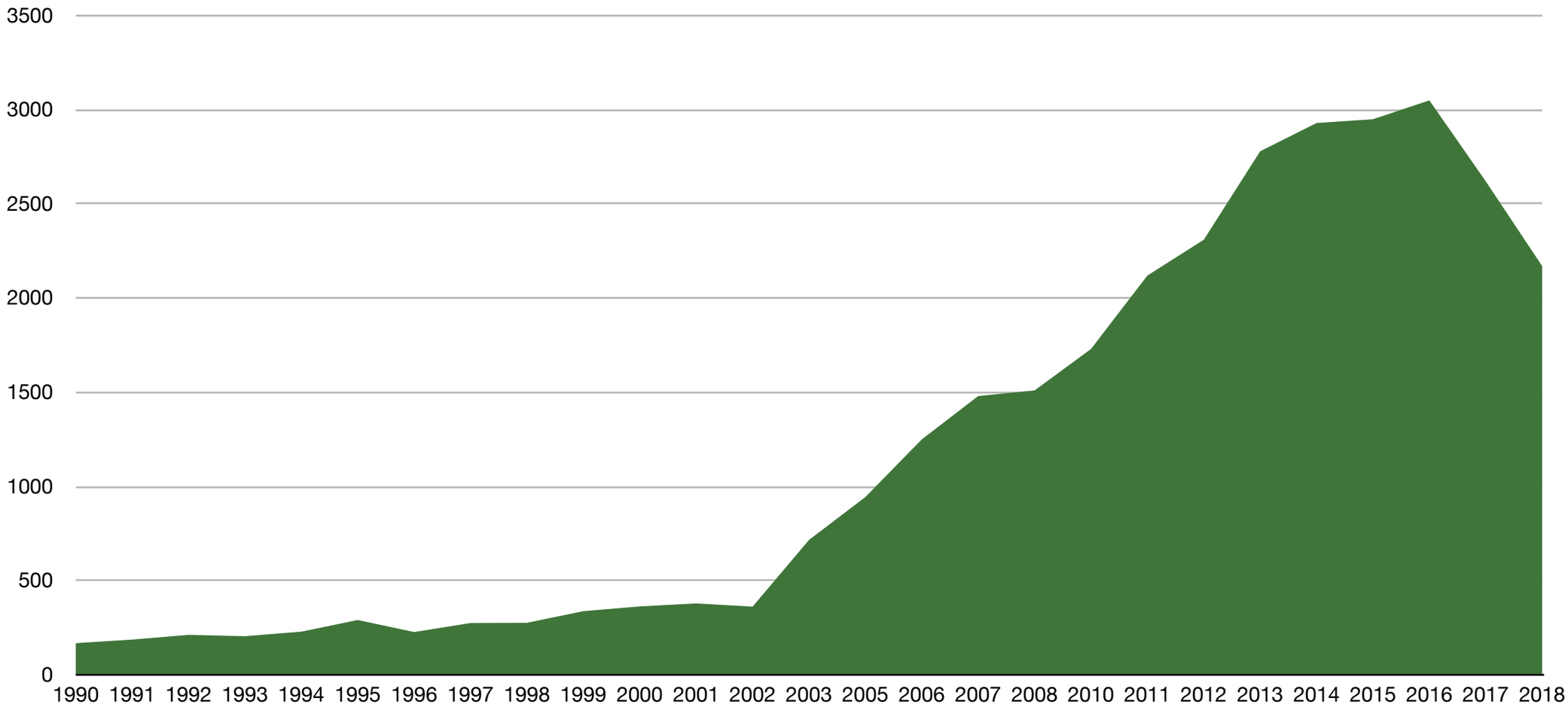
Variability and matching

DAT116, Nov 12 2018

Lena Peterson
Lars Svensson

New problem on horizon!?



- Number of hits in Google Scholar for keywords vlsi variability vs. year of publication

Naah. Old news.

- **Analog** design was ever such!
  - Parameter spread of active components:
    - Transistors
    - Amplifiers
    - ...
  - Temperature, aging, …
- Lately, also a problem for **digital** design
  - => Renewed interest!

Dimensions of variability

- Time **independent** vs time **dependent**
  - Time scales
- **Global** variations vs **local** variations
  - Space scales
- **Deterministic** vs **random**

Division of topic

- Monday Nov 12
- Variability and **matching**
  - Transistors, resistors and capacitors
- Thursday Nov 15 (8-10! due to DATE-IT)
  - Variability and **feedback**

Matching - motivation

Digital-to-analog conversion DAC - resistors

Resistors are R and 2R (which can be implemented as R+R)

How to select R?
How large space will resistors take?

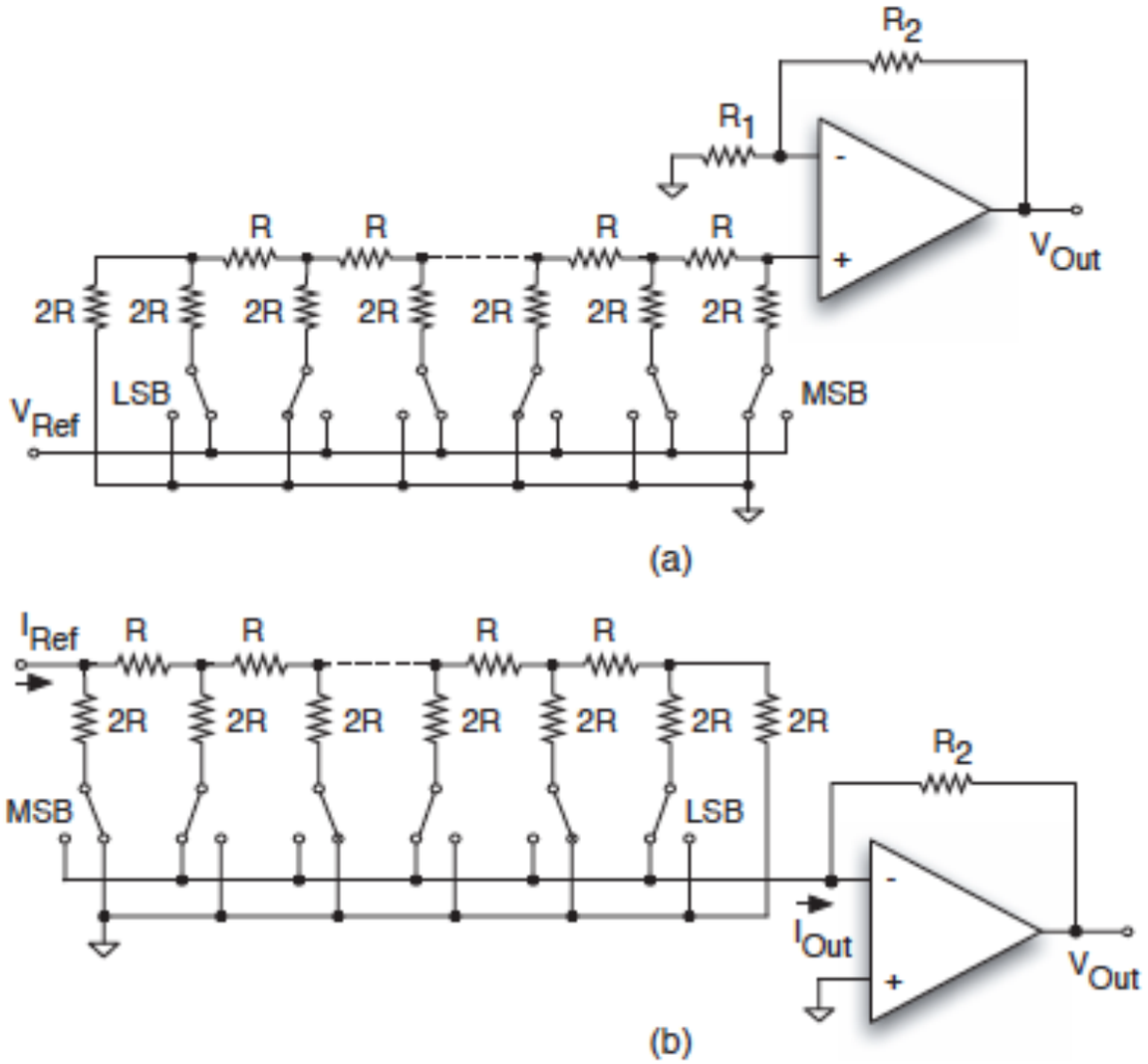


Figure 3.17. (a) Use of the voltage-mode R-2R ladder network. (b) Use of the current-mode R-2R ladder network in an output voltage DAC.

Source: Maloberti Ch. 3

Matching - motivation

Digital-to-analog conversion DAC - capacitors

The spread in C values is due to the number of bits.

How large must the unit capacitor C_u be? And how much space will capacitors take?

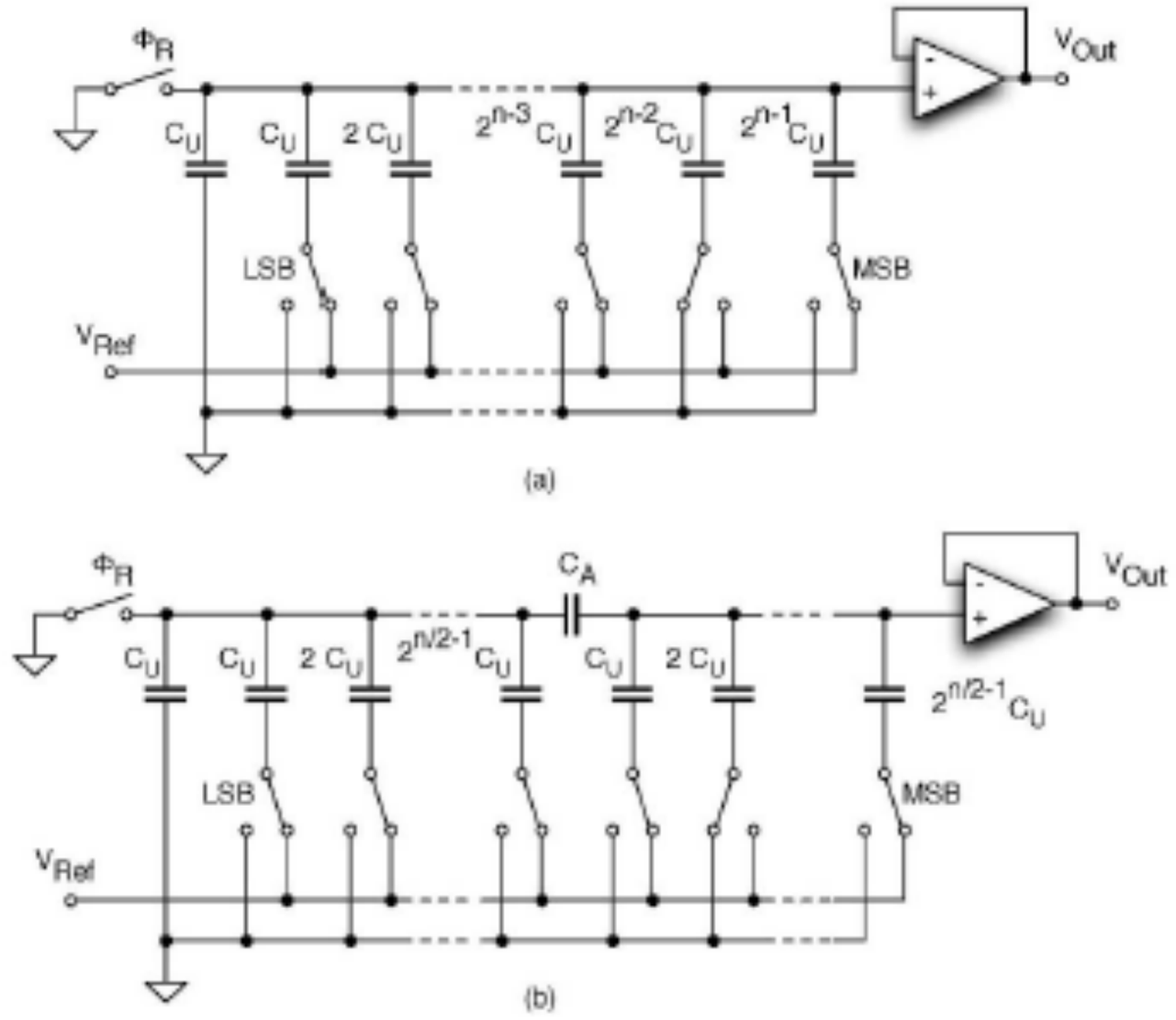


Figure 3.28 (a) n-bit Capacitive divider DAC. (b) The use of an attenuator in the middle of the array reduces the capacitance spread.

