

Tecnología de Fabricación de Microcircuitos: Aspectos Básicos

El presente libro describe las características fundamentales de los procesos tecnológicos que se utilizan en la fabricación de dispositivos semiconductores, especialmente los dispositivos de silicio. También analiza aspectos básicos de la física que le sirven de base a los diferentes procesos. Aspectos generales de fabricación, como características de las áreas limpias, pureza requerida en los diferentes materiales y tipos de equipamiento utilizado en los diferentes procesos tecnológicos, también se discuten. El contenido está orientado a estudiantes de las especialidades de Electrónica y Microelectrónica de las facultades de Física y de Ingeniería. Puede ser utilizado también como libro de consulta. El material recoge la experiencia en investigación y docencia de los autores en este campo, por mas de 40 años.

Tecnología de microcircuitos

Estrada , Escobosa

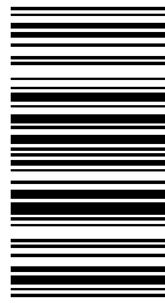
Magali Estrada

Físico, Ph.D., (1977); Profesor Titular de la Universidad de La Habana (1978-1994); Profesor Titular en el Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional México DF, (desde 1995). Desde 1968 trabaja en docencia e investigación en la microelectrónica.

Tecnología de Fabricación de Microcircuitos: Aspectos Básicos

Tecnología de microcircuitos

editorial académica española



978-3-659-00937-2

**Magali Estrada
Arturo Escobosa**

Tecnología de Fabricación de Microcircuitos: Aspectos Básicos

**Magali Estrada
Arturo Escobosa**

**Tecnología de Fabricación de
Microcircuitos: Aspectos Básicos**

Tecnología de microcircuitos

Editorial Académica Española

Impresión

Información bibliográfica publicada por Deutsche Nationalbibliothek: La Deutsche Nationalbibliothek enumera esa publicación en Deutsche Nationalbibliografie; datos bibliográficos detallados están disponibles en internet en <http://dnb.d-nb.de>.

Los demás nombres de marcas y nombres de productos mencionados en este libro están sujetos a la marca registrada o la protección de patentes y son marcas comerciales o marcas comerciales registradas de sus respectivos propietarios. El uso de nombres de marcas, nombre de producto, nombres comunes, nombre comerciales, descripciones de productos, etc. incluso sin una marca particular en estas publicaciones, de ninguna manera debe interpretarse en el sentido de que estos nombres pueden ser considerados ilimitados en materias de marcas y legislación de protección de marcas y, por lo tanto, ser utilizadas por cualquier persona.

Imagen de portada: www.ingimage.com

Editor: Editorial Académica Española es una marca de
LAP LAMBERT Academic Publishing GmbH & Co. KG
Heinrich-Böcking-Str. 6-8, 66121 Saarbrücken, Alemania
Teléfono +49 681 3720-310, Fax +49 681 3720-3109
Correo Electronico: info@eae-publishing.com

Publicado en Alemania

Schaltungsdienst Lange o.H.G., Berlin, Books on Demand GmbH, Norderstedt,
Reha GmbH, Saarbrücken, Amazon Distribution GmbH, Leipzig
ISBN: 978-3-659-00937-2

Imprint (only for USA, GB)

Bibliographic information published by the Deutsche Nationalbibliothek: The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

Any brand names and product names mentioned in this book are subject to trademark, brand or patent protection and are trademarks or registered trademarks of their respective holders. The use of brand names, product names, common names, trade names, product descriptions etc. even without a particular marking in this works is in no way to be construed to mean that such names may be regarded as unrestricted in respect of trademark and brand protection legislation and could thus be used by anyone.

Cover image: www.ingimage.com

Publisher: Editorial Académica Española is an imprint of the publishing house
LAP LAMBERT Academic Publishing GmbH & Co. KG
Heinrich-Böcking-Str. 6-8, 66121 Saarbrücken, Germany
Phone +49 681 3720-310, Fax +49 681 3720-3109
Email: info@eae-publishing.com

Printed in the U.S.A.

Printed in the U.K. by (see last page)

ISBN: 978-3-659-00937-2

Copyright © 2012 by the author and LAP LAMBERT Academic Publishing GmbH & Co. KG
and licensors
All rights reserved. Saarbrücken 2012

***PROCESOS TECNOLÓGICOS DE
FABRICACIÓN DE
MICROCIRCUITOS: ASPECTOS
BÁSICOS.***

Dra. Magali Estrada del Cueto

Dr. Arturo Escobosa Echavarría

Sección de Electrónica del Estado Sólido

Departamento de Ingeniería Eléctrica

CINVESTAV-IPN

Prólogo

El presente libro está concebido como libro de texto para dos cursos de Tecnología que se imparten en los programas de maestría y doctorado en la Sección de Electrónica del Estado Sólido, (SEES), del Departamento de Ingeniería Eléctrica del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, (CINVESTAV-IPN), en México.

Consta de dos partes, la primera corresponde al curso de Tecnología 1 y la segunda al curso de Tecnología 2.

El contenido está orientado a estudiantes de nivel posgraduado en la especialidad de Electrónica y Microelectrónica, tanto en las facultades de Física como de Ingeniería, aunque puede ser utilizado también como libro de consulta. Recoge las características fundamentales de los procesos tecnológicos que hacen posible la fabricación de dispositivos semiconductores, llamada Tecnología Planar.

Los temas que aborda están enfocados desde un punto de vista básico, que permita posteriormente profundizar en aquellos aspectos que sean de interés del lector. Parte del contenido fue tomado de un libro anterior de uno de los autores, Circuitos Integrados: Fundamentos Tecnológicos, por los autores M. Estrada, A. Cerdeira y R. Martínez, Editorial Pueblo y Educación, 1987, La Habana, el cual fue actualizado y complementado con temas que no estaban incluidos anteriormente.

La parte I fue escrita por el Dr. Arturo Escobosa Echavarría y la parte II por la Dra. Magali Estrada del Cueto, ambos Profesores Titulares de la SEES del CINVESTAV y los materiales recogen la experiencia tanto de investigación como docente de cada uno de ellos.

Los autores expresan especial agradecimiento a los alumnos de maestría de la SEES, que entre 2000 y 2002 fueron revisando cuidadosamente las notas de clases que luego permitieron la escritura de este texto y muy especialmente al M.C. Hector Báez Medina, quien de manera voluntaria revisó muy cuidadosamente la versión final del texto, cuando aun era alumno de maestría .

CAPÍTULO 1 INTRODUCCIÓN A LA TERMODINÁMICA	2
1.1 TERMODINÁMICA	2
1.2 DEFINICIONES BÁSICAS	2
1.3 COMPORTAMIENTO DE LOS GASES	4
1.3.1 Ley de Boyle-Mariotte	4
1.3.2 Ley de Charles	5
1.3.3 Ecuación general de los gases ideales	6
1.3.4 Ley De Dalton	7
1.4 PRIMERA LEY DE LA TERMODINAMICA	8
1.4.1 Conceptos básicos	8
1.4.2 Energía interna de un gas ideal	12
1.4.3 Termoquímica	15
1.5 SEGUNDA Y TERCERA LEYES DE LA TERMODINAMICA	18
1.5.1 Segunda ley de la termodinámica	18
1.6 TERCERA LEY DE LA TERMODINÁMICA	25
1.7 CRITERIOS DE EQUILIBRIO QUÍMICO	26
1.8 . LEY DE ACCIÓN DE MASAS	32
1.9 GRÁFICAS DE ARHENIUS	34
1.10 BIBLIOGRAFÍA	38
CAPÍTULO 2 DIAGRAMAS DE FASE	40
2.1 REGLA DE FASES	41
2.2 SISTEMAS UNITARIOS.	42
2.2.1 Diagrama del agua	42
2.3 DIAGRAMAS BINARIOS	43
2.3.1 Regla de la palanca	45
2.3.2 Diagramas isomorfos	47
2.3.3 Sistemas Eutécticos	49
2.3.4 Transformaciones congruentes	54
2.3.5 Reacciones peritécticas	57

2.3.6 Otros ejemplos	57
2.3.7 Diagramas de sistemas de óxidos	60
2.3.8 Solubilidad Sólida	61
2.4 DIAGRAMAS TERNARIOS	62
2.5 BIBLIOGRAFÍA	67
CAPÍTULO 3 CRECIMIENTO DE CRISTALES, (LINGOTES)	68
3.1 INTRODUCCIÓN	68
3.2 SÍNTESIS DE SILICIO	69
3.3 SÍNTESIS DE GaAs	72
3.4 MÉTODO BRIDGMAN	74
3.5 MÉTODO CZOCHRALSKI	77
3.5.1 Encapsulado líquido	80
3.5.2 Síntesis Directa	80
3.5.3 Consideraciones Generales	81
3.5.4 Impurificación de los cristales	82
3.5.5 Superenfriamiento Constitucional	88
3.6 CRISTALIZACIÓN Y PURIFICACIÓN POR ZONA FLOTANTE	89
3.7 IMPURIFICACIÓN POR NEUTRONES	91
3.8 BIBLIOGRAFÍA	92
CAPÍTULO 4 DIFUSIÓN	94
4.1 DIFUSIÓN EN CRISTALES	95
4.1.1 Movimiento Intersticial	95
4.1.2 Movimiento substitucional	97
4.1.3 Movimiento Intersiticial-Substitucional	98
4.2 DIFUSIÓN EN UN GRADIENTE DE CONCENTRACIÓN	99
4.2.1 El Coeficiente de difusión	101
4.2.2 Coeficiente de difusión considerando vacancias cargadas	106
4.2.3 La Ecuación de difusión	109
4.2.4 Difusión dependiente de la concentración	115

4.2.5 Difusión Intersticial-Substitucional	116
4.3 SISTEMAS DE DIFUSIÓN	117
4.4 IMPUREZAS EN SILICIO	122
4.4.1 Difusión de boro.	125
4.4.2 Difusión de fósforo.	128
4.4.3 Impurezas Intersiticial-Substitucional	131
4.5 DIFUSIÓN EN GAAS	132
4.6 INTRODUCCIÓN DE IMPUREZAS A TRAVÉS DE ÓXIDO DE SILICIO	133
4.7 DIFUSIÓN HACIA AFUERA O EXODIFUSIÓN	136
4.8 REDISTRIBUCIÓN DE IMPUREZAS	137
4.8.1 Efectos de interacción	139
4.8.2 Difusión bidimensional	140
4.9 BIBLIOGRAFIA	141
CAPÍTULO 5 EPITAXIA	142
5.1 LOS PROCESOS EPITAXIALES DE ACUERDO AL MEDIO DE TRANSPORTE.	144
5.2 EPITAXIA POR HACES MOLECULARES.	145
5.3 EPITAXIA EN FASE DE VAPOR (VPE)	147
5.3.1 Transporte y Reacción.	148
5.3.2 Nucleación	151
5.3.3 Diseño de sistemas.	155
5.3.4 Procesos epitaxiales de silicio	158
5.3.5 Epitaxia de Arseniuro de Galio.	159
5.4 EPITAXIA EN FASE LÍQUIDA.	163
CAPÍTULO 6 IMPLANTACIÓN IÓNICA	168
6.1 INTRODUCCIÓN	168
6.2 SISTEMAS DE IMPLANTACIÓN	169
6.2.1 Fuente de iones	170

6.2.2 Purificador del haz	171
6.2.3 Aceleración	173
6.2.4 Blanco	174
6.2.5 Deflexión del haz	175
6.2.6 Control de dosis	175
6.3 CARACTERÍSTICAS DE PENETRACIÓN DE LOS IONES EN EL BLANCO	176
6.3.1 Dispersión	176
6.3.2 Frenado electrónico	177
6.3.3 Frenado nuclear	177
6.4 IMPLANTACIÓN A TRAVÉS DE MULTICAPAS	181
6.5 CANALES DE PROPAGACIÓN	184
6.6 DAÑO DE IMPLANTACIÓN	185
6.7 BIBLIOGRAFÍA	186
CAPÍTULO 1_INTRODUCCIÓN A LA TECNOLOGÍA PLANAR	190
1.1 CARACTERÍSTICAS DE LA TECNOLOGÍA PLANAR	190
CAPÍTULO 2_CARACTERÍSTICAS DEL SIO ₂ POR OXIDACIÓN TÉRMICA DEL SILICIO	196
2.1 CINÉTICA DE CRECIMIENTO	197
2.1.1 MODELO DE LA CAPA LIMÍTROFE O FRONTERIZA	198
2.1.2 CÁLCULO DEL FLUJO DE ESPECIE ACTIVA DURANTE LA OXIDACIÓN	204
2.1.3 VELOCIDAD DE CRECIMIENTO	207
2.2 DATOS EXPERIMENTALES QUE CONFIRMAN EL MODELO DE OXIDACIÓN DE GROVE	211
2.3 ESTRUCTURA DEL ÓXIDO DE SILICIO Y CARACTERÍSTICAS DE LA FRONTERA SIO ₂ -SI	216
2.3.1 ESTRUCTURA DEL ÓXIDO DE SILICIO	216

2.3.2 CARACTERÍSTICAS DE LA INTERFAZ SIO ₂ -SI: ESTADOS SUPERFICIALES, CARGA FIJA Y CARGA MÓVIL, SU ORIGEN Y NATURALEZA FÍSICA	218
2.3.3 INFLUENCIA DE LAS CARGAS EN LA INTERFAZ Y EN EL DIELÉCTRICO SOBRE LOS PARÁMETROS DE LOS DISPOSITIVOS MOS	222
2.4 FORMAS DE DISMINUIR LA DENSIDAD DE CARGAS EN LA INTERFAZ SIO ₂ -SI Y EN EL DIELÉCTRICO	226
2.5 MEDIDAS DE CONTROL DE LA CONTAMINACIÓN DURANTE EL PROCESAMIENTO DE LA OBLEA	230
2.5.1 TECNOLOGÍA LIMPIA	230
2.5.2 PROCESO DE LIMPIEZA DE OBLEAS	235
2.6 BIBLIOGRAFÍA	236
CAPÍTULO 3 CURVAS CAPACITANCIA-VOLTAJE (C-V) DE ESTRUCTURAS MDS: FUNDAMENTO TEÓRICO Y APLICACIONES FUNDAMENTALES	238
3.1 FUNDAMENTO TEÓRICO DE LAS CURVAS C-V	238
3.2 APLICACIÓN DE LAS CURVAS C-V DE ALTA FRECUENCIA A LA DETERMINACIÓN DE LAS PROPIEDADES DE LA INTERFAZ SIO ₂ -SI	247
3.2.1 DETERMINACIÓN DEL VOLTAJE DE BANDA PLANA	250
3.2.2 DETERMINACIÓN DE LA MAGNITUD DE LA CARGA MÓVIL	254
3.3 DETERMINACIÓN DE LA CONCENTRACIÓN DE IMPUREZAS EN UN SEMICONDUCTOR	256
3.4 TÉCNICAS EXPERIMENTALES DE MEDICIÓN	258
3.4.1 TÉCNICA DE MEDICIÓN DE CURVAS C-V DE ALTA FRECUENCIA	258

3.4.2 TÉCNICAS DE MEDICIÓN DE CURVAS DE BAJA FRECUENCIA	261
3.5 BIBLIOGRAFÍA	264
CAPÍTULO 4 DEPÓSITO DE CAPAS DIELÉCTRICAS Y SEMICONDUCTORAS: CARACTERÍSTICAS Y MÉTODOS DE DEPÓSITO.	266
4.1 CAPAS DIELÉCTRICAS UTILIZADAS EN LA TECNOLOGÍA PLANAR	266
4.1.1 CAPAS DE ÓXIDO DE SILICIO	267
4.1.2 CAPAS DE VIDRIO DE FOSFOSILICATO (VFS)	269
4.1.3 CAPAS DE VIDRIO DE BOROFOSFOSILICATO	273
4.1.4 CAPAS DE NITRURO DE SILICIO (Si_3N_4)	274
4.2 DEPÓSITO QUÍMICO A PARTIR DE LA FASE, DE VAPOR (CVD)	277
4.2.1 CINÉTICA DE CRECIMIENTO	277
4.3 DEPÓSITO DE SiO_2	281
4.3.1 DEPÓSITO DE SiO_2 POR APCVD	281
4.3.2 DEPÓSITO DE SiO_2 POR LPCVD	283
4.4 DEPÓSITO DE CAPAS DE VIDRIO DE FOSFOSILICATO (VFS)	284
4.5 DEPÓSITO DE CAPAS DE VIDRIO DE BOROFOSFOSILICATO (VBFS)	289
4.5.1 DEPÓSITO DE CAPAS DE VIDRIO DE BOROFOSFOSILICATO POR CVD	289
4.5.2 DEPÓSITO DE CAPAS DE VIDRIO DE BOROFOSFOSILICATO POR LPCVD	290
4.5.3 DEPÓSITO DE CAPAS DE VIDRIO DE BOROFOSFOSILICATO POR PECVD	290
4.6 DEPÓSITO DE CAPAS DE NITRURO DE SILICIO (Si_3N_4)	291
4.6.1 DEPÓSITO DE NITRURO DE SILICIO POR LPCVD	291

4.6.2 DEPÓSITO DE NITRURO DE SILICIO POR CVD CON AYUDA DE PLASMA (PECVD).	293
4.6.1 DEPÓSITO DE NITRURO DE SILICIO POR PULVERIZACIÓN CATÓDICA	294
4.7 CAPAS DE POLISILICIO. CARACTERÍSTICAS Y APLICACIONES	295
4.7.1 DEPÓSITO DE POLISILICIO POR LPCVD	300
4.8 CARACTERÍSTICAS DEL EQUIPAMIENTO UTILIZADO EN EL DEPÓSITO QUÍMICO A PARTIR DE LA FASE VAPOR (CVD)	301
4.9 DESCARGA ELÉCTRICA AUTOMANTENIDA: REGIÓN DEL PLASMA	309
4.9.1 FENÓMENOS FÍSICOS ASOCIADOS A LA INTRODUCCIÓN DE UN ELECTRODO EN UNA DESCARGA GASEOSA DE CD.	313
4.9.2 FENÓMENOS ASOCIADOS A UNA DESCARGA PRODUCIDA POR UN CAMPO ALTERNO O PLASMA DE R.F.	319
4.10 CARACTERÍSTICAS DEL EQUIPAMIENTO UTILIZADO EN EL DEPÓSITO DE CAPAS POR PULVERIZACIÓN CATÓDICA	326
4.11 BIBLIOGRAFÍA	327
CAPÍTULO 5. ATAQUE EN SECO DE CAPAS DIELÉCTRICAS, SEMICONDUCTORAS Y METÁLICAS: MÉTODOS DE ATAQUE Y CARACTERÍSTICAS DE CADA UNO	330
5.1 MÉTODOS DE ATAQUE EN SECO Y PRINCIPIOS FÍSICOS EN LOS QUE SE BASAN	332
5.2 EQUIPOS MÁS UTILIZADOS EN LOS PROCESOS DE ATAQUE EN SECO	334
5.2.1 EQUIPOS DEL TIPO DE BARRIL Y DE PLACAS PARALELAS	334
5.2.2 CARACTERÍSTICAS COMUNES A LOS EQUIPOS DE ATAQUE EN SECO	341
5.3 RESISTENCIA DE LA FOTORRESINA AL ATAQUE EN SECO	342

5.4 ATAQUES ISÓTROPOS Y ANISÓTROPOS; PRINCIPALES CARACTERÍSTICAS Y APLICACIONES	343
5.5 SELECTIVIDAD Y CONTROL DE LAS DIMENSIONES	348
5.5.1 ASPECTOS A CONSIDERAR EN EL ATAQUE EN SECO DE UNA CAPA CUBIERTA CON FOTORRESINA	349
5.5.2 SELECTIVIDAD REQUERIDA RESPECTO A UNA CAPA SUBYACENTE A LA QUE SE ESTÁ ATACANDO.	352
5.6 BIBLIOGRAFÍA	356
CAPÍTULO 6_MEDICIÓN DE RESISTIVIDAD	358
6.1 BIBLIOGRAFÍA	369
CAPÍTULO 7_NUEVOS DIELÉCTRICOS PARA DISPOSITIVOS MOS	370
7.1 LIMITACIONES DEL SIO ₂ CON EL ESCALAMIENTO DE LOS DISPOSITIVOS MOS	370
7.1.1 FENÓMENOS FÍSICOS QUE LIMITAN EL USO DEL SIO ₂ CON ESPESOR MENOR DE 3 NM	373
7.1.2 NUEVOS DIELÉCTRICOS SUBSTITUTOS DEL SIO ₂	378
7.2 OTROS PROBLEMAS ASOCIADOS A LOS ÓXIDOS DE COMPUERTA ULTRAFAINOS	381
7.2.1 PENETRACIÓN DE IMPUREZAS A TRAVÉS DEL DIELÉCTRICO	381
7.2.2 EMPEORAMIENTO DE LAS CARACTERÍSTICAS DE LA INTERFAZ	382
7.2.3 INCREMENTO DEL VALOR DEL VOLTAJE DE UMBRAL POR EFECTOS CUÁNTICOS	383
7.3 METODOS DE OBTENCIÓN DE DIELÉCTRICOS DE ALTA K	385
7.4 NUEVOS DIELÉCTRICOS PARA AISLAMIENTO ENTRE LÍNEAS DE METALIZACIÓN	387

7.4.1 CARACTERÍSTICAS DE LA METALIZACIÓN	
MULTINIVELES	387
7.5 BIBLIOGRAFÍA	394
CAPÍTULO 8 _EL PROCESAMIENTO TÉRMICO RÁPIDO: SU APLICACIÓN EN LA MICROELECTRÓNICA	398
8.1 INTRODUCCIÓN	398
8.2 CARACTERÍSTICAS DE LOS EQUIPOS DE RTP	399
8.3 OXIDACIÓN TÉRMICA RÁPIDA (RTO)	403
8.4 TRATAMIENTOS TÉRMICOS RÁPIDOS (RTA)	404
8.5 CRECIMIENTO DE CAPAS POR RTP (RTCVD)	405
8.6 BIBLIOGRAFÍA	407
CAPÍTULO 9 _PROCESOS FOTOLITOGRÁFICOS APLICADOS A LA MICROELECTRÓNICA	410
9.1 INTRODUCCIÓN	410
9.2 FOTOLITOGRAFÍA ÓPTICA	412
9.2.1 TIPOS DE MÁSCARAS	412
9.2.2 MÉTODOS DE FABRICACIÓN DE MÁSCARAS DE PRIMERA Y SEGUNDA REDUCCIÓN	414
9.2.3 MÉTODO DE FABRICACIÓN DE RETÍCULOS Y MÁSCARA PATRÓN A TRAVÉS DE GENERADOR DE PATRONES	415
9.3 FOTORRESINAS	418
9.3.1 TIPOS DE FOTORRESINAS	418
9.4 TIPOS DE ALINEADORAS PARA FOTOLITOGRAFÍA ÓPTICA	419
9.4.1 ALINEADORAS DE CONTACTO	420
9.4.2 ALINEADORAS DE PROXIMIDAD	422
9.4.3 ALINEADORAS DE PROYECCIÓN	423
9.4.4 SITUACION ACTUAL Y TENDENCIAS.	424
9.5 FOTOLITOGRAFIA CON HAZ DE ELECTRONES	429

9.5.1 GENERADOR DE PATRONES UTILIZANDO HAZ DE ELECTRÓNOS	430
9.5.2 FOTORRESINA ELECTRÓNICA	431
9.5.2.1 FOTORRESINAS ELECTRÓNICAS POSITIVAS	432
9.5.2.2 FOTORRESINAS ELECTRÓNICAS NEGATIVAS	432
9.5.3 LA DISPERSIÓN Y EL EFECTO DE PROXIMIDAD	433
9.5.4 APLICACIONES	434
9.5.4.1 PREPARACIÓN DE RETÍCULOS Y MÁSCARAS POR EXPOSICIÓN DIRECTA	434
9.5.4.2 ESCRITURA DIRECTA	436
9.6 FOTOLITOGRAFIA DE RAYOS X	437
9.6.1 CARACTERÍSTICAS GENERALES	437
9.6.2 FOTORRESINA PARA RAYOS X	438
9.6.3 MÁSCARAS	439
9.7 MÉTODOS NO CONVENCIONALES DE FOTOLITOGRAFÍA	439
9.8 BIBLIOGRAFÍA	441
CAPÍTULO 10 ANEXOS	442
10.1 TABLA A-1	442
10.2 TABLA A-2	444
10.3 TABLA A-3	445
10.4 TABLA A-4	446
10.5 TABLA A-5	448
10.6 TABLA A-6	449
10.7 TABLA A-7	450
10.8 TABLA A-8	451

PARTE I

CAPÍTULO 1

INTRODUCCIÓN A LA TERMODINÁMICA

1.1 TERMODINÁMICA

La termodinámica trata con las propiedades de los sistemas en equilibrio. Es independiente de la cinética de los procesos y sólo le interesan el estado inicial y el estado final del sistema en cuestión. La termodinámica sólo se ocupa de sistemas de materia macroscópicos, o sea, aquellos que contienen un alto número de átomos.

1.2 DEFINICIONES BÁSICAS

Se llama **sistema** a la región del universo en consideración.

Un **sistema macroscópico** se puede describir por medio de criterios sensoriales como presión, temperatura, volumen o masa. En un sistema macroscópico no se conocen la estructura de la materia o como ocurren las reacciones.

Sistema cerrado es aquél que no permite la entrada o salida de materia. En caso contrario se tiene un **sistema abierto**. Ambos sistemas pueden intercambiar trabajo y calor con su medio ambiente.

La materia puede presentarse en tres estados.

Estado gaseoso: En este estado, la materia toma el volumen y la forma del espacio que la limita.

Estado líquido: La materia que se encuentra en este estado toma la forma del recipiente en que se encuentra, pero a diferencia de la materia en estado gaseoso, su volumen está limitado por una superficie.

Estado sólido: La materia tiene forma y volumen definidos.

Un **fluido** puede ser tanto un gas como un líquido.

Si dos sistemas cerrados se ponen en contacto, cambiarán las propiedades de ambos hasta que alcance un equilibrio. **Equilibrio** es entonces, el estado de un sistema, más allá del cual no se presentan cambios.

Ley cero de la termodinámica

Si un sistema A está en equilibrio con un sistema B y A también guarda equilibrio con C entonces B está en equilibrio con C.

Dos sistemas cerrados estarán en equilibrio si tienen la misma temperatura, independientemente de los parámetros internos de cada sistema.

Ecuación de estado. El estado de un sistema depende de su volumen, temperatura y presión, si se desprecian efectos de gravitación y campos eléctricos y magnéticos externos. De acuerdo con la experiencia, cuando se definen dos de estas variables, automáticamente estará dada la tercera. Así la presión será una función de la temperatura y el volumen:

$$p = f(V, T) \quad (1.1)$$

donde f recibe el nombre de ecuación de estado.

Esta regla tiene excepciones cuando para un estado dado se pueden tener dos valores que satisfacen la ecuación. Un ejemplo de esto es el caso del agua cerca de 4 °C, donde se tiene un mismo valor de volumen para dos valores de temperatura diferente.

Cantidades termodinámicas extensivas: son aquellas que dependen del tamaño del sistema, como por ejemplo, el volumen.

Cantidades termodinámicas intensivas: son las que no dependen del tamaño del sistema como la temperatura y la presión.

1.3 COMPORTAMIENTO DE LOS GASES

Se pueden definir varias escalas de temperatura, pero la mejor es aquella basada en el comportamiento de los gases ideales.

1.3.1 Ley de Boyle-Mariotte

A temperatura constante, el volumen de una cierta cantidad de gas decrece cuando la presión es incrementada, esto es, el volumen es inversamente proporcional a la presión aplicada, Fig. 1.1.

$$V = \frac{K_1}{p} \quad (1.2)$$

donde K_1 es una constante.

