John Y. Chen

CMOS DEVICES AND TECHNOLOGY FOR VLSI

PRENTICE HALL, *Englewood Cliffs, New Jersey 07632*

Greenthread Ex. 2059, p. 1 of 47
Semiconductor v. Greenthread
IPR2023-01242

Library of Congress Cataloging-in-Publication Data

CHEN, JOHN Y.
CMOS devices and technology for VLSI / John Y. Chen
p. cm.
Bibliography: p.
Includes index.
ISBN 0-13-138082-6
1. Integrated circuits—Very large scale integration 2. Metal oxide semiconductors, Complimentary I. Title
TK7874.C523 1990 88-38557
621.381'73—dc19 CIP

Editorial/production supervision
and interior design: BARBARA MARTINI
Cover design: DIANE SAXE
Manufacturing buyer: MARY ANN GLORIANDE

© 1990 by Prentice-Hall, Inc.
A Division of Simon & Schuster
Englewood Cliffs, New Jersey 07632

The publisher offers discounts on this book when ordered in bulk quantities. For more information, write:

Special Sales/College Marketing
College Technical and Reference Division
Prentice Hall
Englewood Cliffs, New Jersey 07632

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

ISBN 0-13-138082-6

PRENTICE-HALL INTERNATIONAL (UK) LIMITED, *London*
PRENTICE-HALL OF AUSTRALIA PTY. LIMITED, *Sydney*
PRENTICE-HALL CANADA INC., *Toronto*
PRENTICE-HALL HISPANOAMERICANA, S.A., *Mexico*
PRENTICE-HALL OF INDIA PRIVATE LIMITED, *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*
SIMON & SCHUSTER ASIA PTE. LTD., *Singapore*
EDITORA PRENTICE-HALL OF BRASIL, LTDA., *Rio de Janeiro*

TO MY PARENTS

CONTENTS

PREFACE xi

BIOGRAPHY xv

1 INTRODUCTION TO CMOS, THE VLSI TECHNOLOGY 1

2 CMOS DEVICE PHYSICS 5

2.1 ELECTRONS AND HOLES 6
2.2 MOS CAPACITORS 10
2.3 MOS TRANSISTORS 16
2.4 BURIED CHANNEL DEVICES 27
2.5 SHORT- AND NARROW-CHANNEL EFFECTS 33

3 MOS MODELLING 38

3.1 MODELLING FOR CMOS TECHNOLOGY DEVELOPMENT 38
3.2 MODELS FOR INDIVIDUAL PROCESS MODULES 41
3.3 MODELS FOR PROCESS INTEGRATION 56
3.4 TWO-DIMENSIONAL DEVICE MODELS 61
3.5 DEVICE PARAMETER EXTRACTION FOR CIRCUIT STIMULATION 70

Greenthread Ex. 2059, p. 3 of 47
Semiconductor v. Greenthread
IPR2023-01242

4 CMOS OPERATION 92

4.1 INVERTERS 92
4.2 A CMOS OUTPUT BUFFER 101
4.3 TRANSMISSION GATE AS A SWITCH 103
4.4 CMOS LOGIC GATES 105
4.5 DYNAMIC CMOS LOGIC 106
4.6 CMOS RAMS 110

5 CMOS PROCESS TECHNOLOGY 119

5.1 PROCESS ARCHITECTURE 119
5.2 *P*-WELL PROCESS 122
5.3 *N*-WELL PROCESS 125
5.4 *P*-WELL VERSUS *N*-WELL 127
5.5 TWIN-TUB PROCESS 128
5.6 RETROGRADE-WELL PROCESS 130
5.7 CHOICE OF PROCESS ARCHITECTURES 136
5.8 SOS TECHNOLOGY 136
5.9 SOI TECHNOLOGY 142
5.10 BIPOLAR/CMOS INTEGRATION—BiCMOS 149

6 CMOS TRANSISTOR DESIGN 174

6.1 MOSFET SCALING 175
6.2 NON-SCALABLE DEVICE PARAMETERS 176
6.3 TRANSISTOR DESIGN FOR SHORT-CIRCUIT MOSFETS 182
6.4 HOT CARRIER EFFECTS AND *n*MOS DESIGN FOR RELIABILITY 183
6.5 BURIED-CHANNEL EFFECTS AND *p*MOS DESIGN 211

7 CMOS ISOLATION 233

7.1 BACKGROUND 233
7.2 MOS ISOLATION TECHNIQUES 238
7.3 ISOLATION IN CMOS 249
7.4 NEW ISOLATION TECHNIQUES FOR CMOS 272
7.5 ISOLATION DESIGN RULES IN CMOS 277

8 LATCHUP IN CMOS 285

8.1 INTRODUCTION 285
8.2 PHYSICS AND LUMPED CIRCUIT MODEL 286
8.3 PARASITIC TRANSISTORS AND RESISTORS 288
8.4 LATCHUP CHARACTERIZATION 291

8.5 AVOIDING LATCHUP 301
8.6 LAYOUT CONSIDERATIONS 309
8.7 SUMMARY 317

9 CMOS DESIGN RULES 323

9.1 DESIGN RULES DERIVATION 323
9.2 LAMBDA-BASED DESIGN RULES 328
9.3 LIMITATION OF LAMBDA RULES AND THEIR MODIFICATIONS 331
9.4 SUBMICRON DESIGN RULES AND THEIR IMPACTS 332
9.5 SUMMARY AND FUTURE PERSPECTIVES 334

LIST OF SYMBOLS 338

INDEX 343

Greenthread Ex. 2059, p. 4 of 47
Semiconductor v. Greenthread
IPR2023-01242

2.3.2 MOSFET Operation

When a voltage greater than V_t is applied to a gate, the semiconductor surface is inverted to n-type and a MOSFET channel is formed, as shown in Fig. 2.11. For small drain voltages, the entire channel is inverted and channel conductance is proportional to the inversion charges. This area is the linear region in which the drain current I_d is proportional to V_d and the FET acts as a resistor. As V_d increases and reaches a point such that $(V_d + 2\Phi_B)$ is just larger than the band bending produced by the gate voltage, the inversion disappears at point X near the drain (Fig. 2.11b). The channel is then pinched off at this point and the corresponding drain voltage is defined as V_{dsat} because for $V_d > V_{dsat}$, I_d remains essentially constant. As shown in Fig. 2.11(c), the additional voltage above V_{dsat} is consumed for widening the depletion region and the potential at Point X remains constant. Point X moves toward the source as V_d exceeds V_{dsat}, but the movement is very slight; hence, I_d increases very little. The drain current basically remains at a constant level, I_{dsat}, for $V_d > V_{dsat}$. However, for a device with very short channel length, a slight movement of the pinchoff point can be a significant portion of the entire channel length, thereby causing significant increase of I_{dsat}. Other short channel effects will be discussed in Sec. 2.5.

Linear and saturation regions. This section derives the basic MOSFET characteristics. If a MOSFET is biased with sufficient gate voltage to cause surface inversion, the free charge in the inversion layer at a distance y from the source is

$$Q_n(y) = Q_s(y) - Q_B(y) \tag{2.14}$$

where $Q_s(y)$ is the charge induced in the semiconductor per unit area at Point y. Because

$$Q_s(y) = -C_{ox}[V_g' - \Phi_s(y)] = -C_{ox}\{V_g' - [2\Phi_B + V(y)]\} \tag{2.15}$$

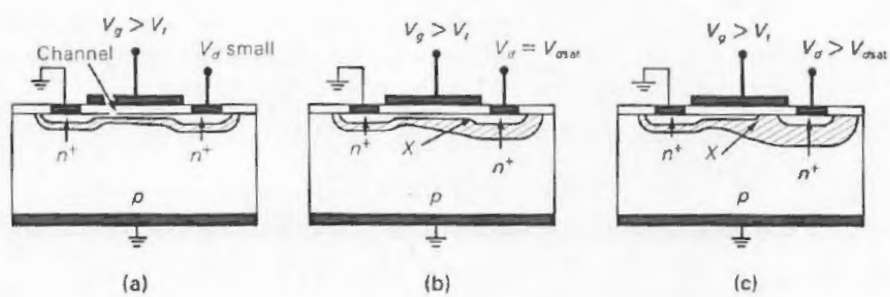


Figure 2.11 Illustration of the operation of a MOSFET for $V_g > V_t$: (a) $V_d \ll V_{dsat}$; (b) $V_d = V_{dsat}$; and (c) $V_d > V_{dsat}$ (Ref. 1).

Greenthread Ex. 2059, p. 5 of 47
Semiconductor v. Greenthread
IPR2023-01242

The discussion of an inverter circuit in Chapter 4 will illustrate the significance of the body effect on circuit performance.

MOSFET types. All derivations previously mentioned use n-channel transistors as an example. Similar results can be obtained for p-channel devices with reverse polarity for all terminal voltages and threshold voltage. Table 2.1 summarizes important equations for n- and p-channel MOSFETs. In both cases, the devices are off unless a sufficient gate voltage is applied. This type of device (normally-off) is called an enhancement mode device. Another type of transistor is normally on unless reverse-polarity V_g (e.g., negative V_g for an n-channel) is applied to turn it off. These devices (normally-on) are called depletion mode devices.

For an n-channel depletion device, the threshold voltage is made negative by adding an arsenic ion implantation during device fabrication. A total of four different types of MOSFETs result, as shown in Fig. 2.17. The corresponding I-V characteristics are also shown. In n-channel cases, a positive V_g must be applied to turn on an enhancement-mode device. But, a negative V_g must be applied to turn a depletion-mode device off. A depletion-mode device delivers current at $V_g = 0$ and is often used as the load device in NMOS logic circuits. In p-channel cases, one applies a negative V_g to turn on an enhancement-mode device, but a positive V_g to turn off a depletion-mode device. However, depletion-mode p-channel devices are normally not used in present MOS technologies. In a CMOS circuit, enhancement mode p-channel MOSFETs are often used as load devices. A depletion-mode device is commonly made by forming a thin n-type surface layer in the p-substrate for n-channel cases or a thin p-type surface layer in the n-substrate for p-channel cases. Thus, source and drain are connected with the same type of semiconductor layer unless this layer is fully depleted by applying the appropriate gate voltage. The presence of this surface layer however changes normal MOSFET operation from surface-channel conduction to buried-channel conduction.

2.4 BURIED CHANNEL DEVICES

Discussions so far have covered device operations in surface-channel MOSFETs in which carriers propagate at the semiconductor surface. However, in some devices such as depletion-mode devices, carriers propagate slightly under the semiconductor surface. This type of device is called a buried-channel device. Fig. 2.18 shows the cross-section of a buried-channel nMOS device. Notice the n-type layer with the doping concentration N_D and junction depth x_j. The device has two depletion regions. The bottom region originated from the p-n junction, and the top region is associated with the MOS capacitor. The depletion width is not uniform because, under the bias

Greenthread Ex. 2059, p. 6 of 47
Semiconductor v. Greenthread
IPR2023-01242

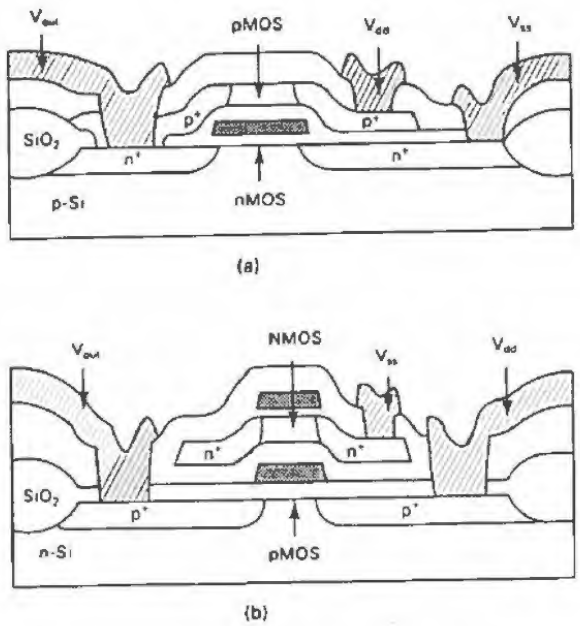


Figure 5.21 Schematic cross-sections of a 3-D CMOS/SOI inverter: (a) Joint-gate (JMOS) type; and (b) stacked type (Ref. 33, © 1983, IEEE).

To date, a 1.1K gate array[36] and a 256 bit SRAM[37] have been demonstrated for 3-D CMOS/SOI. The 8K bit parallel array multiplier in the gate array exhibits complete operation. The 256-bit SRAM is configurated with the NMOS memory cells in the bottom layer and the CMOS peripheral circuits in the top layer as shown in Fig. 5.22. Complete memory operation including the intralayer and interlayer data transfer has been demonstrated. Both these circuits were fabricated with a laser recrystallized SOI layer. An even larger circuit, a 64K SRAM, was demonstrated[38] using *p*-channel loads made in a non-crystallized polysilicon SOI layer. The as-deposited polysilicon was hydrogen passivated rather than beam recrystallized for manufacturability. Although the *p*-channel polysilicon FETs had low mobility, the resultant CMOS configuration provided lower static power than that of a conventional NMOS RAM with poly load resistors. On the other hand, SOI advantages such as latchup free, high density, high alpha-particle immunity also existed in this structure.

Despite the fact that sizable CMOS circuits have been demonstrated in two-layer SOI, 3-D VLSI using CMOS/SOI technology has a long way to go

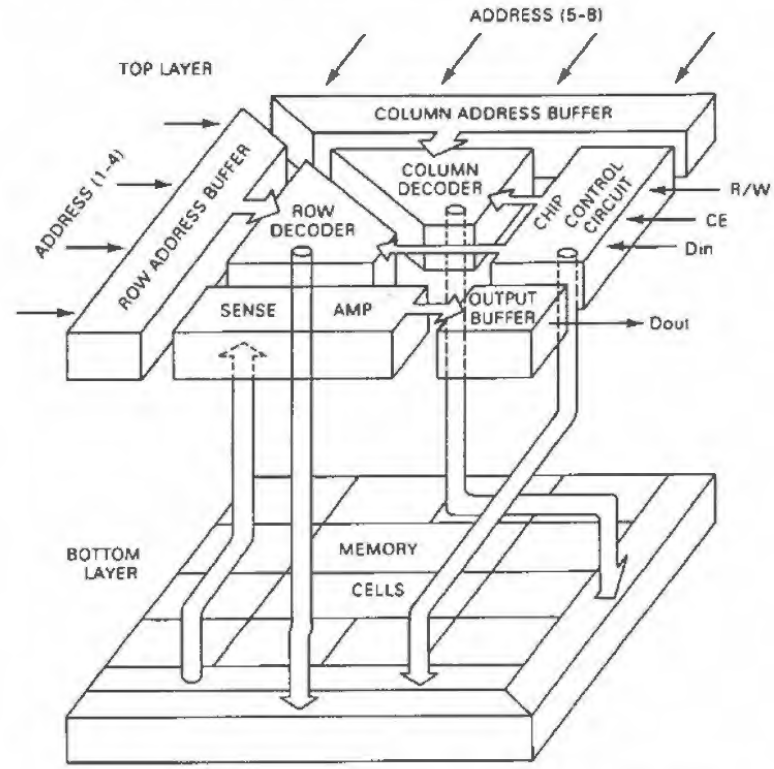


Figure 5.22 Circuit diagram of a 3-D static RAM (Ref. 37, © 1986, IEEE).

for production. A reliable and high-throughput recrystallization process for SOI layers is needed. Intra- and inter-layer interconnection with planarization must be developed. High yield, fast turn-around and redundance circuits are other concerns for VLSI fabrication.

5.10 BIPOLAR/CMOS INTEGRATION—BiCMOS

As described in Chapter 1, CMOS offers low power which is attractive for VLSI, especially in digital applications. However, the speed of CMOS, although comparable with NMOS, is slower than what bipolar can provide. It is particularly true when heavy-loading or long interconnects need to be driven. A bipolar transistor not only can deliver a large current, it also has well-controlled turn-on voltage, good noise margin and small logic swing for

ECL (Emitter-Coupled Logic). ECL also provides a high speed sense amplifier for static RAMs. All these advantages make it suitable for high speed circuits and analog applications. Its major disadvantages are: high power, poor density and limited circuit options.

BiCMOS technology is the integration of bipolar and CMOS devices on a single chip. The intent is to offer high density, low power CMOS arrays with high speed bipolar drivers. It can also provide analog and digital system integration on the same chip. With BiCMOS, the advantage of high density and low power in CMOS can be combined with the speed advantage offered by bipolar. Other improvements include ECL and TTL (Transistor Transistor Logic) interface, high speed I/Os, less sensitivity to fan-out and output load, reduced clock skew and improved internal gate delay. Because bipolar devices offer more current driving ability, for high speed applications, BiCMOS does not have to be down-scaled as much as CMOS. One-to-two micron BiCMOS can offer circuit speed as high as submicron CMOS. Consequently, the 5 V power supply can be maintained and submicron process needs are not as high. The major drawbacks to BiCMOS are higher cost and longer fabrication time.

5.10.1 BiCMOS Operation

Figure 5.23 shows a BiCMOS inverter. When the input (V_{in}) is low, both N0 and N1 are off but the *p*-channel device is on and the base of the upper *npn* bipolar rises to V_{BE} and above, turning the bipolar on and pulling the V_{out} up to $V_{DD} - V_{BE}$. When the input (V_{in}) is switched to high, $Q1$ is cut off, and so is N2; but N1 is turned on and the base of the lower *npn* bipolar is charged (V_{out}) to V_{BE}, turning the bipolar on and pulling the output voltage (V_o) down to V_{BE}.

Because the BiCMOS output does not swing the full rail (V_{DD} to ground), the superior noise margin and zero static power dissipation inherent in CMOS are not fully maintained. As the output is pulled down, V_{out} goes to V_{BE}; although V_{out} eventually can be discharged below V_{BE}, it would take an unreasonably long time. For practical cases, the BiCMOS output does not go to ground, noise margin is therefore degraded and when it drives to another CMOS input gate, power dissipation for the CMOS gate is slightly increased. The amount of power dissipation depends on the threshold of the *n*-channel FET in the CMOS input gate. If 1 μA is used to define the threshold voltage (V_{Tn}), the static current dissipated at the *n*-channel FET of the CMOS gate can be approximated as

$$I_D = 10^{-6} - (V_{Tn} - V_{out})/S_t \text{ in Amp} \tag{5.1}$$

where V_{Tn} is the *n*-channel FET threshold voltage and S_t is the FET subthreshold factor in mV per decade as described in Chapter 2 (Eq. 2.36). The static power dissipation of the CMOS gate, which is proportional to I_D, can

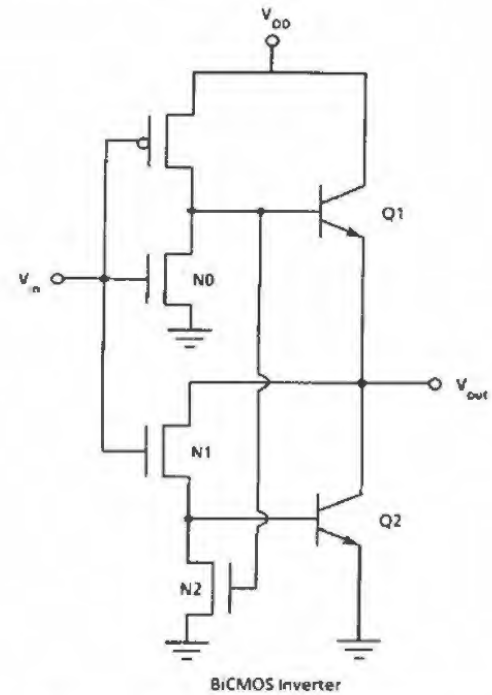


Figure 5.23 Circuit schematic of a BiCMOS inverter.

be reduced by increasing the threshold voltage and decreasing subthreshold swing. However, too high a threshold voltage results in lower current driving ability, hence poorer circuit speed. As an example, if $V_{Tn} = 0.7$ V, $S_t = 100$ mV/decade, then $I_D = 1$ μA because $V_{BE} = 0.7$ V. If V_{Tn} is raised to 1 V, then $I_D = 1$ nA, an improvement of three orders of magnitude in power dissipation.

Figure 5.24 shows delay time for the BiCMOS inverter plotted versus load capacitances (C_L).[39] Notice that the speed improvement is greater as the load capacitance becomes larger. This improvement is due to the fact that a bipolar device normally delivers much more current than a MOS device. Therefore, compared to CMOS, bipolar or BiCMOS is less sensitive to the output load capacitance. In first order approximation, the slope of the line is roughly

$$\begin{aligned} S_L &= t_{pd}/C_L = Q/(I_L C_L) = V_o/I_L \\ &= V_o/I_D \text{ for the CMOS case} \\ &= V_o/I_C = V_o/(\beta I_B) \text{ for the BiCMOS case} \end{aligned} \tag{5.2}$$

Greenthread Ex. 2059, p. 8 of 47
Semiconductor v. Greenthread
IPR2023-01242

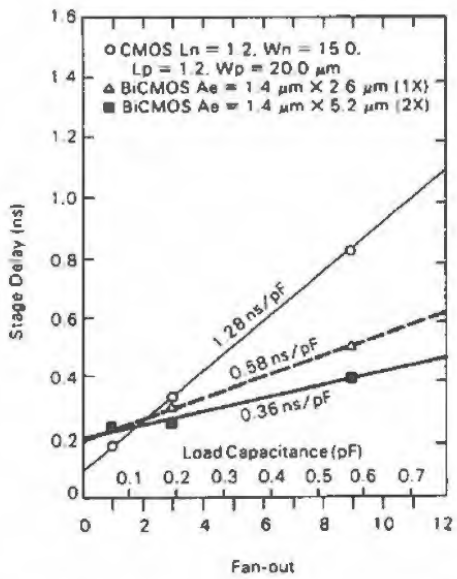


Figure 5.24 Propagation delay versus load capacitance (Ref. 39).

where Q is the charge stored at the load capacitor, V_o is the output voltage, and I_L is the output current, which is the drain current (I_D) for the CMOS case, and is the collector current (I_C) in the BiCMOS case. Although I_D and I_C are geometry-dependent, i.e., $I_D \propto W/L$ and $I_C \propto I_k \propto A_E$, the bipolar current I_C in general is much larger than the drain current I_D of a FET of similar size. So, for gates that need to drive a heavy load ($\geq$0.5 pF), BiCMOS is faster for a factor of two or more. For a very small load ($<$0.1 pF in this example), the BiCMOS circuit is actually slower because it takes time to charge additional bipolar junction capacitances. In integrated circuits, a large fan-out normally corresponds to a high load capacitance. Compared to bipolar for the same fan-out, CMOS has a larger load capacitance to drive due to the higher input capacitance. As shown in Fig. 5.24, delays for CMOS inverters increase as 1.28 ns/pF and this loading effect decreases to 0.58 and 0.36 ns/pF for BiCMOS inverters with two different emitter areas. Notice for fan-out greater than 2, the BiCMOS inverters run faster than the CMOS inverter.

Slightly more complex BiCMOS logic such as NAND and NOR gates are shown in Fig. 5.25. Again, only two bipolar devices are used in spite of the increase in FET counts. With respect to the total device count and extra

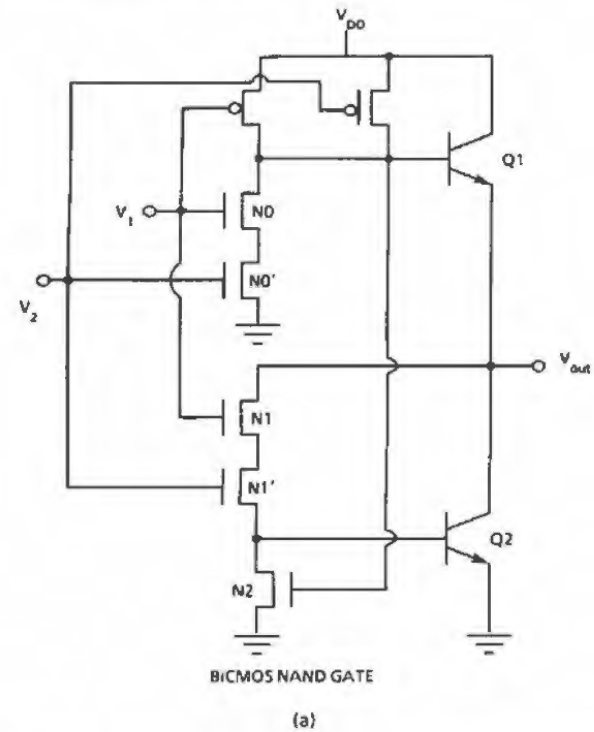


Figure 5.25 Schematics for (a) NAND, and (b) NOR BiCMOS circuits.

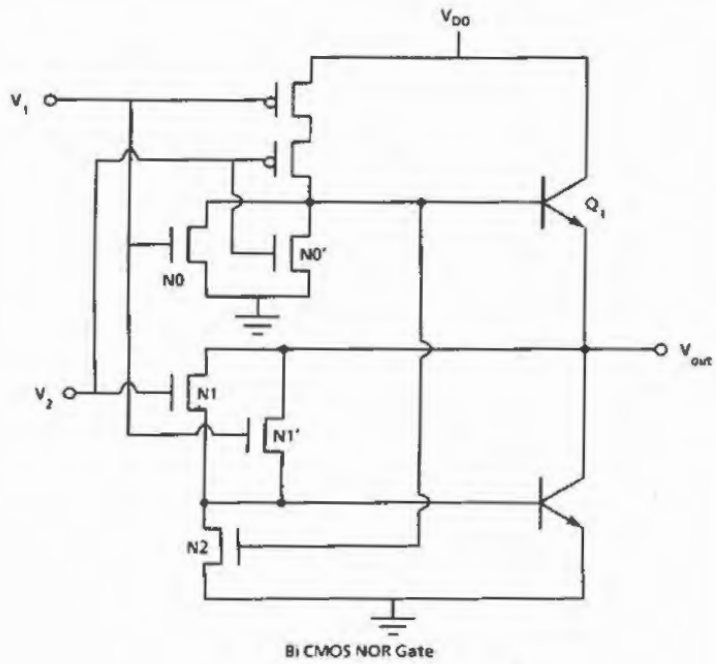
layout area, the penalty of adding bipolar drivers on a CMOS logic circuit is reduced as the CMOS circuit component becomes larger.

