UiO : Department of Informatics
University of Oslo

INF4420

Layout and CMOS processing technology

Spring 2012 Jørgen Andreas Michaelsen (jorgenam@ifi.uio.no)

Outline

CMOS Fabrication overview

Design rules

Layout of passive and active componets

Packaging

Introduction

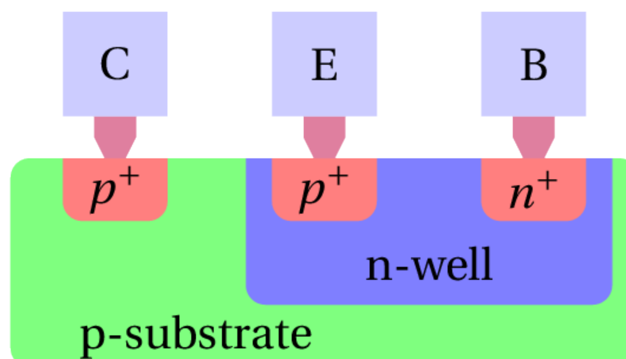
As circuit designers we must carefully consider how to draw layout for critical/sensitive parts of the circuit in order to get robust and predictable performance.

To be sucessfull at this, we must have a basic understanding of how circuits are manufactured, packaged, tested, and even how the circuit eventually is used on a PCB (e.g. external parasitics that we need to drive off chip)

Physical design

The physical circuit is built on a disc of silicon (wafer) layer by layer.

Some layers are implanted in the substrate, other layers are stacked on top.



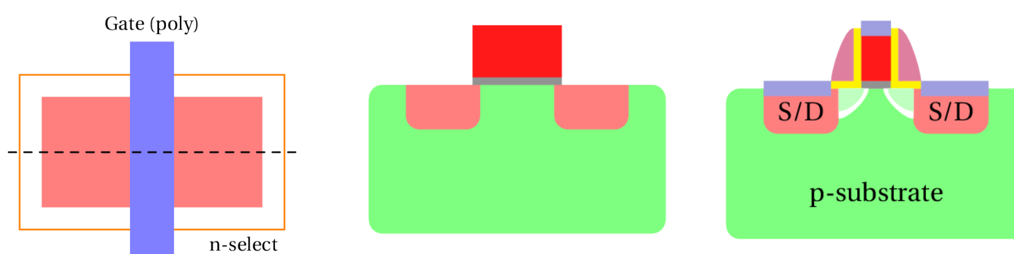
Physical design

How do we go from a layout (GDS2) to a physical circuit?

For each step in the processing, we must get the relevant part of the design onto the wafer, do the processing (implant, etch, or grow), and ready the wafer for the next step.

Physical design

Layout is "encoding" the physical realization of circuits.
CMOS processing (manufacturing) is done in layers, so is layout.



Photolithography

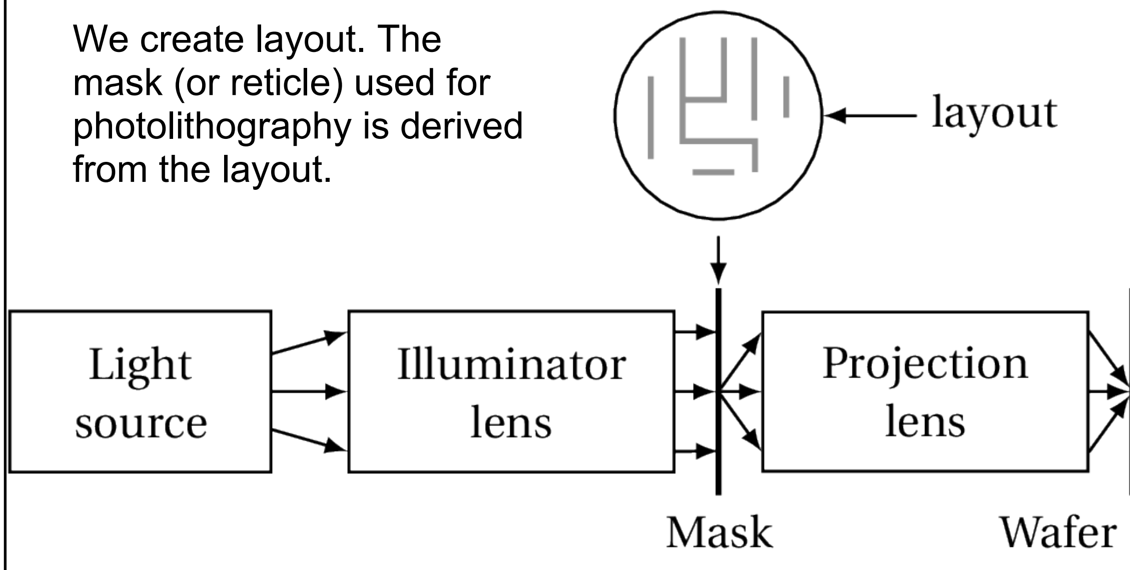
Photolithography (litho) is used to define regions for each layer.

For each processing step, we need to transfer the mask onto the wafer (selectively coat/shield part of the wafer).

