Workshop on Frontiers of Extreme Computing
Santa Cruz, CA
October 24, 2005

ITRS MOSFET Scaling Trends, Challenges, and Key Technology Innovations

Peter M. Zeitzoff

Accelerating the next technology revolution.

Outline

- Introduction
- MOSFET scaling and its impact
- Material and process approaches and solutions
- Non-classical CMOS
- Conclusions

Introduction

- **IC Logic technology: following Moore's Law by rapidly scaling into deep submicron regime**
  - **Increased speed and function density**
  - **Lower power dissipation and cost per function**
- **The scaling results in major MOSFET challenges, including:**
  - **Simultaneously maintaining satisfactory I_{on} (drive current) and I_{leak}**
  - **High gate leakage current for very thin gate dielectrics**
  - **Control of short channel effects (SCEs) for very small transistors**
  - **Power dissipation**
  - **Etc.**
- **Potential solutions & approaches:**
  - **Material and process (front end): high-k gate dielectric, metal gate electrodes, strained Si, …**
  - **Structural: non-classical CMOS device structures**
  - ***Many innovations needed in rapid succession***

International Technology Roadmap for Semiconductors (ITRS)

- **Industry-wide effort to map IC technology generations for the next 15 years**
  - **Over 800 experts from around the world**
    - **From companies, consortia, and universities**
  - **For each calendar year**
    - **Projects scaling of technology characteristics and requirements, based on meeting key Moore's Law targets**
    - **Assesses key challenges and gaps**
    - **Lists best-known potential solutions**
  - **Projections are based on modeling, surveys, literature, experts' technical judgment**
- **This talk is based on both the 2003 ITRS and on preliminary data from 2005 ITRS (not yet released)**

Key Overall Chip Parameters for High-Performance Logic, Data from 2003 ITRS

Year of Production	2003	2004	2005	2006	2007	2008	2009	2010	2012	2013	2015	2016	2018
Technology Node		hp90			hp65			hp45		hp32		hp22	
DRAM ½ Pitch (nm)	100	90	80	70	65	57	50	45	35	32	25	22	18
MPU Physical Gate Length (nm)	**45**	**37**	**32**	**28**	**25**	**22**	**20**	**18**	**14**	**13**	**10**	**9**	**7**
Vdd (V)	1.2	1.2	1.1	1.1	1.1	1	1	1	0.9	0.9	0.8	0.8	0.7
Chip Frequency (MHz)													
On-chip local clock	**2,976**	**4,171**	**5,204**	**6,783**	**9,285**	**10,972**	**12,369**	**15,079**	**20,065**	**22,980**	**33,403**	**39,683**	**53,207**
Allowable Maximum Power													
High-performance with heatsink (W)	**149**	**158**	**167**	**180**	**189**	**200**	**210**	**218**	**240**	**251**	**270**	**288**	**300**
Cost-performance (W)	80	84	91	98	104	109	114	120	131	138	148	158	168
Functions per chip at production (million transistors [Mtransistors])	153	193	243	307	386	487	614	773	1,227	1,546	2,454	3,092	4,908

- **Technology generations defined by DRAM half pitch**
- **Gate length (L_g) ≤ 0.5 X DRAM half pitch**
  - **Rapid scaling of L_g is driven by need to improve transistor speed**
- **Clock frequency, functions per chip (density) scale rapidly, but allowable power dissipation rises slowly with scaling: limited by ability to remove heat**

Outline

- Introduction
- MOSFET scaling and its impact
- Material and process approaches and solutions
- Non-classical CMOS
- Conclusions

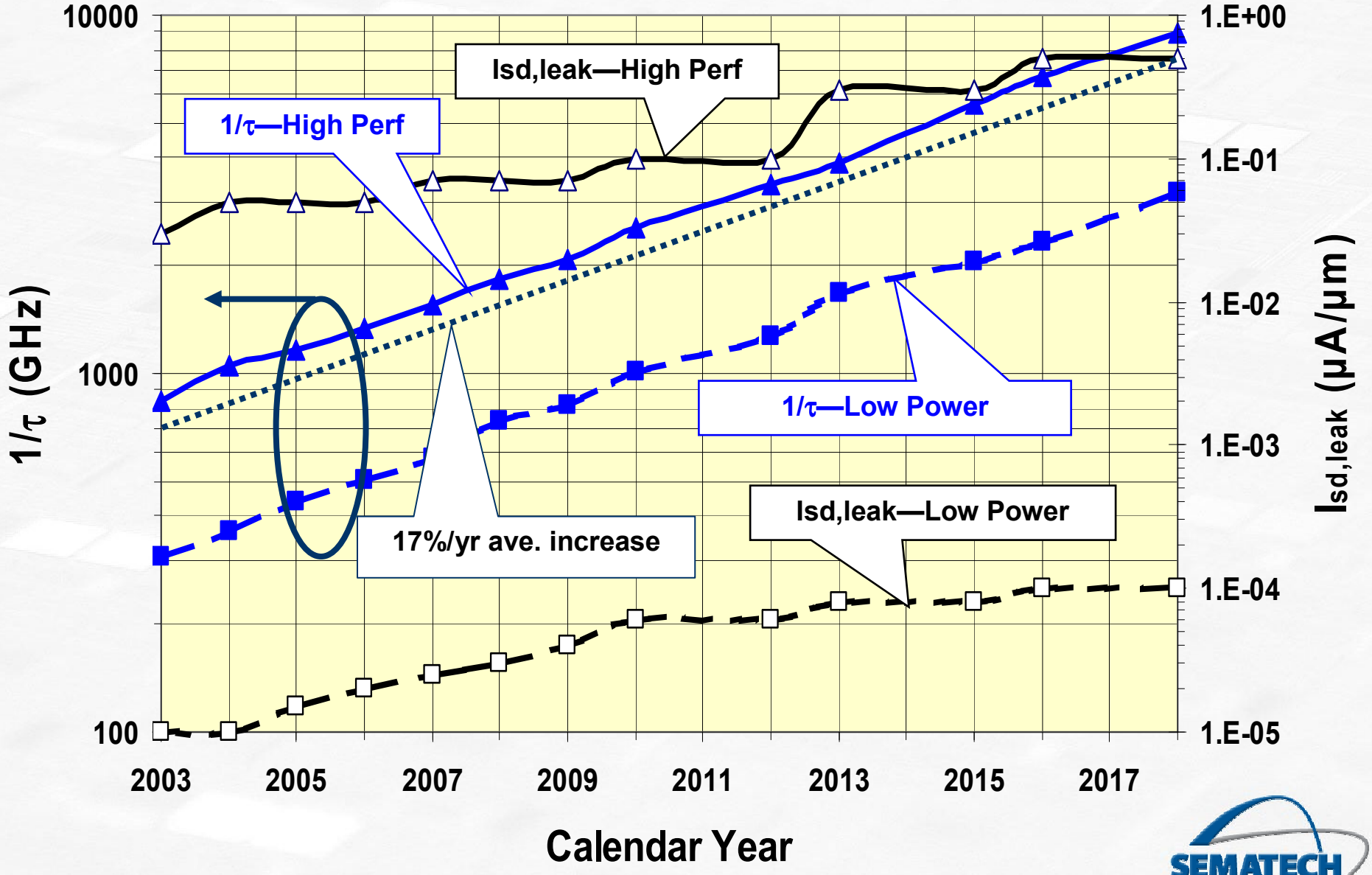
MOSFET Scaling Approach: 2005 ITRS

- **MASTAR computer modeling software is used: detailed, analytical MOSFET models with key MOSFET physics included**
  - **Initial choice of scaled MOSFET parameters is made**
  - **Using MASTAR, MOSFET parameters are iteratively varied to meet ITRS targets for either**
    - **Scaling of transistor speed OR**
    - **Specific (low) levels of leakage current**

