Introduction to Deep Submicron CMOS Device Technology

Alvin Loke
alvin.loke@agilent.com
Agilent Technologies

CSU EE571 VLSI System Design
Fort Collins, CO
May 3, 2005

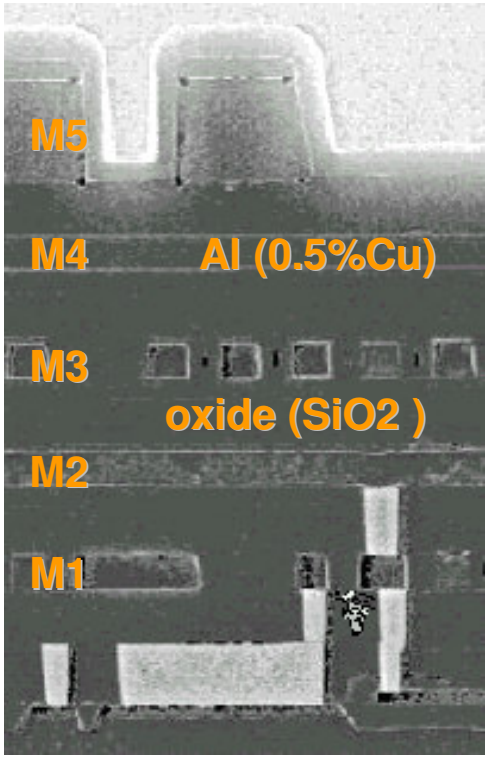
Outline

- CMOS Technology Trends
- Lithography
- Deep Submicron FET Fabrication Sequence
- Enabling Device & Equipment Technologies
- Impact on Circuit Design
- Dealing with Manufacturing Variations
- Conclusions

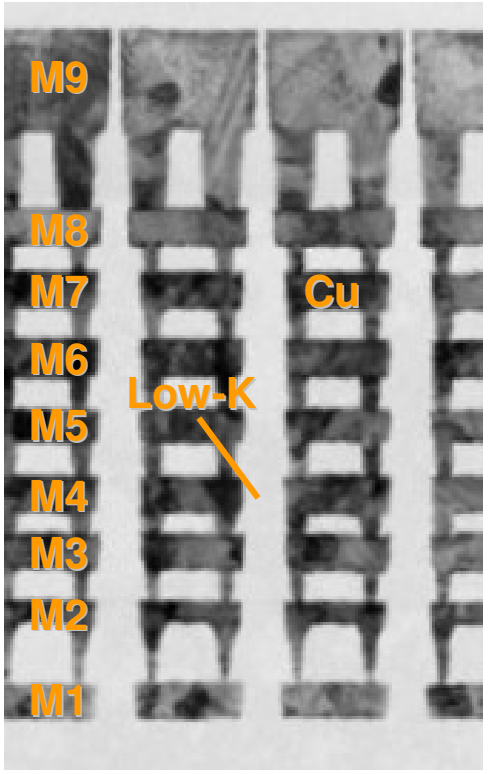
Evolution of IC Technology

I can see Waldo, but where's the transistor?

1st Fabricated IC
(Texas Instruments, 1958)



0.25µm Technology
(Motorola, 1996)

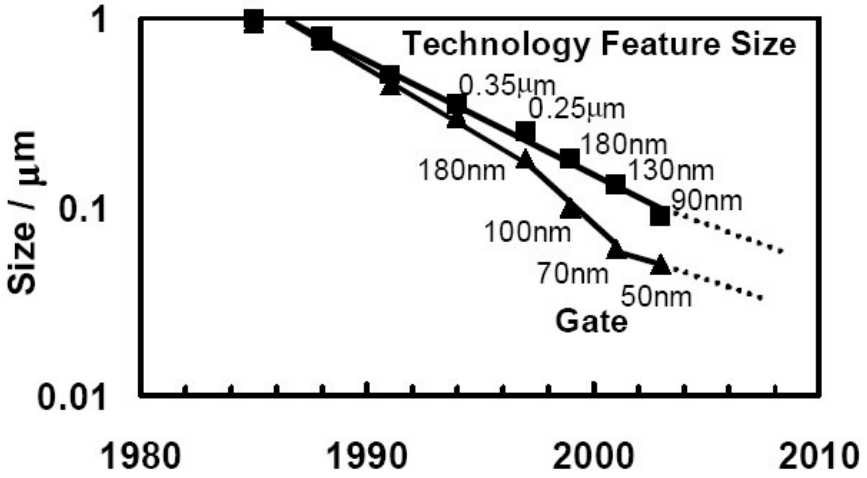


90nm Technology
(TSMC, 2002)

Is the transistor getting less important?

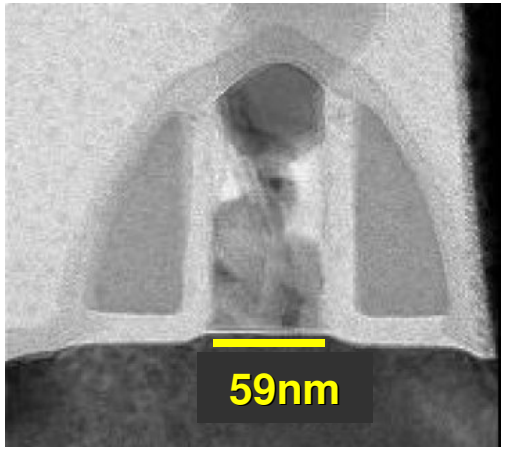
CMOS Scaling is Alive & Well
Transistors Are Picking Up The Slack

- Where are we now?
  - 130nm is standard fare
  - 90nm already in volume manufacturing
  - 65nm pilot production in progress
- Key trends:
  - Gate CD* scaling is more aggressive than interconnect scaling
  - Scaling driven by exclusively by digital circuit needs



Source: Thompson *et al.*, Intel (2002)

90nm Technology



Source: Wu *et al.*, TSMC (2002)

* CD = critical dimension

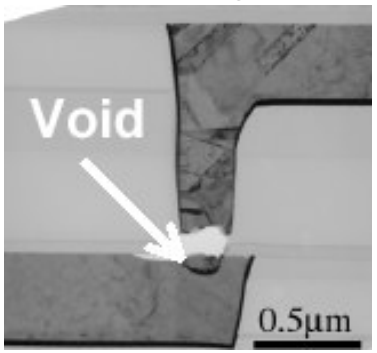
Why Aggressive FET Scaling?

$$t_{delay} \approx \frac{C_{load}\,\Delta V}{I_{FET}}$$

- The road to higher digital performance
  - $C_{load}\downarrow$ → reduce parasitics (largely dominated by interconnect now)
  - $\Delta V \downarrow$ → reduce V_{DD} or logic swing, need for core & I/O FET's
  - $I_{FET} \uparrow$ → all about moving charge quickly
- Hiccups along the way
  - Interconnect scaling much more difficult than anticipated, especially Cu/low-κ reliability
  - FET leakage doesn't go well with V_{DD} scaling
- How to beef up I_{FET}?
  - Tweak with μ, C_{ox}, L & V_T

$$I_{dsat} \approx \tfrac{1}{2}\,\mu C_{ox}\,(W/L)\,(V_{GS} - V_T)^2$$

Stress-Induced Voiding



Got redundant vias?

- Technology upgrades not necessarily compatible with analog design

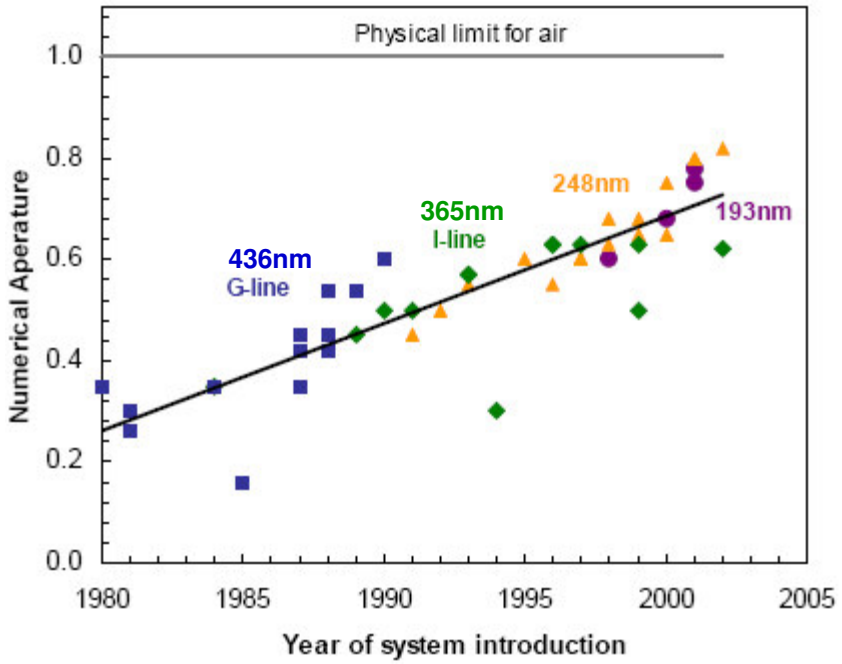
Why Should Designers Care So Much About Understanding Technology?

- Technology dictates performance limitations
  - Layout-related aspects of design has become even more critical
  - Statistical design considerations have become necessary for key circuits to yield
  - Technology-related surprises keep showing up
- Implementation of accurate models is ***ALWAYS*** late
  - Get to live with new technology-related effects & issues before they get accounted for
  - New model implementations not necessarily accurate or reliable
  - Meanwhile, learn to either mitigate or exploit these effects
- Doesn't solving these new challenges make life that much more exciting, at the very least provide a little bit more job security?

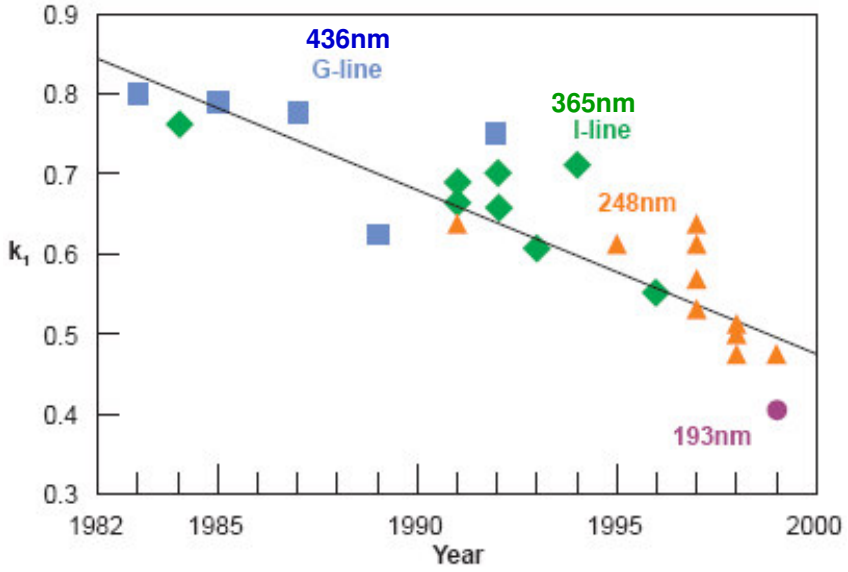
Lithography Trends
What 1µm Barrier???

$$\text{Resolution} = \frac{k_1 \lambda}{NA}$$

- Refractive projection (4X) optics
- More aggressive CD's → shorter λ
- Higher NA lenses → $$$
- Larger field sizes → $$$
- Now 25% of total wafer cost



$k_1 = f$ (resist quality, resolution tricks)



Source: ICKnowledge.com (2003)

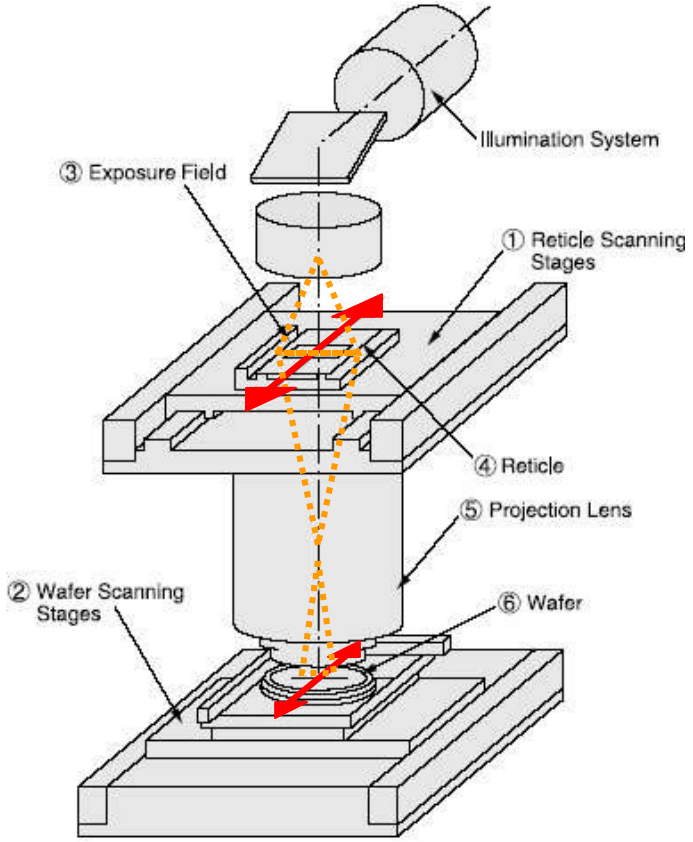
Step-and-Scan Projection Lithography
Beyond 0.35µm Technology