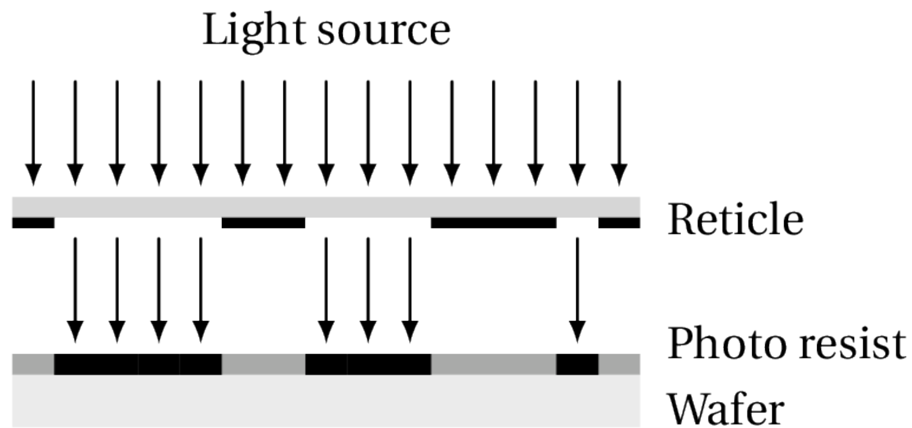
Light source and a mask defines patterns on photoresist. Photoresist hardens when exposed to light.

Lithography system

We create layout. The mask (or reticle) used for photolithography is derived from the layout.



Photoresist



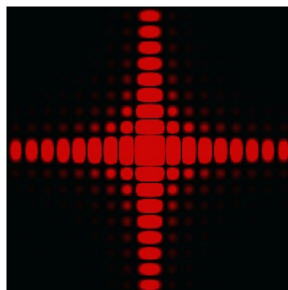
(1) Photoresist hardens when exposed to light (negative photoresist), leaving a developed mask on the wafer. Remaining photoresist is removed. (2) Do processing. (3) After processing step, hardened resist is also removed. Repeat for all processing steps required for full circuit.

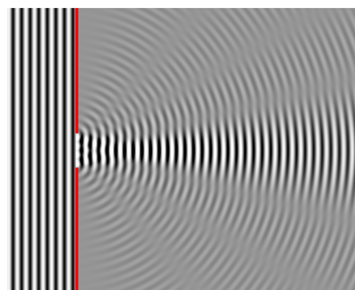
Diffraction

Slits in the reticle cause diffraction (pattern spreads out).

Wavelength of light is a limitation for feature size.

Images illustrating diffraction from Wikipedia. org.





Resolution

Resolution is limited by the wavelength of light and numerical aperture (NA) of the lens (angle of light captured by the lens, and refractive index n).

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

$$\text{Depth of focus: } k_2 \frac{\lambda}{\text{NA}^2}$$

$$\text{NA} = n \sin\theta$$

Depth of focus

As the wafer is built layer by layer, geometry becomes uneven. Wavelength of light and NA will limit the allowed topology difference.

Planarize wafer between processing steps (before imaging) with chemical mechanical polishing (CMP).

Unfortunately, inherent tradeoff between DOF and resolution (better NA, finer pitch, more narrow DOF).

Reducing k1

Optical proximity correction (OPC), sub resolution assist features (SRAF).

Modelling the lithography as a non-linear low-pass 2D spatial filter, try to come up with an inverse.

Reducing k1

Phase shifting masks (PSM)
Instead of "binary" on/off masks, masks alter the phase of the light.

Double patterning
Split layout accross two (or more) masks

Off-axis illumination
Optimizing the shape of the light source

Reducing k1

Result: k1=0.25 instead of k1=0.5

Restricting allowed pitch may be neccessary for pattern fidelity.

Additionally, NA is improved through immersion lithography (water between lens and wafer, higher refractive index, n).

Extreme UV source

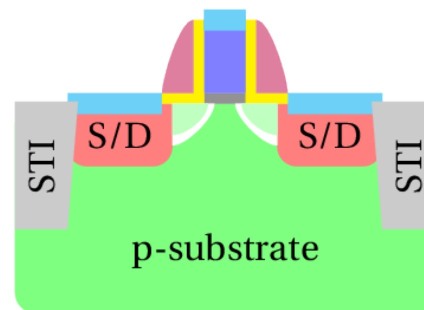
20 nm process with 193 nm light source?
It *can* be done without defying the laws of physics!

Why not use 13.5 nm (EUV) instead?
http://spectrum.ieee.org/semiconductors/devices/euv-faces-its-most-critical-test
http://www.asml.com/asml/show.do?ctx=41905&rid=41906

Front end of line (FEOL)

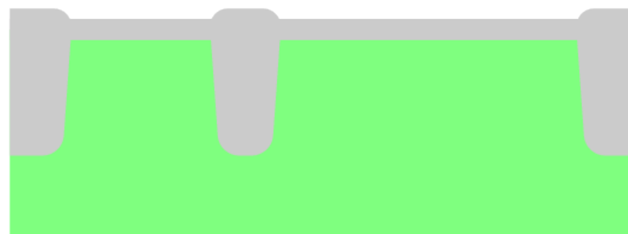
Process modules that form the active devices

- Active area
- Channel doping
- Gate
- Source/drain extension
- Spacer
- Junction
- Silicide



