CMOS Transistor Layout KungFu

Lee Eng Han
Valerio B. Perez
Mark Lambert Cayanes
Mary Grace Salaber

Copyright © 2005 by Lee Eng Han.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, mechanical, photo-copying, recording, or otherwise, without the prior written permission from Lee Eng Han, who is one of the authors of the publication.

Table of Contents

Preface

1. Introduction 1

2. MOS Transistors 2

3. Fabrication of MOS Transistor 5

4. Layout a Single Transistor 11

First Stroke — The basic transistor layout 12

Second Stroke — Compact the transistor layout 13

Third Stroke — Speed up the transistor 17

Fourth Stroke — Clean up the substrate Disturbances 20

Fifth Stroke — Balancing area, speed and noise 26

Sixth Stroke — Relief the stress 29

Seventh Stroke — Protect the gate 30

Eighth Stroke — Improve yield 32

SK hynix Ex. 1016, Page 3 of 39

5. Layout Several Transistors 34

Eighth Stroke Re-visit — Improve yield 35

Ninth Stroke — Close proximity 36

Tenth Stroke — Interdigitated layout 36

Eleventh Stroke — Dummy transistor 41

Twelfth Stroke — Two-dimension interdigitated layout 43

Thirteenth Stroke — Guard ring for the matched transistors ... 45

Fourteenth Stroke — Keep NMOS away from N-well 45

Fifteenth Stroke — Orientate the transistor 46

Sixteenth Stroke — Match the interconnects 47

Seventeenth Stroke — The unmatchable 50

6. Verifying the Transistor Layout 52

Eighteenth Stroke — Physical verification beyond DRC and LVS 61

Bibliography 64

SK hynix Ex. 1016, Page 4 of 39

Preface

Many IC design books emphasize on circuit design theories and there is little coverage on custom circuit layout techniques. Hence this book is specially written to focus on the custom circuit layout techniques. It is an easy book for the custom layout engineer as prior knowledge on circuit design is not necessarily required to understand the content of the book.

This book is written for two groups of audiences – New & Experienced custom layout designers. The new comer to custom layout would benefit from the wider perspectives in implementing the design to a layout. The experienced custom layout designers, on the other hand, will have a better appreciation of the rationale behind the layout practices.

Feedback to the authors

Please feel free to let us have your valuable inputs for future improvement. We would also appreciate if you could let us know if the book is of value to you. We can be contacted at layout@eda-utilities.com. Your feedback is most welcome.

Acknowledgements

The authors wish to thank Hong Siar Lim from Future Technology Design Solutions, and Hwee Ling Goh from STMicroelectronics for reviewing the technical details. The authors also like to thank Karen Phang for her effort in proof-reading the book and giving invaluable suggestions in making the book easier to read.

SK hynix Ex. 1016, Page 5 of 39

Chapter 1

Introduction

Welcome to *Layout KungFu*!

There are many levels of *Layout KungFu*, but we will focus on the fundamental and CMOS transistor layout is what you will find in this KungFu book. If you enjoy the KungFu and want to find out more, we could start a *Layout KungFu* series!

Different design requirements and process technologies pose great challenges for layout implementation. A set of *Layout KungFu* that works well before might be inappropriate for the next project that is targeted for a different application or a different process technology. A better understanding of the fundamental behind the layout techniques will allow you to apply the techniques efficiently.

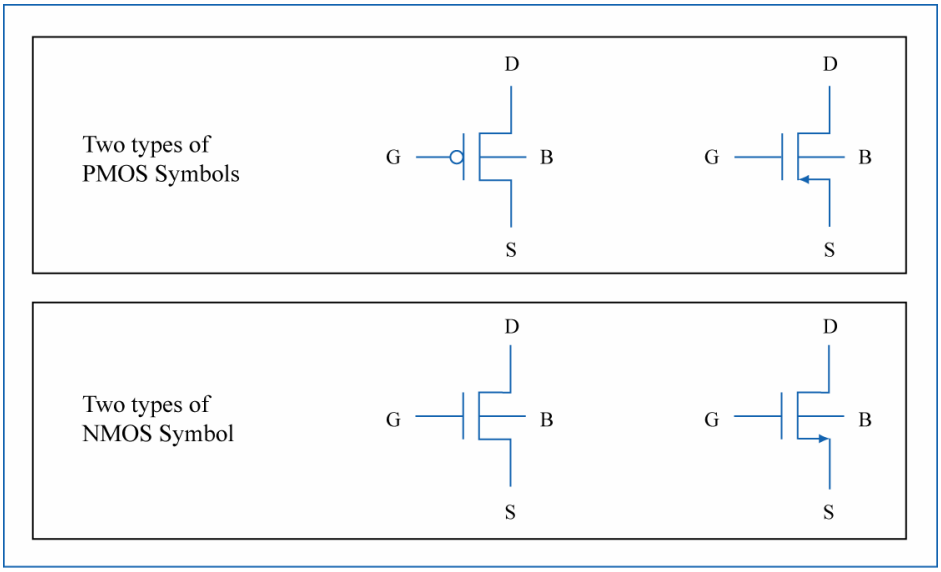
Please place your feet firmly on the floor, bend your knee at 90 degree and take a deep breath. Now, are you sitting comfortably on the chair? We shall begin the KungFu journey.

SK hynix Ex. 1016, Page 7 of 39

Chapter 2

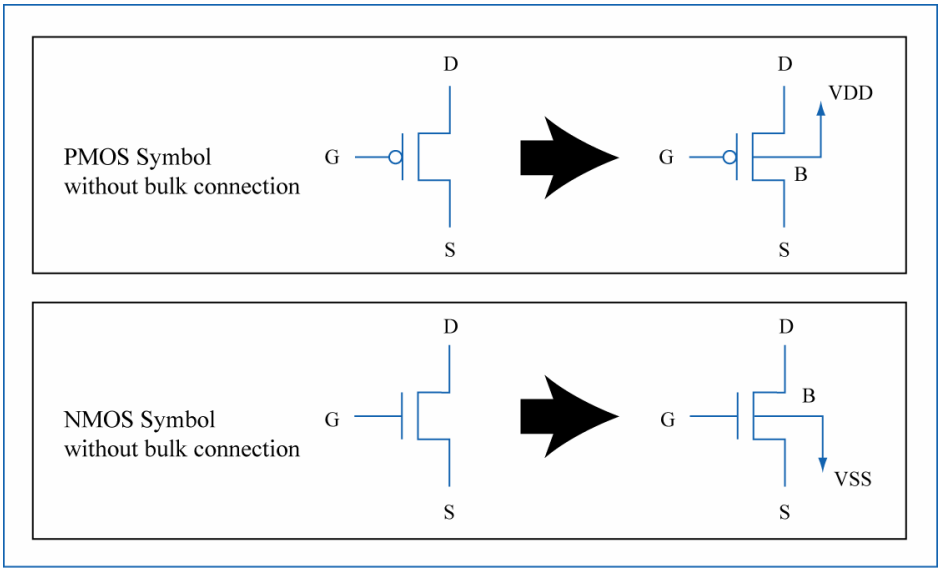
MOS Transistor

There are two types of MOS transistors. They are called n-channel MOS transistor (NMOS) and p-channel MOS transistor (PMOS). Each transistor has 4 terminals, namely drain (D), gate (G), source (S) and bulk (B) as illustrated in the transistor symbols.

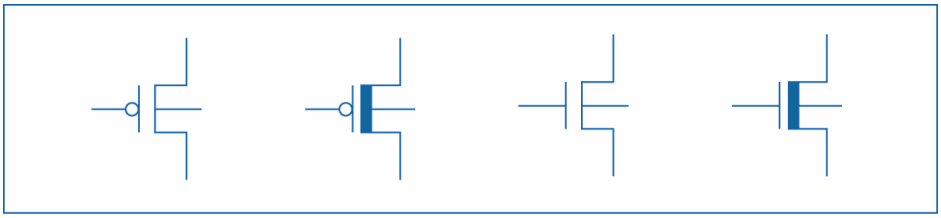


The bulks of the PMOS and the NMOS are usually connected to power and ground respectively. If the bulk terminal is omitted from the schematic symbol, the connections can be assumed to be what is shown in the following diagram.

SK hynix Ex. 1016, Page 8 of 39



A circuit design usually uses more than one type of PMOS or NMOS to cater for different power supply voltages. For example, a design may use thicker gate-oxide transistors to operate in higher power supply at the I/O interfaces. In order to differentiate the various voltage range transistors, sometimes circuit designers conveniently make use of the "depletion-mode" transistor symbol for this purpose. Typically, the schematic symbols for the depletion-mode transistors are represented with a thicker gate drawing.



The MOS transistor's performance varies with its channel length (L) and channel width (W). The drain current (I_D) that flows through the transistor operating in the saturation mode is shown in the following equation.

SK hynix Ex. 1016, Page 9 of 39

$$I_D = K * \left(\frac{W}{L}\right) * \left(V_{GS} - V_t\right)^2 * \left(1 + \lambda V_{DS}\right)$$

where K and λ can be taken as process technology constants.

Note that I_D is proportional to the ratio of W over L. Typically L is kept to the minimum dimension allowed in the design rule, and is to be layout exactly as indicated in the schematic. However, this is not always the case for W. We will elaborate more in chapter 4.

Extra Punch

Transistor Spice Model

The Fab supports different types of transistor. For example, transistors can have different types of V_T to compromise between leakage power and speed. Transistors can also have different types of gate oxide thickness to allow the transistors to operate in different voltage range.

Every type of transistor is associated with its own transistor spice model. PMOS and NMOS also have different transistor model. A transistor model includes a set of parameters that define the electrical performance of the transistor. The design engineers use the transistor spice model and the circuit netlist to simulate their design. The layout engineer has to craft the transistor layout that match the transistor spice model and the designed W and L.