5.10.2 Bipolar Device Characteristics

In this section, one-dimensional bipolar device operation is discussed. Important device characteristics and relevant process parameters are also described to provide a background for the discussion of BiCMOS technology.

Shown in Fig. 5.26 are the cross-sectional view of an *npn* bipolar transistor, its corresponding one-dimensional representation and the impurity doping distribution. It is a three terminal device with a heavily doped *n*-type emitter that emits current, a lightly doped *n*-type collector that collects current and a *p*-type base region. As shown in Fig. 5.26(b), it is like two *p-n* junctions connected back-to-back, but the *p* region is very narrow. During normal operation, the emitter-base *pn* junction is forward biased and the

Greenthread Ex. 2059, p. 9 of 47
Semiconductor v. Greenthread
IPR2023-01242



(b)

Figure 5.25 *Continued.*

collector-base junction is reverse biased. A transistor biased under this condition operates in the so-called active region. Under forward bias, the emitter current (I_E) is caused by the diffusion of electrons from the n^+ emitter to the p-type base (I_{NE}) and the diffusion of holes from the base to the emitter (I_{PE}). The current components governed by diffusion are

$$I_{NE} = A_E q D_{nB}\, dn'_B/dx = A_E q D_{nB}\, n'_B(0)/W_B \tag{5.3a}$$

$$I_{PE} = A_E q D_{pE}\, dp'_E/dx = A_E q D_{pE}\, p'_E(0)/W_E \tag{5.3b}$$

$$I_E = I_{NE} + I_{PE} \tag{5.3c}$$

assuming W_B and W_E are very small so that linear approximation can be applied. A_E is the emitter area, D_n and D_P are the diffusivities for electrons and holes, n'_B and p'_E are the excess minority carrier concentrations in the base and emitter regions due to injection under forward bias at the emitter-

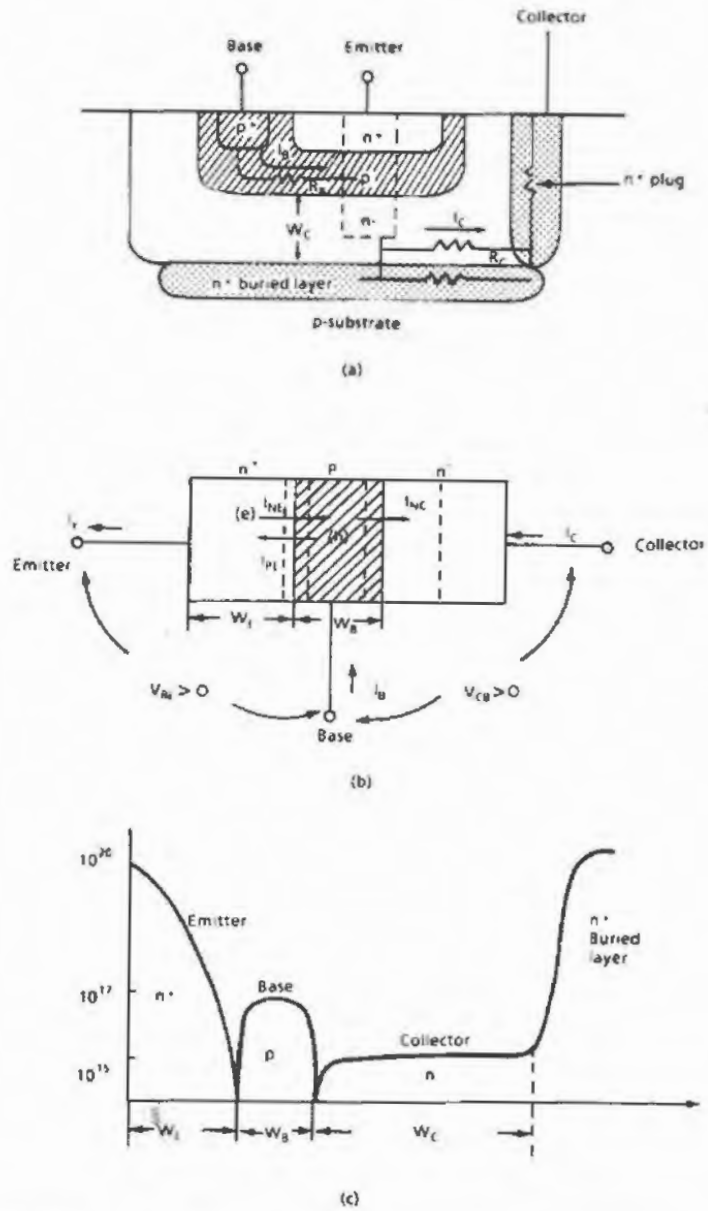


Figure 5.26 An *npn* bipolar transistor: (a) cross-section of an *npn* transistor; (b) One-dimensional representation; and (c) Impurity doping distribution in one dimension.

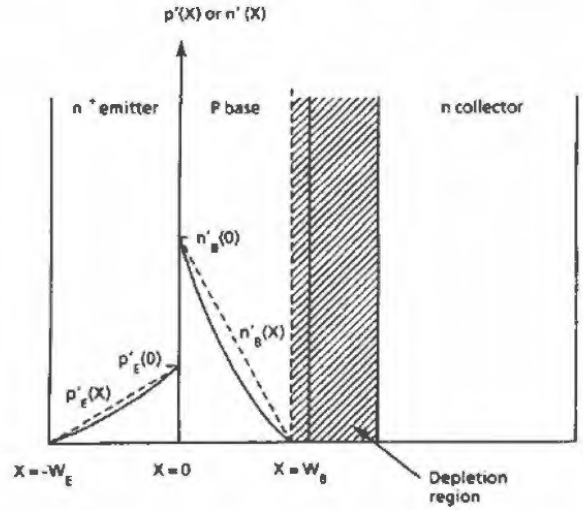


Figure 5.27 Excessive minority carrier distribution in a one-dimensional *npn* transistor.

base junction. They are calculated as

$$n'_B(x) = n_B(x) - n_{B0} = n_{B0}[\exp(qV_{BE}/kT) - 1] \quad \text{at } x = 0 \tag{5.4a}$$

$$p'_E(x) = p_E(x) - p_{E0} = p_{E0}[\exp(qV_{BE}/kT) - 1] \quad \text{at } x = 0 \tag{5.4b}$$

where n_B and p_E are the minority carrier concentrations, n_{B0} and p_{E0} are the minority concentrations at equilibrium, and V_{BE} is the forward bias emitter-base voltage. Figure 5.27 shows the distribution of minority concentrations in the base and emitter regions. The slopes of the curves correspond to current values. The excess electron concentration reaches zero at the base-collector junction because the junction is reverse biased and the electric field in the depletion region across the junction sweeps carriers away. Notice that $n'_B(0)$ is much larger than $p'_E(0)$ because $pn = n_i^2$ and the majority carrier concentration (i.e., the doping concentration) in the base is much less than that in the emitter. As a result, $I_{PE} << I_{NE}$. The ratio of the current injected into the base to the total emitter current is defined as emitter efficiency (γ_e) and is expressed as

$$\gamma_e = I_{NE}/(I_{NE} + I_{PE}) \approx 1 - I_{PE}/I_{NE} \tag{5.5}$$
$$= 1 - [D_p p'_E(0) W_B]/[W_E D_n n'_B(0)]$$

Using Einstein relation ($D = \mu kT/q$), the above equation can be written as

$$\gamma_e = 1 - (\mu_p p_B W_B)/(\mu_n n_E W_E) \approx 1 - R_{SE}/R_{SB} \tag{5.6}$$

where p_B and n_E are majority carrier densities in the base and the emitter. $p_B W_B$ is the total majority charge in the base and is called the Gummel number. The higher the Gummel number, the lower the emitter efficiency. R_{SE} and R_{SB} are the sheet resistances for both the emitter and base, respectively. Because typical values of R_{SE} and R_{SB} are tens and thousands respectively, γ_e is close to but less than unity.

Not all injected electrons are collected at the collector. Some of them are recombined in the base region and the portion which reaches the collector-base junction is defined as the base transport factor (T):

$$T = \frac{I_{NC}}{I_{NE}} = \frac{(dn_B/dx)_{x=W_B}}{(dn_B/dx)_{x=0}} \tag{5.7}$$

Solving the diffusion equation for n_B with boundary conditions, the above equation can also be shown as[40]

$$T = (\text{Cosh} W_B/L_{nB})^{-1} \tag{5.8}$$
$$= 1 - W_B^2/(2L_{nB}^2) \quad \text{since } W_B << L_{nB}$$

where L_{nB} is the diffusion length of minority carriers in the base. $L_{nB} = (D_n \tau_n)^{1/2}$ where τ_n is the minority lifetime.

Because the collector-base junction is reverse biased, the total collector current (I_C) is approximately equal to I_{NC} assuming the junction leakage current is negligible. The common base current gain (α) can then be written as

$$\alpha = \frac{I_C}{I_E} = \frac{I_{NC}}{I_{NE}} \frac{I_{NE}}{I_E} = T\gamma_e = [1 - W_B^2/(2L_{nB}^2)](1 - R_{SE}/R_{SB}) \tag{5.9}$$

Again, α is typically slightly less than one.

Based on Eqs. 5.3a and 5.4a, the current driving ability is

$$I_C = \alpha I_E = \alpha I_{NE}/\gamma_e = \frac{T A_E q D_n n_i^2 [\exp(qV_{BE}/kT) - 1]}{p_B W_B} \tag{5.10}$$

The denominator is the Gummel number. The numerator indicates that the current of a bipolar transistor is exponentially dependent on the voltage across the emitter-base junction. In reality, because of the IR drop from the base contact to the emitter-base junction (Figure 5.26a), the V_{BE} term in the above equation should be replaced by $V_{BE} - I_B R_B$.

The I_C versus V_{BE} curve is known as the Gummel plot. Figure 5.28 shows experimental data for I_C and I_B versus V_{BE} plotted linearly and semilogarithmically. Notice that the curve in Figure 5.28(b) is steep and a turn-on voltage at about 0.7 V can be defined. A larger emitter area, a narrow base and/or a lower base doping concentration can increase bipolar current driving ability. However, a narrow base or a low base doping concentration can cause significant base width modulation, an undesirable effect similar to

Greenthread Ex. 2059, p. 11 of 47
Semiconductor v. Greenthread
IPR2023-01242

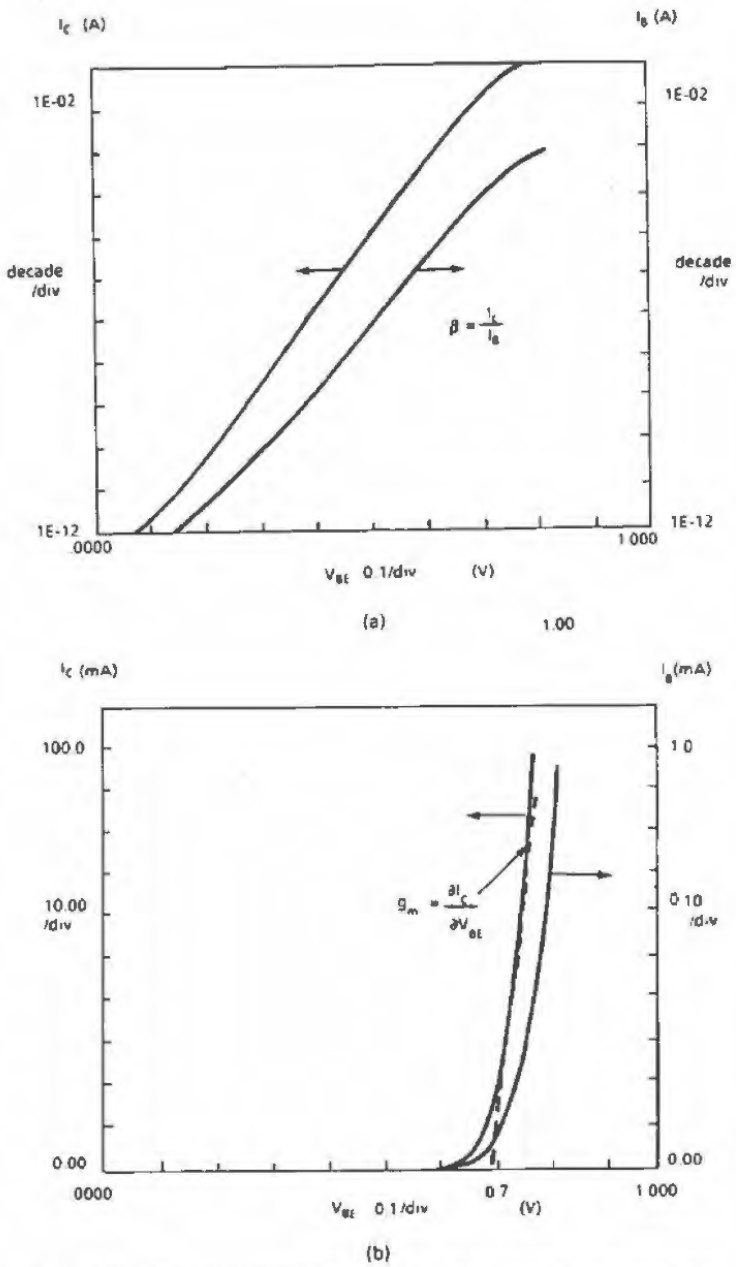


Figure 5.28 Collector and base current (I_c and I_b) versus base-emitter voltage (V_{BE}) for an *npn* transistor: (a) Gummel plot in semi-logarithmic; and (b) Gummel plot in linear.

the short-channel effect in MOSFETs. This base width modulation, known as Early effect, leads to increased output conductance for the bipolar transistor.

The base doping concentration should be much higher than the collector doping concentration so that most of the depletion width extends to the collector region to minimize Early effect and to maintain sufficient collector-base breakdown voltage (BV_{CBO}) and collector-emitter breakdown voltage (BV_{CEO}). However, a lightly doped collector gives a high collector resistance (R_C), resulting in a reduced current at a given bias condition. A common way to accommodate both high breakdown and low collector resistance is to use a heavily doped buried layer (Fig. 5.26a) under the lightly doped collector region so that R_C can be greatly reduced.

The common emitter current gain (β) is defined as I_C/I_B and equals $\alpha/(1-\alpha)$. Depending on α, β values range from tens to hundreds. The higher the α, the higher the β. From Eq. 5.9, it is clear that narrow base width, low emitter sheet resistance and low Gummel number in the base can all increase the common emitter current gain. Again, the Gummel number cannot be too low due to the limitation in output conductance and breakdown voltages.

While current gains are important DC parameters, transit time (τ_t) is important for high frequency operation. τ_t is the time needed for the carriers to travel across the base region. It can be approximated by

$$\tau_t = W_B/\langle v \rangle \tag{5.11}$$

where v is carrier velocity and $\langle\ \rangle$ means average value. Because

$$I_C = A_E q \langle n'_B \rangle \langle v \rangle = A_E q \langle v \rangle n'_B(0)/2 \tag{5.12}$$

and

$$I_C = I_{NF} = q D_n A_E n'_B(0)/W_B \tag{5.13}$$

equating Eq. (5.12) and (5.13),

$$\langle v \rangle = 2D_n/W_B \tag{5.14}$$

and

$$\tau_t = W_B^2/2D_n \tag{5.15}$$

can be obtained.

The above derivation, based on the concepts of average velocity and average minority carrier concentration, is at best an estimation. τ_t is also the time required to remove all minority carriers from the base when the emitter-base voltage (V_{BE}) switches to zero. It is interesting to note that $\tau_n/\tau_t \approx \beta$. The reciprocal of the transit time corresponds to the frequency limitation for the operation of a bipolar transistor. In reality, even before the operating frequency reaches $1/\tau_t$, current gain decreases at high frequen-

Greenthread Ex. 2059, p. 12 of 47
Semiconductor v. Greenthread
IPR2023-01242

cies due to additional RC time constants associated with parasitics. The cutoff frequency f_T, defined as the frequency at which the common emitter current gain is unity, is an important parameter for high frequency operations. f_T can be expressed as[41]

$$f_T = 1/(2\pi\tau_{total}) \tag{5.16}$$

and

$$\tau_{total} = \tau_t + C_{jE}/g_m + C_{jC}/g_m \tag{5.17}$$

where C_{jE} and C_{jC} are the junction capacitances at emitter and collector junctions and g_m is the transconductance defined as $\partial I_C/\partial V_{BE}$. To obtain good high frequency response, the parasitic capacitances and resistances must be reduced. In addition, the intrinsic base transit time must be decreased. Along with the use of the buried layer, a thinner collector layer (small W) and/or a heavily doped collector plug under the collector contact can further decrease R_C (Fig. 5.26a).

The previous discussion assumes that the carrier concentration injected to the base is much less than the majority carrier concentration (p_B) which equals the base doping concentration (N_{AB}). This assumption is often invalid at high current levels in which $n_B >> N_{AB}$. In such cases, the majority carrier concentration must increase to maintain charge neutrality in the base. This condition is called conductivity modulation. As a consequence, the effective base doping concentration becomes higher, resulting in a smaller emitter efficiency and a reduced current gain. High level injection also occurs at the collector and is known as Kirk effect.[42] For *npn* devices, it was assumed that minority carriers (electrons) that reached the depletion edge of the reversely biased collector-base junction were swept across so that electron density (n) in the depletion region was negligible. This assumption is not valid at a high current density J ($J = qnv$ where v is electron velocity). Because v is limited by the saturation velocity due to scattering, a finite amount of n exists when J is large. This electron density J/v adds to the negatively charged ionized acceptors in the depletion region at the base side so that the total charge density increases.

At the collector side, where the depletion region was composed of the positively charged ionized donors, the injected electrons decreased the charge density and even changed the net charge from positive to negative. In fact, the concept of the depletion region is not reflected accurately here due to the presence of a large number of carriers. It should be described as a space charge region. This modulation of charge density, for a constant collector-base voltage, must be accompanied by a narrowing of the depletion region at the base and a widening of the depletion region at the collector side. Consequently, the undepleted (or neutral) base widens, resulting in a reduction of the current gain at high currents and a poorer frequency response.

5.10.3 BiCMOS Process Technology

BiCMOS technology can be developed in two basic ways: adding bipolar on CMOS or adding CMOS on bipolar. Because the discussion, thus far, has been concentrated on CMOS technology, the former method will be used to build BiCMOS. This method has been used for several recent BiCMOS developments.[43-46] It often compromises bipolar performance to be compatible with existing CMOS and minimize added process steps. In a BiCMOS circuit, vertical *npn* transistors are needed because they have higher current gain compared to a *pnp* transistor or a lateral device. Furthermore, each bipolar transistor should be isolated so that it is not limited to emitter-follower type logic only. N-wells are used in CMOS as the isolated collectors for the vertical *npn* transistors. Thus, CMOS with n-well[44] or twin-tub on p-substrate[46] is commonly extended to BiCMOS.

Because buried-layer is normally used for *npn* bipolars to reduce collector resistance, a buried-layer mask is defined and followed by a shallow arsenic implant (Fig. 5.29a). A p-type epi layer is then grown and n-wells are formed above the n^+ buried layer (Fig. 5.29b). After the standard CMOS LOCOS field isolation, another mask and boron implant are then needed to form the *npn* base (Fig. 5.29c). Subsequent to standard gate oxidation, polysilicon deposition and gate definition, the n^+ mask is used to implant arsenic for the nMOS S/D, the *npn* emitter and n^+ contacts for the collector and n-wells (Fig. 5.29d). Similarly, p^+ mask is used to implant boron for the pMOS S/D and base contact. Standard contact and metal masks and associated process steps finish the entire BiCMOS process (Fig. 5.29e). Notice that Masks 1 and 4 are the additional masks needed for making bipolar on a CMOS chip. With buried-layer, it is necessary to grow an epi layer, which is often already included in CMOS for latch-up prevention. Boron base implant is yet another extra process step. The entire process consists of 9 masks: buried layer, n-well, active area, *npn* base, poly, n^+, p^+, contact and metal. This sequence described the basic BiCMOS process in which MOSFETs are not scaled to minimize short-channel effects and *npn* bipolars are not high performance devices.

As discussed in Secs. 5.2 and 5.3, for micron or submicron CMOS, additional masks and implants are necessary for FET threshold adjustment and punchthrough protection, as well as for the double-level metal. The boron implant designed for n-channel punchthrough protection can also be used to form the *npn* base; however, this compromise does not, in general, produce a base well suited for the *npn* bipolar. To incorporate n-well and n^+ buried layer, the BiCMOS substrate must be a lightly-doped p-type. For latchup immunity, p^+ buried-layer can also be included to reduce the substrate resistance (shown as a dashed curve in Fig. 5.24c). The p^+ layer also creates a p/p^+ hi-low junction which is known to reduce α-particle induced soft error

Greenthread Ex. 2059, p. 13 of 47
Semiconductor v. Greenthread
IPR2023-01242

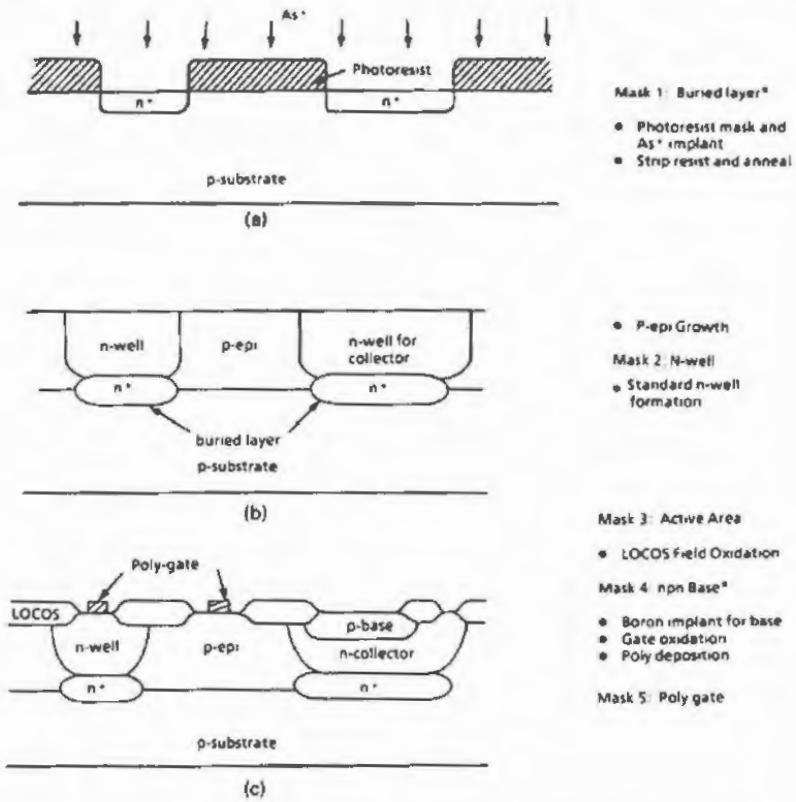


Figure 5.29 An n-well BiCMOS process: (a) buried layer formation; (b) epi growth and n-well formation; (c) active area & base formation and poly gate definition; (d) n^+; and (e) p^+ and contact & metal.

as described in Chapter 4. This addition requires one more mask, making the BiCMOS process three masks more than the twelve masks commonly used for today's CMOS processes. Another significant performance improvement is the addition of a polysilicon emitter[45,46] which is known to increase emitter efficiency, hence current gain. Because polysilicon technology has matured in CMOS, it would not be difficult to add to the bipolar devices. In fact, for SRAMs using poly-load resistors, the same polysilicon layer can be used for the poly emitters. Trench isolation is very attractive for this hybrid technology because several different types of devices need to be isolated with tight isolation space. Figure 5.30 shows an example of the

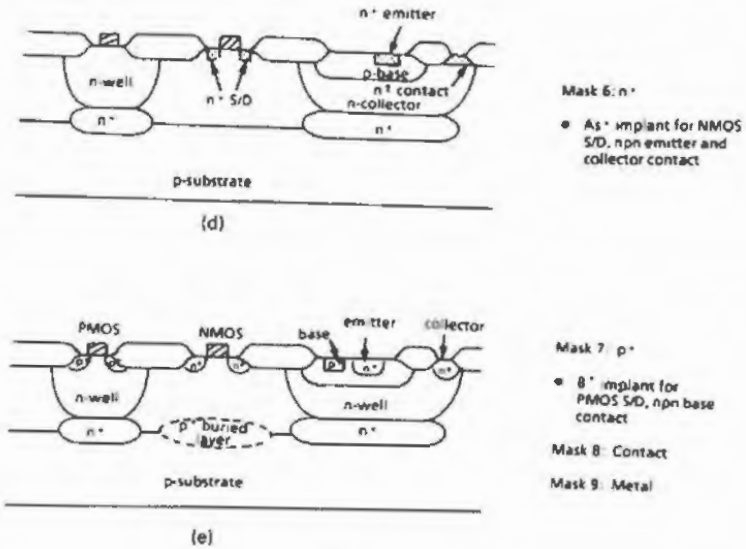


Figure 5.29 *Continued*

advanced BiCMOS device structure with trench isolation, n^+ and p^+ buried layer, and a polysilicon emitter.

