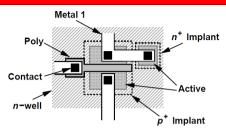
Chapter 19:Layout and Packaging

- 19.1 General Layout Considerations
- 19.2 Analog Layout Techniques
- 19.3 Substrate Coupling
- 19.4 Packaging

Overview

- Analog CMOS circuits have evolved into high-speed, highcomplexity, low-voltage "mixed-signal" systems with digital peripheral circuitry.
- Today's analog circuit design is influenced by interaction, layout and packaging.
- Overview
 - Layout rules and methods
 - Substrate coupling
 - Packaging issues

General Layout Considerations



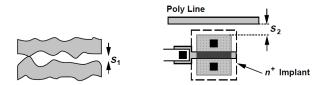
- Layout of IC defines the geometries that appear on the masks used in fabrication
- · Geometry requirements:
 - N-well surrounds the device with enough margin in case of misalignments during fabrication
 - "Active area" is surrounded by a proper implant geometry with enough margin
 - Contact windows provides connection from active and poly regions to the M1

Design Rules(1)

 Minimum Width defines the widths and lengths of geometries must exceed a minimum value imposed by lithography and processing capabilities.

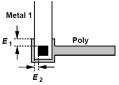


 Minimum Spacing defines different masks must be separated by a minimum spacing.

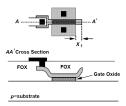


Design Rules(2)

 Minimum Enclosure defines the sufficient margin of geometries surrounded by other masks

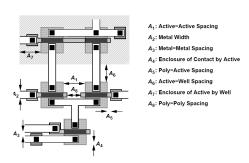


 Minimum Extension defines some geometries mush extend beyond the edge of others by a minimum value.

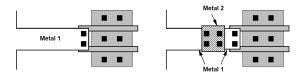


Summary of Design Rules

- Modern CMOS technologies typically involve several hundred of layout design rules
- Some maximum allowable dimensions or antenna effect would also be considered
- E.g., layout of an NMOS differential pair with PMOS current-source load



Antenna Effect



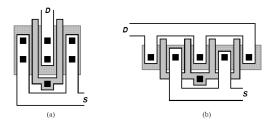
- If metal 1 interconnect having a large area is tied to the small gate of MOSFET. The metal area acts as an antenna during etching, collecting ions and rising in potential.
- Gate oxide may break down. Both happens in metal 1 and poly.
- A discontinuity in metal 1 layer solves the issue.

Analog Layout Techniques

 Analog systems, unlike digital ICs, not aim to maximize the yield but demand more on minimizing effects such as crosstalk, mismatch and noise.

- · Techniques:
 - Multifinger Transistor
 - Symmetry
 - Reference Distribution

Multifinger Transisitors



- Reduce both the S/D junction area (about 2) and the gate resistance(~ a factor of 4) by one fold
- Multiple "fingers" may be suitable for very wide devices.
- Rule of Thumb:

The width of each finger is chosen such that the gate resistance of the finger is less than $_{1/g_m}.$ For low noise applications, 1/5 or 1/10 is fine.

Example 19.1

- W/L= 5um/40nm MOSFET biased at 1mA
- Transconductance of 1/100 Ohm
- The sheet resistance of the gate poly is 30 $_{\Omega/\Box}$ Let gate thermal noise voltage is one-fifth of the gate-referred thermal noise voltage. What is the widest finger that the structure can incorporate?

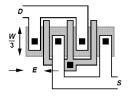
Solution:

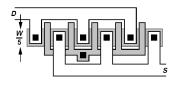
As N parallel fingers, each finger, W/L = 5um/4nnm/N. The sheet resistance:

$$R = \frac{\rho}{t} \frac{L}{W} = R_s \frac{L}{W} \qquad 30 \ \Omega \times (5/0.04)/N$$
 Channel Noise = $\sqrt{4kT\gamma(100)} \ \text{V}/\sqrt{\text{Hz}}$ Gate Noise = $\sqrt{4kT\frac{150}{0.04N^2}\frac{1}{3}} \ \text{V}/\sqrt{\text{Hz}}$ **Ire of resistance fact.**

N = 17.7. Thus a minimum of 18 fingers is required.