- Slide both reticle & wafer across narrow slit of light
- Aberration-free high-NA optics only required along 1-D but now requires high-precision constant-velocity stages
- Still much cheaper than high-NA optics optimized in 2-D
- 6” x 6” physical reticle size (4X reduction)
- 25 x 33mm or 26 x 32mm field size
  → shorter edge limited by slit width
- Relatively weak intensity of deep-UV source required development of very sensitive *chemically-amplified* resists for throughput

Deep-UV *Slit* Source

Excimer Laser

KrF (248nm) or ArF (193nm)



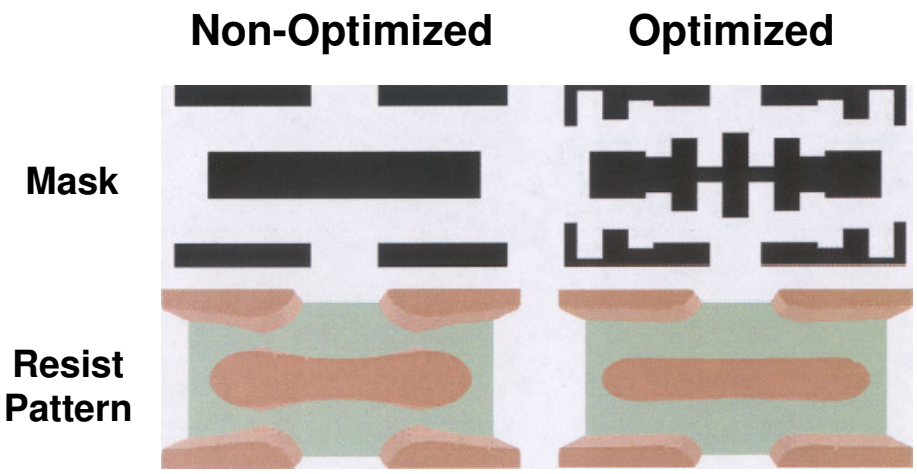
Source: Nikon

Key Resolution Enhancement Tricks

- Sharp features (e.g., corners) are lost because diffraction attenuates & distorts higher spatial frequencies (low-pass optical filtering)
- Compensate for diffraction effects for features much smaller than exposure λ → manage sub-λ constructive & destructive interference
- Software complexity during mask fabrication

Optical Proximity Correction (OPC)

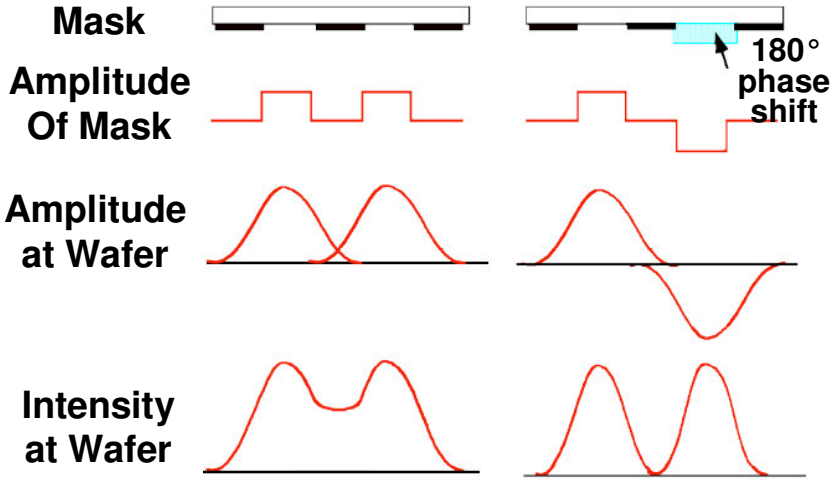
- **Add scattering features to sharpen corners**
- **Used extensively for poly gate definition**



Source: Socha, ASML (2004)

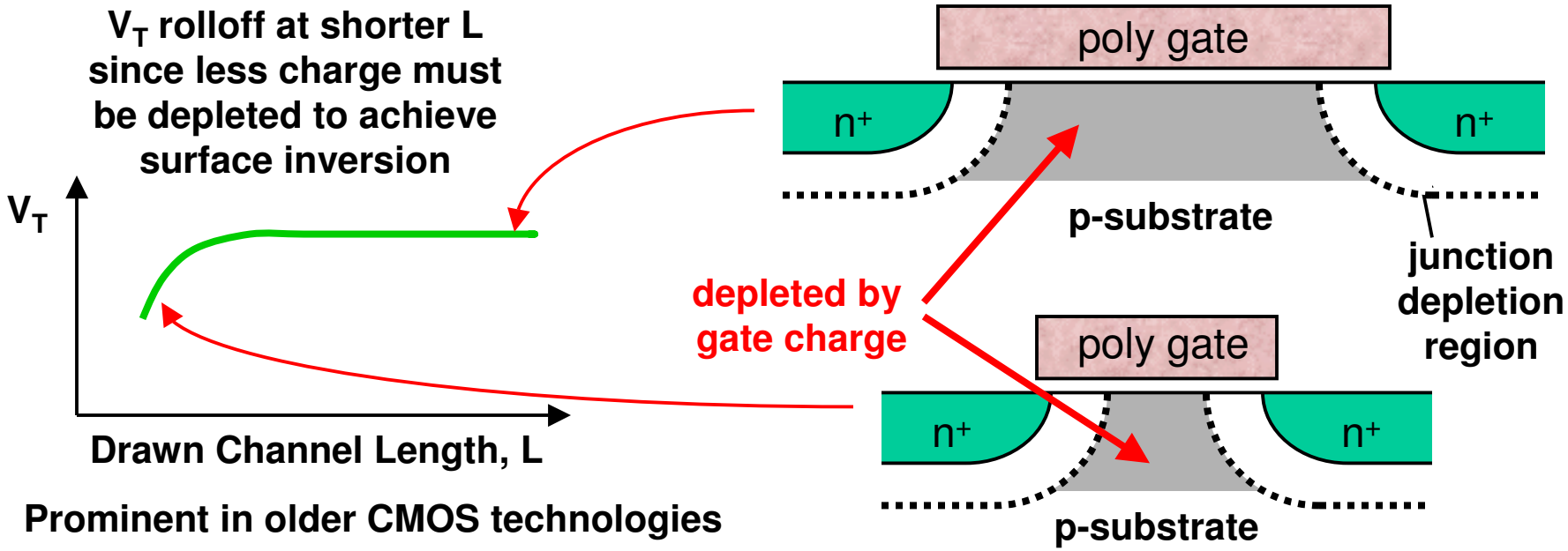
Phase Shift Masking (PSM)

- **Modulate optical path through mask**
- **Used extensively for contacts & vias**
- **Complicated for irregular patterns**



Source: Plummer, Stanford (2004)

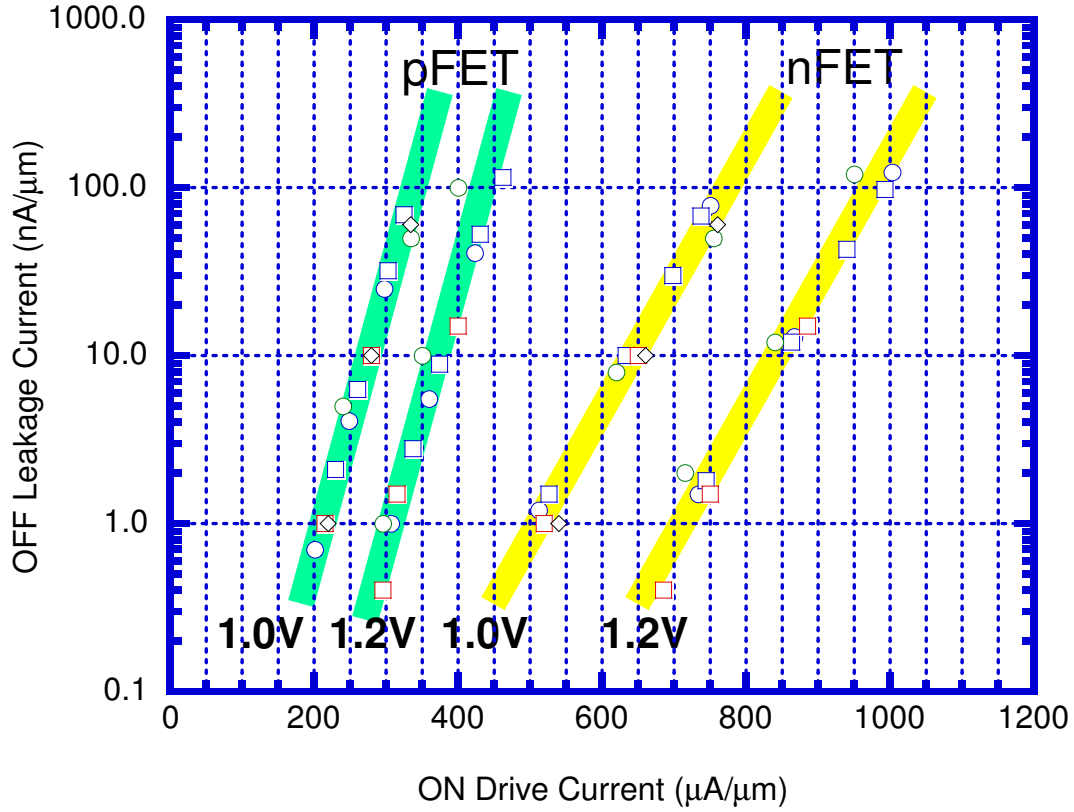
Short-Channel Effect (SCE)



- Prominent in older CMOS technologies
- How to minimize SCE?
  - Minimize volume of charge depleted by source/drain junctions
    - Higher substrate doping for thinner junction depletion regions ($x_{dep} \propto 1/\sqrt{N}$)
      - Higher V_T & junction capacitance → not consistent with scaling
    - Shallower source/drain junctions
      - Higher source/drain resistance → smaller drive currents
  - Tighter gate coupling to surface potential
    - Thinner gate oxide of surface potential → direct tunneling leakage
    - Higher K gate dielectrics
- Other SCE problems: large electric fields → carrier v_{sat} & μ degradation

ON vs. OFF Current Benchmark

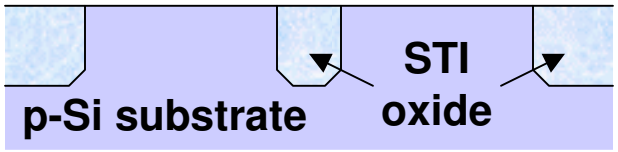
Comparison of 90nm Technology Foundry Vendors



- "No free lunch" principle prevails again: high I_{ON} → high I_{OFF}
- V_T's not scaling as aggressively as V_{DD}
- Technology providers offer variety of V_T's on same die to concurrently meet high-speed vs. low-leakage needs

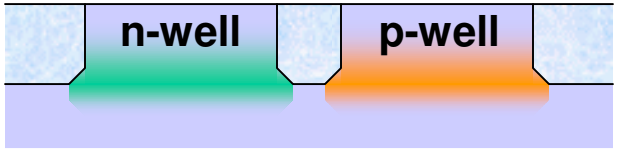
Deep Submicron FET Fabrication Sequence

1



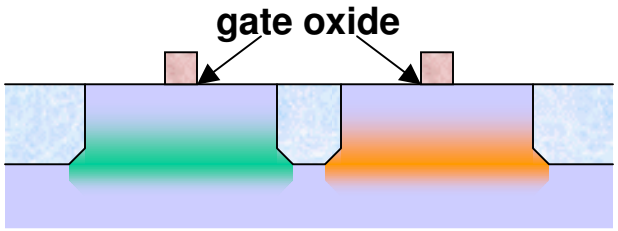
Shallow Trench Isolation

2



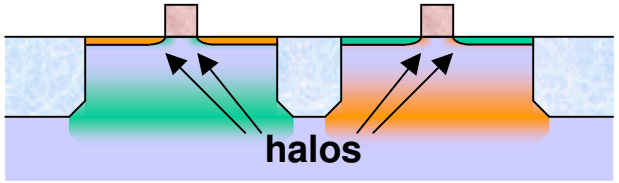
Well Implantation

3



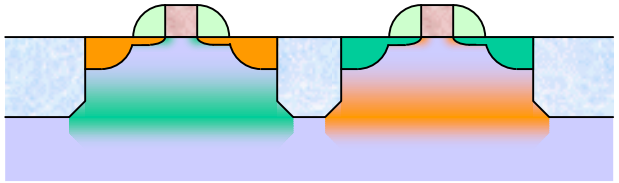
Gate Oxidation &
Poly Definition

4



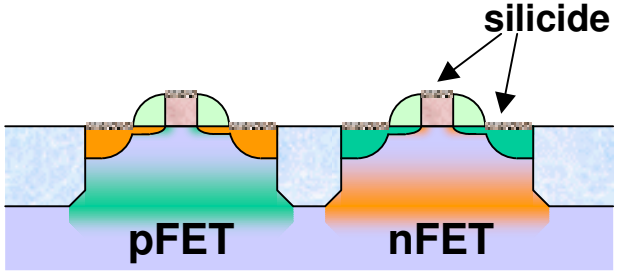
Source/Drain Extension
& Halo Implantation

