B. JAYANT BALIGA

MODERN POWER DEVICES

13·95
E8

MODERN POWER DEVICES

MODERN POWER DEVICES

B. JAYANT BALIGA

General Electric Company
Schenectady, New York

A Wiley-Interscience Publication

JOHN WILEY & SONS

New York • Chichester • Brisbane • Toronto • Singapore

Library of Congress Cataloging in Publication Data:

Baliga, B. Jayant, 1948-
Modern power devices.

"A Wiley-Interscience publication."
Includes bibliographies and index.
1. Power electronics. 2. Power semiconductors.
3. Power transistors. 4. Semiconductor rectifiers.
I. Title.
TK7881.15.B35 1987 621.38 86-15891

Printed in Singapore

10 9 8 7 6 5 4 3 2 1

To

My wife

PREFACE

The development of electronic power systems based on semiconductor devices can be traced to the early 1950s. At that time, the first rectifiers and thyristors capable of operating at high current and voltage levels were introduced into the marketplace. During the ensuing 30 years, the technology for these bipolar devices reached a high degree of maturity. Devices capable of operating at up to 3000 volts and controlling 1000 amperes of current were in production by the early 1970s. The physics and technology of these bipolar power devices have been treated by Professor S. K. Ghandhi in a book titled *Semiconductor Power Devices*, published by John Wiley & Sons in 1977.

Since that time considerable progress has been made in discrete power semiconductor device technology. The most important development is the creation of a new family of power devices with high-input-impedance metal oxide–semiconductor (MOS) gates. These devices greatly reduce the size and complexity of the control circuitry. This has allowed a large reduction in system cost, making the power electronics attractive for many new applications such as control of home appliances and for automotive electronics. In addition, several other new device concepts, such as the field-controlled diode (FCD) and the high-voltage junction field-effect transistor (JFET), have been explored for applications in power circuits. Despite the growing interest in this field, no book has been available that treats the physics and technology of these modern power devices.

This book is written at the tutorial level to fill this need. It can be used for self-study by professional engineers practicing the art of power device design and fabrication. It can also serve as a textbook for a graduate level course on power devices when used together with Professor Ghandhi's book. In writing the book, it has been assumed that the reader is familiar with the fundamental concepts of current transport in semiconductors and has a basic knowledge of process technology.

Each chapter is organized to first introduce the basic device structure and its electrical output characteristics. A detailed discussion of the physics of device operation in the static blocking and conduction states is then provided. Following this, the dynamic switching behavior and frequency response are analyzed.

Throughout this discussion, emphasis is placed on deriving simple analytical expressions that describe the device characteristics, because they provide a good physical understanding of device performance. For a complete characterization of the electrical properties of devices, it is necessary to resort to numerical techniques using computer programs that solve the fundamental semiconductor transport equation in two dimensions (and sometimes in three dimensions), with time dependence included for the transient case. By use of the device analysis presented in each chapter, in conjunction with the silicon material properties provided in Chapter 2, such computer programs can be generated if desired.

No book on power devices would be complete without a detailed discussion of the physics, design, and technology for the realization of high breakdown voltage as discussed in Chapter 3. The device breakdown voltage is usually limited by the edge termination. The importance of this technology has stimulated the conception of a wide variety of termination techniques. The choice of the device termination technique is interrelated to device power rating, chip size, and the fabrication process. The study of this chapter is essential to the design of all the devices discussed in this book.

Chapters 4–7 are devoted to the three-terminal power devices that have evolved since 1975. In Chapter 4, the power JFET is treated. These devices are attractive for very-high-frequency applications. The power FCD is discussed in Chapter 5. These devices possess unusually good radiation tolerance and can be used in high-temperature applications. Chapter 6 provides a detailed analysis of the power MOSFET, whereas Chapter 7 is focused on the MOS–bipolar device concepts. These two developments are the most important aspects of the new device technology. They can be expected to play an increasing role in power circuits in the future.

Concurrent with these revolutionary advances in three-terminal power devices, remarkable strides have been made in improving the performance of power rectifiers. The new concepts that are responsible for the availability of lower forward drop and higher speed in power rectifiers are discussed in Chapter 8. Finally, in Chapter 9, guidelines for making a choice between these devices are provided.

Over the course of the many years that I have worked in this field, it has been my privilege to have the opportunity to discuss the operation of devices and their process development with my many colleagues at the General Electric Research and Development Center, Schenectady, New York. I would like to acknowledge the assistance received from them and the unselfish dedication of the process technicians who were often called upon to perform seemingly impossible tasks. My gratitude goes to Dr. T. S. P. Chow for his assistance with the preparation of the illustrations used in this book, and to Mrs. Y. Nakahigashi for her typing of the manuscript on a demanding schedule.

A special note of gratitude must be expressed to Professor S. K. Ghandhi. He had the foresight to encourage me to start my career in the field of power semiconductor devices at a time when it appeared to be stagnant. In addition

to his encouragement in the preparation of this book, I am indebted to him for his detailed and constructive criticism of the manuscript.

The preparation of this book has been made possible by the generosity of the General Electric Company in conferring on me the Coolidge Fellowship Award. During the one-year sabbatical that is provided for in the award, I found the time to pursue the preparation of this book, which would have been impossible under the pressures of my normal working schedule.

In closing, special thanks are due to my wife, Pratima, and my son, Avinash, for tolerating my preoccupation with this project, which absorbed a large proportion of my time and detracted from my giving them the attention that they certainly deserve.

B. JAYANT BALIGA

Schenectady, New York
December 1985

CONTENTS

MODERN POWER DEVICES

1 INTRODUCTION

Starting with the conception of the bipolar junction transistor (BJT) in 1947, the power handling capability of silicon devices has grown steadily. As the power handling capability and frequency response have improved, new applications for these devices have been created. Today, the market for silicon bipolar power devices exceeds $1 billion. The recent introduction of MOS-controlled power devices and power integrated circuits promises to lead to a further growth in this industry.

Power metal oxide semiconductor field-effect transistors (MOSFETs) were commercially introduced in the 1970s. Because of the high input impedance of the device, a significant reduction in the complexity and cost of the gate drive circuit in power control systems can be achieved by replacing the bipolar transistor with power MOSFETs. During the past 10 years there has been an increasing acceptance of the usage of power MOSFETs. However, the displacement of bipolar transistors in their applications by power MOSFETs has lagged far behind initial expectations because of the relatively high cost of the latter devices for the same power rating. As power MOSFET technology matures, the cost differential is continually diminishing making it more attractive. In 1985 the power MOSFET market was only 10% of the bipolar transistor market.

A new class of power devices has emerged in the 1980s. The operation of these devices is based on the fusion of the physics of the MOS gate structure and bipolar current conduction. The most commercially advanced device of this category is the insulated-gate transistor (IGT). The IGT exhibits the important features of a high input impedance and a very high power handling capability for a given chip size. In combination with power integrated circuits, these devices can be used in power systems to obtain cost reductions by factors of 10 or more over existing systems. This is expected to result in a very rapid growth in the overall market for power transistors in which the MOS-controlled devices will occupy an increasing segment.

The applications for power devices extend over a very large range of power levels and frequencies. Some of these applications are shown in the power–frequency spectrum in Fig. 1.1. From this figure, it can be seen that the power

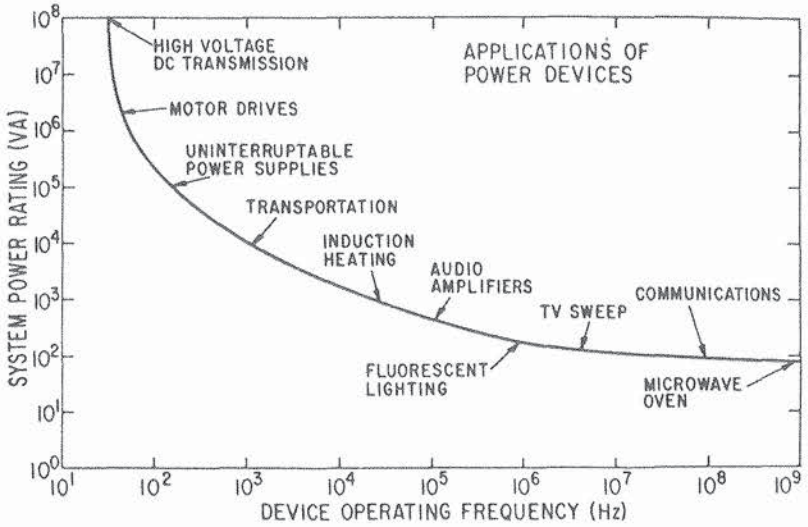


Fig. 1.1. Applications of power semiconductor devices.

levels demanded by systems are generally higher at lower frequencies. For very high power applications, such as high-voltage direct-current (HVDC) transmission systems, the power thyristor is being used very successfully. The recent ability to obtain excellent electrical characteristics combined with light triggering capability for device isolation has ensured the dominance of this technology. The physics of operation, design considerations, and fabrication technology for power thyristors have been adequately treated in other books [1, 2]. These references also provide a good analysis of the operation of bipolar transistors and gate turn-off thyristors that have been used for higher-frequency applications such as motor drives and lighting control. These devices, therefore, are not treated in this book.

During recent years considerable progress has occurred in the power device industry. New power devices, such as the static-induction transistor, FCD, and the IGT, have been conceived. These devices have matured to the point at which they are no longer laboratory curiosities. Their commercial availability creates the threat of displacing the bipolar transistor in many of its traditional applications. The purpose of this book is to elucidate the physics of the operation of these new devices, provide a guide to their design methodology, and describe the unique fabrication technology that has evolved for each device.

The advent of the new power devices discussed in this book is important not merely because these devices may replace the bipolar transistor, but also—and more significantly—because they are forcing a change in the way power circuits are being designed and applied. For example, the availability of MOS gate-controlled devices capable of operating at high switching speeds is changing the way in which motor control is performed. There is an increasing trend toward pulse-width-modulated, higher-frequency, adjustable-speed motor drives because of the significant improvement in efficiency and system performance. This shift in technology was not a viable option with the traditional bipolar

transistor technology because of the high cost of implementing the power electronic circuits for adjustable-speed motor drives. Other examples of areas in which the new device technology is expected to play an important role are automotive electronics, switching power supplies, factory automation, new higher-efficiency lighting systems, and appliance control for the electronic home of the future.

REFERENCES

1. S. K. Ghandhi, *Semiconductor Power Devices*, Wiley, New York, 1977.
2. B. J. Baliga and D. Y. Chen, *Power Transistors: Device Design and Applications*, IEEE Press, New York, 1984.
3. A. Blitcher, *Field Effect and Bipolar Power Transistor Physics*, Academic, New York, 1981.
4. A. Blitcher, *Thyristor Physics*, Springer-Verlag, New York, 1976.

6 POWER METAL-OXIDE–SEMICONDUCTOR FIELD-EFFECT TRANSISTORS

Power metal-oxide–semiconductor field-effect transistors (MOSFETs) are the most commercially advanced devices discussed in this book. These devices have evolved from MOS integrated-circuit technology. Prior to the development of power MOSFETs, the only device available for high-speed, medium-power applications was the power bipolar transistor. The power bipolar transistor was first developed in the early 1950s, and its technology has matured to a high degree, allowing the fabrication of devices with current handling capability of several hundred amperes and blocking voltages of 600 V. Despite the attractive power ratings achieved for bipolar transistors, there exist several fundamental drawbacks in their operating characteristics. First, the bipolar transistor is a current-controlled device. A large base drive current, typically one-fifth to one-tenth of the collector current, is required to maintain them in the on-state. Even larger reverse base drive currents are necessary for obtaining high-speed turn-off. These characteristics make the base drive circuitry complex and expensive. The bipolar transistor is also vulnerable to a second breakdown failure mode under the simultaneous application of a high current and voltage to the device as commonly required in inductive power circuits. Furthermore, it is difficult to parallel these devices. The forward voltage drop in bipolar transistors decreases with increasing temperature. This promotes diversion of the current to a single device unless emitter ballasting schemes are utilized.

The power MOSFET was developed to solve the performance limitations experienced with power bipolar transistors. In this device, the control signal is applied to a metal gate electrode that is separated from the semiconductor surface by an intervening insulator (typically silicon dioxide). The control signal required is essentially a bias voltage with no significant steady-state gate current flow in either the on-state or the off-state. Even during the switching of the devices between these states, the gate current is small at typical operating frequencies (< 100 kHz) because it only serves to charge and discharge the input gate capacitance. The high input impedance is a primary feature of the power MOSFET that greatly simplifies the gate drive circuitry and reduces the cost of the power electronics.

The power MOSFET is a unipolar device. Current conduction occurs through transport of majority carriers in the drift region without the presence of minority carrier injection required for bipolar transistor operation. No delays are observed as a result of storage or recombination of minority carriers in power MOSFETs during turn-off. Their inherent switching speed is orders of magnitude faster than that for bipolar transistors. This feature is particularly attractive in circuits operating at high frequencies where switching power losses are dominant.

Power MOSFETs have also been found to display an excellent safe operating area; that is, they can withstand the simultaneous application of high current and voltage (for a short duration) without undergoing destructive failure due to second breakdown. These devices can also be easily paralleled because the forward voltage drop of power MOSFETs increases with increasing temperature. This feature promotes an even current distribution between paralleled devices.

These characteristics of power MOSFETs make them important candidates for many applications. They are being used in audio/radiofrequency circuits and in high-frequency inverters such as those used in switch mode power supplies. Other applications for these devices are for lamp ballasts and motor control circuits.

6.1. BASIC STRUCTURES AND OPERATION

The fundamental operation of the power MOSFET relies on the formation of a conductive layer at the surface of the semiconductor. The modulation of the charge at the surface of a semiconductor by using the bias on a metal plate with an intervening insulating layer was first demonstrated in 1948 [1]. Through application of this principle, the first surface FET was fabricated in silicon with thermally grown silicon dioxide as the insulator [2]. Since then, the insulated-gate field-effect transistor (IGFET) has been applied to the fabrication of a large variety of integrated circuits. Until recently, the development of discrete devices has followed the basic concept of the lateral channel structure used in these earlier applications. Such devices have the drain, gate, and source terminals on the same surface of the silicon wafer. Although this feature makes them well suited for integration, it is not optimum for achieving a high power rating. As mentioned in Chapter 4, the vertical channel structure with source and drain on opposite surfaces of the wafer is more suitable for a power device because more area is available for the source region and because the electric field crowding at the gate is reduced.

Two discrete vertical-channel power MOSFET structures have evolved. A cross section of these structures is provided in Fig. 6.1. The DMOS structure is fabricated by using planar diffusion technology with a refractory gate such as polysilicon. In these devices, the P-base region and the N^+-source region are diffused through a common window defined by the edge of the polysilicon

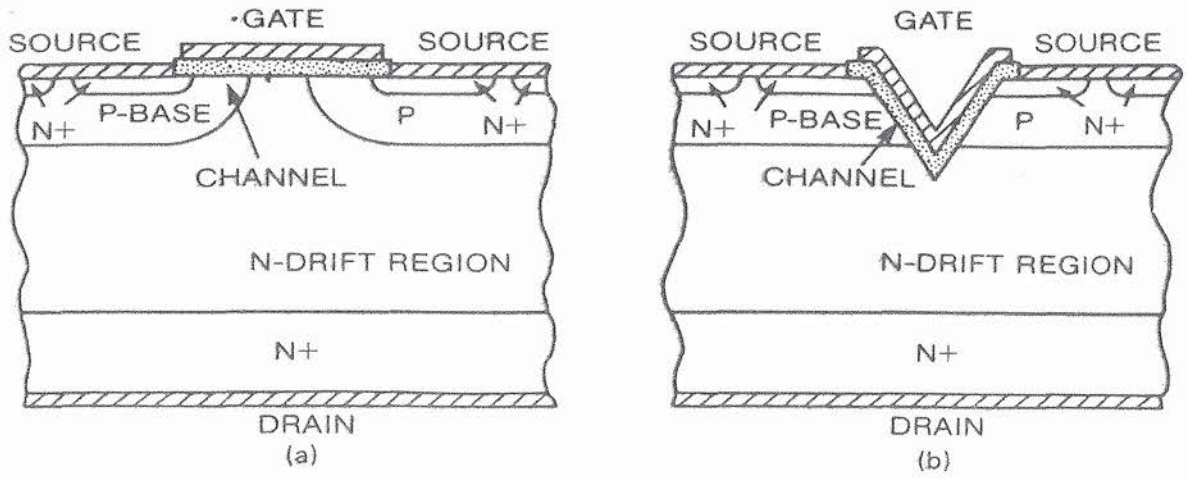


Fig. 6.1. Cross-sectional view of the basic power MOSFET cell structures: (*a*) DMOS structure; (*b*) VMOS structure.

gate. The *P*-base region is driven in deeper than the N^+ source. The difference in the lateral diffusion between the *P*-base and N^+-source regions defines the surface channel region.

The vertical-channel VMOS power FET structure requires a different fabrication process. In this case, an unpatterned *P*-base diffusion is performed followed by the N^+-source region diffusion. A V-shaped groove is then formed extending through these diffusions. The gate electrode is next placed so as to overlap the N^+ source and extend into the groove beyond the bottom of the *P*-base region. The channel region for this structure is formed along the walls of the V groove.

In both of these structures, the $P-N$ junction between the P-base region and the *N*-drift region provides the forward blocking capability. Note that the *P*-base region is connected to the source metal by a break in the N^+-source diffusion. This is important to establishing a fixed potential to the *P*-base region during device operation. If the gate electrode is externally short-circuited to the source, the surface of the *P*-base region under the gate (i.e., the channel region), remains unmodulated at a carrier concentration determined by the doping level. When a positive drain voltage is now applied, it reverse-biases the *P*-base/*N*-drift region junction. This junction supports the drain voltage by the extension of a depletion layer on both sides. Because of the higher doping level of the *P*-base region, the depletion layer extends primarily into the *N*-drift region. Its doping concentration and width must be chosen in accordance with the criteria established in Chapter 3 for avalanche breakdown of $P-N$ junctions. Higher drain blocking voltage capability requires a lower drift region doping and a greater width.

It is important to connect the gate electrode to the source to establish its potential at the lowest point during the forward blocking state. If the gate is left floating, its potential can rise through capacitive coupling to the drain potential. This induces modulation of the channel region, which can produce

an undesirable current flow at drain voltages well below the avalanche breakdown limit. Thus power MOSFETs will not support large drain voltages unless the gate is grounded (i.e., connected to the source during forward blocking).

To carry current from drain to source in the power MOSFET, it is essential to form a conductive path extending between the N^+-source regions and the N-drift region. This can be accomplished by applying a positive gate bias to the gate electrode. The gate bias modulates the conductivity of the channel region by the strong electric field created normal to the semiconductor surface through the oxide layer. For a typical gate oxide thickness of 1000 Å and gate drive voltage of 10 V, an electric field of 10^6 V/cm is created in the oxide. The gate-induced electric field attracts electrons to the surface of the P-base region under the gate. This field strength is sufficient to create a surface electron concentration that overcomes the P-base doping. The resulting surface electron layer in the channel provides a conductive path between the N^+-source regions and the drift region. The application of a positive drain voltage now results in current flow between drain and source through the N-drift region and the channel. This current flow is controlled by the resistance of these regions. Note that the current flow occurs solely by transport of majority carriers (electrons for n-channel devices) along a resistive path comprising the channel and drift regions. No minority carrier transport is involved for the power MOSFET during current conduction in the on-state.

To switch the power MOSFET into the off-state, the gate bias voltage must be reduced to zero by externally short-circuiting the gate electrode to the source electrode. When the gate voltage is removed, the electrons are no longer attracted to the channel and the conductive path from drain to source is broken. The power MOSFET then switches rapidly from the on-state to the off-state without any delays arising from minority carrier storage and recombination as experienced in bipolar devices. The turn-off time is controlled by the rate of removal of the charge on the gate electrode because this charge determines the conductivity of the channel. Turn-off times of under 100 nsec can be achieved with a moderate gate drive current flow arising from the discharge of the input gate capacitance of the device.

An examination of the power MOSFET structures illustrated in Fig. 6.1 reveals the existence of a parasitic N^+–P–N–N^+ vertical structure. This parasitic bipolar transistor must be kept inactive during all modes of operation of the power MOSFET. To accomplish this, the P-base region is short-circuited to the N^+-emitter region by the source metallization as shown in the cross section of the devices. The short circuit between the P-base and the N^+ emitter can be provided within each cell of the device structure as illustrated in Fig. 6.1, or occasionally at selected locations. The latter approach considerably facilitates the fabrication of the cells because it eliminates an alignment step necessary to form the short circuit within the cells. However, the resistance of the P-base region between the short circuits can become large, and any lateral current flow in the P base, as the result of capacitive currents arising at high applied (dV/dt) to the drain, can lead to forward-biasing of the N^+–P junction

at locations remote from the short circuits. Forward-biasing of the N^+–P junction activates the parasitic bipolar transistor and leads to the initiation of minority carrier transport. This can not only slow down the switching of the power MOSFET, but can lead to second breakdown. Because of the high rate of change of voltage observed in the high-frequency applications for which the power MOSFET is particularly well suited, it is common practice to form the short circuit in every cell and minimize the length of the N^+-source region from the edge of the channel to the short circuit.

6.2. BASIC DEVICE CHARACTERISTICS

The power MOSFET is capable of blocking voltage in only one quadrant. For n-channel structures, the devices are operated with a positive voltage applied to the drain. When the gate electrode is short-circuited to the source, the device can support a large drain voltage across the P-base–N-drift region junction. The forward blocking capability is shown in Fig. 6.2 by the lowest trace. Although a finite leakage current is illustrated in the figure, it is very small except at high operating temperatures. The maximum forward blocking voltage is determined by the avalanche breakdown voltage of the P-base–N-drift layer junction. The breakdown voltage is dependent not only on the device termination, but is affected by the internal cell structure of the devices.

When a positive gate bias is applied, the channel becomes conductive. At low drain voltages, the current flow is essentially resistive, with the on-resistance determined by the sum of the channel and drift region resistances. The channel resistance decreases with increasing gate bias, whereas the drift region resistance remains constant. The total on-resistance decreases with increasing gate bias until it approaches a constant value. At large gate bias voltages, the channel resistance becomes smaller than the drift region resistance, and the device on-resistance becomes independent of gate bias. The on-resistance is an important power MOSFET parameter. It is a measure of the current handling capability

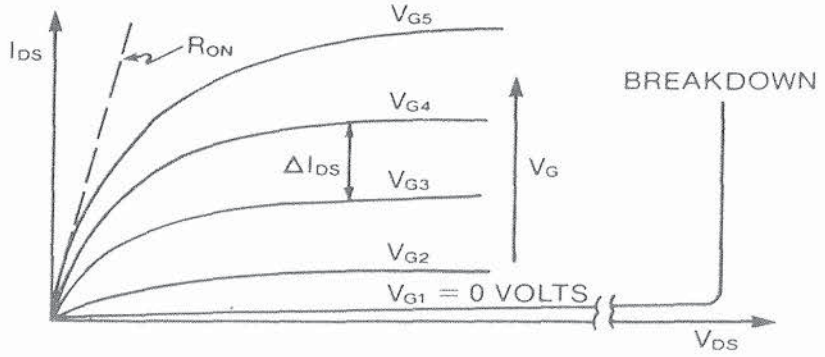


Fig. 6.2. Output characteristics of a power MOSFET.

of the device because it determines the power dissipation during current conduction. The on-resistance is defined as the slope of the output characteristic in the linear region at low drain voltages.

At high drain voltages, the resistance of the power MOSFET increases. Ultimately, the current saturates at high drain voltages as shown on the right-hand side of Fig. 6.2. The current saturation in power MOSFETs can be used to provide a current-limiting function in power circuits as long as the power dissipation in the devices is kept within reasonable bounds.

An important device parameter for power MOSFETs is the transconductance. It is defined by

$$g_m = \left.\frac{\Delta I_{DS}}{\Delta V_{GS}}\right|_{V_{DS}} \tag{6.1}$$

where $\Delta V_{GS} = (V_{G4} - V_{G3})$ is illustrated in Fig. 6.2. A large transconductance is desirable to obtain a high current handling capability with low gate drive voltage and for achieving high frequency response. In the saturated current region of operation, the output characteristics are controlled by the gate-induced channel characteristics. The transconductance is, therefore, determined by the design of the channel and gate structure.

Although the power MOSFET was originally intended for operation in only one quadrant as illustrated in Fig. 6.2 for n-channel devices, it has recently been operated with negative drain voltages as a synchronous rectifier [6]. When a negative drain voltage is applied to the devices illustrated in Fig. 6.1 with their gates connected to the source, the P-base–N-drift layer junction becomes forward-biased. When the drain voltage exceeds a knee voltage V_N of approximately 0.7 V at room temperature, the P–N junction will inject minority carriers and begin to conduct current as illustrated in Fig. 6.3.

If a positive gate voltage is applied to the device to create a channel, an alternate path for current flow between drain and source is created. This current flow occurs by means of majority carriers and is limited by the resistance of

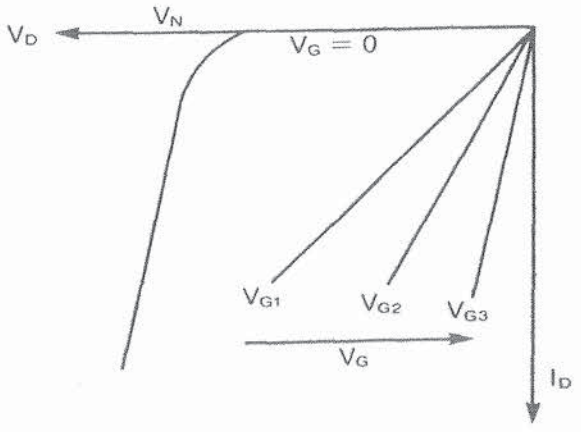


Fig. 6.3. Power MOSFET (n-channel) operated with negative drain voltage.

the channel and drift region. If these resistances are small, the power MOSFET characteristics will resemble those shown in Fig. 6.3. In this figure, the device is shown to carry current with a voltage drop well below the knee point V_N. Under these operating conditions, the P–N junction will not inject any significant concentration of minority carriers. Consequently, the power MOSFET will exhibit a very low forward drop, significantly lower than for a P–N junction rectifier, and retain its high switching speed. For these reasons, it has been suggested that the power MOSFET be used as a replacement for diodes in low-voltage (<5-V), switch-mode power supplies. To accomplish this function, it is essential to provide a gate signal to the power MOSFET in synchronism with the supply voltage to maintain it in the low forward drop on-state, as well as the blocking state, at the appropriate times. The device is, therefore, called a *synchronous rectifier*.

For use as a synchronous rectifier, the power MOSFET must exhibit an extremely low on-resistance if it is to be capable of handling significant drain currents. For example, a device capable of switching 50 A with a forward drop of less than 0.25 V must have an on-resistance of less than 5 mΩ. To achieve this, very-large-area devices are required with breakdown voltages of below 50 V.

6.3. DEVICE ANALYSIS

To understand current flow in the power MOSFET, it is essential to analyze the physics of the MOS structure. Metal-oxide–semiconductor physics controls the characteristics of the channel region, which governs the output characteristics of the devices. The current transport in the channel must then be coupled to the drift region to obtain the total on-resistance of the device. In analyzing the drift region resistance, it is necessary to consider the effect of the pinch-off action arising from the adjacent P-base regions at higher drain voltages.

In many ways, the physics of the operation of power MOSFETs is simpler than that for other power devices because of the absence of minority carrier injection. However, it is necessary to understand the interaction between the cell geometry and the device characteristics before undertaking an accurate device design. In this section, the blocking characteristics are first treated followed by analysis of the on-state characteristics.

6.3.1. Forward Blocking Capability

In the forward blocking mode, the gate electrode of the power MOSFET is externally short-circuited to the source. Under these conditions, no channel forms under the gate at the surface of the P-base region. When a positive drain voltage is applied, the P-base–N-drift layer junction becomes reverse-biased and supports the drain voltage. The breakdown voltage of this junction is not given by the simple parallel-plane analysis described in Chapter 3 for several reasons. First, the power MOSFET structure consists of an N^+–P–N transistor where

the P-base region is short-circuited at selected points to the N^+ emitter by the source metallization. Despite the short-circuiting of the N^+ emitter and P base at some locations, this structure will conduct current as soon as the depletion layer in the P-base punches through to the N^+ emitter because the N^+ emitter then becomes forward-biased and injects electrons into the P-base region. The punch-through breakdown condition is determined by the doping profiles used for device fabrication. The second factor that can alter the breakdown voltage is the presence of junction curvature in the DMOS cell. Its impact on the breakdown voltage is dependent on the cell spacing. The third factor is the presence of sharp corners in the VMOS cell, producing high local electric fields and leading to premature breakdown. These factors are discussed in the following sections.

6.3.1.1. Doping Profile. In the forward blocking mode, the depletion layer of the P-base–N-drift layer junction extends on both sides. The P-base doping must be maintained at a relatively low concentration in power MOSFETs to permit inversion of the surface under the gate during the operation of the power MOSFET in its on-state. A typical diffusion profile for the device is shown in Fig. 6.4. The solid lines indicate the dopant distributions, whereas the dashed lines show the resulting carrier concentration profiles, which differ from the dopant profiles as a result of compensation effects. The surface concentration N_{SP} of the P-base diffusion and the N^+-emitter depth combine to determine the peak doping N_{AP} in the P-base indicated by the arrow in Fig. 6.4. The peak P-base doping is an important parameter for the doping profile because it

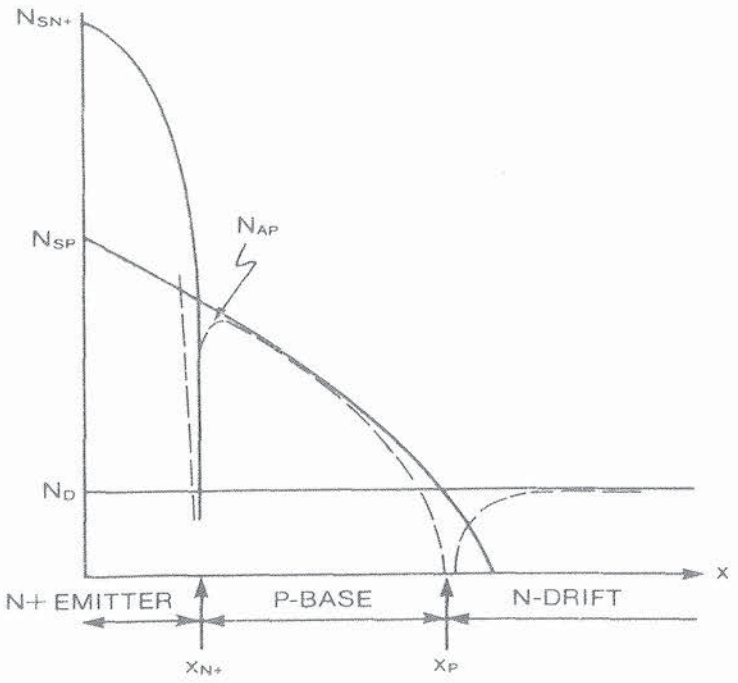


Fig. 6.4. Typical diffusion profile for a power MOSFET.

controls the threshold voltage of the power MOSFET, that is, the minimum gate voltage required to induce a surface channel. For a typical threshold voltage of 2–3 V for an n-channel power MOSFET, the peak base concentration N_{AP} is about $1 \times 10^{17}/\text{cm}^3$.

The channel length is another important design parameter in power MOSFETs because it has a strong influence on the on-resistance and the transconductance. The channel length is determined by the difference between the depths of the P-base and N^+-emitter diffusions [i.e., $(x_P - x_{N^+})$]. (It should be noted that, although a vertical doping profile is measured to characterize devices, the lateral profile under the gate determines the channel properties.) The need to maintain a low peak base concentration and a narrow base width to achieve good on-state characteristics can adversely affect the forward blocking capability. Despite the short-circuiting of the N^+ emitter to the P-base by the source metal, a parasitic N^+–P–N bipolar transistor exists in the power MOSFET. When the P-base–N-drift layer junction is reverse-biased, the depletion layer in the P-base can extend to the N^+-emitter–P-base junction and cause premature punch-through breakdown. It is important to design the P-base diffusion profile so that sufficient charge is resident in it to prevent punch-through of the depletion layer to the N^+ emitter.

