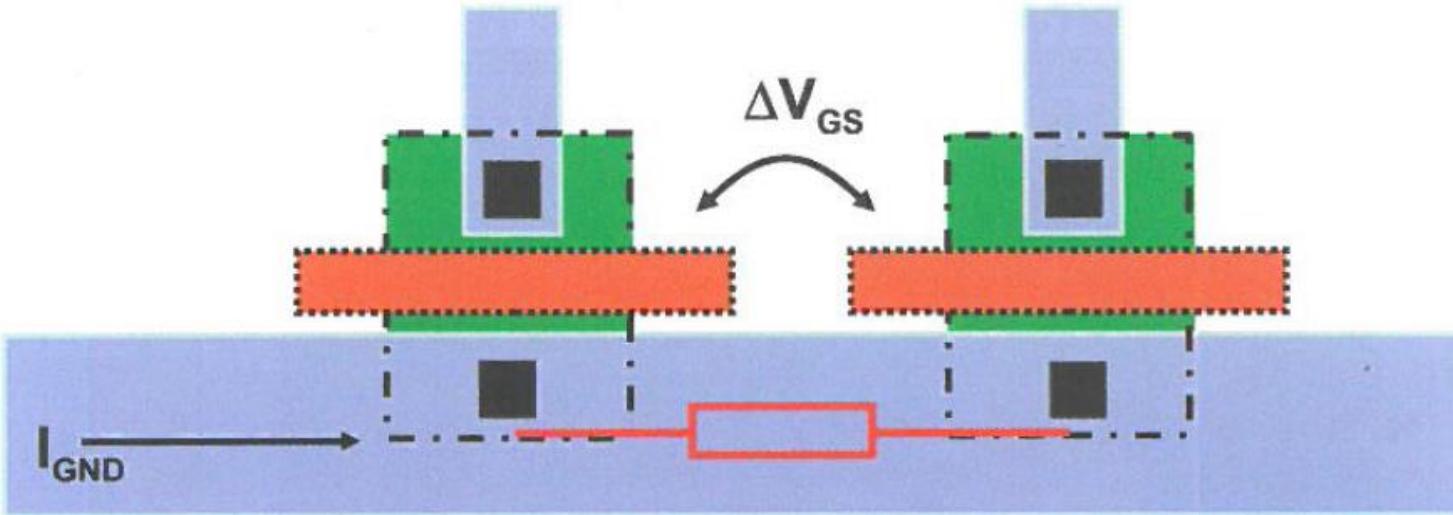


## **Podstawy Mikroelektroniki**

# **VLSI Process Mismatches**

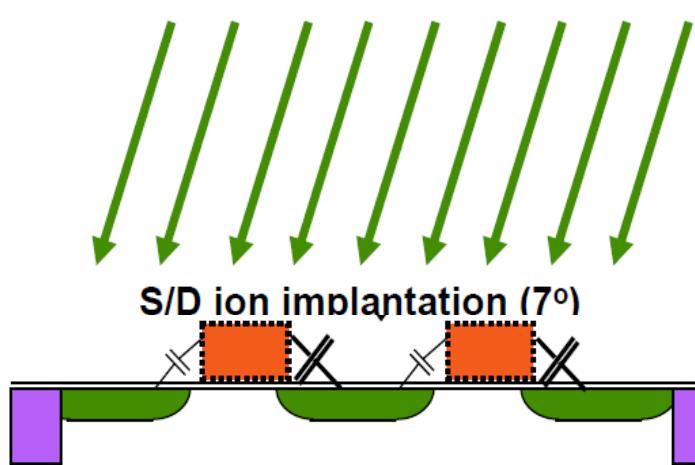
**P. Kmon, R. Kłeczek**

# Reasons of inequality

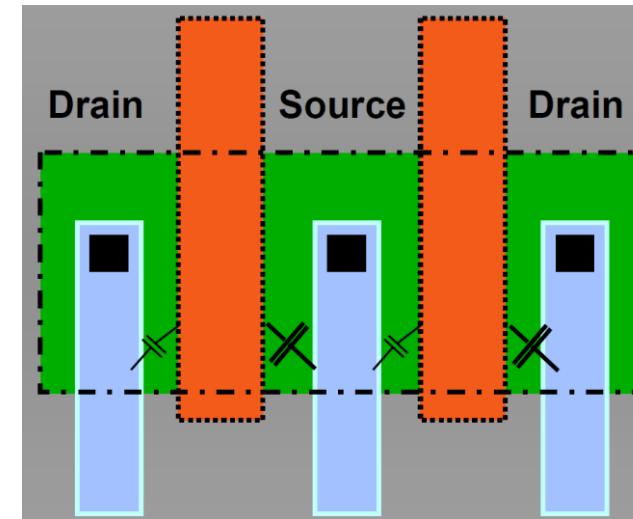


A small resistance in the source connection wire will result in offset in the gate voltage.

Cross-section view

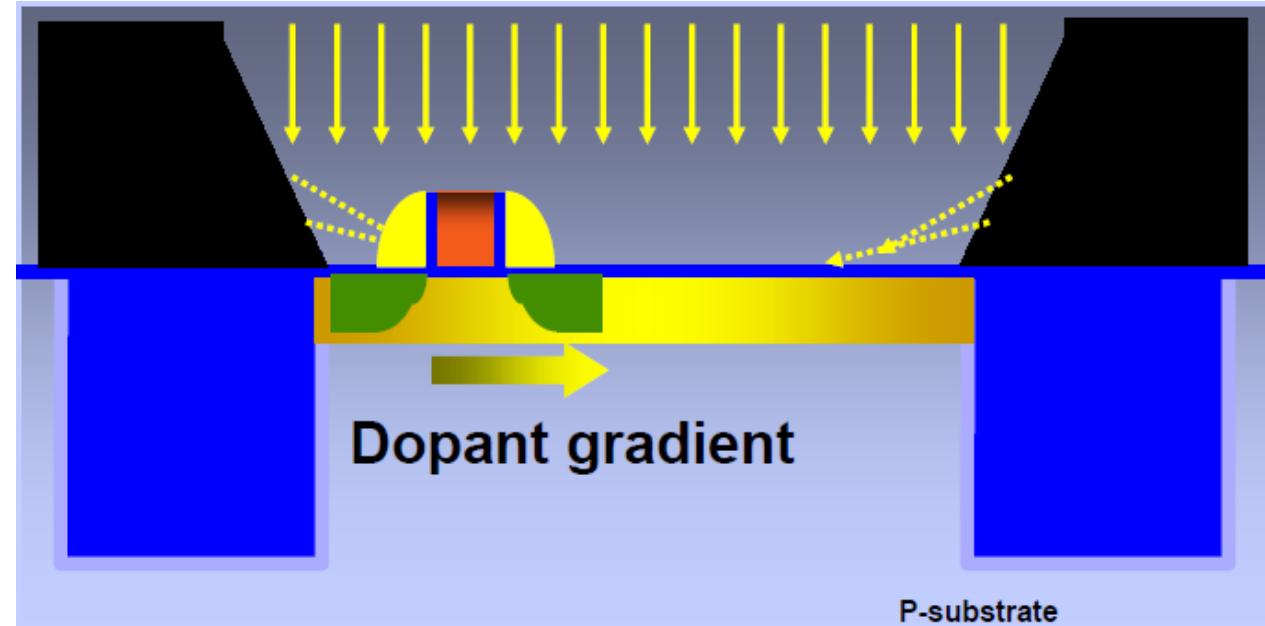


The source and drain diffusions are implanted under an angle. This causes asymmetry.

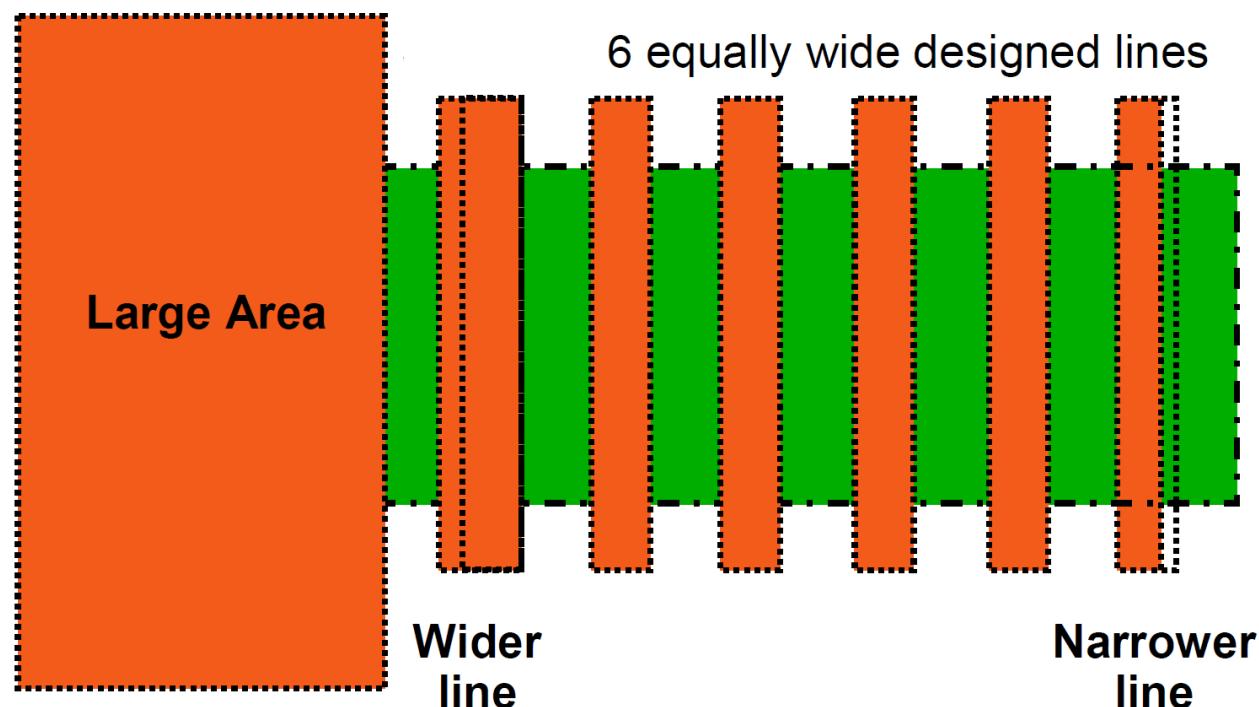


## Reasons of inequality

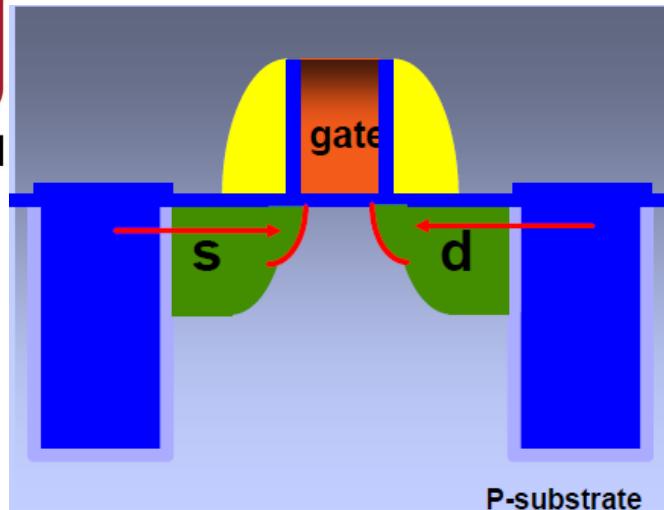
The **well-proximity effect** occurs during the well implant. The drawn transistor is fabricated afterwards, but resides in the well with a horizontal doping gradient due to scattering.



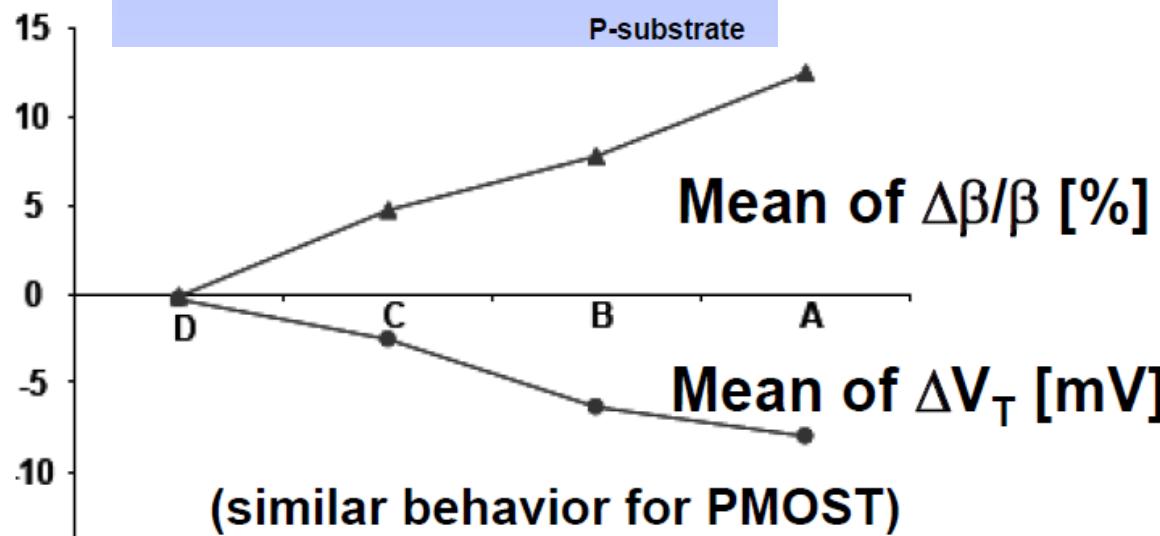
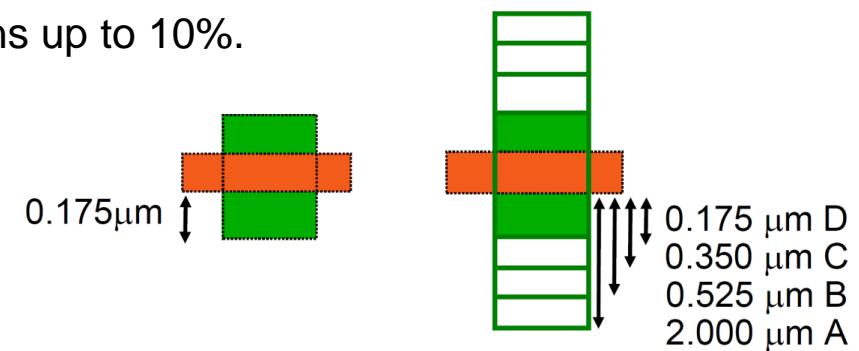
The **proximity effect**: lines with large neighbouring structures grow in size, while lines next to open space shrink. Size of the component depends on the structure density because of photo resist exposure and plasma etching (local loading effects).



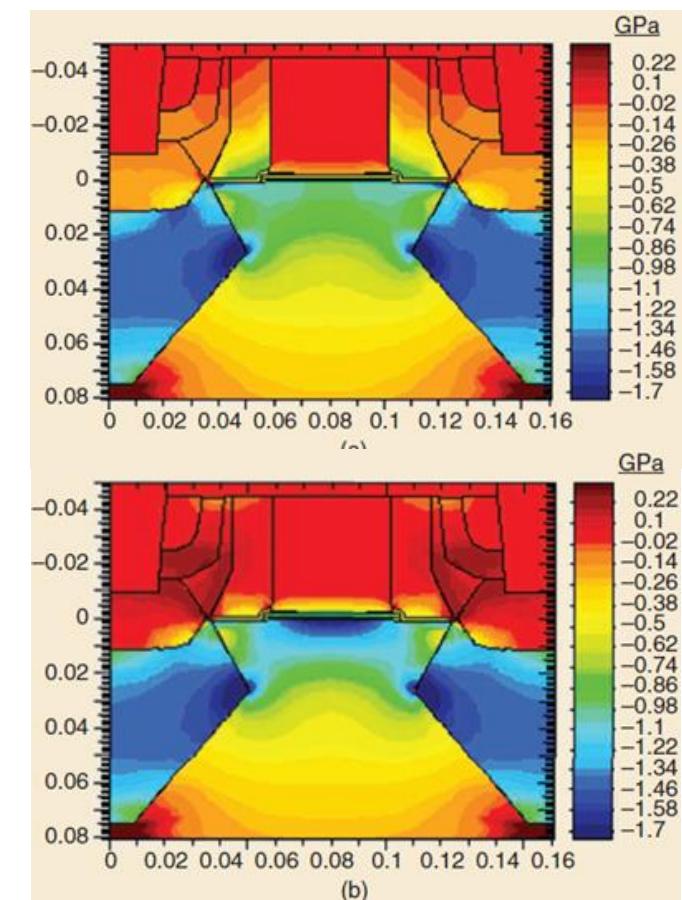
## Reasons of inequality



Stress in integrated circuits is caused by thermal expansion of various materials ( $\text{SiO}_2$ ,  $\text{Si}_3\text{O}_4$ , AL, Cu). The blocks of STI create considerable stress in the active area of the transistor causing current variations up to 10%.



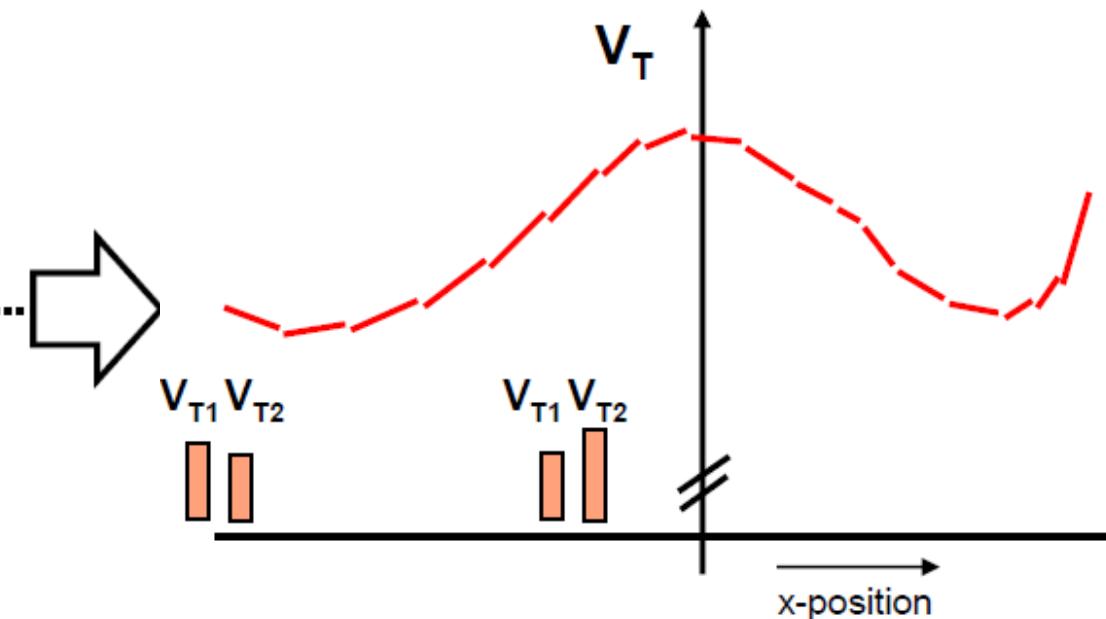
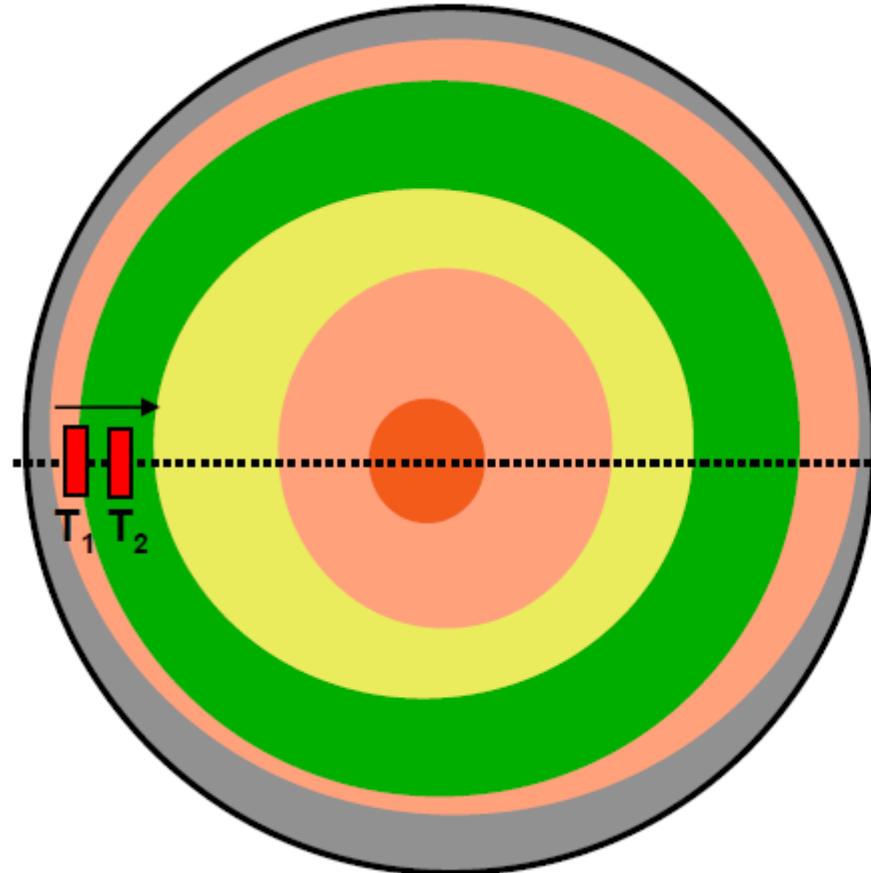
An experiment shows the influence of the STI edge on the drain current and threshold voltage. Top: a 65-nm 2  $\mu\text{m}$ /0.5  $\mu\text{m}$  NMOS reference transistor is designed with the STI edge of the source and drain at 2.0  $\mu\text{m}$ . A second device has a similar STI distance (A) 0.525  $\mu\text{m}$ , (B) 0.35  $\mu\text{m}$ , (C) 0.16  $\mu\text{m}$  (D). On the right stress in the channel w/wo dummy gates.



# Threshold voltage spread for various processes

Voltage threshold variations are due to:

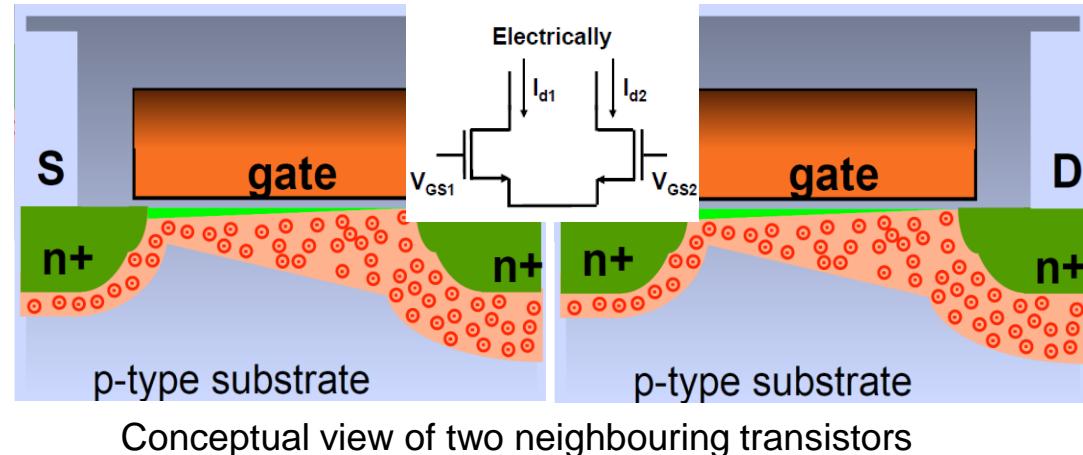
- temperature gradients in furnaces (oxidation, CVD, RTP),
- chemical non-uniformity's (gas or plasma non-uniformity's during deposition & etching),
- photo resist thickness variations,
- ion implantation channeling effects.



Parametric contour plot and voltage threshold measurements across the horizontal line.

# Reasons of inequality

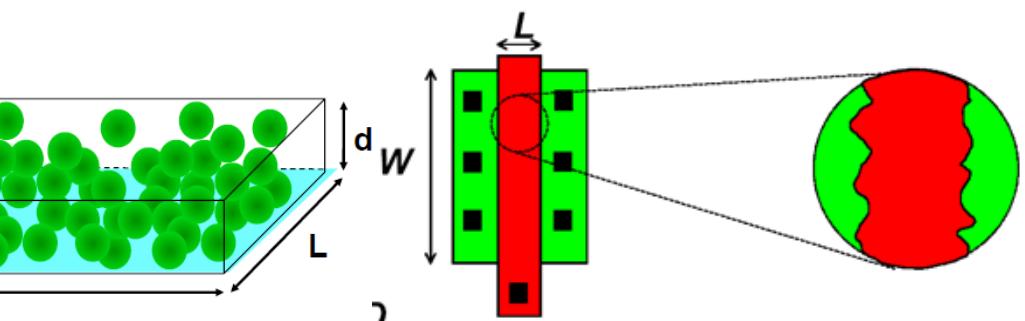
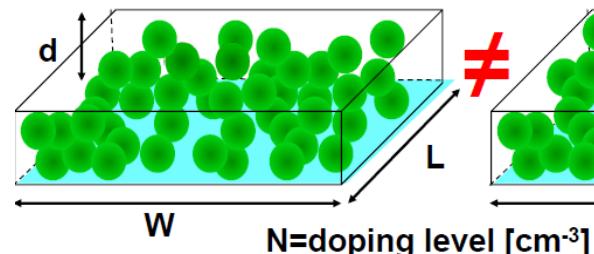
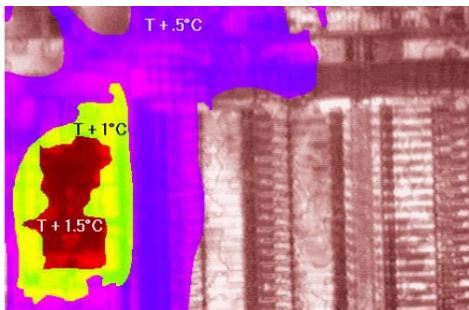
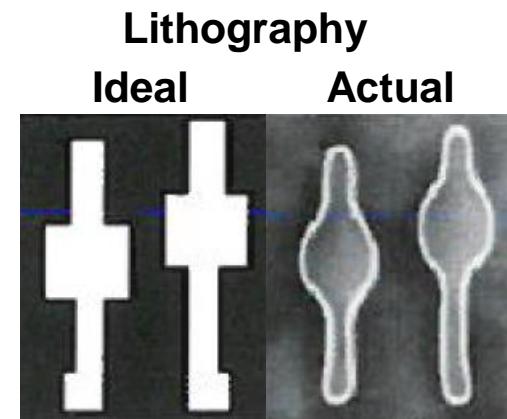
Parameters that reflect the behaviour of devices are the result of combination of a large number of microscopic processes, i.e. the conductivity of resistors and transistors and capacitance of the capacitors are built up of a large number of single events: the presence of ions in the conduction path, the local distances between the polysilicon grains that form capacitor plates, etc. All of these are the reasons of elements' parameters inequality.



