

Podstawy Mikroelektroniki

Piotr Kmon

Budynek **B1**, pokój **210**, e-mail: **piotr.kmon@agh.edu.pl**

Rafał Kłeczek

Budynek **B1**, pokój **210**, e-mail: **rafal.kleczek@agh.edu.pl**

**Wydział Elektrotechniki, Automatyki, Informatyki i i Inżynierii
Biomedycznej**

Katedra Metrologii i Elektroniki



AGH

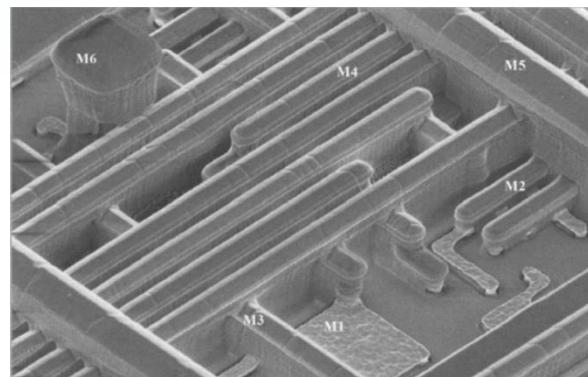
Interconnects

The performance of complex ICs heavily depends on the quality of the available interconnects, requiring more metal layers in new generations of the technology.

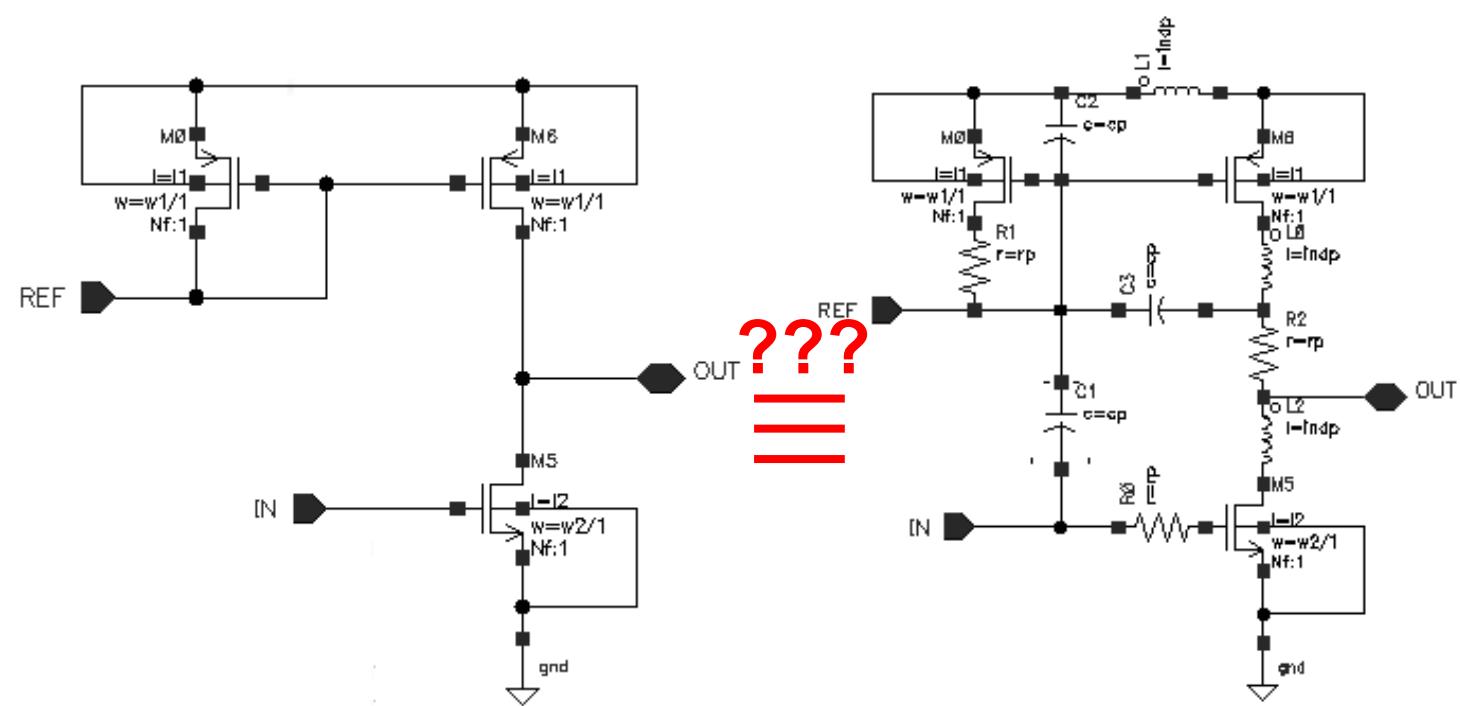
Two properties of interconnects: series resistance and parallel capacitance impact the performance of the circuit.

The series resistance becomes problematic in supply and ground lines, creating dc and transient voltage drops. The resistance of metal wires can be easily estimated at low frequencies, where skin effect is negligible. Typical sheet resistances are $30 \text{ m}\Omega/\square$ for the topmost (thickest) layer and $70 \text{ m}\Omega/\square$ for the lowermost.

For long signal lines the distributed resistance and capacitance of the wire may result in a significant delay.

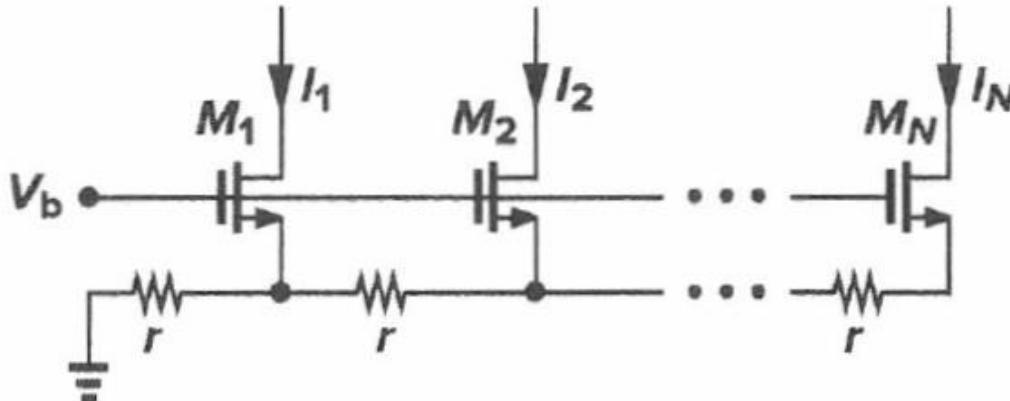


Cross sections of CMOS back end.



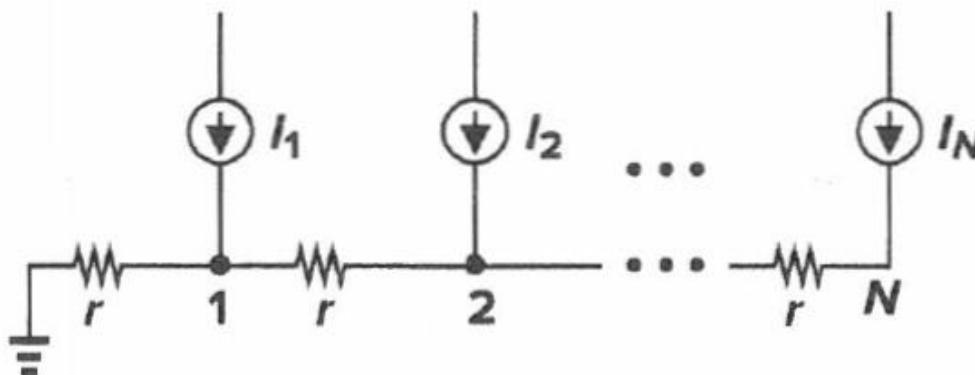
Interconnects

Effect of ground resistance in a D/A converter.



For $I_1 \approx I_2 \approx \dots \approx I_N \approx I$ and r is small:

$$V_N = Ir + I(2r) + \dots + I(Nr) = \frac{N(N+1)}{2} Ir$$



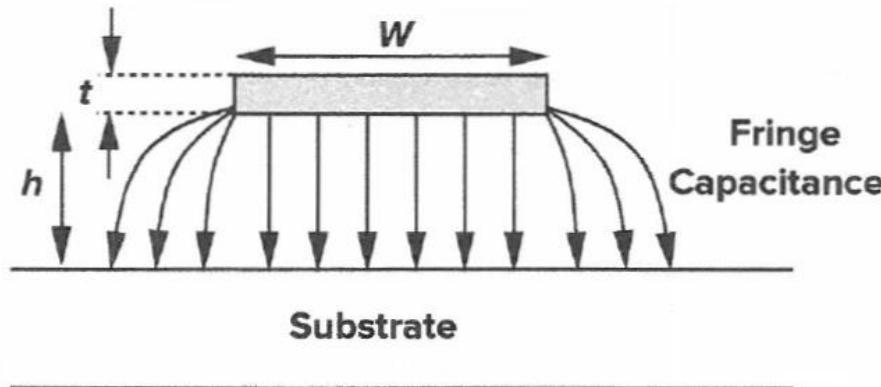
what under assumption that V_N is relatively small leads to the relative mismatch between I_1 and I_N :

$$\left| \frac{I_1 - I_N}{I} \right| = g_m r \frac{N(N-1)}{2}$$

The key point is that the error grows in proportion to N^2 . The ground bus must be therefore sufficiently wide to minimize r .

Interconnects

The problem of interconnect capacitance is much more complicated. In practise, the capacitances between the layers are calculated by „electromagnetic field solver”, measured experimentally, and tabulated in the process design manual.



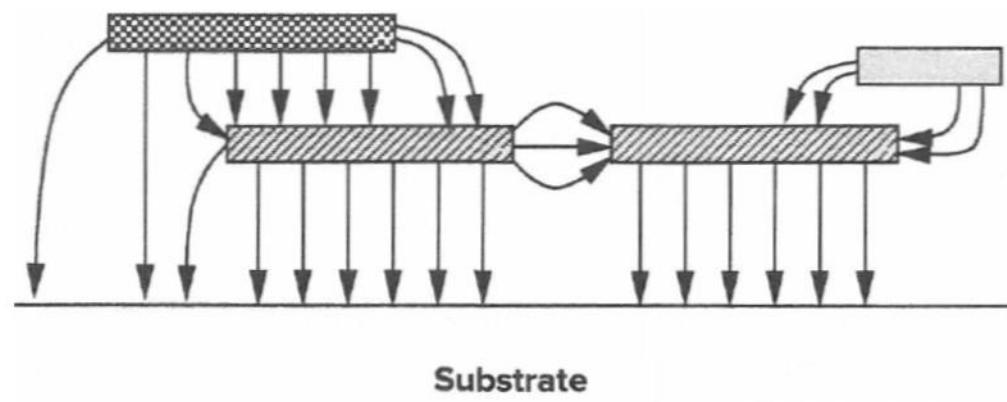
Parallel-plate and fringe capacitance of an interconnect.

Minimum widths and capacitances of interconnects in a 0.25 μm technology.

	Poly	Metal 1	Metal 2	Metal 3	Metal 4
Minimum Width (μm)	0.25	0.35	0.45	0.50	0.60
Bottom-Plate Capacitance ($a\text{F}/\mu\text{m}^2$)	90	30	15	9.0	7.0
Fringe Capacitance (Two Sides) ($a\text{F}/\mu\text{m}$)	110	80	50	40	30

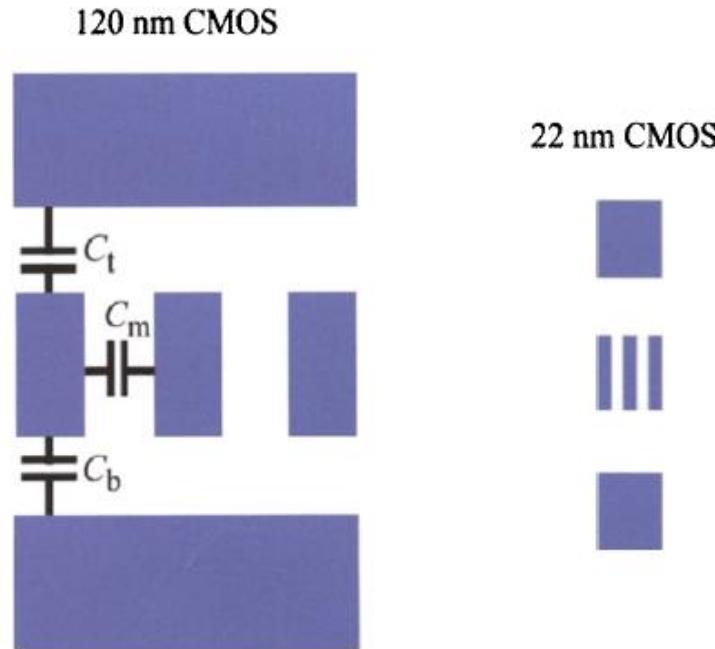
Empirical relationship for calculating the total per unit length on top of a conducting substrate:

$$C = \epsilon \left[\frac{W}{h} + 0.77 + 1.06 \left(\frac{W}{h} \right)^{0.25} + 1.06 \left(\frac{t}{h} \right)^{0.5} \right]$$



Complex interconnect structure.

Interconnects



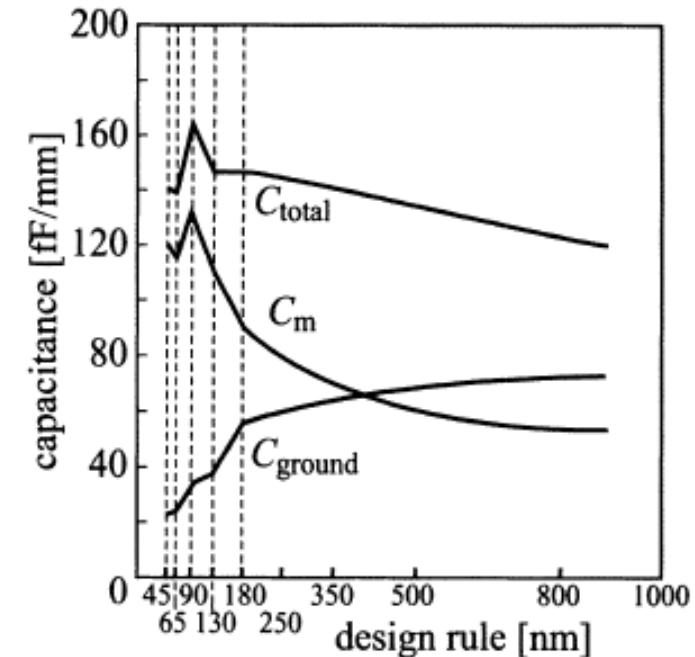
Expected scaling of metal track width and spacing.

As the CMOS technologies are becoming more advanced the bottom (C_b) and top capacitances (C_t) reduce while the mutual capacitances (C_m) increase.

The increase in mutual capacitance (C_m) has dramatic effects on the performance and robustness of integrated circuits because:

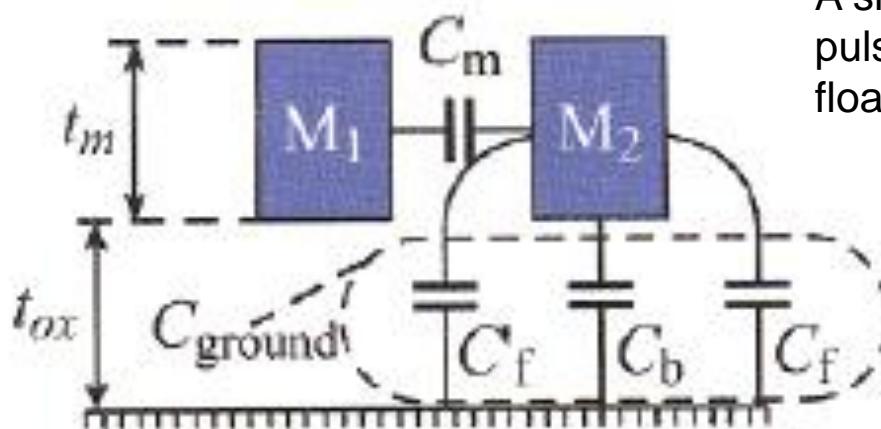
1. Interference between two neighbouring interconnect lines increase, which is usually referred to as crosstalk (**performance**).
2. The growing signal propagation delay across the interconnect because of its increasing RC times (**performance**).
3. The interconnect capacitances also affect the overall IC's power consumption (**robustness**).

The above model refers to the two minimum spaced wires in the same metal layer.



Interconnect capacitances across various technology nodes.

Interconnects



A signal swing ΔV_{M1} on metal track M_1 causes the noise pulse on the metal track M_2 . Whenever the M_2 is a floating metal the noise pulse is given by:

$$\Delta V_{M2} = \frac{C_m}{C_m + C_{\text{ground}}} \times \Delta V_{M1}$$

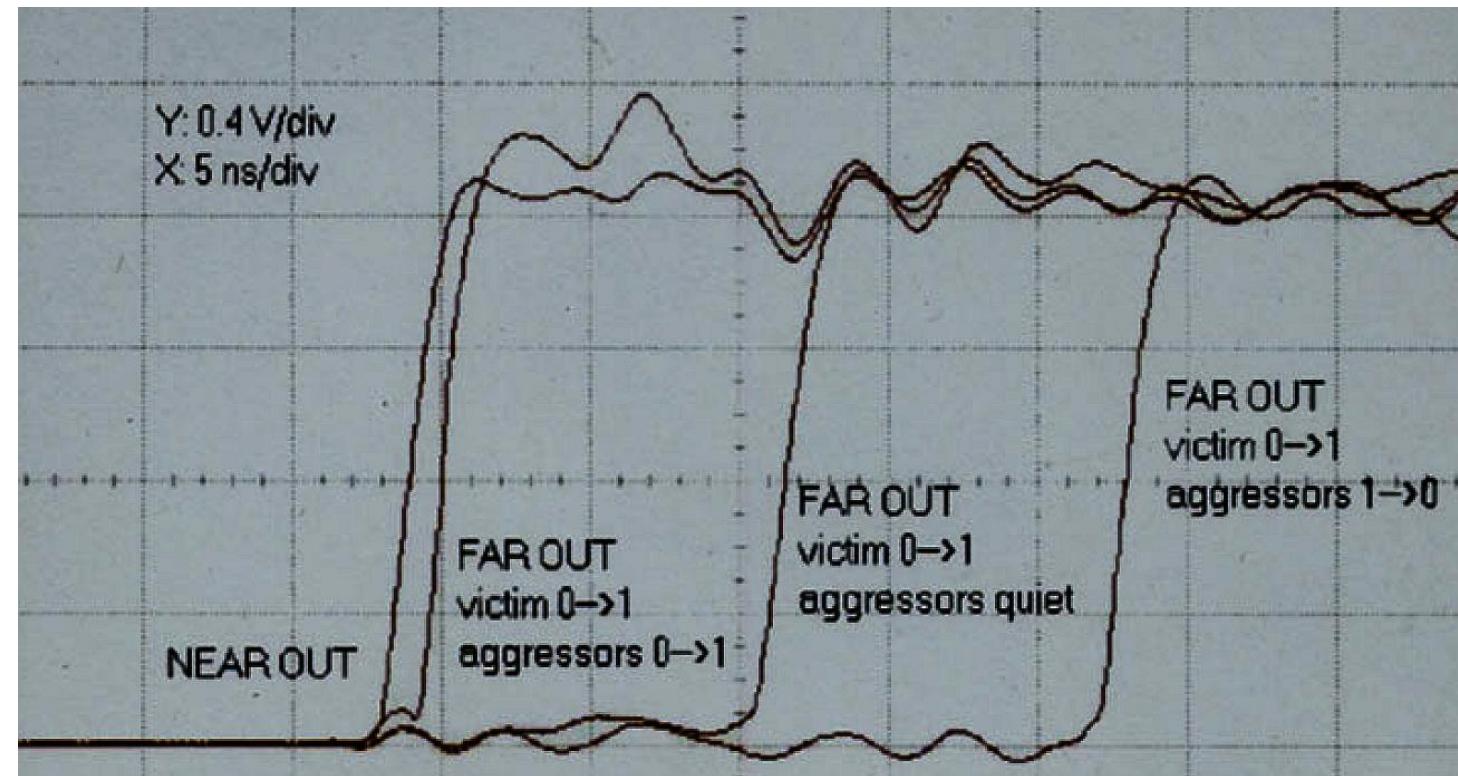
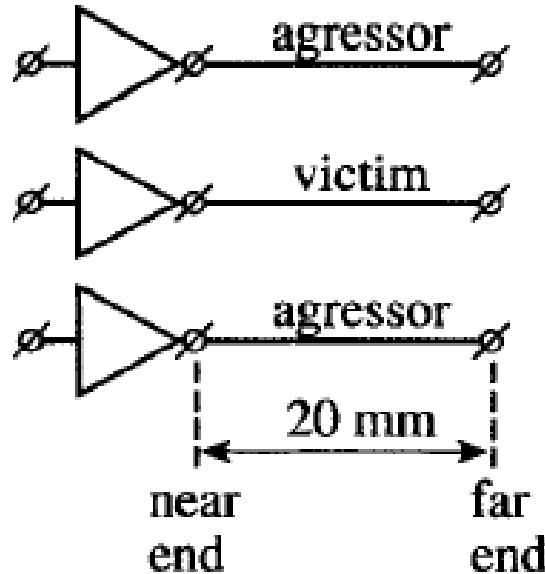
Node	180 nm	130 nm	90 nm	65 nm	45 nm
C_m	89 fF	110 fF	132 fF	115 fF	120 fF
C_{ground}	58 fF	36 fF	32 fF	21 fF	18 fF
C_{total}	147 fF	146 fF	164 fF	136 fF	138 fF
ΔV_{M2}	$0.6 \cdot \Delta V_{M1}$	$0.75 \cdot \Delta V_{M1}$	$0.8 \cdot \Delta V_{M1}$	$0.84 \cdot \Delta V_{M1}$	$0.86 \cdot \Delta V_{M1}$

Capacitive values for second metal layers in different CMOS processes.

The bottom table line shows the amount the one signal propagates to the neighbouring line through the cross-talk. For the 45nm CMOS this means the 86% of that switching signal propagates into its floating neighbours.

Because of this, all floating lines (e.g. precharged memory lines, tri-state lines, etc.) are very susceptible to the cross-talk.

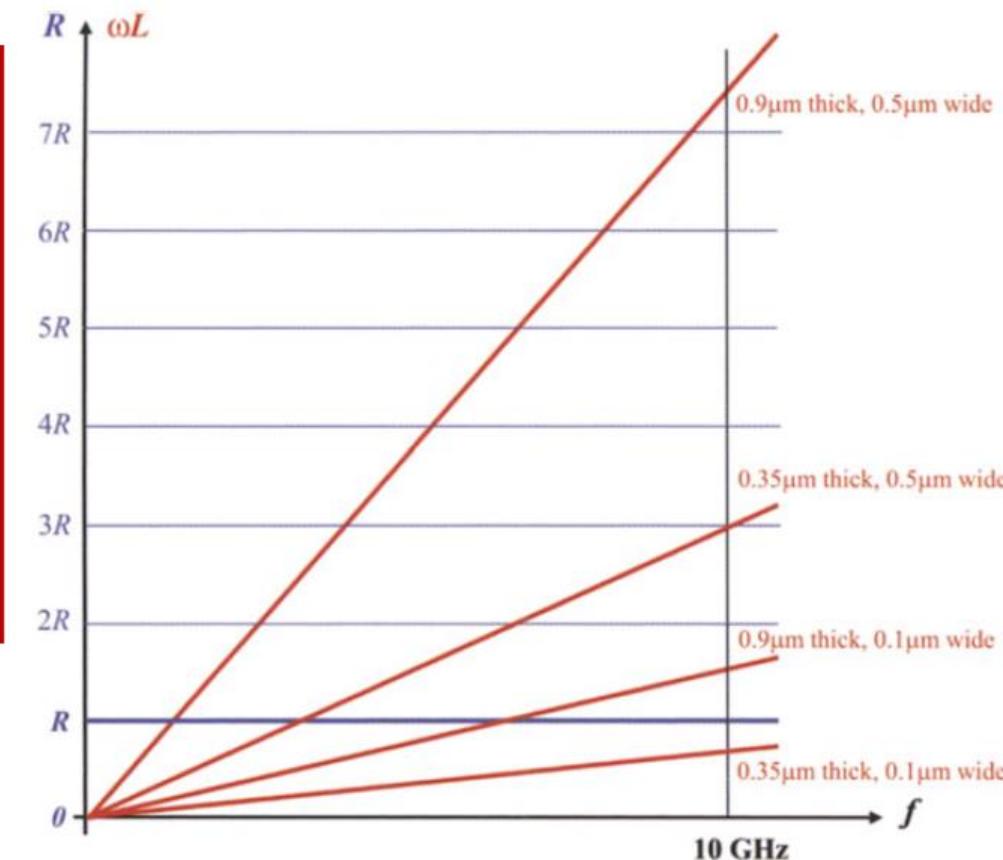
Interconnects



The switching behaviour of both aggressors with respect to the victim causes a large dynamic range in signal propagation across the victim line. In case both aggressors switch opposite to the victim the signal propagation across the victim lasts about sixteen times longer than in case the aggressors and victim all switch in the same direction.

Interconnects

Metal routing inductance influence.



Influence of the frequency on the inductance of the 500 μm long on-chip metal lines.

At a frequency of 10 GHz, the inductance contribution (of a 350 nm thick signal line in a 65nm CMOS process) to the total impedance of a metal wire reaches about two third of the resistance contribution. This means that there is a need to change from an RC interconnect model to an RLC model for designs that exceed 1 GHz (at this frequency the inductance value is about 10% of the resistance value and can thus no longer be neglected).