Source: Maloberti Ch. 3

Matching - motivation

Digital-to-analog conversion DAC - current sources

All current sources identical - how large must I_u be?

The MOS transistor can be used as a current source.

How large (physically) must the transistors be?

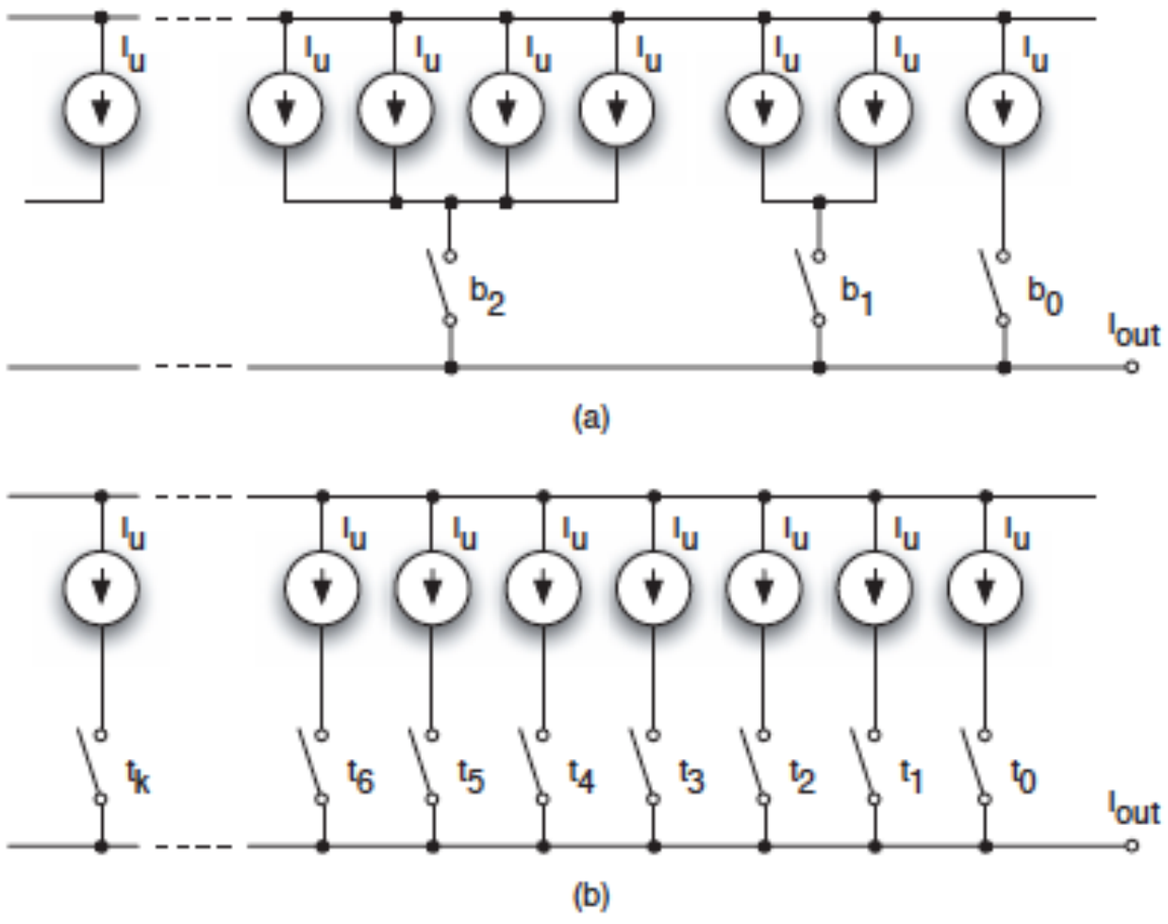


Figure 3.32. (a) Binary weighted control. (b) Unary weighted control.

Source: Maloberti Ch. 3

MOS transistor as current source

- ON or OFF
  - The threshold voltage, V_T, determines state
  - When ON how much the current you get is determined by this equation:

$$I_{DS} = \frac{\beta}{2}(V_{GS} - V_T)^2$$

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

Time scales

Max current CPU clock frequency: 4 GHz => period: 250 ps

fabrication:
lithography
dopants
process corners

power supplies
temperature
LF noise (1/f)

IR drop
temperature
gradients

IR drop
clock jitter
HF noise
(thermal)

static

seconds = 10^{0}

billions of clock cycles

microseconds = 10^{-3}

millions of clock cycles

250 ps

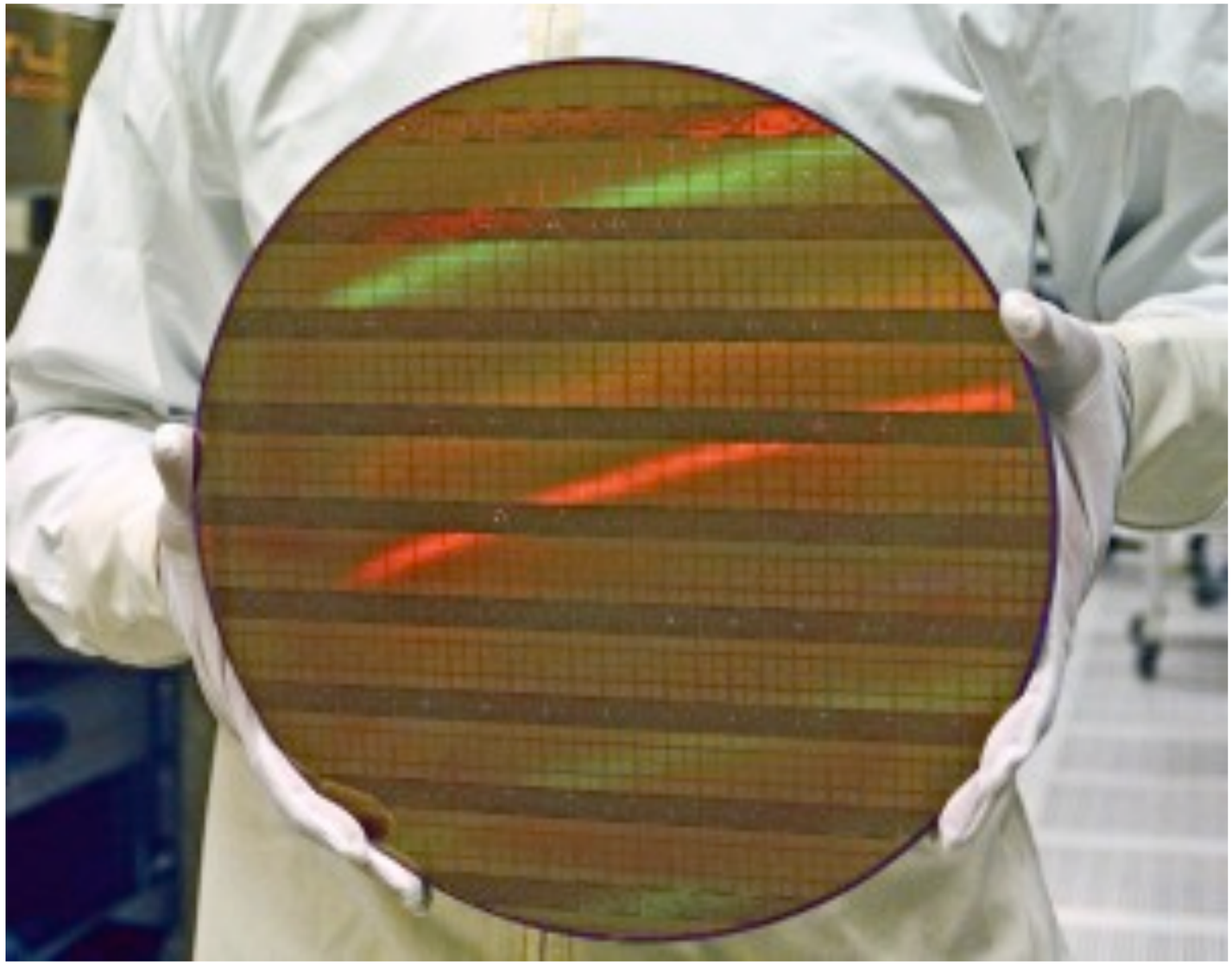
picoseconds = 10^{-12}

less than one clock cycle

Static variability

- Two devices, nominally identical, come out differently
  - Random + predictable
  - Note: more insights move more of variations from r to p
  - What is knowable in principle may still be efficiently handled by statistical methods

Chips are fabricated on wafers

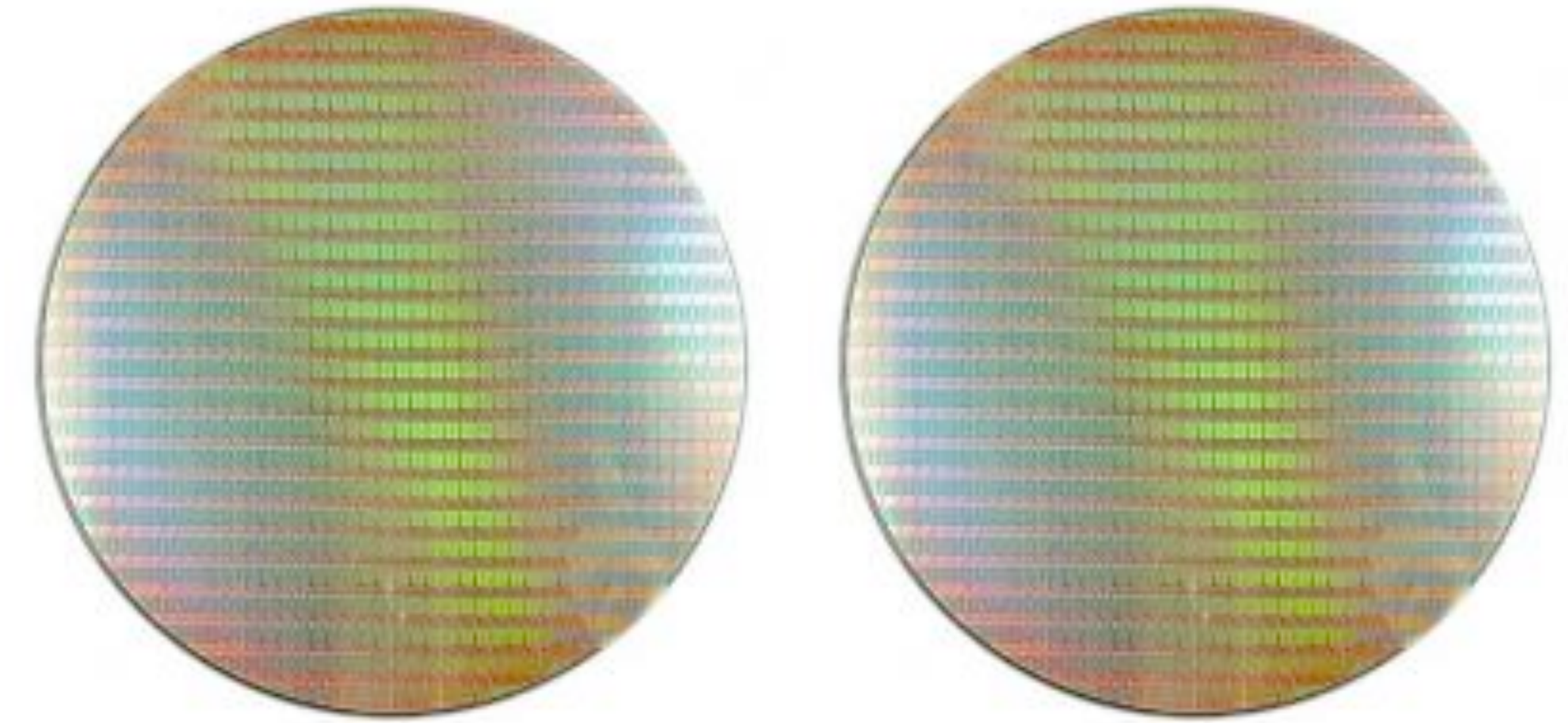


Current Intel 300-mm wafer (photo from Intel)