The depletion width extension on the diffused side of a P–N junction has been calculated at breakdown using numerical techniques [3]. In Fig. 6.5, the

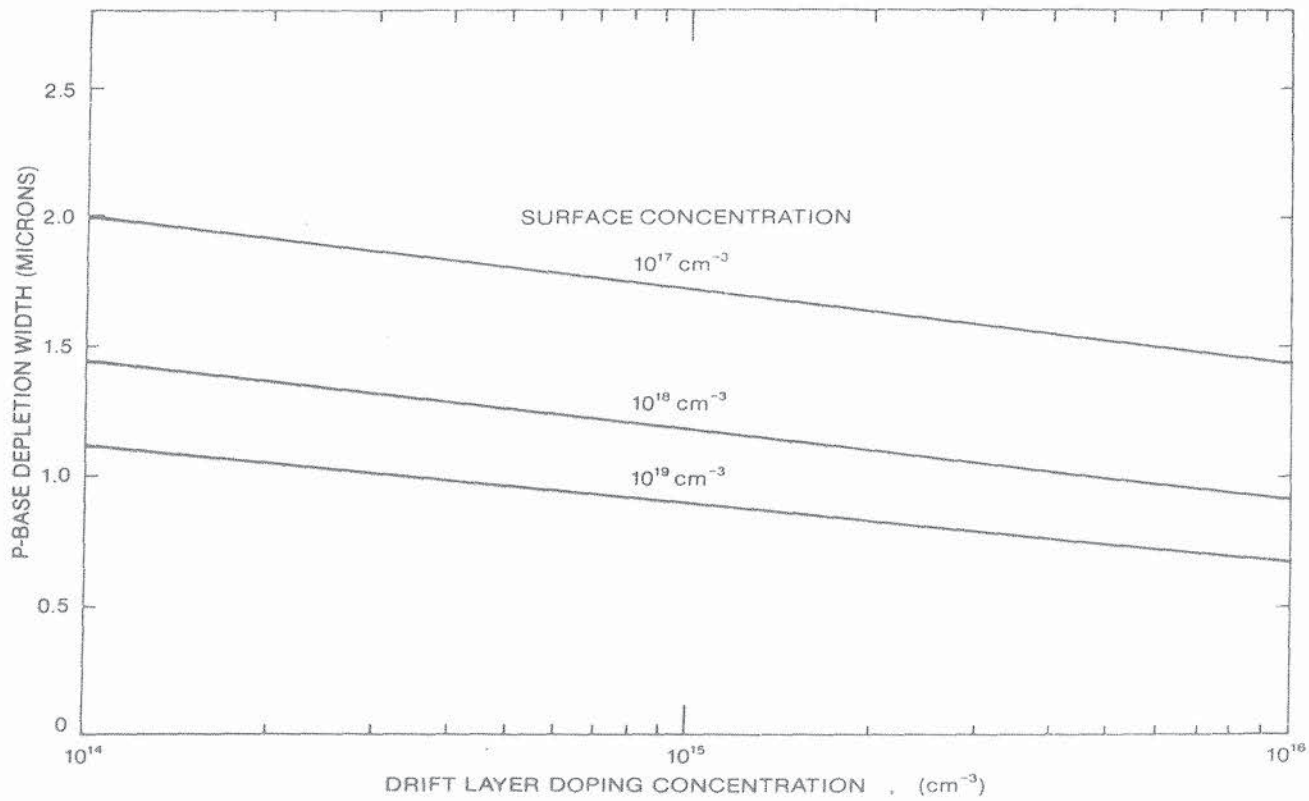


Fig. 6.5. Depletion width extension on the diffused side of a P–N junction for various surface concentrations.

change in the depletion width on the diffused side of the junction is provided as a function of the background doping level and surface concentration of the diffusion. As the drift layer concentration decreases, the junction profile becomes more graded, causing an increase in the depletion width on the diffused side. It should be noted that when the surface concentration is below $10^{18}/cm^3$, the depletion width on the diffused side can extend well over 1 μm, especially for low drift layer concentrations. Care must be taken during device design to prevent punch-through breakdown in these cases. From the data in Fig. 6.5, it can also be concluded that the fabrication of devices with channel lengths of less than 1 μm is difficult to achieve unless the device is required to support only low voltages, that is, for devices with high drift layer concentrations.

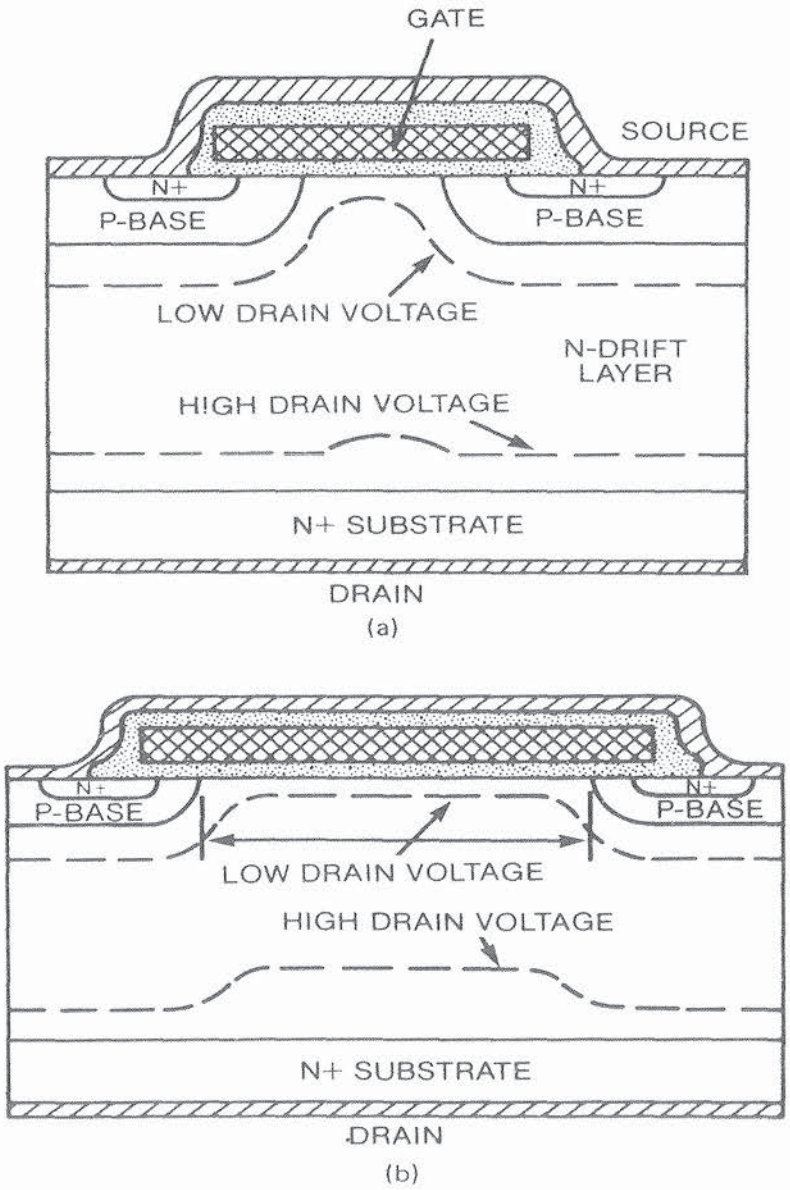


Fig. 6.6. Depletion layer profiles at low and high drain voltages for DMOS devices with (*a*) small and (*b*) large cell spacing.

6.3.1.2. Cell Spacing. To achieve a large channel width for good on-state characteristics, power MOSFETs are fabricated with a repetitive pattern of small cells. In the DMOS cell structure shown in Fig. 6.1*a*, the *P*-base–*N*-drift layer is brought to the surface between the individual cells. Each cell contains a planar junction edge. During forward blocking the depletion layer extends out from the junction as well as from the gate overlap region between the cells. The depletion layer profile at low and high drain voltages is shown for two examples of a very small cell spacing and a large cell spacing in Fig. 6.6. In the case of the DMOS design with small cell spacing, the depletion layer curvature becomes small at high drain voltages. These devices will break down at the edge termination. If the cell spacing is large, however, the curvature of the depletion layer at the edges of the *P*-base regions is significant, resulting in lower breakdown voltage, as was discussed for the case of cylindrical or spherical junctions in Chapter 3.

Numerical analysis of the effect of cell spacing on the breakdown voltage has been performed for the DMOS structure [4]. As an example, the calculated breakdown voltage is shown as a function of the gate width in Fig. 6.7 for several drift region doping levels. It can be seen that the breakdown voltage increases when the cell spacing decreases below 10 μm. Note that the rise in the breakdown voltage occurs at larger cell spacing for the case of lower drift layer doping concentrations. This is consistent with the larger depletion layer

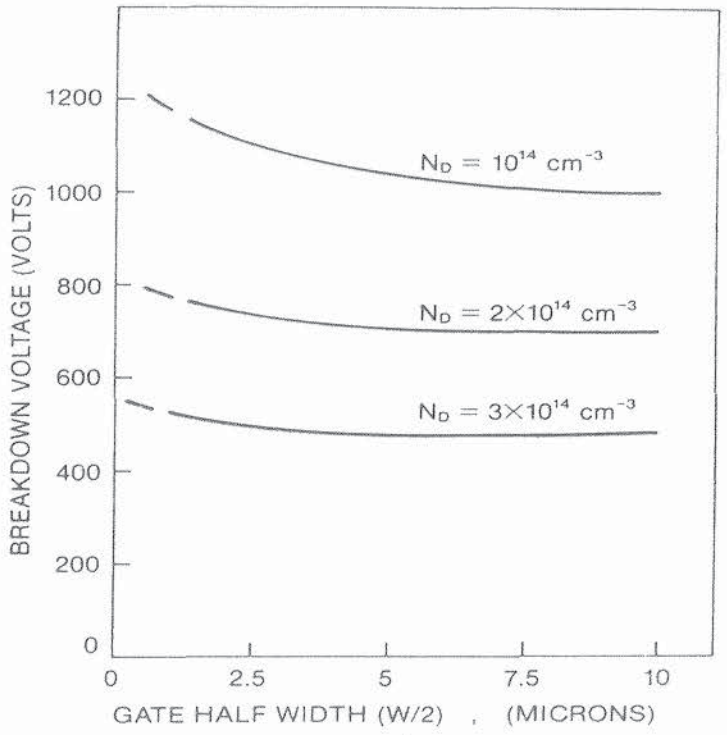


Fig. 6.7. Impact of DMOS cell spacing on the breakdown voltage. (After Ref. 4, reprinted with permission from the IEEE © 1984 IEEE.)

spreading at lower doping levels, which tends to alleviate the junction curvature. In general, a 15% increase in the breakdown voltage can be obtained by bringing the cells close together. However, the cells cannot be placed arbitrarily close together because of an increase in the on-resistance, as discussed later in this chapter.

6.3.1.3. V-Groove Corner. In the case of the VMOS structure shown in Fig. 6.1*b*, the depletion layer from the *P*-base–*N*-drift layer junction expands into the drift region from either side of the V-groove. The depletion layer in the *N*-drift layer extends well beyond the tip of the V groove, as illustrated in Fig. 6.8. A higher electric field is created at the vicinity of the sharp tip of the V groove. The higher local electric fields at the tip of the groove cause avalanche breakdown along the paths indicated in Fig. 6.8 prior to avalanche breakdown in the parallel-plane portion of the junction. To alleviate the increase in electric field at the sharp pointed end of the V groove, a truncated V-groove structure has been developed as illustrated in Fig. 6.9. The truncation of the groove is achieved by using a wider window during etching of the groove and stopping the etch before a sharp point is created. This creates a wide drain overlap region and eliminates the sharply pointed end of the groove. Nevertheless, a discontinuity in the semiconductor surface still remains at point *A*. The impact of this discontinuity on the breakdown voltage has been analyzed by numerical techniques [5]. An example of the reduction in the breakdown voltage due to the groove is provided in Fig. 6.10. As the drain overlap distance x (i.e., the distance over which the gate lies over the drift region) shortens the structure approaches the V-groove case and the breakdown voltage decreases, as shown by the solid line for the VMOS case. In comparison, the breakdown voltage of the DMOS structure (solid line) increases slightly as the drain overlap decreases. These results apply when the gate electrode completely extends between the adjacent cells over the entire drain overlap region.

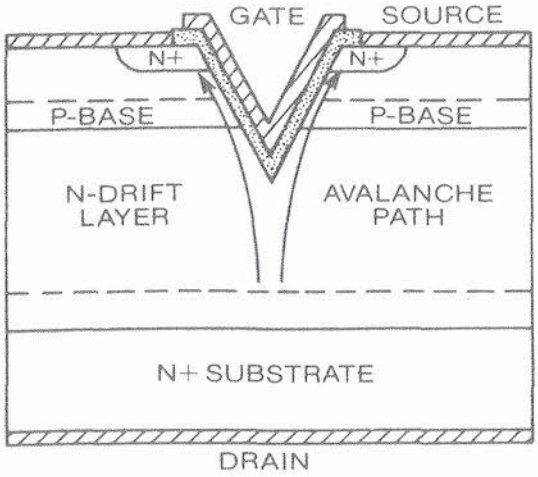


Fig. 6.8. Enhanced avalanche path created in a V-groove power MOSFET structure due to electric field crowding at the sharp point at the end of the groove.

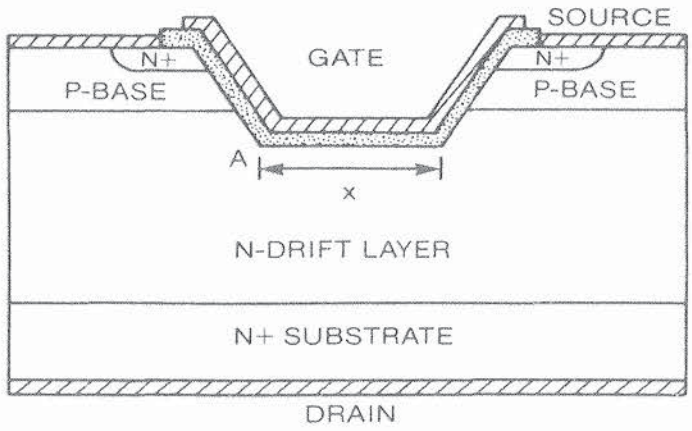


Fig. 6.9. Truncated V-groove power MOSFET structure.

When no overlap of the gate electrode is provided, that is, if the gate extends only to the edge of the junction between the *P*-base and *N*-drift regions, the breakdown voltage is found to rapidly decrease with increasing spacing between the cells for both the DMOS and VMOS structures as shown by the dashed lines in Fig. 6.10. For this reason, it is advisable to let the gate electrode extend completely between the cells during power MOSFET design. An exception to

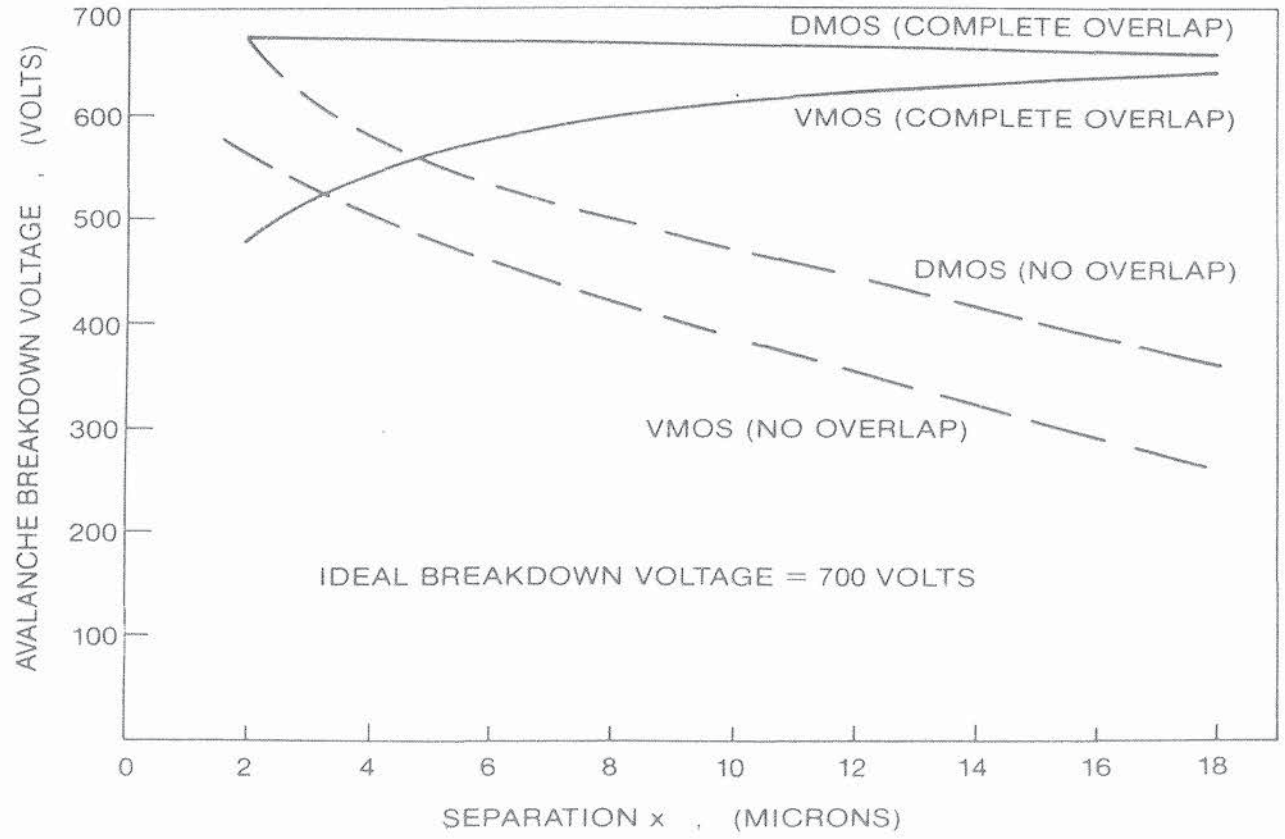


Fig. 6.10. Avalanche breakdown voltages calculated by numerical analysis for VMOS and DMOS power MOSFET structures. (After Ref. 5, reprinted with permission from the IEEE © 1979 IEEE.)

this rule occurs when designing for maximizing the frequency response. The overlap of the gate electrode over the drain region adds significantly to the input capacitance, thus degrading the frequency response. By minimizing the overlap of the gate over the drain and keeping the separation between cells small, it is possible to achieve a reasonable compromise between achieving high breakdown voltage and high frequency response [6]. Another approach to achieving this goal is to use a thicker oxide in the region between the cells.

These results of numerical analysis of the breakdown of VMOS structures indicate that these devices will exhibit a lower avalanche breakdown voltage when compared with DMOS devices for the same epitaxial layer doping level and thickness. This has serious consequences on the minimum on-resistance achievable by use of the VMOS structure. Because of higher specific on-resistance for a given avalanche breakdown voltage for the VMOS structure, most manufacturers have converted to the DMOS structure for commercial fabrication of power MOSFETs.

6.3.2. Forward Conduction Characteristics

Current flow in a power MOSFET during forward conduction is achieved by the application of a positive gate bias voltage for n-channel devices to create a conductive path across the P-base region underneath the gate. The current flow is limited by the total resistance between source and drain. This resistance consists of several components as illustrated in the DMOS cross section in Fig. 6.11. The resistances of the N^+ emitter (R_{N+}) and substrate (R_S) regions are

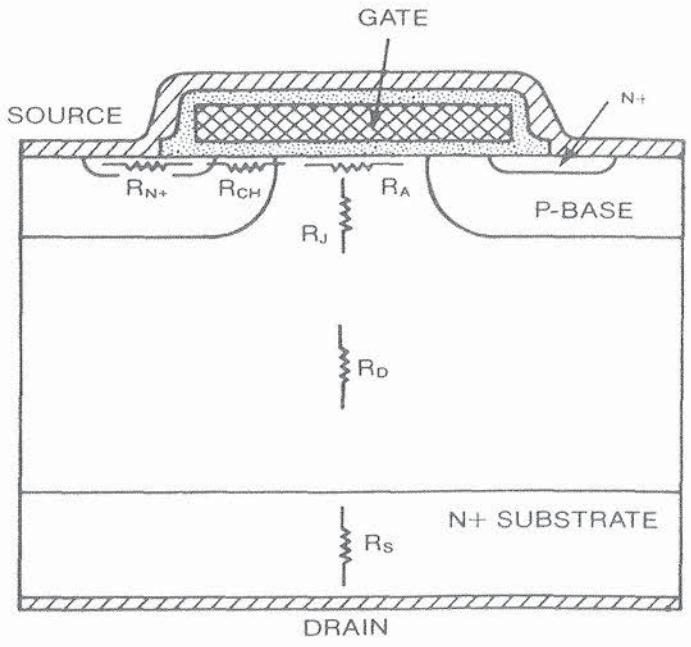


Fig. 6.11. Cross section of DMOS power transistor with internal resistances of the structure indicated.

generally negligible for high-voltage power MOSFETs that have high drift region resistance. They become quite important when the drift and channel resistance become small as in the case of low (<100 V) breakdown voltage devices. The channel (R_{ch}) and accumulation layer (R_A) resistances are determined by the conductivity of the thin surface layer induced by the gate bias. These resistances are a function of the charge in the surface layer and the electron mobility near the surface. The analysis of the charge at the surface is treated in this section. The electron mobility in surface inversion and accumulation layers was treated in Chapter 2.

In addition to these resistances, the drift layer contributes two more components to the total on-resistance. The portion of the drift region that comes to the upper surface between the cells contributes a resistance R_J that is enhanced at higher drain voltages due to the pinch-off action of depletion layers extending from adjacent P-base regions. This phenomenon has been termed the *JFET action*. Finally, the main body of the drift region contributes a large series resistance R_D, especially for high-voltage devices. The analysis of each of these components of the on-resistance is provided in this section.

6.3.2.1. MOS Physics. In this section, the charge created at the surface of the semiconductor as a result of the gate bias applied to the MOS structure is analysed. For this analysis, a one-dimensional structure is assumed with no electric field applied parallel to the surface. To perform the analysis, a P-type semiconductor region is assumed. The treatment is applicable to the analysis of current transport in n-channel power MOSFETs. A similar analysis can be performed for an N-type semiconductor region by appropriate changes in the polarity of the voltage, electric field, and free-carrier charge. In the analysis presented here, the oxide layer is assumed to be a perfect insulator that does not allow the transport of any charge between the gate and the semiconductor.

The energy-band diagrams for an ideal MOS structure with a P-type semiconductor are shown in Fig. 6.12 for four examples of different bias potentials on the metal electrode. Here an ideal MOS structure is defined as one that satisfies the following conditions: (1) the insulator has infinite resistivity, (2) charge can exist only in the semiconductor and on the metal electrode, and (3) there is no energy difference between the work function of the metal and the semiconductor. Under these conditions, there is no band bending in the absence of a gate bias and the band structure assumes the form shown in Fig. 6.12*a*. This is known as the *flatband* condition. In this figure, ϕ_m is the metal work function, χ and χ_o are the semiconductor and oxide electron affinities, ϕ_B is the barrier height between the metal and the oxide, and ψ_B is the potential difference between the intrinsic and Fermi levels in the semiconductor. Under flatband conditions

$$q\phi_m = q\chi + \frac{E_g}{2} + q\psi_B = q\phi_B + q\chi_O \tag{6.2}$$

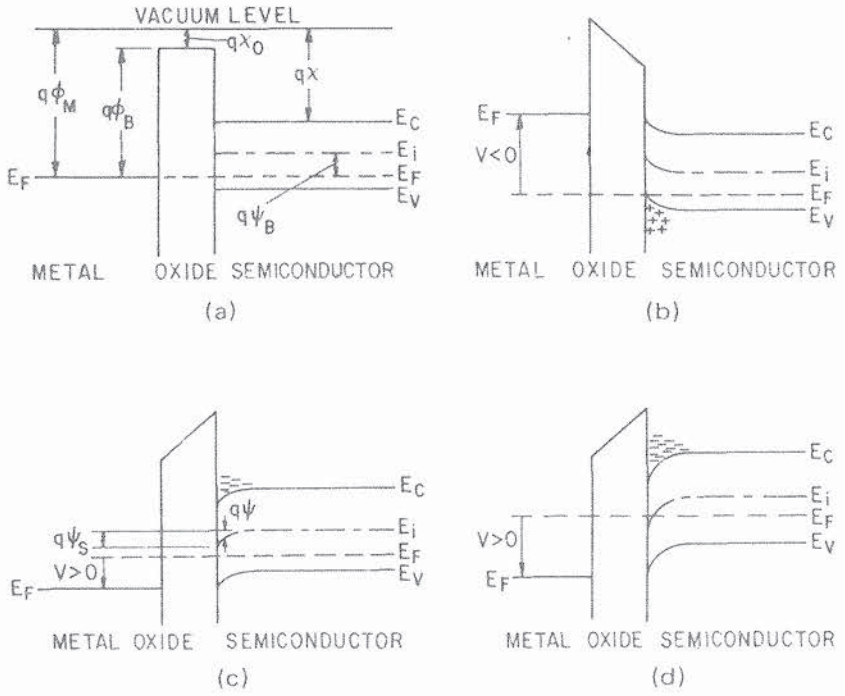


Fig. 6.12. Energy band diagrams for a MOS structure with *P*-type semiconductor region under (*a*) flatband (zero gate bias), (*b*) accumulation (negative gate bias), (*c*) depletion (positive gate bias), and (*d*) inversion (positive gate bias).

When a negative bias is applied to the metal electrode, holes are attracted to the surface, resulting in the bending of the valence band closer to the Fermi level as shown in Fig. 6.12*b*. This condition of excess majority carriers at the surface is referred to as *accumulation*. In contrast, when a positive bias is applied to the metal electrode, the holes are repelled from the surface. At small negative baises, this results in the bending of the valence band away from the Fermi level, resulting in the formation of a surface depletion layer as shown in Fig. 6.12*c*. At larger negative biases on the metal, the band bending increases and the intrinsic level crosses the Fermi level as shown in Fig. 6.12*d*. At this point, the density of electrons at the surface exceeds the density of holes (majority carriers) in the bulk. This condition is defined as *inversion*.

The charge in the inversion layer plays a key role in the determination of current transport in MOSFET devices. To determine this charge, the potential distribution $\psi(x)$ in the semiconductor must first be determined. The electron and hole concentrations can then be obtained from

$$n_p = n_{p0}\, e^{q\psi/kT} \tag{6.3}$$

and

$$p_p = p_{p0}\, e^{-(q\psi/kT)} \tag{6.4}$$

where n_{p0} and p_{p0} are, respectively, the equilibrium electron and hole concentrations in the bulk. The potential distribution $\psi(x)$ can be derived by solving

Poisson's equation:

$$\frac{d^2\psi}{dx^2} = -\frac{\rho(x)}{\epsilon_s} \tag{6.5}$$

where the charge density is given by

$$\rho(x) = q(N_D^+ - N_A^- + p_p - n_p) \tag{6.6}$$

The solution of the Poisson's equation must satisfy the boundary condition that charge neutrality exists in the bulk of the semiconductor far from the surface. In the bulk, charge neutrality requires

$$(N_D^+ - N_A^-) = (n_{p0} - p_{p0}) \tag{6.7}$$

Using Eqs. (6.3), (6.4), and (6.6) in Poisson's Eq. (6.5), it can be shown that:

$$\frac{d^2\psi}{dx^2} = -\frac{q}{\epsilon_s}[p_{p0}(e^{-(q\psi/kT} - 1) - n_{p0}(e^{(q\psi/kT)} - 1)] \tag{6.8}$$

Integration of this equation from the bulk toward the surface gives the electric field distribution [7]:

$$\mathscr{E}(x) = -\frac{d\psi}{dx} = \frac{\sqrt{2}\,kT}{qL_D} F\left(\frac{q\psi}{kT}, \frac{n_{p0}}{p_{p0}}\right) \tag{6.9}$$

where

$$L_D = \sqrt{\frac{kT}{q^2}\frac{\epsilon_s}{p_{p0}}} \tag{6.10}$$

is called the *extrinsic Debye length for holes*, and

$$F\left(\frac{q\psi}{kT}, \frac{n_{p0}}{p_{p0}}\right) = \left\{\left[e^{-(q\psi/kT)} + \left(\frac{q\psi}{kT}\right) - 1\right] + \frac{n_{p0}}{p_{p0}}\left[e^{(q\psi/kT)} - \left(\frac{q\psi}{kT}\right) - 1\right]\right\}^{1/2} \tag{6.11}$$

The space charge required per unit area to create this electric field can be obtained from Gauss's law:

$$Q_S = -\epsilon_s \mathscr{E}_s \tag{6.12}$$

where $\mathscr{E}_s$ is the surface electric field. If the surface potential is called ψ_s, the surface charge can be obtained from Eqs. (6.9) and (6.12):

$$Q_S = \frac{\sqrt{2}\,\epsilon_s kT}{qL_D} F\left(\frac{q\psi_s}{kT}, \frac{n_{p0}}{p_{p0}}\right) \tag{6.13}$$

The variation of the surface charge with shift in the surface potential is illustrated in Fig. 6.13.

For negative values of the surface potential, the bands are bent upward near the surface, creating a positively charged surface accumulation layer. In

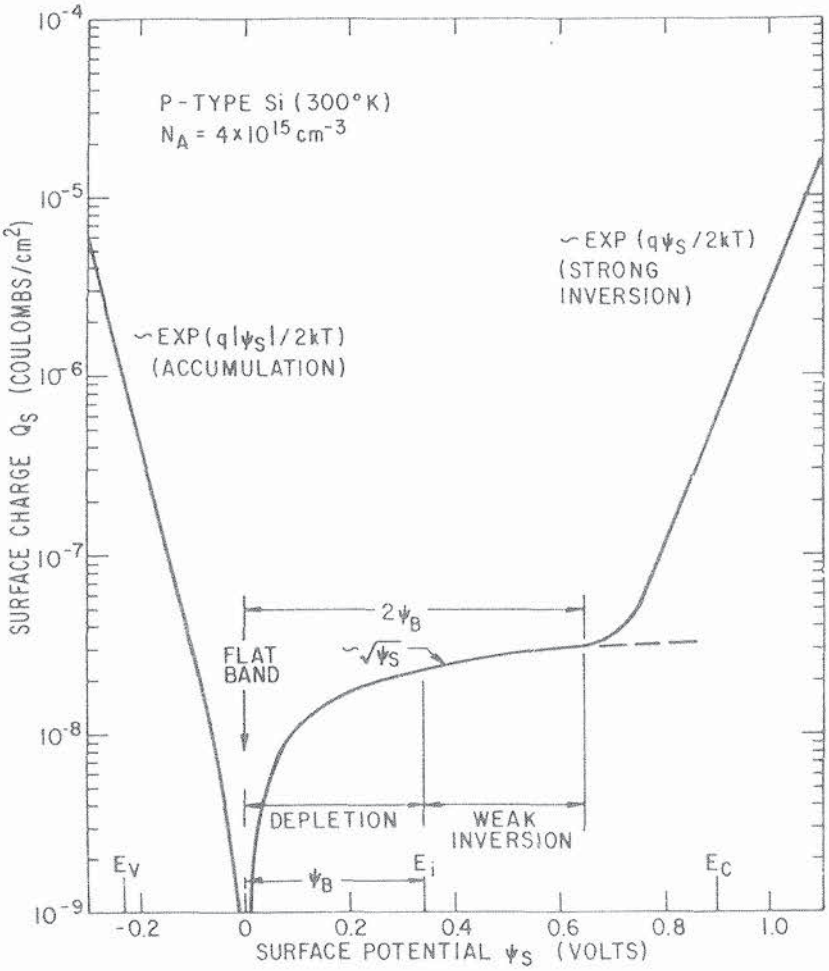


Fig. 6.13. Dependence of the surface space-charge density on the surface potential for *P*-type silicon.

this regime of operation, the charge in the accumulation layer varies exponentially with the surface potential because the first term on the right-hand side of Eq. (6.11) becomes dominant:

$$Q_S \simeq \frac{\sqrt{2}\,\epsilon_s kT}{qL_D}\, e^{(q|\psi_s|/kT)} \tag{6.14}$$

where $|\psi_S|$ is the magnitude of the surface potential. As the magnitude of ψ_S decreases, the accumulation layer charge decreases exponentially until it becomes equal to zero at the flatband condition shown in Fig. 6.12*a*. Note that under accumulation conditions, the electric field is essentially confined to the oxide and the entire gate voltage is supported by the oxide.

When the surface potential increases from zero in the positive direction, the bands are bent downward, as illustrated in Fig. 6.12*c*. As long as the surface potential ψ_S remains below the bulk value ψ_B, the intrinsic level E_i does not cross the Fermi level. The surface is then still p-type and contains a depletion layer. The negative charge in the depletion layer can also be calculated from Eq. (6.13). In this case the second term of the function F [Eq. (6.11)] becomes

dominant and the surface charge is given by

$$Q_S \simeq \frac{\epsilon_s}{L_D}\sqrt{2\frac{kT}{q}\psi_S} \tag{6.15}$$

Using Eq. (6.10) for L_D, we obtain

$$Q_S \simeq \sqrt{2\epsilon_s q p_{p0}\psi_S} \tag{6.16}$$

The surface charge increases gradually as the square root of the surface potential as indicated in Fig. 6.13.

When the surface potential ψ_S exceeds the bulk value ψ_B, the intrinsic level E_i crosses the Fermi level. At this point, a mobile negative charge begins to form at the surface. The concentration of this charge is small as long as ψ_S is less than $2\,\psi_B$. This regime is called *weak inversion.* In this region of operation, the surface charge continues to rise as the square root of the surface potential. When the surface potential exceeds $2\psi_B$, the surface charge begins to rise rapidly as shown on the right-hand side of Fig. 6.13. This occurs because the fourth term in Eq. (6.11) for the function F now becomes dominant. This regime of operation is known as *strong inversion.* It occurs when

$$\psi_S > 2\psi_B = 2\frac{kT}{q}\ln\left(\frac{N_A}{n_i}\right) \tag{6.17}$$

The surface charge under strong inversion can be derived from Eq. (6.13) by assuming that the fourth term in Eq. (6.11) is dominant:

$$Q_S \simeq \sqrt{\frac{2n_{p0}}{p_{p0}}}\frac{\epsilon_s kT}{qL_D}e^{q\psi_s/2kT} \tag{6.18}$$

Substituting for L_D from Eq. (6.10), this expression can be rewritten as:

$$Q_S \simeq \sqrt{2\epsilon_s n_{p0}}\,e^{(q\psi_s/2kT)} \tag{6.19}$$

In the strong inversion region, the surface charge again increases exponentially with surface potential. This charge is an important quantity because it determines the conductivity of the channel in power MOSFETs.