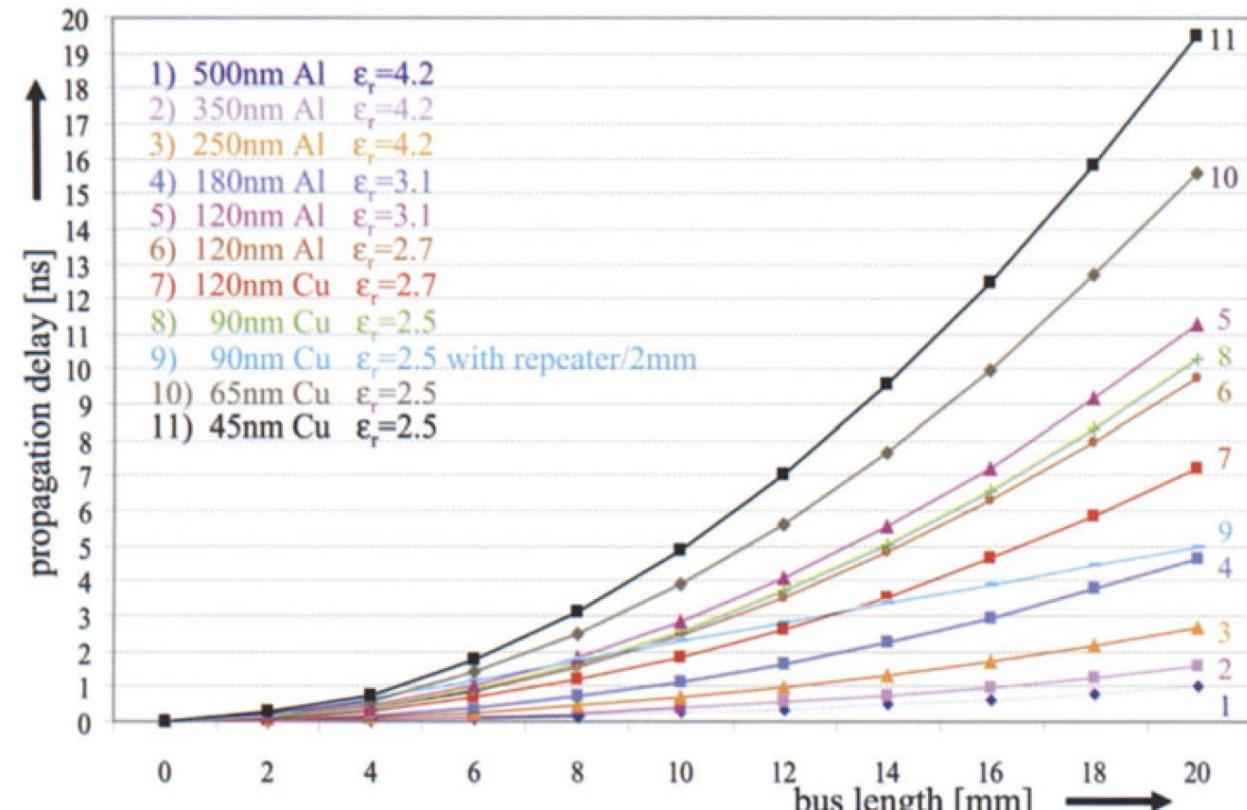
Therefore, even in designs that do not yet reach 1GHz, the wider metal lines, with lower resistance (e.g., in clock distribution networks and upper metal layers) can exhibit significant inductive effects.

The lines in figure represent the relative ωL values with respect to their resistance values. Therefore, the ωL value of a 0.9 μm thick and 0.5 μm wide metal track reaches the level of its resistance value already at a frequency of close to 1.5 GHz.

Interconnects

Technology	Ratio: wire load/fan-in
350 nm	30/70
250 nm	33/67
180 nm	36/64
130 nm	45/55
90 nm	54/46
65 nm	66/34
45 nm	75/25
32 nm	81/19

The increase in the average ratio between wire load and fan-in, for average standard cell blocks, caused by scaling.



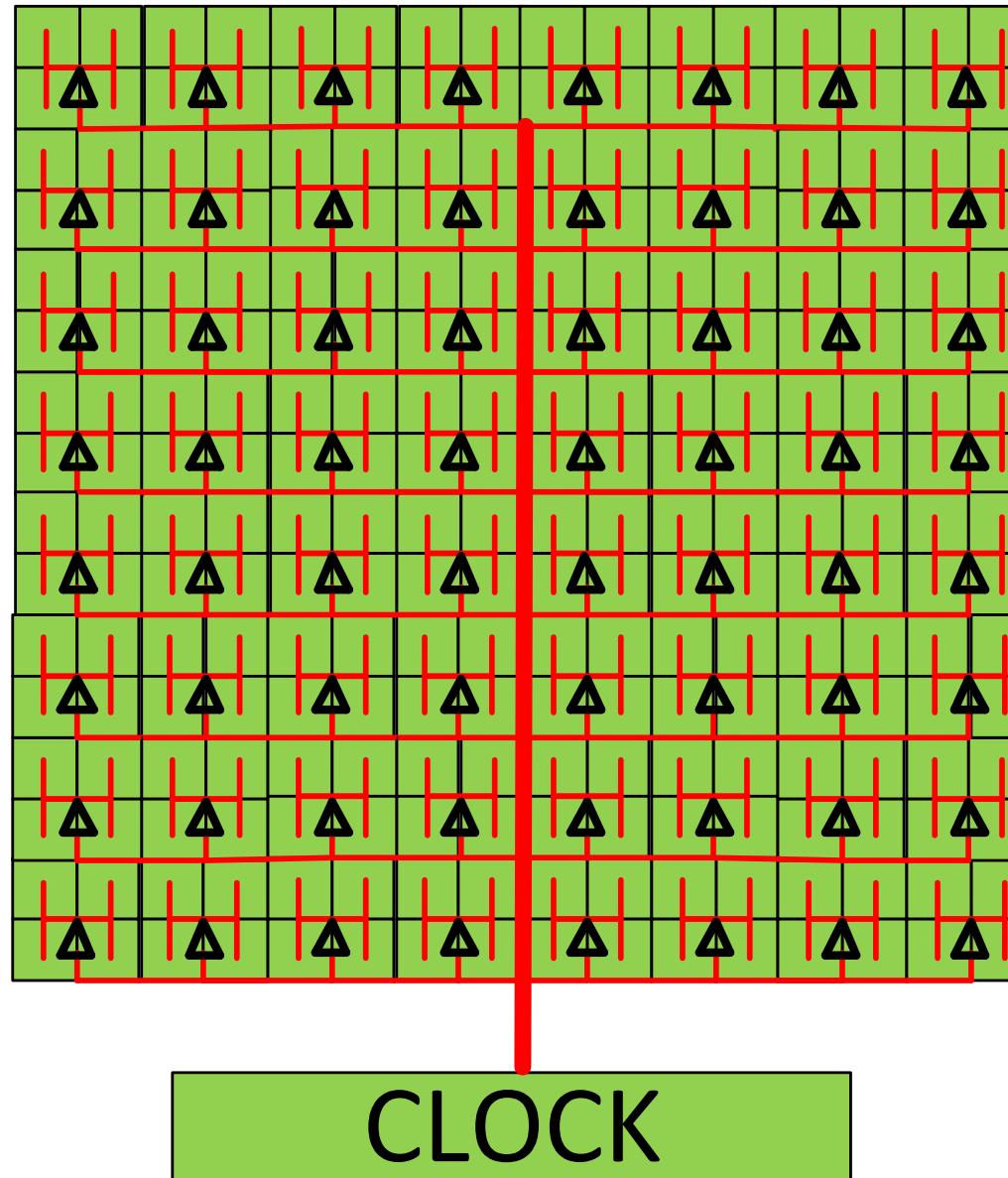
Propagation delay of an embedded track in different technologies

The output load of a logic gate is equal to the total of the fan-in capacitances of its connecting gates and the total wire load of the interconnections.

Although the introduction of copper with the 120 nm node shows some relief in the increase of the propagation delay, it only helped for about one technology node. This means that in the 120 nm node, with an aluminium backend, the interconnect propagation delay would have reached the same order of magnitude as the 90 nm node with a copper backend. Signal propagation delay over a metal wire is proportional to the square of its length. The diagram also shows that the propagation delay will further increase.

These inquires different design architectures, in which the highspeed signals are kept local.

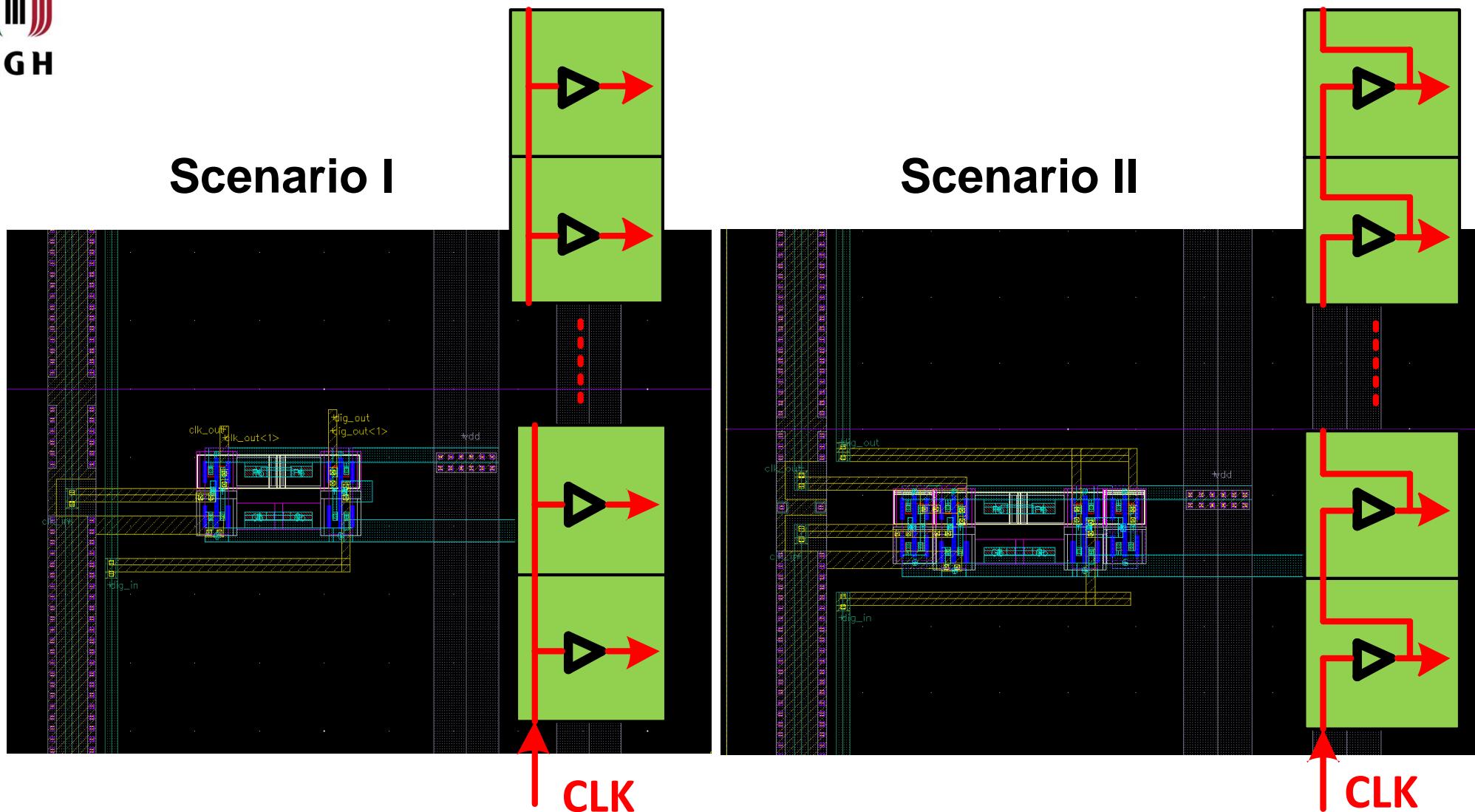
Clock signal distribution



Idea of commonly used clock distribution

Clock signal distribution

96 rows of 50um x 50um pixels are considered



Layout view with its sketch description of two clock distribution scenarios considered

Shielded line parasitic capacitance to VSS → about 10.3 fF per pixel

Unshielded line parasitic capacitance to VSS → about 2.6 fF per pixel

Clock signal distribution – Scenario I - 200 MHz (5 ns)

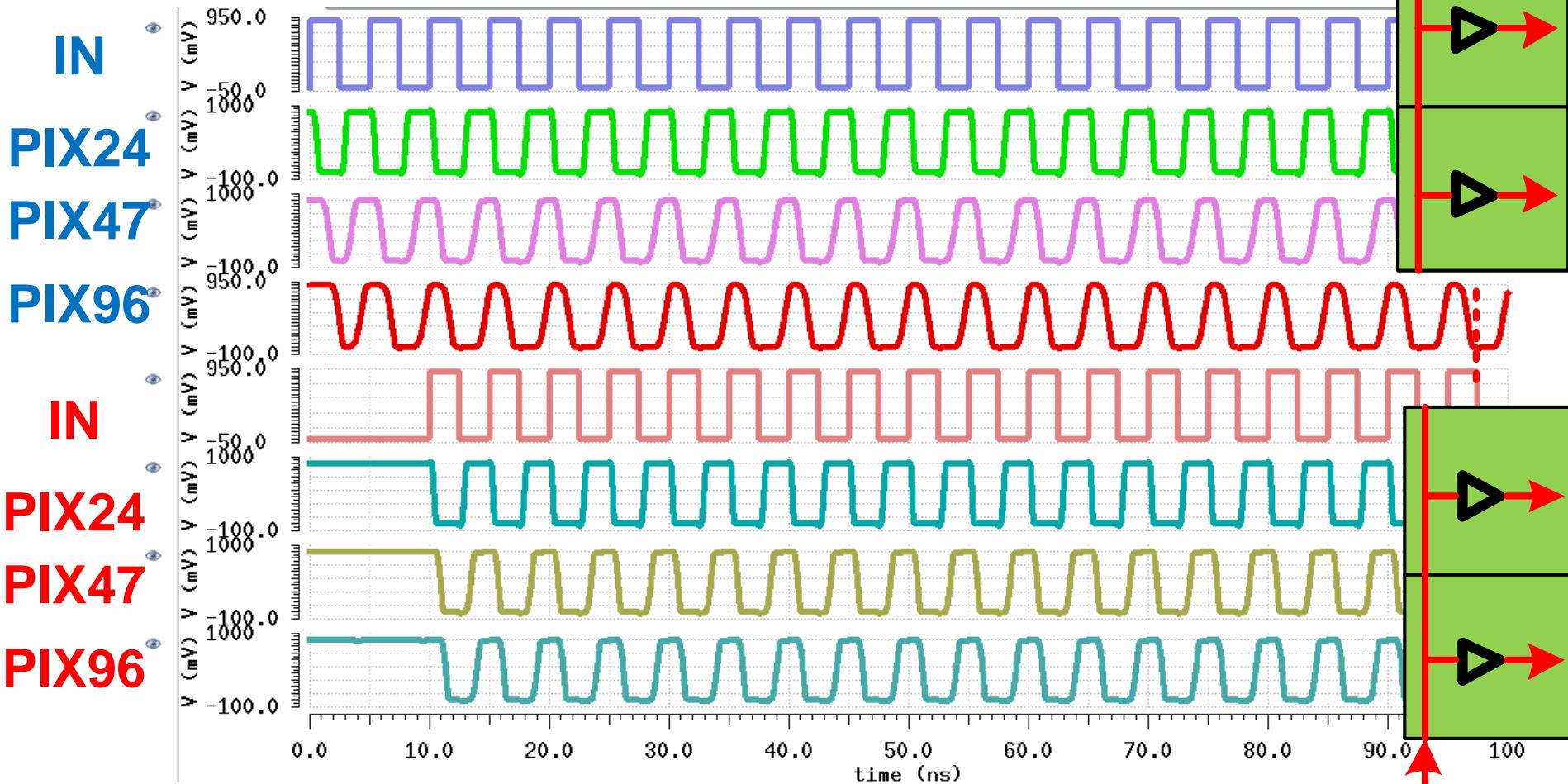


AGH

Power consumption/pixel for shielded line → 0.15 µW (0.17 µA @ 0.9 V)

Power consumption/pixel for unshielded line → 0.15 µW (0.17 µA @ 0.9 V)

SHIELDED LINE

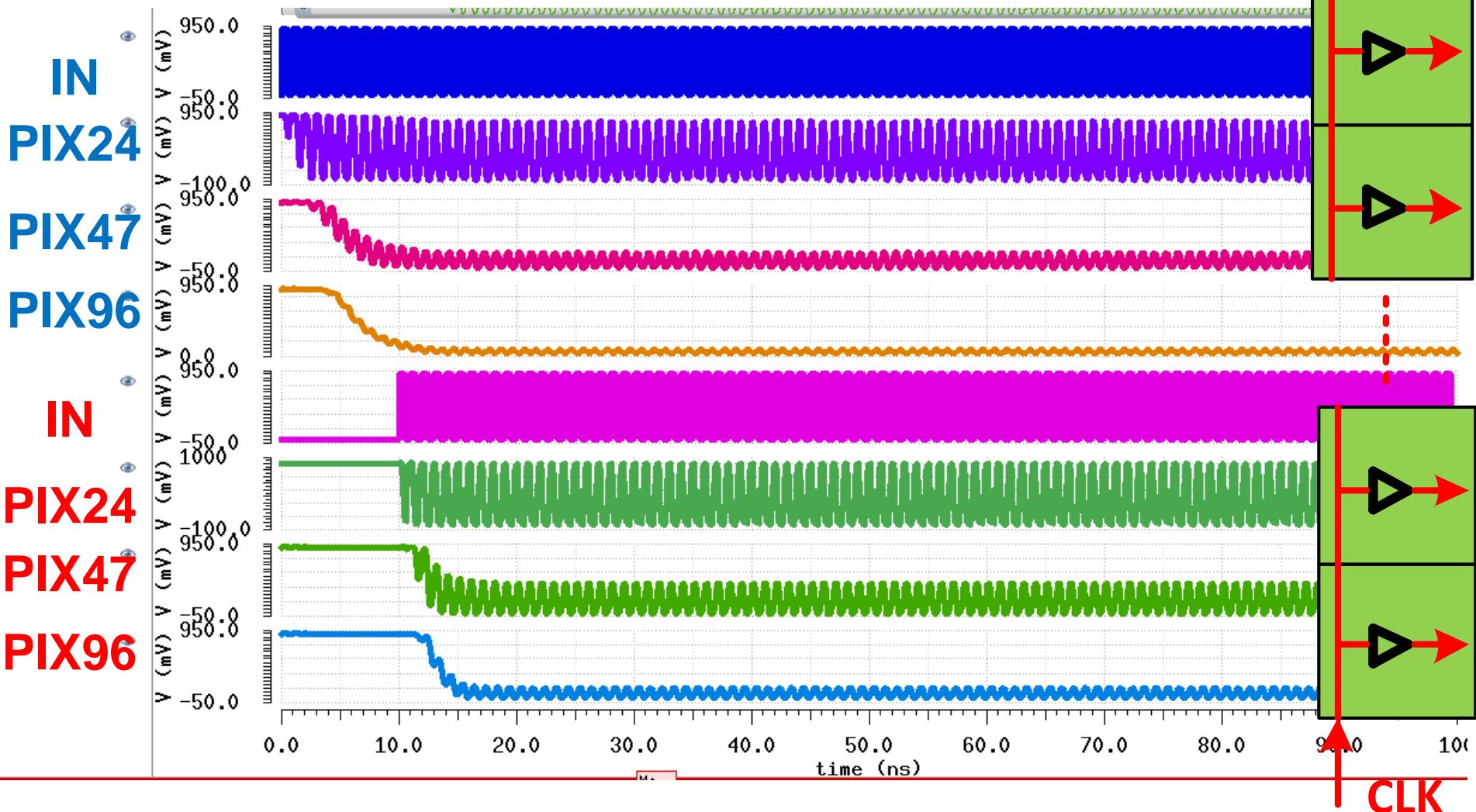


Postlayout simulation results of two signal propagation: clk (through the shielded line) and dig (through the unshielded line)

Clock signal distribution – Scenario I - 1 GHz (1 ns)

Column is driven by ideal buffer.

UNSHIELDED LINE SHIELDED LINE



Postlayout simulation results of two signal propagation: clk (through the shielded line) and dig (through the unshielded line)

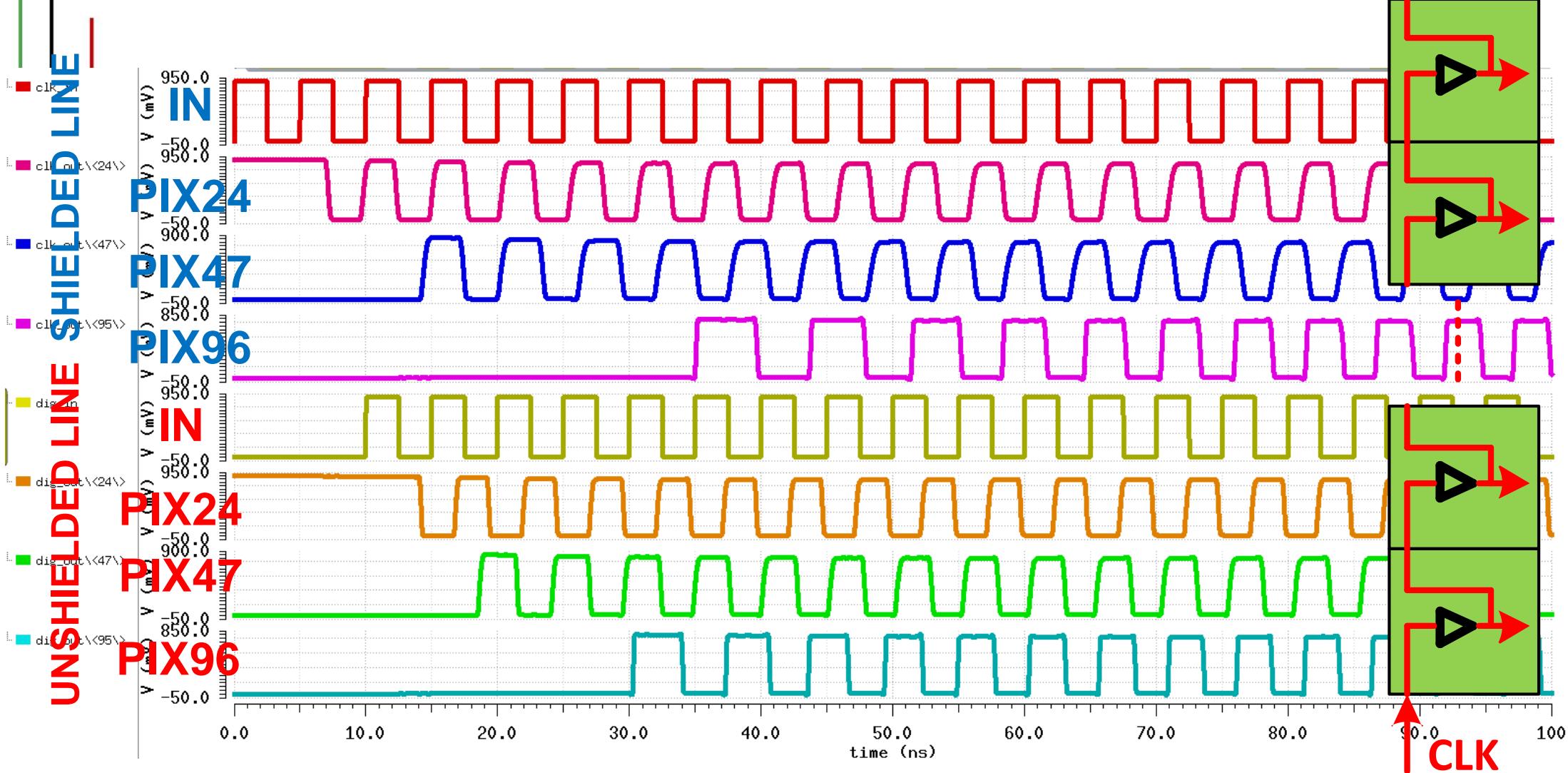
Clock signal distribution – Scenario II - 200 MHz (5 ns)



Power consumption/pixel for shielded line → 2.0 μ W (2.2 μ A @ 0.9 V)

AGH Power consumption/pixel for unshielded line → 1.1 μ W (1.3 μ A @ 0.9 V)

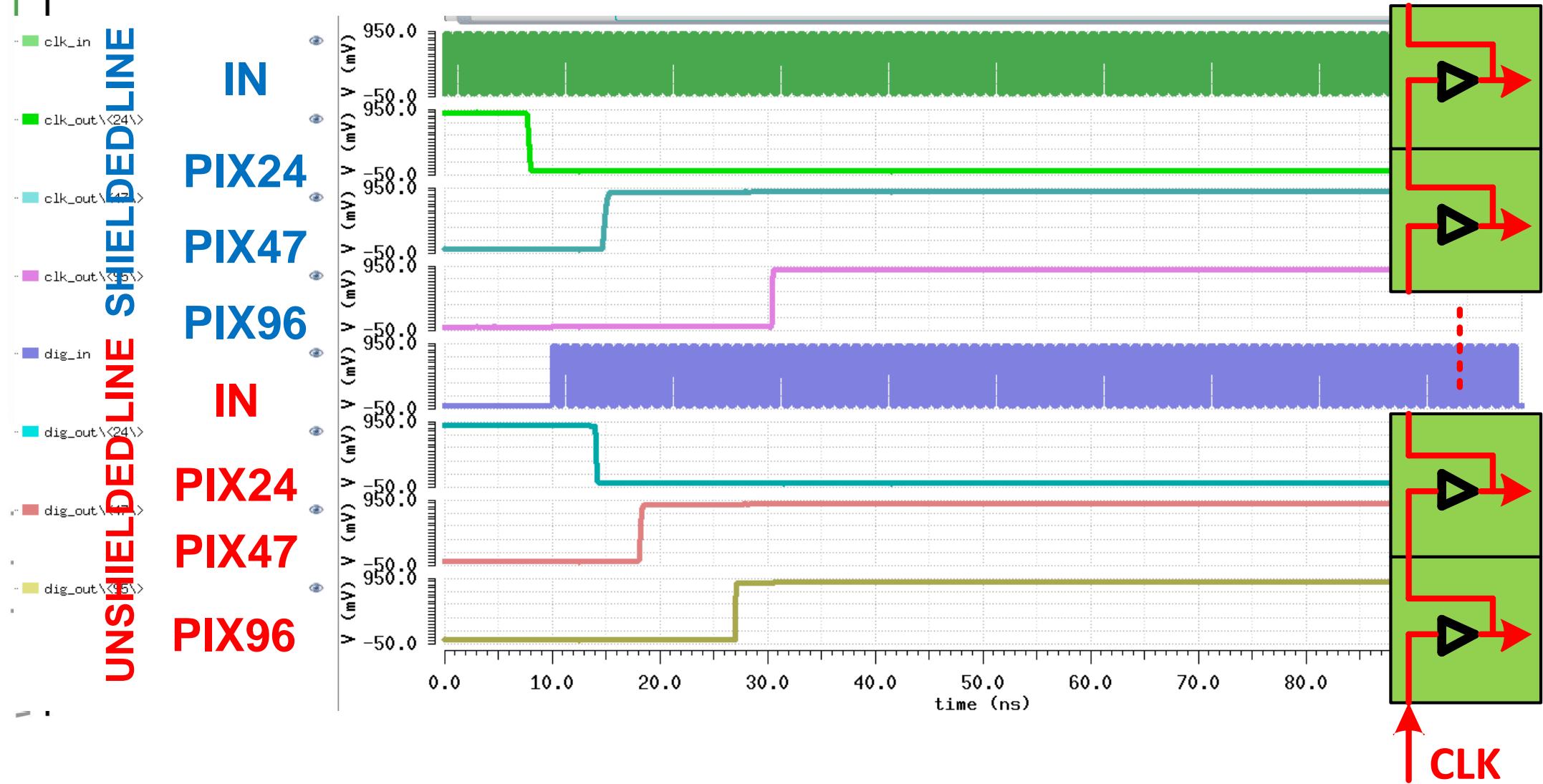
Column is driven by ideal buffer.