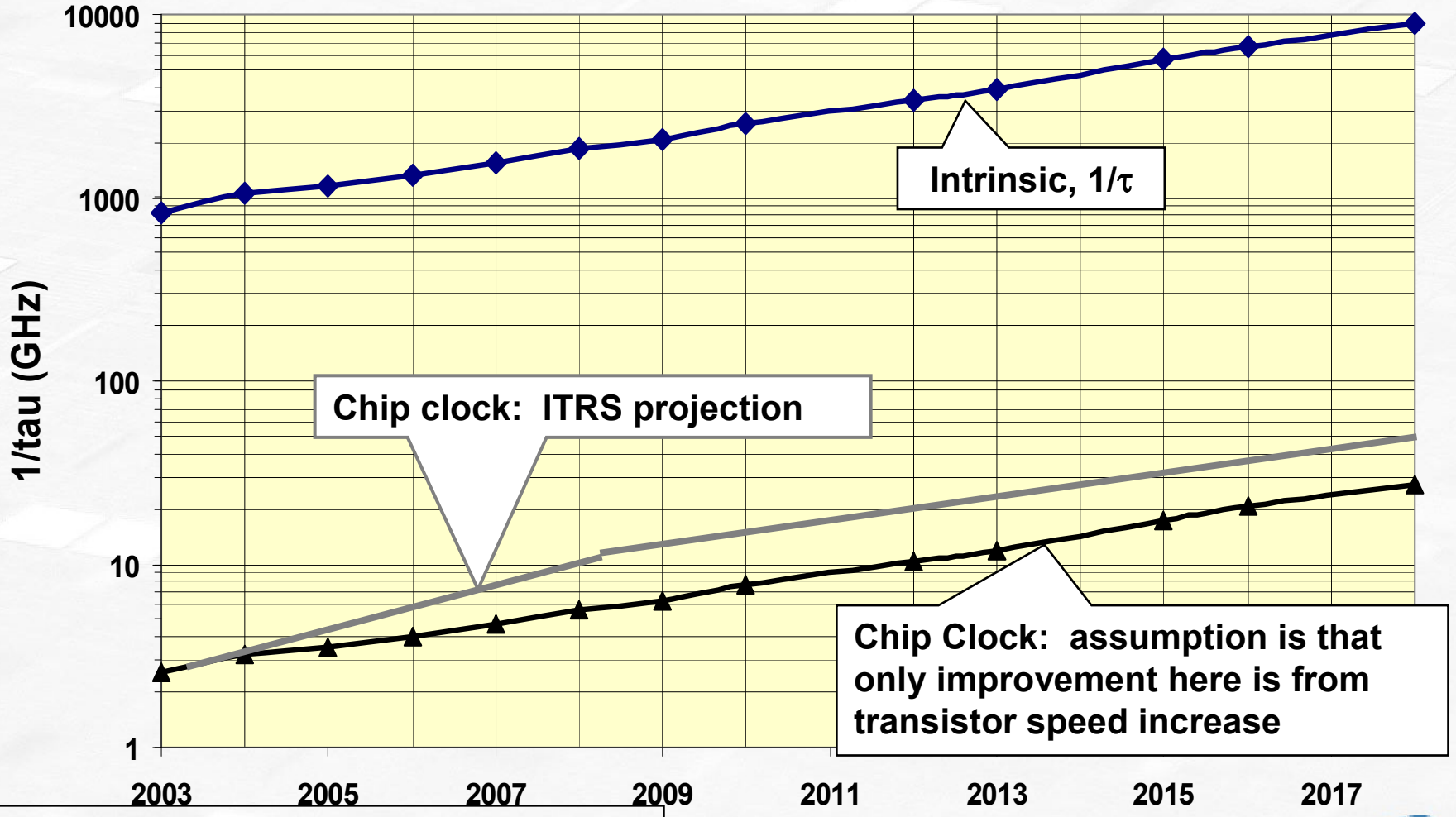
ITRS Drivers for Different Applications

- High-performance chips (MPU, for example)
  - Driver: maximize chip speed→maximize transistor performance (metric: τ, transistor intrinsic delay [or, equivalently, $1/\tau$])
    - Goal of ITRS scaling: $1/\tau$ increases at ~ 17% per year, historical rate
      - Must maximize I_{on}
      - Consequently, I_{leak} is relatively high
- Low-power chips (mobile applications)
  - Driver: minimize chip power (to conserve battery power) →minimize I_{leak}
    - Goal of ITRS scaling: low levels of I_{leak}
      - Consequently, $1/\tau$ is considerably less than for high-performance logic
- This talk focuses on high-performance logic, which largely drives the technology

$1/\tau$ and $I_{sd,leak}$ scaling for High-Performance and Low-Power Logic. Data from 2003 ITRS.

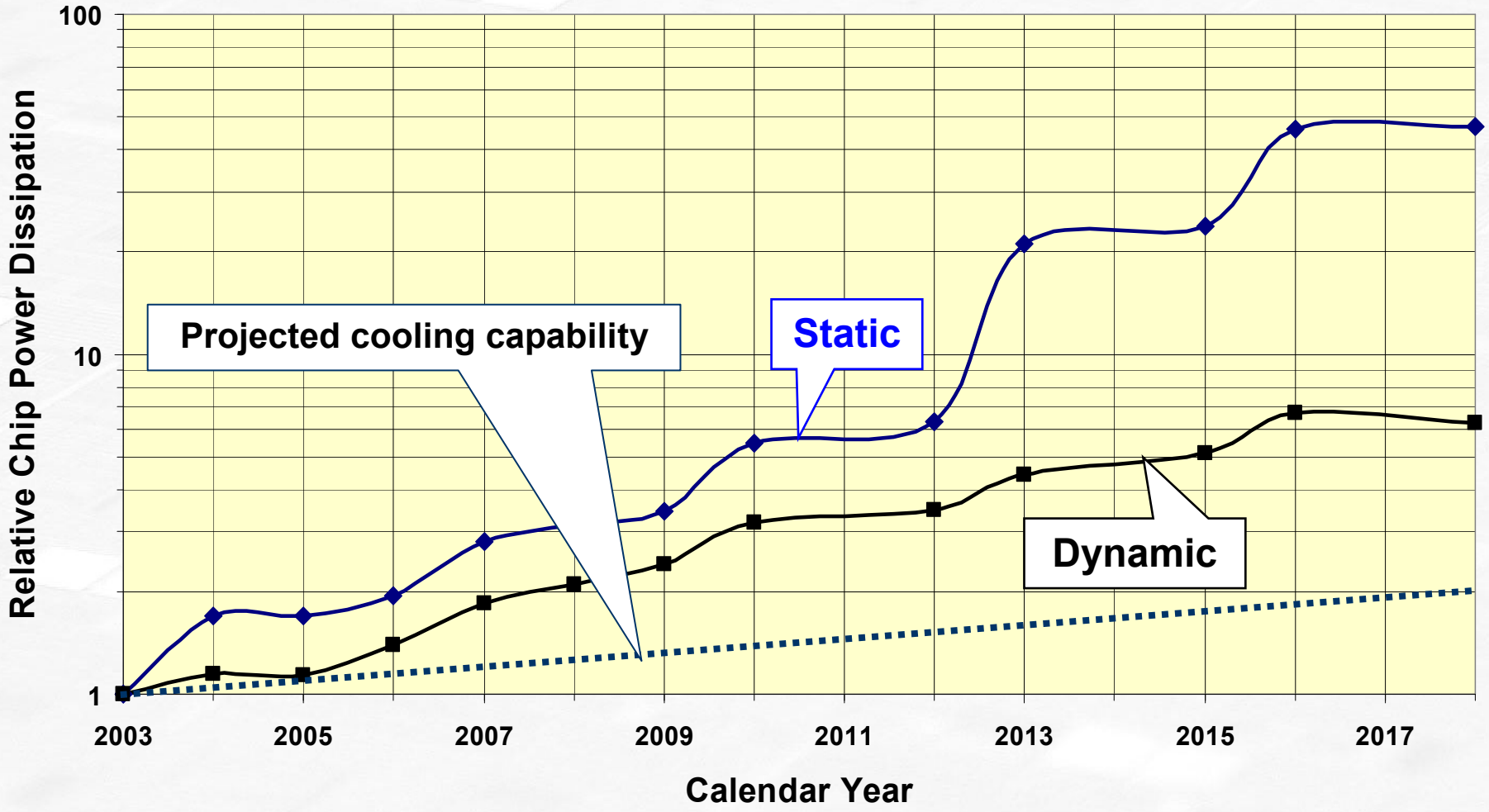


Frequency scaling: Transistor Intrinsic Speed and Chip Clock Frequency for High-Performance Logic. Data from 2003 ITRS.