Entonces, en un sistema a temperatura constante donde se realiza un cambio de estado se cumplirá que:

$$p_i V_i = p_f V_f, \quad (1.3)$$

donde i (indica el valor de las variables en su estado inicial) y f (indica el estado final).

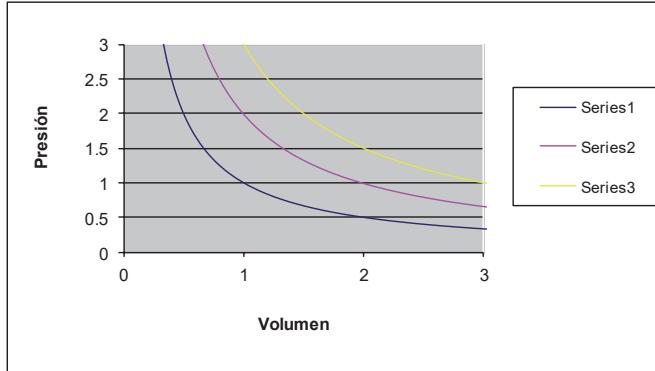


Fig. 1.1 Ley de Boyle-Mariotte, para tres isotermas.

1.3.2 Ley de Charles

Si la presión es constante, el aumento en el volumen de un gas será proporcional al aumento de la temperatura, Fig. 1.2. Es importante seleccionar una escala de temperatura absoluta al aplicar esta ley, que se expresa como.

$$V = K_2 T \quad (1.4)$$

donde K_2 es una constante.

Entonces, en un sistema a presión constante donde se realiza un cambio de estado, se cumplirá que:

$$\frac{V_i}{T_i} = \frac{V_f}{T_f} \quad (1.5)$$

donde el volumen del gas será cero al cero absoluto de temperatura.

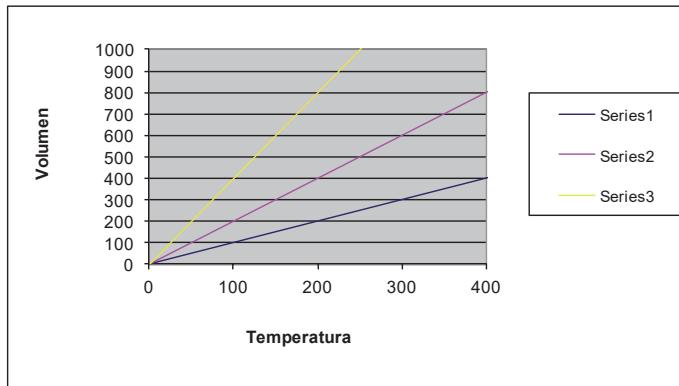


Fig.1.2 Ley de Charles, para tres isobaras

1.3.3 Ecuación general de los gases ideales

Puesto que para un gas ideal el volumen es una función de la presión y la temperatura absoluta, ec. (1.1), el incremento en volumen se puede expresar como:

$$dV = \left(\frac{\partial V}{\partial p} \right)_T dp + \left(\frac{\partial V}{\partial T} \right)_p dT \quad (1.6)$$

De las ecuaciones (1.2) y (1.4):

$$dV = \frac{K_1}{p^2} dp + K_2 dT \quad (1.7)$$

$$\frac{dV}{V} + \frac{dp}{p} = \frac{dT}{T} \quad (1.8)$$

$$\ln p + \ln V = \ln T + \ln(cte) \quad (1.9)$$

$$PV = (cte)T \quad (1.10)$$

y entonces:

$$\frac{P_i V_i}{T_i} = \frac{P_f V_f}{T_f} \quad (1.11)$$

Conociendo la cantidad de materia, expresada en moles (n) podemos escribir la ecuación general de los gases ideales:

$$pV = nRT \quad (1.12)$$

donde R es la constante de los gases ideales o constante de Rydberg, y se ha determinado como:

$$R=8.317 \text{ Joul/K mol}$$

$$R=1.987 \text{ cal/mol}$$

1.3.4 Ley De Dalton

Si tenemos una mezcla de gases ideales con un número total de moles n_T :

$$n_T = n_A + n_B + n_C + \dots \quad (1.13)$$

la presión estará dada por:

$$p_T = n_T \frac{RT}{V} = (n_A + n_B + n_C + \dots) \frac{RT}{V} = n_A \frac{RT}{V} + n_B \frac{RT}{V} + n_C \frac{RT}{V} + \dots \quad (1.14)$$

o sea:

$$p_T = p_A + p_B + p_C + \dots \quad (1.15)$$

esto es: "la presión total de una mezcla de gases ideales es igual a la suma de las presiones parciales de los mismos". Esta es la **Ley de Dalton** y es válida para gases reales que no interactúen entre sí.

1.4 PRIMERA LEY DE LA TERMODINAMICA

La primera ley de la termodinámica se ocupa de las relaciones de los diferentes tipos de energía como calor y trabajo.

1.4.1 Conceptos básicos

Proceso adiabático es aquel que ocurre en un sistema cerrado, sin intercambio de calor.

Una **frontera adiabática** será aquella que no permite la transmisión de calor, sin embargo puede permitir el intercambio de trabajo.

Proceso isotérmico es aquel que se realiza a una temperatura constante.

El **trabajo** es igualmente que el **calor**, una forma de **energía**. Hay diferentes tipos de trabajo, como el de expansión de volumen, el trabajo eléctrico o el trabajo mecánico. Es posible la transformación de un tipo de trabajo a otro. El trabajo mecánico es el producto escalar entre fuerza por distancia:

$$w = \int f \cdot ds \quad (1.16)$$

El trabajo se puede expresar más generalmente como el producto de un factor de intensidad por uno de capacidad.

Algunos tipos de trabajo se expresan en la tabla 1.1:

El trabajo es una cantidad algebraica y se tomará como positiva cuando el sistema realiza el trabajo y negativa cuando el trabajo es ejercido sobre él.

Tabla 1.1 Tipos de trabajo y principales características

TIPO DE TRABAJO	FACTOR DE INTENSIDAD	FACTOR DE CAPACIDAD
MECANICO	FUERZA	DISTANCIA
EXPANSION DE VOLUMEN	PRESION	CAMBIO DE VOLUMEN
INCREMENTO DE SUPERFICIE	TENSION SUPERFICIAL	CAMBIO DE AREA
ELECTRICO	DIFERENCIA DE POTENCIAL	CORRIENTE
GRAVITACIONAL	POTENCIAL GRAVITACIONAL	MASA

Experimento de Joule: Joule investigó el aumento de temperatura de un fluido al cual se le hizo una cantidad determinada de trabajo y encontró que bajo condiciones adiabáticas, el citado aumento era siempre el mismo, independientemente de la forma y trayectoria de trabajo. A la propiedad del sistema que cambia se le llama **Energía Interna** y se expresa para un proceso adiabático como:

$$\Delta E = -w \quad (1.17)$$

Calor. Bajo condiciones no adiabáticas podrá haber un cambio de estado. Por ejemplo, si introducimos un cuerpo caliente al agua, este se enfriará hasta alcanzar la misma temperatura de ésta. En este proceso hay un flujo de "calor" por medio del cual disminuirá la energía interna del cuerpo y aumentará la energía del agua. El calor será entonces, por definición, la energía transferida en el proceso y será positiva cuando el sistema absorbe calor de sus alrededores y negativa en caso contrario.

En un proceso sin intervención de trabajo podemos escribir:

$$\Delta E = q \quad (1.18)$$

Analicemos el caso de un proceso en el que estén involucrados calor y trabajo. Dado que la energía interna es una función de estado, su valor final es independiente de las trayectorias utilizadas, esto es, si se entrega a un sistema unas cantidades de calor y trabajo dadas, el valor final de la energía interna será independiente de si esto se hizo de forma simultánea o no, o del orden en que ocurrieron los procesos. Se verifica, entonces que

$$\Delta E = q - w \quad (1.19)$$

A ésta ley se le llama **primera ley de la termodinámica** y no es otra cosa que la ley de la conservación de la energía.

La primera ley de la termodinámica también se puede expresar en forma diferencial:

$$dE = dq - dw \quad (1.20)$$

donde se usarán los dq y dw para las diferenciales del calor y del trabajo. Estas son diferenciales inexactas, o sea que dependen del camino del proceso.

Entalpía. Algunas veces es conveniente trabajar con variables que se adapten mejor a los procesos que se analizan. Por ejemplo la mayor parte de las reacciones químicas se llevan a cabo a presión constante, esto es a una atmósfera.

Si se tiene un sistema a presión constante, donde el único trabajo será aquel hecho por presión-volumen el cambio de energía interna será:

$$\Delta E = q - p\Delta V = q - p(V_2 - V_1) = E_2 - E_1 \quad (1.21)$$

$$q = (E_2 + PV_2) - (E_1 + PV_1) \quad (1.22)$$

debido a que el calor absorbido está dado por la diferencia de dos cantidades que son funciones de estado, se puede introducir una nueva función de estado que se define como "**Entalpía**". Esta será:

$$H = E + PV \quad (1.23)$$

Debido a esto, en un proceso a presión constante, donde sólo va a haber un trabajo presión-volumen se cumplirá que:

$$\Delta q = H_2 - H_1 = \Delta H \quad (1.24)$$

o bien:

$$\Delta q_p = \Delta H \quad (1.25)$$

La entalpía es una función de estado del sistema. Si se eligen presión y temperatura como variables independientes se puede expresar como:

$$H = f(p, T) \quad (1.26)$$

$$dH = \left(\frac{\partial H}{\partial p} \right)_T dp + \left(\frac{\partial H}{\partial T} \right)_p dT \quad (1.27)$$

A presión constante se expresa como:

$$dq_p = dH = \left(\frac{\partial H}{\partial T} \right)_p dT \quad (1.28)$$

Capacidad calorífica. A la relación entre el flujo de calor y la variación en la temperatura se le llama capacidad calorífica y se expresa como:

$$C = \frac{dq}{dT} \quad (1.29)$$

El valor de la capacidad calorífica puede variar de acuerdo a las restricciones del sistema, así C_V será la capacidad calorífica de un sistema a volumen constante.

Considerando $E(V, T)$ a volumen constante se tiene:

$$dE = \left(\frac{\partial E}{\partial T} \right)_V dT + \left(\frac{\partial E}{\partial V} \right)_T dV = \left(\frac{\partial E}{\partial T} \right)_V dT \quad (1.30)$$

Además, de acuerdo a la primera ley de la termodinámica ($dE = dq - PdV$):

$$dE_V = dq \quad (1.31)$$

Por lo tanto, a volumen constante se cumplirá que:

$$dq = \left(\frac{\partial E}{\partial T} \right)_V dT \quad (1.32)$$

y la capacidad calorífica a volumen constante será:

$$C_V = \left(\frac{\partial q}{\partial T} \right)_V = \left(\frac{\partial E}{\partial T} \right)_V \quad (1.33)$$

y la capacidad calorífica a presión constante será utilizando (1.28)

$$C_p = \left(\frac{\partial q}{\partial T} \right)_p = \left(\frac{\partial H}{\partial T} \right)_p \quad (1.34)$$

Ambas cantidades son, en general, diferentes ya que en el caso de un proceso con volumen constante no habrá trabajo presión-volumen. En el caso de presión constante podrá haber una expansión, la cual consistirá en un trabajo el cual será hecho a costa del calor entregado.

1.4.2 Energía interna de un gas ideal

Joule trató de medir el cambio de energía interna de un gas haciendo el experimento que se muestra en la Fig. 1.3.

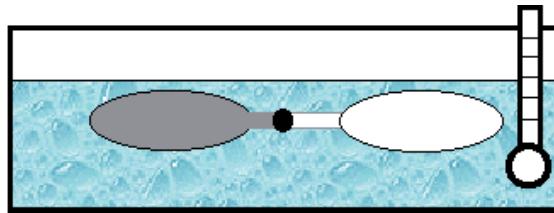


Fig.1.3 Experimento de Joule

Se tienen dos botellas conectadas por una válvula y sumergidas en un calorímetro¹ (éste nos permite medir el flujo de calor hacia un sistema). Una de las mismas se llena de un gas a presión P_1 , mientras que la otra está en vacío. Las botellas se consideran de volumen constante por lo que el sistema no intercambiaba trabajo ($dV = 0$) hacia el calorímetro. Entonces:

$$dw = 0 \quad (1.35)$$

Al abrir la válvula y dejar que el gas de la primera botella se expanda no se observa cambio de temperatura en el calorímetro lo que indica que el flujo de calor de las botellas hacia el calorímetro es cero:

$$dq = 0 \quad (1.36)$$

por consiguiente:

$$dE = 0 \quad (1.37)$$

o bien:

$$dE = \left(\frac{\partial E}{\partial T} \right)_V dT + \left(\frac{\partial E}{\partial V} \right)_T dV = 0 \quad (1.38)$$

¹ Un calorímetro es básicamente un recipiente aislado térmicamente y con una cantidad determinada de un líquido conocido. Al introducir un sistema cerrado, se puede conocer el calor emitido en un proceso mediante el incremento de temperatura del líquido.

Debido a que la temperatura se mantiene constante ($dT = 0$):

$$dE = \left(\frac{\partial E}{\partial V} \right)_T dV = 0 \quad (1.39)$$

$$\text{y} \quad \left(\frac{\partial E}{\partial V} \right)_T = 0 \quad (1.40)$$

Similarmente, si se considera a $E = f(T, p)$ se puede demostrar que:

$$\left(\frac{\partial E}{\partial p} \right)_T = 0 \quad (1.41)$$

Ejemplo 1. ¿Cuanto vale $(\partial H / \partial p)_T$ para un gas ideal?

$$H = E + PV = E + nRT \quad (1.42)$$

$$\left(\frac{\partial (E + nRT)}{\partial p} \right)_T = \left(\frac{\partial H}{\partial p} \right)_T = 0 \quad (1.43)$$

En gases reales el término (1.43) es diferente de cero y en la mayor parte de los casos reales se obtiene un enfriamiento al expandir el gas. Excepciones a esta norma son el hidrógeno y el helio a temperatura ambiente, los cuales se calientan cuando se expanden.

Relación entre C_p y C_v . La capacidad calorífica a presión constante C_p siempre es mayor que la capacidad calorífica a volumen constante C_v , dado que en el primer caso hay trabajo por expansión del sistema.

$$C_p - C_v = \left(\frac{\partial H}{\partial T} \right)_p - \left(\frac{\partial E}{\partial T} \right)_v \quad (1.44)$$

$$C_p - C_v = \left(\frac{\partial (E + PV)}{\partial T} \right)_p - \left(\frac{\partial E}{\partial T} \right)_v \quad (1.45)$$

Como

$$dE = \left(\frac{\partial E}{\partial T} \right)_V dT + \left(\frac{\partial E}{\partial V} \right)_T dV \quad (1.46)$$

a presión constante se tendrá:

$$\left(\frac{\partial E}{\partial T}\right)_p = \left(\frac{\partial E}{\partial T}\right)_V + \left(\frac{\partial E}{\partial V}\right)_T \left(\frac{\partial V}{\partial T}\right)_p \quad (1.47)$$

lo que da:

$$C_p - C_V = \left(\frac{\partial E}{\partial V}\right)_T \left(\frac{\partial V}{\partial T}\right)_p + p \left(\frac{\partial V}{\partial T}\right)_p \quad (1.48)$$

$$C_p - C_V = \left[p + \left(\frac{\partial E}{\partial V}\right)_T \right] \left(\frac{\partial V}{\partial T}\right)_p \quad (1.49)$$

Para un gas ideal $(\partial E / \partial V)_T = 0$ y $(\partial V / \partial T)_p = nR / p$ luego:

$$C_p - C_V = nR \quad (1.50)$$

1.4.3 Termoquímica

Se puede aplicar la primera ley de la termodinámica para calcular el calor absorbido o emitido en una reacción química. En algunos casos el flujo de energía será positivo y en otros negativo.

Energía o entalpía de reacción. El cambio de energía interna ΔE o de entalpía ΔH de una reacción depende del estado en que se encuentren los reactivos y los productos. Por ejemplo el calor de combustión² del grafito es diferente al del diamante.

Por conveniencia se definen los estados estándar de un componente. Normalmente el estado estándar de un gas o un líquido es el que se presenta a una presión de una atmósfera y 20 °C, mientras que para un sólido se define también la forma cristalina.

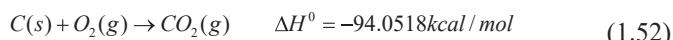
² El calor que se produce al quemarse el elemento

Para una reacción se escribirá:



donde ΔH indica la entalpía de la reacción y ΔH^0 significa que tanto los reactivos como los productos se encuentran en sus estados estándar.

Ejemplo 2. A partir de la combustión del carbono se puede medir el calor de formación del CO_2 .

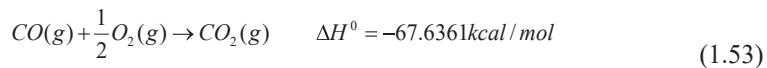


En algunos casos no es posible medir experimentalmente la entalpía (o energía) de una reacción. Esto es común cuando en un proceso hay varios subproductos. En estos casos se pueden hacer cálculos a partir de otras mediciones, ya que la energía de un sistema depende solo de sus variables de estado y no de la forma en que se llega a las mismas.

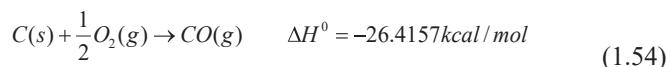
Ejemplo 3. El monóxido de carbono (CO) es producto de la combustión incompleta del carbono. Este se produce cuando la disponibilidad de oxígeno es insuficiente para producir bióxido de carbono (CO_2). En este último caso se producirán ambos compuestos, dificultando el asignar el calor de reacción a cada uno de los procesos de oxidación. Sin embargo, se puede conocer el calor de formación del monóxido de carbono restando, al calor de formación del CO_2 , el de la oxidación del CO (ver tabla 1.2):

Tabla 1.2 Calor de formación y de reacción del CO_2 y CO

Calor de formación del CO_2	-94.0518
Menos calor de reacción de oxidación del CO	-67.6361
Calor de formación del CO	-26.4157



Entonces se puede escribir



Entalpía de formación

No es posible obtener la entalpía absoluta de una sustancia por lo que se definirá la entalpía de formación de un compuesto como la entalpía necesaria para formar el mismo a partir de sus elementos, en sus estados estándar. La entalpía de formación de un compuesto se denota como $\Delta\bar{H}_{f,T}^0$, donde en lugar de T se dará la temperatura de la reacción.

Conociendo las entalpías de formación de los diferentes reactivos y productos se puede conocer el cambio de entalpía de una reacción.

$$\Delta H^0 = \sum v_j H_{f,productos}^0 - \sum v_i H_{f,reactivos}^0 \quad (1.55)$$

donde v_i y v_j son las cantidades molares de las substancias.

Esta no es sino una aplicación directa de la primera ley de la termodinámica.

Energía de enlace. Los átomos en las moléculas están ligados por enlaces entre sí. Cada uno de éstos es de naturaleza diferente y depende del tipo (e.g. covalente, iónico, etc.) La energía de enlace E(A-B) es la contribución total de un enlace determinado en una molécula a la energía total de unión que es la requerida para disociar las moléculas en sus elementos en sus estados estándar.

1.5 SEGUNDA Y TERCERA LEYES DE LA TERMODINAMICA

La primera ley de la termodinámica se ocupa de la conservación de la energía, sin embargo no puede predecir la dirección en que ocurren los cambios, para esto ayuda la segunda ley, la cual está basada en observaciones experimentales pero que no ha podido ser demostrada.

Procesos espontáneos. En la vida diaria se observan una serie de eventos los cuales ocurren en una dirección: Los cuerpos caen, el calor fluye hacia el lado frío o una en mezcla de sal y agua se disolverá la primera. Sin embargo, es difícil hacer los procesos contrarios, o sea que estos no ocurren espontáneamente.

Cambios reversibles e irreversibles. Un proceso reversible es aquel en el cual es posible volver al estado inicial variando infinitesimalmente una variable independiente. Contrariamente en un proceso irreversible es necesaria la introducción de trabajo para volver al estado inicial. Un proceso espontáneo puede ocurrir reversible o irreversiblemente. En el primer caso ocurrirá lentamente y prácticamente en equilibrio y el trabajo obtenido será el máximo mientras que en el segundo caso el trabajo obtenido por el cambio será inferior al trabajo teórico máximo que es posible obtener.

1.5.1 Segunda ley de la termodinámica

El hecho de que no existen procesos reversibles reside en que una parte del trabajo obtenido se convierte en calor. Para poder convertir una cantidad de calor en trabajo es necesario que exista un gradiente de temperatura. La segunda ley de la termodinámica se puede enunciar de

varias formas: "El calor no puede ser convertido completamente en trabajo sin causar otros cambios en el sistema o sus alrededores"; "Es imposible usar un proceso cíclico para transferir calor y convertirlo en trabajo sin al mismo tiempo transferir calor del cuerpo más caliente al más frío"; "Es imposible usar un proceso cíclico para transferir calor de un cuerpo frío a uno caliente, sin al mismo tiempo convertir una cantidad de trabajo en calor".

Ciclo de Carnot

Es posible transformar calor en trabajo si hay dos depósitos de temperatura constante T_1 y T_2 empleando un proceso cíclico en el que un gas se puede usar donde el resultado final es la transferencia de calor de un depósito al otro y la realización de un trabajo.

El ciclo de Carnot es muy similar al proceso que tiene lugar en un motor de combustión interna, este tiene cuatro fases, al cabo de las cuales el sistema está en la misma posición que al principio. En este caso se aplican las ecuaciones sobre el gas, y para algunas consideraciones es conveniente suponer que es ideal.

Dado que el gas volverá a su estado inicial, todas las variables termodinámicas del gas serán las mismas al final del ciclo. Los incrementos de energía interna y de entalpía serán cero.

Las cuatro fases del ciclo, ilustradas en la figura 1.4, son:

Expansión isotérmica: El gas se expande isotérmica y reversiblemente a una temperatura constante T_2' durante este proceso absorbe una cantidad de calor q_2 , proveniente del entorno que se considera un reservorio de calor.

El gas se expande adiabáticamente y la temperatura baja de T2 a T1. En esta fase el sistema está aislado térmicamente y no puede absorber calor, sin embargo entrega trabajo, lo que provoca la reducción de calor.

El gas se comprime isotérmicamente a la temperatura T1 liberando la cantidad de calor q1. En esta fase pone en contacto el sistema con un reservorio a temperatura menor, con lo que se enfriá el gas, a la vez el sistema pierde energía debido al trabajo de compresión de volumen.

Por medio de una compresión adiabática retorna el gas a su estado inicial a T2. En este último paso el gas se sigue comprimiendo para llegar al volumen inicial. El trabajo resultante del la compresión genera un incremento de temperatura del sistema.

	Expansión isotérmica	Temperatura constante T2 Absorción de calor q2	$S = \frac{q_2}{T}$
	Expansión adiabática	Temperatura variable de T2 a T1.	$S = 0$
	Compresión isotérmica	Temperatura Constante T1 Emisión de calor q1	$S = -\frac{q_1}{T}$
	Compresión adiabática	Temperatura variable de T1 a T2.	$S = 0$

Fig. 1.4 Ciclo de Carnot.

Todo el proceso se realiza de manera espontánea, esto es sin aplicación de trabajo externa. El trabajo realizado y entregado hacia el exterior proviene de la transformación del calor absorbido por el gas. En un motor

de combustión interna real el trabajo es realizado por el incremento de volumen provocado por el calentamiento (calor) que se produce al arder el carburante.

La eficiencia del proceso estará dada por la cantidad de trabajo obtenido sobre la cantidad de calor removido del depósito de mayor temperatura, que es el que provee la energía:

$$\eta = \frac{w}{q_2} = \frac{q_2 - q_1}{q_2} \quad (1.56)$$

debido a que, de acuerdo a la primera ley de la termodinámica la cantidad de trabajo obtenida va a ser igual a la cantidad de calor absorbida menos la cantidad de calor liberada por el sistema. Dado que los estados inicial y final del sistema son los mismos, la ley de la conservación de la energía permite sustituir al trabajo por la diferencia entre el calor absorbido y el calor entregado.

Entropía. La entropía es una cantidad termodinámica extensiva, esto es, que depende del tamaño del sistema. Será designada por S y es una función de estado. Se puede definir el aumento de entropía en un proceso como:

$$dS = \frac{dq}{T} \quad (1.57)$$

si el proceso es reversible y como:

$$dS > \frac{dq}{T} \quad (1.58)$$

para procesos irreversibles.

Combinando estas dos ecuaciones se tiene la segunda ley de la termodinámica:

$$dS \geq \frac{dq}{T} \quad (1.59)$$

Entropía en el ciclo de Carnot. Para el ciclo de Carnot se tienen los siguientes cambios de entropía:

$$S_1 = \frac{q_2}{T_2} \quad (1.60a)$$

$$S_2 = 0 \quad (1.60b)$$

$$S_3 = \frac{q_1}{T_1} \quad (1.60c)$$

$$S_4 = 0 \quad (1.60d)$$

Debido a que el estado final es igual al estado inicial el cambio neto de entropía será cero:

$$S = \frac{q_2}{T_2} - \frac{q_1}{T_1} = 0 \quad (1.61)$$

Entonces de las ecuaciones (1.56) y (1.61) se obtiene:

$$\eta = \frac{w}{q_2} = \frac{T_2 - T_1}{T_2} \quad (1.62)$$

En la aplicación de esta ecuación es importante considerar escalas absolutas de temperatura.

Cambios de entropía en procesos reversibles. Es fácil calcular los cambios de entropía en procesos reversibles:

$$S = \frac{q_{rev}}{T} \quad (1.63)$$

Para un proceso en el que se absorbe calor:

$$dS = \frac{CdT}{T} \quad (1.64)$$

si C es independiente de la temperatura:

$$S_2 - S_1 = \int_{S_1}^{S_2} dS = \int_{S_1}^{S_2} \frac{CdT}{T} = C \ln \frac{T_2}{T_1} \quad (1.65)$$

donde C_p o C_v pueden ser usados para C, dependiendo del caso.

Entropía en cambios de estado. Los cambios de estado en substancias puras se pueden realizar a temperatura constante (e.g. el agua pura se funde exactamente a 0 °C). Si este es el caso el cambio de entropía durante la fusión será:

$$dS_f = \frac{dq}{T} = \frac{nC_f}{T_f} \quad (1.66)$$

y para evaporación

$$dS_v = \frac{dq}{T} = \frac{nC_v}{T_v} \quad (1.67)$$

donde C_f y C_v son los calores de fusión³ y vaporización⁴ respectivamente y T_f y T_v las temperaturas a que suceden los procesos. Nótese que el cambio de entropía es negativo si se invierte el proceso (solidificación y condensación).

Si el proceso no se realiza a temperatura constante, es necesario integrar de acuerdo a:

$$S_2 - S_1 = \int_{S_1}^{S_2} dS = \int_{S_1}^{S_2} \frac{CdT}{T} \quad (1.68)$$

Entropía en la expansión de un gas ideal. Tomemos el caso de un gas ideal:

$$dq = C_v dT + pdV = C_v dT + \frac{nRT}{V} dV \quad (1.69)$$

³ Calor de fusión: El calor necesario para fundir un mol de materia.