Mutifinger Capacitance

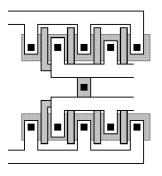




- The capacitance associated with the perimeter of the source/drain varies
- For an odd number finger N

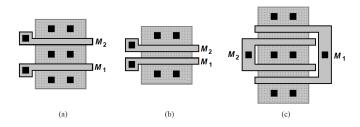
$$\begin{split} C_P &=& \frac{N+1}{2} \left(2E + \frac{2W}{N}\right) C_{jsw} \\ &=& \left[(N+1)E + \frac{N+1}{N}W \right] C_{jsw}. \end{split}$$

Large Number of Gate Fingers



 To avoid long geometries and hence disproportionate dimensions in the layout of the overall circuit

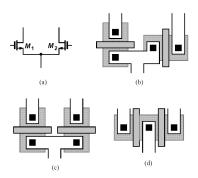
Cascode Circuit Layout Simplification



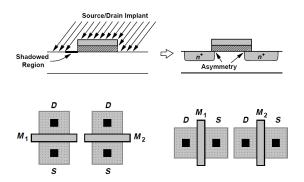
- The layout of a cascade circuit can be simplified if the input device and the cascade device have equal widths.
- The drain of M1 and the source of M2 can share the same junction.
- This junction does not need a contact window, which is quite small to improve HF performance.
- Wide cascade device may be needed to add fingers.

Symmetry Advantages

- Asymmetries may lead to inevitable mismatch, thus introducing input-referred offsets, and limiting the minimum signal level
- Suppress the effect of CM noise and even-order nonlinearity
- Two plausible solutions in (c) and (d)

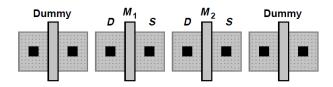


Gate Shadowing

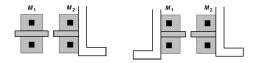


- The implant is tilted by about 7 degree
- A small asymmetry between source and drain side diffusions
- The first topology is preferable

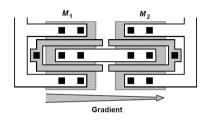
Symmetry by "Dummy" Transistors

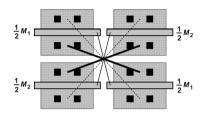


- The asymmetry inherent to the structures can be ameliorated by adding "dummy" transistors.
- In complex circuits, such measures cannot be easily applied.
- We emphasize the importance of the axis of symmetry.



"Common-centroid" Configuration





- Symmetry becomes more difficult to establish if there are gradients along certain(x-axis) axis
- · A "common-centroid" configuration may be used
- Decompose each transistor into two halves that are placed diagonally opposite each other
- · The routing of interconnects is quite difficult

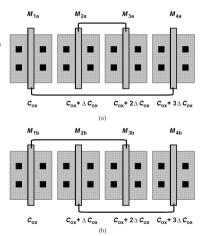
"One-dimensional" Cross Coupling

- All four half transistors are placed along the same axis
- Connect either the near ones and far ones or every other one
- The approach removes the error to a lessor extent

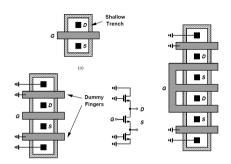
$$I_{D1a} + I_{D4a} = \frac{1}{2} \mu_n (C_{ox} + C_{ox} + 3\Delta C_{ox}) \frac{W}{L} (V_{GS} - V_{TH})^2.$$

 $I_{D2a}+I_{Da3}=\frac{1}{2}\mu_n(C_{ox}+\Delta C_{ox}+C_{ox}+2\Delta C_{ox})\frac{W}{L}(V_{GS}-V_{TH})^2.$ • Dunning translators much

י טעוווווויש וימוווווס וועסווי to the very left and right

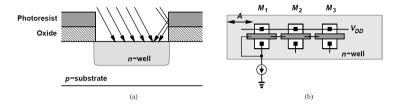


Shallow Trench Isolation Issues



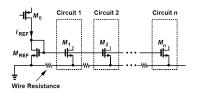
- Modern MOS devices are surrounded by a shallow "trench" (STI).
- This structure exhibits a different thermal expansion coefficient, introducing I/V error.
- To alleviate, we insert two fingers on the two sides of main device. Yet S/D capacitances are increased by dummy gates.

