

**CMOS Emerging Technologies Workshop**  
**July 10-12, 2007, Whistler, BC, Canada**

**Process variability and leakage currents  
in nanometer CMOS circuits**

**W. Kuzmicz, Warsaw University of Technology, Poland**

## Acknowledgments

**Coauthors of the EDA software used in this work:**

**A. Pfitzner, E. Piwowarska, D. Kasprowicz, A. Wojtasik  
(WUT Warsaw, Poland).**

**The work presented here is partially supported by the  
European Union R&D project CLEAN (Controlling  
leakage power in nanoCMOS SOC's,  
<http://clean.offis.de>)**

**Discussions, collaboration and contributions of the  
CLEAN project partners:**

**J. Figueras (UPC Barcelona, Spain)**

**E. Beigne, O. Rozeau (CEA-LETI, Grenoble, France)**

**D. Helms (OFFIS, Oldenburg, Germany)**

**are gratefully acknowledged.**

## Outline

- **Prologue**
- **Origins and consequences of leakage currents**
- **Effects of device design and technology on leakage currents**
- **Variability and leakage currents: why is variability important?**
- **Leakage modeling in logic blocks**
- **Leakage limited parametric yield**
- **Conclusions**

## The power crisis

- **Power density in high performance VLSI chips is already at the same level as in a nuclear reactor**
- **Cooling is a big problem and is becoming expensive**
- **Increasing amount of energy is consumed by electronic devices (Amsterdam: 20% for telecom; USA: 9% for internet,...\*)**
- **Battery maximum power and capacity increase by 10-15% per year, but chip power requirements increase much faster: 35-40% per year.**

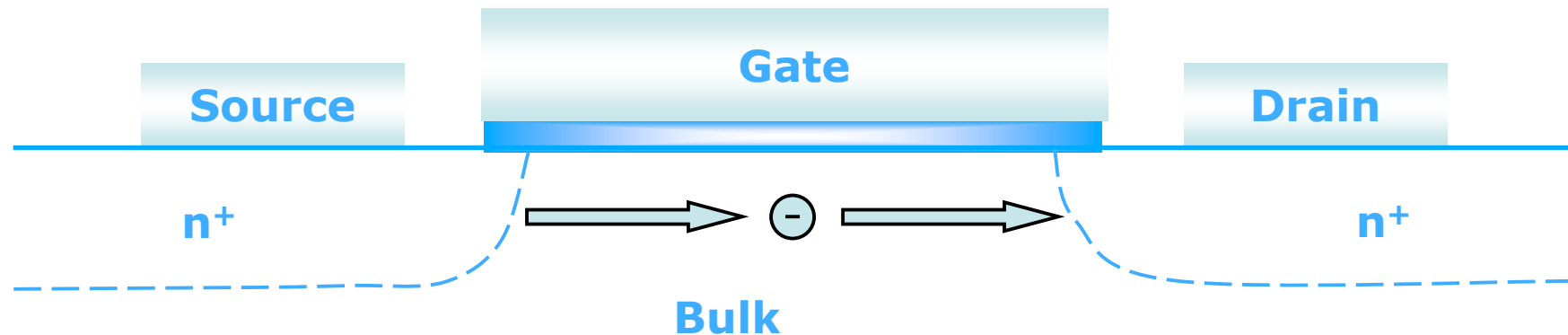
**\*Source: 2000 CO<sup>2</sup> conference, Amsterdam, NL**

## The variability problem

- **How process variabilities affect leakage**
  - considering all kinds of process variabilities,
  - taking all leakage components into account,
  - accounting for all correlations in a realistic way?
- **How variability-enhanced leakage affects manufacturing yield?**  
(we consider here *leakage-related yield loss* only, but neither performance nor defect-limited yield)

# Subthreshold leakage

**Subthreshold current is the current of minority electrons flowing by means of diffusion through p-type substrate from source to drain:**



**Exponential dependence of the minority carrier concentration on the gate and drain voltages leads to exponential I(V) characteristics**

$$I_{subth} = \mu_n C \frac{W_N}{L_N} V_t^2 \exp\left(\frac{V_{GS} - V_{TH}}{n V_t}\right) \left[1 - \exp\left(-\frac{V_{DS}}{V_t}\right)\right]$$

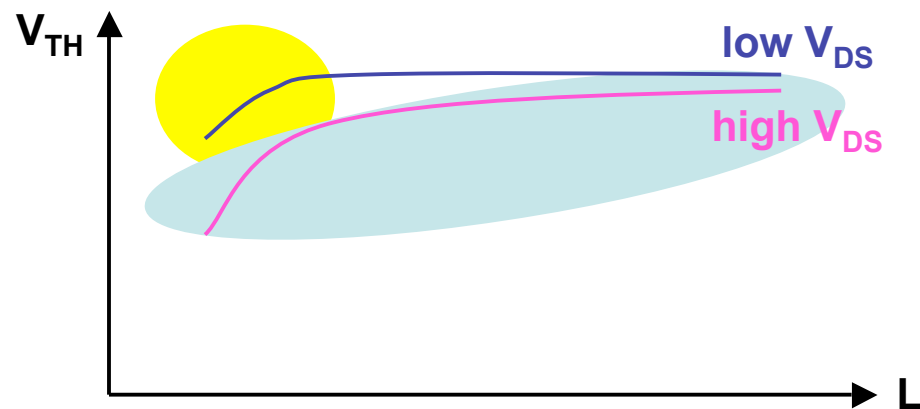
**Source: A. Ferre and J. Figueras, Low Power Electronics Design (C. Piguet, ed.), Chapter 3, CRC Press, 2004.**

# Subthreshold leakage

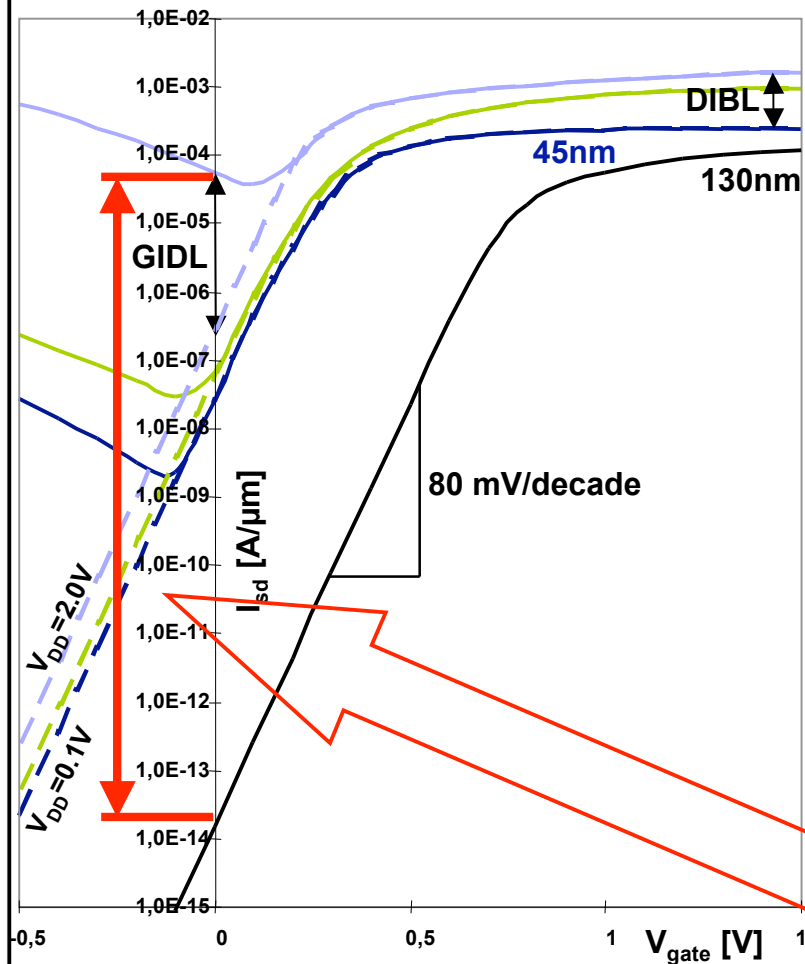
Subthreshold "off" current (for  $V_{GS} = 0$  and  $V_{DS} \gg V_T$ ):

$$I_{subth} = \mu_n C \frac{W_N}{L_N} V_t^2 \exp\left(\frac{-V_{TH}}{n V_t}\right)$$

- increases **exponentially** when threshold voltage is reduced
- increases with decreasing channel length:
  - direct dependence ●
  - short channel effect (shorter channel  $\rightarrow$  lower  $V_{TH}$ ) ●
  - drain induced barrier lowering (DIBL) effect ●
- increases with temperature