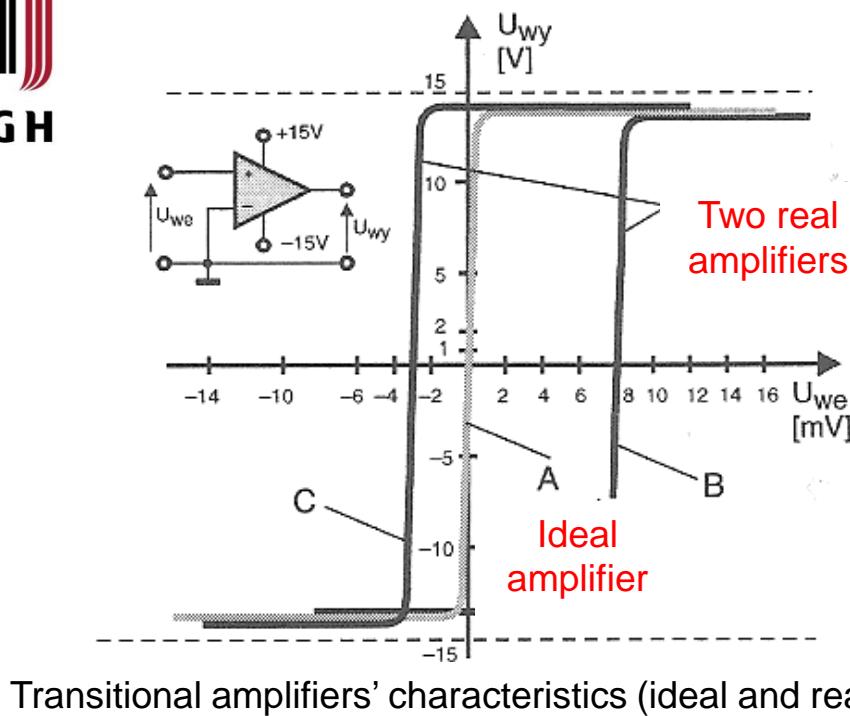
Conceptual view of two neighbouring transistors

There are three main groups of reasons that influence parameters uniformity:

1. **Deterministic** (supply voltage, process, temperature, lithography, stress, STI, IR drops, wiring differences).
2. **Pseudorandom** (cross talk, substrate noise, hot carrier, NBTI).
3. **Random** (dopant fluctuations, line edge roughness, noise, granularity, interface states, work-function fluctuations, fixed-oxide charges).

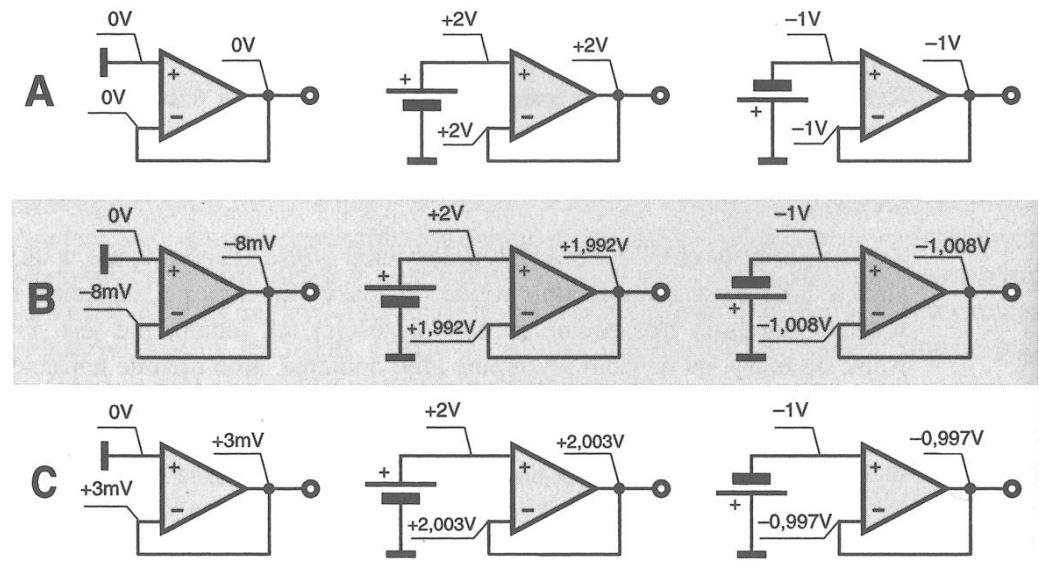


# Results of inequality

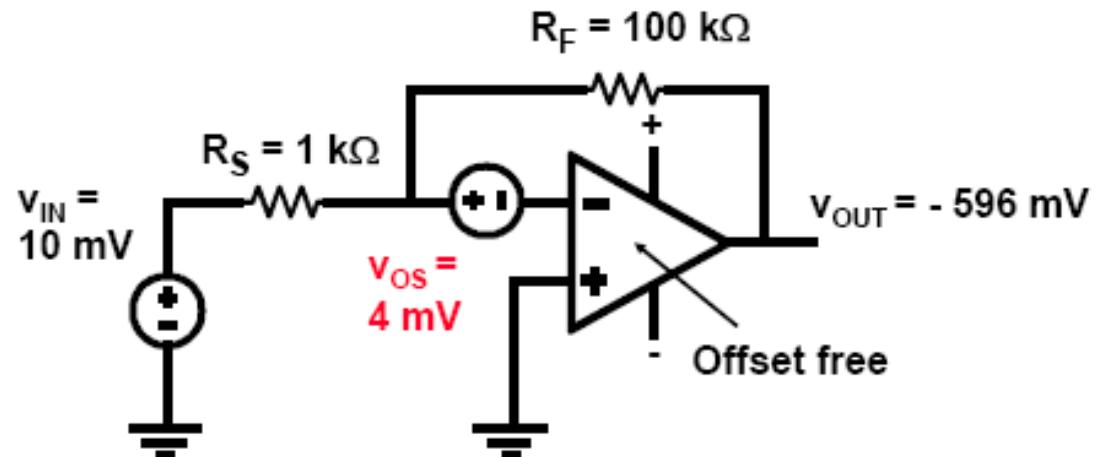


Transitional amplifiers' characteristics (ideal and real).

Real amplifier experiences the voltage offsets that can negatively influence its voltage gain or may be amplified further in the signal path.



Results of voltage offsets: a) zero offset, b) negative offset, c) positive offset.

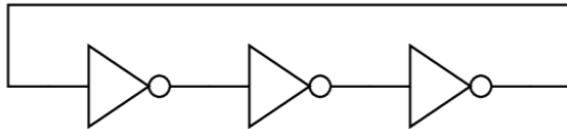


The gain is 59 instead of 100 !

Amplifiers' voltage offset influence on its gain.



AGH



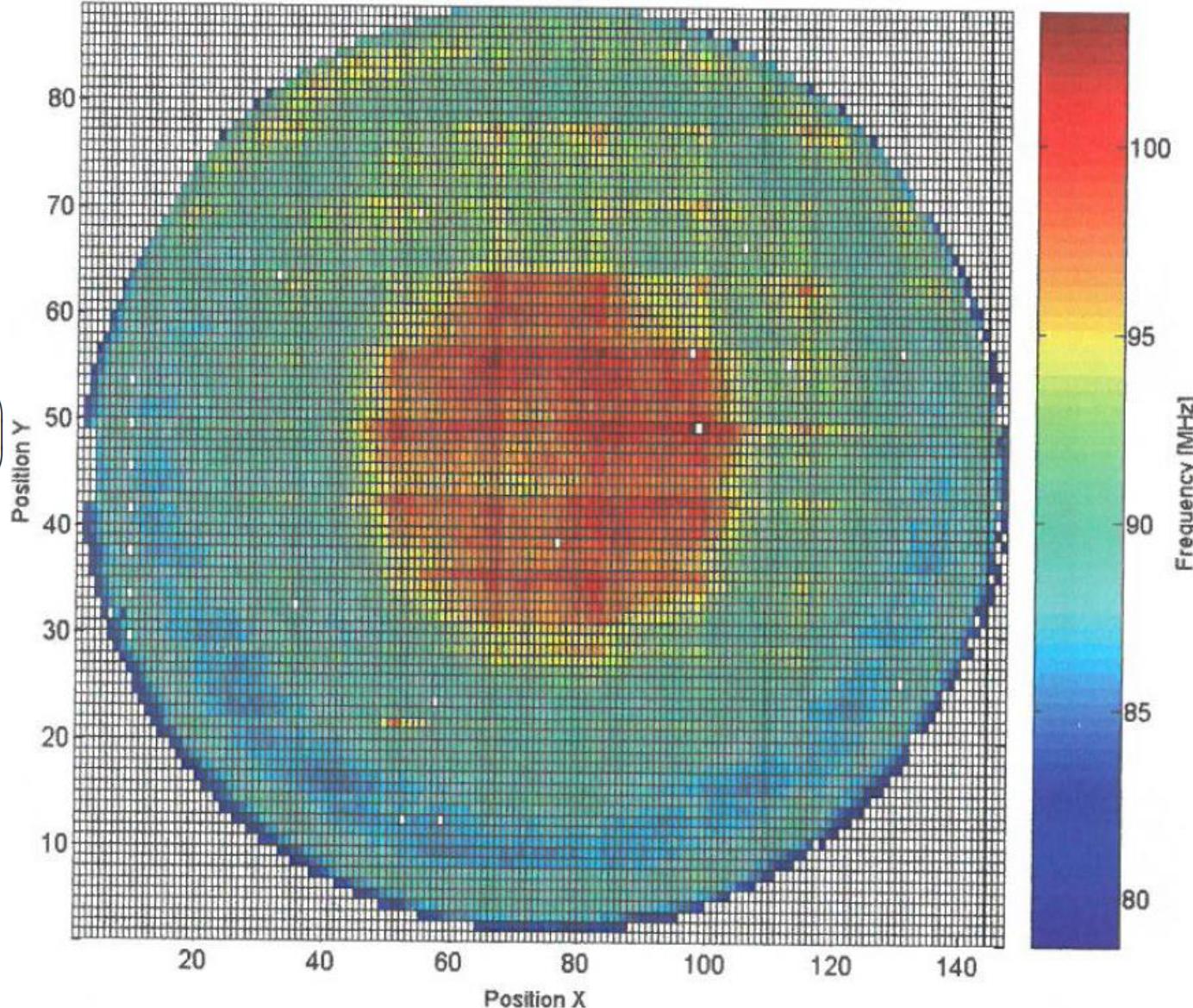
Ring oscillator.

$$T = \frac{2C_{ox}L^2(k+1)(V_{DD} - 2V_T)}{(V_{DD} - V_T)^2} \left( \frac{1}{k \cdot K_{PP}} + \frac{1}{K_{PN}} \right)$$

$T = t_r + t_f$  approximate delay of one inverter, loaded by the inverter with identical W and L sizes. The clock frequency of the oscillator using n-inverters chain is determined as:

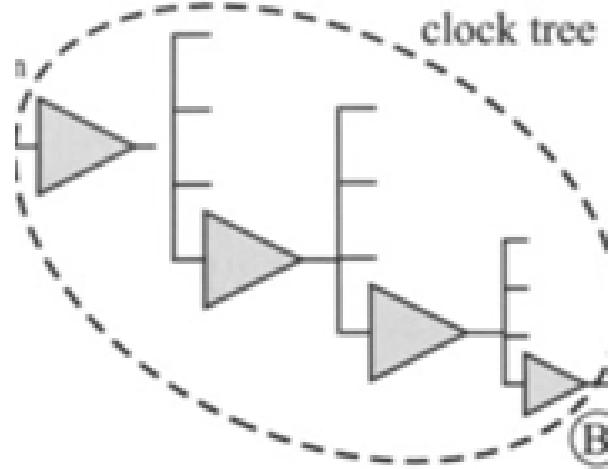
$$F_{CLK} = \frac{1}{n \cdot T}$$

## Results of inequality

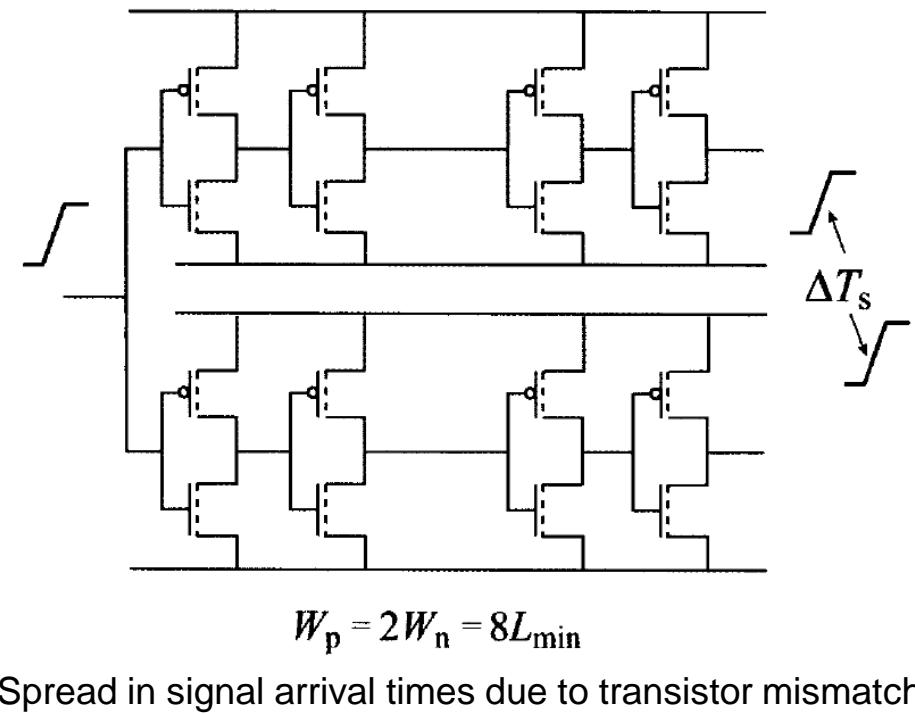


A wafer from a 90-nm CMOS production lot on which a free running oscillator frequency is measured. The frequency deviation (80 – 100 MHz) is largely indication for the variation in gate length, lithography and other processing steps that determine the electrical gate length.

# Results of inequality



For logic circuits, matching of transistors is becoming an important issue, resulting in different propagation delays of identical logic circuits. Due to the  $V_{TH}$  spread, inverter chains show different arrival times of the signals at their output nodes.



Spread in signal arrival times due to transistor mismatch.

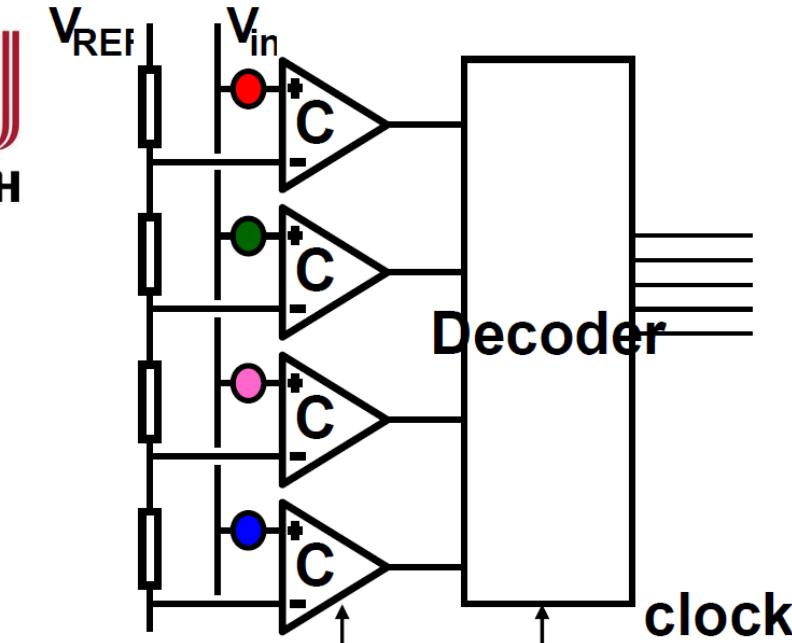
Technology node	250 nm	180 nm	130 nm	90 nm	60 nm	40 nm	32 nm
$\sigma_{\Delta TS}(C_{Load} = 50 \text{ fF})$	16 ps	21 ps	38 ps	68 ps	88 ps	100 ps	110 ps
Clock period T	10 ns	5 ns	2 ns	1 ns	0.5 ns	0.45 ns	0.4 ns

Spread in signal arrival times for different technology nodes.

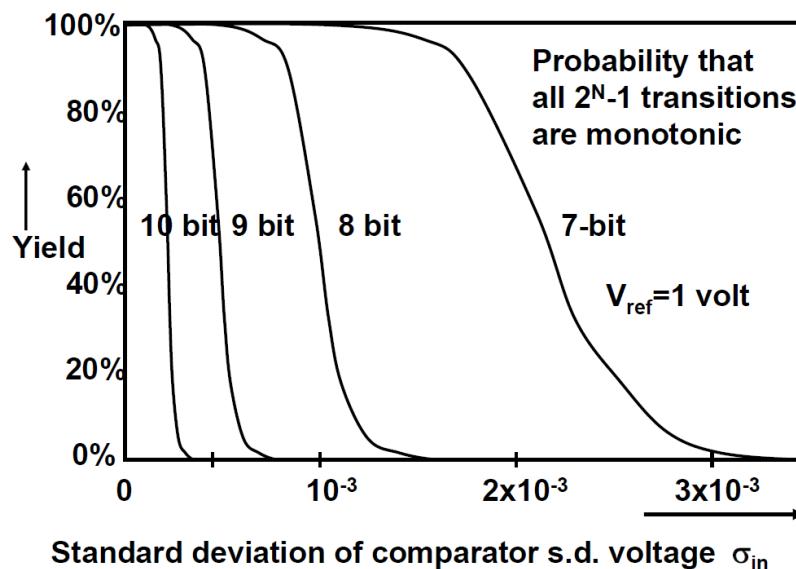
While the difference in arrival times at the second clock-tree stages may be 16 ps in a 0.25  $\mu\text{m}$  CMOS technology, it can be as high as 88 ps in a 65 nm CMOS technology, which is in the order of several gate delays.

Particularly for high-speed circuits, for which timing is a critical issue, transistor matching and its modelling is of extreme importance to maintain design robustness at a sufficiently high level.

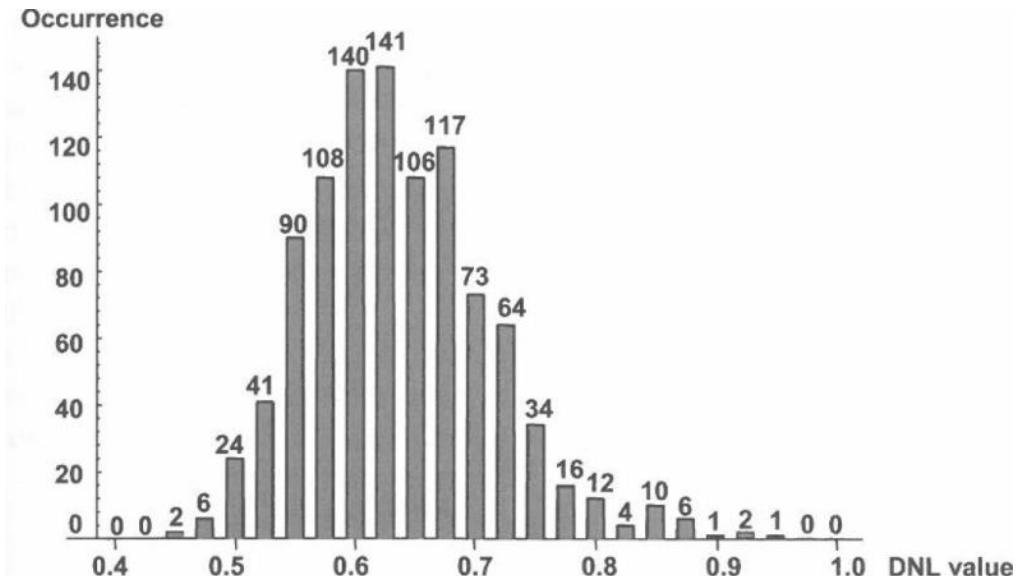
# Results of inequality



Flash converter with comparators voltage offsets.



Yield monotonicity vs the standard deviation of the comparator random offset.



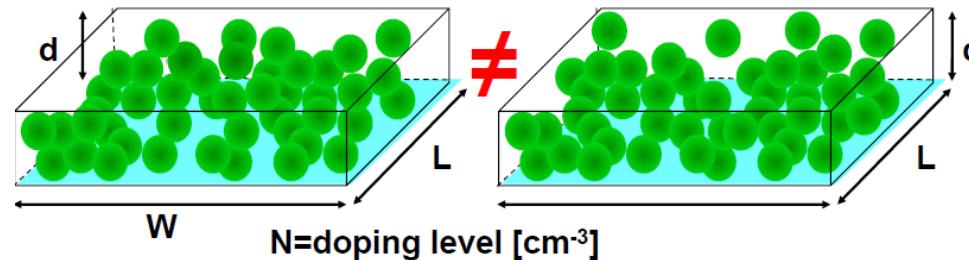
Typical histogram of DNL values for an 1000 8-bit full flash architecture, with  $\sigma=0.15 V_{LSB}$ . The mean DNL in this simulation is  $0.64 V_{LSB}$ .

The flash converter yield refers to the calculation of the probability of all comparators being within the monotonicity limit.

Ideally the  $(V_{j+1} - V_j) - V_{LSB} = 0$  but in real the non-monotonicity occurs, i.e. the comparator  $(j + 1)$  (fed with a rising input signal) switches before the adjacent comparator  $(j)$  that has a lower reference voltage.

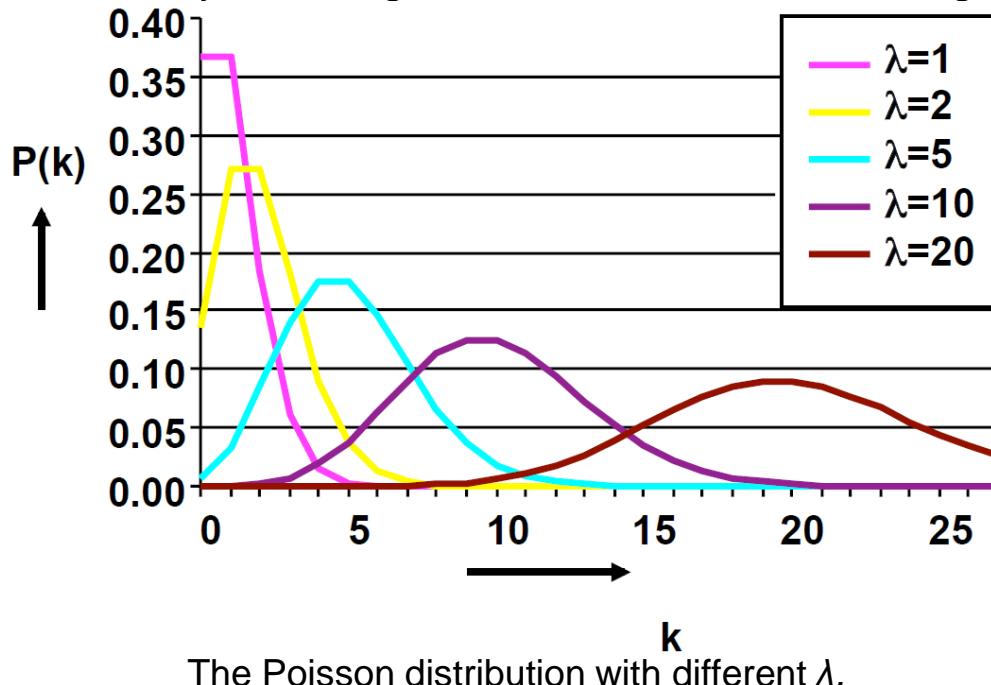
In the presented simulations every comparator offset spread is given by a Gaussian distribution with zero mean and standard deviation  $\sigma$ .

# Statistics for parameters variations



The number of dopants in a certain volume is on average controlled by the average implantation dose  $N_a$ . However, in one specific volume the actual number will fluctuate from one volume to another due to random processes.

Probability of having  $k$  ions in the box if the average is  $\lambda$  ions, is given by a **Poisson** distribution:



$$p(k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

$$\sigma_\lambda = \sqrt{\lambda}$$

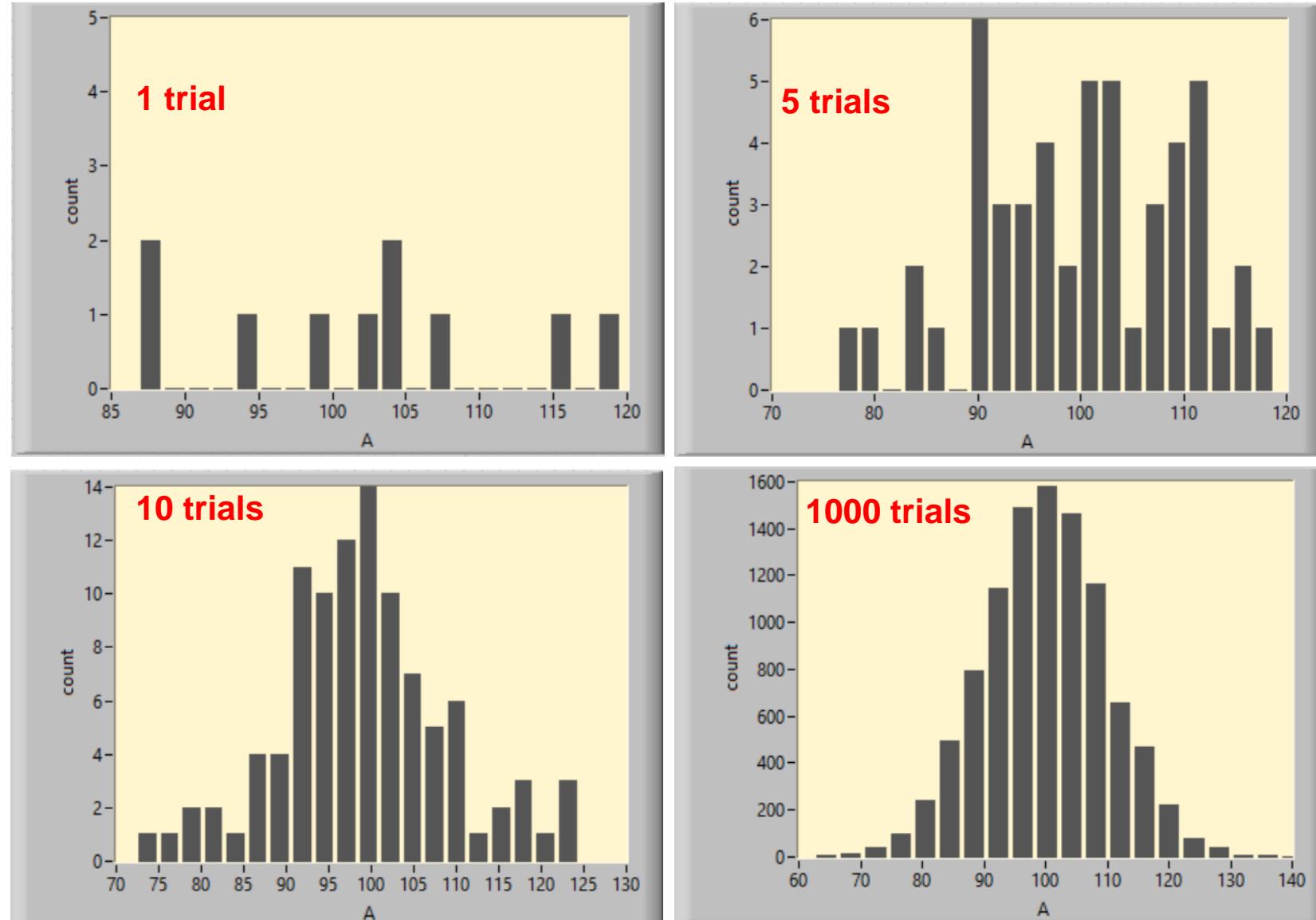
$\lambda$  is the average number of ions and  $\sigma$  is its spread value

Many known processes that cause the parameters spread like ion-implantation, diffusion follows the Poisson distribution.