5



Spacer Formation &
Source/Drain Implantation

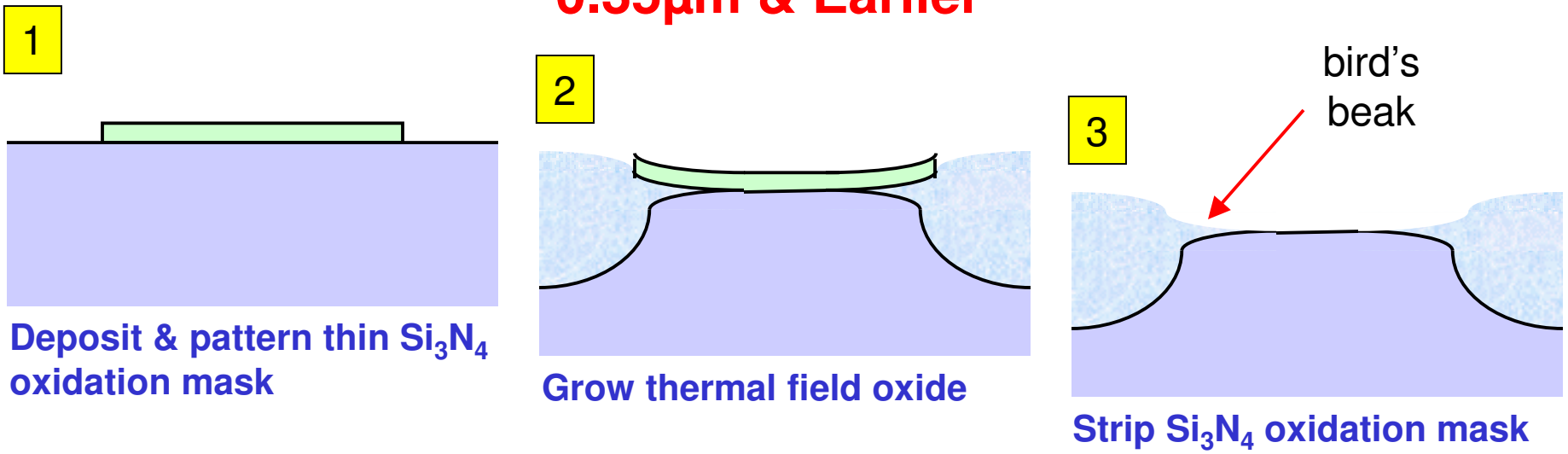
6



Salicidation

Local Oxidation of Silicon (LOCOS) Isolation

0.35µm & Earlier



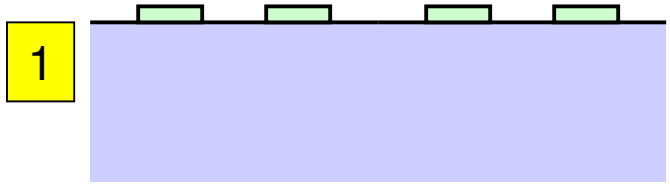
- Industry played lots of tricks to reduce width of bird's beak & make field oxide coplanar with active areas

$$\text{Depth of Focus} \propto \text{Resolution} / \text{NA}$$

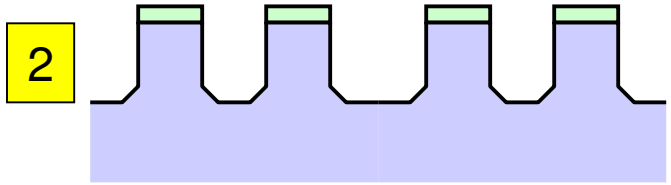
- Required very careful understanding of visco-elastic properties of oxide during thermal oxidation
- LOCOS ran out of gas beyond 0.25µm

Shallow Trench Isolation (STI)

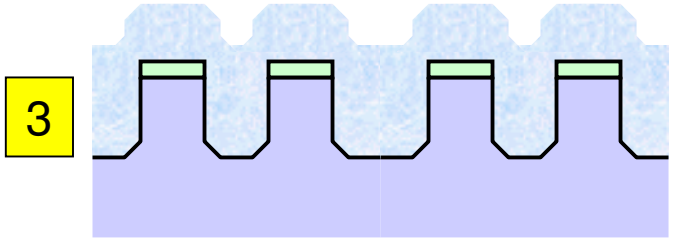
0.25 µm & Beyond



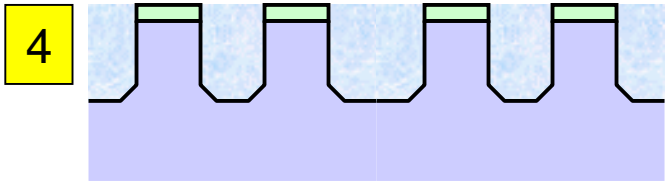
Deposit & pattern thin Si_3N_4 etch mask & polish stop



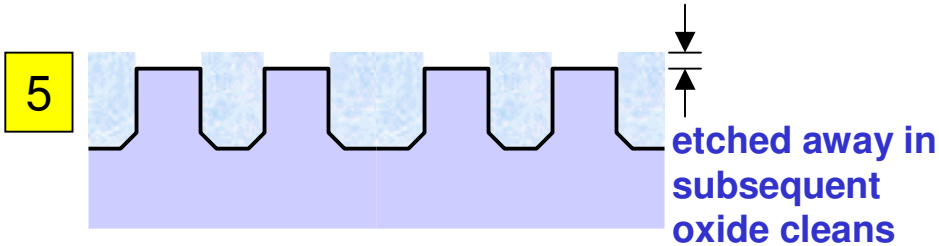
Etch silicon around active area – profile critical to minimize stress



Grow liner SiO_2, then deposit conformal SiO_2 – void-free deposition is critical



CMP excess SiO_2

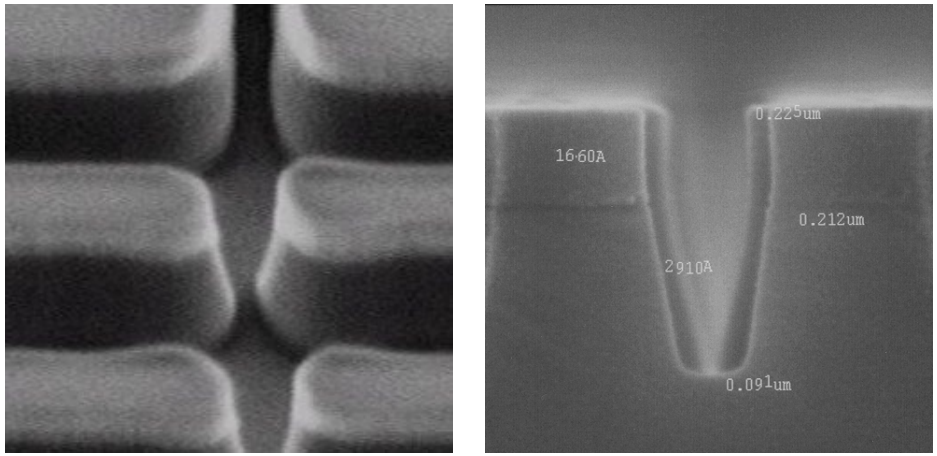


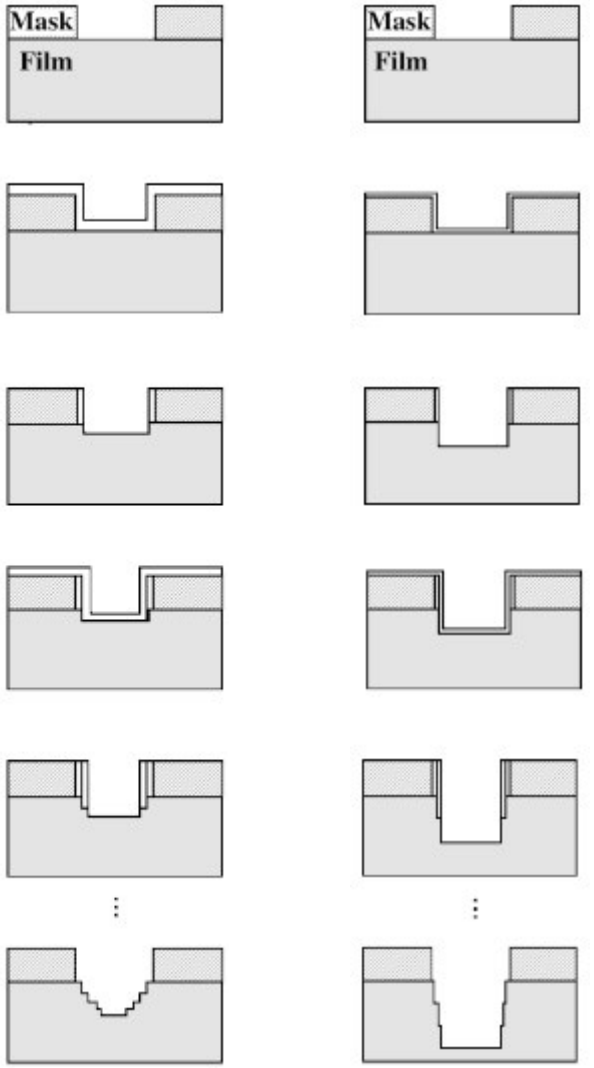
Strip Si_3N_4 polish stop

Advantages over LOCOS technologies

- Reduced active-to-active spacing (no bird's beak)
- Planar surface for gate lithography

STI Etch Profile Control





Reactive ion plasma etching (RIE)

- Etching ions vertically bombard surface of material to be removed
- Very directional
- High sputtering component causes resist to re-deposit inside trench
- Tune etch gas chemistry to keep or remove redeposited inhibitor formation (micromasking)

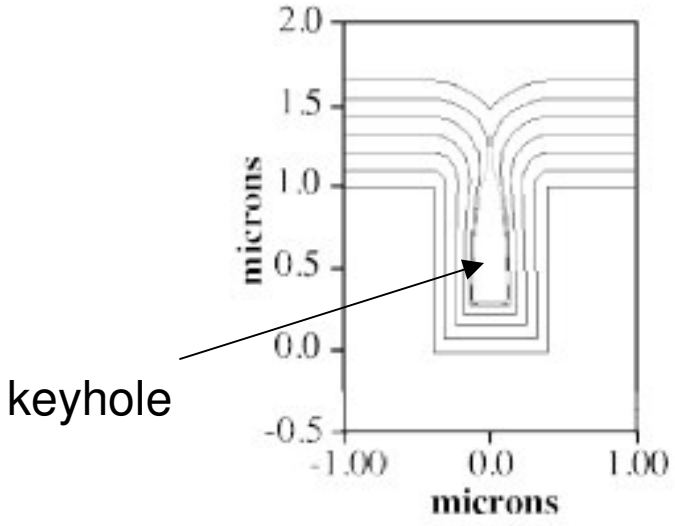
Source: Plummer, Stanford (2004)

STI HDP-CVD Oxide Gapfill

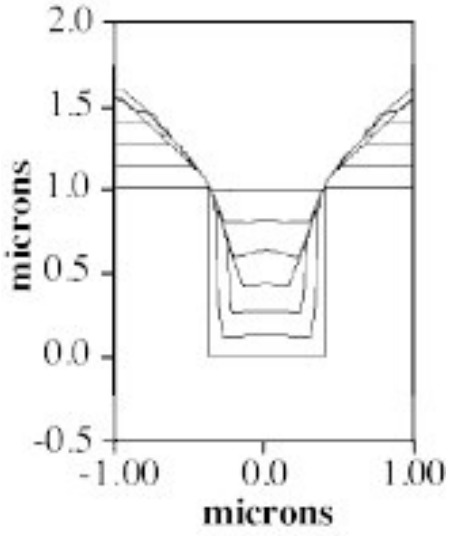
High-Density Plasma (HDP) Chemical-Vapor Deposition (CVD)

- Fills aggressive aspect ratios by cycling between deposition & etch
- Re-sputtering nature results in very compressive oxide films

Conformal CVD

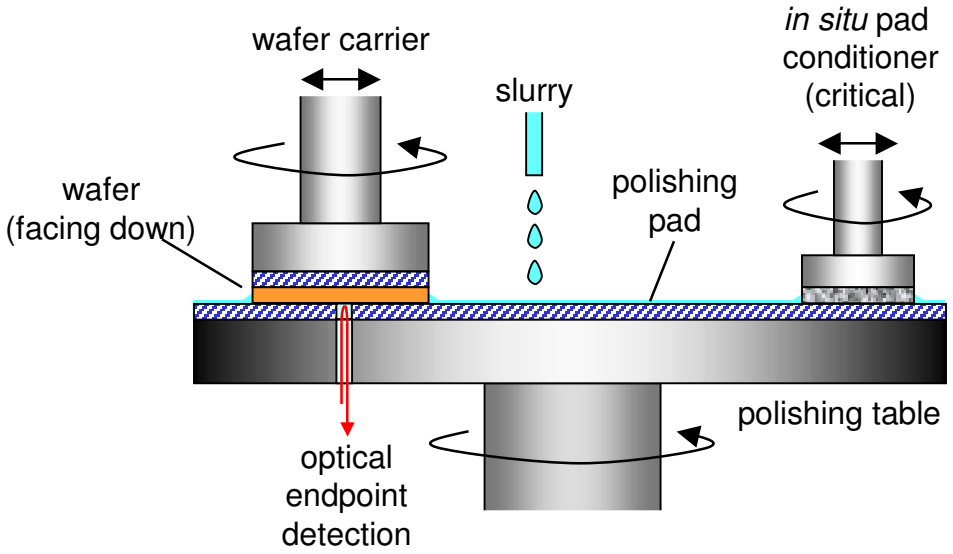


HDP CVD



Source: Plummer, Stanford (2004)