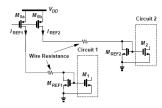
Well Proximity Effects



- Nonuniform doping is due to reflection effect
- The border areas of the wells receive a different doping density
- The PMOS locates at the edge influenced "well proximity" error
- To reduce this effect, the length of A must extend beyond the PMOS devices several microns

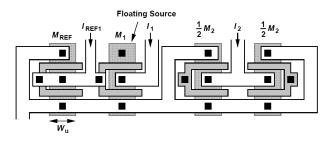
Reference Distribution





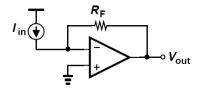
- If the matching is critical, then the voltage drop along the ground line must be taken into account.
- To remedy, the reference can be distributed in the current domain rather than in the voltage domain.
- Local current mirror pair, or even local bandgap reference is necessary.
- · Particular orientation must be paid attention to.

Scaling of Device Dimensions



- $I_{D1} = 0.5I_{REF}$ $I_{D2} = 2I_{REF}$. $W_1 = 0.5W_{REF}$ and $W_2 = 2W_{REF}$
- Layout of scaling comes from one unit width device
- To improve the matching, the array can be surrounded by dummy devices

Passive Devices



- Linearity of resistors is critical
 - e.g. current-to-voltage conversion $V_{out} = -I_{in}R_F$
- IC resistors are usually polysilicon resistor.
 - a silicide block exhibit high linearity
 - low capacitance to the substrate
 - relatively small mismatch
 - linearity depends on the length (on the order of 0.2% for microns width or length)
- Symmetry is also important

Example 19.2

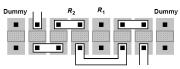
Consider the bandgap circuit shown below. Choose the values of n, R1, and R2 such that Vout exhibits a zero temperature coefficient and the layout can be designed for high precision.



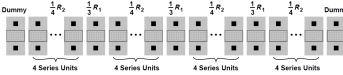
$$V_{out} = V_{BE3} + V_T (R_2/R_1) \ln n$$

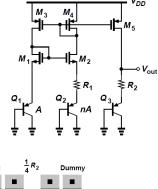
 $(R_2/R_1) \ln n \approx 17.2$

If n = 31, then $R_2/R_1 \approx 5$

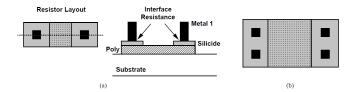


n = 25, obtaining $R_2/R_1 = 5.34$.



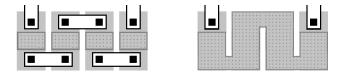


Polysilicon Resistance(1)



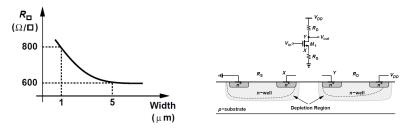
- Consists of two components: unsilicided region and the resistance between contacts
- Double width and length to remain polysilicon resistance while reduce the contact resistance

Polysilicon Resistance(2)



- Resistors are usually decomposed into shorter units that laid out in parallel and connected in series
- From the viewpoint of matching and reproducibility, the structure is preferable to "serpentine" topologies.
- The sheet resistance varies with temperature and process. Around 0.1%/°C and less than 20% process variation.

Other Resistors(1)



- For other resistors, we have N-well, source/drain p+ or n+ material, silicided polysilicon, or metal with R_{\square} decreasing in this order.
- N-well
 - Vary a large fraction, 40% with process
 - Depends on the width of resistor
 - Depends on n-well-substrate voltage difference
 - Vout varies, the sheet resistance of RD changes

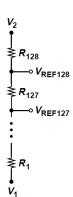
Example 19.3

An A/D converter incorporates a resistor ladder consisting of 128 units made of n-well to generate equally-spaced reference voltage. If the two ends of the ladder are connected to V1 =+1V and V2 = +2V, calculate the ratio R128/R1.

Solution:

The depletion region width

$$\begin{split} x_{d} &= \sqrt{2\epsilon_{si}(\phi_{B} + V_{R})/(qN_{well})} \\ \frac{R_{128}}{R_{1}} &\approx \left[1 + \frac{1}{t_{0}}\sqrt{\frac{2\epsilon_{si}}{qN_{well}}\phi_{B}}\left(1 - \sqrt{1 + \frac{V_{1}}{\phi_{B}}}\right)\right] \left[1 - \frac{1}{t_{0}}\sqrt{\frac{2\epsilon_{si}}{qN_{well}}\phi_{B}}\left(1 - \sqrt{1 + \frac{V_{2}}{\phi_{B}}}\right)\right] \\ &\approx 1 + \frac{1}{t_{0}}\sqrt{\frac{2\epsilon_{si}}{qN_{well}}\phi_{B}}\left(\sqrt{1 + \frac{V_{2}}{\phi_{B}}} - \sqrt{1 + \frac{V_{1}}{\phi_{B}}}\right). \end{split} \tag{19.13}$$