In the depletion region of operation, the applied gate voltage is shared across the gate oxide and the semiconductor depletion layer. With the onset of strong inversion, any further increase in the gate voltage is dropped across the oxide and the voltage across the semiconductor depletion layer remains essentially constant. As a consequence, the depletion layer width in the semiconductor expands until strong inversion occurs, and then remains constant with further increase in gate voltage. The maximum depletion layer width is approximately given by

$$W_m \simeq \sqrt{\frac{2\epsilon_s}{qN_A}(2\psi_B)} \tag{6.20}$$

Use of Eq. (6.17) for ψ_B gives

$$W_m \simeq \sqrt{\frac{4\epsilon_s kT}{q^2 N_A} \ln\left(\frac{N_A}{n_i}\right)} \tag{6.21}$$

A plot of the variation in the maximum depletion depth with background doping concentration is provided in Fig. 6.14. It is worth pointing out that the maximum depletion layer width for the MOS structure is considerably smaller than the maximum depletion layer width in the semiconductor determined by avalanche breakdown.

6.3.2.2. Threshold Voltage. The voltage on the gate electrode at which strong inversion begins to occur in the MOS structure is an important design parameter for power MOSFETs because it determines the minimum gate bias required to induce an *n*-type conductance in the channel. This voltage is called the *threshold voltage*. For proper device operation, its value can be neither too large nor too small. If the threshold voltage is large, a high gate bias voltage will be needed to turn on the power MOSFET. This imposes problems in the design of the gate drive circuitry. It is also important that the threshold voltage not be too low. Because of the existence of charge in the gate oxide, it is possible for the threshold voltage to be negative for *n*-channel power MOSFETs.

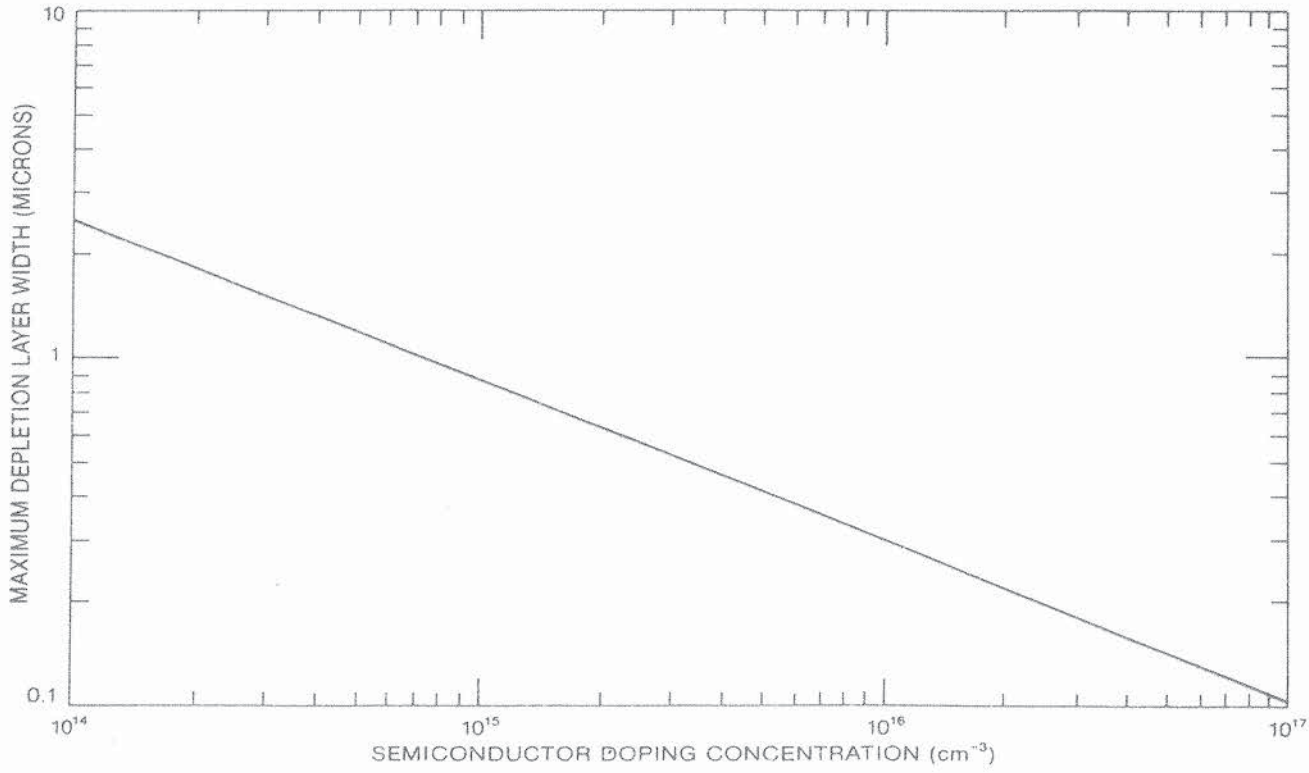


Fig. 6.14. Maximum depletion layer width of an MOS structure as a function of the doping level of the semiconductor.

This is an unacceptable condition because a conductive channel will now exist at zero gate bias voltage; that is, the devices will exhibit normally-on characteristics. Even if the threshold voltage is above zero for an *n*-channel power MOSFET, its value should not be too low because the device can then be inadvertently triggered into conduction either by noise signals at the gate terminal or by the gate voltage being pulled up during high-speed switching. Typical power MOSFET threshold voltages are designed to range between 2 and 3 V.

In the absence of any difference in the work function of the metal and the semiconductor, the applied gate voltage is given by

$$V_G = V_{ox} + \psi_S \tag{6.22}$$

where V_{ox} is the voltage across the oxide. This voltage is related to the surface charge:

$$V_{ox} = \mathscr{E}_{ox} t_{ox} \tag{6.23}$$

$$= \frac{Q_S}{\epsilon_{ox}} t_{ox} \tag{6.24}$$

$$= \frac{Q_S}{C_{ox}} \tag{6.25}$$

At the point of transition into the strong inversion condition, the gate bias voltage is equal to the threshold voltage:

$$V_G = V_T = \frac{Q_S}{C_{ox}} + 2\psi_B \tag{6.26}$$

Using Eq. (6.16) for Q_S at the point of transition into strong inversion, it can be shown that

$$V_T \simeq \frac{\sqrt{4\epsilon_s q N_A \psi_B}}{C_{ox}} + 2\psi_B \tag{6.27}$$

$$\simeq \frac{\sqrt{4\epsilon_s kT N_A \ln(N_A/n_i)}}{(\epsilon_{ox}/t_{ox})} + \frac{2kT}{q} \ln\left(\frac{N_A}{n_i}\right) \tag{6.28}$$

From this expression, it can be concluded that the threshold voltage will increase linearly with gate oxide thickness and approximately as the square root of the semiconductor doping. The calculated threshold voltage for silicon MOS structures with gate oxide thicknesses ranging from 100 to 10,000 Å are provided in Fig. 6.15 for doping levels ranging from 10^{14} to $10^{18}/cm^3$. These curves apply to a uniformly doped semiconductor and do not take into account the presence of any charge in the oxide.

In actual MOS structures, the threshold voltage is altered due to (1) an unequal work function for the metal and the semiconductor—if the barrier

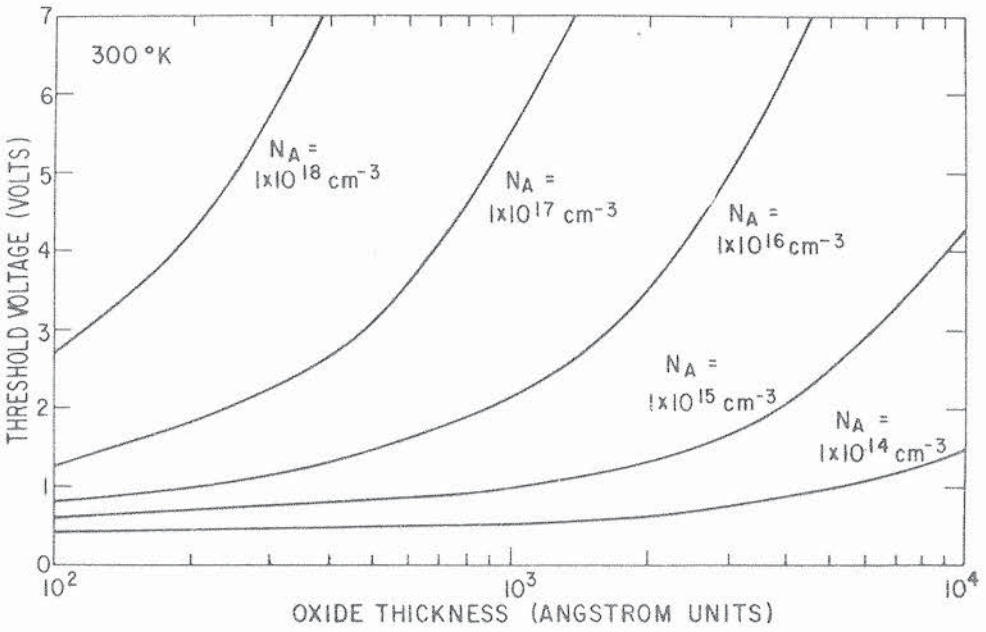


Fig. 6.15. Threshold voltage of ideal silicon MOS structures as a function of oxide thickness.

height between SiO_2 and the metal is ϕ_B, the difference between the metal and the semiconductor work function can be obtained as

$$q\phi_{ms} = q\phi_B + q\chi_0 - \left(q\chi + \frac{E_g}{2} + q\psi_B\right) \tag{6.29}$$

(2) the presence of fixed surface charge Q_{FC} at the oxide–silicon interface, (3) the presence of mobile ions in the oxide with charge Q_I, and (4) the presence of charged surface states at the oxide–silicon interface with charge Q_{SS}. All these charges cause a shift in the threshold voltage:

$$V_T = \phi_{ms} + 2\psi_B - \left(\frac{Q_S + Q_{SS} + Q_I + Q_{FC}}{C_{ox}}\right) \tag{6.30}$$

The fixed surface charge Q_{FC} is located within 20 Å of the oxide–silicon interface and cannot be charged or discharged over a wide range of surface potentials such as those normally encountered in device operation. The density of this charge is dependent on the oxide growth conditions and its subsequent heat treatment. The presence of excess silicon or a deficiency of oxygen has been postulated as the origin of the fixed charge, but the influence of impurities cannot be excluded.

The presence of mobile ions (charge Q_I) has been related to the shift in the threshold voltage under bias stressing at elevated temperatures or under long periods. Sodium ions are the most prominent mobile species in the oxide. The concentration of these ions is dependent on the cleanliness of the wafers during processing and the purity of the gate metal.

Interface states (charge Q_{SS}) are energy levels in the silicon band gap located close to the surface that can charge and discharge with changes in the surface potential. The origin of interface states on a clean free surface in vacuum has been shown to arise from unsaturated bonds at the surface. The interface states at the $Si-SiO_2$ boundary are usually several orders of magnitude lower (10^{10}–$10^{12}/cm^2$) than on the free surface. The origin of these surface states is not completely understood, but the effect of processing on their density has been extensively studied. This is discussed in the section on device fabrication.

The dependence of the threshold voltage of an aluminum gate MOS capacitor fabricated on a 1000- and 15,000-Å oxide layer grown on *P*-type silicon is provided in Fig. 6.16 as a function of the doping level. The 1000-Å oxide thickness is typically used to fabricate the MOS gate of the FETs, whereas the 15,000-Å oxide thickness represents the thick field oxide in the inactive portions of the device where the influence of the gate bias must be minimized. In these figures, curves are shown for various values of the charge in the oxide that may arise from either fixed positive interface charge or positive ions at the oxide–silicon interface or from positively charged surface states. For a typical peak channel doping on the order of 5×10^{16} atoms/cm^3 and a charge density of $10^{11}/cm^2$, a threshold voltage at 2–3 V can be obtained with an oxide thickness of 1000 Å. This threshold voltage is convenient for controlling the device with integrated drive circuits. For this doping level and interface charge, the threshold voltage of the field oxide is over 30 V, which provides sufficient margin for overdriving the gate without inverting the regions outside the device active area.

As indicated by Eq. (6.30), the threshold voltage for inversion in an MOS capacitor is affected by the barrier height between the metal and the oxide.

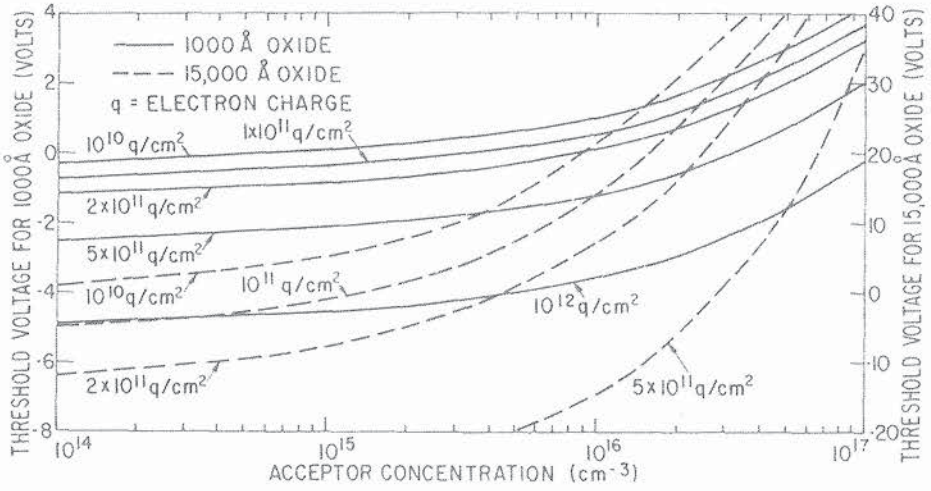


Fig. 6.16. Threshold voltage as a function of *P*-base doping level for the case of a 1000-Å gate oxide and 15,000-Å field oxide. Curves are shown for charge in the oxide ranging from 10^{10} to $10^{12}/cm^2$.

TABLE 6.1. Barrier Heights of Metals on Silicon Dioxide

Metal	Ag	Al	Au	Cu	Mg	Ni	Pd	Pt	W
Barrier height ϕ_m (eV)	4.2	3.2	4.1	3.8	2.3	3.7	4.2	4.4	3.6

Table 6.1 provides a list of metals and the measured barrier height between them and silicon dioxide [8]. As can be seen from this table, the threshold voltage of a device can be tailored to a limited extent by choosing the proper metal.

A greater control over the threshold voltage can be achieved by substituting polycrystalline silicon for the metal. By doping the polycrystalline silicon either N or P type, the work function difference in the poly Si–SiO_2–Si system can be varied. Measurements have shown that the Fermi level in heavily doped polycrystalline silicon corresponds exactly to that in monocrystalline silicon [9]. On this basis, the work function difference between N-type polysilicon and P-type monocrystalline silicon is given by the expression

$$q\phi_{\mathrm{PS-Si}} = \frac{kT}{q} \ln\left(\frac{N_{\mathrm{DPS}} N_{\mathrm{A}}}{n_i^2}\right) - E_g \tag{6.31}$$

where N_{A} is the substrate doping concentration and N_{DPS} is the donor concentration in the polycrystalline silicon.

Most power MOSFETs are fabricated by using heavily doped polycrystalline silicon as the gate electrode. This provides a refractory gate material that is compatible with silicon device processing and allows the fabrication of large-area devices with a two-level "metal" process as illustrated in Fig. 6.11. The conductivity of polycrystalline silicon is an order of magnitude lower than that for metals. In very-high-frequency power MOSFETs, the RC gate charging time constant becomes too large for polycrystalline silicon gate electrodes. For these devices, aluminum has been used as the gate electrode, with source and gate metal being interdigitated on the wafer surface. With the development of refractory gate silicide or polycide technology for integrated circuits, it is expected that very-high-frequency power MOSFETs with two-level electrode structures will be forthcoming.

6.3.2.3. Channel Conductance. Consider a MOSFET having a P-type base region with N^+ source and drain located on either side as shown in Fig. 6.17. The current flow between source and drain is controlled by the electron charge available for transport in the surface inversion layer of the P base as well as the surface mobility of these electrons. According to the MOS analysis, when the surface potential ψ_S exceeds twice the bulk potential ψ_B, a strong inversion layer begins to form. Since the band bending is small beyond this point, the

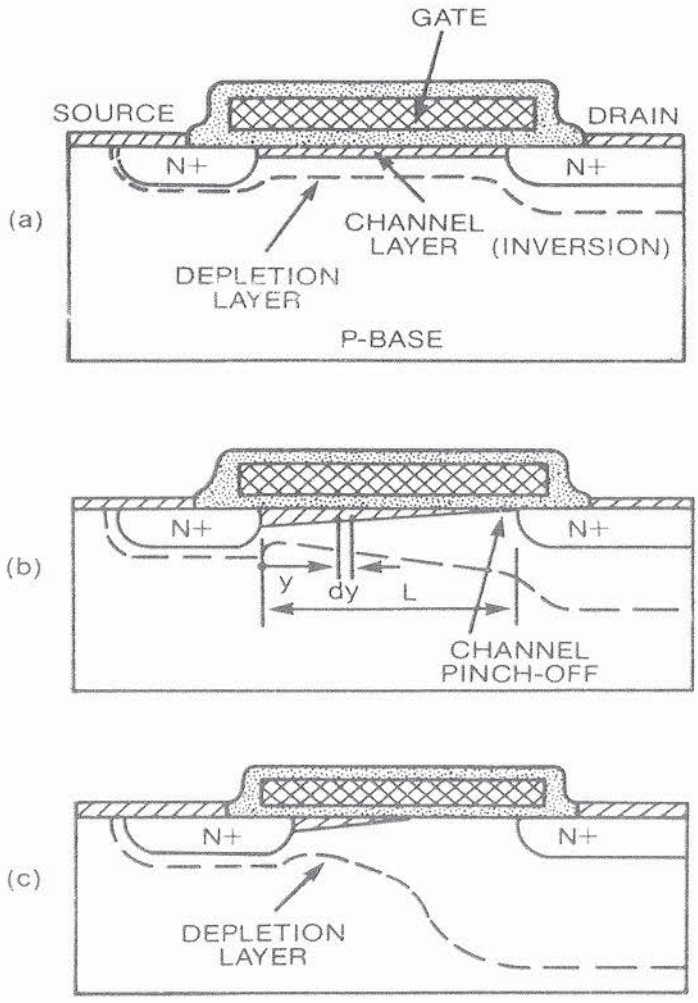


Fig. 6.17. Basic MOS transistor with inversion layer charge transport.

inversion layer charge available for current conduction is

$$Q_n \simeq C_{ox}(V_G - V_T) \tag{6.32}$$

Thus the channel resistance at low drain voltages, where the voltage drop along the channel is negligible, is given by

$$R_{ch} = \frac{L}{Z\mu_{ns}C_{ox}(V_G - V_T)} \tag{6.33}$$

where Z and L are the width and length of the channel, respectively, and μ_{ns} is the surface mobility of electrons. The surface mobility and its dependence on doping, temperature, and surface orientation are discussed extensively in Chapter 2 because of their importance in determining the conductance of the channel in the MOSFET.

As the drain current increases, the voltage drop along the channel between drain and source becomes significant. The positive drain potential opposes the

gate bias voltage and reduces the surface potential in the channel. The channel charge near the drain is reduced by the voltage drop along the channel as illustrated in Fig. 6.17*b*. When the drain voltage becomes equal to $(V_G - V_T)$, the charge in the channel at the drain end becomes equal to zero. This condition is called *channel pinch-off*. At this point, the drain current saturates. Further increase in drain voltage results in no further increase in drain current. The drain voltage is now supported across an extension of the drain under the gate as shown by Fig. 6.17*c*, resulting in a reduction in the effective channel length.

The channel resistance can be derived as a function of the gate and drain voltages under the following assumptions: (1) the gate structure is an ideal MOS structure as discussed earlier; (2) the free-carrier mobility is a constant, independent of the electric field strength; (3) the base region is uniformly doped; (4) current transport occurs exclusively by drift; (5) the leakage current is negligible; and (6) the longitudinal electric field along the surface is small compared with the transverse electric field resulting from the gate bias. Assumption 6 allows treatment of an elemental segment of the channel dy at location y under conditions similar to the gradual channel approximation used to analyze the impedance of JFETs in Chapter 4.

The resistance of an elemental segment dy of the channel is dependent on the inversion layer charge per unit area and the mobility of the free carriers:

$$dR = \frac{dy}{Z\mu_{ns}Q_n(y)} \tag{6.34}$$

The charge in the inversion layer depends not only on the gate voltage, but also on the drain current because of the potential drop along the channel:

$$Q_n(y) = C_{ox}[V_{GS} - V_T - V(y)] \tag{6.35}$$

where $V(y)$ is the voltage drop along the channel. The voltage drop in segment dy is given by

$$dV = I_D\, dR \tag{6.36}$$

From these expressions

$$\int_0^L I_D\, dy = -Z\mu_{ns}C_{ox}\int_0^{V_D}(V_G - V_T - V)\, dV \tag{6.37}$$

Since the drain current I_D must remain constant throughout the channel:

$$I_D = \frac{\mu_{ns}C_{ox}Z}{2L}[2(V_G - V_T)V_D - V_D^2] \tag{6.38}$$

When the drain voltage is small, the drain current increases linearly with drain voltage:

$$I_D \simeq \mu_{ns}C_{ox}\frac{Z}{L}(V_G - V_T)V_D \tag{6.39}$$

In this linear region, the channel resistance is given by

$$R_{ch} = \frac{L}{Z\mu_{ns}C_{ox}(V_G - V_T)} \tag{6.40}$$

which agrees with Eq. (6.33), representing a homogeneous inversion layer extending between drain and source. In this linear regime, the transconductance can be obtained by differentiating Eq. (6.39) with respect to the gate voltage:

$$g_m = \frac{dI_D}{dV_G} = \mu_{ns}C_{ox}\frac{Z}{L}V_D \tag{6.41}$$

As the drain voltage and current increase, the second term in Eq. (6.38) becomes increasingly important, resulting in an increase in the channel resistance. Physically, this corresponds to the reduction in the channel inversion layer charge near the drain with increasing drain voltage. Ultimately, the channel resistance becomes infinite, corresponding to drain current saturation. The drain voltage at which current saturation occurs is called the *channel pinch-off point*:

$$V_{DS} = (V_G - V_T) \tag{6.42}$$

Substituting for V_{DS} in Eq. (6.38), an expression for the saturated drain current can be derived:

$$I_{DS} = \frac{\mu_{ns}C_{ox}}{2}\frac{Z}{L}(V_G - V_T)^2 \tag{6.43}$$

The saturated drain current is an important parameter because it determines the maximum current that the channel will support. In actual devices, the saturated drain current will be lower than that indicated by Eq. (6.43) because the surface mobility μ_{ns} is a function of the longitudinal electric field. The transconductance of the device in the saturated current region of operation is given by

$$g_{ms} = \mu_{ns}C_{ox}\frac{Z}{L}(V_G - V_T) \tag{6.44}$$

The basic features of the current–voltage characteristics of a MOSFET as described by Eq. (6.38) are shown in Fig. 6.18.

According to the preceding analysis, the drain current will saturate above a certain drain voltage. In actual practice, the drain current does not remain constant in the saturation region because when the drain voltage increases beyond V_{DS}, the length of the channel is reduced as illustrated in Fig. 6.17(*c*). This decreases the effective channel length and produces a finite drain output conductance. For devices with higher P-base doping levels, the channel shrinkage ΔL can be approximated by treating the gate–drain depletion region like a N^+–P step junction [10]:

$$\Delta L = \sqrt{\frac{2\epsilon_s}{qN_A}(V_D - V_{DS})} \tag{6.45}$$

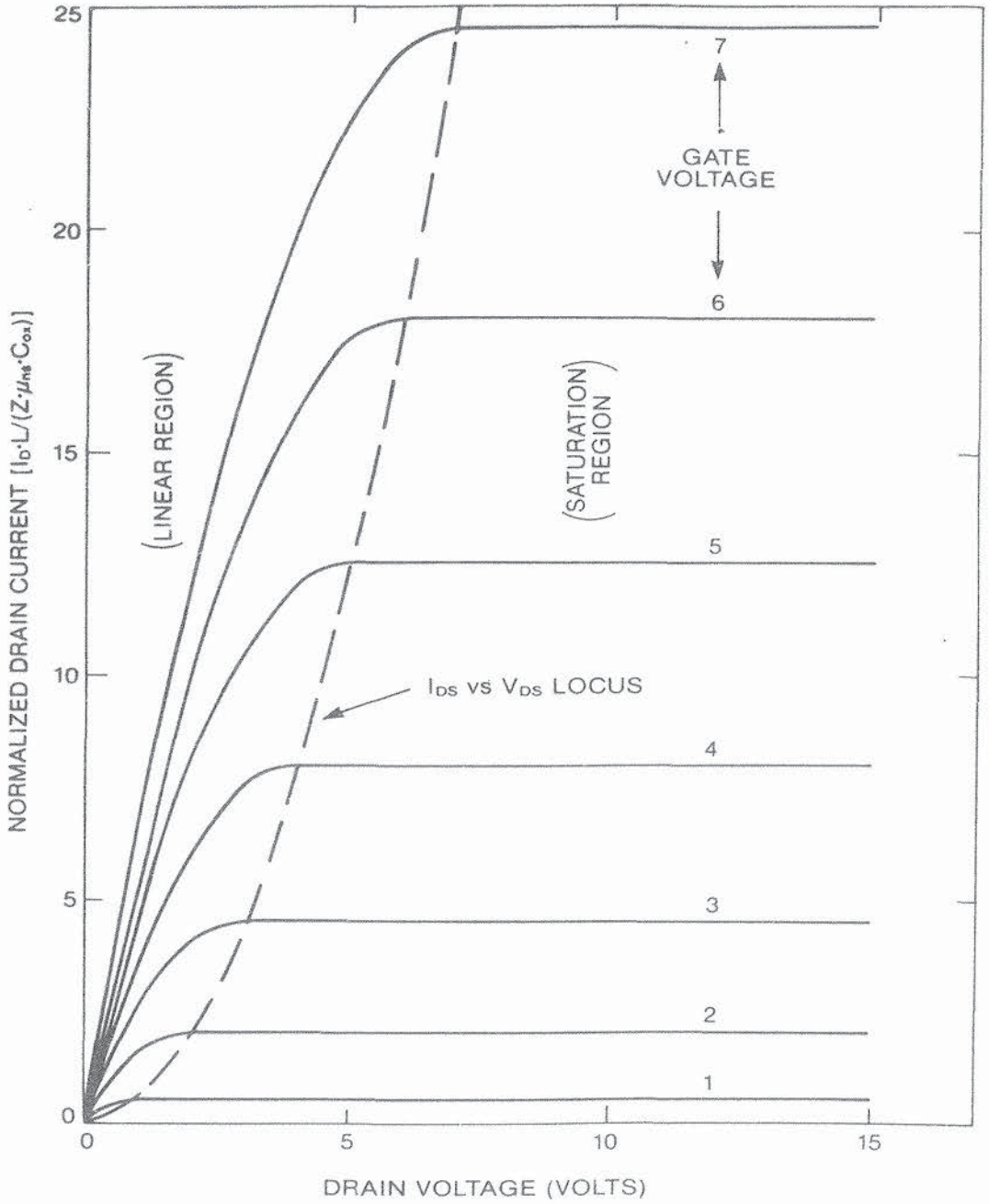


Fig. 6.18. Fundamental current–voltage characteristics of a MOSFET.

The drain output resistance can then be derived:

$$r_{ds} = 12\left[L - \sqrt{\frac{2\epsilon_s(V_D - V_{DS})}{qN_A}}\right]^2 \sqrt{V_D - V_{DS}} \times \left\{Z\mu_{ns}C_{ox}\sqrt{\frac{2\epsilon_s}{qN_A}}\left[(V_{DS} + 2\phi_B)^2 + 2V_G(V_{DS} + 2\phi_B) - 12\,\phi_B\left\{\left(V_G - \phi_B - \frac{4}{3}\sqrt{\frac{\epsilon_s qN_A\phi_B}{C_{ox}^2}}\right)\right]\right\}^{-1} \tag{6.46}$$

A fairly good agreement between the calculated drain output conductance with this expression and the measured output conductance of n-channel devices has been reported. Better results can only be obtained by resorting to a two-dimensional numerical analysis of current transport in these devices [11, 12].

6.3.2.4 On-Resistance. The on-resistance of a power MOSFET is the total resistance between the source and drain terminals in the on-state. The on-resistance is an important device parameter because it determines the maximum current rating. The power dissipation in the power MOSFET during current conduction is given by

$$P_D = I_D V_D = I_D^2 R_{on} \tag{6.47}$$

Expressed in terms of the chip area A, this is

$$\frac{P_D}{A} = J_D^2 R_{on,sp} \tag{6.48}$$

where (P_D/A) is the power dissipation per unit area; J_D is the on-state current density; and $R_{on,sp}$ is the specific on-resistance, defined as the on-resistance per unit area. These expressions are based on the assumption that the power MOSFET is operated in its linear region (low V_D) during current conduction. The maximum power dissipation per unit area is determined by the maximum allowable junction temperature and the thermal impedance. For typical power packages, the maximum power dissipation per unit area is in the range of 100 W/cm^2. The maximum operating current density will thus vary inversely as the square root of the specific on-resistance.

The specific on-resistance of the power MOSFET is determined by several components as illustrated in Fig. 6.11 for the DMOS structure:

$$R_{on} = R_{N+} + R_{ch} + R_A + R_J + R_D + R_S \tag{6.49}$$

where R_{N+} is the contribution from the N^+-source diffusion, R_{ch} is the channel resistance, R_A is the accumulation layer resistance, R_J is the contribution from the drift region between the P-base regions, R_D is the drift region resistance, and R_S is the substrate resistance. Additional resistances can arise from poor contact between the source–drain metal and the N^+ semiconductor regions as well as from the leads used to connect the device to the package.

The contributions from the N^+ source and drain regions are generally negligible for high-voltage power MOSFETs. For low-voltage devices with breakdown voltages below 50 V, however, these regions can contribute significantly to the on-resistance. This is especially true for the substrate because it must be thick to impart adequate strength to the wafers during device fabrication. For a typical substrate with thickness of 20 mils and resistivity of 0.01 Ω·cm, the substrate resistance per unit area is then given by

$$R_{S,sp} = 0.01\ \Omega\cdot\text{cm} \times 5 \times 10^{-2}\ \text{cm} = 5 \times 10^{-4}\ \Omega\cdot\text{cm}^2 \tag{6.50}$$

This resistance can be reduced by using 0.001 $\Omega \cdot$cm arsenic-doped wafers and by lapping the substrate after device fabrication.

To analyze the contribution of the N^+-source resistance, consider the DMOS cell structure shown in Fig. 6.19. In this cell, $2m$ is the cell diffusion window and L_G is the length of the gate electrode between the adjacent cells. The cell diffusion window $2m$ is determined by the photolithographic design tolerances used for device fabrication. These tolerances also determine the N^+-emitter length L_E. If a linear cell is considered with a 1-cm extension perpendicular to the cross section shown in Fig. 6.19, the resistance per square centimeter due to the N^+ emitter regions is given by

$$R_{N^+,sp} = \frac{1}{2}\rho_{\square N^+} L_E(L_G + 2m) \tag{6.51}$$

where $\rho_{\square N^+}$ is the sheet resistance of the N^+ diffusion. In a typical device, $\rho_{\square N^+} = 10\ \Omega \cdot$cm, $L_E = 10\ \mu$m, and the cell repeat spacing $(L_G + 2m) = 40\ \mu$m. The resulting specific resistance of the N^+-emitter regions ($2 \times 10^{-5}\ \Omega \cdot \text{cm}^2$) is negligible compared with all other resistances in the structure.

The channel resistance was discussed in the previous section for the basic MOSFET structure. The contribution from the channel depends on the ratio (L/Z), the gate oxide thickness (through C_{ox}), and the gate drive voltage V_G. The contribution from the channel can be minimized by making the channel length L small and keeping its width Z large. This requires a high cell density and good control over the P-base and N^+-emitter diffusion profiles to keep the channel short without causing punch-through breakdown as discussed earlier in Section 6.3.1. To calculate the contribution from the channel, consider the DMOS cell structure shown in Fig. 6.19. The channel resistance per square centimeter for the linear cell structure is given by

$$R_{ch,sp} = \frac{1}{2}\frac{L(L_G + 2m)}{\mu_{ns} C_{ox}(V_G - V_T)} \tag{6.52}$$

For a typical device with channel length L of 2 μm, a gate oxide thickness of 1000 Å, and a gate drive of 10 V, the channel resistance per square centimeter is found to be $2.5 \times 10^{-3}\ \Omega \cdot \text{cm}^2$. Note that the channel resistance decreases when the cell repeat spacing is reduced. The channel resistance can also be reduced by decreasing the gate oxide thickness while maintaining the gate drive voltage.