5.10.4 BiCMOS Applications

Static random access memory. BiCMOS is penetrating the CMOS market in high speed applications. Because SRAMs are always performance-sensitive, BiCMOS is attractive for SRAMs requiring higher speed, higher density and lower power than what can currently be achieved by either CMOS or ECL. While CMOS cannot compete on speed, pure ECL dissipates too much power and is very costly. With BiCMOS SRAM, sub-10 ns access time and <1 W power dissipation are feasible. These characteristics offer BiCMOS an opportunity in high performance workstation and superminicomputer applications. One example is a second-level cache to the memory architecture that serves as a backup for higher system speeds by supporting the first-level (CPU) cache on misses and reducing the need to access the main memory, which normally consists of DRAMs with much longer (100–150 ns) access times. This second-level cache is best implemented by large (1 Mwords) SRAMs with cache cycle times set at 25 ns. 256K BiCMOS

Greenthread Ex. 2059, p. 14 of 47
Semiconductor v. Greenthread
IPR2023-01242

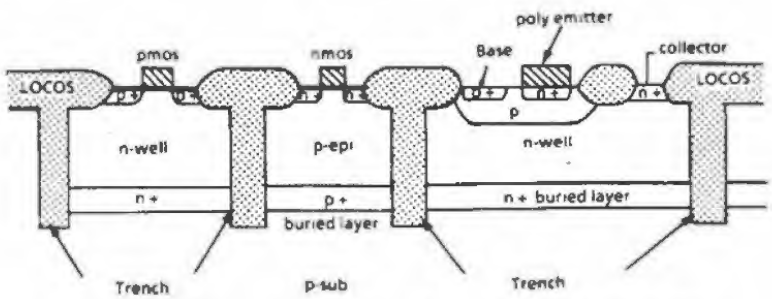


Figure 5.30 Schematic cross-section of an advanced BiCMOS device structure with trench isolation and polysilicon emitter.

SRAMs with 10–15 ns access time and low power dissipation are desirable for this unique application.

Most present BiCMOS SRAMs use conventional 6-T CMOS cells or 4-T resistor-load *n*MOS cells with bipolar transistors only for sense amplifiers and peripheral circuits.The BIT line and the $\overline{\text{BIT}}$ line from a CMOS RAM cell are connected to a bipolar differential pair of the sense amplifier for high speed sensing. The bipolar differential amplifier is similar to the CMOS amplifier discussed in Chapter 4. The two *n*MOS input FETs are now re-

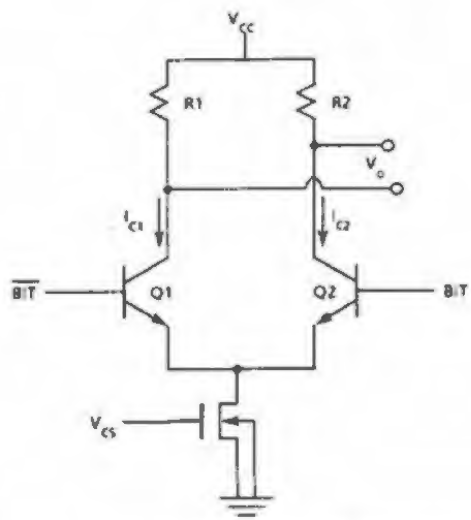


Figure 5.31 Circuit schematic of a BiCMOS differential amplifier.

placed by the two *npn* bipolar devices shown in Fig. 5.31. The load resistors can be implemented by *p*MOSFETs. Compared to the CMOS differential pair in which the FET drain current was linearly proportional to the gate voltage, the bipolar collector current (I_{c1} or I_{c2}) is exponentially dependent on the base-emitter voltage shown in Eq. 5.10. As a result, the bipolar differential amplifier has much higher gain for a very small differential input signal, thereby leading to a very fast sense amplifier.

BiCMOS SRAMs containing CMOS or *n*MOS cells have been made at 256 kbit complexity.[47–49] Table 5.4 summarizes their characteristics and compares them to high speed ECL RAMs.[50–52] BiCMOS access times are 10 ± 2 ns, not much slower than ECL speed, the power dissipation is much less, and the cell sizes are much smaller. The Hitachi ECL uses a Ta_2O_5 capacitor and the Fujitsu ECL uses a *pnp* load cell so that memory cell size can be kept around 500 μm^2. BiCMOS technologies for the 256K SRAMs have current gain (β) ranging from 80–100 and f_T's at 8 ± 1 GHz. For comparison, 1 Mbit CMOS SRAMs made by ~0.8 μm technology have access times of 15–25 ns,[53–56] with cell sizes of ~60 μm^2 for 6-T cells.

BiCMOS SRAM cells have been proposed[57,58] to decrease access time while maintaining low power of CMOS memory arrays. Using a bipolar transistor to deliver more current can reduce the time to charge BIT line capacitance during READ operation, but at the expense of larger cell area for the bipolar device. A two-port BiCMOS SRAM cell that combined ECL level word-line voltage swings and emitter-follower bit line coupling with a static CMOS latch was designed and fabricated.[58]

As shown in Fig. 5.32, CMOS cross-coupled latch is formed by M1, M2, M3 and M4. The READ Word Line (RWL) acts as a power supply for the CMOS latch, and the WRITE Word Line (WWL) controls the pass transistor (M5) just like a CMOS SRAM cell, but with a single-ended WRITE. WRITE Bit Lines (WBLs) for unselected columns have to be biased at a safe level which does not change the state of the cell residing in the selected row. The condition of disturbing the unselected cells is more severe when RWL is low because of the reduced noise margin associated with the small supply on the CMOS latch.

The cell is read by raising RWL. The emitter follower (Q6) and Q7 form a differential pair for the sense amplifier. For a low output, the *npn* bipolar device Q6 is off and the bit line current I_{BI} flows through Q7. The $I_{BI}R$ drop results in a low V_o. For high output, the increase of RWL is coupled to the base of Q6 through the *p*MOS (M2). Q6 is then switching I_{BI} away from Q7 and V_o is at logic high.

An experimental 4K × 1 bit two-port SRAM[58] fabricated by a 1.5 μm 5 GHz BiCMOS process offers a READ access time of 3.8 ns, a power dissipation of 550 mW, and the cell area is 650 μm^2, approximately 35% larger than that of a 6-transistor CMOS cell implemented with the same design rules. The memory can offer independent and simultaneous READ and

Greenthread Ex. 2059, p. 15 of 47
Semiconductor v. Greenthread
IPR2023-01242

TABLE 5.4 256K BiCMOS SRAMs and ECL SRAMs

Source	Architecture	Technology	Design Rule	Cell Size in μm^2	Access time (ns)	Power (W)	β	f_t in GHz
National[47]	256 k × 1	BiCMOS	1.0 μm	96	12	0.72	80	7
TI[48]	256 k × 1	BiCMOS	0.8 μm	117	8	0.75	80	7
Hitachi[49]	256 k × 1	BiCMOS	1.0 μm	54.7	8	0.4	100	9
Hitachi[50]	4 k × 4	ECL	1.0 μm	495	3.5	2		8
Fujitsu[51]	64 k × 1	ECL	1.2 μm	524	10	1.3	100	5
IBM[52]	512 k × 10	ECL	1.2 μm	760	1	2.4	80	

* Using TiN for self-aligned local interconnect
†Using Schottky Barrier Diode

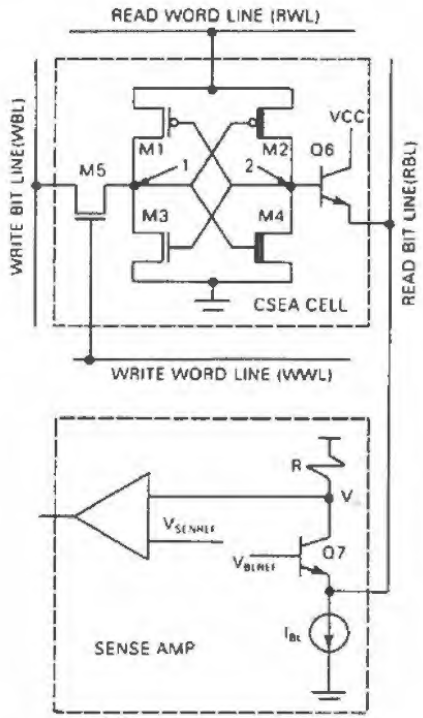


Figure 5.32 Circuit schematic of a two-port BiCMOS SRAM cell (Ref. 58, © 1988, IEEE).

WRITE, making it applicable for the design of video, cache, and other application-specific memories.

Gate array and standard cells. Combining bipolar and CMOS offers I/O flexibility (ECL and TTL compatible) and new architectures, in addition to high speed and high density in gate array applications.[59] Bipolar provides high driving ability with reduced layout area, making it attractive for high density gate arrays. Internal CMOS logic with a bipolar buffer results in negligible static power dissipation and low propagation delay under loading. Because high density gate arrays normally use fan-out of 3–5 or even larger, BiCMOS certainly has an advantage in speed as previously described. For large arrays, BiCMOS has less clock skew because of its high driving ability per unit area; deep clock trees are not needed in general. Large signal swing

Greenthread Ex. 2059, p. 16 of 47
Semiconductor v. Greenthread
IPR2023-01242

in CMOS often results in noise spikes. BiCMOS offers ECL I/Os with small signal swing. Switching noise can then be better controlled. BiCMOS, CMOS RAM cells, ECL buffers, macrocells, and analog functions can all be integrated in application-specific gate arrays.

For a standard cell, BiCMOS also offers a large degree of freedom in cell design and high speed ECL blocks can be developed and integrated with CMOS/BiCMOS logic. Finally, the merge of *npn* bipolar and CMOS transistors can offer high speed and low offset operational amplifiers and comparators. As previously described, a bipolar differential pair offers high gain and more precisely controlled input voltages as compared with its CMOS counterparts. Compared to MOS technology in which threshold voltage is sensitive to manufacturing process and device dimensions, the bipolar emitter-base turn-on voltage (V_{BE}) is much more precisely controlled and well-matched pairs can be more easily obtained. Offset voltages of MOS operational amplifiers are typically an order of magnitude larger than those of bipolar circuits[60] because closely matched MOS input devices are difficult to make and voltage gains are lower. On the other hand, MOS circuits can be made with higher packing density. BiCMOS thus offers the low input offsets and high gains of bipolar circuit components while maintaining high packing density and low power. These components are the basic blocks for high speed and low power A/D and D/A converters. BiCMOS certainly provides more opportunities in the integration of digital and analog functions on a single VLSI chip.

5.10.5 Summary

In general, about three extra masks and associated processes are required for adding bipolar transistors on a CMOS IC. Major gains are: faster speed, higher current driving ability, ECL I/O compatibility, and more flexibility in analog design. Longer turn-around time and higher costs are expected. It is not worth using BiCMOS to replace CMOS for implementing digital systems with low or medium performance. For high performance digital applications or mixed digital/analog systems, additional values offered by BiCMOS justify its extra cost. Moreover, for state-of-the-art CMOS VLSI technology, the mask count is about a dozen and the process steps are already large; another three masks and additional process steps needed for making bipolar devices do not comprise more than 15–20% of the CMOS process steps. In addition, as bipolar and CMOS technology approach submicron dimensions, they share more processing techniques such as epitaxy, polysilicon gate/emitter, trench isolation, and RIE. This technological convergence[61] further justifies the merging of bipolar and CMOS for high performance analog/digital system applications.

REFERENCES

1. K. Lee and A. R. Neureuther, "SIMPL-2 (SIMulated Profiles from the Layout—Version 2)," Technical Digest Symp. on VLSI Technology, Kobe, Japan, p. 64, 1985.
2. L. C. Parrillo, R. S. Payne, R. E. Davis, G. W. Reutlinger, and R. L. Field, "Twin-tub CMOS: A technology for VLSI circuits," in *IEDM Tech. Dig.*, p. 752, 1980; also L. C. Parrillo, L. K. Wang, R. D. Swenumson, R. L. Field, R. C. Melin, and R. A. Levy, "Twin-tub CMOS II: An advanced VLSI technology," in *IEDM Tech. Dig.*, p. 706, 1982.
3. J. Agraz-Guerena, R. Ashton, W. Bertram, R. Melin, R. Sun, and J. T. Clemens, "Twin Tub III—A third generation CMOS technology," in *IEDM Tech. Dig.*, p. 63, 1984.
4. R. D. Rung, C. J. Dell'Oca, and L. G. Walker, "A retrograde *p*-well for high-density CMOS," *IEEE Trans. Electron Devices*, vol. ED-28, p. 1115, 1981.
5. S. R. Combs, "Scalable retrograde *p*-well CMOS technology," in *IEDM Tech. Dig.*, p. 346, 1981.
6. J. Y. Chen, "Quadruple-well CMOS for VLSI technology," *IEEE Trans. on Electron Devices*, vol. ED-31, p. 910, 1984.
7. R. Chwang and K. Yu, "CHMOS: An *n*-well bulk CMOS technology for VLSI," *VLSI Design*, 4th quarter, p. 42, 1981.
8. H. Shichijo, "A re-examination of practical performance limits of scaled *n*-channel and *p*-channel MOS devices for VLSI," *Solid State Electronics*, vol. 26, p. 969, 1983.
9. A. Schmitz and J. Y. Chen, "Design, modelling and fabrication of subhalf-micrometer CMOS transistors," *IEEE Trans. Electron Devices*, vol. EO-33, p. 148, 1986.
10. K. Yu, R. Chwang, M. Bohr, P. Warkentin, S. Stern, and C. N. Berglund, "HMOS-CMOS: A low power high performance technology," *IEEE Journal of Solid State Circuits*, vol. SC-16, p. 454, 1981.
11. R. A. Martin and J. Y. Chen, "Optimized retrograde *n*-well for one micron CMOS technology," *IEEE J. Solid-State Circuits*, vol. 21, p. 286, 1986; also, A. G. Lewis, J. Y. Chen, R. A. Martin, and T. Y. Hunag, "Device isolation in high density CMOS," in the Proc. of the 16th European Solid State Device Res. Conf., Cambridge, U.K., 1986.
12. J. O. Borland, M. Kuo, J. Shibley, B. Roberts, R. Schindler, and T. Dalrymple, "An intrinsic gettering process to improve minority carrier lifetimes in MOS and bipolar silicon epitaxial technology," Semiconductor Processing, ASTM STP 850, C. Gupta, Ed., American Society for Testing and Materials, 1984.
13. L. G. Heller, W. R. Griffin, J. W. Davis, and N. G. Thomas, "Cascade voltage switch logic: A differential CMOS logic family," in *ISSCC*, p. 16, 1984.
14. K. Kokkonen, and R. Pashley, "Modular approach to CMOS technology tailors process to application," *Electronics*, p. 129, 1984.

Greenthread Ex. 2059, p. 17 of 47
Semiconductor v. Greenthread
IPR2023-01242

15. S. Kohyama, J. Matsunaga, and K. Hashimoto, "Directions in CMOS technology," in *IEDM Digest*, p. 151, 1983.

16. Y. Taur, G. H. Hu, R. H. Dennard, L. M. Terman, C. Y. Ting, and K. E. Petrillo, "A self-aligned 1-μm-channel CMOS technology with retrograde *n*-well and thin epitaxy," *J. of Solid-State Circuits*, vol. SC-20, p. 123, 1985.

17. D. C. Mayer, P. K. Vasadev, J. Y. Lee, Y. K. Allen, and R. C. Henderson, "A short channel CMOS/SOS Technology in recrystallized 0.3-μm thick silicon-on-sapphire films," *IEEE Trans. Electron Devices*, vol. EDL-5, p. 156, 1984.

18. D. Leong, "Silicon on sapphire," Silicon on Insulator Technologies for Integrated-Circuit Applications, U.C. Berkeley short course, Oct. 1983.

19. E. Sun, "Performance comparison of MOS transistors and circuits fabricated using bulk and SOS technology," *CICC*, May 1980.

20. K. Bean and W. Runyan, "Dielectric Isolation: comprehensive, current, and future," *J. Electrochem. Soc.*, vol. 124, p. 5C, 1977.

21. Y. Arita and Y. Sunohara, "Formation and properties of porous silicon film," *J. Electrochem. Soc.*, vol. 124, p. 285, 1977.

22. S. Nakashima and K. Okwada, "Electrical characteristics of an upper interface on a buried SiO_2 layer formed by oxygen implantation," *Japan J. Appl. Physics.*, vol. 22, p. 1119, 1983.

23. D. D. Rathman, D. J. Silversmith, and J. A. Burns, "Lateral epitaxial overgrowth of silicon on SiO_2," *J. Electrochem. Soc.*, vol. 129, p. 2303, 1982.

24. L. Jastrzebski, A. C. Ipri, and J. F. Corboy, "Device characterization on monocrystalline silicon grown over SiO_2 by the ELO (Epitaxial Lateral Overgrowth)," *IEEE Electron Device Lett.*, EDL-4, p. 32, 1983.

25. T. I. Kamin and D. R. Bradbury, "Trench-isolated transistors in lateral CVD epitaxial silicon-on-insulator films," *IEEE Electron Device Lett.*, EDL-5, p. 449, 1984.

26. J. Maserjian, "Single-crystal germanium films by microzone melting," *Solid State Electronics*, vol. 6, p. 4477, 1963.

27. G. Zimmer and H. Vogt, "CMOS on buried nitride—a VLSI SOI technology," *IEEE Trans. on Electron Devices*, vol. ED-30, p. 1515, 1983.

28. K. Hashimoto, T.I. Kamins, K. M. Cham, and S. Y. Chiang, "Characteristics of submicrometer CMOS transistors in implanted- buried-oxide SOI films," in *IEDM Dig.*, p. 672, 1985.

29. Y. Omura, S. Nakashima, and K. Izumi, "A 4kb CMOS/SIMOX SRAM," *Dig. of the Sym. on VLSI Technology*, p. 24, 1985.

30. T. I. Kamin, "MOS Transistors in beam-recrystallized polysilicon," *IEDM Digest*, p. 420, 1982.

31. D. B. Rensch and J. Y. Chen, "Recrystallization of Si films on thermal SiO_2—coated Si substrates using a high-speed *e*-beam line source," *IEEE Electron Device Lett.*, vol. EDL-5, p. 38, 1984.

32. M. W. Geis, H. I. Smith, B. Y. Tsaur, S. C. Fan, E. W. Maby, and D. A. Antoniadis, "Zone-melting recrystallization of encapsulated silicon films on SiO_2—Morphology and crystallography," *Appl. Phys. Lett.*, vol. 40, p. 158, Jan. 1982, and references therein.

33. Motoo Nakano, "3-D SOI/CMOS," in *IEDM Dig.*, p. 792, 1984; also S. Kawamura, N. Sasaki, T. Iwai, M. Nakano, and M. Takegi, "Three-dimensional CMOS IC's fabricated by using beam recrystallization," *IEEE Trans. Electron Device Lett.*, vol. EDL-4, p. 366, 1983.

34. J. F. Gibbons and K. F. Lee, "One-gate-wide CMOS inverter on recrystallized polysilicon," *IEEE Electron Device Lett.*, vol. EDL-1, p. 117, 1980.

35. S. Akiyama, M. Yoneda, S. Ogawa, N. Yoshii, and Y. Terui, "Fabrication technologies for multilayer CMOS device," *Dig. Sym. on VLSI Technology*, p. 28, 1985.

36. K. Sakashita, T. Nishimura, S. Kusunoki, Y. Kuramitsu, and Y. Akasaka, "1.1 K-gate CMOS/SOI gate array by laser recrystallization technique," *Dig. of Sym. on VLSI Technology*, p. 32, 1985.

37. Y. Inoue, K. Sugahara, S. Kusunoki, M. Nakaya, T. Nishimura, Y. Horiba, Y. Akasaka, and H. Nakata, "A three dimensional static RAM," *IEEE Electron Devices Lett.*, EDL-7, p. 327, 1986.

38. S. D. S. Malhi, R. Karnaugh, A. H. Shah, L. Hite, P. K. Chatterjee, H. E. Davis, S. S. Mahant-shetti, C. D. Gosmeyer, R. S. Sundarean, C. E. Chen, H. W. Lam, R. A. Haken, R. F. Pinizzotto, and R. K. Hester, "A VLSI suitable 2-μm stacked CMOS process," presented at 1984 Device Research Conf., abstract published in *IEEE Trans. Electron Devices*, vol. ED-31, p. 1981, 1984.

39. M. P. Brassington, M. El-Diwany, P. Tuntasood and R. R. Razouk, "An advanced submicron BiCMOS technology for VLSI applications," in the *Digest of Sym. on VLSI Technology*, p. 89, 1988.

40. A. S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley and Sons, Inc., N.Y. 1967.

41. R. S. Muller and T.T. Kamins, *Device Electronics for Integrated Circuits*, ed. by John Wiley & Sons, N.Y. 1977.

42. C. T. Kirk, "A theory of transistor cutoff frequency (f_t) fall off at high current densities," *IEEE Trans. Electron Devices*, ED-9, p. 164, 1962.

43. H. Higuchi, G. Kitsukawa, T. Ikeda, Y. Nishio, N. Sasaki, and K. Ogiue, "Performance and structures of scaled-down bipolar devices merged with CMOS-FETs," *IEDM Dig.*, p. 694, 1984.

44. J. Miyamoto, S. Saitoh, J. Momose, H. Shibata, K. Kanzaki, and S. Kohyama, "A 1.0 μm *N*-well CMOS/Bipolar technology for VLSI circuits," *IEDM Dig.*, p. 63, 1983.

45. A. R. Alvarez, P. Meller, and B. Tien, "2 micron merged bipolar-CMOS technology," *IEDM Dig.*, p. 761, 1984.

46. T. Ikeda, T. Nagano, N. Momma, K. Miyata, H. Higuchi, M. Odaka, and K. Ogiue, "Advanced BiCMOS technology for high speed VLSI," *IEDM Dig.*, p. 408, 1986.

47. R. A. Kertis, D. D. Smith, and T. L. Bowman, "A 12 ns 256K BiCMOS SRAM,"

Greenthread Ex. 2059, p. 18 of 47
Semiconductor v. Greenthread
IPR2023-01242

ISSCC Digest, p. 186, 1988; also in *IEEE J. Solid-State Circuits*, vol. 23, p. 1048, 1988.

48. H. V. Tran, D. B. Scott, P. K. Fung, R. H. Havemann, R. E. Eklund, T. E. Ham, R. A. Haken, and A. Shah, "An 8 ns Battery Back-up sibmicron BiCMOS 256K ECL SRAM," *ISSCC Digest*, p. 188, 1988.
49. N. Tamba, S. Miyaoka, M. Odaka, K. Ogiue, K. Yamada, T. Ikeda, M. Hirao, H. Higuchi, and H. Uchida, "An 8 ns 256K BiCMOS RAM," *ISSCC Digest*, p. 184, 1988.
50. N. Homma, K. Yamaguchi, H. Nanbu, K. Kanetani, Y. Nishioka, A. Uchida, and K. Ogiue, "A 3.5-ns, 2-W, 20-mm², 16-Kbit ECL bipolar RAM," *IEEE J. Solid State Circuits*, vol. SC-21, p. 675, 1988.
51. Y. Okajima, K. Toyoda, T. Awaya, K. Tanka, and Y. Nakamura, "64Kb ECL RAM with redundancy," *ISSCC Digest*, p. 48, 1985.
52. C-T. Chuang, D. D. Tang, G. P. Li, E. Hackbarth and R. R. Boedeker, "A 1.0-ns 5-kbit ECL RAM," *IEEE J. Solid State Circuits*, vol. SC-21, p. 670, 1986.
53. K. Sasaki, S. Hanamura, K. Ueda, T. Oono, O. Minato, K. Nishimura, Y. Sakai, S. Meguro, M. Tsunematsu, T. Masuhara, M. Kubotera, H. Toyoshima, "A 15 ns 1 Mb CMOS SRAM," *ISSCC Digest*, p. 174, 1988.
54. H. Shimada, Y. Tange, K. Tanimoto, M. Shiraishi, N. Suzuki, and T. Nomura, "An 18 ns 1 Mb CMOS SRAM," *ISSCC Digest*, p. 176, 1988.
55. F. List, S. Bell, S. Chu, J. Dikken, C. Hartgring, J. Raemaekers, B. Walsh, and R. Salters, "A 25 ns low-power full-CMOS 1 Mb SRAM," *ISSCC Digest*, p. 178, 1988.
56. H. Lee, B. El-Kareh, R. Glaker, G. Gravenities, R. Lipa, J. Maslack, J. Pessetto, W. Pokorny, M. Roberge, T. Williams, H. Zelelr, and K. Beilstein, "An experimental 1 Mb CMOS SRAM with configurable organization and operation," *ISSCC Digest*, p. 180, 1988.
57. H. De Los Santos and B. Hoefflinger, "On the analysis and design of CMOS-Bipolar SRAM's," *IEEE J. of Solid-State Circuits*, vol. SC-22, p. 616, 1987.
58. T. S. Yang, M. A. Horowitz, and B. A. Wooley, "A 4 nsec 4K × 1 bit two-port BiCMOS SRAM," *IEEE J. Solid State Circuits*, vol. SC-23, p. 1030, 1988.
59. A. R. Alvarez and D. W. Schucker, "BiCMOS technology for semi-custom integrated circuits," in the Proc. of Custom Integrated Circuits Conf., p. 22.1.1, 1988.
60. D. A. Hodges, P. R. Gray, and R. W. Brodersen, "Potential of MOS technologies for analog integrated circuits," *IEEE J. of Solid State Circuits*, vol. SC-13, p. 285, 1988.
61. A. Wieder, "Submicron bipolar technology—new chances for high speed applications," *IEDM Dig.*, p. 8, 1986.

EXERCISES

1. In a *p*-well CMOS, the *n*-type substrate has a doping concentration of 1×10^{15} cm^{-3} and the *p*-well concentration is 5×10^{16} cm^{-3}; calculate source/drain junction capacitances and body effect coefficients for both *n*- and *p*-channel FETs assuming the gate oxide is 200 Å for both devices.
2. In a standard CMOS process, specify the necessary masking sequence in correct order and also state the number of masks needed for a double-level metal process.
3. To design a retrograde *p*-well with a well depth of 1 μm, how high an implant energy must be selected for the boron implant and what should be the minimum resist thickness if a resist mask is used during the high energy implant?
4. Repeat Exercise 3 and assume a retrograde *n*-well is needed.
5. Describe the major process steps and masking sequence for a CMOS/SOS and compare them with a bulk CMOS process.
6. In an *npn* bipolar transistor, the emitter sheet resistance is 50 Ω/□; the intrinsic base sheet resistance is 5000 Ω/□; the base width is 0.1 μm; the electron diffusion length in the base region is 10 μm; and, the electron mobility is 1000 cm²/V-sec; calculate the common emitter current gain (β) and maximum frequency ($1/\tau_t$).
7. From the Gummel plot shown in Fig. 5.28, calculate the bipolar transconductance (g_m) defined as $\partial I_c/\partial V_{BE}$ and compare it with the transconductance ($g_m = \partial I_d/\partial V_g$) of the *n*-channel MOSFET shown in Fig. 2.13.
8. Describe the additional masking layers and process steps needed for adding *npn* bipolar transistors to an *n*-well CMOS process.