# Subthreshold leakage and GIDL



BSIM4 simulations using BPT models

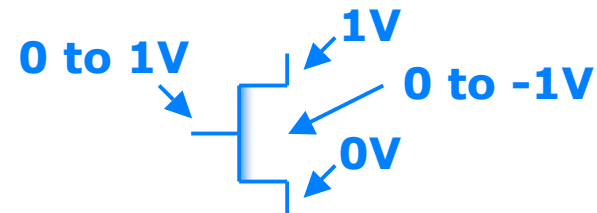
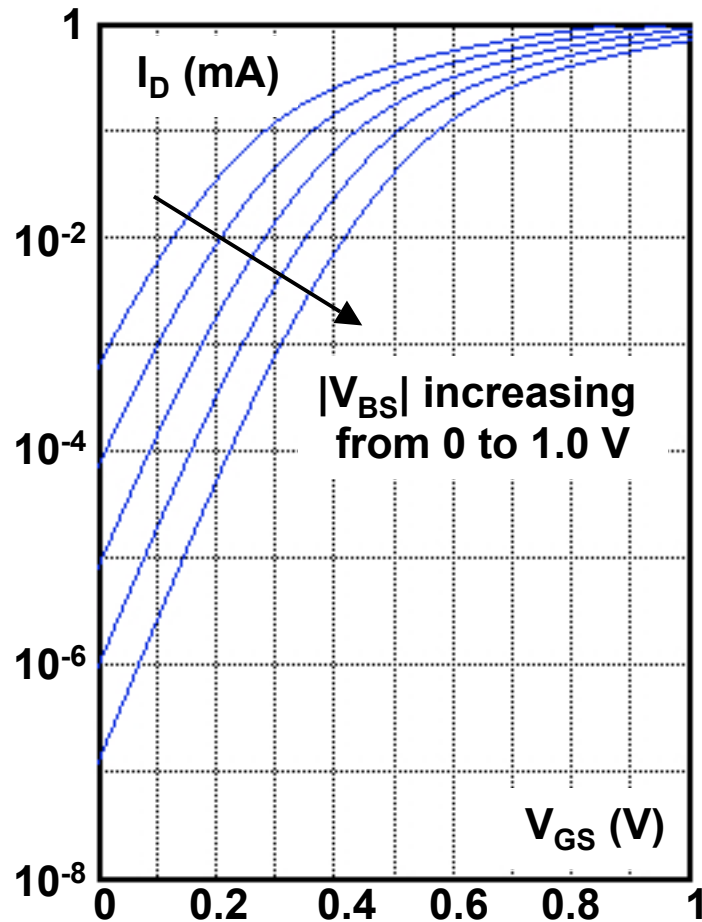
- Classical  $I_{sd}(V_{gate})$  behaviour for 130nm
- For 45nm the off current rises by  $10^6$ .  
→ subthreshold current in off-state  $I_{off}$
- Threshold voltage is  $V_{ds}$  dependent.  
→ increasing  $V_{ds}$  decreases  $V_{th}$   
→ drain induced barrier lowering (DIBL)
- If  $V_d > V_g$ , the drain depletion layer gets thinner due to the drain-gate field.  
→ more electrons pass the lower barrier  
→ the resulting  $I_{db}$  increases  $I_{off}$   
→ gate induced drain leakage (GIDL)

**Overall:  $I_{off}$  increases by  $10^9$**

BPT models: <http://www.eas.asu.edu/~ptm/>

# Subthreshold and body effect

**Body effect:  $V_{TH}$  increases with increasing reverse body bias ( $V_{BS}$ ). This in turn reduces the subthreshold current**

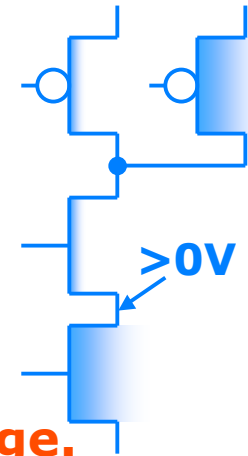
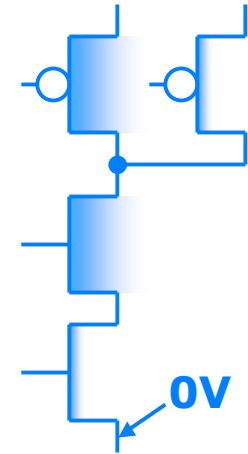


**BSIM4 simulations**  
**NMOS 65 nm process**  
**W=65 nm, L = 1  $\mu$ m**  
 **$V_{DS} = 1$  V**

# Subthreshold leakage

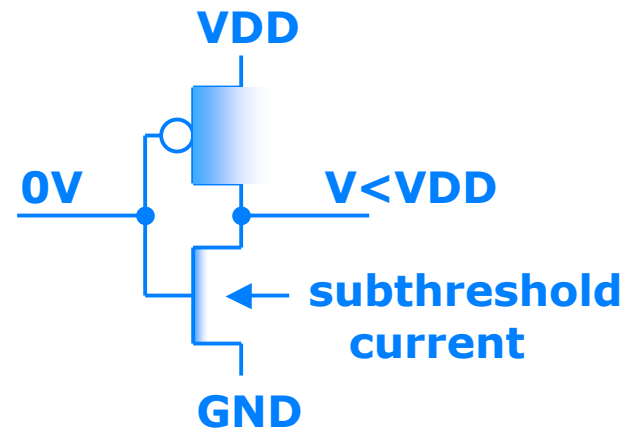
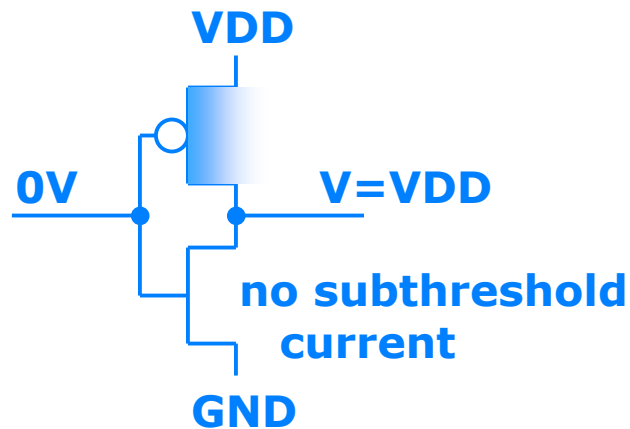
$$V_{TH} = \underbrace{V_{FB}}_{\text{Flatband Voltage}} + \underbrace{\Phi_S(T)}_{\text{Surface Potential}} + \underbrace{\gamma \left( \sqrt{\Phi_S + V_{BS}} - \sqrt{\Phi_S} \right)}_{\text{Body Effect}}$$

- **The body effect is important for stacked transistors.**
- **If the n-transistor connected to GND is closed:**
  - $V_{bs}$  is at 0V → no body effect
- **If the n-transistor connected to the output is closed:**
  - Due to the finite resistance of the open n-transistor, the source of the closed one is higher than 0V.
  - The body effect will result in a higher threshold and thus in a lower leakage current of the closed transistor.
- **This stacking effect causes data dependence of leakage.**



# Subthreshold leakage

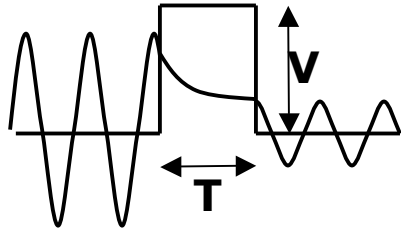
High subthreshold leakage current may affect logic levels



This can be a problem at elevated temperatures ( $I_{on}$  decreases while  $I_{off}$  increases with temperature).

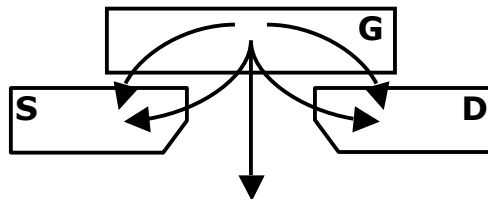
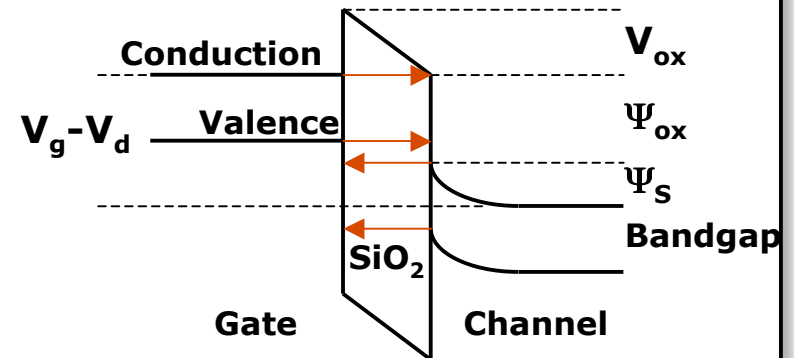
**Change in logic levels affects leakage in gates driven by the "leaking" gate.**