Source 1

- 2 “boats” (wafer batches) will differ
  - Equipment may have changed

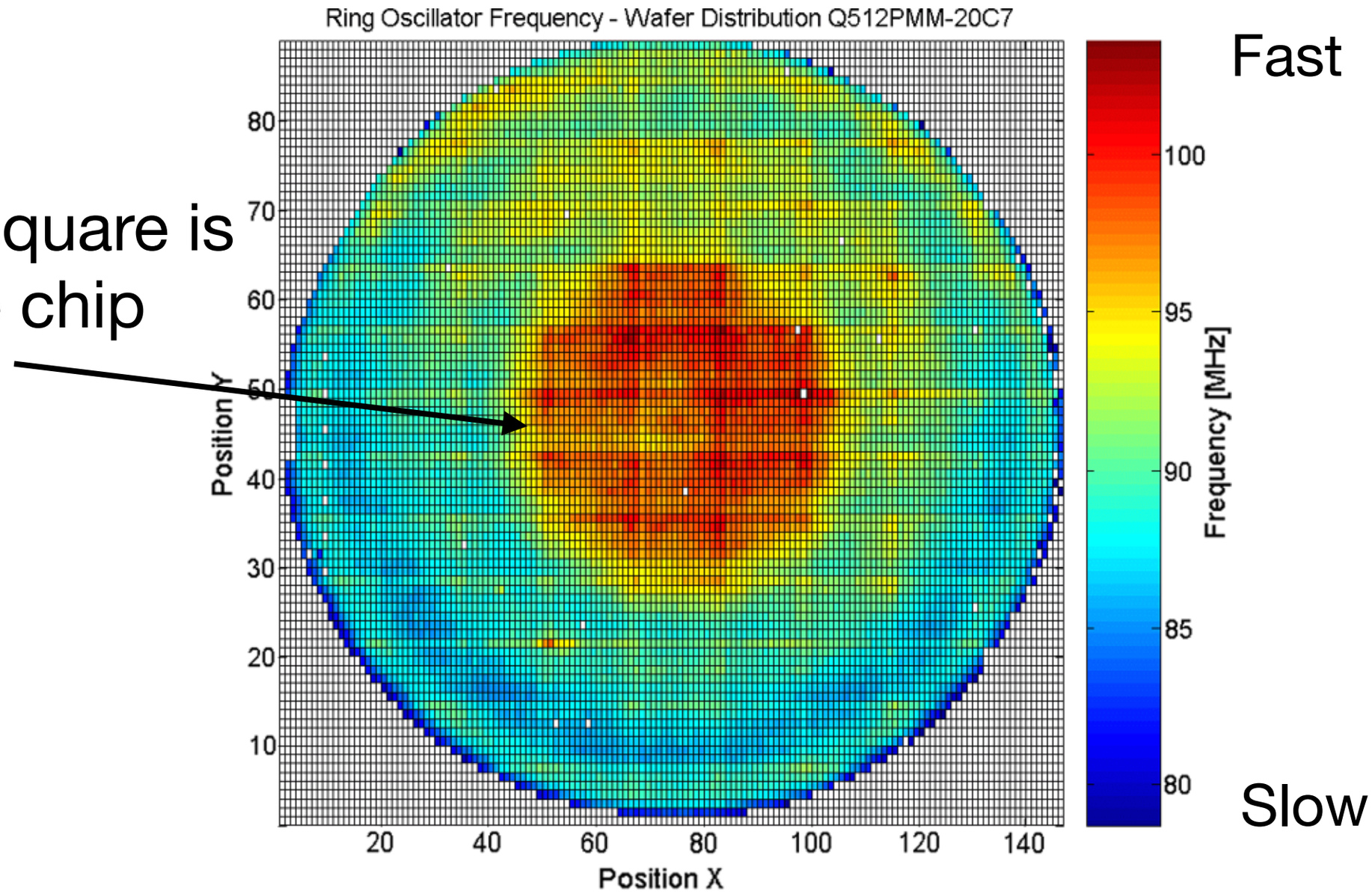
Source 2



- Wafers from same boat will differ
  - Slight differences in processing steps

Source 3 Physical location on wafer

Each square is one chip

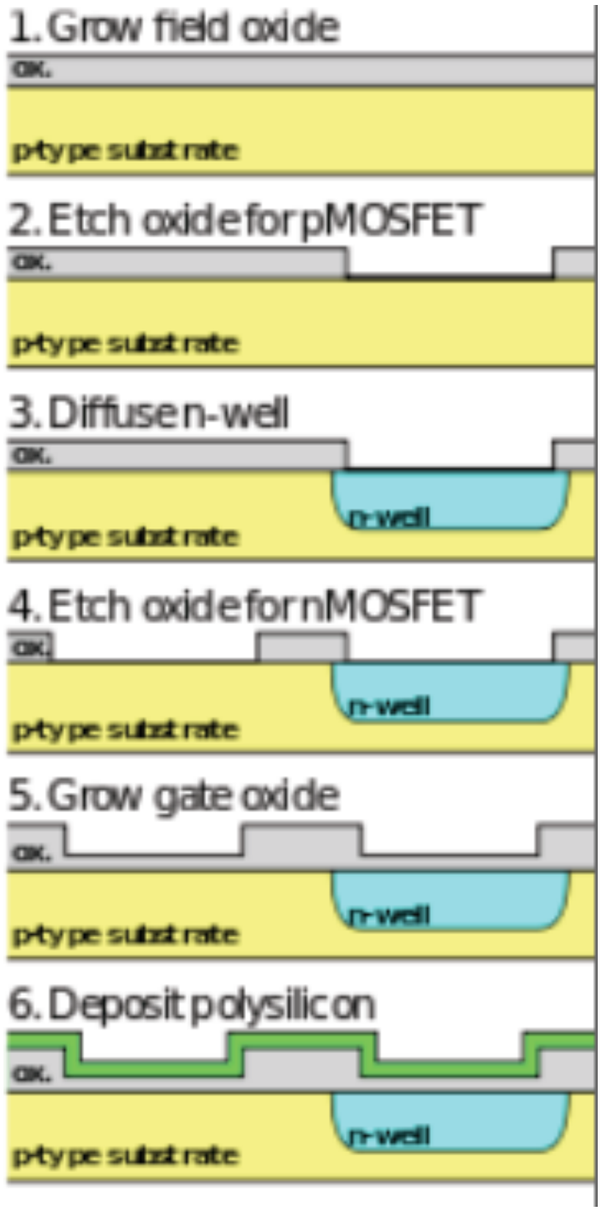


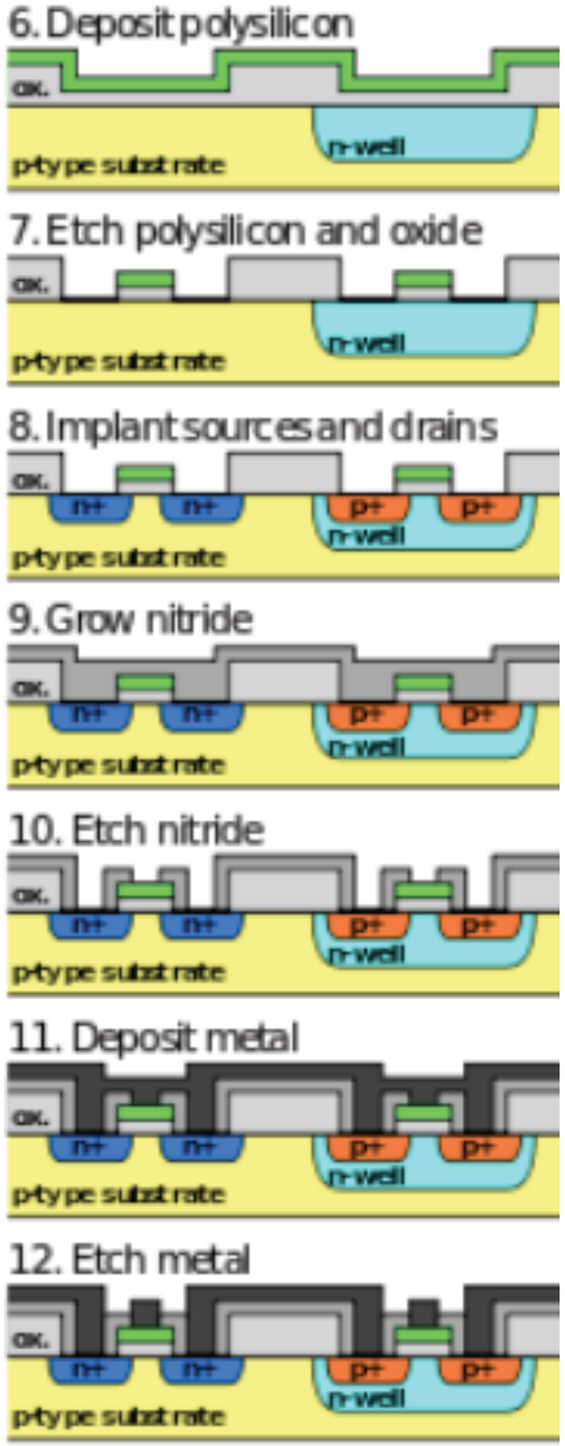
- Chips from same wafer will differ

Source: Pelgrom: Analog-to-Digital Conversion 2010. Springer. Fig. 11.4

Source 4: CMOS fabrication steps

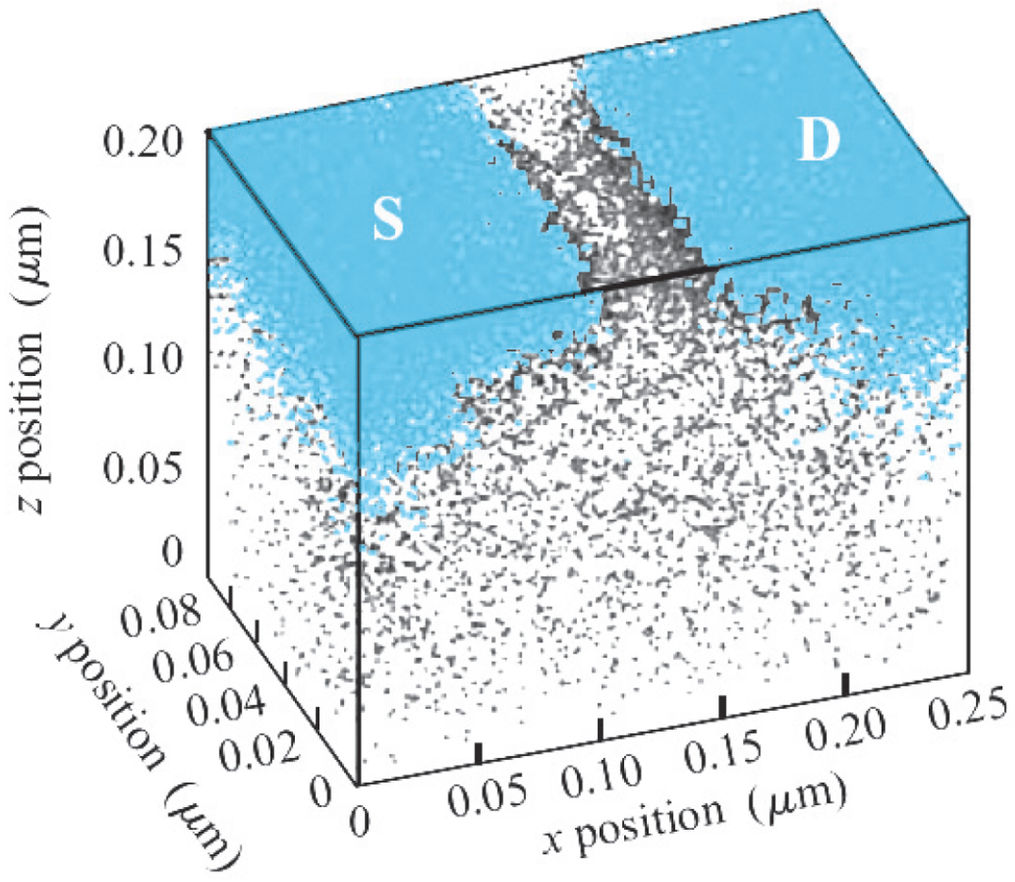
Grow
Etch
Diffuse
Etch
Grow
Deposit
Etch
Implant
Grow
Etch
Deposit
Etch
...





Figures: Wikipedia

Source 5



- Pure random variations
  - i.e. what is still unpredictable …

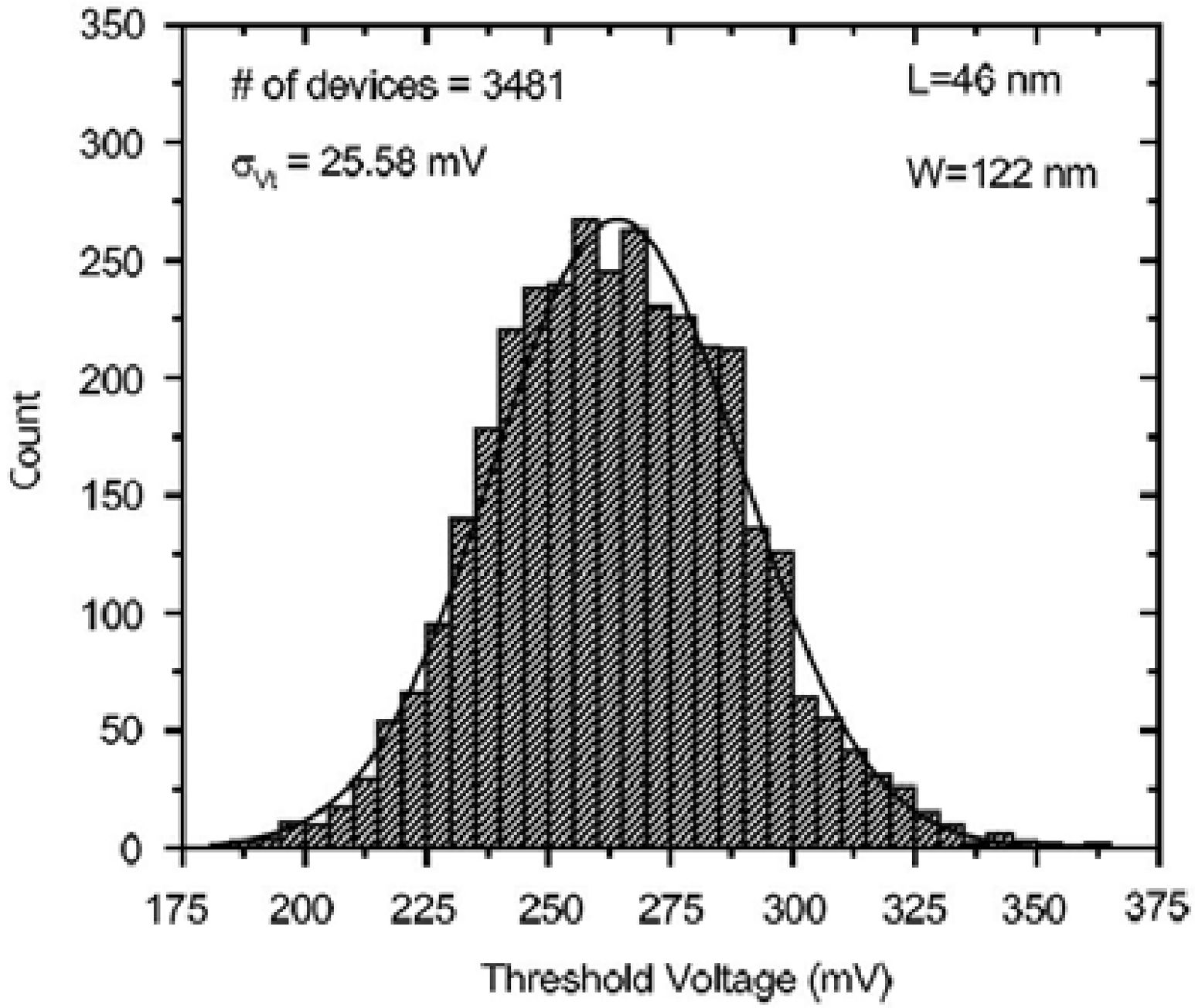
How handle?

- Designs should depend on **ratios**, not on absolute values
  - Usually means selecting different circuit topology.
- **Matching rules:**
  - Use identical sizes
  - Minimize predictable/controllable differences
  - Increase device size to reduce random differences