Greenthread Ex. 2059, p. 19 of 47
Semiconductor v. Greenthread
IPR2023-01242

contact to poly-gate distance is 10 μm, calculate the intrinsic transistor channel conductance (g_d), transconductance(g_m), and total g_d and g_m in which source and drain resistances are included (assuming contact resistance is negligible).

4. Describe punchthrough current and subthreshold current and explain the differences between them.
5. An n-channel MOSFET has 200 Å t_{ox}, 1 μm L_{eff}, 0.2 μm x_j and V_t = 0.6 V. If the device is biased at V_g = 3 V and V_d = 5 V, calculate the saturation voltage V_{dsat} and maximum electric field E_m.
6. (a) Write the E_m expression for an LDD nMOSFET and using the E_m value calculated in Exercise 5, calculate the decrease of E_m if the n^- LDD length is 0.2 μm. (b) If a FET channel width is 10 μm and the S/D sheet resistance in the LDD region is 1000 Ω/□, calculate additional S/D resistance due to the LDD insertion and compare it with the intrinsic FET resistance.
7. Describe the process sequence for an LDD CMOS process and state qualitatively the advantages and disadvantages of an LDD structure.
8. If n^+ polysilicon is used for both n- and p-channel FET gates in a CMOS chip, what are the threshold voltages for the n- and p-channel devices assuming 250 Å gate oxide, 5×10^{16} cm^{-3} substrate doping concentration, and a fixed oxide charge density of 5×10^{10} cm^{-2}?
9. For Exercise 8, what are the threshold voltages if p^+ polysilicon is used for both n- and p-channel devices?

7

CMOS ISOLATION

The preceding chapter discussed device design for individual n- and p-channel transistors. Because as many as hundreds of thousands of transistors are used in a CMOS VLSI chip, they must be properly isolated so that the conduction state for each transistor can be individually controlled. What is unique in a CMOS circuit is the isolation among opposite-type transistors as well as among similar-type transistors. This chapter starts with the physics and techniques of field isolation in MOS technologies. It then discusses isolation in CMOS including physical and electrical isolation, failure mechanisms, dependence on process architecture, and new isolation techniques for future CMOS.

7.1 BACKGROUND

One important aspect of integrated circuit technology is the art of isolation among transistors which are all built in a common silicon substrate. The state (on or off) and conductance of each transistor can only be controlled if proper isolation exists among transistors. If not, leakage current paths occur, causing DC power dissipation, noise margin degradation, and voltage shift on a dynamic node. Crosstalk among transistors can destroy the logic state

Greenthread Ex. 2059, p. 20 of 47
Semiconductor v. Greenthread
IPR2023-01242

of each device. Dynamic RAMs are very sensitive to leakage because it can degrade memory holding time, hence requiring more refresh cycles. In a CMOS circuit, leakage current in the isolation region can escalate latchup, an important phenomenon to be described in the next chapter.

Device isolation is becoming more critical in VLSI because of increased transistor counts and decreased isolation space. For a VLSI chip, a small amount of leakage per transistor can induce significant power dissipation for the entire chip. For example, if a VLSI chip has one million transistors with each transistor producing 1 μA leakage, the total power dissipation for the entire chip would increase on the order of a few watts. Because the transistors are closely packed on a VLSI chip, adequate isolation between neighboring devices is more difficult.

In MOS technologies, isolation is commonly formed using a thick oxide and/or a heavily-doped Si layer in the field region among active MOSFETs (Fig. 7.1). The thick oxide is often called field oxide and the heavily-doped region is often referred to as channel stop. This field isolation works well for isolating active MOSFETs if only the Si surface in the active channel region is allowed to invert. This inversion occurs when the gate voltage reaches the threshold voltage of the active MOSFETs. As described in Chapter 2, threshold voltage increases with the increase of oxide thickness or substrate doping concentration according to Eq. 2.11. Thus, threshold voltage in a field region consisting of a thick oxide and/or heavily doped region is normally around 10 V or more, at least an order of magnitude higher than the threshold of active MOSFETs. As a result, surface inversion does not occur under the field oxide because the polysilicon or metal lines on the oxide is normally at 5 V or below. In normal operations, a conducting channel is allowed to form only in the active area, not in the field region.

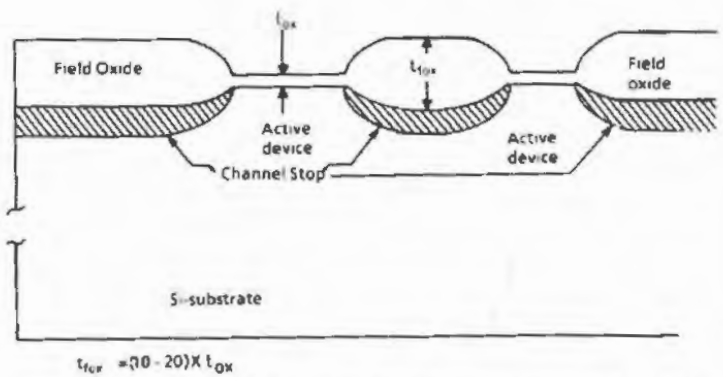


Figure 7.1 Schematic of field isolation with self-aligned channel stop for MOS ICs.

7.1.1 Test Structures

This section considers NMOS technology only. Shown in Fig. 7.2 is an IC structure consisting of an aluminum interconnect line running between two active n-channel MOSFETs. Both a top-view layout and a cross-sectional view are shown. The aluminum line acts as a gate on the thick field oxide and the two n^+ diffusions adjacent to it act as source and drain, forming a parasitic FET. This FET is commonly used as a test structure to detect leakage current in the field region under the aluminum interconnect lines. Because polysilicon lines are also used for interconnects in an IC, a similar structure as shown in Fig. 7.3 is also possible. Notice the subtle difference as the n^+ regions adjacent to the polysilicon gate do not abut to the field oxide. This exact structure in general may not exist in an actual IC layout because polysilicon lines normally do not extend over the active (thin oxide) area (see Fig. 7.4(a)) unless they are also used for the gate electrode of active transistors (Fig. 7.4(b)). Here, the polysilicon line serves as gate electrode as well as the interconnect between the two electrodes. In Fig. 7.3, the

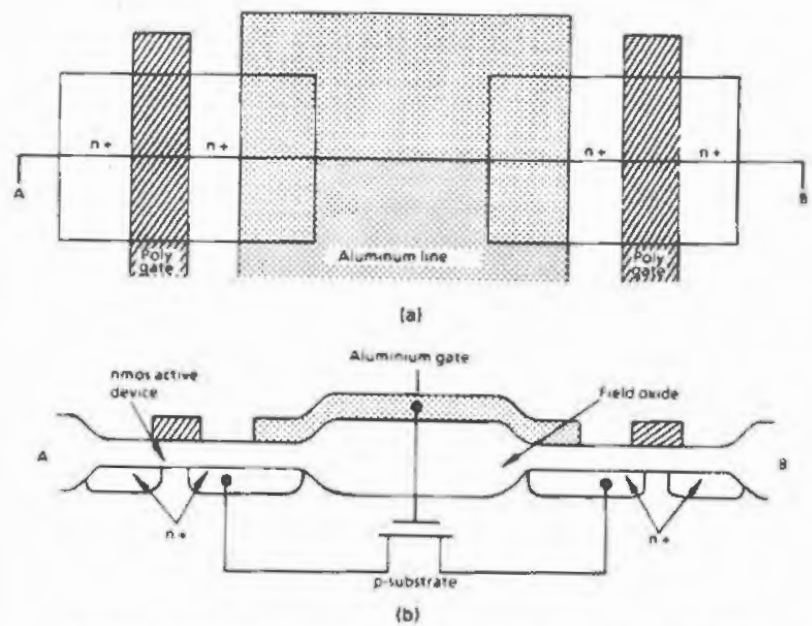


Figure 7.2 Aluminum-gate parasitic MOSFET between two active devices in NMOS technology: (a) layout, and (b) cross-sectional view.

Greenthread Ex. 2059, p. 21 of 47
Semiconductor v. Greenthread
IPR2023-01242

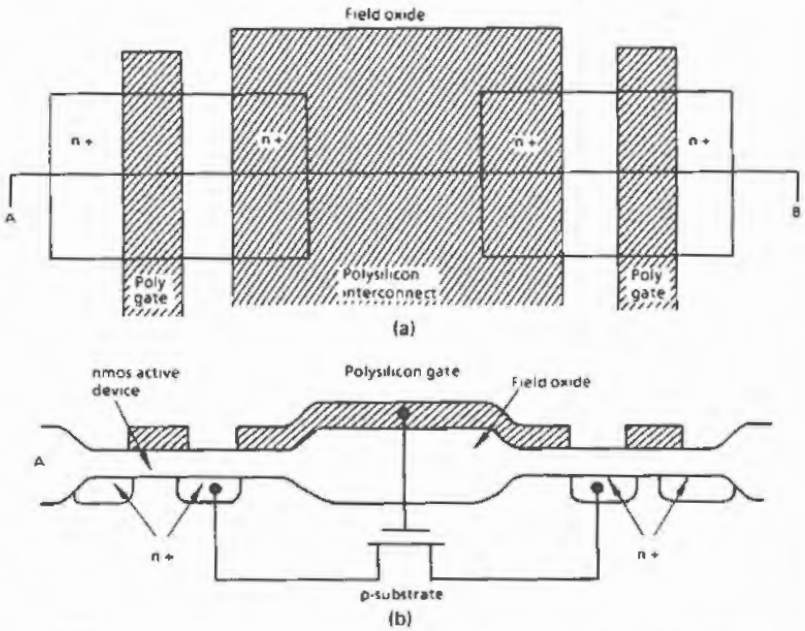


Figure 7.3 Polysilicon-gate parasitic MOSFET between two active devices in NMOS technology: (a) layout; and (b) cross-sectional view.

silicon surface under the thin oxide is normally inverted at a very low gate voltage ($\leq$1 V), resulting in ultra shallow "n^+" regions. If the silicon surface under the field oxide is inverted, the two transistors are basically shortened and isolation between the two transistors is completely broken down. Thus, the test structure shown in Fig. 7.3 is a good test vehicle for the case shown in Fig. 7.4(b), because leakage current measured from the parasitic FET indicates the onset of surface inversion under the field oxide.

For the layout shown in Fig. 7.4(a), the test structure that has polysilicon overlapping the field oxide also tests the worst case because inversion is harder to achieve under field oxide that is not covered with a polysilicon line (see areas pinpointed by arrows in Fig. 7.4(a)). The configuration in Fig. 7.4(a), which is often seen in a real circuit, is unlikely to have an isolation problem if the test structure in Fig. 7.3 passes the isolation test.

Another difference between the aluminum and polysilicon gate parasitic FETs is the thicker oxide in the former devices due to an additional oxide layer deposited to separate polysilicon and aluminum interconnects. The total oxide under the aluminum is typically 1–2 μm, while in the polysilicon device, the gate oxide is composed of thermally grown oxide only, which is

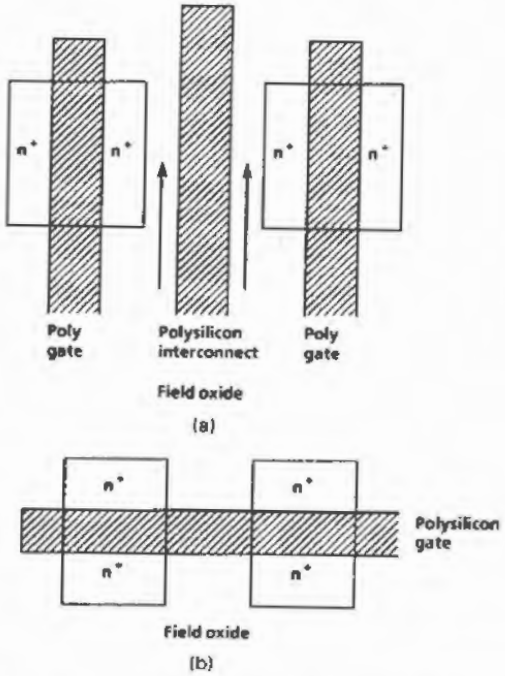


Figure 7.4 Polysilicon lines as transistor gate; and interconnect in real NMOS IC layouts.

about 0.5–1 μm. In both cases, the thick field oxide and the doped silicon surface underneath inhabit surface inversion unless the gate voltage is raised several volts above the 5 V supply.

7.1.2 Field Threshold Voltage

Field oxide threshold voltage (V_{TF}), as cited in the literature to evaluate the rigorousness of field isolation, has often been defined as the voltage required to get 1-μA current in a parasitic FET. This definition is crude and as such V_{TF} of >5 V does not necessarily correspond to an adequate isolation for circuits operated at 5 V power supply. Actually, the subthreshold swing of the parasitic thick-oxide FET is so large (e.g., S_t = 0.5–0.6 V/DEC) that the circuit must operate at 3–4 V below V_{TF} to ensure a <1 pA off-state current. In other words, V_{TF}, so defined, must be at least 8–9 V for a 5 V circuit operation. This condition must be met for minimum isolation spacing in a given technology, because V_{TF} decreases with isolation spacing in a similar

Greenthread Ex. 2059, p. 22 of 47
Semiconductor v. Greenthread
IPR2023-01242

way to the short-channel effect observed in active FETs.[1] Even worse, as temperature increases, V_{TF} decreases and the subthreshold swing (V/DEC) rises.[2] For example, V_{TF} reductions as much as 2 V were noted as the temperature was increased from 25 to 125°C.[1]

Finally, another problem exists if circuits are used in a radiation environment, because the threshold shift, caused by radiation-induced oxide charges, is proportional to the oxide thickness to a power of between 2 and 3.[3] The resulting threshold shift for a field oxide can be 5 to 10 V depending on the total dose of radiation. It is therefore essential to provide an additional margin for radiation-hard circuits in which field threshold voltage as high as ≈ 20 V may be necessary. A more stringent measure of isolation will be discussed in Sec. 7.3.

7.2 MOS ISOLATION TECHNIQUES

An ideal isolation technique must satisfy all of the following requirements. First, leakage current must be negligible between active devices. For MOS technology, the silicon surface under the field oxide cannot be inverted during circuit operation. Second, spacing between active devices should be minimal. This requirement is particularly important for VLSI isolation. Third, isolation should not consume a significant portion of the active device area or result in significant narrow-channel effect as described in Chapter 2. In other words, the transition from an isolation area to the neighboring active area should be abrupt. Fourth, the isolation process should not adversely affect the process parameters required for fabricating the active device. In this way, the performance of an active device can be optimal. Finally, a process should not be too complicated to control. For some applications, the low leakage requirement should be met in high-temperature or radiation environments.

7.2.1 Window or Moat Isolation

This method involves etching windows for the active devices in a uniform thick oxide. This technique was used before LOCOS[4] was invented and was an industrial standard. The process is simple if a channel stop is not incorporated. Although sufficiently high field threshold voltage can be achieved with a thick enough field oxide alone, step coverage becomes a serious problem. Metal or polysilicon lines at the sharp edge of the oxide steps tend to break, especially when they are narrowed down for VLSI interconnects. Techniques such as implanting argon ions near the top portion of the oxide have been used to change the etching rate for the formation of a sloped oxide edge. This method facilitates step coverage, however, at the expense of a

larger layout area to accommodate the gradual transition from the active to isolation region. Obviously, reduced oxide thickness is most effective to ease the step coverage problem, but then a channel stop must be incorporated to produce adequate field threshold voltage.

For p-channel MOS technology, a window isolation structure with self-aligned channel stop was developed.[5] In this structure, a phosphorus implant over the entire wafer is performed after the windows are etched into the field oxide. The phosphorus-doped active region is compensated by a shallow boron implant that goes only through windows patterned in the field oxide. The phosphorus implant forms an n-type channel stop under the field oxide, whereas the boron implant through the windows provides adequate channel doping for active devices. This process is similar to the retrograde n-well CMOS process as described in Chapter 5. Good device characteristics were obtained for MOSFETs with 8-μm channel length. However, the two implant conditions must be controlled precisely to achieve the exact doping compensation for threshold control. A 2 σ variation of 0.35 V was found in earlier work.[5] This control problem becomes more serious for scaled devices, which require higher implant doses and lower threshold voltage.

Similar window isolation has been done for NMOS technology. Active channel regions are first defined by etching windows in a thick field oxide. Boron ions are deeply implanted under the oxide to form a p-type channel stop.[6] The boron impurities in the active regions are counter doped by a shallow phosphorus implant to set a low threshold voltage. Again, this process is like a retrograde p-well process. The cross-sectional view of the final device structure is shown in Fig. 7.5. Although submicron MOSFETs with adequate threshold and punchthrough voltages were made, they had poor body effect and high junction capacitance because of excessive boron impurities in the channel region. Moreover, a significant 2-D fringing field effect existed in the devices with narrow channel widths. Figure 7.6 shows the 2-D contours for potentials and carrier concentrations in a narrow-channel device with its gate biased slightly above threshold voltage. Only half of the channel is shown due to the symmetry of the structure. The cross-section is perpendicular to the direction of current. Even with a precisely-defined field oxide edge, both surface potential and electron concentration decrease significantly near the edge of the channel, resulting in a reduction of effective (or electrical) channel width. This channel-narrowing effect is the origin of threshold increase and transconductance decrease as the channel width is reduced.

In spite of the problems previously mentioned, direct window isolation provides advantages for isolation in high-density ICs. Compared to LOCOS, another isolation technique, direct window (moat) isolation has been shown to save 30%–40% of the silicon area.[1] It was pointed out that increasing doping concentration in short-channel (≈1 μm) MOSFETs relaxed the requirement of N_A (channel stop) > N_A (channel), where N_A represented the

Greenthread Ex. 2059, p. 23 of 47
Semiconductor v. Greenthread
IPR2023-01242

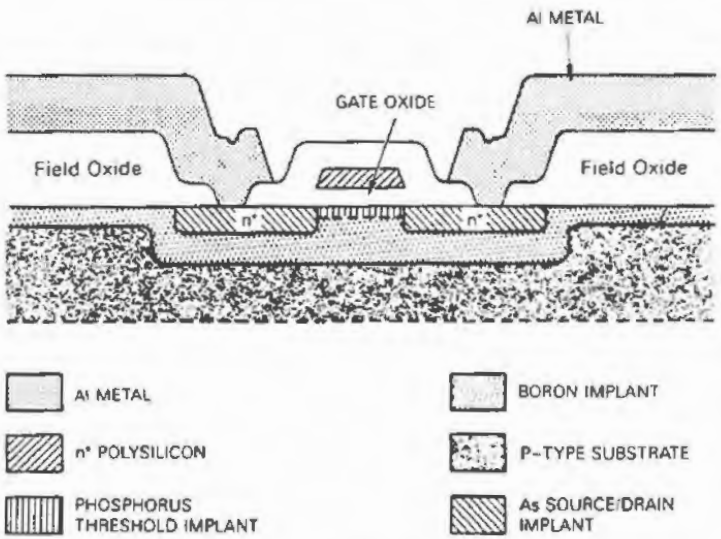


Figure 7.5 Cross-sectional view of an *n*MOS device using window isolation and channel stop.

dopant concentration.[1,7] It was therefore possible, using a single boron implant, to form a channel stop under the field oxide and provide punchthrough protection in the channel region.

Another implant was made to adjust threshold voltage after the active areas were defined. This technique, however, was not completely satisfactory for an optimum NMOS process. Neither channel stop nor channel area can have optimized doping distribution by using only one implant. The boron implants selected in the work[1] may have offered suitable punchthrough protection, but the boron profile was too deep to provide appropriate body effect and junction capacitance for the active devices. Also, the dopant concentration for the channel stop might not have been sufficient.

Another NMOS process that forms channel stops for device isolation, but independently dopes channel regions for active devices, has been proposed.[8] After etching windows into field oxide with a patterned photoresist layer, a layer of metal is deposited and the portion above the field oxide is subsequently lifted off, leaving metal just inside the windows. The metal inside the windows acts as an implant mask to block boron ions while they are implanted through the field oxide to form channel stops. The channel stops are self-aligned to the thick field oxide for high packing density. Meanwhile, the active devices can have an optimum impurity profile. The associated liftoff process is, however difficult for VLSI production.

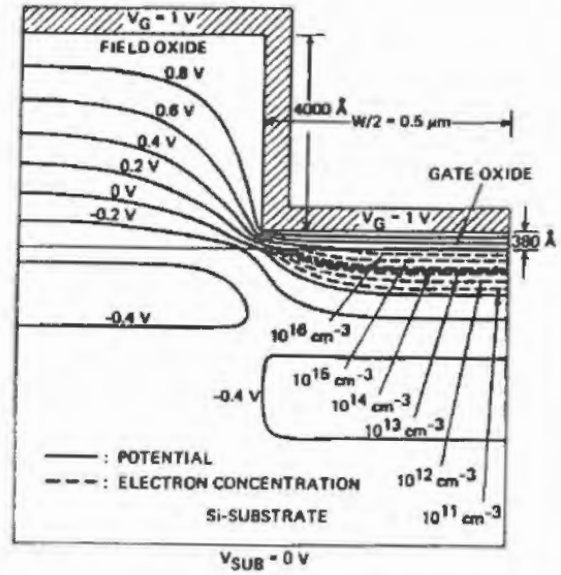


Figure 7.6 Two-dimensional contours of equipotentials and electron concentrations for a non-planar MOS structure in channel width direction (Ref. 13, © 1982 IEEE).

7.2.2 LOCal Oxidation of Silicon (LOCOS)

The LOCOS (LOCal Oxidation of Silicon) process[4] is commonly used for the isolation of active devices in the MOS IC industry. As shown in Fig. 7.7, this process uses a patterned nitride layer together with a thin buffer oxide underneath to mask the active area. This pad oxide releases stress caused by the nitride layer, and is therefore often referred to as SRO for Stress-Relief Oxide. The nitride/oxide mask is used to block ions during channel stop implantation and the same mask is then used for local oxidation. The nitride/oxide masking layer is stripped after LOCOS. The resulting isolation structure consists of a semi-recessed field oxide with the channel stop naturally self-aligned to the oxide edges. This standard LOCOS process can be modified to produce fully recessed field oxide if the Si substrate in the isolation region is etched before LOCOS. Normally, the amount of Si etched is roughly half of the desired field oxide thickness so that an isoplanar surface is formed after LOCOS. This scheme however adds more processing steps and concerns for etch-induced crystal damage and increased oxidation encroachment.

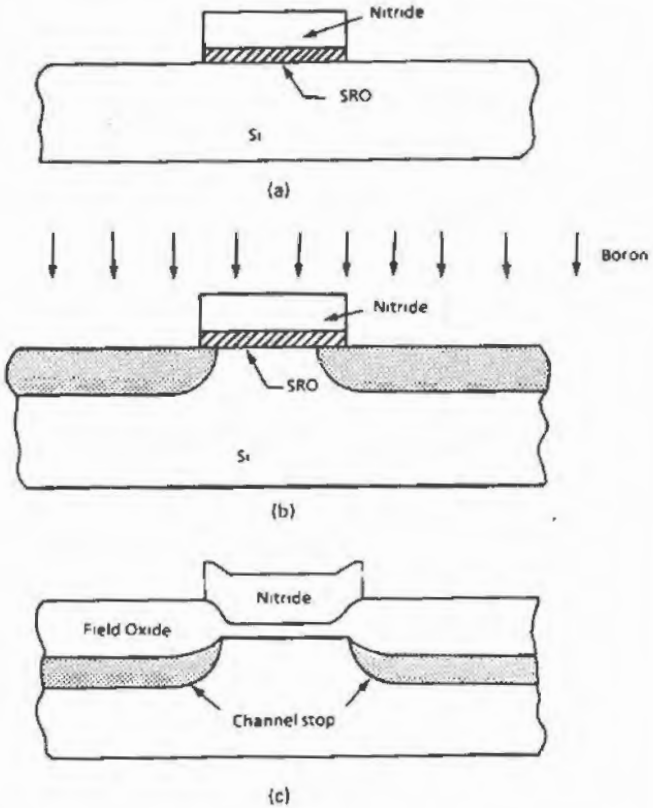


Figure 7.7 Process steps for LOCOS isolation with self-aligned channel stop.

Applying LOCOS to VLSI isolation is severely limited by field oxide encroachment and lateral diffusion of channel-stop dopants into the active device area. Oxidant diffuses through the buffer oxide under the nitride, hence inducing oxidation near the nitride edges. This lateral oxidation encroachment makes the edge of the LOCOS oxide resemble a bird's beak. The impurity atoms in the channel stop region also diffuse laterally during LOCOS and other subsequent high temperature process steps. Figure 7.8 is an SEM micrograph showing the bird's beak structure due to lateral oxidation. As shown in Fig. 7.9, the lateral oxidation rate is a function of the lateral distance from the nitride edge and depends on buffer oxide thickness. Both oxidation encroachment and impurity lateral diffusion overtake the active area near the channel edges. Hence, the physical channel width W_p

Figure 7.8 SEM micrograph showing the cross-section of the bird's beak due to oxide encroachment (Ref. 9. © 1981 IEEE).