⁴ Calor de vaporización: El calor necesario para evaporar un mol de materia.

$$dS = \frac{dq}{T} = C_V \frac{dT}{T} + \frac{nR}{V} dV \quad (1.70)$$

$$\Delta S = C_V \ln \frac{T_2}{T_1} + nR \ln \frac{V_2}{V_1} \quad (1.71)$$

A temperatura constante:

$$\Delta S = nR \ln \frac{V_2}{V_1} \quad (1.72)$$

Entropía de mezcla. Si se mezclan dos gases ideales a temperatura y presión constante, el cambio de entropía será igual al cambio de entropía por la expansión individual de cada uno de ellos, dado que se supone que no hay interacción entre los mismos:

$$\Delta S_1 = n_1 R \ln \frac{V_1 + V_2}{V_1} \quad (1.73)$$

$$\Delta S_2 = n_2 R \ln \frac{V_1 + V_2}{V_2} \quad (1.74)$$

Si se define a:

$$X_1 = \frac{n_1}{n_1 + n_2} \quad (1.75)$$

$$X_2 = \frac{n_2}{n_1 + n_2} \quad (1.76)$$

entonces

$$\Delta S = S_1 + S_2 = -R(n_1 \ln X_1 + n_2 \ln X_2) \quad (1.77)$$

donde X1 y X2 son las fracciones molares de los gases.

Entropía a nivel microscópico. Desde el punto de vista microscópico la entropía puede ser expresada por la siguiente relación:

$$S = k \ln \Omega \quad (1.78)$$

donde Ω es el número de posibles configuraciones de los átomos. Esto es, si un átomo puede tener dos configuraciones posibles, Ω será igual a dos. Debido a la naturaleza estadística de los procesos macroscópicos, las diferentes posibilidades de distribución de estados es muy amplia. Generalmente se habla de distribuciones energéticas, o sea niveles ocupados y desocupados, aunque esta regla se aplica para cualquier caso en que existan varias posibilidades. Es por ello que la entropía también se asocia con el desorden.

1.6 TERCERA LEY DE LA TERMODINÁMICA

Esta ley nos permite calcular la entropía absoluta de un sistema y se postula como:

"La entropía de un sólido monocrystalino a 0 K es cero".

Para calcular la entropía absoluta de un sistema se toman los cambios de forma reversible a partir de temperatura cero hasta llegar al estado final en consideración.

Para comprender esta ley se verá el siguiente ejemplo donde se calcula la entropía del $S0_2$. A partir del estado ideal del material a 0 K se va elevando la temperatura. Utilizando la ecuación:

$$dS = \frac{dq}{T} \quad (1.57)$$

Se toman los valores de la capacidad calorífica medidos experimentalmente, si esto es posible. Para temperaturas muy bajas se aproxima la capacidad calorífica al valor teórico, en este caso está dado por la función de Debye. Así mismo se deben considerar los cambios de fase, usando los calores específicos de fusión y de evaporación del material.

Dado que varios de los parámetros se obtienen de manera experimental, en los casos en que no se tenga una función analítica, el cálculo se puede hacer de manera gráfica o numérica. La entropía del material estará dada por la suma de los incrementos de entropía en cada una de las fases del proceso como se muestra en la tabla 1.3:

Tabla 1.3 valores de S para diferentes T y métodos de obtención

T (K)	Método de cálculo	S
0-15	Función de Debye ($C_p=\text{const.} \cdot T^3$)	0.30
15-197.64	Gráfico, del sólido	20.12
197.64	Fusión	8.95
197.64-263.08	Gráfico del líquido	5.96
263.08	Evaporación	22.06
263.08-298.1	C_p de un gas	1.25
	$\Delta S_{298.1}^0$	59.24 cal/mol K

1.7 CRITERIOS DE EQUILIBRIO QUÍMICO

Sí hay equilibrio, cualquier proceso que tenga lugar será reversible para un proceso irreversible:

$$dS > \frac{dq}{T} \quad (1.78)$$

Entonces:

$$dq - TdS < 0 \quad (1.79)$$

De la primera ley de la termodinámica:

$$dE = pdV - TdS < 0 \quad (1.80)$$

Si el volumen y la entropía son constantes

$$(dE)_{V,S} < 0$$

(1.81)

En un sistema aislado, el volumen y la energía interna pueden ser constantes, entonces:

$$(dS)_{E,V} > 0$$

(1.82)

A volumen constante se tiene:

$$(dE - TdS)_V < 0$$

(1.83)

y:

$$(dE - TS)_{T,V} < 0$$

(1.84)

Si la temperatura es constante, a la cantidad $E - TS$ se le llamará "**Energía**

libre de Helmholtz" y se denotará por F .

$$(dF)_{T,V} < 0$$

(1.85)

También se puede definir a la "**Energía libre de Gibbs**" (G) como $E + PV - TS$ y:

$$(dG)_{T,p} < 0$$

(1.86)

Para procesos reversibles estas desigualdades se convierten en igualdades:

$$(dE)_{V,S} = 0$$

(1.87)

$$(dE)_{T,S} = 0$$

(1.88)

$$(dF)_{T,V} = 0$$

(1.89)

$$(dG)_{T,p} = 0$$

(1.90)

Las energías libres de Helmholtz y de Gibbs son funciones de estado, y las diferenciales absolutas serán:

$$dF = d(E - TS) = dE - TdS - SdT \quad (1.91)$$

$$dG = d(E + PV - TS) = dE + pdV - Vdp - TdS - SdT \quad (1.92)$$

Si se tiene la presión y la temperatura constantes:

$$dG = dE + pdV - TdS \quad (1.93)$$

$$dG = dq - dw + pdV - TdS \quad (1.94)$$

Si el cambio es reversible ($dq = TdS$):

$$-dG - dw - pdV = 0 \quad (1.95)$$

o sea que el cambio en la energía libre de Gibbs será igual al máximo trabajo posible de obtener por un sistema a presión y volumen constante.

De manera similar:

$$-dF - dw = 0 \quad (1.96)$$

o sea que la energía libre de Helmholtz será igual al máximo trabajo posible de obtener por un sistema a temperatura y presión constantes.

Las energías libres son muy útiles para calcular el estado de equilibrio de los sistemas.

Diferenciales de las funciones termodinámicas. Para un proceso reversible donde solo hay trabajo presión volumen, el cambio de energía interna estará dado por:

$$dE = dq - dw \Rightarrow dE = TdS - pdV \quad (1.97)$$

El cambio de entalpía será:

$$\begin{aligned} dH &= d(E + PV) = TdS - pdV + VdP + pdV \Rightarrow \\ dH &= TdS + Vdp \end{aligned} \quad (1.98)$$

El cambio de energía libre de Helmholtz es:

$$\begin{aligned} dF &= d(E - TS) = TdS - pdV - TdS - SdT \Rightarrow \\ dF &= -pdV - SdT \end{aligned} \quad (1.99)$$

y para la energía libre de Gibbs:

$$\begin{aligned} dG &= d(H - TS) = TdS + Vdp - TdS - SdT \Rightarrow \\ dG &= Vdp - SdT \end{aligned} \quad (1.100)$$

Resumiendo:

$$dE = TdS - pdV \quad (1.101)$$

$$dH = TdS + Vdp \quad (1.102)$$

$$dF = -pdV - SdT \quad (1.103)$$

$$dG = Vdp - SdT \quad (1.104)$$

Comparando estas ecuaciones con la expresión matemática de una diferencial exacta:

$$du(x, y) = \left(\frac{\partial u(x, y)}{\partial x} V \right)_y dx + \left(\frac{\partial u(x, y)}{\partial y} \right)_x dy \quad (1.105)$$

Se pueden determinar también los siguientes parámetros:

$$\left(\frac{\partial E}{\partial S} \right)_V = T \quad (1.106) \quad \left(\frac{\partial E}{\partial V} \right)_S = -p \quad (1.107)$$

$$\left(\frac{\partial H}{\partial S} \right)_p = T \quad (1.108) \quad \left(\frac{\partial H}{\partial p} \right)_S = V \quad (1.109)$$

$$\left(\frac{\partial F}{\partial V} \right)_T = -p \quad (1.110) \quad \left(\frac{\partial F}{\partial T} \right)_V = -S \quad (1.111)$$

$$\left(\frac{\partial G}{\partial p} \right)_T = V \quad (1.112) \quad \left(\frac{\partial G}{\partial T} \right)_p = -S \quad (1.113)$$

Cantidades molares parciales. Cada sustancia de una solución contribuye a cada una de las propiedades termodinámicas extensivas. En el caso de soluciones ideales se pueden sumar algunas de éstas como es el caso del volumen, pero esto no siempre es válido, como es el caso de una mezcla de $H_2O + H_2SO_4$ donde los volúmenes no son aditivos.

En el caso de una mezcla, el volumen dependerá además de presión y temperatura del número de moles de cada uno de los componentes. El diferencial del volumen será:

$$dV = \left(\frac{\partial V}{\partial p} \right)_{T,n_1,n_2} dp + \left(\frac{\partial V}{\partial T} \right)_{p,n_1,n_2} dT + \left(\frac{\partial V}{\partial n_1} \right)_{p,T,n_2} dn_1 + \left(\frac{\partial V}{\partial n_2} \right)_{p,T,n_1} dn_2 \quad (1.114)$$

Así el volumen parcial molal será:

$$\overline{V}_i = \left(\frac{\partial V}{\partial n_i} \right)_{p,T,n_j} \quad (1.115)$$

En general, para una cantidad termodinámica extensiva, a temperatura y presión constantes

$$dY = \overline{Y}_1 dn_1 + \overline{Y}_2 dn_2 + \cdots + \overline{Y}_n dn_n, \quad (1.116)$$

donde:

$$\overline{Y}_i = \left(\frac{\partial Y}{\partial n_i} \right)_{p,T,n_j} \quad (1.117)$$

será la cantidad molal parcial de Y para el componente i.

Las cantidades molales parciales son cantidades intensivas de un sistema ya que estas sólo dependerán de la fracción molar del componente, normalizada con respecto al número total de moles del sistema.

Potencial químico. El potencial químico sirve para conocer los criterios de equilibrio de un sistema.

Si se tiene una mezcla de componentes cada cual con un número de moles n_i :

$$dE = TdS - pdV + \left(\frac{\partial E}{\partial n_i} \right)_{V,S,n_j} dn_i + \cdots + \left(\frac{\partial E}{\partial n_n} \right)_{V,S,n_j} dn_n \quad (1.118)$$

se llamará potencial químico (μ_i) del componente i a:

$$\mu_i = \left(\frac{\partial E}{\partial n_i} \right)_{V,S,n_j} dn_i \quad (1.119)$$

Así:

$$dE = TdS - pdV + \sum \mu_i dn_i \quad (1.120)$$

y se puede demostrar que:

$$dG = Vdp - SdT + \sum \mu_i dn_i \quad (1.121)$$

Entonces:

$$\mu_i = \left(\frac{\partial G}{\partial n_i} \right)_{p,T,n_j} = \bar{G}_i \quad (1.122)$$

Dado que el potencial químico es una cantidad intensiva del sistema la diferencia de potencial químico entre dos fases determinará la dirección de una reacción de la misma forma que la diferencia de temperaturas determina la dirección del flujo de calor.

Potencial químico de un gas ideal. A temperatura constante se expresa como:

$$\left(\frac{\partial \bar{G}}{\partial p} \right)_T = \left(\frac{\partial \mu}{\partial p} \right)_T = \bar{V} \quad (1.123)$$

Para un gas ideal $\bar{V} = RT/p$

$$\int_{\mu_0}^{\mu} d\mu = \int_{p_0}^p \frac{RT}{p} dp \quad (1.124)$$

$$\mu - \mu_0 = RT \ln \frac{p}{p_0} \quad (1.125)$$

o sea:

$$p = p_0 e^{\frac{\mu - \mu_0}{RT}} \quad (1.126)$$

1.8 . LEY DE ACCIÓN DE MASAS

En una ecuación química los reactivos y productos se presentan en relaciones constantes, por ejemplo:



De manera general se puede escribir:



donde las letras mayúsculas representan diferentes compuestos y las minúsculas los **coeficientes estequiométricos**, o sea aquellos que dan las relaciones entre las cantidades de las diferentes substancias.

La expresión para el cambio en la energía libre de Gibbs será:

$$dG = Vdp - SdT + \mu_A dn_A + \mu_B dn_B + \mu_C dn_C + \mu_D dn_D \tag{1.129}$$

Si no se añade más material al sistema cada cambio en el número de moles de los reactivos deberá estar acompañado por un cambio en las cantidades de los productos. Si se define:

$$n_i - n_i^0 = v_i \lambda \tag{1.130}$$

donde n_0 y n_i^0 son los números inicial y final de moles de la sustancia i y v_i es el coeficiente estequiométrico de la misma:

$$\begin{aligned} n_A - n_A^0 &= a\lambda \\ n_B - n_B^0 &= b\lambda \\ n_C - n_C^0 &= c\lambda \\ n_D - n_D^0 &= d\lambda \end{aligned} \tag{1.131}$$

entonces

$$dn_i = v_i d\lambda \quad (1.132)$$

y:

$$-\frac{dn_A}{a} = -\frac{dn_B}{b} = \frac{dn_C}{c} = \frac{dn_D}{d} = d\lambda \quad (1.133)$$

Entonces queda a presión y temperaturas constantes:

$$dG = (-a\mu_A - b\mu_B + c\mu_C + d\mu_D)d\lambda \quad (1.134)$$

Para equilibrio químico

$$\left(\frac{dG}{d\lambda} \right)_{T,p} = 0 \quad (1.135)$$

Entonces:

$$c\mu_C + d\mu_D - a\mu_A - b\mu_B = 0 \quad (1.136)$$

Introduciendo el concepto de actividad (a):

$$\mu = \mu_i^0 + RT \ln a_i \quad (1.137)$$

donde μ_i^0 será el potencial químico para la sustancia i con actividad igual a uno y su valor depende de la presión y la temperatura.

$$c\mu_C + d\mu_D - a\mu_A - b\mu_B = \Delta\mu + RT \ln \left(\frac{a_C^c a_D^d}{a_A^a a_B^b} \right) \quad (1.138)$$

Entonces en equilibrio:

$$\Delta\mu = c\mu_C + d\mu_D - a\mu_A - b\mu_B = -RT \ln K \quad (1.139)$$

donde:

$$K = \left(\frac{a_C^c a_D^d}{a_A^a a_B^b} \right) \quad (1.140)$$

Esto significa que para que exista equilibrio en un sistema, la suma de los potenciales químicos multiplicados por los coeficientes estequiométricos en un sistema deberá ser igual a $RT \ln K$, o sea, que las actividades en un sistema se modificarán hasta llegar a ese valor.

Para gases ideales la actividad es igual a la presión parcial del gas.

$$K_p = \frac{p_C^c p_D^d}{p_A^a p_B^b} \quad (1.141)$$

Esto significa que en un proceso la presión parcial de los gases se ajustará hasta que K llegue al valor de equilibrio.

Dado que μ_i es el potencial químico de una sustancia en su estado estándar el cual a su vez es igual a su energía estándar molar de Gibbs.

$$\Delta G^0 = -RT \ln K \quad (1.141)$$

El potencial químico de una sustancia en su estado estándar es fácil de determinar lo cual permite conocer el valor de la constante K . Si ΔG_i^0 tiene un valor negativo K va a ser mayor que uno y habrá una reacción espontánea. Si $\Delta G_i^0 = 0$ habrá equilibrio entre reactivos y productos, por lo que ya no va a haber una reacción. Si $\Delta G_i^0 < 0$ K será negativa y los reactivos no podrán formar productos.

1.9 GRÁFICAS DE ARHENIUS

El cambio de energía libre de una reacción estará dado por:

$$\Delta G^0 = -RT \ln K \quad (1.143)$$

Esto también se puede escribir como una función exponencial del tipo función de distribución:

$$K = e^{-\frac{\Delta G^0}{RT}} \quad (1.144)$$

donde K estará dada por la ley de acción de masas. En general existen varios procesos que muestran este tipo de comportamiento, como los que se muestran en la Tabla 1.5

Tabla 1.5 ejemplos de procesos con diferente comportamiento

Presión de vapor	$p = p_0 e^{-\left(\frac{\Delta G_p}{RT}\right)}$	$\Delta G_p = -RT \ln p / p_0$	(1.145)
Velocidad de reacción	$A = A_0 e^{-\left(\frac{\mu - \mu_0}{RT}\right)}$	$\mu - \mu_0 = -RT \ln A / A_0$	(1.146)
Concentración de electrones	$np = N_v N_c e^{-\left(\frac{E_G}{kT}\right)}$	$E_G = -kT \ln np / N_v N_c$	(1.147)
Coeficiente de difusión	$D = D_0 e^{-\left(\frac{E_A}{kT}\right)}$	$E_A = -kT \ln D / D_0$	(1.148)

En cada uno de ellos el término de la exponencial es una energía asociada al proceso. Esta puede ser la que necesitan las moléculas en un líquido para salir de la superficie y pasar a la fase gaseosa, la barrera de potencial que requieren vencer dos reactivos para reaccionar, el salto que tienen que dar un electrón ligado en un semiconductor para generar un par electrón hueco o la barrera energética que limita la difusión de una especie. Esta se denomina **Energía de Activación** y tiene un valor característico que puede servir para identificar la naturaleza del proceso. Por ejemplo en el caso de un semiconductor si conocemos E_G podemos identificar si es silicio ($E_G = 1.11 \text{ eV}$), germanio ($E_G = 0.7 \text{ eV}$) o ninguno de ellos. Cuando un proceso cumple con este comportamiento se dice que está **térmicamente activado**.

La figura 1.5 muestra la gráfica convencional de los coeficientes de difusión de diferentes impurezas en silicio los cuales muestran un comportamiento como el de la ecuación (1.148).

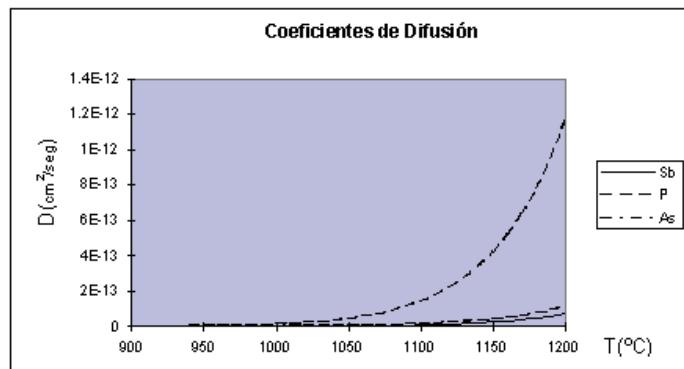


Fig. 1.5 Coeficientes de difusión de diferentes impurezas en silicio

Considerando el comportamiento exponencial de los procesos, se puede escribir en general una ecuación para un proceso activado térmicamente:

$$\frac{E_A}{R} \left(\frac{1}{T} \right) + \ln K - \ln K_0 = 0 \quad (1.149)$$

Análogamente:

$$ax + by + c = 0 \quad (1.150)$$

Si se considera que las variables de la recta son $1/T$ y $\ln K$ para x y y respectivamente, se obtiene una pendiente igual a:

$$m = \frac{b}{a} = \frac{E_A}{R} \quad (1.151)$$

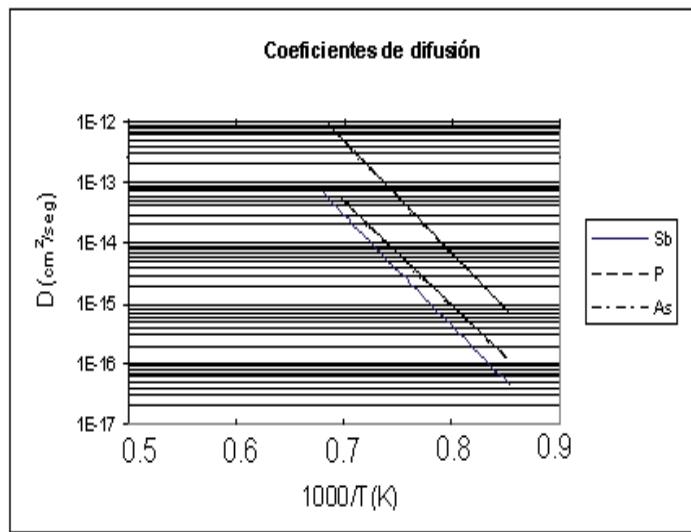


Fig. 1.6 Gráfica de Arhenius de los coeficientes de difusión de impurezas.

Si en lugar de realizar la gráfica convencional se hace D vs. $1000/T$, se tiene la que se muestra en la figura 6, donde en lugar de curvas exponenciales se muestran rectas.

La pendiente de las rectas será igual a la R/E_A o k/E_A según sea el caso y la determinación de ésta permitirá conocer su valor. A este tipo de diagramas se le denomina "**Gráfica de Arhenius**".

Considerando que las mediciones de los procesos en muchos casos son experimentales, y tan solo se dispone de algunos puntos, la utilidad de una gráfica de Arhenius radica en:

- Identificar si un proceso es térmicamente activado.
- Facilitar la identificación energías de activación

- Extrapolar valores.

Para la realización de gráficas de Arhenius es válido realizar mediciones de procesos y graficarlos aún cuando no se ha llegado al equilibrio, por ejemplo midiendo la cantidad de un líquido que se ha evaporado en un determinado lapso de tiempo, para determinar la energía de activación del proceso de evaporación.

La extrapolación es muy útil para la estimación de valores en rangos donde no se puede realizar una medición, por ejemplo determinar la probabilidad de falla de un dispositivo a temperatura ambiente, cuando ésta es muy pequeña, o cuando el tiempo medio de vida del mismo de varios años. En este caso se realizan pruebas de fallas a alta temperatura, se determina el comportamiento del proceso durante tiempos más cortos y se hace la estimación a condiciones normales.

Hay que tener cautela en la identificación de resultados por medio de gráficas de Arhenius y asociarlos con energías de activación de procesos conocidos, porque es común que los procesos reales estén enmascarados.

1.10 BIBLIOGRAFÍA

- F. Daniels and R. A. Alberty. Physical Chemistry, John Wiley and Sons.
- S. Glasstone and D. Lewis. Elements of Physical Chemistry. D. Van Nostrand Company inc.
- H. B. Callen, Thermodynamics, John Wiley and Sons.
- L. García Colín, Introducción a la termodinámica. Ed. Trillas.
- K. Warth, Thermodynamics. Mc. Graw Hill.

CAPÍTULO 2

DIAGRAMAS DE FASE

Los diagramas de fase o diagramas de equilibrio son muy útiles en tecnología para determinar compatibilidad entre diferentes materiales y poder predecir resultados antes de experimentar combinaciones nuevas.

En los procesos tecnológicos de semiconductores los diagramas de fases se utilizan para diferentes propósitos, como: determinar la impurificación máxima de un semiconductor, conocer los productos resultantes de procesos como oxidación en presencia de elementos extraños, compatibilidad entre metales u otras películas empleadas en la tecnología. El conocimiento de los diagramas de fases nos permite prevenir reacciones indeseables y determinar los parámetros óptimos para un proceso determinado.

Una **fase** es una parte de un sistema, uniforme en su composición química y propiedades físicas, y está separada de otras zonas homogéneas por fronteras. Dado que los gases son completamente miscibles entre sí, siempre se presentará una sola fase gaseosa, pero puede haber varias fases líquidas y sólidas.

Cuando se habla del número de fases se entenderá que son fases diferentes, e.g.: agua con hielo contará como dos fases, independientemente si el hielo está fragmentado en muchos pedazos o no.

El equilibrio de las fases no depende de las cantidades relativas de cada una. E.g. Las condiciones para que exista un equilibrio entre agua y hielo son fijas independientemente del volumen de cada una de las fases.

Un diagrama de fases muestra sus composiciones en equilibrio para condiciones termodinámicas (temperatura, presión, composición) definidas. Si se conocen las constantes de equilibrio de los materiales, se pueden aplicar las fórmulas correspondientes ($(dS)_{V,E} \geq 0$, $(dF)_{T,V} \leq 0$, etc.) para la elaboración del diagrama, sin embargo en la mayor parte de los casos los diagramas de fase se elaboran a partir de datos experimentales.

Un sistema puede tener uno o varios *componentes*. El número de componentes es la cantidad mínima de substancias que pueden ser utilizadas para la descripción del mismo. Las substancias pueden ser elementos o compuestos químicos, por ejemplo un diagrama de galio, arsénico y arseniuro de galio tendrá 3 substancias, pero se podrá definir a partir de los dos elementos (número de componentes = 2). Un sistema cloro-oxígeno, podrá mostrar un gran número de substancias (Cl_2O , Cl_2O_3 , Cl_2O_5 , etc.), aunque solo dos componentes.

Para describir un sistema se necesitan especificar algunas variables como temperatura, presión y composición. E. g. para describir un mol de gas se pueden usar temperatura, presión y volumen, sin embargo si se especifican solo temperatura y volumen, la presión ya estará determinada. Se denomina número de grados de libertad a la cantidad de parámetros que pueden ser variados manteniendo las fases del sistema.

2.1 REGLA DE FASES

La relación entre componentes, fases y grados de libertad está dada por la Regla de las Fases:

$$F = C - P + 2$$

Donde F es el número de grados de libertad, C el número de componentes y P el número de fases.

2.2 SISTEMAS UNITARIOS.

Los sistemas unitarios son aquellos que presentan solo un componente.

Este puede ser un elemento o un compuesto estable. De acuerdo a la regla de las fases ($C=1$):

$$F = 1 - P + 2 = 3 - P \quad (2.1)$$

donde $P_{\max}=3$ y por lo tanto $F_{\max}=2$

Un sistema unitario puede tener un máximo de 2 grados de libertad o tres fases coexistentes, luego para hacer su diagrama se requieren solo 2 coordenadas, que pueden ser temperatura y presión.

2.2.1 Diagrama del agua

El agua es un compuesto estable y se puede elaborar un diagrama unitario a partir de ésta. Se le conocen tres fases, hielo, agua y vapor, aunque a altas presiones se presentan otras fases del hielo. Se conocen adecuadamente los valores en equilibrio entre las mismas. El punto de congelación del agua a una atmósfera es 0°C y el de ebullición 100°C .

De acuerdo a la regla de las fases en un diagrama unitario un punto triple, o sea donde coexisten tres fases es invariable y todas las variables termodinámicas tienen un valor definido. Así, el punto triple del agua, que es donde coexisten las fases sólida, líquida y gaseosa, se usa como patrón internacional de temperatura, Fig. 2.1. Este valor se puede reproducir fácilmente en cualquier lugar de la tierra si se toman las precauciones para

que coexistan las fases y se alcance un equilibrio. La temperatura del punto triple del agua es de 0.075°C , o 273.16 K .

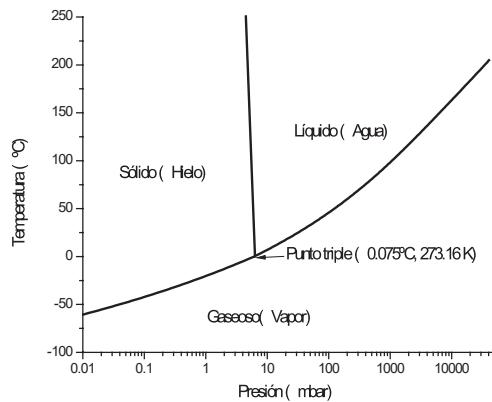


Fig. 2.1 Diagrama de fases del agua.

2.3 DIAGRAMAS BINARIOS

De acuerdo a la regla de las fases, un sistema binario, esto es, de dos componentes, puede tener hasta 3 grados de libertad (para una fase: $F = 2 - P + 2 = 3$). Por ello la diagramación completa de un sistema de este tipo requiere de gráficas en tres dimensiones, donde las coordenadas serían temperatura, presión y composición. La composición del componente A , C_A será la fracción de esta en la mezcla, mientras que la de la componente B será:

$$C_B = 1 - C_A \quad (2.2)$$

Es impráctico trabajar con diagramas tridimensionales, por lo que en realidad los diagramas binarios generalmente se presentan a presión

constante (1 atm). Con esto se sacrifica un grado de libertad y en este caso se grafica la composición contra la temperatura.