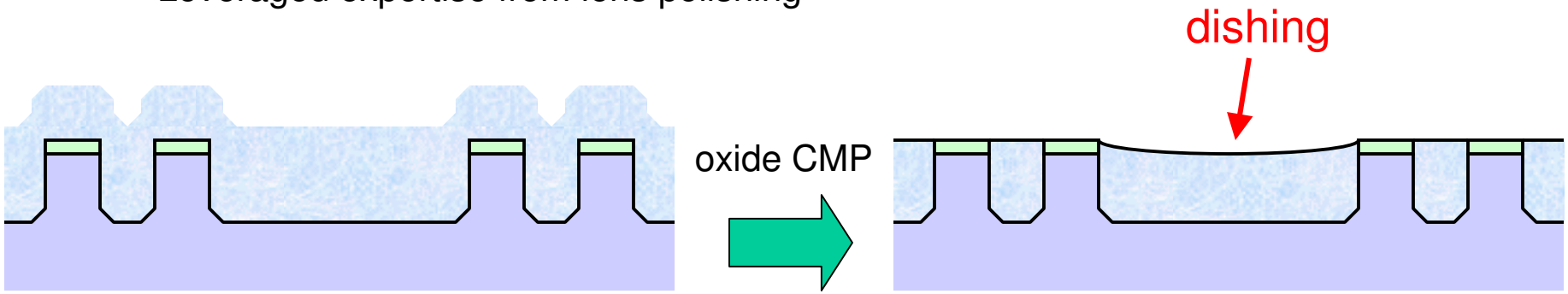
Let's Think a Little Bit More About CMP



CMP technology pioneered by IBM

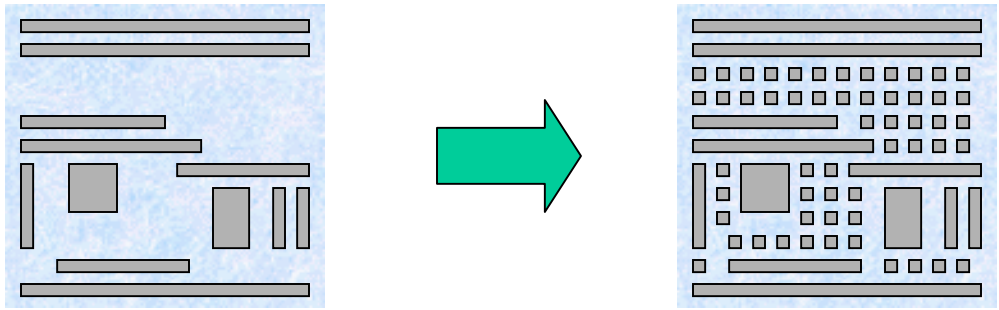
- Leveraged expertise from lens polishing

- Ideal world for CMP: want *perfect* periodicity of patterns throughout wafer
- Need to throw in dummy features to minimize pattern density variations → optimize planarity
- Polishing pad will flex

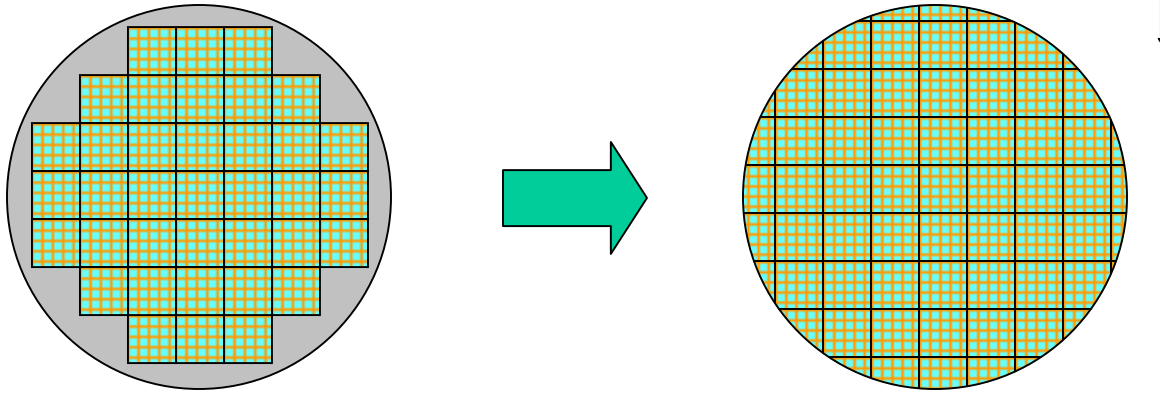


Always Think Dummies in Any CMP Process

- Dummification is key to minimize topography in any CMP process
- Add dummy patterns to open spaces to minimize layout density variations
  → Added design complexity to check layout density & insert dummy patterns

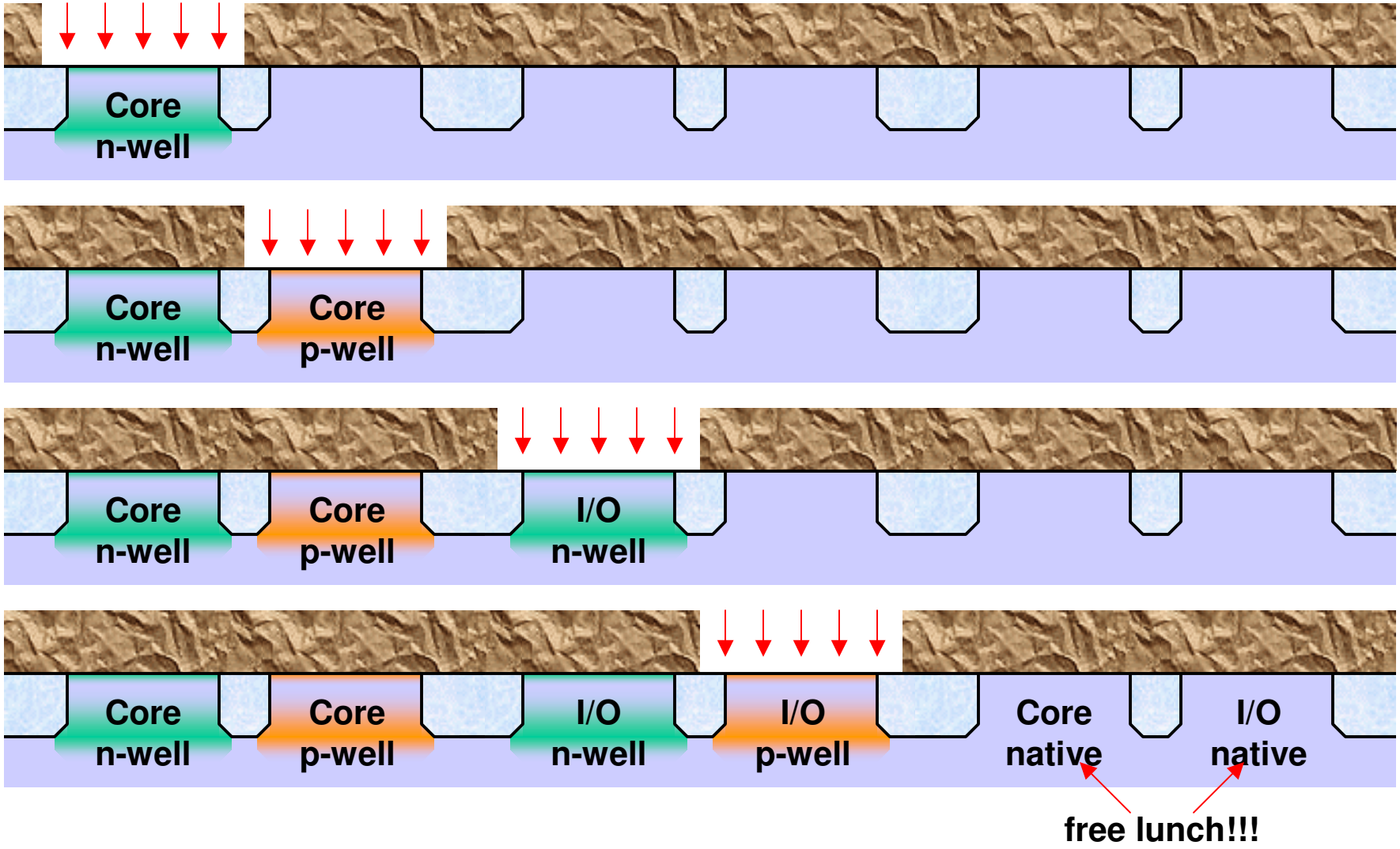


- Also critical to step dummy dies along wafer circumference



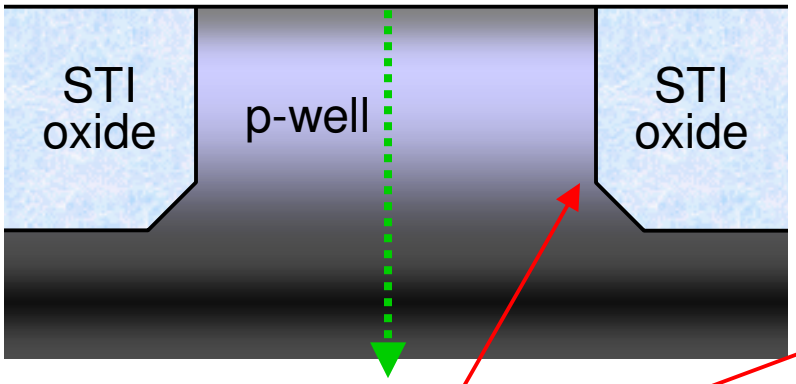
Well Implants – Lots of Transistor Variants

- core vs. I/O FET's, core low-/nom-/high-V_T variants, native vs. implanted

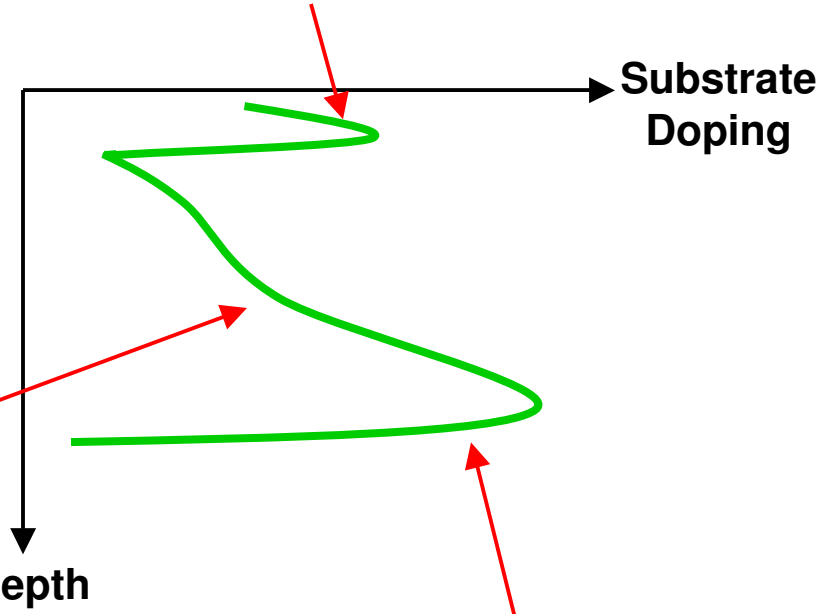


Well Engineering

Retrograded well dopant profile
(implants before poly deposition)



Shallow & steep surface channel implant
- V_T control
- Slow diffusers critical (Ga, Sb)



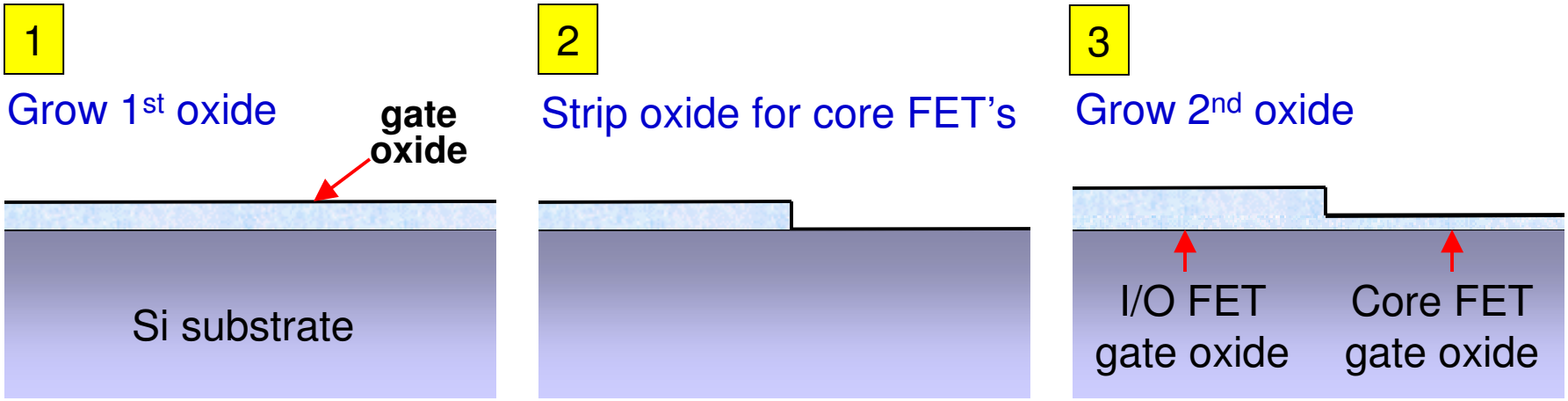
Deeper subsurface implant
- Extra dopants to prevent subsurface punchthrough under halos
- Prevent parasitic channel formation on active sidewall beneath source/drain
- Faster diffusers OK (B, As/P)