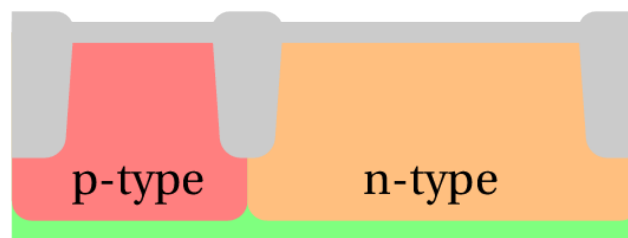
Active area definition

- Shallow trench isoloation (STI)
- Insulation between active devices
- Etch trenches in the substrate
- Filled with SiO2



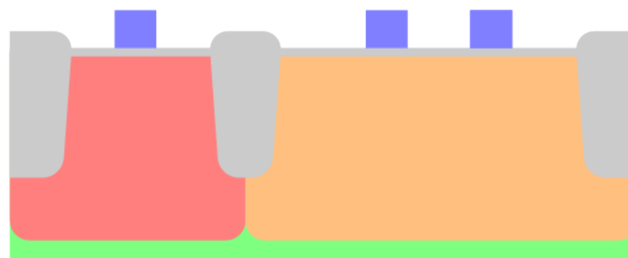
Channel doping

Define p- and n-type regions for NMOS and PMOS



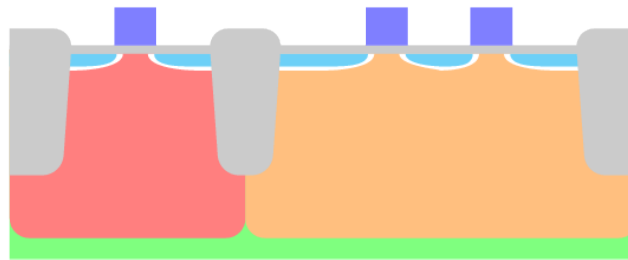
Gate electrode

Gates made of polysilicon



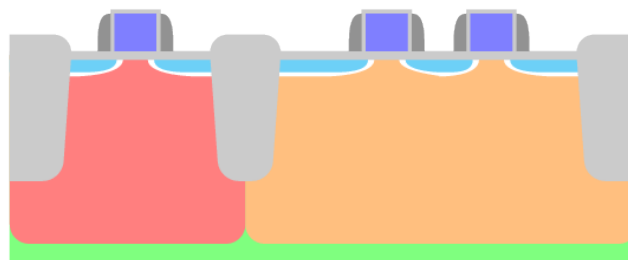
Source/drain extension

Mitigate short-channel effects
source/drain resistance
leakage current, drive current



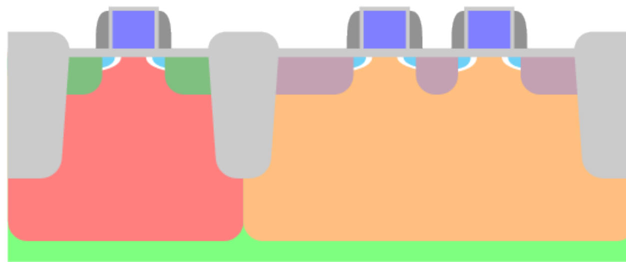
Spacers

Avoid bridging S/D and gate due to silicide
Offset junctions (next step)



Junctions

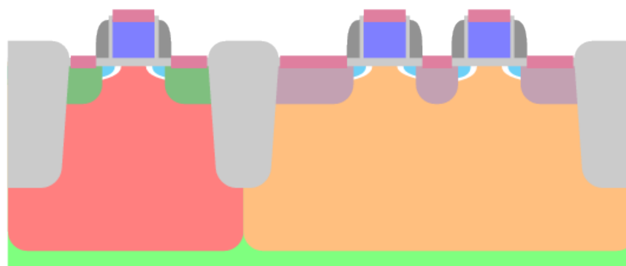
Source and drain junctions
Implant arsenic/phosphore (n-type)
or boron (p-type)



Silicide

Lower resistance, better Ion. Avoid current crowding.

Self aligned silicide = salicide



Back end of line (BEOL)

Back end of line adds connection between devices, contacts and metal layers with vias between layers

Parasitic resistance and capacitance is challenging for scaled technology. In modern CMOS, copper (Cu) metallization and low k dielectric is used.

BEOL modules

Pre Metal Dielectric (PMD)

Contacts to source, drain, and gate (tungsten)

Inter Level Dielectric (ILD)

Vias and metal lines (copper)

Dual Damascene

Metallization used to be etching away aluminum. Impossible with copper.

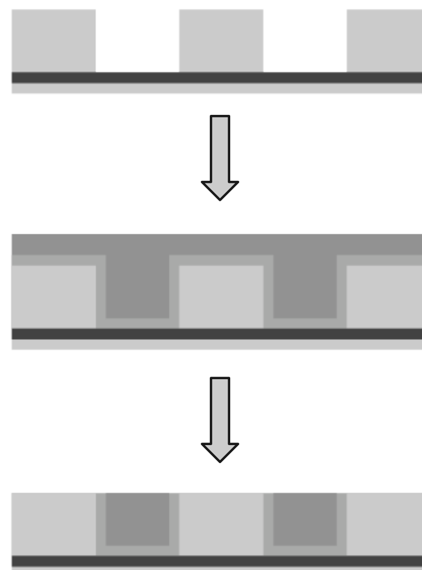
Image: wikipedia.org

Instead: Damascene. Used throughout the BEOL.