The resistance of the accumulation layer R_A determines the current spreading from the channel into the drift region. The accumulation layer resistance is dependent on the charge in the accumulation layer and the mobility for free carriers at the accumulated surface. For the linear cell geometry, the accumulation layer resistance per square centimeter is

$$R_A = \frac{K(L_G - 2x_p)(L_G + 2m)}{\mu_{nA} C_{ox}(V_G - V_T)} \tag{6.53}$$

where K is a factor introduced to account for the two-dimensional nature of the current flow from the accumulation layer into the bulk. Good agreement with experimental results has been observed for $K = 3$ [13]. For the above example, with $L_G = 20\ \mu m$ and $x_p = 3\ \mu m$, the accumulation resistance per square centimeter is found to be $6 \times 10^{-3}\ \Omega \cdot cm^2$. The accumulation layer resistance can be reduced by decreasing the length L_G of the gate electrode between the cells. However, this has an adverse effect on the resistance R_J of the JFET region.

The resistance of the drift region between the P-base diffusions can be calculated if the voltage drop along the vertical direction is neglected. Under the assumption that the depletion layer width of the P-base–N-drift layer junction is negligibly small and the current is flowing uniformly down from the accumulation layer into the JFET region, the resistance of the JFET region can be analyzed as a resistance with increasing cross section when proceeding downward from the surface. The cross section of the JFET region is

$$L_J(x) = L_G - 2\sqrt{x_p^2 - x^2} \tag{6.54}$$

The resistance of a small segment dx of the JFET region per square centimeter is then

$$dR_J = \frac{\rho_D(L_G + 2m)\,dx}{L_J(x)} \tag{6.55}$$

Using Eq. (6.54) and integrating from $x = 0$ to point x_A [13], the total JFET region resistance is obtained:

$$R_J = 2\rho_D(L_G + 2m)\left[\frac{1}{\sqrt{1 - (2x_p/L_G)^2}}\tan^{-1}(0.414)\sqrt{\frac{L_G + 2x_p}{L_G - 2x_p}} - \frac{\pi}{8}\right] \tag{6.56}$$

where ρ_D is the resistivity of the drift region. In deriving this expression, it has been assumed that the JFET region extends to a depth x_A at which angle $\theta = 45°$.

The drift region is assumed to begin beyond this point. The current spreading into the drift region is shown by the dashed lines in Fig. 6.19. The cross section of the drift region increases when proceeding down from the JFET region. For the current spreading at an angle α shown in Fig. 6.19, the drift region resistance per square centimeter is given by [13]

$$R_D = \rho_D \frac{(L_G + 2m)}{\tan(\alpha)} \ln\left[1 + 2\frac{h}{a}\tan(\alpha)\right] \tag{6.57}$$

where the dimensions h and a are indicated in Fig. 6.19. It has been found that a good approximation for α is provided by [13]

$$\alpha = 28° - \left(\frac{h}{a}\right) \qquad \text{if} \quad h \geq a \tag{6.58}$$

and

$$\alpha = 28° - \left(\frac{a}{h}\right) \qquad \text{if} \quad h < a \tag{6.59}$$

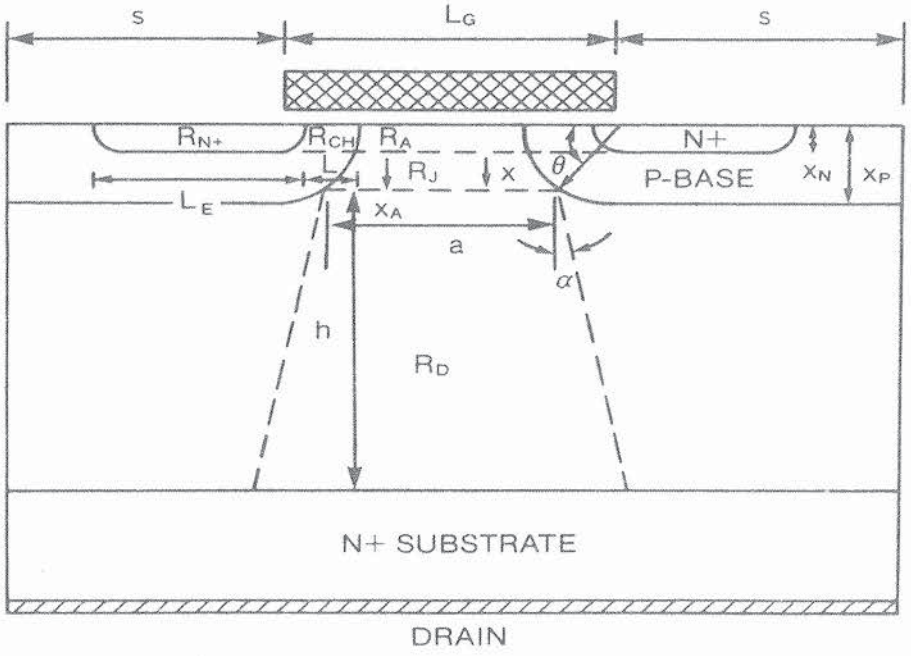


Fig. 6.19. Double-diffused MOSFET cell structure with each component of the on-resistance indicated.

The preceding expression for the drift region resistance is based on the condition of no overlap in the current flow paths. If the drift region thickness is large and the cell window $2m$ is small, the current flow paths will overlap as illustrated in Fig. 6.20. The resistance of the lower region can be calculated by using a uniform current density in this portion. Note that the resistance of the JFET and drift region decrease as the gate length L_G is increased.

Consider the ideal case where the resistances of the N^+-emitter, the N^+ substrate, the n-channel, the accumulation region, and the JFET region are negligible. The specific on-resistance of the power MOSFET will then be determined

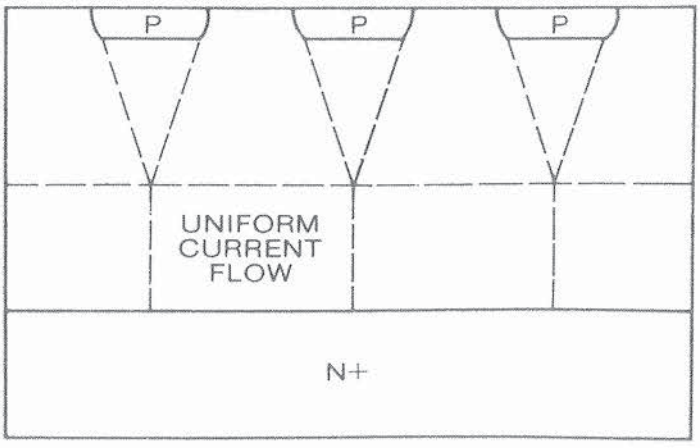


Fig. 6.20. MOSFET cell structure with overlapping current flow in the drift region.

by the drift region. This resistance is identical to that derived for the ideal JFET:

$$\text{(ideal)} \qquad R_{on,sp} = 5.93 \times 10^{-9}(BV_{pp})^{2.5} \tag{6.60}$$

for n-channel devices and

$$\text{(ideal)} \qquad R_{on,sp} = 1.63 \times 10^{-8}(BV_{pp})^{2.5} \tag{6.61}$$

for p-channel devices.

This treatment is applicable to the DMOS structure. A similar analysis can be performed for the VMOS structure. In this case, no JFET region exists. Instead, the current spreads out from an accumulation layer formed at the tip of the V groove as illustrated in Fig. 6.21. The drift region resistance must now be modified to contain two portions R_{D1} and R_{D2}. The portion at the tip of the V groove has been analyzed [13]:

$$R_{D1,sp} = (0.477)\rho_D(L_G + 2m) \tag{6.62}$$

The resistance due to component R_{D2} is the same as in the case of the DMOS structure for the a and h as illustrated in Fig. 6.21.

Returning to the DMOS device structure, the impact of changing the cell geometry on the on-resistance can best be illustrated by considering some specific cases. Consider the case of a relatively low voltage (100-V) power MOSFET with a drift region thickness W of 4.6 μm and a channel length L of 0.3 μm [14]. The impact of increasing the width of the accumulation region L_G upon the various components of the on-resistance is provided in Fig. 6.22*a* for two cell window sizes $S = 7$ μm and $S = 30$ μm). It can be seen that the channel and accumulation layer resistances grow as the accumulation layer width L_G increases. Meanwhile, the resistance of the drift region decreases because of improved current spreading. A pronounced minimum in the total on-resistance is observed at an optimum accumulation region width. When the cell

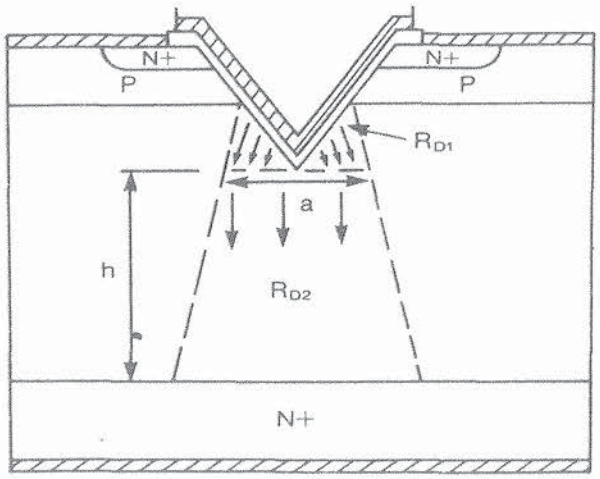
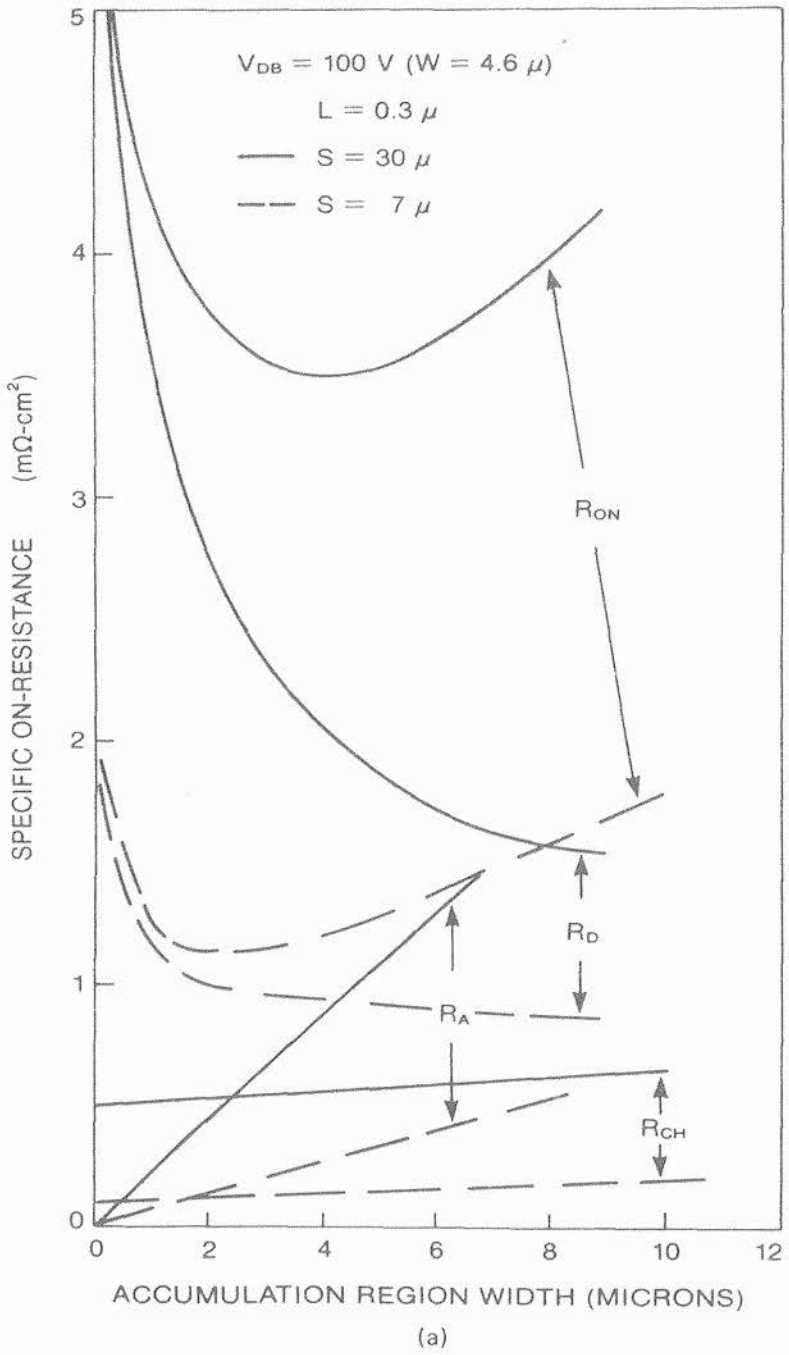


Fig. 6.21. V-Groove MOSFET cell structure with current flow path indicated.



(a)

Fig. 6.22. Calculated components of the specific on-resistance for (*a*) 100-V and (*b*) 1000-V power MOSFET structures. (After Ref. 14, reprinted with permission from the IEEE © 1979 IEEE.)

window size is reduced, all components of the resistance undergo a drastic reduction and the point of minimum total resistance shifts to smaller accumulation layer width L_G. This emphasizes the importance of using high-resolution photolithography to fabricate these low-breakdown-voltage power MOSFETs.

Another example of the impact of cell geometry on the on-resistance is provided in Fig. 6.22*b* for the case of a high-voltage (1000 V) device. In this case, the resistance of the drift region becomes predominant and the cell window

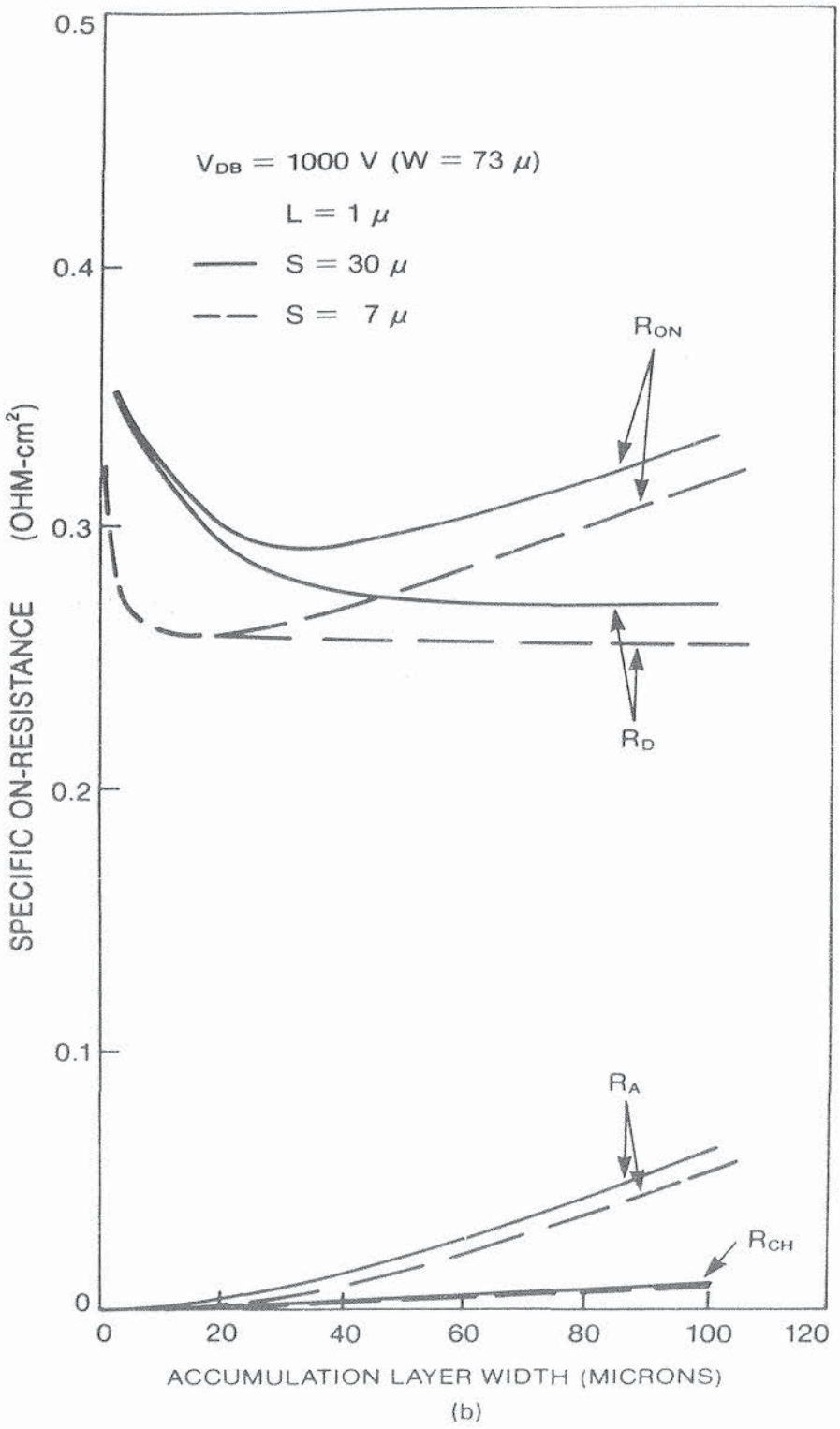


Fig. 6.22. (Continued)

size has only a small impact on the specific on-resistance. It is also significant that the minimum on-resistance is observed to occur at a much larger accumulation layer width L_G than for the low-voltage devices. The high-voltage power MOSFET can, consequently, be fabricated with less stringent photolithographic requirements than in the case of low-voltage devices.

For a given lithographic design rule, power MOSFETs with on-resistances approaching the ideal case can be realized at higher breakdown voltages. The specific on-resistance for state-of-the-art (circa 1981) power MOSFETs are compared with the ideal case in Fig. 6.23 [15]. High-voltage devices with breakdown voltages above 500 V approach within a factor of 2 of the ideal breakdown voltage. In contrast, the low voltage devices with breakdown voltages below 100 V exhibit a specific on-resistance of an order of magnitude higher than the ideal. Recent progress with the application of higher-resolution processing to low-voltage power MOSFETs is improving their on-resistance, making it closer to the ideal case.

The preceding analysis of on-resistance is based on a homogeneously doped drift region. All power MOSFETs are fabricated by using uniformly doped epitaxial layers. It has been proposed that a lower on-resistance can be obtained by using a nonuniform epitaxial doping profile to obtain the same breakdown

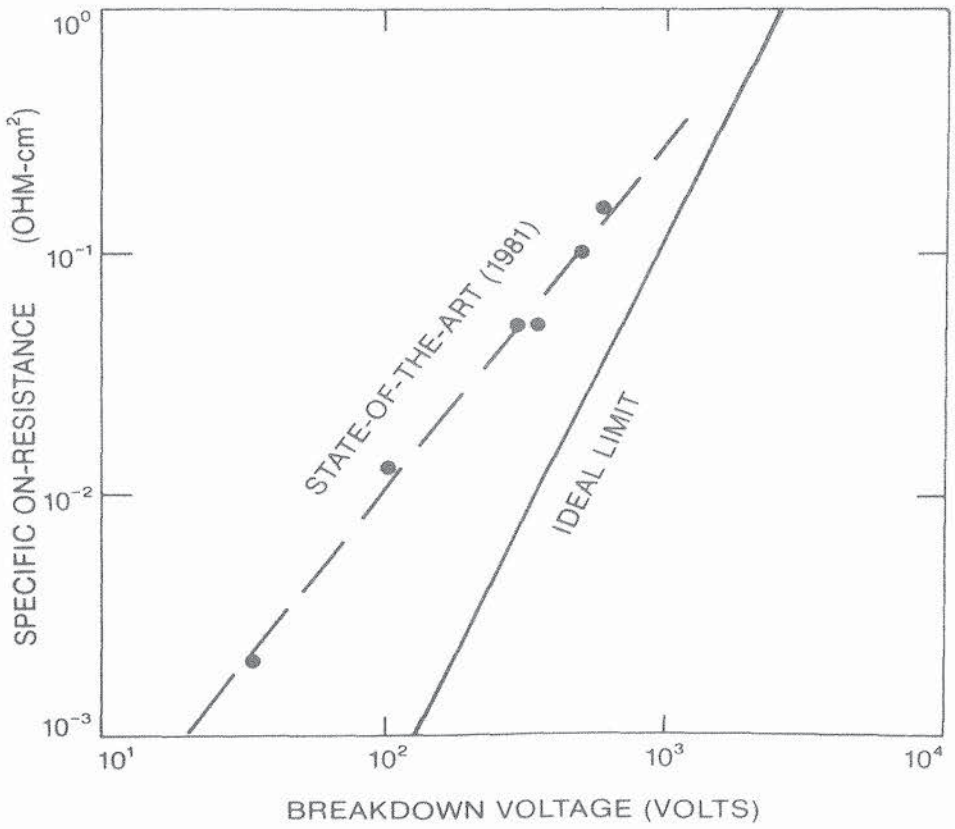


Fig. 6.23. Comparison of actual power MOSFET on-resistances with the ideal case. (After Ref. 15, reprinted with permission from the IEEE © 1981 IEEE.)

voltage [16]. Consider a doping profile $N(x)$ that will minimize the on-resistance of the drift layer

$$R_{on} = \int \frac{1}{q\mu N(x)}\, dx \tag{6.63}$$

under the condition that the electric field distribution satisfy the breakdown voltage requirement

$$V_B = \int \mathscr{E}\, dx \tag{6.64}$$

Using Poisson's equation to relate the doping and electric field, it can be shown that

$$\frac{d\mathscr{E}}{dx} = \frac{qN}{\epsilon_s} \tag{6.65}$$

These equations can be solved to obtain the optimum doping profile:

$$N(x) = \frac{\epsilon_s \mathscr{E}_C^2}{3qV_B\sqrt{1 - (2\mathscr{E}_C x/3V_B)}} \tag{6.66}$$

A comparison of the optimum profile with the uniformly doped case for the same breakdown voltage is provided in Fig. 6.24. According to this doping profile, the specific on-resistance is

$$R_{on,min} = \frac{3V_B^2}{\epsilon_s \mu_n \mathscr{E}_C^3} \tag{6.67}$$

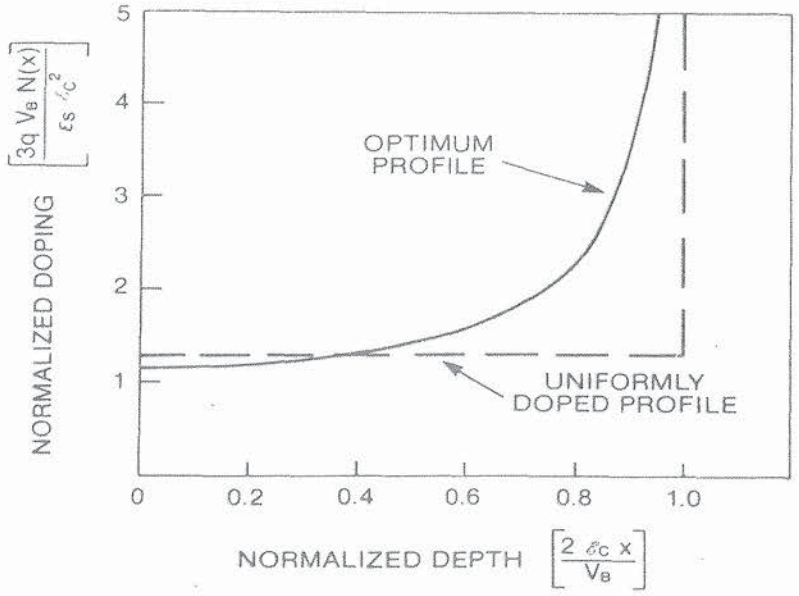


Fig. 6.24. Optimum doping profile for minimizing drift region resistance. (After Ref. 16, reprinted with permission from the IEEE © 1979 IEEE.)

where $\mathscr{E}_C$ is the critical electric field for breakdown. This specific on-resistance is only 12.5% lower than that for the uniformly doped profile. As a result of out-diffusion from the substrate during processing and epitaxial growth with uniformly doped epitaxial layers, the doping profile of low-voltage (100-V) power MOSFETs approaches the ideal profile without the need to take a special effort to tailor the epitaxial layer doping.

6.4. FREQUENCY RESPONSE

The power MOSFET is inherently capable of operation at high frequencies because of the absence of minority carrier transport. Its frequency response can be treated in a manner similar to that used for the power JFET in Chapter 4. Two limits to operation at high frequencies are set by the transit time across the drift region and the rate of charging of the input gate capacitance. The transit time limited frequency response was already treated in Chapter 4. It is a function of breakdown voltage:

$$f_T = \frac{6.11 \times 10^{11}}{(1 + L/d)(BV_{PP})^{7/6}} \tag{6.68}$$

In this formula, L is the channel length and d is the drift region thickness, including the JFET region. To achieve a high frequency response, it is important to keep the channel length small. As in the case of power JFETs, the frequency response diminishes with increasing breakdown voltage. This limit to the frequency response is seldom encountered since it lies in the gigahertz range.

Another limit to high-frequency operation arises from the need to charge and discharge the input gate capacitance. A simple equivalent circuit for the power MOSFET is shown in Fig. 6.25. In addition to the gate–source capacitance, a significant gate–drain capacitance must be included in the analysis because of the overlap of the gate electrode over the drift region. This capacitance is amplified by the Miller effect into an equivalent input gate capacitance of

$$C_M = (1 + g_m R_L) C_{GD} \tag{6.69}$$

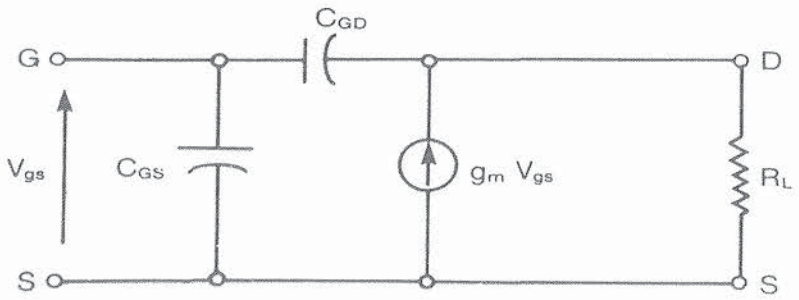


Fig. 6.25. Simple equivalent circuit of the power MOSFET.

The total input capacitance is

$$C_{in} = (C_{GS} + C_M) \tag{6.70}$$

To analyze the frequency response, it is necessary to evaluate each of the components of the input capacitance arising in the power MOSFET structure. Consider the conventional DMOS structure, shown in Fig. 6.26, with the gate electrode extending between adjacent DMOS cells. The input gate–source capacitance for this structure contains several components: (1) the capacitance C_{N+} arising from the overlap of the gate electrode over the N^+ emitter region, (2) the capacitance C_P arising from the MOS structure created by the gate electrode over the P-base region, and (3) the capacitance C_M arising from running the source metal over the gate electrode. The total gate–source capacitance is given by

$$C_{GS} = C_{N+} + C_P + C_M \tag{6.71}$$

The capacitance between the source and gate electrodes (C_M) is determined by the dielectric constant and thickness of the intervening insulator

$$C_M = \frac{\epsilon_I A_o}{t_I} \tag{6.72}$$

where A_o is the area of the overlap between the source and gate electrodes. To reduce this capacitance, a thick insulating layer is generally used during device fabrication. The other components of the capacitance require analysis of the MOS structure.

6.4.1. MOS Capacitance

Consider the basic MOS structure shown in Fig. 6.27. This structure contains two capacitances in series, the oxide capacitance and the semiconductor

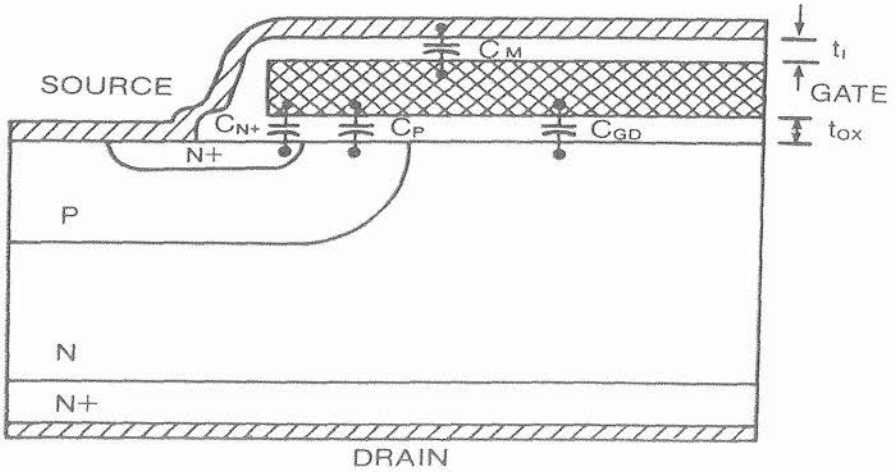


Fig. 6.26. Conventional DMOS structure with capacitances indicated.

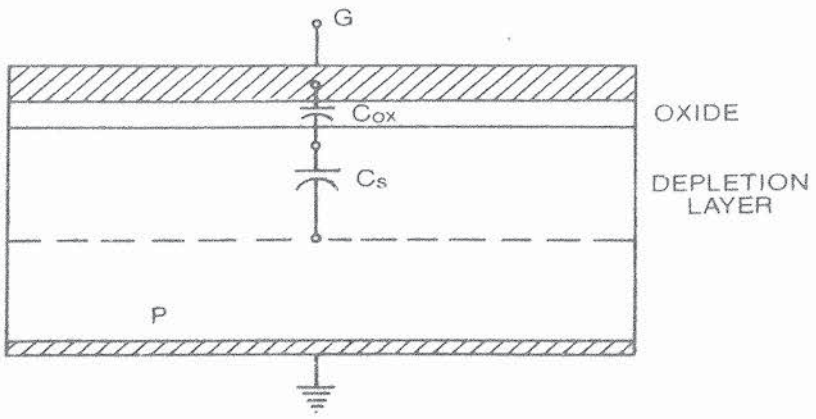


Fig. 6.27. Basic MOS structure with capacitances indicated.

capacitance:

$$\frac{1}{C_G} = \frac{1}{C_{ox}} + \frac{1}{C_S} \tag{6.73}$$

The oxide capacitance per unit area is determined by the thickness of the gate oxide:

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} \tag{6.74}$$

The semiconductor capacitance is determined by the width of the depletion layer. For a *P*-type substrate, if a negative gate bias is applied, an accumulation layer will form at the surface. Any changes in the gate voltage cause a corresponding change in the accumulation layer charge with a response time corresponding to the dielectric relaxation time. The capacitance is then equal to the oxide capacitance:

$$C_{G,ACC} = C_{ox} \tag{6.75}$$

When a positive bias is applied, a depletion layer forms in the semiconductor. The capacitance of the depletion layer is related to the space charge in the semiconductor:

$$C_S = \frac{dQ_S}{d\psi_S} \tag{6.76}$$

Using Eq. (6.13), it can be shown that

$$C_S = \frac{\epsilon_s}{\sqrt{2}\,L_D} \frac{[1 - e^{-(q\psi_S/kT)} + (n_{p0}/p_{p0})[e^{(q\psi_S/kT)} - 1]}{F(q\psi_S/kT, n_{p0}/p_{p0})} \tag{6.77}$$

The MOS gate capacitance will decrease with increasing positive gate voltage as a result of widening of the depletion layer. Beyond a certain voltage, a surface inversion layer will form. Once the inversion layer forms, the depletion layer

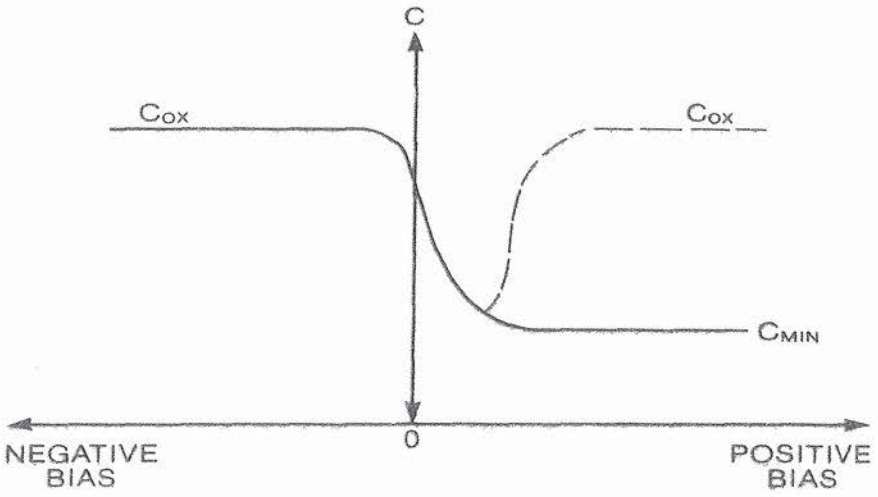


Fig. 6.28. The $C-V$ characteristics of a MOS structure with p-type substrate.

reaches its maximum value and the semiconductor capacitance attains its lowest value. The resulting $C-V$ curve is illustrated in Fig. 6.28 by the solid line. This curve is called the *high-frequency response curve* (typically measured at 1 MHz) because it is assumed that the inversion layer charge cannot follow the AC signal used to measure the capacitance. If either the measurement frequency or the semiconductor lifetime is reduced until the inversion layer charge can respond to the AC signal, the capacitance will become equal to the oxide capacitance as illustrated by the dashed line.