Conclusion: transistor speed improvement is a critical enabler of chip clock frequency improvement

Potential Problem with Chip Power Dissipation Scaling: High-Performance Logic, Data from 2003 ITRS



Unrealistic assumption, to make a point about P_{static}: all transistors are high performance, low V_t type

Potential Solutions for Power Dissipation Problems, High-Performance Logic

- Increasingly common approach: multiple transistor types on a chip→multi-V_t, multi-T_{ox}, etc.
  - Only utilize high-performance, high-leakage transistors in critical paths—lower leakage transistors everywhere else
  - Improves flexibility for SOC
- Circuit and architectural techniques: pass gates, power down circuit blocks, etc.
- Improved heat removal, electro-thermal modeling and design
- Electrical or dynamically adjustable V_t devices (future possibility)

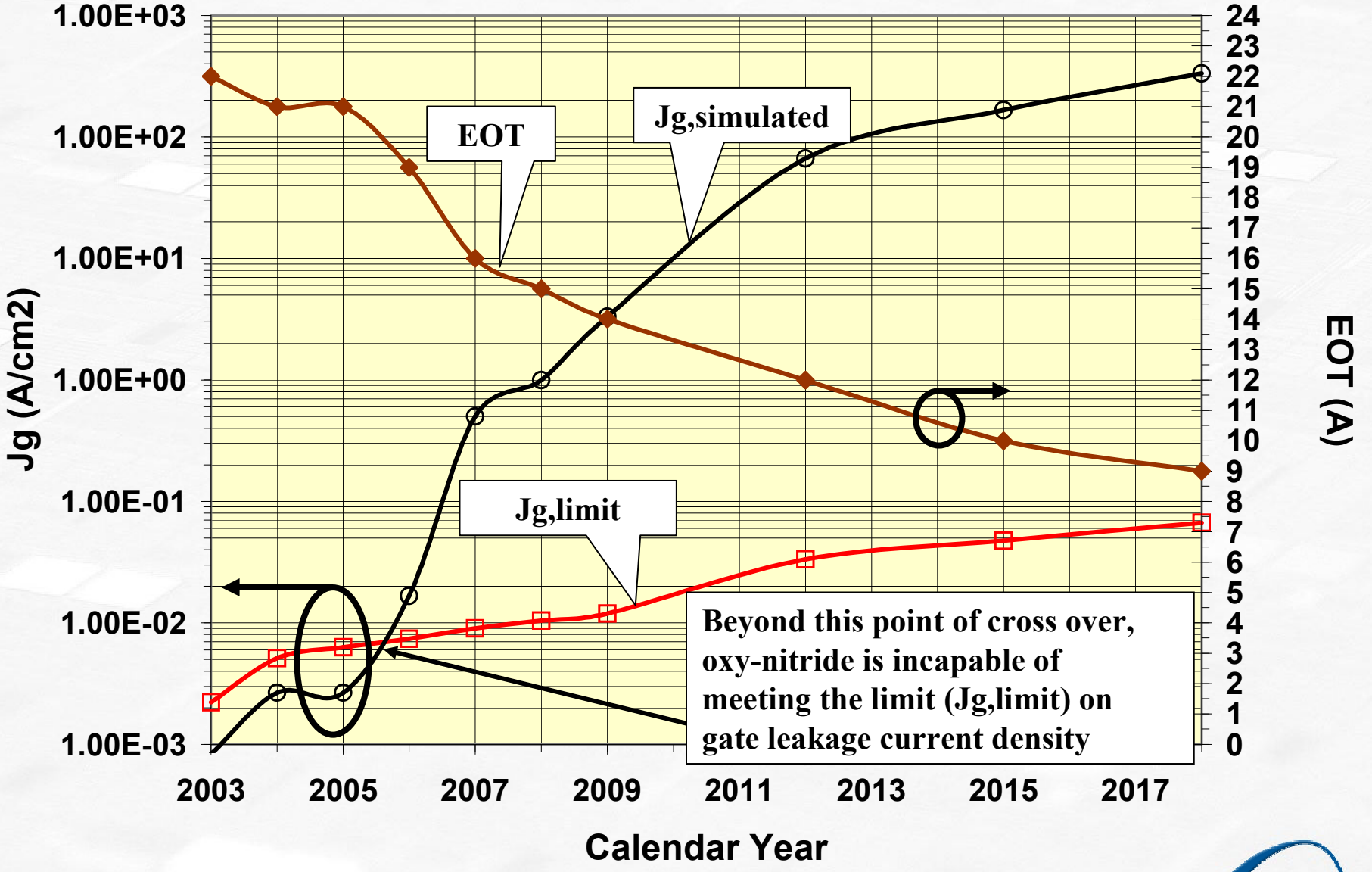
Outline

- Introduction
- MOSFET scaling and its impact
- Material and process approaches and solutions
- Non-classical CMOS
- Conclusions

Difficult Transistor Scaling Issues

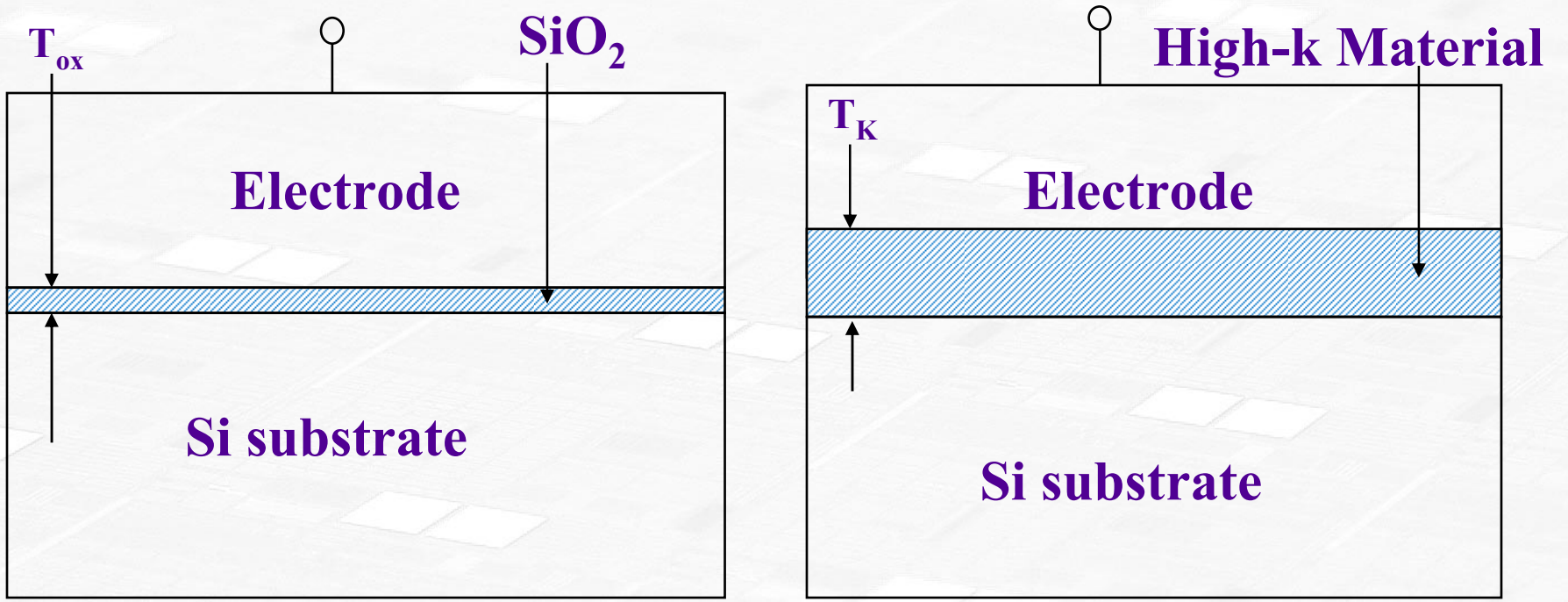
- **Assumption: highly scaled MOSFETs with the targeted characteristics can be successfully designed and fabricated**
- **However, with scaling, meeting transistor requirements will require significant technology innovations**
  - ***Issue: High gate leakage*** → **static power dissipation**
    - **Direct tunneling increases rapidly as T_{ox} is reduced**
    - **Potential solution: high-k gate dielectric**
  - ***Issue: Polysilicon depletion in gate electrode*** → **increased effective T_{ox}, reduced I_{on}**
  - ***Issue: Need for enhanced channel mobility***
  - **Etc.**

For Low-Power Logic, Gate Leakage Current Density Limit Versus Simulated Gate Leakage due to Direct Tunneling. Data from 2003 ITRS.