En un diagrama binario podremos encontrar regiones de 1 y dos fases, y puntos de tres fases. Las fronteras entre regiones de dos fases pueden ser una temperatura constante o de transición o un punto triple. No se pueden ser contiguas en composición, porque esto significaría que podrían existir tres fases al mismo tiempo, violando la regla de las fases. (El punto triple es invariable si la presión es constante).

En una región de dos fases las condiciones de coexistencia de las mismas están dadas por las composiciones en las fronteras. Si la composición varía, se va a presentar descomposición, como disolución de una de las fases hasta alcanzar el equilibrio. Una vez alcanzado el equilibrio las cantidades que se tengan de cada una de las fases puede variar. Estas se pueden calcular mediante la regla de la palanca.

Una región de dos fases separará a dos regiones de una fase. Sus fronteras se denominarán: *solidus* si delimita una fase sólida de una región sólido + líquido ($\alpha + L$); *liquidus* cuando la línea separe a esta última de una fase líquida; y *solvus* a la frontera entre una fase sólida y una región de dos fases igualmente sólidas, Fig. 2.2.

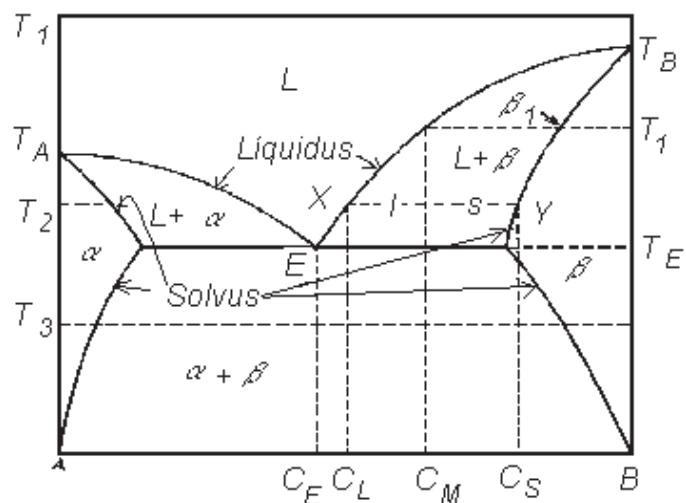


Fig. 2.2 Diagrama binario típico.

2.3.1 Regla de la palanca

La regla de la palanca permite calcular las cantidades respectivas de cada una de las fases de acuerdo a la composición global de la mezcla. Si se tiene una mezcla de dos componentes a una temperatura dada, y la composición corresponde a una región de dos fases, tendremos, en equilibrio, una porción de la fase A y otra de la fase B. En las ecuaciones a continuación se considerarán fases líquida y sólida respectivamente, donde:

W_L : Peso de la fase líquida.

W_{LA} : Peso del componente A en la fase líquida.

W_{LB} : Peso del componente B en la fase líquida.

W_s : Peso de la fase sólida.

W_{SA} : Peso del componente A la fase sólida.

W_{sB} : Peso del componente B en la fase sólida.

W_M : Peso de la mezcla.

C_L : Composición de líquido.

C_s : Composición del sólido.

C_M : Composición de la mezcla

Si se aplican las leyes de conservación de masa:

$$W_s = W_{sA} + W_{sB}$$

$$W_L = W_{LA} + W_{LB}$$

Se tendrá para cada uno de los componentes:

$$W_A = W_{LA} + W_{sA} = W_M (1 - C_M) \quad (2.3)$$

$$W_B = W_{LB} + W_{sB} = W_M C_M \quad (2.4)$$

$$W_{LA} = W_L (1 - C_L) \quad (2.5)$$

$$W_{sA} = W_s (1 - C_s) \quad (2.6)$$

$$W_{LB} = W_L C_L \quad (2.7)$$

$$W_{sB} = W_s C_s \quad (2.8)$$

Reagrupando y resolviendo:

$$W_L (1 - C_L) + W_s (1 - C_s) = W_M (1 - C_M) \quad (2.9)$$

$$W_L C_L + W_s C_s = W_M C_M \quad (2.10)$$

$$\frac{W_L (1 - C_L) + W_s (1 - C_s)}{W_L C_L + W_s C_s} = \frac{W_M (1 - C_M)}{W_M C_M} \quad (2.11)$$

$$\frac{W_L (1 - C_L) + W_s (1 - C_s)}{W_L C_L + W_s C_s} = \frac{(1 - C_M)}{C_M} \quad (2.12)$$

$$[W_L (1 - C_L) + W_s (1 - C_s)] C_M = (1 - C_M) [W_L C_L + W_s C_s] \quad (2.13)$$

$$W_L [(1 - C_L) C_M - ((1 - C_M) C_L)] = W_s [C_s (1 - C_M) - C_M (1 - C_s)] \quad (2.14)$$

$$\frac{W_s}{W_L} = \frac{C_M - C_L C_M - C_L + C_L C_M}{C_s - C_s C_M - C_M + C_s C_M} \quad (2.15)$$

Con lo que se llega a una relación sencilla:

$$\frac{W_s}{W_L} = \frac{C_M - C_L}{C_s - C_M} = \frac{l}{s} \quad (2.16)$$

2.3.2 Diagramas isomorfos

Los diagramas binarios más sencillos son los que muestran una sola fase. Esto se presenta casos cuando los componentes son completamente solubles entre si.

Se ha encontrado que solo es posible si los componentes tienen un radio atómico similar que no difiere en mas del 15%, tienen la misma valencia, estructura cristalina y una diferencia mínima en electronegatividad. Por ello no hay muchos sistemas binarios que presenten este comportamiento. Algunos ejemplos son: Cu-Ni, Ag-Pd, Au-Pt. Es mucho más común el caso de solubilidad limitada.

En un diagrama isomorfo se presentan los puntos de fusión de los componentes A y B, además del comportamiento de las líneas solidus y liquidus. Las zonas de una fase están separadas por la región intermedia de dos fases ($\alpha + L$). Un diagrama isomorfo se puede reproducir si se conocen estos valores.

El sistema Si-Ge, Fig. 2.3, presenta este comportamiento. Ambos tienen elementos tienen propiedades similares, y se puede tener un cristal con cualquier composición de éstos, que cristalizará en la estructura del diamante. En este caso se hablará de aleación, ya que los átomos podrán tomar aleatoriamente cualquier sitio de la red, a diferencia de un compuesto

donde los sitios de la red cristalina son diferentes para cada uno de los elementos.

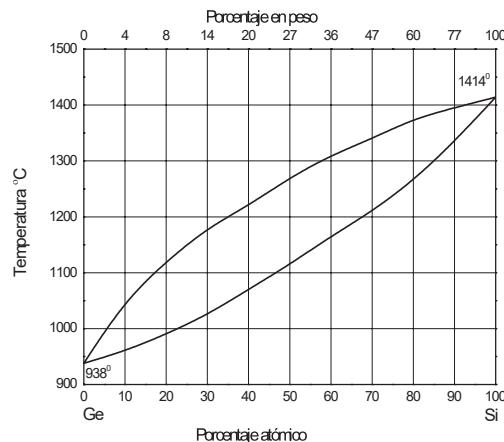


Fig. 2.3 Diagrama Si.-Ge.

Como en el caso del agua, cuando se tienen compuestos estables se pueden dar diagramas de fase entre compuestos. Los semiconductores de la familia III-V presentan varios sistemas isomorfos, donde los componentes serán los semiconductores binarios. Podemos mencionar entre otros a los sistemas AlAs-GaAs y GaAs-GaP, Fig. 2.4. Las propiedades de las mezclas serán, por lo general, valores intermedios y su comportamiento será función de la composición. Por ejemplo el ancho de banda prohibida variará en el caso del sistema GaAs-GaP entre 1.4 y 2.3 eV.

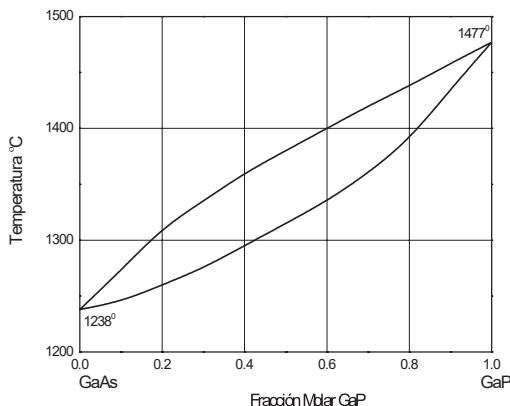


Fig. 2.4 Diagrama de GaP-GaAs.

2.3.3 Sistemas Eutécticos

Un diagrama eutéctico mostrará por un punto triple de dos fases sólidas y una líquida. Se caracterizará por la reacción: $\alpha + \beta \leftrightarrow L$. El punto triple o eutéctico será invariable (asumiendo presión constante) y tendrá temperatura T_E y composición C_E definidas. La temperatura eutéctica será inferior a los puntos de fusión de los componentes A y B (T_A y T_B). Los sistemas eutécticos se presentan generalmente cuando hay solubilidad completa entre los componentes en la fase líquida, pero presentan fases sólidas diferentes, con solubilidad parcial entre si. En estos sistemas no hay formación de compuestos químicos.

Se puede caracterizar un diagrama eutéctico por el punto triple o eutéctico, con su composición y temperatura. Asimismo se deben determinar los límites de solubilidad de los componentes en las fases sólidas, los cuales están dados por las líneas solidus y solvus. La máxima solubilidad de un componente en otro generalmente se da a la temperatura del eutéctico.

El sistema Pb-Sn, Fig. 2.5, es un eutéctico clásico con aplicación en la industria electrónica, dada su importancia en soldaduras. Cuando se requiere un punto de fusión bajo, generalmente para proteger dispositivos y tablillas de circuito, se busca que la composición sea la eutéctica. Si por el contrario, como es el caso de los dispositivos de potencia, lo que se quiere es una mayor resistencia a la temperatura, lo que se utilizan son soldaduras ricas en plomo. Los puntos importantes en un diagrama Pb-Sn se muestran en la Tabla 2.1:

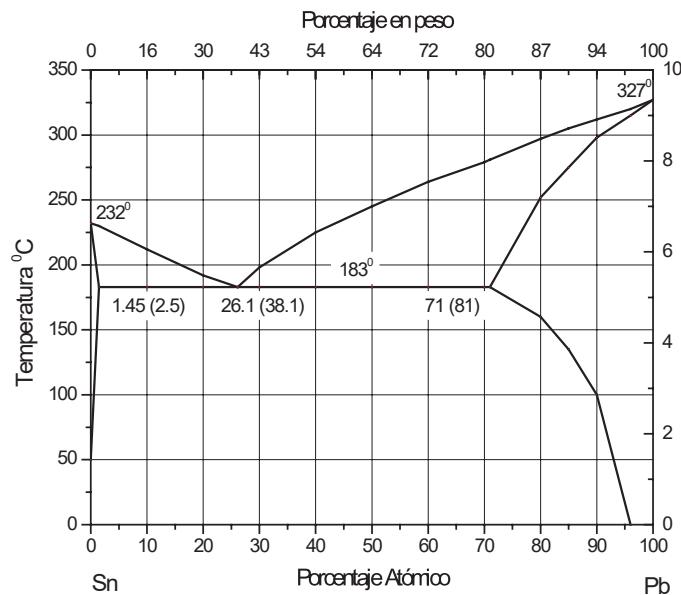


Fig. 2.5 Diagrama Pb-Sn.

En algunos sistemas eutécticos la solubilidad de un componente en otro es tan pequeña que la fase sólida solo se ve como una línea. Este es el caso del Al en Si (menor al 0.1 %), Fig. 2.6. La solubilidad del silicio en

aluminio también es reducida (1.59% at.). Este es un sistema de interés en procesos tecnológicos ya que el aluminio se utiliza en metalizaciones, generalmente esta se realiza después de los procesos de alta temperatura y queda aluminio en contacto con el silicio La película de aluminio es por lo general muy delgada en comparación del espesor de la oblea.

Tabla 2.1 Puntos importantes de un diagrama de fase de Pb-Sn

Punto de fusión del estaño	232 °C
Punto de fusión del plomo	327 °C
Temperatura del eutéctico	183 °C
Composición del eutéctico (atómica)	26.1%
Composición del eutéctico (en peso)	38.1%
Solubilidad máxima de plomo en estaño (atómica)	1.45% a T _E
Solubilidad máxima de plomo en estaño (en peso)	2.5% a T _E
Solubilidad máxima de estaño en plomo (atómica)	71% a T _E
Solubilidad máxima de estaño en plomo (en peso)	81 % a T _E

El diagrama de fases nos servirá para la prevenir fenómenos indeseables como la aleación del silicio con aluminio o favorecer reacciones intencionales, al impurificar una porción de silicio: Si se eleva la temperatura sobre el punto eutéctico se formará una fase líquida. Esto implica fusión tanto del metal, como de una parte de la oblea. Esto puede implicar la destrucción de una estructura superficial. Al solidificar se formará un aglomerado de ambas fases, tanto la del metal como la del semiconductor. La fusión del metal puede favorecer la adherencia, pero dadas las condiciones, se sacrifica la homogeneidad. Si el cristal (silicio)

tiene defectos o dislocaciones, la fusión se puede favorecer alrededor de éstos y en algunos casos el metal puede penetrar más profundamente en algunos puntos.

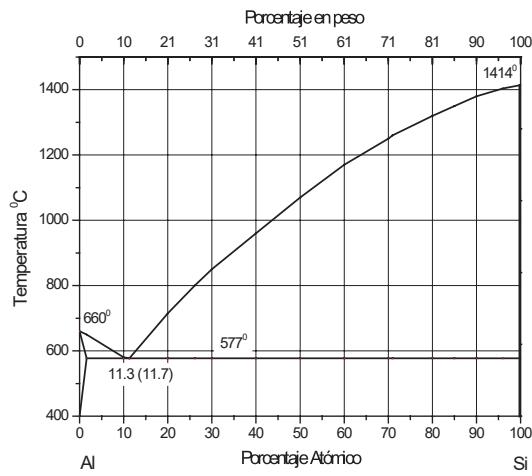


Fig.2. 6. Diagrama Al-Si

En el sistema oro silicio, Fig. 2.7, las solubilidades en ambos casos son tan reducidas que las dos fases sólidas se representan por solo una línea. En este caso la temperatura del eutéctico es muy inferior a las de los componentes, y se deben de tomar consideraciones similares a las expuestas para el aluminio.

En algunos casos las metalizaciones se realizan a partir de aleaciones, el diagrama Au-Ge, Fig. 2.8, nos muestra un eutéctico a 12% en peso de germanio y 363 °C. Cuando este sistema de utiliza para la formación de contactos óhmicos en GaAs es necesario fundir una película superficial del semiconductor, para que recristalice altamente impurificada con germanio, (Así la resistividad del contacto se reduce). Es por ello que se utiliza

precisamente esta composición y de esta manera reducir el punto de fusión de la mezcla.

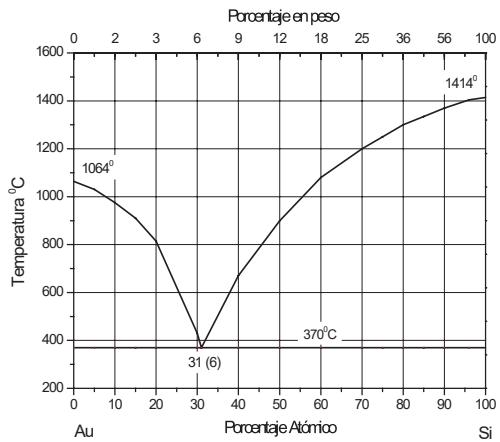


Fig. 2.7 Diagrama Au-Si.

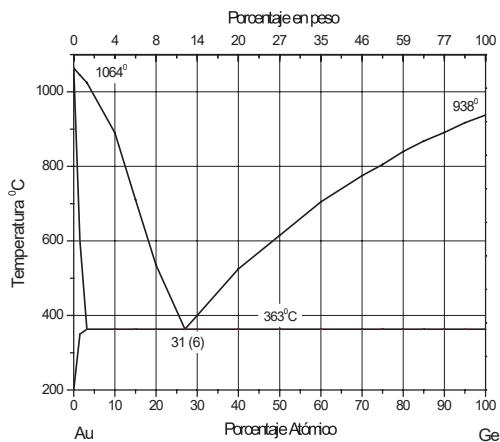


Fig. 2.8 Diagrama Au-Ge.

En algunos casos el punto eutéctico está tan cercano al punto de fusión de alguno de los componentes, que no se nota la depresión en temperatura ni se detecta la composición del punto triple. Un ejemplo de este caso lo muestra el sistema Ge-Ga, Fig. 2.9.

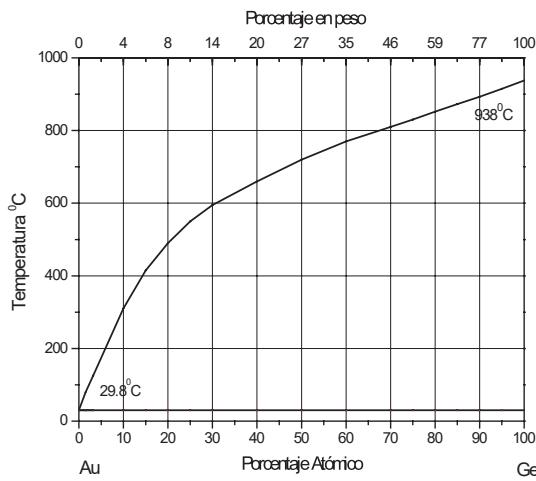


Fig. 2.9 Diagrama Ga-Ge.

2.3.4 Transformaciones congruentes

En los diagramas eutécticos, no hay formación de compuestos, y las fases corresponden a los componentes del sistema. Cuando se forma un compuesto, como el caso del GaAs, aparece una nueva fase. Una de las conveniencias de que los diagramas de fase se realicen en notación atómica, es que permite identificar perfectamente las posiciones de los compuestos, así si las cantidades de los componentes son idénticas, el compuesto de localizará al centro del diagrama. Si tenemos una sustancia A_2B_3 , este tendrá una composición de 60% (B).

Si la fase se presenta como una sola línea hasta su punto de fusión se le llama “*Transformación Congruente*”. Una transformación congruente

puede dividir el diagrama de fases en dos partes, donde se pueden identificar subdiagramas con comportamiento eutéctico, o de algún otro tipo.

El sistema galio-arsénico, Fig. 2.10, presenta este comportamiento y se caracteriza por los parámetros que se muestran en la Tabla 2.2:

Tabla 2.2 Parámetros de un sistema Ga-As

Punto de fusión de galio	29.5 °C
Punto de fusión de GaAs	1238 °C
Punto de fusión de As	810 °C
No se detectan solubilidades entre las fases sólidas	

La parte izquierda del diagrama, será la región rica en galio y será independiente de la región rica en arsénico. Esto es el diagrama se puede dividir en dos partes iguales.

El conocimiento de este diagrama de fase es importante en procesos como el de epitaxia en fase líquida, donde se ponen en contacto un sustrato con un líquido, ambos en equilibrio. Al disminuir la temperatura se genera una sobresaturación en el líquido tendiendo a expulsar el excedente. La forma de hacerlo con la menor energía es en los sitios de la red cristalina del sustrato, con lo que se obtiene una capa cristalina. Dado que la temperatura de obtención del cristal es inferior al punto de fusión del GaAs, el número de defectos que tenga esa capa será también menor.

Una ampliación de la zona de fusión del GaAs, Fig. 2.11, muestra que el sistema si permite desviaciones de la estequiometría que aunque sean muy pequeñas pueden afectar el punto de fusión del semiconductor. Estas desviaciones permitirán la incorporación de impurezas anfotéricas en una u otra subred del semiconductor. Aquí tampoco se ha tomado en consideración la presión, sin embargo en procesos reales será la presión

parcial de arsénico la que determine las vacancias el tipo de estas, así como la desviación de la estequiométría.

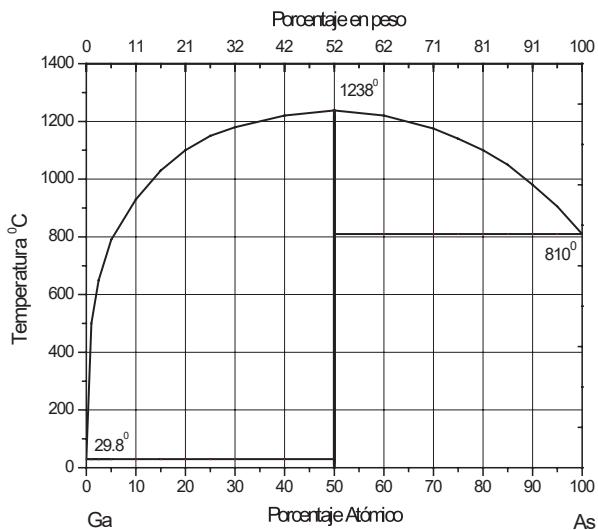


Fig. 2.10 Diagrama galio-arsénico.

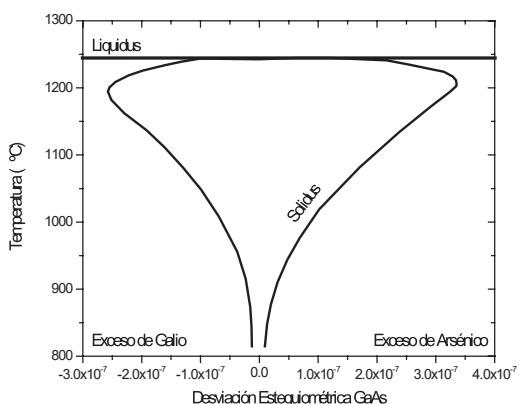


Fig. 2.11 Detalles del diagrama Ga-As.

2.3.5 Reacciones peritécticas

Un compuesto binario puede descomponerse antes de llegar a su punto de fusión, dando a un líquido mas un sólido ($\alpha \leftrightarrow \beta + L$). A esta se le llamará reacción peritéctica. Un ejemplo de este comportamiento lo muestra el sistema Si-As, donde el compuesto SiAs_2 se descompone a 944 °C formando SiAs y líquido. El diagrama silicio-arsénico muestra además dos eutécticos, Fig. 2.12.

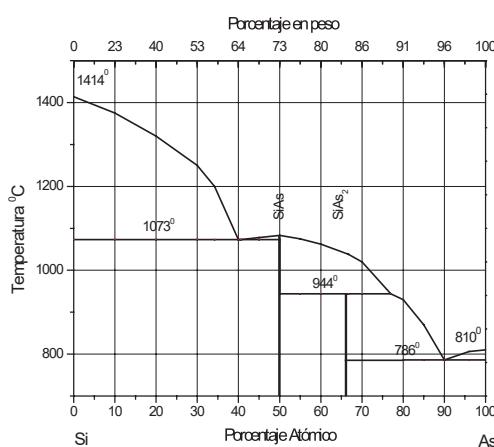


Fig. 2.12 Diagrama de Si-As.

En el ejemplo de la Fig. 2.12, al enfriar un líquido con más del 66% de arsénico, inicialmente se formará SiAs , sin embargo al llegar a los 944 °C, se comenzará a formar SiAs_2 . Esto es el líquido ya no puede estar en equilibrio con el SiAs . Los granos de este material se recubrirán de SiAs_2 , formando una película que prevendrá su descomposición.

2.3.6 Otros ejemplos

El sistema aluminio-oro es de gran importancia en los procesos tecnológicos de semiconductores. Ambos son ampliamente usados en

metalizaciones y en algunos casos se les combina. Estos dos metales forman varios compuestos, todos con diferentes propiedades, Fig. 2.13. El AuAl_2 es de color púrpura, de fuerte adherencia, altamente conductor. El Au_2Al tiene color bronce, es quebradizo y tiene mala conductividad. La presencia de este último es indeseable, ya que propicia la falla de los circuitos. Aún cuando su punto de fusión es de 624 °C, su formación puede realizarse a en reacciones en el sólido a temperaturas del orden de 300 °C. A este efecto se le llama “*plaga púrpura*” debido a que en el proceso también se forma AuAl_2 , y cuya presencia oculta al causante de los problemas.

Los diagramas de fase son importantes en la síntesis de siliciuros. Estos son compuestos refractarios con buena conductividad eléctrica y tienen la ventaja de que pueden ser sometidos a altas temperaturas sin su destrucción. En algunos casos se pueden usar como mascarillas, para procesos autoalineados.

El sistema platino-silicio presenta tres compuestos y tres eutécticos, Fig. 2.14. En este caso se busca la formación del compuesto Pt-Si, la cuál, como se ha determinado experimentalmente se realiza a 600 °C. El silicio de platino sirve también como capa adherente para posteriores metalizaciones.

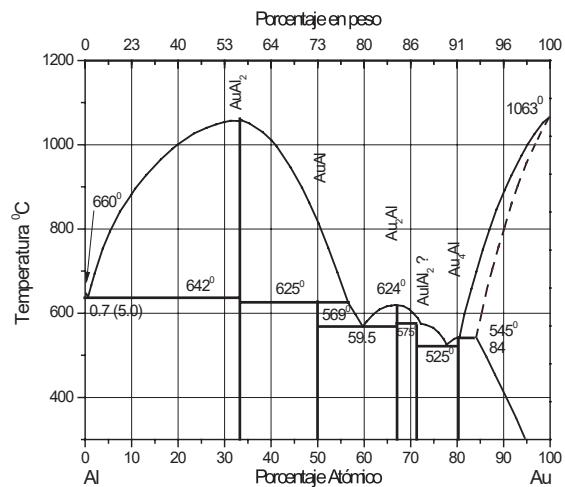


Fig. 2.13 Diagrama Al-Au.

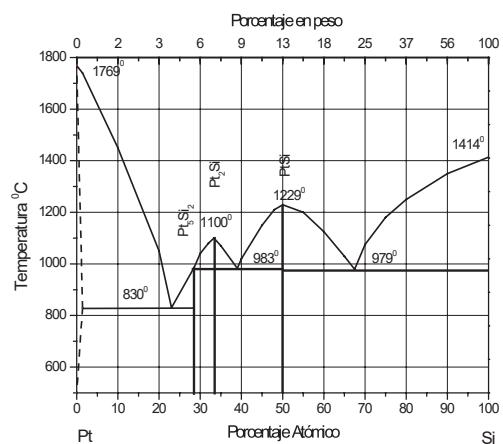


Fig. 2.14 Diagrama Pt-Si.

2.3.7 Diagramas de sistemas de óxidos

Los sistemas de óxidos son especialmente importantes en la tecnología de circuitos integrados. Algunas veces se utilizan los óxidos como fuentes de impurificantes, otras veces se utilizan las combinaciones de óxidos para películas de protección o mascarillas.

La figura 2.15 muestra el diagrama de fases de óxido de silicio con óxido de boro. El sistema conocido como Borosilicato tiene diversas aplicaciones como la elaboración de recipientes refractarios. En microelectrónica la estabilidad de la mezcla es la que permite el uso de óxido de silicio como mascarilla.

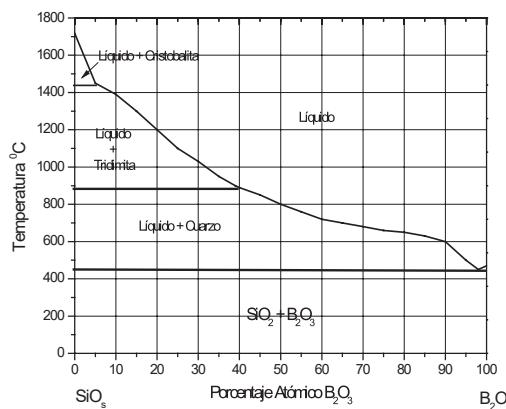


Fig. 2.15 Sistema $B_2O_3-SiO_2$.