Very deep high-dose implant
- Latchup prevention
- Noise immunity
- Faster diffusers OK (B, As/P)

Implant order matters to prevent ion channeling, especially for the shallow implant

Gate Oxidation

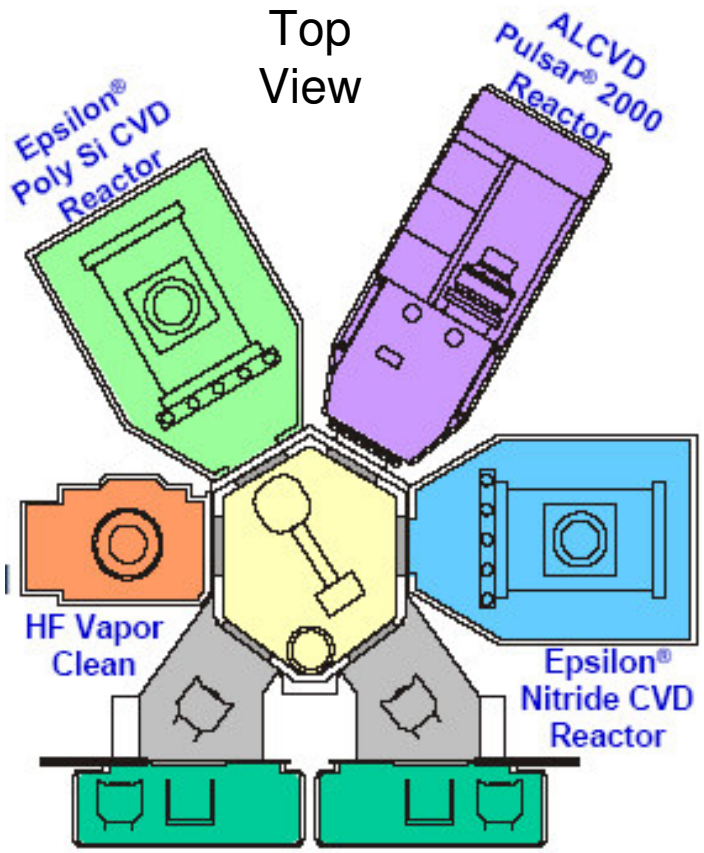
- Need two gate oxide t_{ox}'s – thin for core FET's & thick for I/O FET's



- Oxide is grown, not deposited
  - Need high-quality Si-SiO_2 interface with low Q_f & D_{it}
- Gate oxide is really made of silicon oxynitride (SiO_xN_y)
  - N content prevents boron penetration from p+ poly to channel in pFET's
  - Improves GOI (gate oxide integrity) reliability
  - Side benefit – increased ε_{ox}
- TSMC now offers triple gate oxide (TGO) processes with two I/O FET varieties

Aside on Equipment Technology Evolution

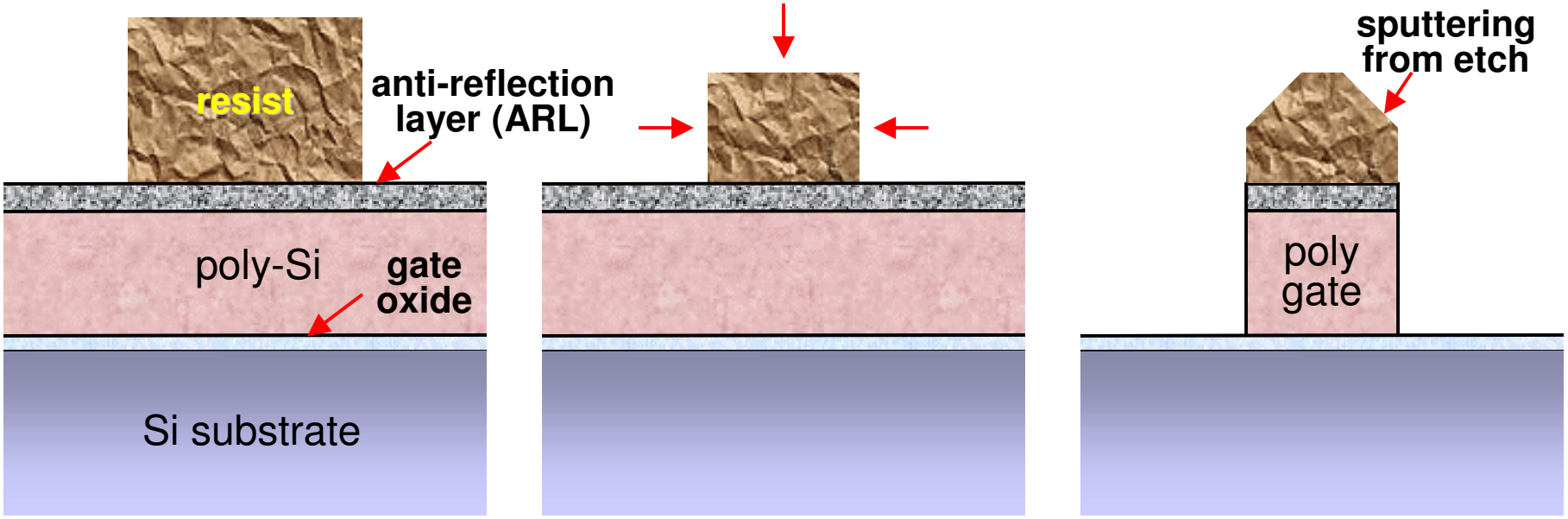
- Gate oxide no longer furnace grown
- Multi-chamber *cluster tools* now ubiquitous
- Pre-oxidation clean, gate oxidation & poly/ARL deposition performed in separate chambers without breaking vacuum
- Better thickness & film compositional control (native SiO_2 grows instantly when exposed to air)
- Fast – minutes-seconds per wafer vs. hours per wafer batch



Source: Maex, IMEC (2002)

Poly Gate Definition

- Gate CD way smaller than lithography capability, even with mask tricks



1 Pattern resist

2 Trim resist (oxygen ash)

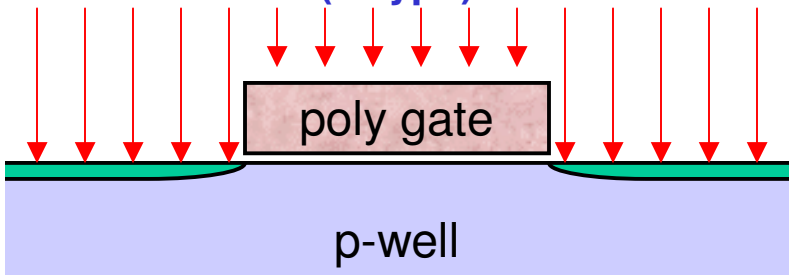
3 Etch gate stack

- *Process control is everything* – resist & poly etch chamber conditioning is critical (lesson to remember: don't clean those residues in tea cups or woks)
- Way to get smaller CD's to trim more (requires tighter control)
- Dummification also necessary for poly mask

Source/Drain & Channel Engineering

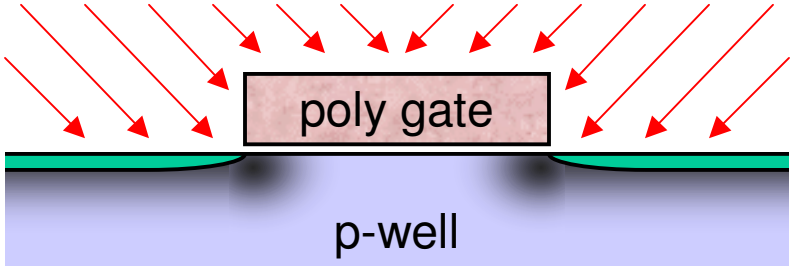
self-aligned source/drain extension implant (n-type)

1



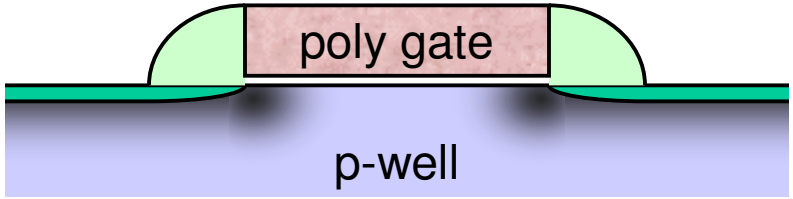
self-aligned high-tilt halo/pocket implant (p-type)

2



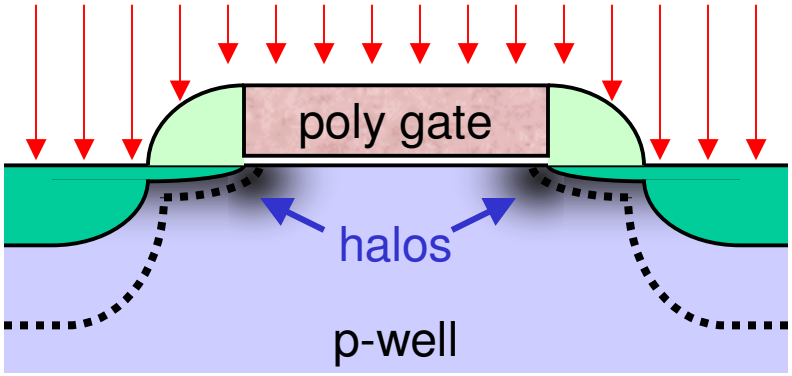
3

dielectric spacer formation



self-aligned source/drain implant (n-type)

4



Resulting structure has:

- **Smaller SCE**
- **Shallow junction where needed most**
- **Low junction capacitance**

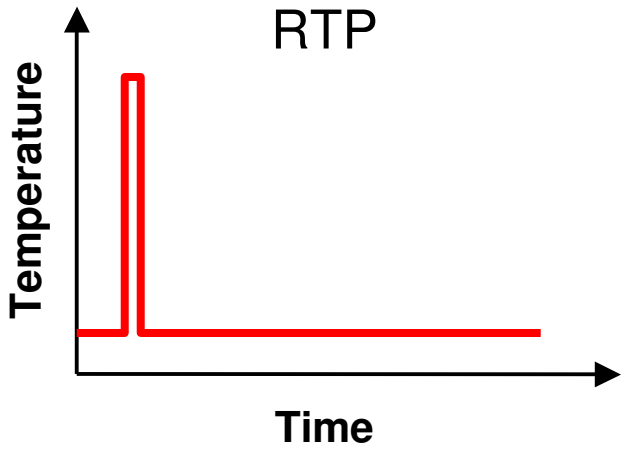
Not to be confused with LDD's in I/O FET's

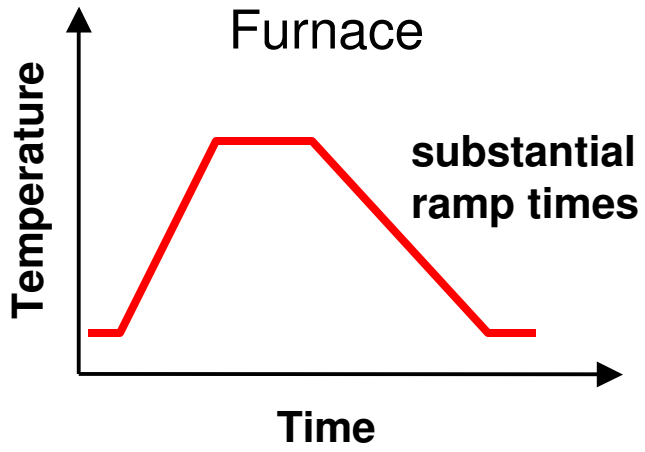
- **Same process with spacers but lightly doped drain (LDD) is used for minimizing peak *E* fields that cause hot carriers & breakdown**
- **Extensions need to be heavily doped to minimize series resistance**

Different halo & extension/LDD implants for each FET variant

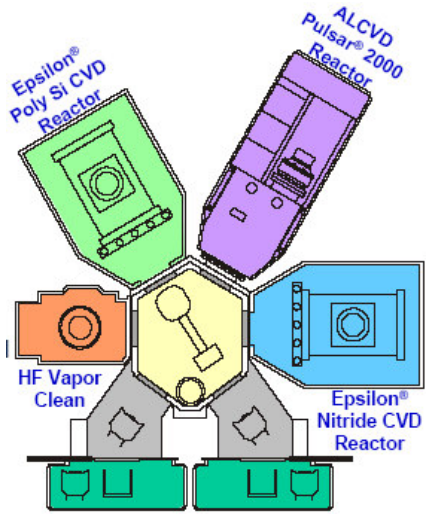
Rapid Thermal Processing (RTP)