2006, EOT = 1.9 nm, $J_{g,max}$ ~ 0.007 A/cm^{-2}

High K Gate Dielectric to Reduce Direct Tunneling

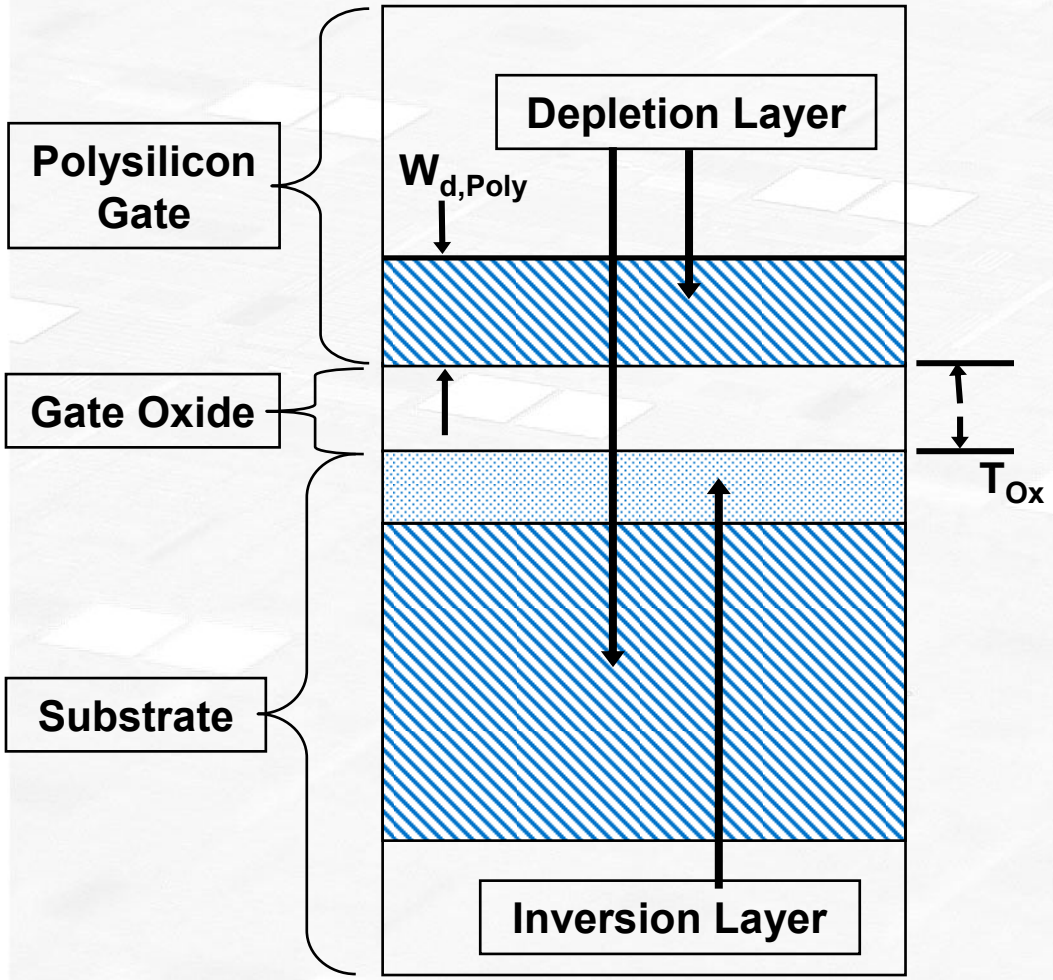


- Equivalent Oxide Thickness = EOT = T_{ox} = T_K * (3.9/K), where 3.9 is relative dielectric constant of SiO2 and K is relative dielectric constant of high K material
  - $C = C_{ox} = \varepsilon_{ox}/T_{ox}$
  - To first order, MOSFET characteristics with high-k are same as for SiO2
- Because $T_K > T_{ox}$, direct tunneling leakage much reduced with high K
  - If energy barrier is high enough
- Current leading candidate materials: HfO_2 (K_{eff}~15 - 30); $HfSiO_x$ (K_{eff}~12 - 16)
  - Materials, process, integration issues to solve

Difficult Transistor Scaling Issues

- **With scaling, meeting transistor requirements requires significant technology innovations**
  - *Issue: High gate leakage* → static power dissipation
    - Potential solution: high-k gate dielectric
  - ***Issue: polysilicon depletion in gate electrode*** → **increased effective T_{ox}, reduced I_{on}**
    - **Potential solution: metal gate electrodes**
  - *Issue: Need for enhanced channel mobility*
  - Etc.

Polysilicon Depletion and Substrate Quantum Effects



- $T_{ox,electric} = T_{ox} + (K_{ox}/K_{si})^* (W_{d,Poly})$
  - $K_{ox} = 3.9$
  - $K_{si} = 11.9$
- $T_{ox,electric} = T_{ox} + (0.33)^* (W_{d,Poly})$
  - $W_{d,Poly} \sim 1/(\text{poly doping})^{0.5}$ → increase poly doping to reduce $W_{d,Poly}$ with scaling
  - But max. poly doping is limited → can't reduce $W_{d,Poly}$ too much
- Poly depletion become more critical with T_{ox} scaling
  - Eventually, poly will reach its limit of effectiveness

Metal Gate Electrodes

- **Metal gate electrodes are a potential solution when poly "runs out of steam": probably implemented in 2008 or beyond**
  - **No depletion, very low resistance gate, no boron penetration, compatibility with high-k**
  - **Issues**
    - **Different work functions needed for PMOS and NMOS==>2 different metals may be needed**
      - **Process complexity, process integration problems, cost**
    - **Etching of metal electrodes**
    - **New materials: major challenge**

Difficult Transistor Scaling Issues

- **With scaling, meeting transistor requirements requires significant technology innovations**
  - *Issue: High gate leakage* → static power dissipation
    - Potential solution: high-k gate dielectric
  - *Issue: Poly depletion in gate electrode* → increased effective T_{ox}, reduced I_{on}
    - Potential solution: metal gate electrodes
  - ***Issue: Need for enhanced channel mobility***
    - ***Potential solution: enhanced mobility via strain engineering***
  - Etc.

Uniaxial Process Induced Stress for Enhanced Mobility

NMOS: uniaxial tensile stress from stressed SiN film

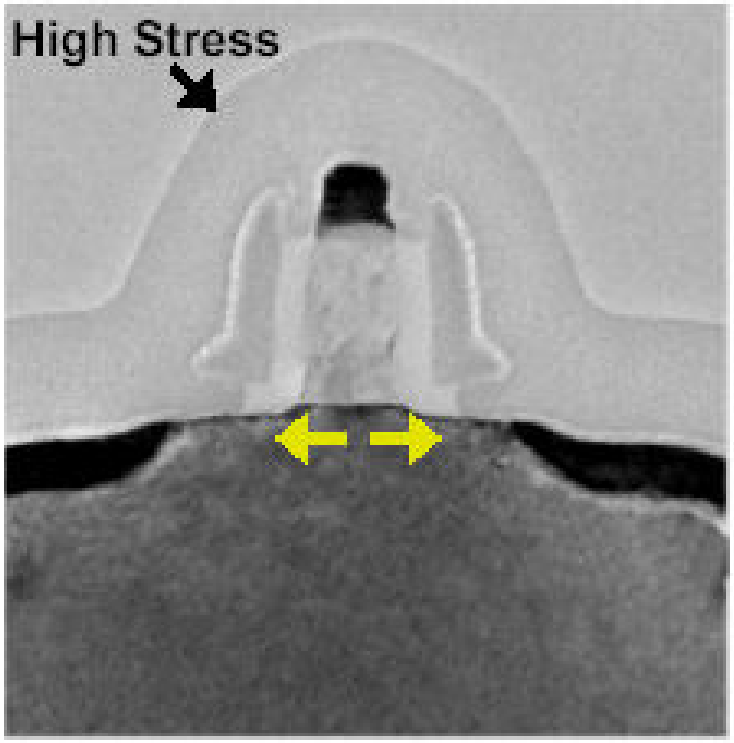


Fig. 3 TEM of NMOS transistor showing high tensile stress nitride overlayer.