Además de los sistemas eutécticos y peritécticos hay otras reacciones similares, aunque son menos comunes. En resumen:

- | | |
|-----------------|---|
| 1.- Monotéctica | $\alpha + L_2 \leftrightarrow L_1$ |
| 2.- Eutéctica | $\alpha + \beta \leftrightarrow L$ |
| 3.- Eutectoide | $\alpha + \beta \leftrightarrow \gamma$ |

4.- Sintéctica	$\beta \leftrightarrow L_1 + L_2$
5.- Peritéctica	$\beta \leftrightarrow L + \alpha$
6.- Perictectoide	$\beta \leftrightarrow \alpha + \gamma$

En estas reacciones se muestra la parte de menor temperatura al lado izquierdo y la de mayor temperatura al derecho.

2.3.8 Solubilidad Sólida

Con excepción de los sistemas isomorfos, hay un límite de la cantidad máxima de un componente que puede aceptar contener una fase determinada. A este se le llama solubilidad. Para que dos fases diferentes estén en contacto se requiere que en ambos casos se tenga el límite máximo de solubilidad. Se llamará solvente al componente o sustancia dominante, y el soluto será la que se le incorpore. El solvente puede ser sólido o líquido. En el primer caso se hablará de solubilidad sólida

La solubilidad depende fuertemente de la temperatura y generalmente se incrementa con ésta, sin embargo, en el caso de los sólidos cuando se acercan al punto de fusión esta disminuye. A esta característica se le llamará *solubilidad retrograda*, o sea cuando la solubilidad disminuye con la temperatura. Este efecto lo vemos en los sistemas eutécticos como el Pb-Sn (Fig 2.5) .

En algunos casos, como es el caso de las impurezas en el silicio, la solubilidad no es visible en un diagrama de fases convencional, para esto hay que ampliar el mismo, Fig. 2.16. En procesos como difusión se aprovechan los valores máximos de solubilidad para lograr una reproducibilidad en la impurificación. Hay que hacer notar que la incorporación de impurezas estará determinada por la temperatura del proceso y al enfriar a valores ambientales los cambios en la estructura hacia

los nuevos valores de equilibrio impuestos por la termodinámica son tan lentos, por lo que se pueden despreciar.

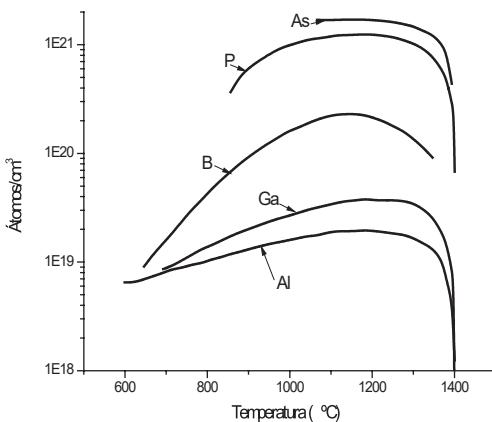


Fig. 2.16 Solubilidad de impurezas en silicio.

2.4 DIAGRAMAS TERNARIOS

Los diagramas de tres componentes son más complicados y se les ha estudiado en mucho menor grado que los diagramas binarios. Algunas veces se extrapolan los resultados obtenidos de los diagramas binarios para determinar el sistema ternario. La principal razón es la dificultad experimental de determinar los diagramas de fases, dado que en la mayor parte de los casos esto se hace a partir de datos experimentales. El desarrollo de computadoras de alta capacidad ha sido un factor clave para el cálculo de los diagramas a partir de consideraciones teóricas.

El grafico de los diagramas de fase se complica debido a que la especificación de la composición debe de incluir los porcentajes de cada uno de los componentes. De tres componentes es necesario conocer las

fracciones de dos de ellos, ya que: $X_A + X_B + X_C = 1$. Por ello para determinar un diagrama de fases ternario se requerirán al menos dos fracciones, la temperatura y la presión. Esto es de acuerdo a la regla de las fases el diagrama podrá tener hasta 4 grados de libertad, o variables que deban ser definidas.

En el caso de graficar composición y temperatura a presión constante, el diagrama será un prisma con base triangular, donde cada una de sus caras será el diagrama binario regular de los componentes correspondientes.

Debido a la dificultad de trabajar con gráficas en tres dimensiones, la solución empleada es realizando cortes a temperatura constante o de secciones isotérmicas. Para graficar los sistemas ternarios se usa como base un triángulo equilátero. Este permite la diagramación de la composición especificando solo dos componentes. A continuación algunas reglas para facilitar su comprensión y calcular las composiciones:

- 1.- Cada vértice del triángulo equilátero representa 100% del componente correspondiente
- 2.- Cada lado del triángulo corresponderá a un sistema binario, donde no habrá componentes de lado opuesto.
- 3.- Cada línea paralela al lado opuesto a un vértice corresponderá a concentración constante del componente en cuestión.
- 4.- Cada línea recta que llegue a un vértice representará una relación constante entre los otros dos componentes.

La figura 2.17 muestra algunos ejemplos del cálculo de la composición

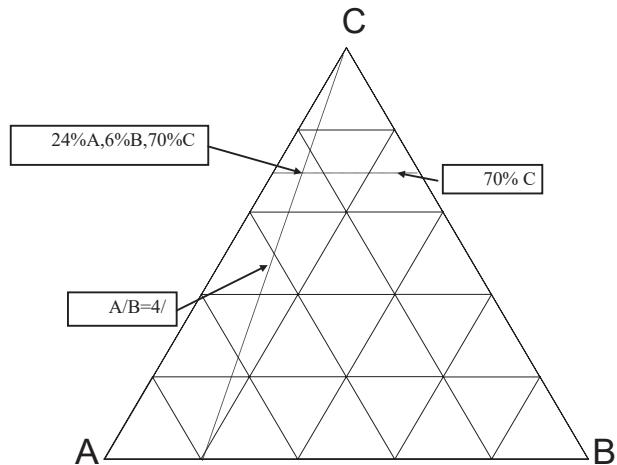


Fig. 2.17 Ejemplos de composición.

Es común sobreponer los cortes isotérmicos de los diagramas de fase en un solo triángulo. En estos casos se marcarán las líneas de líquidos, sólidos o solvus con la temperatura a la que corresponden. Los puntos especiales, como eutécticos, peritécticos u otros se marcarán junto con la temperatura a la que correspondan.

La figura 2.18 muestra las líneas de liquidus del sistema Ga-As-Au para diferentes temperaturas. Se puede identificar una fase sólida del GaAs para todos los cortes y la del Au para los cortes a temperaturas menores a 1000°C, ya que los puntos de fusión son de 1238°C y 1063°C respectivamente. Las regiones de dos fases son aquellas que se encuentran entre la línea de liquidus y la fase sólida, quedando la fase líquida al centro de los diagramas.

En la figura 2.19 se muestra la sobreposición de los diferentes cortes. En el sistema Ga-As-Au no se detectan fases adicionales a la del GaAs y los elementos del sistema. Además en el área reportada se identifica una

fase líquida, lo que debe tomarse en consideración si se van a realizar metalizaciones basadas en oro sobre el semiconductor.

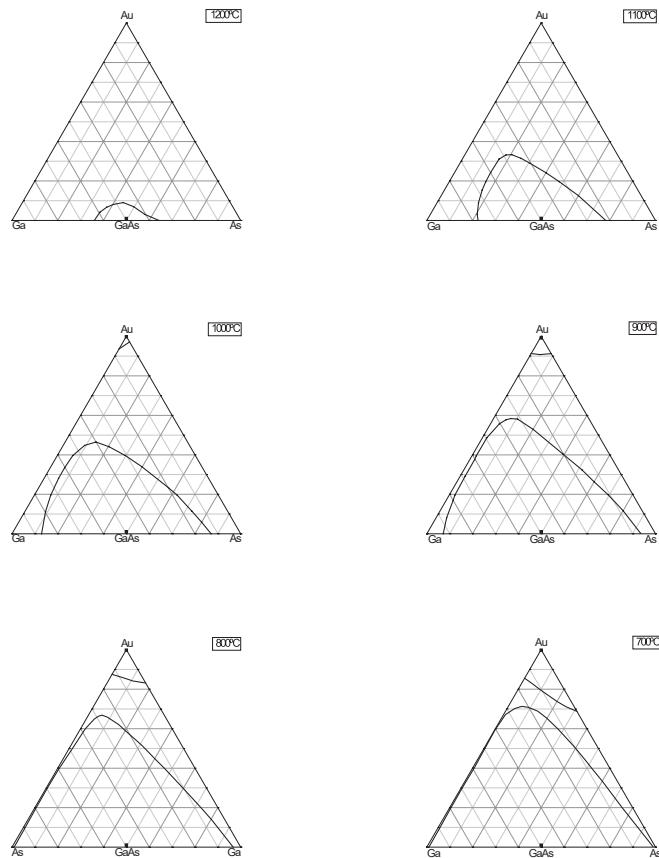


Fig. 2.17 Cortes isotérmicos del sistema Ga-As-Au.

La figura 2.20 muestra el diagrama ternario del sistema Ga-As-S. En este se marca la existencia de diferentes compuestos, así como puntos eutécticos. De igual forma se debe considerar en la tecnología de los semiconductores la posible formación de los mismos o la eventual liquefacción al subir la temperatura con la presencia de los tres elementos.

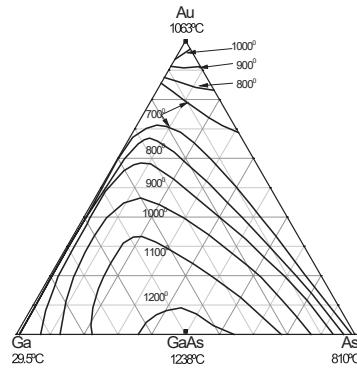


Fig. 2.19 Diagrama ternario del sistema Ga-As-Au.

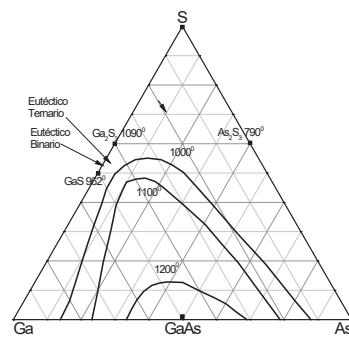


Fig. 2.20 Diagrama ternario del sistema Ga-As-S.

Como consecuencia de la regla de las fases, los sistemas ternarios deben cumplir una serie de normas:

1.- Las regiones de cuatro fases tienen 4 fronteras: dos de ellas estarán opuestas y representarán fases sencillas, las otras dos podrán ser lados del triángulo, o bien fronteras con regiones de tres fases. En el caso de los

sistemas de Ga-As son diferentes fronteras las corren a lo largo de la línea GaAs y que unen la fase del GaAs con la línea de líquidos hacia el lado del Ga, y la línea que corre hacia el lado del arsénico. Además, aunque la fase del GaAs sea prácticamente puntual, esta se considerará una fase adicional

2.- Las regiones de tres fases son triangulares. Cada uno de los lados colindará con una región de dos fases, mientras que cada vértice será una región de una fase. Las composiciones de cada una de las tres fases serán invariables.

3.- Es posible encontrar cuatro fases en equilibrio, esto se observa en el corte correspondiente a T_5 , donde hay un eutéctico ternario en equilibrio con 3 fases sólidas.

En las regiones de dos fases se tendrá un rango de composiciones para cada una de estas, pero cada composición de una de las fases corresponderá a otra definida para la otra fase. Estrictamente estas correspondencias deben de ser determinadas experimentalmente, sin embargo, dado que se espera una variación gradual se pueden aproximar las mismas por regla un cálculo proporcional.

2.5 BIBLIOGRAFÍA

- [1] Sorab K. Ghandhi, “VLSI fabrication principles: silicon and gallium arsenide”, New York : Wiley, c1983.
- [2] Max Hansen, Kurt Anderko, Constitution of binary alloys / . 2nd ed. New York : McGraw-Hill Book, c1958.
- [3] Constitution Binary alloys, first supplement, Rodney P. Elliot, Max Hansen, New York: McGraw-Hill, c1965..
- [4] Francis A. Shunk, Max Hansen, Constitution Binary alloys, second supplement, New York : McGraw-Hill, 1969.

CAPÍTULO 3

CRECIMIENTO DE CRISTALES, (LINGOTES)

3.1 INTRODUCCIÓN

Los semiconductores son materiales cuyas propiedades son extremadamente sensibles a los defectos del material. La simple introducción de impurezas en rangos del orden de una parte en mil millones puede implicar cambios en la resistividad eléctrica de varios órdenes de magnitud. Los defectos cristalinos y las fronteras de grano modifican la movilidad de los portadores, y los tiempos de vida. La compensación eléctrica del material tiene repercusiones negativas en la frecuencia máxima de operación de los dispositivos o del ruido que estos generan.

El material con la pureza adecuada para la fabricación de dispositivos deberá tener “*pureza electrónica*”, esto se aplica tanto al cristal semiconductor, como a los reactivos empleados y materiales empleados como precursores. El “*Silicio de Grado Electrónico*” (EGS: Electron Grade Silicon) solo podrá tener algunas partes por mil millones o ppb (parts per billion) ($1 \text{ ppb} = 1/10^9$) de impurezas activas, las impurezas inactivas no deberán superar las partes por millón.

La determinación de la perfección de un material generalmente no se realiza de forma directa, ya que los sistemas para hacerlo son costosos y tardados, sino a través de mediciones de resistividad y movilidad. Un sistema con muchas impurezas activas tendrá una resistividad baja. Además la movilidad de los portadores es también afectada por las impurezas inactivas y otros centros de dispersión.

3.2 SÍNTESIS DE SILICIO

Los primeros transistores comerciales se hicieron a base de germanio. Sin embargo este material fue substituido por el silicio una vez que fue posible la obtención de material monocrystalino de alta calidad. Sus principales ventajas radican en el mayor ancho de banda, lo que resulta en diodos de menor corriente de fuga, y a las propiedades del óxido de silicio, como material fundamental para el desarrollo de la tecnología planar.

El silicio existe en gran abundancia en la corteza terrestre, ocupando el segundo lugar en peso después de oxígeno. Generalmente se encuentra en forma de silicatos, o sea óxidos de silicio en conjunto con otros materiales. Los silicatos tienen varias aplicaciones, en la fabricación de vidrio, el uso de dieléctricos (mica), en joyería (jade, zirconia, cuarzo), o el empleo de cristales de cuarzo para relojería y como oscilador a frecuencia estable.

El primer paso hacia la realización de obleas de alta calidad es la obtención de silicio metalúrgico (MGS), ya que, a pesar de su abundancia, el silicio puro no se encuentra en la naturaleza, sino en forma de óxidos. Generalmente se selecciona con material de partida para la síntesis alguna arena de dióxido de silicio con la menor cantidad posible de impurezas.

Para separar el oxígeno de la molécula es necesario romper los enlaces. Para ello se funde la arena en presencia de partículas de carbono o madera, a la vez que se le aplica un campo eléctrico intenso para formar una descarga (el dióxido de silicio es aislante). El arco disocia la molécula, y el oxígeno liberado puede reaccionar, ya sea con el silicio de nuevo o con el carbono. En este último caso el monóxido de carbono escapa hacia el exterior, con lo que la mezcla va perdiendo el oxígeno. El silicio puro es

más pesado y se deposita en el fondo del crisol. Este último es drenado para obtener silicio metalúrgico (MGS), Fig. 3.1.

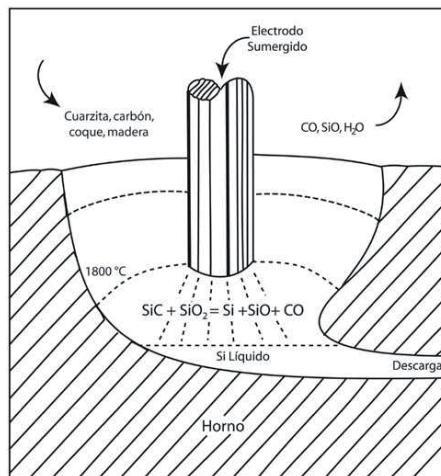
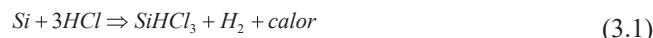


Fig. 3.1 Sistema para producción de silicio metalúrgico.

El costo de obtención del silicio es alto debido a la cantidad de energía necesaria para el proceso, que es del orden de 13 kWh/kg. Una planta de silicio metalúrgico bien puede necesitar la energía eléctrica de una comunidad pequeña.

El silicio metalúrgico tiene una pureza del orden del 98% y debe ser purificado para poder ser útil para las aplicaciones en microelectrónica. La purificación se simplifica si se hace reaccionar el silicio con ácido clorhídrico para formar triclorosilano (SiHCl_3):



En la reacción se pueden formar otros compuestos como tetracloruro de silicio o cloruros de las impurezas. El triclorosilano es líquido a temperatura ambiente siendo su punto de ebullición de 32 °C. Este se puede purificar mediante destilación fraccionada. Esta técnica consiste en calentar

el líquido con control estricto de temperatura condensando y recogiendo el gas que se produce a un valor específico de temperatura. Si este proceso se realiza varias veces sucesivamente se incrementa la pureza del líquido recolectado.

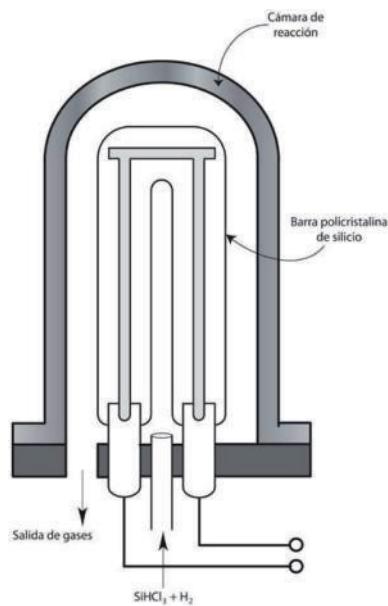


Fig. 3.2 Sistema de descomposición del triclorosilano.

Una vez purificado el triclorosilano es necesario descomponerlo para sintetizar el silicio. La reacción de la descomposición sería:



Esta se realiza en un sistema como el mostrado en la figura. El triclorosilano mezclado con hidrógeno se introduce al reactor, el cual es calentado eléctricamente. El silicio inicial sirve como centro de nucleación, esto es, una superficie donde se favorece la condensación del silicio. El proceso puede durar varias horas para realizarse. El material obtenido es de

pureza electrónica pero todavía no puede ser usado para dispositivos debido a ser policristalino.

3.3 SÍNTESIS DE GaAs

Para sintetizar el arseniuro de galio se parte de galio y arsénico puros. Dado que la reacción de estos elementos al punto de fusión del GaAs (1238 °C) es exotérmica y violenta esta se debe de realizar de manera controlada. Además es necesario considerar las diferencias de presión de vapor de los dos elementos para realizar la síntesis.

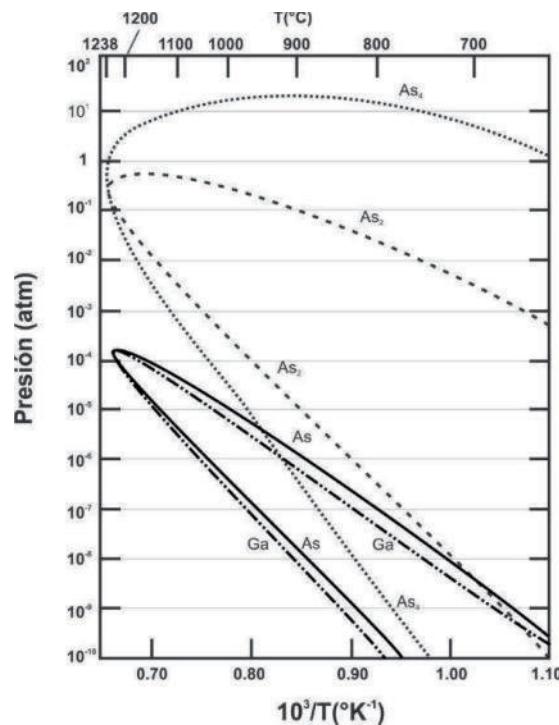


Fig. 3.3 Presión de vapor de As y Ga sobre GaAs.

La figura 3.3 muestra las presiones de vapor de diferentes moléculas sobre GaAs, As y Ga. La presión de vapor del galio al punto de fusión del GaAs (1238°C) es del orden de 10^{-4} atmósferas, esto es, la evaporación del metal es lenta y no implica problemas tecnológicos o precauciones especiales para su manejo a altas temperaturas. Por el contrario, la presión de vapor del arsénico puro a 1238°C puede ser superior a las 50 atm y de 1 atm sobre arseniuro de galio. Esto es: si se calienta el elemento a este valor, antes de reaccionar, la cámara deberá estar debidamente presurizada.

Un método para la síntesis de GaAs que toma en consideración los argumentos anteriores se muestra en la figura 3.4 y se describe a continuación.

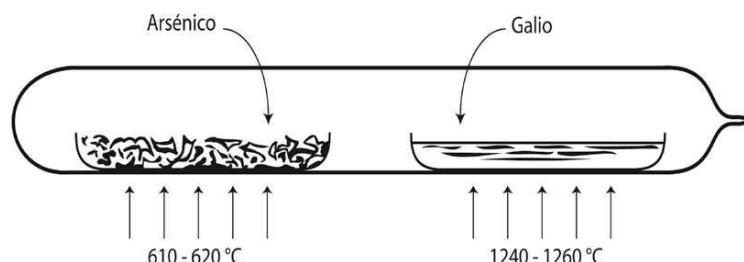


Fig. 3.4 Sistema para la síntesis de GaAs.

El sistema consiste en una ampolla de cuarzo sellada donde se encuentran dos botes conteniendo uno arsénico y otro galio. Los botes pueden ser de cuarzo o de grafito cubierto de algún material inerte, como carburo de silicio, nitruro de boro o grafito pirolítico. El recubrimiento tiene como función prevenir la contaminación del GaAs.

La ampolla se introduce en un horno con dos zonas de temperatura. La temperatura de la fuente de galio se eleva ligeramente arriba del punto de fusión del GaAs ($1240 - 1260^{\circ}\text{C}$), mientras que la del bote de arsénico

se ajusta para producir una presión de vapor de una atmósfera (620 -640 °C). Al evaporarse el arsénico ocupa todo el espacio del bote y es capturado por el galio para formar GaAs.

La síntesis siempre deberá realizarse sobre el punto de fusión del GaAs, ya que de lo contrario, al solidificarse formará una costra sobre el galio que actuará como barrera para la reacción entre los dos elementos.

Con este método el transporte de arsénico se realiza de manera lenta y controlada por medio de la fase gaseosa. La reacción puede tomar varias horas por kilogramo de arseniuro de galio.

La temperatura de la fuente de arsénico será determina la presión parcial del elemento. Al inicio del proceso ésta se podrá incrementar a un valor mayor para favorecer el transporte, tomando en consideración la resistencia de la ampolla a la sobrepresión y el control de la reacción con el galio. Posteriormente deberá controlarse a valores muy precisos para obtener el material con la mayor estequiometría

Si bien, las consideraciones teóricas son de gran ayuda en el diseño de los procesos, la optimización de los parámetros se realiza por medio de prueba y error. Esto debido a que generalmente los modelos no consideran todos los parámetros involucrados en el proceso, muchos de los cuales, como la distribución de la temperatura al interior de la ampolla, son dependientes de la geometría del sistema. El material obtenido por éste método es, similarmente que con el silicio EGS, de alta pureza, aunque policristalino e inadecuado para la fabricación de dispositivos.

3.4 MÉTODO BRIDGMAN

El estado cristalino es el de menor energía, y por lo tanto el más estable. La policristalinidad en la síntesis de materiales se presenta debido a la

formación de varios centros de nucleación⁵ independientes, lo que provoca que el material se solidifique alrededor de éstos. Un recocido a altas temperaturas tenderá a corregir el material reduciendo los defectos, sin embargo los procesos son lentos y por lo tanto no son prácticos. La obtención de monocristales se realiza fundiendo el material policristalino y solidificarlo de manera controlada induciendo la nucleación alrededor de una semilla. Un método sencillo es el Bridgman, el cual es usado para la obtención de lingotes de GaAs [1]

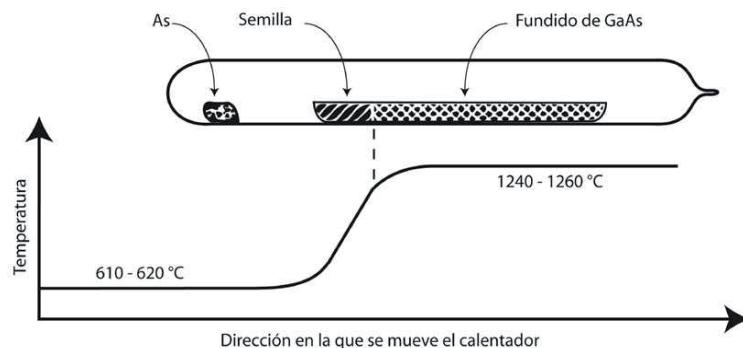


Fig. 3.5 Método Bridgman para la obtención de monocristales de GaAs.

El sistema Bridgman recuerda al de síntesis de GaAs, dado que también se realiza en una ampolla de cuarzo cerrada con dos zonas de temperatura. Una de ellas a 610-620 °C para estabilizar la presión de arsénico y la otra a 1240-1260 °C por encima de la fusión del GaAs. Las diferencias con el de síntesis son:

⁵ Centro de nucleación: al condensarse la materia lo hace preferentemente alrededor de núcleos ya existentes. Un centro de nucleación puede ser un cristal introducido intencionalmente al sistema para favorecer la cristalización, puede ser un núcleo formado espontáneamente debido a sobresaturación o perturbaciones del medio, o bien impurezas o partículas presentes en el sistema. Al condensarse los átomos tenderán a hacerlo siguiendo la dirección cristalina del centro de nucleación.

- En el crisol de alta temperatura se encuentra GaAs fundido.- A baja temperatura sólo se tiene una cantidad pequeña de arsénico para estabilizar la presión. No se usa como fuente de material.
- Parte del crisol de GaAs atraviesa la zona en donde comienza a bajar la temperatura. Aquí se encuentra una interfaz líquido-sólido.
- En el extremo del crisol se encuentra un monocrystal (semilla), que actúa como centro de nucleación.
- Ya sea el tubo de cuarzo, o el horno, se mueven para que la interfaz líquido-sólido se recorra hacia la derecha.

El proceso en sí consiste en introducir los componentes (bote con semilla y GaAs policristalino, y arsénico para estabilizar) en una ampolla de cuarzo cerrada en un horno. La semilla, o sea un pequeño monocrystal se encuentra en un extremo. Se debe tener cuidado que al llegar el bote a la zona de alta temperatura se funda el arseniuro de galio sin que lo haga la semilla, por ello ésta debe quedar en la zona de transición de temperatura.

Una vez estabilizado el sistema, se comienza a recorrer el gradiente de temperatura, ya sea moviendo el horno o la misma ampolla de cuarzo, de tal forma que se enfrie el extremo donde está la semilla. Al enfriarse el fundido solidifica en "*centros de nucleación*", o sea superficies, donde por sus características se favorece la condensación. Esto se da de manera preferencial sobre la superficie de la semilla, la cuál determinará la orientación cristalográfica del cristal.

Es posible la combinación de síntesis y crecimiento Bridgman en un mismo reactor, en este caso se introducen cantidades grandes de arsénico y el movimiento del gradiente de temperatura se realiza una vez que se ha completado la síntesis.