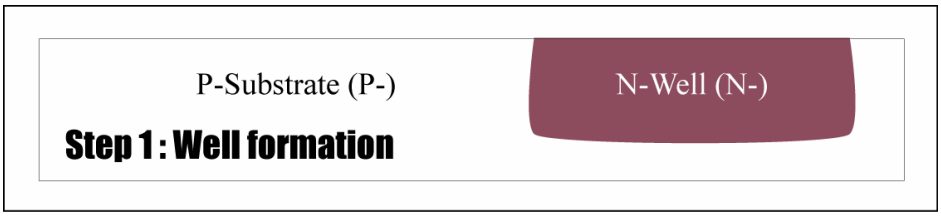
SK hynix Ex. 1016, Page 10 of 39

Chapter 3

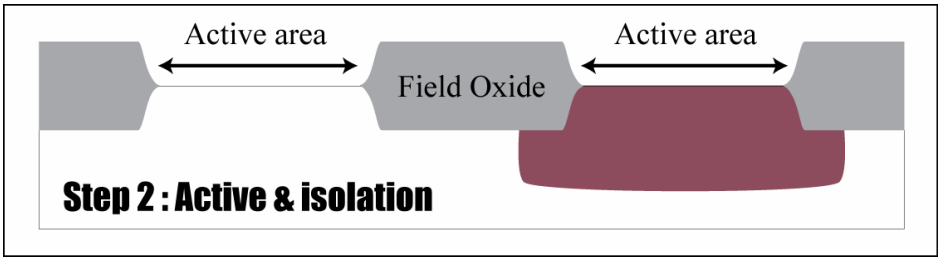
Fabrication of MOS Transistor

This chapter briefly describes a simplified version of the fabrication of a transistor on the silicon wafer. The ability to visualize the cross-section of a layout is a basic skill that all layout designers should master.

Step 1 : Well formation stage. Implants n-type impurities into the wafer followed by diffusing the impurities deep into the substrate to form the N-Wells. For CMOS process, the silicon substrate is usually p-type.



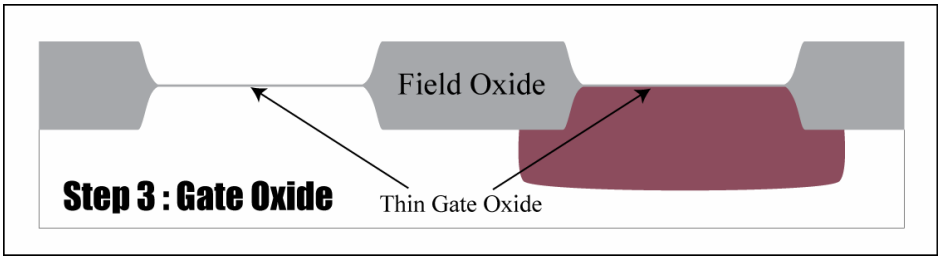
Step 2 : Active & isolation stage. Thick oxide is grown outside the active areas. Active areas are defined as areas where the CMOS transistors are fabricated. Thick oxide is also known as field oxide. Field oxides isolate the transistors from one another.



SK hynix Ex. 1016, Page 11 of 39

The first two steps describe a formation of a conventional well. The depth and doping profile of a conventional well are controlled by the diffusion drive-in at high temperature. A better way to form the well, known as retrograde well, is usually used in 0.25um[1] and smaller process technologies. Retrograde well is formed by very high energy implantation. The depth and doping profile of a retrograde well are controlled by implantation energy and impurity dose. Retrograde well is formed AFTER the field oxide. Since retrograde well does not require diffusion drive-in, it has smaller lateral diffusion and a more ideal doping profile.

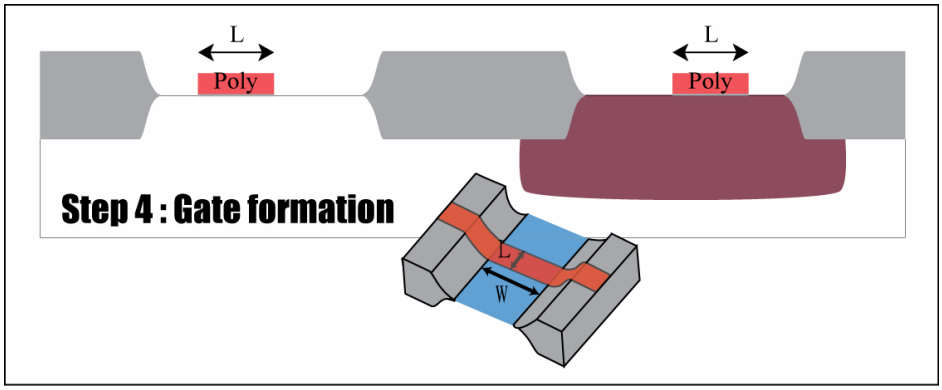
Step 3 : Gate oxide formation stage. A thin gate oxide is grown across the wafer. Gate oxide of only tens of silicon oxide atoms thick is created during the fabrication process with the current technology. Gate oxide is the insulator between the transistor's gate and its channel. Gate oxide refers to the "O" in "MOS" which stands for **M**etal-**O**xide-**S**emiconductor.



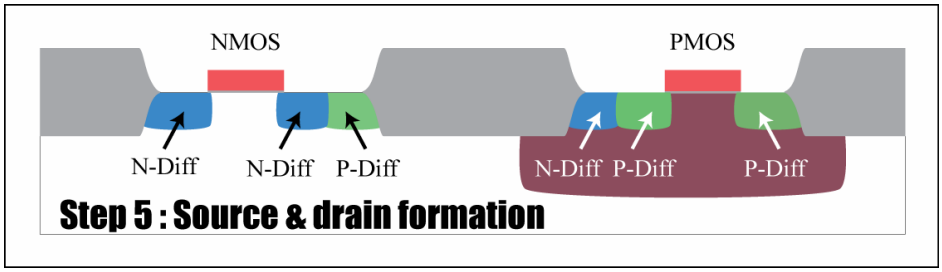
Step 4 : Gate formation stage. Poly (i.e. poly-silicon) is deposited on the wafer. The poly that are deposited on the gate oxides are the gates of the transistors which are usually known as gate poly. The gate poly will incline upward when it extends over the field oxide. The gate oxide in the active area that are not covered by the gate poly will be etched away to form the source and the drain of the transistor.

[1] A process technology of 0.25um means that the shortest channel length (L) of a transistor is 0.25um.

SK hynix Ex. 1016, Page 12 of 39

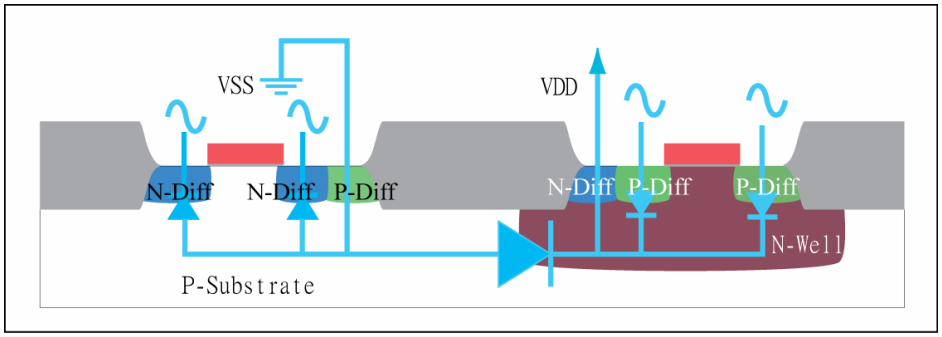


Step 5 : Source and drain formation stage. P-type and n-type impurities are implanted into the active areas. The impurities are diffused into the silicon to form the source terminals and the drain terminals. As the impurities diffuse both vertically and laterally, the gate poly will slightly overlap the sources and the drains which will result in gate overlap capacitances. The diffusions for the sources and the drains of NMOS and PMOS are N-diffusion (N-diff) and P-diffusion (P-diff) respectively.

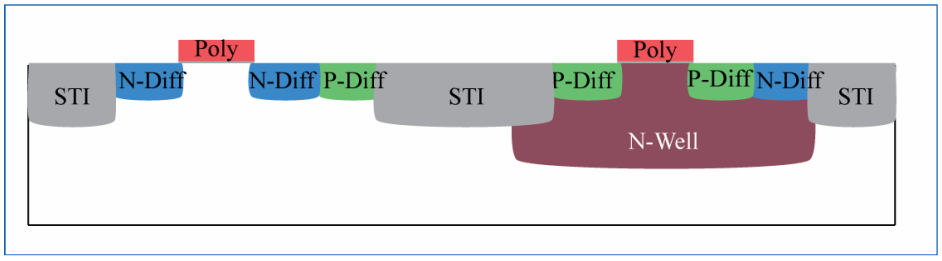


P-diff in p-substrate is known as p-tap, while n-diff in N-well is known as n-tap. Connections from the metal routings to the substrate and the N-wells are made through the p-tap and the n-tap. This is necessary to ensure the wells are properly tied down and the transistors are isolated. The p-substrate should be biased to the lowest voltage potential while the N-well should be biased to the highest voltage potential. In this way, all the P-N junctions are reverse biased and hence the transistors are electrically isolated from one another as shown in the diagram below.

SK hynix Ex. 1016, Page 13 of 39



Isolating the transistors with the thick field oxide is commonly found in 0.35um and larger process technologies. For 0.25um and smaller process technologies, shallow trench isolation (STI) shown in the diagram below is more commonly used to isolate the transistors. In STI fabrication, trenches are etched into the wafer and filled with silicon oxide to isolate the islands of transistor active area.