PMOS: uniaxial compressive stress from sel. SiGe in S/D

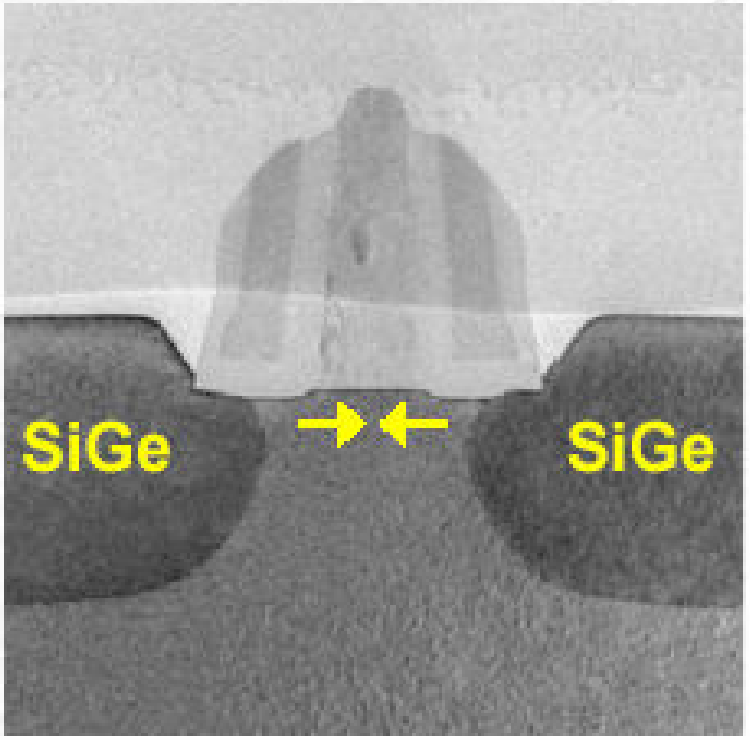


Fig. 4 TEM of PMOS showing SiGe heteroepitaxial S/D inducing uniaxial strain.

From K. Mistry et al., "Delaying Forever: Uniaxial Strained Silicon Transistors in a 90nm CMOS Technology," *2004 VLSI Technology Symposium*, pp. 50-51.

Results from Uniaxial Process Induced Stress

NMOS $I_{d,sat}$

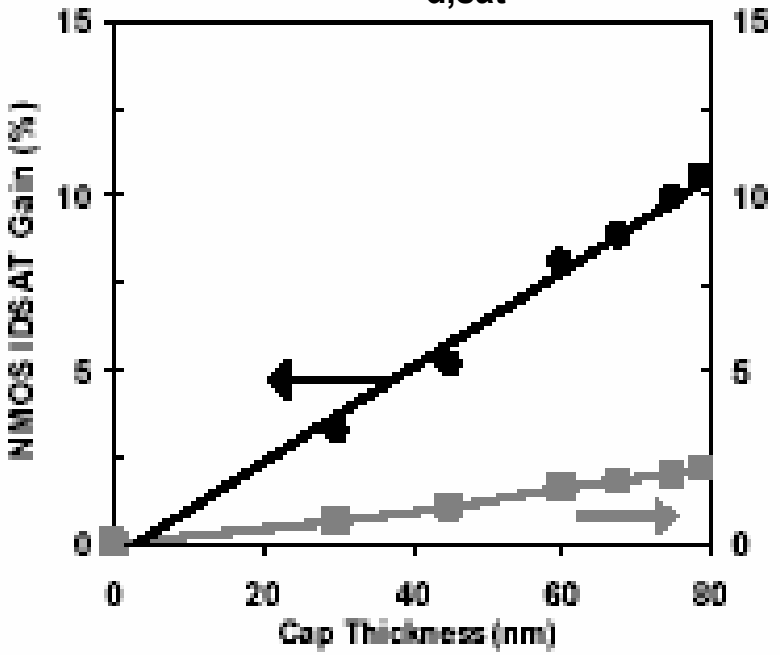


Fig. 7 N & P I_{DSAT} vs. nitride capping layer thickness.

PMOS $I_{d,lin}$

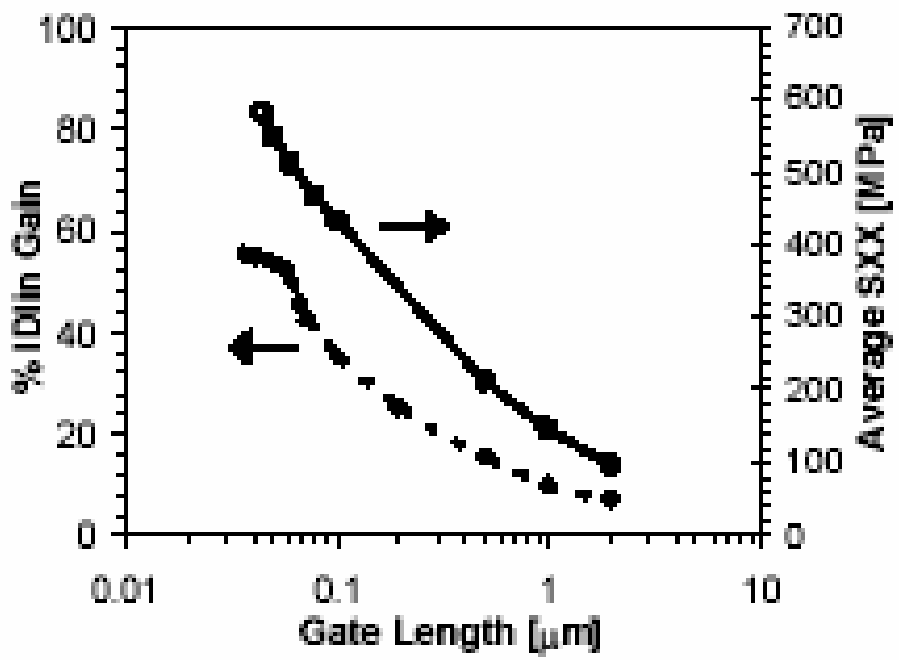


Fig. 8 Avg. PMOS channel stress (SXX) and % I_{DLIN} gain [corrected for (V_G-V_T)] vs. L_{GATE}

From K. Mistry et al., "Delaying Forever: Uniaxial Strained Silicon Transistors in a 90nm CMOS Technology," *2004 VLSI Technology Symposium*, pp. 50-51.

Outline

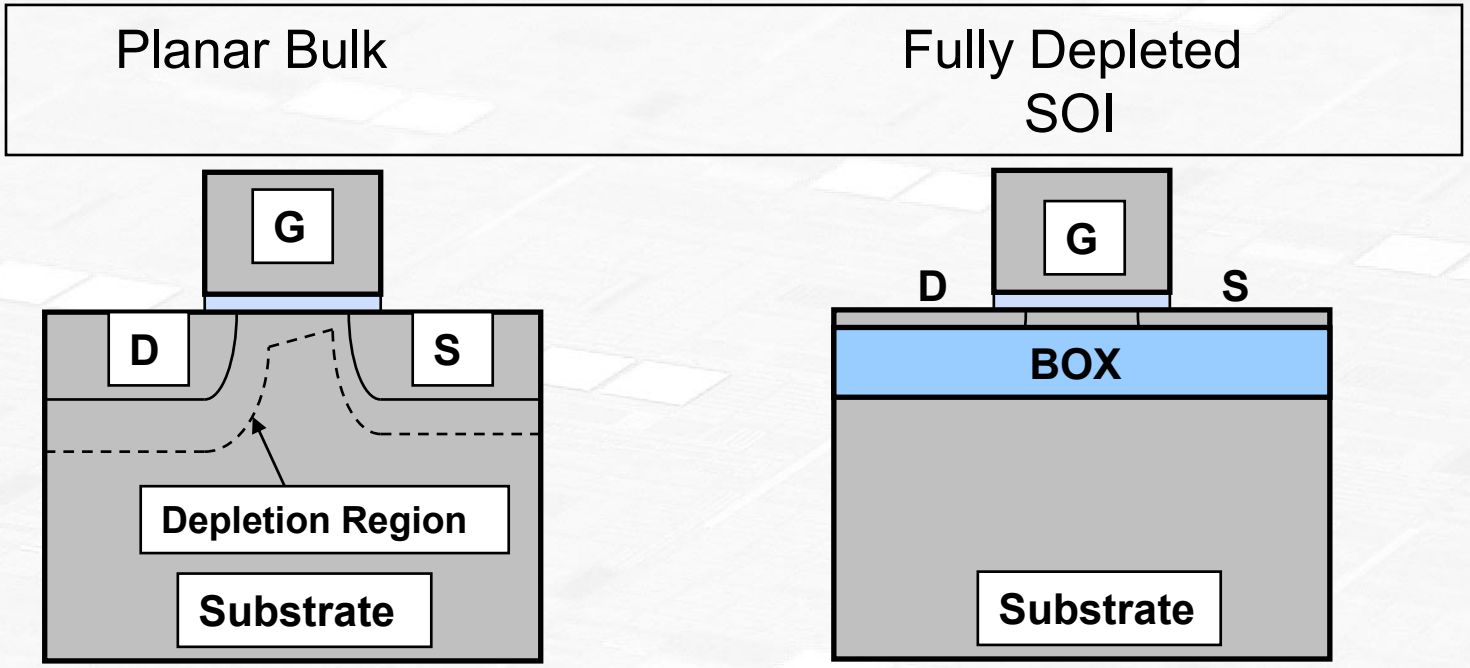
- Introduction
- Scaling and its impact
- Material and process approaches and solutions
- Non-classical CMOS
- Conclusions