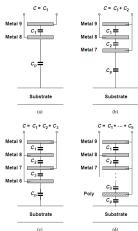
Other Resistor(2)

- P+/N+ source/drain region has a sheet resistance of 20 to 30 ohms per square. Silicided S/D regions are suited to low-value resistor. The junction to the bulk introduces capacitance and voltage dependence.
- Silicided polysilicon has a sheet resistance of 20 to 30 ohms per square for low resistor values. Its sheet resistance is process-dependent.

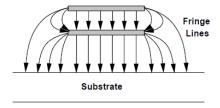
• Metal layers can provide very low resistor values. e.g. high speed A/D converters, the ladder may be constructed as simply a long metal line having equally spaced taps.

Capacitors

- Linear capacitors are designed using sandwiches made of the available conductive layer.
- Topology is determined by two factors:
 - the area occupied by the capacitor
 - Cn/C
- C between consecutive metal layers about 35 to 49 aF/um2 between metal 1 and poly about 60aF/um2.
- The ratio of bottom-plate C and interplate C reaches a minimum about 5 to 10% for the structure(b) or (c) and increases to 20% in (d).
- Gate-oxide is well controlled, thus
 (d) typically suffers from less variation.



Fringe Capacitance



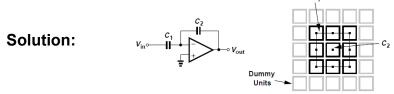
- The electric field lines terminate on the edge of other plate.
- Fringe capacitance must be taken into account

MOS Transistor Capacitor

- With source and drain tied together, one MOSFET is a simple capacitor
- Voltage dependence of the capacitance leads to nonlinearity issues.
- Layout of capacitor for high-precision circuits may follow transistors and resistors. E,g., dummy devices.

Example 19.4

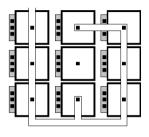
The circuit is designed for a nominal gain of C1/C2 = 8. How should C1 and C2 be laid out to ensure precise definition of the gain?



We form C1 as 8 unit capacitor, each equal to C2, and Place all of the units in a square array.

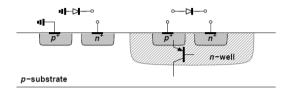
Dummy capacitor units are placed around the main array, creating approximately the same environment.

Layout of Capacitors with Interconnections



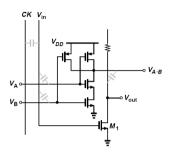
- For large capacitors, the cross-coupling techniques may apply.
- Capacitors are sensitive to interconnections, thus some error may adds to our desired value.

Diodes



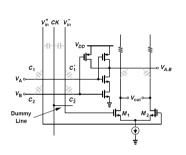
- Two types of PN junctions can be formed in standard CMOS technology
 - in the p substrate(reverse biased as a varactor)
 - in an n-well
- In an n-well, it may form a pnp transistor produces current to the substrate
- Analog CMOS circuits rarely incorporate forwardbiased diodes(except in bandgap circuits)

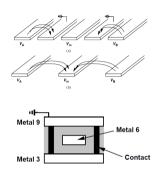
Interconnects Capacitance



- The clock signal distributed over long wires experiences significant line capacitance.
- Capacitance between lines introduces substantial coupling of signals. E.g., even though the coupling cap may be small, the voltage swing is big on clk or digital input

Crosstalk Reduce

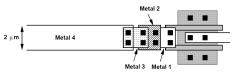




- Differential signaling converts most of the crosstalk to common-mode disturbance
- Sensitive signals can be "shielded" in the layout
 - shielding by ground is better than spacing
 - introduce capacitance and complex wiring

Interconnect Resistance

- The sheet resistances may introduce substantial thermal noise for low-noise applications.
- The contacts and vias also suffer from high R. E.g., Example 19.5, a 100um metal 4 line is connected to vias and contacts. Calculate the thermal noise of that.