# Gate leakage



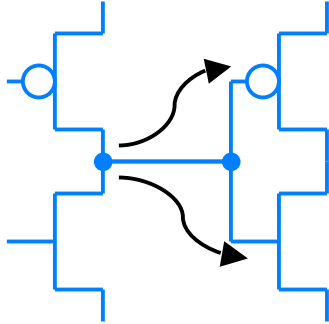
- **Tunneling: Electrons can pass potential barriers, higher than their energy (classically impossible).**
- **Tunneling current exponentially depends on barrier height  $V$  and width  $T$  and on electron's mass  $m_{\text{eff}}$ .**

- **Energy diagram of a Poly-Si transistor:**
- **Leakage current can be carried by tunneling electrons or holes.**
  - direct tunneling: from gate to channel
  - Fowler-Nordheim: from gate to oxide



- **Carriers leak to source, drain and channel**
- **Channel leakage is sub-divided into source, drain and bulk leakage**

# Gate leakage

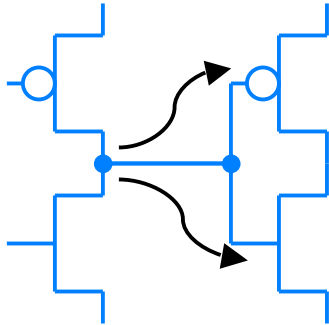


Unlike other leakage components, gate leakage changes the way CMOS static gates work: DC current flows from gate output nodes to input nodes of other gates (the “loading effect”).

## Some consequences:

- logic levels and noise margins may be affected,
- gate fanout is limited by the DC load,
- **total leakage in a digital circuit can no longer be calculated as a sum of leakages of all gates treated independently.**

# Gate leakage



**The loading effect may either increase or reduce the total leakage of the gate.**

**R. Mendoza, R. Ferre, L. Balado and J. Figueras, in Proc. IEEE Int. Conf. on Design and Test of Integrated Systems in Nanoscale Technology, Tunis, Sept. 2006.:**

**1% error** in estimated total leakage of a logic block ( $t_{ox}=1.85$  nm)

**S. Mukhopadhyay, S. Bhunia and K. Roy, IEEE Trans. on CAD, vol. 25, no. 8, 2006, pp. 1486-1495:**

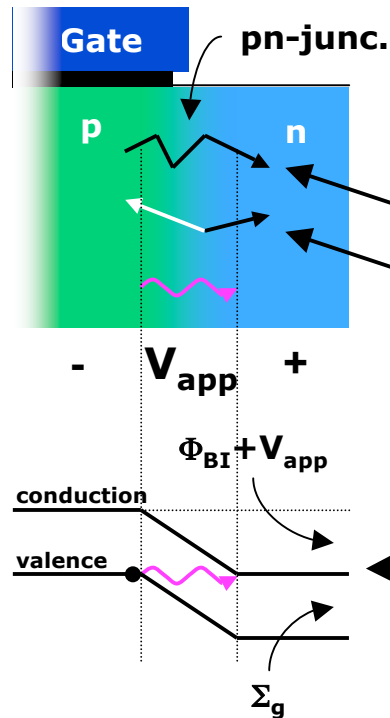
**5% to 8% error** for a single gate ( $t_{ox}=1.0$  nm) but lower for bigger logic blocks

**A. Rastogi, W. Chen and S. Kundu, in Proc. DAC 2007, San Diego, June 2007, pp. 712 - 715:**

**from 1.9% up to 17.2% error** for various ISCAS benchmark blocks

# Junction leakage

## pn-junction in reverse bias



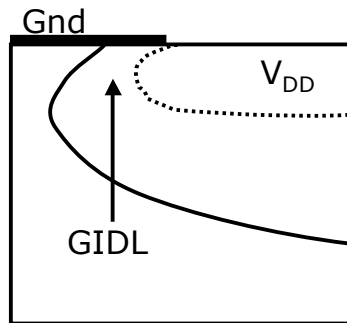
- **As known from diodes: small currents are carried by**

- minority carriers drifting across the junction
- electron-hole generation in junction
- impact ionization at high reverse bias (normally exceeding max. operating voltage)

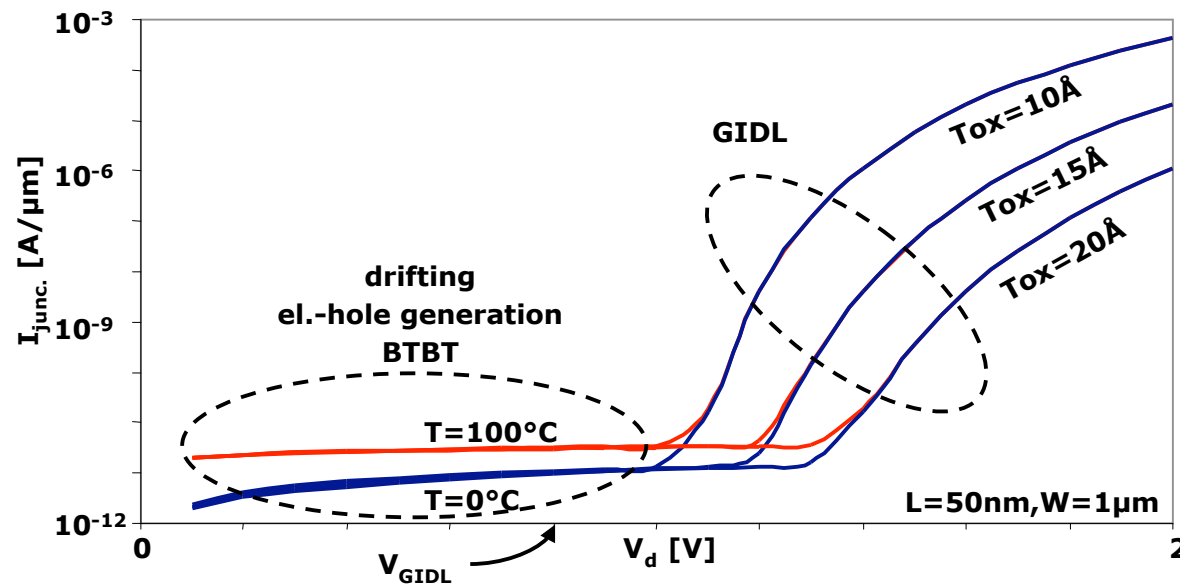
▪ **not the dominant leakage components in DSM CMOS**

- **An electron in the p-side's valence band sees a potential as shown**
- **Electrons can pass this triangular barrier by tunneling through it → Band-To-Band-Tunneling (BTBT) current**
- **can be a significant leakage component**

# Junction leakage and GIDL



- **Smaller barrier means exponentially higher BTBT:**
  - tech scaling: steeper doping profiles
  - Gate Induced Drain Leakage (GIDL)
- **GIDL: field from drain to gate reduces junction width**

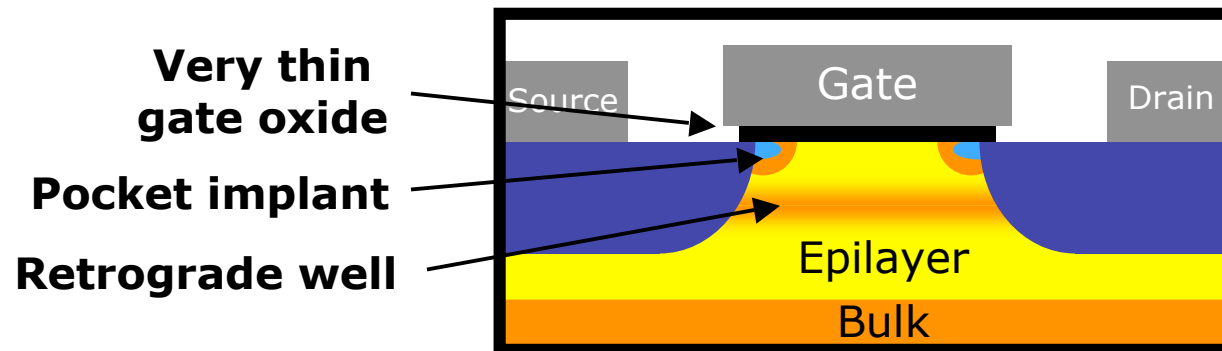


## Technology and device design

**Can leakage currents be eliminated or reduced by better device design and manufacturing technology?**

- **Subthreshold “off” current is not a “parasitic” phenomenon and as such cannot be eliminated.**
- **Higher  $V_{TH}$  and longer channels help to reduce it at a price of reduced performance:**
  - **“Low  $V_{TH}$ /high performance” and “High  $V_{TH}$ /low leakage” device versions**
  - **Channel length can be selectively increased for some devices/gates in a circuit to reduce “off” current without excessive performance penalty**
  - **Gate and channel engineering (thin  $T_{ox}$ , pocket implants) reduces short channel effect on  $V_{TH}$  (but may increase other leakage components)**