Limits of Scaling Planar, Bulk MOSFETs

- **65 nm tech. generation (2007, L_g = 25nm) and beyond: increased difficulty in meeting all device requirements with classical planar, bulk CMOS (even with high-k, metal electrodes, strained Si…)**
  - **Control of SCE**
  - **Impact of quantum effects and statistical variation**
  - **Impact of high substrate doping**
  - **Control of series S/D resistance ($R_{series,s/d}$)**
  - **Others**

- **Alternative device structures (non-classical CMOS) may be utilized**
  - **Ultra thin body, fully depleted: single-gate SOI and multiple-gate transistors**

Transistor Structures: Planar Bulk & Fully Depleted SOI



+ Wafer cost / availability
- SCE scaling difficult
- High doping effects and Statistical variation
- Parasitic junction capacitance

+ Lower junction cap
+ Light doping possible
+ Vt can be set by WF of Metal Gate Electrode
- SCE scaling difficult
- Sensitivity to Si thickness (very thin)
- Wafer cost/availability

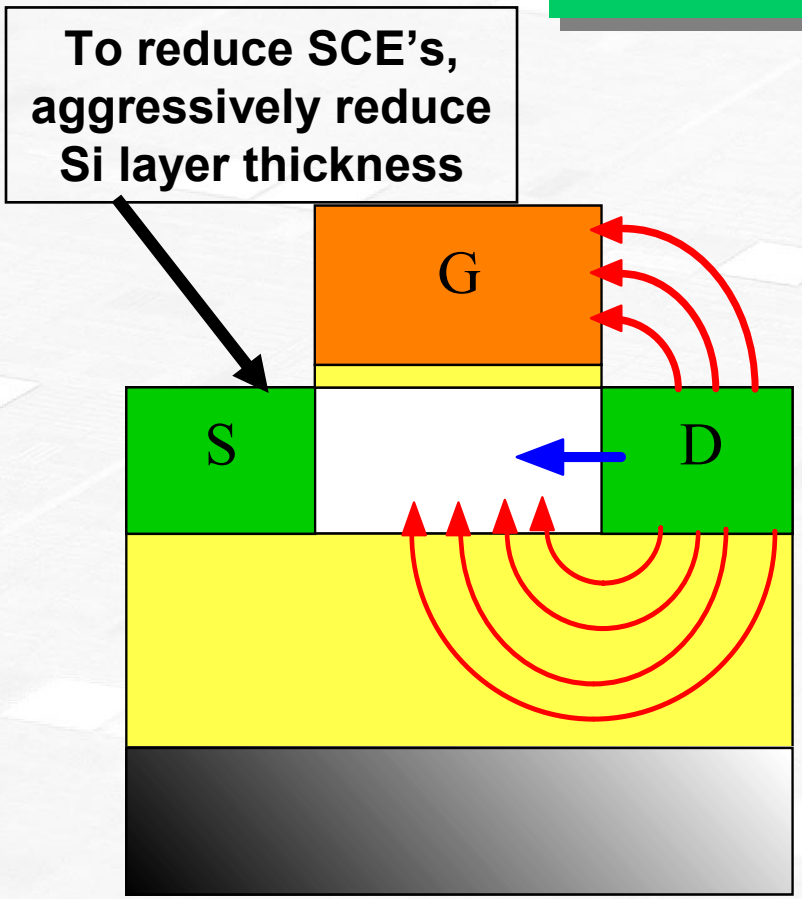
REFERNCES

*1. P.M. Zeitzoff, J.A. Hutchby and H.R. Huff, MOSFET and Front-End Process Integration: Scaling Trends, Challenges, and Potential Solutions Through The End of The Roadmap, International Journal of High-Speed Electronics and Systems, **12**, 267-293 (2002).*

*2. Mark Bohr, ECS Meeting PV **2001-2**, Spring, 2001.*

Field Lines for Single-Gate SOI MOSFETs

E-Field lines



Single-Gate SOI

Courtesy: Prof. J-P Colinge, UC-Davis

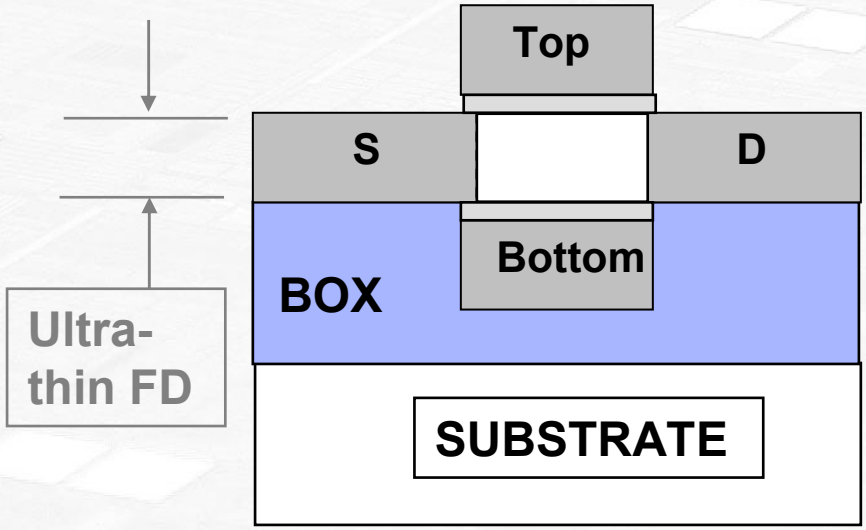
Double Gate Transistor Structure

REFERENCES

1. P.M. Zeitzoff, J.A. Hutchby and H.R. Huff, MOSFET and Front-End Process Integration: Scaling Trends, Challenges, and Potential Solutions Through The End of The Roadmap, International Journal of High-Speed Electronics and Systems, ***12****, 267-293 (2002).*

2. Mark Bohr, ECS Meeting PV ***2001-2****, Spring, 2001.*

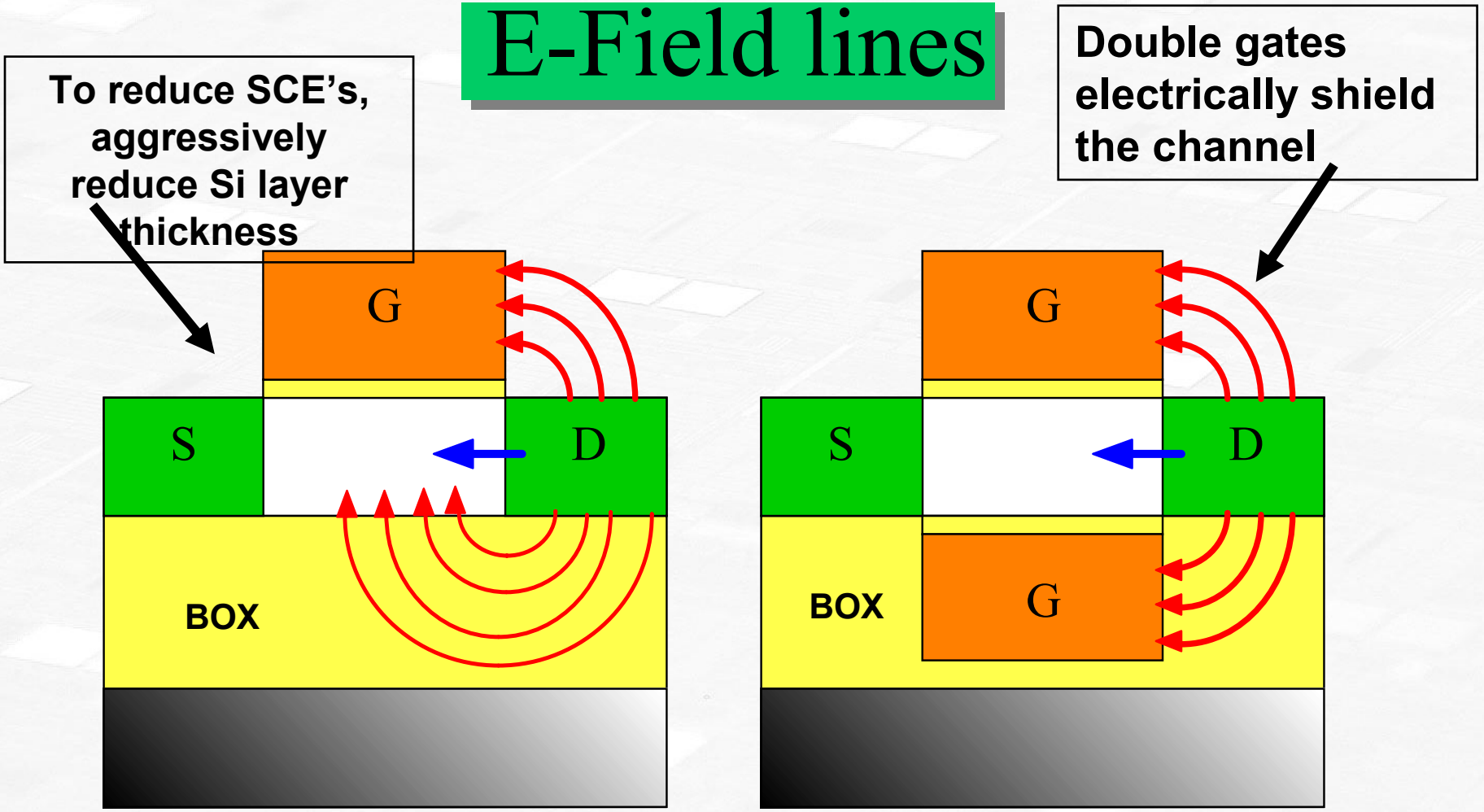
Double-Gate SOI:



- + ***Enhanced scalability***
- + **Lower junction capacitance**
- + **Light doping possible**
- + **Vt can be set by WF of metal gate electrode**
- + **~2x drive current**
- - **~2x gate capacitance**
- - **High $R_{series,s/d}$→raised S/D**
- - **Complex process**

Summary: more advanced, optimal device structure, but difficult to fabricate, particularly in this SOI configuration

Field Lines for Single and Double-Gate MOSFETs



Single-Gate SOI | **Double-Gate**

Courtesy: Prof. J-P Colinge, UC-Davis

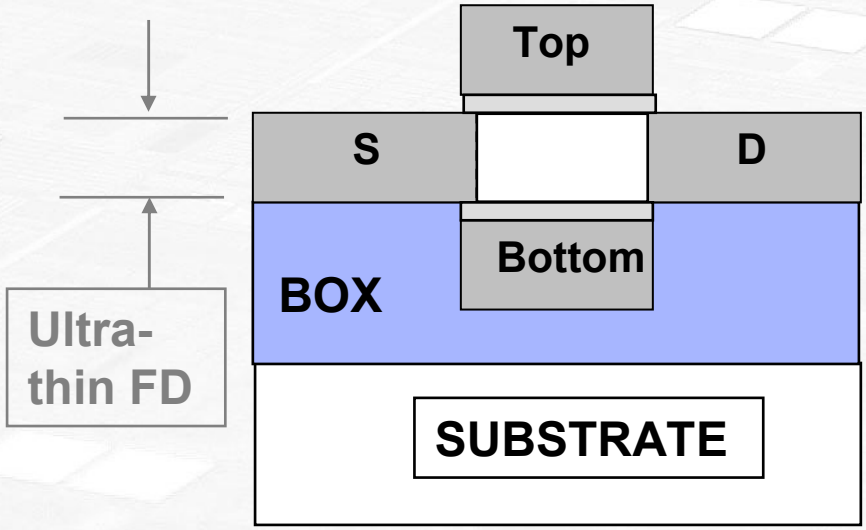
Double Gate Transistor Structure

REFERENCES

1. P.M. Zeitzoff, J.A. Hutchby and H.R. Huff, MOSFET and Front-End Process Integration: Scaling Trends, Challenges, and Potential Solutions Through The End of The Roadmap, International Journal of High-Speed Electronics and Systems, ***12****, 267-293 (2002).*

2. Mark Bohr, ECS Meeting PV ***2001-2****, Spring, 2001.*

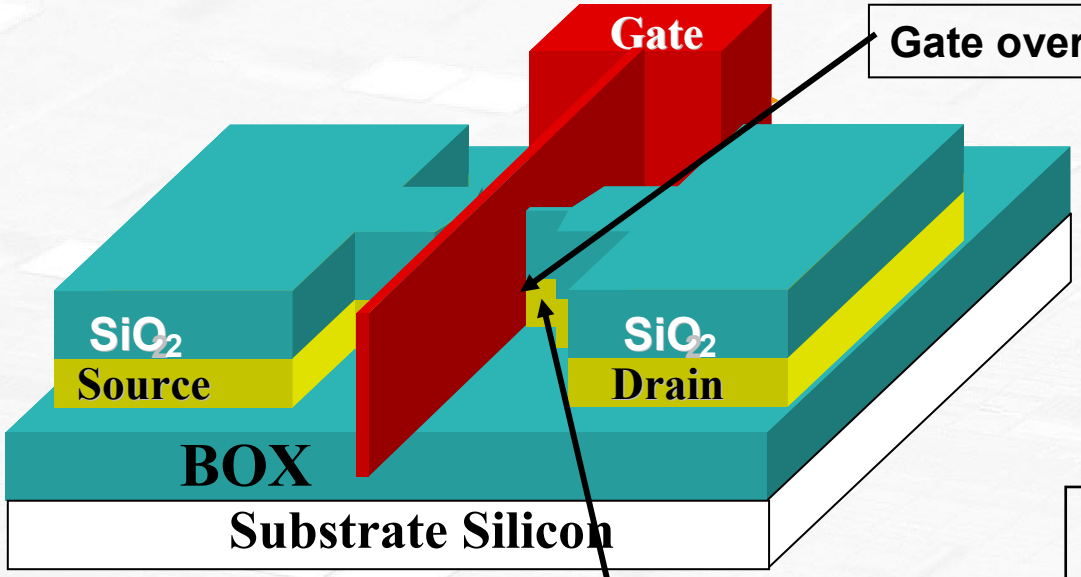
Double-Gate SOI:



- \+ ***Enhanced scalability***
- \+ Lower junction capacitance
- \+ Light doping possible
- \+ Vt can be set by WF of metal gate electrode
- \+ ~2x drive current
- \- ~2x gate capacitance
- \- High $R_{series,s/d}$→raised S/D
- \- Complex process

Summary: more advanced, optimal device structure, but difficult to fabricate, particularly in this SOI configuration

Other Double-Gate Transistor Structures (FinFET)

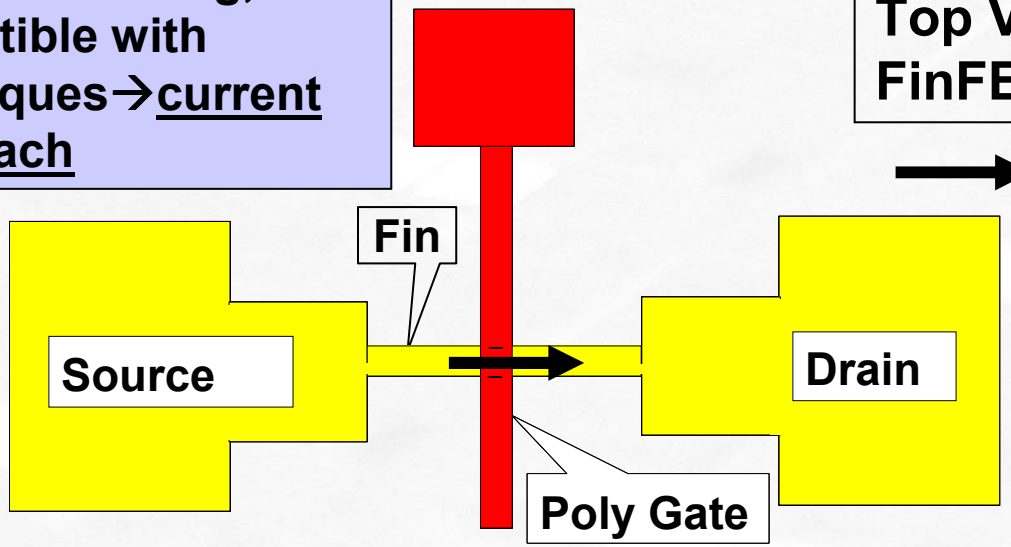


Perspective view of FinFET. Fin is colored yellow.