(1) Etch trenches in oxide, (2) deposit copper, (3) polish away the overfill (CMP)

Dual Damascene

1. Trenches are etched in the oxide (to the barrier)
2. Metal (Cu) is deposited through electroplating (leaving excess Cu)
3. CMP to remove excess metal.



Dual = via and metal

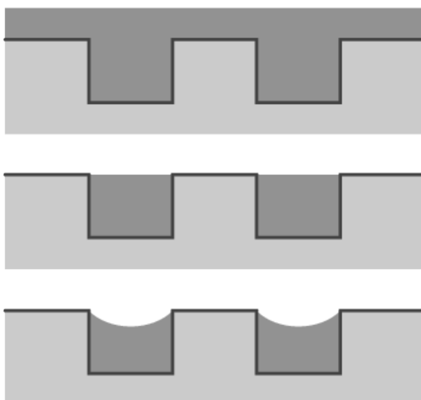
Dual Damascene

Why should we as designers care?

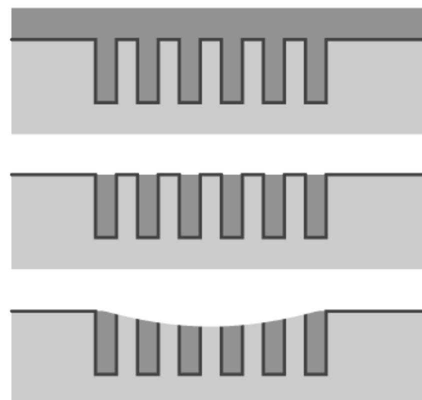
CMP polish rate is pattern dependent, i.e width and spacing of metal lines matter.

Poor layout results in metal lines that are too thin and/or less dielectric separation of metal layers. **Layout dependent delay**. Post-layout simulation does not neccessarily reveal these problems.

Dishing and erosion



Dishing affecting wide metal lines (Cu polishes faster than dielectric)



Erosion affecting high density metal pattern

Drawing layout

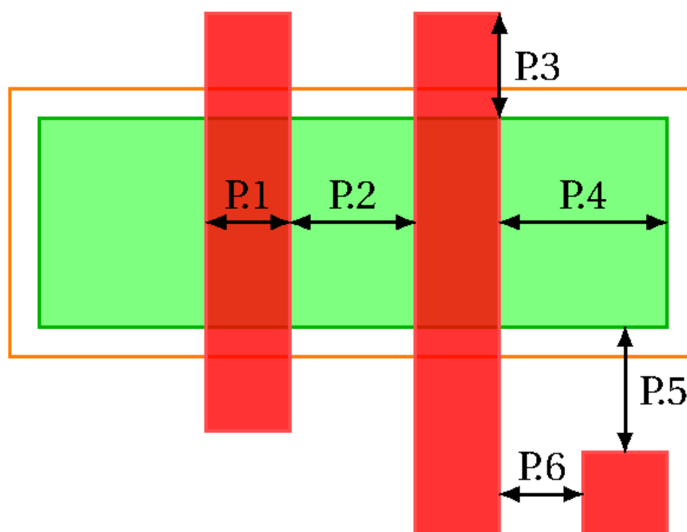
Layout is drawing the masks used in the manufacturing process.

As we have seen, the layout we draw is not perfectly reproduced on the wafer.

We must comply with a set of rules to ensure that the layout we draw is manufacturable.

Design rule examples

	Rule name (minimum)
P.1	Poly width
P.2	Space poly and active
P.3	Poly ext. beyond active
P.4	Enc. active around gate
P.5	Spc. field poly to active
P.6	Spc. field poly



Design rule examples

Poly rules example (FreePDK45)

	Rule name (minimum)	Length
P.1	Poly width	50 nm
P.2	Space poly and active	140 nm
P.3	Poly extension beyond active	55 nm
P.4	Enclosure active around gate	70 nm
P.5	Space field poly to active	50 nm
P.6	Space field poly	75 nm

Design rule examples

Metal1 rules example (FreePDK45)

	Rule name (minimum)	Length
M1.1	Metal1 width	65 nm
M1.2	Space metal1	65 nm
M1.3	Enclosure around contact (two opposite sides)	35 nm
M1.4	Enclosure around via1 on two opposite sides	35 nm
M1.5	Space metal1 wider than 90 nm and longer than 900 nm	90 nm
M1.6	Space metal1 wider than 270 nm and longer than 300 nm	270 nm
M1.7	Space metal1 wider than 500 nm and longer than 1.8 um	500nm
...	...	...

Density design rules

In addition to spacing and area rules. There are density rules.

Ref. dishing and erosion resulting from CMP.

Typically, the layout will undergo dummy fill to comply with density rules (automatic). Neccessary for manufacturability, but increases capacitance.

Antenna design rules

Large area metal connected to a MOSFET gate can collect ions during manufacturing and irreversibly break down gate oxide.

Design rule switches

Different set of rules can be invoked for different parts of the circuit.

E.g. Minimum rules for high density generic digital circuitry

Analog or DFM rules for sensitive circuits.

Design Rule Check, DRC

Design Rule Check (DRC)

Large number of rules to comply with. Difficult to keep track of.

Automated by design tools with foundry rule set.

Used to be pass/fail, more recently reporting level of severity. Some rules can be waived.