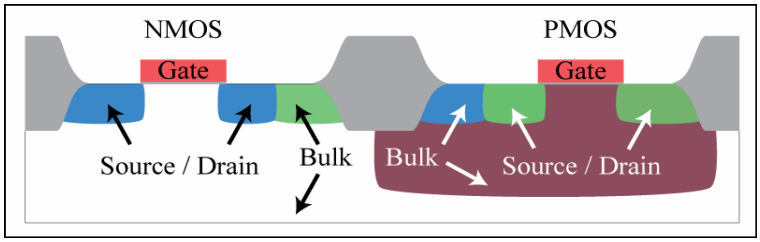


SK hynix Ex. 1016, Page 14 of 39

Source, Drain Gate and Bulk

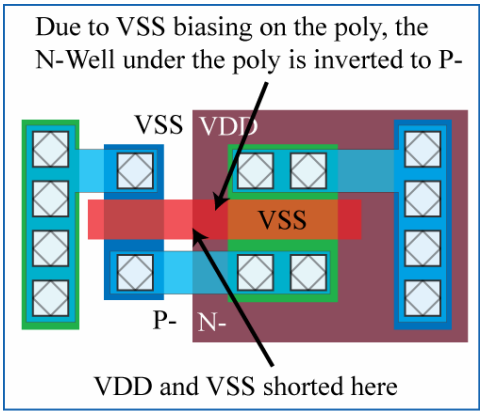
The drains, sources, gates and bulks of the NMOS and the PMOS are illustrated in the figure below. Observe the cross section of the layout and you will find the followings.

- The drain and the source are fabricated in the same way.
- The bulk of all the NMOS are connected together.
- The bulk of all the PMOS in the same N-well are connected together.



Quiz I

Some layout books may have suggested that the poly should not cross the N-well boundary. However, the following inverter layout that is commonly used shown otherwise. As indicated in the layout, a VSS bias on the ploy will not only form a P-channel under the gate, it will also cause the N-well under the poly to invert to P-type. Hence, it is possible to short VDD to VSS. However, this conclusion is incorrect. What is wrong with the conclusion? If the answer is not obvious to you, you may try to draw the cross-section of the layout at the location where the ploy crosses the N-well boundary. If you still need help, you can download the explanation at www.eda-utilities.com/layout-kungfu-quiz-I.pdf



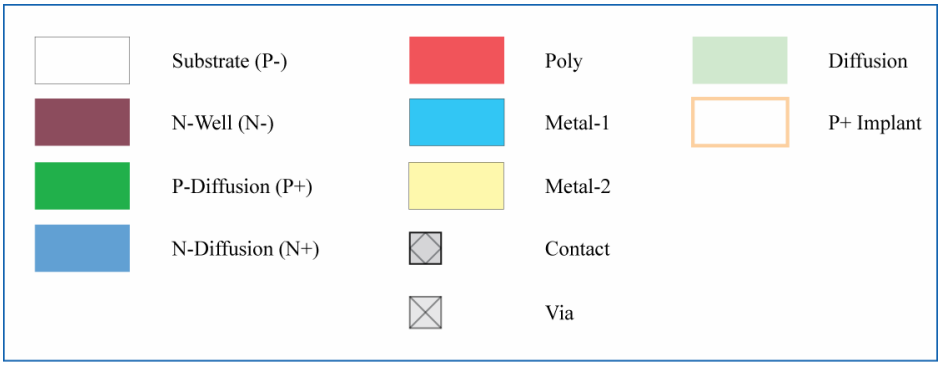
SK hynix Ex. 1016, Page 16 of 39

Chapter 4

Layout a Single Transistor

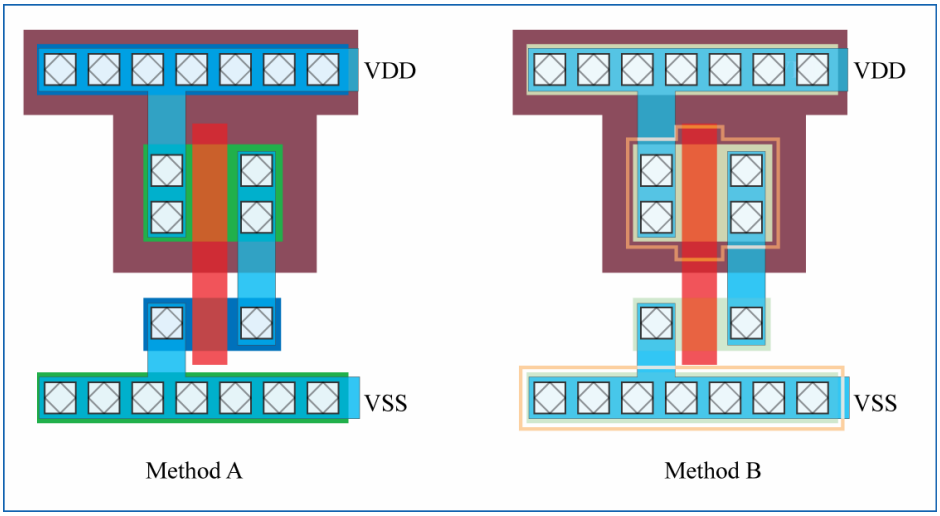
Before we can start to layout the transistor, we have to settle a few logistics.

Firstly, the following legends are adopted.



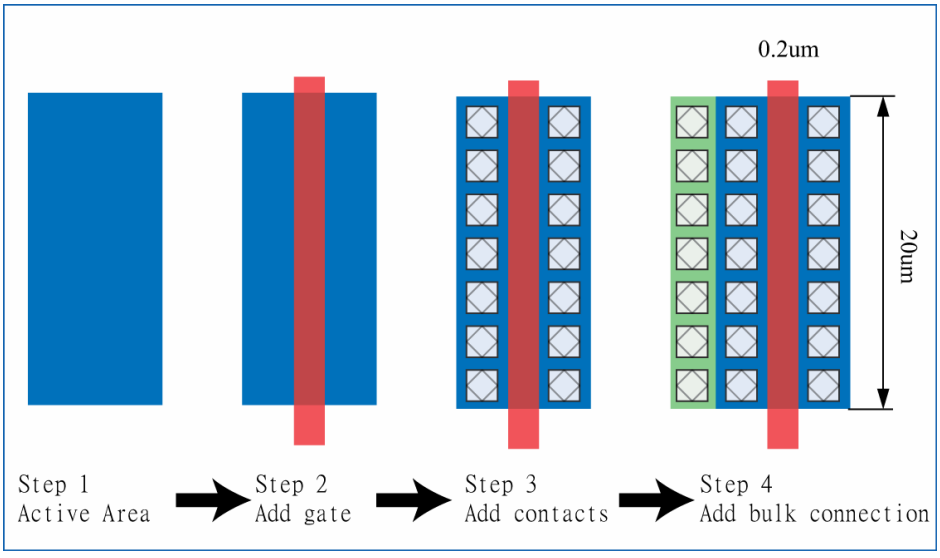
Secondly, there are more than one ways to draw an identical layout. For example, the two layouts in the following diagram are the same. Layout in “Method A” uses P-diffusion layer and N-diffusion layer. Layout in “Method B” uses diffusion layer and implant layer. The illustrations in this book use “Method A”.

SK hynix Ex. 1016, Page 17 of 39



First Stroke. The basic transistor layout

The basic transistor layout as illustrated has a channel length (L) of 0.2um and a channel width (W) of 20um. The source diffusion and the drain diffusion should be filled with the maximum number of contacts to reduce the resistance of the connection from the metal to the diffusion, and to maximize the amount of current that can flow through the contacts.



SK hynix Ex. 1016, Page 18 of 39

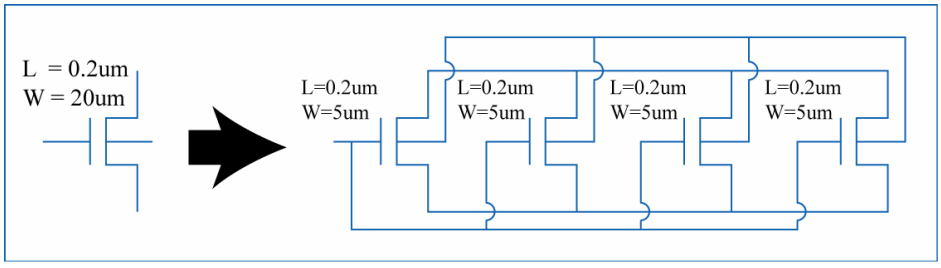
Extra Punch

Bulk Connection

The substrate and the N-well are doped lightly. A direct connection from the metal routing layer to the bulk is not allowed. The connection should be made through a higher doped diffusion such as the p-diffusion and the n-diffusion in order to establish a good contact. In some fabrication process, two additional layers are used for these connections. These layers have much higher doping than the diffusions for the source and the drain.