Wasabi vs. Curry



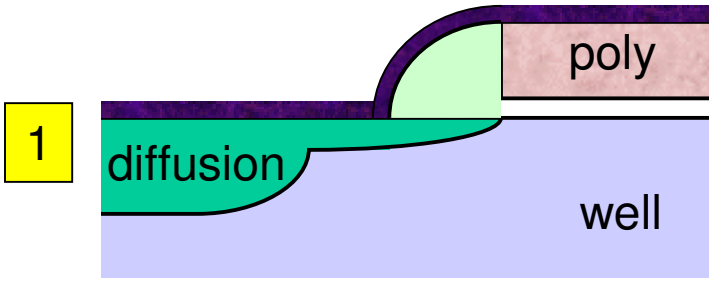


- Initially developed for short anneals
  - Impossible to control short thermal cycles in furnaces
  - Want minimum diffusion for shallow & abrupt junctions
- Process steps:
  - Annealing → repair implant damage
  - Oxidation → gate oxide
  - Nitridation → spacers, ARL
  - Poly deposition → gate
- RTP in single-wafer multi-chamber cluster tools

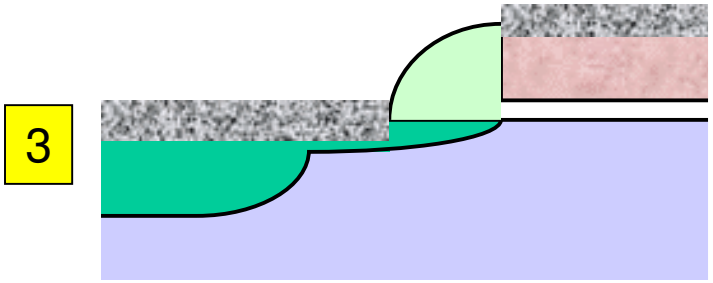


Self-Aligned Silicidation (Salicidation)

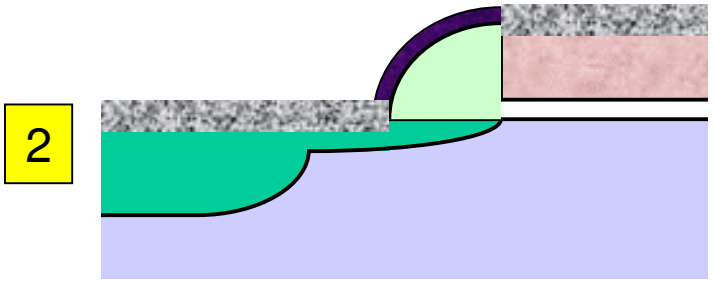
- **Need to reduce poly & diffusion R_s, otherwise get severe IFET degradation due to voltage drops from contacts to intrinsic FET (source degeneration)**



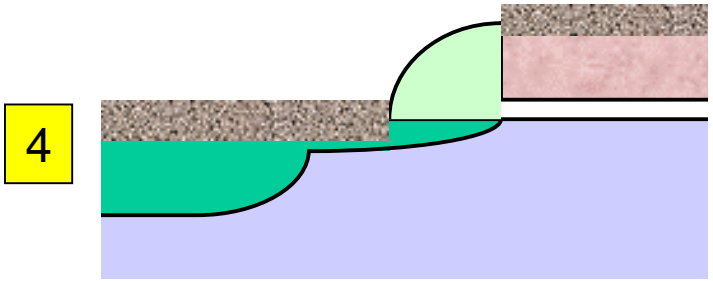
Deposit sicilide metal (Ti, Co, Ni)



Strip unreacted metal



RTA1 (low temperature)
***Selective* formation of metal silicide from direct metal reaction with Si**

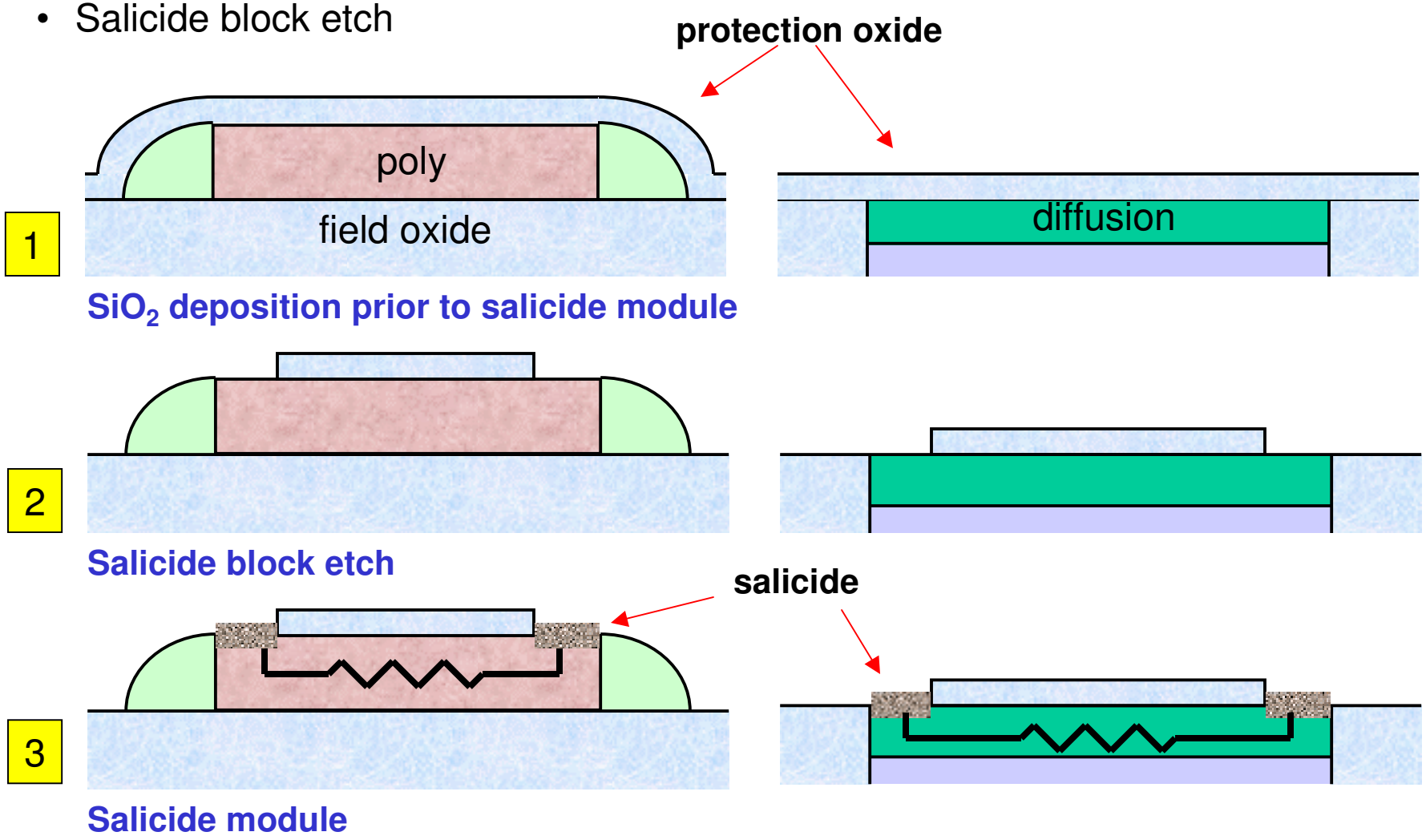


RTA2 (high temperature)
Transforms metal silicide into low-ρ phase by consuming more Si

- Selectivity degraded at RTA2, i.e., metal over SiO_2 & Si_3N_4 will form silicide
- Technology progression: $TiSi_x$ → $CoSi_x$ → $NiSi_x$
  - Scaling requires smaller silicide grain size to minimize R_s variations

Monolithic Resistors

- Only one extra mask can buy you unsalicided poly & diffusion resistors
- Salicide block etch



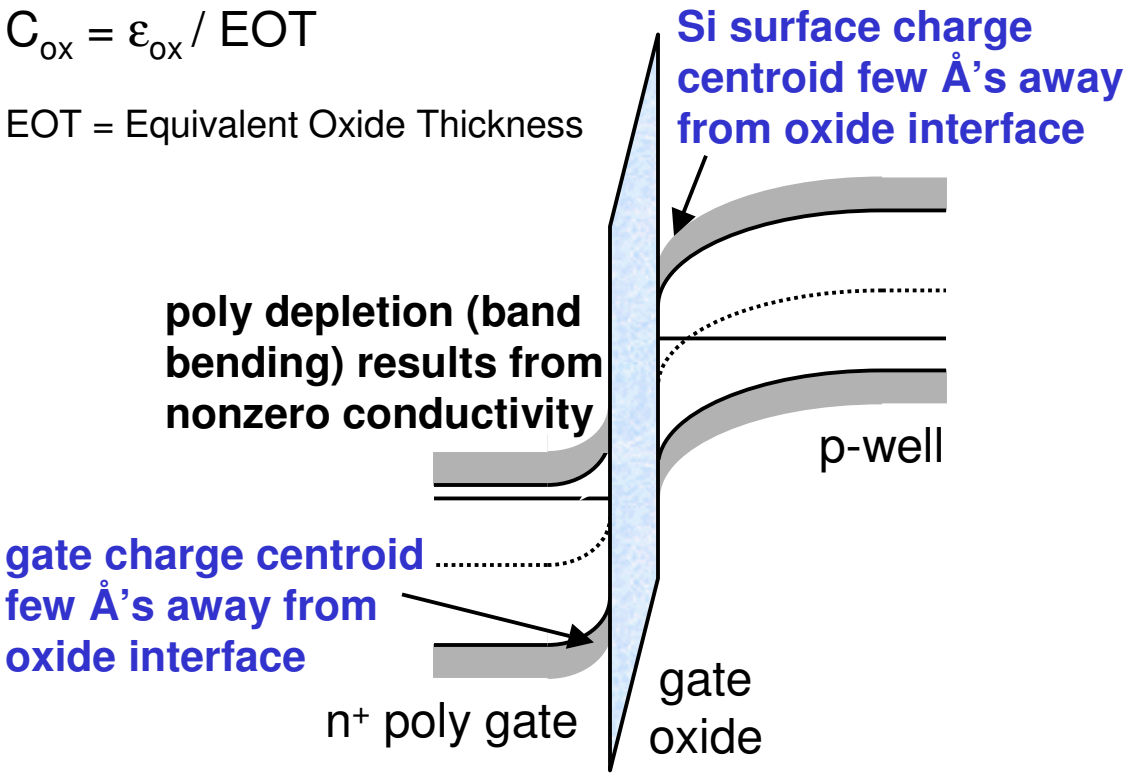
So What Do All These Cool Process Upgrades Mean To Designers?

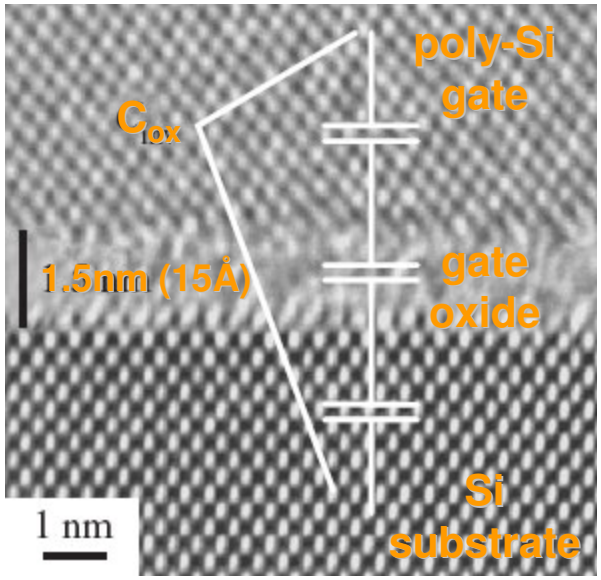
- Short answer…
  few extras like speed but plenty more non-idealities to worry about, *especially* in analog land !!!
- Examples:
  - Poly depletion & charge centroid effects
  - Gate direct tunneling
  - Impact of halos on output resistance
  - Well proximity effect
  - Active area mechanical stress effect
  - Manufacturing variations

Poly Depletion & Charge Centroid Effects

$C_{ox} = \varepsilon_{ox} / EOT$

EOT = Equivalent Oxide Thickness



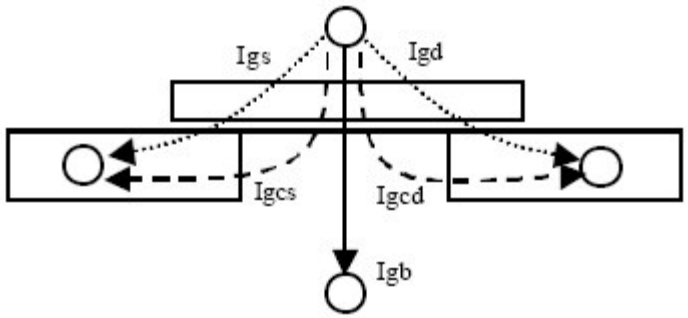


Source: Wong, IBM (2002)