In the case of the power MOSFET, the $C-V$ curve shown by the solid line in Fig. 6.28 is not observed because the inversion layer formation in the P-base region is accompanied by transport of carriers between the N^+ emitter and the inversion layer. Consequently, even at high operating frequencies, the inversion layer charge can respond to the applied high-frequency signal by the rapid transport of electrons from the N^+ emitter into the inversion layer. Consequently, the measured input gate $C-V$ curve of power MOSFETs follows the dashed lines rather than the solid line shown in Fig. 6.28.

6.4.2. Input Capacitance

From the preceding analysis of the MOS capacitance for the gate structure, the various components of the input capacitance can be determined as follows. For the heavily doped N^+-emitter region, the capacitance is determined by the gate oxide thickness

$$C_{N^+} = \frac{\epsilon_{ox} A_{N^+o}}{t_{ox}} \tag{6.78}$$

where A_{N^+o} is the area of overlap of the gate electrode over the N^+ emitter. For the DMOS structure with a channel width Z, the overlap area over the

N^+ emitter is given by

$$A_{N^+o} = X_{N^+}Z \tag{6.79}$$

where X_{N^+} is the depth of the N^+-emitter diffusion. It is assumed here that the lateral diffusion of the emitter is equal to its depth and that the gate fringing capacitance can be neglected. When taken together, these are reasonable approximations because they compensate for each other.

The capacitance of the gate over the P base (C_P) is dependent on the gate bias. It goes through a minimum with increasing gate bias as shown in Fig. 6.28 by the dashed lines. This capacitance can be reduced by keeping the channel length small. It represents a fundamental contribution to the input capacitance of the power MOSFET.

The gate–drain capacitance also varies with gate and drain voltage. It has a high value during the on-state because the surface of the drift region is under accumulation. As the drain voltage increases and the device supports high voltages, the gate–drain capacitance decreases. As a result of the amplification of the gate–drain capacitance by the Miller effect, this capacitance can severely reduce the frequency response.

The gate–drain overlap capacitance can be drastically reduced by altering the gate structure by eliminating the gate overlap over the drift region as shown in Fig. 6.29. In this structure, the source electrode is also confined to the diffusion window to eliminate the capacitance C_M. A significant improvement in the high-speed switching performance of power MOSFETs has been reported by taking these measures during device design and fabrication [6]. It is noteworthy that elimination of the gate–drain overlap does adversely impact the on-resistance.

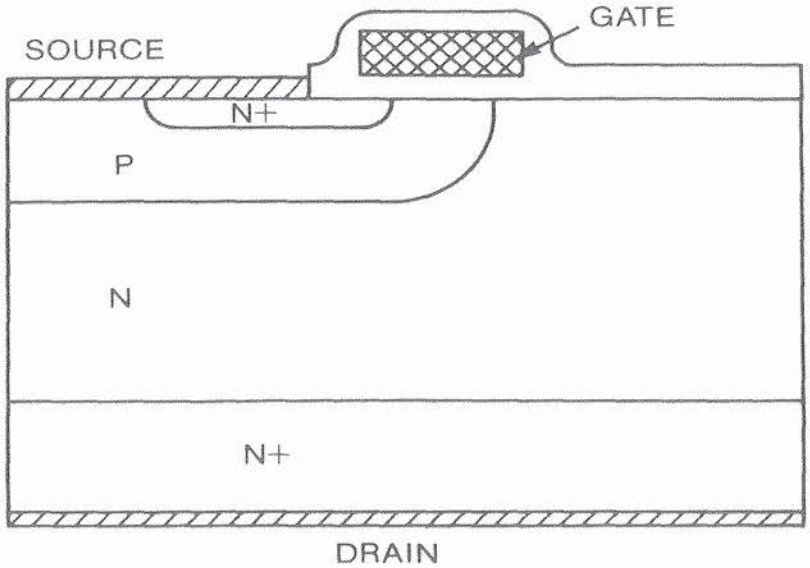


Fig. 6.29. High-frequency DMOS structure.

6.4.3. Gate Series Resistance

In the power MOSFET equivalent circuit shown in Fig. 6.25, no input gate series resistance was assumed to exist. In practice, a series gate resistance is always present, arising from either the resistance of the gate electrode internal to the device or from the gate drive circuit. The frequency response limited by the RC charging time constant of the input gate circuit is given by

$$f_{in} = \frac{1}{2\pi C_{in} R_G} \tag{6.80}$$

In the case of conventional polysilicon gate power MOSFETs, the sheet resistance of the gate electrode is typically over 10 Ω/□. This has been found to seriously lower the frequency response [17]. By changing to a molybdenum gate structure, an increase in frequency response by an order of magnitude has been reported, demonstrating that this is the limiting factor to device high-frequency performance. Power MOSFETs capable of delivering 100 W at 900 MHz have been developed by using a molybdenum silicide gate process [18].

The maximum frequency of operation at which the input current becomes equal to the load current is given by

$$f_m = \frac{g_m}{2\pi C_{in}} \tag{6.81}$$

This limit to operation is generally avoided because of the high power dissipation in the gate circuit.

6.5. SWITCHING PERFORMANCE

Because of the inherent high-speed turn-on and turn-off capability of power MOSFETs, these devices are often used as power switches. When used in this manner, they are maintained in either the on-state or forward blocking state for most of the time and must rapidly switch between these modes of operation. The power dissipation is determined by the on-resistance during the conducting state and the leakage current in the forward blocking state. These parameters have been discussed earlier. In this section, the parameters that govern the transition between the on and off states are analyzed. In addition, high-speed switching is accompanied by a very high rate of change of the drain voltage, which can cause an undesirable turn-on of the parasitic bipolar transistor in the power MOSFET. This can limit the switching speed and safe operating area of power MOSFETs.

6.5.1. Transient Analysis

A simple analytical procedure for predicting the switching transients experienced with power MOSFETs can be achieved by using the circuit model shown in Fig. 6.30 [19]. As a result of the interaction between the device and the circuit,

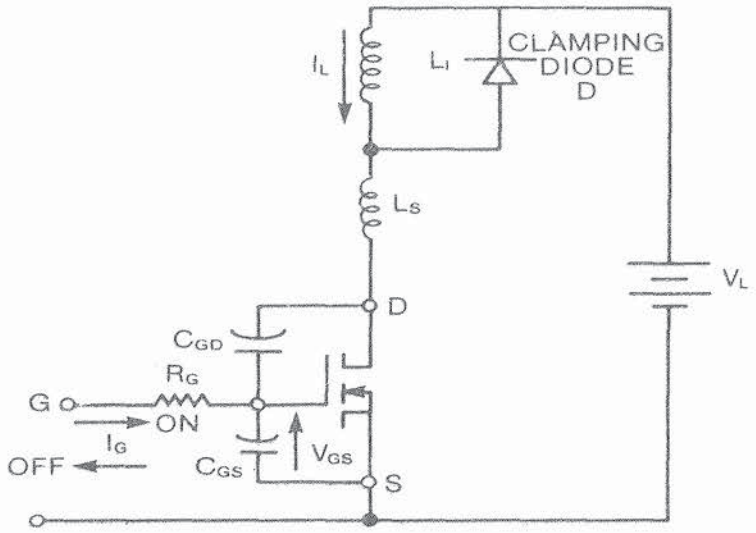


Fig. 6.30. Inductive switching circuit using power MOSFET for controlling load current.

it is necessary to consider a specific type of load for analysis. Here a clamped, inductive load L_1 is considered with a steady-state current I_L flowing through it. The inductance L_S is the stray circuit inductance not clamped by the diode D.

First, consider the case of a step voltage V_{GA} applied at the gate terminal G. The gate voltage V_{GS} that controls the drain current flow is determined by the voltage across the gate–source capacitance C_{GS}. This voltage is governed by the charging of the capacitances C_{GS} and C_{GD} by resistor R_G. As long as the gate voltage V_{GS} remains below the threshold voltage V_T, no drain current will flow. The time taken for the gate voltage V_{GS} to reach the threshold voltage V_T represents a turn-on delay period. During this period, the input capacitance is simply $(C_{GS} + C_{GD})$ because the Miller effect is operative only when the transistor is in its active region. The gate voltage then rises exponentially:

$$V_{GS} = V_{GA}\{1 - e^{-[t/R_G(C_{GS}+C_{GD})]}\} \tag{6.82}$$

The turn-on delay time obtained from this expression is

$$t_1 = R_G(C_{GS} + C_{GD}) \ln\left[\frac{1}{1 - (V_T/V_{GA})}\right] \tag{6.83}$$

Drain current flow begins to occur beyond the turn-on delay time. Consider the case of a MOSFET with linear transfer characteristics as shown in Fig. 6.31. In this case, the drain current will increase in proportion to the gate voltage. Since the device is now in its active region, the Miller effect will determine the gate–drain capacitance that is being charged by the gate circuit. If the stray inductance L_S is small, the Miller effect is small and the gate voltage will continue to rise exponentially as described by Eq. (6.82). The drain current will then take the form

$$I_D = g_m\{V_{GA}[1 - e^{-[t/R_G(C_{GS}+C_{GD})]}] - V_T\} \tag{6.84}$$

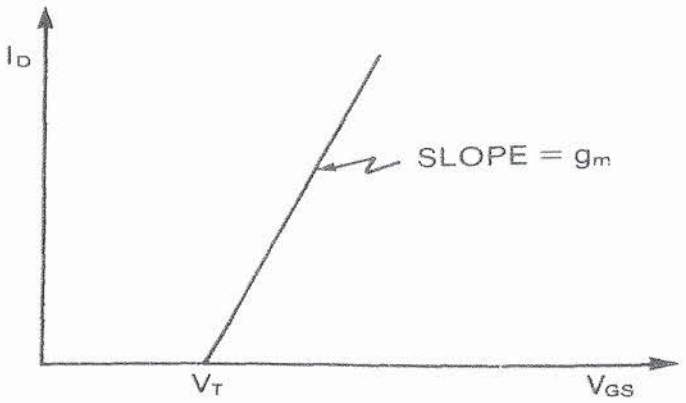


Fig. 6.31. Linear transfer characteristics of a power MOSFET used for transient analysis.

The drain voltage will be nearly constant during this time. This period will end when the drain current reaches the load current I_L, that is, when all the circulating diode current is transferred to the power MOSFET. Beyond this point the drain voltage will fall from V_L to the on-state voltage of the power MOSFET. Since the drain current is now constant, the gate voltage must be

$$V_{GS} = V_T + \frac{I_L}{g_m} \tag{6.85}$$

Since the gate voltage V_{GS} is constant, all the input current flows into the Miller capacitance C_{GD} during this period. The input current is given by

$$I_G = \frac{V_G - V_{GS}}{R_G} \tag{6.86}$$

Thus

$$\frac{dV_{GD}}{dt} = \frac{I_G}{C_{GD}} = \frac{V_G - (V_T + I_L/g_m)}{R_G C_{GD}} \tag{6.87}$$

The rate of change of the drain–source voltage is equal to the rate of change of the gate–drain voltage because the gate–source voltage is constant during this period. Thus

$$\frac{dV_D}{dt} = \frac{V_G - (V_T + I_L/g_m)}{R_G C_{GD}} \tag{6.88}$$

Integration of this equation yields

$$V_D = V_L - \left[\frac{V_G - (V_T + I_L/g_m)}{R_G C_{GD}}\right] t \tag{6.89}$$

The drain voltage will then decrease linearly with time. The composite gate and drain waveforms are shown in Fig. 6.32. Note that the gate voltage may continue to rise beyond time t_3, but this will have no influence on the drain current or voltage because they have reached their steady-state values.

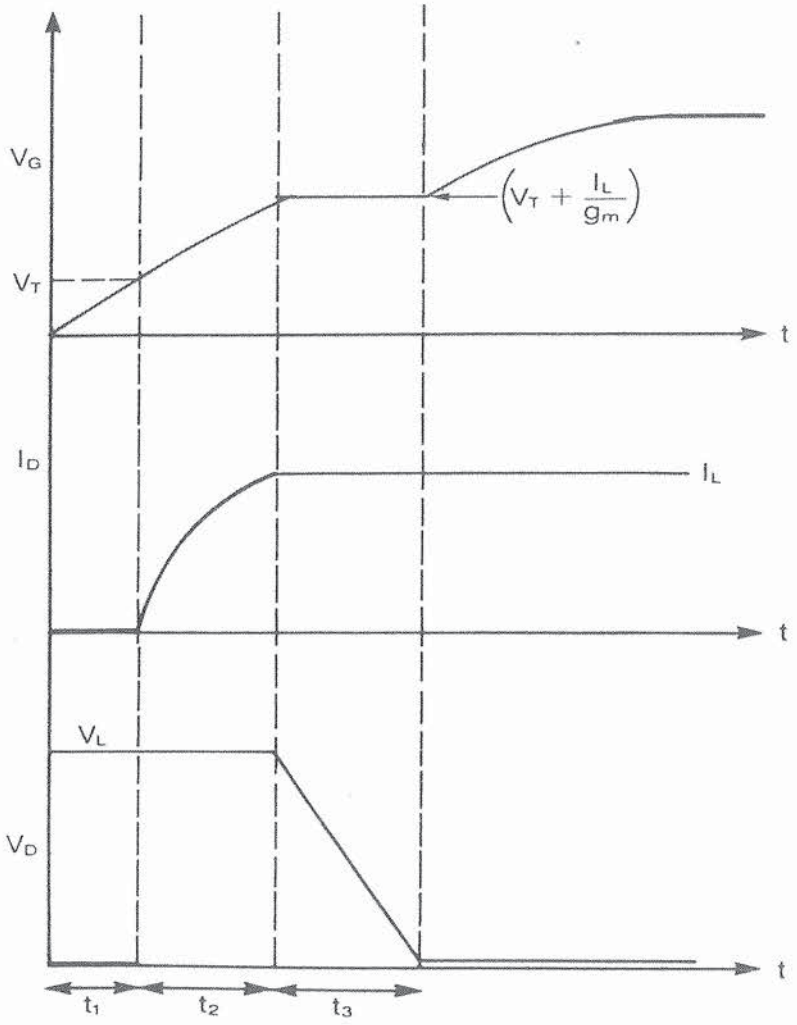


Fig. 6.32. Turn-on transients of a power MOSFET for the case of small stray inductance [low L_S/R_G) ratio].

The preceding analysis is valid only when the ratio (L_S/R_G) is small. Other cases can be treated in a similar manner [19]. The high-power dissipation due to the device turn-on transient occurs during the interval t_2 and t_3, where both high current and voltage are sustained by the device simultaneously. During this period the device current–voltage locus must stay within its safe operating area.

A corresponding analysis can be performed for device turn-off. In this case, an abrupt drop in applied gate voltage from V_G to zero is assumed to occur at the terminal G. The voltage V_{GS} then decreases exponentially with time as a result of discharging of the gate capacitance, until the gate voltage reaches the value required to obtain a saturated drain current equal to the load current I_L. During this period

$$V_{GS} = V_G\, e^{-(t/R_G C_G)} \tag{6.90}$$

The duration of this turn-off delay phase is given by the condition

$$V_{GS}(t_4) = V_T + \frac{I_L}{g_m} = V_G\, e^{-(t_4/R_G C_G)} \tag{6.91}$$

From this equation, the turn-off delay time is obtained:

$$t_4 = R_G C_G \ln\left[\frac{V_G}{V_T + (I_L/g_m)}\right] \tag{6.92}$$

Beyond this time, the drain current will remain at I_L whereas the drain voltage will begin to rise toward V_L because the load current cannot be diverted from the MOSFET into the diode until V_D exceeds V_L. As long as the drain current is constant, the gate voltage V_{GS} will also remain constant and the gate current I_G will drive the gate–drain capacitance C_{GD}. During this period

$$V_{GS} = V_T + \frac{I_L}{g_m} \tag{6.93}$$

and

$$I_G = \frac{V_{GS}}{R_G} = \frac{V_T + (I_L/g_m)}{R_G} \tag{6.94}$$

Since this current charges the gate–drain capacitance C_{GD} and the rate of change of drain–source voltage is equal to the rate of change of the drain–gate voltage, it follows that

$$\frac{dV_{DS}}{dt} = \frac{dV_{DG}}{dt} = \frac{I_G}{C_{GD}} \tag{6.95}$$

Thus

$$V_{DS} = V_{on} + \frac{I_G}{C_{GD}}\, t \tag{6.96}$$

$$= V_{on} + \frac{1}{R_G C_{GD}}\left(\frac{I_L}{g_m} + V_T\right) t \tag{6.97}$$

The drain voltage will rise linearly from the on-state voltage drop V_{on} to the load voltage during this period. The duration of this period can be obtained by equating V_{DS} to V_L:

$$t_5 = \frac{R_G C_{GD}(V_L - V_{on})}{(I_L/g_m) + V_T} \tag{6.98}$$

At this point the freewheeling diode D will turn on. To achieve this, it is essential that the MOSFET drain voltage exceed the load supply voltage V_L. If the stray inductance L_S is small, the overshoot in the drain voltage will be

small. However, if L_S is large and the rate of change of the drain current becomes large, the voltage on the MOSFET drain can exceed its breakdown voltage. If the stray inductance is small, the drain voltage can be assumed to remain relatively constant. The gate voltage will then continue to decrease exponentially:

$$V_G = \left(\frac{I_L}{g_m} + V_T\right) e^{-(t/R_G C_G)} \tag{6.99}$$

The drain current will follow this change in the gate voltage:

$$I_D = (I_L + g_m V_T)\, e^{-(t/R_G C_G)} - g_m V_T \tag{6.100}$$

This period will extend until the gate voltage reaches the threshold voltage V_T and the drain current is reduced to zero. The time interval for this period is

$$t_6 = R_G C_G \ln\left(\frac{I_L}{g_m V_T} + 1\right) \tag{6.101}$$

After this interval, the gate voltage will continue to decay exponentially to zero. Since the drain current and voltage are at their steady-state voltages (off state), this will have no influence on them.

The composite waveforms corresponding to the preceding discussion of turn-off are shown in Fig. 6.33. These waveforms apply to the case of a small ratio (L_S/R_G). Other cases can be treated in a similar manner [19]. The power dissipation during the turn-off transient occurs primarily during the periods t_5 and t_6, where both the current and the voltage are large. Once again the current–voltage locus must remain within the safe operating area of the device. The drain voltage overshoot during period t_6 must especially be noted. Its magnitude depends on the size of the stray inductance. If the stray inductance is large, the MOSFET can be forced into avalanche breakdown, causing destructive failure. The analysis presented here also indicates the need to keep the gate–drain capacitance small. This can be achieved by avoiding the overlap of the gate electrode with the drift region as shown in Fig. 6.29 or by using a thick oxide in the portion where the drift region comes to the surface.

6.5.2 (dV/dt) Capability

It is well known that thyristors are prone to turn-on under the application of a high rate of change of the anode voltage. When a high (dV/dt) occurs at the drain terminal, power MOSFETs can also be forced into current conduction. In certain cases, this will lead to destructive failure of the devices. The various mechanisms that can lead to (dV/dt)-induced turn-on are discussed here. Circuit and device design approaches to raising the (dV/dt) are provided.

To analyze the (dV/dt)-induced turn-on in power MOSFETs, consider the equivalent circuit of the device shown in Fig. 6.34 with a ramp applied between drain and source [20]. In addition to the device capacitances, the equivalent circuit shows the parasitic N–P–N bipolar transistor with a base–emitter

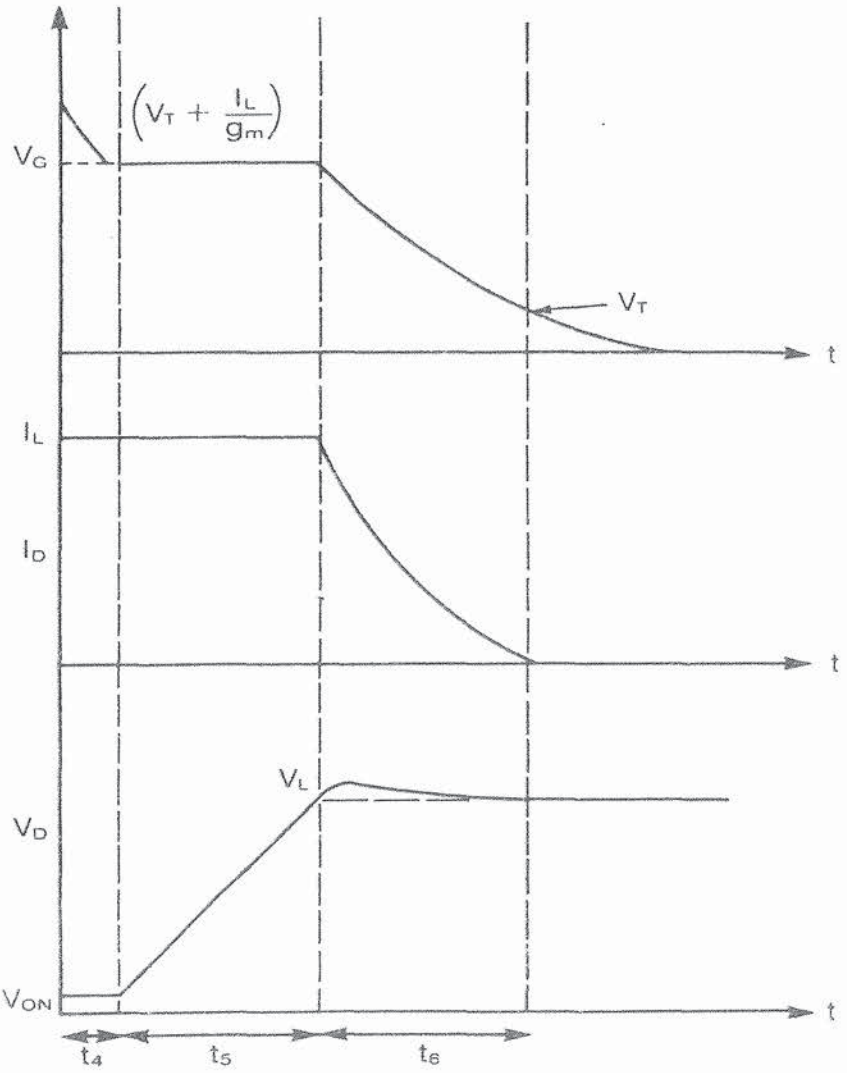


Fig. 6.33. Turn-off transients of a power MOSFET for the case of small stray inductance [low L_S/R_G) ratio].

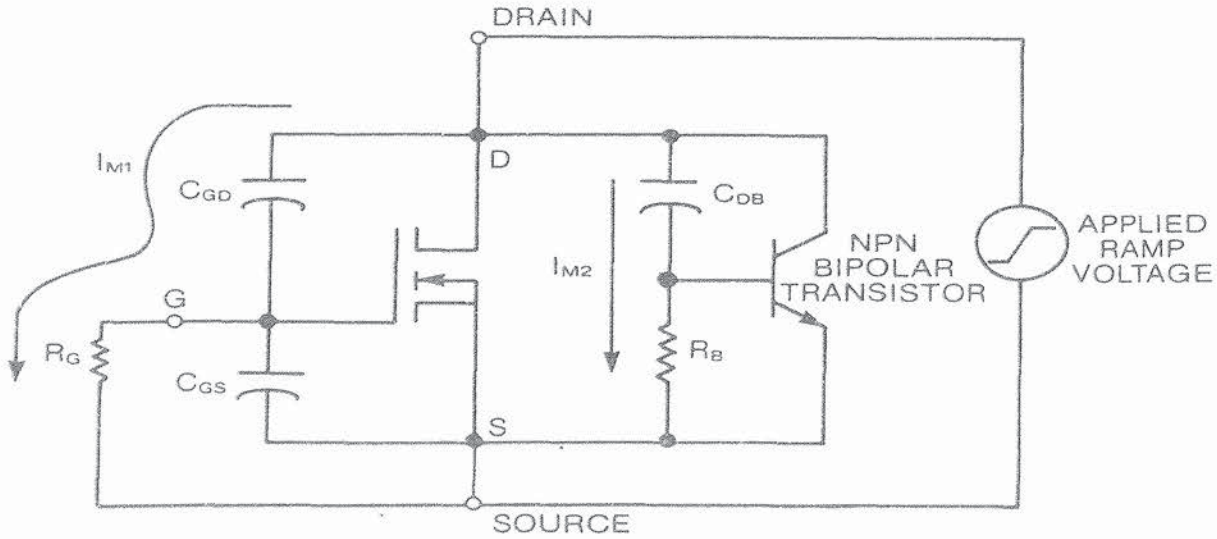


Fig. 6.34. Equivalent circuit used for analysis of (dV/dt)-induced turn-on in power MOSFETs.

shunting resistance R_B. This shunting resistance is present in the device structure as a result of the overlapping of the source metal over both the N^+-emitter and the P-base regions.

Mode 1: For the first mode of (dV/dt)-induced turn-on, consider the current I_{M1} flowing by means of the gate-to-drain capacitance C_{GD} into the gate circuit resistance R_G. If the drain voltage is much larger than the gate voltage, the voltage drop across the gate resistance will be approximately given by

$$V_{GS} = I_{M1} R_G \tag{6.102}$$

$$= R_G C_{GD} \left[\frac{dV}{dt}\right] \tag{6.103}$$

If the gate voltage V_{GS} exceeds the threshold voltage of the power MOSFET, the device will be forced into current conduction. The (dV/dt) capability set by this mode is

$$\left[\frac{dV}{dt}\right] = \frac{V_T}{R_G C_{GD}} \tag{6.104}$$

The (dV/dt) capability can be increased by using a very-low-impedance gate drive circuit and raising the threshold voltage. Since the threshold voltage decreases with increasing temperature, this mode of turn-on can become aggravated with increasing power dissipation within the device. In general, it is nondestructive because the gate voltage does not rise much above the threshold voltage and the device current is limited by the high device resistance.

Mode 2: This mode of (dV/dt)-induced turn-on occurs as a result of the existence of the parasitic bipolar transistor. At high rates of change of the drain voltage, a current I_{M2} flows by means of the capacitance C_{DB} into the base short-

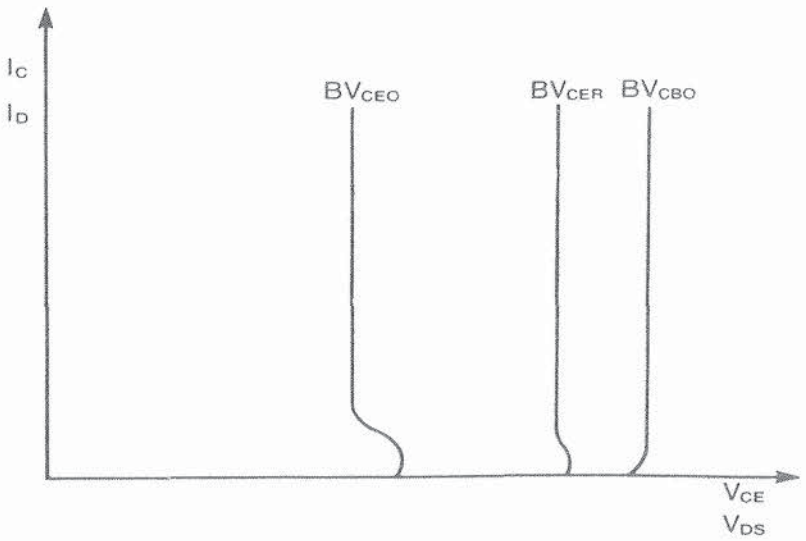


Fig. 6.35. Breakdown characteristics of the power MOSFET induced by turn-on of the parasitic N–P–N bipolar transistor.

circuiting resistance R_B. If this current is sufficient for the base–emitter junction of the bipolar transistor to become forward-biased, it will turn on the transistor. For low values of R_B, when the emitter junction is inactive, the breakdown voltage of the transistor, and hence the power MOSFET, approaches the collector–base breakdown voltage BV_{CBO} as shown in Fig. 6.35. But under a high (dV/dt) and large values of R_B, the emitter–base junction becomes strongly forward-biased. The breakdown voltage then collapses to the level of the open-base transistor breakdown voltage BV_{CEO}, which is about 60% of the collector–base breakdown voltage BV_{CBO}. If the applied drain voltage is greater than BV_{CEO}, the device will enter avalanche breakdown or may be destroyed by second breakdown if the drain current is not externally limited.

The (dV/dt)-induced turn-on in this mode is dependent on the internal device structure. Consider the DMOS structure shown in Fig. 6.36 with one end of the N^+ emitter short-circuited to the P-base region. The applied (dV/dt) creates a displacement current flow in capacitance C_{DB}. This current must flow laterally in the P-base region to the source metal. It produces a voltage drop along the resistance R_B that forward-biases the edge A of the N^+ emitter. When the edge A of the N^+ emitter becomes forward-biased by the lateral current flow, the bipolar transistor will turn on, thus precipitating further current flow. An estimate of the (dV/dt) at which this will occur can be obtained by assuming that the bipolar transistor will turn on at an emitter–base forward-bias voltage of V_{BE}:

$$\frac{dV}{dt} = \frac{V_{BE}}{R_B C_{DB}} \tag{6.105}$$

where R_B is a distributed base resistance. To obtain a high (dV/dt) capability, it is important to keep the resistance R_B small. This can be achieved by increasing the doping level of the P base and keeping the length L_{N+} of the N^+ emitter

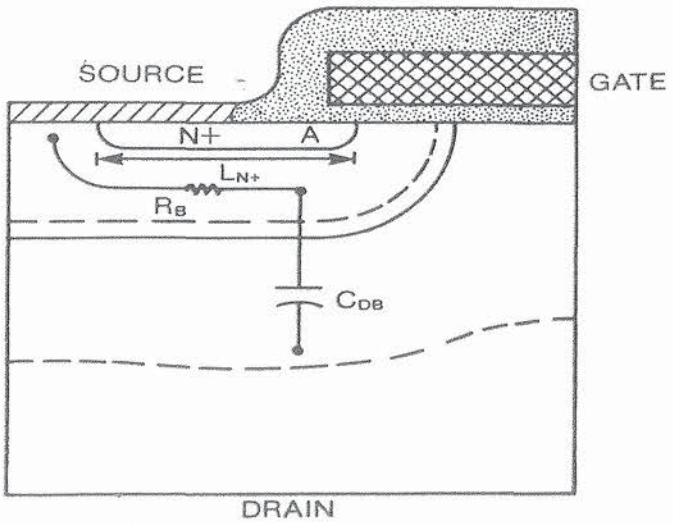


Fig. 6.36. Cross section of power MOSFET with capacitance C_{DB} and base resistance R_B shown.

as small as possible within the constraints of the lithography used for device fabrication.

It should be noted that the P-base sheet resistance increases with increasing drain voltage as result of the extension of the depletion layer, resulting in lowering of the dV/dt capability. In addition, increasing the temperature will lower the voltage V_{BE} at which the emitter will begin to inject. The P-base resistance will also increase with temperature because of a reduction in the mobility. These factors will cause a reduction in the (dV/dt) capability as temperature increases. Experimental confirmation of this mode of (dV/dt)-induced breakdown has been obtained on special structures, where the P base was not short-circuited to the N^+ emitter, allowing the use of large external base resistance R_B to make the effect pronounced [21]. In the case of practical power MOSFETs designed for operation at high switching speeds, the (dV/dt) capability is over 10,000 V/μsec.

6.6. SAFE OPERATING AREA

The safe operating area defines the limits of operation of the device. It is well known that the maximum current at low drain voltages is limited by power dissipation if the leads are sufficiently rated to prevent fusing. The maximum voltage at low drain currents is determined by the avalanche breakdown phenomena as discussed in Section 6.3.1. Under the simultaneously application of high current and voltage, the device may be susceptible to destructive failure even if the duration of the transient is too small to prevent excessive power dissipation. This failure mode has been referred to as *second breakdown.*

6.6.1. Bipolar Second Breakdown

The term *second breakdown* refers to a sudden reduction in the blocking voltage capability when the drain current increases. This phenomenon has been observed in power MOSFETs. It originates from the presence of the parasitic bipolar transistor in the device structure. When the drain voltage is increased to near the avalanche breakdown voltage, current flows into the P-base region in addition to the normal current flow within the channel inversion layer. The avalanche current collected within the P-base flows laterally along the P-base to its contact. The voltage drop along the P base forward-biases the edge of the N^+ emitter furthest from the base contact. When the forward bias on the emitter exceeds 0.6–0.7 V, it begins to inject carriers. The parasitic bipolar transistor is no longer capable of supporting the P-base–N-drift layer breakdown voltage BV_{CBO}. Its breakdown voltage is instead drastically reduced to BV_{CEO}, which is typically 60% of BV_{CBO}.