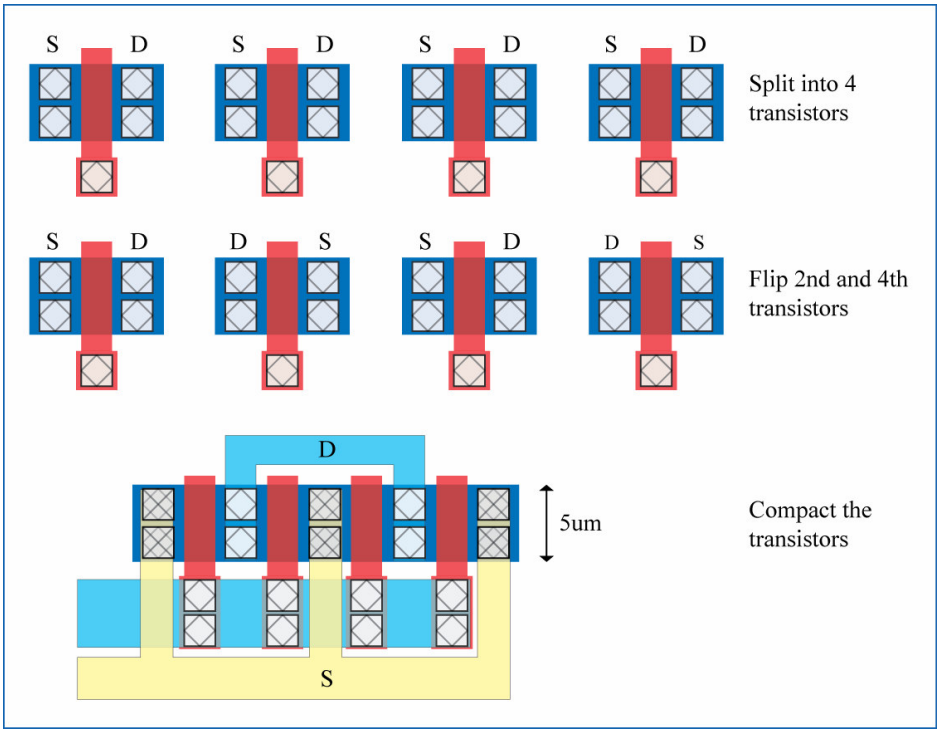
Second Stroke. Compact the transistor layout

The basic transistor layout from the first stroke has a rather awkward aspect ratio. Putting together transistors with fixed aspect ratio will not result in a compact layout. Fortunately, the aspect ratio of the transistor can be modified by using the transistor current equation shown in page 3. For example, the transistor with a width of 20um and a length of 0.2um is similar to having four transistors connected in parallel, each with a width of 5um and a length of 0.2um.



The next layout shows the transistor with four fingers. The layout has a better aspect ratio than the one in the first stroke. Note that the connection to the bulk is omitted in the layout so as to simplify the drawing. Bulk connection will be discussed later.

SK hynix Ex. 1016, Page 19 of 39

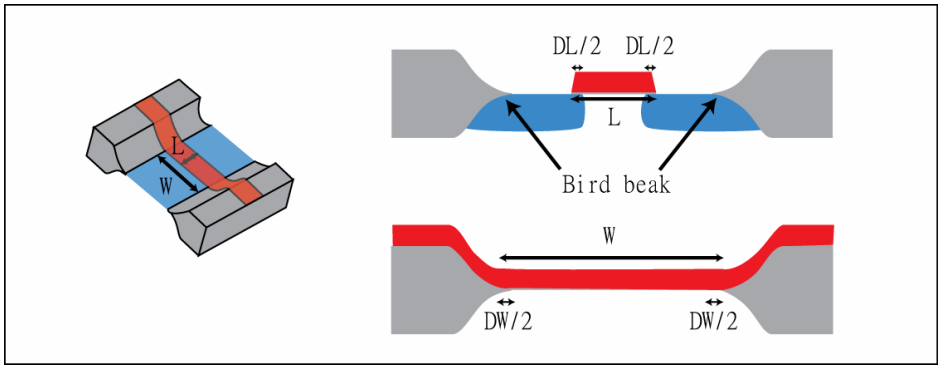


So far everything looks fine. **But it is not**. The transistor with a width of 20um and a length of 0.2um is not exactly the same as having four transistors that are connected in parallel, each with a width of 5um and a length of 0.2um.

Layout cannot be fabricated exactly as drawn in the layout due to the limitations in the manufacturing process, such as process tolerances and mask misalignment. Some of the manufacturing limitations are captured in the Spice transistor model. Two of the main parameters in the Spice transistor model are DW and DL. DW shows the delta difference of drawn W from effective W. DL shows the delta difference of drawn L from effective L.

The next diagram shows the transistor effective channel length can be affected by under-etching or over-etching of the poly, as well as the amount of lateral diffusion under the gate. The effective channel width of the transistor is affected by the “bird peak” of the isolation scheme. In addition, the inclination of the gate poly up to the field oxide makes it difficult to determine the exact width of the transistor.

SK hynix Ex. 1016, Page 20 of 39



Considering an example where DL is 0.015um and DW is 0.045um. To simulate a "fast" corner transistor, the transistor is modeled to have a narrower L and a wider W. On the contrary, a "slow" corner transistor is modeled with a wider L and a smaller W. Hence, at the fast corner DL is negative (i.e. -0.015um) and DW is positive while at the slow corner DL is positive and DW is negative (i.e. -0.045um). The following table shows L and W of the transistors from the first stroke and the second stroke, and at slow, typical and fast corners to emulate manufacturing process tolerances.

		L (um)	W (um)
Transistor from first stroke	Fast corner	0.185	20.045
	Typical corner	0.200	20.000
	Slow corner	0.215	19.955
Transistor from second stroke	Fast corner	0.185	20.180
	Typical corner	0.200	20.000
	Slow corner	0.215	19.820

The width of the transistor from the first stroke differs by 0.045um between the slow or the fast corner from the typical corner. However, the width of the transistor from the second stroke differs by 0.18um between the slow or the fast corner from the typical corner. This could pose circuit performance deviation for the circuit designer if left unaccounted for.

The layout in the second stroke can be statistically significant. The circuit designer would perform Monte Carlo simulation to "center the design" so as to improve manufacturing yield. In Monte Carlo simulation, a small statistical variation is added to W and L of every

SK hynix Ex. 1016, Page 21 of 39

transistor in the circuit. Folding a transistor to four fingers means that a larger variation in the circuit performance since the variation made to the four fingers is four times larger than the same variation made to a single transistor. The layout can be challenging for the circuit designer to optimize the design for better yield.

Extra Punch

Operating Corners

Variations in fabrication process, ambient temperature and supply voltage affect the electrical performance of the transistors. For example, a higher temperature and a lower supply voltage make the transistor operate slower. Many companies require the circuit designers to verify the operation of the circuit design by simulating the design in slow (SS) corner, typical corner (TT) and fast corner (FF). Some companies also require the circuit designers to simulate the design with fast NMOS and slow PMOS corner (FS), and slow NMOS and fast PMOS corner (SF). Some companies require the circuit to work correctly within a 3 sigma spread from the typical corner (FF3, SS3).

At 0.13um and smaller technologies where transistor leakage current is significant, two more corners become important. They are the ML (max leakage) corner and TL (typical leakage) corner. ML happens at faster process, maximum supply voltage and maximum temperature. TL happens at typical process, typical supply voltage and maximum temperature. Leakage at ML corner can be more than double the leakage at TL corner, and one to two orders larger than the leakage at TT corner!

Is folding a transistor into multiple fingers a bad idea? **Actually, it is an excellent idea**. Everyone uses it. Second stroke makes the layout compact. Third stroke is a continuation of the second stroke and it will enable the design to run at a higher speed!

SK hynix Ex. 1016, Page 22 of 39

Before we leave the second stroke, remember that whatever tricks you applied in the layout, it must be communicated to the circuit designer. The circuit designer needs to update the schematic to reflect the layout implementation. Using the same layouts from the first stroke and the second stroke as an example, the original spice netlist (i.e. the first stroke) of the transistor is

```
M1 D G S B N_model L=0.2u W=20u
```

The 'N_model' in the netlist is the name of the NMOS transistor model. The spice netlist for the layout in the second stroke is

```
M1 D G S B N_model L=0.2u W=5u M=4
```

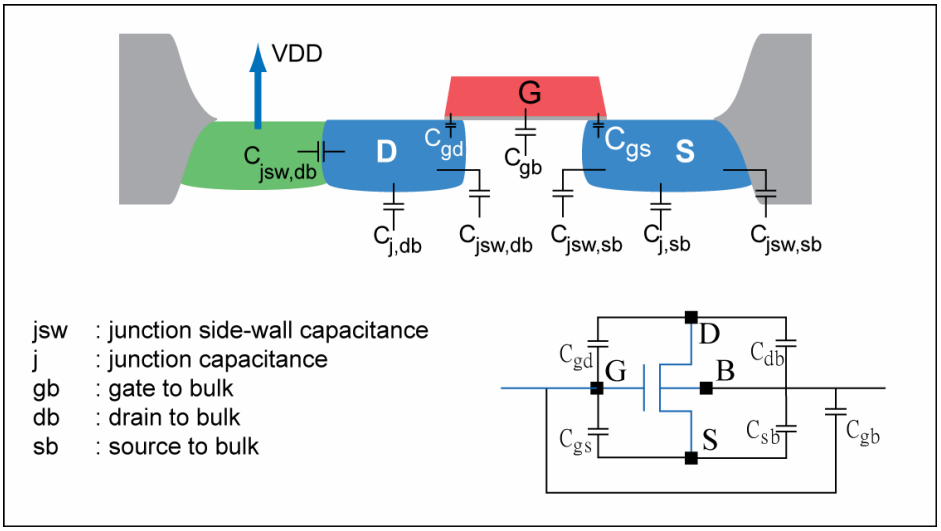
The circuit designer simulates the design using the spice netlist. Thus it is important for the spice netlist to represent the layout implementation as closely as possible.

Third Stroke. Speed up the transistor

What can you do in the layout to make the transistor operate faster?

Reducing the parasitic capacitance and resistance would increases the speed of the transistor. We need to understand where the parasitic capacitances and resistances are. The following diagram shows the simplified capacitances associated with the drain, gate and source of a transistor to the bulk.