How large should they be?


IEEE GUSTAV ROBERT KIRCHHOFF AWARD
2017 IEEE Gustav Robert Kirchhoff Award
The Pelgrom Law or "Pelgrom Model"
Marcel J.M. Pelgrom
IEEE.tv
00:01
28:47
HD

Matching example V_T



Parameter fluctuation (1)

- Parameter *P* describes physical property of device
  - Deterministic + random function
  - Different values at different locations on wafer (x,y): *P*(x,y)

Parameter fluctuation (2)

Describe the **difference** in *P* between two points, *ΔP*, mathematically:

$$\Delta P(x_{12}, y_{12}) = P(x_1, y_1) - P(x_2, x_2)$$

which can also be written as:

$$\Delta P(x_{12}, y_{12}) = \frac{1}{\text{area}} \left[\iint_{\text{area}(x_1, y_1)} P(x', y') dx' dy' - \iint_{\text{area}(x_2, y_2)} P(x', y') dx' dy' \right]$$

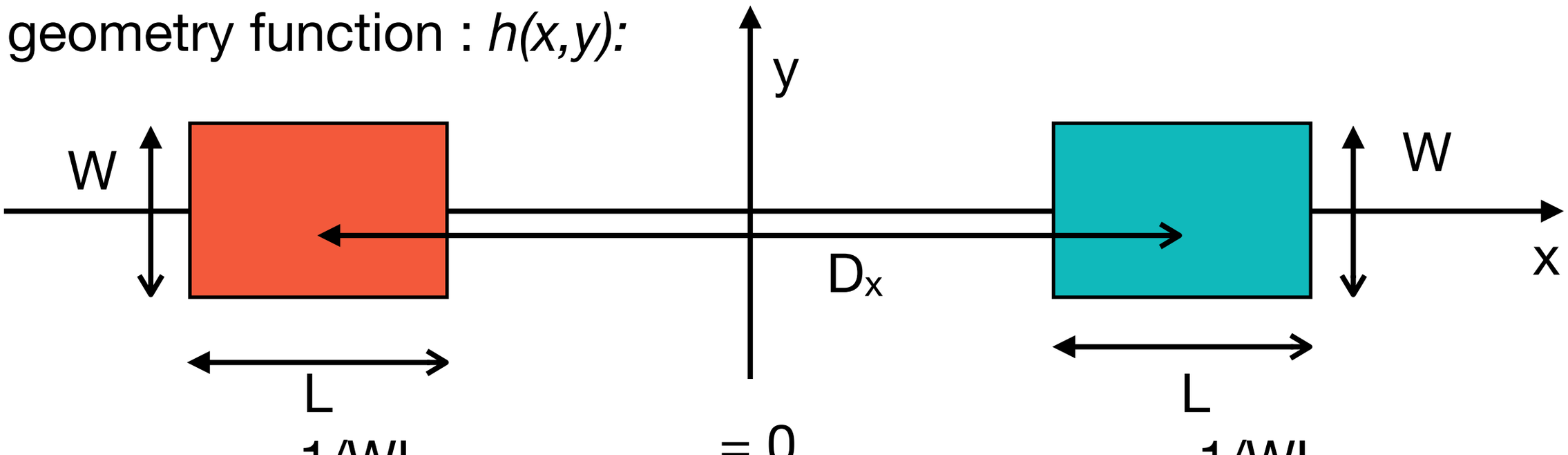
In the Fourier domain this convolution becomes a multiplication:

$$\Delta \mathcal{P}(\omega_x, \omega_y) = \mathcal{G}(\omega_x, \omega_y) \mathcal{P}(\omega_x, \omega_y)$$

Can analyze **geometry** and **mismatch generating source** separately!