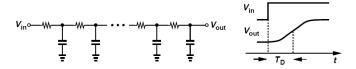


Solution: Assuming sheet resistance 40mohm per square for metal 4, a via resistance of 50hm and a poly contact of 30ohm. Rtotal = 24.50hm

The thermal noise voltage is equal to

 $0.64 \text{ nV}/\sqrt{\text{Hz}}$

Distributed RC in Long Interconnects



- The distributed RC of long interconnects may introduce significant delay and "dispersion" in signals. $T_D = \frac{1}{3}R_uC_uL^2$
- E.g. if the delay from left to right is unequal, the levels sampled are distorted.

 Or if CK is no longer rectangular.
 - Or if CK is no longer rectangular, the signals are also distorted.

"Dispersion" over a Line

- "Dispersion" refers to the significant increase in the transition time of the signal as it propagates through a line, a particularly troublesome effect if a clock edge is to define a sampling point.
- The clock edge can be sharpened by inserting an inverter
- · IC chips attention,

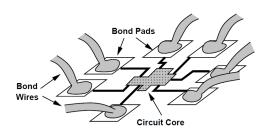


- The DC or transient voltage drops along the

buses may be significant

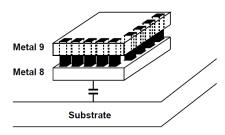
- Electromigration calls for a minimum line width
- Series resistance can be reduced by connect multiple layer, yet making the routing more difficult.
- Tapered ground for constant voltage drop

Bonding Pads to a Chip



- Large "pads" are placed on the perimeter of the chip
- Pad dimensions and structure are dictated by the reliability issues and margin for manufacturing tolerances
- Pad dimensions must be minimized so as to reduce both the capacitance of the pad to the substrate and the total die area

"Lift-off" during Boding

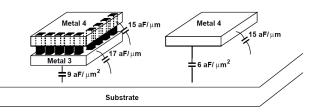


- A simple pad would consist of only a square made of the top metal layer. However, such a structure may "lift-off".
- A modified structure is formed by the two topmost metal layers, connected to each other by many small vias. Yet it suffers from a larger capacitance.

Example 19.6

Calculate the capacitance of a metal-4 pad and a metal-4/metal-3 pad. Assume dimensions of 75um*75um.





For metal-4

For metal-4/me

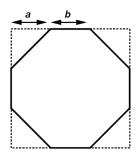
$$C_{tot} = 75^2 \times 6 + 75 \times 4 \times 15$$

= 38.25 fF.

$$C_{tot} = 75^2 \times 9 + 75 \times 4 \times (17 + 15)$$

= 60.22 fF.

High-frequency Signals Pads

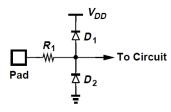


- The pads carrying HF signals can be configured as octagons to reduce capacitance
- The area is reduced by 20%
- Removing the corner of a square pad is applicable in the CMOS process

Electrostatic Discharge(ESD)

- When an external object having a high potential touches one of the connections to the circuit, the ESD produces a large voltage to the devices.
- E.g., If ICs are handled by human beings
- ESD may occur even without actual contact
- MOS devices sustain two types of permanent damage as a result of ESD.
 - The gate oxide may break down, leading to a very low resistance between the gate and the channel
 - The S/D junction diodes may melt if they carry a large current in forward or reverse bias, creating a short to bulk

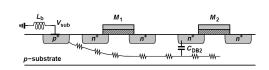
ESD Protection

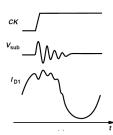


- Devices clamp the external discharge to ground or VDD, limiting the potential applied to the circuit.
- Resistor R1 is necessary so as to avoid damaging D1,D2 due to large currents
- ESD protection issues:
 - Introduce substantial capacitance, degrading speed and impedance matching, or noise figure
 - Parasitic capacitance of the ESD may couple noise
 - May lead to latch up in the circuit design

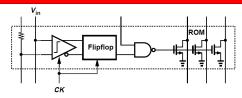
Substrate Coupling

- CMOS technologies use a heavily-doped p+ substrate to minimize latchup susceptibility, but it creates unwanted paths between various devices due to the low resistivity and produces "substrate coupling".
- Substrate coupling influences body effect, varying the threshold voltage of M1
- Many "noise" generators exist





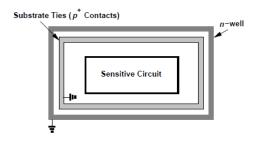
Reduce Substrate Coupling



- Increasing physical spacing between sensitive building blocks may not be applicable. The system performance may degraded due to long interconnects.
- Minimize the effect of substrate noise
 - Use differential operation
 - Digital signals and clocks should be distributed in complementary form
 - Critical operations, e.g., sampling or charge transfer should be performed after clock transition
 - Minimize the bond wire inductance
 - Use PMOS differential input