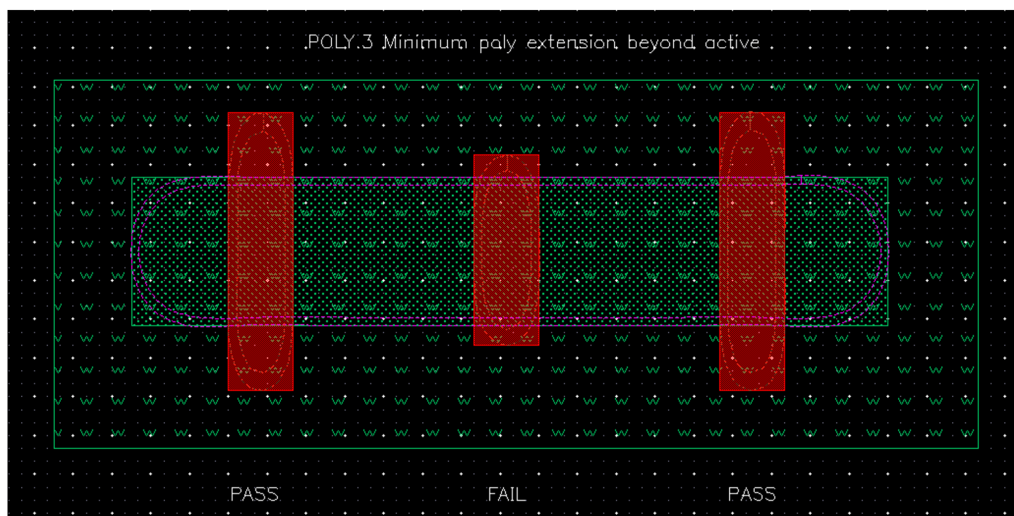
Litho friendly design, LFD

Design rules does not guarantee a robust design or good yield.

Possible to simulate and analyze how the layout will print on the wafer.

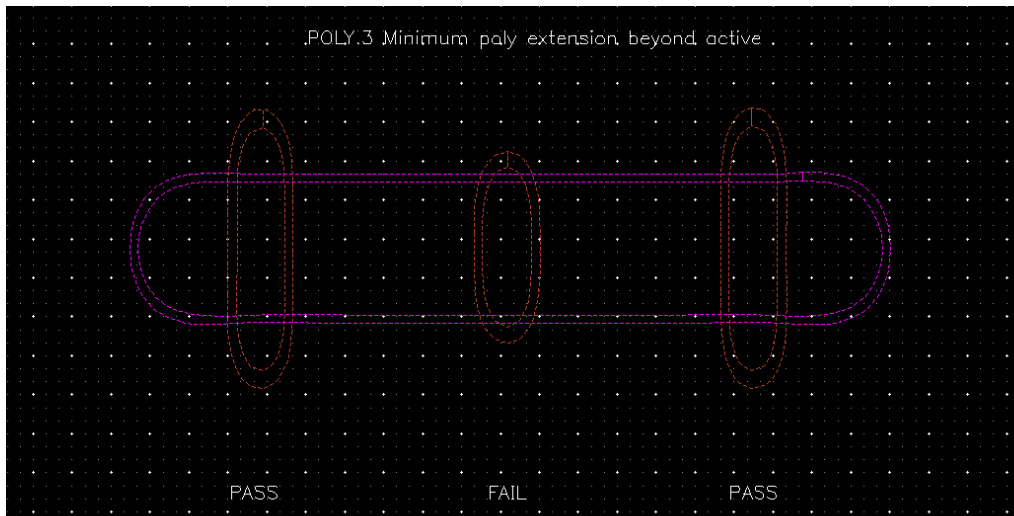
Difficult to get access to data.
Non-linear 2D spatial filter.

Lithography simulation



Litho simulation using FreePDK45 and Calibre (http://www.eda.ncsu.edu/wiki/FreePDK)

Lithography simulation



Litho simulation using FreePDK45 and Calibre (http://www.eda.ncsu.edu/wiki/FreePDK)

Layout vs. schematics, LVS

Recognizing shapes in the layout (transistors and passive devices), and how they are connected.

Comparing layout netlist to schematics.

Post layout simulation

Extracts layout dependent parasitics (capacitance and resistance), and some layout dependent transistor parameters (e.g. LOD which we will discuss in the short-channel lecture).

More accurate simulation results (but does not include all effects). Also, parasitics have fast/slow corners, temperature dependence. Results in slow simulation due to large netlists.

Interconnect

Drawing metal "wires" in the layout is not like wires in the schematic.

Must think about resistance, capacitance, and inductance. (E.g. 0.1 Ω/□ for metal, and 10 Ω for via)

Crosstalk and ground bounce. Decoupling.

Interconnect

Single via approximately 10 Ω. Worse, single via failure.

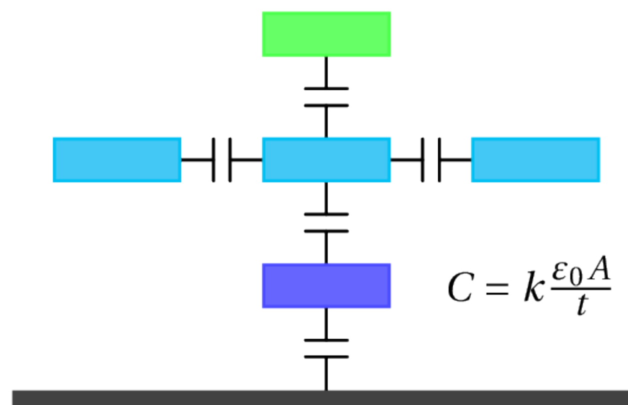
Not only resistance, but limitaions on current capability (electromigration).

