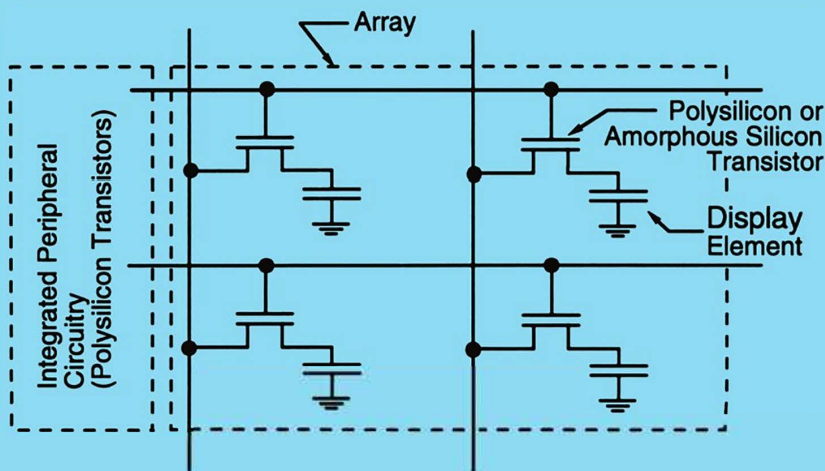


POLYCRYSTALLINE SILICON FOR INTEGRATED CIRCUITS AND DISPLAYS



SECOND EDITION

Ted Kamins

YMTC Ex. 1023
YMTC v. Micron
IPR2025-00099 Page 1

POLYCRYSTALLINE SILICON FOR INTEGRATED CIRCUITS AND DISPLAYS

Second Edition

POLYCRYSTALLINE SILICON FOR INTEGRATED CIRCUITS AND DISPLAYS

Second Edition

by

Ted Kamins
Hewlett-Packard Laboratories



SPRINGER SCIENCE+BUSINESS MEDIA, LLC

ISBN 978-1-4613-7551-7 ISBN 978-1-4615-5577-3 (eBook)
DOI 10.1007/978-1-4615-5577-3

Library of Congress Cataloging-in-Publication Data

A C.I.P. Catalogue record for this book is available
from the Library of Congress.

Copyright © 1998 by Springer Science+Business Media New York
Originally published by Kluwer Academic Publishers, New York in 1998
Softcover reprint of the hardcover 2nd edition 1998

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher, Springer Science+Business Media, LLC.

Printed on acid-free paper.

Contents

1 Preface	xi
1 Deposition	1
1.1 Introduction	1
1.2 Thermodynamics and kinetics	2
1.3 The deposition process	3
1.4 Gas-phase and surface processes	5
1.4.1 Convection	5
1.4.2 The boundary layer	6
1.4.3 Diffusion through the boundary layer	8
1.4.4 Reaction	10
1.4.5 Steady state	11
1.5 Reactor geometries	15
1.5.1 Low-pressure, hot-wall reactor	15
1.5.2 Single-wafer, cold-wall reactor	24
1.5.3 Cold-wall batch reactor	33
1.6 Reaction	34
1.6.1 Decomposition of silane	35
1.6.2 Surface adsorption	37
1.6.3 Deposition rate	39
1.6.4 Rate-limiting surface process	42
1.7 Deposition from disilane	44
1.8 Deposition of doped films	45
1.8.1 <i>n</i> -type films	46
1.8.2 <i>p</i> -type films	49
1.8.3 Electrostatic models	51
1.9 Conformal deposition	51
1.10 Enhanced deposition techniques	53

1.11	Summary	56
2	Structure	57
2.1	Nucleation	57
2.1.1	Amorphous surfaces	58
2.1.2	Single-crystal surfaces	64
2.2	Surface diffusion and structure	65
2.3	Evaluation techniques	70
2.4	Grain structure	72
2.5	Grain orientation	76
2.5.1	Films formed by thermal CVD	79
2.5.2	Effect of plasma on structure	84
2.5.3	Evaporated and sputtered films	85
2.5.4	Other mechanisms controlling structure	86
2.6	Optical properties	86
2.6.1	Index of refraction	87
2.6.2	Absorption coefficient	88
2.6.3	Ultraviolet surface reflectance	90
2.6.4	Use of optical properties for film evaluation	91
2.7	Thermal conductivity	96
2.8	Mechanical properties	97
2.9	Oxygen contamination	99
2.10	Etching	100
2.11	Structural stability	102
2.11.1	Recrystallization mechanisms	103
2.11.2	Undoped or lightly doped polycrystalline films	103
2.11.3	Heavily doped polycrystalline films	105
2.11.4	Amorphous films	110
2.12	Hemispherical-grain (HSG) polysilicon	116
2.13	Epitaxial realignment	117
2.14	Summary	121
3	Dopant Diffusion and Segregation	123
3.1	Introduction	123
3.2	Diffusion mechanism	124
3.2.1	Diffusion along a grain boundary	124
3.2.2	Diffusion in polycrystalline material	128
3.3	Diffusion <i>in</i> polysilicon	129
3.3.1	Arsenic diffusion	134

3.3.2	Phosphorus diffusion	137
3.3.3	Antimony diffusion	138
3.3.4	Boron diffusion	138
3.3.5	Limits of applicability	139
3.3.6	Heavy doping	140
3.3.7	Nitrogen	140
3.3.8	Implant channeling	140
3.4	Diffusion <i>from</i> polysilicon	141
3.5	Interaction with metals	144
3.5.1	Aluminum	145
3.5.2	Other metals	146
3.5.3	Silicides	147
3.6	Dopant segregation at grain boundaries	149
3.6.1	Theory of segregation	149
3.6.2	Experimental data	152
3.7	Computer modeling of diffusion	159
3.8	Summary	162
4	Oxidation	163
4.1	Introduction	163
4.2	Oxide growth on polysilicon	164
4.2.1	Oxidation of undoped films	164
4.2.2	Oxidation of doped films	166
4.2.3	Effect of grain boundaries	174
4.2.4	Effect of device geometry	177
4.2.5	Oxide-thickness evaluation	180
4.3	Conduction through oxide on polysilicon	181
4.3.1	Interface features	184
4.3.2	Deposition conditions	185
4.3.3	Oxidation conditions	188
4.3.4	Dopant concentration and annealing	189
4.3.5	Carrier trapping	191
4.3.6	CVD dielectrics	192
4.4	Summary	193
5	Electrical Properties	195
5.1	Introduction	195
5.2	Undoped polysilicon	196
5.3	Amorphous silicon	198

5.4	Moderately doped polysilicon	199
5.4.1	Carrier trapping at grain boundaries	200
5.4.2	Carrier transport	205
5.4.3	Trap concentration and energy distribution . . .	213
5.4.4	Thermionic-field emission	218
5.4.5	Grain-boundary barriers	219
5.4.6	Limitations of models	221
5.4.7	Segregation and trapping	224
5.4.8	Summary: Moderately doped polysilicon	225
5.5	Grain-boundary modification	225
5.5.1	Grain-boundary passivation	226
5.5.2	Recrystallization	228
5.6	Heavily doped polysilicon	230
5.6.1	Solid solubility	231
5.6.2	Method of doping	232
5.6.3	Stability	235
5.6.4	Mobility	237
5.6.5	Trends	237
5.7	Minority-carrier properties	238
5.7.1	Lifetime	238
5.7.2	Switching characteristics	240
5.8	Summary	243
6	Applications	245
6.1	Introduction	245
6.2	Silicon-gate MOS transistor	246
6.2.1	Complementary MOS	248
6.2.2	Threshold voltage	249
6.2.3	Silicon-gate process	251
6.2.4	Polysilicon interconnections	254
6.2.5	Gate-oxide reliability	255
6.2.6	Limitations	257
6.2.7	Process compatibility	259
6.2.8	New structures	260
6.3	Nonvolatile memories	261
6.4	Polysilicon resistors	262
6.5	Fusible links	267
6.6	Gettering	268
6.7	Polysilicon contacts	268

6.7.1	Reduction of junction spiking	268
6.7.2	Diffusion from polysilicon	269
6.8	Vertical <i>npn</i> bipolar transistors	270
6.8.1	Fabrication: Polysilicon contacts	270
6.8.2	Physics of the polysilicon-emitter transistor . . .	274
6.9	Lateral <i>pnp</i> bipolar transistors	279
6.10	Device isolation	281
6.10.1	Dielectric isolation	281
6.10.2	Poly-buffered LOCOS	283
6.10.3	Trench isolation	284
6.11	Dynamic random-access memories	285
6.11.1	Trench capacitor	287
6.11.2	Stacked capacitor	289
6.12	Polysilicon diodes	292
6.13	Polysilicon thin-film transistors	296
6.13.1	Device physics	298
6.13.2	Methods of improving polysilicon for TFTs . . .	304
6.13.3	TFTs for active-matrix, liquid-crystal displays .	308
6.13.4	TFTs for static random-access memories	310
6.14	Microelectromechanical Systems	311
6.14.1	Integrated sensors	312
6.14.2	Polysilicon for MEMS	313
6.15	Summary	315

Chapter 5

Electrical Properties

5.1 Introduction

Polycrystalline-silicon films formed by chemical vapor deposition are used in a wide variety of ULSI applications requiring very different electrical properties. High-value load resistors for static random-access-memory (RAM) cells utilize the high resistance of lightly doped polysilicon to provide a convenient and stable resistor that limits the current flowing in the cell. At the other extreme, the excellent technological compatibility of polysilicon with high-temperature, integrated-circuit processing allows straightforward fabrication of self-aligned gates and convenient interconnections in ULSI circuits. Although a resistivity of less than about $10^{-3} \Omega\text{-cm}$ — eight orders of magnitude less than for static RAM load resistors — is routinely achieved, the lower bound on the resistivity of polysilicon can limit the performance of silicon-gate integrated circuits that use polysilicon interconnections to conduct signals long distances across a chip [5.1]. As feature sizes become smaller and intrinsic transistor delays decrease on chips of increasing overall dimensions, the resistance of polysilicon interconnections is becoming a more serious limitation on integrated-circuit performance.

In this chapter we will examine the wide range of electrical properties that can be achieved in CVD polysilicon films and show that the high resistance of lightly and moderately doped polysilicon is related to the grain boundaries. Methods of modifying the electrical properties by controlling the number of grain boundaries or their electrical activity will be examined. For highly doped films, we will see that the resistivity

is limited by the solid solubility of the dopant; *ie*, by the amount of dopant that can be incorporated in substitutional sites within the grains.

5.2 Undoped polysilicon

The range of deposition parameters accessible in a low-pressure reactor is limited, and the resistivity can be studied over a much wider range of deposition temperatures in films deposited in cold-wall reactors. For thin polysilicon films deposited in a polycrystalline form over the temperature range from 650 to 1050°C in an atmospheric-pressure, cold-wall reactor, the resistivity parallel to the surface is in the mid- 10^5 Ω -cm range and is not sensitive to the deposition temperature [5.2]. The value of mid- 10^5 Ω -cm is close to the resistivity of intrinsic silicon.

Although the crystal structure varies widely over the range in which polysilicon films are deposited, the resistivity of undoped films is quite insensitive to the structure. For example, the grain size increases with increasing thickness, but the resistivity varies only slightly with film thickness, changing from about 5×10^5 Ω -cm for films 0.5 μm thick to about 3×10^5 Ω -cm for films 25 μm thick. Similar values of resistivity are also observed in layers deposited in different types of reactors.

Undoped films deposited in the amorphous form at lower temperatures and subsequently crystallized by annealing at a higher temperature have a resistivity about 2 or 3 times higher than films deposited in the polycrystalline form. In films deposited at a higher rate, the discontinuity in resistivity occurs at a higher temperature, consistent with the increased transition temperature from an amorphous deposition to a polycrystalline deposition at a higher deposition rate.

At higher temperatures, epitaxial films can be formed on single-crystal wafers at the same time that polycrystalline films are deposited on oxide-covered wafers. Measurement of the conductivity type (using a *hot-point-probe*) shows that the single-crystal films are *n*-type while the polycrystalline films are *p*-type. The *n*-type behavior of single-crystal samples is attributed to impurities in the silane used to deposit the films. The different conductivity type of the polycrystalline films implies that structural defects influence the resistivity enough to compensate for the small amount of *n*-type dopant in the silane.

In thick films, the resistivity can be measured perpendicular to the film plane, as well as parallel to it, and is somewhat lower in the per-

pendicular direction. In films about $13\text{ }\mu\text{m}$ thick, the resistivity is about $3 \times 10^5\text{ }\Omega\text{-cm}$ in the film plane, but only $1.5 \times 10^5\text{ }\Omega\text{-cm}$ in the perpendicular direction. This anisotropy is probably related to the columnar crystal structure normal to the film plane (Fig. 2.10) [3.12]. Similar anisotropic conductivity is seen in thicker samples with a columnar grain structure deposited from a silicon halide at a high temperature (1100°C) [2.57]; the resistivity is $5 \times 10^5\text{ }\Omega\text{-cm}$ parallel to the film plane, but only $5 \times 10^4\text{ }\Omega\text{-cm}$ perpendicular to it.

For moderate electric fields, the current flow is ohmic, although nonlinear current-voltage behavior is observed for fields above 10^4 V/cm in the vertical samples, in which high fields can be readily applied. Measurements with different electrode materials and sample geometries suggest that the nonlinearity is not dominated by the electrodes.

Over the temperature range from -25 to 200°C , the conductivity of undoped films varies with the measurement temperature T according to the relation

$$\sigma = \sigma_0 \exp\left(-\frac{E_a}{kT}\right) \quad (5.1)$$

with an apparent activation energy E_a above room temperature of approximately half the bandgap of silicon. However, a weaker temperature dependence is sometimes observed below room temperature, indicating a different conductivity-limiting mechanism at low temperatures. The conduction may arise from hopping between localized states [5.3] within the grain boundaries [2.34, 5.4, 5.5], with a temperature dependence of the conductivity of the form

$$\sigma\sqrt{T} = \sigma_0 \exp\left[-\left(\frac{T_0}{T}\right)^{1/4}\right] \quad (5.2)$$

Irreproducibly low values of resistivity are sometimes observed on films with unprotected surfaces, probably because of surface conduction from adsorbed moisture or contamination. Thin films deposited in the amorphous form and subsequently crystallized are less likely to exhibit anomalously low resistivity, even without surface protection. The resistivity of all the films is stable if a thermally grown oxide is used to passivate the surface.

5.3 Amorphous silicon

The electrical properties of amorphous silicon that are not crystallized by subsequent heat treatment can be very different from those of polysilicon because of their greater structural disorder. In polysilicon, the basic conduction process within the grains is similar to that in crystalline silicon, with conduction of electrons and holes in well-defined conduction and valence bands. At all energies above the lower edge of the conduction band, electrons can readily move through the crystal. The properties of polysilicon differ from those of single-crystal silicon primarily because of the grain boundaries.

In amorphous silicon, on the other hand, the basic crystalline structure is absent, and the concept of a band structure becomes somewhat tenuous. Instead of a well-defined bandgap, there is a range of energies with few allowed states. The edges of this region are not sharply separated from regions with allowed states in which the carriers can readily move through the material. Instead, near the edges of the “bandgap,” a high density of *localized* states is present. Carriers in these states cannot readily move through the material, and they do not contribute significantly to the conduction. At higher energies, carriers can move through the material, but the absence of a well-defined crystal structure greatly reduces the ability of carriers to move (*ie*, reduces their mobility).

In amorphous material the energy corresponding to the lower edge of the conduction band in a crystalline material is called the *lowest unoccupied molecular orbital* (LUMO), and the energy corresponding to the upper edge of the valence band is called the *highest occupied molecular orbital* (HOMO). Although the structure of amorphous material is very different from that of crystalline material, applying the familiar concepts of bandgap and mobility to amorphous material is useful.

The effective bandgap of amorphous silicon can be much greater than that of polycrystalline silicon; values near 1.8 eV are often found, although the value depends strongly on the deposition conditions, the bonding, and the amount of hydrogen incorporated. As discussed in Sec. 2.11.4, the amorphous silicon used in very low-temperature processes can be deposited in the 300°C temperature range by plasma-enhanced CVD and contains a large amount of hydrogen. The relative amounts of hydrogen bound as Si-H and Si-H₂ can affect the electrical properties, as well as the stability of the material.

Conduction occurs primarily by hopping between localized states, with the temperature dependence given by Eq. 5.2, rather than by conduction in well-defined bands. The “bandgap” is important because it influences carrier generation and, therefore, the leakage current and the energy dependence of the photosensitivity. Because the bandgap is greater in amorphous silicon than in crystalline silicon, the leakage current may be lower because fewer free carriers are generated; however, the larger bandgap may lead to a greater temperature dependence of the leakage current than in polysilicon with its smaller bandgap.

5.4 Moderately doped polysilicon

For static RAMs and other devices in which high resistivity is needed, the magnitude of the resistivity and its control are important. At low dopant concentrations, the resistivity changes only slowly with increasing dopant concentration. The measured resistivity remains in the mid- $10^5 \Omega\text{-cm}$ range and can be as much as six orders of magnitude greater than those of similarly doped, single-crystal silicon, as shown in Fig. 5.1. At intermediate dopant concentrations, the resistivity decreases rapidly and approaches that of single-crystal silicon. Control of the resistance of a high-value resistor is best in lightly doped material; achieving accurate values of resistance in the intermediate dopant-concentration range is difficult because of the rapid variation of resistance with dopant concentration. The large difference between polysilicon and single-crystal silicon is observed in material deposited over a wide range of deposition conditions, with deposition temperatures varying from less than 600°C to more than 1100°C . Similar behavior is seen when the dopant is added during deposition as when it is added after deposition by ion implantation or from a gas-phase dopant source.