Postlayout simulation results of two signal propagation: clk (through the shielded line) and dig (through the unshielded line)

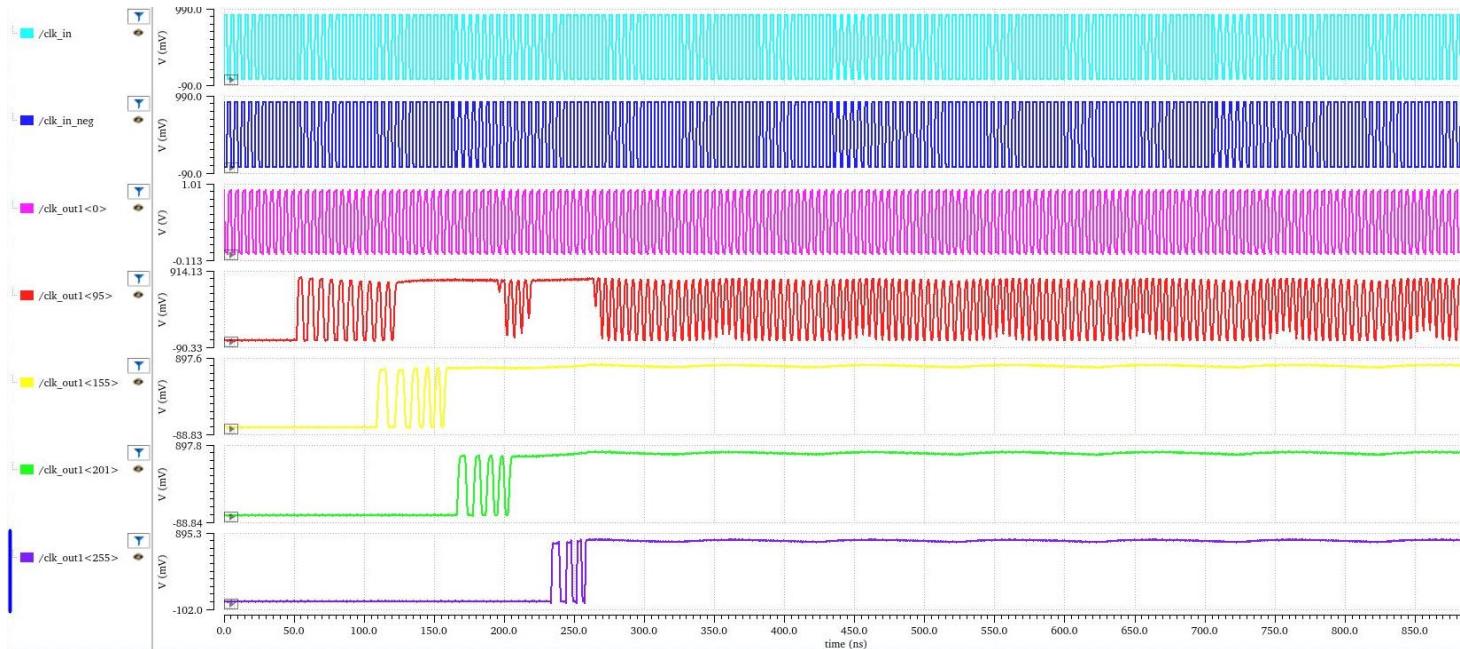
Clock signal distribution – Scenario II - 1 GHz (1 ns)

Column is driven by ideal buffer.

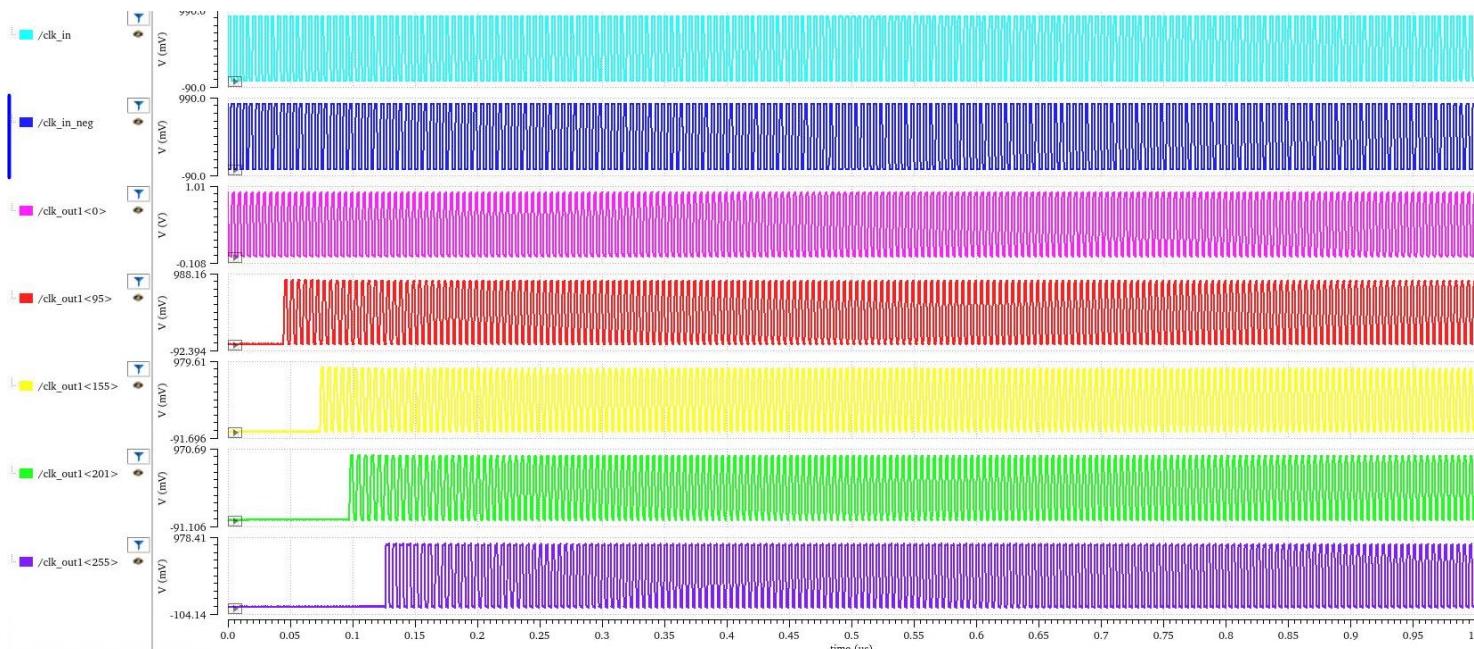


Postlayout simulation results of two signal propagation: clk (through the shielded line) and dig (through the unshielded line)

Clock distribution – postlayout simulations (200 MHz)



1 μm ME4 supply line (VSS, VDD)



5 μm ME4-ME7 supply line (VSS, VDD)

Interconnects

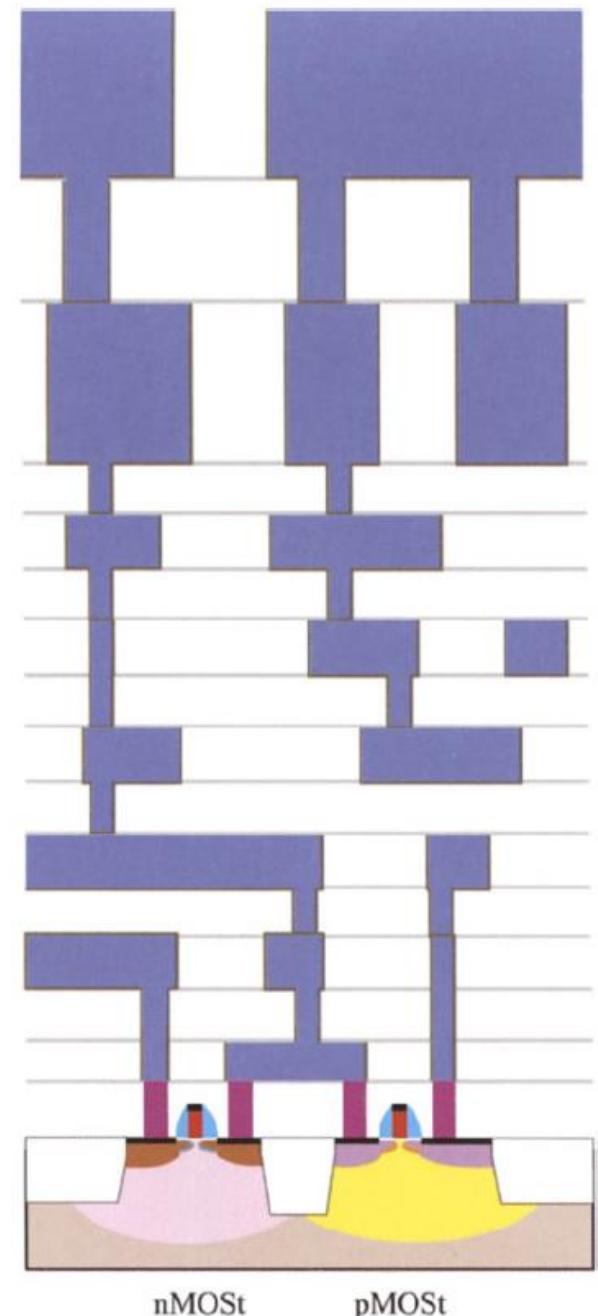
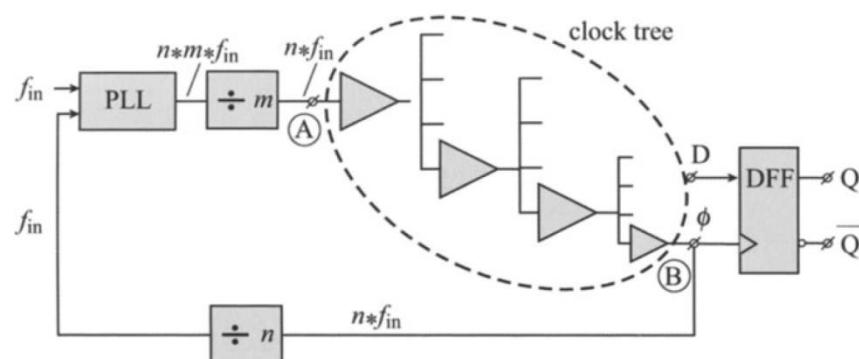
How to minimize negative effects of interconnections?

1. Modify the materials used for processing:

- a) current values for the dielectric constant are between 2.5 and 3.5 (however technology roadmap says that it will be lowered by the 2020 down to 2, beware of mechanical stress as the lower dielectric constant means the more porosity).
- b) reduce the resistance (the sheet resistance of conventional aluminium alloys is around $3 \mu\Omega cm$, while that of copper is about $1.8 \mu\Omega cm$. However, the potentials of the reduced copper resistance cannot fully be exploited as to its encapsulation by the protective layer. Its effective resistance depends mainly on the barrier material)

2. Minimize the area of the standard cells blocks.

3. As shown before the signal propagation delay over a metal wire is proportional to the square of its length. Therefore, use signal repeaters that reducing the propagation delay to linear dependence on length.



Example cross-section of 65nm CMOS process.

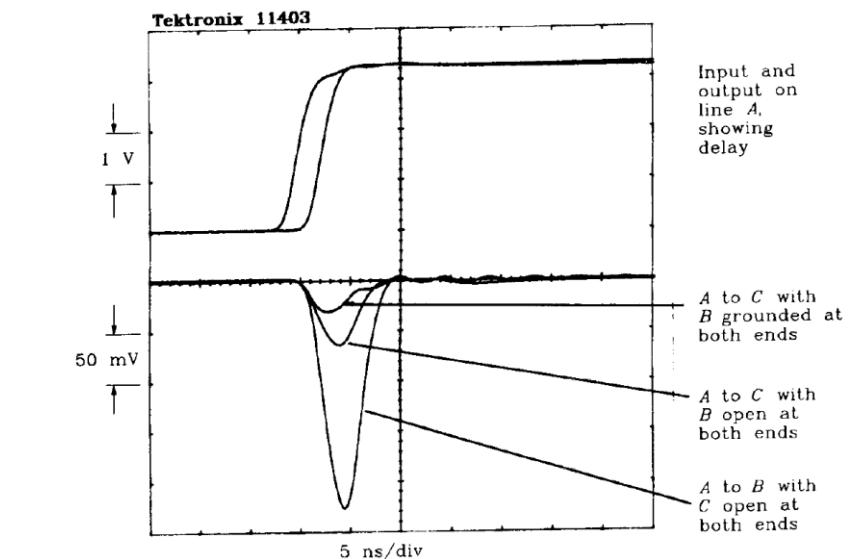
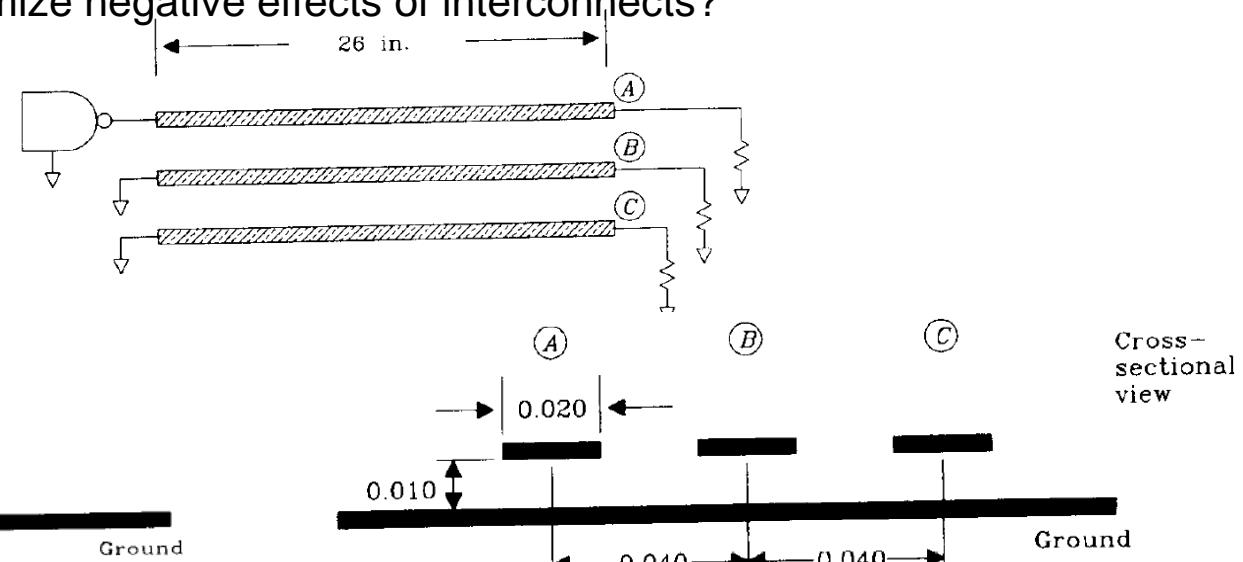
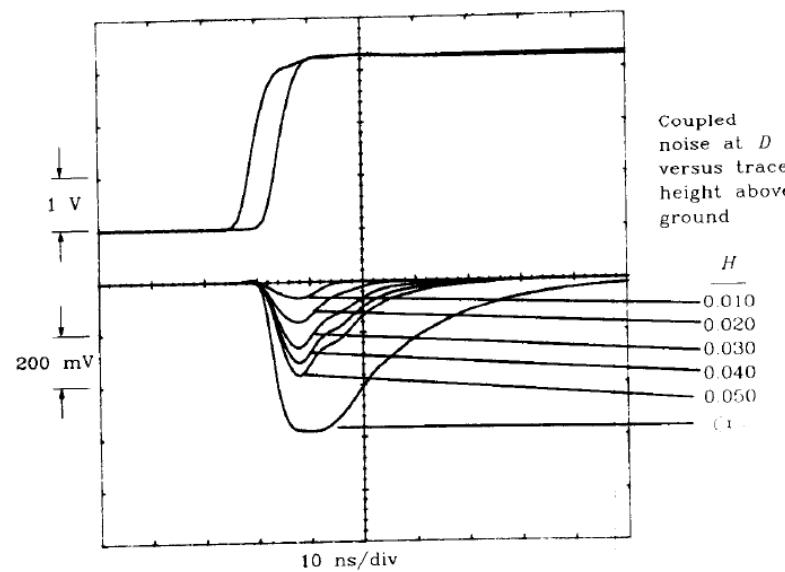
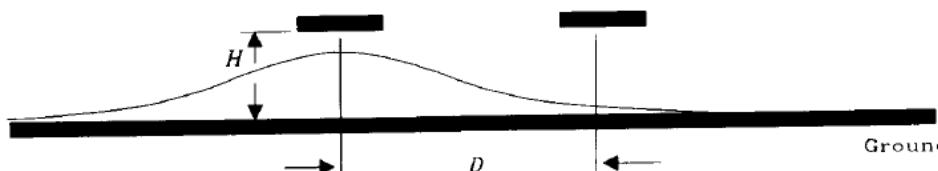


AGH

Interconnects

How to minimize negative effects of interconnects?

$$\text{Crosstalk} = \frac{K}{1 + (D/H)^2}$$



Measurement results of the cross-talks in different wire-to-wire configurations.

Separated crucial traces with neighbouring traces connected to ground.

The induced noise coupled to the neighbouring traces falls off with the square of traces mutual distance

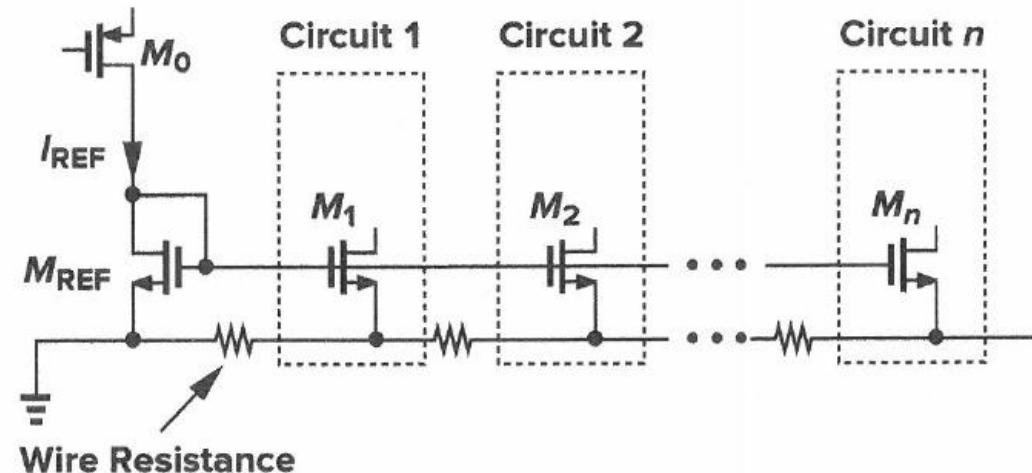
Interconnects – reference distribution

If the matching between $I_{D1}-I_{Dn}$ and I_{REF} is critical, then the voltage drop along the ground line must be taken into account.

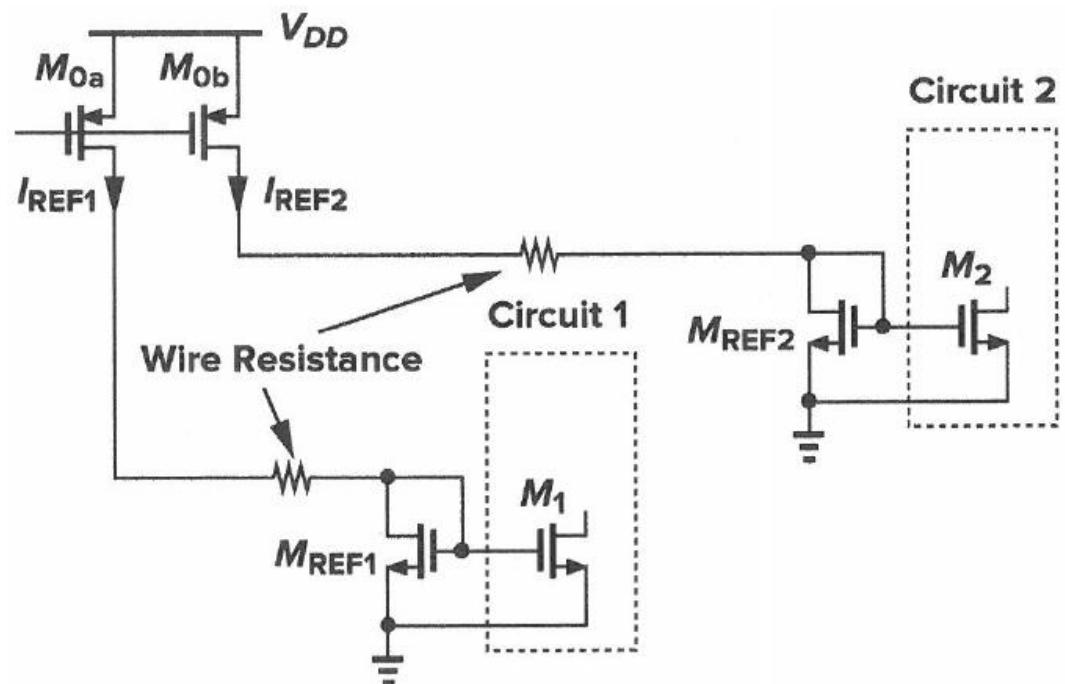
Placing the interconnect resistance in series with current sources lowers systematic error if the building blocks appear in dense groups in different region on the chip.

Mismatches between I_{REF1} and I_{REF2} and between M_{REF1} and M_{REF2} introduce error.

In large systems, it may be advantageous to employ several local bandgap reference circuits to alleviate routing problems.



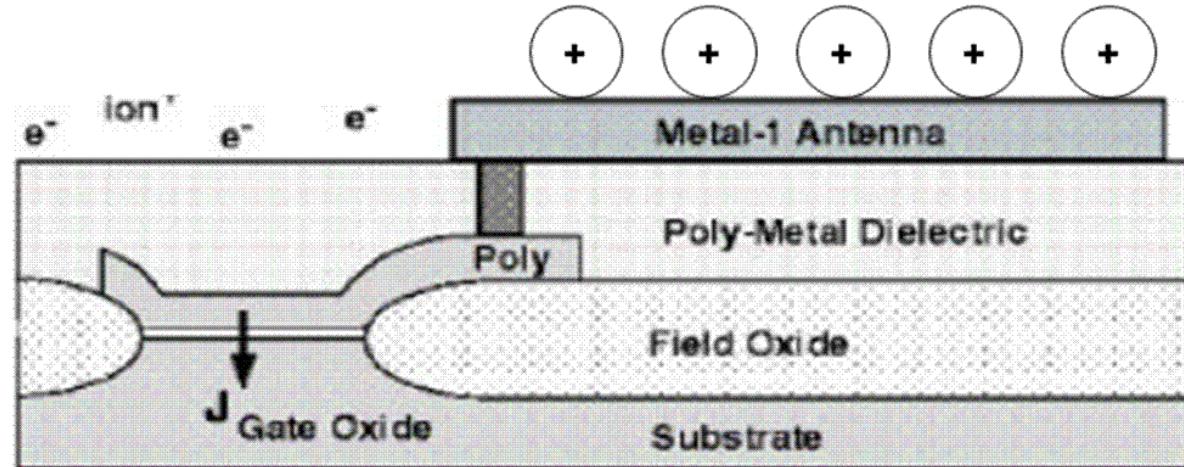
Distribution of a reference voltage for current-mirror biasing.



Distribution of current to reduce the effect of interconnect resistance.

Antenna effect

Modern wafer processing uses **Reactive Ion Etching (RIE)** or **Dry Etching** (often called ‘plasma etching’). Plasma is an ionized/reactive gas used to etch. It allows super control of pattern (shaper edges / less undercut) and also allows several chemical reactions that are not possible in traditional (wet) etch.



This effect is called as an „antenna efect” as the charge particles are beeing collected by the conductive layer (polysilicon, metal).

As a result there may be large charge deposited that finally:

- may shift the threshold voltage of the MOS transistors – affects matching of transistors,
- may brake the gate oxide irreparably – reliability is no longer guaranteed.

Due to the trend in oxide-thickness scaling the apperance of the antenna efect is expected to have greater consequences for the design.

Antenna effect

The conditions that lead to antenna formation depend on the technology used to fabricate the chip and must be determined empirically for each process. Then this is used in the DRC analysis.