Geometry function in x,y plane

geometry function : *h(x,y):*



= -1/WL
for all (x,y)
within red box

= 0
everywhere
else

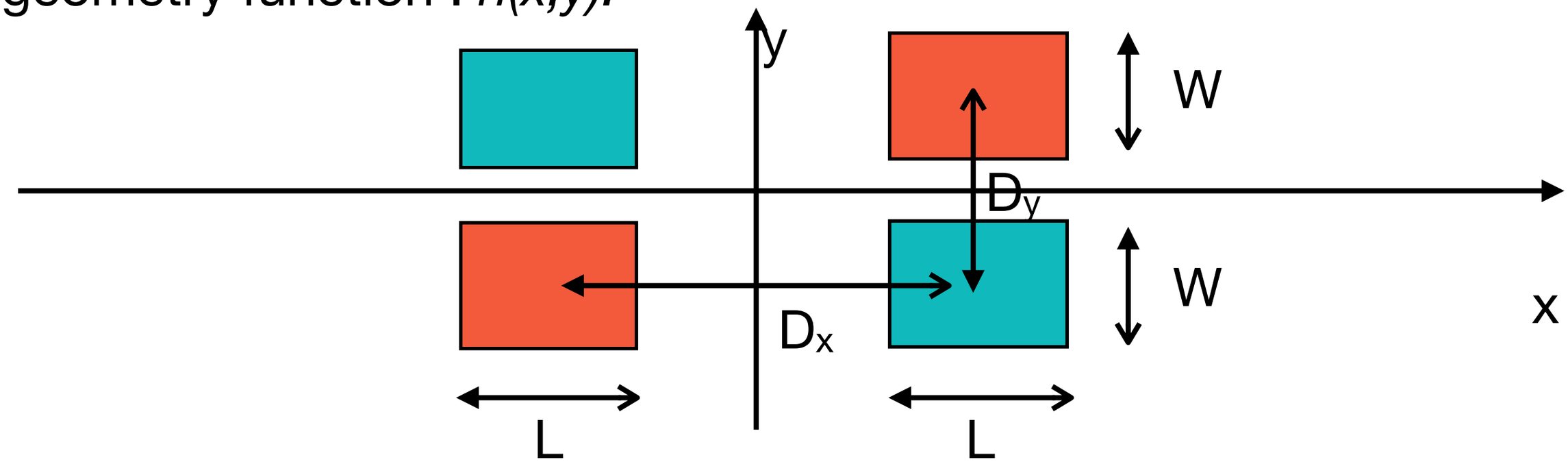
= 1/WL
for all (x,y)
within blue box

Fourier transform (after some manipulation):

$$\mathcal{G}(\omega_x, \omega_y) = \frac{\sin(\omega_x \frac{L}{2})}{\omega_x \frac{L}{2}} \frac{\sin(\omega_y \frac{W}{2})}{\omega_y \frac{W}{2}} \left[2 \sin(\omega_x \frac{D_x}{2}) \right]$$

Another geometry function in x,y

geometry function : *h(x,y):*



= -1/WL
for all (x,y)
within red boxes

= 0
everywhere
else

= 1/WL
for all (x,y)
within blue boxes

Fourier transform (after some more manipulations):

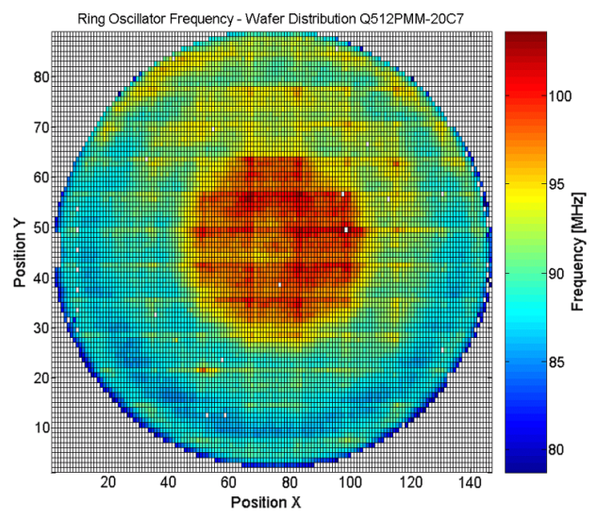
$$\mathcal{G}(\omega_x, \omega_y) = \frac{\sin(\omega_x \frac{L}{2})}{\omega_x \frac{L}{2}} \frac{\sin(\omega_y \frac{W}{2})}{\omega_y \frac{W}{2}} \left[\cos(\omega_x \frac{D_x}{2}) - \cos(\omega_y \frac{D_y}{2}) \right]$$

Mismatch generation source: *P*(x,y)

- *P*(x,y) of “class 1”:
  - Total mismatch of parameter *P* is due to mutually independent events
  - Effects on the parameter are so small that contributions are linear
    - The correlation distance between events is small compared to device size
- Result is a Possion process which converges to Gaussian with 0 mean.
- Corresponds to spatial “white noise”. Characterized by **one constant** for all spatial frequencies.

Mismatch generation source *P*(x,y)

- *P*(x,y) of “class 2”:
  - Total mismatch of parameter *P* is due to mutually independent events
  - Effects on the parameter are so small that contributions are linear
  - The correlation distance between events is **large** compared to device size



Finding the power (= variance) of Δ*P*

Sum the power contributions =
Integrate the squares over all spatial frequencies:

$$\sigma^2(\Delta P) = \frac{1}{4\pi^2} \int_{\omega_y=-\infty}^{\infty} \int_{\omega_x=-\infty}^{\infty} |\mathcal{G}(\omega_x, \omega_y)|^2 \, |\mathcal{P}(\omega_x, \omega_y)|^2 \, d\omega_x d\omega_y$$

With two rectangular devices and
a mismatch generation source of class 1 we get:

$$\sigma^2_{\Delta P} = \frac{A_P^2}{WL}$$

Here A_P is the proportionality constant for parameter Δ*P*

Finding the power (= variance) of ΔP

With two rectangular devices and
a mismatch generation source of class 2
(large correlation distance)
we get:

$$\sigma^2_{\Delta P} = \frac{A^2_P}{WL} + S^2_P D^2_X$$

Here S_P describes the variation of parameter P with the spacing

Average (relative) or absolute quantity?

This relation holds for *averaged* or *relative* values of parameter *P* (for example threshold voltage of MOS transistor)*:*

$$\sigma^2_{\Delta P} = \frac{A^2_P}{WL}$$

Absolute number of events are proportional to WL
(for example number of charges in MOS transistor channel).
The relation is then:

$$\sigma^2_{\Delta P} = A^2_P WL$$

Conclusion of derivation

The result in short:

Variance

$$\sigma^2_{\Delta P} = \frac{A_P^2}{WL}$$

Standard deviation

$$\sigma_{\Delta P} = \frac{A_P}{\sqrt{WL}}$$

Holds under what conditions? (3)
(Added after lecture: see slide 25)

For what types of parameters? (2)
(Added after lecture: see slide 29)

MOS transistor matching

$$I_{DS} = \frac{\beta}{2}(V_{GS} - V_T)^2$$

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

	Variance	Standard deviation
Threshold voltage matching - absolute (since averaged quantity)	$\sigma^2_{\Delta VT} = \frac{A^2_{VT}}{WL}$	$\sigma_{\Delta VT} = \frac{A_{VT}}{\sqrt{WL}}$
Current factor matching - relative (since absolute quantity)	$\frac{\sigma^2_{\Delta\beta}}{\beta^2} \approx \frac{A^2_{\beta}}{WL}$	$\frac{\sigma_{\Delta\beta}}{\beta} \approx \frac{A_{\beta}}{\sqrt{WL}}$

Has to analyze contribution from μ, Cox, W and L

Finding proportionality constant from measurements

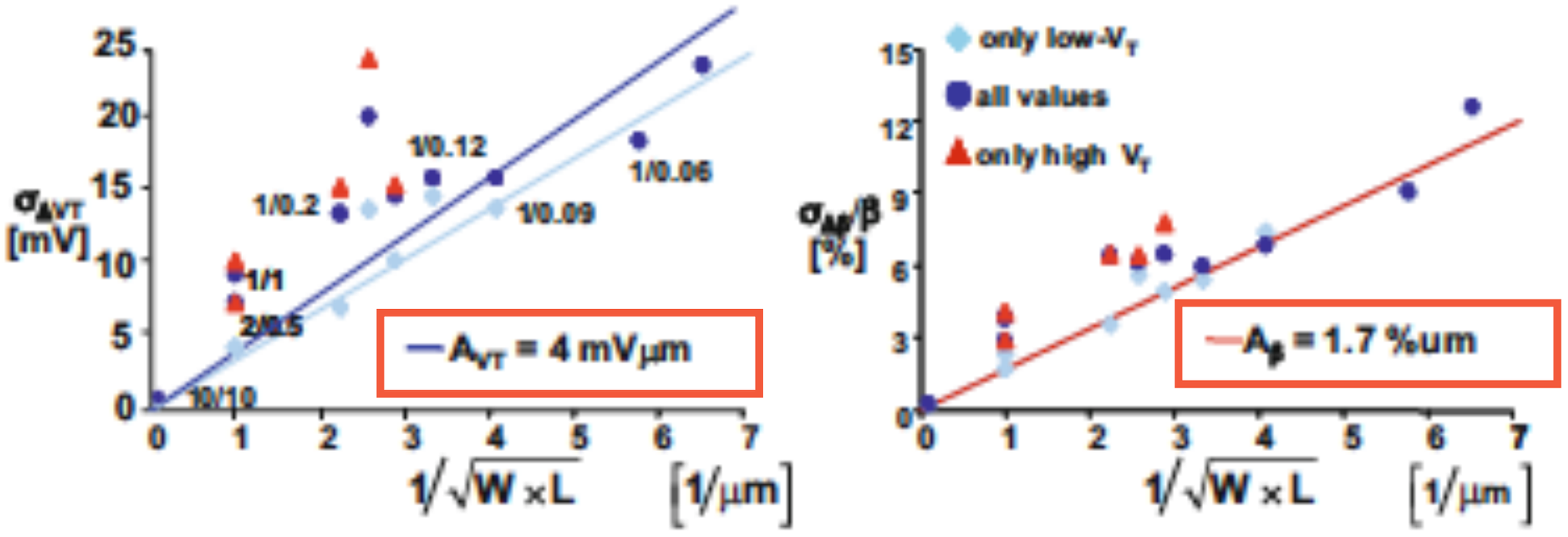
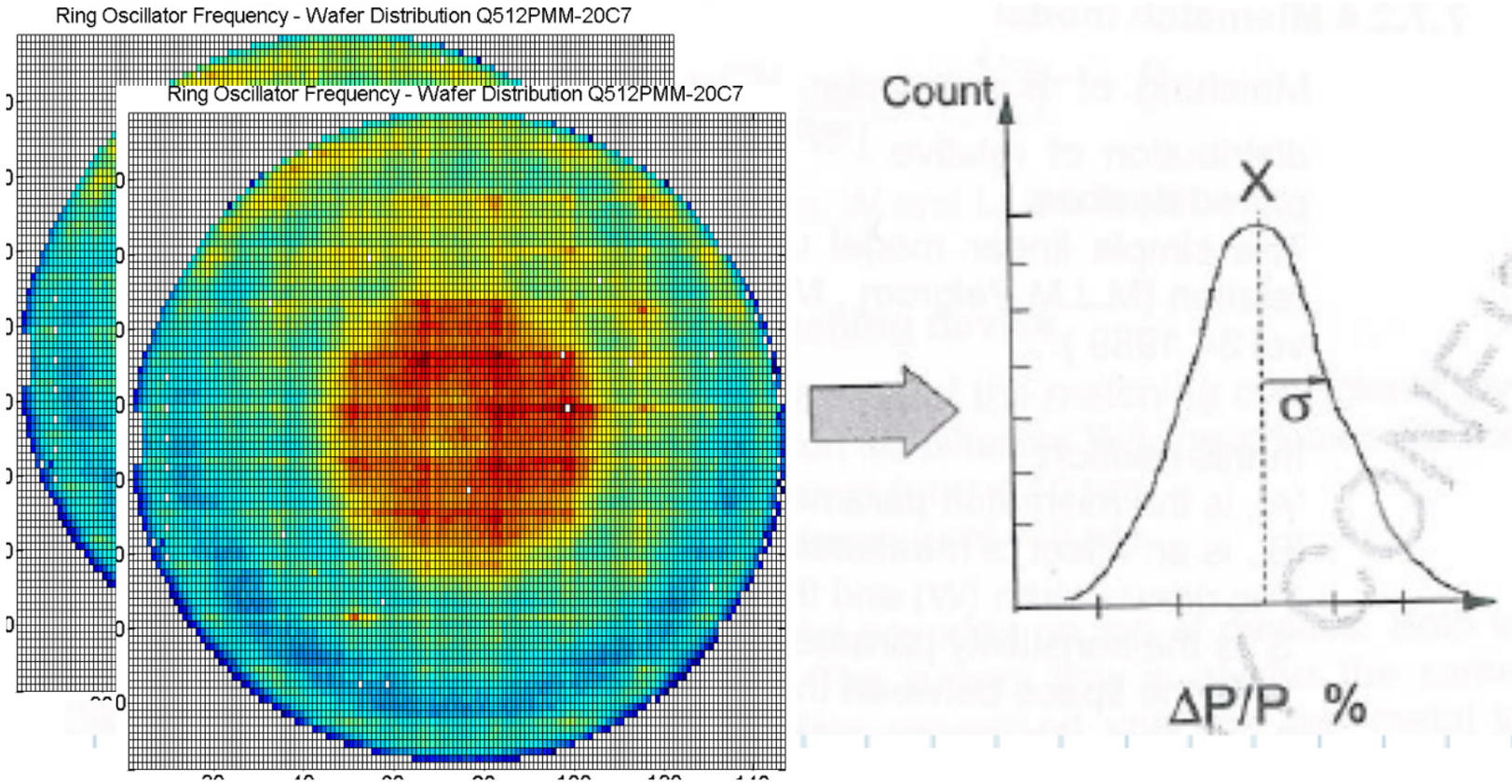