To analyze second breakdown in power MOSFETs, consider the equivalent circuit of the device with the parasitic bipolar transistor as shown in Fig. 6.37*a*. A base–emitter resistor R_B is shown in this figure corresponding to the lateral resistance in the P-base region illustrated in Fig. 6.37*b*. The current flow in the

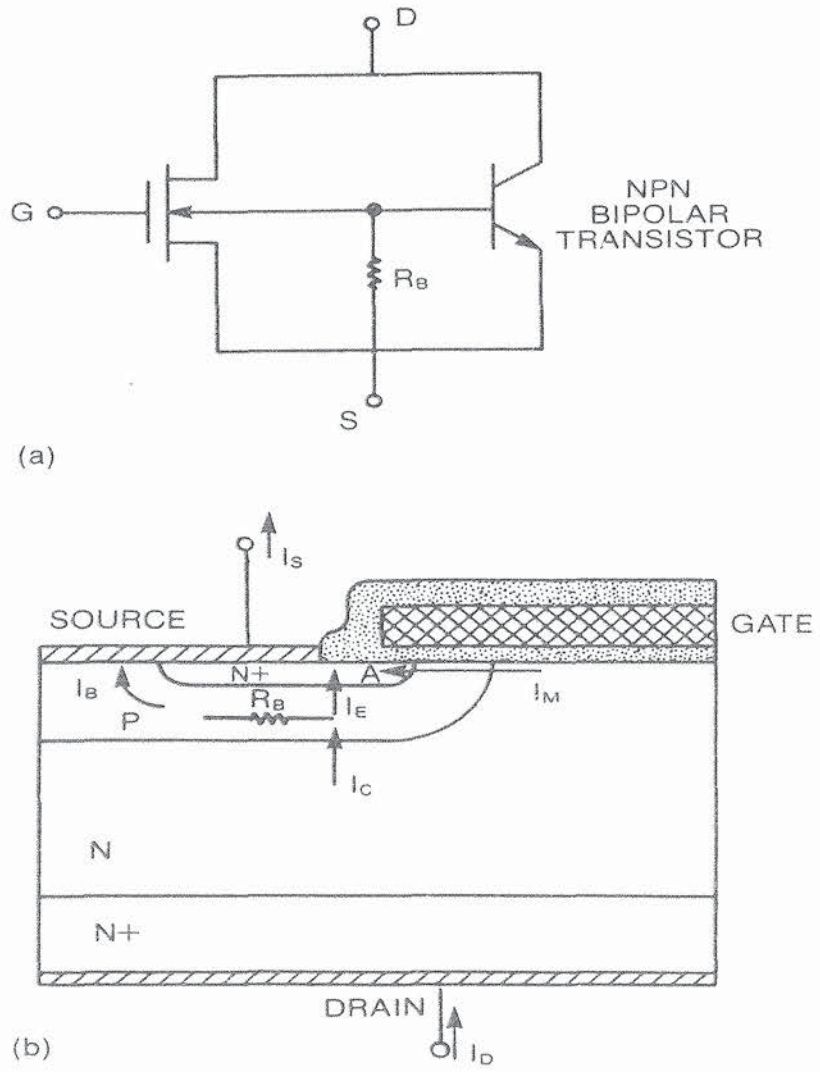


Fig. 6.37. (*a*) Power MOSFET equivalent circuit; (*b*) corresponding device cross section.

device takes two paths—one through the MOS channel and the other through the active bipolar transistor. These currents must satisfy the following conditions:

$$I_D = I_C + I_M \tag{6.106}$$

$$I_S = I_E + I_M \tag{6.107}$$

and

$$I_B = I_C - I_E \tag{6.108}$$

The emitter and collector currents are related by

$$I_C = \gamma_E \alpha_T M I_E \tag{6.109}$$

where γ_E is the injection efficiency, α_T is the base transport factor, and M is the avalanche multiplication factor. The first two parameters can be assumed to be equal to unity for a typical power MOSFET structure. Furthermore, during

second breakdown, the emitter current is caused by the forward bias at point A:

$$I_E = I_0 \, e^{(qV_B/kT)} \tag{6.110}$$

where V_B is the forward bias caused by the lateral base current flow:

$$V_B = R_B I_B \tag{6.111}$$

Combination of these equations gives

$$I_E = I_0 \exp\left[\frac{qR_B}{kT}(M-1)I_E\right] \tag{6.112}$$

If the first-order expansion of the exponential is used to evaluate the initiation of the second breakdown effect, it can be shown that

$$I_E = \frac{I_0}{[1-(qR_B/kT)(M-1)I_0} \tag{6.113}$$

The multiplication factor is related to the drain voltage by

$$M = \frac{1}{[1-(V_D/BV)^n]} \tag{6.114}$$

with $n \simeq 4$ for electrons. As the drain voltage increases, the emitter current, and hence the source current, can rise catastrophically in accordance with Eq. (6.113). The voltage at which this will occur can be obtained from the preceding equations:

$$V_{D,SB} = \frac{BV}{[1+qR_B I_0/kT]^{1/n}} \tag{6.115}$$

A reduction of the second breakdown voltage is predicted with increasing base resistance R_B by this analysis.

6.6.2. MOS Second Breakdown

Another phenomenon that causes second breakdown occurs because of the effect of the lateral voltage drop in the P base on the channel current. This is called the *MOSFET body bias effect* [22]. In this case, it is necessary to define a body bias coefficient:

$$\gamma = \frac{\Delta I_D}{\Delta V_B} \tag{6.116}$$

where

$$V_B = R_B I_B \tag{6.117}$$

The source current is then given by

$$I_S = I_M + \gamma V_B = I_M + \gamma R_B I_B \tag{6.118}$$

At high drain voltages, the large electric field causes avalanche multiplication of the channel current. The base current is now given by

$$I_B = I_D - I_S = (M - 1)I_S \tag{6.119}$$

Using Eq. (6.118), it can be shown that

$$I_B = (M - 1)(I_M + \gamma R_B I_B) \tag{6.120}$$

and from this equation

$$I_B = \frac{(M - 1)I_M}{1 - \gamma R_B(M - 1)} \tag{6.121}$$

and

$$I_S = \frac{I_M}{1 - \gamma R_B(M - 1)} \tag{6.122}$$

which is of the same form as Eq. (6.113) derived earlier. As the drain voltage increases, the multiplication factor increases and causes a catastrophic increase in the source current. The drain voltage at which this second breakdown occurs is

$$V_{D,SB} = \frac{BV}{(1 + \gamma R_B)^{1/n}} \tag{6.123}$$

Once again a reduction of the second breakdown voltage occurs with increasing base resistance. An exaggerated example of a power MOSFET I–V characteristic is shown in Fig. 6.38 with the safe operating area indicated.

As temperature increases, the breakdown voltage BV and the coefficient n will increase. This will tend to raise the second breakdown voltage. However,

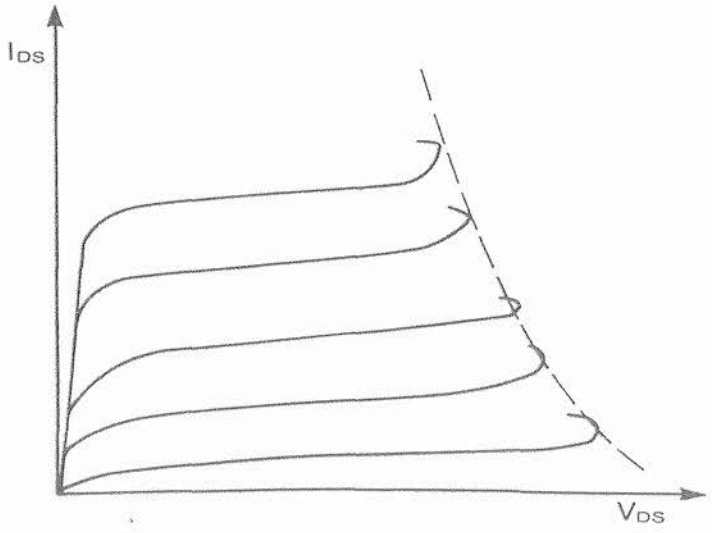


Fig. 6.38. Output characteristics of a power MOSFET with exaggerated second breakdown limit shown by the dashed lines.

the P-base resistance R_B also increases with temperature as a result of a reduction in the mobility. This has a compensating effect. It has been found that the second breakdown voltage and current are only weak functions of temperature [22]. In addition, p-channel power MOSFETs exhibit a smaller collapse in voltage in comparison with n-channel devices because of the lower N-base resistance and the lower hole impact ionization coefficient.

In conclusion, unless the emitter–base junction of the parasitic bipolar transistor in the power MOSFET structure is adequately short-circuited, the devices can exhibit second breakdown. Although this problem was observed in early devices, advances in design and process technology have eliminated this problem in modern devices.

6.6.3. Modern Device Characteristics

Because of the rugged design with very efficient short-circuiting of the emitter–base junction of the parasitic bipolar transistor, modern commercial power MOSFETs are rated with excellent safe operating area. As an example, the safe operating area of a typical 500-V power MOSFET [23] is shown in Fig. 6.39. The curves shown in this figure are simply based on a power dissipation that will keep the junction temperature below 150°C. The ability of a power

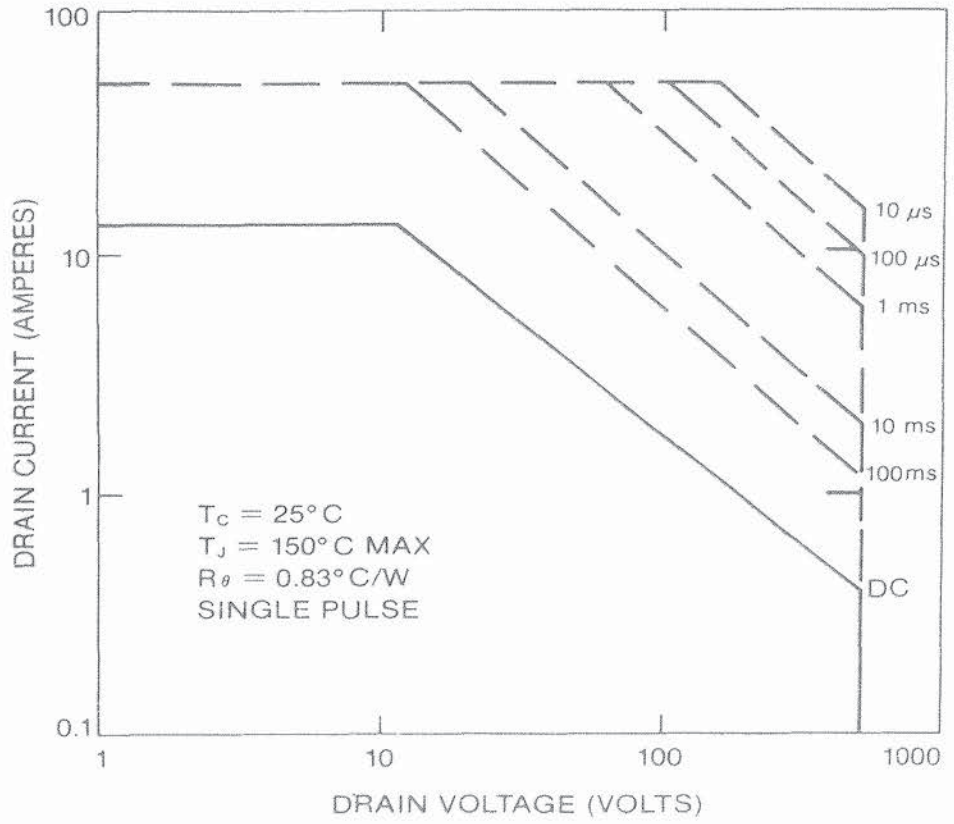


Fig. 6.39. Typical commercial power MOSFET safe-operating-area limits.

MOSFET to operate along these constant power contours substantiates the absence of second breakdown, by adequate short-circuiting of the internal parasitic bipolar transistor, in modern commercially available devices.

6.7. INTEGRAL DIODE

Some power switching circuits require reverse current flow through the active power switching devices. Even though the power switching devices do not have to exhibit any reverse blocking capability, in these applications they must be capable of reverse conduction. Some examples of these types of circuit are DC to AC inverters for adjustable-speed motor drives, switching power supplies, and DC to DC choppers for motor speed control with regenerative breaking. When using bipolar transistors it is necessary to use a diode connected across the transistor (often called an *antiparallel diode*) to conduct the reverse current. This diode must exhibit good reverse recovery characteristics with a small reverse recovery charge to keep the power dissipation low and a reverse recovery waveform that does not contain an abrupt change in current (i.e., snap-off). Any abrupt current changes create very high transient voltages that can damage the power devices.

When the power MOSFET is used in place of the bipolar transistor as the power switching device, the possibility of utilizing the reverse conducting diode inherent in the structure becomes attractive because it eliminates the complexity and cost of adding an external diode. The existence of the integral reverse conducting diode is illustrated in Fig. 6.40. This diode is formed across the

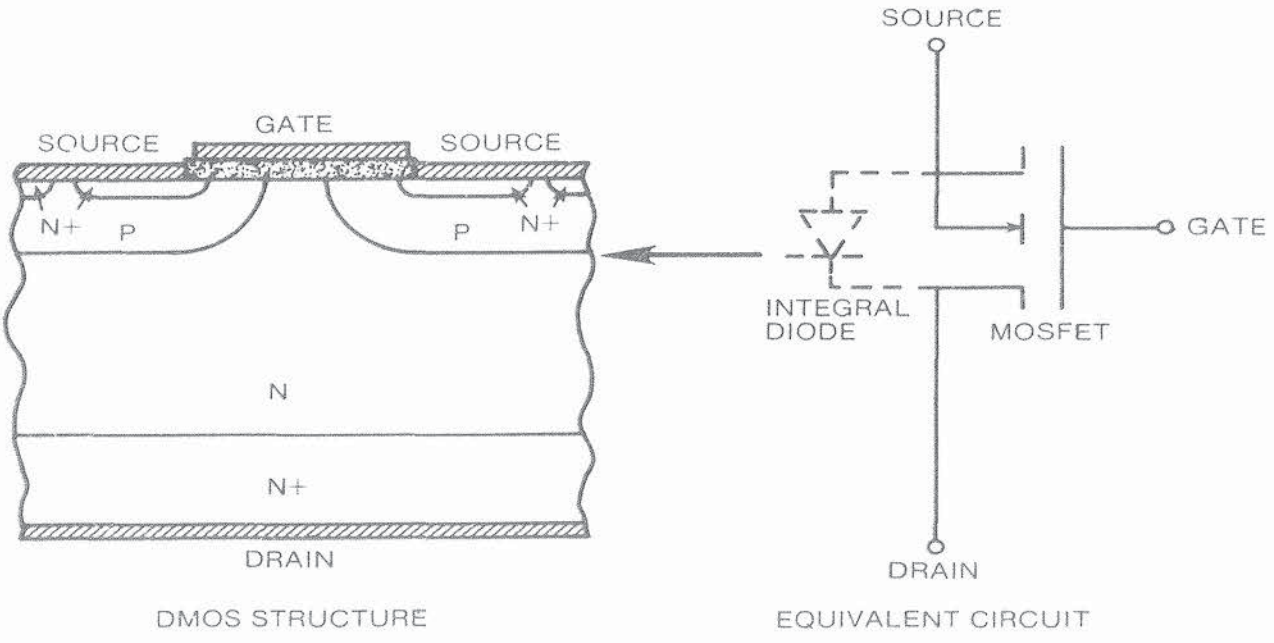


Fig. 6.40. Power MOSFET structure with integral diode formed by *P*-base and *N*-drift layer junction.

P-base–*N*-drift layer junction with its anode current flowing through the source contact to the *P*-base region. It should be noted that the applications under consideration here require high-voltage operation. The power MOSFET forward drop in these cases exceeds the junction knee voltage discussed earlier with reference to Fig. 6.3, and it is not practical to gate the power MOSFET and utilize it as a synchronous rectifier.

In the as-fabricated form, the lifetime in the *N*-drift region of power MOSFETs is generally high. The integral diode can be treated as a *P*–*i*–*N* rectifier between the source and drain terminals. It will conduct current very efficiently with a forward voltage drop in the range of 1 V as the result of the injection of a high concentration of minority carriers into the drift region. The theory of current conduction in a *P*–*i*–*N* rectifier has been discussed in Chapter 5. In the case of the power MOSFET integral diodes, the maximum current carried by the diode must match the power MOSFET rated current because in these applications, the current in one MOSFET invariably flows through the integral diode of another device in the circuit. Because of the high on-resistance of high-voltage power MOSFETs, their forward conduction current density is limited to below 100 A/cm^2. At these relatively low current densities for the integral diode, the forward voltage drop will be low (about 1 V). It can, therefore, easily sustain the current requirements of the intended applications.

The primary difficulty with using the integral diodes of power MOSFETs is with their reverse recovery characteristics. Because of the high lifetime in the drift region of as-fabricated devices, these diodes exhibit very slow reverse recovery with large recovery charge, so that a very large reverse recovery current is observed. The peak reverse recovery current I_p will increase with increasing (di/dt) (i.e., with higher switching speed) as illustrated in Fig. 6.41. Unfortunately,

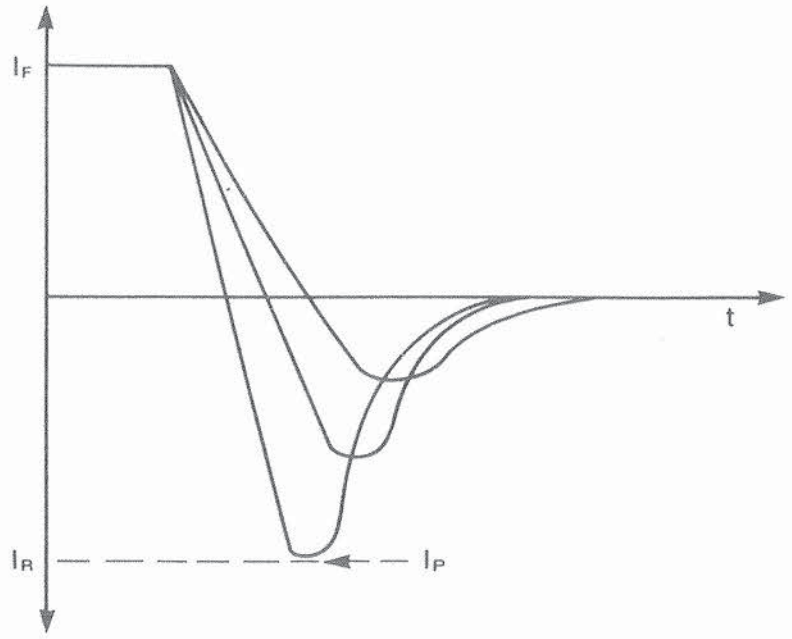


Fig. 6.41. Increase of peak reverse recovery current of *P*–*i*–*N* rectifier with increasing (di/dt).

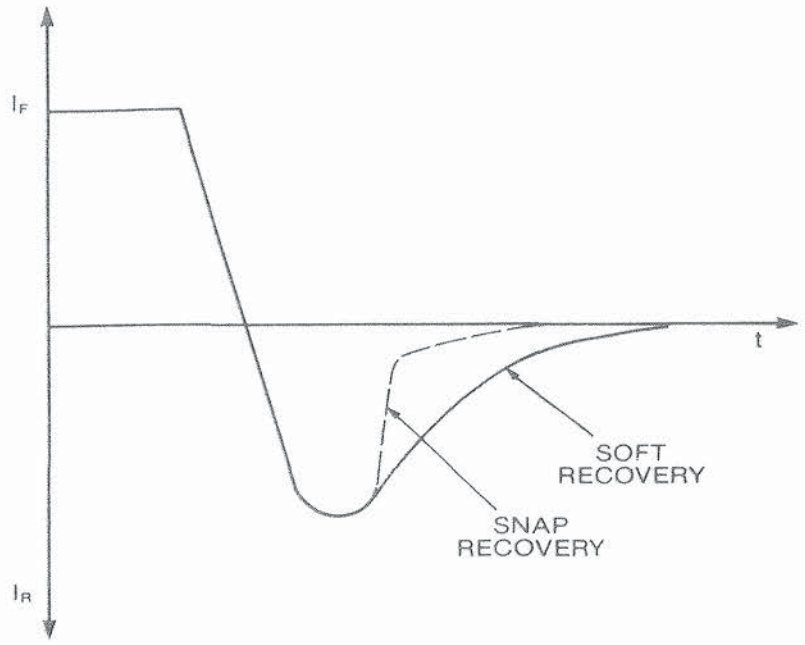


Fig. 6.42. Comparison of the snap-reverse recovery of a diode with soft reverse recovery.

this current flows through the transistors in the circuit imposing power dissipation and thermal stress on them. In addition, some diodes can go through the reverse recovery with an abrupt change in current. This behavior is called *snappy recovery* and is compared with the preferred soft recovery characteristics in Fig. 6.42. The high rate of change of current during the snap–reverse recovery generates a very high voltage ($L_S\, di/dt$) across the stray inductance L_S, which was discussed in Section 6.5.1. This voltage appears across the power MOSFET terminals and can take the device outside its safe operating area.

6.7.1. Control of Switching Speed

Improvement of the reverse recovery characteristics of the integral diode in power MOSFETs was first accomplished by using electron irradiation [24]. Electron irradiation can be used to introduce recombination centers into the drift region as discussed in Chapter 2. During reverse recovery, the resulting lifetime reduction can greatly reduce both the reverse recovery charge and the recovery time. The improvement that can be achieved by using the electron irradiation process is illustrated for a typical device in Fig. 6.43. An important feature of the process is the excellent control over the reverse recovery characteristics obtained by the ability to choose the radiation dose precisely. The reverse recovery time can be tailored from 350 nsec for the as-fabricated devices to less than 100 nsec at a dose of 16 Mrads. Since even good discrete P–i–N rectifiers exhibit a reverse recovery time of over 200 nsec, the radiation process can improve the power MOSFET integral diode to the level that it can replace the external flyback diode. Another important observation with regard to the

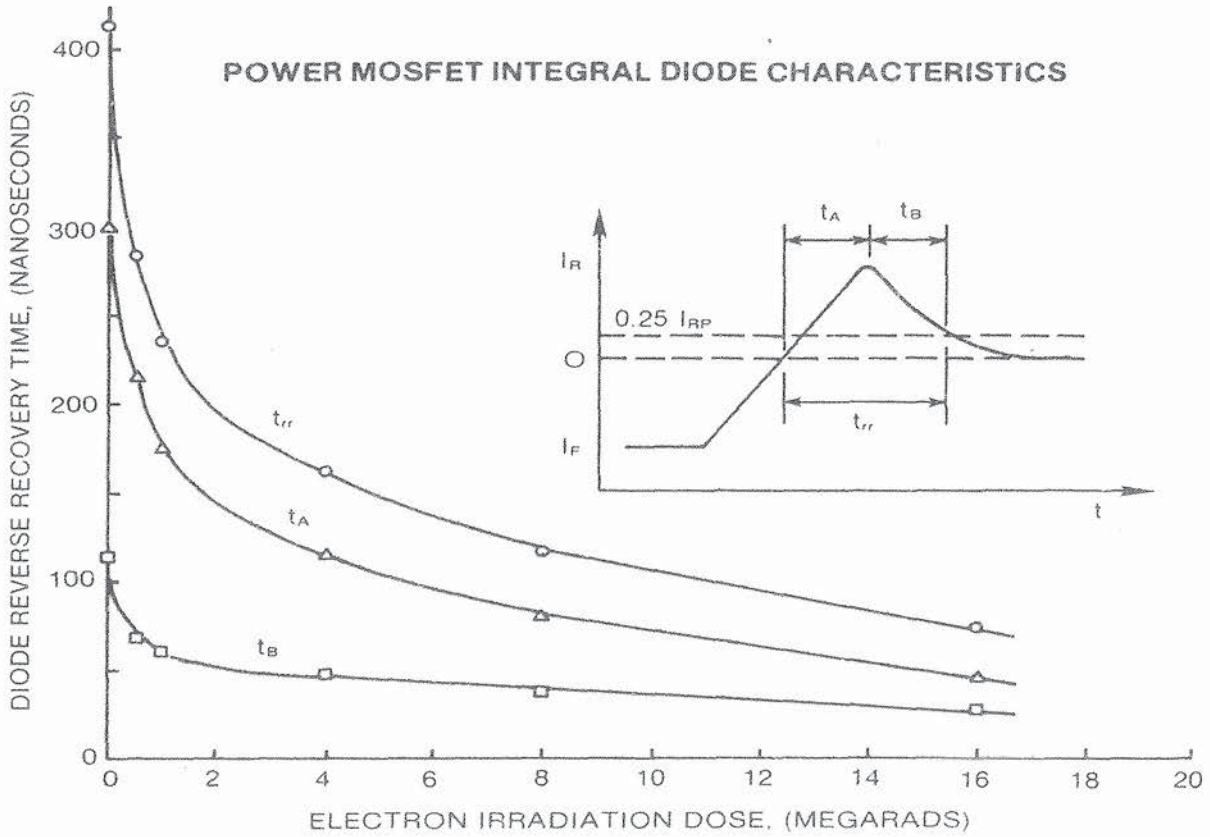


Fig. 6.43. Measured improvement in reverse recovery time with electron irradiation. The inset shows the reverse recovery waveform with the times t_A and t_B indicated. (After Ref. 24, reprinted with permission from *Solid State Electronics*.)

electron irradiation process is that the ratio of the time intervals for the two segments (t_A/t_B) shown in the inset of Fig. 6.43 remains approximately constant at a value of less than 2 for all radiation doses. This is an indication that no snappy recovery phenomenon occurs as the diode speed is increased by the electron irradiation.

It should be noted that the electron irradiation process introduces a positive charge in the gate oxide [24]. This causes a reduction in the threshold voltage of *n*-channel devices and an increase in the threshold voltage of *p*-channel devices. It has been observed that the threshold voltage can become negative for *n*-channel devices after electron irradiation, causing a drastic reduction in the forward blocking capability. Fortunately, the oxide charge due to electron irradiation can be annealed out at relatively low temperatures (150–200°C). At these temperatures, the recombination centers in the bulk that improve the reverse recovery of the integral diode are stable [25]. Other power MOSFET parameters remain essentially unaffected by the electron irradiation process. Consequently, it is an excellent method for obtaining power MOSFETs with high-quality integral diodes because electron irradiation, and subsequent annealing, can be performed after complete device fabrication in a highly controlled manner.

The speed of the integral diode of power MOSFETs can also be improved by gold or platinum doping [26]. A problem observed when using this process is an increase in the on-resistance at high gold and platinum diffusion temperatures. This is caused by compensation because of deep levels arising from the gold or platinum doping. The compensation becomes apparent when the concentration of the deep level becomes comparable to the background doping level. Since the capture cross section for the platinum deep level responsible for minority carrier lifetime reduction is larger than for gold, it produces less compensation for a given reverse recovery speed. Experimental comparison of gold and platinum doping [26] indicates that platinum doping will provide a 50% lower reverse recovery time when compared with gold for the same on-resistance. It is worth pointing out that the gold or platinum doping processes must be performed during device fabrication prior to contact metallization. This provides much less control over the minority carrier lifetime compared with electron irradiation leading to a wider spread in device characteristics. Furthermore, the platinum and gold atoms tend to segregate to regions of the device with strain such as the N^+ substrate and under the gate oxide. This phenomenon can severely degrade the MOS characteristics. The absence of the segregation effect in the case of electron irradiation makes the compensation effect negligible.

6.7.2. Bipolar Transistor Conduction

A potential problem with the use of the integral diode in the power MOSFET arises from the existence of the bipolar transistor in the structure. The equivalent circuit for the integral diode contains this bipolar transistor as shown in Fig. 6.44. The primary diode current flow path is shown by I_D in this figure. This current flow produces a voltage drop across the base resistance R_B that forward-biases the emitter–base junction of the bipolar transistor. If the base resistance

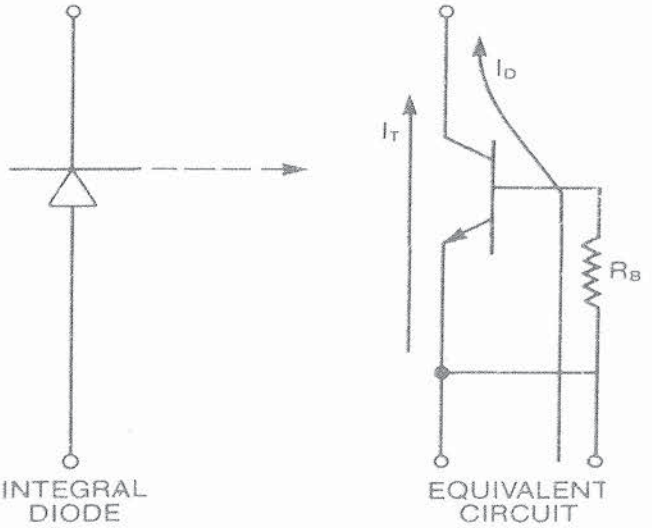


Fig. 6.44. Equivalent circuit of the integral diode showing the bipolar transistor.

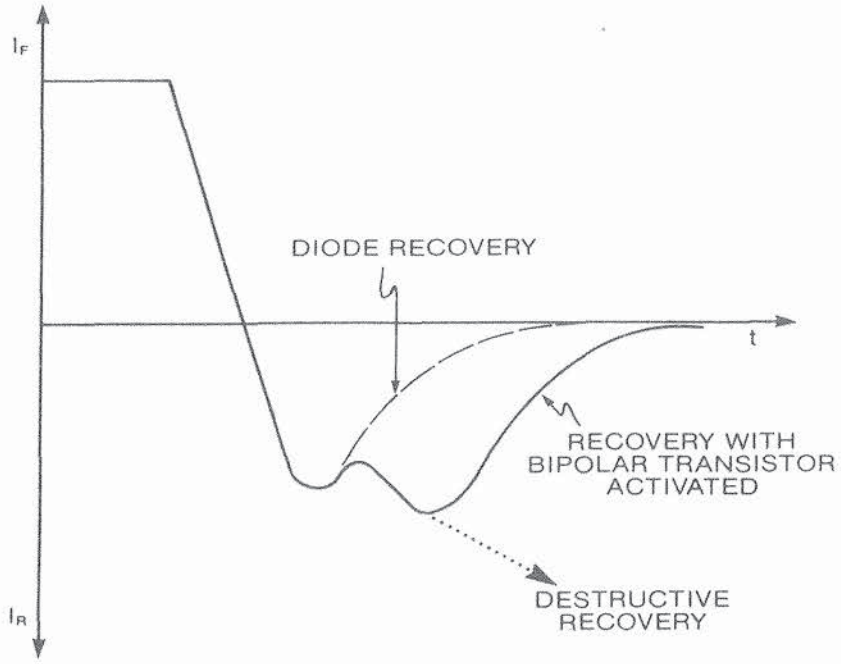


Fig. 6.45. Reverse recovery waveforms observed when the parasitic bipolar transistor is activated.

is not sufficiently low, the bipolar transistor will turn on and a significant current can flow along path I_T because even in the inverse mode the bipolar transistor has some gain. When the diode in the structure goes through its reverse recovery, the high voltage across the transistor can result in its going into second breakdown, leading to catastrophic failure.

The activation of the bipolar transistor during reverse recovery of the integral diode can be observed in the reverse recovery waveform. A typical example of the reverse recovery current waveform with an activated bipolar transistor is provided in Fig. 6.45. If second breakdown occurs, the recovery will proceed according to the dotted line shown in Fig. 6.45, leading to destructive failure. This mode of failure is aggravated by the voltage surge caused by snap recovery of the integral diode because this not only produces a high voltage at the drain but imposes a high (dV/dt) across the device terminals. This high (dV/dt) can precipitate further bipolar action as discussed in Section 6.5.2, leading to destructive failure.

6.8. NEUTRON RADIATION TOLERANCE

Bipolar transistors exhibit severe degradation in electrical performance after being subjected to neutron radiation. In these bipolar devices, the severe reduction in the minority carrier lifetime as a result of the neutron radiation damage, destroys the current gain and drift region modulation. In comparison, power MOSFETs are majority carrier devices for which the minority carrier lifetime is of consequence only in terms of the leakage current. Consequently, these devices can be expected to exhibit much superior neutron radiation tolerance.

6.8.1. On-Resistance

The influence of high energy or fast neutron irradiation on the on-resistance of two typical power MOSFETs is shown in Fig. 6.46 [23]. The on-resistance remains relatively constant until the dose exceeds 10^{13} neutrons/cm^2. In contrast, even a highly neutron radiation resistant bipolar device, such as the field-controlled diode, exhibits a sharp increase in forward voltage drop at neutron influences about 10 times lower than those observed for the power MOSFETs. The reason for this difference in performance is because the on-resistance of power MOSFETs increases only when the deep levels caused by the neutron damage begins to compensate the donors in the drift region and raise its resistivity. The deep-level concentration required to achieve compensation is comparable to the drift region doping level. As a result, power MOSFETs with lower breakdown voltage will exhibit higher neutron radiation tolerance as shown in Fig. 6.46 because their drift region is more heavily doped. In contrast, for a bipolar device such as the FCD, the forward drop will increase sharply when the lifetime (diffusion length) begins to decrease below a critical value. This occurs at neutron fluences well below the point at which compensation is significant.