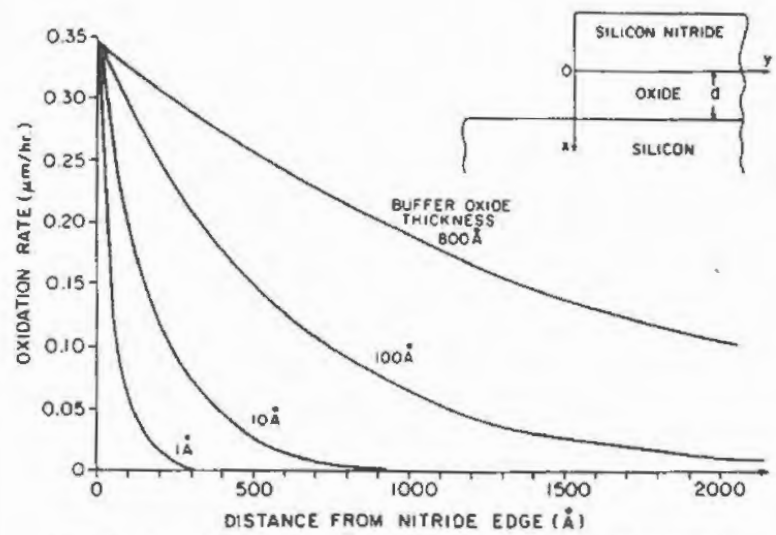


Figure 7.9 Lateral oxidation rate as a function of the lateral distance from nitride edge for various buffered oxide thicknesses used in the LOCOS process (Ref. 9, © 1981 IEEE).

Greenthread Ex. 2059, p. 25 of 47
Semiconductor v. Greenthread
IPR2023-01242

becomes less than the designed channel width W_d. The difference ($W_d - W_p$) becomes a significant portion of W_d when devices are scaled down for VLSI applications. Furthermore, the electrical channel width (W_e) is even less than W_p because the 2-D narrow-channel effect applies to LOCOS as well as other non-planar isolation schemes, such as window isolation. Both physical and electrical channel-narrowing effects increase the FET threshold voltage and reduce its current driving capability. To maintain the desired channel width, the corresponding mask dimension must be drawn oversized, resulting in wasted layout area. Other considerations are nitride-induced crystal defects and step coverage on non-planar surface.

Several methods have been tried to improve LOCOS isolation in terms of reducing the bird's beak. Sealed-Interface Local Oxidation (SILO)[9] uses a thin (<300 Å) nitride film directly on Si to seal the interface, shutting off lateral oxidant diffusion during local oxidation. As predicted in Fig. 7.9, the lateral oxidation rate decreases sharply as the buffer oxide thickness is reduced. Consequently, a steeper step angle and reduced bird's beak[9] can be obtained (Fig. 7.10). A thicker nitride layer can reduce the bird's beak even further, but at the expense of generating more stress, more crystal defects, and higher leakage. To achieve low stress and a small bird's beak, a sandwich masking layer (Fig. 7.11) consisting of a thin nitride (100–200 Å), an LPCVD (Low-Pressure Chemical Vapor Deposited) oxide (300 Å) and a thicker nitride (1000 Å) has shown basically defect-free field oxide with encroachment of only 25% of the oxide thickness (Fig. 7.12). A conventional LOCOS generally has lateral encroachment of ≈80–100%.

Another attempt to improve LOCOS is the use of the SideWAll Masked Isolation (SWAMI) process,[10] which offers basically defect-free and near-zero bird's beak, but at the expense of process complexity. As shown in Fig. 7.13, the process starts with forming and patterning a composite masking layer with stress relief oxide (SRO I), and nitride (Nitride I) on active regions. The Si substrate is then etched with a sloped sidewall and channel stop is implanted (Fig. 7.13a). A second composite nitride/oxide layer is then formed

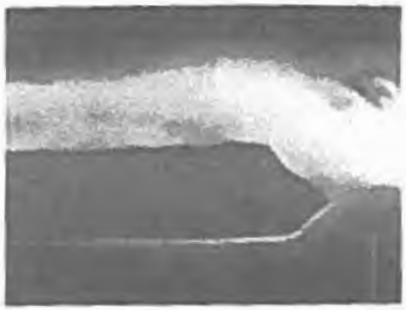


Figure 7.10 An SEM micrograph showing the cross-section of reduced bird's beak in a SILO structure (Ref. 9, © 1981 IEEE).

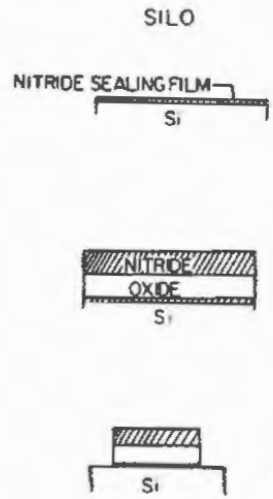


Figure 7.11 Process steps for a sandwiched SILO structure.

and an LPCVD oxide is deposited (Fig. 7.13b). After RIE etching of the CVD oxide, an oxide sidewall spacer is left (Fig. 7.13c). This sidewall spacer is then used as a mask to etch the nitride II, then the mask is stripped (Fig. 7.13d). LOCOS is then performed with the Si sidewall protected from lateral oxidation encroachment by the thin nitride spacer (Fig. 7.13e). All masking materials are then stripped and gate oxide is grown (Fig. 7.13f). The process employs two nitride masks, anisotropic oxide and nitride etching. The slope of the sidewall (≈54°) must be controlled during Si etching. A near isoplanar structure is achieved as shown in Fig. 7.14 and good electrical characteristics have been obtained.[10]

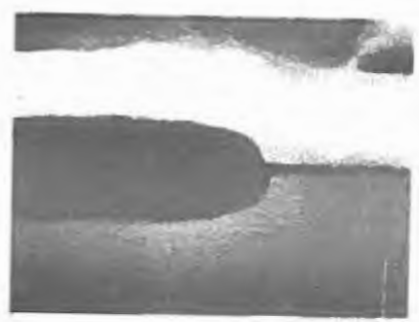


Figure 7.12 SEM micrograph showing the cross-section of minimum encroachment using the sandwich structure shown in Fig. 7.11 (Ref. 9, © 1981, IEEE).

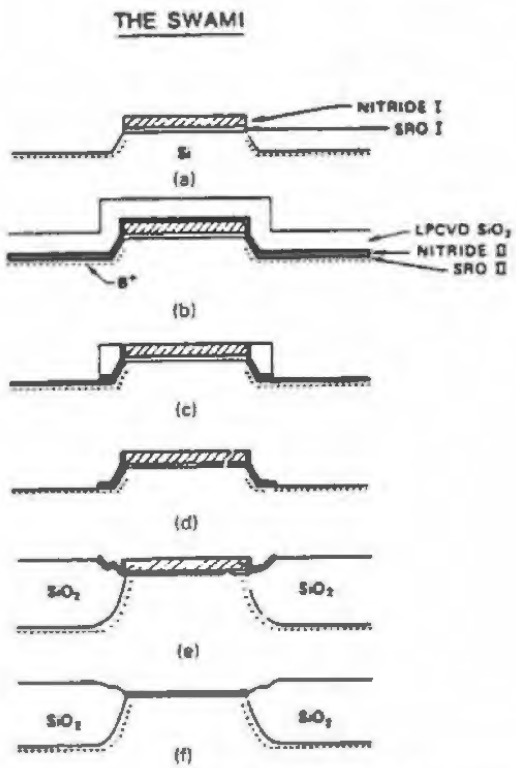


Figure 7.13 The major process steps of SWAMI (SideWAll Masked Isolation) (Ref. 10, © 1982 IEEE).

7.2.3 Isoplanar Field Oxide

Another isolation approach involves etching a shallow (several tenths of a micron) step in a silicon substrate and filling it with a depositing oxide. This isolation has fully recessed isoplanar configuration and is bird's beak-free. It was introduced as BOX for Buried OXide.[11] A liftoff technique using aluminum[11] or molybdenum[12] as a stencil was employed for this approach. As shown in Fig. 7.15, SiO_2 was plasma-deposited at 300°C for the field oxide. A wet etching before liftoff was done to remove oxide at the sidewall of the Si groove. Because the oxide at the sidewall was highly stressed, its etch rate was much higher, allowing preferential etching. The oxide on the metal stencil was then lifted off by resolving the metal. To fill

Figure 7.14 SEM micrograph showing the cross-section of a SWAMI structure (Ref. 10, © 1982 IEEE).

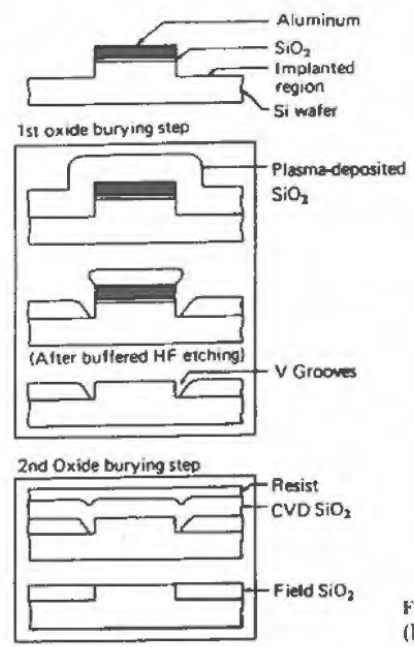


Figure 7.15 Process steps in BOX (Buried OXide) technology (Ref. 11, © 1981 IEEE).

Figure 7.16 SEM micrograph showing the cross-section of isoplanar PHOTOX structure (Ref. 13, © 1982 IEEE).

in the V-grooves at the step, another oxide was deposited and RIE etched to form a planar surface.

A similar technique using photoresist to lift off low-temperature (100°C) photo-CVD oxide (Photox) was also proposed for field isolation.[13] The idea again is to provide a fully isoplanar structure as shown in Fig. 7.16. The

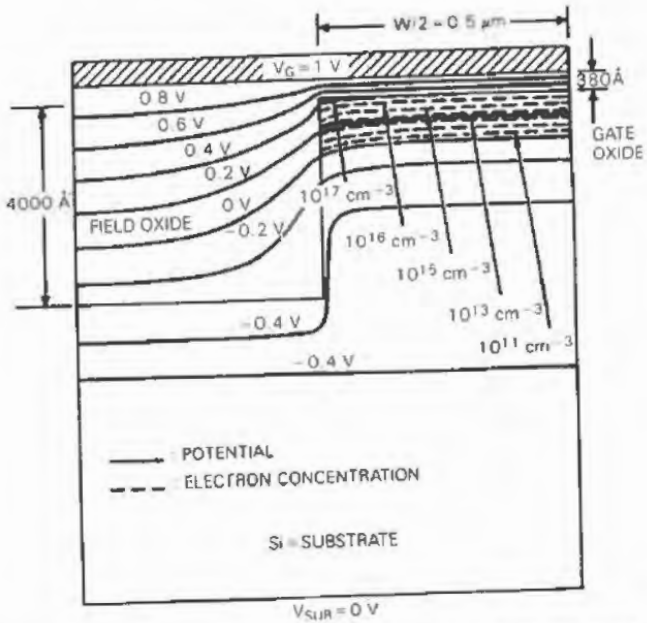


Figure 7.17 Two-dimensional contours of equipotentials and electron concentrations for a fully-recessed isoplanar MOS structure in channel width direction (Ref. 13, © 1982 IEEE).

planar isolation with fully-recessed field oxide not only eliminates physical encroachment and surface topology, but also avoids the electrical 2-D fringing field effect illustrated in Fig. 7.6. Fig. 7.17 shows the 2-D contours of potential lines and carrier concentrations for the planar isolation structure. Notice that both potentials and electron concentrations are uniform across the entire channel width, indicating negligible electrical channel-narrowing effect.

Despite all of the advantages of the isoplanar structure, the associated liftoff process is, so far, doubtful for production. The use of preferential etching can facilitate liftoff, but even then the process may be too complicated for manufacturing VLSI circuits. Other processes, such as trench, are being developed for realizing isoplanar isolation. Section 7.4 will discuss these techniques.

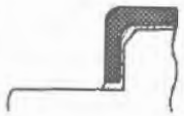
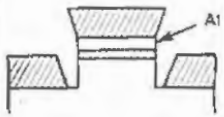
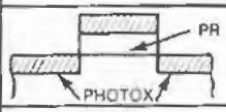
Table 7.1 summarizes the various MOS isolation techniques discussed. Key process steps, distinct features, and schematics are included. For the present MOS IC fabrication, standard LOCOS or improved LOCOS such as SILO, is still the dominant isolation technique, even though scaling isolation spacing for VLSI becomes severe.

7.3 ISOLATION IN CMOS

Layout density has been a major problem in bulk CMOS because both *n*- and *p*-channel devices must be made on the same chip and the two devices must be properly isolated. The two types of devices must be separated physically by at least one well, which is normally formed by diffusion. Because diffusion is an isotropic process, a large layout space is required to accommodate lateral diffusion during well formation. Another problem is the implementation of channel stop(s) for CMOS devices. In NMOS technology, the *p*-type channel stop can be formed easily under the LOCOS oxide to raise field threshold voltage well above the 5 V power supply. In the case of CMOS, a *p*-type channel stop is still necessary to provide isolation among *n*-channel transistors. Moreover, it must be spatially restricted to the *n*-channel area only. Similar requirements apply for *n*-type channel stops needed in the *p*-channel area. These additional constraints require extra layout spacing between the *n*- and *p*-channel areas unless some sort of self-aligned structure can be realized. Even worse, the channel-stop doping concentration at the edge of the well drops substantially due to lateral diffusion and the compensation between acceptor and donor impurity atoms. As an example, Fig. 7.18 shows this effect through 2-D simulation for the twin-tub process.[14] The solid curves show the reduction in net impurity concentration at the well boundary when both boron and phosphorus implants are used. The dashed line represents a case in which only boron implant was used. The doping reduction leads to isolation problems associated with the tight spacing between an *n*- and *p*-channel device.[14]

Greenthread Ex. 2059, p. 28 of 47
Semiconductor v. Greenthread
IPR2023-01242

TABLE 7.1 Comparison of Various MOS Field Isolation Technologies

TECHNIQUE		KEY PROCESS STEPS	DISTINCTIVE FEATURES
LOCOS	NITRIDE ON OXIDE		BIRD'S BEAK SEMI-RECESSED
IMPROVED LOCOS	ELIMINATING BUFFER OXIDE		• REDUCED BIRD'S BEAK LENGTH • DIFFICULT TO CONTROL DEFECT FORMATION • SURFACE NOT COMPLETELY FLAT (PLANAR)
	USING NITRIDE PROTECTION OF SIDEWALL		• EXCELLENT SURFACE PLANARITY • VIRTUALLY NO BIRD'S BEAK • DEFECT FORMATION (?) • COMPLICATED
WINDOW ISOLATION			• VIRTUALLY NO BIRD'S BEAK • SIMPLICITY OF PROCESS • SURFACE NOT FLAT (PLANAR) • DIFFICULT TO APPLY TO SUBMICRON ISOLATION • STEP COVERAGE
GROOVE	BURIED OXIDE		• NO BIRD'S BEAK • PERFECTLY FLAT (PLANAR) SURFACE • NO IMPURITY REDISTRIBUTION • REQUIRE Al MASK AND PLANARIZATION BY RIE
	PHOTOX		• NO BIRD'S BEAK • SUPER-PLANAR • NO Al OR RIE • NOT MATURE
TRENCH			• EFFECTIVE FOR BIPOLAR OR CMOS WELL ISOLATION • INCREASED PARASITIC INTERCONNECT CAPACITANCE • COMPLICATED • ≥2μm WIDE?

Si_3N_4 POLY-Si
SiO_2 Si

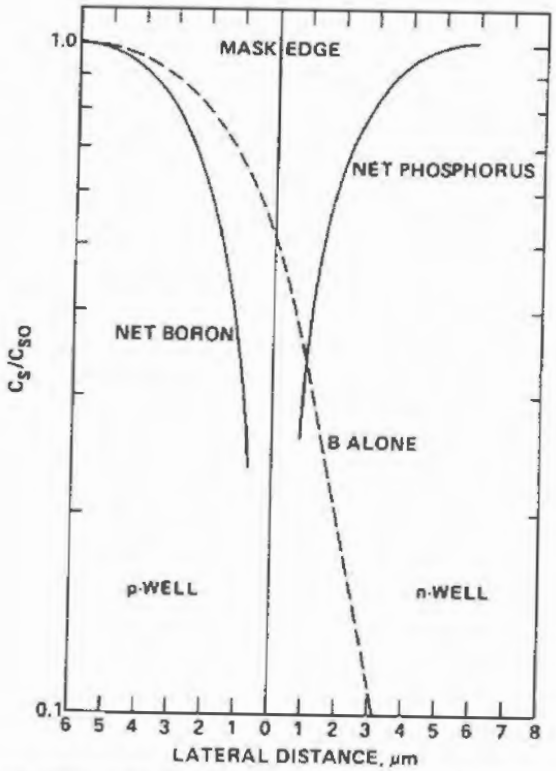


Figure 7.18 Impurity surface concentration near the border of the two tubs in twin-tub CMOS (Ref. 14, © 1980 IEEE).

In a CMOS circuit, proper isolation must be provided among opposite-type devices as well as between similar devices.[15] The following sections first describe isolation among similar devices, followed by the discussion of isolation among opposite-type devices, then the enhancement of isolation by using retrograde-well technologies and finally, vertical isolation between diffusion in the well and the substrate.

7.3.1 Isolation Among Similar Transistors

This section will discuss isolation between two *n*-channel transistors and two *p*-channel transistors. These similar transistors must be isolated properly to avoid leakage current occurring in parasitic FETs. Parasitic FETs are

Greenthread Ex. 2059, p. 29 of 47
Semiconductor v. Greenthread
IPR2023-01242

formed by polysilicon or aluminum interconnect lines on relatively thick field oxide and adjacent *pn* junctions.

Test devices for isolation between two *n*-channel active transistors were described in Sec. 7.1.1. Problems associated with the conventional definition of field threshold voltage were discussed in Sec. 7.1.2. Instead of defining threshold at 1 μA, a stringent way to test field isolation is to measure the field threshold voltage (V_{TF}), defined as the gate voltage that generates a small and acceptable drain current at drain voltage equal to power supply voltage, i.e., currently a 5 V V_{DD}. The V_{TF} values measured at drain currents of 10 pA, 1 nA and 0.1 μA in a 20 μm wide FET are shown in Fig. 7.19.[16] The test structures used for this measurement are shown in Figs. 7.2 and 7.3. The LOCOS field oxide thicknesses for polysilicon and aluminum gate FETs are about 0.5 and 1.0 μm, respectively in this example. A boron field implant of 4×10^{13} cm^{-2} at 25 KeV is used for the channel stop formation before the LOCOS oxidation. The higher current (0.1–1.0 μA) corresponds to a threshold similar to the threshold defined at the onset of strong inversion. This definition is suitable for the active MOS transistors but not particularly useful for parasitic FETs in which a very low level of leakage current (10 pA – 1 nA) is a serious concern for maintaining isolation. Notice that the V_{TF}s are higher for the aluminum gate because of the associated thicker field oxide. The thicker oxide also gives poorer subthreshold slope, causing V_{TF} curves to separate further. The V_{TF} decreases as the channel length (n^+-to-n^+ separation) is reduced in a way that is analogous to the short-channel effect commonly observed in active MOSFETs (see Chapter 2, Sec. 2.5). This effect should be taken into account when setting n^+-to-n^+ layout design rules. For the example shown in Fig. 7.19, V_{TF} values are 10 V or above even at 1 μm separation between two *n*-channel active devices.

Parasitic FETs among *p*-channel active devices are identical to those shown in Figs. 7.2 and 7.3 except that the doping types are reversed. The symmetric parasitic pMOSFETs are, in general, harder to turn on because the net oxide charge density (Q_{fc} term) is commonly positive, making the V_{TF} even more negative. As shown in Fig. 7.20,[16] the absolute V_{TF} values are well above 5 V. The test devices used here are fabricated from the same CMOS process which made the devices used for obtaining data in Fig. 7.19. The channel stop is formed by a shallow phosphorus implant inside the *n*-well. The very high V_{TF} values in the polysilicon gate devices are not always measurable due to dielectric breakdown in the thin oxide region between the polysilicon gate and the source. Notice that the aluminum gate devices exhibit greater short-channel effect, due to thicker oxide.

Leakage can also be observed in the parasitic FETs if punchthrough between the two n^+ (or two p^+) regions occurs. Due to the relatively high doping concentration in the field region and shallow n^+ or p^+ diffusion regions, punchthrough usually does not take place unless the separation between the two diffusion regions is scaled below 1 μm. Breakdown is normally

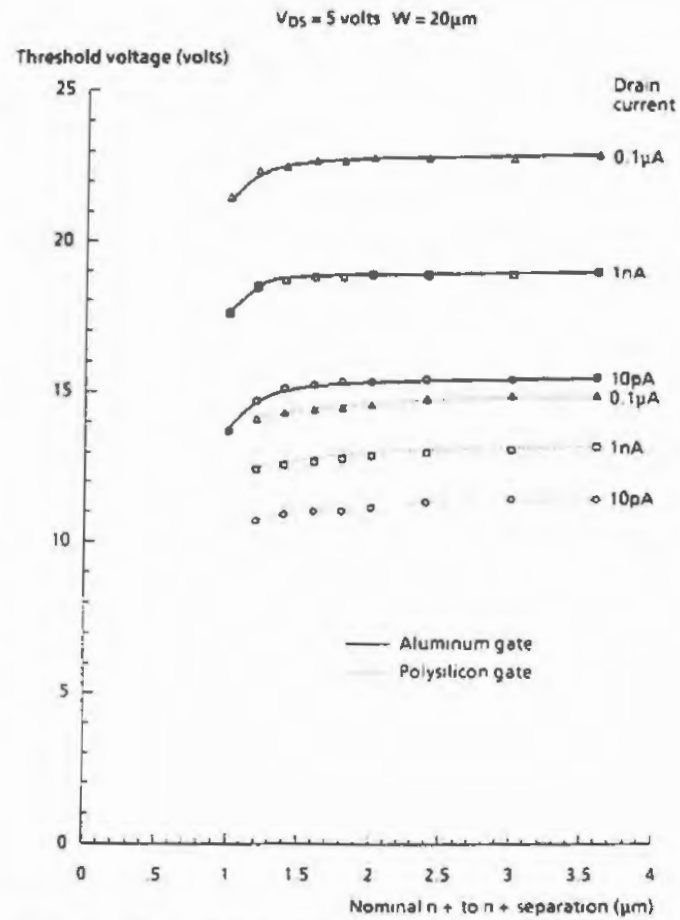


Figure 7.19 *n*-Channel field threshold voltages for symmetrical parasitic FETs (Ref. 16).

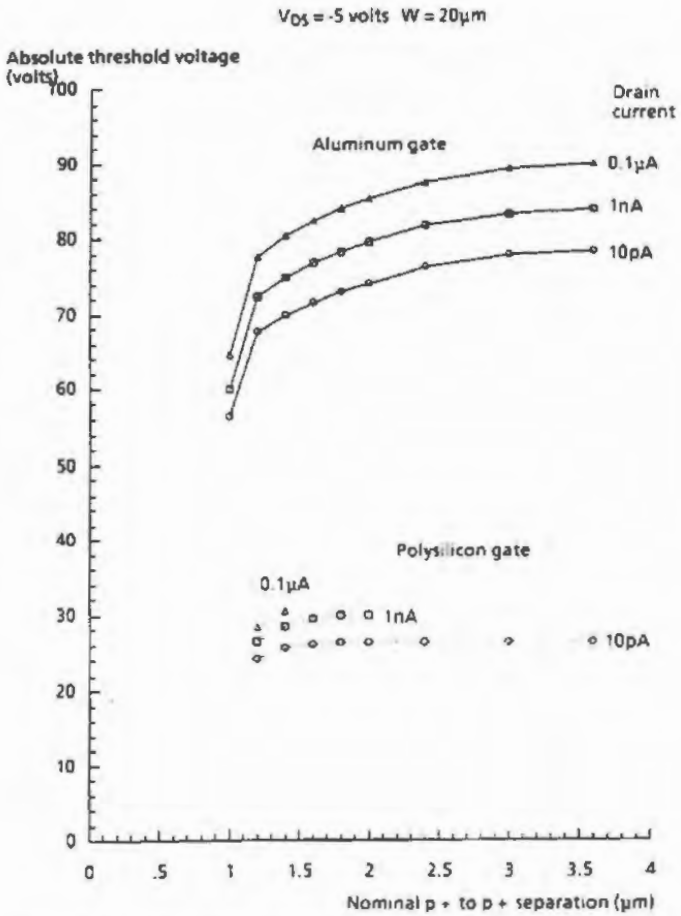


Figure 7.20 *p*-Channel field threshold voltages for symmetrical parasitic FETs (Ref. 16).

limited by avalanche breakdown which, in general, occurs at voltages well above 5 V.

7.3.2 Isolation Among Opposite-Type FETs

In addition to physical separation and the precise definition of channel stops, electrical isolation between *n*-channel and *p*-channel active devices is just as important. Inadequate isolation leads to leakage between opposite-type devices. This leakage corresponds to the subthreshold leakage measured in parasitic FETs.

Test structures. Fig. 7.21 shows two parasitic FETs formed at the edge of an *n*-well by aluminum or polysilicon lines. The n^+ and the *n*-well form source and drain for the parasitic *n*MOSFET while the p^+ and *p*-substrate serve as source and drain for the *p*MOSFET. Here, the source and drain are not symmetrical and the *n*-well and *p*-substrate are used as drains because a deeper drain junction and a highly conductive source always produce more subthreshold current, thereby representing a worst case condition for isolation. Similar to the test structures described in Sec. 7.3.1, the two structures have two differences between them. The field oxide in the aluminum gate structure is thicker than that in the polysilicon gate. Also, the

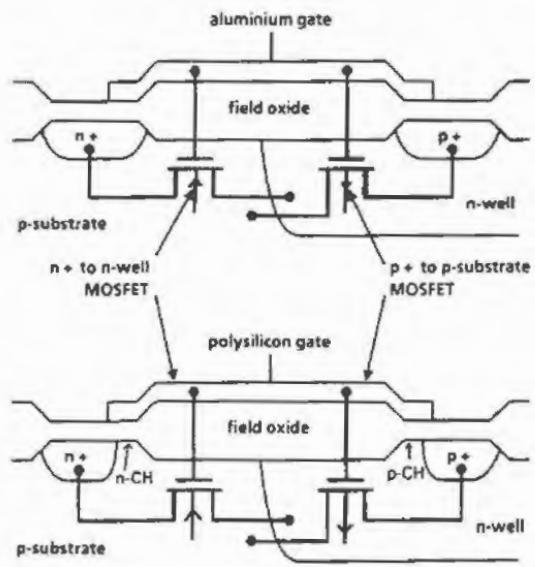


Figure 7.21 Asymmetric parasitic FETs formed at *n*-well edge.

n^+ and p^+ in the polysilicon gate structure are self-aligned to the polysilicon gate instead of butting against the LOCOS edges. This test structure simulates the real CMOS inverter layout shown in Fig. 7.22,[17] although the actual p^+ and n^+ are not along the same direction of the polysilicon gate. When the gate is biased at 5 V, the n-channel surface under the polysilicon gate (indicated by n-CH) is inverted to n-type but is isolated from the n-well by the p-substrate as long as the field region in the p-substrate is not inverted. Similarly, as the gate is turned to 0 V, the p-channel surface under the polysilicon gate (indicated by p-CH) is inverted to p-type but is isolated to the p-substrate unless the field region inside the n-well is inverted.