In this section, we will examine the causes of the large difference between the resistivity of polysilicon and that of single-crystal silicon, and in a subsequent section we will discuss techniques to achieve better control with intermediate dopant concentrations. Two mechanisms — both involving grain boundaries — can cause the electrical behavior of polysilicon to depart from that of single-crystal silicon at low and moderate dopant concentrations. In Sec. 3.6 we saw that n -type dopant atoms physically segregate at grain boundaries because of their lower energy in the disordered, grain-boundary region [3.23, 5.6] and, therefore, do not contribute to the conduction process. In this section, we

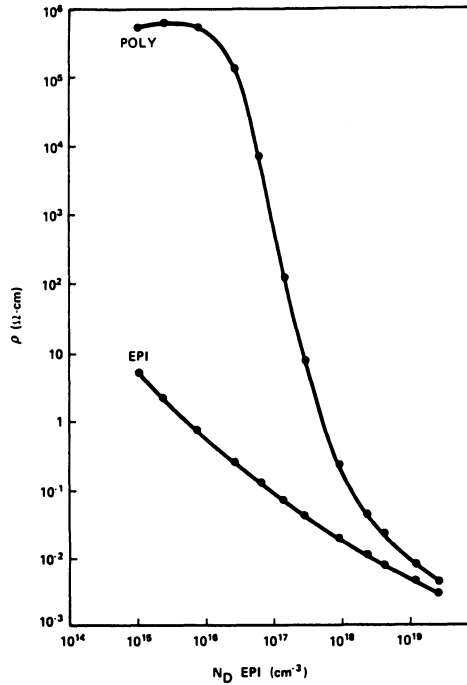


Figure 5.1: The resistivity of polysilicon is much greater than that of similarly doped, single-crystal, epitaxial silicon. It changes slowly at low dopant concentrations, but decreases rapidly at intermediate dopant concentrations. At high dopant concentrations it approaches the resistivity of single-crystal silicon. (Large-grain, $\sim 15 \mu\text{m}$ -thick films deposited in a cold-wall reactor at $\sim 1000^\circ\text{C}$.)

will see that carriers contributed by substitutional dopant atoms can be trapped in deep energy levels at the grain boundaries [5.7–5.11], again reducing the number of carriers available for conduction.

5.4.1 Carrier trapping at grain boundaries

To determine the number of free carriers in polysilicon films, the substitutional dopant concentration must first be found by considering the loss of dopant atoms that segregate to the grain boundaries and are electrically inactive. Even after this loss is considered, however, the conductivity of polysilicon is still much less than that of single-crystal silicon containing the same concentration of substitutional dopant atoms. Free

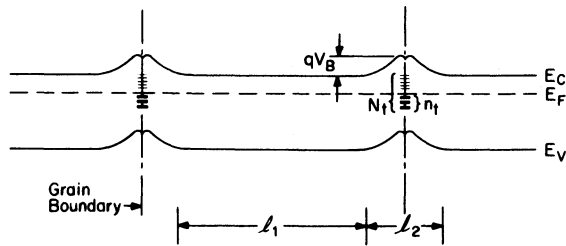


Figure 5.2: Many of the free carriers contributed by substitutional dopant atoms in the grains are quickly immobilized at the grain boundaries at traps of density N_T per unit area with energies within the bandgap.

carriers (either electrons or holes) are contributed to the conduction or valence band by substitutional dopant atoms located within the grains, just as in single-crystal silicon. However, many of these free carriers are quickly trapped at low-energy positions at the grain boundaries, as schematically shown in Fig. 5.2 [5.7] and, consequently, cannot contribute to the conduction [5.7, 5.8].

Dangling bonds at the grain boundaries and defects within the grains both lead to allowed states within the bandgap of the polysilicon (Fig. 5.3). The *deep states* associated with the dangling bonds have energies near the middle of the forbidden energy gap [5.9], where they are most effective in allowing generation and recombination of carriers. (The maximum density is slightly below mid-gap, as discussed below.) In addition to states near mid-gap resulting from broken bonds, strained bonds cause a high density of shallow *tail states* near the band edges.

In our discussion of carrier trapping, barrier formation, and conduction, we will consider *n*-type silicon. However, unlike dopant segregation, carrier trapping is similar in *n*-type and *p*-type polysilicon; barriers form in polysilicon of both conductivity types. (On the other hand, in germanium the grain boundaries are strongly *p*-type so that barriers form only in *n*-type material.) Some minor differences in the electrical behavior of *n*-type and *p*-type polysilicon will be considered later.

Potential barriers

The charge trapped at the grain boundaries is compensated by oppositely charged depletion regions surrounding the grain boundaries

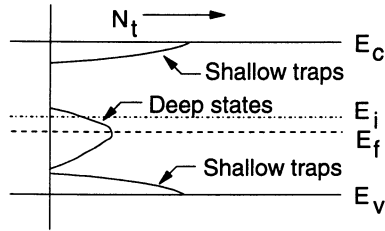


Figure 5.3: Traps are primarily located at grain boundaries, although defects within the grains also create some states. The shallow *tail* states are associated with strained bonds, and the deep states near mid-gap are caused by broken bonds.

(Fig. 5.4b) [5.10]. From Poisson's equation, the charge in the depletion regions causes curvature in the energy bands, leading to potential barriers that impede the movement of any remaining free majority carriers from one grain to another [5.7, 5.8]. The barrier height V_B (Fig. 5.4c) can be expressed in terms of the dopant concentration N and the width of the depletion region x_d using Poisson's equation:

$$\frac{d^2V}{dx^2} = \frac{qN}{\epsilon} \quad (5.3)$$

where q is the magnitude of the charge on an electron, ϵ is the permittivity of silicon, and a one-dimensional case is considered. Solving for V_B , we find

$$V_B = \frac{qN}{2\epsilon} x_d^2 \quad (5.4)$$

The barrier height depends strongly on the substitutional dopant concentration and the trap density and energy. Barrier formation in both n -type and p -type polysilicon indicates that both electron and hole traps exist at the grain boundaries. At low dopant concentrations the total number of carriers NL per unit area in a grain of length L is much less than the number of traps N_T per unit area at the grain boundary. If the energy levels of the traps are deep enough, virtually all the free carriers contributed to the conduction band by the substitutional dopant atoms are trapped at grain boundaries, and very few are free to contribute to conduction. Because the dopant concentration is low, the depletion regions surrounding the grain boundaries extend through the entire grain, and no neutral region remains. The low dopant concentration also means that the energy bands within the grain have little

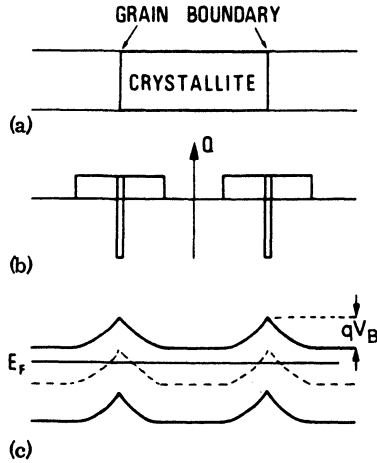


Figure 5.4: Depletion regions form in the grains surrounding the grain boundaries to compensate the charge trapped at the grain boundaries (b). The charged depletion regions cause curvature in the energy bands (c) and the resulting potential barriers. From [5.10]. Reprinted with permission.

curvature (Fig. 5.5a); the barriers to conduction are small, and any free carriers that are present can readily move from one grain to another. The small barrier and low carrier density are consistent with an observed resistivity of lightly doped polysilicon close to that of intrinsic silicon.

When the grains are completely depleted, the depletion region associated with a grain boundary extends a distance $L/2$ on each side of the boundary, and the barrier height can be expressed as

$$V_B = \frac{qN}{2\epsilon} \left(\frac{L}{2} \right)^2 = \frac{qNL^2}{8\epsilon} \quad (5.5)$$

As the dopant concentration increases, more carriers are trapped at the grain boundary; the curvature of the energy bands and the height of the potential barrier both increase (Fig. 5.5b), making carrier transport from one grain to another more difficult.

In the simplest version of the carrier-trapping model, the energy of the grain-boundary traps is assumed to be deep enough that they are completely filled when the dopant concentration exceeds a critical value $N^* = N_T/L$ [5.8, 5.10]. As the concentration of dopant atoms

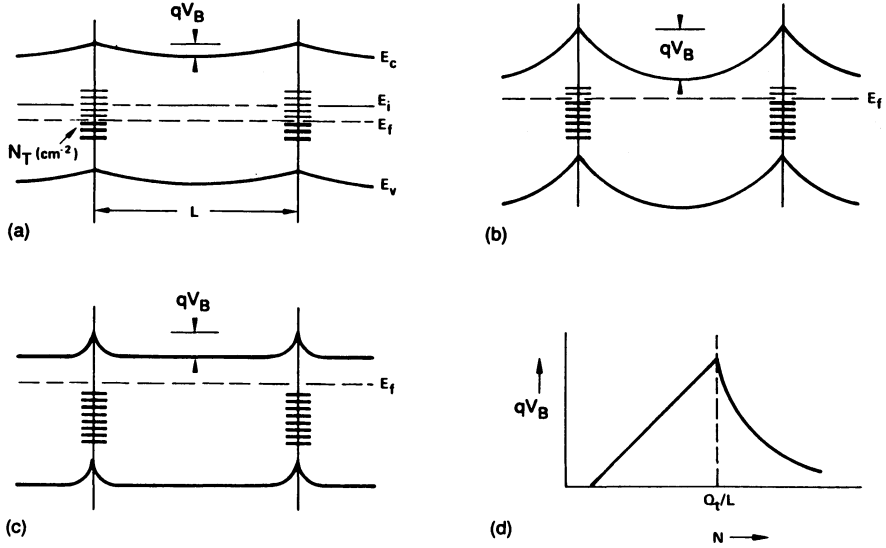


Figure 5.5: The barrier height V_B increases with increasing dopant concentration until all the traps are filled. Above the critical dopant concentration $N^* = N_T/L$, it decreases as neutral regions form within the grains.

increases above N^* , the number of trapped carriers per unit area of grain boundary remains constant at the value N_T . The added carriers, which are not trapped, form neutral regions within the grains, as seen in Fig. 5.5c; and, from charge neutrality, the width of the depletion regions decreases according to the relation

$$x_d = \frac{N_T}{2N} \quad (5.6)$$

The barrier height is then found from Eq. 5.4 to be

$$V_B = \frac{qN}{2\epsilon} \left(\frac{N_T}{2N} \right)^2 = \frac{qN_T^2}{8\epsilon N} \quad (5.7)$$

and the barrier to carrier transport between grains decreases as the dopant concentration increases above its critical value N^* . Thus, as shown in Fig. 5.5d, when dopant atoms are added to polysilicon, the potential barrier first increases, reaches a maximum, and then decreases. This nonmonotonic behavior has important consequences for the electrical conduction and can explain the markedly different electrical properties of polysilicon and single-crystal silicon seen in Fig. 5.1.

5.4.2 Carrier transport

At low dopant concentrations, the width of the depletion region on each side of the grain boundary is half the grain size. At moderate dopant concentrations, it decreases with increasing dopant concentration, but even at a dopant concentration of 10^{17} cm^{-3} , it is several tens of nanometers wide (for a typical grain-boundary trap density $\sim 10^{12} \text{ cm}^{-2}$). Tunneling is not significant for barriers of this width, and carriers usually travel from one grain to another by thermionic emission over the barrier. At high dopant concentrations the width of the depletion regions can be less than about 10 nm, and tunneling through the barrier can also contribute to the current flow. However, the barrier height decreases with increasing dopant concentration, so conduction across the barrier may no longer limit the current flow.

When thermionic emission dominates, the conduction process can be understood by considering the effect of an applied field on the barrier height. The thermionic-emission current density J can be written as

$$J = qnv_c \exp \left[-\frac{q}{kT}(V_B - V) \right] \quad (5.8)$$

where n is the free-carrier density, v_c is the collection velocity ($v_c = \sqrt{kT/2\pi m^*}$), V_B is the barrier height with no applied bias, and V is the applied bias across the depletion region. With no applied bias, the barriers to carrier transport in the forward and reverse directions are equal. The forward current J_F of electrons emitted over the barrier by thermionic emission equals the reverse current J_R in the opposite direction, and no net current flows.

When a voltage is applied, the barrier to carrier transport in one direction decreases, while it increases in the opposite direction. In general, the applied voltage divides nonuniformly between the two sides of a grain boundary [5.12]; however, for small applied biases, approximately half the applied voltage appears across each depletion region. The barrier in the forward direction decreases by an amount $V \approx \frac{1}{2}V_G$ where V_G is the bias across one grain boundary (*ie*, the applied bias across the sample divided by the number of grains, assuming all grain boundaries are equivalent). The barrier increases by the same amount in the reverse direction [5.12].

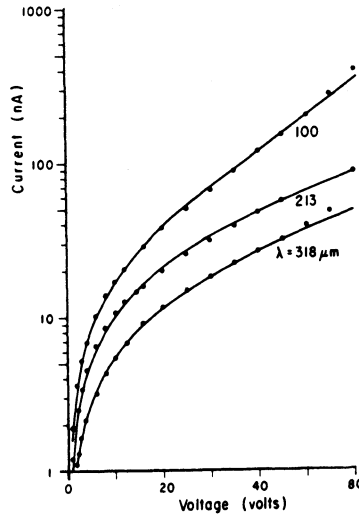


Figure 5.6: Current-voltage characteristics for 3 μm thick, polysilicon films doped with about $2 \times 10^{16} \text{ cm}^{-3}$ phosphorus atoms; the samples have lengths λ of 100, 213, and 318 μm , as indicated. From [5.13]. ©1978, Pergamon Journals, Ltd. Reprinted with permission.

By considering thermionic emission in the two directions, the net current density $J = J_F - J_R$ can be found [5.13]. With $V \approx \frac{1}{2}V_G$

$$J_F = qnv_c \exp \left[-\frac{q}{kT} \left(V_B - \frac{1}{2}V_G \right) \right] \quad (5.9)$$

and

$$J_R = qnv_c \exp \left[-\frac{q}{kT} \left(V_B + \frac{1}{2}V_G \right) \right] \quad (5.10)$$

The net current density is then given by

$$J = qnv_c \exp \left(-\frac{qV_B}{kT} \right) \left[\exp \left(\frac{qV_G}{2kT} \right) - \exp \left(-\frac{qV_G}{2kT} \right) \right] \quad (5.11)$$

or

$$J = 2qnv_c \exp \left(-\frac{qV_B}{kT} \right) \sinh \left(\frac{qV_G}{2kT} \right) \quad (5.12)$$

Current flow in polysilicon films fits this expression well, as shown in Fig. 5.6 [5.13].

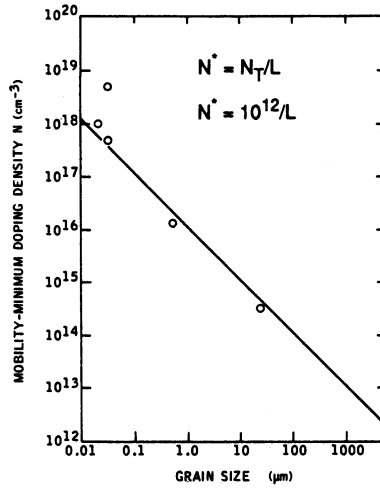


Figure 5.7: The critical dopant concentration N^* varies inversely with the grain size. From [5.14]. Reprinted with permission.

At low applied voltages the voltage drop across each grain boundary is small compared to the thermal voltage kT/q , and the general expression of Eq. 5.12 can be simplified by using the relation

$$\sinh\left(\frac{qV_G}{2kT}\right) \approx \frac{qV_G}{2kT} \quad (5.13)$$

to obtain a linear relation between current and applied voltage:

$$J = \frac{q^2 n v_c}{kT} \left[\exp\left(-\frac{qV_B}{kT}\right) \right] V_G \quad (5.14)$$

From Eq. 5.14 an expression for the average conductivity $\sigma = J/\mathcal{E} = JL/V_G$ can be obtained:

$$\sigma = \frac{q^2 n v_c L}{kT} \exp\left(-\frac{qV_B}{kT}\right) \quad (5.15)$$

Thus, conduction in polysilicon is an activated process with an activation energy of approximately qV_B , which depends on the dopant concentration and the grain size.

From this simple model we see the importance of the critical dopant concentration N^* , at which all the grain-boundary traps are filled with

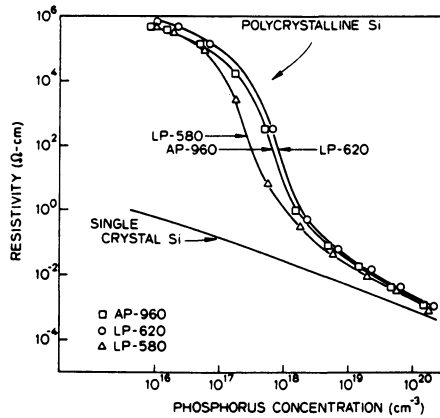


Figure 5.8: The resistivity of phosphorus-implanted polysilicon films is lower for films with larger grains, especially at intermediate dopant concentrations.

carriers from the substitutional dopant atoms. Below N^* , the free-carrier concentration is low, and the resistivity is high. As dopant is added, the increasing free-carrier concentration is compensated by the increasing barrier height, and the resistivity remains high. At dopant concentrations greater than N^* , the free-carrier concentration increases and the barrier height decreases, allowing the resistivity to decrease rapidly with increasing dopant concentration. Thus N^* divides the high resistivity region from the region with rapidly decreasing resistivity. From the definition of N^* ($N^* = N_T/L$), we expect it to vary inversely with grain size L . The inverse proportionality between critical dopant concentration and grain size has been observed for grain sizes varying from 20 nm to over 20 μm (Fig. 5.7) [5.14].