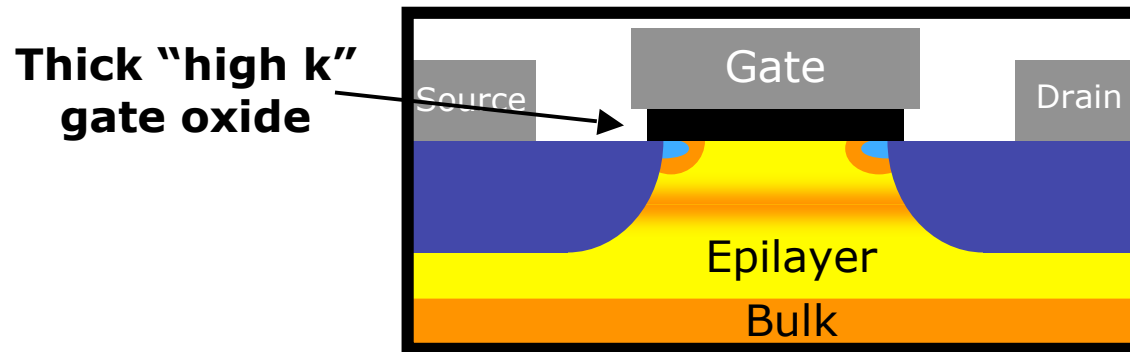
## Technology and device design



**A tradeoff: careful device engineering allows to maximize the  $I_{on}/I_{off}$  ratio and minimize the short channel effect, but some other leakage current components may increase:**

- **very thin gate oxide -> gate tunneling**
- **high doping at the channel/drain junction -> GIDL and BTBT**

## Gate leakage



**New gate dielectric materials with higher dielectric constant (SiON, HfSiON) are sought. Such a gate dielectric could be thicker without performance penalty, thus reducing gate tunneling current.**

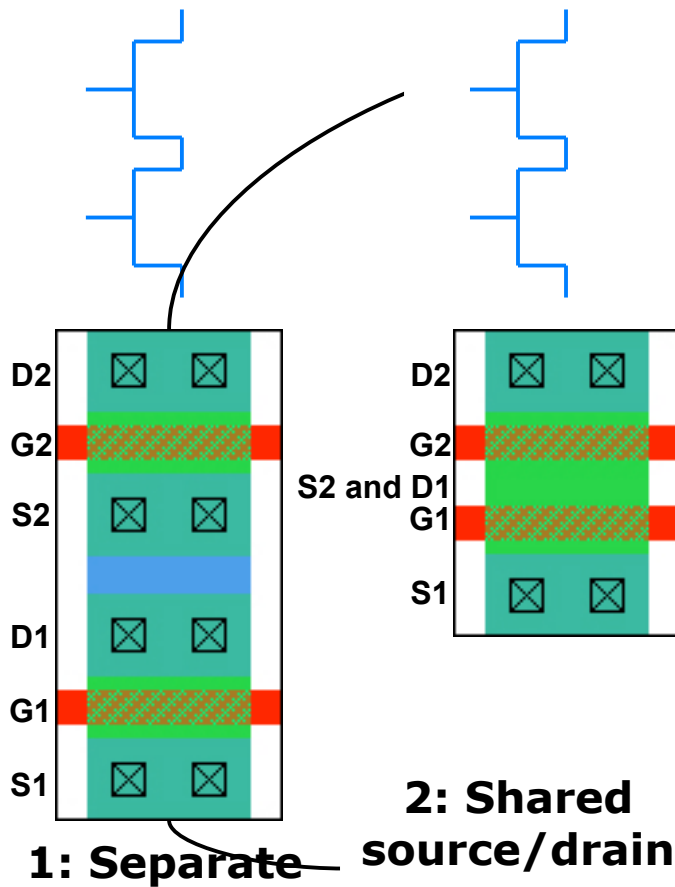
## Device design: scaling factors

Leakage component	Scaling factor				
	W	W/L	W*L	AD (AS)	PD (PS)
Subthreshold current					
GIDL					
Gate tunneling current					
Junction leakage + BTBT					

**For accurate leakage modeling channel aspect ratio (W/L) is not sufficient, complete device layout must be known**

# Leakage and device layout

**Two MOS transistors connected in series**



**Pre-layout simulation:**

→ MN1 1 2 3 0 Nchan W=325E-9 L=65E-9  
+PD=1040E-9 AD=6.34E-14

→ MN2 3 4 5 0 Nchan W=325E-9 L=65E-9  
+PD=1040E-9 AD=6.34E-14

**OK for layout 1**

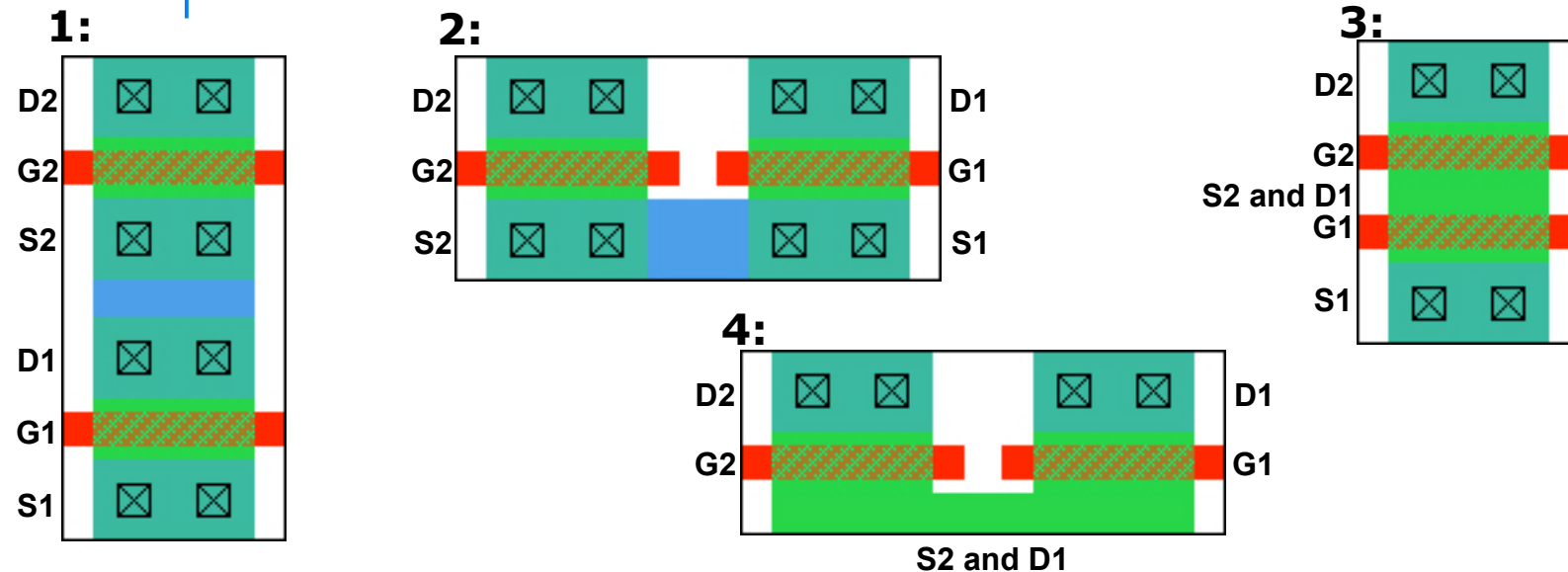
**WRONG for layout 2: actual area and perimeter of D1 and S2 much lower than the sum for individual drain/source junctions**

**The same schematic, the same W and L, but different S1 and D2 areas and perimeters -> different leakage**

# Leakage and device layout



**Two MOS transistors connected in series:  
1,2: separate devices; 3,4: shared source/drain**



**Post-layout simulation (after extraction): to avoid ambiguities and either missing or duplicated S/D areas, transistors and S/D diodes should be extracted separately -> PSP and JUNCAP2 models**

## Leakage physics: a summary

- **Leakage currents depend in a complex way on the properties of *device structure* (doping, gate oxide thickness, channel dimensions etc.) and *layout* (-> *intra-device correlations of compact model parameters*)**
- **Leakage currents not only increase the overall power consumption, but also may change logic levels (-> *leakage is data dependent in a complex way*)**
- **Total leakage in a logic block is not a simple sum of leakage currents in gates treated individually, *leakage currents are correlated***

# Why is variability important

The problem: **highly nonlinear** dependences of some leakage currents on parameters that are subject to process variability.