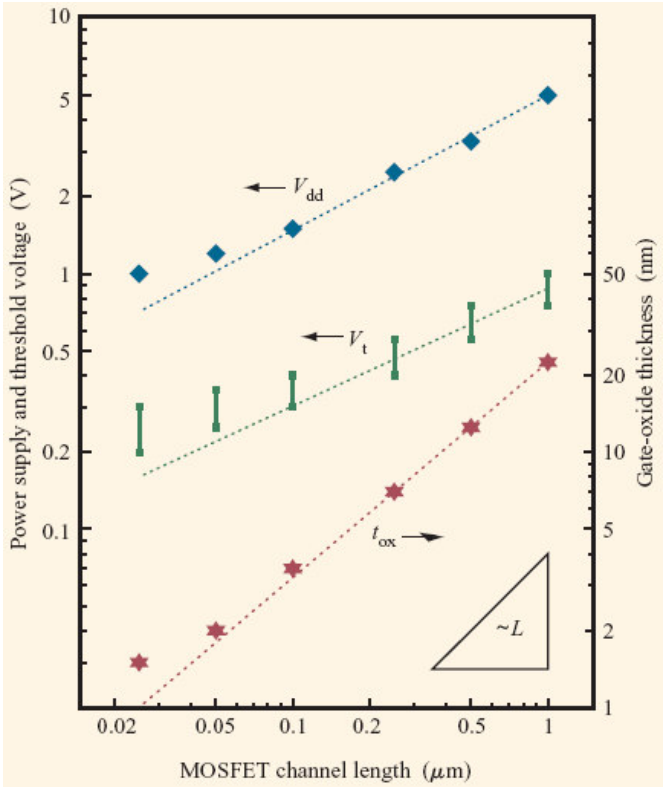
- **Increasing discrepancy between electrical & physical gate oxide thicknesses since charges are not intimately in contact with oxide interface → C_{ox} not as small**
- **C_{ox} must account for gate & Si surface charge centroids as well as ε_{Si}**
- **Modeled in BSIM4 → more accurate I-V & C-V calculations**
- **Mitigated with high-K gate dielectrics & metal/silicided gate technology**
  - ***Oh my, are we regressing to non-self-aligned Al gate technology?***

Gate Direct Tunneling Leakage

- t_{ox} has been scaling aggressively with L_{min}
  - higher I_{FET}
  - tighter gate control → less SCE
- Significant direct tunneling for t_{ox} < 2nm
- Gate leakage = $f(t_{ox}, V_G)$
  - Tunneling probability ∝ exp ($-\alpha\, t_{ox}$)
  - V_G dependence from Fermi-Dirac statistics & density-of-states considerations
- Hole & electron tunneling in CB & VB
- High-K gate dielectric achieves same C_{OX} with much thicker t_{ox}
- Modeled in BSIM4 not BSIM3



Historical Trends

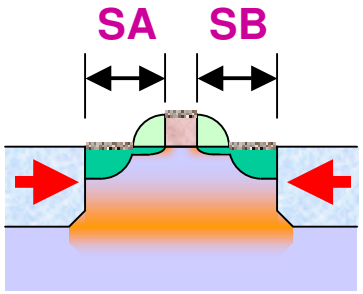


Source: Taur, IBM (2002)

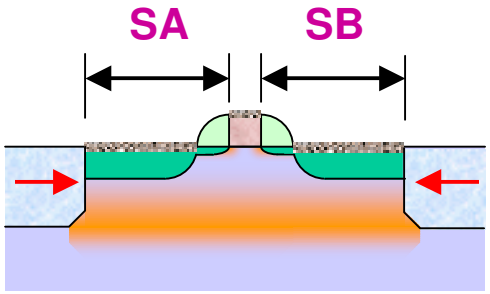
Active Area Mechanical Stress Effect

- Silicon is *piezoelectric* – electrical properties depends on mechanical stress state
  - Stress affects m*, μ, E_g, V_T, body effect, DIBL, …
  - Compression → slower nFET, faster pFET
  - Tension → faster nFET, slower pFET
- STI → compression in Si channel due to 10x CTE mismatch between Si & SiO_2
  - I_{dsat}'s can change by 10-15% → affects digital device ratios
  - Channel stress is strong function of distance from poly to active edge

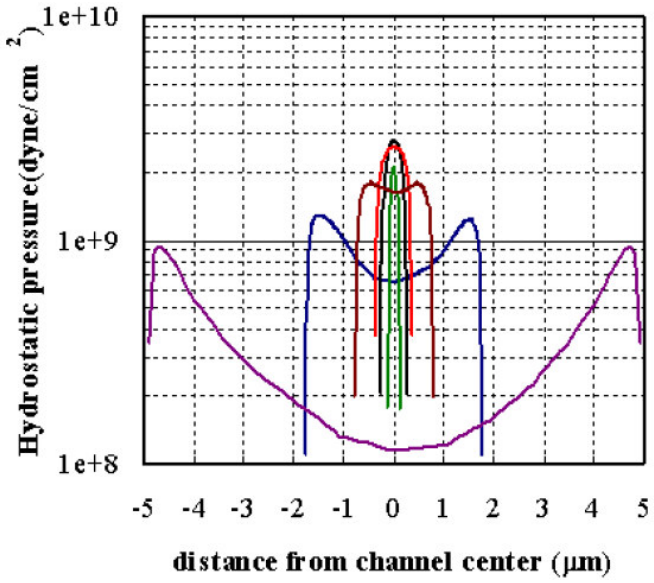
high compression



lower compression



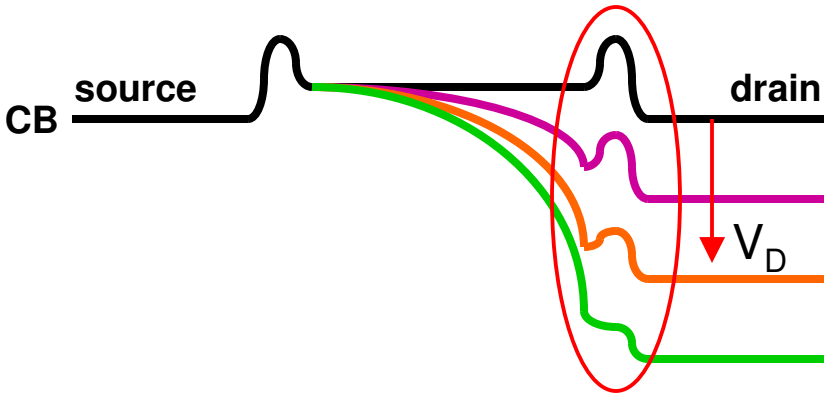
- Known for some time
  - Can play tricks with tensile spacer & silicide films to relieve channel stress
  - Basis of high-μ_n strained-Si technology
- Modeled in BSIM4 not BSIM3

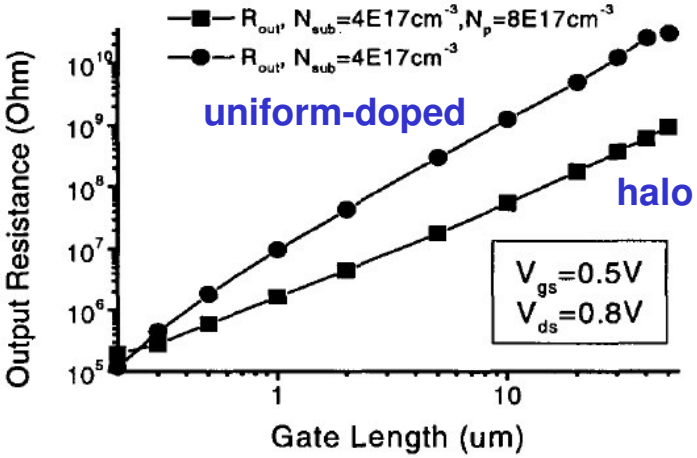


Source: Xi *et al.*, UC Berkeley (2003)

Impact of Halos on Output Resistance

- Halo at source side suppresses SCE & DIBL in *short-channel devices*
- Halo at drain side creates *Drain-Induced Threshold Shift* (DITS) in *long-channel devices*
  - Drain bias very effective in modulating drain halo barrier → $V_T \downarrow$ → $I_{ds} \uparrow$
  - Worse DIBL compared to uniform-doped FET
  - Can degrade FET r_{out} by 10-100x!!!
  - Critical limitation for building current sources (cascoding difficult with low V_{DD})
  - Asymmetric FET's with only source-side halo shown to improve r_{out} significantly



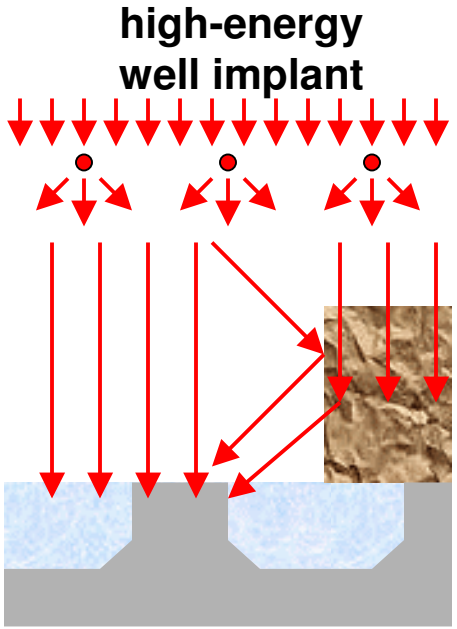


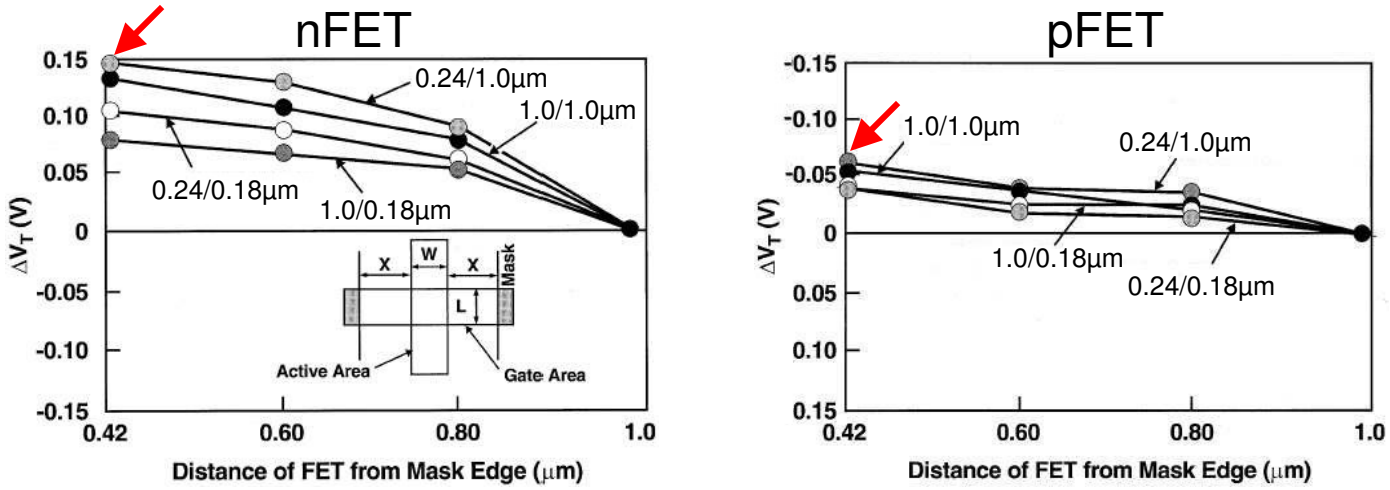
Source: Cao *et al.*, UC Berkeley (1999)

- Less degradation for devices with weaker halos, e.g., long-channel low-V_T FET's
- Modeled in BSIM4 not BSIM3, but still need improvement

Well Proximity Effect

- $|V_T| \uparrow$ if FET is too close to edge of well resist mask due to extra channel dopants from lateral scattering off resist sidewall into active area during well implants
- $|\Delta V_T|$ depends on:
  - FET distance to well mask edge
  - FET orientation
  - implanted ion species/energy
- Well mask matching critical especially in analog layout
- Not modeled in BSIM4