Process documentation should list actual values for a given process.

Interconnect

Overlap capacitance

low-k dielectric helps reduce interconnect capacitance



Additionally, fringe capacitance also important.

Interconnect

Sizing metal lines is a tradeoff between capcitance vs resistance (and current handling capability).

Wide lines, fewer squares, less resistance, but potentially more overlap capacitance.

Finally, to make it even more complicated, resistance and capacitance will vary due to dishing and erosion.

Passive components

Mixed signal and analog require passive components (resistors, capacitors). RF needs inductors.

Why not use parasitic resistance and capacitance? Possible in some cases.

Resistors

Several possibilities. Need to consider:

- Ω/□ (area, practical limit for large R)
- Temperature dependence (TC)
- Voltage dependence (linearity)
- Mismatch (ΔR/R, abs value +/- 20 %)
- Parasitic capacitance

The TC and voltage dependence is not only linear, but also quadratic in the simulator.
E.g. R(T) = R(T0) [1 + TC1(T-T0) + TC2(T-T0)^2]. Similar for voltage dependency.

Resistors

Realistic alternatives for large resistors:

N-well: Large R, poor TC (> 2000 ppm/C), poor linearity (< 1 %), low mismatch, parasitic capacitance from pn-depletion. Always available.

Poly with silicide block: Large R, good TC (~ 100 ppm/C), reasonable linearity (< 0.1 %), low mismatch. Extra layer needed.

Capacitors

Need to consider:

- F/m^2
- Temperature dependence (TC)
- Voltage dependence (linearity)
- Mismatch (ΔR/R)
- Cost

Capacitors

MOSCAP, using a mosfet as a capacitor (Cox). High capacitance per area, very non-linear, good e.g. for decoupling, but gate leakage current is problematic.

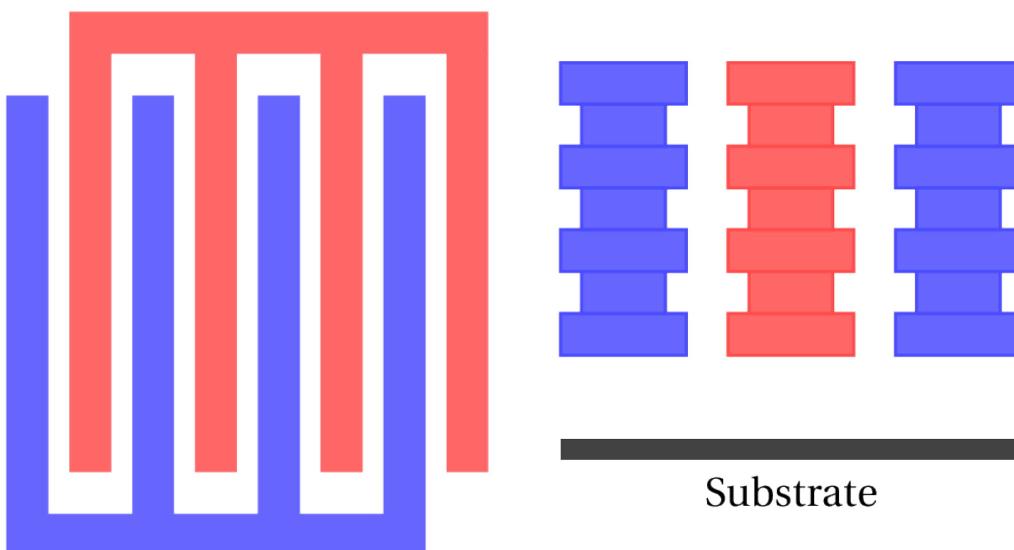
PiP, using two poly layers. Usually not available in modern CMOS processes.

Capacitors

MiM (Metal-insulator-Metal). Requires extra mask. ~ 1 or few fF/um^2. Good option if available. Thin separation of metal layers and special dielectric. Usually available in RF process flavours. Cost issue.

MoM (Metal-oxide-Metal). Exploit fine pitch in CMOS. No extra processing required.

MoM Capactiors



Matching passives

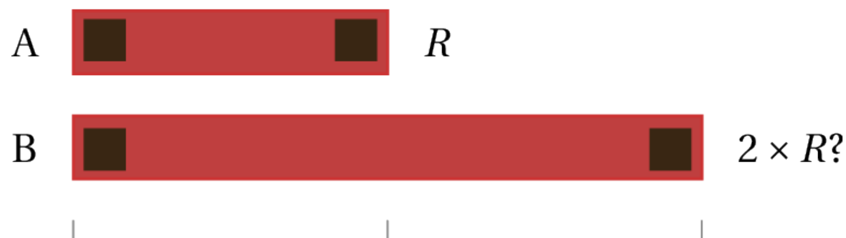
Systematic vs. random

- Different absolute values between runs
- Layout dependent problems
- Stress, thermal, or doping gradients