SK hynix Ex. 1016, Page 23 of 39

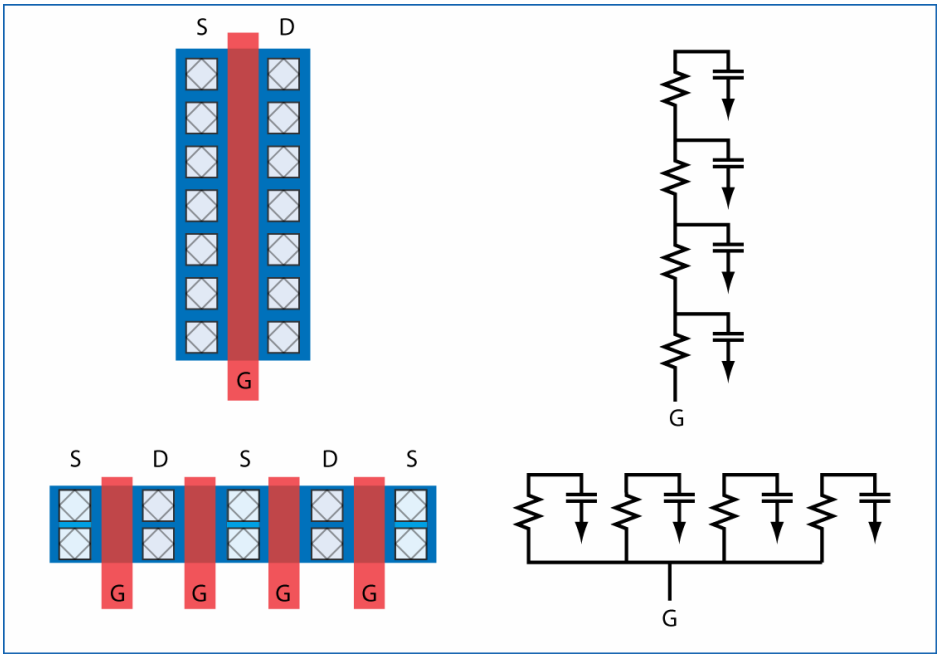


Overall C_{sb} is dependent on the area of source (AS) and perimeter of source (PS). Similarly, overall C_{db} is dependent on the area of drain (AD) and perimeter of drain (PD). Both C_{sb} and C_{db} have components that are dependent on the diffusions in the proximity. The values of AS, AD, PS and PD of a transistor can be extracted from the layout. Post-layout spice netlist should include these parameters.

The frequency response of the transistor can be improved if the source capacitance and drain capacitance are reduced. Study the transistors from the first stroke and the second stroke. Can you see that the transistor from the second stroke has the drain area reduced by half and the source area reduced by a quarter as compared to the transistor from the first stroke?

The folded transistors have smaller gate resistance as shown in the diagram below. This will make the transistors turn on and off faster than the transistor in the first stroke.

SK hynix Ex. 1016, Page 24 of 39

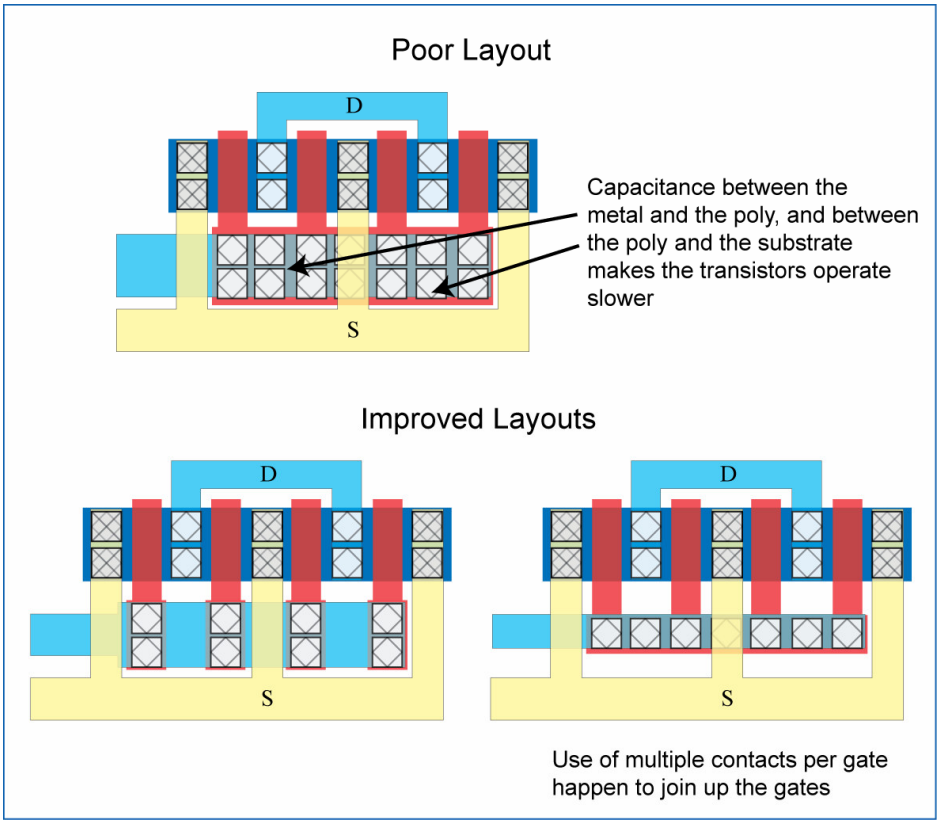


In general, try to fold a transistor to an even number of fingers. For example, the transistor from the second stroke is folded to four transistors, and is drawn to optimize the frequency response of the drain.[1]

Resistivity of the poly is a few orders higher than the resistivity of the metal. The parasitic capacitances between the poly and the substrate, and between the metal and the poly, are very much larger than the parasitic capacitance between the metal to the substrate. Hence, using poly for interconnect could degrade the frequency response of the transistor if the poly routing is not optimized carefully. Refer to the following layout. The layout at the top uses both poly and metal to connect the gates. The layout at the bottom-left uses only metal to connect the gates. The layout at the bottom-right is a popular method to have multiple contacts per gate. To improve yield, the contacts for the gates in this layout are placed slightly further from the transistors so as to increase the distance between the diffusion and the poly that are running in parallel to the diffusion[2]. The layout at the top will have poorer frequency response due to additional parasitic capacitances from the metal to the poly, and from the poly to the substrate.

[1] AD and PD are smaller than AS and PS respectively.
[2] More details on this in the eight stroke in both chapter 4 and chapter 5.

SK hynix Ex. 1016, Page 25 of 39



Fourth Stroke. Clean up the Substrate Disturbances

Analog design performances are sensitive to electrical disturbance. Disturbance in the substrate should be minimized as much as possible. Two common types of substrate disturbance are

- Disturbance from minority carrier
- Substrate coupling noise

Disturbance from minority carrier

Minority carriers are injected into the substrate from the source diffusions and the drain diffusions when

- The source potential or the drain potential of NMOS is below the substrate potential

SK hynix Ex. 1016, Page 26 of 39

- The source potential or the drain potential of PMOS is above the N-well potential

There are several possibilities for the above conditions to happen. Examples are

- Inductive ground path causes the ground in the substrate to bounce

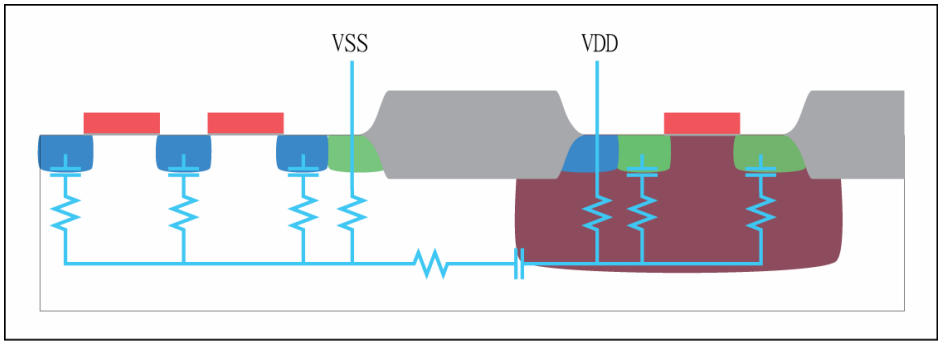
- Resistive power and ground path from the power pins to the substrate and the N-well

- Fast switching signal with significant overshoot

The drifting of the minority carriers in the substrate and the N-well create a potential difference that can affect the performance of the circuit, or trigger a latch-up.

Substrate coupling noise

A reverse biased diode has the electrical properties of a capacitor. Circuit signals can be coupled through the substrate as illustrated in the diagram below.



To reduce disturbances from minority carrier, you may use guard ring in the following configuration around noisy transistors.

- Surround NMOS in the p-substrate with N-well guard ring. Tie the N-well guard ring to VDD. The N-diffusions from the NMOS could inject stray electrons into the substrate. These stray

SK hynix Ex. 1016, Page 27 of 39

electrons could be collected efficiently by the N-well guard ring that is biased to VDD to attract the electrons.

- Surround the PMOS in the N-well with P-diffusion guard ring. Tie the P-diffusion guard ring to ground. P-diffusions from the PMOS inject stray holes into the N-well. These stray holes could be collected efficiently by the P-diffusion guard ring that is biased to ground to attract the holes.

For the guard rings to be effective, the resistance in the path from the straying minority carrier to the guard ring and then to the voltage source must be kept as low as possible. Hence, the minority carrier noise guard rings are made wider so as to decrease it resistance. Ideally, the guard rings should be placed as closely to the likely noise sources as possible. The guard rings are also placed around the critical transistors to minimize stray electrons and stray holes from affecting the critical transistors.