Examples:

- subthreshold current as a function of the threshold voltage
- gate tunneling current as a function of gate oxide thickness

Subthreshold “off” current:

$$I_{subth} = \mu_n C \frac{W_N}{L_N} V_t^2 \exp\left(\frac{-V_{TH}}{nV_t}\right)$$

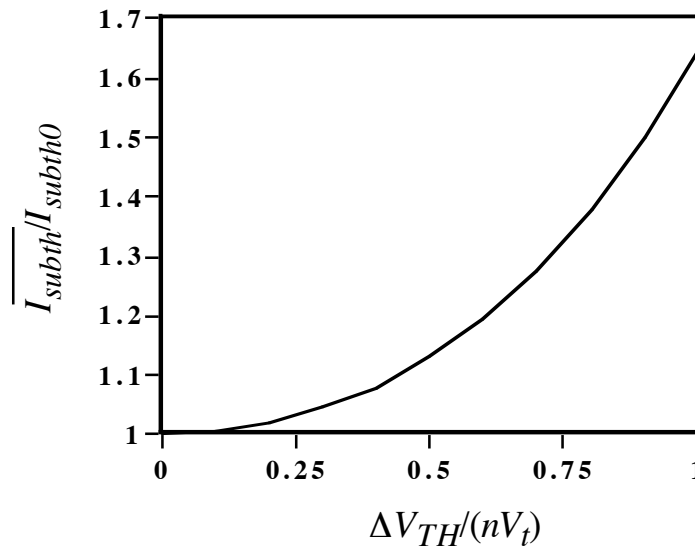
If  $V_{TH}$  is a random variable distributed normally with mean value  $V_{TH0}$  and standard deviation  $\Delta V_{TH}$ , the distribution of the “off” current  $I_{subth}$  is log-normal with mean value

$$\overline{I_{subth}} = I_{subth0} \exp\left[\frac{1}{2}\left(\frac{\Delta V_{TH}}{nV_t}\right)^2\right] \quad (I_{subth0} = I_{subth} \text{ for } V_{TH} = V_{TH0})$$

## Why is variability important

For  $V_{TH}$  variability of the order of  $nkT/q$  ( $n \cdot 26$  mV;  $n$  between 1 and 2) the mean value of the "off" current increases by about 65%.

If this is local (intra-die) variability, the sum of "off" currents of all devices increases accordingly.



**The mean value and the total subthreshold "off" current increases with increasing variability**

# Sources of process variability

## Litography: variability of channel dimensions

- directly affects  $I_{on}$ ,  $I_{off}$
- indirectly affects  $I_{on}$ ,  $I_{off}$  via  $V_{TH}$  variations (short channel effect)

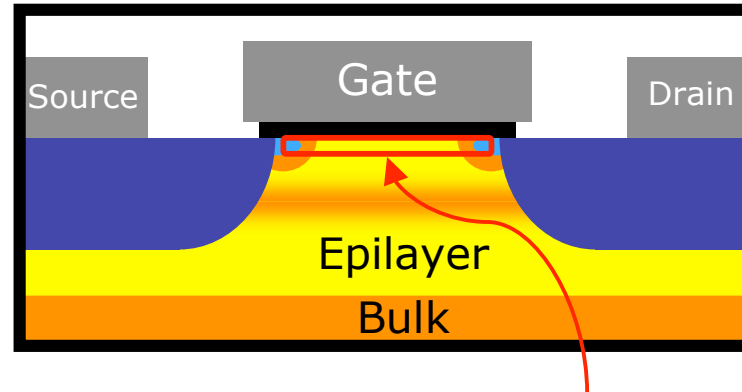
## Variability of gate oxide thickness

- directly affects  $I_{on}$ ,  $I_{off}$  via  $C_{ox}$
- indirectly affects  $I_{on}$ ,  $I_{off}$  via  $V_{TH}$  variations ( $V_{TH}$  depends on  $C_{ox}$ )

## Variability of doping concentrations and profiles

- indirectly affects  $I_{on}$ ,  $I_{off}$  via:
  - $V_{TH}$  variations
  - variations of actual channel length
  - variations of body effect factor

## Process variability: channel doping



Threshold voltage is a function of doping in the channel

However, doping in the channel means **single atoms** in nanometer CMOS: the average number of boron atoms in the channel of a 32 nm device is **3.5**.

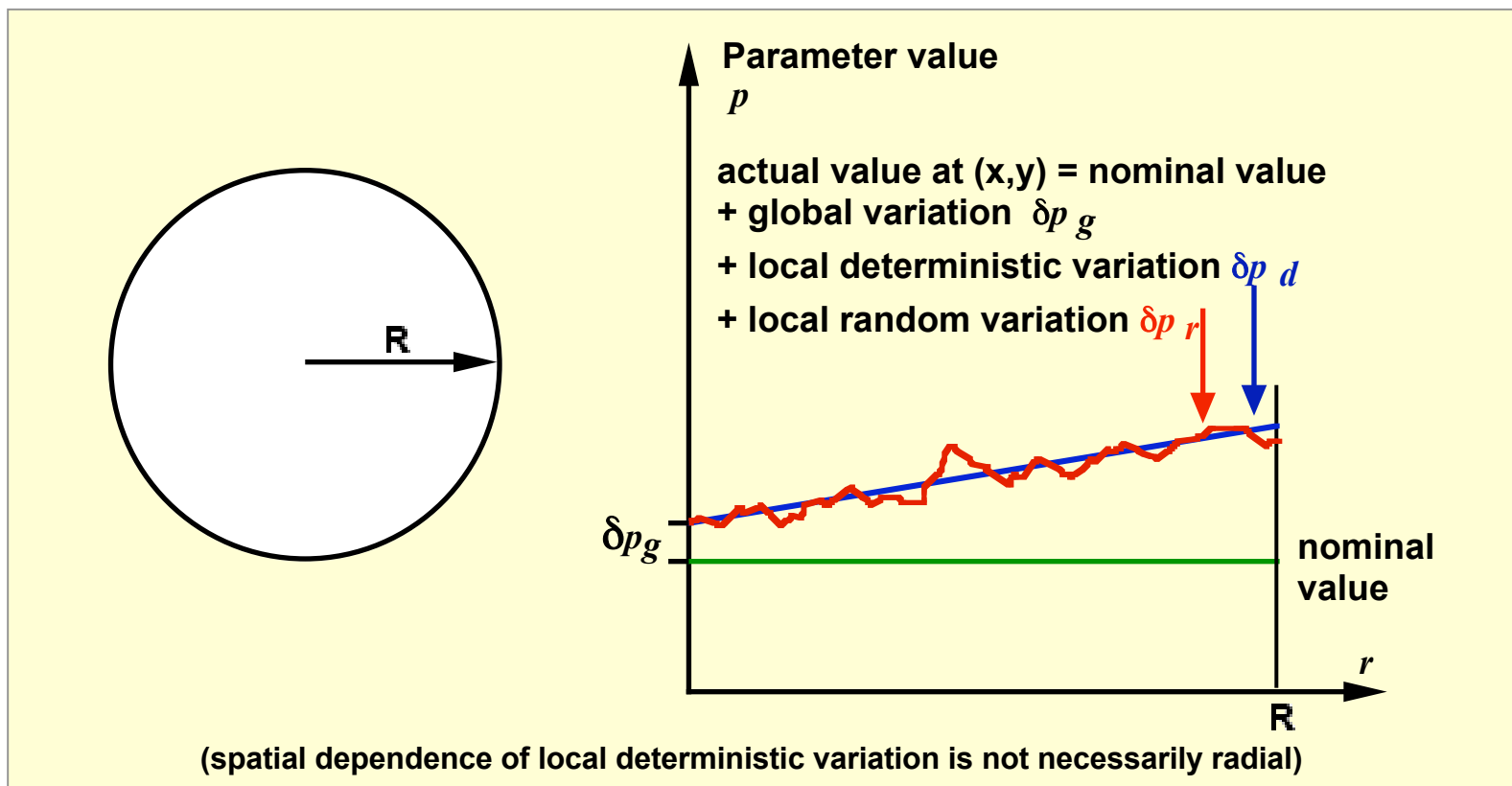
-> 3 atoms in one device and 4 atoms in another means **25% difference of the doping dose.**

Source: T. Skotnicki, "Nano-CMOS & Emerging Technologies—Myths and Hopes", Plenary presentation PL-1, 2006 Int. Conf. on Solid State Devices and Materials, Yokohama, Sept. 2006.