The present test structure serves as a good vehicle to test the field inversion problem under the polysilicon lines. However, the true source and drain diffusions in the polysilicon gate case are kept further apart than in the aluminum gate case (Fig. 7.21). This instance tends to inhibit punchthrough and hence does not test the worst punchthrough problem.

As n^+-to-n-well or p^+-to-p-substrate spacing is reduced, the magnitude of the field threshold voltage for either nMOS or pMOS parasitic FET decreases, again similarly to the short-channel effect commonly observed in an active MOSFET. This threshold roll-off effect in parasitic asymmetric FETs is very serious because of the associated deep drain (n-well or p-substrate). The deep drain is also more likely to cause drain-to-source punchthrough, which is another failure mode in CMOS isolation. Moreover, the two failure modes, field inversion and punchthrough, are coupled through 2-D effects.[17] For worst-case considerations, field inversion must be measured at maximum drain bias and punchthrough must be measured at highest gate voltage.

Modelling. Two-dimensional process and device modelling is carried out using the SUPRA and GEMINI programs discussed in Chapter 3. The input parameters for the SUPRA program are adjusted to achieve agreement with the measured profiles in the vertical direction. The GEMINI program

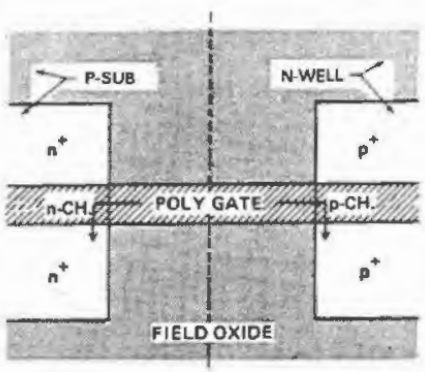


Figure 7.22 Layout top view in CMOS isolation region between n- and p-channel active devices (Ref. 17, © 1986 IEEE).

is basically a Poisson's equation solver, but is quite adequate for isolation modelling in which only very low levels of current are concerned.

Fig. 7.23(a) and (b) show the 2-D impurity distribution and the corresponding potential contours for the n^+-to-n-well aluminum gate asymmetric parasitic FET.[18] The device is formed using a conventional diffused n-well technology. The nominal separation of n^+ diffusion from the n-well (that

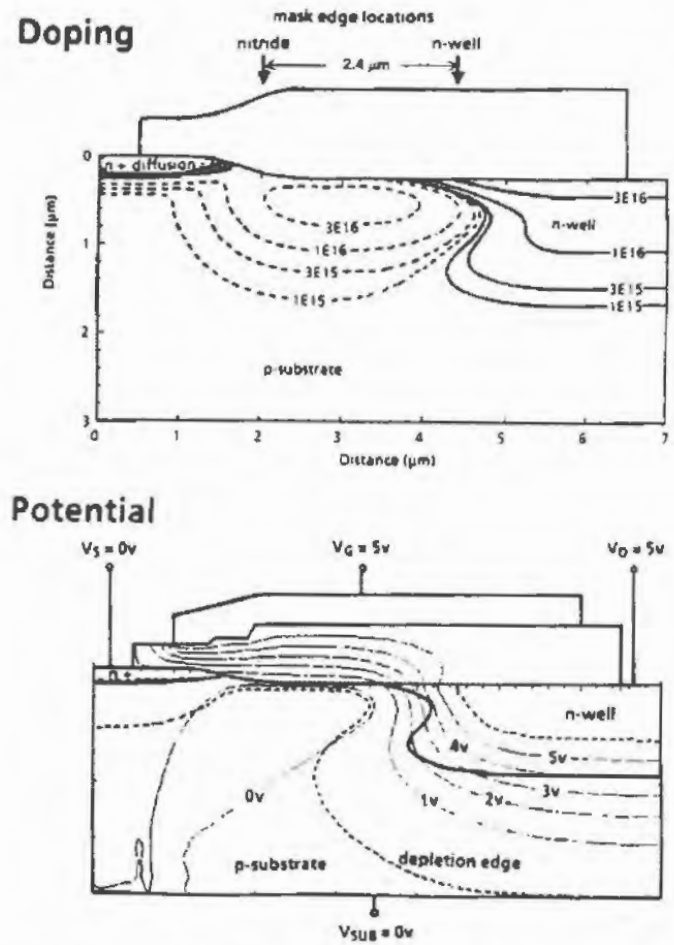


Figure 7.23 2-D simulated impurity doping and potential contours in the n^+-to-n-well parasitic MOSFET (Ref. 18, © 1987 IEEE).

is, the separation of the nitride LOCOS mask and the n-well implant resist mask) is 2.4 μm; the locations of the mask edges are indicated on Fig. 7.23(a).[18] Notice the fairly high p-type field doping level between the n^+ diffusion and the well.

Fig. 7.23(b) illustrates a solution to Poisson's equation for the structure depicted in Fig. 7.23(a). The n-well and parasitic gate are biased to 5 V with respect to the p-substrate and n^+ source; this instance represents the worst case for leakage with a 5 V supply. In this case, the parasitic FET is off as indicated by the 0 V surface potential in the middle of the channel. The suppression of lateral spread of the well-substrate depletion region by the p-type field dopants can be clearly seen.

Figs. 7.24(a) and (b) show simulated doping and potential contours, respectively for a p^+-to-p-substrate aluminum gate parasitic FET. In this case, a device formed using the retrograde n-well technology is taken as an example. The nominal separation of p^+ diffusion from the p-substrate (that is, the separation of the nitride LOCOS mask and the n-well implant resist mask) is 1.2 μm; the locations of the mask edges are indicated on Fig. 7.24(a).[18] Note that the retrograde well is shallower under the field oxide than under the active area,[19,20] and also that the elimination of an n-well drive-in again limits lateral diffusion of the well dopants. The solution to Poisson's equation illustrated in Fig. 7.24(b) is obtained assuming that the p-substrate and parasitic gate are biased to -5 V with respect to the n-well and p^+ source, again representing the worst case for leakage. The potential contours at 0 V or below are all well inside the oxide, indicating that the Si surface underneath is at positive voltage. This indication means that sufficient voltage margin exists before inversion occurs. Also, notice that the depletion region associated with the drain junction is mostly extended to the p-substrate rather than the n-well, which is much more heavily doped than the p-substrate. Because the n-well serves as the channel region in this structure and only a small portion is taken by the drain depletion region, lateral punchthrough between p^+-to-p-substrate is less likely to happen.

n^+-to-n-well parasitic FETs. Fig. 7.25(a) shows field inversion voltages defined by drain currents of 10 pA, 1 nA, and 0.1 μA, measured at a 5 V drain voltage for both aluminum and polysilicon gate n^+-to-n-well parasitic FETs.[18] The threshold lowering effect below 2 μm nominal n^+ to n-well separation is clear. The n-well in these devices acts like a deep drain, and the associated depletion region extends into the substrate toward the n^+ source. This extension causes short-channel threshold lowering and drain-induced barrier lowering similar to normal active devices.

The thresholds of the thinner gate oxide of the polysilicon gate parasitic MOS transistors are lower than those of the aluminum gate devices, and the curves corresponding to the different subthreshold drain currents are closer together due to the improved subthreshold swing. As described in Chapter

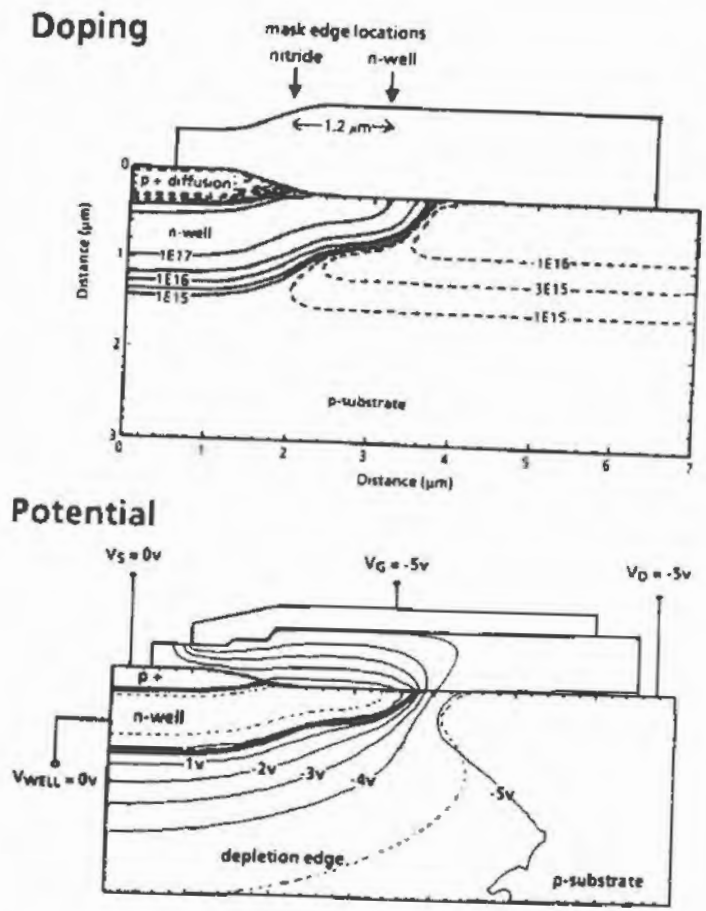


Figure 7.24 2-D simulated impurity doping and potential contours in the p^+-to-p-substrate parasitic FET (Ref. 18, © 1987 IEEE)

2, the subthreshold swing (S_t) decreases as the gate oxide is thinned. Moreover, threshold lowering is less for polysilicon gate FETs, which is again due to the thinner gate oxide. At the smallest n^+-to-n-well separations (1–1.5 μm), the larger threshold reduction in the aluminum gate FETs causes their thresholds to be lower than those of polysilicon gate devices. Consequently, it is the aluminum gate structures which ultimately limit the minimum separation. Furthermore, the oxide quality and thickness control of the alu-

Greenthread Ex. 2059, p. 33 of 47
Semiconductor v. Greenthread
IPR2023-01242

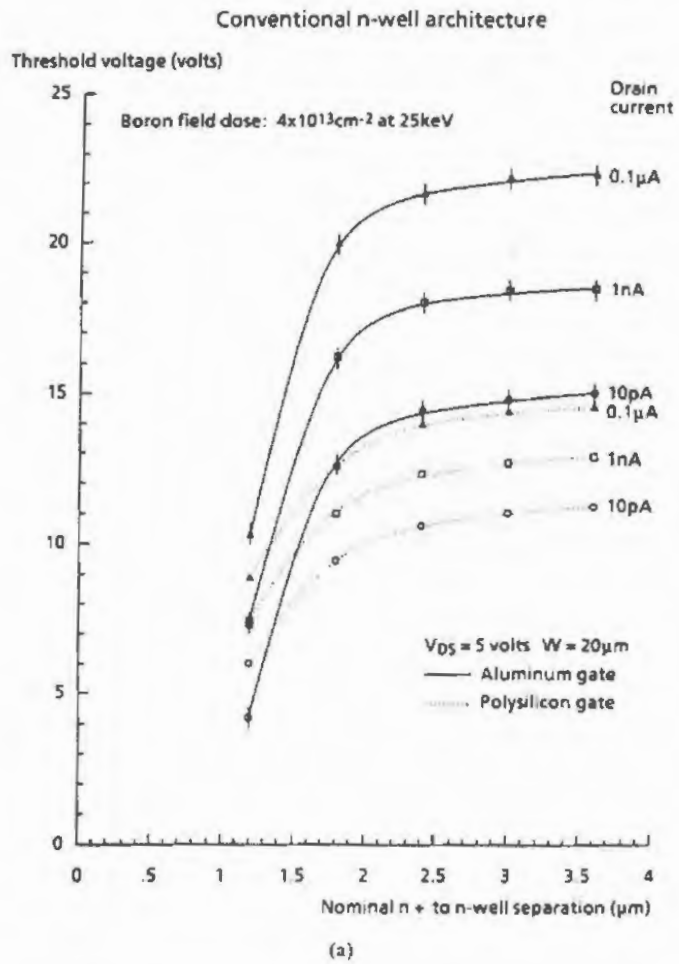


Figure 7.25 (a) Threshold voltages; (b) breakdown voltages for the asymmetric n^+-to-n-well parasitic FETs (Ref. 18. © 1987 IEEE).

minum gate oxide often make its threshold less reproducible from run to run, and so more safety margin must be given.

The breakdown voltages of n^+-to-n-well parasitic FETs also indicate that the aluminum gate FETs ultimately determine minimum separation. Fig. 7.25(b) shows that at 1.2 μm n^+ to n-well spacing, the polysilicon gate

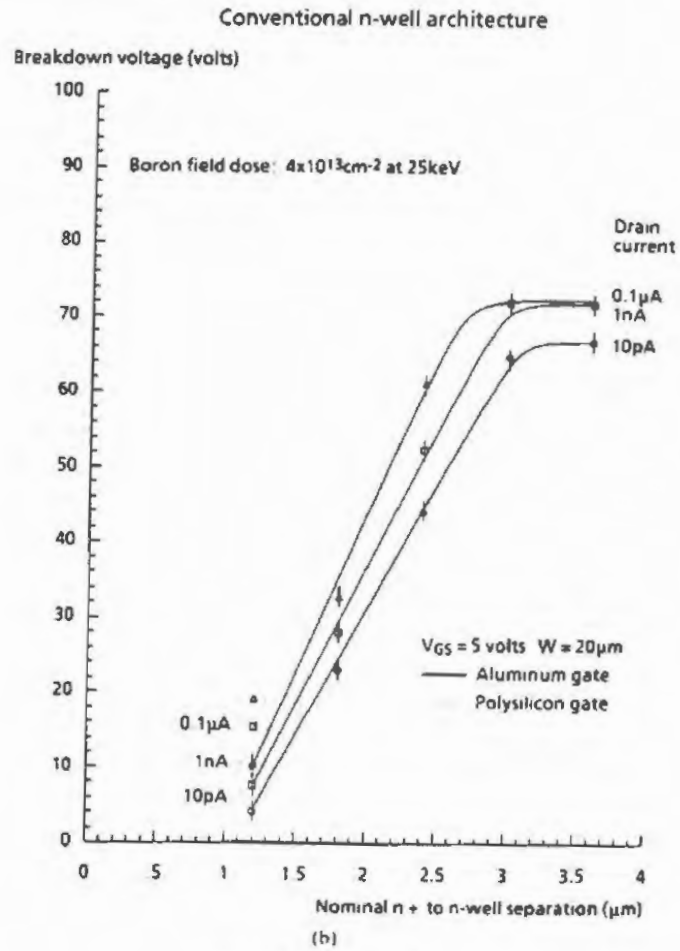


Figure 7.25 *Continued.*

punchthrough voltages measured at 10 pA, 1 nA, and 0.1 μA, with V_{GS} = 5 V are all about twice the corresponding values for aluminum gate devices. At larger separations, breakdown in the polysilicon gate devices is limited by the dielectric strength of the thin gate oxide, where the polysilicon overlaps the active area in the n-well (indicated as p-CH in Fig. 7.21(b)), and so punchthrough or avalanche breakdown cannot be measured. As compared with the symmetric parasitic FETs previously described, punchthrough in asymmetric FETs is much more severe because of the deeper drains in these

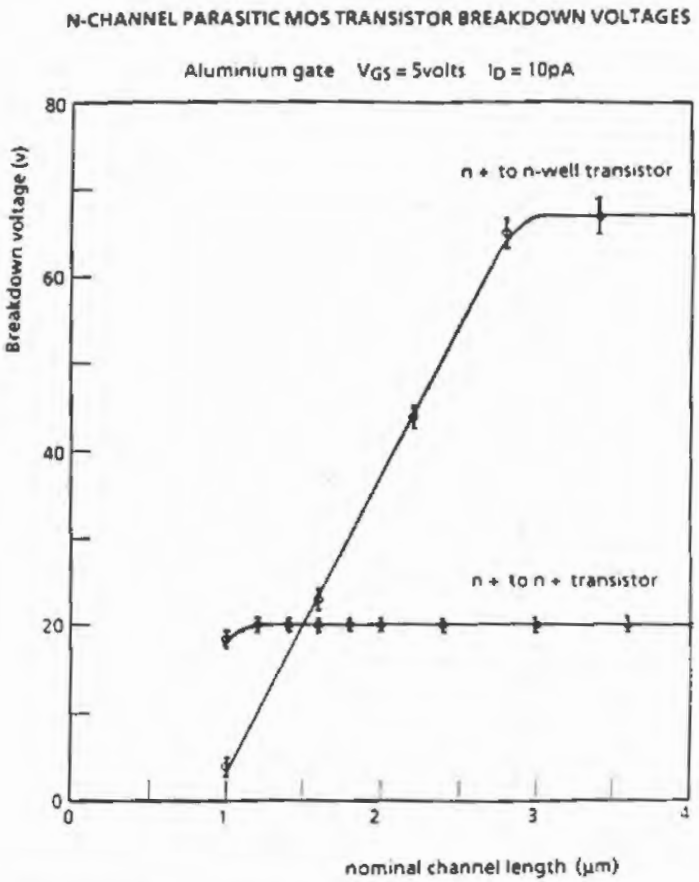


Figure 7.26 Avalanche and punchthrough breakdown voltages for the symmetric (n^+-to-n^+) and asymmetric (n^+-to-n-well) parasitic nMOSFETs (Ref. 21, © 1987 IEEE).

devices. Fig. 7.26 shows an obvious difference by comparing an n^+-to-n^+ parasitic FET to an n^+-to-n-well device.[21]

p^+-to-p-substrate parasitic FETs. Field inversion voltages for p^+-to-p-substrate parasitic FETs are shown in Fig. 7.27 for both aluminum and polysilicon gate devices.[18] Because nominal p^+-to-p-substrate separation

represents the separation of the nitride LOCOS mask and the n-well resist mask, and the actual p^+ in the polysilicon device is abutted to the polysilicon edge rather than the nitride edge, the actual p^+ to p-substrate separation is larger than its nominal value by the length of the poly-overlapping thin oxide. Threshold values can only be obtained at 0.6 μm or less, again due to the dielectric breakdown in the thin oxide region. At nominal zero separation in the polysilicon devices, the actual p^+ to p-substrate separation is about the

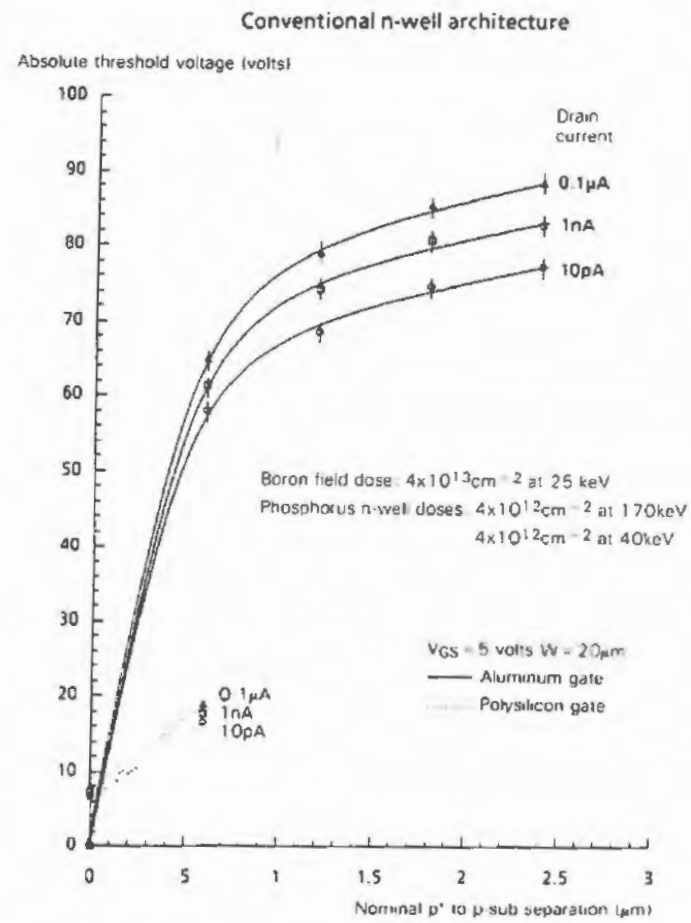


Figure 7.27 Threshold voltages for the asymmetric p^+-to-p-substrate parasitic FETs (Ref. 18, © 1987 IEEE).

Greenthread Ex. 2059, p. 35 of 47
Semiconductor v. Greenthread
IPR2023-01242

same as the length of the poly-overlapping thin oxide, that is, 1.2 μm. For the aluminum gate case in which the p^+ is abutted to the LOCOS edge, the devices fail to operate.

p^+ to p-substrate lateral punchthrough is not observed until the nominal separation is reduced to 0.6 μm. As previously described, this minimum punchthrough is mostly due to the fact that the majority of the drain depletion region is extended into the lightly doped p-substrate, which acts as the drain in this structure. Vertical punchthrough between p^+ to p-substrate can be a more serious problem especially when the n-well depth is scaled down for CMOS VLSI. In general, vertical punchthrough presents a problem if leakage current flows between a p^+ and p-substrate through an n-well. In normal CMOS operation, the n-well is at V_{dd}, the p-substrate is at V_{ss} and the p^+ is biased between V_{dd} and V_{ss}. However, during transient overshoot, the p^+ or p-substrate can become negatively biased, so the condition should be considered.

The problem can be illustrated by using an example of a conventional but shallow (about 1 μm) n-well with its profile shown in Fig. 7.28(a).[22] The

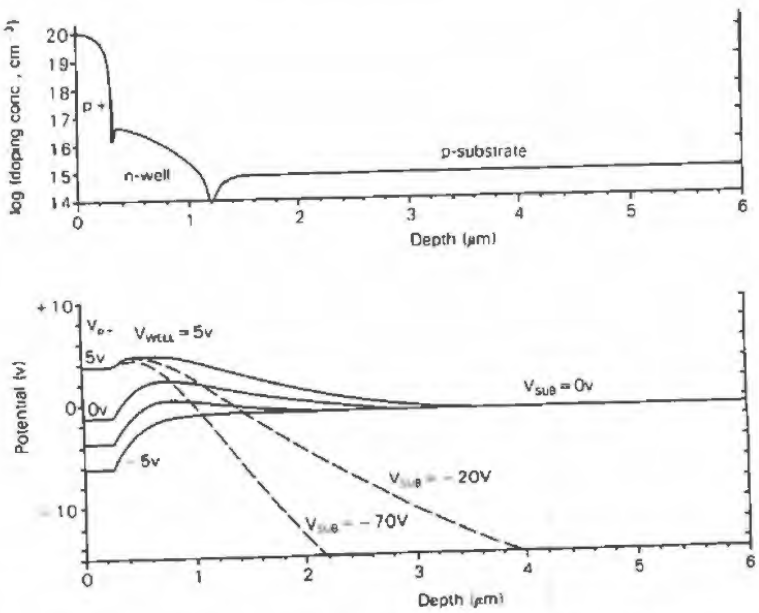


Figure 7.28 Vertical doping and potential profiles for a 1.2 μm CMOS structure. (a) Net doping, and (b) potential profiles, solid lines: p^+ voltage varies with p-substrate grounded, dashed lines: substrate bias changes with p^+ fixed at 5 V. The n-well is always at 5 V (Ref. 22, © 1987 IEEE).

corresponding potential distribution is shown in Fig. 7.28(b), in which the p^+ is biased between +5 V and −5 V while the n-well and p-substrate are biased at normal operation voltages shown as solid lines.[22] As the p^+ is swept from 5 V to 0 V, the p^+/n-well depletion region extends further into the well; however, the potential barrier remains even when the well is fully depleted. Consequently, no vertical punchthrough current is present during normal operational conditions. The barrier disappears when the p^+ potential reaches −5 V, thus punchthrough current flows from the substrate up to the p^+. Punchthrough current can also flow from the p^+ down to the substrate if it is negatively biased (shown as dashed lines). Because most of the well-to-substrate depletion region is into the lightly-doped substrate rather than into the well, the p^+-to-substrate potential barrier cannot be lowered significantly until the substrate is at a large negative bias (< −70 V). The situation is altered however if the p-substrate is replaced by a p-epi layer on a p^+ substrate. Because the heavily-doped substrate prevents the extension of the well-substrate depletion region down to the substrate, the depletion region is forced up into the well. As a result, the V_{SUB} needed to lower the potential barrier is reduced, thus making it easier to cause vertical punchthrough. As shown in Fig. 7.29,[22] the V_{SUB} magnitude at the breakdown of vertical isolation

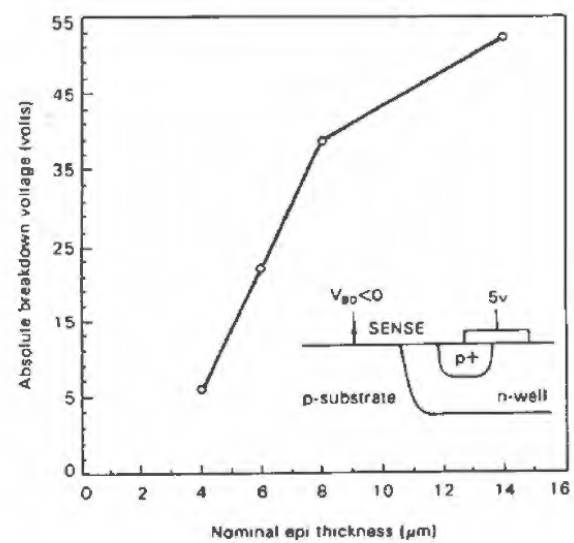


Figure 7.29 Vertical punchthrough voltage versus nominal (i.e., as-grown) epi thickness in a CMOS structure with 1.1 μm well depth. Sense current = 1 nA (Ref. 22, © 1987 IEEE).