Fig. 5.35 The threshold mismatch and the relative current factor mismatch for NMOS transistors in 65-nm technology. Measurements by N. Wils/H. Tuinhout

Typical values for a 65-nm process

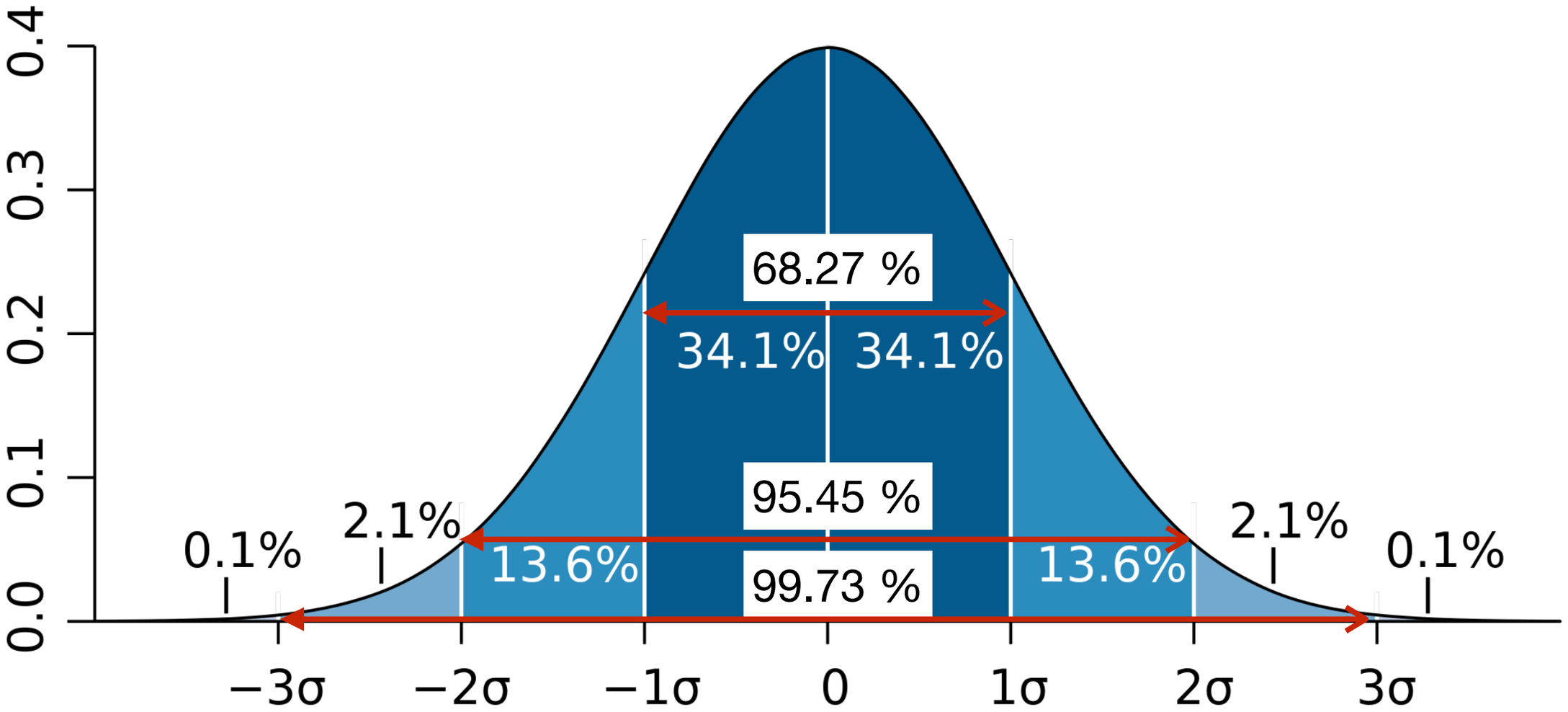
From Pelgrom Analog-to-Digital Conversion 2017

From measurements to statistics



Measured mismatch translated into statistical model
Assumption: Gaussian distribution

Translation continued



The percentage of chips that are within the limit af acceptance, depending on the number of standard deviations the variations are kept within.

Source: Wikipedia

An example

- We want to match V_T of two nMOS transistors in our 65 nm process:
  - The difference in V_T should be at most 10 mV and we know that $A_{V_T} = 5.4\,[mV\,\mu m]$

Variance

- How wide do we have to make them?
  - Assume the length L = 1 μm
  - Assume that we do all the layout perfect.
  - Assume that staying within 3σ is “enough”.

Solution to example

- We need 10 mV max deviation in threshold voltage, V_T, for matched transistor pair with lengths L = 1 μm. Determine required width, W for transistors.

- Translation: 3σ = 10 mV gives us 99.7 % within 10 mV so σ=10/3= 3.3 mV
- Use equation $\sigma^2_{\Delta VT} = \frac{A^2_{VT}}{WL}$
- 3.33^2 [mV^2] = 5.4^2 [$mV^2\mu m^2$] /(W*1[μm])
- W = 5.4^2 [$mV^2\mu m^2$]/(3.33^2 [mV^2]*1[μm]) = **2.7 μm**

MOS transistor as current source revisited

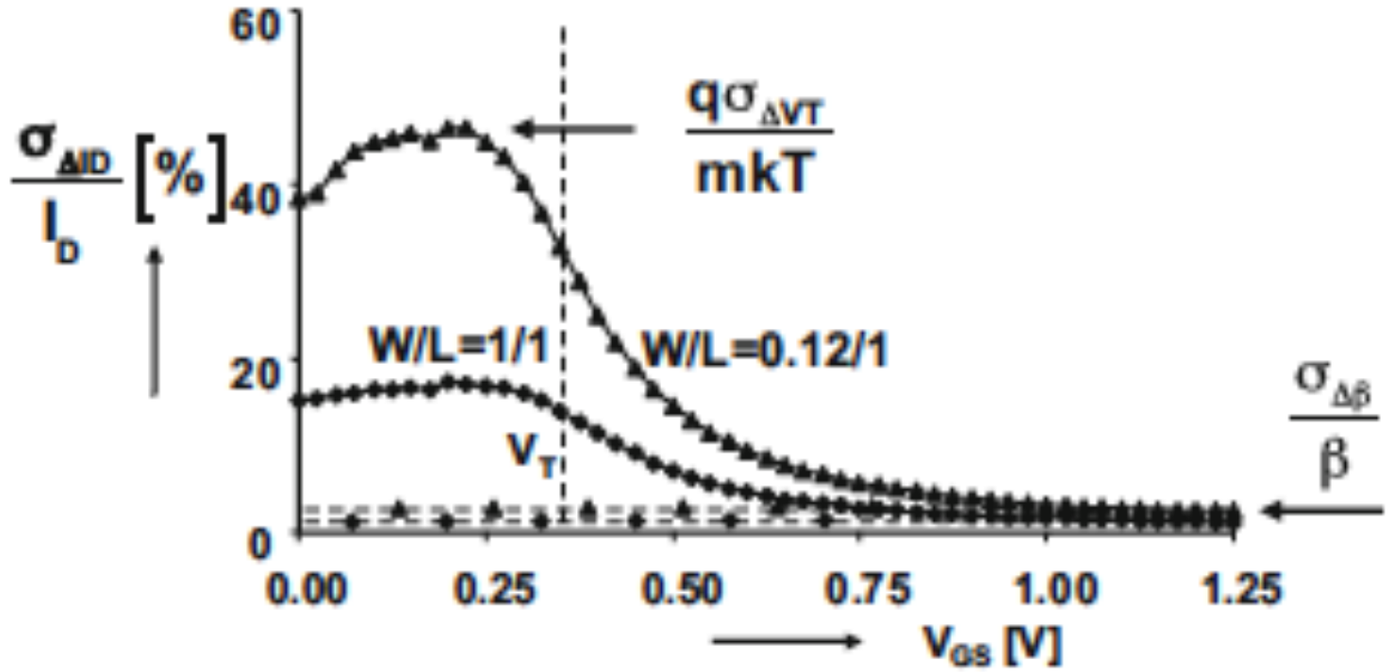


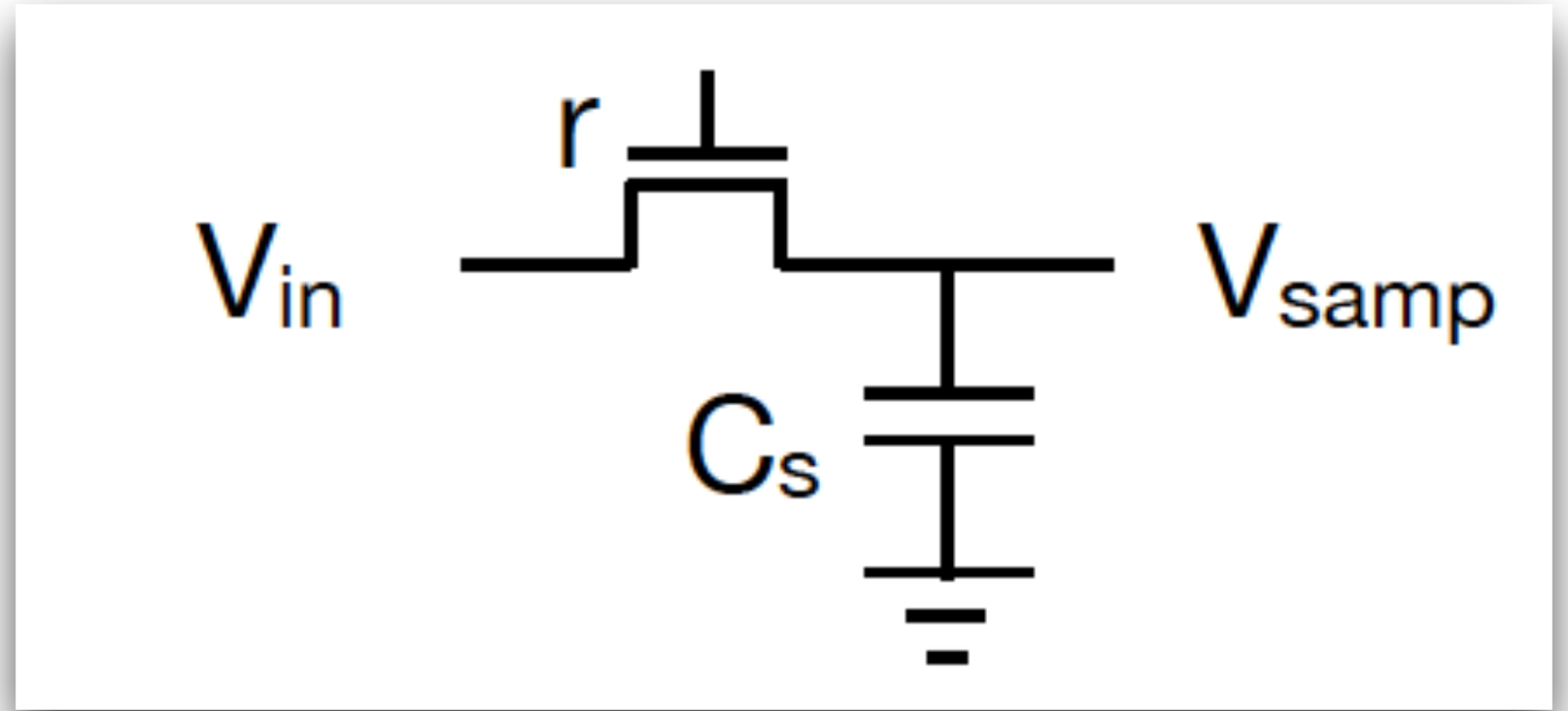
Fig. 5.36 The relative current mismatch for two 65-nm technology transistor geometries swept over the full voltage range. Measurements by N. Wils/H. Tuinhout

Note that one reason that the relative mismatch is lower around and below VT is that the current is much lower!