# Process variability: variation types

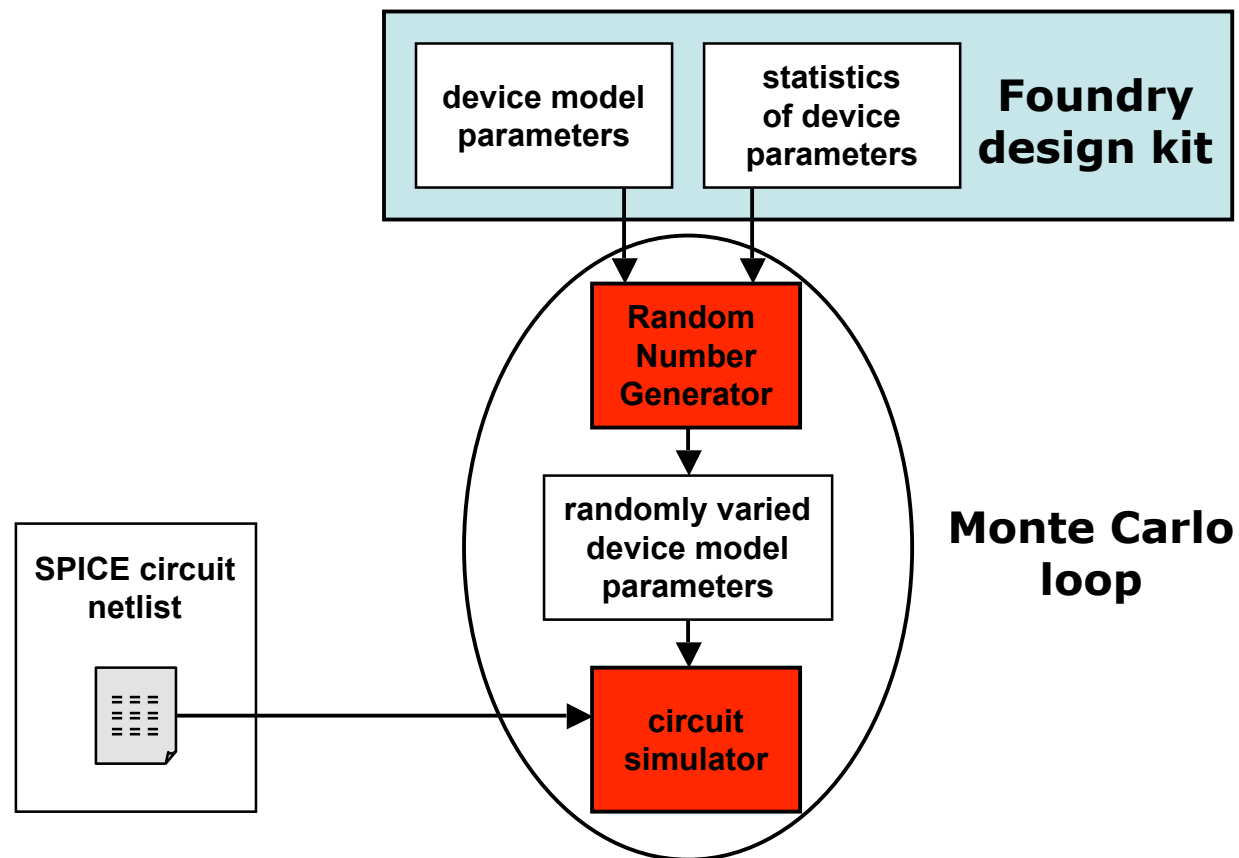
## Process variations

Local value of a parameter  $p$ :  $p = p_{nom} + \delta p_g + \delta p_d(x, y) + \delta p_r$



# Process variability: modeling methods

## Compact model based methodology: statistical Monte Carlo simulation in a circuit simulator



## **Process variability: modeling methods**

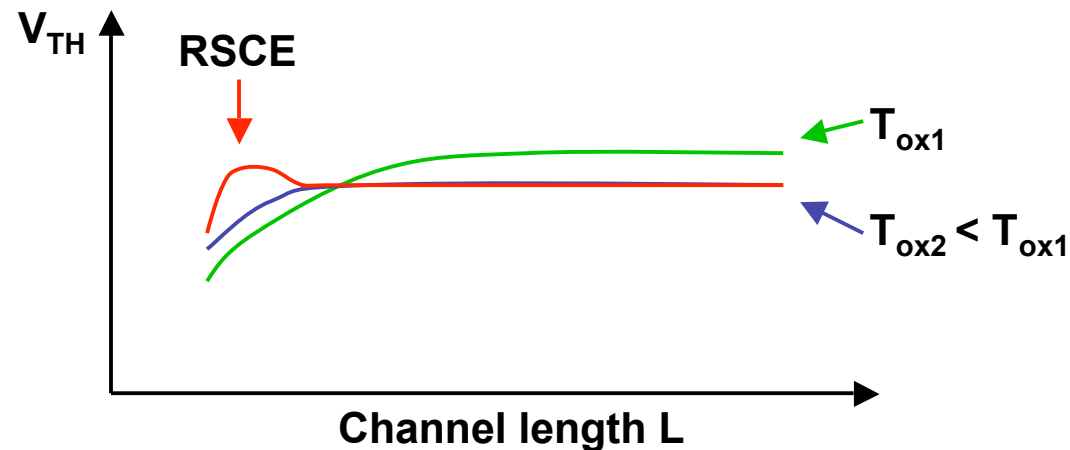
### **Compact model based methodology: statistical Monte Carlo simulation in a circuit simulator**

- **Advantages:**
  - **Well established**
  - **Only standard EDA tools needed**
  - **Limited computational complexity**
- **Disadvantages**
  - **It is difficult to include correlations (between model parameters in a single device and between parameters of various devices in the circuit)**
  - **No direct link to physical design and to manufacturing**
  - **Expensive and time consuming experimental characterization of test devices needed**

## Process variability: modeling methods

### Compact model based methodology: the problem of **intra-device** correlations

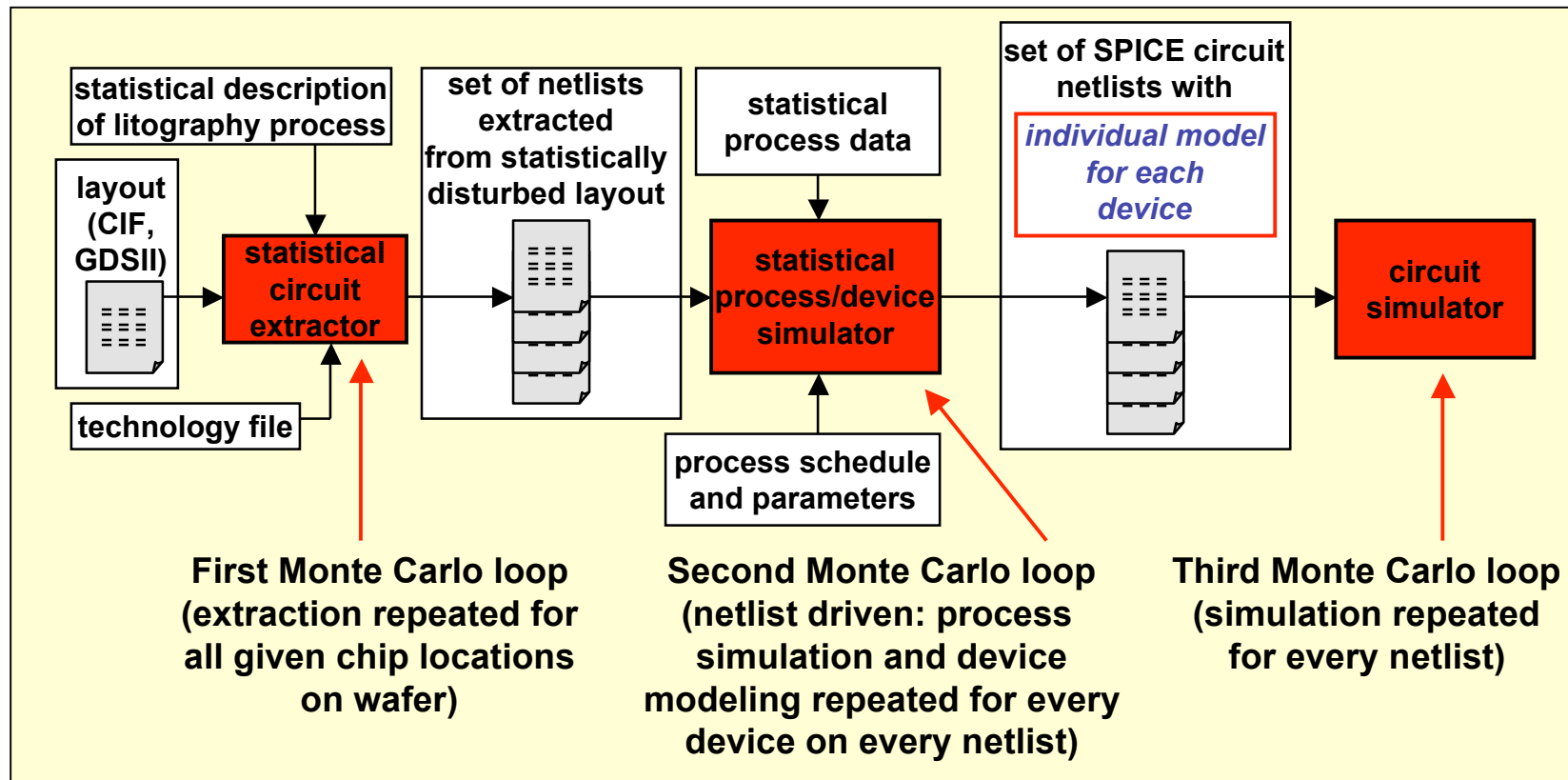
**Intra-device correlations:** e.g. threshold voltage is correlated with gate oxide thickness and channel length



**These correlations are not accounted for in compact model based Monte Carlo simulation.** The mathematical technique is known (K. S. Eshbaugh, IEEE Trans. on CAD, vol. 11, no. 10, 1992, pp. 1198-1206) but computationally infeasible in most cases.