The observed increase in the power MOSFET on-resistance due to neutron radiation produces an increase in conduction losses for power switching application but has little consequence in linear applications. The sizing of the device

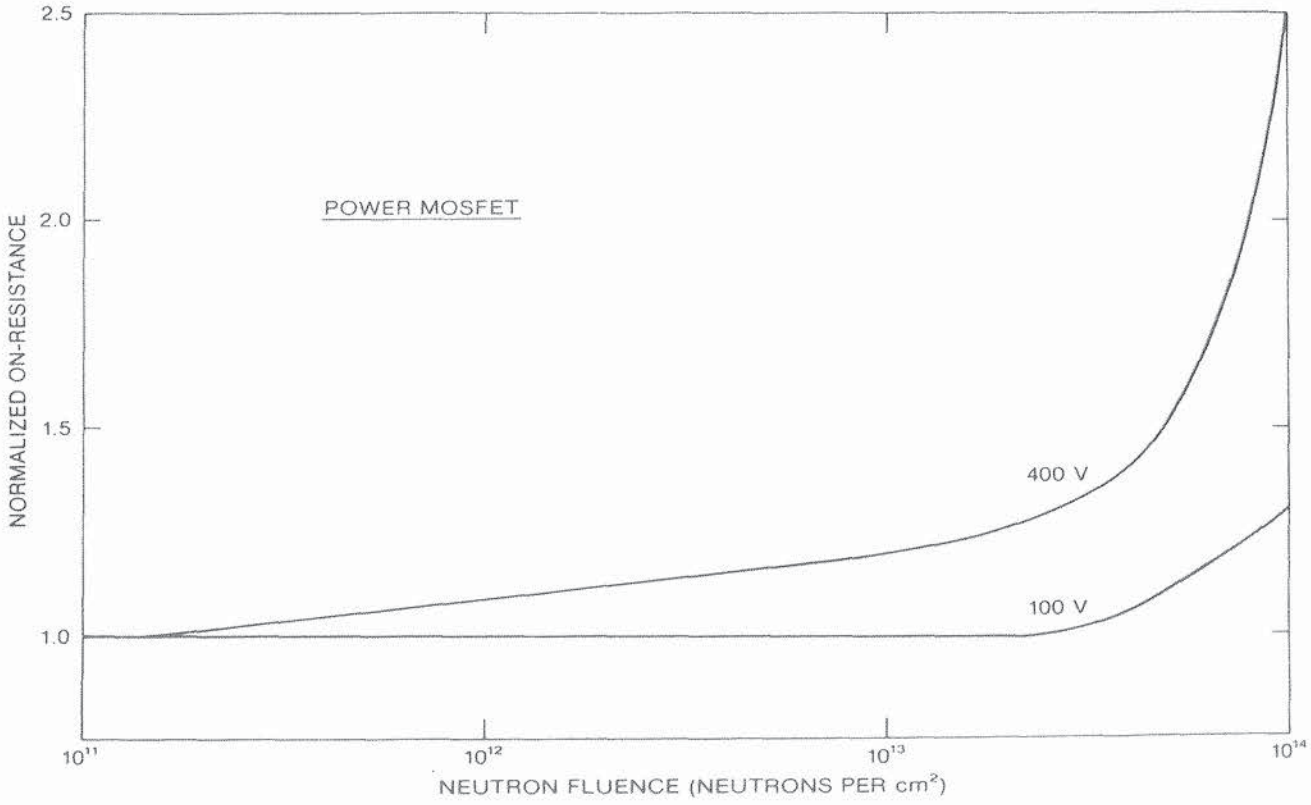


Fig. 6.46. Impact of neutron radiation on the on-resistance of power MOSFETS.

and its heak sink must be initially designed to ensure operation within the maximum safe operating area after radiation. In linear applications, the transconductance is of greater importance. This parameter is affected only when the mobility begins to degrade. The neutron fluences at which the mobility is reduced are about an order of magnitude higher than those at which the on-resistance increases.

6.8.2. Threshold Voltage Shift

The major problem with irradiation of power MOSFETs arises from the introduction of positive charge in the gate oxide by the fast neutron radiation damage. The positive charge in the gate oxide causes decrease in the threshold voltage of *n*-channel power MOSFETs and increase in the threshold voltage of *p*-channel power MOSFETs. A typical example of the variation of the threshold voltage with radiation dose is provided in Fig. 6.47. Following the radiation, *n*-channel power MOSFETs can exhibit negative threshold voltage, that is, normally-on characteristics. It is possible to operate these devices by designing the gate power supply to provide a negative gate drive voltage during the blocking state. Similarly, for *p*-channel devices, the impact of the radiation on the threshold voltage can be accounted for by providing a high gate drive voltage.

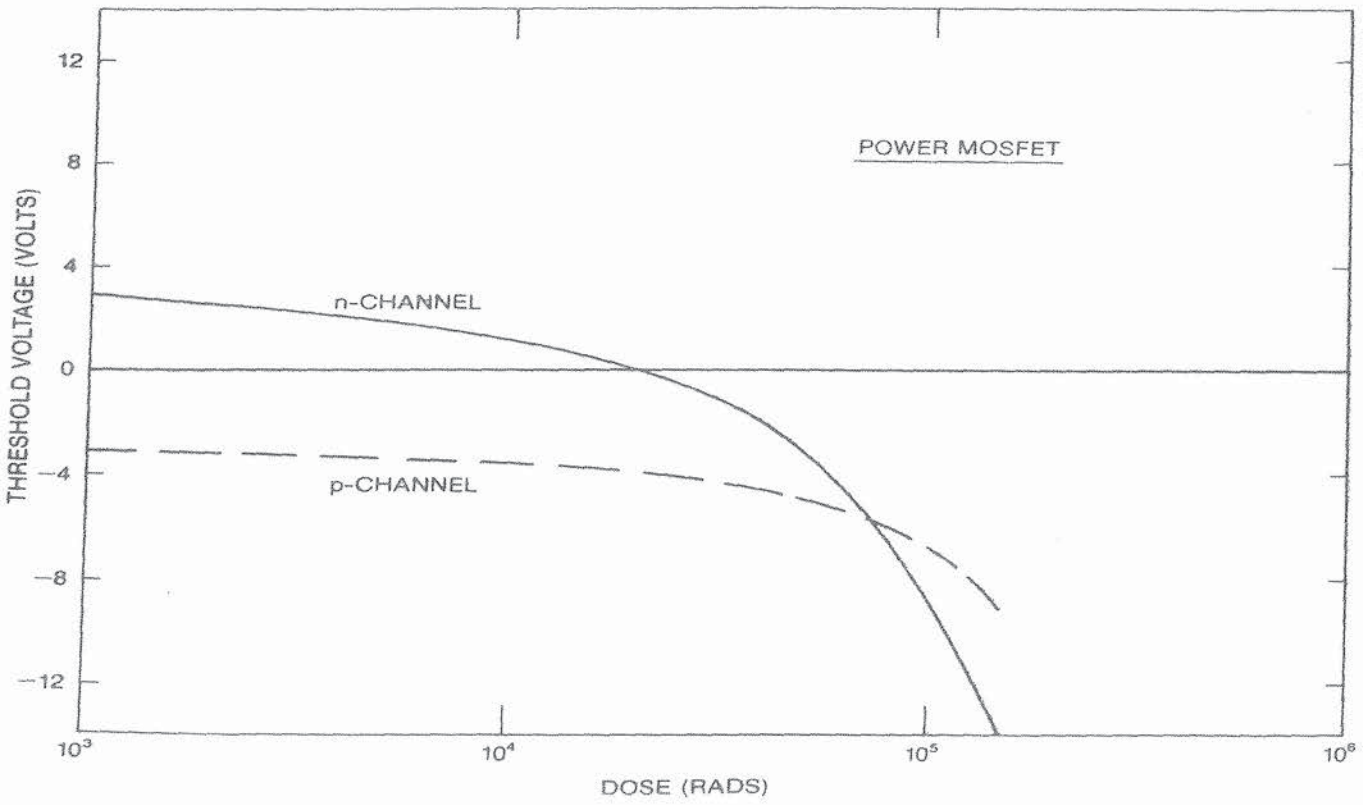


Fig. 6.47. Change in the threshold voltage of power MOSFETs due to radiation.

6.9. HIGH-TEMPERATURE PERFORMANCE

Power MOSFETs exhibit very good high-temperature operating characteristics. Devices are commercially available with peak junction temperatures ratings as high as 200°C. The ability of a power MOSFET to operate at elevated temperatures is related to the absence of minority carrier injection. The effect of increasing temperature on various device parameters is discussed in this section.

6.9.1. On-Resistance

An important merit of the power MOSFET is an increase in its on-resistance with increasing temperature. Although this may appear at first sight to be an undesirable feature because of an increase in power dissipation, it imparts an important benefit in terms of device stability and allowing the operation of devices in parallel. Local variations in drift region resistivity are unavoidable during the fabrication of power devices. Some locations inside the device will contain a lower on-resistance compared with the rest of the device. This promotes localization of current along the paths of lowest resistance. Localization of current produces additional heating. Fortunately, in the case of the power MOSFET, the mobility for holes and electrons decreases with temperature, causing the local resistivity to increase. This will tend to homogenize the current distribution and prevent thermal runaway. Furthermore, when power MOSFETs are connected in parallel, the increase in their on-resistance with temperature promotes good current sharing without the need for external ballasting, which is required for bipolar transistors.

A typical example of the increase in the on-resistance of a power MOSFET with temperature is provided in Fig. 6.48. This curve closely follows the mobility change with temperature discussed in Chapter 2. A general expression that can be used to predict the variation of the on-resistance of p- and n-channel power MOSFETs is

$$R_{on}(T) \simeq R_{on}(25^{\circ}\text{C}) \left(\frac{T}{300}\right)^{2.3} \tag{6.124}$$

where T is the absolute temperature.

6.9.2. Transconductance

The reduction in the mobility with increasing temperature adversely affects the transconductance of power MOSFETs. A typical example of the variation of the transconductance is provided in Fig. 6.49. The transconductance will follow the mobility variation because the other terms that determine its value are relatively unaffected by temperature:

$$g_m(T) \simeq g_m(25^{\circ}\text{C}) \left(\frac{T}{300}\right)^{-2.3} \tag{6.125}$$

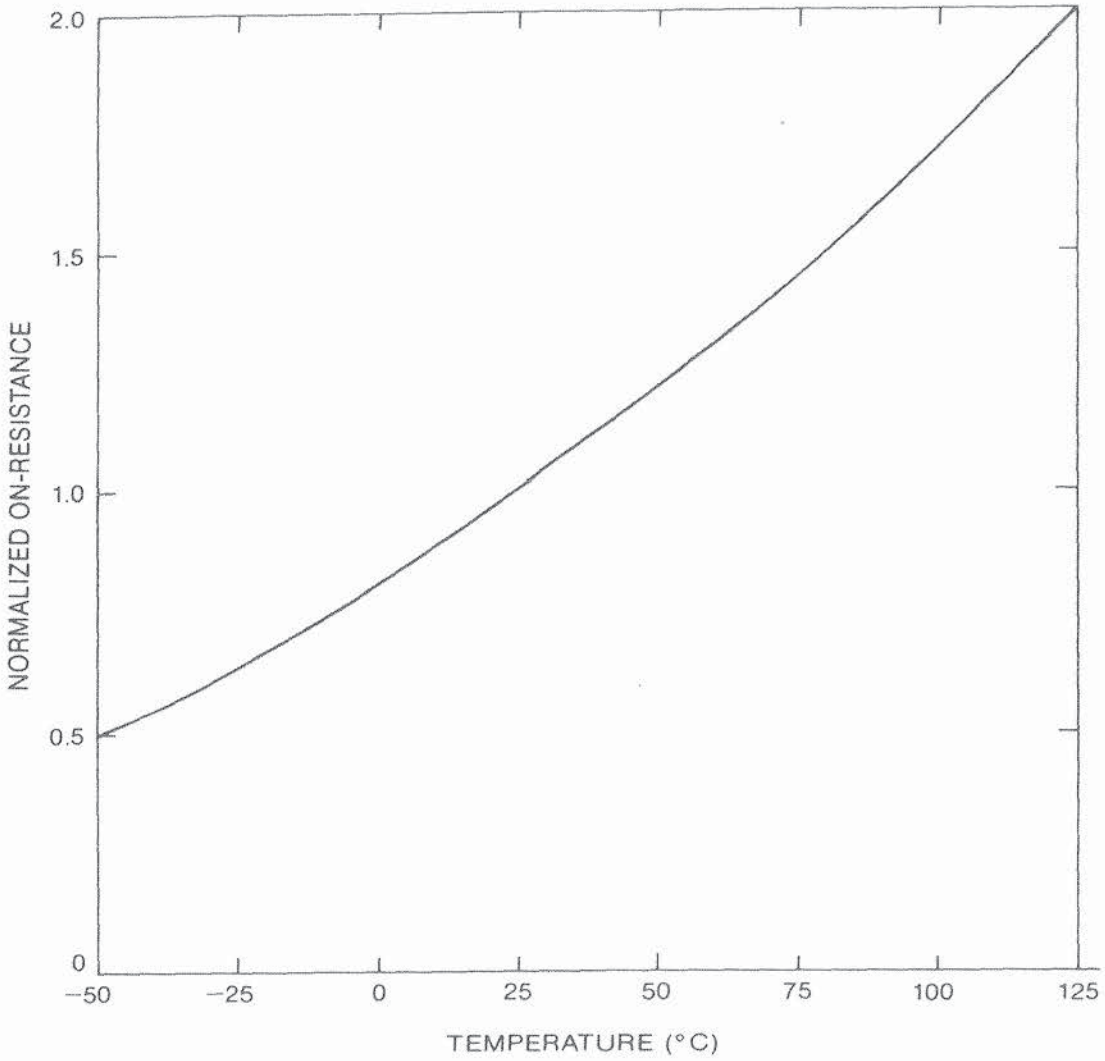


Fig. 6.48. Increase in power MOSFET on-resistance with temperature.

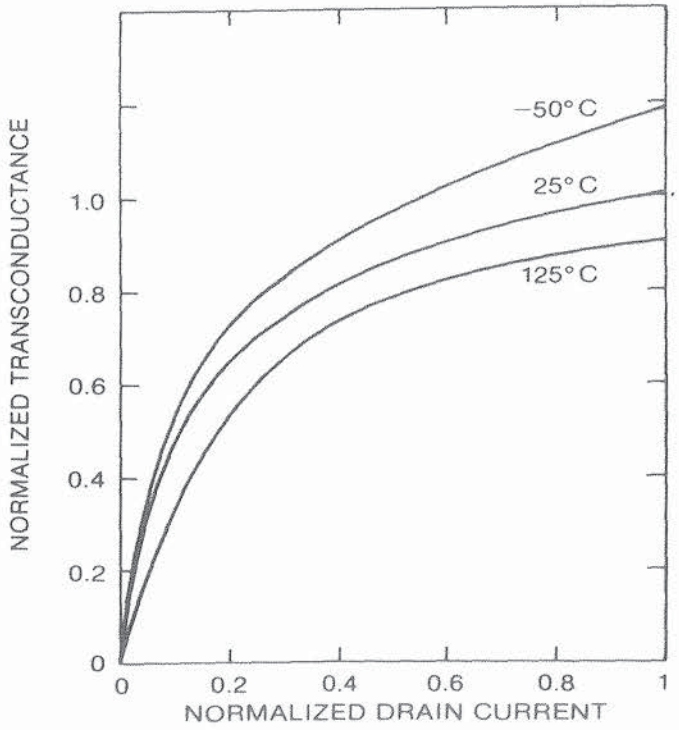


Fig. 6.49. Decrease of power MOSFET transconductance with increasing temperature.

Care must be taken during device design to provide sufficient channel periphery to allow for the degradation in the transconductance if high-temperature operation is anticipated.

6.9.3. Threshold Voltage

In Section 6.3.2 it was shown that the drain current of a power MOSFET will vary as the square of the gate drive voltage ($V_G - V_T$) in the saturated current regime. A plot of the square root of the drain current as a function of the gate voltage can, therefore, be used to obtain a measurement of the threshold voltage. A typical example of this transfer characteristic is shown in Fig. 6.50. The measured drain current indeed varies as the square of the gate voltage, resulting in the straight lines. The intercept of these lines at $I_D = 0$ provides the threshold

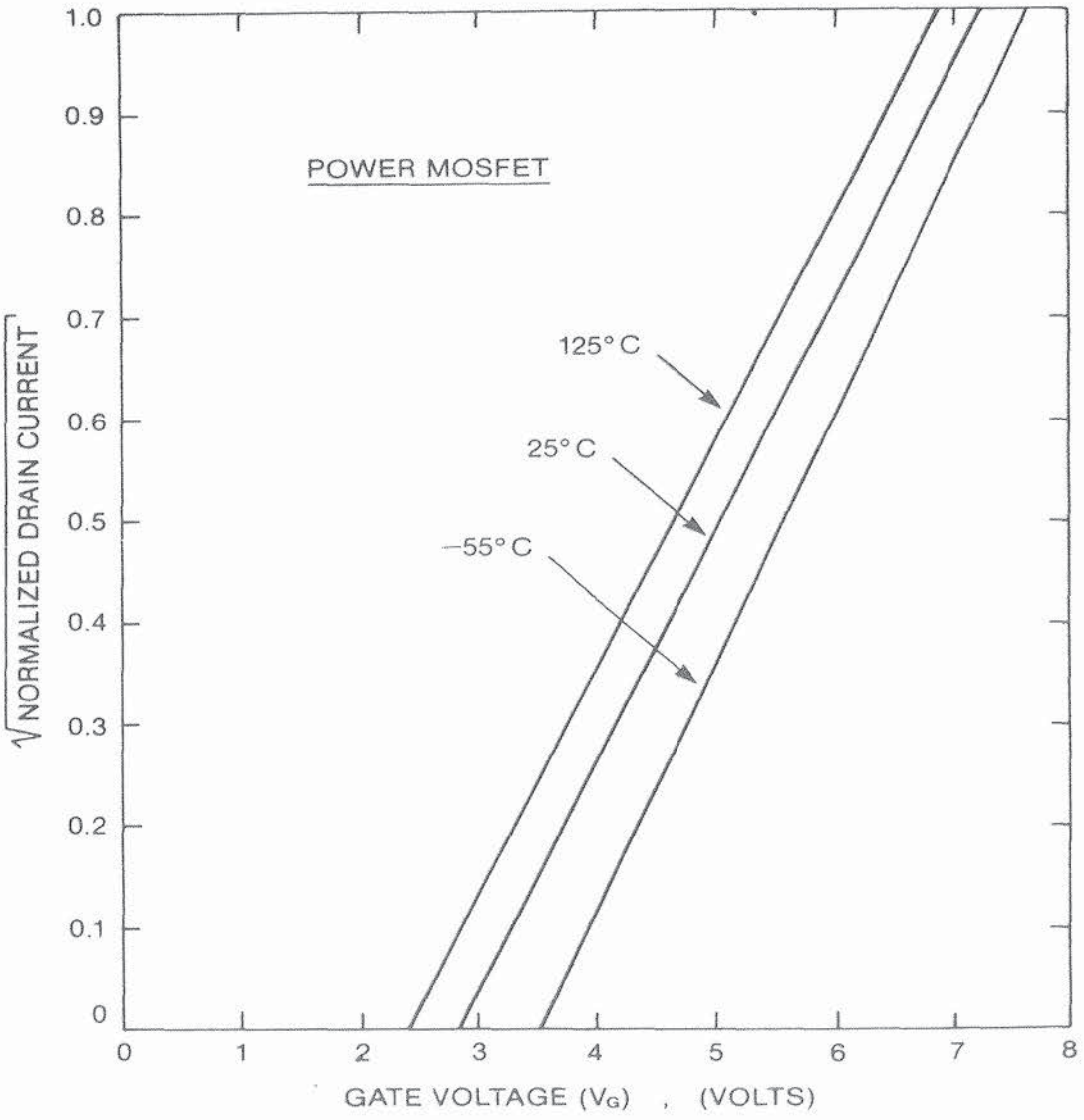


Fig. 6.50. Transfer characteristics of power MOSFET at different temperatures.

voltage. It can be seen that the threshold voltage decreases with increasing temperature.

In the absence of a work-function difference and the oxide charge, it was shown in Section 6.3.2.2 that the threshold voltage is given by

$$V_T = \frac{\sqrt{4\epsilon_s q N_A \psi_B}}{C_{ox}} + 2\psi_B \tag{6.126}$$

Differentiating this expression with respect to temperature, it can be shown that

$$\frac{dV_T}{dt} = \frac{d\psi_B}{dt}\left[\frac{1}{C_{ox}}\sqrt{\frac{\epsilon_s q N_A}{\psi_B}} + 2\right] \tag{6.127}$$

This expression is valid even in the presence of a work-function difference and oxide charge because these parameters are not strongly affected by the temperature. The bulk potential ψ_B varies with temperature because the energy gap changes [27]:

$$\frac{d\psi_B}{dt} \simeq \frac{1}{T}\left\{\frac{E_g(T=0)}{2q} - |\psi_B(T)|\right\} \tag{6.128}$$

From these expressions, the rate of variation of the threshold voltage with temperature can be calculated as a function of the background doping level and oxide thickness. A plot of (dV_T/dT) as a function of background doping is provided in Fig. 6.51 for the $Si–SiO_2$ system using gate oxide thickness as a parameter [27]. Note that the threshold voltage will vary at a higher rate for thicker gate oxides and higher substrate doping levels. During the design and

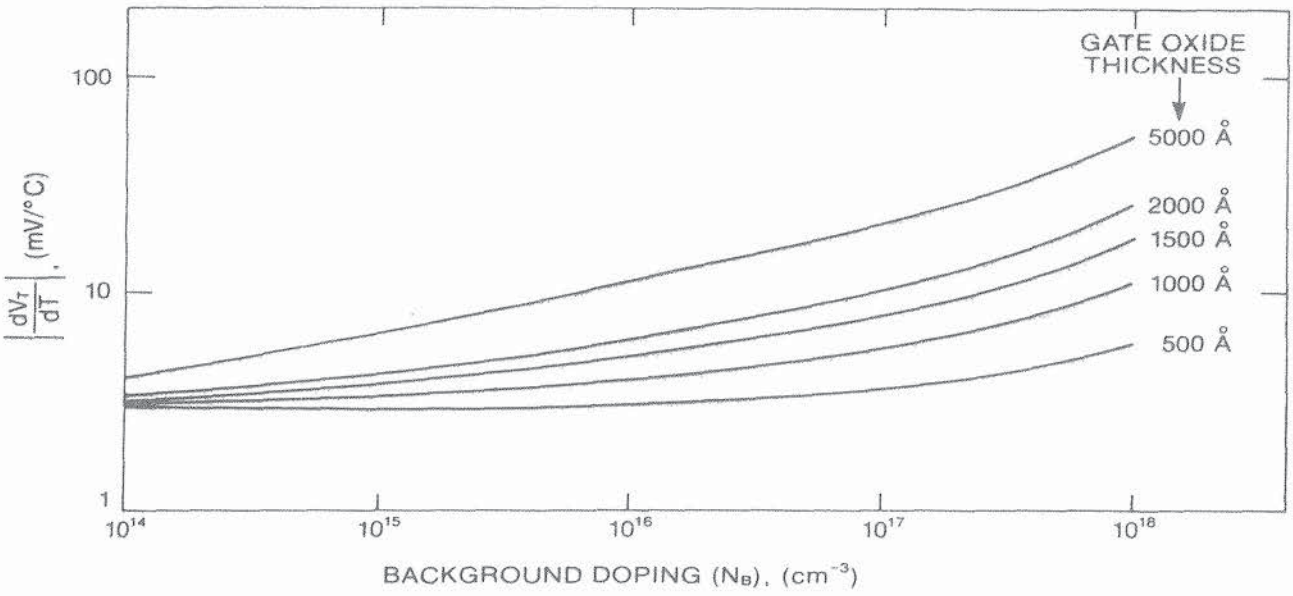


Fig. 6.51. Effect of substrate doping N_B and gate oxide thickness on the rate of change of threshold voltage with temperature. (After Ref. 27, reprinted with permission from the IEEE © 1971 IEEE.)

fabrication of power MOSFETs, it is important to provide an adequate room temperature threshold voltage which will allow operation at high temperatures with sufficient margin to ensure noise immunity and protection from inadvertent turn-on at high (dV/dt) transients.

6.10 DEVICE STRUCTURES AND TECHNOLOGY

Two fundamentally different processes have evolved for the fabrication of discrete vertical-channel power MOSFETs. One of these processes utilizes chemical etching of the silicon surface to form the V-groove structure shown in Fig. 6.1*b*, whereas the other processes relies on the planar diffusion of the P base and N^+ emitter from a common window defined by the refractory gate electrode to form the DMOS structure shown in Fig. 6.1*a*. Because of the anisotropic etch used to form the VMOS device structure, the device topology is restricted to the $\langle 110 \rangle$ directions along the surface of the (100)-oriented wafers. This restriction does not apply to the DMOS structure, allowing optimization of surface layout. In this section, the impact of surface topology is discussed after providing a description of the device processing technology.

6.10.1. DMOS (Planar) Structure

For the fabrication of the planar power MOSFET structure shown in Fig. 6.1*a*, the wafer orientation can be chosen to provide the highest surface mobility to achieve a low channel contribution to the on-resistance. For this reason, these devices are fabricated with the use of (100)-oriented wafers. Because of the extremely rapid increase in on-resistance with increasing breakdown voltage, power MOSFETs are confined to blocking voltages of less than 1000 V. The ideal depletion width for these devices is less than 100 μm. They must be fabricated by growing lightly doped epitaxial layers (typically 5–50 μm thick) on heavily doped substrates. The discussion here is confined to the case of n-channel devices. The fabrication of complementary (p-channel) devices is similar. In the case of n-channel devices, the N^+ substrates are antimony doped with a resistivity of 0.01 $\Omega \cdot$cm. Antimony is chosen as the dopant because its autodoping effects are small during epitaxial growth. For low on-resistance devices, arsenic-doped substrates with a resistivity of 0.001 $\Omega \cdot$cm are preferred.

After epitaxial growth, a thick field oxide is grown. This oxide is patterned to create the desired planar diffused edge termination. Sometimes the processing of the diffusion for the edge termination is combined with the DMOS cell fabrication. Following the termination diffusion, the device active area (central region where the DMOS cells are to be fabricated) is opened up. The gate oxide is then grown to a typical thickness of 1000 Å. Recently, device manufacturers have been scaling down the gate oxide thickness to obtain an increase in the transconductance so as to make it possible to drive the power MOSFETs at voltages compatible with low-voltage logic circuits.

A layer of heavily doped polysilicon (or other refractory gate material) is deposited next. After oxidizing the polysilicon, photolithography is used to pattern the polysilicon. A cross section of a small segment of the active region is shown at this point in the process in Fig. 6.52*a*. At this point in the process, the cell window has been defined. Boron is now implanted through the gate oxide to a dose (typically $10^{14}/cm^2$) that is necessary to obtain the desired threshold voltage in conjunction with the N^+-emitter diffusion. The boron is driven in with a wet oxidation step to obtain the structure shown in Fig. 6.52*b*. An oxide island is now patterned at the center of each cell window to mask the N^+ emitter diffusion. This oxide island leaves a portion of the *P*-base ac-

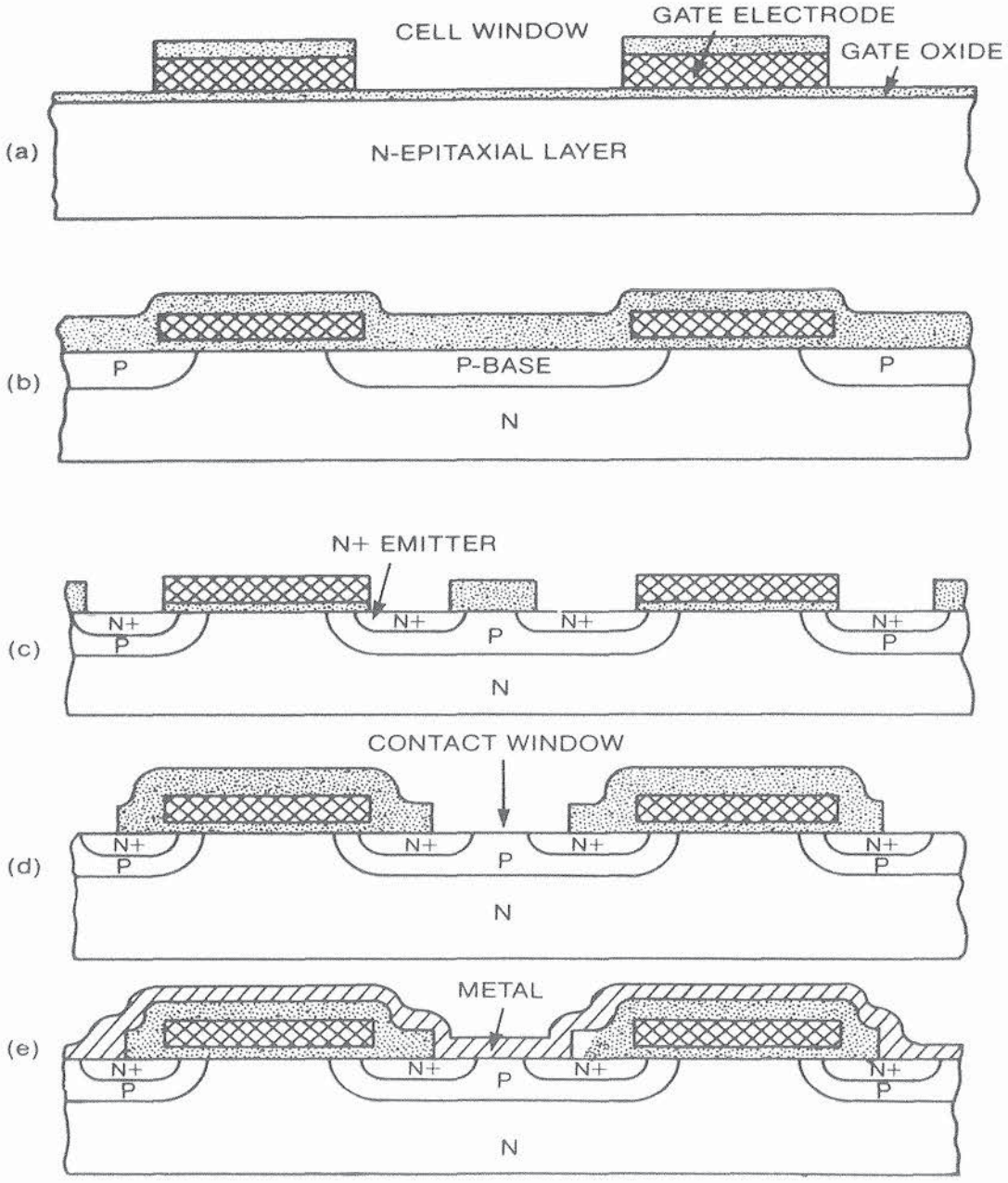


Fig. 6.52. Double-diffused MOSFET process sequence.

cessible at the surface in each cell, allowing good contacts to be subsequently made to it.

The patterning of the oxide island is one of the critical alignment steps during the DMOS process. If the oxide island is misaligned with respect to the edges of the polysilicon in the cell window, the length of the N^+ emitter will increase, causing degradation of the (dV/dt) and the safe operating area. After the oxide island is formed, the N^+ emitter is diffused by using either phosphorus predeposition or ion implantation. The latter process provides greater control over the emitter profile and threshold voltage. A typical dose is $1 \times 10^{15}/\text{cm}^2$. The device cross section at this point is shown in Fig. 6.52*c*. After the N^+ emitter, is driven in a thick layer of oxide is deposited by low pressure chemical vapor deposition (LPCVD). This provides a conformal insulating film over the top and sides of the polysilicon gate electrode.

The oxide layer is now patterned to form the contact windows as illustrated in Fig. 6.52*d*. This is another critical photolithographic step in the DMOS process because misalignment can cause either poor contact to the *P*-base region or lead to an overlap of the contact with the polysilicon at the cell edges. This overlap will produce a short circuit between source and gate. The final step in the device process consists of metallization to create the structure shown in Fig. 6.52*e*.

The DMOS process described here provides the fundamental basis for the fabrication of modern power MOSFETs. Many variations of the process have been explored in an attempt to either eliminate some of the critical alignment steps or to provide self-alignment of the N^+-emitter and contact windows.

6.10.2 VMOS (Nonplanar) Structure

Historically, the VMOS process was applied toward the fabrication of discrete, vertical-channel, power MOSFETs before the development of the DMOS process. This process relies on the anisotropic etching of silicon. It has been found that in certain chemical solutions, the (111) planes in silicon etch at a much slower pace when compared with other surfaces. If (100)-oriented silicon is etched through a window at the surface, these etches will create a V-shaped (or truncated V-shaped) groove as a result of the difference in etch rate between the (100) and (111) planes.

The VMOS (*n*-channel) process begins with the growth of a lightly doped *N*-type epitaxial layer on a (100)-oriented N^+ substrate. A boron diffusion is now performed across the entire wafer surface if the device termination is formed by using the V-groove etch step [28]. Alternately, the boron diffusion can be performed over the entire active area but patterned at the edges of the chip to form the desired device termination. A cross section of the active region at this point in the process is shown in Fig. 6.53*a*. The oxide layer grown during the *P*-base diffusion step is next patterned to define the N^+ emitter regions as illustrated in Fig. 6.53*b*. Note that the N^+ emitters extend between the cells where the V groove is be etched.