For the typical submicrometer-thick polysilicon films used in most integrated-circuit applications, the grain size is usually somewhat less than 100 nm. Using this value with a trap density somewhat greater than 10^{12} cm^{-2} of grain-boundary area, we find the critical dopant concentration to be of the order of 10^{17} - 10^{18} cm^{-3} , as observed for thin polysilicon films deposited over a wide range of conditions (Fig. 5.8) [5.15]. For films deposited at a high temperature in a cold-wall reactor, the critical dopant concentration is slightly lower than for polysilicon films deposited at a much lower temperature in the low-pressure reactor. The films deposited at the higher temperature have a slightly larger grain size and, consequently, a lower critical dopant concentration. The

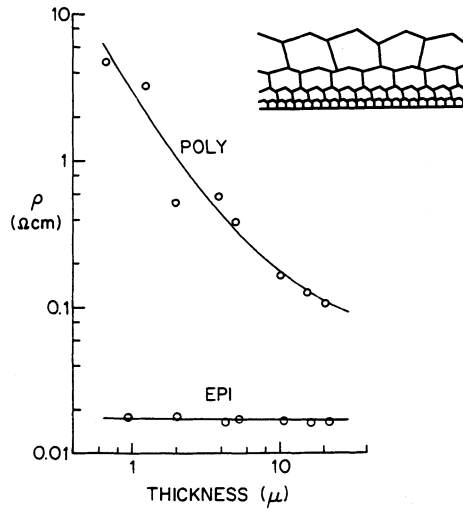


Figure 5.9: The *resistivity*, as well as the sheet resistance, of polysilicon films decreases as the thickness increases.

critical dopant concentration for low-pressure films deposited in an initially amorphous form and subsequently crystallized is appreciably lower than for either of the other two types of polysilicon, consistent with its larger grain size, as discussed in Sec. 2.11.4. The resistivity is most sensitive to the grain size in the intermediate-concentration range. It does not vary significantly with grain size for lightly or heavily doped films.

The effect of grain size is also seen by considering the variation of resistivity with increasing film thickness. The grain size increases as the film thickness increases [2.2], and the *resistivity* decreases, as seen in Fig. 5.9 [5.7]. The resistivity of single-crystal films deposited at the same time is independent of film thickness, as expected. Because the resistivity of the polysilicon films decreases with increasing thickness t , the sheet resistance decreases more rapidly than $1/t$.

Effective mobility

To express the conductivity in Eq. 5.15 in the usual linear form of Ohm's law, $\sigma = qn\mu$, an *effective mobility* μ_{eff} can be defined as

$$\mu_{\text{eff}} = \frac{qv_c L}{kT} \exp\left(-\frac{qV_B}{kT}\right) \quad (5.16)$$

The quantity μ_{eff} is a mobility in the sense that it describes the ease of carrier movement from one grain to another in the polycrystalline material. It is not the familiar microscopic mobility related to carrier scattering in a homogeneous piece of semiconductor. Instead, it describes the restriction of current flow by the grain-boundary barrier.

Because of its strong dependence on barrier height, the mobility depends sensitively on dopant concentration. As dopant atoms are added to undoped polysilicon, the barrier height increases (Fig. 5.5d), and the mobility decreases until the dopant concentration reaches its critical value N^* . Above this value the barrier height decreases, and the mobility increases. The predicted minimum in the mobility with increasing dopant concentration is observed, as shown in Fig. 5.10 [5.8]. (Note that Fig. 5.10 only shows the dependence of the mobility on dopant concentration over a limited range. At even higher dopant concentrations, grain-boundary barriers no longer limit the mobility, and ionized impurity scattering dominates, as in single-crystal silicon. In this region the mobility decreases with increasing dopant concentration, paralleling the behavior in single-crystal silicon; however, even at high dopant concentrations, the mobility is about a factor of two less than that in single-crystal silicon, as will be discussed in Sec. 5.6.)

Hall mobility

In the discussion above, we developed an expression for the effective mobility, which describes the ease with which a carrier moves through a material composed of grains and high-resistance grain boundaries. This macroscopic quantity differs fundamentally from the microscopic mobility characterizing carrier-scattering mechanisms. Although the *effective mobility* has been compared with the measured *Hall mobility*, the theoretical relation between these two quantities is not obvious.

In an inhomogeneous film both the grains and the space charge regions surrounding the grain boundaries contribute to the Hall voltage [5.16–5.20]. Detailed analysis of such an inhomogeneous structure shows that the carrier concentration corresponding to the measured Hall constant is always less than the carrier concentration in the grains [5.19]. The discrepancy is greatest when the barrier is high or the depletion regions occupy a large portion of the grain. At very high carrier concentrations, the barriers are small, and the measured Hall concentration approaches the actual carrier concentration. Similarly, in very lightly

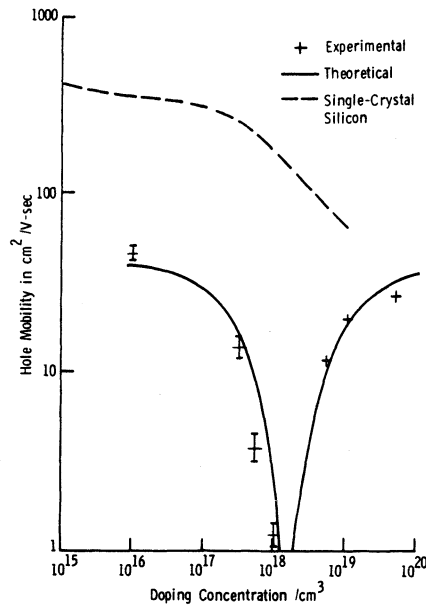


Figure 5.10: The room-temperature, Hall hole mobility as a function of boron concentration measured in approximately $1\ \mu\text{m}$ thick polysilicon films. From [5.8]. Reprinted with permission.

doped material, the entire grain is depleted, and the barriers are again small so that the Hall concentration approaches the average carrier concentration. With certain approximations, the Hall mobility is related to the barrier height by the expression [5.20]

$$\mu_H \sim \exp\left(-\frac{qV_B}{kT}\right) \quad (5.17)$$

which has the same dependence on the barrier height as the effective mobility (Eq. 5.16) used in the expression for the conductivity.

Temperature dependence

From the expression for the low-field conductivity in Eq. 5.15, we can find the temperature dependence of the resistivity

$$\rho = \frac{kT}{q^2 n v_c L} \exp\left(\frac{qV_B}{kT}\right) \quad (5.18)$$

for dopant concentrations both below and above the critical dopant concentration N^* . At low dopant concentrations $NL < N_T$, and the

grains are completely depleted. The maximum carrier concentration is then

$$n = n_i \exp\left(\frac{qV_B + E_F}{kT}\right) = N_c \exp\left(\frac{qV_B + E_F - \frac{1}{2}E_G}{kT}\right) \quad (5.19)$$

where n_i is the intrinsic carrier concentration, E_F is the Fermi level measured from the intrinsic Fermi level at the grain boundary, and N_c is the effective density of states at the conduction band edge. Using this expression in Eq. 5.18, we find the resistivity to be

$$\rho = \frac{kT}{q^2 N_c v_c L} \exp\left(\frac{\frac{1}{2}E_G - E_F}{kT}\right) \quad (5.20)$$

Assuming that the traps pin the Fermi level at the grain boundary near mid-gap, $E_F \approx E_i \equiv 0$, and the dominant temperature dependence of the resistivity has the form

$$\rho \propto \exp\left(\frac{E_G}{2kT}\right) \quad (5.21)$$

Thus, the resistivity exhibits an activated behavior, with an apparent activation energy of approximately half the bandgap.

For dopant concentrations above the critical dopant concentration, $NL > N_T$, and neutral regions exist within the grains. The carrier concentration n is approximately equal to the dopant concentration N , and the resistivity can be written

$$\rho = \frac{kT}{q^2 N v_c L} \exp\left(\frac{qV_B}{kT}\right) \quad (5.22)$$

In this region

$$V_B = \frac{qN_T^2}{8\epsilon N} \quad (5.23)$$

(Eq. 5.7), and the resistivity is proportional to

$$\rho \propto \frac{1}{N} \exp\left(\frac{q^2 N_T^2}{8\epsilon kTN}\right) \quad (5.24)$$

The resistivity is again an activated process; however, in this region the apparent activation energy E_a is no longer constant, but varies inversely with the dopant concentration N :

$$E_a = \frac{q^2 N_T^2}{8\epsilon N} \quad (5.25)$$

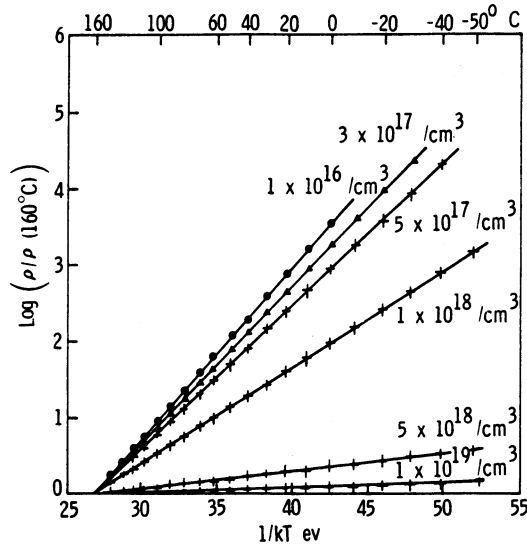


Figure 5.11: Logarithm of the normalized resistivity as a function of reciprocal temperature, showing a well-defined activation energy. From [5.8]. Reprinted with permission.

From Eqs. 5.21 and 5.25, we expect that the activation energy is about half the bandgap energy for undoped polysilicon, and remains at approximately this value for dopant concentrations below the critical dopant concentration. Above the critical dopant concentration, the resistivity and activation energy decrease until other conduction mechanisms limit the conduction process at very high dopant concentrations. Consistent with the model, the resistivity does, indeed, exhibit a well-defined activation energy which decreases with increasing dopant concentration, as shown in Fig. 5.11 [5.8]. (More detailed analysis [5.21] shows that the activation energy obtained from an Arrhenius plot differs from the barrier height because of the shift of the Fermi level with temperature. Calculations show that $E_a \approx qV_{B0} = qV_B(1 + \gamma T)$ with $\gamma \approx 1.5 \times 10^{-3}/\text{K}$ [6.114].) The rapid variation of resistivity with temperature in lightly doped films makes using such films for precise resistors difficult.

5.4.3 Trap concentration and energy distribution

The similar behavior of electrons and holes implies that either type of majority carrier can be trapped at the grain boundaries and that

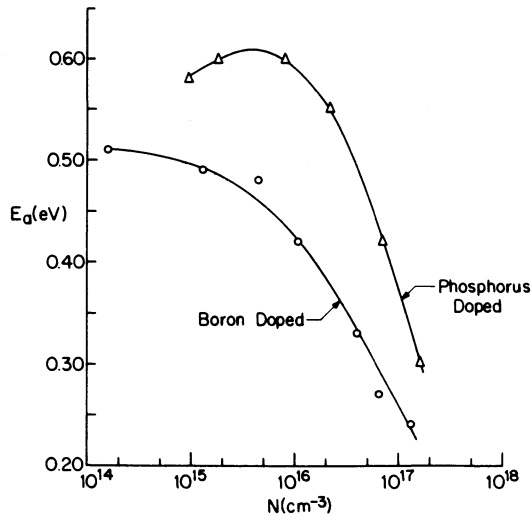


Figure 5.12: Dopant-concentration dependence of the activation energy of the conductivity in thick polysilicon films deposited in a cold-wall reactor.

the grain-boundary traps are located near mid-gap or are distributed symmetrically around mid-gap. More detailed observation reveals some subtle differences. When the *p*-type dopant boron is added, both the resistivity and the activation energy of the resistivity decrease monotonically with increasing dopant concentration, as shown in Fig. 5.12 [2.4]. However, for the common *n*-type dopants, the resistivity and its activation energy initially increase as dopant is added, before reaching a maximum and then decreasing [2.4, 5.22]. This asymmetry suggests that the dominant traps are located somewhat below mid-gap, pinning the Fermi level there in lightly doped polysilicon and is consistent with the conductivity-type measurements discussed in Sec. 5.2, which show that lightly doped polysilicon is slightly *p*-type even if small amounts of *n*-type dopant are added.

Other studies also indicate that the trap density is highest near mid-gap [5.23]; the Fermi level at the grain boundaries appears to be located slightly below mid-gap at about 0.62 eV below the conduction band edge [5.24]. This energy level is consistent with an electronic state related to dangling silicon bonds about 0.65 eV below the conduction band edge [5.25]. (Dangling bonds are related to trivalent silicon, which has been found at grain boundaries in polysilicon [5.26] and at individual

grain boundaries in silicon bicrystals.) In contrast to the behavior in polycrystalline silicon, the dominant traps in lightly doped, plasma-enhanced CVD amorphous silicon formed by decomposition of silane in a glow discharge appear to be located slightly above mid-gap [5.27]; undoped films are *n*-type, rather than *p*-type, and the resistivity and its activation energy initially increase as *p*-type dopant is added, before reaching a maximum and then decreasing.

From the measured electrical behavior and the carrier-trapping model, a typical grain-boundary density of traps is found to be about $\text{mid-}10^{12} \text{ cm}^{-2}$ of grain-boundary area. As the grain size increases, the number of traps per unit volume decreases. We saw in Sec. 2.11.4 that silicon layers deposited in amorphous form and subsequently crystallized can have larger grains than layers deposited in polycrystalline form. The electrical properties reflect the larger grains formed by crystallization and also the increase in grain size with increasing film thickness [5.28]. The resistivity is lower, the mobility is higher, and the fraction of active dopant is also higher.

An independent measure of the the density of grain-boundary states can be obtained by electron spin resonance measurements. The density is lower in films with a greater degree of preferred crystalline orientation [2.34], possibly because of the improved structure of grain boundaries between grains with similar orientation perpendicular to the film plane. As the grain-boundary tilt angle increases, the structure of the grain boundary becomes more complex, and the conduction becomes increasingly nonlinear [5.29].

The simplest carrier-trapping model suggests that a discrete level exists at the grain boundary and that the Fermi level is pinned close to this level until all the carrier traps there are filled as more dopant is added. These assumptions lead to a well-defined maximum in the barrier height as the dopant concentration increases. These approximations are appropriate for a moderate density of traps located near mid-gap. However, if the trap density is high or the traps are located far from mid-gap, the traps may be only partially filled when a neutral region begins to form. The bending of the energy bands may raise the trap level at the grain boundary above the Fermi level so that carriers from additional dopant atoms create a neutral region, with some of the trap levels remaining unfilled.

A further complication arises because of the varying properties of different grain boundaries: The traps can be distributed in energy within

the bandgap, allowing the number of trapped carriers to continue increasing as the Fermi level moves through the trap distribution. In this case, the dopant concentration at which the traps are completely filled need not coincide with the formation of a neutral region within the interior of the grain [5.10]. A neutral region can form within a grain even though the traps are not completely filled, and there may be a range of dopant concentrations over which the barrier height varies only weakly with dopant concentration.

The basic model can be extended to consider both high trap concentrations, trap energies away from mid-gap, and traps distributed in energy. Of these more complex cases, we first consider the effect of increasing the density N_T of traps per unit area located at a discrete energy E_T within the band gap at the grain boundary [5.10]. In this case, the probability of a trap being occupied arises from the requirement that the combination of the grain boundary and compensating depletion regions must be neutral:

$$2Nx_d = N_T \left[1 + \frac{1}{2} \exp \left(\frac{E_T + qV_B - E_F}{kT} \right) \right]^{-1} \quad (5.26)$$

The critical dopant concentration N^* is then found by letting $x_d = \frac{1}{2}L$:

$$N^* = \frac{8\epsilon}{q^2L^2} \left[E_F - E_T + kT \ln \left(\frac{2N_T}{N^*L} - 2 \right) \right] \quad (5.27)$$

which can be solved iteratively. At low trap concentrations, all the traps are filled before any neutral region forms, and $N^* \propto L^{-1}$, as we found before. However, for higher trap concentrations, a neutral region forms before all the traps are filled, and $N^* \propto L^{-2}$.