Las principales desventajas del método Bridgman radican en el hecho de que el fundido, el lingote y la interfaz están determinados por el crisol. Se tienen problemas de contaminación por las paredes del crisol, mojado o adherencia del lingote a las paredes, tensiones y generación de dislocaciones causados por la diferencia en los coeficientes de expansión, también hay que considerar el hecho de que los lingotes no son cilíndricos, sino que toman la forma del crisol, por lo que tienen que ser rectificados para la obtención de obleas redondas que sean adecuadas para producción, con la consiguiente pérdida de material.

3.5 MÉTODO CZOCHRALSKI

En el método Czochralski la interfaz no está en contacto con el contenedor, sino que “flota” en el centro del fundido. Además los lingotes son prácticamente cilíndricos y la cantidad de material que se pierde al rectificarlos es menor.

La figura 3.6 muestra un esquema de un sistema de crecimiento Czochralski [2]. El crisol, generalmente de cuarzo, contiene el material (silicio) y se encuentra dentro de una funda de grafito, cuya función es homogeneizar la temperatura.

El calentamiento debe ser local, dado que se requieren perfiles de temperatura abruptos y controlados para no fundir el lingote sólido y mantener la fase líquida en el crisol. Esto se puede realizar por medio de: a) Inducción de radiofrecuencia para calentar la funda (susceptor) de grafito; b) Calentamiento resistivo, haciendo pasar directamente una corriente por el susceptor, para lo que es necesario diseñar el mismo en forma de filamento, o; c) Radiación de lámparas luminosas de potencia que son enfocadas sobre el grafito.

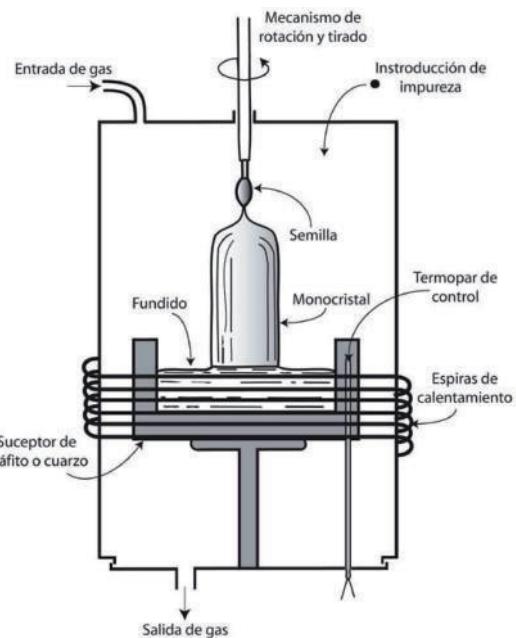


Fig. 3.6 Dibujo esquemático de los componentes de un sistema Czochralski para el crecimiento de lingotes monocristalinos.

El control de temperatura se realiza por medio de termopares que se introducen en el interior del susceptor. La señal recolectada retroalimenta al sistema de control de los calefactores. La temperatura medida difiere de la que se encuentra en la interfaz, debido a los gradientes propios del sistema.

El sistema está cerrado, las paredes pueden ser de cuarzo, las partes que no se calientan, de algún metal. Se hace fluir algún gas inerte como nitrógeno, helio o argón durante todo el proceso, con el fin de impedir la presencia de oxígeno. Por la parte superior entra una barra en cuyo extremo se ha colocado pedazo de cristal, o semilla.

Para iniciar el proceso se calienta el sistema, primero para degasar, o sea eliminar gases atrapados tanto en el silicio como en crisol y susceptor. Posteriormente se incrementa la temperatura para fundir el silicio,

estabilizándose cerca de su punto de fusión. Antes de iniciar el crecimiento se introduce la punta de la semilla, permitiendo que su extremo se funda y de esta forma presentar una superficie limpia. El crecimiento se inicia reduciendo ligeramente la potencia del calefactor, para permitir la nucleación alrededor de la semilla. Esta tendrá un movimiento de giro, con el fin de tener un crecimiento uniforme, a la vez que se aplicará un tirado, o movimiento vertical. La velocidad de tirado es del orden de cm/hr, o sea que la obtención de un lingote de varios metros puede llevar algunos días.

El diámetro del lingote estará dado por:

$$q_i - q_o = -C_f \rho A \frac{dx}{dt} \quad (3.3)$$

Donde:

q_i, q_o calores de entrada (calefactor) y salida (radiación, conducción).

C_f calor específico de fusión.

ρ densidad

A Área del cristal ($2\pi r^2$).

dx/dt Velocidad de tirado.

Si las pérdidas de calor son mayores que la potencia del calefactor la diferencia de energía se tomará del líquido mediante la solidificación del mismo. El volumen solidificado por unidad de tiempo estará dado por $A(dx/dt)$ su masa se conocerá al multiplicarlo por la densidad ρ y el calor absorbido en el proceso estará dado por la masa multiplicada por el calor específico de fusión. Esto es el diámetro del lingote se podrá controlar variando la potencia del calefactor o modificando la velocidad de tiramiento.

3.5.1 Encapsulado líquido

Para la obtención de lingotes de GaAs por la técnica Czochralski es necesario prevenir las fugas de arsénico. Dado que las paredes del sistema se encuentran a baja temperatura, se tendría condensación del mismo. Para solucionar este problema se desarrolló el “*encapsulado líquido*” que consiste en usar una capa de líquido flotando sobre el fundido, que impida la evaporación de arsénico.

Para seleccionar un líquido encapsulante se busca un compuesto que tenga compatibilidad con el proceso, que no contamine, que sea más ligero que el material del fundido y que no reaccione con los materiales del crisol. El compuesto más utilizado para esto ha sido B_2O_3 , aunque también se pueden emplear $BaCl_2$ o $CaCl_2$.

El líquido encapsulante, cubre el fundido, impidiendo la evaporación del arsénico, siempre y cuando la presión de vapor sea inferior a la de la cámara. Para otros sistemas, como GaP, donde la presión de fósforo es muy alta, es necesario además presurizar la cámara. El encapsulante también protege al fundido de la contaminación, y en el caso del B_2O_3 se queda una película muy delgada sobre el lingote solidificado que previene la descomposición de su superficie mientras todavía está caliente.

3.5.2 Síntesis Directa

Cada proceso al que se somete un semiconductor implica mayor contaminación y el GaAs no es la excepción. El desarrollo de sistemas Czochralski que operan bajo presión ha permitido sintetizar el material dentro del reactor antes de iniciar el tirado del cristal a partir de galio y arsénico puros permitiendo evitarse el proceso previo de síntesis en otro sistema. En este caso se carga el equipo de crecimiento con estos elementos

además del encapsulante líquido. Primero es necesario degasar los materiales calentándolos en vacío antes de iniciar la reacción. Posteriormente se introduce un gas inerte en el reactor y se funde el B_2O_3 a 450 °C. Seguido se presuriza la cámara alrededor de 60 atm, y se incrementa la temperatura muy lentamente, ya que la reacción es muy violenta, una sintetizado el GaAs la temperatura se estabiliza alrededor del punto de fusión, se puede reducir la presión e iniciarse el proceso de tirado.

El material obtenido por síntesis directa puede ser de muy alta resistividad (10^8 - $10^9 \Omega \text{ cm}$)⁶, y es bastante más costoso que el que se obtiene a partir de las otras técnicas, principalmente por la necesidad de que el sistema de tirado se pueda presurizar.

3.5.3 Consideraciones Generales

La orientación de la semilla determina la dirección del crecimiento. Igualmente los defectos de la misma, como las dislocaciones se propagarán a lo largo del cristal, para disminuirlas, se hace un “cuello” o sea la parte inicial del crecimiento se realiza a una mayor velocidad de tirado, con el fin de adelgazar el cristal. Las dislocaciones tienden a desaparecer o propagarse hacia el exterior.

El crecimiento se realiza mas fácilmente a lo largo de los planos $<111>$, por lo que generalmente se usarán semillas con esta orientación. Si embargo en algunos casos se necesitan obleas $<100>$ ⁷ por lo que se emplearán semillas de esta dirección, en este caso hay que tener mucho mas cuidado para prevenir perturbaciones.

⁶ Recordar que una resistividad alta está asociada a alta pureza.

⁷ Dispositivos MOS, Laseres de GaAs

Las perturbaciones en el crecimiento, como vibraciones, defectos en el cristal, gradientes de temperatura, u otras, pueden dar lugar a “*gemelamiento*”⁸, este es más probable cuando la orientación cristalina de crecimiento no es preferencial. En algunos casos, el cristal *gemelado* se autocorrige, esto es, con el crecimiento posterior se va haciendo mas pequeño, en otros casos el *gemelo* se propaga a lo largo de todo el lingote, como aquellos crecidos con orientación <110>, por ello la realización de lingotes en esta dirección es extremadamente difícil y la obtención de obleas <110> generalmente se realiza mediante cortes inclinados de los lingotes. Un lingote gemelado tenderá a tener una alta cantidad de dislocaciones, por ello, y dado el dominio de la tecnología, cualquier cristal de silicio que presente gemelos generalmente se rechaza.

Se deben optimizar experimentalmente las razones de crecimiento, razón de giro, forma de la interfaz y distribución de temperatura. Esta última no es constante, ya que el sistema se calienta desde afuera y debe de estar mas frío en el centro. El crecimiento es más homogéneo cuando la interfaz es plana, pero en el caso del silicio ésta puede estar ligeramente curvada hacia arriba en forma de Ω , con lo que se alcanza una mayor velocidad de tirado. En el GaAs, debido a su mala conductividad térmica esto no es posible. Una mayor velocidad de giro aumenta la planaridad de la interfaz, pero también los efectos de disolución del crisol.

3.5.4 Impurificación de los cristales

Generalmente se requiere que las obleas tengan un valor y tipo (n o p) de conductividad determinados para su empleo en dispositivos. Las impurezas se agregan al fundido, generalmente como polvo de

⁸ gemelamiento: cambio de orientación cristalina.

semiconductor altamente impurificado, para así controlar la resistividad. Los elementos mas usados son fósforo y boro para silicio y zinc, cadmio, selenio, telurio, silicio y cromo para impurificar el GaAs.

La concentración de impurezas en el líquido C_L siempre será mayor que la concentración en el sólido C_s , de acuerdo a los diagramas de fase. Se definirá el coeficiente de distribución o segregación como:

$$k = \frac{C_s}{C_L} \quad (3.4)$$

donde k siempre será menor a la unidad. Debido a esto las impurezas irán quedándose en el fundido, incorporándose solo una pequeña parte de las mismas en el sólido, por lo que la concentración en el líquido aumentará a lo largo del crecimiento. La tabla 1 muestra los coeficientes k para el silicio. Los valores son similares para el GaAs.

Al aumentar la concentración de impurezas en el líquido, necesariamente se incorporará una mayor cantidad en el sólido.

Tabla 1. Coeficientes de distribución o segregación de impurezas en silicio.

	P	As	Sb	B	Al	Ga	In	Au
Coeficiente de distribución [k]	0.23	0.27	0.02	0.72	1.8×10^{-3}	7.2×10^{-4}	3.6×10^{-5}	2.25×10^{-5}
Peso atómico	30.97	74.92	121.75	10.81	26.98	69.72	114.82	196.97
Tipo de impureza	tipo n			tipo p				profunda

Tabla 2. Coeficientes de distribución o segregación de impurezas profundas en silicio.

Impureza	P	As	Sb	B	Al	Ga	In	Au
Coeficiente de distribución (K)	0.23	0.27	0.02	0.72	1.8×10^{-3}	7.2×10^{-4}	3.6×10^{-5}	2.25×10^{-5}
Peso atómico	31.0	74.9	121.8	10.8	27.0	69.7	115.0	197.0

La distribución de impurezas no es necesariamente uniforme en el fundido, ya que estas se acumulan en la interfaz, solo si se mezclan rápido con el resto del líquido hablaremos de una concentración homogénea. Esto se logra, por ejemplo, cuando la velocidad de giro es suficientemente rápida para “revolver” la solución. En este caso se hablará de condiciones de “mezclado rápido”.

Si se tiene:

W_M : Peso inicial del fundido.

C_M : Concentración inicial del soluto (impureza) en el fundido (en peso).

C_L : Concentración del soluto en el líquido.

C_S : Concentración del soluto en el cristal.

S Peso del soluto en el fundido.

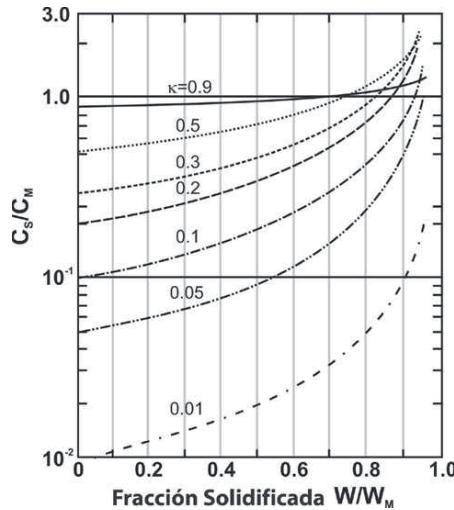


Fig. 3.7 Incorporación de impurezas con respecto a la fracción solidificada del lingote.

Al solidificar un elemento de fundido dW se tendrá también una pérdida de soluto igual a $C_s dW$:

$$-dS = C_s dW \quad (3.5)$$

y el peso del fundido será $W_M - W$:

$$C_L = \frac{S}{W_M - W} \quad (3.6)$$

o sea que al inicio del crecimiento ($W=0$ y $C_M=S/W_M$) $C_L=C_M$

como $C_s/C_L=k$:

$$\frac{dS}{S} = -k \frac{dW}{W_M - W} \quad (3.7)$$

$$\int_{C_M W_M}^S \frac{dS}{S} = k \int_0^W \frac{dW}{W_M - W} \quad (3.8)$$

$$C_s = k C_M \left(1 - \frac{W}{W_M}\right)^{k-1} \quad (3.9)$$

Esto es la concentración de impurezas en el lingote C_s será variable con respecto a la cantidad solidificada (W), y aumentará al ir creciendo la mezcla como se muestra en la figura 3.8. Al final del crecimiento el fundido queda excesivamente impurificado. Si se sobrepasa el límite de solubilidad sólida del lingote este puede crecer como policristal a partir de este punto.

En el caso real de un proceso Czochralski, las velocidades de giro son lentas y el impurificador no se distribuye homogéneamente en el fundido, sino que se acumula en una capa cercana a la interfaz como se muestra en la figura 3.8.

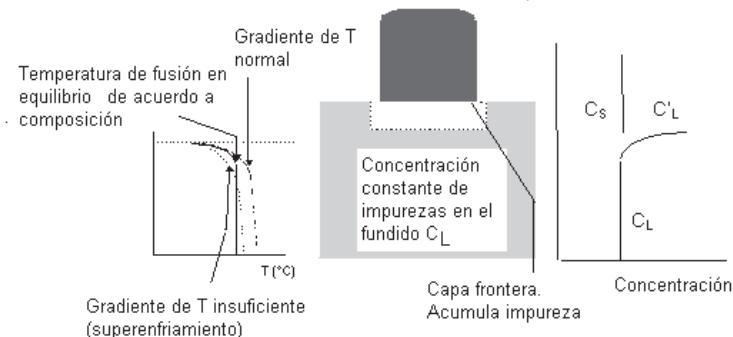


Fig. 3.8 Formación de capa frontera y temperatura de liquidus en un sistema Czochralski.

A esta capa se le llama capa frontera, y se caracterizará por un espesor δ , el cuál será inversamente proporcional a la velocidad de giro. La concentración en el resto del líquido fuera de esta capa no se modifica y continúa siendo la misma que tenía al inicio del proceso. En el interior de

esta capa hay procesos de difusión⁹ y de movimiento de la interfaz por el crecimiento del lingote. Si tenemos:

R: Razón de crecimiento del lingote.

D: Coeficiente de difusión.

La ecuación que determina la concentración en el líquido estará dada por:

$$D \frac{d^2 C}{dx^2} + R \frac{dC}{dx} = 0 \quad (3.10)$$

resolviendo:

$$C = Ae^{-Rx/D} + B \quad (3.11)$$

$$\frac{dC}{dx} = -\frac{AR}{D}e^{-Rx/D} \quad (3.12)$$

Las condiciones de frontera para determinar las constantes son: las concentraciones $C(x=0)=C_L'$, $C(x=\delta)=C_L$ y que la suma de los flujos en la interfaz (por difusión e incorporación al sólido) es igual a cero (esto es la materia se conserva). Esta última se puede escribir de la siguiente forma:

$$D \left(\frac{dC}{dx} \right)_{x=0} + (C_L' - C_S)R = 0 \quad (3.13)$$

y:

$$\left(\frac{dC}{dx} \right)_{x=0} = -\frac{R}{D}(C_L' - C_S) \quad (3.14)$$

Substituyendo en (3.11):

$$\frac{C_L' - C_S}{C_L' - C_S} = e^{-R\delta/D} \quad (3.15)$$

⁹ Las ecuaciones de difusión se tratarán con detalle posteriormente.

y reordenando, podemos definir:

$$k_e = \frac{C_s}{C_L} = \frac{k}{k + (1-k)e^{-R\delta/D}} \quad (3.16)$$

Donde k_e será el coeficiente de distribución efectivo y se aplicará a la ec. (3.9):

$$C_s = k_e C_M \left(1 - \frac{W}{W_M}\right)^{k_e-1} \quad (3.17)$$

Si el término $R\delta/D$ es grande, el coeficiente de distribución efectivo tenderá a la unidad (Ec. 3.16), y se tendrá una concentración uniforme en el cristal. De esta manera, si la velocidad de giro es lenta la concentración del lingote será uniforme, hay que considerar, sin embargo, que el coeficiente de distribución efectivo es valido una vez que se ha estabilizado el proceso, o sea que en la parte inicial del lingote se tendrá un gradiente de concentración. Similarmente, al final del crecimiento tendremos un incremento de la concentración debido a que ya no será válida la aproximación de que $C_L=C_M$.

3.5.5 Superenfriamiento Constitucional

La presencia de impurezas en la capa frontera tiene efectos sobre el punto de fusión de la mezcla. A mayor impurificación menor temperatura de fusión, como se puede ver en un diagrama de fases convencional. En el caso de un sistema Czochralski con mezclado lento, el gradiente de concentración implica que el punto de fusión de es dependiente de la posición dentro de la capa frontera. Esto se muestra en la figura 3.9. Como se ve, el punto de fusión mas bajo es precisamente en la interfaz, siendo mayor entre mas se aleje.

Dado que la interfaz está en equilibrio, en este punto su temperatura corresponderá exactamente al punto de fusión, esta deberá aumentar hacia el interior del líquido. Si el gradiente no es lo suficientemente abrupto habrá parte del líquido cuya temperatura es inferior al punto de fusión de acuerdo a su composición (Fig. 3.9.). Si se presenta esta situación puede haber nucleación en esta región (formación de pequeños núcleos sólidos), provocando a su vez policristalinidad en el lingote.

El superenfriamiento constitucional se presenta principalmente cuando la velocidad de crecimiento del cristal es muy alta, porque las impurezas se acumulan más rápidamente en la interfaz, o cuando la concentración de impurezas en el fundido es muy grande.

3.6 CRISTALIZACIÓN Y PURIFICACIÓN POR ZONA FLOTANTE

Tanto en las técnicas de Czochralski como Bridgman se requiere que el semiconductor fundido esté sobre un crisol de otro material como puede ser cuarzo. A la temperatura del proceso siempre hay algo de contaminación, aunque ésta puede ser mínima y despreciable para casi todas las aplicaciones. Cuando se requiere material de muy alta pureza se puede usar el método de cristalización por zonas [3].

En este sistema se coloca un lingote de silicio policristalino verticalmente, a la vez de que se le sujetó de ambos extremos. El lingote se encuentra encerrado y en una atmósfera inerte. Uno de los extremos tendrá una semilla monocristalina, en éste lado se aplica un campo de radiofrecuencia lo suficientemente intenso para fundir el material. Lentamente se mueve el generador de radiofrecuencia, para que a la vez que solidifica una parte se funde la siguiente, de esta forma cristaliza el material a la vez que va recorriendo la zona fundida o “*zona flotante*”.

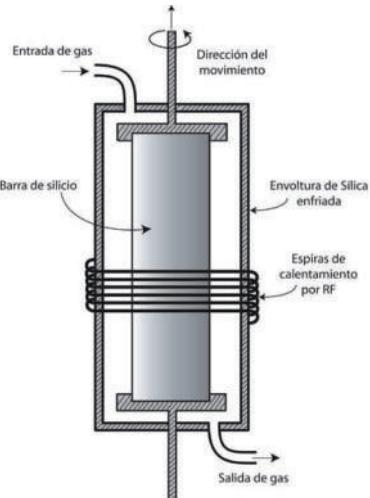


Fig. 3.9 Sistema para purificación por zona flotante.

Al moverse la zona flotante, las impurezas disueltas no se incorporan en la misma concentración que tienen en el líquido debido a que el coeficiente de distribución de impurezas siempre es menor que la unidad. Esto es, las impurezas se acumularán en la zona líquida y eventualmente al final del lingote. La distribución de impurezas estará dada por:

$$C_s = C_m [1 - (1 - k_e) e^{-k_e x/L}] \quad (3.18)$$

Donde:

L: es la longitud de la zona fundida;

X, la distancia del extremo de la barra;

k_e , el coeficiente de distribución efectivo;

y C_m , la concentración inicial de impurezas.

Si se repite el proceso, las impurezas continuarán transportándose hacia el extremo, purificándose el cristal adicionalmente con cada paso. La figura

3.10 muestra la concentración a lo largo del para diferentes pasos de purificación.

En el método de zona flotante no hay contacto entre el semiconductor líquido y su contenedor, de ahí que se evita la contaminación. El lingote de partida también puede ser monocristalino. Su principal desventaja radica en el costo, además, debido a las características de la zona flotante la calidad cristalina es inferior a la que se obtiene por los otros métodos.

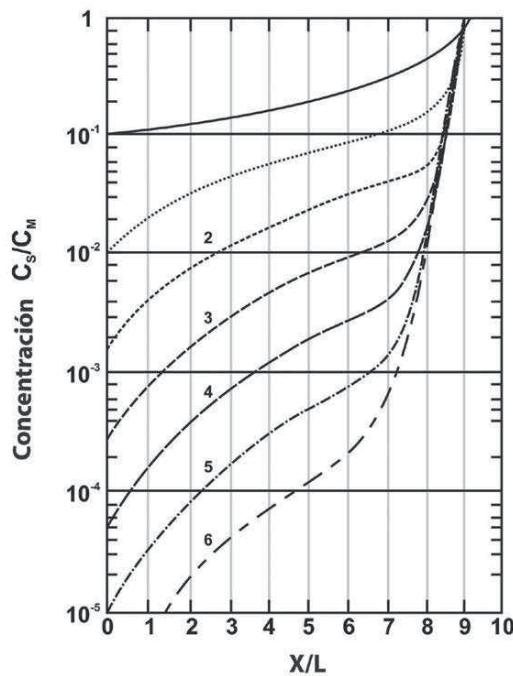
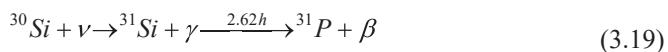


Fig. 3.10 Perfil de impurezas para pasos múltiples y $k=0.1$.

3.7 IMPURIFICACIÓN POR NEUTRONES

En algunos casos la homogeneidad de los cristales no es buena, sino que se encuentran variaciones importantes a nivel microscópico de la

resistividad. Estas se presentan mas comúnmente en material obtenido por zona flotante. Estas variaciones pueden provocar ruptura prematura, campos eléctricos no homogéneos o degradación de los dispositivos. Para homogeneizar la concentración se puede usar la técnica de transmutación por neutrones, esto es irradiando los lingotes:



Esto es, los neutrones convertirán al isótopo ^{30}Si en ^{31}Si con la liberación de radiación gamma¹⁰. Dado que este último es inestable y tiene una vida media de 2.62 horas, se descompondrá generando fósforo. La importancia de la técnica radica en que los neutrones pueden atravesar prácticamente todo el material de manera uniforme, con lo que la impurificación resultante también lo será. Esta técnica es útil cuando se requiere que la distribución del campo eléctrico en el material sea uniforme.

3.8 BIBLIOGRAFÍA

- [1] H.J. Sell, “Melt growth processes for semiconductors,” Key Eng. Materials, 58 (1991) 169.
- [2] W. Zuhlehner y D. Huber, “Czochralski Grown Silicon”, Cristal 8, Springer-Verlag Berlina 1982.
- [3] W. C. Dash, “Growth of silicon crystals free from dislocations,” J. Appl Phys, 30 (1959) 459.

¹⁰ El silicio tiene normalmente dos isótopos: ^{28}Si y ^{30}Si encontrándose que este último es el 3% del total.

CAPÍTULO 4

DIFUSIÓN

El control de los procesos de difusión para impurificar semiconductores en conjunto con los de oxidación y las técnicas de litografía ha sido fundamental en el desarrollo de la tecnología del silicio, y con ello de toda la electrónica moderna.

Se puede definir difusión como el flujo que se realiza cuando existe un gradiente de concentración de una especie dada. En el lenguaje cotidiano se habla de medios de difusión como aquellos que dan a conocer la información, o sea que esta fluye de los que tienen el conocimiento hacia los que no. Difusión es el fenómeno que propaga una epidemia, una plaga o simplemente el desorden. La difusión también se presenta en la materia, por ejemplo cuando se tiene una gota de tinta en un vaso de agua veremos que lentamente se teñirá toda el agua.

La difusión de una especie se presenta cuando ésta se puede mover de manera aleatoria. La información confidencial no se difunde, porque hay barreras que lo impiden. El contagio de una enfermedad se debe al contacto que tienen personas sanas con enfermos, o sea que hay mecanismos de transmisión de microbios. La materia se puede mover debido a las oscilaciones propias del medio.

Aunque el proceso de difusión es estadístico y se basa en movimientos aleatorios, cuando existe un gran número de eventos el cálculo matemático permite predecir con gran exactitud el resultado. Esto se ha aprovechado en la tecnología de semiconductores para introducir de manera controlada impurezas en el interior de un cristal para la formación de dispositivos.

4.1 DIFUSIÓN EN CRISTALES

En una red cristalina hay lugares substitucionales e intersticiales. Un átomo colocado en un intersticio puede saltar a un lugar vecino, si por efectos térmicos adquiere la energía necesaria para vencer la barrera que lo tiene en su lugar, en este caso se hablará de “**difusión interstitial**”.

Cuando los saltos en el cristal se realizan de un sitio de la red hacia una vacancia, hablaremos de “**difusión substitucional**”. En este caso se requiere conjugar la existencia de vacancias con el fenómeno.

En algunos casos los átomos substitucionales saltarán a intersticios y de ahí iniciarán el proceso de difusión para al final tomar de nuevo un lugar substitucional. Este se llama “**mecanismo disociativo**”.

“**Difusión de Intercambio**” se presenta cuando dos átomos substitucionales participan intercambiando sus lugares.

Algunas especies pueden difundirse de manera combinada, o sea una parte substitucionalmente mientras que otra por la vía intersticial. También puede haber difusión a través de los defectos como son dislocaciones o fronteras de grano.

4.1.1 Movimiento Intersticial

La gran mayoría de los intersticios de la red cristalina están vacíos, o sea que lo más probable es que un átomo intersticial tenga libres a sus vecinos. Para que un átomo salte a otro intersticio es necesario vencer la barrera de potencial presente, la cual estará caracterizada por una energía determinada.

Una red cristalina se encuentra oscilando constantemente por efecto de la temperatura. Cada vez que los átomos que limitan un intersticio se

alejan, se abrirá un hueco y la impureza tendrá una mayor probabilidad de saltar.