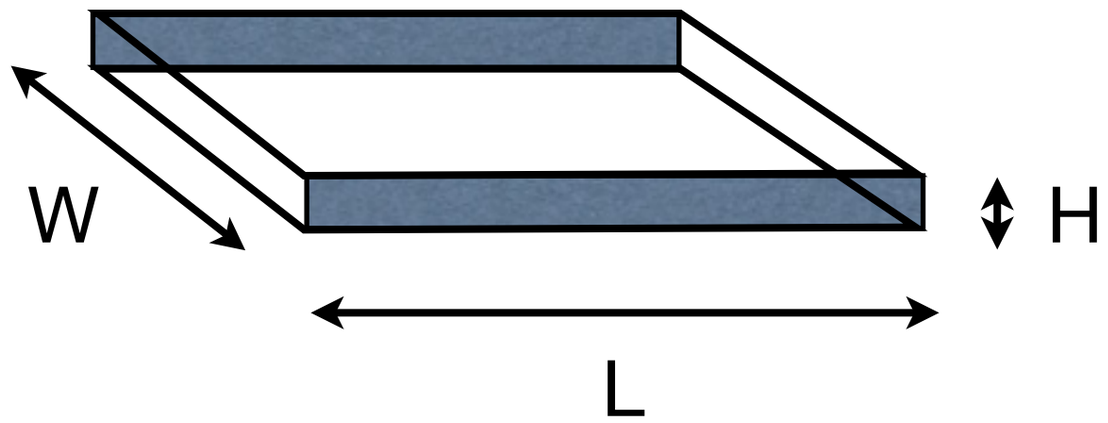
Passive components

- Better matching than for transistors
  - When setting gain!
    - => Feedback and variability next time.
- Also used in DAC and ADCs
  - R2R or C2C ladders, for example.

From last week - sampling



On-chip resistor (ideal)



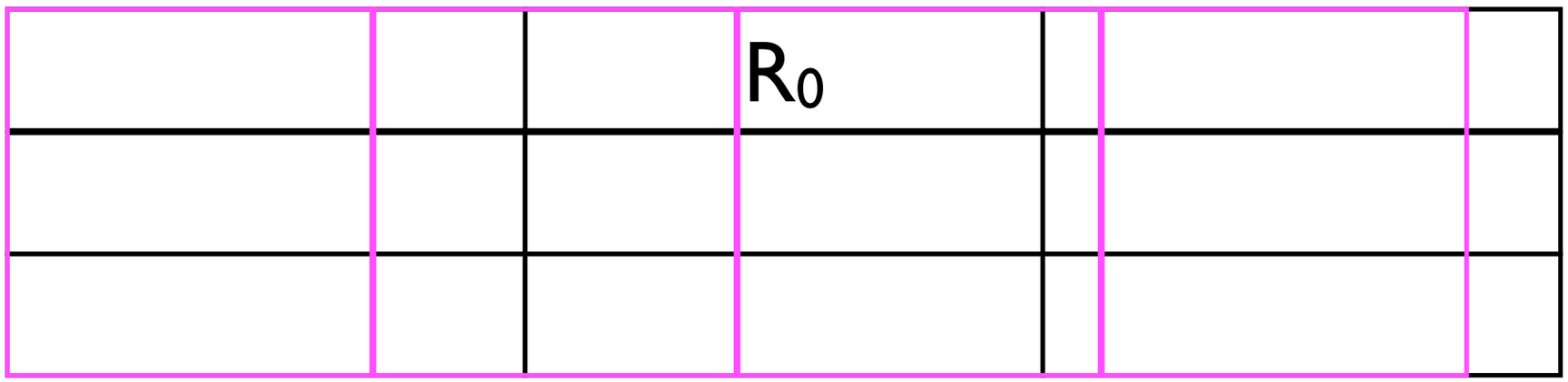
- Rectangular slab of some material

$$R = r \cdot L / (W \cdot H)$$

- Height given by process
- Value determined by length and width

Top view

direction of current



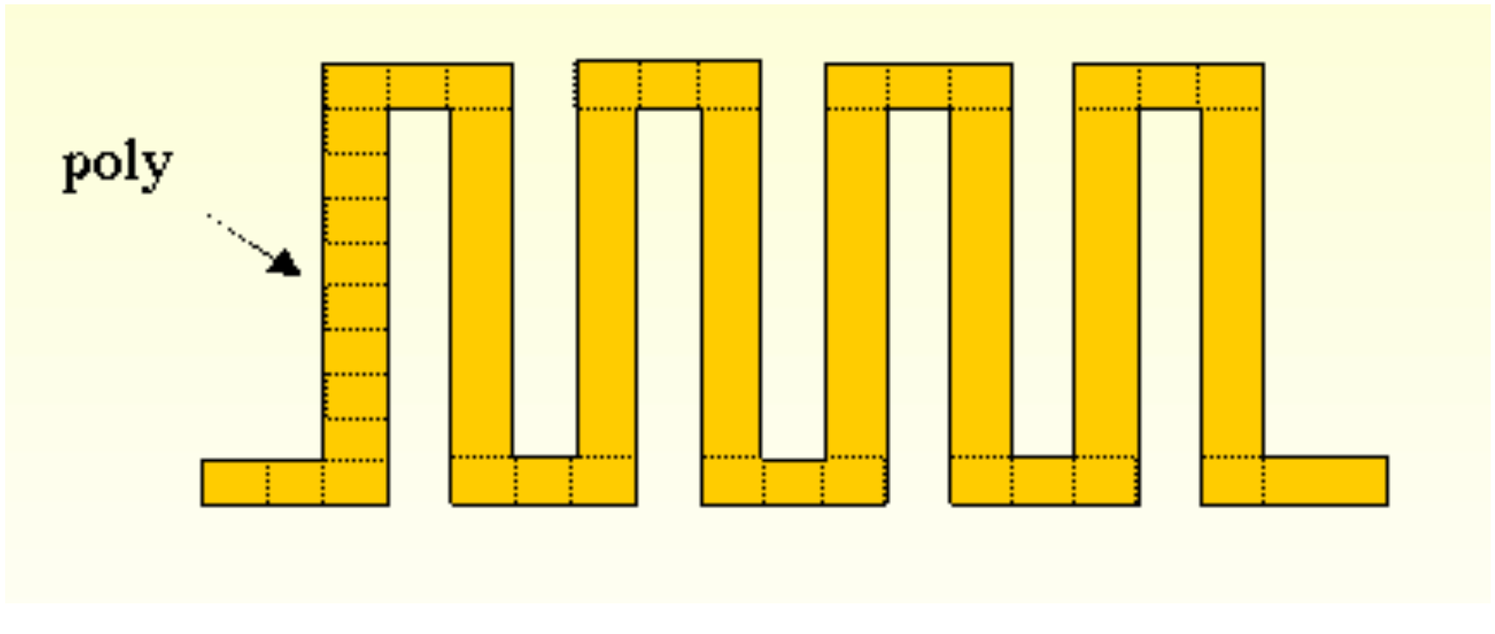
- Nominal resistance set by aspect ratio, L / W
- $R = r \cdot L / (W \cdot H) = (r / H) \cdot (L / W)$

resistivity per square

Small values?

- Inaccuracies at ends

Large values?



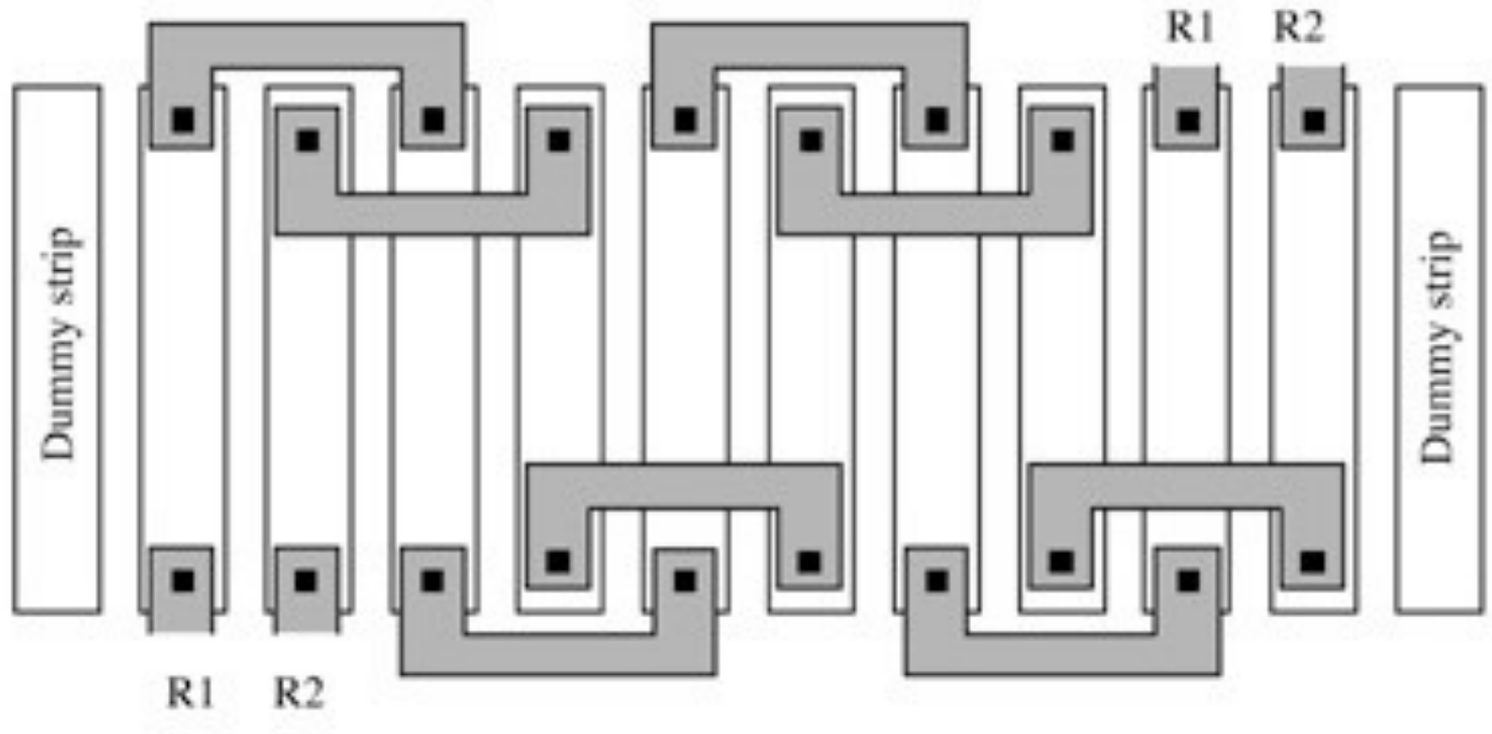
- Need large L / W
- Minimum W set by process design rules
- L is “only” parameter!
- Long and narrow…

Corners count for 0.5 square

Accuracy?