For $N < N^*$, the corresponding barrier height is

$$qV_B = \frac{q^2NL^2}{8\epsilon} \quad (5.28)$$

as before. For $N > N^*$, however,

$$qV_B = E_F - E_T + kT \ln \left[\frac{qN_T}{(2q\epsilon NV_B)^{1/2}} - 2 \right] \quad (5.29)$$

Figure 5.13 [5.10] shows the barrier height for a grain size of 100 nm, typical of polysilicon films used in integrated circuits. For low trap densities, the barrier height reaches a maximum at the critical dopant

concentration and then decreases, as we found earlier. However, for higher trap concentrations, the neutral region begins to form before all the traps are filled. As additional dopant atoms are added, the corresponding carriers fill the remaining unoccupied traps. Because a neutral region exists for dopant concentrations greater than N^* , the intrinsic Fermi level in the center of the grain remains fixed in relation to the Fermi level. At the grain boundary, the Fermi level is pinned near the discrete mid-gap trap level, and the barrier height remains nearly constant. Not until all the traps are filled can the Fermi level move above the trap level to decrease the barrier height. In the intermediate region, the barrier height is approximately [5.10]

$$qV_B \approx \frac{1}{2}E_G - E_T + kT \ln \left[\frac{qN^{1/2}N_T}{N_c(2q\epsilon V_B)^{1/2}} \right] \quad (5.30)$$

which varies only slowly with dopant concentration.

If the grain-boundary states are distributed in energy, rather than being located at a discrete energy, the behavior is more complex. For a trap density D_T per unit area per unit energy ($\text{cm}^{-2} \text{eV}^{-1}$) distributed uniformly in energy, the charge-neutrality condition can be written [5.10]

$$2Nx_d = (E_F - qV_B)D_T \quad (5.31)$$

The critical dopant concentration for this case again varies as L^{-1} for small grains and low trap densities, while it changes as L^{-2} for large grains and high trap densities. Another analysis considered a Gaussian energy distribution of trap states at the grain boundary [5.30]. Detailed comparison of experimental data with models which assume either discrete or uniformly distributed trap levels indicates that the trap distribution appears to be strongly peaked near mid-gap, rather than being broadly distributed within the band gap [5.10].

Key points of the basic carrier-trapping models

The models predict the resistivity to be high below the critical dopant concentration and to decrease rapidly above this value, as observed (Fig. 5.1). Below the critical dopant concentration the free-carrier concentration is low, and the resistivity is high. As the critical dopant concentration is approached, the barrier height increases; the ease of carrier motion from one grain to another (*ie*, the *effective mobility*)

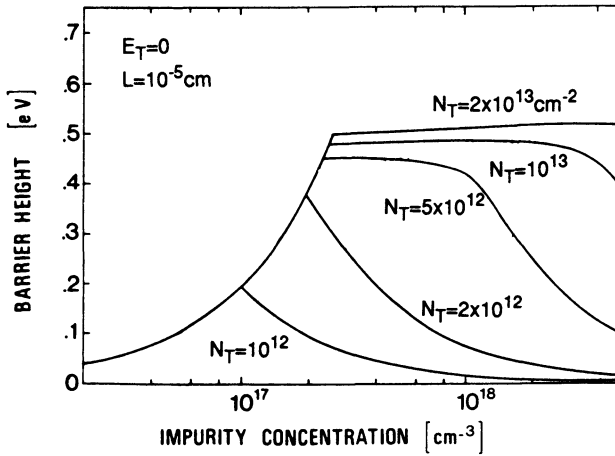


Figure 5.13: Barrier height as a function of dopant concentration when a neutral region forms before all the grain-boundary traps are filled. From [5.10]. Reprinted with permission.

decreases; and the resistivity remains high. The barrier height is maximum near the critical dopant concentration, and the effective mobility is minimum. Above the critical dopant concentration the free-carrier concentration increases and the barrier height decreases; in this range the resistivity decreases rapidly with increasing dopant concentration until it approaches that of single-crystal silicon.

5.4.4 Thermionic-field emission

The basic model presented above considers conduction only by thermionic emission of energetic carriers over the top of a parabolic potential barrier. However, near its top the barrier is narrow, and tunneling through the top portion of the barrier is likely for those carriers with energy slightly below the top of the barrier [5.31, 5.32].

The combination of thermionic emission and tunneling, called *thermionic-field emission*, can be modeled by calculating the number of carriers approaching the barrier with energy E and the probability $T(E)$ of each carrier tunneling through the portion of the potential barrier

$$V(x) = V_B \left(1 - \frac{|x|}{x_d} \right)^2 \quad (5.32)$$

for which $V(x) > E$.

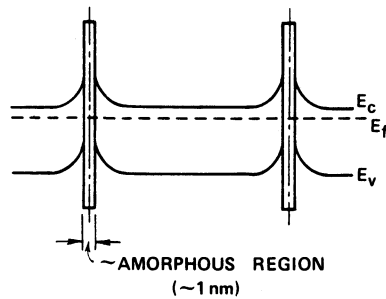


Figure 5.14: The impedance of the grain boundary is modeled by considering a high, narrow barrier at the grain boundary, in addition to the parabolic, depletion-region barriers surrounding the grain boundary.

For some ranges of dopant concentration, grain size, and temperature, thermionic-field emission can contribute a significant portion of the total current. For example, at room temperature with a dopant concentration of $5 \times 10^{17} \text{ cm}^{-3}$ and a grain size of 120 nm, the thermionic-field-emission current is estimated to be about 70% as large as the thermionic-emission current [5.32]. Thermionic-field emission becomes more important at lower temperatures, where fewer carriers have energies greater than the barrier height. This conduction mechanism will be discussed in more detail in the next section.

5.4.5 Grain-boundary barriers

In the basic carrier-trapping model the grain-boundaries themselves are assumed to be very narrow compared to the grains. Their only effect is to change the number of active dopant atoms and free carriers in the grains by acting as sites for dopant segregation and carrier trapping. However, the disordered nature of the grain boundaries and the discontinuities that they introduce into the periodic structure of the crystallites can also directly reduce the current flow. This additional impedance can be modeled by adding a high, narrow, potential barrier at the grain boundary, as shown in Fig. 5.14. This grain-boundary barrier is assumed to be present in addition to the parabolic, depletion-region potential barriers created by the uncompensated dopant atoms that neutralize carriers trapped at the grain boundary. The grain-boundary barrier can be modeled as a region with a width similar to that of the physical grain boundary (a few atom spacings) and a band gap different

from that of crystalline silicon. To move from one crystallite to another, the carriers must either tunnel through the grain-boundary barrier or be sufficiently energetic to be thermally excited over the barrier.

To treat the problem over a fairly wide range of dopant concentrations, we consider the composite barrier shown in Fig. 5.14, consisting of both the parabolic potential barrier and the high, grain-boundary barrier. Because the composite barrier can be wide and high, we consider thermionic-field emission, in which a carrier is thermally excited before tunneling through a portion of the barrier. This intermediate case can be treated by finding the number of carriers with an energy E incident on the composite barrier from each side and the probability that each carrier can tunnel through the barrier [5.33]. The number of carriers having energy within the range dE_x incident on the barrier per second per unit area (the *supply function*) is given by [5.34].

$$N(T, \psi, E_x)dE_x = \frac{4\pi m^* kT}{h^3} \ln \left[1 + \exp \left(-\frac{E_x + \psi}{kT} \right) \right] dE_x \quad (5.33)$$

where $\psi = E_c - E_F$, E_x is the energy component in the x -direction, and h is Planck's constant. The probability that a carrier with energy E_x can tunnel through the barrier is given in the WKB approximation by

$$T(E_x) = \exp \left\{ -\frac{4\pi}{h} \int_{x_1}^{x_2} [2m^*(qV(x) - E_x)]^{1/2} dx \right\} \quad (5.34)$$

where zero energy is taken as the bottom of the conduction band in the neutral regions of the grains, x_1 and x_2 are the classical turning points at which $E_x = V$, and m^* , the effective mass of the carriers, is assumed to be independent of the carrier energy.

The high, grain-boundary potential barrier is most important at intermediate and high dopant concentrations. At low dopant concentrations, the parabolic, depletion-region barrier limits the conduction. In the dopant-concentration region where both barriers must be included in the analysis, three different ranges of carrier energies are considered:

1. The carriers have energy less than that required to surmount the parabolic potential barrier in the depleted regions of the grains: Carrier transport occurs by tunneling through a portion of the parabolic potential barrier plus the grain-boundary barrier.
2. The carriers have sufficient energy to surmount the parabolic barrier in the depletion regions but not sufficient energy to travel

over the grain-boundary barrier: Transport occurs by tunneling through the grain-boundary barrier.

3. The carriers are sufficiently energetic to surmount the grain-boundary barrier: Carriers freely pass over the barrier.

The total current can be found by considering the energy distribution of the carriers and using the appropriate forms of the potential $V(x)$ in the expression for the net current — the difference between the currents traversing the barrier from the left and from the right. By comparing the model to experimental data, reasonable values of about 1 nm have been obtained for the width of the grain-boundary barrier [5.35].

The detailed behavior of the grain-boundary barrier can depend on the dopant being added [5.36]. For boron, the width of the barrier is independent of the dopant concentration, with a value in the range 0.7-0.9 nm. For phosphorus, however, the width appears to decrease with increasing dopant concentration above about 10^{18} cm^{-3} , changing from 1.1 nm at a dopant concentration of about 10^{18} cm^{-3} to about 0.5 nm for dopant concentrations in the mid- 10^{19} cm^{-3} range. This different behavior of boron and phosphorus is attributed to their different tendencies to segregate to the grain boundaries. The incorporation of phosphorus atoms into the disordered sites at a grain boundary is postulated to decrease its effective width. Since boron does not tend to segregate to grain boundaries, the grain-boundary width is independent of dopant concentration for boron-doped films. As we will see in Sec. 5.6, the added barriers at the grain boundaries can model the higher resistivity of polycrystalline silicon compared to single-crystal silicon observed for very high dopant concentrations, at which the parabolic, depletion-region barriers are small.

5.4.6 Limitations of models

The carrier-trapping model can be further refined, but these more detailed treatments often require that physical parameters be the same for all grains and grain boundaries in the layer. For example, the density of carrier traps N_T at a grain boundary depends on the structure of the grain boundary, which can vary from one grain boundary to another [5.37]. Photoconductivity measurements suggest that the potential barriers are larger for high-angle grain boundaries than for low-angle grain boundaries. The properties can even vary from point to point along

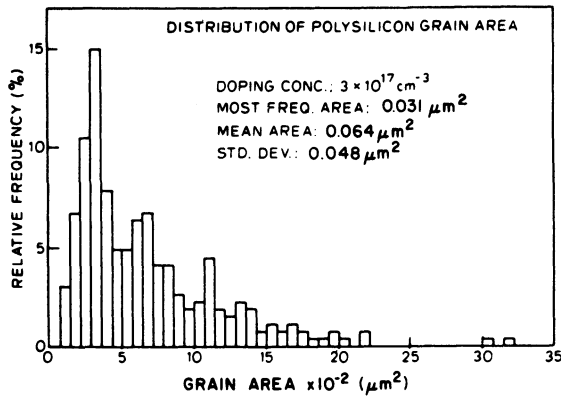


Figure 5.15: Distribution of grain sizes in a $1 \mu\text{m}$ thick polysilicon film. From [5.39]. ©1984, Pergamon Journals, Ltd. Reprinted with permission.

the same grain boundary. Boundaries with low densities of trap states, and therefore low barriers, are preferential current paths [5.38]. These variations limit the quantitative use of refined models.

Grain-size variations also limit detailed quantitative modeling. Although the “grain size” used to find the voltage across one grain boundary is consistent with the average measured grain size in some investigations, it differs in other studies and is often used as an adjustable parameter to match the experimental data; the models then only provide the functional form of the equation describing the conduction.

The number of grains N_G in the current path should be obtainable from the expression for the current flowing at high electric fields:

$$J = J_0 \sinh \left(\frac{qV}{2kTN_G} \right) \quad (5.35)$$

where V is the voltage applied across the sample. If a uniform grain size is assumed, N_G should be just the sample length divided by the grain size L . However, values of N_G found from Eq. 5.35 appear to vary with dopant concentration [5.39] while transmission electron microscopy shows that the grain size does not change appreciably with dopant concentration for moderate doping.

Most models assume that all grains are the same size both along the current-flow direction and perpendicular to the film plane. However, experimental data indicate that the sizes of grains in a given sample are

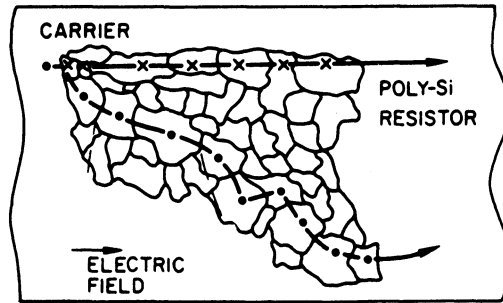


Figure 5.16: The nonuniform grain size in a polysilicon film leads to nonuniform current flow, making quantitative modeling difficult. From [5.39]. ©1984, Pergamon Journals, Ltd. Reprinted with permission.

fairly widely distributed, as shown in Fig. 5.15 [5.39], and the barrier height between grains can also vary markedly from one grain boundary to another. Consequently, the current flow is nonuniform, with the current traveling preferentially along lower resistance, but possibly longer paths, as indicated in Fig. 5.16 [5.39]. We also saw in Sec. 2.4 that the grain size varies through the thickness of the film, being greater near the top than near the bottom, so that the grain-boundary density is lowest near the top of the film. Thus, the resistance should be lower for paths closer to the film surface, and the current is expected to flow nonuniformly over the film thickness. If the current is a nonlinear function of the voltage, the current path can even vary as the applied voltage changes. The nonuniform current flow changes the effective number of grains in the path; the apparent variations in grain size inferred from electrical measurements are consistent with the effect of a distribution of grain sizes, both in the film plane and through the thickness of the film [5.39]. The effect of these nonuniformities is most noticeable near the critical dopant concentration, where some grains can be completely depleted while others contain a neutral region.

The nonuniform grain size in polysilicon films limits the usefulness of refined models which assume a uniform grain size. While differentiating between discrete and uniformly distributed energy levels for the traps seems to be possible, extracting more detailed information, such as the energy distribution of these levels may not be possible. Similarly, using experimental conduction data to differentiate between slightly different models is difficult. These refined models are only useful when an ad-

equate statistical description of the grain-size distribution is available. However, this distribution depends sensitively on deposition conditions. Thus, the carrier-trapping model is a useful, semi-quantitative tool for describing conduction in polysilicon, but overly detailed refinement of this model requires more accurate values of physical parameters than can realistically be obtained.

5.4.7 Segregation and trapping

Although we have considered dopant segregation and carrier trapping independently, possible interactions can occur. We saw in Sec. 3.6 that dopant atoms become electrically inactive when they segregate at grain boundaries. However, more subtle modification of the electrical properties of the grain boundaries can be caused by the segregated dopant [3.24]. Dopant atoms segregating to the grain boundaries may interact with the dangling bonds or other defects which cause the grain-boundary traps, satisfying some of the unsaturated bonds and reducing the trap density and, consequently, the resistivity. For example, phosphorus segregation has been postulated to decrease the trap density and lower the barrier height [5.38]. The lower vibrational entropy implied by the decrease of A in Eq. 3.7 as dopant segregates suggests that segregation [3.24], as well as the heat treatment itself [5.11], causes the grain boundaries to become more ordered, possibly decreasing the number of carrier traps there [3.24]. Thus, as the dopant segregates, dopant loss from substitutional sites within the grains increases the resistivity, but reduced carrier trapping at grain boundaries tends to decrease it [5.40]. (These effects are most readily observed in moderately doped films, in which carrier trapping affects the electrical behavior most strongly.)

On the other hand, if dopant atoms at the grain boundary are readily ionized, they may cause additional Coulomb scattering and reduce the majority-carrier mobility. They may possibly change the minority-carrier transport as well. (These effects are most visible in highly doped samples.) However, loss of dopant atoms by segregation usually dominates, and modification of the carrier traps or scattering is secondary. Dopant segregation may also promote grain-boundary reconstruction, again changing the number of traps at the grain boundary. Unlike the other effects discussed, however, reconstruction is not reversible.

Although the disorder at grain boundaries makes complete bonding of dopant atoms there possible, magnetic resonance spectroscopy

of phosphorus segregated at the grain boundaries in polysilicon suggests that the phosphorus is four-fold coordinated [5.41]. Four-fold coordination implies that the dopant is electrically inactivated by carrier trapping, rather than by changes in the number of bonds.

5.4.8 Summary: Moderately doped polysilicon

In this section, we have seen that three interrelated effects lead to the marked differences between the electrical properties of polysilicon and single-crystal silicon: First, atoms of some dopant species physically segregate to the grain boundaries and do not contribute to the conduction. Second, some of the carriers provided by substitutional dopant atoms within the grains are immobilized by carrier traps at the grain boundaries and are lost to the conduction process. Third, potential barriers at the grain boundaries impede the movement of the remaining free carriers from one grain to another.

As the dopant concentration increases beyond its critical value, the barrier height decreases until transport across the barrier no longer limits the conduction. Other conduction mechanisms then must be considered to understand the resistivity of polycrystalline silicon. For example, when the grain-boundary barriers are low and ionized impurity scattering within the grains limits the conduction, the conductivity decreases with increasing temperature. This temperature dependence contrasts with the increase in conductivity with increasing temperature found when thermionic emission over the grain-boundary barriers limits conduction. In Sec. 6.4 we will see that the change in the temperature coefficient of resistivity (TCR) from negative to positive with increasing dopant concentration is important for the low-value resistors often fabricated in polysilicon. At even higher dopant concentrations, other mechanisms limit the conduction. Thus, the carrier-trapping models explain the electrical behavior over an important dopant-concentration range, but do not apply to very lightly or very heavily doped polysilicon.