La frecuencia con la cual los átomos intersticiales pueden saltar a otro lado de la red sería [1]:

$$\nu_i = 4\nu_0 e^{-E_m/kT} \quad (4.1),$$

donde

E_m es la energía necesaria para saltar a otro intersticio. ($\approx 0.6\text{-}1.2$ eV)

T es la temperatura en [K].

ν_0 : es la frecuencia de oscilación de la red del orden de $10^{13}\text{-}10^{14}$ Hz.

El factor 4 proviene del hecho de que en la red zincblenda habrá 4 intersticiales vecinos. Este podrá ser diferente para otras redes cristalinas.

La frecuencia de saltos por cada átomo intersticial será del orden de 1 salto por minuto a 300 K. Hay que notar en la ecuación (4.1) que la dependencia de la frecuencia con la temperatura será de forma exponencial.

En la Fig. 4.1 se muestra la trayectoria de estas impurezas dentro de la red.

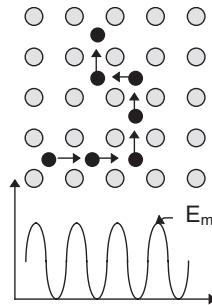


Fig. 4.1 Movimiento de impurezas intersticiales dentro de una red cristalina.

4.1.2 Movimiento substitucional

En el caso de los semiconductores, las impurezas de interés son de tipo substitucional. Para poder saltar a otro lado de la red será necesario que, además de tener la energía necesaria para romper los enlaces que a la que está sujeta, exista una vacancia contigua adonde pueda llegar. Esto es, la difusión substitucional requiere también de la difusión de vacancias, y dependerá directamente de la concentración de las mismas. En este caso la probabilidad para que se realice un salto se determinará en base a las energías de formación de vacancias y la barrera de potencial necesaria para liberar los enlaces [1].

$$\nu_s = 4\nu_0 e^{-(E_n + E_S)/kT} \quad (4.2),$$

dónde:

E_S es la energía de formación de una vacancia.

E_n es la energía de la barrera de potencial.

En la Fig. 4.2 se muestra el desplazamiento de una impureza substitucional dentro de una red cristalina.

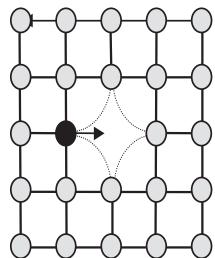


Fig. 4.2 Desplazamiento de una impureza sustitucional dentro de una red cristalina.

En el caso del silicio En+Es es del orden de 3 a 4 eV (considerablemente mayor que el caso de las impurezas intersticiales) y la frecuencia de saltos será del orden de 1 cada 1045 años. La dependencia exponencial del proceso permite que estos sean notables a altas temperaturas mientras que en condiciones ambientales las impurezas estarán prácticamente inmóviles.

4.1.3 Movimiento Intersiticial-Substitucional

En algunos casos las impurezas substitucionales pueden saltar a intersitios, difundiéndose rápidamente para situarse posteriormente en algún lugar substitutional. En este caso se considerarán dos procesos independientes y la difusión será la suma de ambos procesos.

La frecuencia de saltos será:

$$\nu_{eff} = \frac{\nu_s N_s}{N_i + N_s} + \frac{\nu_i N_i}{N_i + N_s} \quad (4.3)$$

donde

N_s es el número de impurezas en sitios substitucionales.

N_i es el número de impurezas en sitios intersticiales.

$\frac{N_i}{N_i + N_s}$ es la fracción de átomos en sitios intersticiales.

$\frac{N_s}{N_i + N_s}$ es la fracción de átomos en sitios substitucionales.

y v_s , v_i son las frecuencias de saltos de impurezas substitucionales e intersticiales, respectivamente.

Considerando que $N_s \gg N_i$:

$$\nu_{eff} \cong \frac{v_i N_i}{N_s} \quad (4.4)$$

O sea que la frecuencia de saltos estará determinada por la capacidad de los átomos a saltar a los intersticios de la red. La relación N_i/N_s será diferente en cada caso y algunas veces dependerá de la posición del nivel de Fermi, como es el caso del Zn en GaAs.

4.2 DIFUSIÓN EN UN GRADIENTE DE CONCENTRACIÓN

En un proceso de difusión de semiconductores se coloca una cantidad determinada de impurezas sobre la superficie de la oblea. Estas pueden saltar aleatoriamente hacia los diferentes lugares vecinos de acuerdo a su frecuencia característica de salto. Debido a que hay una mayor cantidad de impurezas en la superficie que en el interior, esto resultará en un flujo neto de impurezas en dirección contraria al gradiente de concentración.

Considerando una barra de silicio con orientación <100> y de área transversal A. En el centro se seleccionan dos diferenciales de longitud dx contiguos entre los que puede haber saltos de impurezas. El tamaño de cada uno de estos, de acuerdo a consideraciones geométricas de la estructura del

diamante, será $d / \sqrt{3}$. Sean N_1 y N_2 las concentraciones en cada una de las regiones respectivas y n_1 y n_2 las cantidades totales de impurezas en cada una de las áreas.

$$N_1 = \frac{\sqrt{3}n_1}{Ad} \quad (4.5)$$

$$N_2 = \frac{\sqrt{3}n_2}{Ad} \quad (4.6)$$

Cada una de las impurezas podrá saltar con una frecuencia característica hacia cada uno de los 4 lugares vecinos, de los cuales una mitad estarán a la derecha y la otra a la izquierda. El número neto de saltos por unidad de tiempo que atraviesen la interfaz entre las dos regiones será la diferencia entre los que brinquen de 1 a 2 menos los que vayan de 2 a 1, Fig. 4.3:

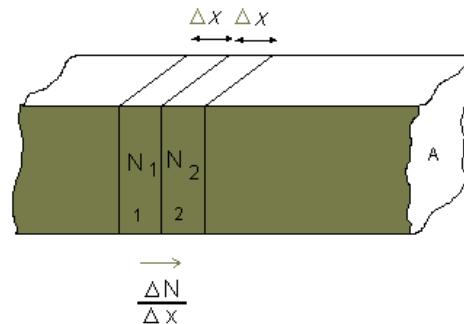


Fig. 4.3 Movimiento de impurezas debido a un gradiente de concentración.

Este desplazamiento puede expresarse como:

$$\frac{\Delta n}{\Delta t} = \frac{(n_1 - n_2)/2}{1/\nu} = \frac{\nu}{2} \frac{Ad}{\sqrt{3}} (N_1 - N_2) \quad (4.7)$$

El gradiente de concentración estará dado por:

$$\frac{\Delta N}{\Delta x} = \frac{N_2 - N_1}{d / \sqrt{3}} \quad (4.8)$$

Con lo que la cantidad de impurezas que atraviesan la interfaz por unidad de tiempo será:

$$\frac{\Delta n}{\Delta t} = -A \frac{\nu d^2}{6} \frac{\Delta N}{\Delta x} \quad (4.9)$$

o:

$$j = -\frac{\nu d^2}{6} \frac{\Delta N}{\Delta x} \quad (4.10)$$

Se define el coeficiente de difusión como:

$$D = \frac{\nu d^2}{6} \quad (4.11)$$

y se tendrá:

$$J = -D \frac{\partial N}{\partial x} \quad (4.12)$$

$$J = -D \nabla \cdot N \quad (4.13)$$

Las ecuaciones (4.12) y (4.13) constituyen la **Primera ley de Fick** para una y tres dimensiones respectivamente [1].

4.2.1 El Coeficiente de difusión

El coeficiente de difusión tendrá la forma:

$$D = \frac{\nu d^2}{6} = \frac{4 \nu_0 d^2}{6} e^{-E_D/kT} \quad (4.14)$$

donde E_D será E_m o $(E_n + E_s)$ según corresponda, y:

$$D_0 = \frac{4V_0 d^2}{6} \quad (4.15)$$

entonces el coeficiente de difusión estará dado por:

$$D = D_0 e^{-E_D/kT} \quad (4.16)$$

Para generalizar esta ecuación hay que considerar que tanto D_0 como E_D estarán determinados por los mecanismos de dispersión del medio. Esto es, la difusión no solo se presenta en medios cristalinos.

Típicamente los procesos de difusión en silicio se realizan de 900 a 1200 °C, el rango de concentración intrínseca de portadores varía entre 10^{17} a 10^{20} cm⁻³, ver Fig. 4.4. O sea, el semiconductor puede ser o no ser intrínseco cuando se realiza el proceso. En el primer caso, o sea cuando el semiconductor es intrínseco, el nivel de Fermi no varía, mientras que en el segundo va a ser dependiente de la concentración misma de las impurezas que se están difundiendo. Este efecto es mas notable en GaAs ya que las difusiones se realizan a menor temperatura y la concentración intrínseca de portadores varía de 10^{13} cm⁻³ (600 °C) a 10^{16} cm⁻³,(1000 °C).

Cuando la difusión se realiza en un material semiconductor extrínseco las impurezas se ionizarán liberando portadores de carga móviles (electrones o huecos). Estos a su vez se difundirán mas rápido generando un campo eléctrico debido a la separación de las cargas. El siguiente modelo se planteará para impurezas donadoras y liberación de electrones, sin embargo también será válido para difusión de aceptores con liberación de huecos.

Un campo eléctrico generará una velocidad y una corriente en las impurezas ionizadas que estarán dadas por:

$$v_d = \mu \bar{E} \quad (4.17)$$

y

$$j_{\bar{E}} = \mu N \bar{E} \quad (4.18)$$

Considerando el efecto del campo independiente de la difusión:

$$j = j_D + j_{\bar{E}} = -D \frac{\partial N}{\partial x} + \mu N \bar{E} \quad (4.19)$$

Estas mismas ecuaciones son las que se han aplicado para el transporte de los portadores libres en semiconductores. Si la aplicamos a los

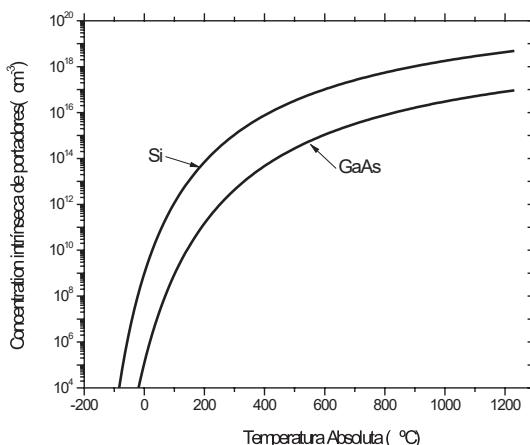


Fig. 4.4 Concentración intrínseca de portadores de Silicio y Arseniuro de Galio.

electrones liberados por las impurezas y consideramos que no hay corriente eléctrica neta podremos calcular el campo eléctrico resultante¹¹:

¹¹ Notar que las ecuaciones están multiplicadas por q dado que se trata corriente eléctrica.

$$j_n = q\mu\bar{E} + qD\frac{\partial n}{\partial x} = q\mu\left(n\bar{E} + \frac{kT}{q}\frac{\partial n}{\partial x}\right) = 0 \quad (4.20)$$

$$\bar{E} = -\frac{kT}{q} \frac{1}{n} \frac{\partial n}{\partial x} \quad (4.21)$$

Aplicando la relación de Einstein: ($D = (kT/q)\mu$), y el valor del campo eléctrico (eq. 4.21) a la (4.19) tenemos que:

$$j = -D\left(1 + \frac{dn}{dN}\right) \frac{\partial N}{\partial x} \quad (4.22)$$

Podemos definir un coeficiente de difusión efectivo D_{eff} y un factor de enriquecimiento por campo h :

$$D_{eff} = D\left(1 + \frac{dn}{dN}\right) = hD \quad (4.23)$$

o:

$$h = \left(1 + \frac{dn}{dN}\right) \quad (4.24)$$

Para conocer el valor de h se aplica la ecuación de neutralidad de carga:

$$n - p + N_A^- - N_D^+ = 0 \quad (4.25)$$

considerando que la concentración de aceptores es despreciable ($N_A \approx 0$), que el material es extrínseco y todos los átomos donores están ionizados ($N_D^+ \approx N_D$) y aplicando la ley de acción de masas para concentración de electrones y huecos ($np = n_i^2$) podemos resolver la ec. (4.25):

$$n = \frac{N_D^+ + \sqrt{N_D^{+2} + 4n_i^2}}{2} \quad (4.26)$$

resolviendo para (dn/dN)

$$\frac{\partial n}{\partial N_D^+} = \frac{1 + (N_D^+)^2 + 4n_i^2}{2}^{-1/2} 2N_D^+ = \frac{1 + \sqrt{1 + (2n_i / N_D^+)^2}}{2} \quad (4.27)$$

el valor de h será:

$$h = 1 + \frac{1}{\sqrt{1 + (2n_i / N_D^+)^2}} \quad (4.28)$$

o sea $1.5 < h < 2$.

Para llegar a la eq. (4.28) se usó la aproximación de que todas las impurezas estaban ionizadas ($N_D^+ \approx N_D$). Cuando N_D tiene valores cercanos a n_i esta aproximación ya no es válida ya que $n \approx p \approx n_i$ y entonces la ecuación de neutralidad de carga quedará:

$$N_D^+ = n - p \approx 0 \quad (4.29)$$

por lo que la concentración de portadores será independiente de la de impurezas ($dn / dN_D \approx 0$), y el valor de h será:

$$h = \left(1 + \frac{dn}{dN} \right) \approx 1 \quad (4.30)$$

De manera general se puede afirmar que $1 < h < 2$.

4.2.2 Coeficiente de difusión considerando vacancias cargadas

Las vacancias en semiconductores pueden estar cargadas eléctricamente de acuerdo a la forma en que se acomoden los enlaces. En la red cristalina del diamante se podrán presentar vacancias con una carga positiva, o una, dos o tres cargas negativas como se muestra en las figuras de la derecha. Una impureza substitucional ionizada tendrá una interacción diferente cuando la vacancia presente un estado de carga debido a las fuerzas coulombianas.

Considerando que los eventos de difusión son independientes entre si para cada mecanismo se tendrá para silicio intrínseco [1]:

$$D_i = D_i^0 + D_i^+ + D_i^- + D_i^{2-} + D_i^{3-} \quad (4.31)$$

Esto es habrá un coeficiente de difusión individual para cada uno de los procesos. Los estados de carga de las vacancias estarán determinados por las ecuaciones de equilibrio, y si el material es intrínseco su concentración solo dependerá de la temperatura. Para material extrínseco cada uno de los coeficientes de difusión se modificará de acuerdo a las concentraciones de cada uno de los estados de las vacancias:

$$D^0 = D_i^0 \frac{[V^0]}{[V^0]_i}; D^+ = D_i^+ \frac{[V^+]}{[V^+]_i}; D^- = D_i^- \frac{[V^-]}{[V^-]_i} \dots \quad (4.32)$$

donde $[V^k]$ representa la concentración de vacancias cargadas en el estado k y el subíndice i denota el valor propio del material intrínseco. El coeficiente de difusión para cada uno de estos procesos se multiplicará por

la variación de la concentración de cada estado con respecto al valor respectivo del semiconductor intrínseco de esta forma se puede determinar el coeficiente de difusión para silicio extrínseco como:

$$D = D_i^0 \frac{[V^0]}{[V^0]_i} + D_i^+ \frac{[V^+]}{[V^+]_i} + D_i^- \frac{[V^-]}{[V^-]_i} + D_i^{2-} \frac{[V^{2-}]}{[V^{2-}]_i} + D_i^{3-} \frac{[V^{3-}]}{[V^{3-}]_i} \quad (4.33)$$

Dado que la porción de vacancias ionizadas es muy pequeña la concentración de vacancias neutras no se modifica apreciablemente si el semiconductor es intrínseco o extrínseco por lo que se puede aproximar $[V^0] \approx [V^0]_i$.

Los demás estados de carga dependerán de la posición del nivel de Fermi y/o de la concentración de portadores. Para que una vacancia se cargue es necesaria la intervención de un portador de carga (electrón o hueco). En el caso de vacancias cargadas positivamente se tendrá la siguiente reacción:



aplicando la ley de acción de masas para los estados extrínsecos e intrínsecos se tendrá:

$$K_1 = \frac{[V^+]}{p[V^0]} = \frac{[V^+]_i}{n_i[V^0]_i} \rightarrow \frac{[V^+]_i}{[V^+]_i} = \frac{p}{n_i} \quad (4.35)$$

Similarmente para cargas negativas:



$$K_{rn} = \frac{[V^{r-}]}{n^r[V^0]} = \frac{[V^{r-}]_i}{n_i^r[V^0]_i} \rightarrow \frac{[V^{r-}]_i}{[V^{r-}]_i} = \left(\frac{n}{n_i}\right)^r \quad (4.37)$$

donde r denota el número de cargas negativas. Entonces el coeficiente de difusión será:

$$D = D_i^0 + D_i^+ \left(\frac{p}{n_i} \right) + D_i^- \left(\frac{n}{n_i} \right) + D_i^{2-} \left(\frac{n}{n_i} \right)^2 + D_i^{3-} \left(\frac{n}{n_i} \right)^3 \quad (4.38)$$

Esto es, cuando la concentración de impurezas (N_D, N_A) es inferior a n_i a la temperatura del proceso (Fig. 4.4) el coeficiente de difusión será constante. Cuando se sobrepasa este valor, habrá un incremento de D por los términos relacionados con vacancias cargadas incrementándose al hacerlo la concentración. Dado que en los procesos normales existe un gradiente, D tendrá un valor mayor cerca de la superficie. Conforme mayor sea el estado de carga, más rápido lo hará el coeficiente de difusión.

En la Fig. 4.5 se muestran los diferentes estados de carga de las vacancias.

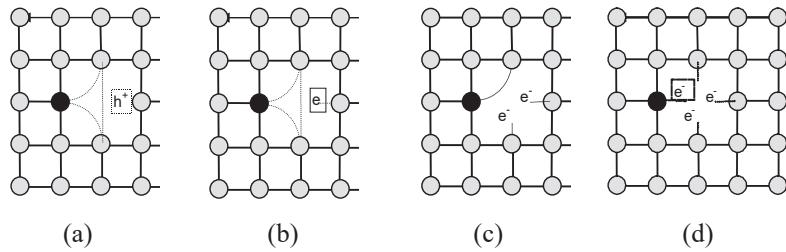


Fig. 4.5 Diferentes estados de carga de vacancias: (a) Vacancia cargada positivamente; (b) vacancia con una carga negativa; (c) vacancia con 2 cargas negativas; (d) vacancia con 3 cargas negativas.

Por otro lado hay que notar que no todos los estados de carga contribuirán al coeficiente de difusión. En el caso de aceptores solo serán importantes las vacancias cargadas positivamente, ya que la concentración

de vacancias negativas será despreciable. En el caso de donores tendremos la situación inversa. Para el caso práctico del silicio el nivel de carga triple negativo está dentro de la banda de conducción por lo que no habrá vacancias con este estado de carga y solo se presentarán los coeficientes D^- y D^{2-} . Este último solo será importante en la difusión de fósforo.

De manera general considerando el efecto de campo eléctrico el coeficiente de difusión se podrá escribir como:

$$D_{eff} = h \left[D_i^0 + D_i^+ \left(\frac{p}{n_i} \right) + D_i^- \left(\frac{n}{n_i} \right) + D_i^{2-} \left(\frac{n}{n_i} \right)^2 + D_i^{3-} \left(\frac{n}{n_i} \right)^3 \right] \quad (4.39)$$

4.2.3 La Ecuación de difusión

En tecnología se requiere calcular el perfil de distribución de las impurezas después de un proceso de difusión. Consideremos de nuevo la barra de material. En este caso se tendrá un flujo j a la izquierda de un elemento, y $j+dj$ en un plano a la salida con una separación dx , ver Fig. 4.6.

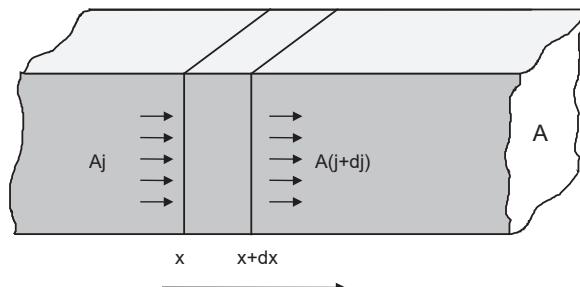


Fig. 4.6 Proceso de difusión de impurezas.

Si los flujos de entrada y de salida son diferentes habrá una acumulación o empobrecimiento de impurezas, esto es la concentración N variará con el tiempo ($\partial N / \partial t$) de acuerdo a:

$$A \frac{\partial N}{\partial x} dx = -Adj \quad (4.40)$$

aplicando la primera ley de Fick para dj se tendrá:

$$\frac{\partial N}{\partial x} = \frac{\partial}{\partial x} \left(D \frac{\partial N}{\partial x} \right) \quad (4.41)$$

Esta es la segunda ley de Fick para una dimensión. Generalizando en tres dimensiones:

$$\frac{\partial N}{\partial x} = \nabla \cdot D \nabla \cdot N \quad (4.42)$$

Si el coeficiente de difusión no depende de la posición (x), esto es no depende de la concentración N , las ec. (4.41) y (4.42) se pueden simplificar:

$$\frac{\partial N}{\partial x} = D \frac{\partial^2 N}{\partial x^2} \quad (4.43)$$

y

$$\frac{\partial N}{\partial x} = D \nabla^2 N \quad (4.44)$$

Si la difusión se realiza en un campo eléctrico, como puede ser el caso de portadores de carga en semiconductores, la segunda ley de Fick se puede escribir como¹²:

$$\frac{\partial N}{\partial x} = D \frac{\partial^2 N}{\partial x^2} - \mu E \frac{\partial N}{\partial x} \quad (4.45)$$

¹² En el caso analizado anteriormente para el campo eléctrico producido por la ionización de las impurezas se integró el efecto del campo eléctrico al coeficiente de difusión, por lo que ya no se aplicaría esta ecuación. En el caso general no siempre es posible relacionar el campo eléctrico con la concentración N .

Los procesos realizados en situaciones reales conllevan una serie de consideraciones prácticas que permiten simplificar la segunda ley de Fick. Normalmente la difusión se realiza en dos partes: Predepósito y redistribución. En el primero se expone la oblea de semiconductor a una concentración de impurezas constante, casi siempre impuesto por el límite de solubilidad sólida. Este último es dependiente de la temperatura del proceso, y muy alto para fines prácticos, por lo que es necesario un segundo paso de redistribución, que bajará la concentración superficial y causará que las impurezas penetren hacia el interior de la oblea.

Aplicando las condiciones de frontera para la ec. (4.41) que son: concentración superficial constante (N_0), concentración para $x \rightarrow \infty = 0$ (la al interior de la oblea será despreciable) y $N(t=0)=0$ para $x > 0$ (al inicio del proceso no hay impurezas dentro de la oblea) se puede encontrar la solución [2]:

$$N(x,t) = N_0 \operatorname{erfc} \left(\frac{x}{2\sqrt{Dt}} \right) \quad (4.46)$$

donde $\operatorname{erfc}(x / (2\sqrt{Dt}))$ representa la función error complementaria ($\operatorname{erfc}(z) = 1 - \operatorname{erf}(z)$) y

$$\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-y^2} dy \quad (4.47)$$

Las soluciones particulares de la ec. 4.44 para diferentes tiempos se encuentran graficados en las Fig. 4.7.

Para estructuras de dispositivos es importante determinar la profundidad de la unión pn (x_j), o sea el punto donde se verifica el cambio de

conductividad. Si se conoce la concentración del sustrato (N_s), es posible conocer ésta donde se igualan las concentraciones $N(x,t)$ y N_s :

$$N_0 \operatorname{erfc} \left(\frac{x_j}{2\sqrt{Dt}} \right) = N_s \quad (4.48)$$

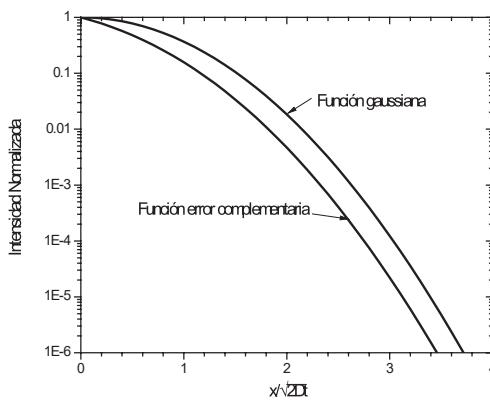


Fig. 4.7 Forma de un perfil de distribución de impurezas tipo función error complementario y tipo gaussiano.

Para la solución del proceso de redistribución habría que partir de la solución anterior. Sin embargo, si el tiempo de este segundo paso es suficientemente grande y dado que generalmente éste se realiza a una temperatura mayor con lo que el coeficiente de difusión se incrementa considerablemente, la influencia del perfil inicial sobre la distribución final es despreciable. En este caso la solución estará dada por [2]:

$$N(x,t) = \frac{Q_0}{\sqrt{\pi Dt}} e^{-(x/2\sqrt{Dt})^2} \quad (4.49)$$

Donde Q_0 será la cantidad total de impurezas que se hayan introducido en el primer paso y que puede determinarse integrando la ec. (4.46).

$$Q_0 = 2N_{01} \left(\frac{D_1 t_1}{\pi} \right)^{1/2} \quad (4.50)$$

y:

$$N(x, t) = \frac{2N_{01}}{\pi} \left(\frac{D_1 t_1}{D_2 t_2} \right) e^{-(x/2\sqrt{Dt})^2} \quad (4.51)$$

donde $N(x, t)$ será la distribución final siempre que se cumpla la condición

$$\sqrt{D_1 t_1} \ll \sqrt{D_2 t_2} \quad (4.52)$$

Las funciones de distribución gaussiana y error complementarias se muestran en la Fig. 4.7 Es importante notar que el argumento z determina el orden de magnitud de los perfiles., dado que ambas funciones decaen rápidamente cada vez que z se incrementa en una unidad. A su vez: $z = x / 2\sqrt{Dt}$, por lo que el producto en la raíz cuadrada dará idea de los órdenes de magnitud en que se manifestarán los perfiles de los procesos. Esto es, la profundidad de unión generalmente estará en el rango de 1.5 a 3 veces el valor de $2*(Dt)^{1/2}$.

Si luego de un proceso de introducción de impurezas se realiza un proceso térmico a alta temperatura, garantizando de alguna manera que no haya flujo de impurezas hacia fuera del material, la impureza se redistribuye hacia el interior del material. Para conocer la nueva distribución se resuelve la ec. 4.43 bajo la condición de concentración total de impurezas constante. La distribución de impurezas en este caso vendrá

dada por la expresión (4.49), que describe un perfil del tipo gaussiano. En este caso, para una temperatura dada del proceso de redistribución, la concentración superficial va disminuyendo y el perfil penetrando, como se muestra en la Fig. 4.8.

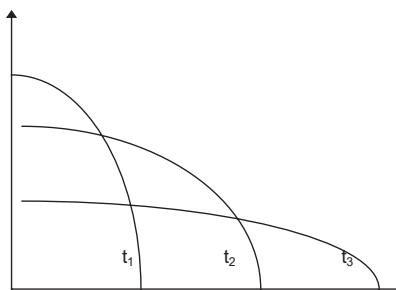


Fig. 4.8 Perfil de distribución de impurezas bajo la condición de cantidad de impurezas constante.