# Central Limit Theorem

The **central limit theorem (CLT)** states that for the most commonly known processes, when independent random variables are added, their sum tends toward a normal distribution even if the original variables themselves are not normally distributed.

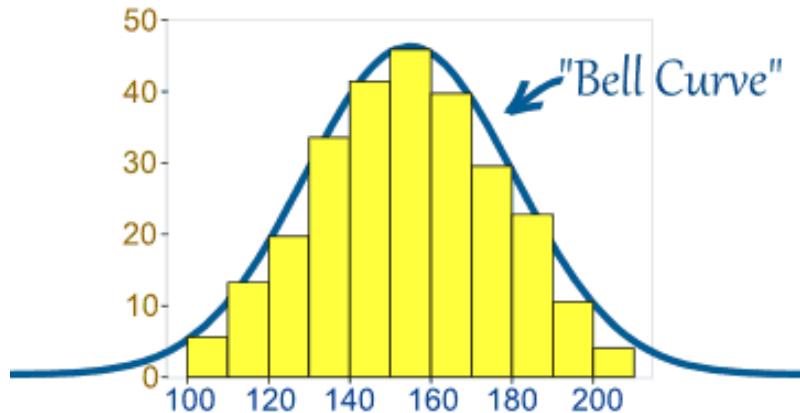
Now suppose that during the ion implantation the ions numbers are implanted with time step resulting in the histogram named „1 trial”. Whenever the process is repeated again and again the final histogram approaches the normal distribution.



Histograms of the hypothetical process with its different repetition number.

# Normal Distribution

Many things closely follow a Normal Distribution, i.e. heights of people, size of things produced by machines, errors in measurements, blood pressure, marks on a test, etc., etc., etc. ...



The yellow histogram shows some data that follows normal distribution closely, but not perfectly (which is usual) and its fit to normal distribution.

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The cumulative distribution function:

$$p(x_1 < x < x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = [1 + erf(\frac{x_2 - \mu}{\sigma\sqrt{2}})] - [1 + erf(\frac{x_1 - \mu}{\sigma\sqrt{2}})]$$

$\sigma$  - sigma

$\mu$  - mean value

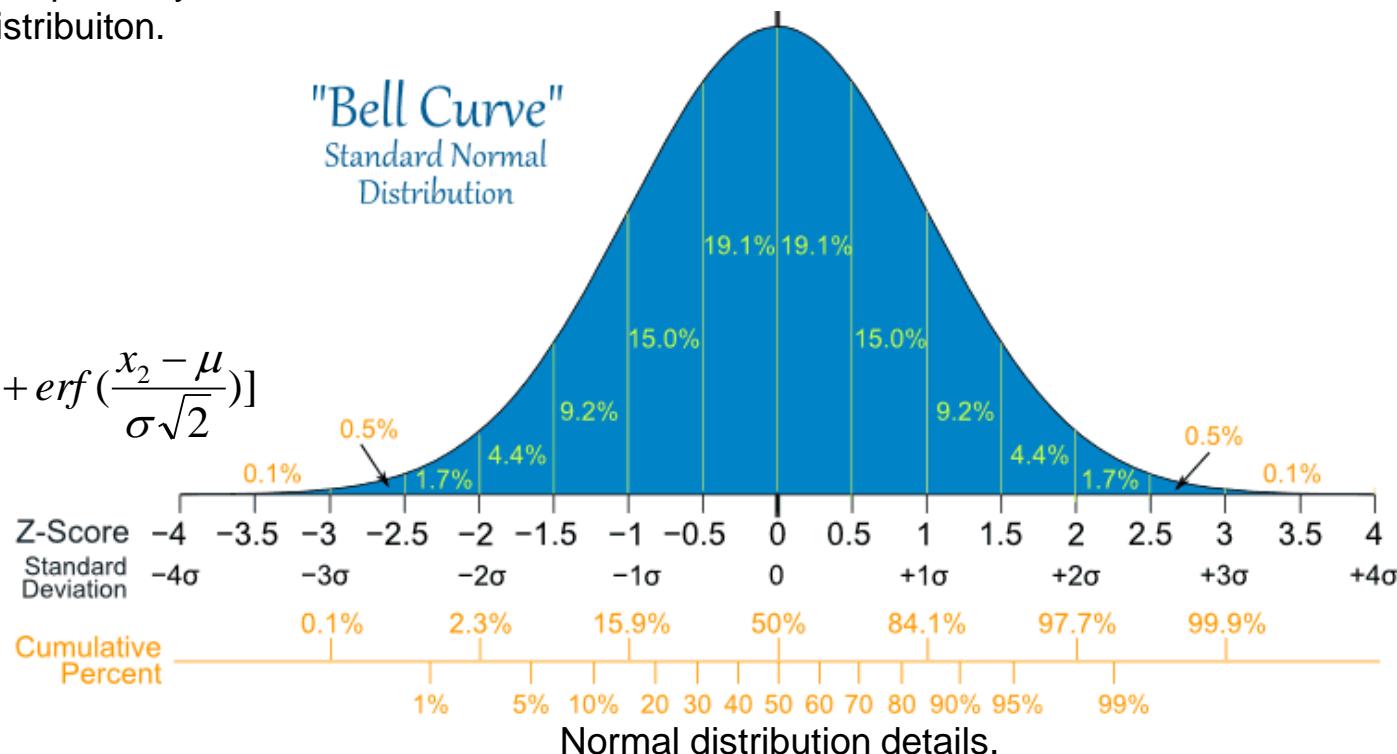
The Normal Distribution has:

1. mean = median = mode
2. symmetry about the center
3. 50% of values less than and 50% greater than the mean

**68% of values are within 1 standard deviation of the mean**

**95% of values are within 2 standard deviations of the mean**

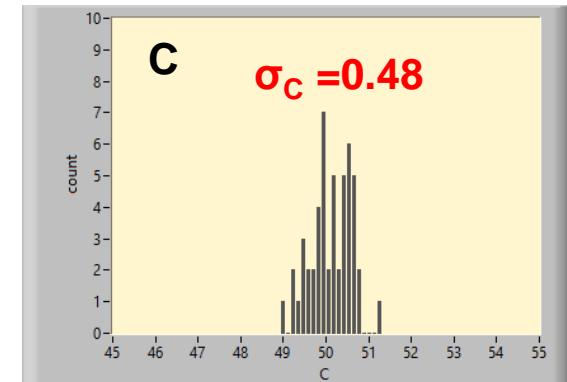
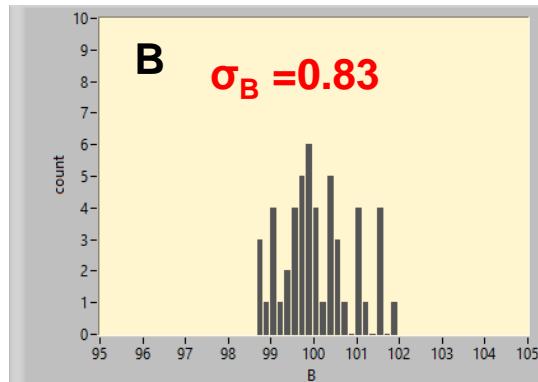
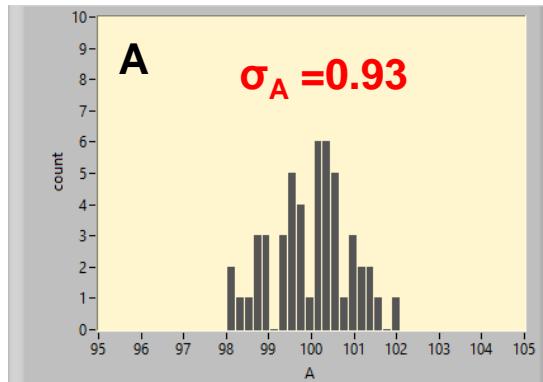
**99.7% of values are within 3 standard deviations of the mean**



Normal distribution details.

# The components spread influence on the overall spread

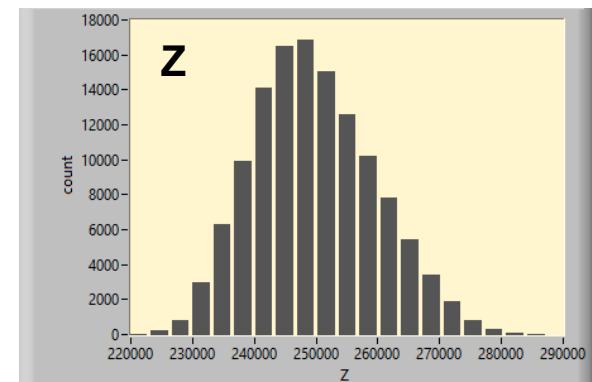
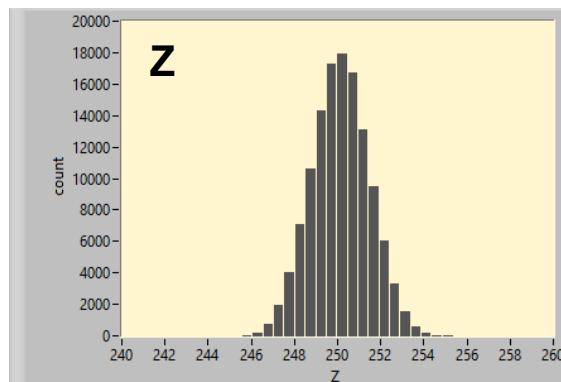
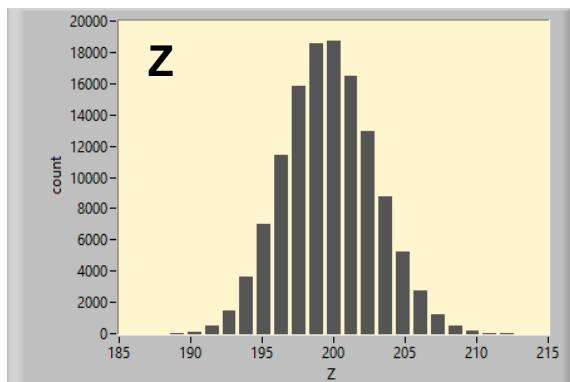
Simulation



$$Z = A + B + C$$

$$Z = (A \times B) / C$$

$$Z = A \times (B - C)^2$$

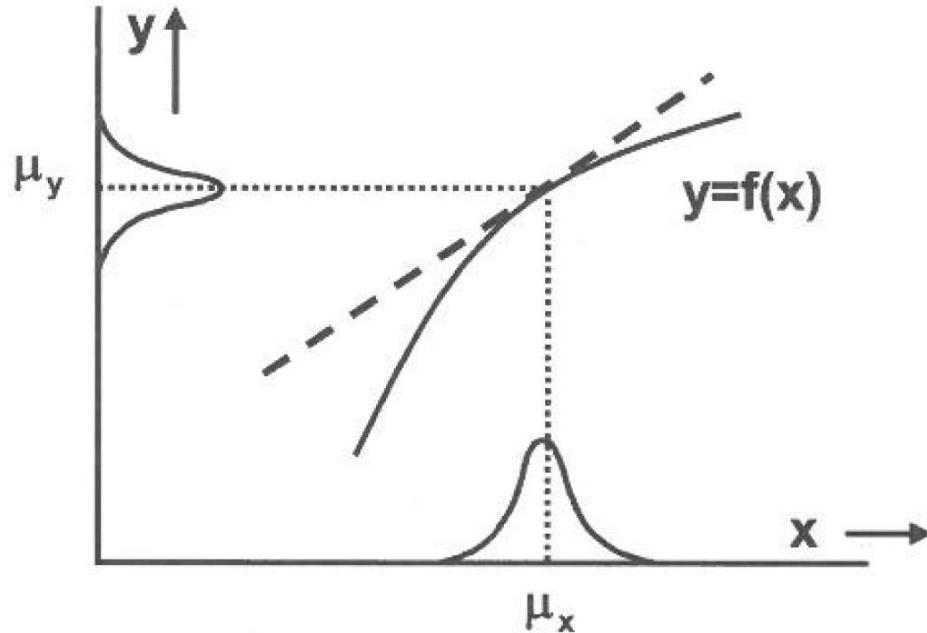


The particular components spread may influence the overall spread in different ways, i.e. it may be minor or it may be a considerable spread contributor. This, however depends on the function that components work with.

Function	Mean	Sigma	Sigma/Mean [%]
$Z=A+B+C$	250.1	1.32	<b>0.53</b>
$Z=(A \times B) / C$	199.7	3.13	<b>1.56</b>
$Z=A \times (B-C)^2$	249628	9837	<b>3.94</b>

# The components spread influence on the overall spread

In the circuit design the function  $y=f(x)$  is often a transfer function of a current or voltage into another current or voltage. If the function  $y=f(x)$  is smooth and differentiable, mean value of  $y$  and its sigma value can be approximated using the Taylor series expansion.



$$\mu_y \approx f(\mu_x) + \left( \frac{d^2 f(x)}{dx^2} \right) \frac{\sigma_x^2}{2}$$

$$Var(y) = \sigma_y^2 \approx \left( \frac{df(x)}{dx} \right)^2 \sigma_x^2$$

The transformation of one stochastic variable in another via a function  $y=f(x)$  uses the Taylor expansion.

If there are more than one variables of the function  $y$  the overall spread of value  $y$  is given as:

$$\sigma_y^2 \approx \left( \frac{df(x_1)}{dx_1} \right)^2 Var(x_1) + \left( \frac{df(x_2)}{dx_2} \right)^2 Var(x_2) + \left( \frac{df(x_1)}{dx_1} \frac{df(x_2)}{dx_2} \right) Cov(x_1, x_2)$$

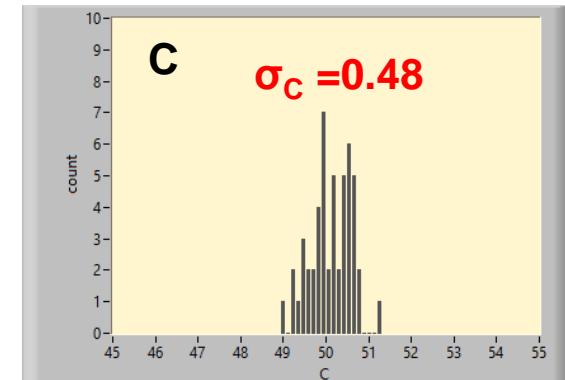
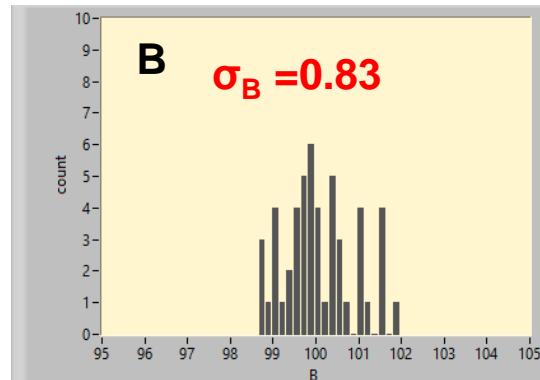
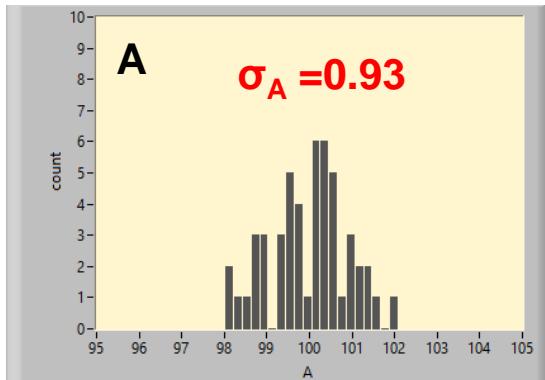
If the values of  $x_1$  and  $x_2$  are uncorrelated or independent the above result reduces to:

$$\sigma_y^2 \approx \left( \frac{df(x_1)}{dx_1} \right)^2 \sigma_{x1}^2 + \left( \frac{df(x_2)}{dx_2} \right)^2 \sigma_{x2}^2$$

**Note that this summation favors the largest value**

# The components spread influence on the overall spread

## Calculations



$$Y = A + B + C$$

$$\sigma_Y^2 = \sigma_A^2 + \sigma_B^2 + \sigma_C^2$$

$$\frac{\sigma_Y^2}{Y^2} = \frac{\sigma_A^2}{(A+B+C)^2} + \frac{\sigma_B^2}{(A+B+C)^2} + \frac{\sigma_C^2}{(A+B+C)^2}$$

$$Y = (A \times B) / C$$

$$\sigma_Y^2 = \left(\frac{B}{C}\right)^2 \sigma_A^2 + \left(\frac{A}{C}\right)^2 \sigma_B^2 + \left(\frac{AB}{C^2}\right)^2 \sigma_C^2$$

$$\frac{\sigma_Y^2}{Y^2} = \frac{\sigma_A^2}{A^2} + \frac{\sigma_B^2}{B^2} + \frac{\sigma_C^2}{C^2}$$

$$Y = A \times (B - C)^2$$

$$\sigma_Y^2 = (B - C)^4 \sigma_A^2 + (2A(B - C))^2 \sigma_B^2 + (2A(B - C))^2 \sigma_C^2$$

$$\frac{\sigma_Y^2}{Y^2} = \frac{\sigma_A^2}{A^2} + \frac{4\sigma_B^2}{(B - C)^2} + \frac{4\sigma_C^2}{(B - C)^2}$$

One is often interested in the relative spread value instead of its direct value. However, keep in mind that in calculations it is sometimes worth of using the relative spread value instead of direct value.

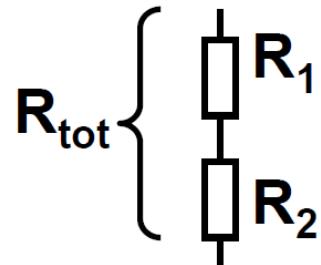
# The components spread influence on the overall spread

## EXAMPLE

Consider two types of resistors' connection, i.e. series and parallel, and calculate what is a standard deviation of particular configuration? Assume the  $R_1 = R_2 = 1\text{k}\Omega$  have standard deviations  $\sigma_1 = \sigma_2 = 5\%$  respectively and these are independent.

$$\sigma_{R1,R2} = 5\% \Rightarrow \sigma_{R1,R2} = 50\Omega$$

$$R = R_1 + R_2$$

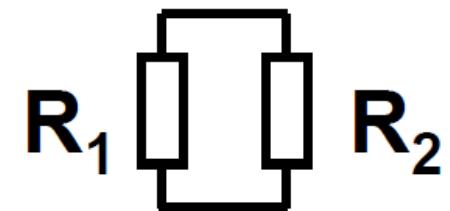


$$\sigma_R^2 = \left(\frac{dR}{dR_1}\right)^2 \sigma_{R1}^2 + \left(\frac{dR}{dR_2}\right)^2 \sigma_{R2}^2 = \sigma_{R1}^2 + \sigma_{R2}^2$$

$$\sigma_R = \sqrt{2 \times 50^2} = 70.7\Omega$$

$$\frac{\sigma_R}{R} = 3.5\%$$

$$R = \frac{R_1 \times R_2}{R_1 + R_2}$$



$$\sigma_R^2 = \left(\frac{R_2}{R_1 + R_2}\right)^4 \sigma_{R1}^2 + \left(\frac{R_1}{R_1 + R_2}\right)^4 \sigma_{R2}^2$$

$$\sigma_R = \sqrt{2 \times (0.5)^4 \times 50^2} = 17.67\Omega$$

$$\frac{\sigma_R}{R} = 3.5\%$$

If one would like to employ relative calculation instead of direct, for parallel resistors connection we would get:

$$\frac{\sigma_R^2}{R^2} = \frac{{R_2}^2 \sigma_{R1}^2}{{R_1}^2 (R_1 + R_2)^2} + \frac{{R_1}^2 \sigma_{R2}^2}{{R_2}^2 (R_1 + R_2)^2}$$

$$\frac{\sigma_R}{R} = \sqrt{2} \frac{50 \times 1k}{1k(1k + 1k)} = 0.035 = 3.5\%$$

## Mismatches

The standard way to define the value of the parameters spread is to show the sigma of the parameters distribution **between the two instances**.

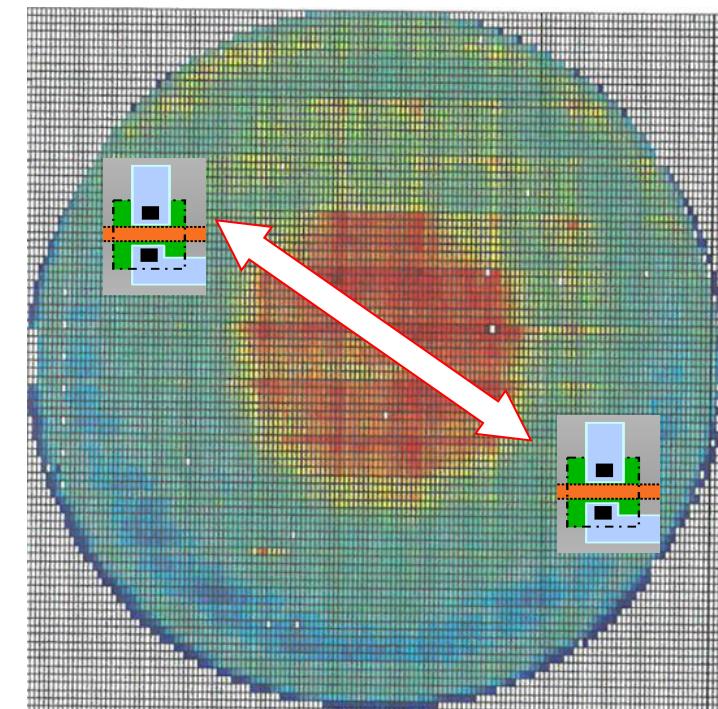
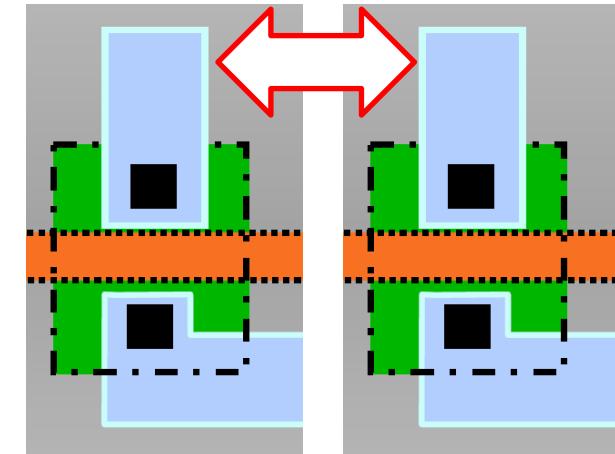
$$\sigma^2(\Delta P) = \frac{A_P^2}{\text{AREA}}$$

where  $A_P$  is area proportionality constant for parameter P.

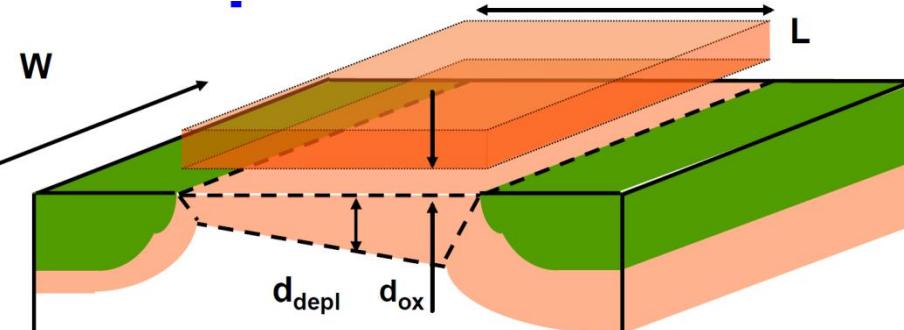
Another approach includes the distance effects in the stochastic model. As the original placement of dies on a wafer is unknown after packaging, the effect of this parameter value distribution is modeled as an additional stochastic process with a long corellation distance.

$$\sigma^2(\Delta P) = \frac{A_P^2}{\text{AREA}} + S_P^2 D_x^2$$

where  $S_P$  describes the variation of parameter P with the spacing.



# Threshold voltage spread of the MOS transistor



A cross section of the transistor indicating the depletion region.

$$V_{TH} = V_{T0} + \gamma \left( \sqrt{2|\phi_F| + V_{SB}} - \sqrt{2|\phi_F|} \right)$$

The threshold voltage and the body effect are determined by oxide thickness and dopant concentration of the depleted (substrate) channel region.