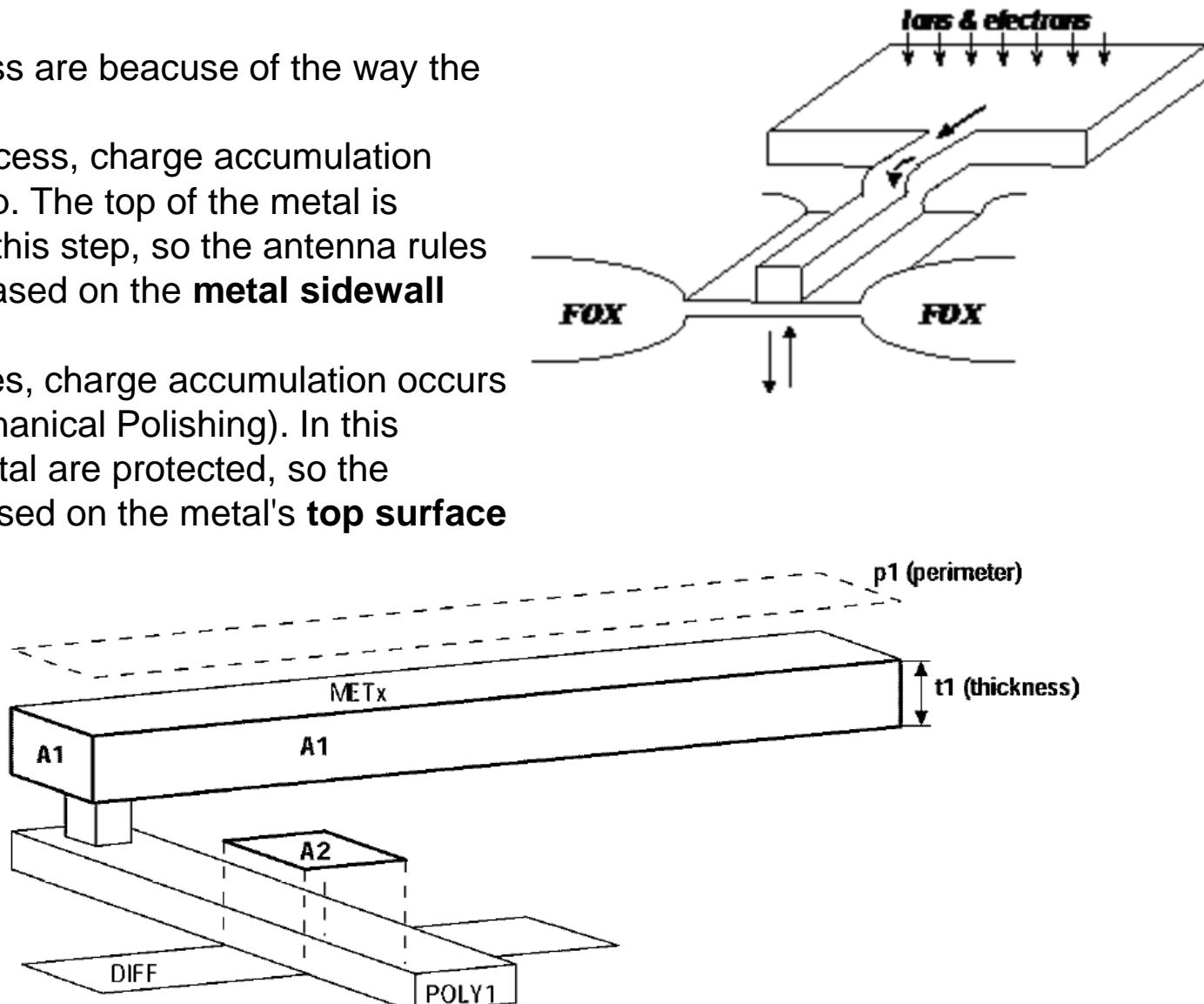
Differences among processes are because of the way the connections are formed:

- In an **aluminum**-based process, charge accumulation occurs during the ETCH step. The top of the metal is protected by a resist during this step, so the antenna rules for this process should be based on the **metal sidewall area**.
- In **copper**-base technologies, charge accumulation occurs during CMP (Chemical-Mechanical Polishing). In this process, the sides of the metal are protected, so the antenna rules need to be based on the metal's **top surface area**.

$$ratio = \frac{A_1}{A_2}$$

$$A_1 = p_1 t_1$$

$$A_2 = gate_area$$

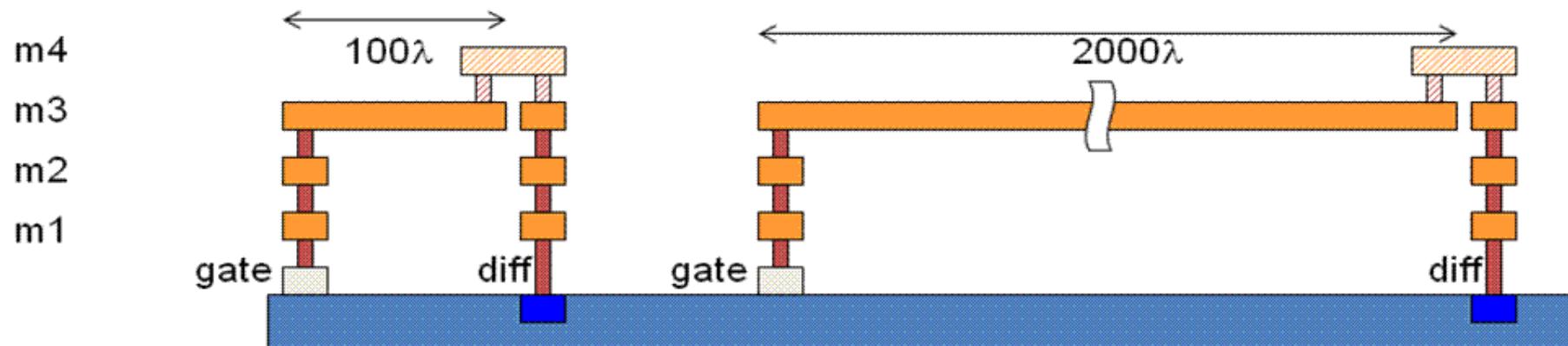
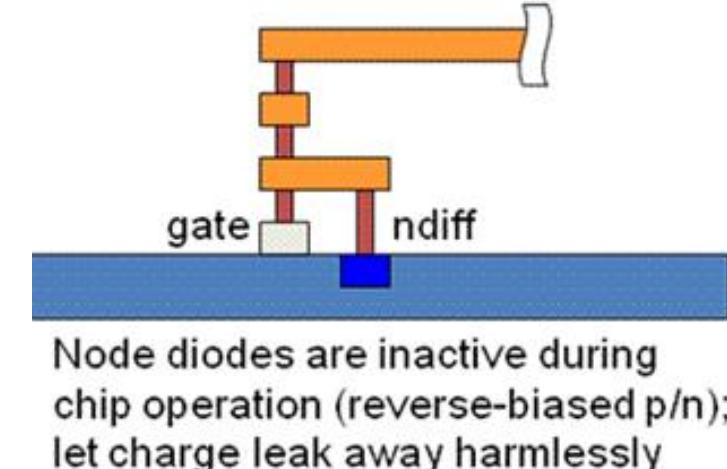


Antenna effect

How to avoid antenna effect?

Methods to avoid antenna effects:

1. Use additional diffusion – connect reverse biased diodes to the gate of the transistor.
2. Use bridges - break signal wires and route to upper metal layers by jumper insertion.
3. Use additional dummy transistors - addition of extra gates will reduce the area to area ratio.



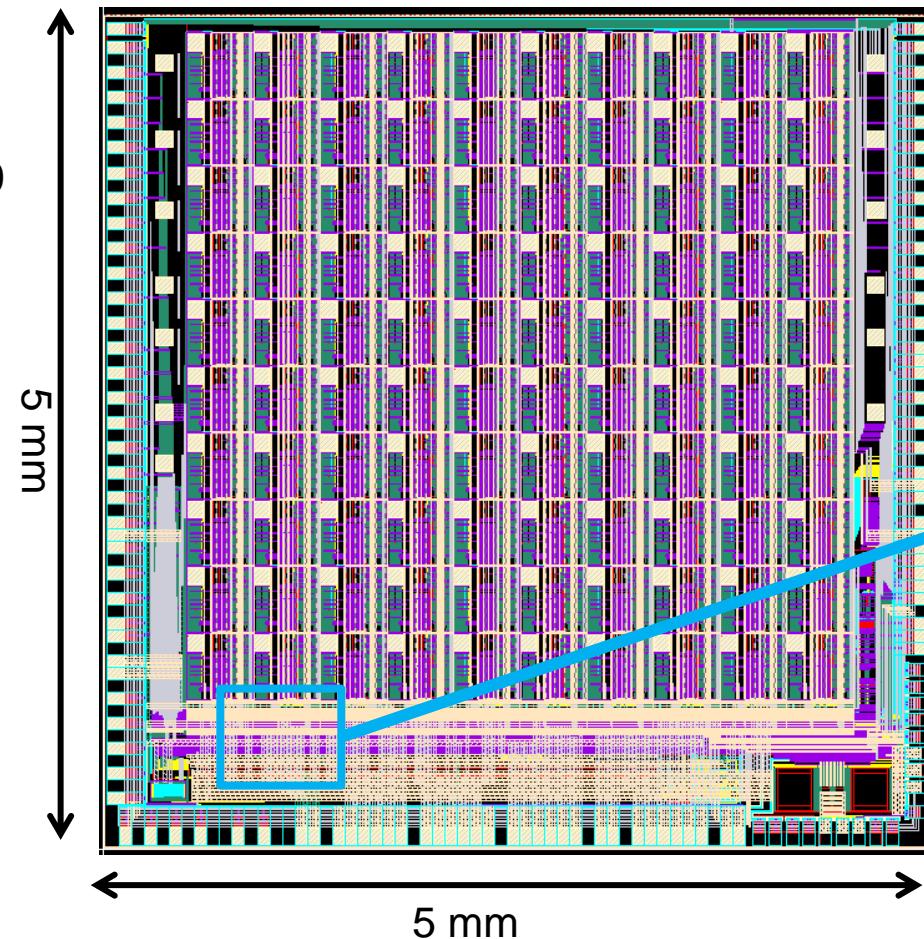
Safe: m3 is too short to accumulate very much charge; won't kill gate

Dangerous: lots of m3; will probably accumulate lots of charge and then blow oxide

Antenna effect

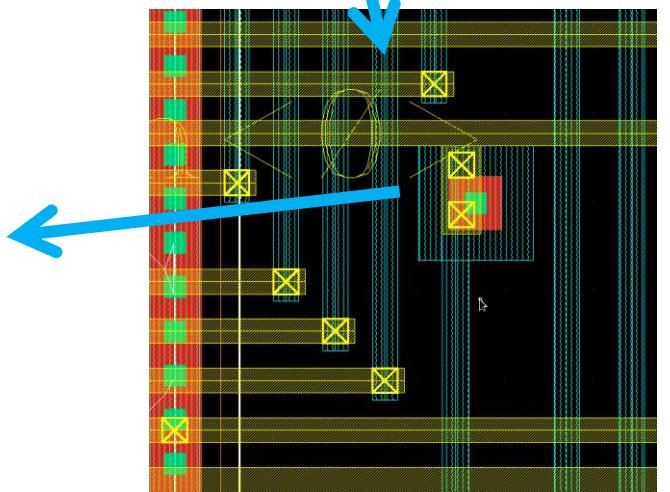
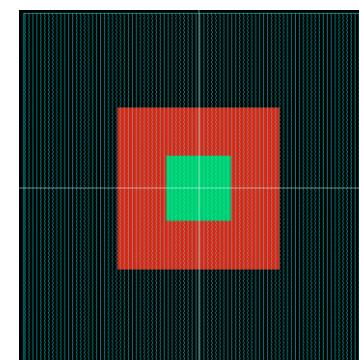
How to avoid antenna effect?

N100 – Neuro 100
10x10 matrix
combined of
recording and
stimulation
UMC180nm
CMOS



The antenna violation is avoided by the substrate contact build of:

1. DIFF
 2. **NPLUS !!!**
 3. ME1
 4. CON

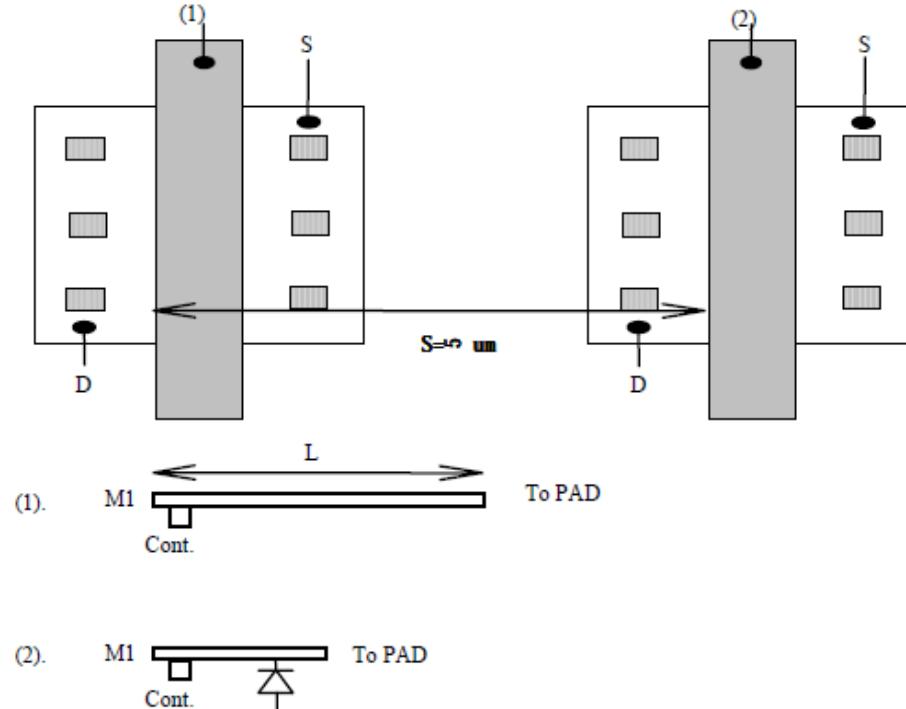


Antenna effect

What happens if the antenna effects are waved?

The best scenario is deterioration of voltage threshold uniformity from transistor to transistor.

The UMC180nm datasheet shows that these may be more than 10 times worse.



W/L=10/0.34

	rDVth(mV)	rDId@Vg =Vd=-0.7V(%)	rDId@Vg =Vd=-1.0V(%)	rDId@Vg =Vd=-1.2V(%)	rDId@Vg =Vd=-1.8V(%)
10/0.34 L=2um	4.90	4.28	1.84	1.19	1.23
10/0.34 L=100um	7.11	5.44	2.31	1.58	1.09
10/0.34 L=1000um	15.05	16.0	8.74	7.38	5.14

Both transistors with protection diode

10/0.34	2.22	2.03	0.95	0.73	0.42
---------	------	------	------	------	------

Antenna effect

The antenna ratio (AR) of the perimeter is defined as the ratio of the perimeter of a poly or metal layer to the active poly gate perimeter.

1.A. If the protection diode is not used, the maximum antenna ratio of

- a. poly layer: AR 200
- b. single metal layer for M1~M5: AR 800
- c. single metal layer for M6: AR 400

1.B. If the protection diode of single polarity (N+/P-Sub or P+/N_WELL) is used, the maximum antenna ratio of

- a. single metal layer for M1~M5: AR 10000
- b. single metal layer for M6: AR 5000

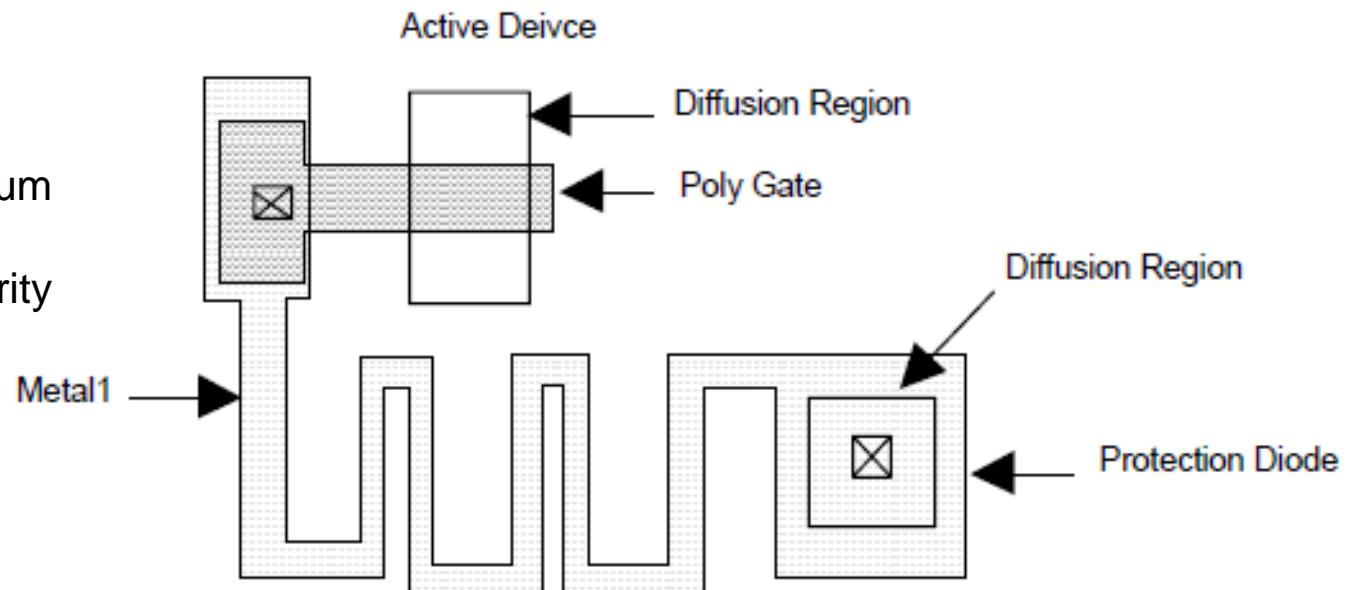
1.C. If the protection diode of dual polarity (N+/P-Sub and P+/N_WELL) is used, the maximum antenna ratio is unlimited.

Minimum DIFFUSION diode area

$0.6 \times 0.6 \mu\text{m}^2$

Be aware that there are minimum protective diodes areas.

Also the both diodes of dual polarity has the best protection.



Electromigration

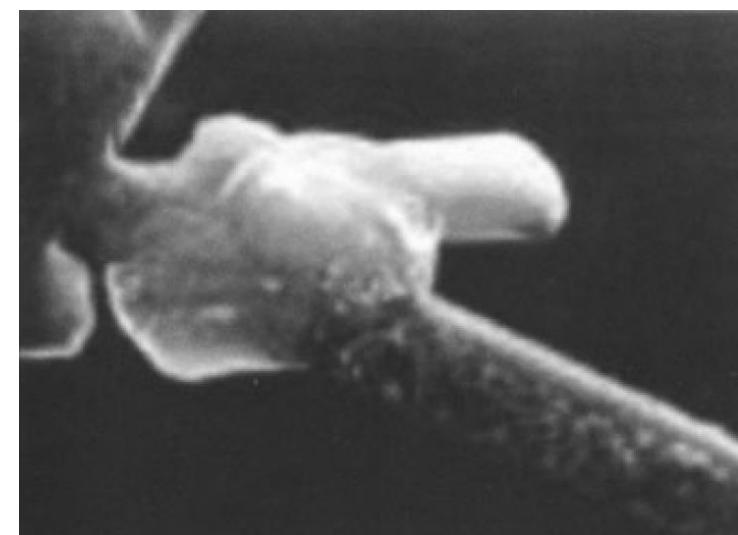
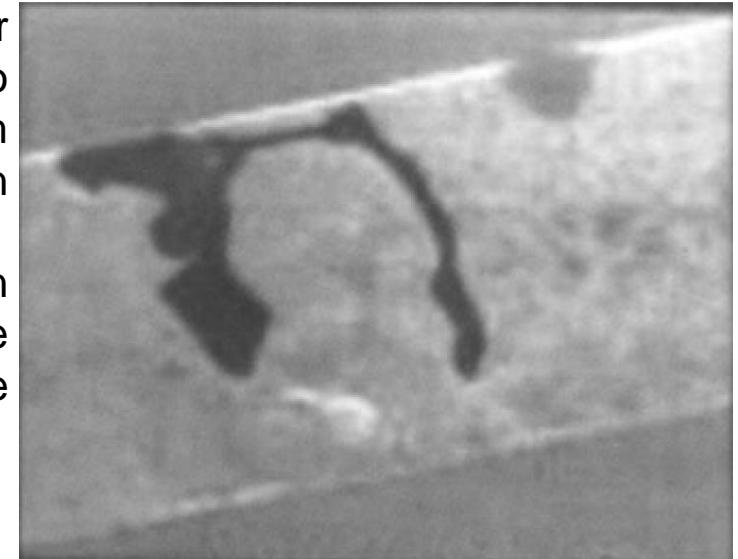
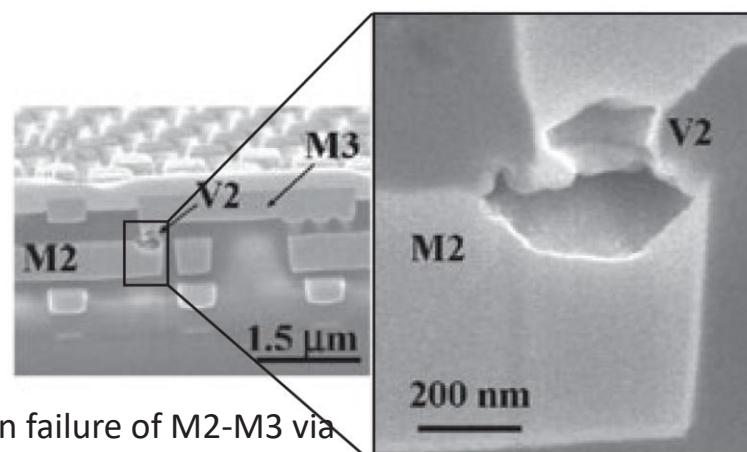
At high current densities the metal atoms (Aluminium or Copper) tend to „migrate” that eventually grows to discontinuity. Therefore, in order to assure the long term IC reliability the proper metal widths have to be taken into account.

A rule of thumb is the current-density of 1 mA per micron of width of a DC current. AC current requirements are eased from DC current (almost one order of magnitude larger) as these flow in two directions.

The minimum allowed width W_{em} of a metal wire with height H to carry current I , according to electromigration requirement, is equal to:

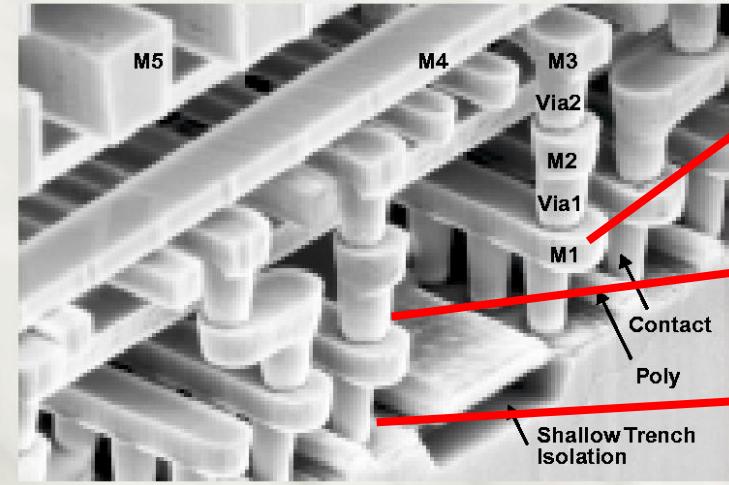
$$W_{em} = \frac{I}{J_{MAX} H}$$

J_{MAX} is maximum allowed current density



Electromigration damage in metal interconnect lines: voids (a); hillocks (b)

Electromigration



One should be also aware the electromigration may be problematic issue in the vias and that it depends on the circuits' temperature.

Also, despite the fact the modern processes use lower supply voltages and its standard cells become smaller the power density increases as these blocks' current consumption does not decrease a lot.

DC Rules:

- A0. METAL1 Jmax_dc (width < 1um)
- A. METAL1 Jmax_dc (1um ≤ width < 10um)
- B. METAL1 Jmax_dc (10um ≤ width < 20um)
- C. METAL1 Jmax_dc (width ≥ 20um)

(unit : mA/um)

	80°C	100°C	125°C
Jmax_dc (width < 1um)	2.18	1.18	0.55
Jmax_dc (1um ≤ width < 10um)	1.48	0.87	0.44
Jmax_dc (10um ≤ width < 20um)	2.66	1.56	0.80
Jmax_dc (width ≥ 20um)	2.66	1.56	0.80

- MVIA1 Jmax_dc
- MVIA2 Jmax_dc
- MVIA3 Jmax_dc
- MVIA4 Jmax_dc
- MVIA5 Jmax_dc

(unit : mA/Via)

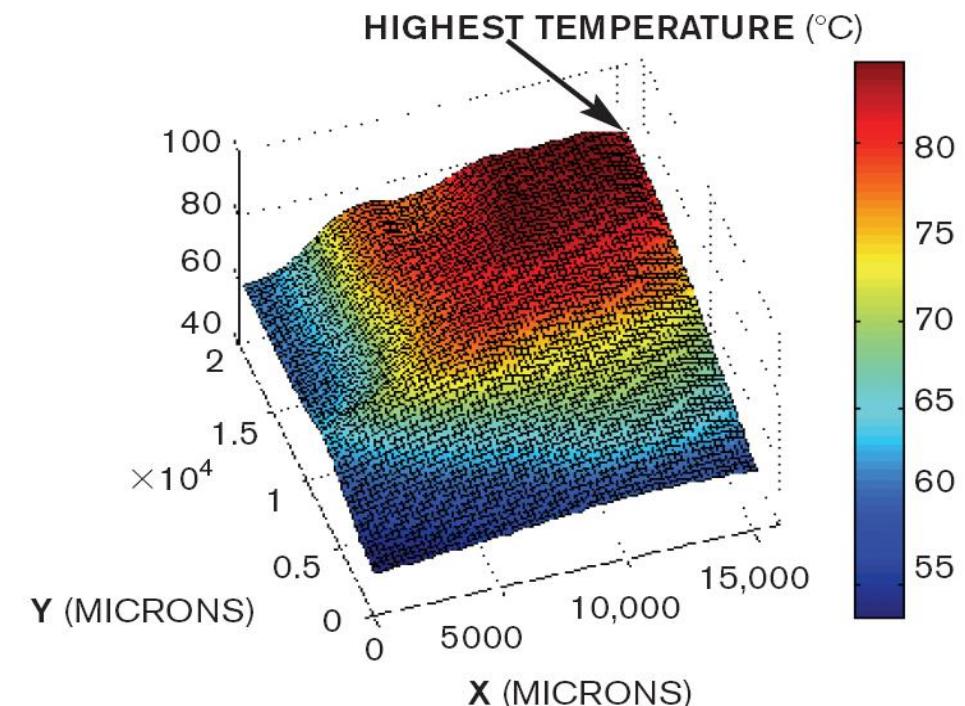
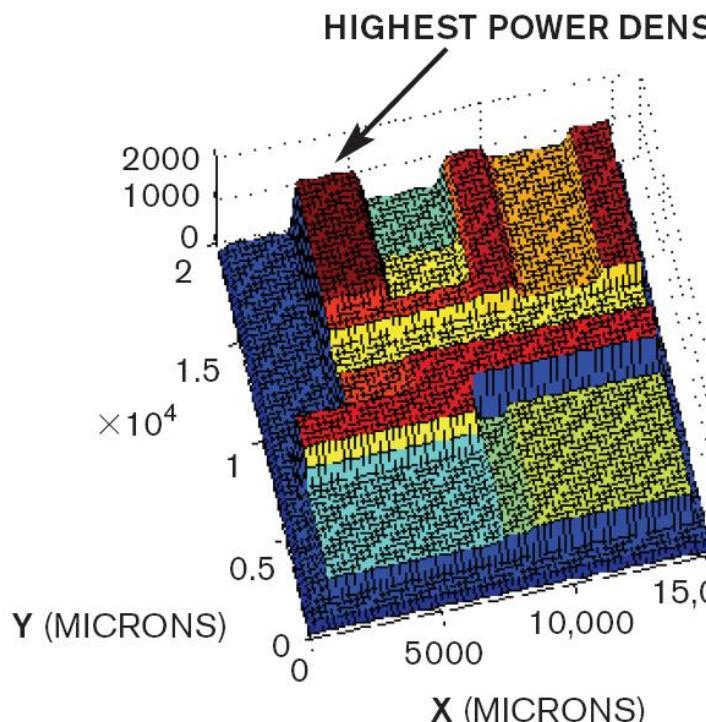
	80°C	100°C	125°C
Jmax_dc (width < 1um)	0.76	0.41	0.21
Jmax_dc (1um ≤ width < 10um)	0.76	0.41	0.21
Jmax_dc (10um ≤ width < 20um)	0.76	0.41	0.21
Jmax_dc (width ≥ 20um)	0.76	0.41	0.21

CONTACT Jmax_dc

Table 1. Design Rules for AC Current

AC operation status	@80°C	@100°C	@125°C
Uni-directional current	Average current density $ J _{ave}$ replaces J_{max_DC}	Average current density $ J _{ave}$ replaces J_{max_DC}	Average current density $ J _{ave}$ replaces J_{max_DC}
Bi-directional current	$ J _{ave} \leq 20 \text{ mA}/\mu\text{m}$ for metal line; $ J _{ave} = 6 \text{ mA}/\text{Contact}; 7 \text{ mA}/\text{Via}$	$ J _{ave} \leq 20 \text{ mA}/\mu\text{m}$ for metal line; $ J _{ave} = 3 \text{ mA}/\text{Conatct}; 4 \text{ mA}/\text{Via}$	$ J _{ave} \leq 20 \text{ mA}/\mu\text{m}$ for metal line; $ J _{ave} = 3 \text{ mA}/\text{Contact}; 4 \text{ mA}/\text{Via}$
Single pulse peak current	$\leq 20 \text{ mA}/\mu\text{m}$ for metal line; $6 \text{ mA}/\text{contact}; 7 \text{ mA}/\text{VIA}$	$\leq 20 \text{ mA}/\mu\text{m}$ for metal line; $3 \text{ mA}/\text{contact}; 4 \text{ mA}/\text{Via}$	$\leq 20 \text{ mA}/\mu\text{m}$ for metal line; $3 \text{ mA}/\text{Contact}; 4 \text{ mA}/\text{Via}$

Exemplary IC Power and Temperature Profiles



At 90-nm-process nodes, leakage accounts for 25 to 40% of total power. At 65-nm processes, leakage accounts for 50 to 70% of total power.