Greenthread Ex. 2059, p. 36 of 47
Semiconductor v. Greenthread
IPR2023-01242

decreases as the epi layer is thinned. At 4 μm nominal (as-grown) epi thickness, isolation breaks down at V_{SUB} of about −5 V. Therefore, vertical isolation may limit the use of the thinnest epi layer.

7.3.3 Retrograde Well Technologies to Improve Isolation

As discussed in Chapter 5, retrograde wells are formed by high-energy implantation after the LOCOS oxidation cycle. Lateral diffusion is greatly reduced resulting in a significant saving of layout space. Because a well(s) is basically used to separate *n*- and *p*-channel transistors, isolation among these opposite-type devices are expected to improve. The retrograde *p*-well technology[23] was originally proposed to reduce the well depth and corresponding layout area. Although this technology has shown sufficiently high field threshold voltages and p^+-to-*p*-well punchthrough voltages, *n*-well is used as an example to compare the differences between a retrograde and conventional well technology in CMOS isolation.[16,18] Moreover, the comparison will be restricted to aluminum gate structures.

To make a fair comparison, conventional *n*-wells in this comparison are formed by minimum diffusion to achieve well depths in the active area as shallow as the retrograde wells made by high-energy implantation. The self-aligned feature of the *n*-well and field implants is preserved in a conventional well to achieve rigorous isolation at minimum n^+ to p^+ spacings. In other words, the conventional well technology is optimized to compete with the retrograde-well approach.

The important process steps in forming the *n*-wells for both technologies are shown in Fig. 7.30.[18,20] Both technologies start with the definition of active areas by nitride masking. Next, the *n*-well is aligned to it. Blanket boron field implant is used to form a *p*-type channel stop for *n*-channel transistors. The boron impurity concentration inside the *n*-well is compensated by the phosphorus implant used for the *n*-well formation. The major difference is that a conventional well is formed prior to LOCOS whereas the formation of the retrograde well is done after LOCOS. Because a retrograde well is formed by high-energy ions implanting through the field oxide, the well depth in the field region is only about half of the well depth in the active region. This depth is ideal in the sense of providing a shallow drain for the n^+-to-*n*-well parasitic FET, thereby improving isolation. Meanwhile the well in the active area is kept relatively deep so that the performance of the active transistors in the well are not significantly degraded. Furthermore, in the active area, the implanted impurity profile peaks near the bottom of the well, which is desirable for vertical isolation. At the same time, the profile peaks near the Si surface in the field region, which again is ideal for forming a channel stop and for preventing p^+-to-*p*-substrate lateral punchthrough.

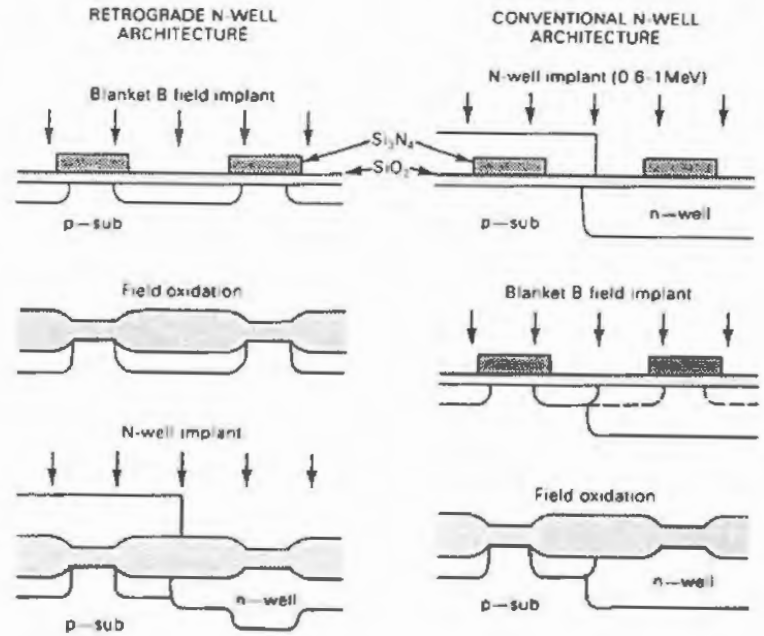
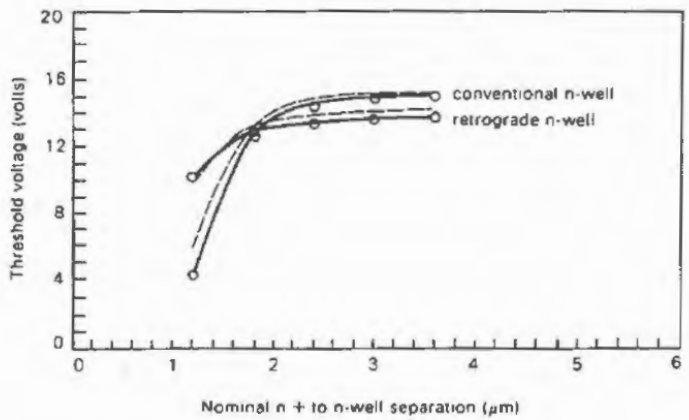


Figure 7.30 Process architectures for retrograde and conventional *n*-well CMOS structures (Ref. 16).

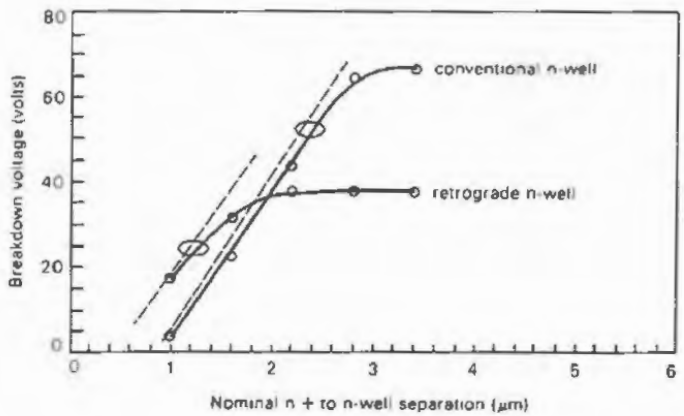
Figs. 7.31(a) and (b) show measured and modelled aluminum gate field inversion voltages and breakdown voltages, all defined at 10 pA drain current, for n^+ to *n*-well parasitic transistors with two different *n*-well technologies.[18] At large n^+ to *n*-well separations, avalanche breakdown at the well-substrate junction occurs before punchthrough takes place. The lower avalanche breakdown voltage observed with retrograde wells is attributed to higher well doping. As previously described, shallower retrograde wells should have less short-channel effects in terms of threshold lowering and drain-induced barrier lowering. These expectations are confirmed from the threshold results (Fig. 7.31(a)) and punchthrough data (Fig. 7.31(b)).

Fig. 7.32(a) shows field inversion voltages for lateral p^+-to-*p*-substrate parasitic devices.[18] The absolute values are much higher than those for n^+-to-*n*-well devices. These higher values are due in part to the higher field doping needed in the *n*-well to ensure adequate compensation of the boron. They are also partly due to the fixed charge at the Si–SiO_2 interface, which is likely to be positive and therefore increase the magnitude of the threshold voltage of *p*-channel transistors. It is also clear that short-channel effects

Greenthread Ex. 2059, p. 37 of 47
Semiconductor v. Greenthread
IPR2023-01242

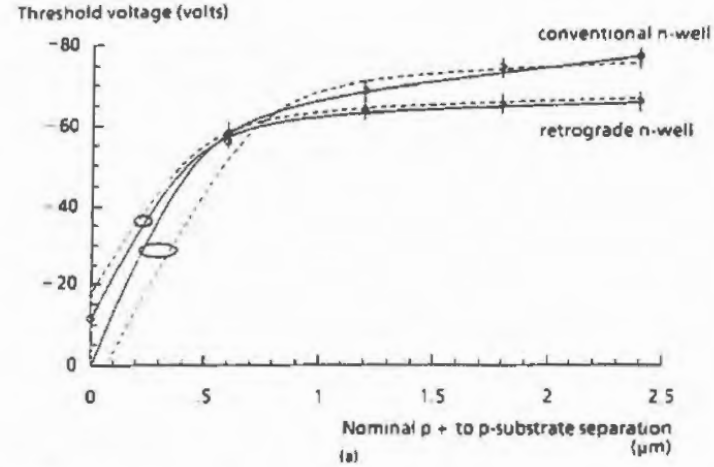


(a)

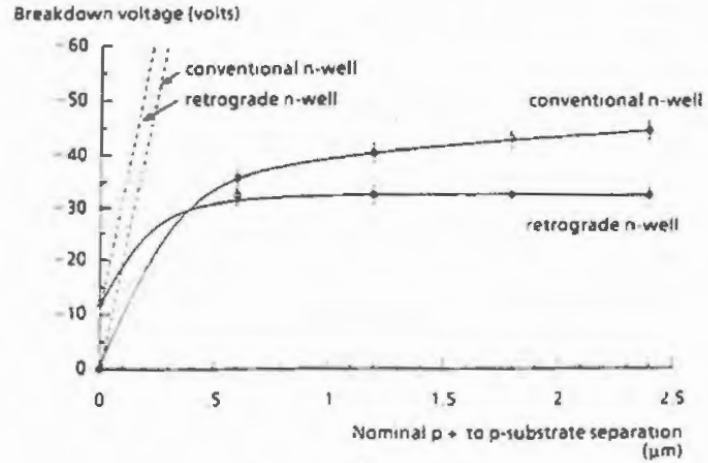


(b)

Figure 7.31 Comparison of retrograde and conventional well structures for n^+ to n-well parasitic FETs: (a) field threshold voltage; and (b) n^+ to n-well breakdown voltage. Solid lines: experimental data, dashed lines: modelled results, breakdown voltage is modelled by punchthrough only (Ref. 18, © 1987 IEEE).



(a)



(b)

Figure 7.32 Comparison of retrograde and conventional well structures for p^+ to p-substrate parasitic FETs: (a) field threshold voltage; and (b) p^+ to p-substrate breakdown voltage. Solid lines: experimental data, dashed lines: modelled results, breakdown voltage is modelled by punchthrough only (Ref. 18, © 1987 IEEE).

are much less severe in these p-channel parasitic devices, particularly for the retrograde structures; the thresholds do not fall until the channel length reaches about 0.5 μm. This effect is due to the higher channel (n-well) doping in these parasitic transistors, and to the drain (substrate) doping being lower than the channel (n-well) doping.

Fig. 7.32(b) shows breakdown voltages obtained with the same p^+ to p-substrate devices, and comparison with Figs. 7.31(a) and (b) again shows the superior performance of these structures.[18] In these types of parasitic FETs, the p-substrate, which is regarded as the drain, is not only deep but wraps around under the p^+ source (Fig. 7.21). Thus, the possibility of both lateral and vertical leakage paths between the p^+ diffusion and the p-substrate exists. The limiting values of breakdown voltage for retrograde n-well structures is set by well to substrate avalanche breakdown, as in the case of the n^+ to n-well transistors. However, with conventional wells, the limit is set by p^+-to-p-substrate vertical punchthrough breakdown. Thus, the maximum breakdown voltage shown on Fig. 7.32(b) for these devices is not the same

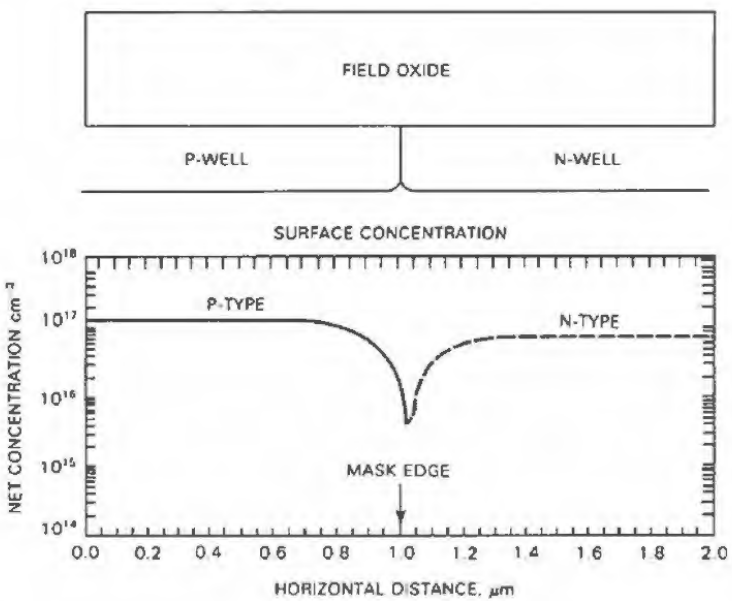
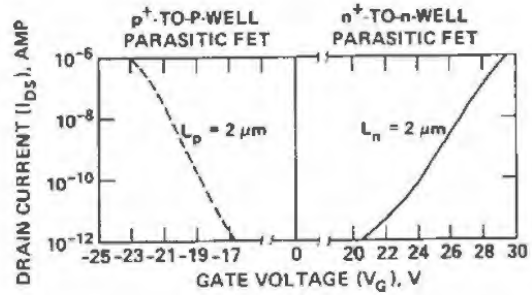


Figure 7.33 Surface impurity concentration at the boundary of the p- and n-retrograde wells in quadruple-well CMOS (Ref. 24, © 1984 IEEE).



*L_p: p^+-TO-p-WELL SPACING, L_n: n^+-TO-n-WELL SPACING

Figure 7.34 Subthreshold characteristics of asymmetric parasitic FETs in quadruple-well CMOS (Ref. 24, © 1984 IEEE).

as that shown on Fig. 7.31(b), even though the drain-channel junction is in fact the same in both cases.

The use of retrograde wells for both n- and p-wells has been applied to twin-tub CMOS technology and was called quadruple wells to emphasize the fact that for each type (n- or p-well), two wells were needed, one in the active area and another, shallower one in the field region. The two shallow wells serve as self-aligned channel stops with the n- and p-type impurity concentration tightly restricted on different sides of the well boundary. Fig. 7.33 shows that the reduction of net impurity concentration occurs only within a few tenths of a micron near the mask edge rather than a few microns, as observed in a conventional twin-tub structure (Fig. 7.18). Experimental data show that at a 4 μm n^+-to-p^+ separation, subthreshold leakage in the isolation region is <1 pA for gate voltages between −15 V and 20 V (Fig. 7.34).[24] Modelling results suggest that a 2 μm n^+-to-p^+ separation is sufficient for providing adequate isolation.[17]

In spite of all the advantages associated with the retrograde well in device isolation, specific concerns for this technology have been expressed. As discussed in Chapters 5 and 6, the active devices made inside a retrograde well generally suffer more from junction capacitance and body effect. The former leads to a decrease in speed of about 5–10% in a typical CMOS logic circuit. Another major concern is the availability of the high-energy implanter and the lack of understanding in using such a machine in a production environment.

7.4 NEW ISOLATION TECHNIQUES FOR CMOS

7.4.1 Trench Isolation

Trenches in the Si substrate have been proposed for CMOS isolation.[25,26] This isolation technique has several advantages. First, it eliminates oxidation encroachment (known as bird's beak) in LOCOS. The associated narrow-channel effect is then prevented. Second, trenches filled with oxide provide a planar surface, thus avoiding a step-coverage problem. Finally, when trenches of several microns depth are used in a CMOS chip, they provide a physical separation between opposite-type devices. Consequently, isolation can be achieved with very narrow n^+-to-p^+ spacing. Fig. 7.35(a) shows the schematic of an n-well CMOS structure with trenches which can be as narrow as 1 μm separating the n^+ and p^+. Notice that the trenches here are deeper than the n-well to effectively isolate the p-channel device.

Trench isolation also has several problems associated with it, making this technology difficult in production. First, fabrication is very complicated

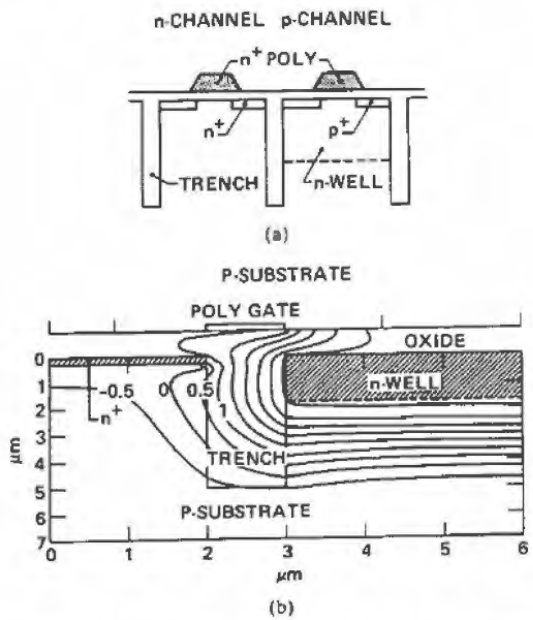


Figure 7.35 Trench isolation and potential distribution. (a) Cross-sectional view of an n-well CMOS with trench isolation; (b) the corresponding 2-D equipotential contours; the simulation is done at $Q_{fc} = 5 \times 10^{10}$ cm^{-2}, $N_A = 6 \times 10^{14}$ cm^{-3}, $V_{well} = 3$ V, $V_G = 0$ V, $V_{sub} = -1$ V (Ref. 27, © 1983 IEEE).

because it involves anisotropic etching, conformal oxide or poly deposition, and planarization. All these processes become even harder if trenches with variable widths are needed. Second, process-induced damage and deposited films always produce fixed charges (Q_{fc}) inside and at the boundary of the trenches. Hence, extra leakage appears along trench sidewalls. Finally, sidewall inversion is a major problem that can place a limitation in layout. Sidewall inversion is caused by the horizontal parasitic MOS device with the well acting as a gate electrode and the trench dielectric as MOS gate oxide with thickness equal to the trench width. The voltage across this MOS device is 5 V under normal CMOS operating conditions. This 5 V gate voltage, and the narrow (e.g., 1 μm) trench width can easily cause inversion at the sidewall outside, but facing the well. In the case of n-well technology, the situation is worse because both lightly-doped p-substrate and Q_{fc} (normally positive) escalate sidewall inversion. The 2-D potential contours in Fig. 7.35(b) indicate that the surface potential at the sidewall opposite the n-well exceeds 0.5 V, causing inversion there.[27] Once sidewall inversion occurs, n-channel devices with n^+ butted to the same sidewall will short along the sidewall. An obvious solution is to leave a separation between the n^+ and the sidewall or to increase trench width; however, both are done at the expense of layout space, thereby defeating the purpose of this technology. More appropriate approaches would be to reduce Q_{fc} through better control in trench etching and filling, and/or to dope the sidewall surface to a higher concentration, i.e., forming the channel stop at the sidewall. The former approach is difficult because it requires extensive process refinements and detail characterization of Q_{fc} at the sidewall. The latter approach is effective in preventing sidewall inversion as shown in Fig. 7.36.

In this example, the Si potential at the sidewall decreases from 0.5 V (Fig. 7.36a) to negative voltages (Fig. 7.36b) when the sidewall is boron-doped to 5×10^{17} cm^{-3}. Low subthreshold leakage has been measured from an n-channel MOSFET, with its source and drain abutted to the sidewall.[28] However, it is extremely difficult to implement a CMOS structure in which only one side of the trench is doped. In the n-well example shown in Fig. 7.35a, only the sidewall in the p-substrate needs to be doped to form a p-type channel stop, the other side of the trench is in the n-well and should not be counter-doped by p-type impurity atoms. The payoff for trench isolation technology in CMOS may not be worth the extra complexity and associated cost[29] unless minimum n^+-to-p^+ separation is absolutely critical to achieve very high packing density.

7.4.2 Selective Epi Growth

Selective epitaxial growth (SEG) has recently been developed for CMOS isolation.[30,31] In this technology, Si is first etched away from the well areas using an oxide mask, then a layer of insulator is formed at the sidewall, and

Greenthread Ex. 2059, p. 40 of 47
Semiconductor v. Greenthread
IPR2023-01242

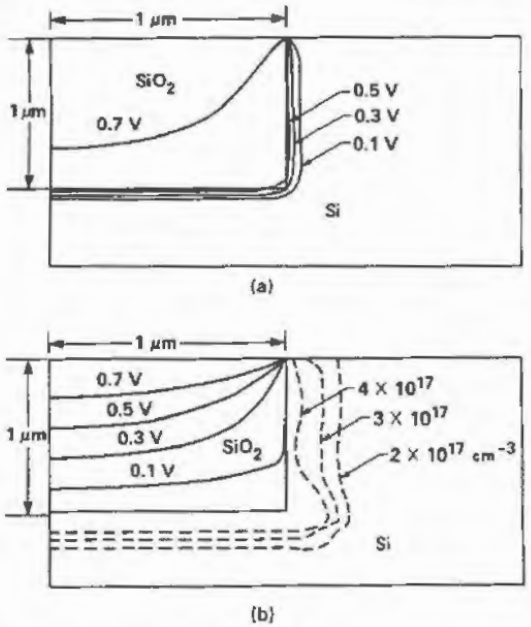


Figure 7.36 2-D potential contours simulated for a trench structure in a p-type substrate with $N_A = 8 \times 10^{16}$ cm^{-3}: (a) no sidewall doping; (b) with sidewall doping. Dashed lines correspond to the doping contours at the sidewall.

finally the well is filled by selective epi growth. Fig. 7.37 shows the schematic of the processing steps for an n-well CMOS. Here phosphorus ions (P^+) are implanted after the sidewall oxide is formed.[31] The SEG technique is then used to grow Si only on the Si substrate, but not on the oxide surface. The n^+ at the bottom improve vertical isolation and lower well sheet resistance. With this etch-and-regrowth technique, doping sidewalls outside the n-well are implemented by depositing a boron-doped film at the sidewall and driving boron into the sidewall before the formation of the sidewall insulator. It is also advantageous to dope the sidewall inside the n-well to higher n-type concentration, but it is not necessary because the n-well is already highly doped and, with the positive Q_{fc}, the n-type Si next to the trench oxide (see arrows in Fig. 7.37d) is less likely to invert. In this process, LOCOS is still needed to form field oxide with various isolation widths.

One way to eliminate LOCOS completely is illustrated in Fig. 7.38.[31] A thick oxide layer is formed and windows are etched in the oxide using the active-area mask. Selective epi growth is then applied to fill the windows followed by an n-well diffusion. The sidewall doping is difficult to achieve

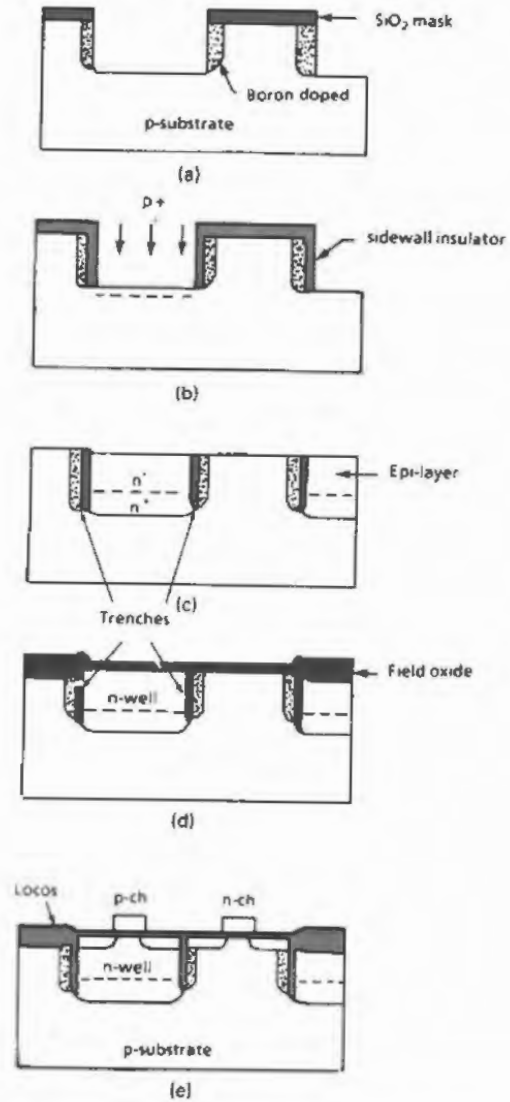


Figure 7.37 SEG process steps for making CMOS with trench and LOCOS isolation (Ref. 31).

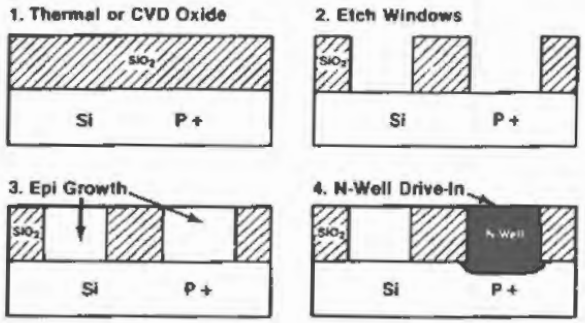


Figure 7.38 SEG process steps for making CMOS with variable isolation widths (Ref. 31).

with this method. A technique to achieve both sidewall doping and variable oxide widths can be done by applying SEG twice to form n- and p-wells independently. As shown in Fig. 7.39,[31] sidewall doping can be added immediately after Step 5 by depositing a boron-doped film, filling p-type SEG epi, and driving boron atoms into p-well sidewalls.

Although the SEG technology appears attractive in CMOS isolation, it is still in its infancy. More work needs to be done to understand material quality at the sidewall of the SEG growth epi layer. Obviously, good selectivity and a clean substrate are key issues in this technology.