If the  $V_{SB}=0$  and keeping in mind that the threshold voltage is proportional to the charge in the depletion region divided by the gate oxide capacitance per unit area :

$$V_{TH} = V_{T0} \propto \frac{Q_B}{C_{OX}} \quad \Delta V_{TH} \propto \frac{\Delta Q_B}{C_{OX}}$$

The total number of active ions in the transistor channel and the depletion layer depth are given by:

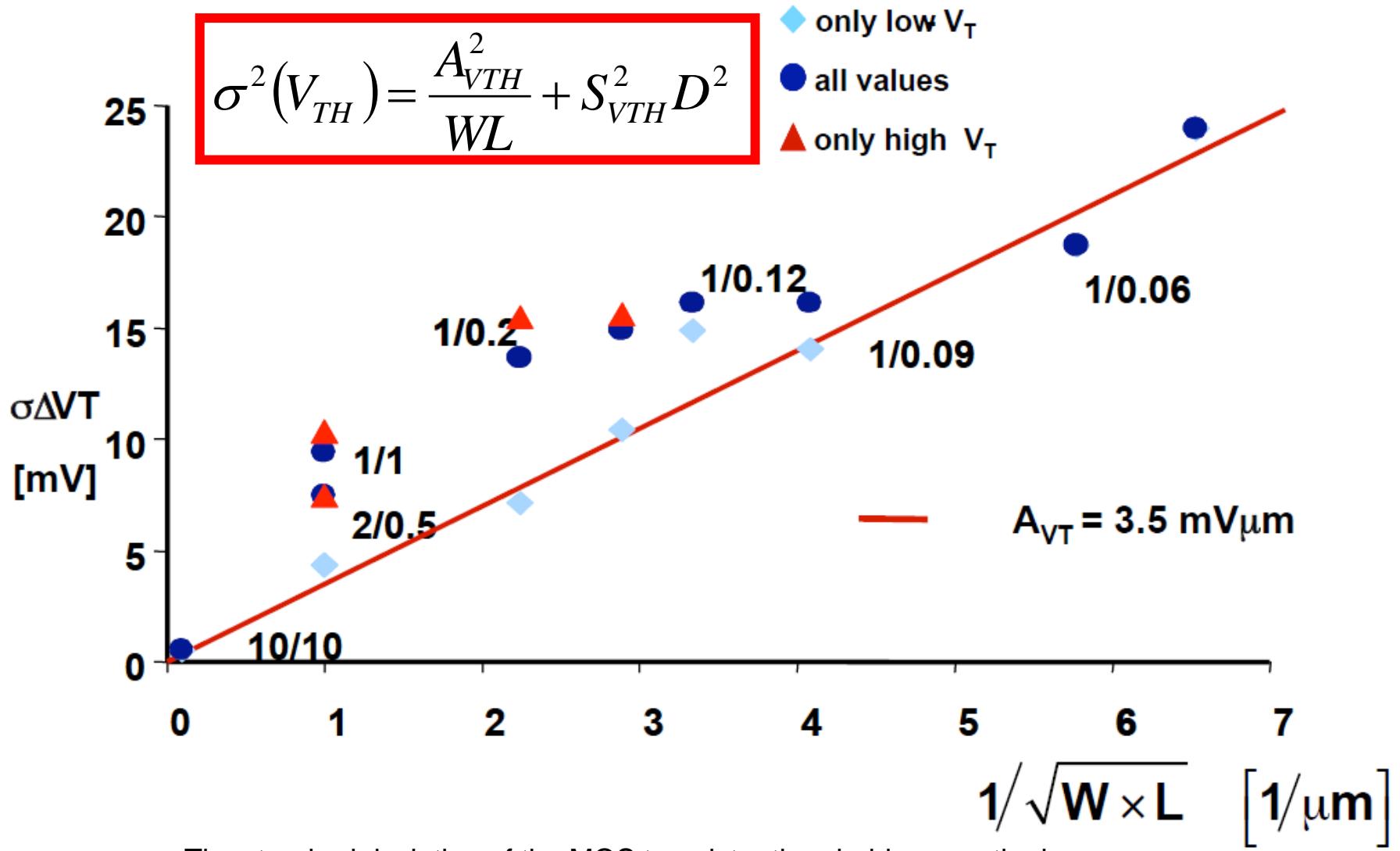
$$N_t = N_B \times W \times L \times t_d \quad t_d = \sqrt{\frac{2\epsilon_{Si}}{q \times N_B} \Psi_s}$$

Having in mind that this process follows the Poisson distribution (the doping spread of  $N_t$  equals  $\sqrt{N_t}$ ) the  $C_{ox} = \epsilon_{Si}/t_{ox}$ :

$$\sigma_{VTH} = \sqrt{2} \frac{1}{\sqrt{WL}} \frac{t_{ox}}{C_{OX}} (2q^3 \epsilon_{Si} N_B \Psi_s)^{\frac{1}{4}} = \frac{A_{VTH}}{\sqrt{WL}}$$

$$\sigma^2(V_{TH}) = \frac{A_{VTH}^2}{WL} + S_{VTH}^2 D^2$$

# Threshold voltage spread of the MOS transistor



The standard deviation of the MOS transistor threshold versus the inverse square root of the channel area.

The slope of the line equals the parameter  $A_{V_T}$ . For the smaller sizes the effective gate area is smaller due to under-diffusion and channel encroachment and for very small MOS devices one reaches the lithographical limits. Also short channel transistors with halo-implants (higher dope) give enhanced mismatch.

# Threshold voltage spread for various processes

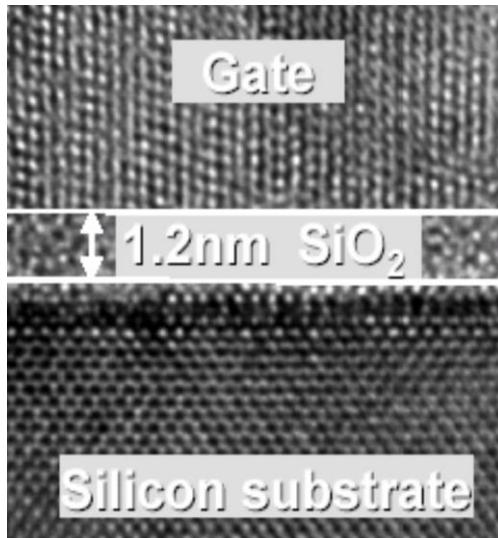
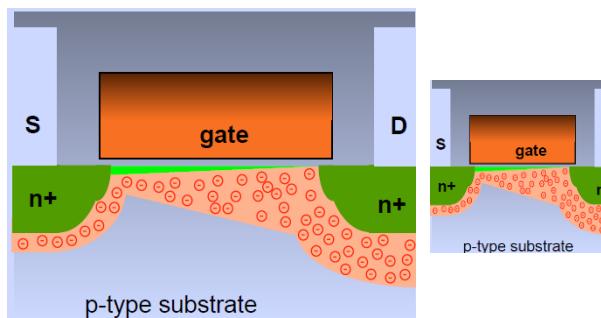


Photo of the single transistor processed in the nanometer CMOS process.



Conceptual view of the minimum MOS channel dimensions for different processes.

Granularity on molecular level is reached what means that there is no continuous change in doping, but in discrete portions.

0.25 μm / 0.25 μm transistor = 1000 doping atoms

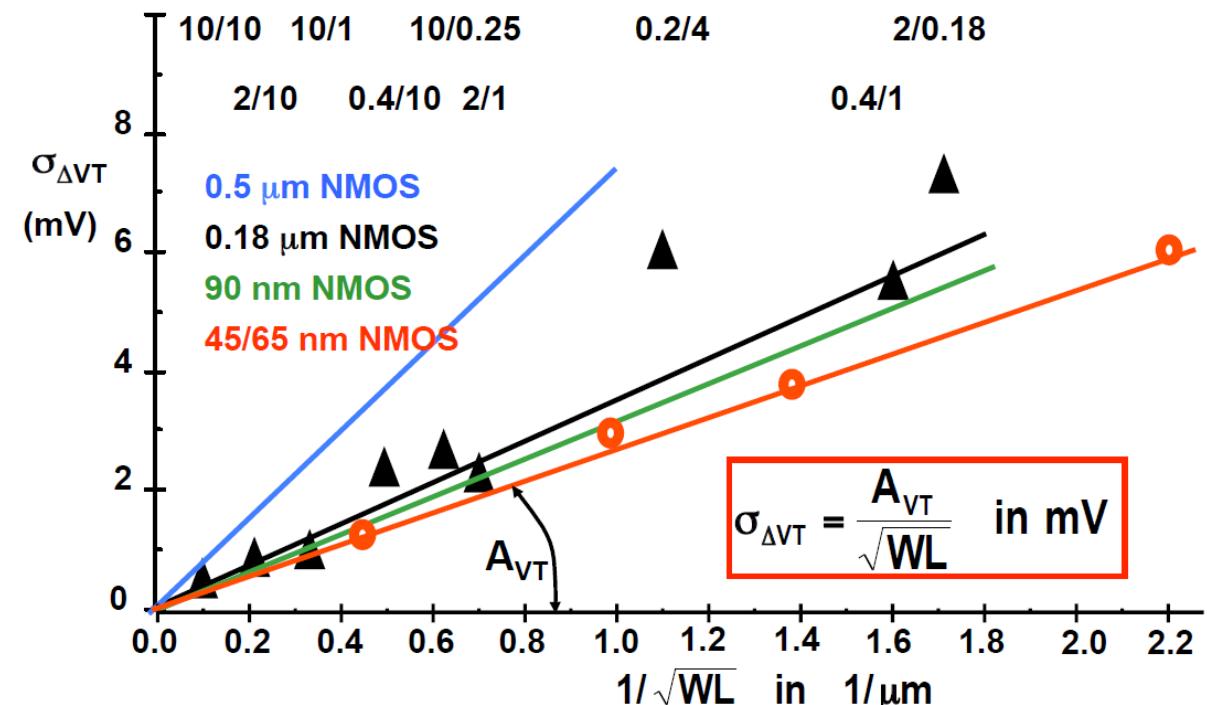
0.1 μm / 0.065 μm transistor = 60-80 atoms

Transistor scaling practice forcing the doping level under the gate to increase, hence the depletion depth  $d_{depl}$  reduces with  $N_a^{-0.5}$ . To first order, the overall dependence of  $A_{VT}$  with the doping is proportional to  $N^{0.25}$ , which means that  $A_{VT}$  will decrease less than the predicted oxide scaling.

$$\sigma_{\Delta VT} \propto \frac{d_{ox} \sqrt[4]{N_a}}{\sqrt{W \times L}}$$

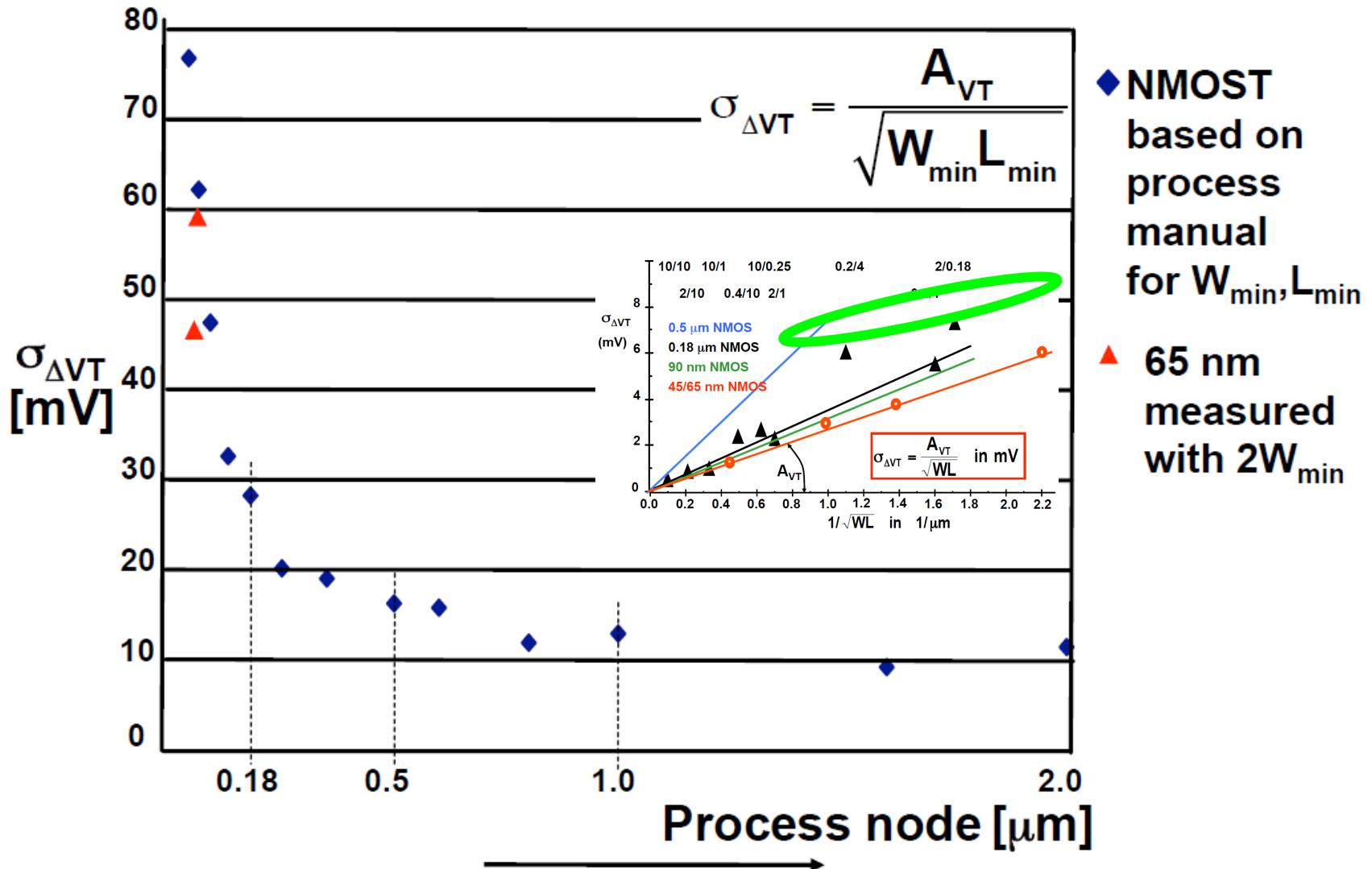
$d_{ox}$

$\sqrt[4]{N_a}$



The standard deviation of the MOS transistor threshold versus the inverse square root of the channel area for different processes.

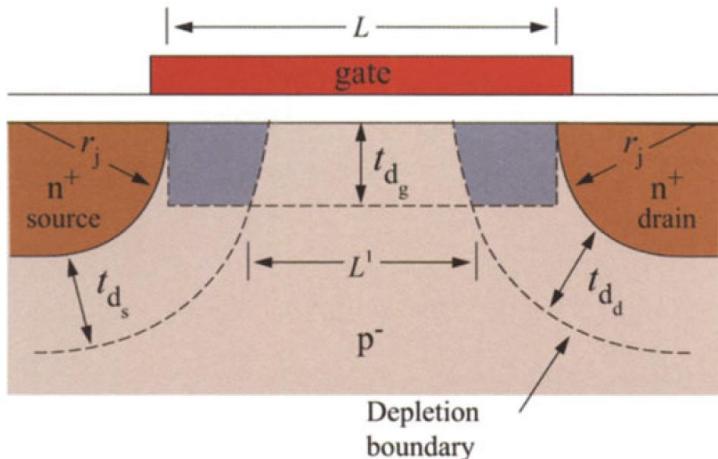
# Threshold voltage spread of the MOS transistor



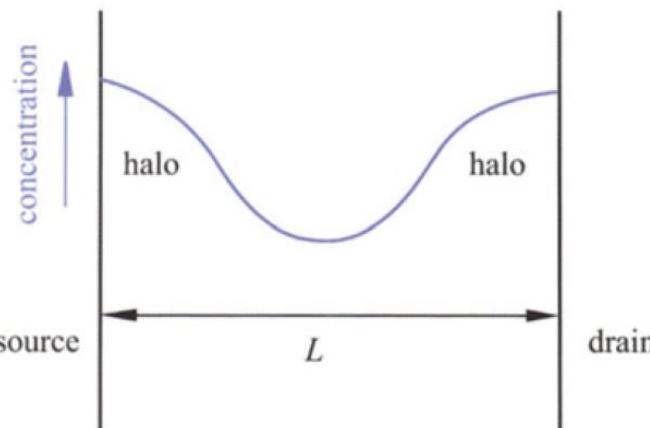
The standard deviation of the MOS transistor threshold versus the process node for minimum size channel dimensions.

It is clearly visible that the minimum size channel dimensions should be avoided whenever the low threshold spread is an aim. Otherwise the threshold spread for modern processes rises drastically.

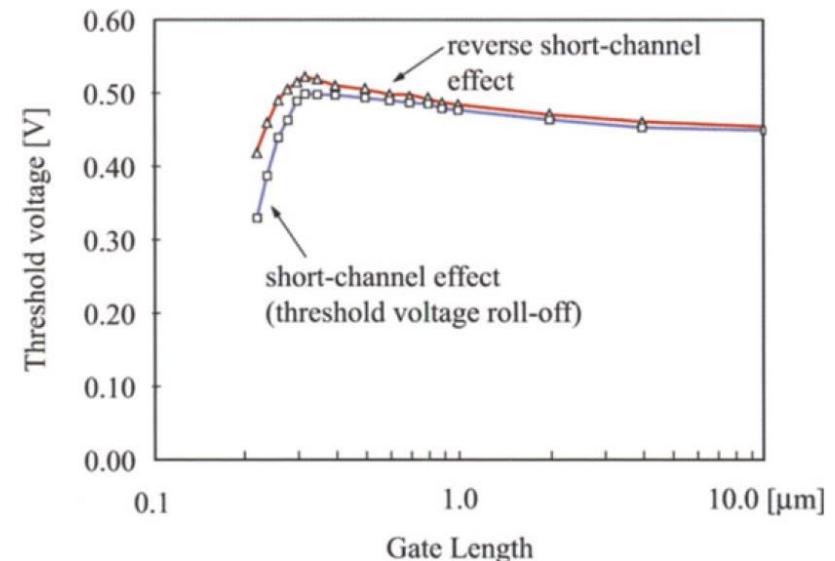
# Threshold voltage spread for various processes



Cross-section of a short-channel transistor, showing several depletion areas that affect each other



Potential doping profile in the channel of a MOS device including the halos



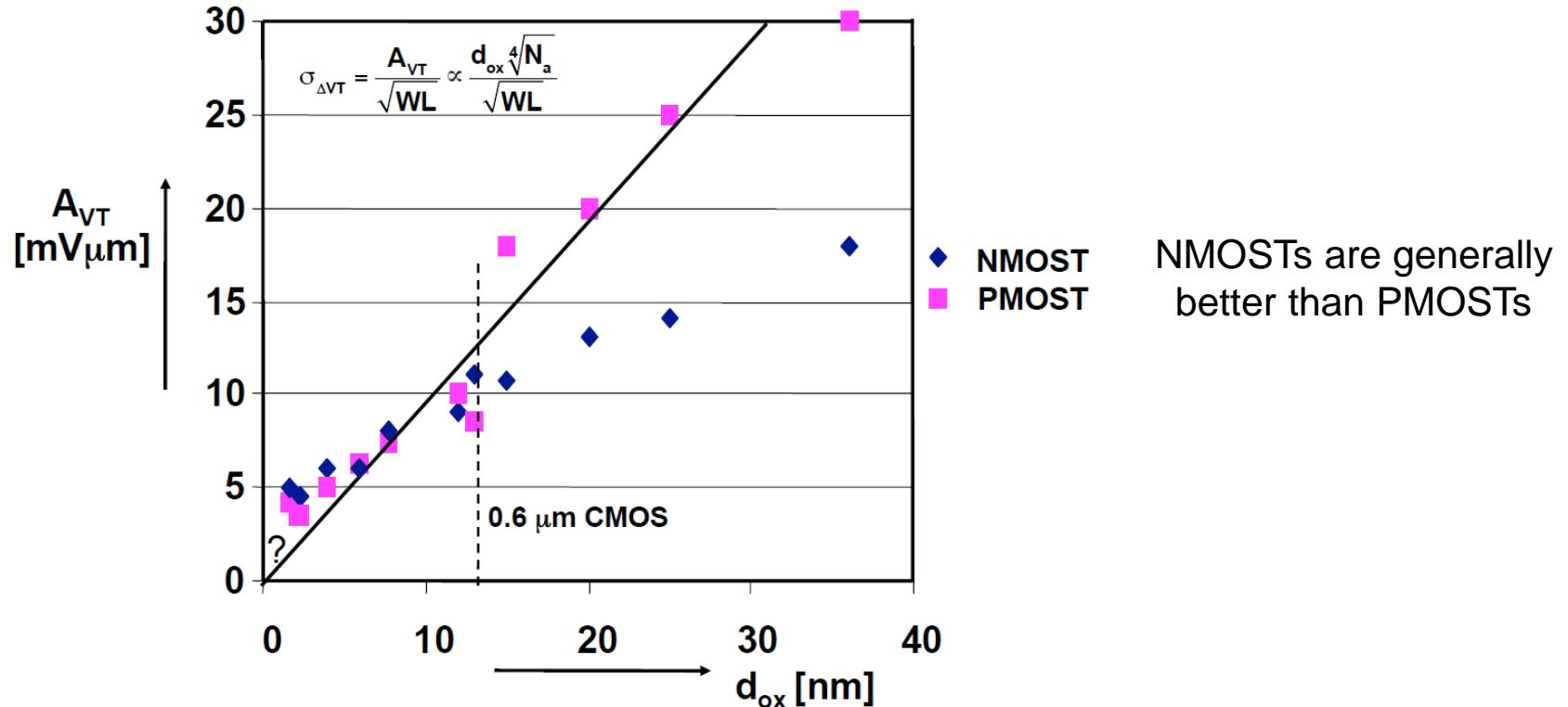
Short-channel and reverse short-channel effect on the threshold voltage  $V_{TH}$  of an nMOS transistor

In the deep-submicron processes the short channel effects in channel are controlled by means of „halo” or „pocket” implants. Next to their own variations, the self-aligned prints any line-edge roughness in the doping profile. The pocket implants defy the uniform dopant hypothesis for the calculation of the threshold mismatch. Therefore the additional term can be included for the threshold variation due to the pocket implant:

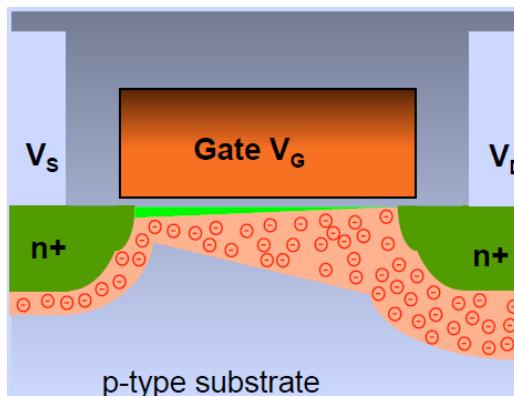
$$\sigma^2(V_{TH}) = \frac{A_{VTH}^2}{WL} + \frac{B_{VTH}^2}{f(WL)}$$

where the function  $f(WL)$  still need to be established

# Threshold voltage spread for various processes

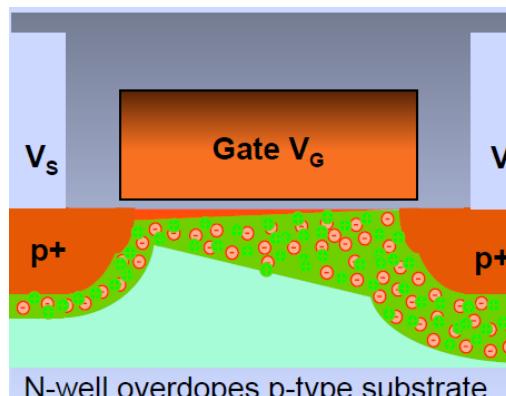


The standard deviation of the MOS transistor threshold versus the process node for NMOS and PMOS transistors.



$$V_T \propto WL_y \times N_a$$

$$\sigma_{\Delta VT} \propto \sqrt{WL_y \times N_a}$$

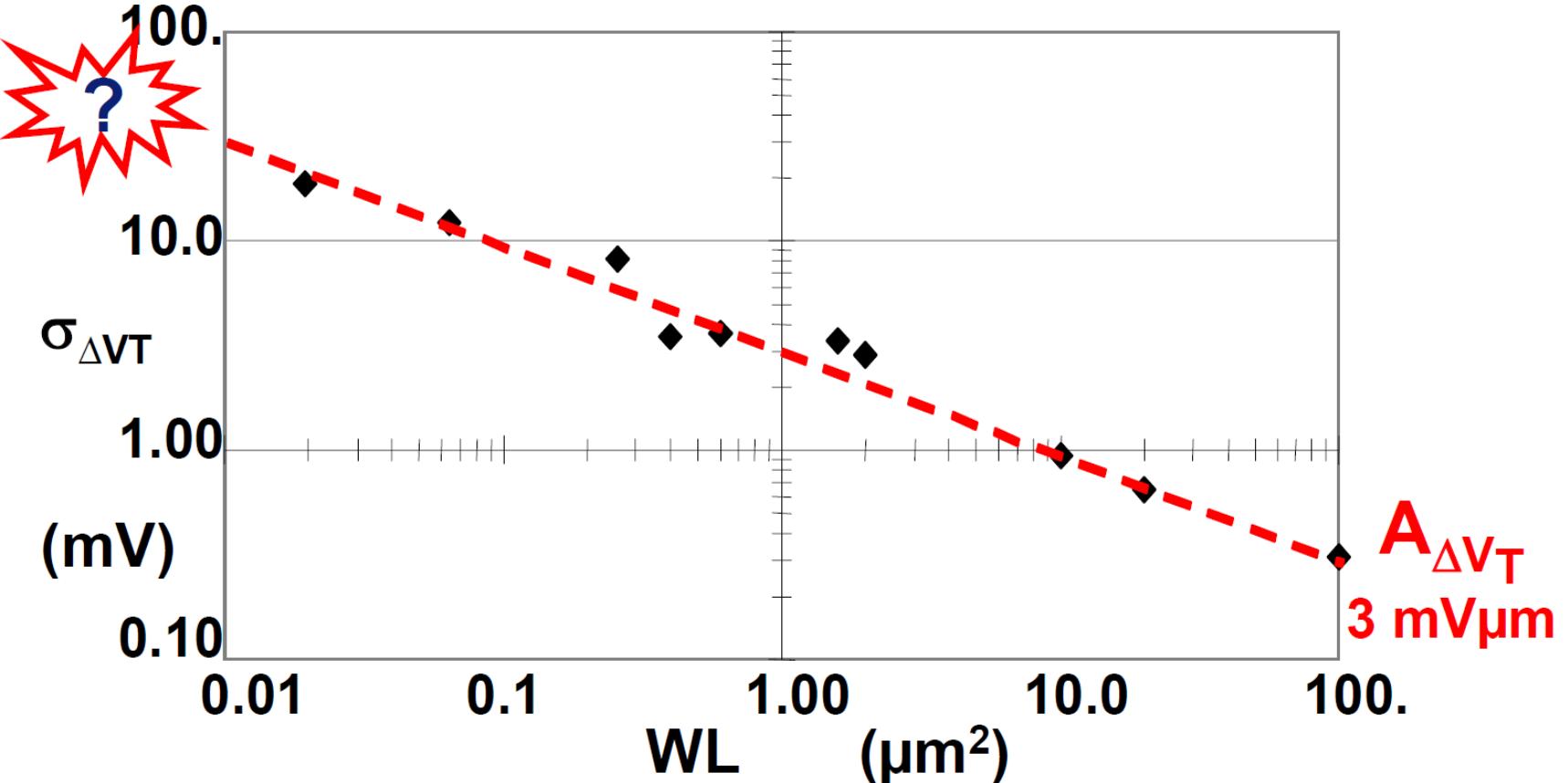


$$V_T \propto WL_y \times (N_a - N_d)$$

$$\sigma_{\Delta VT} \propto \sqrt{WL_y \times (N_a + N_d)}$$

The statistical variation in  $(N_a + N_d)$  determines matching while control of the  $(N_a - N_d)$  determines the threshold voltage. Beyond the  $0.6 \mu m$  node a twin well construction with a dedicated well implant for the PMOS transistor is used that avoids compensating charges.