# Process variability: modeling methods

## “Virtual manufacturing” methodology: statistical (Monte Carlo) process, device and circuit simulation



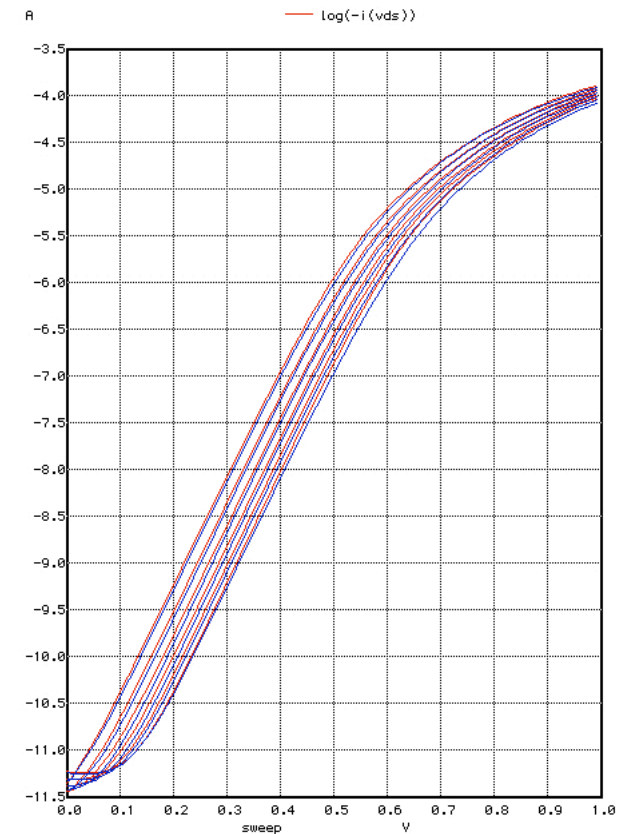
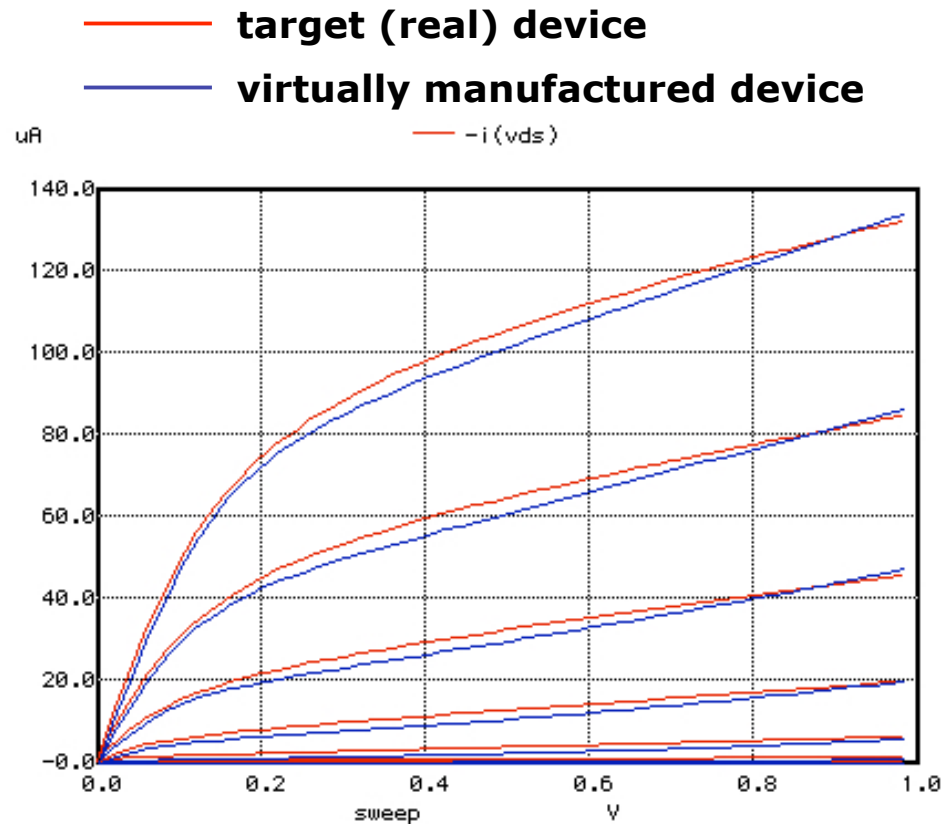
## Process variability: modeling methods

### Virtual manufacturing methodology: statistical (Monte Carlo) process, device and circuit simulation

- **Advantages:**
  - Based on statistical process and device simulations, with **process variabilities accounted for**
  - **All correlations** taken into account
  - **Actual device layout** taken into account
  - **Allows “what if” experiments** with processing parameters, may show directions to process and device design optimization
- **Disadvantages**
  - **Special software needed**
  - **Process must be known or at least reasonable assumptions must be made; fitting to nominal characteristics of real devices needed**
  - **Significant computational resources needed for large statistical samples**