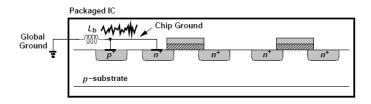
Guard Rings

- On lightly-doped substrates, "guard rings" can be employed to isolate the sensitive sections from the substrate noise
- Substrate ties or n-well with its large depth



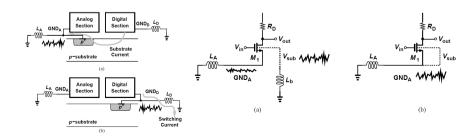
Avoid Substrate "Bounce"

- The ground and the substrate should be connected on the chip and brought out through a single wire
- Two difficulties
 - Most circuits have a digital ground and an analog ground. To which should we refer.
 - Define a specific reference potential for the input signals is hard



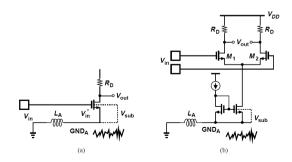
Which Ground to Refer to?

- If the analog ground is used, the substrate noise current must flow through LA, creating noise on GNDA, and if the digital ground is used, the substrate voltage is heavily disturbed by the large noise on GNDD
- Connecting both is not acceptable
- (a) is usually preferable for stable drain current, but simulation is needed for realistic situation

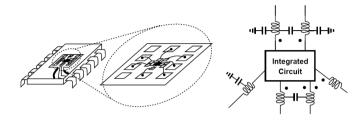


Define a Stable Reference Potential

- The reference point changes from the off-chip ground to the on-chip ground
- The effect is less pronounced in the differential structure but in high-precision application.
- Asymmetries in the circuit and interconnections convert a fraction of the common-mode noise to a differential component

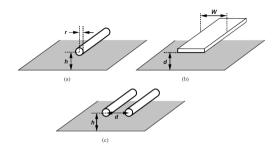


Packaging



- The parasitics associated with the package and connections to the chip introduce difficulties in the high speed or high accuracy.
- Trace self-inductance, mutual inductance, capacitance limits the performance of the package
- Packaging continues to limit the achievable performance

Common Geometries in Packaging



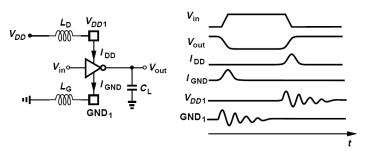
- Circuit model for self- and mutual- inductance
- A round wire $L \approx 0.2 \ln \frac{2h}{m}$ nH/mm,
- A flat trance $L \approx \frac{1.6}{K_f} \cdot \frac{d}{W} \text{ nH/mm},$
- Two round wires $L_m = 0.1 \ln \left[1 + \left(\frac{2h}{d} \right)^2 \right]$ nH/mm. • Requiring electromagnetic near simulation for
- Requiring electromagnetic liefu simulation for proper modeling

Self-Inductance

• Self-inductance with a total value between 2nH and 20nH. E.g., CL = 0.5 pF $C\Delta V/\Delta t = 3$ mA. Voltage drop: $L\Delta I/\Delta t = 6 \times 10^6 L$ Around 30mV if LD

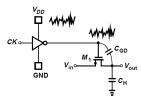
and LG = 5nH

- Differential pair could reduce noise and bounce
- Thus, we must separate analog and digital supply



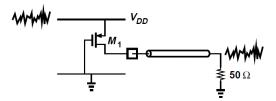
Analog/Digital Supply Separation

- Separating analog and digital supply is not always straightforward
- E.g., supposing a sampling circuit is clocked by an inverter, which power supply we should choose?
 - from digital power lines, the noise would couple
 - from analog power lines, many these circuit would send noise due to transient currents.
 - May have third power line. Signal and P/G line ratio is around 1:1 in today's design



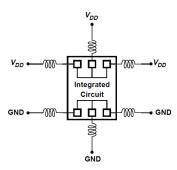
Monitor the Supply Noise

- It is desirable to monitor the supply noise for characterization and troubleshooting purposes.
- A PMOS senses the noise between the on-chip supply and ground lines injects a current into an external 50ohm transmission line.



Multiple Pads, Bonding Wires and Pins

- In cases where a single connection to the chip sustains a prohibitively large transient voltage, multiple pads, bond wires, and package pins are used.
- The equivalent inductance are reduced by connected in parallel.