# Threshold voltage spread for various processes



Observations:

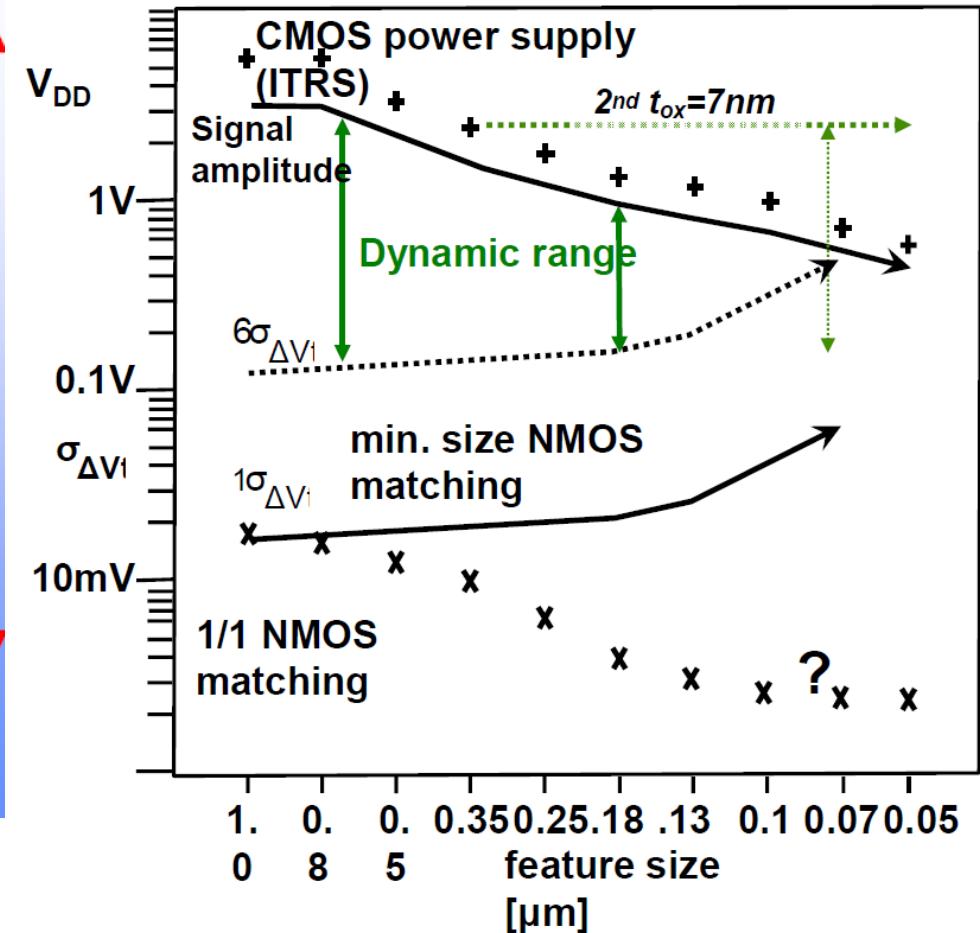
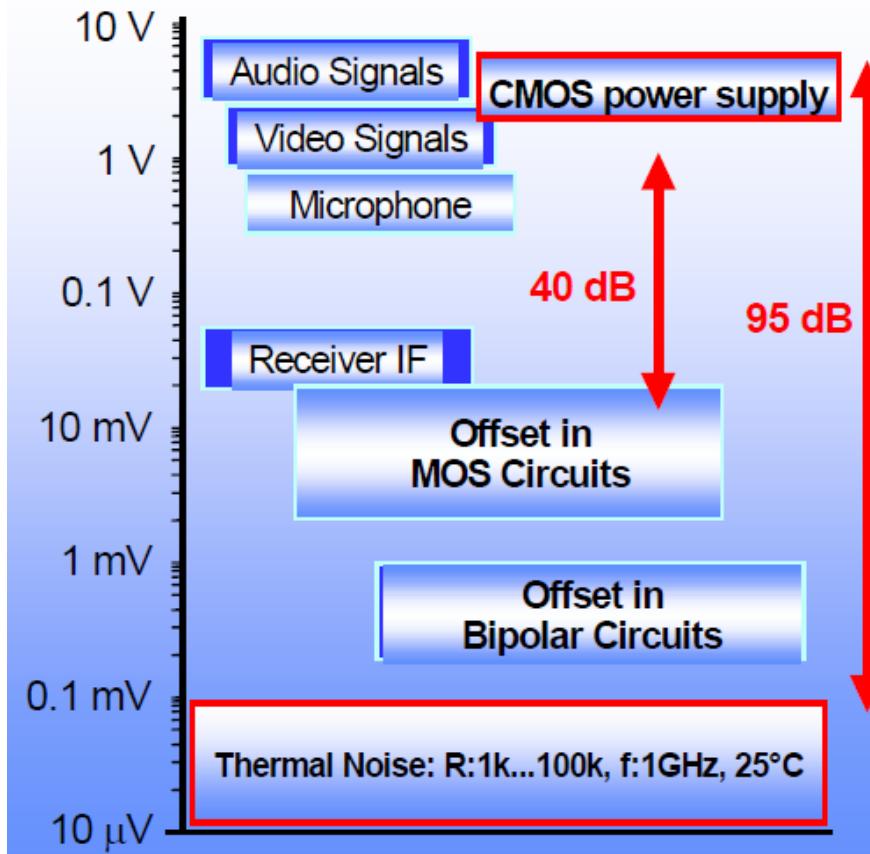
- $\sigma_{V_{TH}}$  ranges from 300  $\mu\text{V}$  for  $W/L = 10/10$  to 20 mV for  $W/L = 0.15/0.13$ ,
- 1/square-root area law seems applicable over 4 decades geometry range.

The next generation technology will allow for MOS channel effective area  $\approx 0.0024 \mu\text{m}^2$ .

The key figure of merit for MOSFET matching characterization:

$A_{V_T} \approx 1 \text{ mV}\mu\text{m}$  per nm gate oxide thickness (1 mV $\mu\text{m}$  x nm)

# Threshold voltage spread for various processes



Development of power supply voltage and the measured NMOS threshold matching factor  $A_{VT}$  through various process generations. The available signal swing (upper solid line) is derived by taking 90% from the nominal power supply and subtracting a threshold voltage and  $4kT/q$  gate overdrive.

The  $6\sigma_{VT}$  line indicates that for  $0.1 \mu\text{m}$  processes threshold voltage matching will also eliminate noise margins in basic digital circuits for multimillion transistor ICs.  
 Matching dominates the circuit performance potential in submicron.

# Current factor mismatch of the MOS transistor

$$I_D = \frac{1}{2} \mu C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2$$

$$\beta = \mu C_{ox} \frac{W}{L}$$

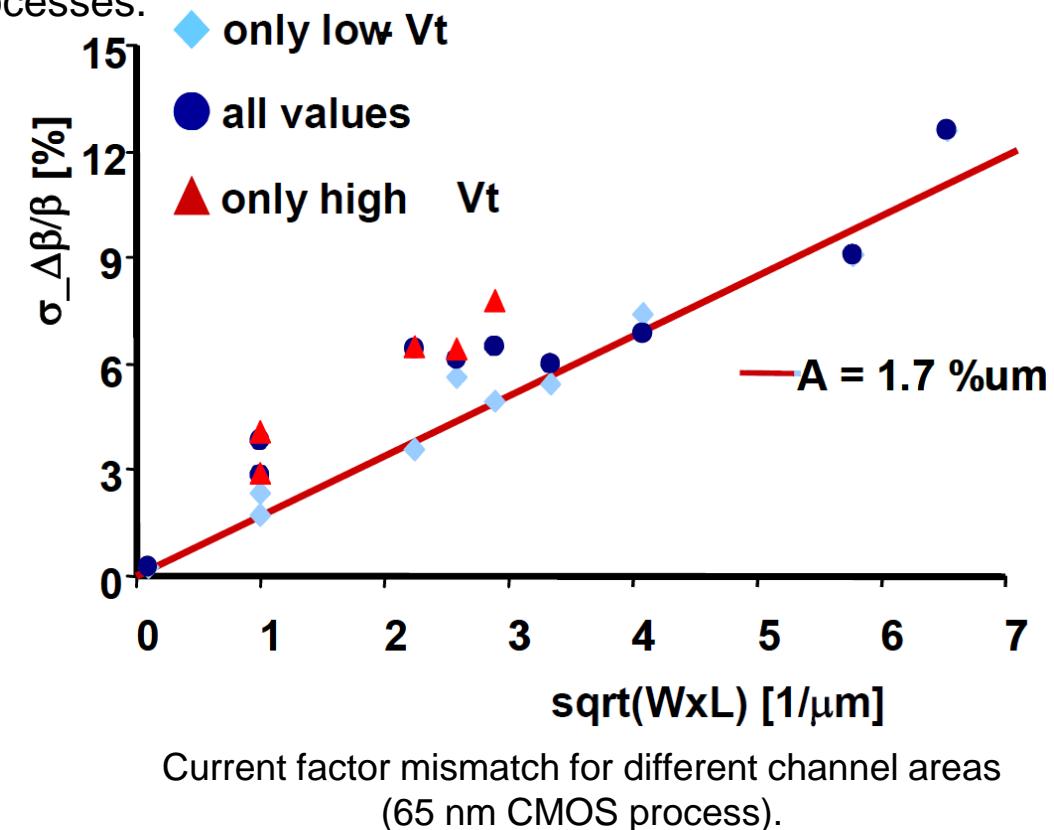
$$\frac{\sigma^2_{\Delta\beta}}{\beta^2} = \frac{A_\mu^2}{WL} + \frac{A_{Cox}^2}{WL} + \frac{A_W^2}{W^2L} + \frac{A_L^2}{WL^2}$$

$A_W$  and  $A_L$  can be neglected for most matured processes.

$$\frac{\sigma^2_{\Delta\beta}}{\beta^2} \approx \frac{A_\beta^2}{WL}$$

Still one should be aware while using very small MOS channel area that the edge roughness can not be neglected for modern processes in that conditions (due to near the resolution limits of the photo lithography and in the presence of halo-implants).

$$[A_\beta] = \% \mu m$$



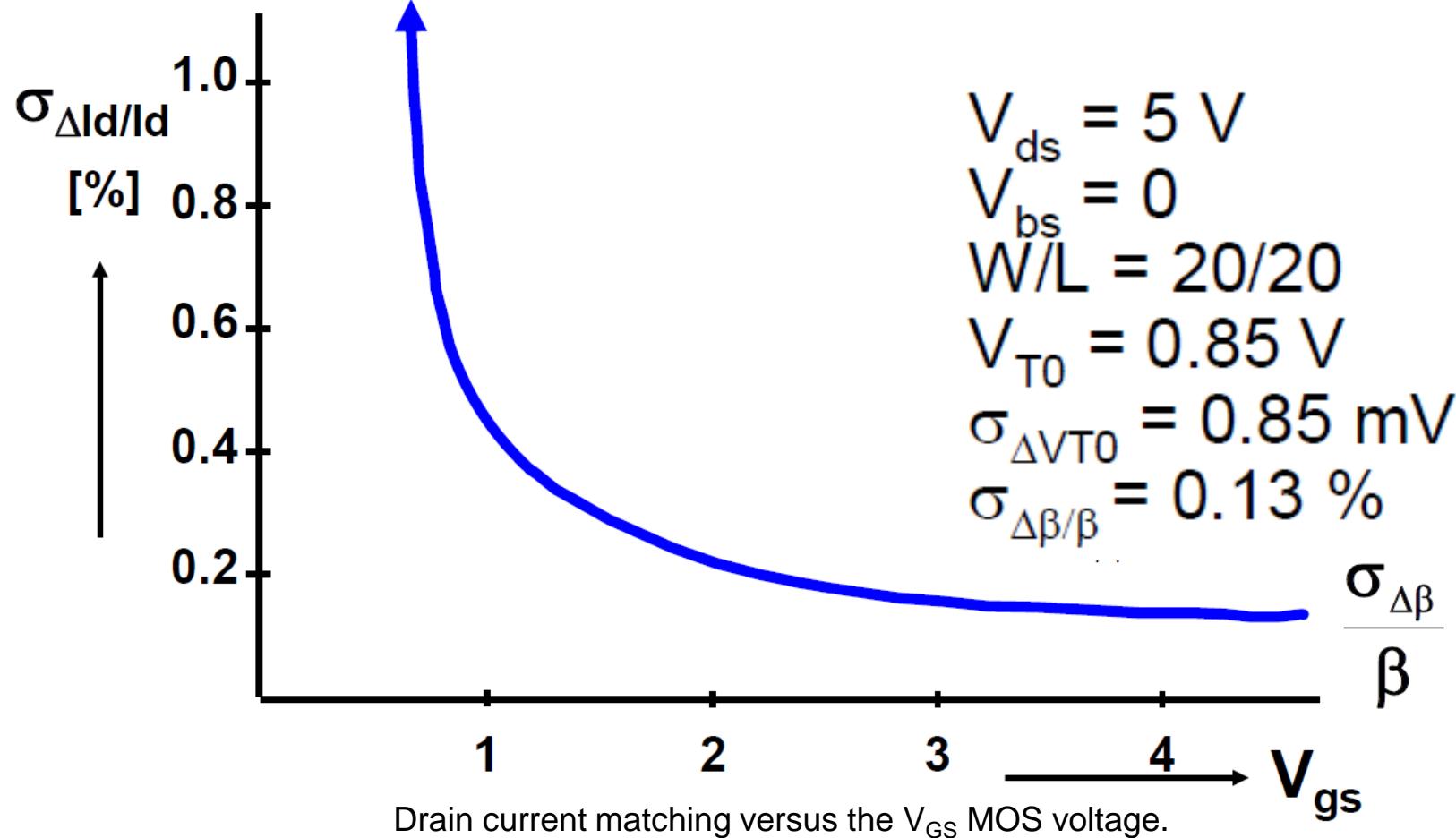
# Drain current matching in 0.5 μm

$$I_D = \frac{1}{2} \mu C_{OX} \frac{W}{L} (V_{GS} - V_{TH})^2$$



$$\frac{\sigma_{\Delta I_d}^2}{I_d^2} = \frac{4\sigma_{\Delta VT}^2}{(V_{GS} - V_T)^2} + \frac{\sigma_{\Delta \beta}^2}{\beta^2}$$

**BENG**



Drain current mismatch is dominated by the current factor mismatch if  $V_{GS} - V_T > 1 \text{ V}$ .

# Current factor mismatch of the MOS transistor

**Weak inversion**

$$I_d = I_0 e^{\frac{q(V_{GS} - V_T)}{mkT}}$$

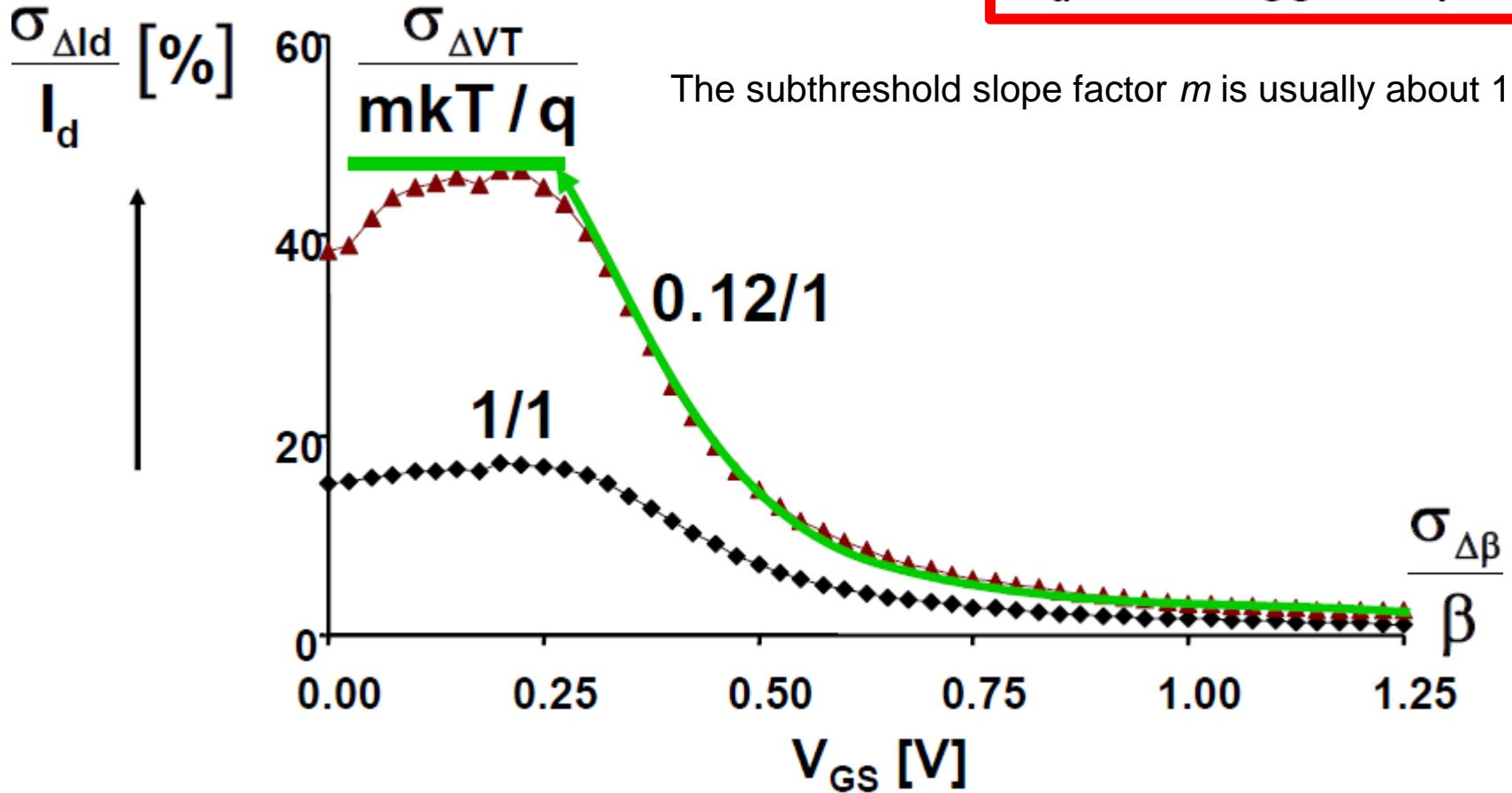
$$\frac{\sigma_{\Delta I_d}}{I_d} = \frac{\sigma_{\Delta VT}}{mkT / q}$$

**Strong inversion**

$$I_D = \frac{1}{2} \mu C_{OX} \frac{W}{L} (V_{GS} - V_{TH})^2$$

$$\frac{\sigma_{\Delta I_d}^2}{I_d^2} = \frac{4\sigma_{\Delta VT}^2}{(V_{GS} - V_T)^2} + \frac{\sigma_{\Delta \beta}^2}{\beta^2}$$

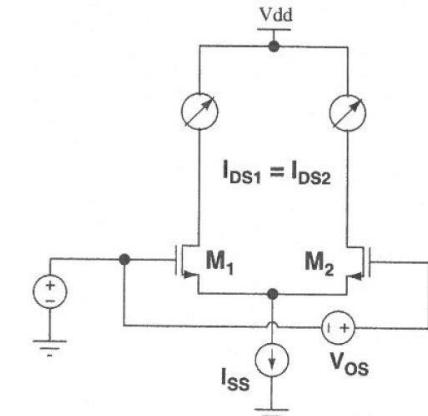
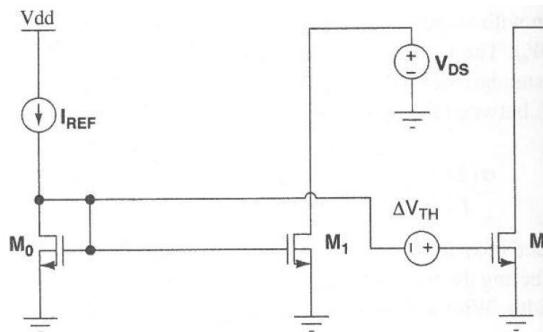
The subthreshold slope factor  $m$  is usually about 1.1 to 1.3.



Drain current matching versus the  $V_{GS}$  MOS voltage in the full gate-source voltage range.

# Mismatches - examples

$$I_D = \frac{1}{2} \mu C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2$$

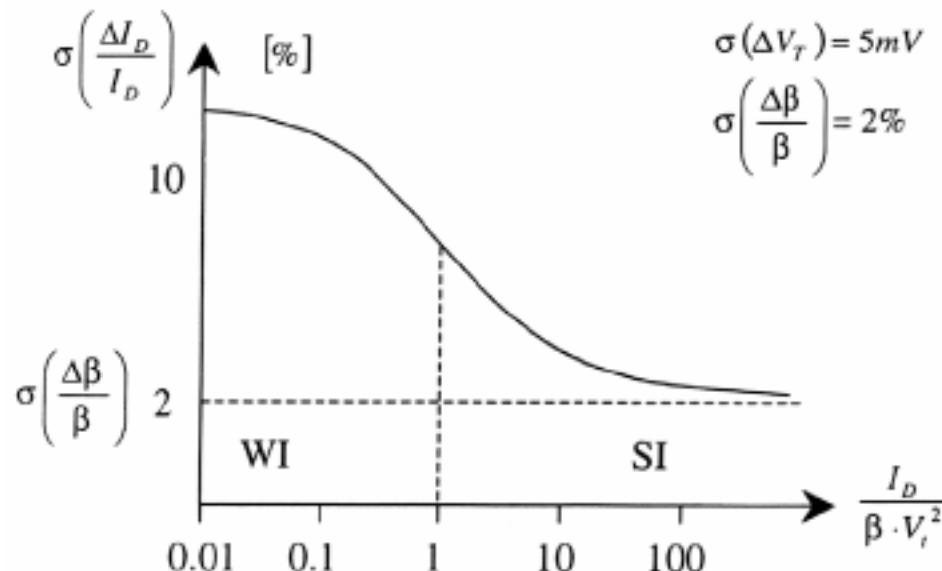


Analysis of the current (in currnt source) and voltage spread (in differential pair).

$$\frac{\sigma_{\Delta I_d}^2}{I_d^2} = \frac{4\sigma_{\Delta V_T}^2}{(V_{GS} - V_T)^2} + \frac{\sigma_{\Delta \beta}^2}{\beta^2}$$

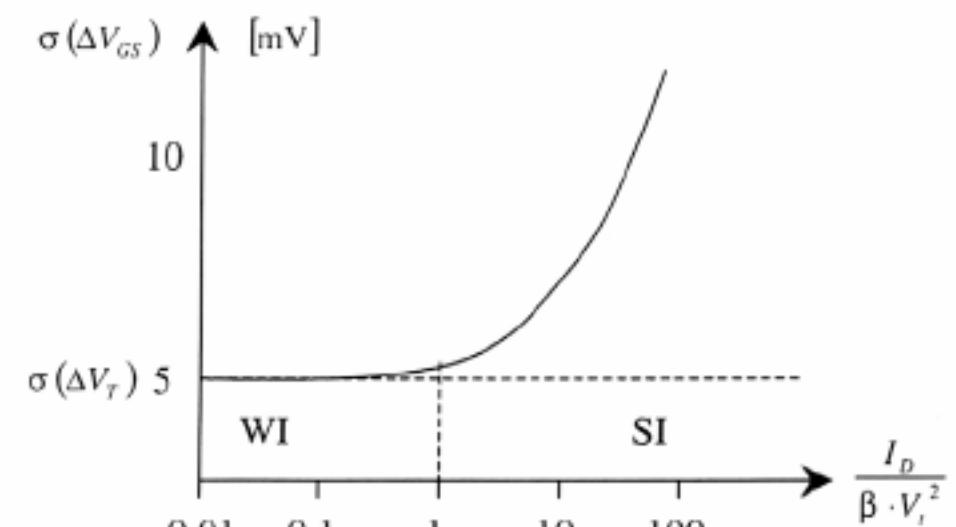
$$\sigma_{I_{DS}} = \sqrt{\left(\frac{\sigma_\beta}{\beta}\right)^2 + \left(\frac{g_m}{I_{DS}} \sigma_{VTH}\right)^2}$$

$$\sigma_{VGS} = \sqrt{\frac{\sigma_{I_{DS}}^2}{g_m^2}} = \sqrt{\left(\frac{I_{DS}}{g_m} \frac{\sigma_\beta}{\beta}\right)^2 + \sigma_{VTH}^2}$$



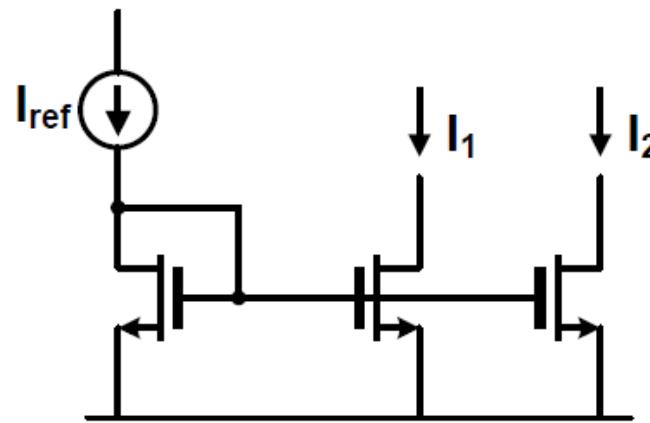
$$\sigma(\Delta V_T) = 5 \text{ mV}$$

$$\sigma\left(\frac{\Delta \beta}{\beta}\right) = 2\%$$



Drain current and voltage threshold matching versus the MOS channel inversion coefficient.