5.5 Grain-boundary modification

Because the traps located at grain boundaries strongly degrade the electrical properties of polysilicon and devices fabricated in it, attempts are made to reduce the density of active traps associated with the grain

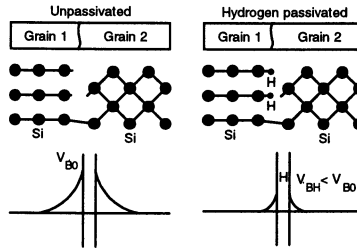


Figure 5.17: As the dangling bonds at the grain boundary are “passivated” with hydrogen, the depletion barrier surrounding the boundary decreases. From [6.122]. ©1985 IEEE. Reprinted with permission.

boundaries. Either the effectiveness of the trapping sites can be reduced or the grain boundaries themselves can be removed.

5.5.1 Grain-boundary passivation

In the first approach, the defects are *passivated* so that the states within the forbidden gap are not active and no longer trap carriers. As we have seen, traps are associated with the dangling bonds at the grain boundaries arising from the lattice discontinuities which form when differently oriented grains join. Dangling bonds at a Si/SiO₂ interface are often passivated by terminating them with hydrogen atoms. Similarly, hydrogen can passivate dangling bonds and other defects at grain boundaries [5.42] and reduce the number of active traps there. The passivation of grain boundaries has been directly observed by electron spin resonance [5.43]. The residual spin density, corresponding to unpassivated grain-boundary states, varies with the passivation temperature and is lowest after *hydrogenation* at 350°C. As the number of trapped carriers decreases, the potential barrier associated with the grain boundary also decreases (Fig. 5.17) [6.122]. The resistivity of moderately doped polycrystalline silicon is markedly reduced by incorporating hydrogen [5.44], and it becomes less sensitive to small variations in the dopant concentration. The properties of devices fabricated in the hydrogenated polysilicon are also improved [6.132], as we will discuss in Sec. 6.13.

Passivation of grain-boundary traps has received increasing attention because of the use of polysilicon thin-film transistors in active-matrix, liquid-crystal displays (Sec. 6.13).

The atomic hydrogen needed to passivate the dangling bonds at grain boundaries cannot be readily obtained by annealing in molecular hydrogen or in forming gas (a N_2/H_2 mixture), but several other techniques can introduce atomic hydrogen into polysilicon.

A plasma dissociates molecular hydrogen; the resulting atomic hydrogen can then diffuse into the grain boundaries, interact with the defects, and effectively passivate the grain-boundary traps, lowering the grain-boundary-limited resistance [5.44, 5.45] and improving device characteristics [6.132]. Only a fraction of the added hydrogen passivates grain-boundary states [5.43] so that the amount of hydrogen added must be considerably greater than the number of grain-boundary states to be passivated. The efficiency of the hydrogenation process increases with increasing dissociation of H_2 or H_2^+ into atomic H and H^+ [5.46]. In addition to the concentration of hydrogen ions, the effectiveness of plasma hydrogenation depends on the temperature of the substrate, and the duration of the treatment [5.47]. The effectiveness of hydrogenation may possibly also be increased by coupling acoustic energy into the system [5.48]. Plasma hydrogenation is the most widely used hydrogenation technique, but it is time consuming.

In addition to passivating grain-boundary defects, plasma hydrogenation can also neutralize dopant atoms, especially boron, reducing the electrically active dopant concentration [5.49, 5.50]. For moderately doped films, the decreased dopant concentration may increase the grain-boundary barrier, adversely affecting the resistivity.

Hydrogen can also be added effectively by diffusing it from a hydrogen-containing, plasma-deposited, nitride layer ($SiN_x:H$) of the type often used to passivate integrated circuits [5.51]. The hydrogen content in these nitride layers is often 10-20%, and the hydrogen is weakly bound; it diffuses readily and can move from the nitride into an adjacent polysilicon film during a mild heat treatment — often at $\sim 400^\circ C$. The effectiveness of adding hydrogen by this technique depends on the annealing temperature and time. The technique is simple, but control and subsequent device stability are questionable.

Hydrogen passivation can also be obtained by implanting hydrogen into the polysilicon film. Subsequent annealing at a low temperature moves the hydrogen to the grain boundaries, where it can interact with

the defects to reduce the effectiveness of the grain-boundary traps [5.45]. The hydrogen also appears to activate some of the dopant atoms segregated to grain boundaries [5.52], causing somewhat different interactions with dopants that segregate and those that do not. The effectiveness of hydrogenation by ion implantation depends on the hydrogen dose and the subsequent annealing temperature. The technique is very controllable, and throughput can be high, but implant damage can possibly degrade the devices. To reduce implant damage, the hydrogen can be implanted into the surrounding dielectrics and diffused into the polysilicon, but much higher hydrogen doses are then needed.

Although the hydrogen added to polysilicon can passivate the grain boundaries, it is not firmly bound. The hydrogen, and consequently the passivation, are rapidly lost from the grain boundaries by annealing at temperatures in the 400-500°C range [5.52, 6.132]. Therefore, it must be added after all high-temperature heat treatments in the device-fabrication process are complete. The hydrogenation process is repeatable, and hydrogen can be re-incorporated by another exposure to active atomic hydrogen. Positively charged hydrogen species may possibly move in electric fields within devices [5.53], causing concern about long-term reliability of hydrogen-passivated devices. The limited stability of hydrogen has led to the investigation of deuterium for similar passivation of grain boundaries, but with greater stability [5.54]. In addition to hydrogen and deuterium, lithium can also passivate grain boundaries [5.55] and neutralize recombination centers [5.56]. Although lithium is less compatible with integrated-circuit processing than hydrogen, lithium has been suggested for use in solar cells.

5.5.2 Recrystallization

Even when the grain boundaries are passivated with hydrogen to the maximum extent practical, the number of active grain-boundary states is still appreciable. To obtain devices of markedly higher quality, the grain boundaries themselves must be removed or at least reduced in number. As we saw in Sec. 2.11.2, solid-phase grain growth of moderately doped polysilicon requires extremely high temperatures for long times and is incompatible with integrated-circuit processing. However, large grains can be obtained by rapidly melting and solidifying the fine-grain polysilicon film with a scanned heat source such as a cw laser, electron beam, graphite strip-heater, or high-intensity incoher-

ent lamp [2.119]. When the silicon solidifies, only a relatively small number of grains nucleate; these grains grow laterally into the rapidly cooling molten silicon to produce large-grain material containing few grain boundaries with their associated traps. The entire thickness of the fine-grain silicon must be melted so that the solidifying silicon does not regrow on the underlying small grains to again produce a fine-grain structure. The moving heat source provides the temperature gradient needed for lateral growth from a few nucleating grains [5.57]. Alternatively, the lateral thermal gradient can be obtained by locally varying the amount of power absorbed by partially covering the surface with layers of varying reflectivity [5.57]. The heat loss into the underlying substrate may also be varied locally by placing oxides of different thicknesses under different regions of the polysilicon [5.57].

Controlled lateral temperature gradients are not readily obtained with a pulsed laser. However, the availability of high-power lasers, especially excimer lasers, with relatively uniform light intensity over areas of $\sim 1 \text{ cm}^2$ makes these lasers useful for significantly improving the properties of polysilicon films [5.58]. Although melting the layer with a pulsed laser may not increase the grain size optimally, it can reduce the defect density within the grains [5.59]. The grain size can be increased further by locally varying the power absorption by placing patterned layers on the surface to change the optical absorption [5.57, 5.60].

Crystallizing a film deposited in the amorphous form with a pulsed laser can be especially attractive when the film is deposited on a substrate with a low melting or softening temperature, such as inexpensive glass. Because the laser pulse is short, the silicon film stays at the melting temperature for $\leq 100 \text{ ns}$, limiting the heating of the underlying glass substrate.

In moderately doped, fine-grain polysilicon, the resistivity varies rapidly with dopant concentration as the grain-boundary traps become saturated with carriers. Because most of the grain boundaries are removed in recrystallized polysilicon, the resistivity is much less sensitive to variations in the dopant concentration. For example, for a resistivity in the intermediate range, melting with a laser can reduce the resistivity variation from a factor of ten to a factor of two for a 30% change in the implanted dose [5.61] if the resistor dimensions are large enough to sample many grains. If the dimensions are small, however, only a few grains are included, and significant statistical resistivity variations can occur.

The remaining grain boundaries in the recrystallized silicon film arise from random nucleation. In critical applications even these few grain boundaries must be removed. This can be accomplished by opening windows periodically in the oxide above the single-crystal substrate before the fine-grain polysilicon is deposited. As the molten silicon solidifies during recrystallization, the crystal structure of the substrate propagates into the solidifying material, first vertically, and then laterally over the oxide, to produce grain-boundary-free, single-crystal silicon over the insulating oxide. However, the distance that the grain-boundary-free material can propagate laterally is limited by competing random nucleation, restricting the use of this *seeding* technique.

As an alternate approach, transparent material of varying thickness can be placed above the silicon before recrystallization so that the amount of power absorbed from a narrow-band optical source varies from one region to another. The regions that absorb the least power are coolest and solidify first so that any grain boundaries that subsequently form are outside of these regions. Because the location of the grain boundaries can be controlled photolithographically using this technique, they can be placed outside of the active device regions (*eg*, in the isolation regions between devices). This technique can be coupled with the seeding technique discussed above to control the orientation of the grain-boundary-free material.

5.6 Heavily doped polysilicon

The major use of polysilicon films in ULSI circuits is for the gate electrode and one layer of interconnections in MOS integrated circuits. The material used for interconnections must have the lowest possible resistivity to increase circuit speed and avoid voltage drops along long lines. However, as increasing concentrations of dopant atoms are added to polycrystalline silicon (for example, by ion implantation), the resistivity decreases to a limiting value and then remains constant, as shown in Fig. 5.18 [5.62]. Because boron does not segregate to grain boundaries, it is the most effective and controllable dopant in moderately doped films; however, at high dopant concentrations the resistivity of phosphorus-doped polysilicon is lower. The limiting resistivities are about $400\ \mu\Omega\text{-cm}$ in phosphorus-doped polysilicon and about $2000\ \mu\Omega\text{-cm}$ in arsenic- or boron-doped polysilicon after annealing at 1000°C .

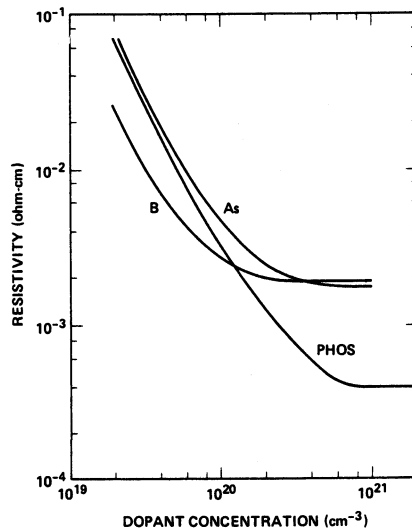


Figure 5.18: The resistivity of polysilicon after annealing at 1000°C is limited to approximately 2000 $\mu\Omega\text{-cm}$ for boron and arsenic doping and about 400 $\mu\Omega\text{-cm}$ when the polysilicon is doped with phosphorus. From [5.62]. Reprinted with permission.

Hall measurements [5.62–5.64] suggest that the electrically active carrier concentration saturates as the dopant concentration increases and remains constant as more dopant atoms are added. The dopant concentrations at which the number of electrically active carriers saturates are somewhat less than $1 \times 10^{21} \text{ cm}^{-3}$ for phosphorus, $\sim 2 \times 10^{20} \text{ cm}^{-3}$ for boron and $\sim 6 \times 10^{20} \text{ cm}^{-3}$ for arsenic [5.62, 5.63]. However, the number of electrically active carriers is considerably less than the dopant concentration needed to reach saturation in polysilicon, as in single-crystal silicon [5.65]. The carrier mobilities in heavily doped polysilicon are about half those in single-crystal silicon of the same dopant concentration. Similar behavior has been found for films deposited under widely varying deposition conditions, as discussed below.

5.6.1 Solid solubility

While the resistivity in moderately doped polysilicon is much greater than that in similarly doped single-crystal silicon, the resistivity in heavily doped polysilicon is only about twice that in correspondingly doped single-crystal silicon. Values of resistivity approaching those in

single-crystal silicon suggest a common resistivity-limiting mechanism. In single-crystal silicon at high dopant concentrations, the amount of substitutional dopant that can be incorporated into the lattice is limited by the solid solubility of the dopant species in silicon [5.65, 5.66]. Solid solubility also appears to limit the amount of dopant that can be actively incorporated into polysilicon [5.67], with the same value of solid solubility in the crystalline grains of polysilicon as in single-crystal silicon [5.68]. At high phosphorus concentrations the carrier concentration saturates at that corresponding to solid solubility [5.69]; for arsenic, the saturation value appears to be less than the reported solid-solubility carrier concentration [5.70], possibly because of uncertainty in the value of arsenic solubility in silicon [5.63], as well as significant arsenic precipitation and segregation. The limiting resistivity is greater in arsenic-doped films than in phosphorus-doped ones for polysilicon, as for single-crystal silicon [5.67]. The solid solubility depends more strongly on temperature for boron than for arsenic over the 800–1000°C temperature range [5.71].

Solid solubility appears to limit the conductivity in films deposited over a wide range of conditions, including those deposited in cold-wall reactors over the temperature range from 700 to 1000°C [1.33], as well as films deposited in the more common, hot-wall, low-pressure reactors.

In somewhat less heavily doped polysilicon, the observed carrier concentration is less than the dopant concentration, probably because of dopant segregation at the grain boundaries. For phosphorus, about 80% of the dopant atoms are electrically active, while for arsenic, which segregates more strongly, only about 60% are active, contributing to the lower conductivity in arsenic-doped films even below solid solubility [5.63].

5.6.2 Method of doping

Gas-phase doping: Dopant atoms can be added to polysilicon during deposition by introducing a phosphorus-containing gas such as phosphine PH_3 into the deposition chamber or after deposition either by diffusion from a gaseous source such as PH_3 or POCl_3 or by ion implantation. When phosphorus is added to the films from a gaseous source at a relatively high temperature (so that thermal equilibrium is attained in a relatively short time), the amount of phosphorus incorporated appears to equal its solid-solubility value at the doping temperature. Auger

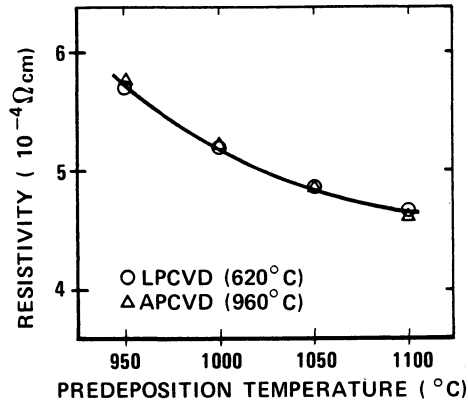


Figure 5.19: After doping from a gaseous source, the resistivity of polysilicon is slightly lower for higher doping temperatures.

analysis of films doped from a POCl_3 source suggests that the total phosphorus concentration increases monotonically with increasing doping temperature, corresponding to its increased solid solubility at higher temperatures. Relatively little additional dopant is incorporated within the grain boundaries or as precipitates *during the deposition process*. The amount of dopant needed to saturate the dopant segregation sites at grain boundaries appears to be only a small fraction of the total dopant concentration in very heavily doped films.

The electrical properties also reflect the increasing solid solubility with increasing temperature in the normal temperature range [5.65]. The resistivity decreases slightly with increasing doping temperature, as shown in Fig. 5.19 [5.72]. Figure 5.20 [5.72] shows that, even in single-crystal silicon, the limiting carrier concentration n_{SS} is lower than the solid-solubility dopant concentration C_{SS} [5.65, 5.66]. If the electrically active carrier concentration corresponding to solid solubility at each doping temperature is combined with the resistivity of the corresponding film (Fig. 5.19), a nearly constant mobility of about $32 \text{ cm}^2/\text{V-s}$ is calculated over the entire doping-temperature range considered. This value is similar to that obtained from Hall measurements and is about half the single-crystal mobility at the same dopant concentration.

Ion implantation: Although little excess phosphorus is incorporated when polysilicon films are doped from a gaseous source, average concentrations of dopant atoms well in excess of solid solubility can be added by

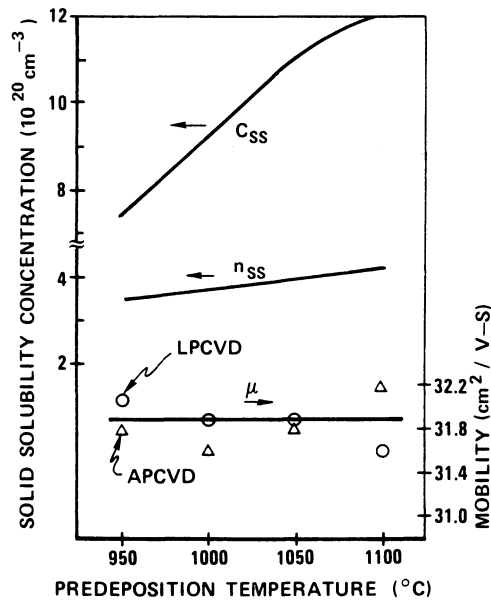


Figure 5.20: The resistivity is consistent with a free-carrier concentration n_{SS} corresponding to the solid solubility C_{SS} of phosphorus at the doping temperature and a mobility μ equal to about half that in similarly doped, single-crystal silicon.

ion implantation. However, dopant atoms at concentrations above solid solubility are not usually electrically active [5.71]. For at least phosphorus, the excess dopant atoms can precipitate at grain boundaries and may seriously degrade the integrated circuit. For example, as we saw in Sec. 4.2.3, phosphorus-containing precipitates can oxidize more rapidly than polysilicon to form a low-quality oxide [4.7]. Subsequent etching of the oxide can leave voids extending through much or all of the thickness of the polysilicon in these regions and attack the underlying gate oxide. For high dopant concentrations, conventional ion implantation from mass-separated sources can be slow. Alternative techniques are being investigated to add large amounts of dopant rapidly, especially for large-area polysilicon devices [5.73, 5.74].