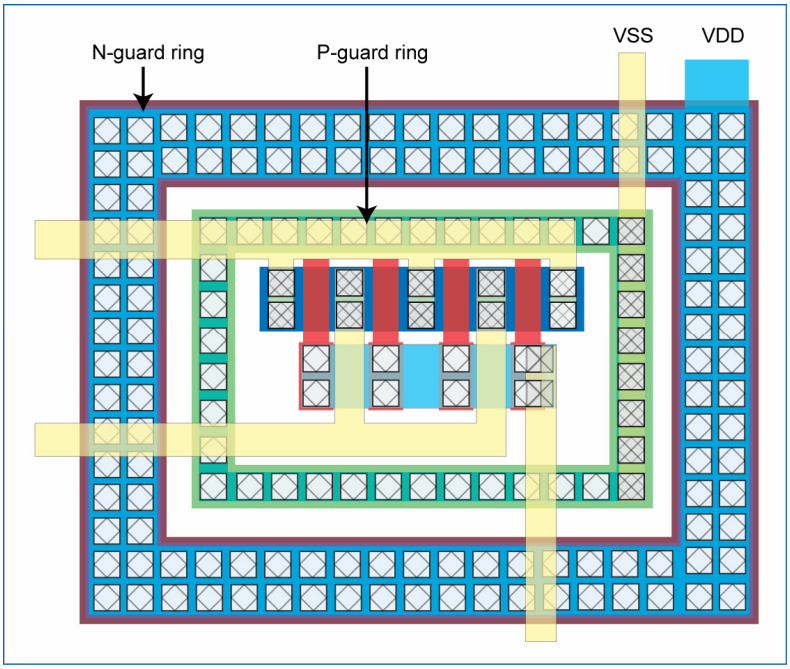
To reduce substrate coupling noise, you may use guard ring in the following configuration around critical transistors.

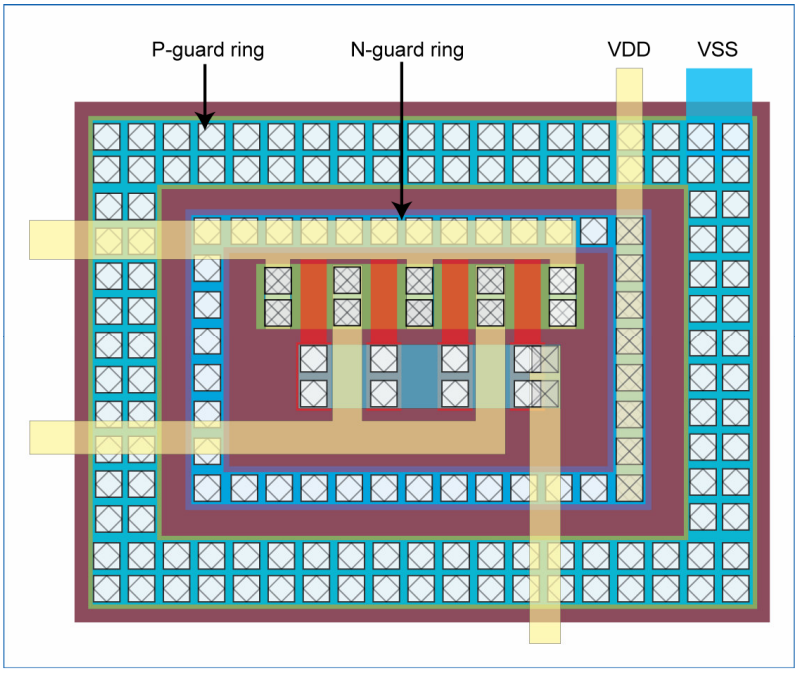
- Surround NMOS in the p-substrate with p-tap guard ring that is connected to ground.

- Surround PMOS in the N-well with n-tap guard ring that is connected to VDD.

The following layouts show both PMOS and NMOS surrounded with double guard rings.

SK hynix Ex. 1016, Page 28 of 39





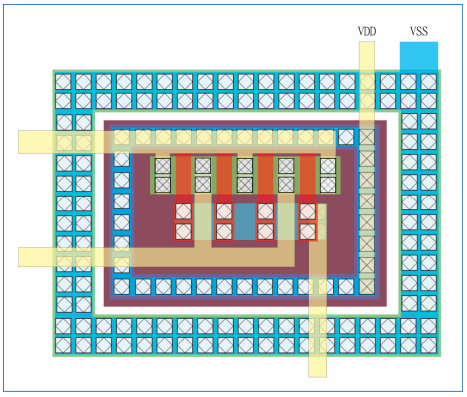
It is generally believed that N-well guard ring in P-substrate and P-diffusion guard ring in N-well are not of much use. The stray electrons and holes travel deep into the substrate and are not collected by the

SK hynix Ex. 1016, Page 29 of 39

guard rings. However, putting a guard ring is better than leaving the space empty since we still have to keep the noisy transistors at a distance away from the other transistors.

Extra Punch

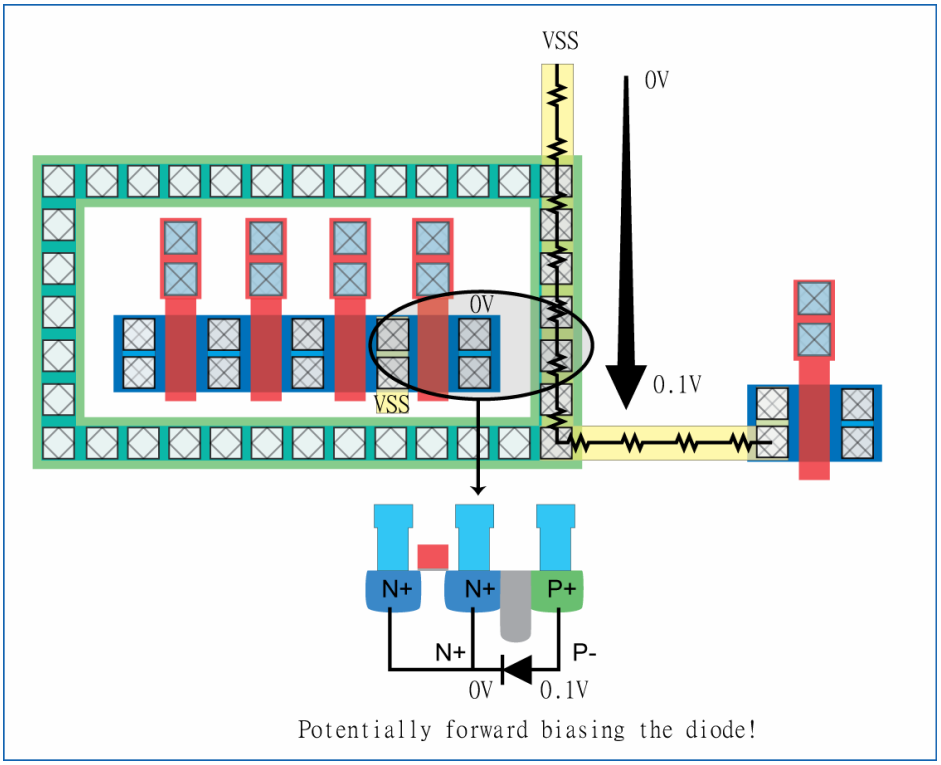
Double Guard Ring for PMOS



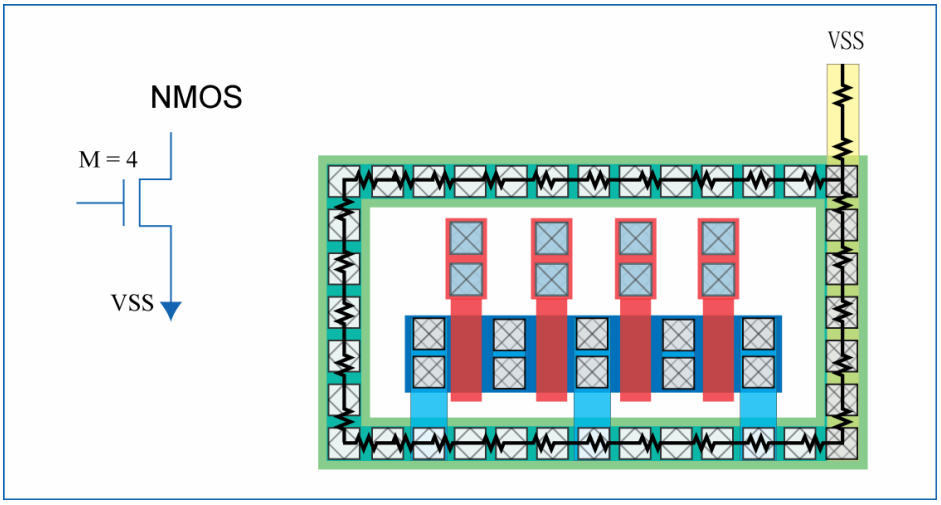
The most common way to add a double guard ring to PMOS is shown here. The outer P-guard ring acts as both a guard ring for PMOS, and as p-tap for the substrate. However, the P-guard ring might be too far from the PMOS to be effective in collecting stray hole from the PMOS.

An important layout practice is to ensure that there is no (or very little) current flowing through any part of the guard ring. Consider the layout in the diagram below. A current flowing in the p-type guard ring raises its potential above VSS. If the potential of the n-diffusion next to the p-type guard ring (shaded in the diagram) is at VSS, the PN junction potentially becomes forward biased, and results in holes injected from the guard ring into the substrate!