## cf) Current Source Mismatch



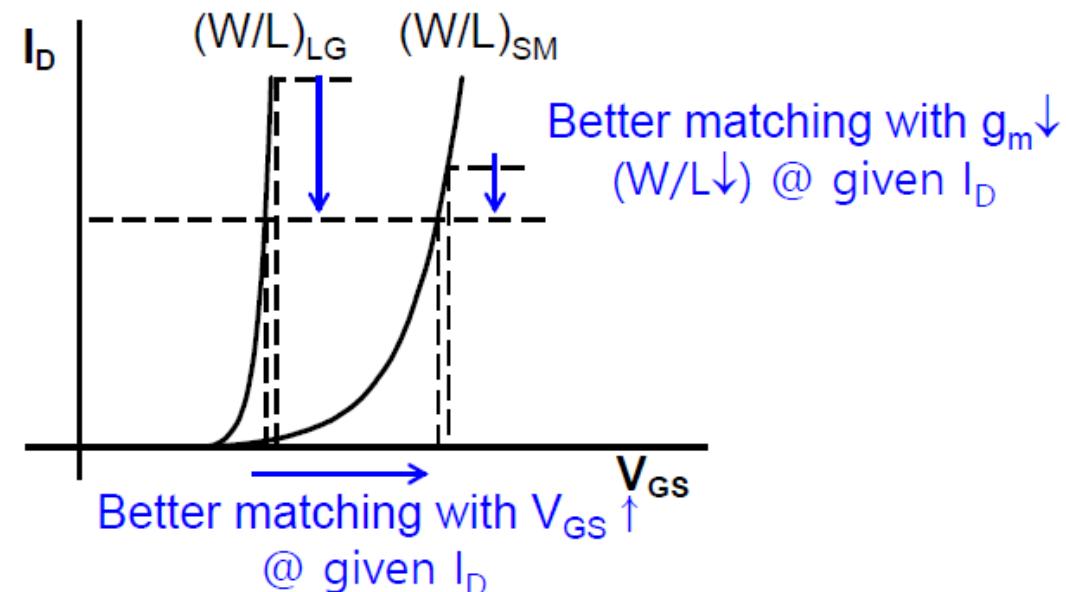
$$I_D = \frac{1}{2} \beta (V_{gs} - V_{th})^2 \quad \text{where } \beta = \mu C_{ox} (W/L)$$

$$\Delta I_D = \frac{1}{2} \Delta \beta (V_{gs} - V_{th})^2 - \beta (V_{gs} - V_{th}) \Delta V_{th}$$

$$\frac{\Delta I_D}{I_D} = \frac{\Delta \beta}{\beta} - \frac{2 \Delta V_{th}}{(V_{gs} - V_{th})}$$

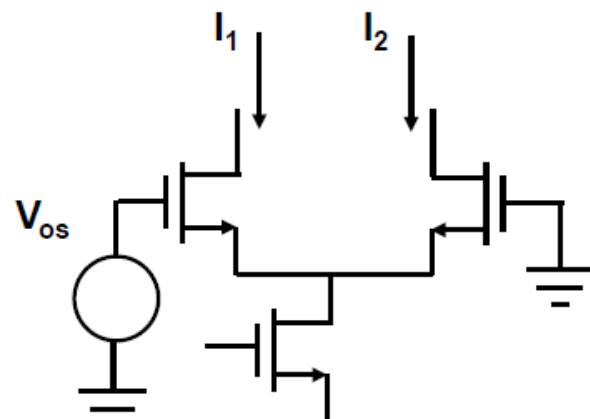
$$\sigma^2 \left( \frac{\Delta I_D}{I_D} \right) = \frac{1}{WL} \left( A_\beta^2 + \frac{4A_{VT}^2}{(V_{gs} - V_{th})^2} \right)$$

For the same WL



## Comparator Offset

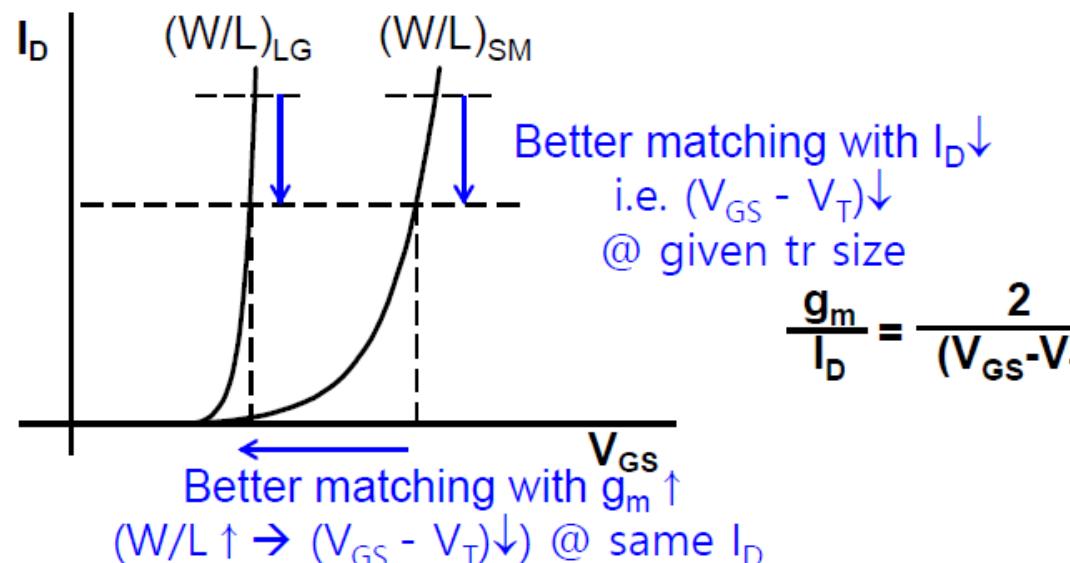
- Differential pair mismatch (offset)



$$V_{os} = V_{GS1} - V_{GS2}$$

$$\sigma^2(V_{os}) = \frac{1}{WL} \left( A_{VT}^2 + \frac{A_\beta^2 (V_{GS} - V_T)^2}{4} \right)$$

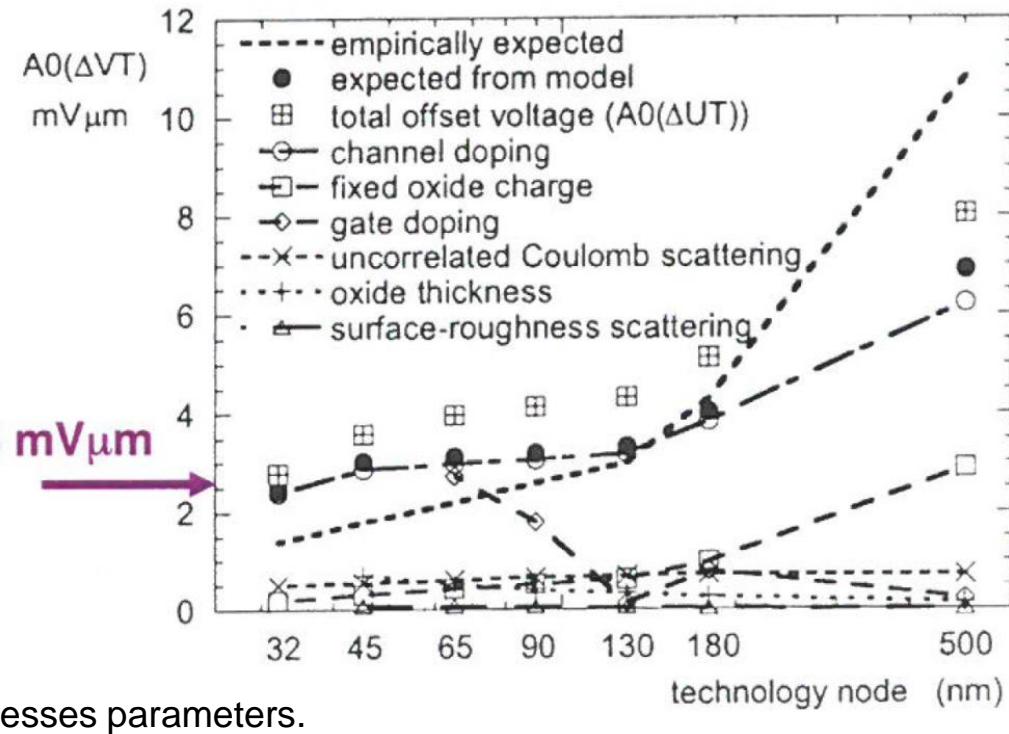
For the same WL



$$\frac{g_m}{I_D} = \frac{2}{(V_{GS} - V_T)}$$

# Mismatches in different CMOS nodes

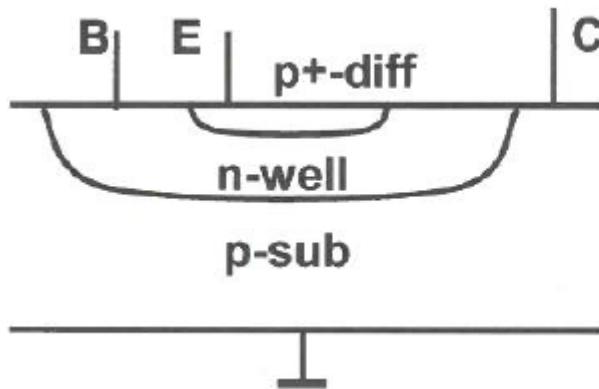
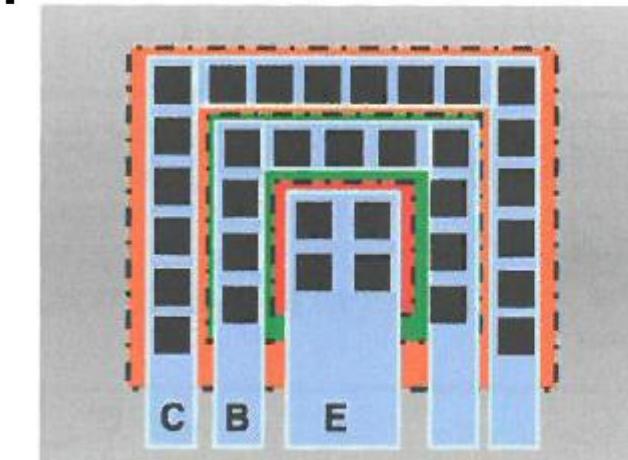
Despite there is still development in process and different technologies barriers are broken, there are different mismatches contributors which originate in different technology steps.



Values of main CMOS processes parameters.

Process	$V_{DD}$	$d_{ox}$	$C_{ox}$	$V_{T,n}$	$V_{T,p}$	$\beta_{\square,n}$	$\beta_{\square,p}$	$A_{V_{T,n}}$	$A_{V_{T,p}}$	$A_{\beta,np}$
Unit	Volt	nm	fF/ $\mu\text{m}^2$	Volt	Volt	$\mu\text{A}/\text{V}^2$	$\mu\text{A}/\text{V}^2$	mV $\mu\text{m}$	mV $\mu\text{m}$	% $\mu\text{m}$
0.8 $\mu\text{m}$	3.3	15	2.3	0.6	-0.65	125	55	10.7	18.0	4
0.6 $\mu\text{m}$	3.3	13	2.7	0.65	-0.8	150	50	11.0	8.5	
0.5 $\mu\text{m}$	3.3	12	2.9	0.6	-0.6	130	36	9	10	1.8
0.35 $\mu\text{m}$	3.3	7.7	4.3	0.63	-0.6	190	46	8	7.4	2
0.25 $\mu\text{m}$	2.5	6	6.9	0.57	-0.53	235	53	6	6	1.5
0.18 $\mu\text{m}$	1.8	4	8.3	0.48	-0.5	300	80	6	5	1.6
0.13 $\mu\text{m}$	1.2	2.5	11	0.34	-0.36	590	135	5	5	1.6
90 nm (LP)	1.2	2.3	11.7	0.37	-0.39	550	160	4.5	3.5	
65 nm (LP)	1.2	2.2	12.6	0.32	-0.36	450	200	4.5	3.5	1.2
45 nm (LP)	1.1	1.7	16	0.39	-0.42	300	100	4	4	1.2
28 nm	1.0	1.0		0.35	-0.35			2	2	

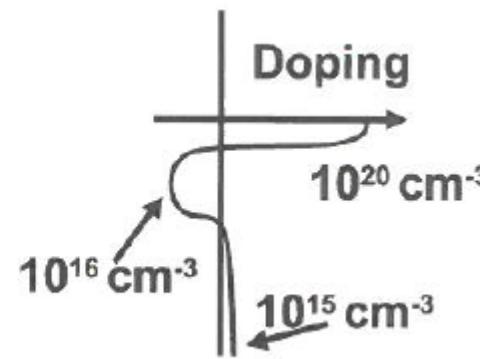
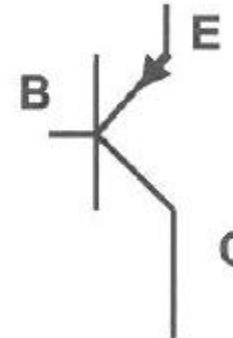
# Mismatches of bipolar transistors



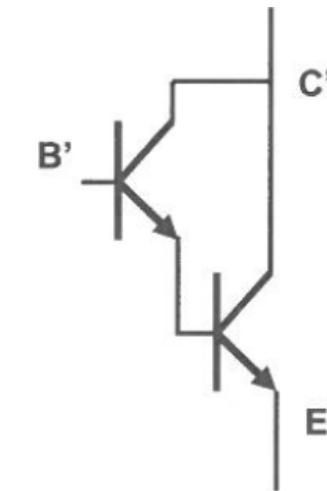
$$\sigma_{\Delta V_{be}} = \frac{A_{V_{be}}}{\sqrt{\text{area of emitter}}}.$$

Data for vertical pnp bipolar transistor.

$h_{fe}$	5–10
$V_A$	150 V <sup>-1</sup>
$V_{be}$ at $J_E = 1 \mu\text{A}/\mu\text{m}^2$	0.8 V
$\sigma_{V_{be}}$ emitter = 4 $\mu\text{m}^2$	0.1 mV



The bipolar transistor can be formed in a CMOS process from the n-well and the p-diffusion: layout, cross-section and the doping profile.



The Darlington circuit.

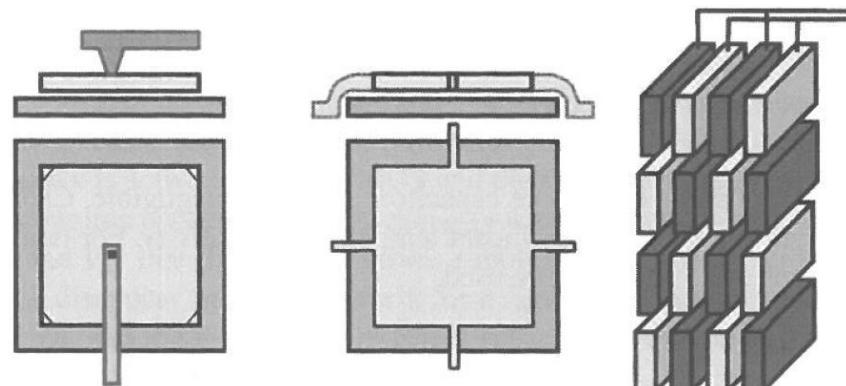
For a high-speed bipolar process with thin base regions there is relatively high impact of base dopant fluctuations and  $A_{VBE} = 0.3 - 0.5 \text{ mV}\mu\text{m}$ . The bipolar transistors processed in a CMOS process exhibits much better matching performance which is  $A_{VBE} = 0.1 - 0.2 \text{ mV}\mu\text{m}$ .

# Mismatches of capacitors

Mutual differences in the capacitors are mainly due to variations in edge definitions and in the dielectric thickness.

$$\frac{\sigma_{\Delta C}}{C} = \frac{A_{C,\text{area}}}{\sqrt{\text{area}}} = \frac{A_{C,\text{area}} \sqrt{\epsilon_0 \epsilon_r / d_{\text{ox}}}}{\sqrt{C}} = \frac{A_C}{\sqrt{C}}$$

The area is expressed in  $\mu\text{m}^2$ . With  $A_C=0.5\%\text{fF}$  a 400 fF capacitor has a mismatch of 0.025% which is 0.1fF



Three capacitors implementations. The left-side configuration uses the fact that a top contact is allowed. The middle configuration connects on the sides whenever the top contact is forbidden. The right one configuration uses fringe based capacitor in high metal density process.

Capacitors parameters in different processes (from 0.18  $\mu\text{m}$  to 90nm).

Material stack	Capacitance $\text{fF}/\mu\text{m}^2$	Voltage coeff $\text{V}^{-1}$	Temp coeff $^\circ\text{C}^{-1}$	Matching $\%/\sqrt{\text{fF}}$
Diffusion	0.5		$3 \times 10^{-4}$	
MOS gate 0.18 $\mu\text{m}$	8.3	$3-5 \times 10^{-2}$		
Fringe capacitors	1.5			0.3
MIM capacitors	4-15	$10^{-5}$		0.3
Plate capacitors (7-9 layers)	0.8			0.5
Poly-poly 0.35 $\mu\text{m}$	0.8	$5-10 \times 10^{-4}$	$-8 \times 10^{-5}$	

# Mismatches in different CMOS nodes

$$\frac{\sigma_{\Delta R}}{R} = \frac{A_R}{\sqrt{\text{area}}}$$

$A_R$  is expressed in the % $\mu\text{m}$  and area in  $\mu\text{m}^2$ .

Mutual differences in the geometrical dimensions and in the composition of the material determines resistors uniformity. Also an environment effects, like proximity of other structures, contacts, heat sources and stress influence the final resistors resistance.

Resistors parameters in different processes.

Material	Square resistance $\Omega/\square$	Voltage coeff $\text{V}^{-1}$	Temp coeff $\text{K}^{-1}$	Matching $A_R$ % $\mu\text{m}$
<i>n/p</i> Diffusion	75...125	$1 \times 10^{-3}$	$1..2 \times 10^{-3}$	0.5
<i>n</i> -Well diffusion	1,000	$80 \times 10^{-3}$	$4 \times 10^{-3}$	
<i>n</i> -Polysilicon	50...150		$-1...+1 \times 10^{-3}$	2
<i>p</i> -Polysilicon	50...150		$0.8 \times 10^{-3}$	2
Polysilicon (silicide)	3...5		$3 \times 10^{-3}$	
Aluminum	0.03...0.1		$3 \times 10^{-3}$	

# Offset in Differential Amplifiers

Component mismatch

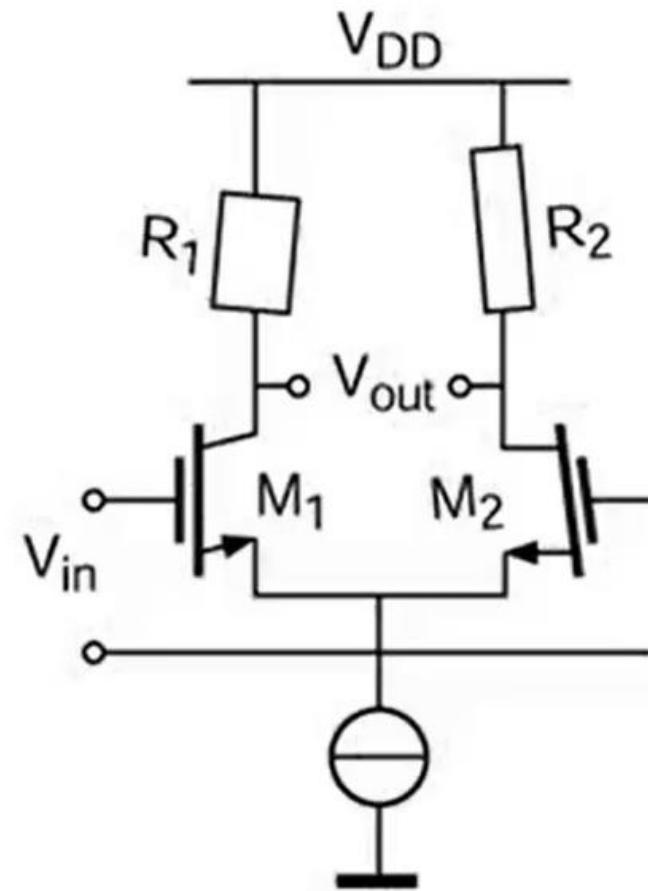
e.g.  $R_1 \neq R_2$ ,  $M_1 \neq M_2 \Rightarrow \text{offset}$

Mismatch is mainly due to

- Doping variations
- Lithographic errors
- Packaging & local stress

All things being equal

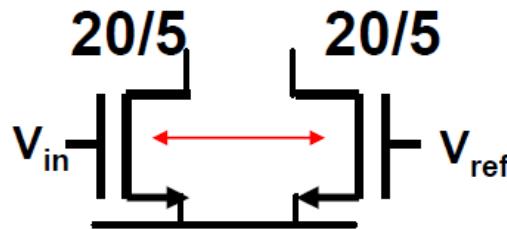
- Bipolar  $\Rightarrow V_{os} \sim 0.1\text{mV}$
- CMOS  $\Rightarrow V_{os}$  is 10 -100x worse!



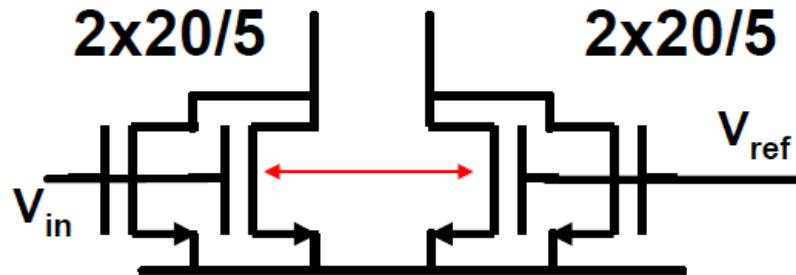
# Mismatches - examples

Calculate the the input-voltage-referred standard deviation if the  $V_{REF}$  is fixed ( $A_{VT} = 10\text{mV}\mu\text{m}$ ).

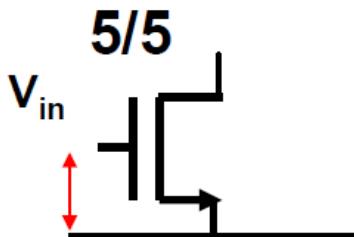
$$\sigma_{VT} = \frac{A_{VT}}{\sqrt{WL}}$$



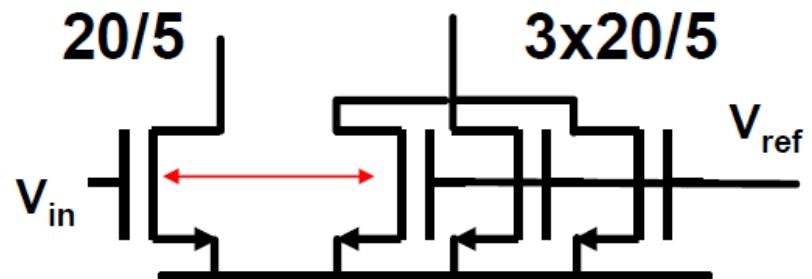
$$\sigma_{VT} = \frac{10}{\sqrt{5 \times 20}} = 1\text{mV}$$



$$\sigma_{VT} = \frac{10}{\sqrt{2 \times 5 \times 20}} = 0.7\text{mV}$$



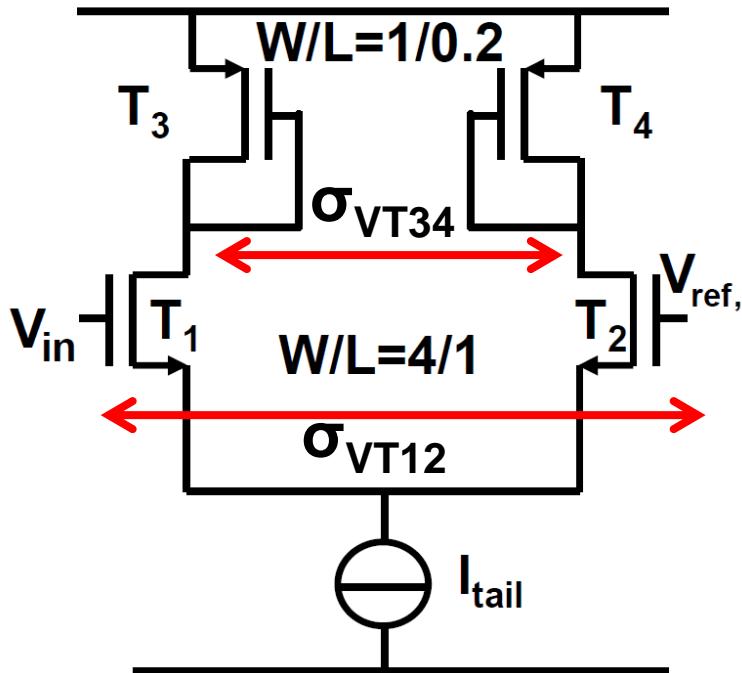
$$\sigma_{VT} = \frac{10}{\sqrt{5 \times 5}} \times \frac{1}{\sqrt{2}} = 0.7\text{mV}$$



$$\sigma_{VT} = \sqrt{\left( \frac{10}{\sqrt{20 \times 5}} \times \frac{1}{\sqrt{2}} \right)^2 + \left( \frac{10}{\sqrt{3 \times 20 \times 5}} \times \frac{1}{\sqrt{2}} \right)^2} = 0.81\text{mV}$$

# Mismatches - examples

Calculate the the input-voltage-referred mismatch if the  $A_{VT} = 4 \text{ mV}\mu\text{m}$ .



Mismatch on input due to  $\sigma_{VT12}$

Mismatch on input due to  $\sigma_{VT34}$

$$\sigma_{in} = \sigma_{VT12}$$

$$\sigma_{in} = \frac{\sigma_{VT34}}{A_v} = \frac{\sigma_{VT34} gm_{3,4}}{gm_{1,2}}$$

Total “input referred mismatch”:

$$\sigma_{in}^2 = \sigma_{VT12}^2 + \left( \frac{gm_{3,4}}{gm_{1,2}} \right)^2 \sigma_{VT34}^2$$

$$\sigma_{VT12} = \frac{A_{VT}}{\sqrt{W_1 L_1}} = \frac{4}{\sqrt{4 \times 1}} = 2 \text{ mV}$$

$$\sigma_{VT34} = \frac{A_{VT}}{\sqrt{W_3 L_4}} = \frac{4}{\sqrt{1 \times 0.2}} = 9 \text{ mV}$$

$$\sigma_{in}^2 = \sigma_{VT12}^2 + \left( \frac{gm_{VT34}}{gm_{VT12}} \right)^2 \sigma_{VT34}^2 = 2^2 + \frac{2 \times (1/0.2) \times 80 \times I}{2 \times (4/1) \times 350 \times I} 9^2 = 5.2^2$$

$\sigma = 5.2 \text{ mV}$

Note the PMOS diodes voltage contribuiton dominantes the overall input referred voltage offset.