# Virtual manufacturing

## Virtual technology fitted to an industrial 65 nm technology: 65 nm NMOS device

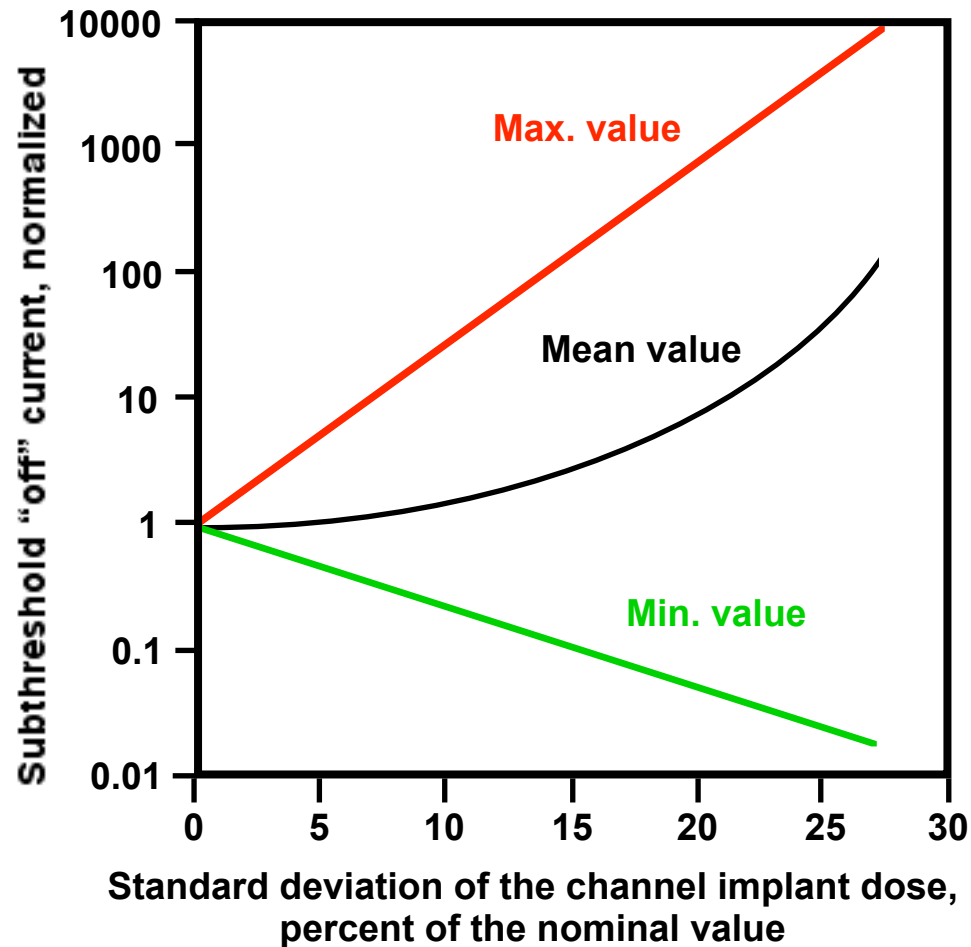


# Virtual manufacturing

## Virtual technology: variability of "off" drain current vs. doping variability

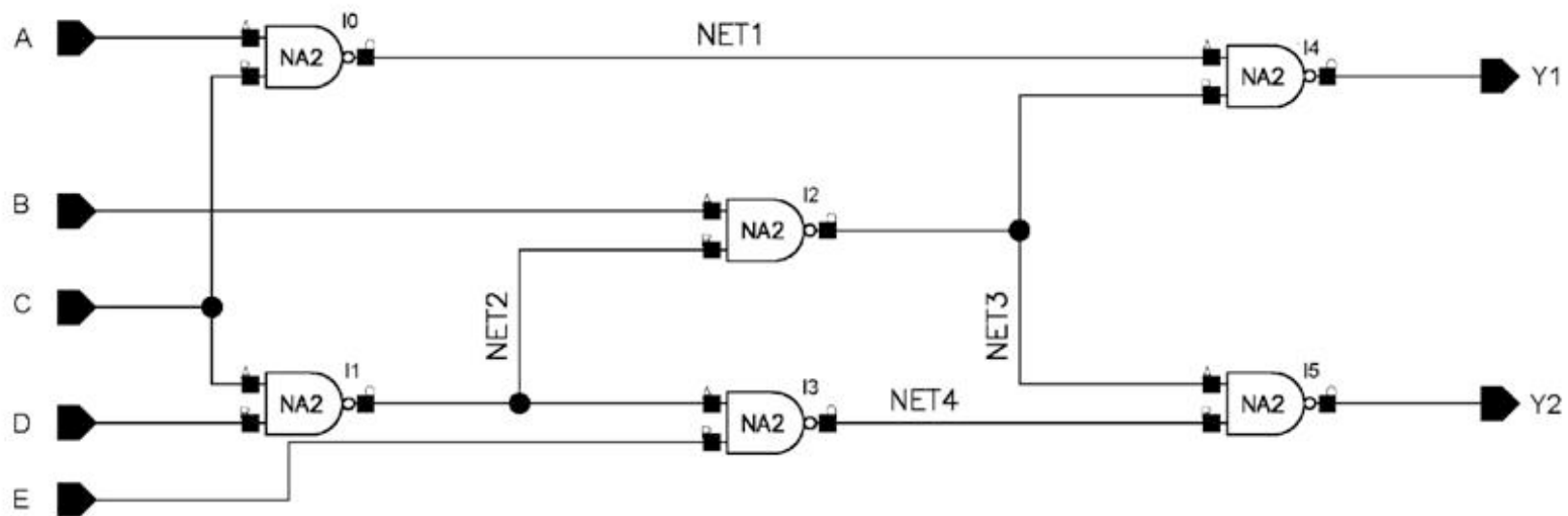
nMOSFET,  
L=65 nm

"Off" drain current  
vs. variability  
of channel implant  
dose



# Examples

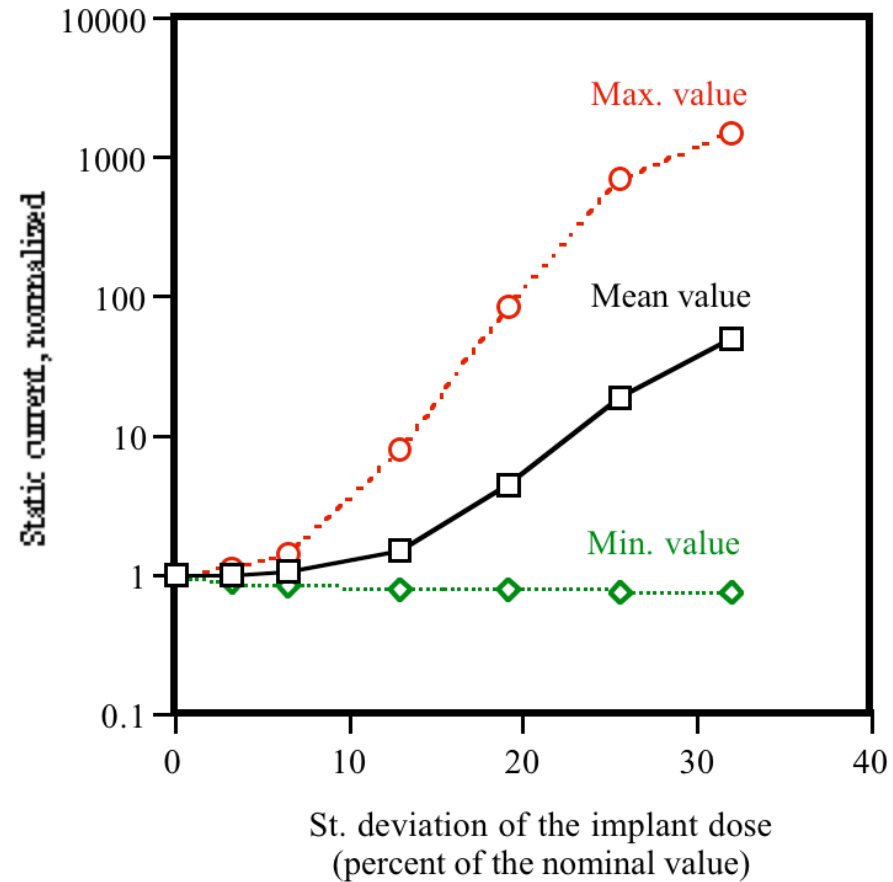
## Example: a small digital block (c17 ISCAS benchmark)



# Examples

## C17 benchmark, version 2 (NMOS W/L = 10, PMOS W/L = 10)

**Total leakage power vs. local variability of the channel doping**

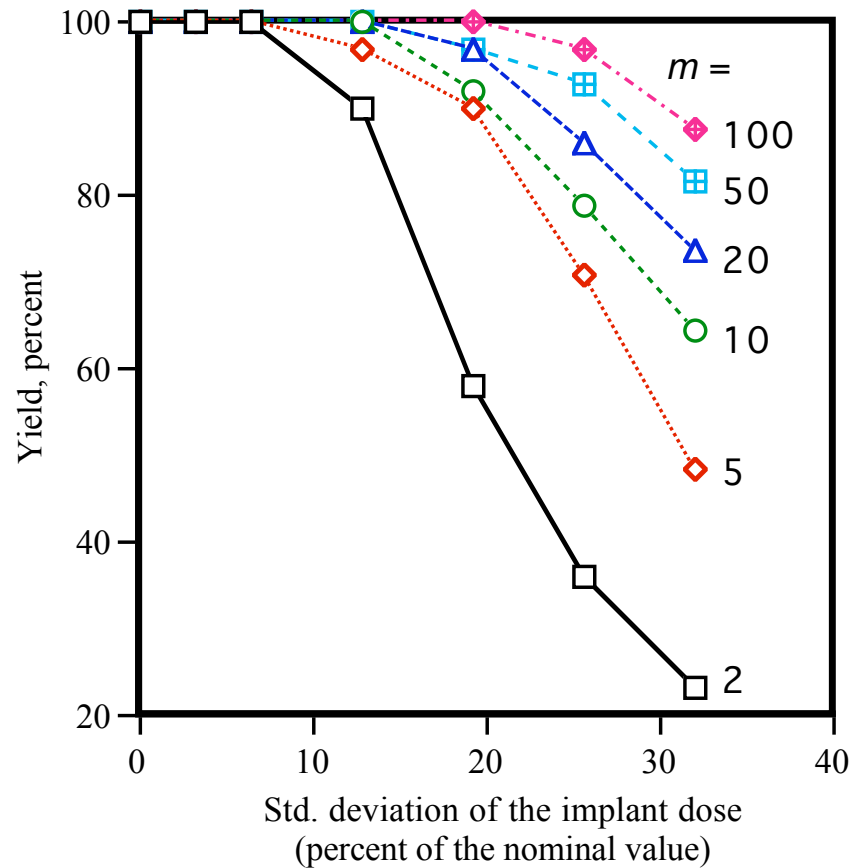


# Examples

## C17 benchmark, version 2 (NMOS W/L = 10, PMOS W/L = 10)

**Yield vs.  
local variability  
of the channel doping**

**$m$  is the ratio of maximum  
acceptable static power  
to the nominal (i.e. zero  
variability) static power**



## Conclusions (1)

- **Leakage effects are greatly enhanced due to process variabilities**
- **Some of these variabilities are of fundamental nature and cannot be reduced by means of better equipment and/or process control**
- **Variability effects are not easy to be simulated, simple Monte Carlo simulation may sometimes be misleading**
- **Process and device simulation are theoretically more accurate but require some knowledge of process details**

## Conclusions (2)

- **Local (intra-die) process variabilities increase **total** static power consumption**
- **Neglecting variability in circuit and system level modeling may significantly underestimate the total leakage power**
- **Although variability may dramatically increase the average total leakage power, yield loss due to excessive leakage is moderate**
- **... but this does not mean that leakage is not a problem!**

## Conclusions (3)

- **High leakage in nanometer CMOS devices is unavoidable**
- **Circuit and system level solutions are needed, such as:**
  - **Threshold voltage optimization (leakage vs. performance tradeoff -> several device families)**
  - **Variable threshold voltage (adaptive body biasing, DTCMOS)**
  - **Power gating**
  - **Drowsy and sleep circuits**
  - **Leakage aware logic synthesis**