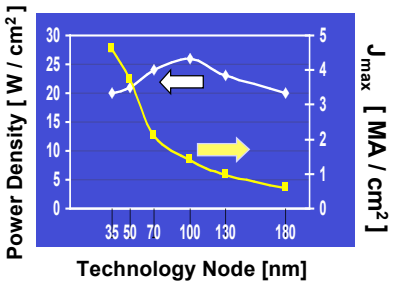


- Dynamic Power: CV^2f
- Leakage power: devices
- Short circuit power during switching
- Static power, e.g., analog components (sense amps etc.)

Power increasingly becoming the performance bottleneck for high-end microprocessors

Thermal Behavior in ICs

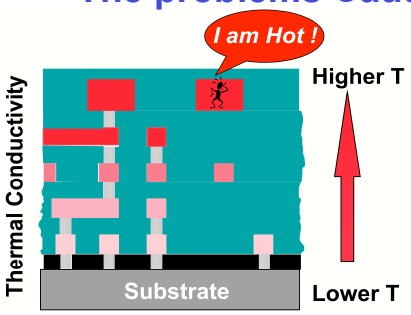




- Thermal conductivity of low-k insulators is poor
- Thermal impedance increases
- Energy dissipated (CV^2f) is increasing as performance improves
- Average chip temperature is rising

Source: ITRS

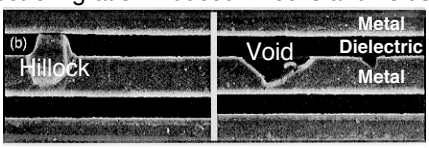
The problems Caused by Increased Power



>100A will flow on these wires

RELIABILITY

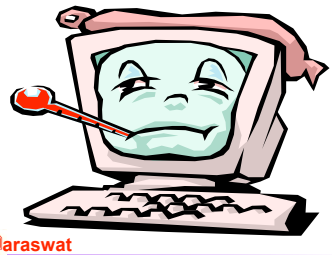
Electromigration induced hillocks and voids



Mean time to failure

$$MTF = \frac{A}{r^m J^n} \exp\left(\frac{E_a}{kT}\right)$$

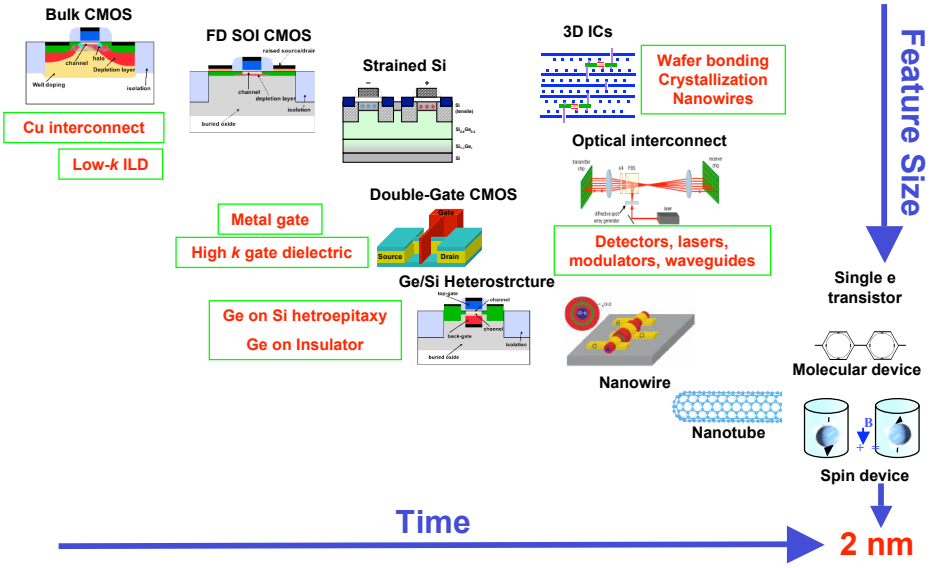
10˚C ⇑ , MTF ⇓ 50%



PERFORMANCE

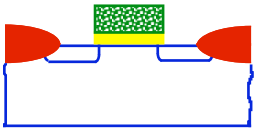
As T ⇑ R ⇑, RC delay ⇑
10˚C ⇑ , Speed ⇓ 5%

Conclusion: Technology Progression



Summary

MOS Transistor in 2010

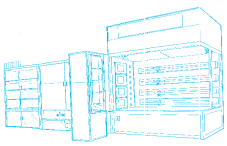


Gate oxide thickness < 1nm
Channel Length < 2nm
Junction depth < 1-2nm
Size of an atom ~ 5 Å

A Circuit in 2010

10^{10} components
Integrated digital, analog, sensors

A Factory in 2010



Approaching $10 billion

Questions we are trying to answer

- How can we continue the Moore's law
- What will be new materials, devices, circuits, sensors, equipment, simulators, etc.
- How will we design them?
- How will we manufacture them?