# Mismatches Simulations

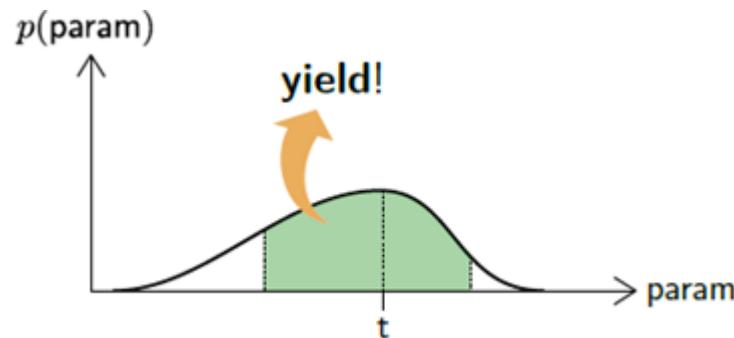
Particulars process elements parameters suffer from its substantial variation from chip to chip, and especially from wafer to wafer and a lot to lot. Therefore, there is a need to cope with that problem with the use of proper elements models.

There are two main ways to verify the circuits sensitivity on its elements variations:

- Monte Carlo simulations,
- Corner simulations.

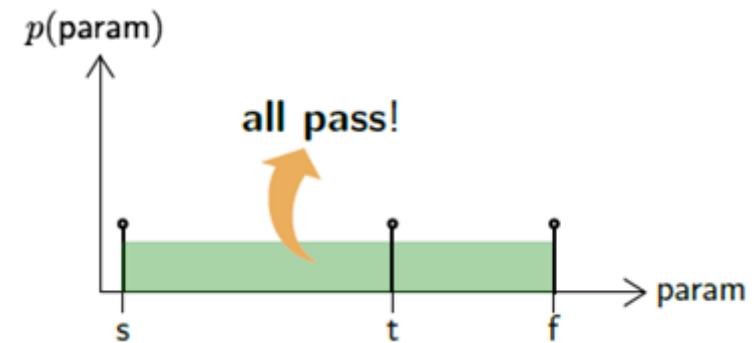
## Monte Carlo Simulations

**Local** deviations of model parameters mean the **statistical** change in all devices.



## Corner Simulations

**Global** deviations of model parameters mean the **same** change in all devices.



Results of Monte Carlo and corner simulations.

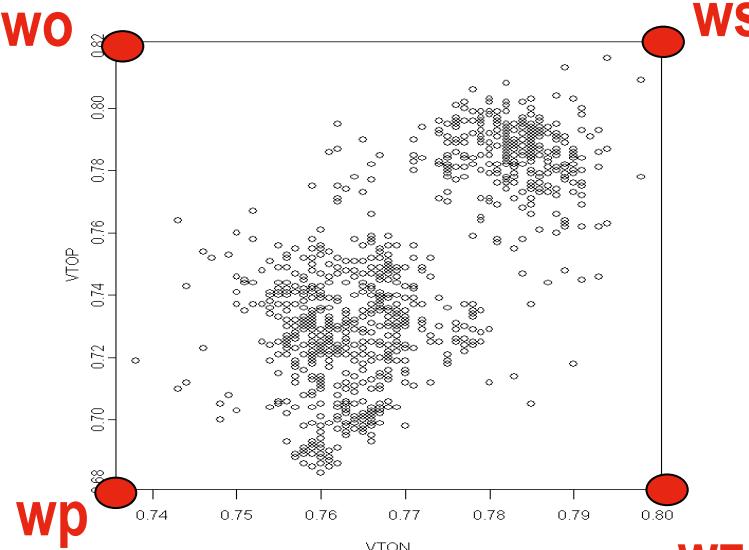
# Mismatches Simulations

MC vs. Corner analysis **WO**

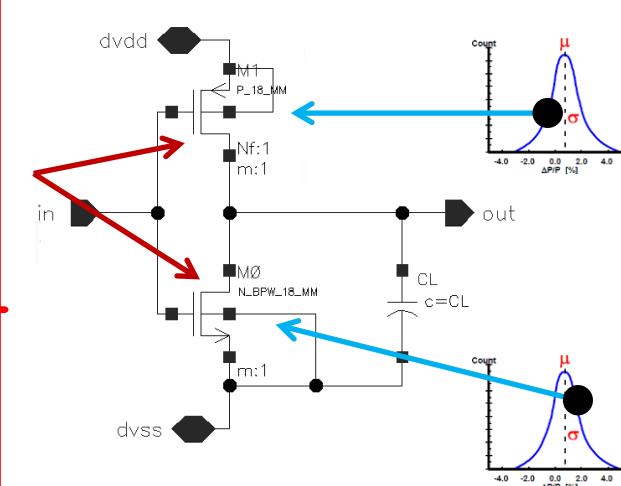
- MOS transistors (BSIM3V3)
  - WP** (fast NMOS, fast PMOS)
  - WS** (slow NMOS, slow PMOS)
  - WO** (fast NMOS, slow PMOS)
  - WZ** (slow NMOS, fast PMOS)
- Bipolar (VBIC, SGP)
  - HS** (high speed, high beta),
  - LB** (low speed, low beta)
  - HB** (low speed, high beta)

Each parameter variation for different corners.

Worst Case Type	NMOS Transistor			PMOS Transistor		
	$V_{TH}$	$\mu_0$	$t_{OX}$	$V_{TH}$	$\mu_0$	$t_{OX}$
TM typical mean	↔	↔	↔	↔	↔	↔
WP (ff) worst power	↓	↑	↓	↓	↑	↓
WS (ss) worst speed	↑	↓	↑	↑	↓	↑
WZ (sf) worst zero	↑	↓	↔	↓	↑	↔
WO (fs) worst one	↓	↑	↔	↑	↓	↔

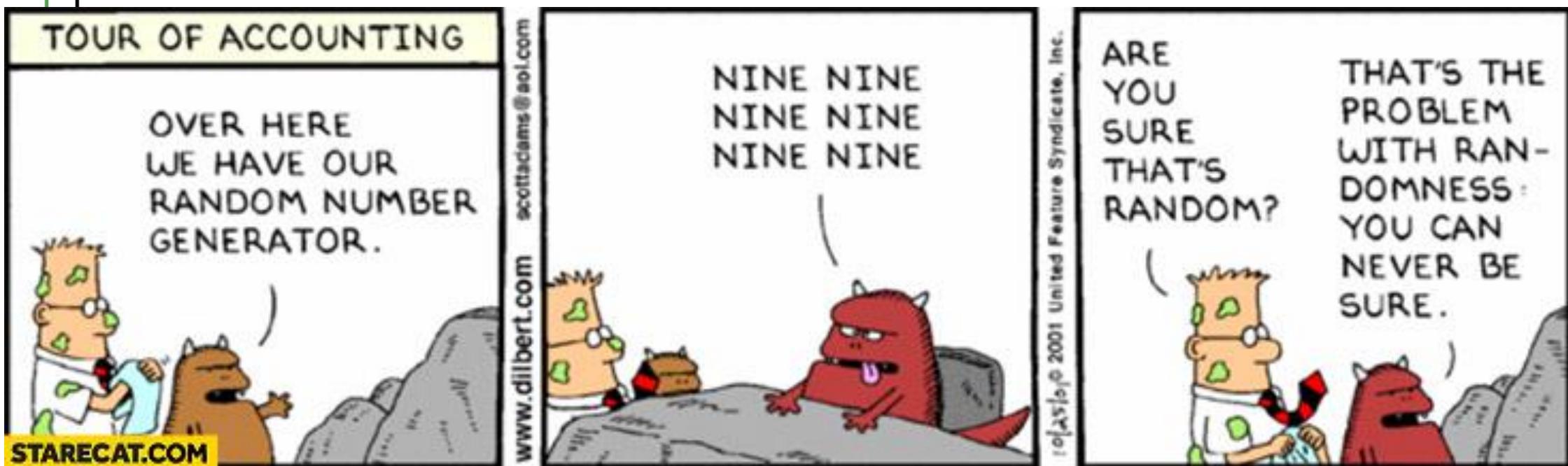


Threshold voltage distribution for MC simulation and borders marked by the process corners.



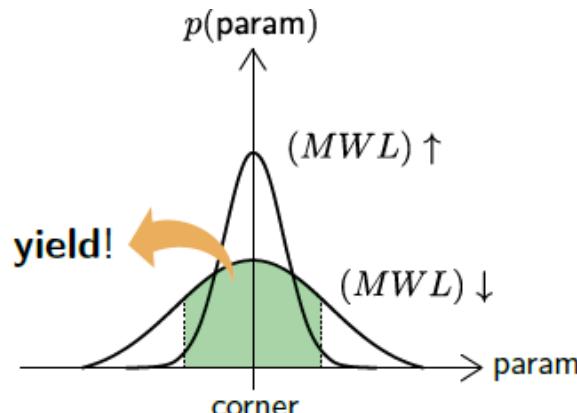
Inverter and its parameter visualization for MC and corner simulations.

# What is a parameter generator?



# Mismatches Simulations

Monte Carlo and corner simulations are performed by proper elements parameter changings. A standard way is to provide by the process factory models of elements with its main parameters variations included. However, if these are not available one need to develop its own statistical models based on the process documentation.



Simulations results of Monte Carlo for a given corner.

```
.lib model
.model cnm25modn nmos LEVEL = 2
+ TOX = 380E-10 VTO = {vton} NSUB = 2.64E16 UO = {uon}
+ UCRIT = 1E4 UEXP = 6.86E-2 NFS = 7.11E11
+ DELTA = 2.20 RS = 93.8 LD = 9.13E-7 XJ = 8.24E-8
+ VMAX = 5.96E4 NEFF = 1.48 CJ = 3.50E-4 MJ = .40
+ CJSW = 5.95E-10 MJSW = .29 PB = .65
+ AF = 1.33 KF = 1e-29
.model cnm25modp pmos LEVEL = 2
+ TOX = 380E-10 VTO = {vtop} NSUB = 1.36E16 UO = {uop}
+ UCRIT = 1E4 UEXP = 1.16E-1 NFS = 6.62E11
+ DELTA = 1.82 RS = 134.9 LD = 8.10E-7 XJ = 2.78E-9
+ VMAX = 1.20E5 NEFF = 6.67E-2 CJ = 3.82E-4 MJ = .35
+ CJSW = 7.38E-10 MJSW = .39 PB = .56
+ AF = 1.33 KF = 1e-29
.model cnm25cpoly c CJ= 4.227E-4 CJSW=0.0
.endl
```

Listening of the MOS models.

```
.param avto=30e-9
.param rauo=5e-8
.param rac=7.3e-8

.subckt cnm25modn d g s b param: w=3u l=3u
+ ad=0 as=0 pd=0 ps=0 m=1
.param vton=.942+rndgauss(avto/sqrt(m*w*l))
.param uon=648*(1+rndgauss(rauo/sqrt(m*w*l)))
.model modnlocal nmos level=2 vto={vton} uo={uon} tox=...
mn d g s b modnlocal w={w} l={l}
+ ad={ad} as={as} pd={pd} ps={ps} m={m}
.ends

.subckt cnm25cpoly t b param: w=30u l=30u m=1
.param cj=4.227E-4*(1+rndgauss(rac/sqrt(m*w*l)))
.model cpolylocal c cj={cj} cjsw=0.0
cpip t b cpolylocal w={w} l={l} m={m}
.ends
```

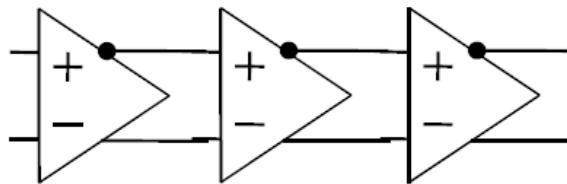
Listening of the MOS MC models.

<pre>.lib tt .param vton=.942 .param vtop=-1.139 .param uon=648 .param uop=213 .lib 'cnm25proc.lib' model</pre>	<pre>.lib ff .param vton=0.7 .param vtop=-0.9 .param uon=881 .param uop=295 .lib 'cnm25proc.lib' model</pre>	<pre>.lib ss .param vton=1.1 .param vtop=-1.3 .param uon=415 .param uop=131 .lib 'cnm25proc.lib' model</pre>
---	--	--

Listening of the MOS corner models.



AGH



3-stage balanced DC-coupled VGA

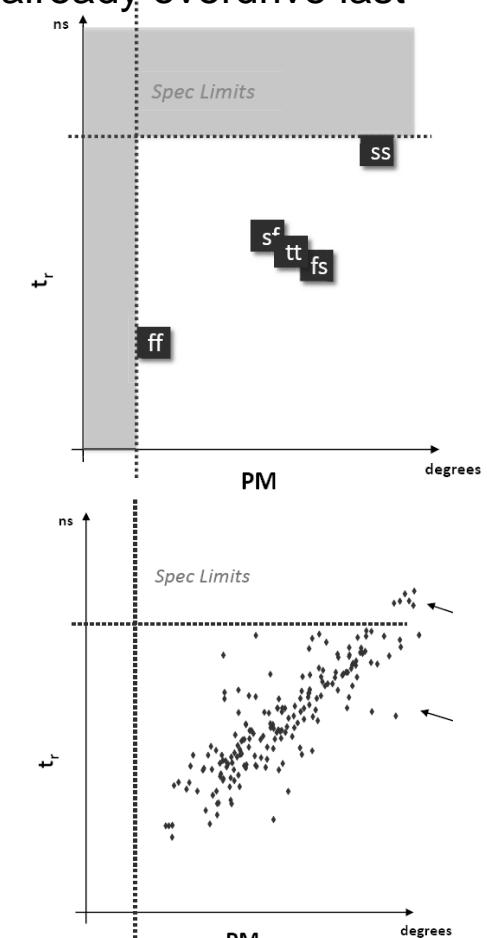
## Mismatches Simulations

Which one is better, MC or corner?

Each stage has gain of 0...20dB => max. total gain = 60dB

Corner simulations show good performance & no offset voltage 😊

But in reality even a small input offset can already overdrive last stage resulting in yield almost zero. 😥



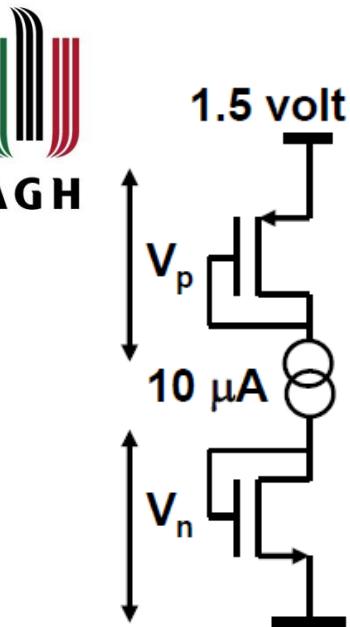
**Corners** → Stresses the design at extreme process points

- Good for quickly finding problems
- Usually 4-20 process corners, usually combined with temperature, voltage.
- Gives unfortunately no real yield prediction & does not treat mismatch

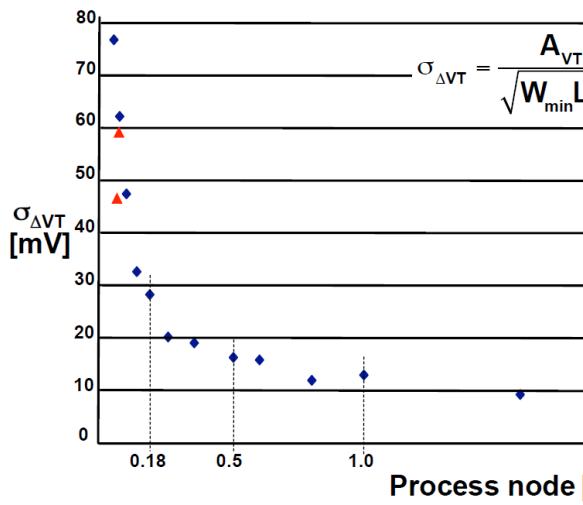
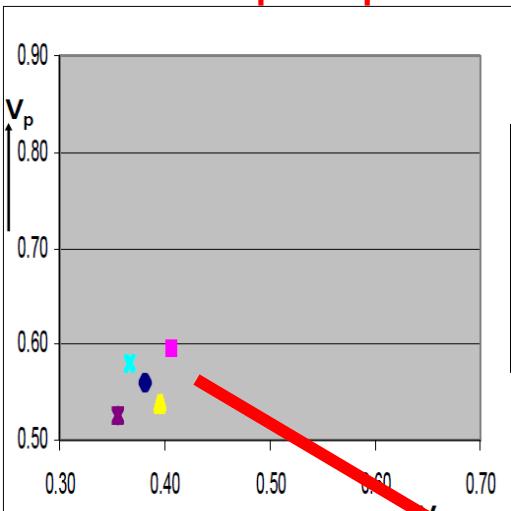
**Monte Carlo** → Mimics the production environment

- Requires statistical models
- Needs usually 50 to 200 runs to get stable mean & sigma
- Can detect failure regions not found by corners
- Can treat mismatch
- Get correlations, histograms, yield and CPK (process capability measure)
- MC needs many points for verification of higher yields, like 2500-8000 for  $3\sigma$

# Missmatches in nanometer CMOS process

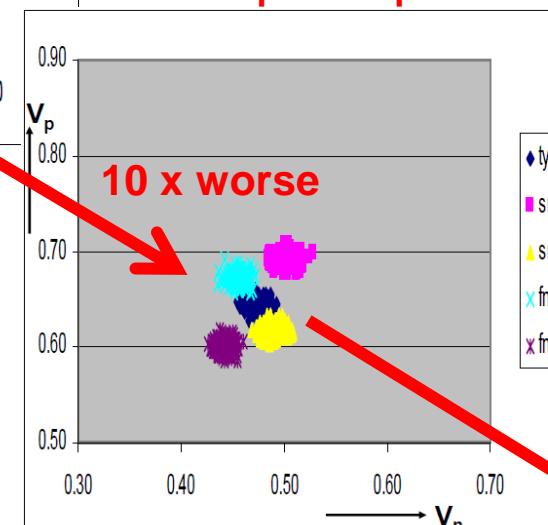


$W/L = 10 \mu\text{m} / 5 \mu\text{m}$



The standard deviation of the MOS transistor threshold versus the process node for minimum size channel dimensions.

$W/L = 1 \mu\text{m} / 0.5 \mu\text{m}$

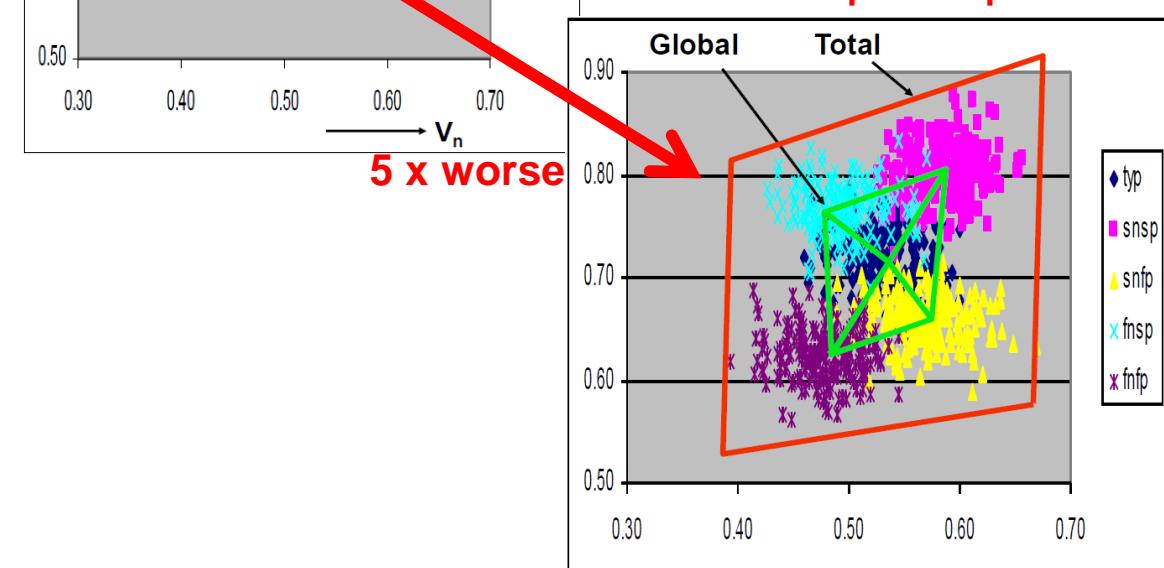


Voltage threshold mismatches in CMOS 90nm process.

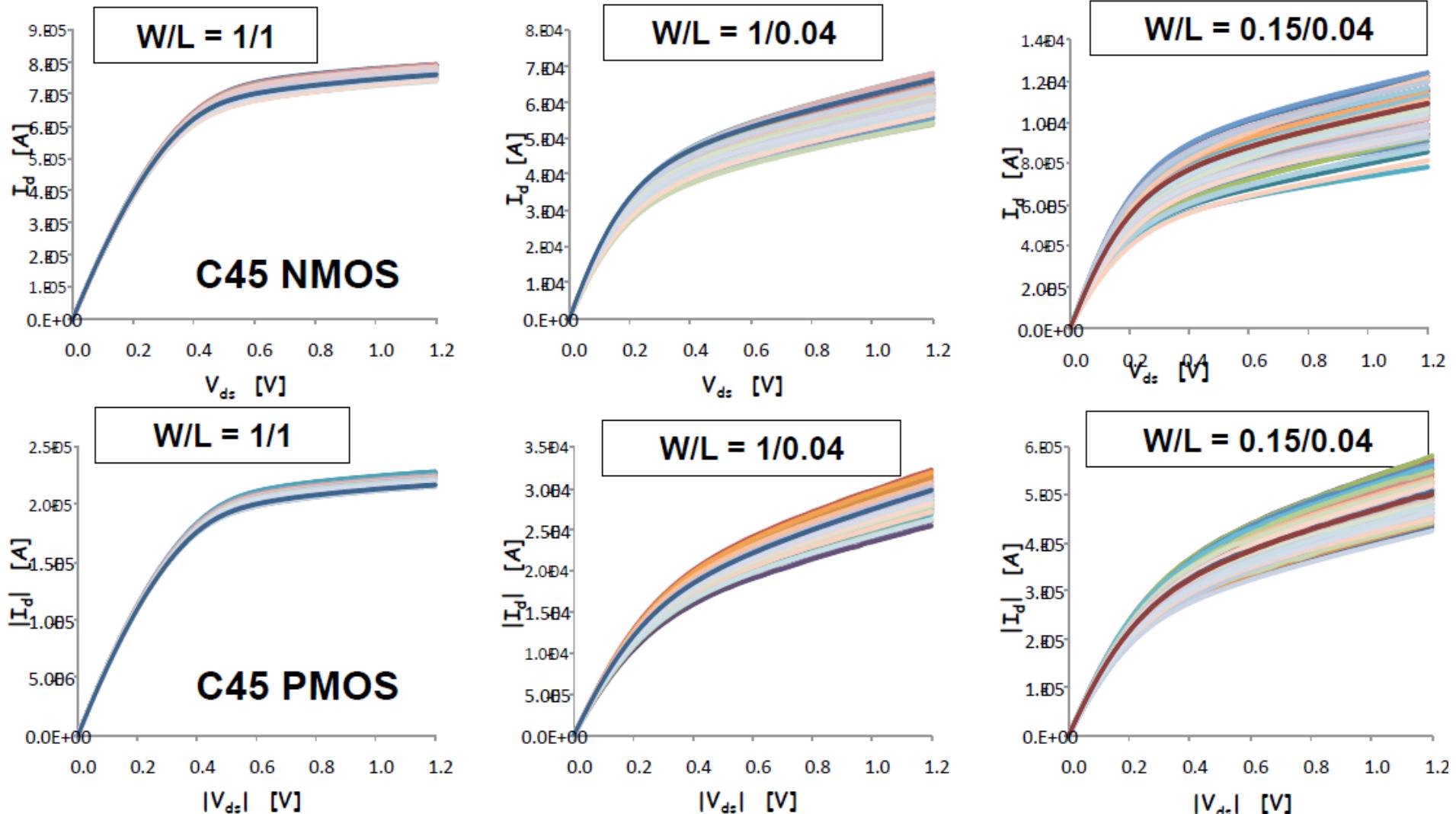
As it can be seen one should be aware of using minimum size transistors especially in modern nanometer processes.

5 x worse

$W/L = 0.2 \mu\text{m} / 0.1 \mu\text{m}$



# Missmatches in nanometer CMOS process



PMOS and NMOS output characteristics for different transistors channel dimensions ( $V_{GS} = 1.1$  V, CMOS 45nm).