Good layout practice helps

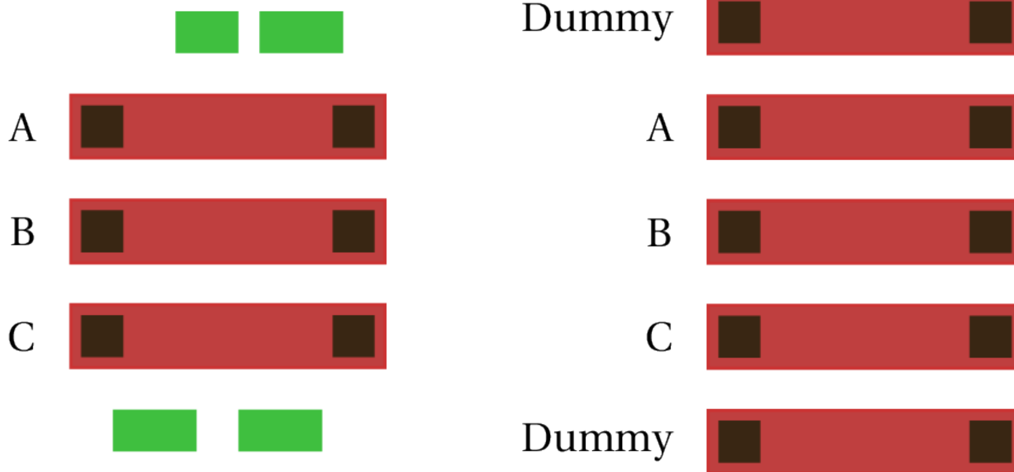
- Unit elements
- Dummies (each unit element should see the exact same surroundings)
- Interdigitation or common centroid

Unit elements



Instead, make identical unit elements. Less systematic mismatch.

Dummy elements

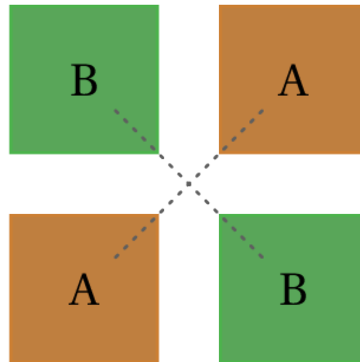


Dummies to make sure matching elements see the same surroundings

Interdigitated layout

Process gradient are spread more evenly between the two elements. Proximity helps with matching!

Common centroid



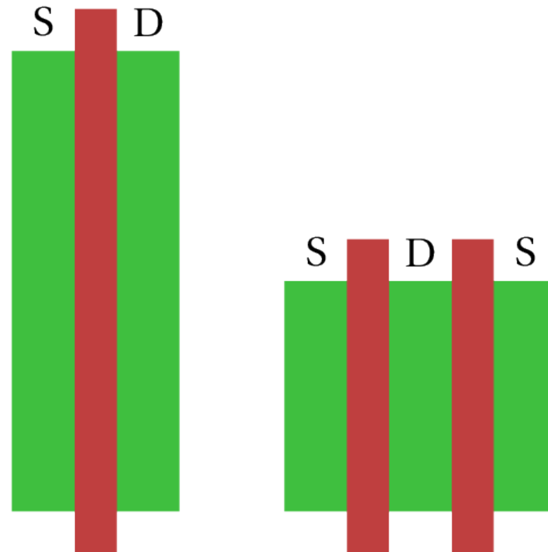
Better process gradient cancelation than interdigitated layout. Perfect cancelation of linear gradients.

Drawing transistors

Unit elements, dummy, and common centroid also applies to layout of transistors. However, there are additional issues that need attention when laying out transistors.

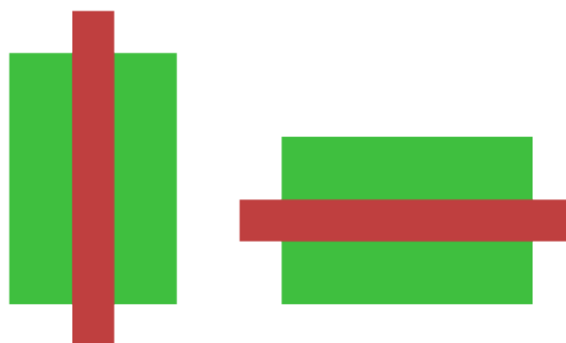
- Multi-finger devices
- S/D symmetry
- WPE and LOD
- Latch-up

Multi-finger devices



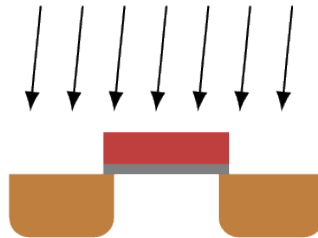
Device orientation

Devices with different orientation do not match!



Source/drain asymmetry

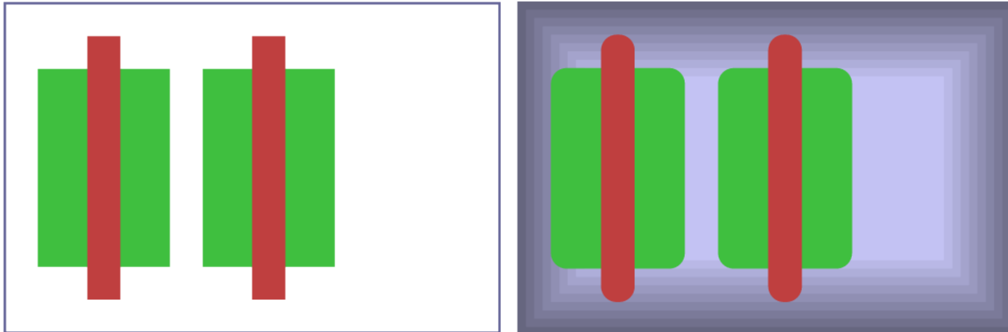
Source and drain may not be symmetric due to ion implantation angle, neccessary to avoid implant depth issues (channeling).