Diffusion from doped oxides: Polysilicon films can also be doped by diffusion from a doped oxide formed by chemical vapor deposition or from a spin-on glass containing a high concentration of the dopant species [5.75]. Using a spin-on glass requires careful evaporation of the

solvent before heating to the dopant diffusion temperature to preserve the film integrity. The portion of the doped oxide or glass adjacent to the polysilicon must be heavily doped; a thin undoped region can block the diffusion of boron. For phosphorus, the phosphorus-containing region can “melt through” a thin undoped region.

***In situ* doping:** The kinetics of *in situ* doping were discussed in Sec. 1.8. Excess dopant can be incorporated when the dopant is added during deposition. At the low temperatures at which polysilicon films are usually deposited, the time needed for the dopant to reach equilibrium can be much longer than the time to deposit a few atom layers of silicon, and nonequilibrium dopant concentrations in excess of solid solubility can be incorporated. The excess dopant may be electrically active immediately after deposition, contributing carriers to the conduction process [5.68, 5.76].

Because of the nonunity segregation coefficient of the dopant from the gas phase into the depositing silicon film, the dopant-gas concentration can decrease at a different rate than the silicon-gas concentration along the length of the reaction chamber. Therefore, the concentration of dopant incorporated into the film can vary along the length of the chamber. When the dopant concentration is above solid solubility, the amount of excess dopant in different films can vary even though the carrier concentration of all the films corresponds to solid solubility. Consequently, the amount of precipitation and the related film degradation occurring later in the device-fabrication process can vary with the position of the wafer in the reactor.

5.6.3 Stability

Although carrier concentrations corresponding to high solid solubilities can be introduced at high doping temperatures, the dopant may not remain electrically active after further, lower-temperature, heat treatments. Because the solid solubility is usually lower at the lower temperature, the dopant is rejected from active substitutional sites. It may precipitate within the grains, or more likely at the grain boundaries, where the disordered structure provides sites for easy precipitation. If the time of the lower-temperature heat treatment is long enough for thermal equilibrium to be reached, the dopant in excess of solid solubility at the lower temperature precipitates and no longer contributes to the conductivity of the film. However, during very short heat treat-

ments, the dopant may not have time to reach equilibrium at the new temperature, and active dopant concentrations in excess of solid solubility can be obtained.

The effect of annealing on dopant deactivation can be seen for polysilicon films doped with phosphorus from a gas-phase source at 1000°C and subsequently annealed at lower temperatures for a limited time [5.63]. High active dopant concentrations can be obtained by rapidly heating to a high temperature with a laser and cooling quickly. Subsequent furnace annealing can reduce the active carrier concentration [5.63]. At very low annealing temperatures, the carrier concentration changes little, although the dopant concentration is substantially above solid solubility at these temperatures; the annealing time is not long enough for excess phosphorus to leave solution. At intermediate annealing temperatures, less time is needed, and the carrier concentration decreases as excess phosphorus is able to leave solution during the heat treatment. Some of the excess dopant is likely to leave solution even during the final integrated-circuit annealing process in the 400°C temperature range. At high temperatures, the carrier concentration again increases because of the increased solid solubility. Assuming that the amount of dopant added during gas-phase doping equals the solid solubility at the doping temperature, the phosphorus concentration should be below solid solubility at temperatures higher than the doping temperature, and the carrier concentration should not depend strongly on the annealing temperature.

In addition to the nominal annealing temperature and time, the rate of temperature transitions can affect the final resistivity. The cooling rate is especially important after annealing heavily doped films at high temperatures. At high temperatures the dopant solid solubility is high. As the films are cooled the equilibrium solid solubility decreases. If the cooling is slow, much of the dopant has time to leave solution; the active carrier concentration decreases; and the resistivity increases, as discussed in Sec. 3.6.2 [3.92]. The slow cooling used to avoid wafer deformation increases the loss of dopant. The fast cooling used after rapid thermal processing minimizes the dopant loss.

Competing changes during annealing can lead to complex changes in the resistivity. *In situ* doped films deposited at low temperatures can contain dopant concentrations above solid solubility at the deposition temperature. When they are annealed at higher temperatures, the grains may grow, with a corresponding increase in mobility and de-

crease in resistivity. At the same time, excess dopant can leave solution, increasing the resistivity.

5.6.4 Mobility

Although the barrier to conduction created by the depletion regions surrounding the grain boundaries decreases at high dopant concentrations (as expected from Poisson's equation), the mobility at the highest dopant concentrations is only about half that in single-crystal silicon. Similar values are found from Hall-mobility measurements [5.15, 5.62] and inferred from the limiting resistivity and the carrier concentration corresponding to solid solubility [5.72] (Fig. 5.20). The high, narrow, grain-boundary barrier, through which the carriers must tunnel [5.32, 5.33], can explain this difference. A barrier thickness of about 1 nm and a barrier height corresponding to amorphous silicon are consistent with the experimental data [5.35, 5.36].

The mobility is also different for different *n*-type dopant species [5.67, 5.68], and is lower in arsenic-doped films than in phosphorus-doped films [5.63, 5.64]. A value close to 30 cm²/V-s is obtained for the highest phosphorus concentrations that can be achieved by furnace annealing, but the mobility is only about 20 cm²/V-s for arsenic-doped polysilicon [5.15, 5.63, 5.64]. The lower mobility in arsenic-doped films is attributed both to stronger ionized-impurity scattering by arsenic, as in single-crystal silicon [5.63], and to more effective grain-boundary scattering in arsenic-doped polysilicon than in phosphorus-doped films. Melting the film with a laser reduces the strong grain-boundary scattering in arsenic-doped polysilicon [5.63]. The resistivity of furnace-annealed films containing high concentrations of both arsenic and phosphorus is higher than that for films doped only with phosphorus. This behavior, attributed to the increased grain-boundary scattering caused by arsenic, degrades the performance of some MOS integrated circuits in which the arsenic implantation used to dope the source and drain regions also enters the polysilicon gate that has previously been doped with phosphorus.

5.6.5 Trends

The trend toward lower processing temperatures for ULSI fabrication decreases the conductivity that can be achieved in polysilicon.

Using short, high-temperature, heat cycles near the end of the device-processing sequence can improve the conductivity [5.77]. Because grain boundaries and other defects limit the mobility in even heavily doped polysilicon, the mobility can be improved somewhat by reducing the number of grain boundaries. Larger grains can potentially be obtained by solid-state grain growth or by liquid-phase recrystallization, but the improvement in mobility is limited to approximately a factor of two for very heavily doped polysilicon.

Although the limited conductivity of polysilicon is becoming an increasingly severe constraint on the performance of small-geometry, high-performance integrated circuits, polysilicon is often augmented by an overlayer of a more highly conductive material, such as a refractory metal silicide, as we saw in Sec. 3.5.3. Retaining the superior electrical properties and excellent stability of MOS transistors with polysilicon directly over the gate insulator is strong motivation for augmenting, rather than replacing, polysilicon in the near future.

However, as will be discussed in Sec. 6.2.6, the modulation of the carrier concentration by the gate voltage within the polysilicon near the underlying gate dielectric can create a depletion region that decreases the effective coupling of the gate voltage to the transistor channel and degrades the maximum transistor current. This limitation may eventually lead to the replacement of polysilicon with a metal [6.11]. Even in this case, however, the gate may be initially formed from polysilicon — to utilize its compatibility with high-temperature processing — and replaced with metal later in the process [6.12].

5.7 Minority-carrier properties

5.7.1 Lifetime

In our discussion of conduction in polycrystalline silicon, we saw that the current was limited by potential barriers that impede the flow of majority carriers. Minority carriers, on the other hand, see an attractive, rather than a repulsive, potential, and can readily move to the grain-boundary defect states and recombine with trapped majority carriers. Because of the efficient recombination at grain boundaries, the minority-carrier lifetime in polysilicon is expected to be markedly less than in single-crystal silicon. In addition, the defect concentration within the grains of polysilicon is higher than in high-quality, single-crystal sili-

con, and recombination at defects within the grains again reduces the minority-carrier lifetime.

The effective lifetime τ_{eff} can be written in terms of the volume lifetime τ_b in the grains and the grain-boundary recombination velocity s [5.24, 5.78, 5.79]. Alternatively, it can be expressed as [5.80]

$$\tau_{\text{eff}} = \frac{2d \exp(-qV_B/kT)}{3\sigma v D_T (E_{fn} - E_{fp})} \quad (5.36)$$

where D_T is the interface trap density ($\text{cm}^{-2} \text{ eV}^{-1}$), σ is the capture cross section, d is the grain size, v is the thermal velocity of the carriers, V_B is the height of the potential barrier at the grain boundary, and E_{fn} and E_{fp} are the quasi-Fermi levels for electrons and holes, respectively.

Another approach estimates the lifetime by treating the surface recombination sites as if they were distributed uniformly throughout the bulk of the material [5.14]. In this approximation

$$\tau_{\text{eff}} = \frac{1}{\sigma v N_{sr}} \quad (5.37)$$

where N_{sr} is the effective density of recombination centers (per unit volume). For cubic grains, the number of recombination centers per unit volume is given by $N_{sr} = 3N_{it}/d$, where N_{it} is the concentration of recombination centers per unit area at the grain boundary.

Using the latter approach, the lifetime can be found from diffusion length measurements, from photoconductive decay, or from the reverse characteristics of p - n junction diodes. The reverse characteristics of diodes formed within material with submicrometer grain size are dominated by generation within the space-charge region of the p - n junction, and the corresponding lifetime is of the order of 20–100 ps [6.106]. In thicker films with larger grains, the lifetime inferred from forward diode characteristics is about 100–300 ps in polysilicon, much less than the microsecond lifetime found in similarly doped single-crystal films.

Because the total grain-boundary area per unit volume depends on the grain size, the lifetime also varies as the grain size changes. Figure 5.21 [5.14] shows the predicted linear dependence of lifetime on grain size. From the empirical relation $\tau_{\text{eff}} = 5 \times 10^{-6} d$ found from Fig. 5.21, and $v = 10^7 \text{ cm/s}$, the product $\sigma N_{it} = 6.7 \times 10^{-3}$. With the value of $\sigma = 2 \times 10^{-16} \text{ cm}^2$ measured for the capture cross section of surface states, N_{it} is found to be $3 \times 10^{13} \text{ cm}^{-2}$ — about an order of magnitude

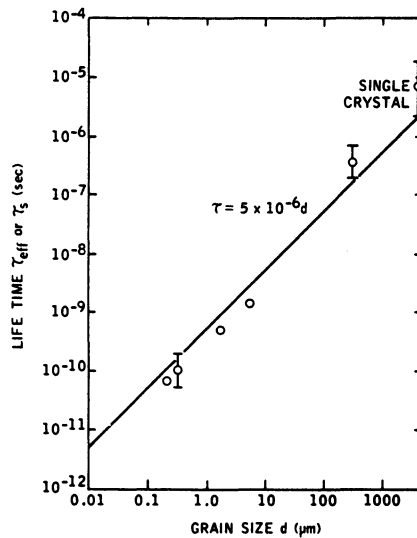


Figure 5.21: The minority-carrier lifetime in polysilicon increases approximately linearly with increasing grain size. From [5.14]. Reprinted with permission.

greater than the density of carrier traps found from the majority-carrier conductivity. However, the different effect of lifetime in neutral regions and in depletion regions should also be considered in interpreting the experimental data [5.81].

5.7.2 Switching characteristics

The properties of the parabolic, depletion-region barriers between grains not only depend on the physical nature of the grain boundary, but they can also be modulated by the electrical bias applied and the current flowing. Applying a high voltage can cause the current to increase abruptly. The change may persist only while the voltage is applied (*threshold switching*), or it can be permanent (*memory switching*).

Temporary threshold switching is usually caused by changes in the net charge in the grain-boundary traps and the surrounding depletion regions by excess carriers traveling across the sample [5.82–5.85]. As the charge that causes the depletion-region barriers is neutralized by injected charge, the depletion-region barriers decrease, and the majority-carrier current flowing in the sample increases. The resulting abrupt change in the conductivity with applied voltage leads to the observed

threshold switching. The dynamics of charge trapping and emission can also cause anomalous capacitive current [5.86].

In the simplest case, *minority carriers* injected into the grains cause threshold switching by neutralizing some of the majority carriers trapped at the grain boundaries, with the consequent barrier-height decrease and current increase. Consider a p^+-n junction within a polysilicon layer [5.87]. At low voltages, only a few holes are injected from the p -type polysilicon into the n -type polysilicon. The trapped charge at the grain boundary is not appreciably changed, and the resistivity of the n -type region remains high. However, when a large number of holes is injected, the holes can interact strongly with the grain-boundary traps. The holes can recombine with trapped electrons or themselves be trapped, depending on the trap energies. In either case, the net negative charge in the grain boundary decreases, and the amount of compensating charge in the surrounding space-charge region also decreases, lowering the barrier height. This reduction of the barrier height increases the conductivity of the polysilicon, so that more voltage is available to forward bias the p^+-n junction. More holes are then injected into the polysilicon, leading to regenerative feedback above a critical *switching voltage*. When sufficient charge has been injected to neutralize the grain boundaries in the lightly doped region, only a small *holding voltage* is required to provide the current needed to maintain the grain boundaries in their low-resistance, neutral state. Similar behavior can occur if charge is injected from the electrodes into the polysilicon. In either case, the resistance of the polysilicon appears to “switch” from a high-impedance state to a low-impedance state. When the voltage is removed, the excess minority carriers recombine; majority carriers are again trapped; and the resistance of the polysilicon returns to its initial high value determined by the zero-bias barrier height.

Injection of excess *majority carriers* can also cause threshold switching. Initially, some of the injected carriers are trapped at the grain boundaries (assuming that not all of the traps were previously filled), and the barrier height increases. However, once the grain-boundary traps are filled, additional majority carriers are not trapped at the grain boundaries, but remain in the grains to neutralize the uncompensated dopant atoms near the grain boundaries, decreasing the barrier height.

In addition to charge injection, the applied voltage itself can cause threshold switching. When a voltage is applied, a portion of the voltage appears across the depletion region on each side of each grain bound-

ary. The resulting high electric field across the reverse-biased side of the barrier can narrow and lower the barrier. Some of the trapped carriers can then escape from the grain-boundary traps either by tunneling through the narrowed barrier, or by thermionic emission over the barrier, which has been lowered by the applied voltage (or by thermionic-field emission). As the traps at one grain boundary empty, the barrier height decreases, allowing more of the applied voltage to appear across other grain boundaries, and similar *field-assisted trap emptying* becomes likely at the other grain boundaries. This regenerative effect can lead to a rapid increase in the conductivity of the polysilicon as the voltage increases.

The switching behavior depends strongly on the temperature, with the probability of switching increasing rapidly with increasing temperature. The switching voltage and the holding voltage both increase significantly when the number of grains across which the voltage is applied increases, but the switching and holding currents are relatively insensitive to the number of grains. In fine-grain polysilicon, the applied voltage is divided across many grain boundaries so that any effect of the statistical inhomogeneities from one grain to another is small. However, in large-grain material, formed by laser recrystallization for example, only a few grain boundaries appear in the conduction path, and multiple, discrete, negative-resistance switching voltages are observed [5.84].

At higher currents, a large, permanent reduction in the resistivity of the polysilicon can occur. This *memory switching* is probably related to the formation of a permanent, relatively conductive filament by resistive heating of the polysilicon when a high current flows [5.82]. Dopant atoms from the heavily doped contact regions can diffuse rapidly along a very hot solid path or a molten region to form a heavily doped, conductive channel that remains when the applied voltage is removed [5.88, 5.89]. Memory switching can decrease the resistance by a factor of 10^6 or more because of the highly nonlinear dependence of the resistivity of polysilicon on the dopant concentration. More complex memory switching behavior is observed in polysilicon doped with both boron and phosphorus. It has been suggested that *n*-type regions within the grains are converted into *p*-type regions as the phosphorus content within the grains is reduced by the tendency of phosphorus to segregate to grain boundaries [5.90].

5.8 Summary

Grain-boundary effects cause the electrical properties of polycrystalline silicon to differ from those of single-crystal silicon. At low and moderate dopant concentrations, dopant segregation to and carrier trapping at grain boundaries reduces the conductivity of polysilicon markedly compared to that of similarly doped, single-crystal silicon. Not only is the number of carriers reduced, but movement of the carriers from one grain to another is impeded by potential barriers surrounding the grain boundaries. Because the properties of moderately doped polysilicon are limited by grain boundaries, modifying the carrier traps at the grain boundaries by introducing hydrogen to saturate dangling bonds improves the conductivity.