Example 19.7

In a 600MHz, 2V CMOS microprocessor containing 15 million transistor, the supply current varies by 25A in approximately 5ns. If the processor provides 200 bond wires for ground and 200 for VDD, estimate the resulting supply bounce.

Solution:

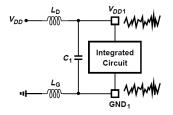
Assuming a total inductance of 5nH for each bond wire and its corresponding package trace and pin, we have

$$\Delta V = L \frac{\Delta I}{\Delta t} = \frac{5 \times 10^{-9}}{200} \cdot \frac{25}{5 \times 10^{-9}}$$

In t... = 125 mV supply and ground bounce add in phase, a total noise is 250mV. An external decoupling capacitance is placed on top.

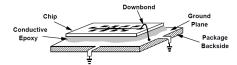
On-chip Capacitor

- High transient currents drawn from the supply make it difficult to maintain a small bounce
- A large on-chip capacitor may be used to stabilize the difference between VDD and ground
- · On-chip decoupling problem
 - May resonate at the CK freq. or its harmonics and amplify the noise(need series R or MOS C)
 - Insufficient decoupling leads to slower ringing
 - The yield suffers from larger area



Flip Chip

- Some modern package contains a metal ground plane to which the die can attached by conductive epoxy.
- Such packages avoid bond wires and long, narrow traces in the substrate connection.
- Even in some high-end design, the board ground can be directly attached to the board ground.
- "Downbonded" to the underlying plane minimizes the inductance, at the expense of cost.



Effect of Self-inductance

- The inductance along with the pad capacitance and the circuit's input capacitance forms a lowpass filter, attenuating HF components.
- E.g., the multiply-by-two circuit's settling speed may be limited.
 - The package inductance degrades the settling behavior for some accurate voltage reference
- Impact the performance of digital buffers

Example 19.8(1)

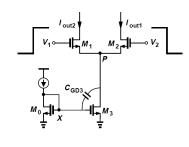
The circuit routes its tail current to either of the outputs according to the large swings controlling the gates of M1 and M2. Explain what happens at node X. If the tail current of a large number of these pairs feed from node X, should voltage be provided externally?

Solution: When M1 is off,

$$V_{P1} = V_2 - \sqrt{2}(V_{GS} - V_{TH})_{eq}$$
 III equilibrium,

$$V_{P2} = \frac{V_1 + V_2}{2} - (V_{GS} - V_{TH})_{eq}$$

$$V_{P2} = V_2 - \left(1 + \frac{\sqrt{2}}{2}\right)(V_{GS} - V_{TH})_{eq}$$



the voltage is coupled to node X

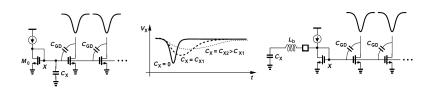
$$\mathbf{1}_{(1-\sqrt{2}/2)(V_{GS}-V_{TH})_{eq}}$$
drain

Example 19.8(2)

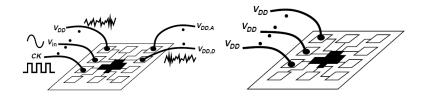
If the tail current of a large number of these pairs feed from node X, should voltage be provided externally?

Solution:

The disturbance may be quite significant, demanding that a decoupling capacitor be connected from X to ground. However, it may possibly degrade overall speed. Also, a large CX is placed off-chip may introduce inductance, too. In some cases, X would be just left agile.

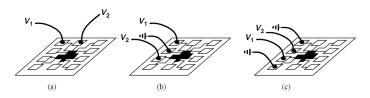


Mutual Inductance



- Power noise also shows through mutual inductance of bond wires and package traces.
- The design of the pad frame and position plays a critical role.
- In parallel lines, mutual inductance also manifests itself.

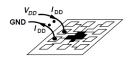
Reduce Mutual Inductance



- Two methods can reduce the mutual coupling
 - First, the wires can be connected such that they are

perpendicular to each other

- Ground or supply lines can be interposed between critical bond wires
- Reduce inductance by carrying opposite current



Self- and Mutual Capacitance

- The capacitance seen from each trace of the package to ground may limit the input bandwidth of the circuit or load the preceding stage
- LC may yield a finite resonance frequency stimulated by various transient currents
- The high Q LC would give rise to a sharp resonance and amplify the noise considerably