On-chip temperature impacts timing. Every 15C increase causes delay of approximately 10 to 15%.

EM increases exponentially with temperature increases and reduces the life of products by four times.

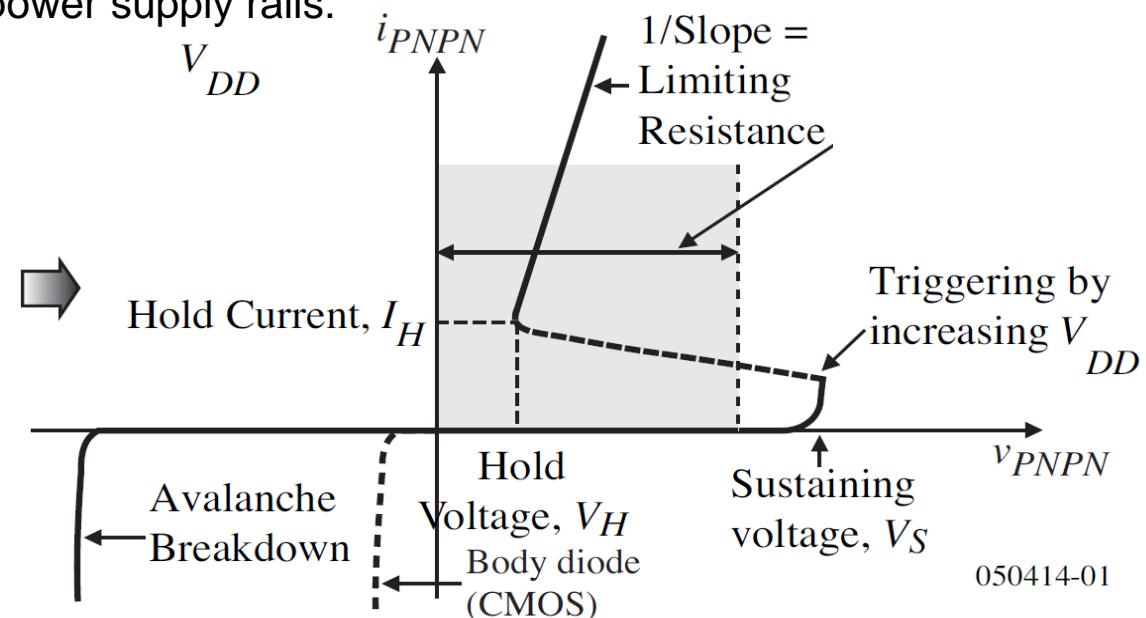
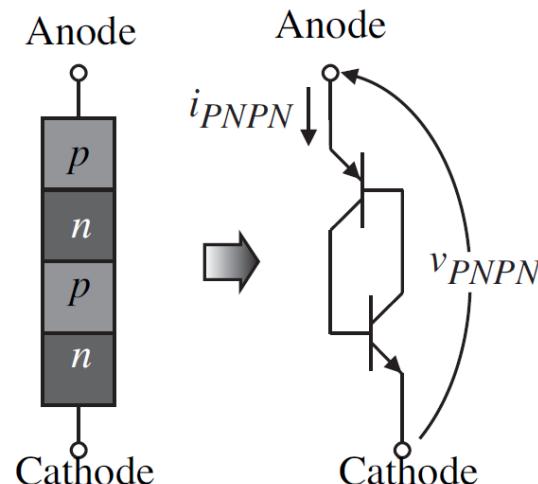
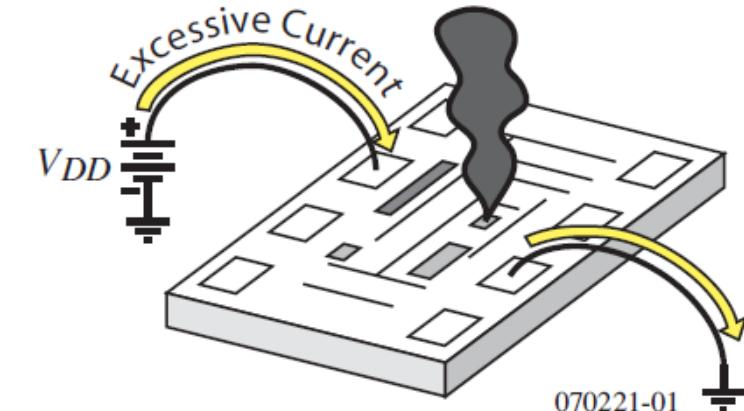
Resistance is a linear function of temperature, affecting IR drop. A change of 15C increases resistance by 10%.

Latch-up

Latch-up is the creation of a low impedance path between the power supply rails.

Latch-up is caused by the triggering of parasitic bipolar structures within an integrated circuit when applying a current or voltage stimulus on an input, output, or I/O pin or by an over-voltage on the power supply pin.

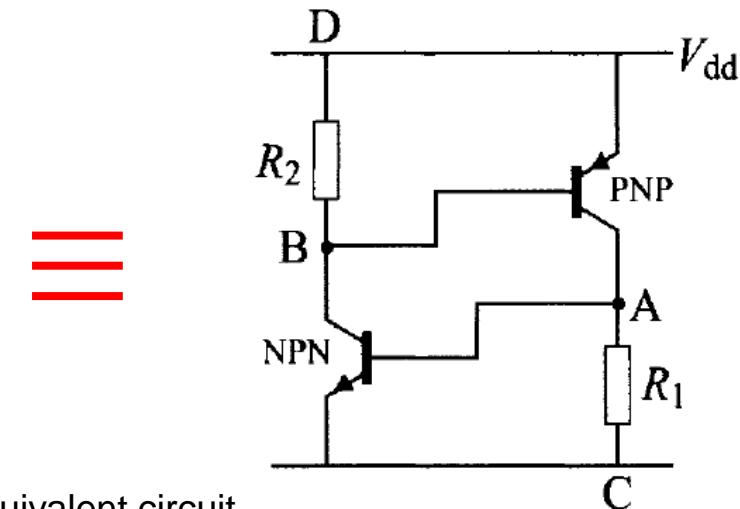
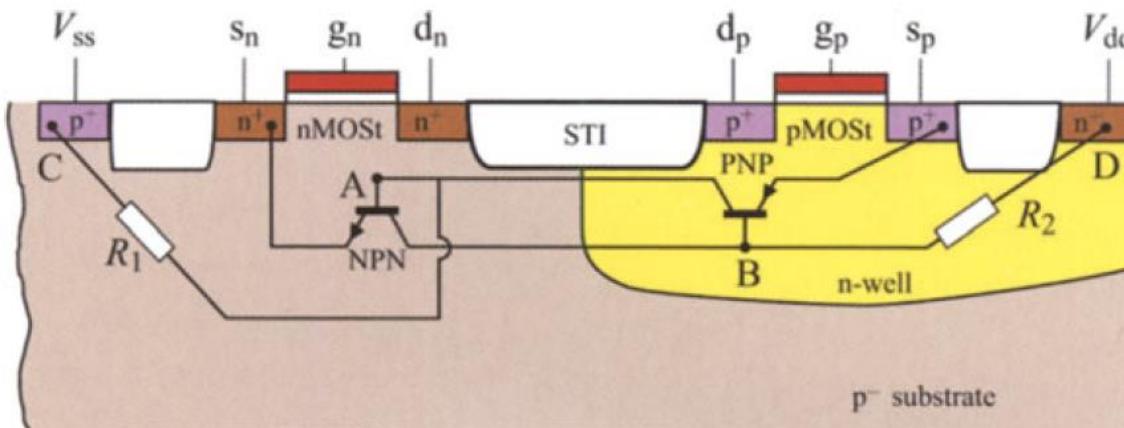
A true latch-up remains after the stimulus has been removed and requires a power supply shut down to remove the low impedance path between the power supply rails.



Latch-up origin explanation.

Is the V_S limit an only way to induce the latch-up??

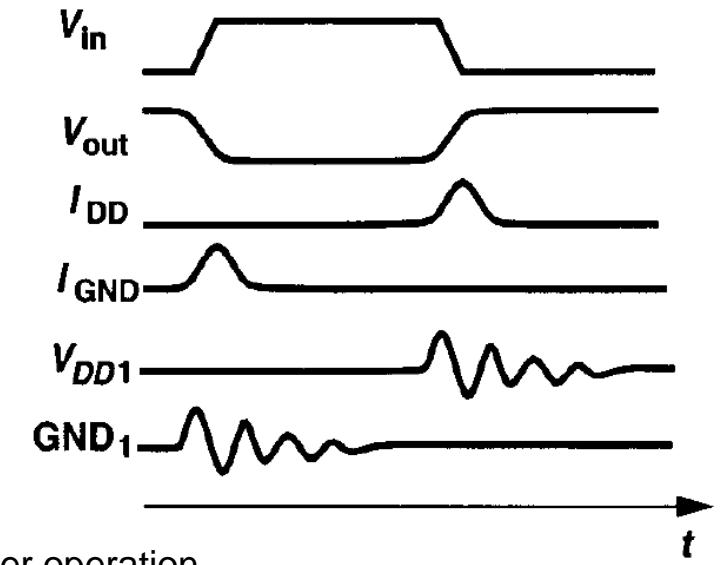
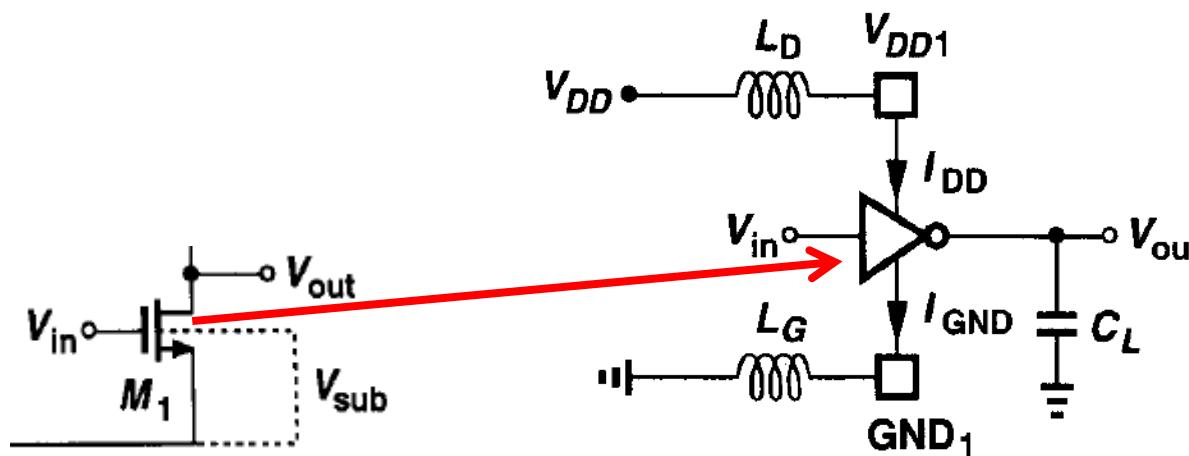
Latch-up



Cross section of the inverter and its parasitics equivalent circuit

Latchup may be triggered by the voltage bounces on the supply lines.

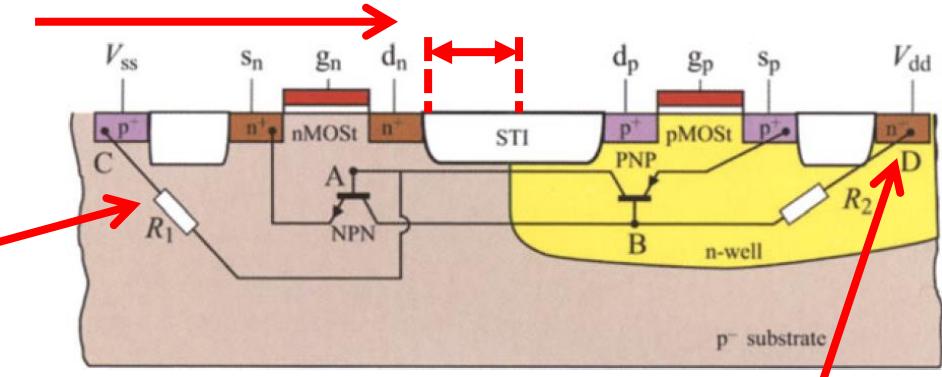
It also may be caused by the signals provided to the input/outputs of the IC pads (especially this is a common case in large digital outputs buffers that inject high currents into the substrate trough the large drain junction capacitances).



Exemplary latch-up origin as a result of the inverter operation.

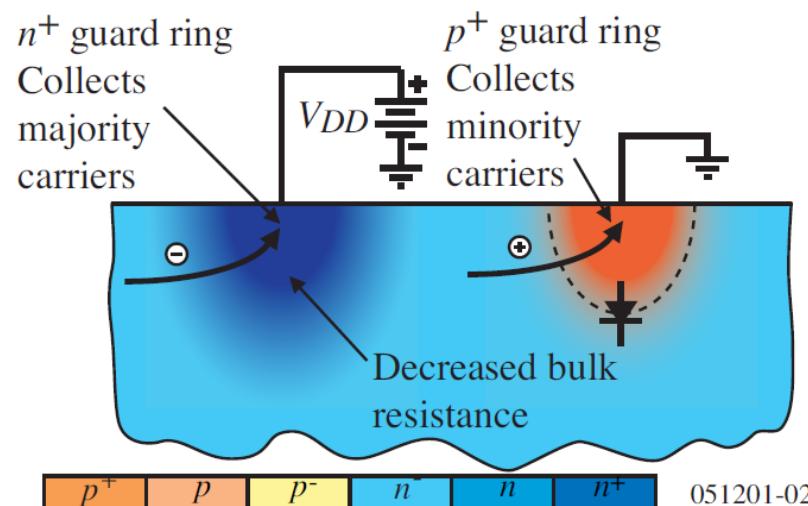
Latch-up Prevention

1. Keep the source/drain of the MOS device not in the well as far away from the well as possible. This will lower the value of the BJT betas.
2. Reduce the values of R_1 and R_2 . This requires more current before latch-up can occur.

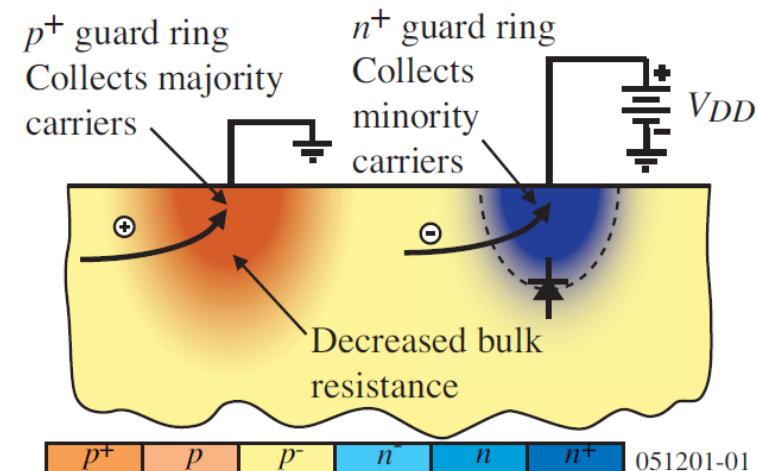


3. Surround the transistors with guard rings. Guard rings reduce transistor betas and divert collector current from the base of parasitic bipolar transistors.

Guard rings in *n*-material:



Guard rings in *p*-material:



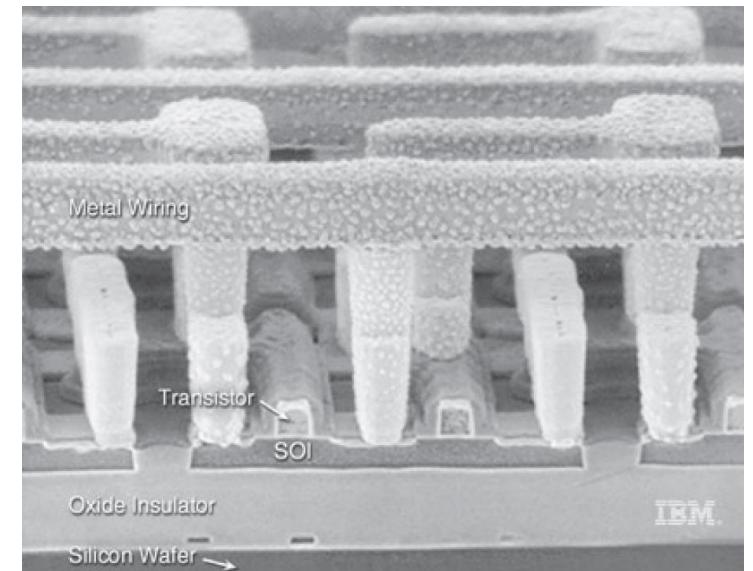
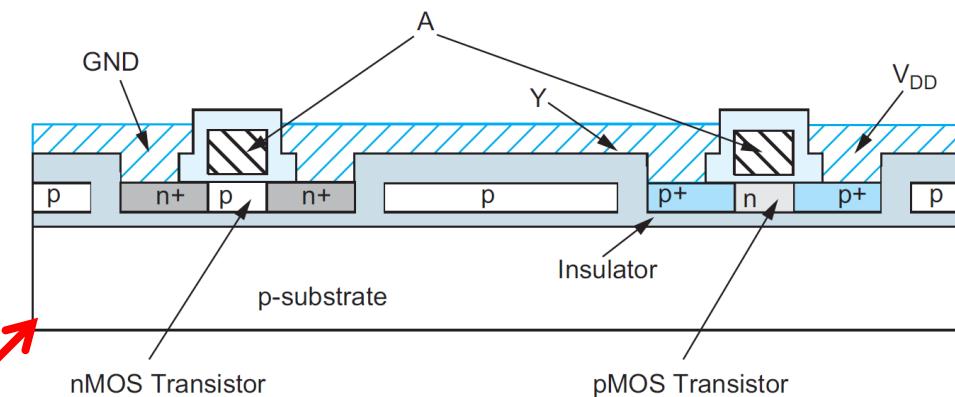
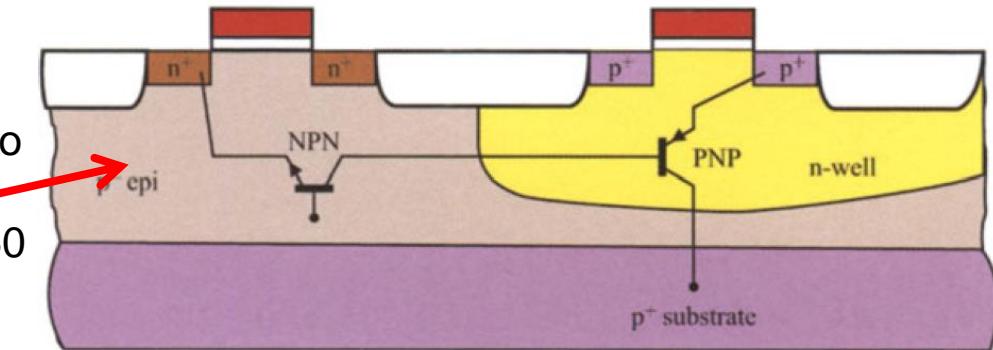
Guard rings are used to collect carriers flowing in the silicon. They can be designed to collect either majority or minority carriers.

AGH 4. The use of so-called epitaxial wafers (technology modification). The epitaxial layer, in which the devices are formed has thickness of 1 to 5 μm , is doped and has a resistivity of $\sim 10\text{-}20 \Omega\text{cm}$. Because the wafer thickness is typically 750 μm , the p+-substrate is relatively thick and has a low resistivity ($\sim 5\text{-}10 \text{m}\Omega\text{cm}$). Such low-ohmic substrates show very low values for R_1 . Because the latch-up effect decreases with reducing voltages, CMOS in 120 nm and beyond are most commonly processed on high-ohmic wafers.

5. Use Silicon-On-Insulator technology to completely isolate the nMOS transistors from the pMOS transistors (technology modification). In this technology the NPN and PNP transistors are completely isolated from one another and so the connections to create latching thyristor circuits are missing.

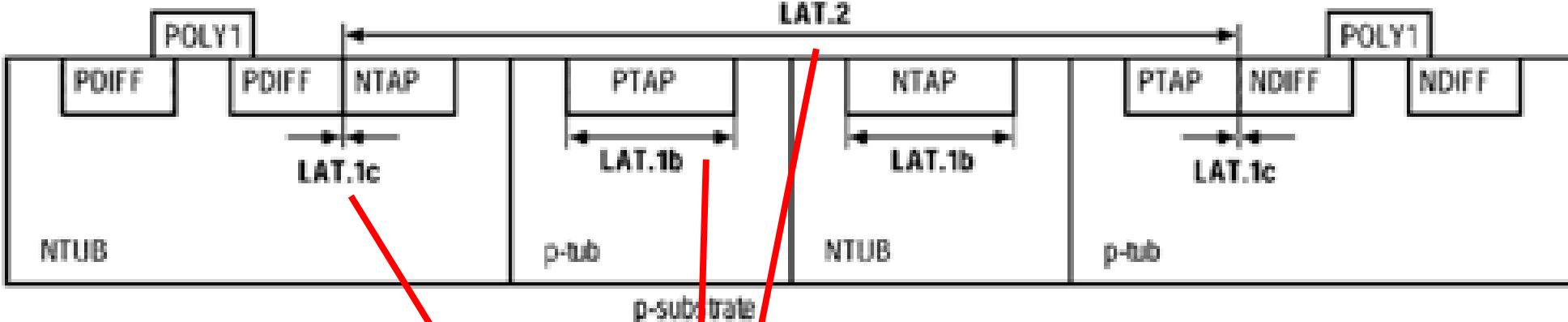
6. Lowering the supply voltage in the modern processes helps in latch-up prevention however there are often input/output circuits supplied from higher voltages.

Latch-up Prevention

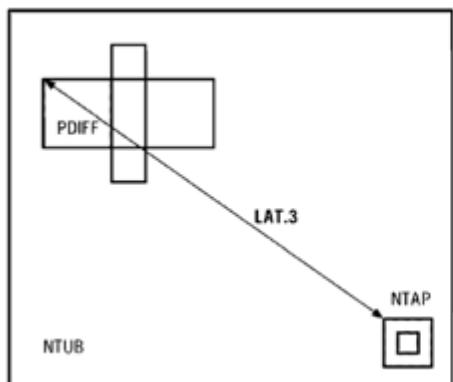


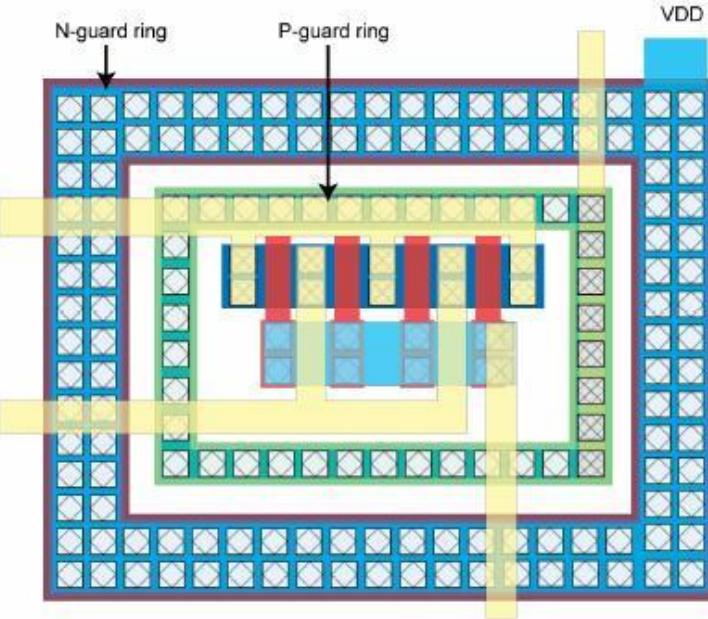
IBM SOI process electron micrograph.

Latch-up Prevention

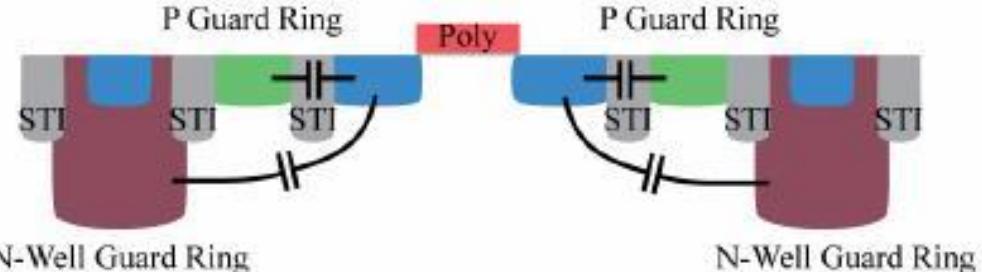


Guideline	Description	Value [um]
LAT.1a	A double guard ring structure should be inserted in between NMOS and PMOS of I / O buffers	
LAT.1b	Minimum PTAP and NTAP guard ring width for I / O buffers	3
LAT.1c	Maximum distance from PTAP or NTAP guard ring to source DIFF for I / O buffers	2
LAT.2	Minimum NMOS to PMOS spacing for I / O buffers and ESD devices Active DIFF area in this spacing is not allowed.	40
LAT.3	Maximum distance from any point inside source / drain DIFF to the nearest TAP DIFF of the same NTUB or PSUB.	20
LAT.4	A guard ring structure with NTUB pseudo-collector and PTAP should be inserted between I / O buffers and internal circuit area	
LAT.5	Minimum I / O buffer to internal circuit spacing	50
LAT.6	Any HOT_NDIFF area connecting to I / O pads should be surrounded by double guard ring.	
LAT.7	Any NTUB without direct connection to VDD and with HOT_NDIFF inside it should be surrounded by double guard ring.	
LAT.8	For special devices such as bipolar transistor, diode, resistor, or special circuits such as charge pump, power regulator, high noise or high power circuitry, a double guard ring should be inserted surrounding and between them.	
LAT.9	All the guard rings and pickups should be connected to VDD / VSS with very low series resistance. That is, NTUB should be tied together with NTAP, and DIFF should be tied together with contacts and metal to VDD / VSS. As many as possible CONT should be used.	