When polysilicon is used as a gate electrode or an interconnection in integrated circuits, its limited conductivity can degrade circuit performance. At high dopant concentrations, the active carrier concentration is limited by solid solubility of the dopant species in crystalline silicon. The amount of active dopant is influenced by the time and temperature of the last fabrication step at which the dopant is mobile enough to interact with grain boundaries or approach its equilibrium solid-solubility value.

The minority-carrier lifetime in polysilicon is markedly lower than that in single-crystal silicon, but it increases with increasing grain size. High applied voltages can lead to temporary or permanent reduction in the resistance of a polysilicon structure.

Chapter 6

Applications

6.1 Introduction

In previous chapters we considered the properties of polysilicon that make it useful in integrated circuits. In this chapter, we want to apply that information to specific applications. Because of the wide range of integrated-circuit devices in which polysilicon is now used, this discussion cannot be all inclusive. A few of the more important applications are considered in detail, and others are briefly mentioned.

The dominant application of polysilicon is for gate electrodes in CMOS (*complementary metal-oxide-semiconductor*) integrated circuits, and that technology will be discussed in some detail. Related devices, such as nonvolatile, floating-gate, memory elements and high-value load resistors will also be examined. In addition to CMOS applications, polysilicon is used in most bipolar integrated circuits, where it both provides compact, low-capacitance structures and modifies the basic transistor physics. As use of polysilicon for bipolar integrated circuits became more widespread, the technologies used to fabricate MOS and bipolar integrated circuits have tended to converge, and many of the processes developed for one technology are applied to the other.

Polysilicon is also widely used in dynamic, random-access memories (DRAMs). In some DRAMs multiple layers of polysilicon are placed above the surface of the silicon substrate to form complex *stacked* capacitor cells; in other cases, a *trench* is formed in the substrate, coated with a dielectric, and then filled with polysilicon to form the storage capacitor. Polysilicon-filled trenches are also being increasingly used for

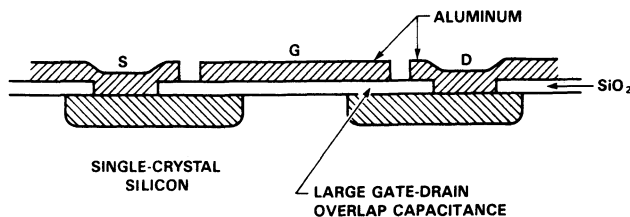


Figure 6.1: In the aluminum-gate MOS transistor, the gate must overlap the source and drain by one alignment tolerance to ensure that a continuous channel is formed when the transistor is turned on.

isolation between adjacent devices of high-performance circuits.

We will also examine structures in which the active portion of the device is placed *within* the polysilicon layer itself. These devices include diodes with both the *p*-type and *n*-type regions within the polysilicon and MOS transistors with their *channels* in polysilicon films. Thin-film transistors are being increasingly used in active-matrix displays. Novel use of polysilicon to form structural elements of *microelectromechanical systems* (MEMS) is seeing increasing attention. This application requires careful control of the mechanical properties of polysilicon.

Use of polysilicon in optical devices, such as solar cells, is not considered extensively here, primarily because of space limitations, but also because the polysilicon used for solar cells often differs significantly from that used in integrated circuits.

6.2 Silicon-gate MOS transistor

Development of polysilicon technology was motivated by the potential use of polysilicon as a gate electrode for MOS integrated circuits.¹ In the mid-1960s most gate electrodes were made from aluminum (Fig. 6.1), which was deposited after the source and drain regions were doped. The aluminum gate must overlap the source and drain regions by at least one alignment tolerance to insure that a continuous conducting channel is formed from the source to the drain when the gate is biased to turn on the transistor. The required alignment tolerance causes a significant

¹The term *Metal-oxide-semiconductor* (MOS) transistor is used to describe the device even when the gate is made from heavily doped polysilicon.

Miller feedback capacitance between the gate and the drain, reducing the circuit speed.²

By employing a gate-electrode material that allows the source and drain regions to be formed using the gate electrode as a mask (*ie*, to be *self-aligned*), this capacitance can be greatly reduced. Self-aligned gates not only reduce the capacitance, but they also make it more uniform from one device to another; this device matching is often critical for improved circuit performance. The material selected for a self-aligned electrode must be compatible with the high temperatures required to activate the source and drain dopant and with other high-temperature processes that follow gate-electrode formation.

Although other high-temperature materials, such as tungsten and molybdenum were investigated, the compatibility of polycrystalline silicon with integrated-circuit fabrication led to its rapid adoption. In the initial attempts to use thin films of silicon for gate electrodes, the silicon layers were deposited by evaporation. However, evaporated films could not adequately cover the steps between the gate oxide and the field oxide on the integrated-circuit surface. At the time silicon-gate technology was being developed, chemical vapor deposition (CVD) was routinely used to form the epitaxial silicon needed for bipolar integrated circuits. The epitaxial CVD technique was readily extended to polysilicon deposition by depositing on an oxide-coated substrate, instead of exposed single-crystal silicon, and by using a silicon source gas that did not contain chlorine. Because of the readily available technology and the conformal step coverage of films produced by chemical vapor deposition, this method of forming the silicon layers was quickly adopted [6.1] for the silicon-gate MOS transistor shown in Fig. 6.2. The ability of polysilicon to be oxidized and its compatibility with further high-temperature processing has allowed development of complex device structures employing several layers of polysilicon. In addition, polysilicon can be used to interconnect devices, allowing more flexible layout than obtained with only aluminum interconnections. Polysilicon can also be used for high-value resistors, further reducing chip area.

Initial silicon-gate integrated circuits used *p*-channel transistors because the turn-on voltage or *threshold voltage* was easier to control both in the active device and in the isolation regions. As more experience was

²The *Miller effect* increases the effective capacitance by multiplying the geometrical capacitance by the gain of the transistor.

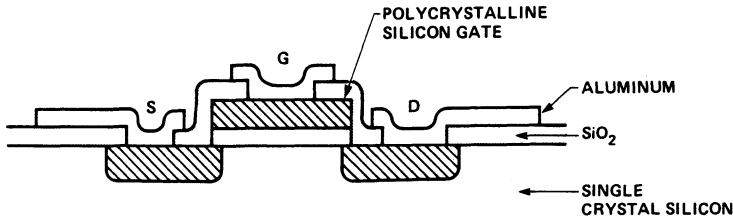


Figure 6.2: In a silicon-gate transistor, the gate electrode can be used to define the location of the source and drain dopant atoms, thereby eliminating the need for a large overlap of the gate electrode. This self-aligned structure reduces the overlap capacitance.

gained, n -channel circuits became more popular because of the higher mobility of electrons than holes. Having both n -channel and p -channel transistors available on the same chip allows building logic elements with very low dc current flowing when the circuit is not switching. Today, most integrated circuits are made using this complementary MOS (CMOS) technology.

6.2.1 Complementary MOS

In the basic CMOS circuit, an n -channel and a p -channel transistor are in series, with the same gate voltage applied to both. The source of the n -channel transistor is grounded; the source of the p -channel transistor is connected to the supply voltage (V_{DD}), as shown in Fig. 6.3. Under dc conditions either the n -channel transistor or the p -channel transistor should be normally off regardless of the gate voltage applied to the pair of transistors to limit the dc current flowing and, therefore, power dissipation.

To satisfy this requirement, the threshold voltage of the n -channel transistor should be positive, and that of the p -channel transistor should be negative. When the gate voltage is lower than the threshold voltage of the n -channel transistor, that transistor is nonconducting (OFF), and no dc current flows in the circuit even though the p -channel transistor is conducting (ON). When the gate voltage is higher than the threshold voltage of the n -channel transistor, that transistor conducts. However, the gate voltage is now close to the supply voltage so that the effective magnitude of the gate voltage on the p -channel transistor ($V_{DD} - V_G$) is small (less than the magnitude of its threshold voltage). The p -channel

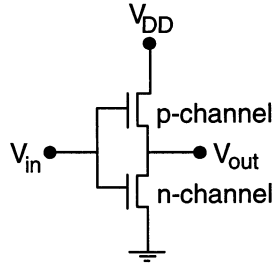


Figure 6.3: The n -channel and p -channel transistors of a CMOS circuit are in series so that one of the transistors is nonconducting when the circuit is not switching to limit the dc current.

transistor is nonconducting, and again negligible current flows in the circuit.

Appreciable current only flows when the transistor is switching. The reduced power dissipation of *complementary* MOS (CMOS) circuits is critical when integrated circuits contain millions or tens of millions of transistors. For maximum current flow when the transistor is switching the magnitude of the threshold voltage should be fairly low. The threshold voltage magnitude is made as low as practical, consistent with the requirement for low *leakage current* when the gate voltage is zero. It is typically ~ 0.5 V and decreases with more advanced processing. Unfortunately, as the threshold voltage decreases, the leakage current at zero gate voltage increases, so the leakage current increases as CMOS technology advances.

6.2.2 Threshold voltage

In addition to reducing the parasitic capacitance of an MOS transistor, using polysilicon provides a threshold voltage more compatible with the voltages used to bias integrated circuits. The threshold voltage V_T of an MOS transistor can be written [6.2]

$$V_T = V_{FB} \pm 2 |\phi_B| \pm |Q_d| / C_{ox} \quad (6.1)$$

where the upper signs refer to n -channel transistors and the lower ones refer to p -channel transistors. ϕ_B is the bulk potential [$\phi_B = (kT/q)(E_f - E_i)$, where E_f is the Fermi level and E_i is the intrinsic Fermi level]; Q_d is the charge in the depletion region separating the

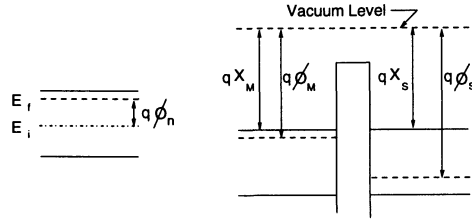


Figure 6.4: Definitions of terms used in transistor threshold voltage.

channel and the neutral bulk material; and C_{ox} is the capacitance per unit area of the gate dielectric.

The *flatband voltage* V_{FB} is the gate voltage required to remove the curvature from the energy bands in the silicon [ie, the gate voltage needed to compensate for the metal-semiconductor work function difference ϕ_{MS} between the gate material and the substrate material ($\phi_{MS} = \phi_M - \phi_S$) and the interface charge Q_f]:

$$V_{FB} = \phi_{MS} - Q_f/C_{ox} \quad (6.2)$$

The work function ϕ_S of a semiconductor depends on its doping concentration: $\phi_S = \chi_S + (E_c - E_f)/q$, where χ_S is the electron affinity of the semiconductor (ie, the energy between the vacuum level and the bottom of the conduction band), and E_c is the energy of the bottom of the conduction band.

For aluminum the work function is about 4.1 V. For silicon the electron affinity is about 4.05 V. To obtain maximum conductivity in the gate electrode, the polysilicon is usually doped as heavily as practical, so that the Fermi level is near the edge of the conduction or valence band for n -type or p -type polysilicon, respectively. The gate work function is then about 4.1 V for n -type polysilicon and about 5.2 V for p -type polysilicon.

The ability to have different work functions for the gate electrode of the n -channel and p -channel transistors can be useful in providing the desired threshold voltages for each type of transistor. For the n -channel transistor, a small positive threshold voltage (typically ~ 0.5 V) is desired, while the threshold voltage of the p -channel transistor should be similarly small in magnitude, but negative.

For an n -channel transistor with a p -type silicon substrate containing a dopant concentration of about 10^{16} cm^{-3} , the work function of the substrate is 4.9 V. With either aluminum or n -type polysilicon as the gate electrode, the gate work function is about 4.1 V, so ϕ_{MS} contributes about -0.8 V to the threshold voltage V_{Tn} of the n -channel transistor, partially compensating the other terms in Eq. 6.1 and reducing the threshold voltage to approximately the desired value.

For the p -channel transistor, the work function of the n -type substrate is about 4.3 V, only moderately different from the work function of either an aluminum or an n -type polysilicon gate. The magnitude of ϕ_{MS} is small and does not significantly compensate the other terms in Eq. 6.1. The threshold voltage V_{Tp} is more negative than desired, making the transistor difficult to turn on. To reduce the magnitude to the desired value, the surface must be *counterdoped* with p -type dopant, degrading the mobility and process control and causing other difficulties. If the gate of the p -channel transistor is made from heavily doped, p -type polysilicon, the gate work function is about 5.2 V. ϕ_{MS} is then significantly positive, compensating the other terms in the expression for the threshold voltage, and reducing its magnitude to the slightly negative value desired.

6.2.3 Silicon-gate process

Basic process: A basic silicon-gate MOS transistor can be built with four masking steps. However, building both n - and p -channel transistors on the same chip complicates the process and increases the number of masking steps needed. Further refinements needed for transistors with very small dimensions and for multiple layers of metal to interconnect them can increase the number of masking levels to well over twenty and add significant complexity to the fabrication process.

The four basic mask levels shown in Fig. 6.5 are

1. *Isolation* to separate the individual transistors. Usually LOCOS (LOCal Oxidation of Silicon) or trench isolation (Sec. 6.10).
2. *Polysilicon definition* to form the gate of each transistor and polysilicon interconnections.
3. *Contact openings* to the source, drain, and gate of each transistor.
4. *Metal definition* to form the interconnections between transistors.

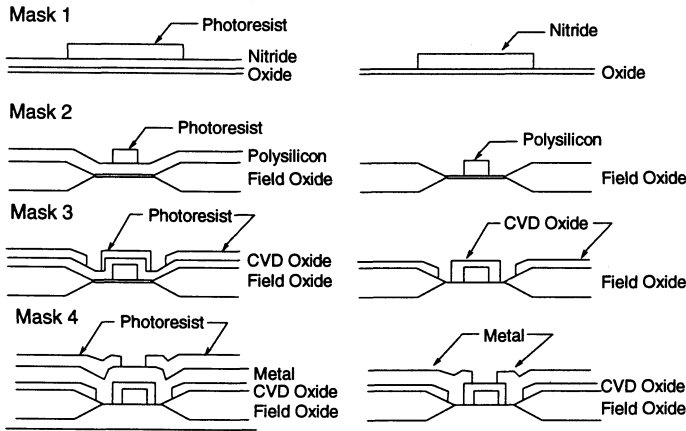


Figure 6.5: The four basic masks needed to form an MOS transistor: (1) Isolation. (2) Polysilicon. (3) Contact openings. (4) Metal.

For complementary transistors, selected regions of the substrate must be doped to form a *well* of the opposite conductivity type that acts as an effective substrate for the second type of transistor. Defining this well adds a masking step. To dope the source and drain regions of each type of transistor with the appropriate dopant element, one or two further masking steps must be added to block (or overcompensate) the other dopant species. Adjusting the threshold voltage of each type of transistor by selectively adding dopant to the channel adds one or two more masks. Another mask may be needed to form lightly doped regions in the source and drain adjacent to the channel to reduce the electric fields and improve the reliability (the *lightly doped drain* or LDD structure). Other masks include one to add extra dopant near the surface of the isolation (field) region surrounding the *n*-channel transistor to prevent inversion of this lightly doped, *p*-type region from shorting adjacent transistors together.

Finally, additional metal layers are needed to effectively interconnect transistors. Two masks are needed for each additional metal level: One to open *vias* through the dielectric separating two metal layers, and one to pattern the metal itself. With five or even more metal interconnection layers, the total number of masks increases significantly. For a complex CMOS process, the number of masks can exceed 20.

Conductivity type of gate electrode

We saw in Sec. 5.6 that the resistivity of n -type polysilicon doped with phosphorus can be lower than that of p -type polysilicon doped with boron. For maximum conductivity, the polysilicon for both n -channel and p -channel transistors can be doped with phosphorus, with the phosphorus added before the polysilicon is patterned. Phosphorus can be added to the polysilicon either during deposition or after deposition by diffusion from a gas-phase dopant source (usually POCl_3) or by ion implantation. Adding dopant during deposition is especially attractive if the polysilicon is to be covered with a layer of chemically vapor deposited silicide to lower the resistance of interconnections. The silicide can be deposited in another chamber of a cluster tool without exposing the surface of the polysilicon to air, eliminating native-oxide formation, as discussed in Sec. 1.5.2.

With n -type polysilicon and a p -type substrate, the threshold voltage of the n -channel transistor is somewhat positive, as desired. However, with n -type polysilicon and an n -type substrate, the threshold voltage of the p -channel transistor is more negative than desired for high conductivity when it is turned on. To obtain a threshold voltage closer to zero, the effective dopant concentration near the surface must be reduced by *counterdoping* with p -type boron. The heavy counterdoping needed lowers the carrier mobility and increases the effective distance of the controlled charge from the gate electrode.

On the other hand, as described above, the desired, slightly negative, threshold voltage can be readily obtained for the p -channel transistor by using p -type polysilicon as the gate electrode. The work function of the heavily doped, p -type polysilicon is higher than that of heavily doped n -type polysilicon by approximately the 1.1 eV bandgap of silicon, making the threshold voltage of p -channel transistors less negative. Thus, the threshold-voltage magnitude can be reduced without the otherwise-required, counterdoping of the channel region. Therefore, high-performance CMOS processes use n -type polysilicon for the gate of the n -channel transistor and p -type polysilicon for the gate of the p -channel transistor.