SK hynix Ex. 1016, Page 30 of 39



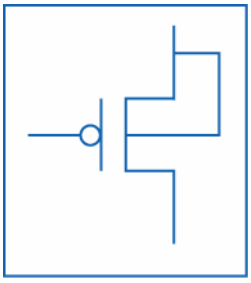
A more subtle case is shown below. In the circuit design, the source of the transistor is connected to VSS. It is very convenient to tap the VSS from the guard ring as shown in the layout below. If the peak current going through the transistor is very small, the layout may be acceptable. Otherwise, tapping power from the guard ring is not allowed.



SK hynix Ex. 1016, Page 31 of 39

Extra Punch

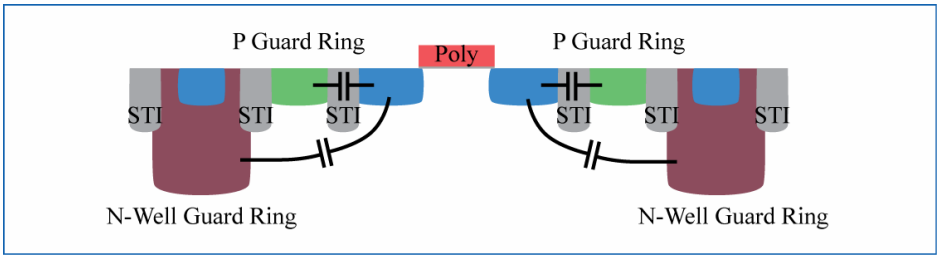
Floating Well



Wells that are tied to and varied with the source terminal of the transistor are called floating well, as illustrated on the right. Floating well is quite commonly used in analog design to reduce the "body effect" of the transistor. The PMOS with this type of bulk connection needs an N-well on it own, and hence the layout has a huge area overhead. It is important to identify all the PMOS that sit on a floating well before floor plan the layout.

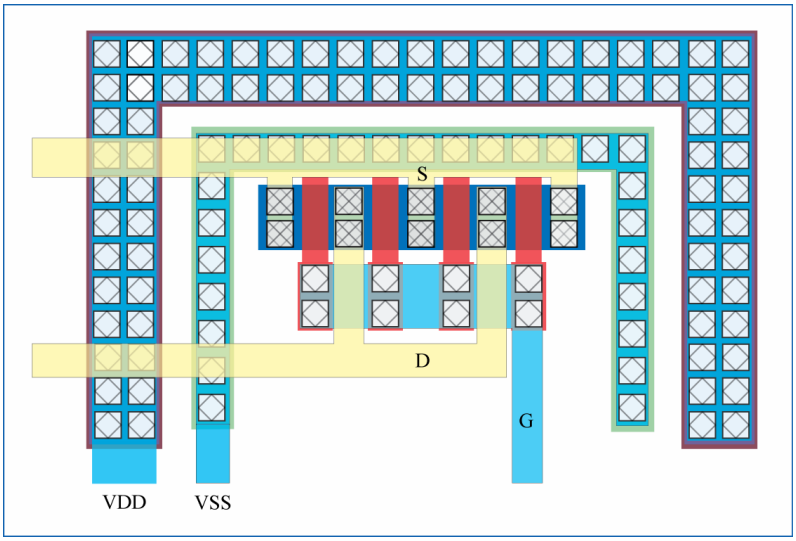
Fifth Stroke. Balancing Area, Speed and Noise

The guard rings in the fourth stroke take up a lot of area. The guard rings also add capacitive load to the transistor as illustrated in the diagram below.



Instead of using a full guard ring, you may consider using a U-shape guard ring. Some circuit designers do not favor the use of U-shape guard ring while some circuit designers use U-shape guard ring only for the p-tap and the n-tap. The following layout shows a NMOS with U-shape guard rings.

SK hynix Ex. 1016, Page 32 of 39

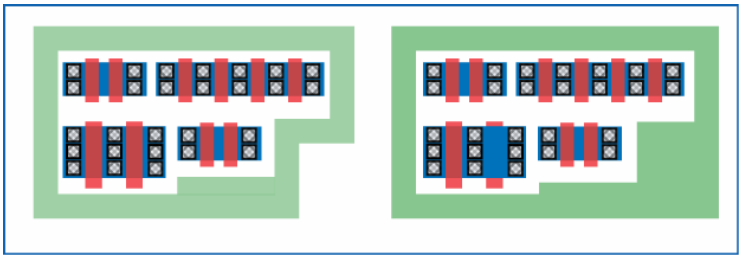


If reducing the area of the layout is of utmost priority and you do not expect much minority carrier noise in the silicon, then you may consider reducing the numbers of local N-well guard rings and P-diffusion guard rings.

Extra Punch

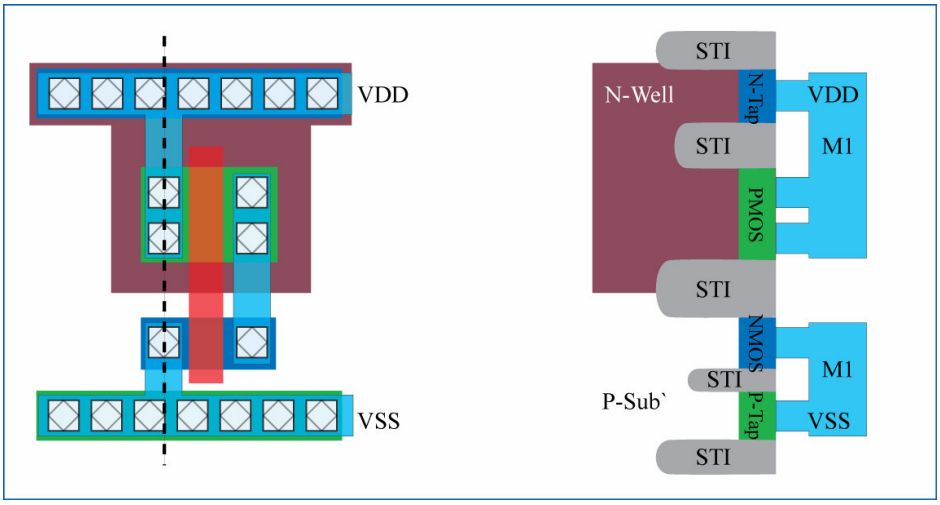
Guard Ring

The guard rings need not be rectangular. Two ways to add a guard ring around a group of transistors is shown here.

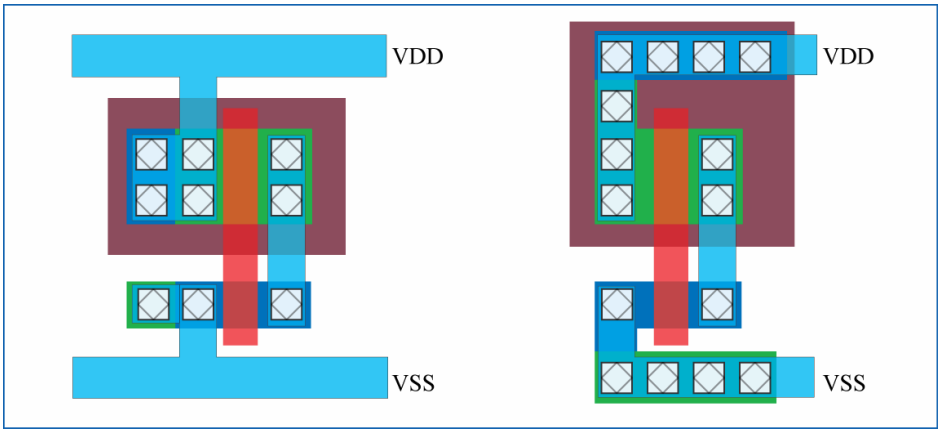


SK hynix Ex. 1016, Page 33 of 39

The inverter layout in page 12 does not optimize the placement of the p-tap and the n-tap. The layout is shown here again. Without looking at the cross-section on the right, study the layout for a moment. Can you spot the problem in the layout? Now, study the cross-section of the layout along the dotted line. Can you see that the n-tap and the p-tap are isolated on four sides by the isolation oxide and hence are electrically far away from the transistors?



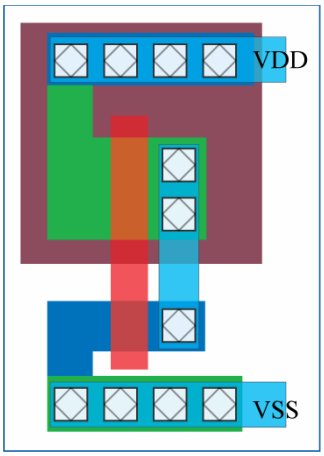
Two alternative layouts are shown below. The p-tap is butted with the n-diffusion, and the n-tap is butted with the p-diffusion. The trade-off in the new layouts is a larger area.



SK hynix Ex. 1016, Page 34 of 39

Extra Punch

Connection via the silicide

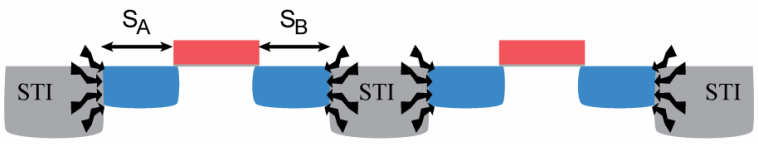


Are the connections to the sources of the transistors in the inverter layout on the right missing? There is nothing wrong with the layout. The Fab that this layout is to be fabricated supports connection between butting diffusions by the silicide.

Silicide is added to the gate, drain and source to reduce its resistivity. In the silicide process, a deposited metal is reacted thermally with the poly and the diffusion to form silicide. When the p-diffusion and the n-diffusion are butted, they are connected by the layout of silicide above them!