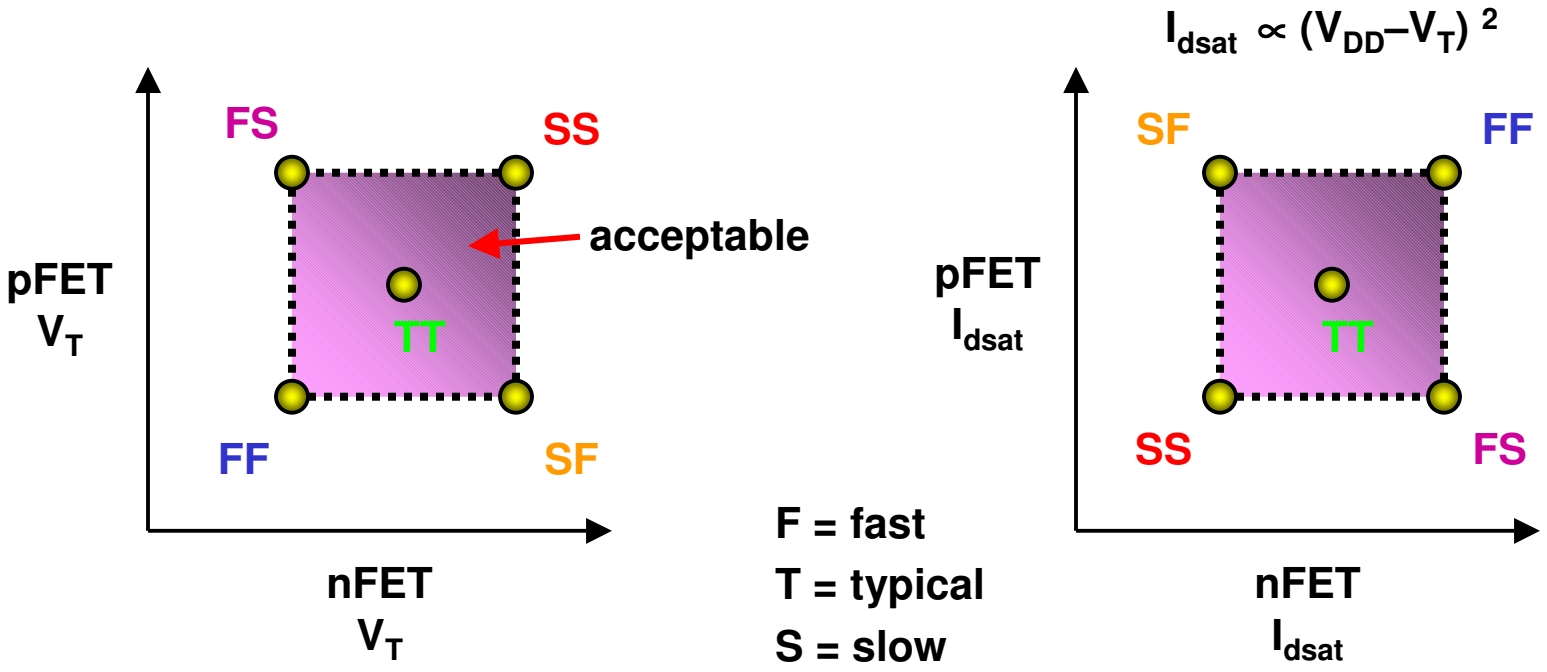




Source: Hook *et al.*, IBM (2003)

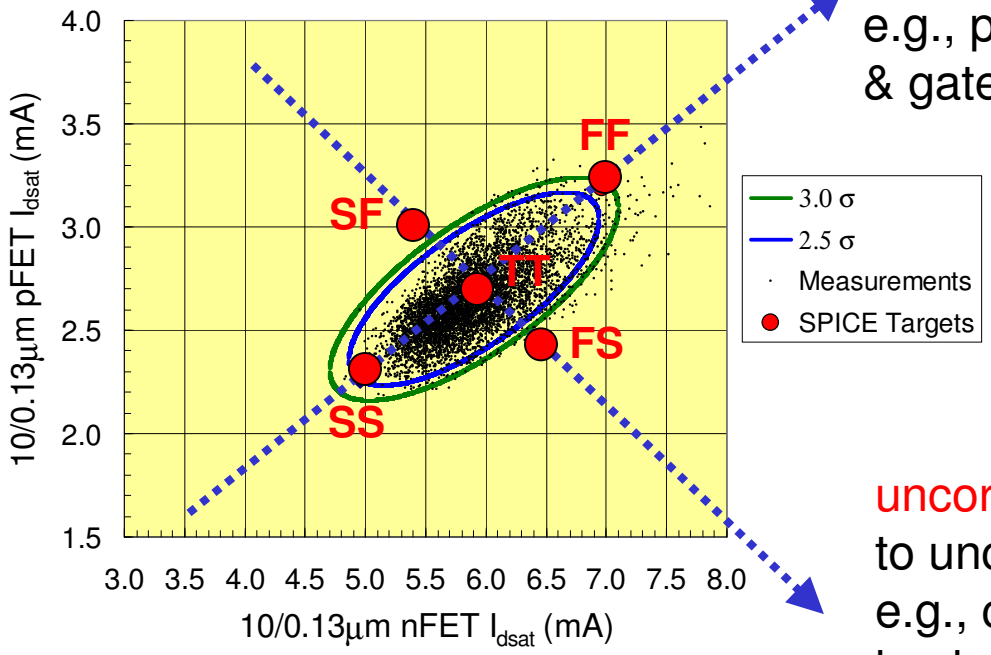
Characterizing Process Variations

- Statistical variations in IC manufacturing → variations in FET characteristics
- Circuits must function across operating V_{DD} & temperature but also across statistically acceptable process tolerances
- Statistical variations summarized by spread in nFET & pFET V_T's, or in I_{dsat}'s, i.e., use V_T or I_{dsat} to summarize cumulative effect of ALL process variances
- Consider die-to-die, wafer-to-wafer & lot-to-lot variations



Correlated vs. Uncorrelated Variations

- Elliptical 2-D Gaussian distribution from natural variations with no deliberate retargetting of process parameters



correlated variations due to common processes e.g., poly photo/etch CD & gate oxide thickness

uncorrelated variations due to uncommon processes e.g., channel & well ion implants

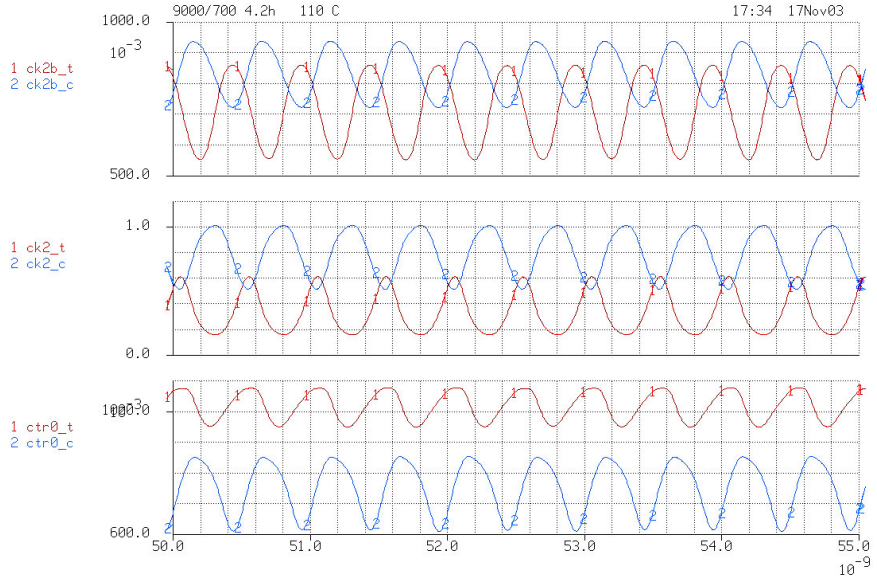
- Tougher to control poly CD than implant doses

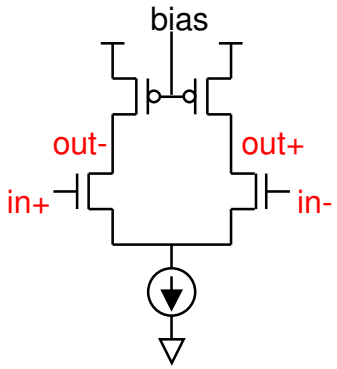
Additional Commentary
Manufacturing Impact on Design

- Importance of poly gate orientation – align FET gate along x or y?
  - Preferred orientation will likely exist due to step-and-scan nature of gate lithography (active area lithography to a smaller extent)
- FET mismatch
  - Symmetric circuits that are built asymmetric may fail
  - DFM (design for manufacturability) is a big buzz word now
    →need for Monte Carlo simulations to statistically validate design
- Exposure λ is *not* scaling as aggressively as L_{min}
  - Roadmaps are just guidelines
  - Huge industry resistance to move to 157nm
    - Requires *reflective* optics → $$$
  - Expect more gate resist trimming & relatively worse CD/overlay control at 65nm node
- Statistical design considerations more critical than ever

Impact of Statistical Variations
130nm High-Speed CML Buffers

- Monte Carlo simulations of initial CML buffer design
- Simulate random variations in L, W & V_T
- Waveforms should be differential, let alone cross each other



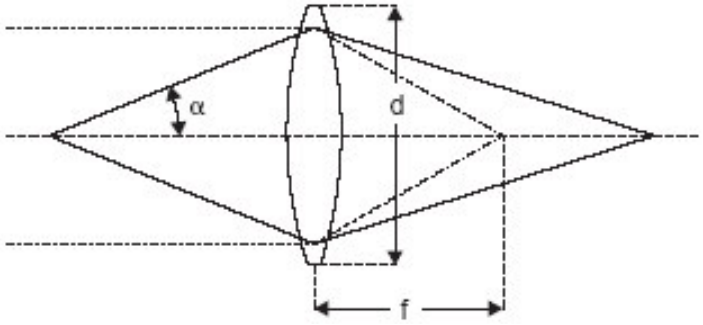


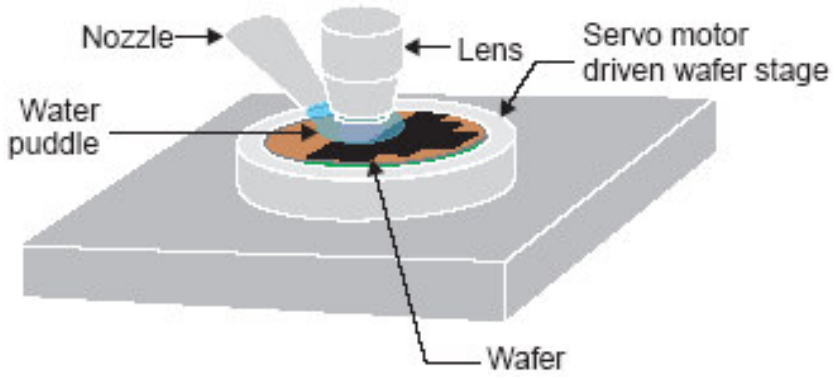
The Near Future of Patterning
Immersion Lithography

- Remember oil immersion microscopy in biology class?
- Extend resolution of refractive optics by squirting water puddle on wafer surface prior to exposure
  - n_{water} ~1.45 vs. n_{air} ~ 1
  - Tedious but less painful (i.e., cheaper) than resorting to reflective optics
- Very likely introduced at 45nm

$$\text{Resolution} = \frac{k_1 \lambda}{NA}$$

$$NA = n \sin \alpha = d / 2 f$$





Source: ICKnowledge.com (2003)

Conclusions

- CMOS scaling continues to be driven by digital circuit needs
- Should expect incremental changes in 65nm CMOS technology
  - Hopefully no major surprises
- Increasing mechanical engineering opportunities in IC technology
- Always a time lag for SPICE models to account for new effects
  - Pressing issue since tapeout mistakes are costlier than ever
  - Need to work closely with technology providers to quickly find out about these new effects
  - Account for them or avoid them!!!
- Designers with intimate knowledge of technology constraints will be best positioned to avoid pitfalls & to turn bugs into features

References

- International Technology Roadmap for Semiconductors, *Front End Processes* (2003 Edition), 2003.
- Technology Backgrounder: Immersion Lithography, ICKnowledge.com, 2003.
- Technology Backgrounder: Atomic Layer Deposition, ICKnowledge.com, 2003.
- H.-S. P. Wong, "Beyond the Conventional Transistor," *IBM Journal of Research & Development*, vol. 46, no. 2/3, pp. 133-168, Mar. 2002.
- C. R. Cleavelin, "Front End Manufacturing Technology," *IEEE IEDM Short Course on The Future of Semiconductor Manufacturing*, Dec. 2002.
- D. Harame, "RF Device Technologies," *IEEE IEDM Short Course on RF Circuit Design for Communication Systems*, Dec. 2002.
- S. Thompson *et al.*, "A 90nm Logic Technology Featuring 50nm Strained Silicon Channel Transistors, 7 Levels of Cu Interconnect, Low k ILD, and 1µm^2 SRAM Cell," *IEEE IEDM Tech. Digest*, pp. 61-64, Dec. 2002.
- C. C. Wu *et al.*, "A 90nm CMOS Device Technology with High-Speed , General-Purpose, and Low-Leakage Transistors for System on Chip Applications," *IEEE IEDM Tech. Digest*, pp. 65-68, Dec. 2002.
- R.A. Bianchi *et al.*, "Accurate Modeling of Trench Isolation Induced Mechanical Stress Effects on MOSFET Electrical Performance," *IEEE IEDM Tech. Digest*, pp. 117-120, Dec. 2002.
- T. B. Hook *et al.*, "Lateral Ion Implant Straggle and Mask Proximity Effect," *IEEE Trans. Electron Devices*, vol. 50, no. 9, pp. 1946-1951, Sept. 2003.
- Y. Taur, "CMOS Design Near the Limit of Scaling," *IBM Journal of Research & Development*, vol. 46, no. 2/3, pp. 213-222, Mar. 2002.
- J. D. Plummer *et al.*, *Silicon VLSI Technology– Fundamentals, Practice and Modeling*, Prentice-Hall, 2000.
- K. M. Cao *et al.*, "Modeling of Pocket Implanted MOSFETs for Anomalous Analog Behavior," *IEEE IEDM Tech. Digest*, pp. 171-120, Dec. 1999.
- A. Chatterjee *et al.*, "Transistor Design Issues in Integrating Analog Functions with High Performance Digital CMOS," *IEEE Symp. VLSI Technology Tech. Digest*, pp. 147-148, June 1999.
- S.M. Sze, *Physics of Semiconductor Devices (2nd ed.)*, John Wiley & Sons, 1981.