- $R = (r / H) \cdot (L / W)$
- r: material property
- H: set by process for each layer
- Geometrical inaccuracies affect mostly W
  - Predictable error + random variations
  - Large W is better (at an area cost…)

Relative accuracy?



- Affected by environment
- For close matching, strive for identical environments!

Resistance values?

- 65 nm process (typical values):
  - P+ diff: R_{SH} = 244 Ω/□
  - N+ diff: R_{SH} = 130 Ω/□
  - P+ poly: R_{SH} = 712 Ω/□
  - N+ poly: R_{SH} = 180 Ω/□
  - High-res P+ poly: R_{SH} = 6 kΩ/□

High resistance requires special layer

Resistance matching

$$\frac{\sigma^2_{\Delta R}}{R^2} \approx \frac{A^2_R}{WL} \quad \text{or} \quad \frac{\sigma_{\Delta R}}{R} \approx \frac{A_R}{\sqrt{WL}}$$

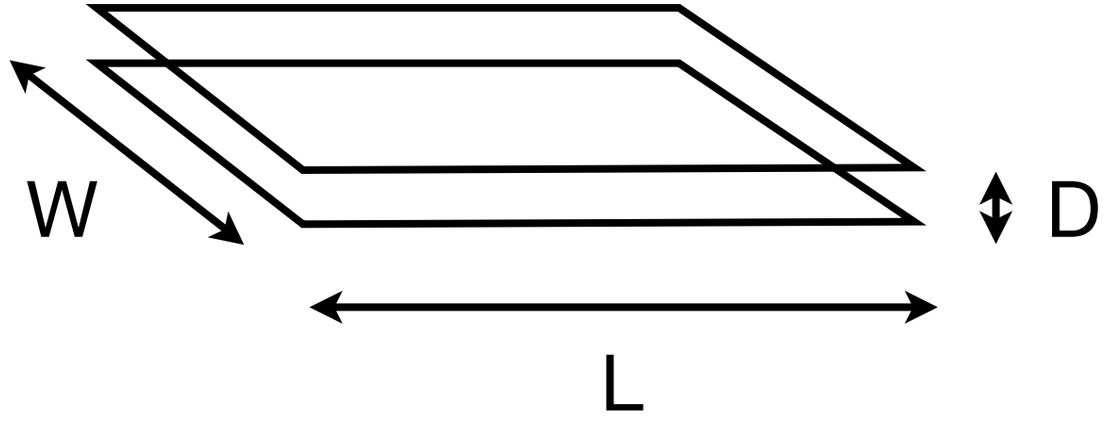
For diffused resistors typical values:

$$A_R = 0.5\ \%\ \mu m$$

For poly resistors typical values:

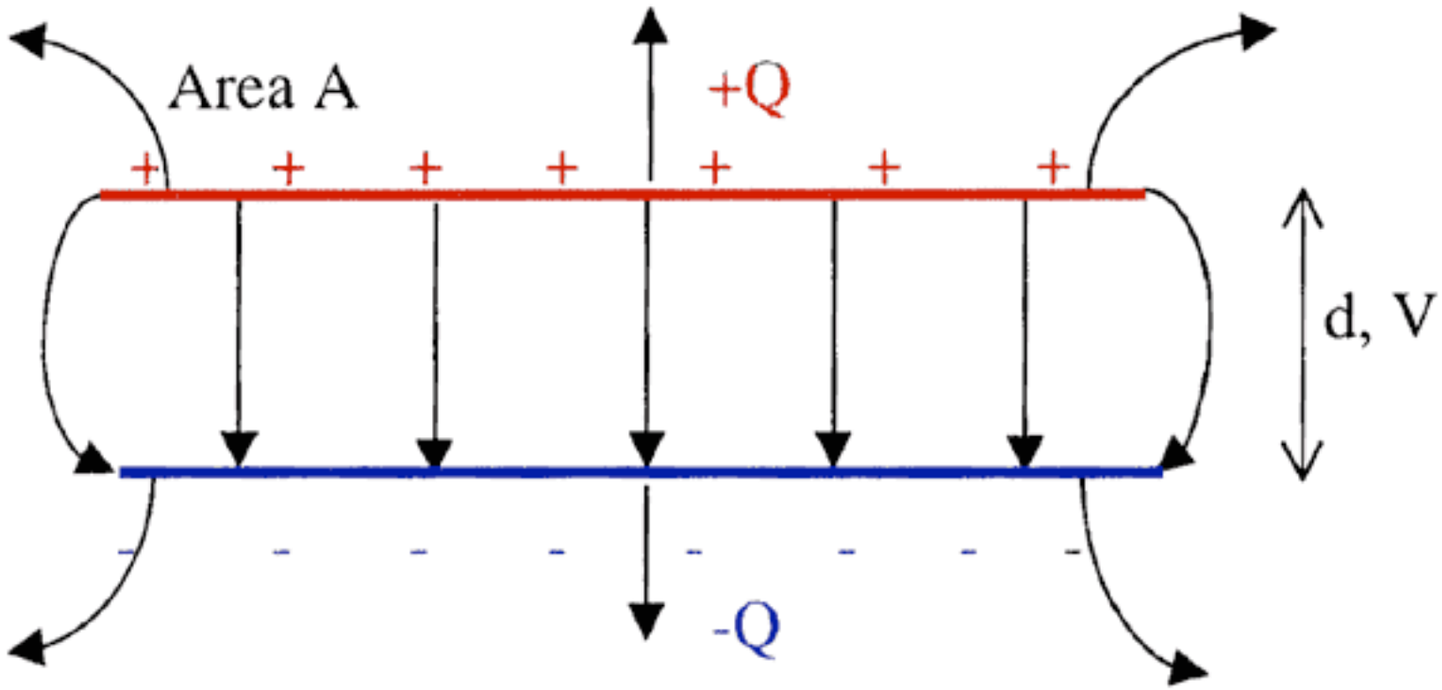
$$A_R = 5\ \%\ \mu m$$

On-chip capacitors



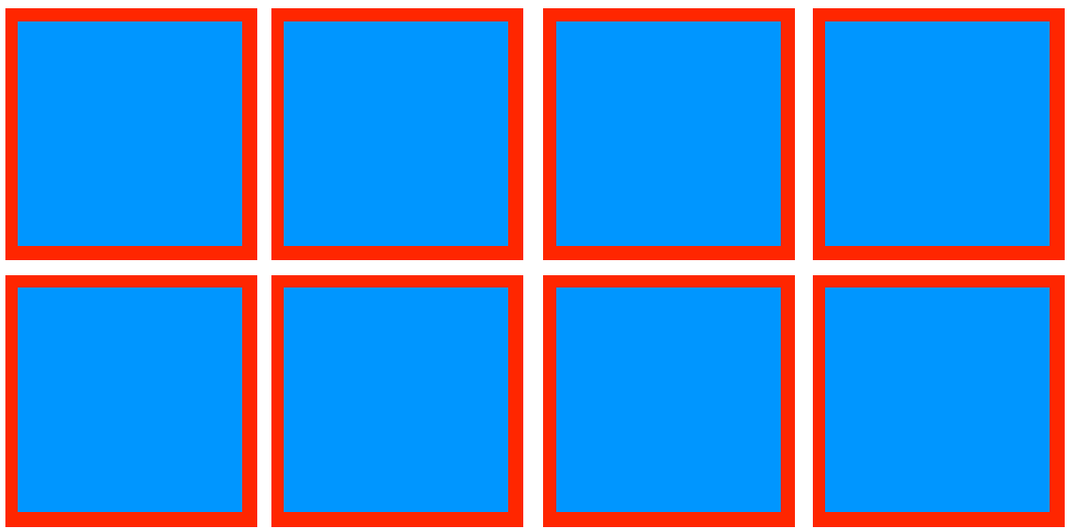
- Two conductor sheets
- One isolator sheet
- D given by process; W, L set by designer
- $C = c \cdot W \cdot L / D = (c / D) \cdot (WL)$ Area

Accuracy?



- Simple equation assumes uniform field
  - Edge fields unaccounted for!
  - Better absolute accuracy: use less edge
  - Better relative accuracy: use constant edge per area

Best relative accuracy



- Repeat unit capacitance
- Dummies around edges

Capacitance values?

- 65 nm process: no extra layer for capacitors
  - Fringe capacitor: 1.6 fF/μm^2
  - Striped stacked M1-M5 capacitor: 0.75 fF/μm^2

Capacitance matching

$$\frac{\sigma^2_{\Delta C}}{C^2} = \frac{A^2_C}{WL}$$

Assuming capacitors large enough that area effects dominate

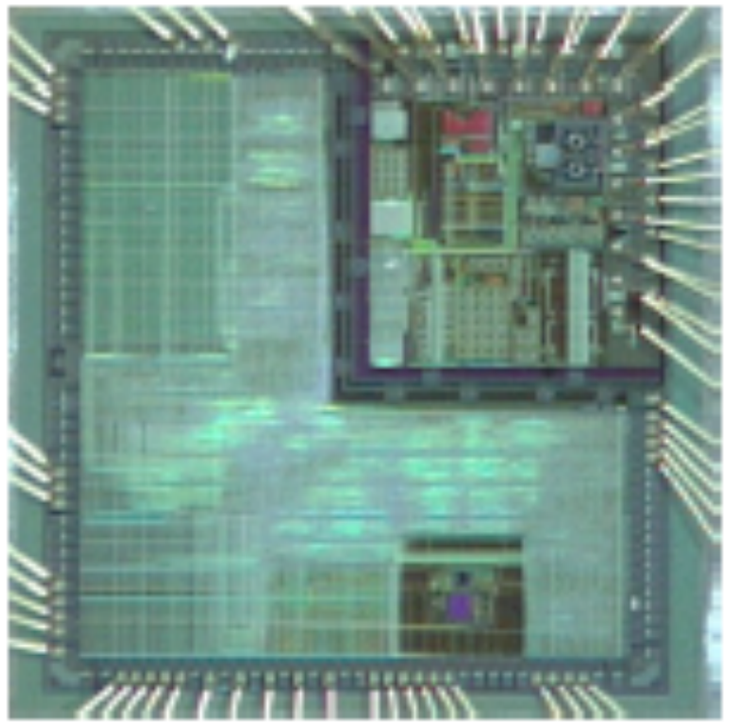
But since the capacitance is always proportional to the capacitor area we can also express the matching as

$$\frac{\sigma^2_{\Delta C}}{C^2} = \frac{A^2_C}{C} \quad \text{that is} \quad \frac{\sigma_{\Delta C}}{C} = \frac{A_C}{\sqrt{C}}$$

Typical values: $A_C = 0.3\ \%\sqrt{fF}$

Physical separation in mixed-mode system

- Example:
- Single-chip bluetooth transceiver Ericsson 2001
- 25% is RF part 75% is digital electronics
- Keep them apart!
- Use separate supplies.



Example due to
Sven Mattisson Ericsson research

Summary

- Passives:
- Pick C over R for on-chip use
  - Less bias current, less edge uncertainty
    - Absolute component precision on chip abysmal (±20%, etc)
      - Relative precision not so bad
        - ±1% “easily” attainable

Summary, cont.

- Precision difficult for extreme values
- Large values and high relative precision cost area => $ and also W
- Mixed-mode systems: Separate analog and digital parts on the same chip as much as possible
- Clocked analog systems a problem!

Thursday

- Variability & feedback