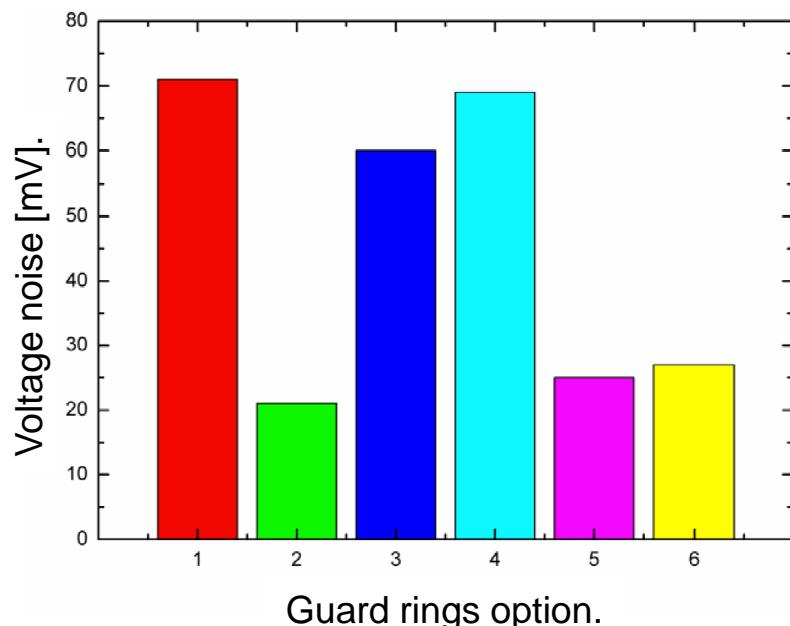


Guard Rings



Top view and a cross section of the transistor equipped with the guard rings.

An important layout practise is to ensure that there is no (or very little) current flowing through any part of the guard ring. There are some disadvantages of the guard rings like, guard rings take up a lot of area and they also add capacitive load to the transistor. So, if the area in the design is utmost importance, the guard rings should be avoided.



Voltage noise dependence for different guard rings configurations: 1 – no guard rings, 2 – p+ guard ring connected to negative supply line, 3 – N - well guard ring connected to GND, 4 - N - well guard ring connected to the positive supply line, 5 – two guard rings (p+ connected to GND, N - well connected to GND), 6 – two guard rings (p+ connected to GND, N - well connected to positive supply line).

Electrostatic Discharge - ESD

ICs consist of components that are very sensitive to excess current and voltage above the nominal power supply.

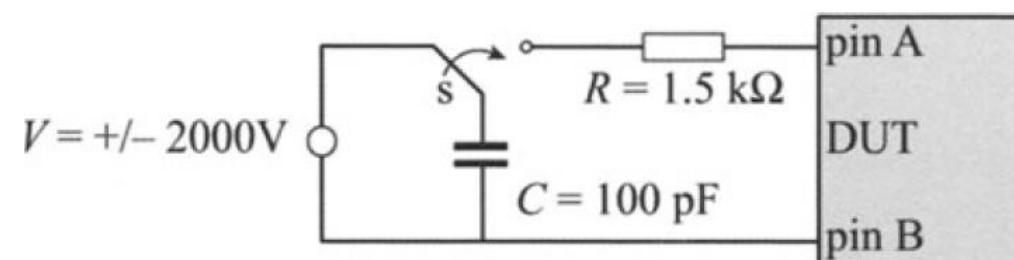
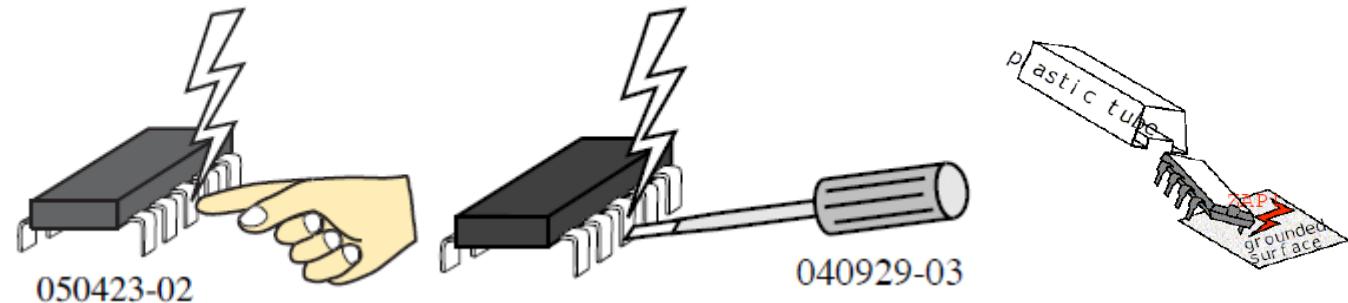
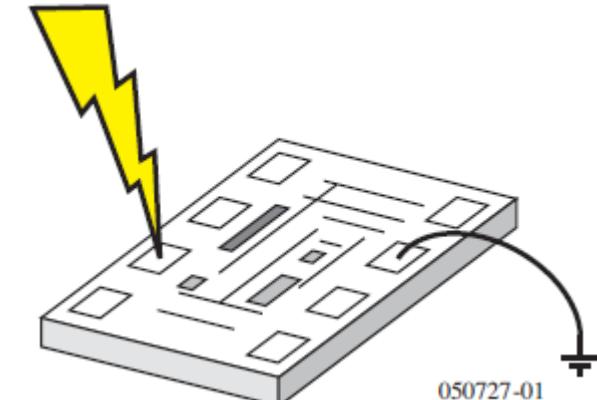
Any path to the outside world is susceptible to ESD.

ESD damage can occur at any point in the IC assembly and packaging, the packaged part handling or the system assembly process.

Note that power is normally not on during an ESD event.

To develop protection different verification models are being build:

1. Human body model (HBM).
2. Machine model (MM).
3. Charge device model (CDM).



A typical equivalent circuit based on the human-body model.

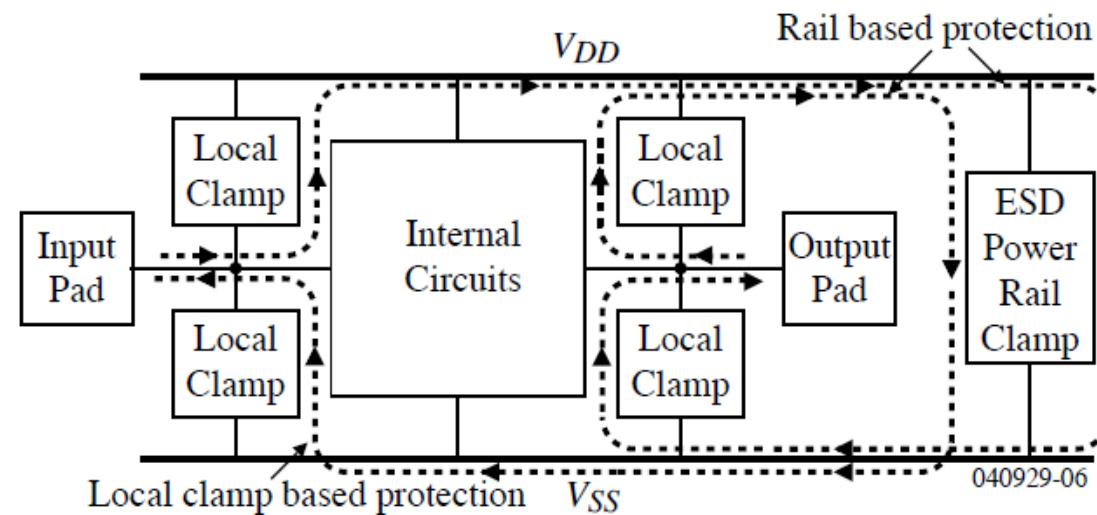
Depending on the environment the voltage across the capacitance ranges from hundred volts to several thousands volts. As a result an ESD event typically creates very high values of current (1-10A) for very short periods of time (150 ns) with very rapid rise times (1ns).

Therefore, components experience extremely high values of current with very little power dissipation or thermal effects.

Resistors, Capacitors, Diodes and Transistors may be permanently damaged.

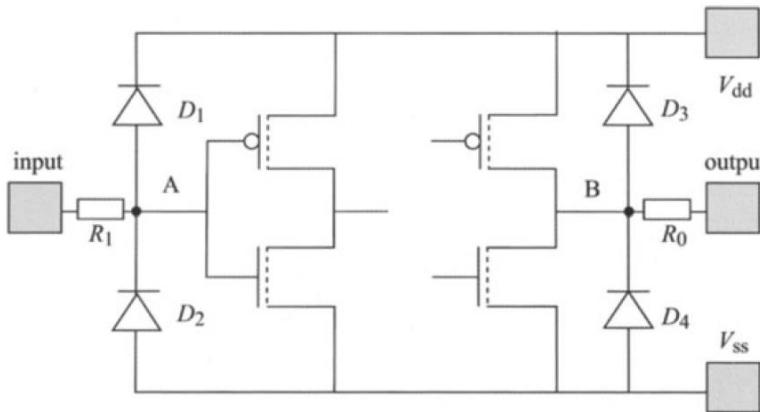
MOS devices sustain two types of permanent damage:

- Gate oxide may breakdown if the electric field exceeds roughly 10^7 V/cm (e.g. 10 V for oxide thickness of 100 Å, i.e. 10 nm) → leads to the very low gate resistance,
- source/drain junctions may melt if they carry large reverse/forward current → short to bulk is created.

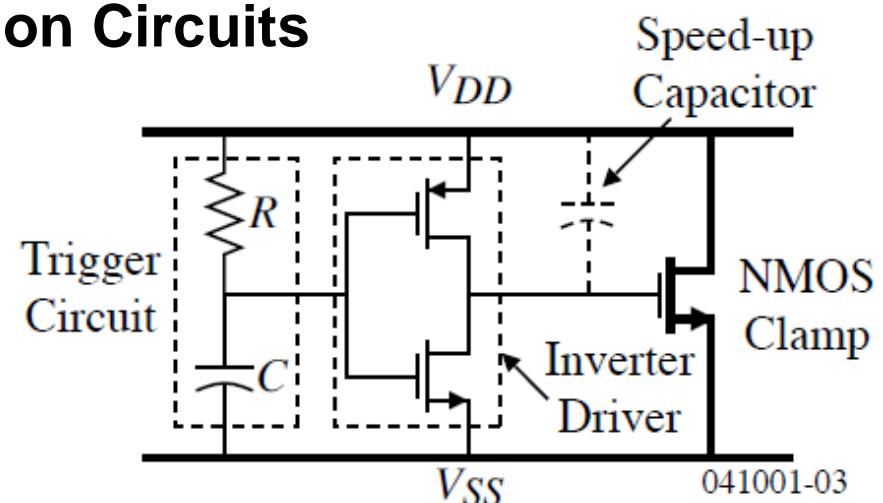


There is a need to provide low impedance path for this charge to protect internal electronics.

ESD Protection Circuits



Simplest form of a MOS input and output protection circuit.



Power rail clamp.

In its simplest form, a protection circuit consists of a spike filter and a set of diodes.

In the figures above:

- resistors R_1, R_2 forms a voltage filter with the parasitic capacitances on the A, B nodes,
- diodes are responsible for providing the low impedance path whenever input/output voltage exceeds V_{DD}/V_{SS} .
- very fast and high voltages develop voltage drop-out on the resistance R and in that way the NMOS Clamp is switched ON.

Values for the input and output resistors are in the order of 100Ω and $5 k\Omega$, respectively.

Not always the ouput ESD protection is necessary, however one needs to take a closer look if the ouput fullfills the electromigration requirements.

The ESD protection circuits involves three critical issues:

- large parasitic capacitances introducion (even several picofarads) → speed degradation, impedance lowering,
- its parasitic capacitances may couple the noise on the supply lines to the critical analog inputs,
- if not properly designed, may cause latch-up.



AGH

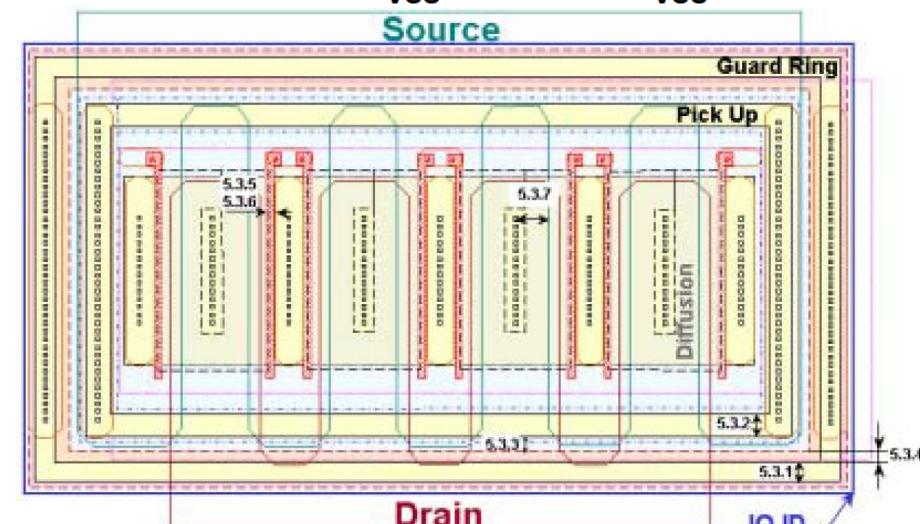
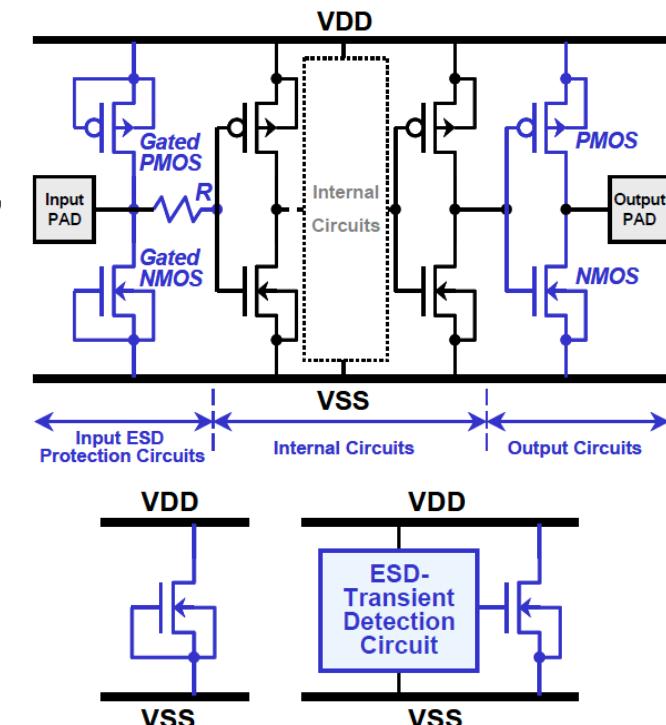
ESD Protection Circuits

How to use them properly?

General Guidelines:

- make sure the current flows where desired and is uniformly distributed,
- series resistance is used to limit the current in the protected devices,
- minimize the resistance in protecting devices,
- use distributed (smaller) active clamps to minimize the effect of bus resistance,
- use guard rings to prevent latchup,
- check the ESD path between every pair of pads,
- check for ESD protection between the pad and internal circuitry,
- check for low bus resistance,
- check for sufficient contacts and vias in the ESD path (uniform current distribution).

Item \ Device (unit : um)	(unit : um)	
	3.3V	1.8V
*5.1.1	Minimum NMOS Total Channel Width	320
*5.1.2	Minimum PMOS Total Channel Width	400
5.1.3	Minimum NMOS Channel Length	0.29
5.1.4	Minimum PMOS Channel Length	0.28
* 5.1.5	Minimum Poly Resistance Width	3
* 5.1.6	Minimum Poly Resistance	200 ohm
* 5.1.7	Minimum Number of First Poly Contacts	4
5.1.8	Minimum Diffusion Contact to Poly Gate Spacing	
5.1.8.1	at Drain Side	2.15
5.1.8.2	at Source Side	0.5
5.1.9	Minimum Drain Side Diffusion Contact to Field Edge Spacing	2



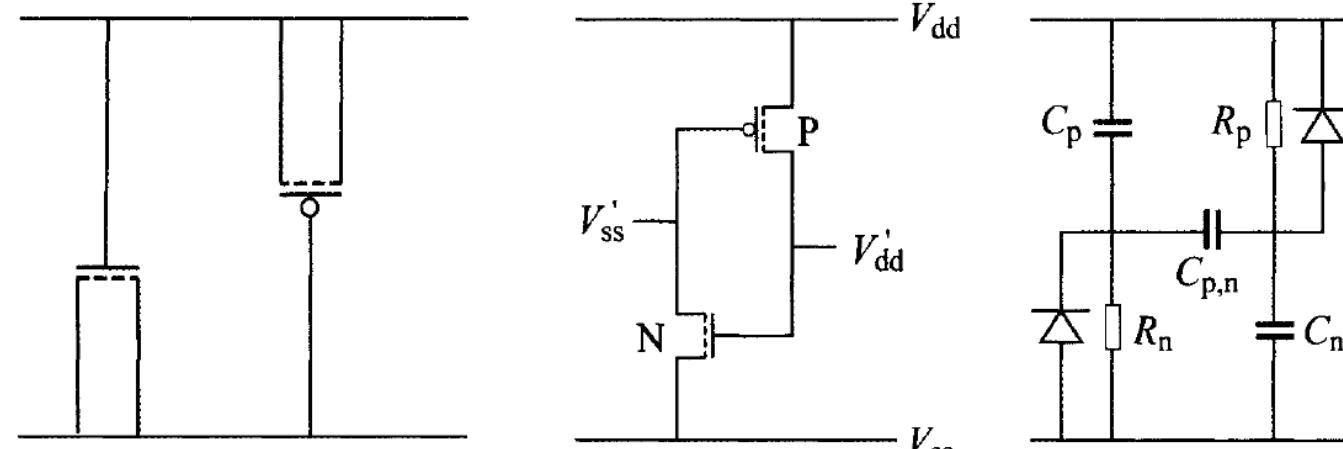
Exemplary schematics, layout and dimensions of the protection circuits for a 180nm CMOS process.

ESD Protection Circuits/Decoupling Caps

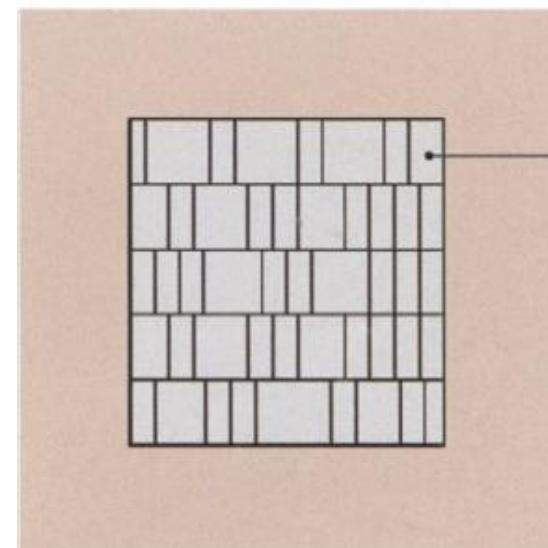
In several applications a tie-off cell supplies dummy V_{dd}' and V_{ss}' potentials to inputs of circuits, which, for reasons of electro-static discharge (ESD), are not allowed to be directly connected to the V_{dd} and V_{ss} rails.

These decoupling capacitor cells are designed as standard cells and are usually available in different sizes. The amount of decoupling capacitance that needs to be added in each core depends on the number of flip-flops in it and on the switching activity of its logic. The switching activity α is defined as the average number of gates that switch during a clock cycle.

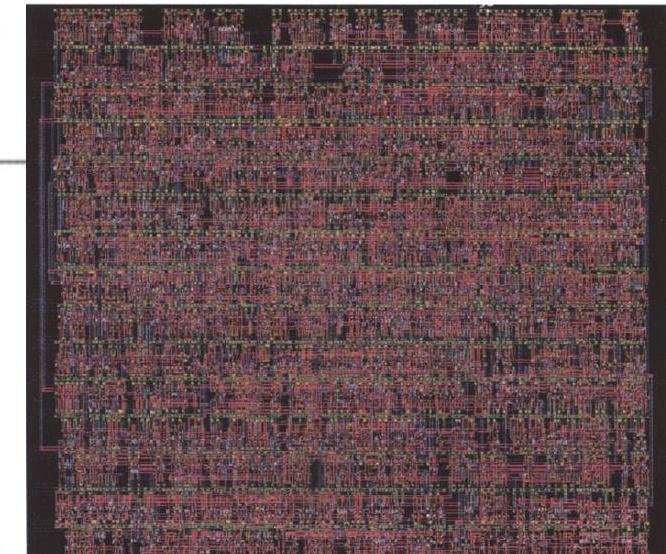
As an example, the total additional decoupling capacitance in a logic block, performing a video algorithm, running at 1 GHz in a 65 nm CMOS core in a digital chip, may occupy about 10 to 20 % of its total area.



Normal decoupling capacitor (a), tie-off cell decoupling capacitor (b), and equivalent circuit (c)



Basic standard-cell layout



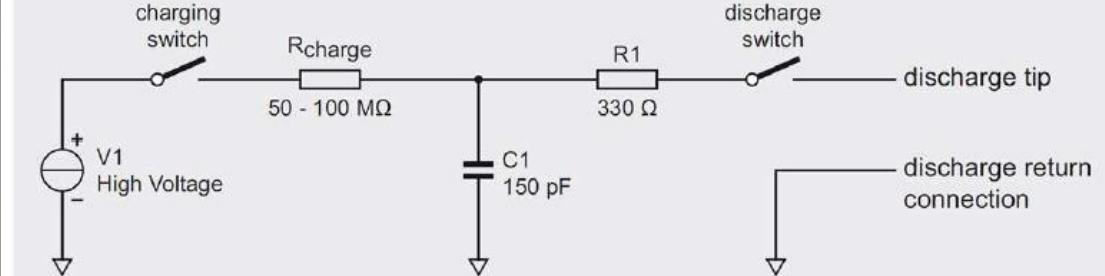
Standard cell implementation of potentiometer

ESD TESTS

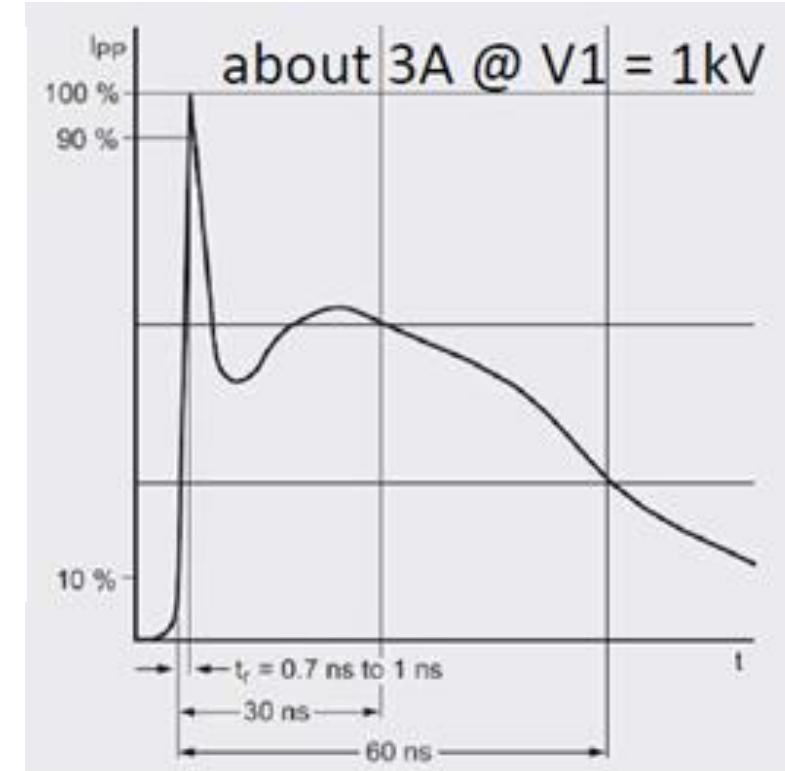
“ESD gun” from emtest :

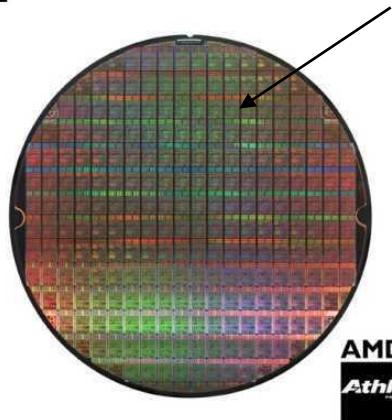


Internal circuit:

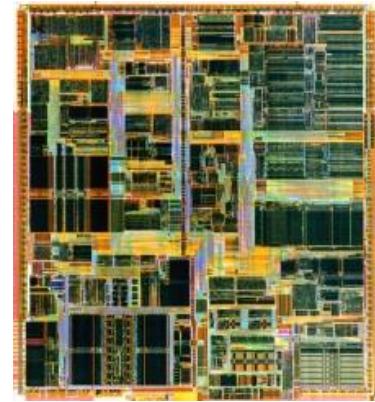


Typical waveform of ESD generator output current.

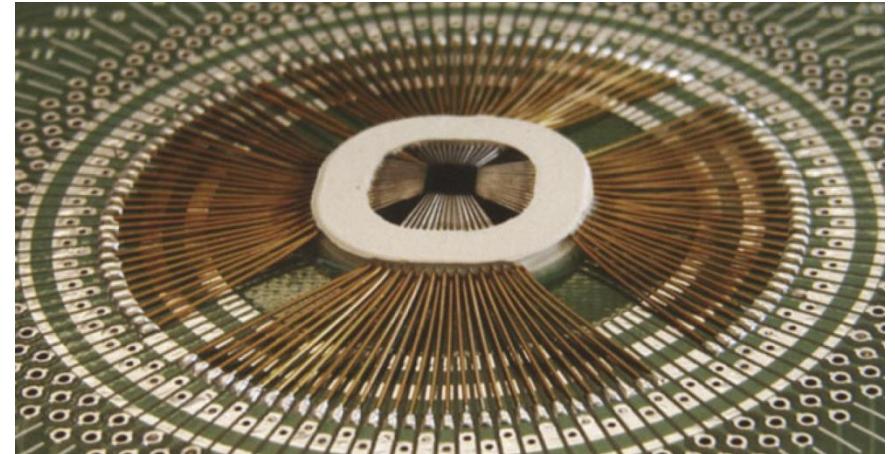




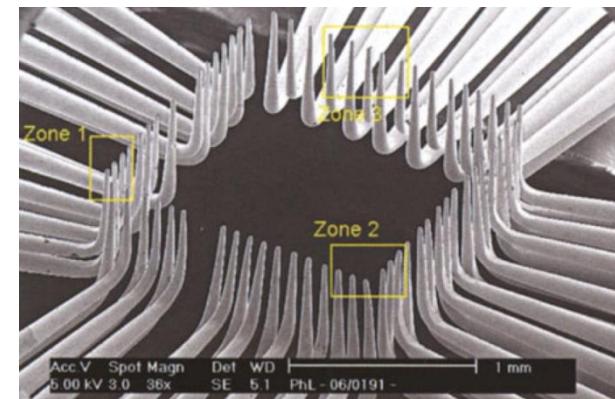
Wafer

**AMD**
Athlon

Packaging



Example of a probe card



Picture of cantilever probe card.