Well proximity effect

High energy ion implants to form the well. Scattering from the edge of the photoresist mask, and embedding in the silicon surface (near well edge). Transistors close to the well edge will therefore have different properties. This is known as the well proximity effect (WPE). Important for matching.

Well proximity effect

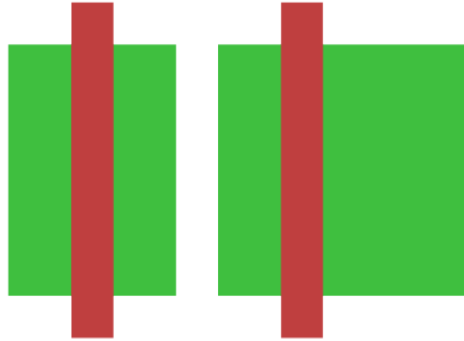


As with S/D, implantation angle may render the scattering and doping asymmetric

STI stress (LOD)

Shallow trench isolation strains the active area of the transistor. Influcences mobility and threshold voltage (stress induced enhancement or suppression of dopant diffusion). Distance between gate and STI impacts perfomance. Important for matching. (Parameters SA and SB in BSIM). Also known as LOD (length of diffusion), LOD = SA + SB + L

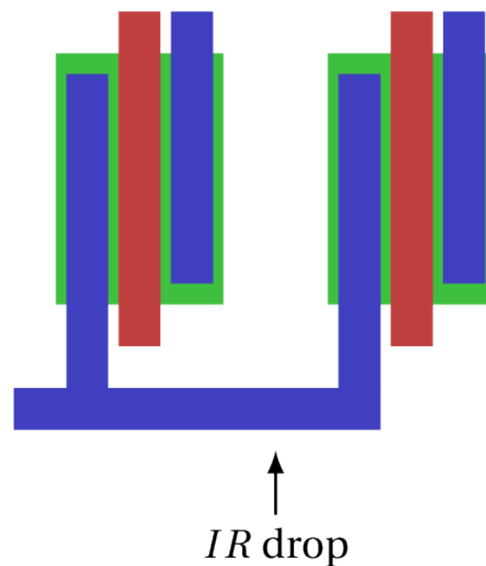
STI stress (LOD)



Transistor interconnect

Unbalanced metal routing will cause the transistors to see different source voltage.

Also, distribute reference as current, not bias voltage.



Matching

This discussion about matching was about minimizing *systematic mismatch*. We will discuss random mismatch later.

Shielding

Substrate ties circuits together. Digital switching couples to the substrate.

- Guard rings around the circuit: Substrate ties and n-well (preferably deep n-well)
- Separate Vdd for digital and analog
- Fully differential signals

See sect. 18.3 in Razavi's book.

Latch-up

As we saw with bandgap references. Parasitic BJTs are readily available in CMOS. Parasitic BJTs may inadvertently turn on due to a large injection of current into the substrate.

Typical design rules make sure that substrate and n-well contacts have sufficiently small spacing. However, latchup is an important problem, and requires careful consideration.

Bond pads

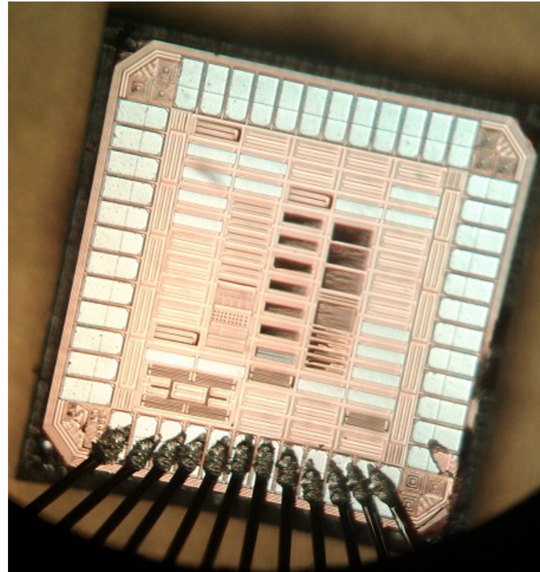
Used for connecting bond wires between die and package.

- Mechanical stability
- ESD protection (very important, and adds C)
- Aluminum (while other metal is Cu)

Pad frame usually contains supply nets (also used by ESD circuitry)

Bond pads

Bonding gone wrong ...



Seal ring

A seal ring is a structure to enclose the die (outside the pad frame). Protects the die from moisture and sawing. Also contains scribe (where to saw the die).

Packaging

Packaging adds very significant parasitics.

Bond wires introduce inductance. Rule of thumb is 1 nH/mm.

Inductors like to keep current constant. Voltage will change to make this happen. Important to balance with decoupling capacitors. However, transients will remain.

References

Hastings, *The Art of Analog Layout*, Prentice Hall, 2001

Orshansky, et al., Design for Manufacturability and Statistical Design, Springer, 2008

Wong, et al., Nano-CMOS Circuit and Physical Design, Wiley, 2005