Courtesy: T-J. King and C. Hu, UC-Berkeley

Key advantage: relatively conventional processing, largely compatible with current techniques→current leading approach

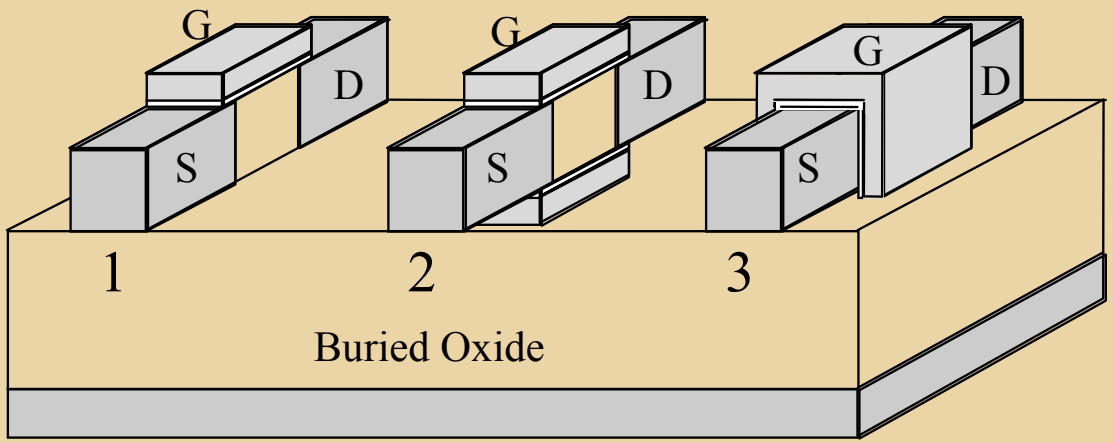
Top View of FinFET

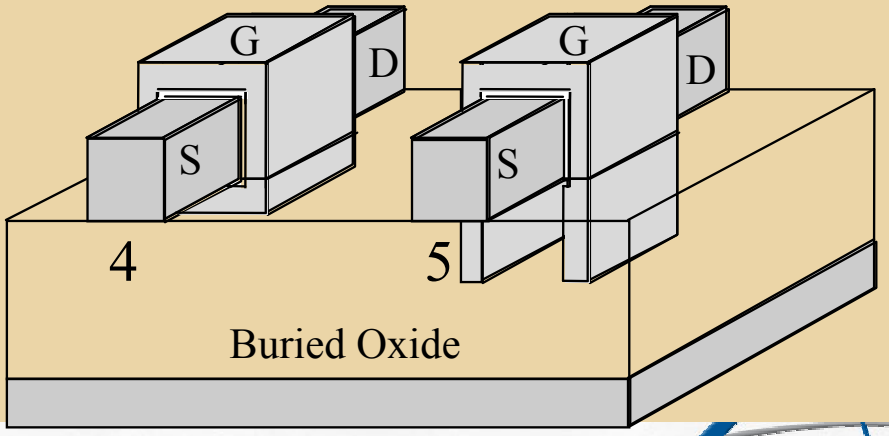


Arrow indicates Current flow

Types of Multiple-Gate Devices

Courtesy: Prof. J-P Colinge, UC-Davis





Increasing process complexity, increasing scalability

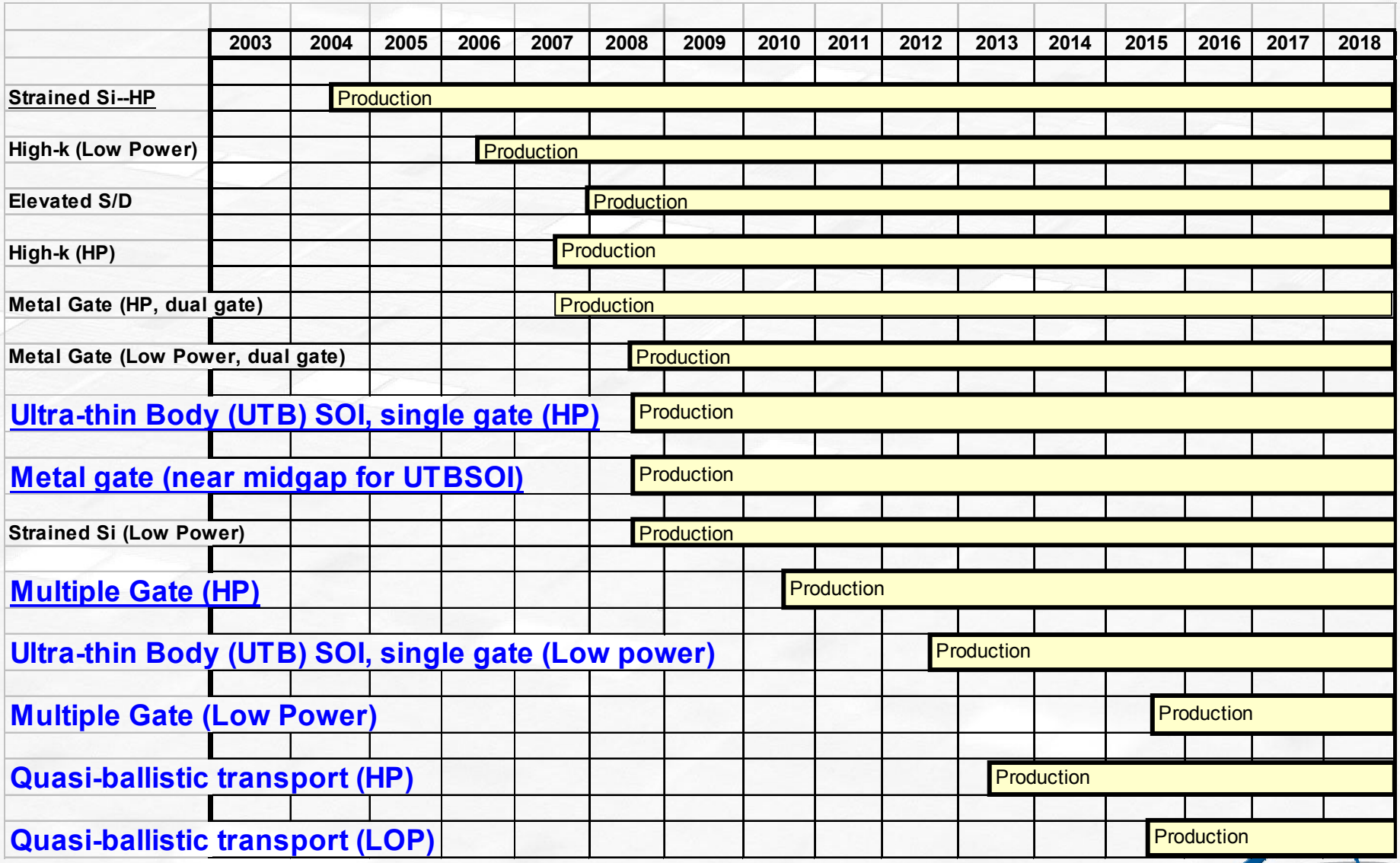
1: Single gate
2: Double gate
3: Triple gate
4: Quadruple gate (GAA)
5: Πgate

Outline

- Introduction
- Scaling and its impact
- Material and process approaches and solutions
- Non-classical CMOS
- Conclusions

Timeline of Projected Key Technology Innovations from '03 ITRS, PIDS Section

This timeline is from PIDS evaluation for the 2003 ITRS



Conclusions

- Rapid transistor scaling is projected to continue through the end of the Roadmap in 2020
  - Transistor performance will improve rapidly, but leakage & SCEs will be difficult to control
    - Transistor performance improvement is a key enabler of chip speed improvement
  - Many technology innovations will be needed in a relatively short time to enable this rapid scaling
    - Material and process innovations include high-k gate dielectric, metal gate electrodes, and enhanced mobility through strained silicon
      - High-k and metal gate electrode needed in 2008
    - Structural potential solutions: non-classical CMOS
- Non-classical CMOS and process and material innovations will likely be combined in the ultimate, end-of-Roadmap device
  - Well under 10nm MOSFETs expected by the end of the Roadmap
- Power dissipation, especially static, is a growing problem with scaling: integrated, innovative approaches needed