Packaging process flow:

- functionality and electrical tests,
- backgrindging and sawing,
- packaging.

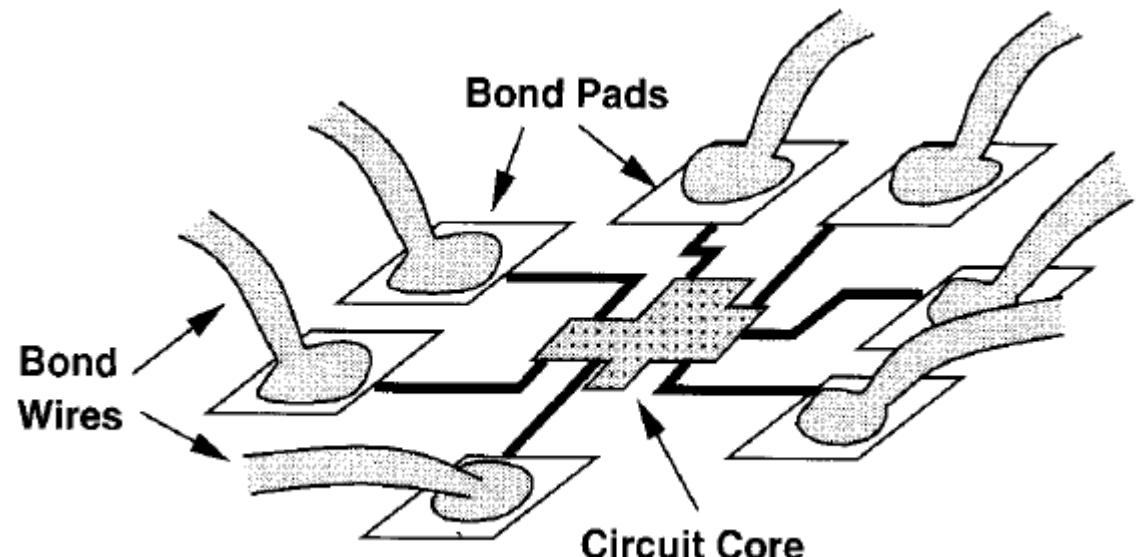
Before the actual packaging starts, each wafer will have to be back grinded to the optimal thickness. Typical thicknesses are 280 µm and 380 µm, while 100 µm is more common for very thin packages. Once back grinded to the right thickness, the wafer has to be separated into individual dies. This is typically done by means of a diamond saw, although laser separation is an up coming trend. Laser dicing has a couple of advantages compared to diamond sawing. Its is faster, it causes less material stress, it requires a smaller scribe line and is able to dice devices with different form factors on the same wafer. To allow this dicing scribe lanes of 50 µm to 200 µm are designed around each die.

PADS

PADS are interface between the IC and external world.

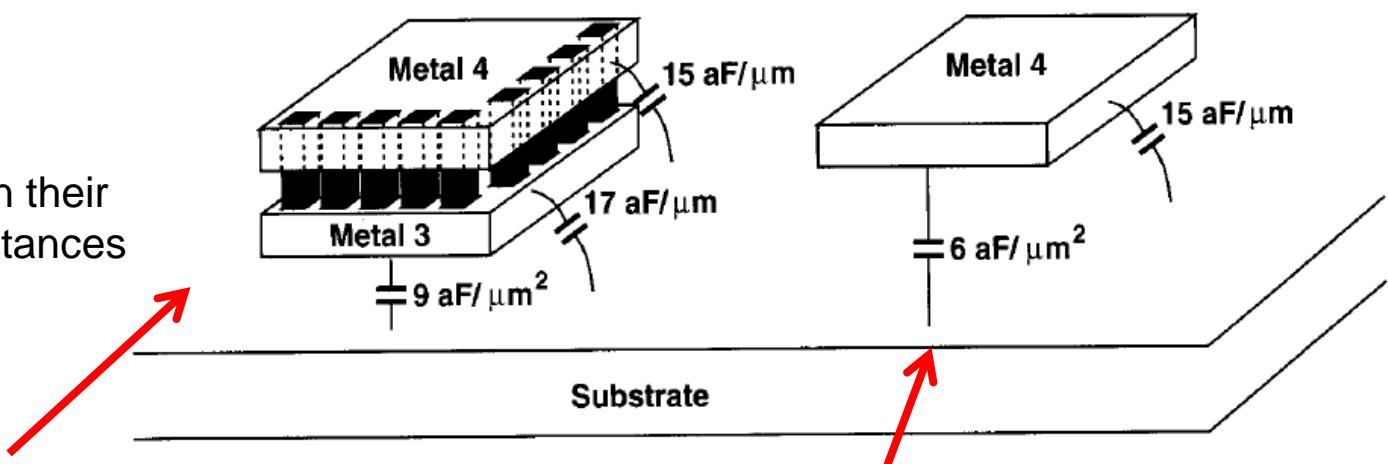
The dimensions are dictated mainly by the bonding requirements and may be equal to $20 \mu\text{m} \times 20 \mu\text{m}$ up to $100 \mu\text{m} \times 100 \mu\text{m}$.

A simple pad would consist of only a one metal layer however because there exist a „lift off” problem during bonding it is usually stacked of at least two metal layers (top most),



Conceptual view of bonding pads.

Compare two different pads in their terms of their parasitic capacitances ($75 \mu\text{m} \times 75 \mu\text{m}$).



$$C_{\text{TOT}} = 75^2 \times 9 + 75 \times 4 \times (17 + 15) \text{ fF} = 60.22 \text{ fF}$$

$$C_{\text{TOT}} = 75^2 \times 6 + 75 \times 4 \times 15 \text{ fF} = 38.25 \text{ fF}$$

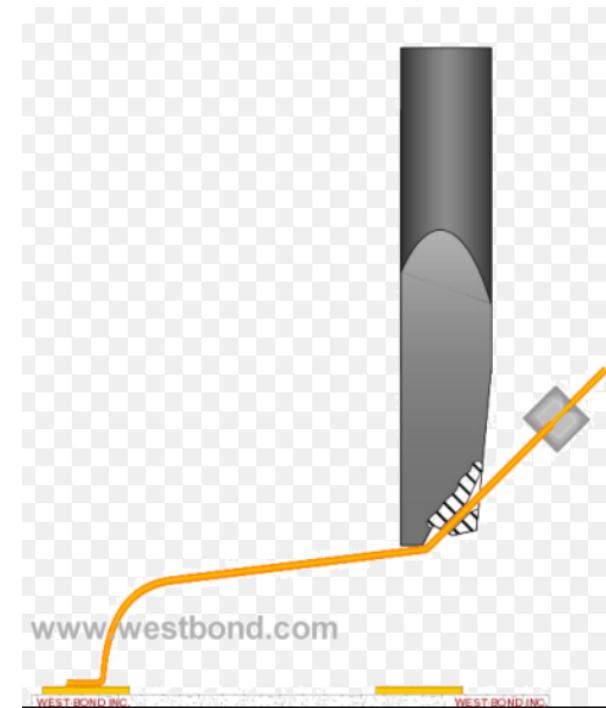
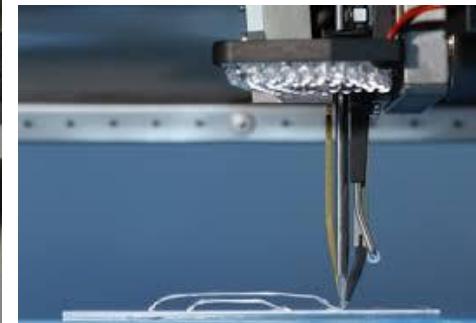


AGH

Packaging



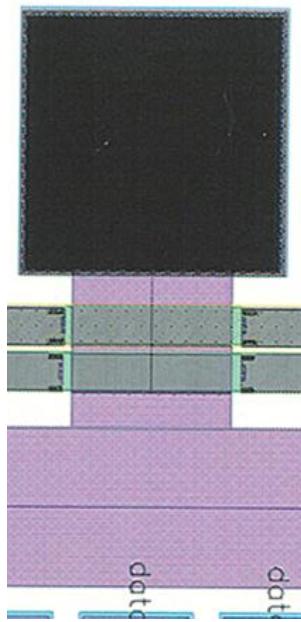
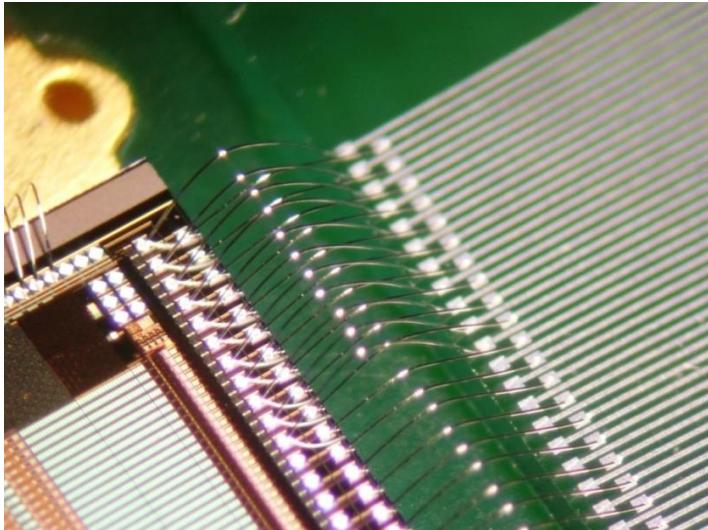
Photos of the wire bonder (Department of Metrology and Electronics)



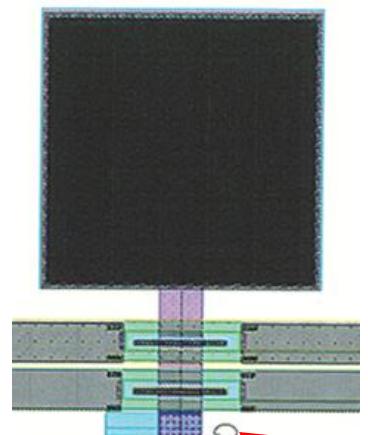
Wire bonders' wedge.



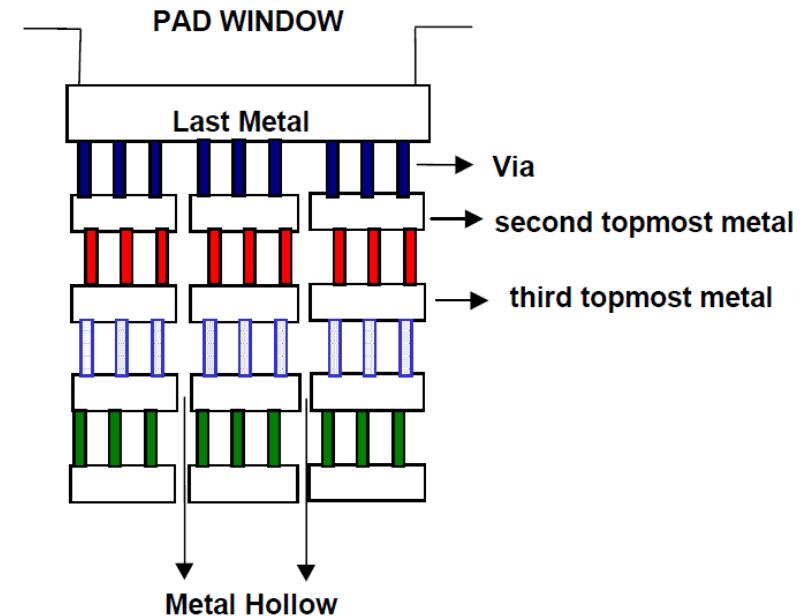
PADS



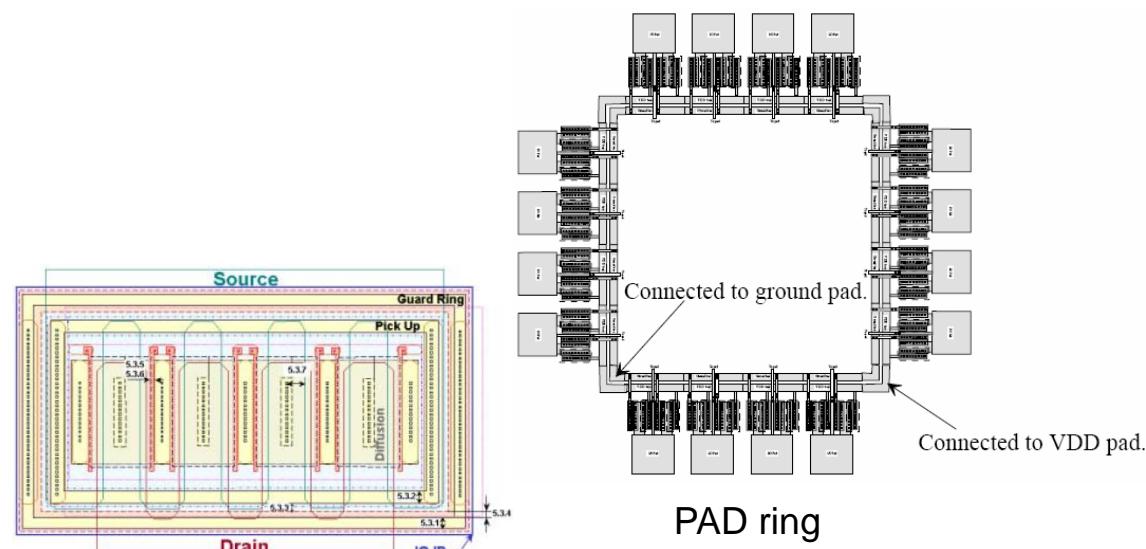
PAD dedicated for supplying.



Input PAD with protecting diodes



PADs' cross section



PAD ring

Packaging

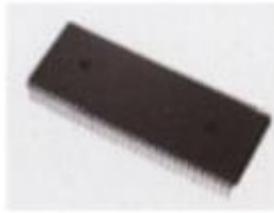
The package choice is very much related to the electrical, thermal and size requirements dictated by the application domain.

The package supports various important functions:

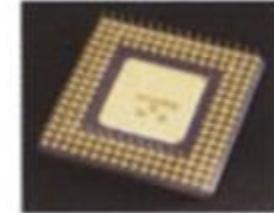
- allow an IC to be handled for PCB assembly and protect it during further PCB production,
- mechanical and chemical protection against the environment,
- mechanical interface to the PCB,
- good electrical connection (signals and power supply) between PCB and chip,
- enhance thermal properties to improve heat transport for IC to environment,
- allow standardization.



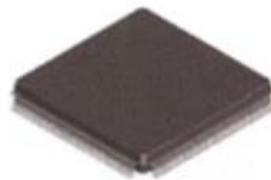
SIL



DIL



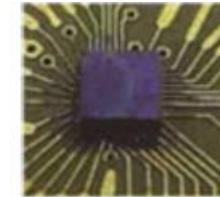
PGA



QFP

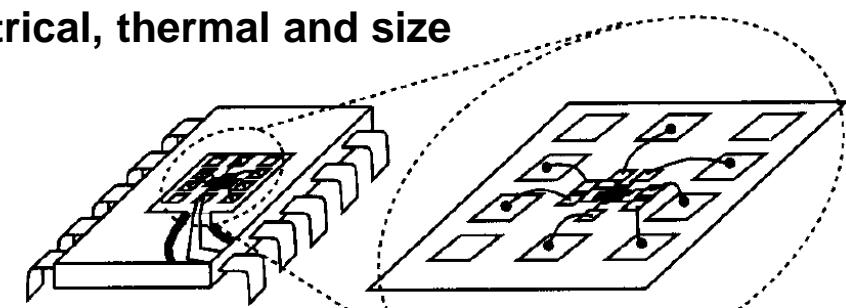


MLF/QFN

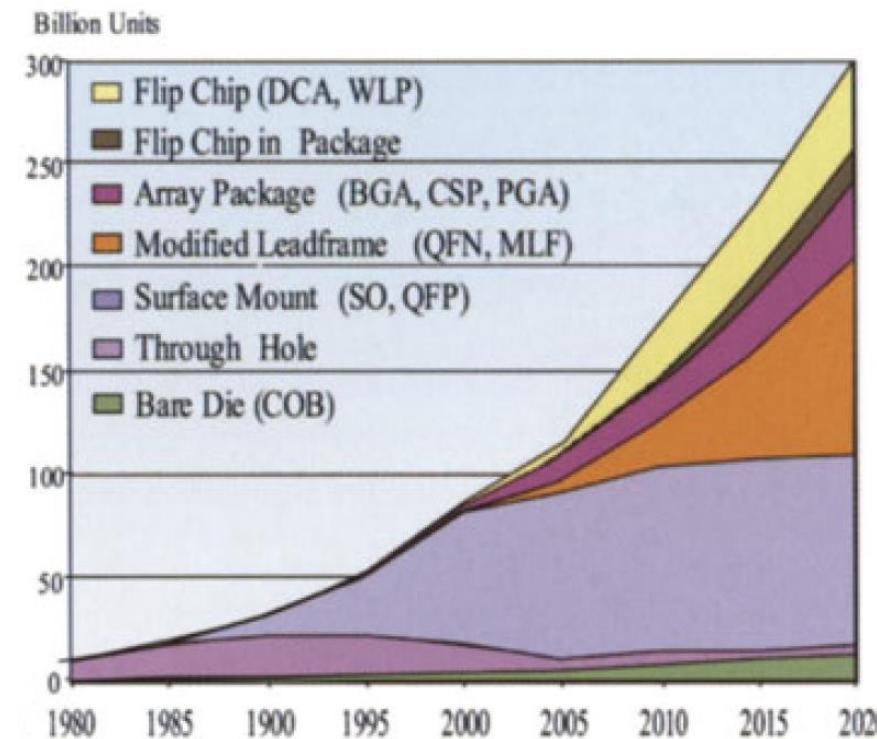


FLIP-CHIP

Various package images.



Conceptual view of the dual in line (DIL) package

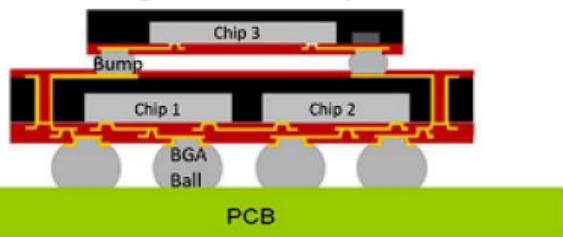


Current and expected market penetration of the different package categories .

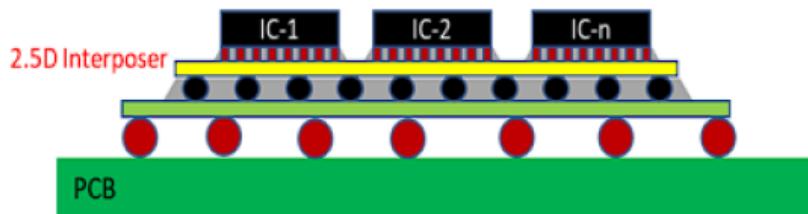
Packaging

Heterogeneous Integration Platforms

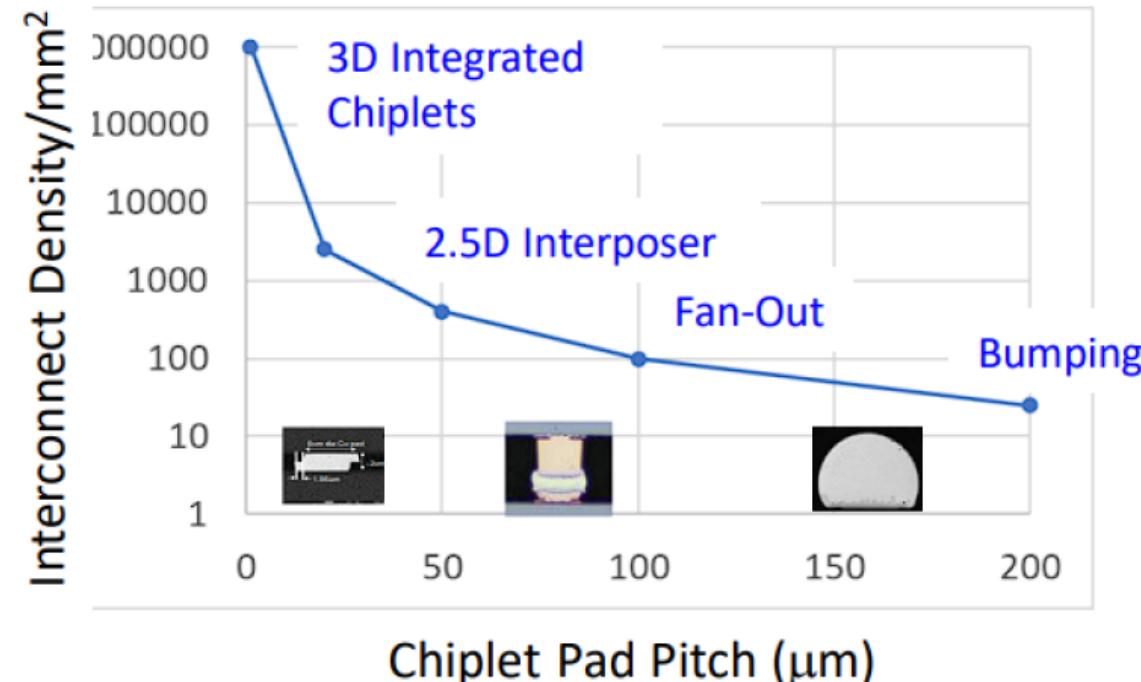
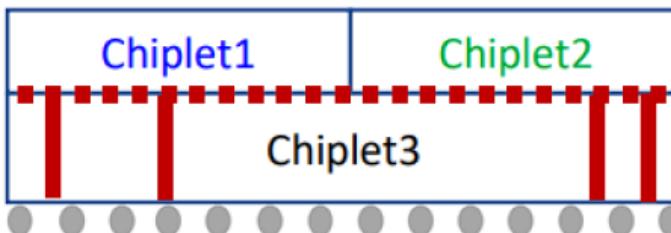
Embedded High Density Fan-Out WLP



2.5D Interposer



3D-Integrated Chiplets

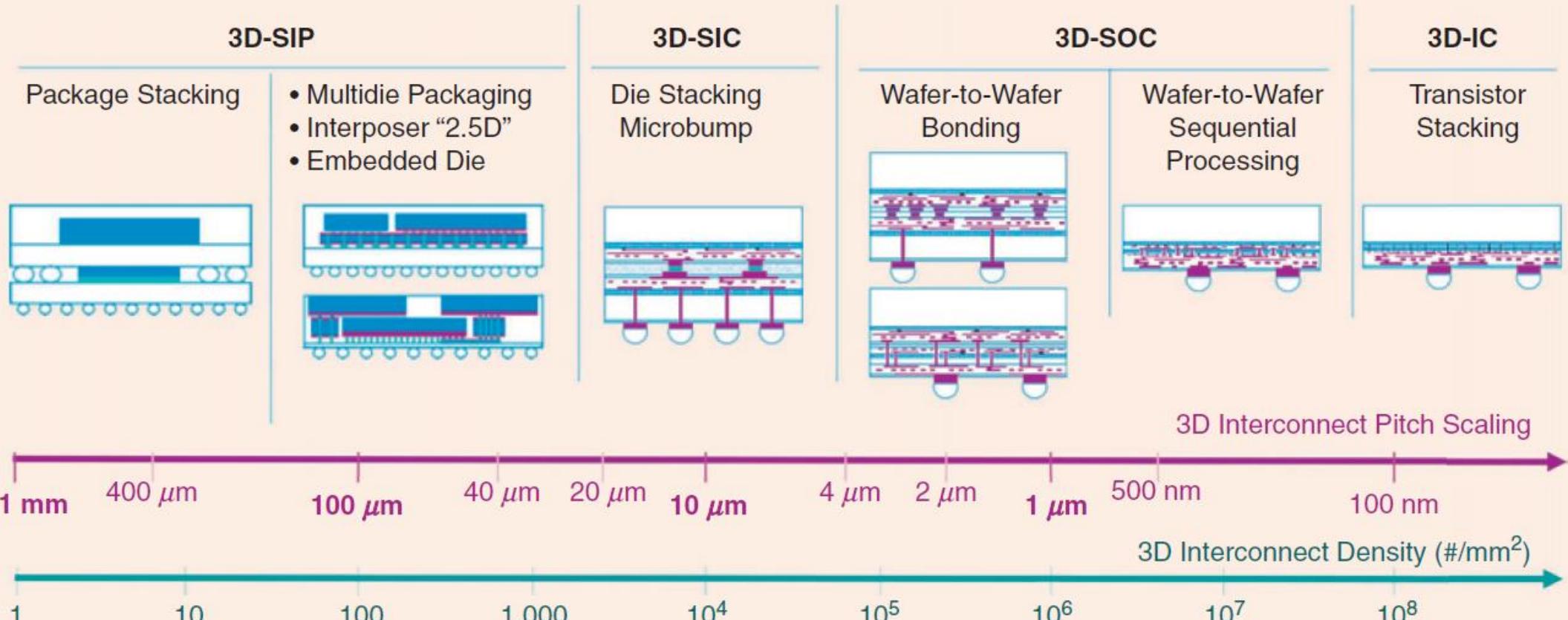


Inter-chiplet wiring density increases

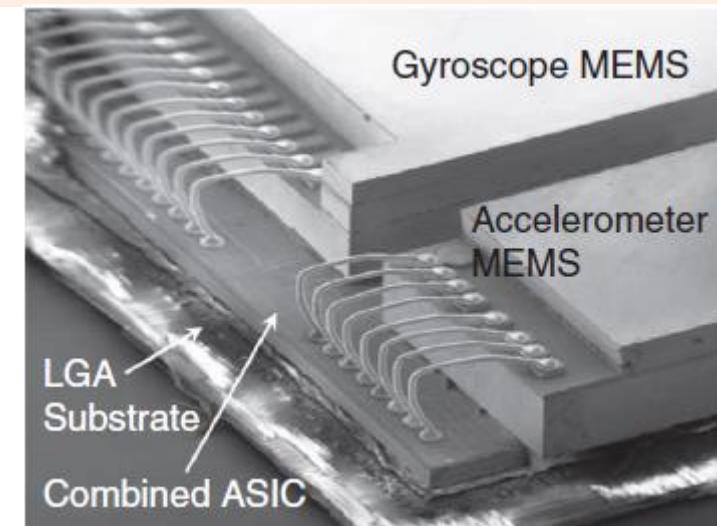
- 5x-10x from Bumping to FOWLP
- 10x from FOWLP to 2.5D Interposer
- >100x from 2.5D Interposer to 3DIC (hyb. Bonded)

CHALLENGES: Package Architecture, Design, Materials, Process Optimization

Packaging



The landscape of 3D interconnect technologies can be categorized based on their increasing 3D interconnect density capability. System-in-package (SIP) integration technologies and stacked-IC (SIC) technology with die-to-die interconnects offer relatively lower-density capabilities. On the other hand, system-on-chip (SOC) integration enabled by advanced codesign EDA tools, wafer-to-wafer bonding technologies, and sequential 3D stacking offer higher-density capabilities. The highest density is offered by transistor-level stacking of devices, allowing for 3D integration at the standard cell level, which is referred to as 3D-IC. Reprinted, with permission, from S. B. Samavedam et al., "Future logic scaling: towards Atomic Channels and Deconstructed Chips," IEDM Tech. Digest, 2020.