Complementary doped gates can be conveniently fabricated. The initially undoped polysilicon is patterned before doping, and the n -channel and p -channel transistors are doped separately. The same ion-implantation used to dope the source and drain regions also dopes the

gate electrode. This is especially beneficial for the p -channel transistor: When an n -type gate is used on a p -channel transistor, the p -type source-drain dopant counterdopes the n -type gate electrode, reducing its conductivity. Using p -type doping in the gate electrode avoids this conductivity reduction. However, the arsenic implantation that forms the source and drain of the n -channel transistor is also used to dope the gate, and the dopant concentration is usually lower than when a separate doping cycle is used. With the lower doping, the Fermi level may not be at the polysilicon conduction-band edge, changing the effective work function of the gate electrode. For reasonably heavy source-drain implant doses, however, the flat-band voltages obtained are comparable to those of devices in which the polysilicon is doped from the gas phase in a separate cycle [6.3].

6.2.4 Polysilicon interconnections

In metal-gate, MOS integrated circuits the signals can only propagate in diffused single-crystal silicon regions and in the metal. In silicon-gate circuits the polysilicon can also be used as an additional partial level of short interconnections. Placing the polysilicon interconnections over thick field oxide also reduces the interconnection capacitance. As integrated circuits become more complex, interconnecting the device elements becomes more critical, and the additional flexibility provided by polysilicon interconnections becomes more important. However, this additional level of interconnections is not totally flexible because the processing sequence of a typical silicon-gate integrated circuit prevents a heavily doped region of single-crystal silicon from crossing under the polysilicon line.

In addition, the resistance of polysilicon interconnections can be high. In Sec. 5.6.1 we discussed the limited conductivity of polysilicon resulting from the solid solubility of the dopant in silicon. Even with the maximum electrically active dopant concentration, the limiting resistivity of polysilicon is about $400\ \mu\Omega\text{-cm}$, much higher than the $3\ \mu\Omega\text{-cm}$ resistivity of aluminum.

When the polysilicon is doped by the source and drain implant, the amount of dopant introduced is limited by the requirement for a shallow source and drain in the single-crystal silicon. The limited amount of dopant restricts the conductance of the polysilicon interconnections. To increase the conductance, a silicide or other low-resistance conducting

material is usually placed over the polysilicon gates, both to lower the interconnection resistance and to short the p^+-n^+ diode in the polysilicon between the gates of the p -channel and the n -channel transistors. As discussed in Sec. 3.5.3, a *salicide* (self-aligned silicide process) can be used to form silicide simultaneously over the source, drain, and gate electrodes of an MOS transistor. An oxide or nitride spacer is first formed on the sidewall of the gate electrode to prevent shorting of the gate to the other electrodes as the silicide is formed. A thin layer of the metal to be used to form the silicide (usually titanium or cobalt) is deposited. The metal is then annealed at a moderate temperature so that it reacts with the exposed regions of single-crystal or polycrystalline silicon to form silicide. The unreacted metal over oxide or nitride regions is then selectively removed by wet chemical etching without removing the silicide. The silicide is then annealed at a higher temperature to lower its resistance. This method of forming silicide lowers the resistance of the source, drain, and gate electrodes considerably and is stable during further processing at moderate temperatures [6.4]. However, heavy n -type doping in the polysilicon retards silicide formation. Forming TiSi_2 also becomes more difficult as the width of the polysilicon lines decreases. Titanium silicide is most frequently used. However, because of the limitations of TiSi_2 , CoSi_2 has been investigated as an alternative [6.5].

Even when the conductivity of the polysilicon is augmented by a silicide, the conductance is much less than that of typical aluminum metallization. Because the resistance of a long polysilicon interconnection can limit high-performance circuits, polysilicon is primarily used for short interconnections.

6.2.5 Gate-oxide reliability

In addition to the advantages of polysilicon for design and fabrication, using a polysilicon gate electrode can improve long-term device reliability. During operation of an MOS integrated circuit, high fields are applied across the gate dielectric, leading to possible charge injection and trapping in the dielectric, which increases the internal electric field. When the internal field exceeds the dielectric breakdown field, catastrophic device failure occurs. The *time-to-breakdown* of the gate oxide is higher for structures which use polysilicon as the gate electrode than for those with aluminum gates [6.6, 6.7]. The increased time-to-breakdown

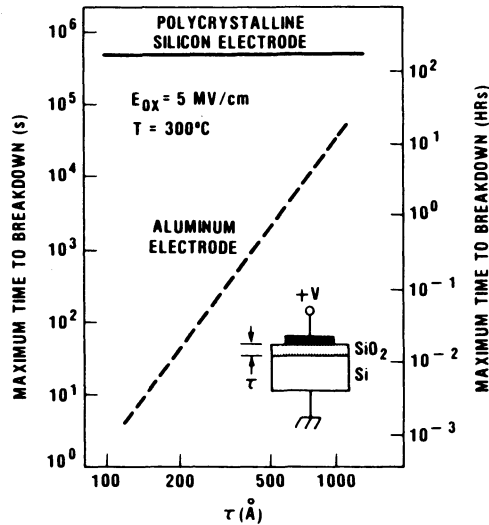
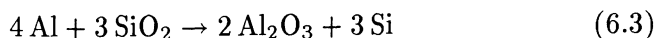


Figure 6.6: The reliability of polysilicon-gate MOS transistors is greater than that of aluminum-gate devices, and the difference increases as the gate-oxide thickness decreases. From [6.8]. ©1979 IEEE. Reprinted with permission.

is primarily observed with positive bias on the gate electrode; only a small improvement is seen with negative bias. For positive gate bias the *wear-out* time at 300°C can 3 to 4 orders of magnitude greater with a polysilicon gate electrode. The difference is larger for thinner gate oxides, as shown in Fig. 6.6 [6.8]. While the wear-out time is independent of oxide thickness with polysilicon electrodes, it decreases as the oxide thickness decreases with aluminum electrodes, making polysilicon more attractive in scaled devices. In addition, the activation energy for wear-out is greater with polysilicon (2.4 eV) than with aluminum (1.4 eV), so that the reliability is improved even more near room temperature.

The improved reliability is postulated to be related to the different interactions of polysilicon and aluminum with silicon dioxide [6.6]. Aluminum is thought to reduce SiO_2 by the reaction



to form stable Al_2O_3 . Alternate explanations involve the hydrogen introduced during polysilicon deposition or the formation of a thin phosphosilicate-glass layer during doping of the polysilicon.

6.2.6 Limitations

The silicon-gate MOS transistor has allowed unprecedented transistor density to be achieved during the three decades that it has been used as the gate electrode of MOS transistors. However, the limitations of polysilicon gate electrodes are becoming more serious as transistor dimensions decrease. Both n -channel and p -channel transistors become more difficult to build as the gate-oxide thickness decreases along with the lateral device dimensions, and the limited conductivity of polysilicon increasingly restricts routing signals long distances in polysilicon.

n -channel transistors: The n -channel transistor is limited by the difficulty of heavily doping the entire thickness of the polysilicon gate electrode. When a metal is used as the gate electrode of an MOS transistor, the high density of states near the Fermi level ensures that adequate carriers are always present adjacent to the gate dielectric to prevent carrier depletion and the associated voltage drop in the gate electrode. When n -type polysilicon is used, the carrier density in the polysilicon near the gate dielectric may be low enough that electrons can be pulled away from the bottom of the polysilicon adjacent to the gate dielectric at the positive biases used to turn on the transistor, forming a depletion region in the polysilicon. For a given gate voltage, the voltage drop across this depletion region reduces the voltage across the gate dielectric and the number of electrons induced in the channel of the transistor, degrading the maximum current that can flow in the channel of the transistor.

The depleted region of polysilicon effectively acts as an additional dielectric in series with the gate dielectric, increasing the gate voltage needed to turn on the transistor (the threshold voltage) and decreasing the drain current by decreasing the coupling of the gate voltage to the channel. The drain current I_D in the saturation region can be expressed by the simplified equation

$$I_D = k C_G (V_G - V_T)^2 \quad (6.4)$$

where k depends on the transistor layout and carrier mobility, C_G is the effective gate dielectric capacitance and V_T is the transistor threshold voltage. Both the increase of V_T and decrease of C_G degrade the saturation current. The depletion also increases the switching time of the circuit [6.9, 6.10] and increases the energy per cycle for the same gate

delay. Short-channel effects also increase because the bulk charge under the channel is reduced.

This carrier depletion is especially serious when the gate electrode of the n -channel transistor is doped with arsenic at the same time as the source and drain junctions are formed. The amount of arsenic that can be added to the polysilicon is limited by the need for a shallow source and drain, so the amount is less than optimum for doping the polysilicon. In addition, to form shallow source and drain junctions, the heat cycle after implantation is limited, restricting the diffusion of the dopant through the thickness of the polysilicon. Even though dopant diffusion in polysilicon is much greater than in single-crystal silicon (Sec. 3.3), high concentrations of arsenic may not have adequate time to diffuse through the entire thickness of the polysilicon to dope the bottom of the polysilicon heavily enough that it cannot be depleted by an applied gate voltage. Avoiding gate depletion requires higher implanted arsenic concentrations (consistent with the amount that can be added to the source and drain regions), maximizing the diffusion heat cycle (consistent with forming shallow source and drain junctions), and optimizing the structure of the polysilicon to allow rapid diffusion. This *polysilicon depletion* can be important for arsenic concentrations less than about $\text{mid-}10^{19} \text{ cm}^{-3}$, with the effective dopant concentration somewhat less than the total dopant concentration, as discussed in Sec. 5.6.1.

Carrier depletion is more serious for the thinner gate dielectrics used in advanced CMOS circuits. As the gate dielectric becomes thinner, its effective coupling capacitance to the channel increases, as desired. A given series capacitance associated with the depleted region of the gate electrode then causes a greater fractional decrease in the effective gate capacitance. With gate dielectrics only a few nanometers thick, even slight depletion at the bottom of the polysilicon seriously degrades the driving current capability of the transistor.

Because carrier depletion is an increasingly important limitation in n -channel transistors, alternate materials are being investigated for the gate electrode [6.11]. However, most metals are less compatible with silicon processing than is polysilicon. To reduce process incompatibility, in one proposed technique [6.12] the polysilicon gate is formed in the normal manner and retained during most of the device processing. After the high-temperature steps are completed, the polysilicon is removed and replaced by a metal.

***p*-channel transistors:** While carrier depletion can limit *n*-channel transistors, a different limitation seriously degrades *p*-channel transistors. Boron diffuses rapidly through polysilicon, and can possibly diffuse through the underlying gate dielectric to change the dopant concentration in the channel region of the transistor. Boron diffusion through the gate oxide occurs readily if hydrogen or fluorine is present. Avoiding a hydrogen ambient during the diffusion reduces the penetration. However, boron is often added to the source, drain, and gate regions of the *p*-channel transistor by implanting BF_2 (to obtain a shallow junction in the source and drain regions). The fluorine incorporated into the polysilicon near the gate oxide enhances boron diffusion through the gate oxide and into the channel of the transistor, changing its threshold voltage [6.13]. Using B^+ as the implant source eliminates the fluorine, but at the expense of a lower implant current and an implant voltage that is harder to control. Alternatively, nitrogen can be added to all or a portion of the thickness of the gate dielectric to reduce the boron diffusion even when fluorine is present [6.14–6.16]. Fluorine can also degrade reliability by promoting interface-state generation during operation [6.15].

Adding nitrogen to the polysilicon near the gate dielectric has also been suggested [6.17–6.19]. Although the nitrogen in the polysilicon can effectively block the diffusion-enhancing effect of fluorine, it also reduces the electrical activity of boron. The carrier concentration may then not be adequate to avoid carrier-depletion near the bottom of the polysilicon, leading to reduced coupling of the gate voltage to the channel of the transistor.

The structure of the polysilicon can also be modified to reduce boron penetration of the gate dielectric. The columnar structure, which allows rapid vertical diffusion, can be avoided (although at the expense of less effective doping of the gate electrode of the *n*-channel transistor). For example, depositing amorphous silicon and crystallizing it creates larger, non-columnar grains. A two-step polysilicon deposition process can also decrease the diffusion if the grain boundaries are not continuous through the thickness of the polysilicon.

6.2.7 Process compatibility

The compatibility of polysilicon with high-temperature processing allows complex fabrication sequences involving multiple layers of polysil-

icon deposition, along with implantation and annealing processes. The surface can also be smoothed to allow better coverage of steps by aluminum and improve fabrication yield. When density and shallow junctions do not limit the heat treatment, a silicon-dioxide layer containing phosphorus and possibly boron can be deposited over the polysilicon gate electrodes. It can be heated until its viscosity decreases and it “flows,” reducing the angle of the steps that the aluminum must cover. This phosphosilicate glass (PSG) or borophosphosilicate glass (BPSG) can be *flowed* before the contact windows are opened, or it can be *re-flowed* after the windows are opened to provide a more easily covered step at the edges of the contact windows.

The flexibility of polysilicon processing can be used in an additional way. As device dimensions decrease, the gate dielectric thickness must decrease also. Fragile gate dielectrics of 5–10 nm are often used in advanced processes. If additional processing is needed between the formation of the gate dielectric and the deposition of the polysilicon gate electrode (*eg*, masking and implantation to adjust the threshold voltage of one type of transistor), a thin, protective layer of polysilicon can be placed over the gate dielectric after its formation. After the intermediate processing, the remainder of the thickness of the polysilicon gate electrode is deposited.

6.2.8 New structures

The high-temperature capability of polysilicon also permits the development of more complex device structures that provide denser integrated circuits with the same minimum feature size. For example, the size of the one-transistor, dynamic, random-access-memory (RAM) cell to be discussed in Sec. 6.11 cell can be reduced significantly by using two layers of polysilicon. The first level is patterned and oxidized, and then the second is deposited so that the effective spacing between these electrodes is only the thickness of the oxide on the sidewall of the first level of polysilicon.

Independent, closely spaced polysilicon electrodes above single-crystal silicon are also used in charge-coupled devices. As in the dynamic RAM cell, different voltages are applied to adjacent electrodes, and no potential barrier can be tolerated in the substrate between these electrodes when they are biased to allow charge transfer. Two or more layers of polysilicon are typically used to form these electrodes. Although the

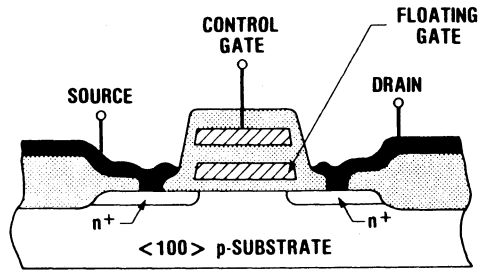


Figure 6.7: A nonvolatile memory element can be constructed by using a *floating* polysilicon gate between the silicon and a second *control* gate. Courtesy of J. Andrews

electrodes are laterally separated by the thickness of the oxide grown on the sidewall of the first layer of polysilicon, fringing fields reduce the barrier between the two electrodes when both are biased to induce inversion layers in the substrate.

6.3 Nonvolatile memories

Using two layers of polysilicon can be advantageous in many other devices also. Nonvolatile, electrically erasable, programmable, read-only memories (EEPROMs or E²PROMs) can be fabricated by using one level of polysilicon above another in the gate region of the device, as shown in Fig. 6.7. The first level of polysilicon is not electrically connected and “floats.” The conductance of the underlying channel between the source and drain regions depends on the charge stored on this *floating gate*. If it contains excess electrons, it repels electrons from the channel and turns off an *n*-channel transistor. A positive charge on the gate induces a channel and allows current to flow between source and drain. Substantial amounts of this type of memory are frequently included in specialized microprocessor chips to allow storage of code specific to a particular application and other information that must be infrequently modified or must remain when power is removed from the system. Similar devices are used in *flash memory*, although the devices are organized in the circuit differently.

Applying a voltage of the appropriate potential to the top *control gate* allows charge to be transferred through the gate oxide to charge the *floating gate* and write the cell. The charge can be generated by

avalanching the drain-substrate junction to create free carriers, which are then attracted to or repelled from the floating gate, depending on the potential of the control gate. The charge can be removed from the floating gate by tunneling through the oxide separating the two layers of polysilicon. When this method of erasing the cell is used, the surface texture of the lower layer of polysilicon is critical. We have already seen in Sec. 4.3 that conduction through oxide grown on polysilicon depends strongly on the surface roughness. Consequently, controlling the deposition conditions and the resulting structure of the polysilicon is probably more critical for this application than for any other. To reduce stringent control of the structure formed during deposition, the floating silicon layer can be deposited in the amorphous form and subsequently crystallized to provide a smooth surface. The surface can then be controllably roughed by oxidizing and removing the oxide before the tunnel oxide is grown [6.20].

Charge may also be removed from the floating gate by forcing it back through the underlying gate oxide into the substrate. The discharging characteristics depend on the structure of the polysilicon [6.21]. Erasing may be more effective along the phosphorus-doped oxide under the grain boundaries [6.21], making the grain size and grain-boundary density critical. Erasing a number of devices uniformly may depend on minimizing statistical variations between devices, favoring a fairly large number of grains within each device to minimize statistical variations. As the device size decreases, the grain size should also decrease to retain a certain number of grains (perhaps ~ 20 in each device); this need may limit scaling.

Using hemispherical-grain polysilicon (HSG) with its rough surface as the floating electrode increases the capacitive coupling to the control gate and improves device performance [6.22]. Separating the floating gate into small, individual nuclei, rather than having a continuous layer of polysilicon, may allow operation at lower voltages and increase endurance [6.23]. Each small island of silicon contains one or a very few electrons.

6.4 Polysilicon resistors

After silicon-gate technology was developed for MOS devices, polysilicon found other uses in integrated circuits. Because it is readily available,