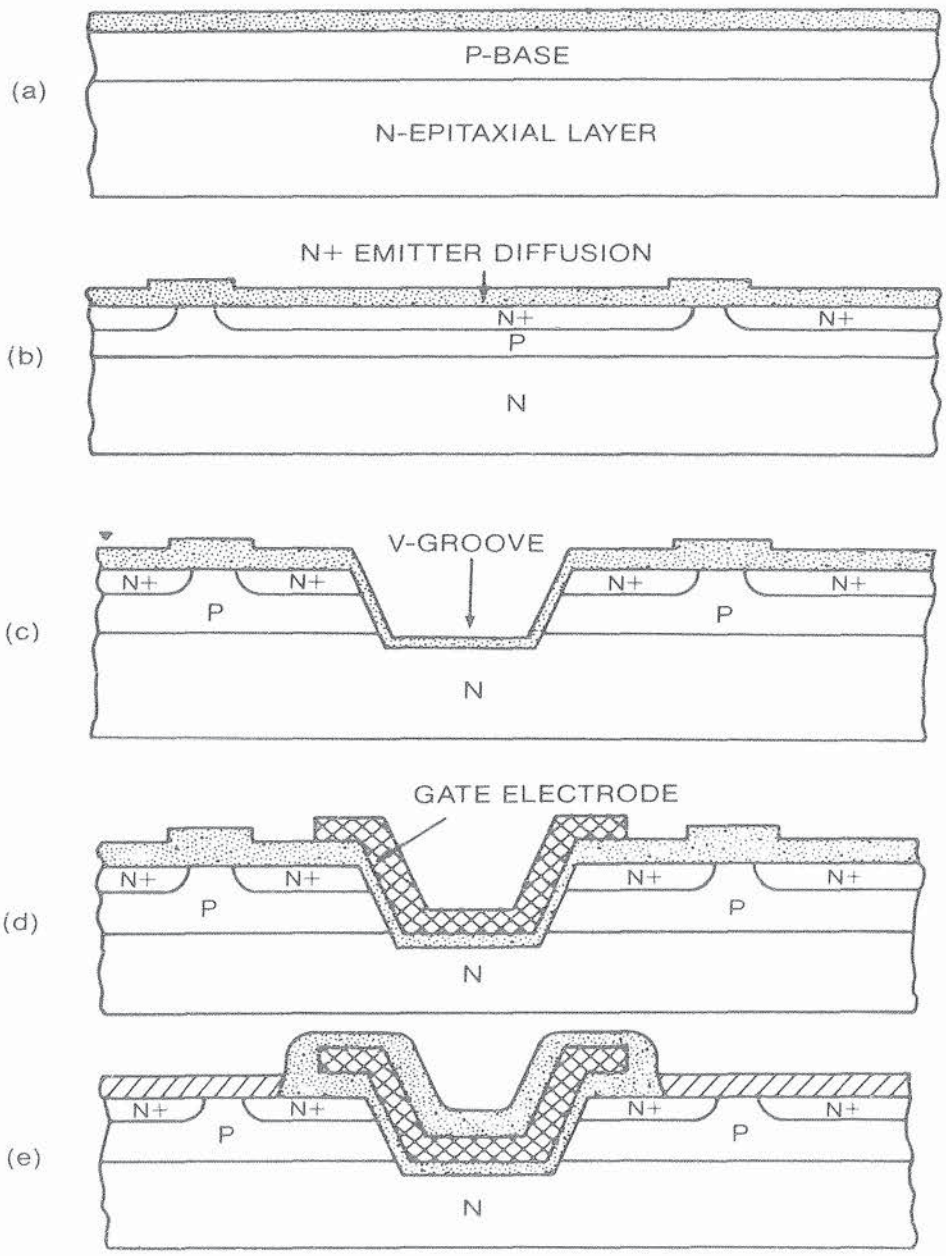


Fig. 6.53. V-Groove MOSFET process sequence.

The oxide layer grown over the wafer surface is now patterned to expose the silicon surface in areas to be etched. This patterning is an important alignment step because it determines the length of the N^+-emitter region and the relative position of the V groove with respect to the P-base contacts. The V groove is formed by using anisotropic etching to create the structure shown in Fig. 6.53*c* and must extend beyond the bottom of the P-base–N-drift layer junction. The etching is usually performed by using a mixture of potassium hydroxide and isopropanol [29]. Note that the window used to etch the grooves must be aligned along the ⟨110⟩ directions on the wafer surface because the (100) and (111) planes intersect along these lines. The V grooves extend down from the surface at an angle of 57.7° to the surface.

After forming the V grooves, the gate oxide must be grown in them. Since the presence of potassium in the gate oxide can cause severe instability of the threshold voltage, it is important to boil the wafers in hydrochloric acid (HCl) prior to gate oxide growth until all the potassium residue remaining from the V-groove etch step is removed. The polysilicon gate electrode is now deposited and patterned to form the structure shown in Fig. 6.53*d*. Another important alignment step is involved here because misregistration of the polysilicon with respect to the groove can lead to loss of channel periphery and an increase in the source–gate capacitance. Following the patterning of the polysilicon, an oxide layer is deposited by LPCVD. This layer is patterned to create the contact windows. These windows must also be carefully aligned to the cell structure as described for the DMOS structure. The source metallization is now evaporated and patterned.

Because of the nonplanar surface created by the V groove, the patterning of the polysilicon layer, contact windows and source metallization are more difficult than in the DMOS process. Poor resist coverage at the groove edges can lead to short circuits between source and gate if the source metal overlaps the gate (not shown in Fig. 6.53). It is also mandatory to eliminate even small traces of potassium from the wafer surface after etching the grooves to avoid instabilities in threshold voltage and drift in the breakdown voltages. These technological problems, together with the high electric field at the groove edges, have placed the VMOS structure at a serious disadvantage when compared with the DMOS structure. Although the first commercial power MOSFETs were manufactured with the VMOS process, most modern power MOSFETs are no longer fabricated with this process.

As in the case of the DMOS process, the VMOS process described here is intended to only provide a fundamental sequence for device fabrication. Many innovative techniques for simplifying the process and reduce the alignment problems have been attempted. One interesting variation of the process [30] utilizes isotropic etching of silicon to create a large oxide overhang as illustrated in Fig. 6.54. This overhang is utilized to obtain a selective deposition of the gate contact metal over the channel with a minimum overlapping of the source and drift regions. This is accomplished by evaporating a three-metal system (Cr–Ni–Au) at an angle to the wafer as indicated by the dashed lines in Fig. 6.54. This metal system was chosen because the evaporation of aluminum at

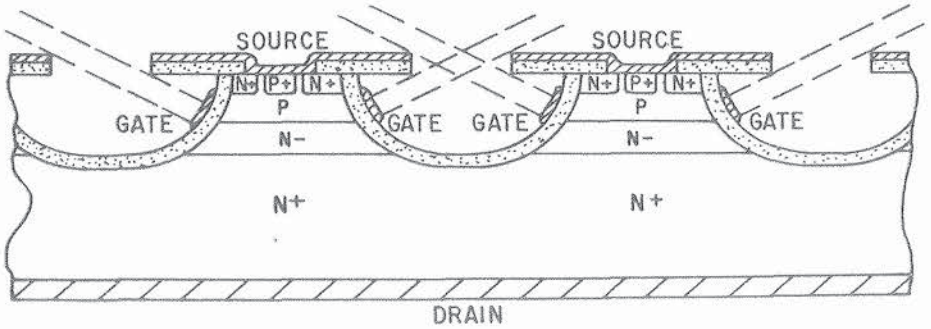


Fig. 6.54. Non-planar power MOSFET structure fabricated by using isotropic etching.

low angles of incidence has been observed to produce a "dull fibrous" film with an electrical resistivity substantially higher than that of normal evaporated aluminum films. Because of the low capacitance in this device structure, very high frequency performance has been achieved.

6.10.3. Gate Oxide Fabrication

Since the electrical characteristics of the MOSFET are modulated by the application of an electric field across the gate oxide, its properties play an important role in determining the device characteristics. A large amount of effort has, consequently, been expended in the semiconductor industry to develop an understanding of oxidation and annealing conditions on the oxide properties. The important process criteria for the growth of the gate oxide are provided in this section.

The silicon dioxide that is used as the insulating barrier between the gate metal and the channel can be grown by subjecting the silicon surface to an oxygen ambient at elevated temperatures. The ambients that have been most commonly used are either dry oxygen or oxygen containing water vapor. Lower interface state densities have been reported when the oxidation ambient contains water vapor [31]. However, the slower oxidation rate in dry oxygen allows greater control over the gate oxide thickness. The higher interface density in this case can be reduced by increasing the oxidation temperature and subsequently annealing the oxide, as described below.

Before considering the effects of annealing on the oxide–silicon interface, it is worth pointing out that the lowest fixed interface charge Q_{FC} and surface state density have been observed for the (100) plane [31, 32]. For the devices made by the anisotropic etching technology, the gate oxide must be formed on a (111) surface since this surface is exposed by the preferential etch. In contrast, the devices made by planar diffusion techniques (DMOS) can be fabricated from (100)-oriented wafers to obtain a lower fixed interface charge and surface state density.

Low-temperature annealing of the oxide films in either a mixture of hydrogen and nitrogen or a nitrogen ambient containing water vapor has proved to be very effective in decreasing the interface state density. A 1000°C anneal in dry nitrogen must be performed prior to the preceding low-temperature annealing to achieve a minimum interface state density [33]. It is believed that during the subsequent low-temperature anneal, the hydrogen in the oxide diffuses to the oxide–silicon interface and ties up the dangling bonds, reducing the surface state density. The optimum annealing temperature is about 400°C, with a low surface state density observed after 60 min of annealing time. A similar annealing of the interface states can be accomplished after the aluminum gate metallization has been applied. A faster annealing rate has been observed in this case. This is believed to arise from the generation of hydrogen at the aluminum by the reduction of water vapor in the oxide. This process has the drawback of

the possible introduction of mobile ions into the silicon dioxide from the aluminum, which can lead to an instability in the threshold voltage.

A drift in the threshold voltage due to the migration of mobile ionic species in the oxide has been one of the most serious problems in MOS gated FETs. This is particularly true with devices made by the preferential etching of the V groove with a potassium hydroxide–isopropanol solution. In this case, the complete removal of potassium ions from the silicon surface prior to the growth of the gate oxide is imperative. The high mobility of alkali ions in silicon dioxide at relatively low temperatures (in the range of 200–300°C) has been demonstrated conclusively [34, 35]. The migration of these alkali ions can be suppressed by either doping the oxide with phosphorus or capping it with a layer of silicon nitride. However, polarization effects [36] in the phosphosilicate glass films and the trapping of charge at the oxide–nitride interface have prevented the application of these techniques to the gate area of power MOSFET devices.

A more promising process innovation has been the addition of halogens to the gas stream prior to and during oxidation. The exposure of the oxidation tube to a mixture of HCl and dry O_2 has been found to be effective in reducing the mobile ion contamination of oxides grown subsequently in the tube [37]. The mobile ion contamination in the oxide has also been found to decrease with increasing HCl concentration in the gas stream during oxidation. However, it is important to keep in mind that HCl will etch silicon at the oxidation temperature if its concentration exceeds a few percent. Typically, an HCl concentration of about 1% by volume in the gas stream is effective in suppressing ionic contamination. The addition of chlorine during oxidation has also been reported to lower the interface state density and the fixed oxide charge density and to improve the dielectric breakdown strength of the oxide [37–39]. It is interesting to note that these improvements are observed only during dry oxidation and that the addition of HCl during oxidation with water vapor has been found to have no effect on the oxide properties.

In addition to the effect of the oxide thickness on the gate threshold voltage, the growth of the oxide over the channel can also affect the threshold voltage by the redistribution of the boron in the channel during oxidation. Experiments conducted on the thermal oxidation of silicon doped with gallium, boron, and indium have found that the surface concentration of these impurities decreases after oxidation [40]. This occurs because these impurities segregate into the oxide, thus depleting the silicon surface. Impurity segregation can seriously alter the diffusion profile of boron in the *P* base of both the DMOS and VMOS devices and, thus, influence the threshold voltage as well as the penetration of the depletion layer into the *P* base when the device is blocking current flow. Several models have been developed to allow the calculation of the diffusion profile of impurities with inclusion of the impurity redistribution during oxidation [41]. These models have also been applied to the calculation of the effect of oxidation on the threshold voltage of MOSFETs transistors [42]. In addition, the strain at the surface under the polysilicon gate of the DMOS structure

alters the lateral diffusion of impurities. It has been found that the boron diffusion is retarded whereas the phosphorus diffusion is enhanced. This effect can severely alter the channel length of both *n*- and *p*-channel DMOS power MOSFETs.

6.10.4. Cell Topology

In previous discussions of power MOSFETs in this chapter, cross sections of devices were shown without consideration of the cell layout on the surface. For the VMOS device structure, the anisotropic etching of the grooves can be accomplished only when the grooves are oriented along the ⟨110⟩ directions on the (100) wafer surface. This constraint allows only two design options for the surface topology of VMOS devices: long stripes or rectangular cells with V grooves running perpendicular to each other. Even the latter option is impractical because of the poor contour at the corners where the V grooves intersect at right angles [43]. Therefore, V-groove power MOSFET cells are confined to long fingers with interdigitation of source and gate [15].

In contrast, the DMOS structure allows any conceivable cell topology as long as it meets all technological constraints such as alignment tolerances. The cell windows that are commonly used for the design of power MOSFETs have square, circular, and hexagonal shapes. These windows can be located in either a square or hexagonal cell pattern as illustrated in Fig. 6.55. The impact of the layout of these cells, as well as the linear cell shown in Fig. 6.56(a), on the resistance has been analyzed [44]. It has been shown that the on-resistance of all

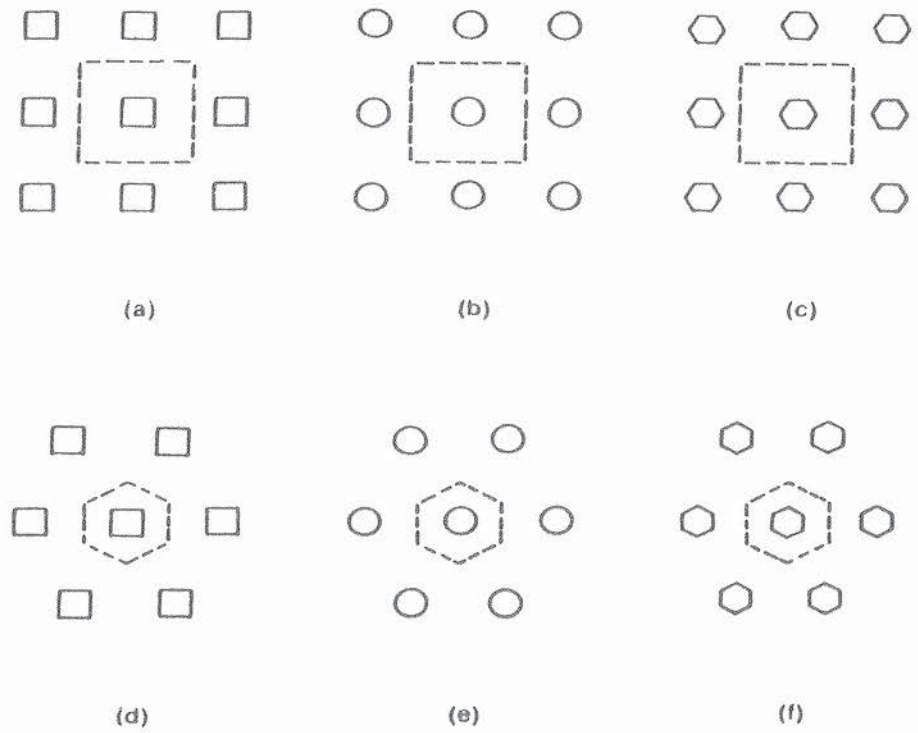


Fig. 6.55. Various cellular designs for power MOSFETs: (*a*) square window, square cell; (*b*) circular window, square cell; (*c*) hexagonal window, square cell; (*d*) square window, hexagonal cell; (*e*) circular window, hexagonal cell; (*f*) hexagonal window, hexagonal cell.

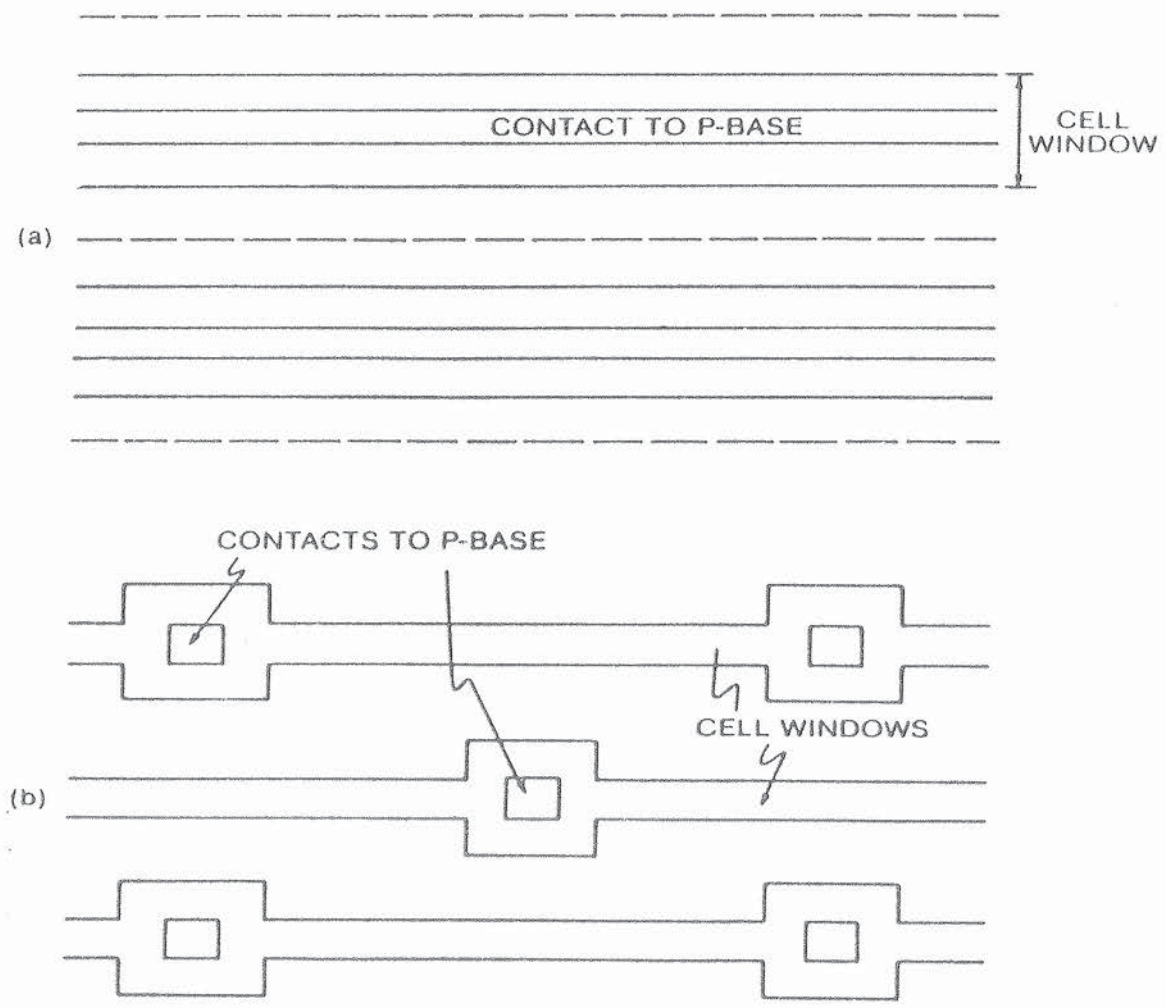


Fig. 6.56. Comparison of linear cell design with (*a*) continuous contact and (*b*) broken contact to *P*-base region.

cellular designs are equal if the size of the cell window and the ratio of the area of the cell window to the total cell area are kept the same. The on-resistance of the cellular designs shown in Fig. 6.55 is smaller by a factor of 1.6 than a linear cell with the same cell window size. This difference can be made up by breaking up the linear cell into regions where contact to the *P* base is made as shown in Fig. 6.56*b*. In all cases, maintenance of the smallest possible cell window size produces a lower on-resistance as long as the cell size is also scaled down to keep the cell window area:cell area ratio constant. Although early power MOSFETs were primarily fabricated by using linear cell geometries, most modern power MOSFETs utilize a cellular topology.

6.11. TRENDS

Power MOSFETs are the most commercially advanced devices discussed in this book. Their attributes are a voltage-controlled characteristic with low input gate power, a stable negative temperature coefficient for the on-resistance, wide

safe operating area, and high switching speed. Because of these features, when first introduced, the devices were expected to replace power bipolar transistors in many applications. This displacement has not occurred at a rapid pace as a result of the high on-resistance of power MOSFETs and their relatively high manufacturing cost. With rapid progress along the learning curve, the manufacturing cost of power MOSFETs is now beginning to rival that of bipolar transistors. In addition, for low voltage (< 100 V) and in high-frequency circuits, the power MOSFET clearly offers superior performance. Although power MOSFETs with breakdown voltage of up to 1000 V have been reported, most of the applications for these devices are expected to lie in the low-voltage area such as for automotive circuits.

REFERENCES

1. W. Shockley and G. L. Pearson, "Modulation of conductance of thin films of semiconductors by surface charges," *Phys. Rev.*, **24**, 232–233 (1948).
2. D. Kahng and M. M. Atalla, *IRE–AIEE Solid State Device Research Conference*, 1960.
3. M. S. Adler and V. A. K. Temple, *Semiconductor Avalanche Breakdown Design Manual*, General Electric Technology Marketing Operation, New York, 1979.
4. M. N. Darwish and K. Board, "Optimization of breakdown voltage and on-resistance of VDMOS transistors," *IEEE Trans. Electron Devices*, **ED-31**, 1769–1773 (1984).
5. V. A. K. Temple and P. V. Gray, "Theoretical comparison of DMOS and VMOS structures for voltage and on-resistance," *IEEE International Electron Device Meeting Digest*, Abstract 4.5, pp. 88–92 (1979).
6. Y. Shimada, K. Kato, and T. Sakai, "High efficiency MOS–FET rectifier device," *IEEE Power Electronics Specialists Conference Proceedings*, pp. 129–136 (1983).
7. C. G. B. Garrett and W. H. Brattain, "Physical theory of semiconductor surfaces," *Phys. Rev.*, **99**, 376–387 (1955).
8. B. E. Deal, E. H. Snow, and C. A. Mead, "Barrier energies in metal–silicon dioxide–silicon structures," *J. Phys. Chem. Solids*, **27**, 1873–1879 (1966).
9. W. M. Werner, "The work function difference of the MOS system with aluminum field plates and polycrystalline silicon field plates," *Solid State Electron.*, **17**, 769–775 (1974).
10. V. G. K. Reddi and C. T. Sah, "Source to drain resistance beyond pinch-off in metal-oxide–semiconductor transistors (MOST)," *IEEE Trans. Electron Devices*, **ED-12**, 139–141 (1965).
11. C. T. Sah and H. C. Pao, "The effects of fixed bulk charge on the characteristics of metal-oxide–semiconductor transistors," *IEEE Trans. Electron Devices*, **ED-13**, 393–409 (1966).
12. T. L. Chiu and C. T. Sah, "Correlation of experiments with a two-section model theory of the saturation drain conductance of MOS transistors," *Solid State Electron.*, **11**, 1149–1163 (1968).

13. S. C. Sun and J. D. Plummer, "Modelling of the on-resistance of LDMOS, VDMOS, and VMOS power transistors," *IEEE Trans. Electron Devices*, **ED-27**, 356–367 (1980).
14. C. Hu, "A parametric study of power MOSFETs," *IEEE Power Electronics Specialists Conference Record*, pp. 385–395 (1979).
15. B. J. Baliga, "Switching lots of watts at high speed," *IEEE spectrum*, **18**, 42–48 (1981).
16. C. Hu, "Optimum doping profile for minimum ohmic resistance and high-breakdown voltage," *IEEE Trans. Electron Devices*, **ED-26**, 243–244 (1979).
17. H. Ikeda, K. Ashikawa, and K. Urita, "Power MOSFETs for medium-wave and short-wave transmitters," *IEEE Trans. Electron Devices*, **ED-27**, 330–334 (1980).
18. H. Esaki and O. Ishikawa, "A 900 MHz, 100 watt, VD-MOSFET with silcide gate self-aligned channel," *IEEE International Electron Devices Meeting Digest*, Abstract 16.6, pp. 447–449 (1984).
19. S. Clemente and B. R. Pelly, "Understanding the power MOSFET switching performance," *Proceedings of IEEE Industrial Application Society Meeting*, Abstract 32.B, pp. 763–776 (1981).
20. R. Severns, "*dV/dt* effects in MOSFETs and bipolar junction transistor switches," *IEEE Power Electronics Specialists Conference Record*, pp. 258–264 (1981).
21. D. S. Kuo, C. Hu, and M. H. Chi, "*dV/dt* breakdown in power MOSFETs," *IEEE Electron Device Lett.*, **EDL-4**, 1–2 (1983).
22. C. Hu and M. H. Chi, "Second breakdown of vertical power MOSFETs," *IEEE Trans. Electron Devices*, **ED-29**, 1287–1293 (1982).
23. B. R. Pelly, "Power MOSFETs—a status review," *1983 International Power Electronics Conference Record*, pp. 19–32 (1983).
24. B. J. Baliga and J. P. Walden, "Improving the reverse recovery of power MOSFET integral diodes by electron irradiation," *Solid State Electron.*, **26**, 1133–1141 (1983).
25. A. O. Evwaraye and B. J. Baliga, "The dominant recombination center in electron irradiated semiconductor devices," *J. Electrochem. Soc.*, **124**, 913–916 (1977).
26. Y. Ohata, "New MOSFETs for high power switching applications," *Proc. Powercon 11*, Abstract C5, pp. 1–11 (1984).
27. R. Wang. J. Dunkley, T. A. DeMassa, and L. F. Jelsma, "Threshold voltage variations with temperature in MOS transistors," *IEEE Trans. Electron Devices*, **ED-18**, 386–388 (1971).
28. K. P. Lisiak and J. Berger, "Optimization of non-planar power MOS transistors," *IEEE Trans. Electron Devices*, **ED-15**, 1229–1234 (1978).
29. J. B. Price, "Anisotropic etching of silicon with KOH–H_2O–isopropyl alcohol," *Electrochemical Society Meeting Digest*, pp. 339–353 (1973).
30. J. G. Oakes, R. A. Wickstrom, D. A. Tremere, and T. M. S. Heng, "A power silicon microwave MOS transistor," *IEEE Trans. Microwave Theory Techniques*, **MTT-24**, 305–311 (1976).
31. P. V. Gray D. M. Brown, "Density of SiO_2–Si interface states," *Appl. Phys. Lett.*, **8**, 31–33 (1966).
32. E. Arnold, J. Ladell, and G. Abowitz, "Crystallographic symmetry of surface state density in thermally oxidized silicon," *Appl. Phys. Lett.*, **13**, 413–416 (1968).

33. Y. T. Yeow, D. R. Lamb, and S. D. Brotherton, "An investigation of the influence of low-temperature annealing treatments on the interface state density at the $Si-SiO_2$ interface," *J. Phys. D: Appl. Phys.*, **8**, 1495–1506 (1975).
34. E. Yon, W. H. Ko, and A. B. Kuper, "Sodium distribution in thermal oxide on silicon by radiochemical and MOS analysis," *IEEE Trans. Electron Devices*, **ED-13**, 276–280 (1966).
35. S. R. Hofstein, "Stabilization of MOS devices," *Solid State Electron.*, **10**, 657–670 (1967).
36. E. H. Snow and B. E. Deal, "Polarization phenomena and other properties of phosphosilicate glass films on silicon," *J. Electrochem. Soc.*, **113**, 263–269 (1966).
37. R. J. Kriegler, Y. C. Cheng, and D. R. Colton, "The effect of HCl and Cl_2 on the thermal oxidation of silicon," *J. Electrochem. Soc.*, **119**, 388–392 (1972).
38. M. C. Chen and J. W. Hile, "Oxide charge reduction by chemical gettering with trichloroethylene during thermal oxidation of silicon," *J. Electrochem. Soc.*, **119**, 223–225 (1972).
39. C. M. Osburn, "Dielectric breakdown properties of SiO_2 films grown in halogen and Hydrogen containing environments," *J. Electrochem. Soc.*, **121**, 809–815 (1974).
40. A. S. Grove, O. Leistiko, and C. T. Sah, "Redistribution of acceptor and donor impurities during thermal oxidation of silicon," *J. Appl. Phys.*, **35**, 2695–2701 (1964).
41. W. G. Allen and C. Atkinson, "Comparison of models for redistribution of dopants in silicon during thermal oxidation," *Solid State Electron.*, **16**, 1283–1287 (1973).
42. H. G. Lee, J. D. Sansbury, R. W. Dutton, and J. L. Moll, "Modelling and measurement of surface impurity profiles of laterally diffused regions," *IEEE J. Solid State Circuits*, **SC-13**, 445–461 (1978).
43. K. E. Bean, "Anisotropic etching of silicon," *IEEE Trans. Electron Devices*, **ED-25**, 1185–1193 (1978).
44. C. Hu, M. H. Chi, and V. M. Patel, "Optimum design of power MOSFET's," *IEEE Trans. Electron Devices*, **ED-31**, 1693–1700 (1984).

PROBLEMS

6.1. The *P*-base region of a power MOSFET is fabricated by using a diffusion with a surface concentration of $10^{18}/cm^3$ and a depth of 3 μm. Determine the minimum achievable channel length for a drift layer doping concentration of $10^{15}/cm^3$.

6.2. Consider an *N*-type source diffusion into the *P*-base region described in Problem 6.1 so that the peak base doping concentration is $10^{17}/cm^3$. Determine the threshold voltage of the resulting power MOSFET, assuming that a gate oxide of 1000 A is used. Ignore any work function differences and fixed charge in the oxide.

6.3. What is the effect of a positive fixed charge density of 10^{11} and $10^{12}/cm^2$ on the threshold voltage for the device described in Problem 6.2?

6.4. Consider the power MOSFET structure described in Problem 6.2 with a channel length of 2 μm. Determine the channel width required to achieve a channel resistance of 1 mΩ, assuming that a gate drive voltage of 15 V is used.

6.5. Calculate the drift layer doping concentration and thickness for a power MOSFET designed with a breakdown voltage of 100 V. Assume that ideal parallel-plane breakdown is achievable.

6.6. What is the ideal specific on-resistance for the device described in Problem 6.5?

6.7. Recompute the drift region parameters when the breakdown voltage is limited to 80% of the ideal value by the edge termination of the device.

6.8. A power MOSFET is fabricated by using the drift region obtained in Problem 6.7 with the DMOS process with a linear cell geometry. The polysilicon gate width is 10 μm and the polysilicon window 16 μm. The source diffusion depth is 1 μm and the length is 5 μm. Its sheet resistance is 10 $\Omega/\square$. The base diffusion depth is 3 μm, resulting in the channel parameters obtained in Problem 6.4. Calculate the specific on-resistance of the device neglecting accumulation layer and substrate contributions.

6.9. What are the percentage contributions to the on-resistance from the channel and drift regions for the device described in Problem 6.8?

6.10. Compare the specific on-resistance of the device structure in Problem 6.8 with the ideal specific on-resistance computed in Problem 6.6.

6.11. What is the impact of using an antimony doped substrate with resistivity of 0.01 $\Omega \cdot$cm and thickness of 20 mils on the specific on-resistance?

6.12. What is the impact of using an arsenic doped substrate with resistivity of 0.002 $\Omega \cdot$cm and thickness of 10 mils on the specific on-resistance?

6.13. Estimate the frequency response of a power MOSFET with the DMOS structure as described in Problem 6.8 with an active area of 0.1 cm^2. Assume that the gate capacitance is determined by the oxide alone and that the gate drive circuit has a resistance of 50 Ω.

6.14. The device described in Problem 6.13 is operated in a switching circuit with a gate drive voltage of 15 V. Determine the turn-on delay times for a drive circuit impedance of 50 Ω.

6.15. Calculate the transconductance of the device described in Problem 6.8 in the current saturation region.

This Wiley International Edition is part of a continuing program of paperbound textbooks especially designed for students and professional people overseas. It is an unabridged reprinting of the original hardbound edition, which is also available from your bookseller.

Wiley International Editions include titles in the fields of:

Agricultural Engineering & Agriculture
Anthropology
Biochemistry
Biology
Business Administration
Chemistry
Civil Engineering
Chemical Engineering
Computers & Data Processing
Earth Sciences
Economics
Education
Electrical Engineering
Engineering Mechanics
Geography
Home Economics
Industrial Engineering
Mathematics
Materials Engineering
Mechanical Engineering
Medicine
Physics
Physical Chemistry
Polymer Science & Technology
Probability & Statistics
Psychology
Sociology
Vocational-Technical