Inside a multidie inertial sensor. LGA: land grid array.

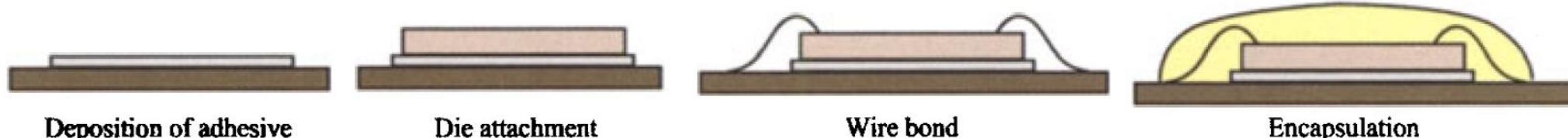
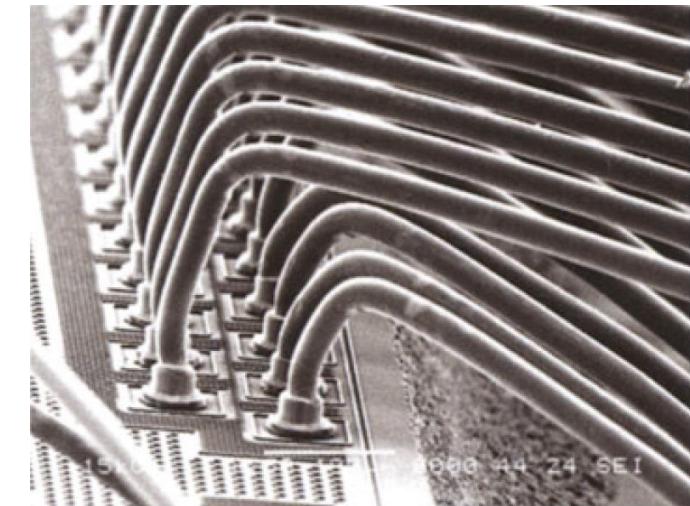
Packaging

IC connection to the outside world.

The most common method is *wire bonding (WB)*, which is still responsible for about 90% of all chip interconnects.

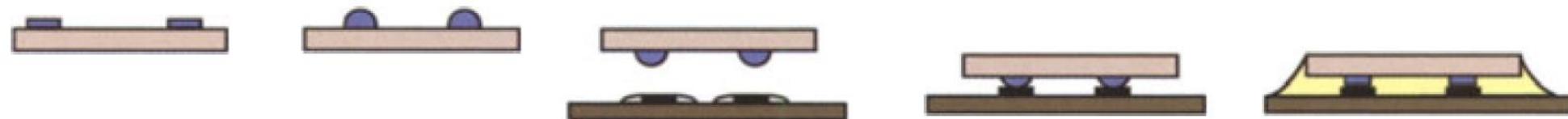
Diameters of the wires range from 15 µm for fine pitch applications to 150 µm for high power devices.

There are two common wirebond processes, depending on the applied wire material: Au ball bonding and Al wedge bonding.



Overview of the wire bonding process.

In flip-chip bonding (FOB), which is the second interconnect technology, the die is assembled face down directly onto the circuit board with solder, Au or Au/Ni bumps. Compared to wire bond, this technology comes with less area overhead, because there is no additional area needed for contacts on the sides of a chip. It enables the final packaged chip to be only marginally larger than the original die.



placing of solder bumps

reflowing of bumps

**alignment of chip
bumps and substrate pads**

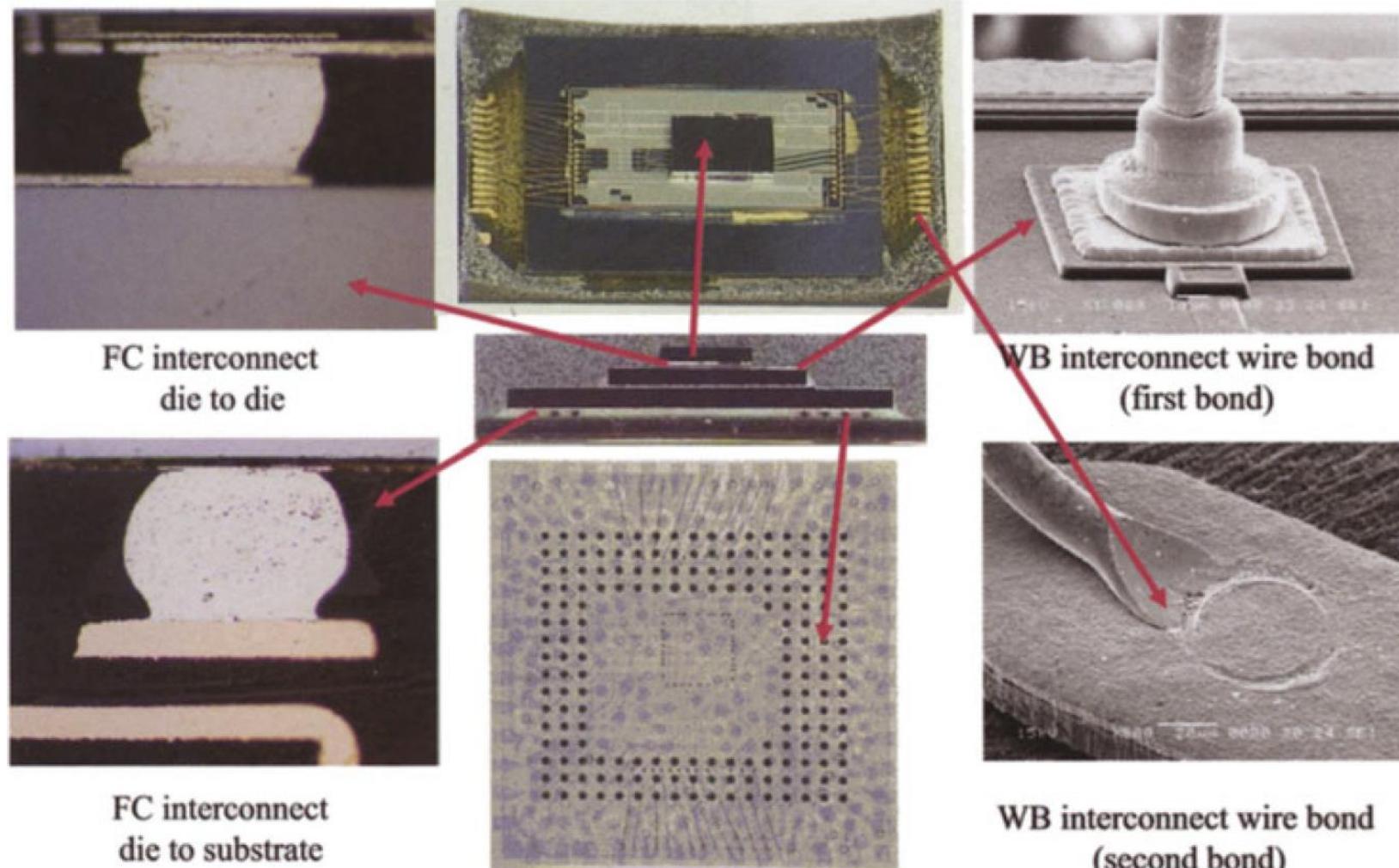
pressing and reflowing

underfilling and curing

Overview of flip-chip bonding process.

Packaging

Flip-chip bonding accommodates dies that may have several hundred bond pads placed anywhere on their top surface. In many cases an additional redistribution metal layer is required to transfer periphery wire bonding pad connections into an area array of connections for flip-chip bonding. Compared to other interconnection techniques, FCB results in very short connections and exhibits improved performance in high-speed applications. **Flip-chip connection can be found both in a silicon to substrate and silicon to silicon bumping.**



Package with both flip-chip (Fe) and wirebond (WB) interconnections.

Packaging

Thermal aspects of packaging.

Another dominating parameter in the performance and reliability of an integrated circuit is the physical temperature of the die inside the package, which is determined by the power consumption of the IC in combination with the thermal behaviour of the package.

$$R_{JA} = \frac{T_{IC_J} - T_{AMBIENT}}{P} [^{\circ}C/W]$$

T_{IC_J} and $T_{AMBIENT}$ represents the temperature difference between the chip (junction) and its environment (ambient).

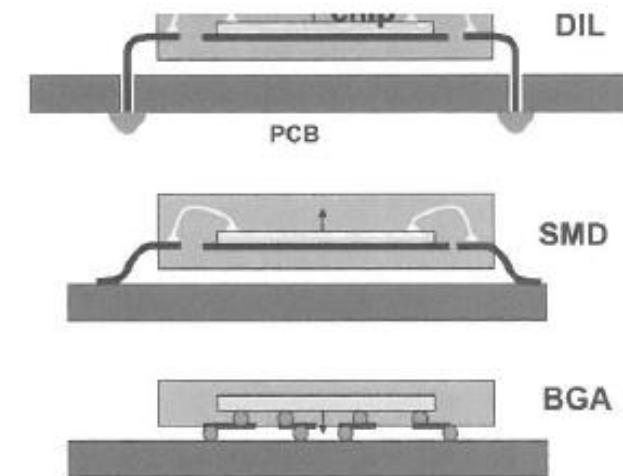
In many applications the maximum junction temperature is defined as 125 "C. If we assume a consumer application with an ambient temperature of 70"C, the maximum allowed power consumption of an IC, packaged with a 20 pins PLCC under still-air conditions (on a reference board of the supplier) , is then equal to:

$$P = \frac{T_{IC_J} - T_{AMBIENT}}{R_{JA}} = \frac{125 - 70}{80} = 687mW$$

The other parameter is defined as the junction-to-case thermal resistance R_{JC} , and represents the ability of a package to conduct heat from the junction (die) to the surface (top or bottom) of the case (package) and is expressed as follows:

$$R_{JC} = \frac{T_{IC_J} - T_{CASE}}{P} [^{\circ}C/W]$$

Package-Pins	Package Designator	Package Outline Code	RTH(J-A) °C/W	RTH(J-C) °C/W
DQFN-14	BQ	SOT762	61	32
DQFN-20	BQ	SOT764	50	23
HVQFN-16	BS	SOT629	40	18
HVQFN-24	BS	SOT616	40	18
LFBGA-96	EC	SOT536	60	16
PLCC-20	A	SOT380	80	32
TSSOP-24	PW	SOT355	128	32



Packaging

Electrical packaging aspects.

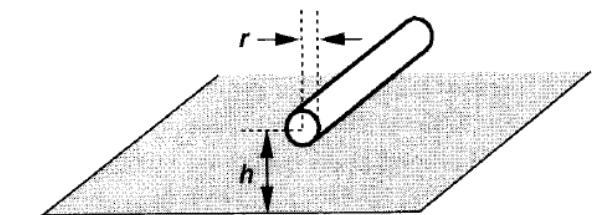
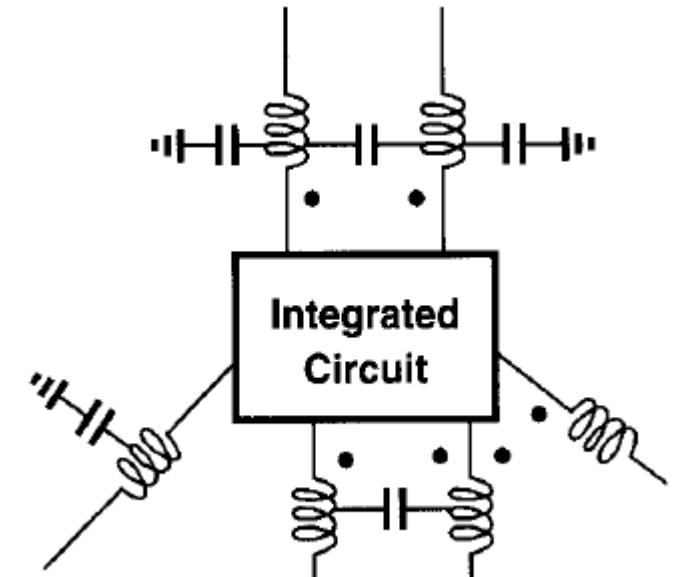
Conventional package types using lead frames have relatively large lead inductances (1-50 nH), because of longer lead lengths. They also tend to have a high mutual coupling.

The use of flip-chip bonding improve electrical performance by minimizing the lengths of the connections between the die and the substrate, resulting in inductances of 0.5-1 nH.

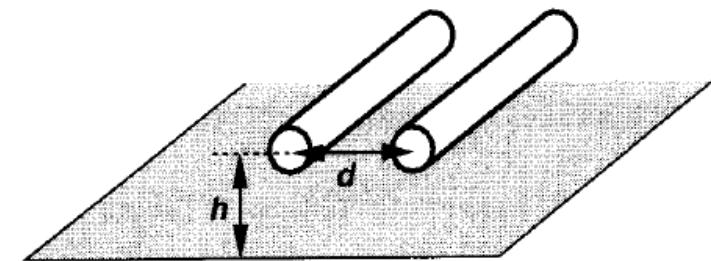
Package Type	Capacitance (pF)	Inductance (nH)
68 Pin Plastic DIP	4	35
68 Pin Ceramic DIP	7	20
256 Pin Pin Grid Array	5	15
Wire Bond	1	1
Solder Bump	0.5	0.1

Parasitics of the different IC connetions.

Typical bond wires inductance is equal to 1nH/mm.



$$L \approx 0.2 \ln \frac{2h}{r} [nH / mm]$$



$$L \approx 0.1 \ln(1 + (\frac{2h}{d})^2) [nH / mm]$$

Packaging

Electrical packaging aspects.

Lets assume the $C_L=0.5 \text{ pF}$, transient times should be at least 0.5 ns , $L_D=L_G= 5 \text{ nH}$, and $V_{DD}=3 \text{ V}$. How big then the supply bounces are?

$$I = C \frac{dV}{dt} = 0.5 \text{ pF} \frac{3\text{V}}{0.5\text{ns}} = 3mA$$

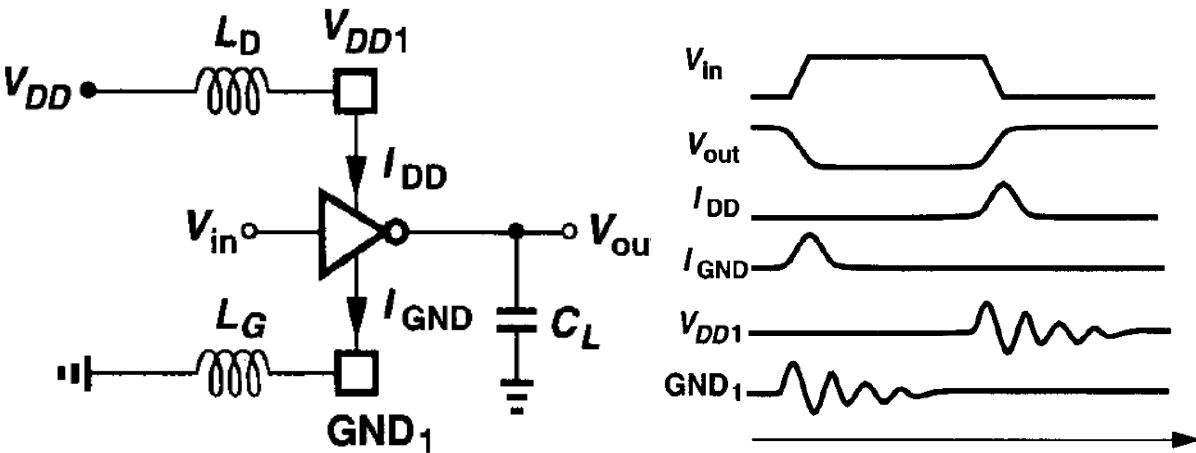
Which means that the inductance drop-out voltage is given by:

$$dV_L = L \frac{dI}{dt} = 6 \times 10^6 L = 30mV$$

Lets assume the 15 milion 600 MHz processor draws 25 A current in 5 ns during the switching (the supply voltage is 2 V). Calculate the supply bounce if both the V_{DD} and GND are connected with 2×200 parallel bonds of 5 nH each.

$$dV_L = \frac{5 \times 10^{-9}}{200} \times \frac{25}{5 \times 10^{-9}} = 125mV$$

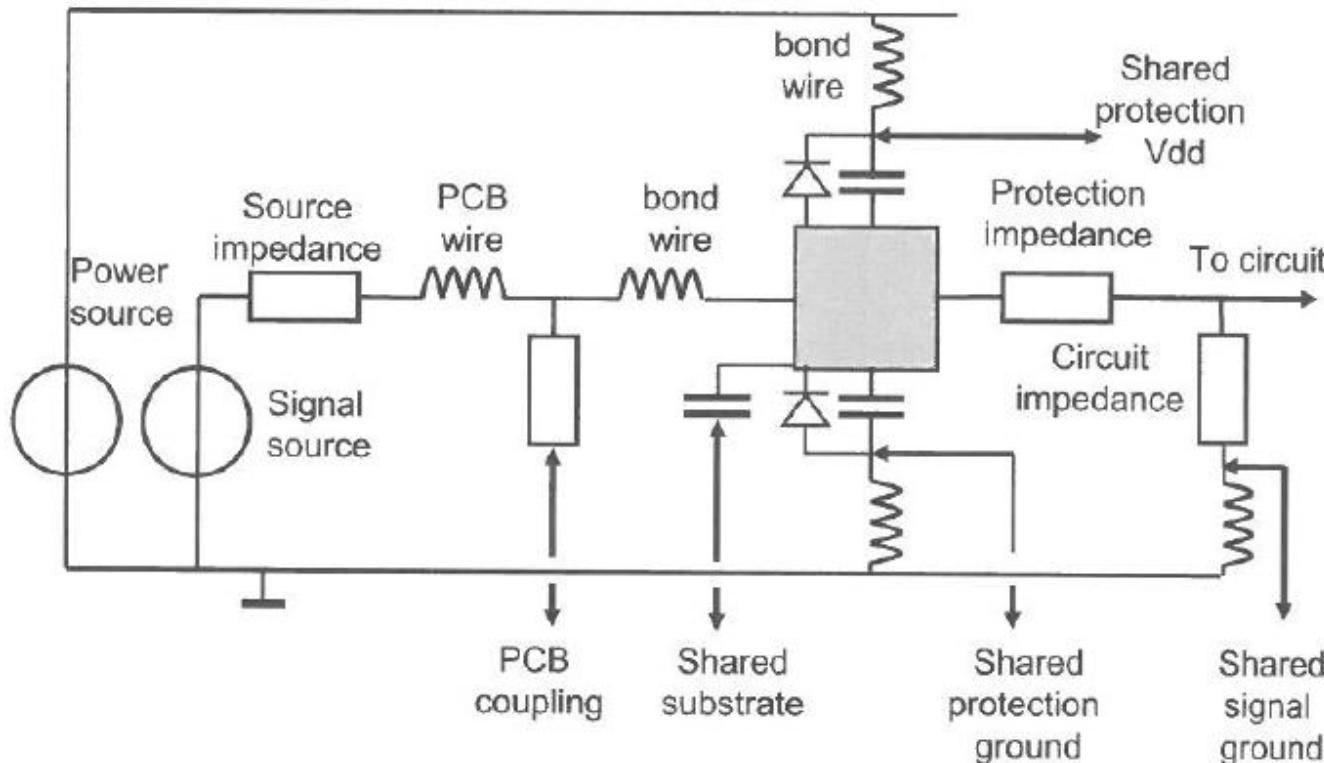
In the worst case both bounces may add in phase and one would get 250 mV of supply variations which is more than 10 % of its value (2 V).



CMOS inverter driving a capacitance.

Packaging

Electrical packaging aspects.



Equivalent circuit diagram showing the most important parasitic elements in connecting a signal to the IC

Bond wire	$L \approx 1 \text{ nH/mm}$, $L = 1 \text{ nH}(\text{CSP})-10 \text{ nH}(\text{DIL})$
Bond wire	$R \approx 0.3-1 \Omega$
Bondpad capacitance to substrate	0.8–2 pF for $0.1 \times 0.1 \text{ mm}^2$
Diode capacitance to supply	0.3–0.7 pF
Protection resistor	100–400 Ω

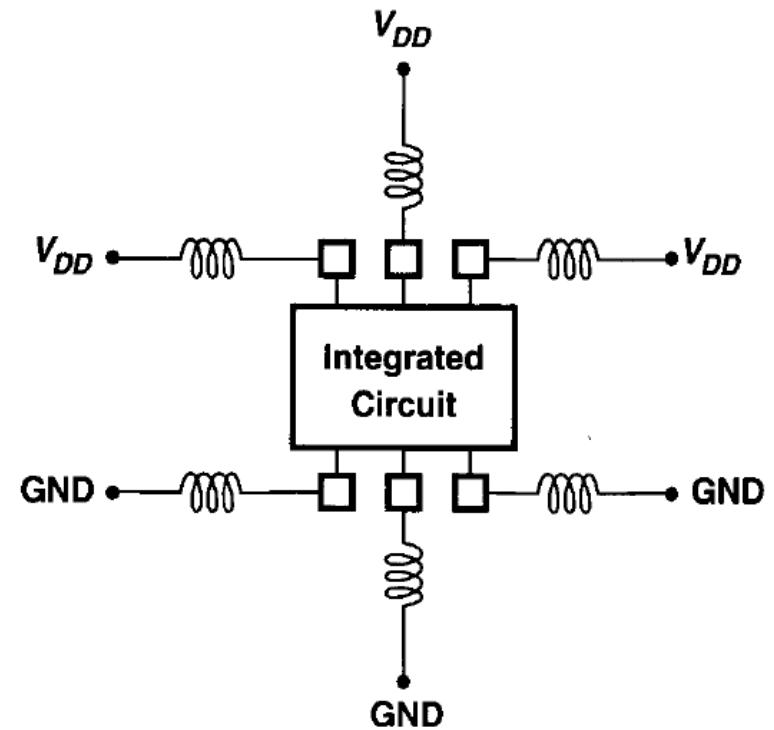
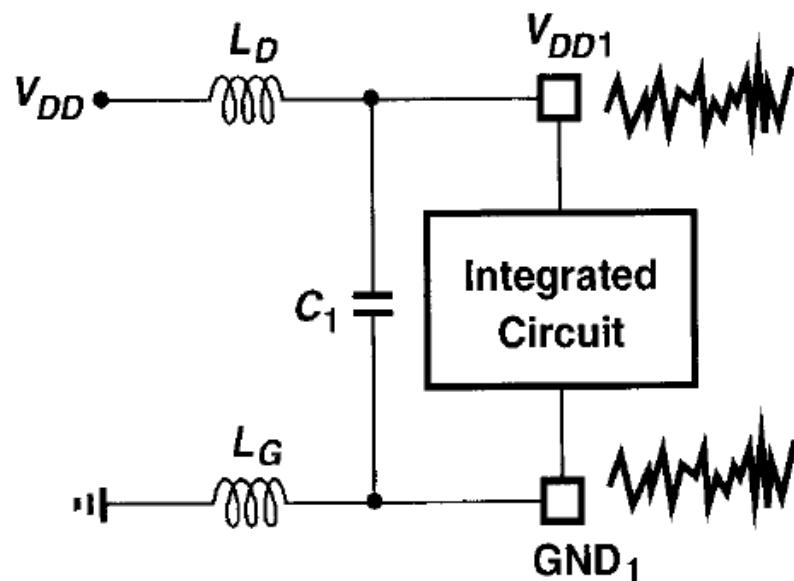
Some values for parasitic components shown in the figure above (CSP – chip scale package, DIL-Dual in Line packages).

Packaging

Electrical packaging aspects.

How to minimize above problems?

1. Use on-chip decaps but be aware of its self resonance and the IC yield (more transistors need to be added as caps therefore degrading the yield).
2. Decrease the bond inductance.



Idea of the on-chip decoupling capacitor and the multiple bonds reducing the overall inductance.

Clean room



Class 10,000

*Printed circuit boards,
Electronic packaging,
Medical devices*



Class 1,000

*MEMS,
Electronic packaging,
Hard disk drives*



Class 100

*MEMS,
RF/Photonic ICs*



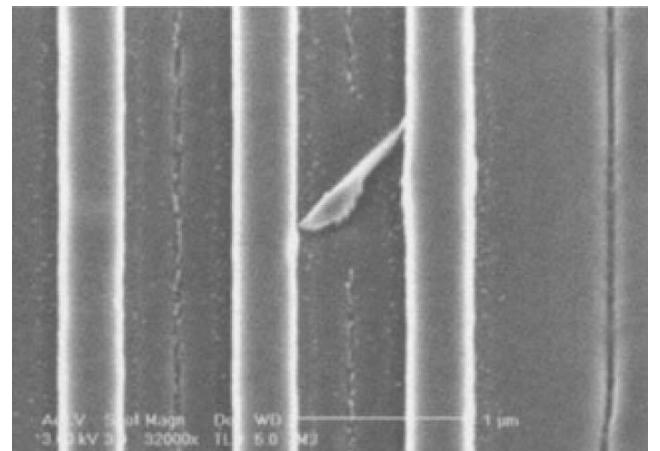
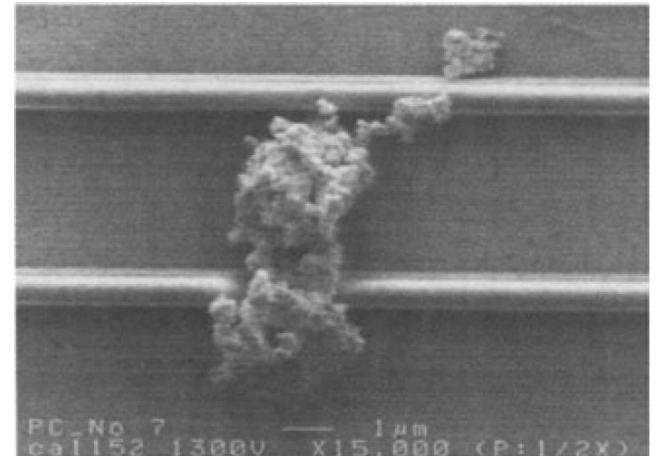
Class 10

Integrated Circuits



Semiconductor clean room facility, today

- Class 1000: fewer than 1,000 particles ($>0.5 \mu\text{m}$) in 1 cubic foot of air
- Class 100: Fewer than 100 particles ($>0.5 \mu\text{m}$) in 1 cubic foot of air



Example of random failures: particles causing a potential short