Sixth Stroke. Relief the Stress

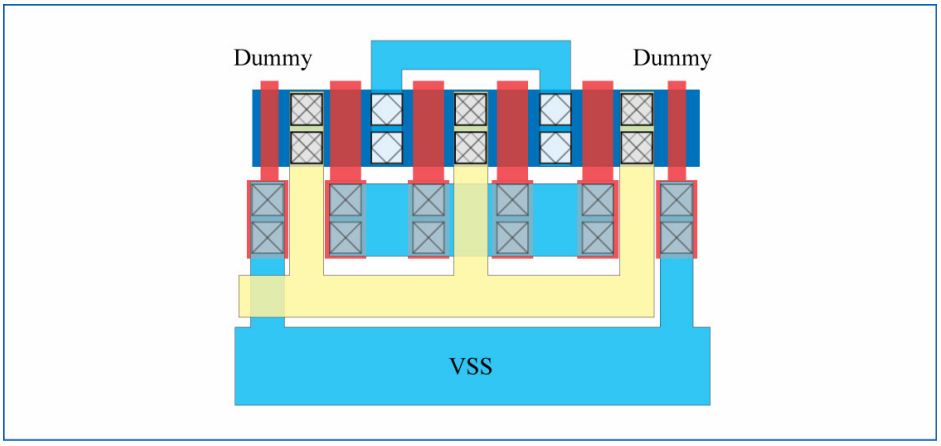
The stress from the STI onto the drain and source has an effect on the performance of NMOS and PMOS. The impact of the STI stress depends on the source and drain overhang (which are indicated as S_A and S_B in the following diagram) of the transistor active island.



Mechnical Stress from STI on Diffusion

SK hynix Ex. 1016, Page 35 of 39

To reduce the effect of STI stress, the source and drain diffusions require to be extended when they are next to the STI. However, a large diffusion also increases parasitic capacitance and layout area. A better approach is to insert one or more dummy transistors at each end of the transistor as illustrated in the diagram below. Note that the dummy transistor must share the diffusion with the non-dummy transistor.



Seventh Stroke. Protect the Gate

The gate oxide underneath the poly is incredibly thin. If the charges accumulated on the poly is sufficiently large, the charges accumulated can damage the gate oxide. This is known as process antenna effect.

The maximum amount of charges that can be accumulated on the poly is proportional to the area of the poly[1]. Thus, an effective layout practice to prevent process antenna violation is to stay within the antenna ratio design rule of the respective technology. Some general guidelines are

- Minimize the use of poly for routing
- Minimize the use of poly to connect the gates together

[1] It is more correct to say that charges are accumulated on the perimeter side-wall area of the poly, which can be calculated as the perimeter of the poly multiple by the thickness of the poly.

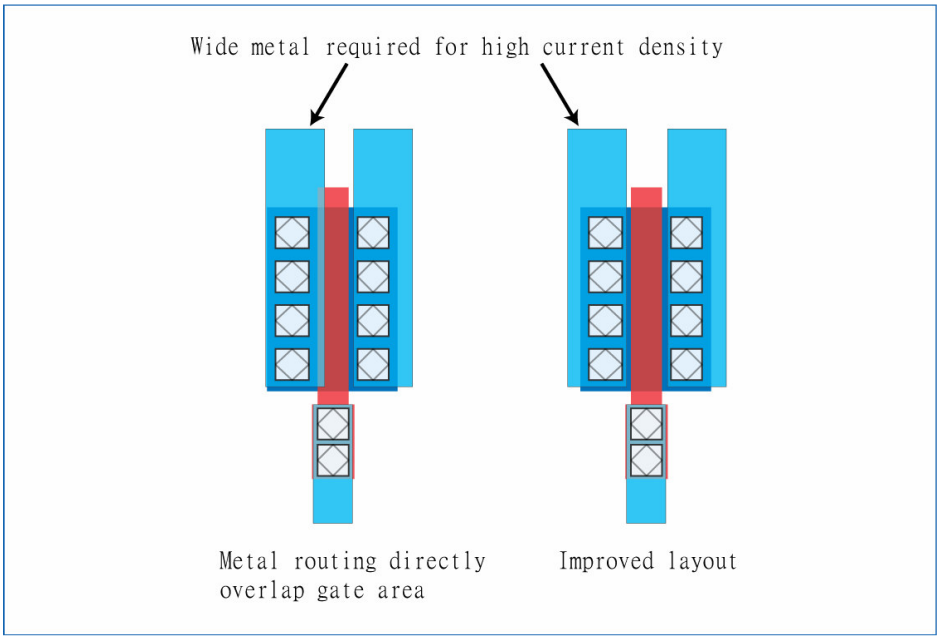
SK hynix Ex. 1016, Page 36 of 39

Some knowledgeable readers may suggest using diodes to protect the poly from antenna ratio violation. Antenna diode is only effective in preventing antenna violation from metal routing, and does not help in antenna violation due to poly. The reason is simple. The diodes are made from diffusions, but the poly is deposited onto the wafer before the diffusions are implanted into the wafer. Hence, the diode does not exist at the time the poly is fabricated on the wafer!

Besides protecting the gates from process antenna effect, other measures to protect the gate are

- Do not place contact and via directly on top of the transistor's gate.
- Avoid routing over the gates of critical transistors. Refer to the following diagram for an example.
- Avoid routing over active areas of critical transistors.

An exception to the guidelines for an analog design is to allow routing over decoupling MOS capacitors. This is done to compromise for a shorter routing. It is rather common to route over the gates in the layout of a digital design.



SK hynix Ex. 1016, Page 37 of 39

Eighth Stroke. Improve Yield

The most compact layout does not give the best manufacturing yield. Use the layout practices as discussed here to enhance yield. The practices are also illustrated in the following layouts.

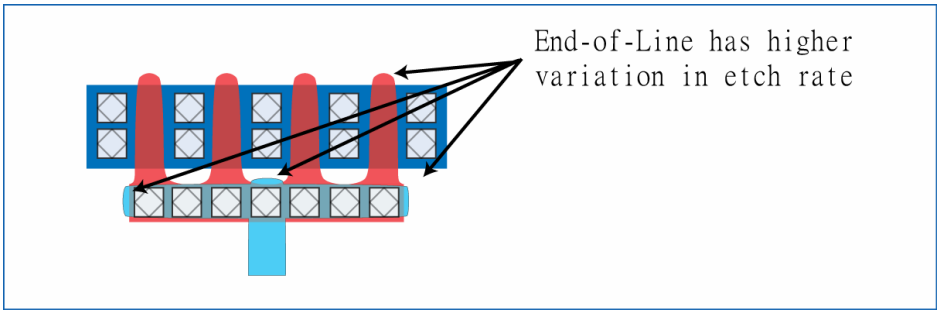
Avoid using of single contact or via

Avoid using single contact or via. A high percentage of IC manufacturing defects is related to faulty contact and via.

Use at least double contact or double via whenever possible.

Metal coverage of contacts and vias

Always give additional metal coverage on the contacts and the vias especially if they are located at the end of the metal. A larger metal coverage reduce contact or via resistance variation and also reduce the chance of an open contact or open via.



Poly extension from diffusion

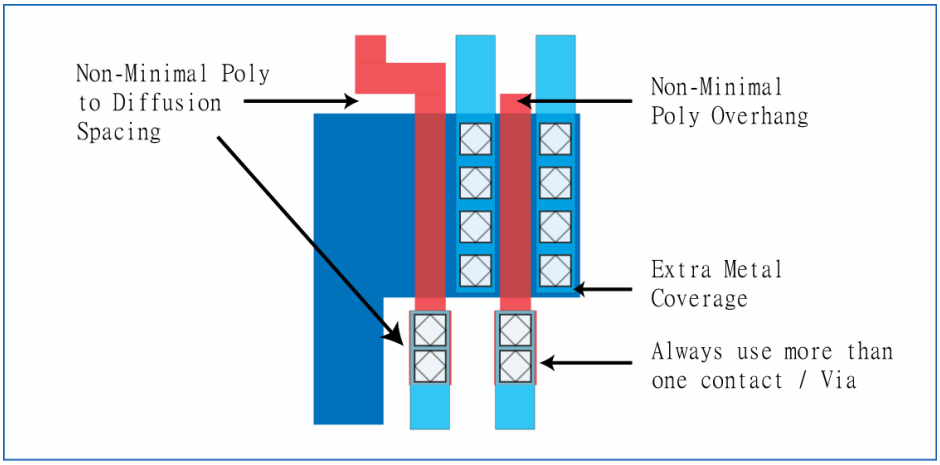
Exceed the minimum design rule requirement for poly overhang. The poly extension provides a margin for any misalignment in the diffusion mask or the poly mask. If the poly extension is less than the sum of the misalignment in the diffusion mask and the poly mask, a short between the drain and the source may happen and result in lower yield.

SK hynix Ex. 1016, Page 38 of 39

Most of the design rules provide both minimum and recommended poly extension values. It is more common in digital cell layouts that are used by Place and Route software to use the minimum value so as to optimize the timing performance and cell area. For custom layout, it is a balance between the priorities.

Spacing from poly to diffusion

Similar to the poly extension, most of the design rules also provide both minimum and recommended spacing from the poly to the diffusion. The minimum spacing from the poly to the diffusion may be less aggressive than the poly extension as any overlapping between the poly and the diffusion due to misalignment in the masks does not create functional error. However, the performance of the circuit can be degraded due to an increase in the capacitive coupling between the poly and the diffusion, and a change in the length and width of the transistor.



Note

Chapter 5 and chapter 6 will not be included
in the free PDF release.

SK hynix Ex. 1016, Page 39 of 39