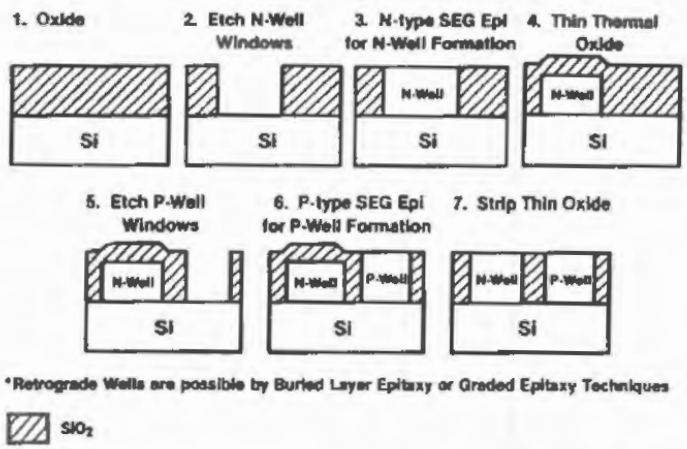


Figure 7.39 SEG process steps for independent p- and n-well formation (Ref. 31).

7.4.3 SOS and SOI Isolation

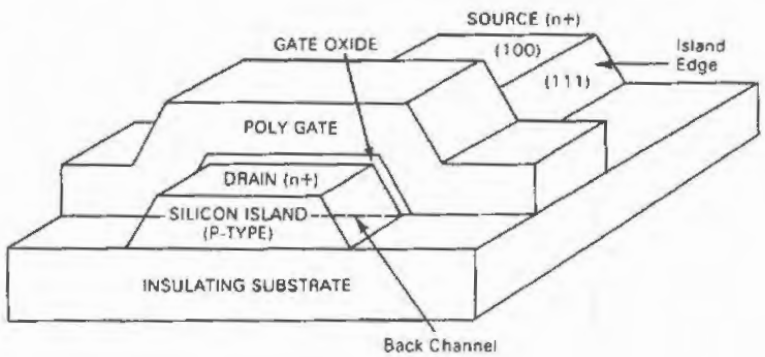
Silicon-on-insulator (SOI) CMOS structures eliminate all previously-mentioned isolation problems by building n- and p-type islands on an insulating substrate for p- and n-channel transistors, respectively. The SOS and SOI process technologies were discussed in Chapter 5, Sec. 5.8 and Sec. 5.9. Similar or opposite type transistors are isolated by insulator and possibly by air. Field inversion and punchthrough do not exist in an SOI or SOS technology, but new parasitic transistors inherent in an SOS or SOI structure can produce leakage current.

As shown in Fig. 7.40(a), a silicon island is formed on an insulator substrate such as sapphire or oxide.[32] A thin gate oxide is grown and the polysilicon line crossing-over the island serves as the gate of the MOS transistor. Two parasitic devices can be found in this structure. First, the substrate contact, the insulating substrate, and the backside of the island form a Metal-Insulator-Semiconductor (MIS) FET, referred to as a back-channel MISFET. Second, the polysilicon gate over the island sidewall forms a sidewall parasitic FET in parallel with the active MOS transistor. Both parasitic devices can result in leakage current, although the leakage is intra-device rather than inter-device in its nature. Because defect density at the Si island and insulating substrate is normally high due to lattice mismatch and film contamination, a back channel is easily formed because of inversion. Current flowing in this back channel cannot be switched off by applying voltage on the poly gate at the front surface. Therefore, a residual off-state leakage current is normally observed as shown in Fig. 7.40(b).[32] Another leakage current, edge leakage, is also shown in the figure. This leakage component arises from the parasitic sidewall FET. For a ⟨100⟩ SOS or SOI film, crystal orientation at the sidewall is ⟨111⟩ which normally results in higher fixed oxide charge (Q_{fc}) after oxide growth. Consequently, threshold voltage at the sidewall is lower than at the active transistor, thereby resulting in a leakage component in the subthreshold characteristics of the active device. Higher leakage reduces noise margin and increases power dissipation in CMOS circuits built on SOS or SOI technology.

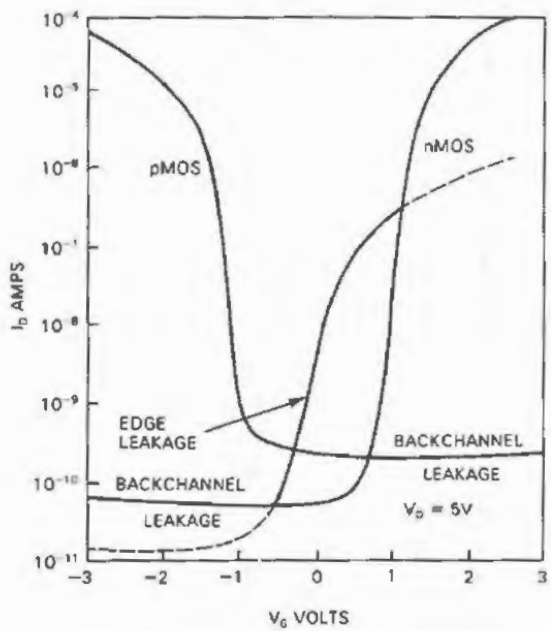
7.5 ISOLATION DESIGN RULES IN CMOS

The results shown in Fig. 7.31 indicate that it is possible to maintain adequate isolation between n^+ diffusions and n-wells at about 1.5 μm nominal separation by using conventional wells, and down to 1 μm separation with retrograde wells. At the same time, the results in Fig. 7.32 show that p^+-to-p-substrate separations down to 0.6 μm can be achieved with either well type. Thus it should be possible to realize full CMOS structures with about 2 μm n^+-to-p^+ separation using conventional n-wells, and about 1.6 μm separation with a retrograde well technology. However, these results are obtained with

Greenthread Ex. 2059, p. 42 of 47
Semiconductor v. Greenthread
IPR2023-01242



(a)



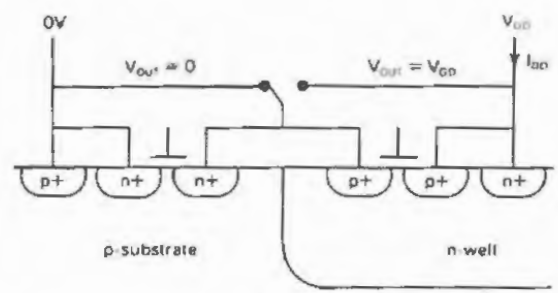
(b)

Figure 7.40 SOS and typical electrical performance: (a) the 3-D representation of an SOS island structure; (b) the corresponding subthreshold characteristics showing edge and back channel leakage for both *n*MOS and *p*MOS (Ref. 32).

isolated n^+-to-*n*-well and p^+-to-*p*-substrate parasitic transistors. To establish isolation design rules, it is necessary to demonstrate such small separations in full CMOS devices.

Fig. 7.41 shows supply currents drawn by full CMOS structures with 1.8 μm n^+-to-p^+ separations (1.2 μm n^+-to-*n*-well edge and 0.6 μm p^+-to-*n*-well edge).[18] Fig. 7.41(a) shows the measurement arrangement. The output node, connecting the n^+ and p^+ diffusions closest to the well edge, is either shorted to ground ($V_{OUT} = 0$) or to the supply line ($V_{OUT} = V_{DD}$). This instance simulates an inverter with its output in either a Low or High state. Ideally the supply current drawn by the CMOS device should be zero in both cases, but any leakage due to inadequate isolation between the *n*- and *p*-channel devices will be detected as a component of the supply current.

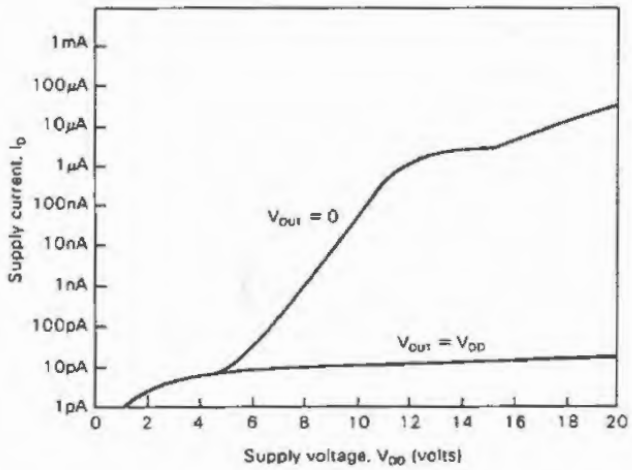
Fig. 7.41(b) shows results for conventional well devices. When $V_{OUT} = V_{DD}$, no potential difference exists between the n^+ diffusion and the *n*-well, so no leakage current flows between them. The limiting mechanism in this case is p^+ to *p*-substrate breakdown, and according to Fig. 7.32, no breakdown should be observed below 20 volts. Clearly, the worst case for leakage is with the output node at ground. Under this condition, the full supply voltage appears between the n^+ diffusion and the *n*-well, and the mechanism limiting the maximum operating voltage is n^+-to-*n*-well punchthrough leakage. The results shown in Fig. 7.41(b) for $V_{OUT} = 0$ are consistent with the punchthrough breakdown voltages shown in Fig. 7.31(b). Fig 7.41(b) shows that significant leakage occurs at $V_{DD} = 5$ volts or above. The CMOS device shown does not provide isolation with sufficient safety margin for 5 V operation. Increasing the n^+ to *n*-well spacing to 1.8 μm (n^+ to p^+ separation



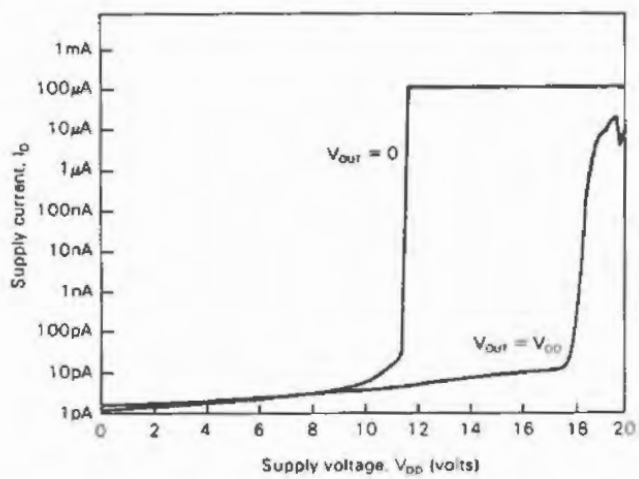
(a)

Figure 7.41 Test structures and I-V curves for a full CMOS isolation structure: (a) test structure and measurement arrangement; (b) leakage current for conventional well; and (c) leakage current for retrograde well devices (Ref. 18, © 1987 IEEE).

Greenthread Ex. 2059, p. 43 of 47
Semiconductor v. Greenthread
IPR2023-01242



(b)



(c)

Figure 7.41 *Continued.*

= 2.4 μm) gives a maximum operating voltage limited by junction breakdown (just over 20 volts, in this case).

Fig. 7.41(c) shows results for retrograde well devices with 1.8 μm n^+-to-p^+ separation. In these cases, the leakage current remains below 10 pA for supply voltages as high as 10 V, providing a good safety margin for 5 V operation. Again, the condition with $V_{OUT} = 0$ sets the lowest breakdown voltage, although p^+-to-n-well junction avalanche breakdown occurs at only a few volts more than the onset of n^+-to-n-well punchthrough leakage.

In a standard LOCOS-isolated CMOS IC, isolation among similar-type transistors can be achieved with 1 μm isolation spacing. Isolation among opposite-type transistors is more involved, however modelling and measurement results show that it is possible to achieve isolation between an n- and a p-channel transistor with approximately 2.4 μm isolation spacing. The use of retrograde wells can further reduce this isolation dimension to below 2 μm. At such small isolation dimensions, latchup becomes a major problem and ultimately may be the factor limiting the isolation spacing between an n- and a p-channel transistor. Latchup in CMOS will be discussed in the next chapter.

REFERENCES

1. K. L. Wang, S. A. Sadler, W. R. Hunter, P. K. Chatterjee, and P. Yang, "Direct moat isolation for VLSI," *IEEE Trans. Electron Devices*, vol. ED-29, pp. 541–547, 1982.
2. S. M. Sze, Physics of Semiconductor Devices, 2nd ed. New York: Wiley, 1981.
3. T. V. Nordstrom and C. F. Gibbon, "The effect of gate oxide thickness on the radiation hardness of silicon-gate CMOS," *IEEE Trans. Nucl. Sci.*, vol. NS-28, pp. 4349–4353, 1981.
4. E. Kooi, J. G. Van Lierop, and J. A. Appels, "Formation of silicon nitride at a Si–SiO_2 interface during local oxidation of silicon and during heat treatment of oxidized silicon in NH_3 gas," *J. Electrochem. Soc.*, vol. 123, p. 1117, 1976.
5. R. A. Monline and G. W. Reutlinger, "Self-aligned maskless chan-stops for IGFET integrated circuits," *IEEE Trans. Electron Devices*, vol. ED-20, pp. 1129–1132, 1973.
6. J. Y. Chen and R. C. Henderson, "Radiation effects of e-beam fabricated submicrometer NMOS transistors," *IEEE Electron Device Lett.*, vol. EDL-3, pp. 13–15, 1982.
7. W. G. Oldham, "Isolation technology for scaled MOS VLSI," in *IEDM Tech. Dig.*, pp. 216–219, 1982.
8. J. Y. Chen, R. C. Henderson and D. E. Snyder, "A novel self-aligned isolation process for VLSI," *IEEE Trans. Electron Devices*, vol. ED-30, p. 1521, 1983.
9. J. Hui, T. Y. Chiu, S.-W.S. Wong, and W. G. Oldham "Sealed-interface local oxidation technology," *IEEE Trans. Electron Devices*, vol. Ed-29, pp. 554–561,

Greenthread Ex. 2059, p. 44 of 47
Semiconductor v. Greenthread
IPR2023-01242

1982; also, J. Hui, T. Y. Chiu, S. Wong, and W. G. Oldham, "Selective oxidation technologies for high density MOS," in *IEEE Electron Device Lett.*, vol. EDL-2, p. 244, 1981.

10. K. Y. Chiu, J. L. Moll, and J. Manoliu, "A bird's beak free local oxidation technology feasible for VLSI circuits fabrication," *IEEE Trans. Electron Devices*, vol. ED-29, pp. 536–540, 1982; K. Y. Chiu, R. Fang, J. Lin, and J. L. Moll, "The SWAMI—A defect free and near-zero bird's beak local oxidation technology for VLSI," in *Proc. 1982 Symp. VLSI Technol.*, pp. 28–29, 1982.
11. K. Kurosawa, T. Shibata, and H. Izuka, "A new bird's beak free field isolation technique for VLSI Devices," in *IEDM Tech. Dig.*, pp. 384–387, 1981.
12. K. Ehara, T. Morimoto, S. Muramoto, T. Hosoya, and S. Matsuo, "Perfect planar technology for VLSIs," in *Proc. 1982 Symp. VLSI Technol.*, pp. 30–31, 1982.
13. J. Y. Chen, R. C. Henderson, J. T. Hall, and E. M. Yee, "A fully recessed field isolation technology using photo-CVD oxide," in *IEDM Tech. Dig.*, pp. 233–236, 1982.
14. L. C. Parrillo, R. S. Payne, R. E. Davis, G. W. Reutlinger, and R. L. Field, "Twin-tub CMOS: A technology for VLSI circuits," in *IEDM Tech. Dig.*, p. 752, 1980; also L. C. Parrillo, L. K. Wang, R. D. Swenumson, R. L. Field, R. C. Melin, and R. A. Levy, "Twin-tub CMOS II: An advanced VLSI technology," in *IEDM Tech. Dig.*, p. 706, 1982; also, J. Agraz-Guerena, R. Ashton, W. Bertram, R. Melin, R. Sun and J. T. Clemens, "Twin Tub III—A third generation CMOS technology," in *IEDM Tech. Dig.*, p. 63, 1984.
15. J. Y. Chen, "CMOS—The Emerging VLSI Technology," *IEEE Circuits and Devices*, vol. 2, p. 16, March 1986.
16. A. G. Lewis, R. A. Martin, and J. Y. Chen, "Retrograde and conventional *n*-well CMOS technologies: A comparison," Internal Report, EIL-87-17, Xerox Palo Alto Research Center, Oct. 1986.
17. J. Y. Chen and D. E. Snyder, "Modeling Device Isolation in High Density CMOS," *IEEE Electron Device Lett.*, vol. EDL-7, p. 64, 1986.
18. A. G. Lewis, J. Y. Chen, R. A. Martin, and T. Y. Huang, "Device Isolation in High Density LOCOS-Isolated CMOS," *IEEE Trans. Electron Devices*, vol. ED-34, p. 1337, June 1987.
19. R. A. Martin and J. Y. Chen, "Optimized Retrograde *N*-Well for 1-μm CMOS Technology," *IEEE J. of Solid State Circuits*, vol. SC-21, p. 286, 1986.
20. R. A. Martin, A. G. Lewis, T. Y. Huang, and J. Y. Chen, "A new process for one micron and finer CMOS," *IEDM Tech. Dig.*, p. 403, 1985.
21. J. Y. Chen and A. G. Lewis, "Parasitic transistor effects in CMOS VLSI," *IEEE Circuits and Devices Magazine*, Vol. 4, p. 8, May 1988.
22. A. G. Lewis, R. A. Martin, J. Y. Chen, T.-Y. Huang, and M. Koyanagi, "Vertical isolation in shallow *n*-well CMOS circuits," *IEEE Electron Device Lett.*, EDL-8, p. 107, 1987.
23. R. D. Rung, C. J. Dell'Oca, and L. G. Walker, "A retrograde *p*-well for high-density CMOS," *IEEE Trans. Electron Devices*, vol. ED-28, p. 1115, 1981.
24. J. Y. Chen, "Quadruple-Well CMOS for VLSI Technology," *IEEE Trans. Electron Devices*, vol. ED-31, p. 910, 1984.

25. R. D. Rung, H. Momose and Y. Nagakubo, "Deep trench isolated CMOS devices," *IEDM Tech. Dig.*, p. 6, 1982.
26. T. Yamaguchi, S. Morimoto, G. Kawamoto, H. K. Park, and G. C. Eiden, "High-speed latch-up-free 0.5 μm-channel CMOS using self-aligned Ti-Si and deep-trench isolation technologies," in *IEDM Tech. Dig.*, p. 522, 1983.
27. K. M. Cham, S. Chiang, D. Wenocur, and R. D. Rung, "Characterization and modelling of the trench surface inversion problem for the trench isolated CMOS technology," in *IEDM Tech. Dig.*, p. 23, 1983.
28. G. Chang, private communication.
29. R. D. Rung, "Trench isolation prospects for application in CMOS VLSI," in *IEDM Tech. Dig.*, p. 574, 1984.
30. T. I. Kamins, S. Y. Chiang, D. R. Bradbury, and D. B. Rao, "A CMOS Device isolation structure using the selective-etch-and-refill-with-epi (SEREPI) Process," in *IEDM Tech. Dig.*, p. 757, 1985.
31. J. Borland, "Submicron CMOS/Epi Technology," presented in the Seminar on Innovations in VLSI/USLI CMOS Technology, Santa Clara, Dec. 5, 1986.
32. D. Leong, "Silicon-on-sapphire," in the Lecture Notes of the short course on Silicon-on-Insulator (SOI) Technologies for Integrated-Circuit Applications, U.C. Berkeley, Oct. 1983.

EXERCISES

1. Suppose the field oxide under a n^+ polysilicon line is 6000 Å and the *p*-type channel stop under the oxide is doped to 1×10^{17} cm^{-3}, calculate the threshold voltage and the subthreshold swing factor if the oxide charge density is 1×10^{11} cm^{-2}.
2. If the measured threshold voltage of an aluminum-gate field oxide test device is 8 V and the measurement is done at 1 μA, what is the leakage current when the aluminum line is at 5 V (assuming the subthreshold swing is 0.5 V/dec at room temperature)? If the threshold voltage decreases to 6.5 V when temperature is increased to 125°C, what is the leakage current at 5 V?
3. Based on the I-V relation of an MOSFET, measured current is proportional to the electric channel width (W_e). Describe how one can extract W_e by measuring a series of test devices with different drawn channel widths.
4. Using a planar isolation technology such as fully-recessed LOCOS, assume that electrical channel narrowing effect is negligible. Supposing the difference between the actual physical channel width and the drawn channel width is 0.8 μm and the transconductance of a minimum size FET must be at least 0.7 mS, what should be the minimum drawn channel width? If the physical separation between two adjacent FETs must be 2 μm to provide proper isolation, what should be the minimum drawn separation? And what is the minimum drawn and actual pitch in the active area? (Assume that the FET has t_{ox} = 200 Å, m_n = 500 cm²/V-sec, L_{eff} = 1 μm.)
5. Based on Figs. 7.19 and 7.20, obtain the subthreshold swing (S_t) for the parasitic *n*- and *p*-channel FETs (both Al- and poly-gate devices) with a 2 μm separation between the two n^+ or two p^+ diffusion regions.

LIST OF SYMBOLS

Symbol	Description
α	bipolar common base current gain
α_n	electron impact ionization rate
α_p	hole impact ionization rate
A_E	emitter area
β	bipolar common emitter current gain
β_{npn}	common emitter current gain for an npn bipolar
β_{pnp}	common emitter current gain for a pnp bipolar
C_{ox}	oxide capacitance per unit area
C_d	depletion capacitance per unit area
C_g	gate capacitance per unit area
C_m	Miller capacitance
C_s	stray capacitance
C_L	load capacitance
C_{gd}	gate-to-drain capacitance
C_{gs}	gate-to-source capacitance
C_{gb}	gate-to-bulk capacitance
C_{ov}	overlapping capacitance
C_{fr}	fringing field capacitance
C_j	junction capacitance
D	diffusivity
D_i	intrinsic diffusivity
D_n	electron diffusion constant

Symbol	Description
d	depletion width
d_{max}	maximum depletion width
d_{inv}	inversion layer thickness
ξ_i	threshold energy for impact ionization
ξ_h	threshold energy for overcoming Si/SiO_2 energy barrier
E	electric field
E_m	maximum electric field
E_{eff}	effective electric field
E_{crit}	critical electric field
E_a	activation energy
E_c	conduction band edge
E_v	valence band edge
E_i	intrinsic energy level
E_F	Fermi level
E_{Fn}	quasi Fermi level for electrons
E_{Fp}	quasi Fermi level for holes
ε_{ox}	oxide dielectric constant
ε_{Si}	silicon dielectric constant
Φ_B	bulk potential
Φ_s	surface potential
Φ_{ms}	work function difference
Φ_m	work function of a metal
Φ_s	work function of a semiconductor
f_t	bipolar cutoff frequency
G	generation rate due to impact ionization
g_d	MOSFET channel conductance
g_m	MOSFET transconductance
I_E	emitter current
I_{NE}	electron current injected from emitter
I_{PB}	hole current injected from base
I_B	base current
I_C	collector current
I_{NC}	electron current at base-collector junction
I_{sub}	substrate current
I_d (I_D)	drain current
I_g (I_G)	gate current
I_S	source current
I_{trig}	latchup trigger current
I_{hold}	latchup holding current
J_n	electron current density
J_p	hole current density
k	Boltzman constant
λ	dimensionless unit for lamda design rules
λ_r	hot electron mean free path
l	characteristic channel length
L	MOSFET channel length
L_{eff}	MOSFET effective channel length
L_{nB}	electron diffusion length in an *npn* base
N_A	acceptor impurity density
N_D	donor impurity density

Greenthread Ex. 2059, p. 46 of 47
Semiconductor v. Greenthread
IPR2023-01242

Symbol	Description
n_i	intrinsic free carrier concentration
n	free electron concentration
n_B	excess electron density in an npn base
n_{BO}	electron density in an *npn* base at equilibrium
p	free hole concentration
p_{EO}	hole density in an *npn* emitter at equilibrium
p_E	excess hole density in an *npn* emitter
ρ	charge density
ρ_c	contact resistivity
Q_m	total charge at MOS gate
Q_n	free charge at MOS inversion layer
Q_B	total depletion charge
Q_{fc}	fixed charge density
R	recombination and generation rate
R_d	drain resistance
R_s	source resistance
R_S	substrate resistance for latchup characterization
R_W	well resistance for latchup characterization
R_σ	transistor resistance
R_{SE}	emitter sheet resistance
R_{SB}	base sheet resistance
R_c	contact resistance
R_p	projected range in implant profile
γ	body effect coefficient
γ_e	emitter efficiency
σ_p	standard deviation (straggle) in an implant profile
S	MOSFET scaling factor
S_t	subthreshold swing factor
T_e	electron temperature
t_{ox}	gate oxide thickness
t_{fox}	field oxide thickness
t_r	rise time
t_f	fall time
t_{in}	inverter delay time
τ_t	bipolar base transit time
μ	mobility
μ_n	electron mobility
μ_p	hole mobility
v	carrier velocity
V_{BE}	base emitter bias voltage
V_{ben}	base-emitter voltage for an *npn* bipolar
V_{bep}	base-emitter voltage for a *pnp* bipolar
V_{bi}	*p-n* junction built-in voltage
V_s (V_S)	source voltage
V_d (V_D)	drain voltage
V_{dd} (V_{DD})	power supply voltage
V_{dsat}	saturation drain voltage
V_g (V_G)	gate voltage
V_{sub}	substrate voltage
V_t	MOS threshold voltage

Symbol	Description
V_{tn} (V_{Tn})	*n*-channel threshold voltage
V_{tp} (V_{Tp})	*p*-channel threshold voltage
V_{TE}	threshold voltage for an enhancement-mode FET
V_{TD}	threshold voltage for a depletion-mode FET
V_{TF}	threshold voltage for a field oxide FET
V_{FB}	flat-band voltage
V_{pt} (V_{PT})	punchthrough voltage
W	MOSFET channel width
W_{eff}	MOSFET effective channel width
W_B	bipolar base width
W_E	bipolar emitter width
x_j	junction depth
χ	electron affinity