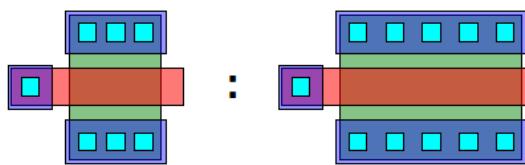
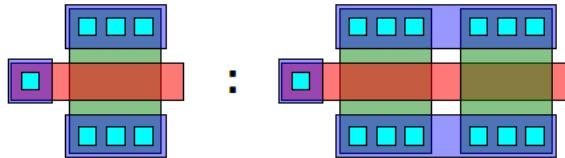
**Large degradation of uniformity and output MOS resistance is visible as the MOS channel area is decreased.**

# General Matching Rules

e.g. 1 : 2 ratio

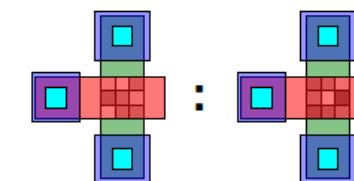
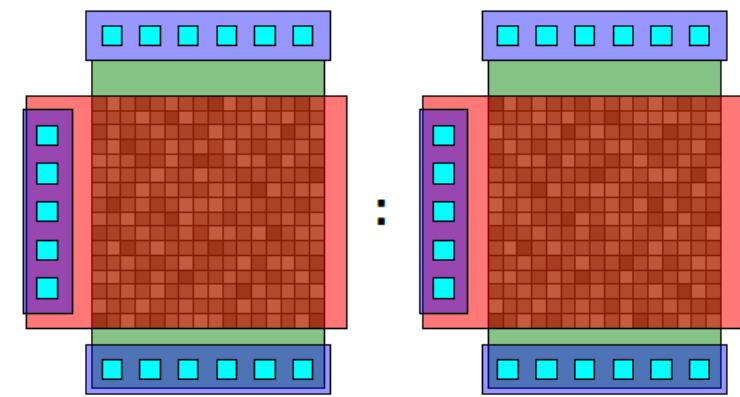
## ► Unitary elements

- Play with **multiplicity** only
- Same ratio for **second order** effects also  
(e.g. area and perimeter ratio in caps)
- **Larger area** for the overall array



## ► Large **area** devices

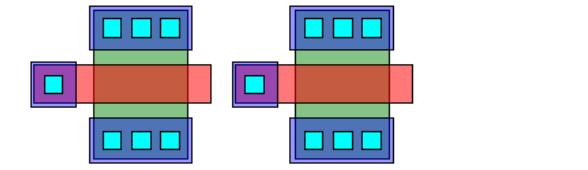
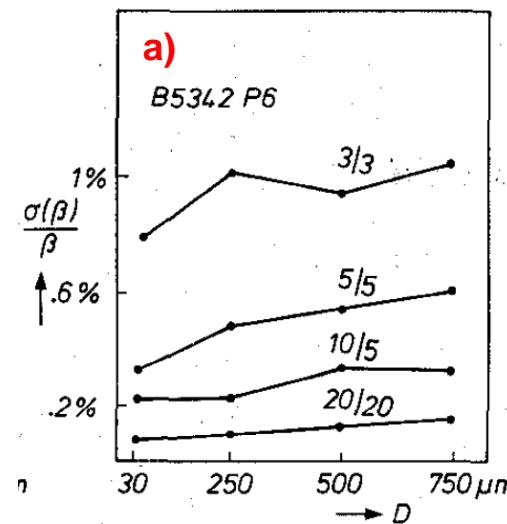
- Technology local **granularity**  
(e.g. distribution of dopants)
- **Pelgrom Law**  $\propto \frac{1}{\sqrt{WL}}$



# General Matching Rules

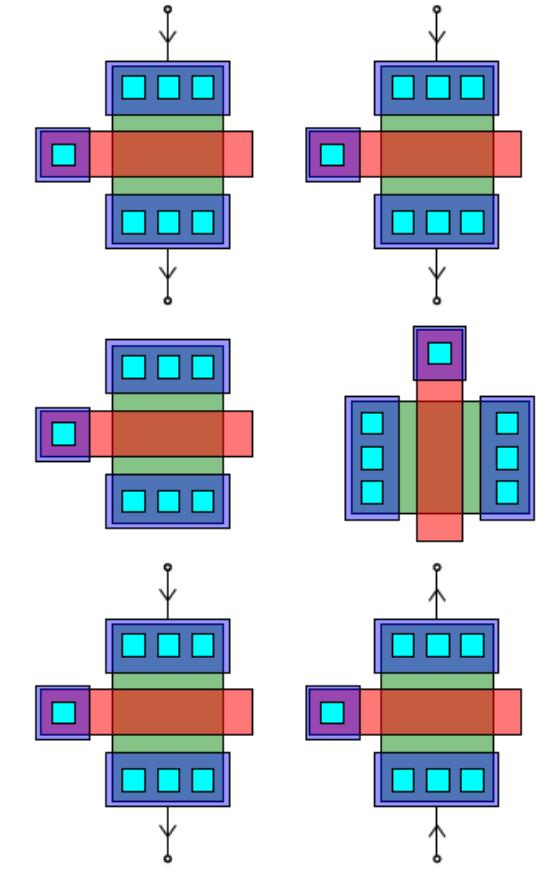
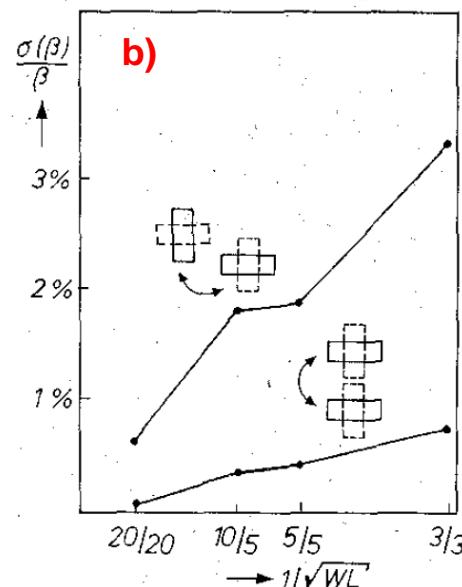
## ► Minimum distance

- Technology global drifts  
(e.g. gate oxide thickness slope)
- Design rule spacing limits



## ► Same orientation

- Anisotropic materials  
(e.g. wafer crystal lattice orientation)
- Longer routing may be required...

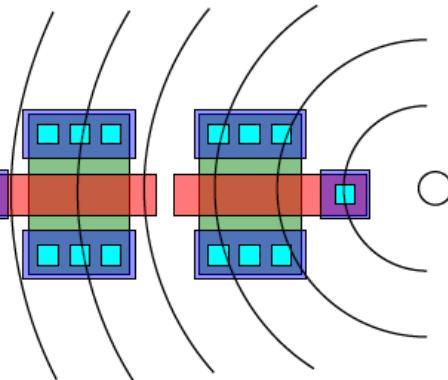
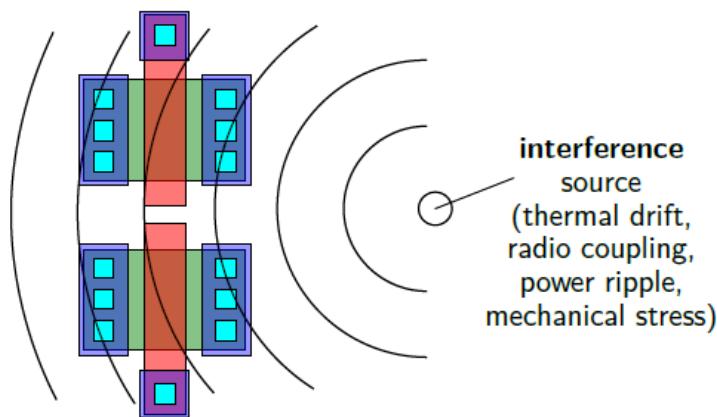


Standard deviation of  $\beta$  for: a) different transistors distance, b) parallel and rotated placement.

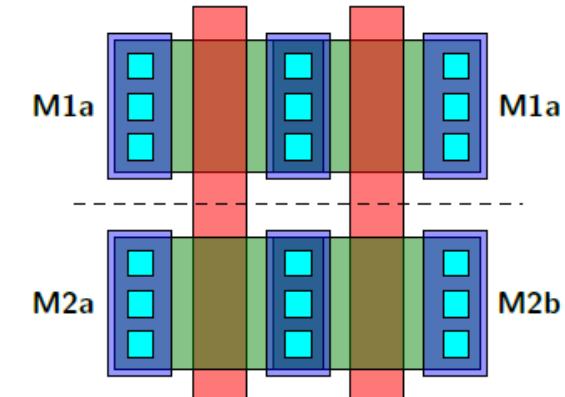
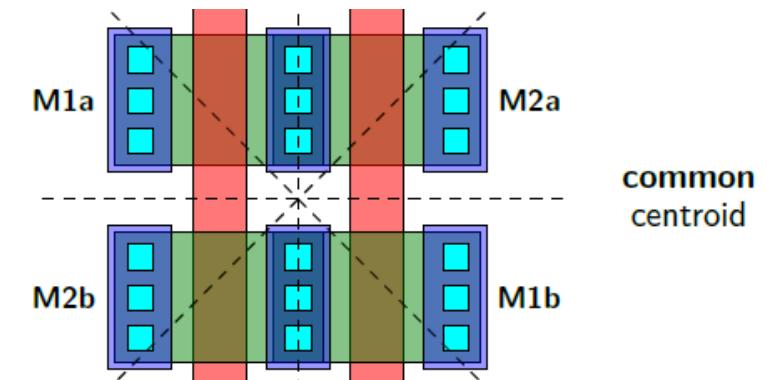
# General Matching Rules

## ► Same **symmetry**

- Differential circuits (common mode interference)
- Complex **floorplan**



- Compensation of **linear** gradients  
(and non-linear at short distance)
- **Longer routing** is usually required...



# General Matching Rules

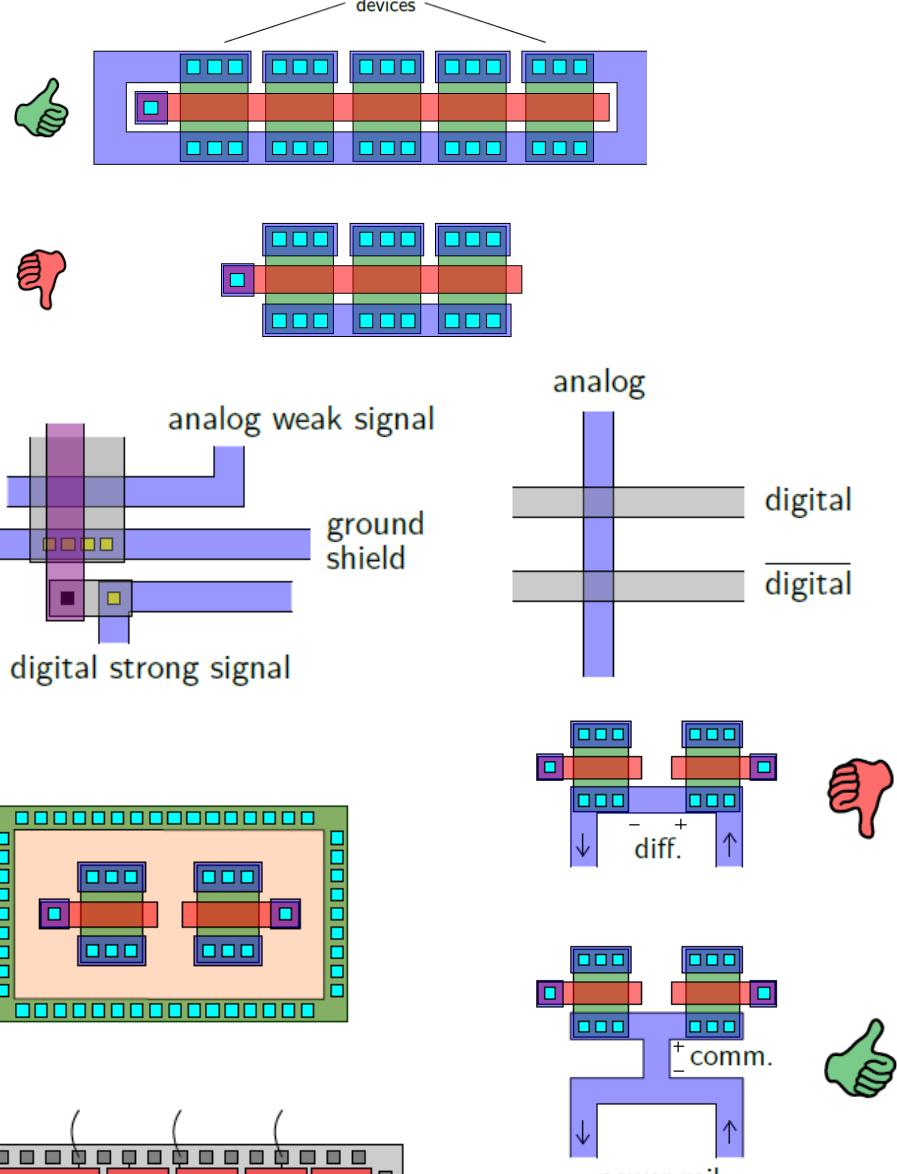
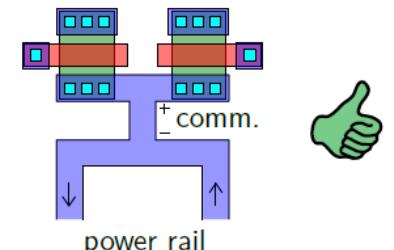
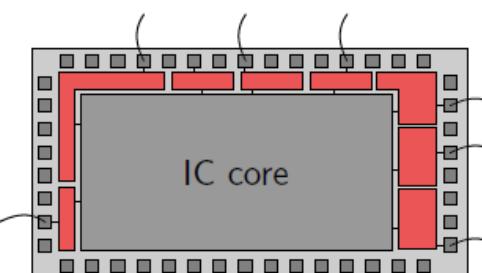
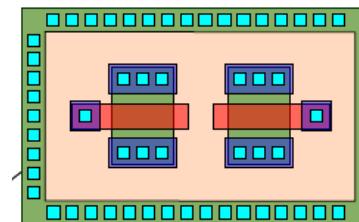
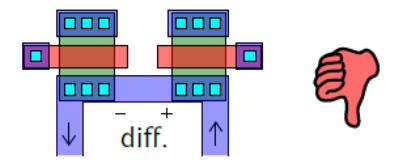
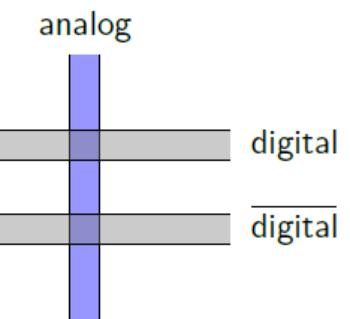
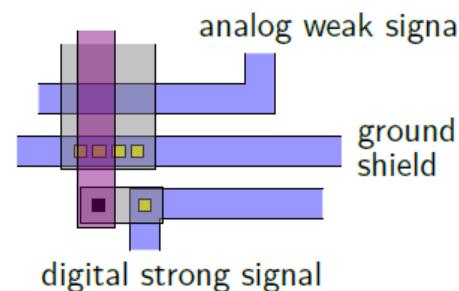
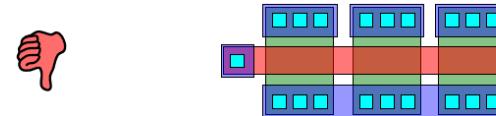
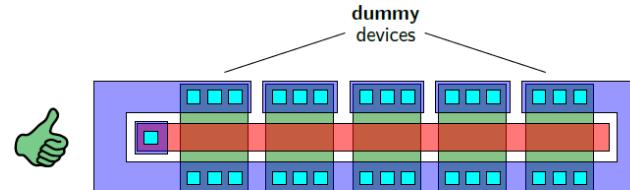
► Same surround

- Inter-device second order effects  
(e.g. parasitic RLC)
- Larger area for the overall array

► Signal integrity between analog, digital, RF... domains

► Avoiding signal coupling through power rails

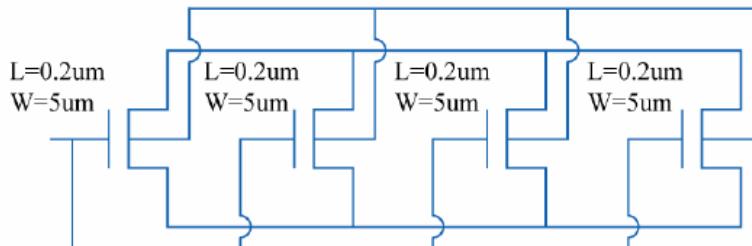
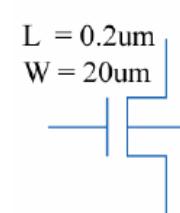
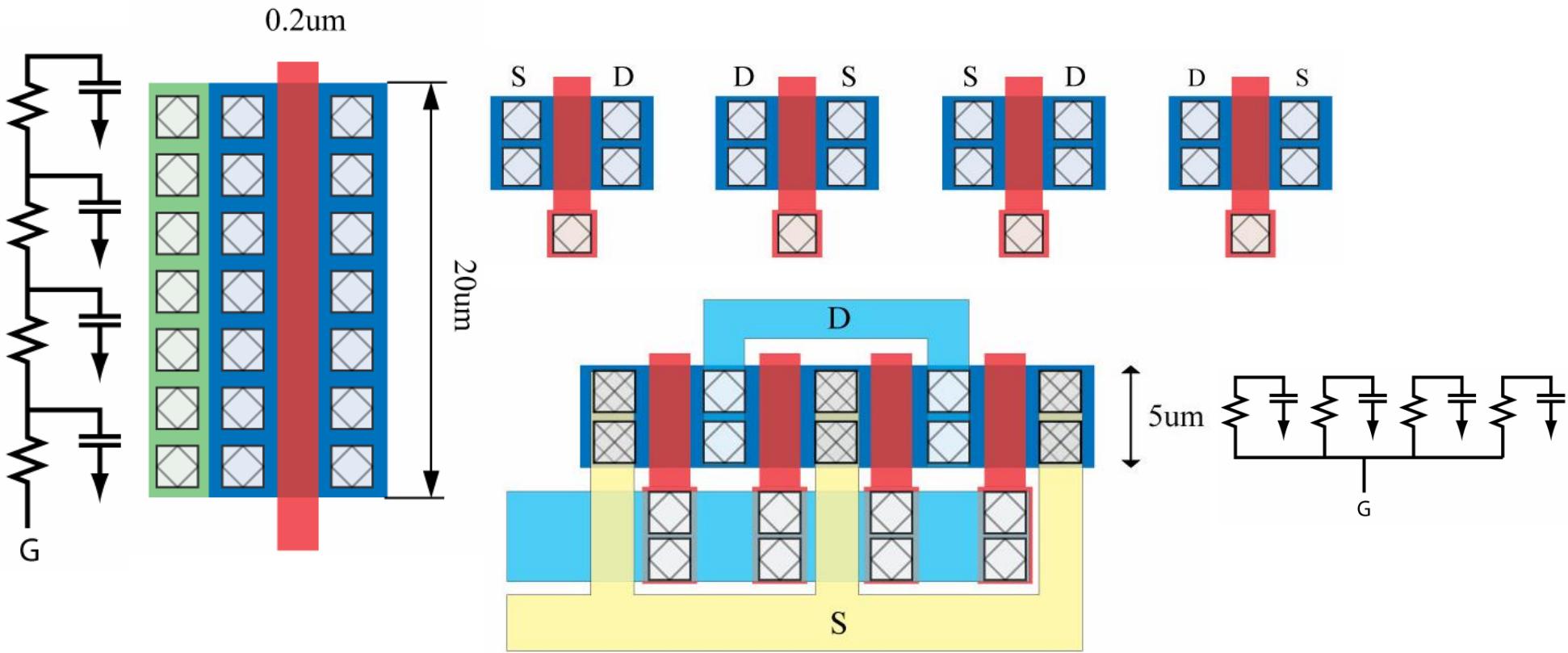
► Investing peripheral free area for on-chip decoupling capacitors





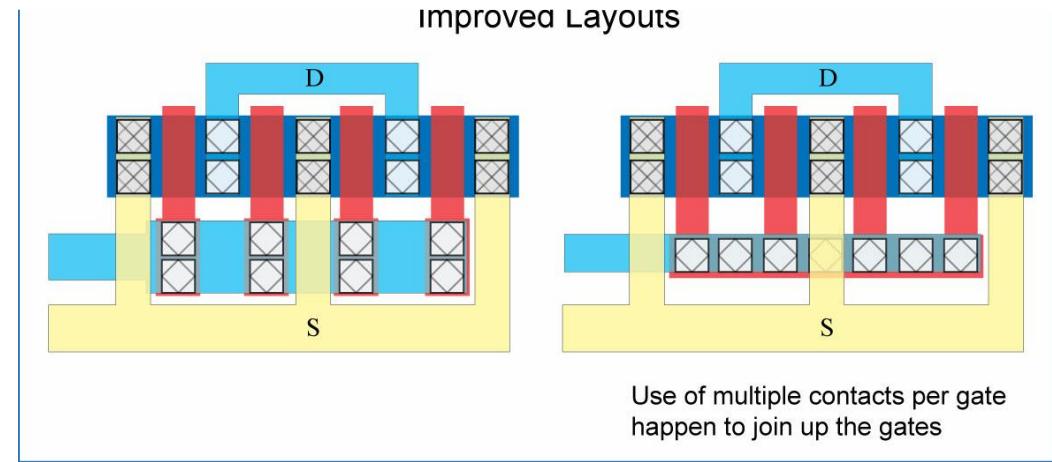
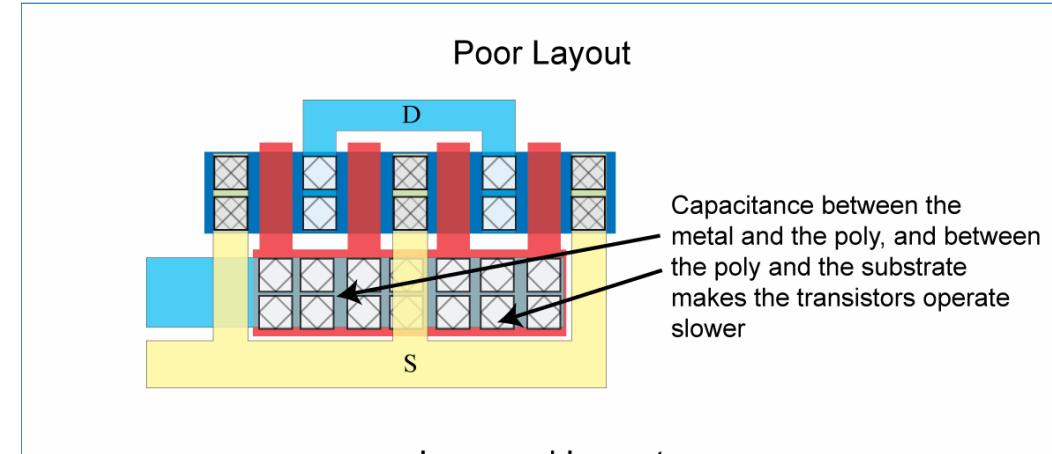
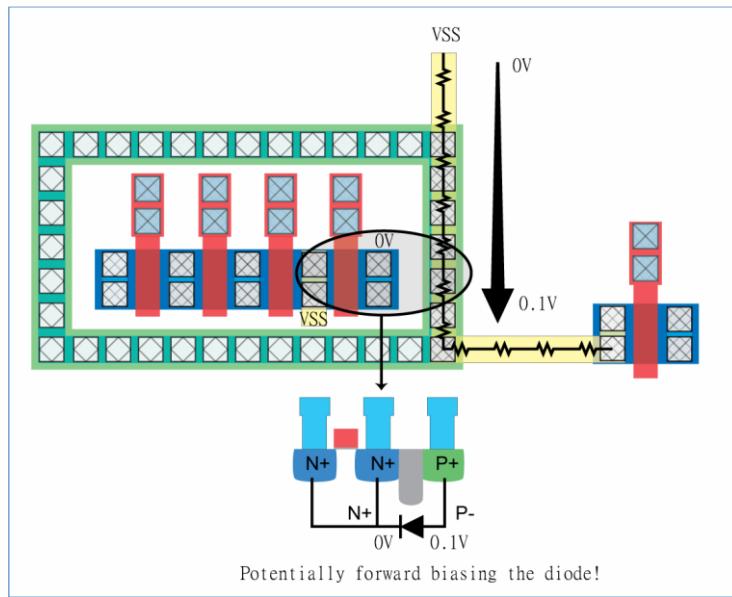
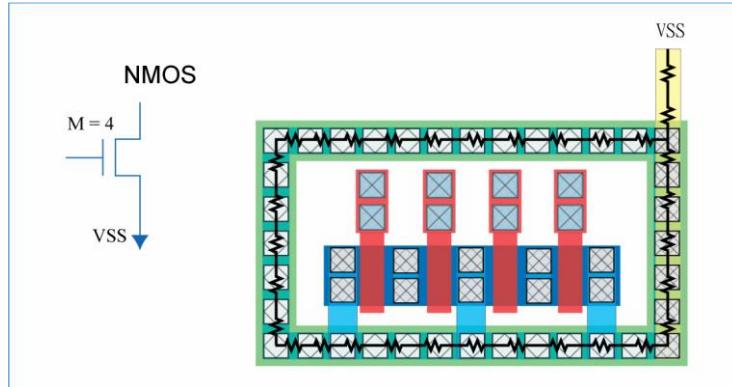
AGH

## General Matching Rules



In case the high frequency operation is required consider using the finger shape transistors. In that way the parasitic capacitances (gate to diffusion) and resistances (gate) can be significantly reduced. However be aware of slight different operating points among these two laid out transistors.

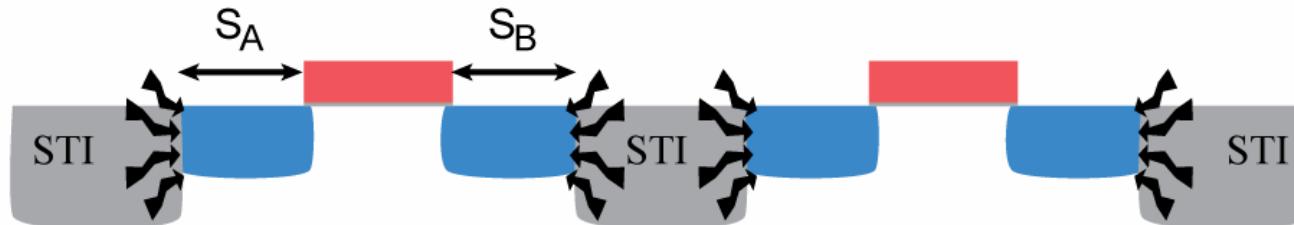
# General Matching Rules



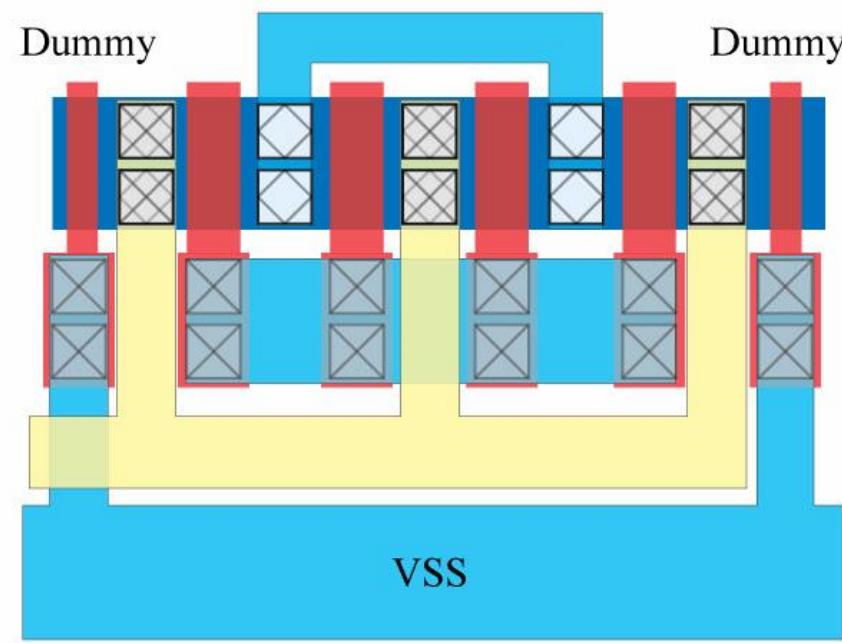
Avoid using guard rings contacts for supplying the circuits. If low current is expected it can be accepted. Beware of any current spikes that may show up during transistors switching.

Avoid using poly for routing as it adds more capacitance parasitics and may generate antenna violations (hard to mitigate as poly is created prior diffusion → protecting diodes than do not exist yet).

## General Matching Rules



Mechanical Stress from STI on Diffusion



Use dummy transistors to minimize the STI stress on transistors channel.