Cuando no se cumple la condición de la ec. (4.52) se aplica la solución general a la segunda ley de Fick:

$$N(x, t_1, t_2) = \frac{2N_{01}}{\pi} \int_0^U \frac{e^{-\beta(1+U^2)}}{1+U^2} dU \quad (4.53)$$

donde:

$$U = \left(\frac{D_1 t_1}{D_2 t_2} \right)^{\frac{1}{2}} \quad (4.54)$$

y

$$\beta = \left(\frac{x}{2\sqrt{D_1 t_1 + D_2 t_2}} \right)^2 \quad (4.55)$$

y la concentración superficial estará dada por:

$$N_{02} = \frac{2N_{01}}{\pi} \tan^{-1} U \quad (4.56)$$

Las soluciones a las ecuaciones anteriores son numéricas.

En muchos casos una difusión es sometida a varios tratamientos térmicos posteriores, por la tecnología que se está empleando, como es el caso de la difusión de base de un transistor, que también será sometida al proceso de difusión del emisor. En el caso de difusiones sucesivas se puede aproximar:

$$(Dt)_{\text{eff}} = \sum D_1 t_1 + D_2 t_2 + D_3 t_3 + \dots \quad (4.57)$$

4.2.4 Difusión dependiente de la concentración

Cuando el coeficiente de difusión depende de la concentración (vacancias cargadas), éste puede ser directamente proporcional a la concentración de impurezas al cuadrado o la tercera potencia de la de la misma. En estos casos D tendrá un valor mayor cerca de la superficie y se reducirá al interior de acuerdo a alguna de las siguientes ecuaciones, según sea el caso [1]:

$$D = K_1 D_0 N = D_{\text{sup}} \left(\frac{N}{N_{\text{sup}}} \right) \quad (4.58)$$

$$D = K_2 D_0 N^2 = D_{\text{sup}} \left(\frac{N}{N_{\text{sup}}} \right)^2 \quad (4.59)$$

$$D = K_3 D_0 N^3 = D_{\text{sup}} \left(\frac{N}{N_{\text{sup}}} \right)^3 \quad (4.60)$$

En estos casos no habrá una solución analítica, y esta se podrá calcular numéricamente. El parámetro mas importante es la profundidad de unión y,

debido a lo abrupto de los perfiles de difusión, esta podrá ser aproximada por alguna de las siguientes ecuaciones:

$$x_j = 1.616(D_{\text{sup}}t)^{1/2} \quad D \propto N \quad (4.61)$$

$$x_j = 1.092(D_{\text{sup}}t)^{1/2} \quad D \propto N^2 \quad (4.62)$$

$$x_j = 0.872(D_{\text{sup}}t)^{1/2} \quad D \propto N^3 \quad (4.63)$$

4.2.5 Difusión Intersticial-Substitucional

Existe el caso de impurezas como el oro en silicio o el zinc en GaAs que se difunden por este mecanismo. En este caso se consideran dos procesos independientes que transcurren en paralelo [1]:

$$\frac{\partial N_s}{\partial x} = \frac{\partial}{\partial x} D_s \frac{\partial N_s}{\partial x} \quad (4.64)$$

$$\frac{\partial N_i}{\partial x} = \frac{\partial}{\partial x} D_i \frac{\partial N_i}{\partial x} \quad (4.65)$$

Combinando (4.64) y (4.65):

$$\frac{\partial(N_s + N_i)}{\partial x} = \frac{\partial}{\partial x} \left(D_s \frac{\partial N_s}{\partial x} + D_i \frac{\partial N_i}{\partial x} \right) \quad (4.66)$$

si $N_s \gg N_i$ y $D_i \gg D_s$:

$$\frac{\partial N_s}{\partial x} = \frac{\partial}{\partial x} \left(D_s + D_i \frac{\partial N_i}{\partial N_s} \right) \frac{\partial N_s}{\partial x} \quad (4.67)$$

donde el coeficiente de difusión se podrá aproximar por:

$$D \approx D_s + D_i \frac{\partial N_i}{\partial N_s} \approx D_i \frac{\partial N_i}{\partial N_s} \quad (4.68)$$

4.3 SISTEMAS DE DIFUSIÓN

El diseño de los sistemas de difusión requiere tomar en cuenta varios requisitos. Considerando que los procesos se deben de realizar a altas temperaturas y dadas las exigencias de perfección de los materiales en la fabricación de dispositivos semiconductores son aspectos que deben de ser tomados en cuenta.

La **limpieza** es requisito fundamental, en una planta de difusión hay varias fuentes de contaminación: el polvo de la atmósfera, las secreciones humanas y las impurezas adsorbidas en la superficie de las obleas, entre otras. Para evitar contaminación en los reactores se utilizarán tubos de sílice o cuarzo¹³, o, cuando las temperaturas de difusión sean altas se emplearán tubos de silicio policristalino o de carburo de silicio. Los tubos de cuarzo tienen la desventaja de que son permeables a los iones alcalinos y que el cuarzo se devitrifica a altas temperaturas, perdiendo varias de sus propiedades. Los otros tubos son mucho más caros, pero su durabilidad compensa la diferencia en costo. Los tubos de carburo de silicio son una alternativa adicional.

El proceso de introducción de impurezas o predifusión se realiza generalmente colocando las obleas de silicio en un horno de alta temperatura, entre 900 y 1250 °C, a través del cual fluye un gas inerte arrastrando la impureza deseada. Estos hornos reciben el nombre de hornos de difusión.

¹³ En realidad el cuarzo es el estado cristalino del silice, sin embargo en lenguaje cotidiano se emplea la esta palabra para designarlo.

Las impurezas puedes ser introducidas al horno de difusión de varias formas:

1) El material que sirve de fuente es sólido a temperatura ambiente y se coloca en una zona del horno a una temperatura dada, para la cual presenta una tensión de vapor es suficientemente alta para que sea arrastrado por el gas portador en las concentraciones requeridas.

2) El material que sirve de fuente es líquido a temperatura ambiente y se coloca en el exterior del horno, a una temperatura dada para alcanzar la tensión de vapor requerida. Esto se realiza introduciéndolo en termostatos.

La relación entre la temperatura y la tensión de vapor de la fuente se puede determinar mediante el uso de tablas y gráficos, como el que se muestra en la Tabla 4.

El gas de arrastre lleva a la especie activa hasta la superficie de la muestra, donde la presión parcial se puede determinar usando la ecuación de transporte de masa y la ley de gases ideales.

Por otro lado, a cada presión parcial de la especie activa junto a la superficie del substrato corresponderá una concentración superficial de la impureza dentro de él. La dependencia entre la concentración de la especie activa en el sólido y su presión parcial junto a la superficie se muestra en la Fig. 4.9 para el caso de una fuente de B_2O_3 [3]. A medida que aumenta la presión parcial, la solubilidad de la especie en el sólido aumenta y alcanza un valor máximo para cada temperatura que corresponde con su valor de equilibrio según el diagrama de fase. Esta concentración recibe el nombre de solubilidad máxima.

De acuerdo a la Fig. 4.9. se pueden diferenciar dos regímenes de trabajo.

Régimen a solubilidad máxima: corresponde a la zona de saturación del la curva. El valor de concentración superficial de la especie C_s se obtiene de las curvas de solubilidad de la especie activa en función de la temperatura de predifusión.

Este régimen se alcanza cuando la tensión de vapor de la fuente es suficientemente alta.

Régimen lineal: corresponde a la zona lineal de la curva y se expresa a través de la ley de Henry, donde H es la constante de Henry. Cuando la concentración superficial de la especie depende linealmente de la tensión de vapor de la fuente para una temperatura dada.

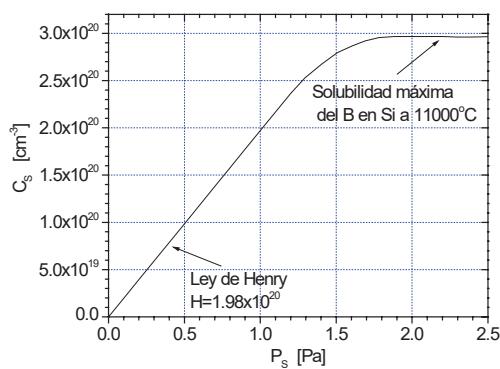


Fig. 4.9 Concentración superficial de boro en silicio en función de la presión parcial de B_2O_3 a la temperatura de 110 °C.

En el primer régimen, los valores de concentración superficial que se obtienen son mas repetibles pues no dependen de la temperatura de la fuente, pero solo se pueden obtener altas concentraciones de impurezas en el sólido.

En el segundo régimen, los valores de concentración superficial se controlan variando la temperatura de la fuente. Se pueden lograr menores concentraciones superficiales. Para lograr una buena repetibilidad de los

valores de C_s se requiere un alto grado de control de la temperatura de la fuente.

La dependencia entre la solubilidad máxima y la temperatura se obtiene del diagrama de fases, para un sistema de dos componentes. En este caso se utiliza el gráfico de composición en función de la temperatura para una presión constante.

Para garantizar **reproducibilidad** en los procesos se deberá contar con sistemas adecuados de control de temperatura, ya que el coeficiente de difusión depende exponencialmente de su valor. Los hornos deberán tener una distribución uniforme de temperatura en la zona donde se encuentren las obleas. A esta región se le denominará “*zona plana*” y la tolerancia será menor a 1°C. El tamaño de esta puede variar desde 10 cm para sistemas experimentales hasta más de un metro para producción. También es importante el control de flujos de los gases para garantizar la reproducibilidad.

En las líneas de producción es conveniente el procesamiento de un número grande de obleas de un área grande. Existe un compromiso entre el tamaño máximo y el costo y esto se determinará por los ingenieros de sistemas.

La difusión puede llevarse a cabo en sistemas abiertos o **sistemas cerrados**. Estos últimos consisten de tubos de cuarzo sellados al vacío y no son muy populares debido al costo que implica el sellado y la destrucción de los tubos para cada proceso. Estos solo se emplean para difusiones prolongadas. Su principal ventaja es la limpieza y estabilidad por otro lado la concentración superficial depende de la presión residual en los tubos

Los **sistemas abiertos** son más populares ya que son más aptos para el procesamiento en masa. En estos la difusión se realiza en ambiente de un

gas inerte como el argón, helio o nitrógeno, mismo que está fluyendo desde uno de los extremos del tubo. El gas impide la contaminación desde el extremo abierto, purgando el sistema. Las obleas se introducen en el extremo opuesto. En estos sistemas se debe considerar que hay un tiempo en que las obleas se introducen hasta llegar a la zona plana, el cual debe ser considerado ya que hay una difusión parcial.



Fig.4.10 Esquema de un sistema de difusión a partir de una fuente sólida.

Para estructuras sencillas una alternativa es la aplicación de **pinturas** directamente sobre la oblea, o bien centrifugar “*spin on*” una solución similarmente a la forma en que se aplican las resinas fotosensibles.

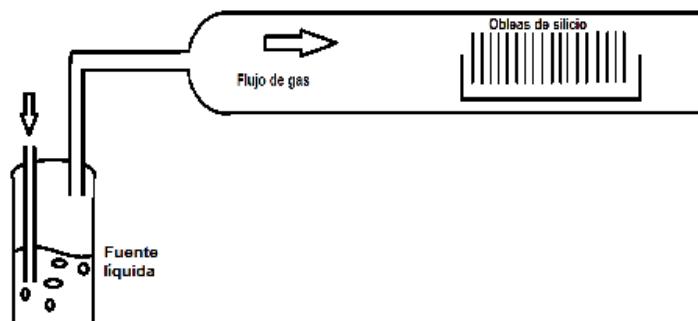


Fig. 4.11 Esquema de un sistema de difusión a partir de una fuente líquida de impurezas.

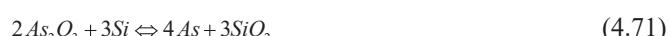
Las **fuentes sólidas** se encuentran en el interior de los sistemas de difusión. La impureza se transporta por sublimación hacia las obleas. En algunos casos estas se encontrarán en un crisol, aunque también son populares las fuentes en forma de obleas que se intercalan entre las de material semiconductor. La difusión en estos casos es muy uniforme.

Las **fuentes líquidas** se colocan fuera del reactor y su flujo se controla mediante burbujeo. Esto es, su temperatura se mantiene constante y se introduce un gas inerte mediante un tubo al interior del recipiente. El gas se satura con el líquido al salir. Conociendo la presión de vapor del compuesto se puede determinar con precisión la cantidad que se introduce al sistema de difusión. Una ventaja de este tipo de fuente sobre las anteriores es que permite iniciar y terminar el proceso abriendo o interrumpiendo los flujos.

También en principio pueden usarse **fuentes gaseosas** que se almacenan en tanques presurizados y se controlan por flujo de masa. Sin embargo su manipulación es compleja al ser productos muy tóxicos y generalmente no se emplean en hornos abiertos.

4.4 IMPUREZAS EN SILICIO

La difusión de impurezas substitucionales generalmente se realiza a partir del óxido:



Para ello se puede introducir oxígeno o vapor de agua durante la oxidación o bien, en el caso de fuentes sólidas, oxidarse ante de proceso.

También se da el caso de hacer un predepósito de la fuente para posteriormente oxidarlo a alta temperatura.

Cuando la impurificación es baja y el material es intrínseco a la temperatura del proceso (Fig. 4.4) el coeficiente de difusión se comporta idealmente. En este caso es fácil calcular los perfiles de impurezas a partir de las ecuaciones simplificadas con buena concordancia con los resultados experimentales. La figura 4.12 muestra una gráfica de Arhenius del comportamiento de impurezas substitucionales y la 4.13 para impurezas intersticiales en Si.

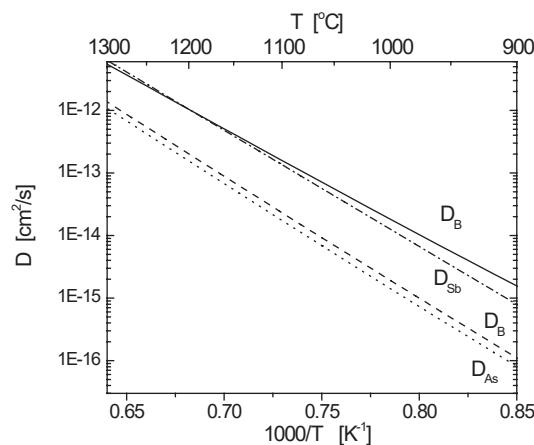


Fig. 4.12 Coeficientes de difusión de impurezas substitucionales en silicio [4].

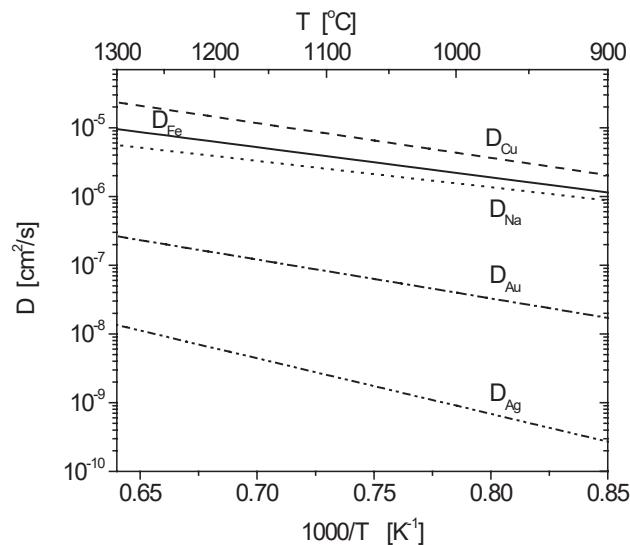


Fig. 4.13 Coeficientes de difusión de impurezas substitucionales en silicio [4].

Tabla 4.1 Coeficientes de difusión de impurezas substitucionales en silicio [5]

		P	As	Sb	B	Al	Ga
D_i^0	D_0 [cm/s]	3.85	0.066	0.214	0.037	1.385	0.374
	E_0 [eV]	3.66	3.44	3.65	3.46	3.41	3.39
D_i^+	D_0 [cm/s]	-	-	-	0.76	2.480	28.5
	E_0 [eV]	-	-	-	3.46	4.20	3.92
D_i^-	D_0 [cm/s]	4.44	22.9	13	-	-	-
	E_0 [eV]	4.0	4.1	4.0	-	-	-
D_i^{2-}	D_0 [cm/s]	44.2	-	-	-	-	-
	E_0 [eV]	4.37	-	-	-	-	-

A altas concentraciones la difusión puede estar dominada por efecto de vacancias cargadas. Las tablas 4.1 y 4.2 muestran los factores que permiten calcular los diferentes coeficientes a partir de

$$D^i = D_0^i \exp - (E_D / kT) \quad (4.72)$$

Tabla 4.2 Coeficientes de difusión de impurezas intersticiales en silicio

Elemento	D ₀ [cm ² /s]	E _a [eV]	Ref.
Na	1.6x10 ⁻³	0.76	[6]
Cu (800-1100 °C)	0.04	1	[6]
Ag (800-1200 °C)	2x10 ⁻³	1.6	[6]
Au	1.1x10 ⁻³	1.12	[6]
Fe	6.2x10 ⁻³	0.87	[6]

4.4.1 Difusión de boro.

La difusión de boro puede ser anisotrópica, esto es puede depender de la orientación cristalina. Este efecto se presenta cuando el proceso se realiza en una atmósfera oxidante y se atribuye a la generación de defectos (vacancias) por la oxidación de la superficie¹⁴.

En atmósfera neutra el coeficiente de difusión de boro presenta participación de vacancias cargadas positivamente de la forma:

$$D = h \left[D_i^0 + D_i^+ \left(\frac{p}{n_i} \right) \right] \quad (4.73)$$

¹⁴ Recordar que el proceso de difusión involucra la participación de vacancias.

en este caso la solución es numérica. La profundidad de unión se puede aproximar por:

$$x_j \approx 2.45 \left(\frac{N_{\text{sup}}}{n_i} \right)^{1/2} (Dt)^{1/2} \quad (4.74)$$

donde $D = 3.17e^{-3.59/kT}$, y el perfil de impurezas se puede aproximar como:

$$N \approx N_{\text{sup}} (1 - Y)^{2/3} \quad (4.75)$$

donde

$$Y = \left(\frac{x^2}{6D_{\text{sup}}t} \right) \quad (4.76)$$

El exceso de boro durante el proceso da lugar a formación de compuestos de boro (piel de boro), difíciles de remover, que son aislantes y pueden ser perjudiciales para los dispositivos.

Fuentes de Boro.

Como fuentes sólidas se utilizan el B_2O_3 y el nitruro de boro. El B_2O_3 se coloca en la zona posterior del horno de difusión a una temperatura del orden de 1050 °C, para que se funda. El gas de arrastre transporta la especie activa hasta la zona central del horno donde se encuentran colocadas las obleas que se desean difundir. Debe señalarse que a pesar de la relativamente alta temperatura de la fuente, el proceso se realiza en régimen lineal (Ley de Henry), pues la tensión de vapor a esa temperatura es menor de 0.0133 Pa (10^{-4} Torr). Se logra una concentración de boro del orden de 10^{19} cm^3 . Es importante humedecer el gas de arrastre para evitar que se ataque la superficie del silicio durante la predifusión.

El nitruro de silicio se usa en forma de láminas que se colocan alternándose con las obleas de silicio. Se oxidan previamente a ser utilizadas, para que la difusión se realice a partir del óxido. Con esta

técnica se alcanza una concentración superficial de impurezas del orden de 10^{20} cm^{-3} .

Como fuente líquida de boro se puede usar el BBr_3 , así como el tribromuro de boro BBr_3 . Esta última tiene la ventaja de actuar como pasivante, ya que el bromo resultante puede atrapar los iones alcalinos, eventualmente puede manchar y atacar la oblea a altas concentraciones debido al efecto del bromo. Igualmente se introduce oxígeno durante el predepósito. Requiere de manejo adecuado de los desechos de bromo.

También puede usarse el trimetilborato $(\text{CH}_3\text{O})_3\text{B}$, que es un compuesto muy volátil, por lo que la fuente se refrigerá. Se introduce oxígeno al horno durante el predepósito para formar óxido.

También en procesos menos exigentes puede utilizarse pinturas y centrifugados, “Spin-On”. Se utilizan mezclas de B_2O_3 y SiO_2 en alcohol polivinílico, carboranos o alquilsiloxanos como solventes. Una vez depositados se les da un tratamiento térmico para evaporar o descomponer el solvente y convertir el compuesto en borosilicato.

Como fuente gaseosa puede usarse el diborano B_2H_6 . Sin embargo este gas es extremadamente venenoso, por lo que se usa disuelto en argón (al 1/1000) para disminuir los riesgos. Se puede oxidar a partir de oxígeno o de dióxido de carbono la reacción lo descompone casi en su totalidad, sin embargo el gas a la salida del horno se hace pasar en HCl para eliminar los posibles restos de diborano. De todas maneras esta técnica no se usa en hornos abiertos. Si se utiliza en crecimiento epitaxial y depósitos por CVD.

En la Tabla del Anexo 4 se muestra la tensión de vapor de diferentes fuentes de impurezas en función de la temperatura de las fuentes.

4.4.2 Difusión de fósforo.

A altas concentraciones el coeficiente de difusión es complejo y estará dominado por diferentes mecanismos cuando el nivel de Fermi atraviesa los niveles de carga de las vacancias:

$$D_{\text{eff}} = h \left[D_i^0 + D_i \left(\frac{n}{n_i} \right) + D_i^2 \left(\frac{n}{n_i} \right)^2 \right] \quad (4.77)$$

A altas concentraciones D es proporcional a N^2 , es muy rápido y está dominado por las vacancias negativas cargadas doblemente. En esta zona el perfil es casi plano. Al interior, debido a una menor concentración, cuando el nivel de Fermi atraviesa E_V^{2-} el mecanismo lo dominarán las vacancias $[V^-]$ y D será proporcional a N. Por último, en la cola se observa un coeficiente de difusión constante. La figura 4.14 muestra el comportamiento idealizado de la difusión de fósforo,. Hay que considerar que a altas concentraciones de fósforo el perfil eléctrico no coincide con el químico, esto se debe a que siendo el material degenerado, no todos los átomos de fósforo están ionizados [7].

4.4.2.1 Fuentes de fósforo

Pinturas y Spin-On. Compuestos orgánicos (trifenilfosfato). Similarmente que en el caso del boro se les da un tratamiento térmico para evaporar el solvente y convertir la película a fosfossilicato.

Fuentes sólidas. Dado que el pentóxido de fósforo es higroscópico, esto es absorbe la humedad, la utilización de éste es impráctica en sistemas de difusión. Los materiales mas usados son difosfato y monofosfato de amonio. También pueden ser usados en forma de disco, donde los compuestos se integran a obleas de cerámica.

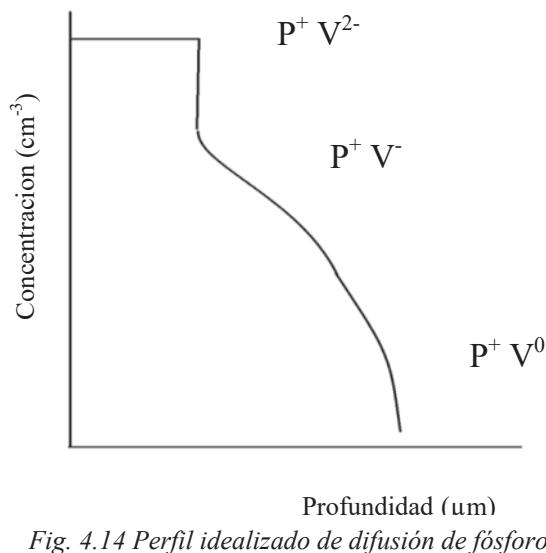


Fig. 4.14 Perfil idealizado de difusión de fósforo.

Las fuentes de fósforo mas utilizadas son las líquidas, como el oxicloruro de fósforo ($POCl_3$), el tribromuro de fósforo (PBr_3) y el tricloruro de fósforo PCl_3 . La fuente de PBr_3 tiene la ventaja de que el bromo resultante atrapa iones alcalinos.

Como fuente gaseosa se utiliza la fosfamina (PH_3), que es altamente tóxico, por lo que se usa diluido y en aplicaciones como se indicó en el caso del diborano.

En la Tabla del Anexo 4 se muestra la tensión de vapor de diferentes fuentes de impurezas en función de la temperatura de las fuentes.

4.4.2.2 Difusión de arsénico y de antimonio

Su coeficiente de difusión es aproximadamente un décimo de los valores de boro o fósforo. Por ello se emplea cuando es necesario que una difusión quede relativamente inmóvil en los procesos térmicos siguientes. Solo presenta interacción con vacancias negativas cargadas unitariamente.

$$D_{\text{eff}} = h \left[D_i^0 + D_i^{-} \left(\frac{n}{n_i} \right) \right] \quad (4.78)$$

A altas concentraciones se puede usar la siguiente aproximación para D:

$$D \approx 2D_i^{-} \left(\frac{N}{n_i} \right) \quad (4.79)$$

y

$$D_{\text{sup}} \approx 2D_i^{-} \left(\frac{N_{\text{sup}}}{n_i} \right) \quad (4.80)$$

y para la profundidad de unión:

$$x_j = 1.616(D_{\text{sup}} t)^{1/2} \quad (4.81)$$

El perfil se puede aproximar por el polinomio:

$$N = N_{\text{sup}} \left(1 - 0.87Y - 0.45Y^2 \right) \quad (4.82)$$

donde

$$Y = x \left(4D_{\text{sup}} t \right)^{1/2} \quad (4.83)$$

Como fuente sólida se puede usar Arsénico elemental disuelto en SiO₂

El manejo de las fuentes es similar al boro y al silicio y se usan arsenosiloxanos para pinturas y Spin-On. y de fuente gaseosa se utiliza arsenamina, (AsH_3). Esta última es también extremadamente tóxica. La difusión de arsénico se realiza mucho en tubo cerrado, ya que a veces los procesos son muy largos. El arsénico se evapora fácilmente provocando empobrecimiento de la superficie.

El antimonio no es tan popular como impureza y no está tan estudiado como el boro o el fósforo. Como fuentes de antimonio también se utilizan fuentes líquidas de SbBr_3 , SbCl_3 . Se puede utilizar trióxido de antimonio (Sb_2O_3) o tetróxido de antimonio (Sb_2O_4) como fuente sólida y pentacloruro de antimonio (Sn_3Cl_5) como fuente líquida.

En la Tabla del Anexo 4 se muestra la tensión de vapor de diferentes fuentes de impurezas en función de la temperatura de las fuentes.

4.4.3 Impurezas Intersiticial-Substitucional

Los donores o aceptores en silicio no presentan este comportamiento. El oro se utiliza para disminuir los tiempos de recuperación de las uniones pn. Este es un difusor extremadamente rápido y se evapora en la parte opuesta de la oblea a la que tiene la estructura de dispositivos. La temperatura empleada para el oro oscila entre 800 a 1050 °C y los tiempos del proceso tan solo de 10 a 15 minutos. Esto es suficiente para atravesar homogéneamente la oblea.

La tabla 4.3 muestra los coeficientes aproximados de otras impurezas en silicio que muestran este mecanismo.

Tabla 4.3 Coeficientes de difusión de impurezas intersiticial-substitucional en silicio.

	Ag	Au	Co	Cu	Fe	Ni	O	S	Zn
D ₀	2x10 ⁻³	1.1x10 ⁻³	0.16	4x10 ⁻²	6.2x10 ⁻³	1.3x10 ⁻³	0.21	0.92	0.1
[cm ² /s]	1.6	1.12	1.12	1.0	0.87	1.4	2.44	2.2	1.4
E ₀									

4.5 DIFUSIÓN EN GAAS

La difusión es un proceso que se emplea muy poco en GaAs. una de las excepciones es el zinc, que sirve para la fabricación de uniones pn. El zinc es un difusor rápido y presenta un mecanismo substitucional-intersticial.

El zinc substitucional es un aceptor y ocupa un lugar de galio quedando ionizado negativamente. Este puede saltar a un sitio intersticial generando una vacancia de galio a la vez que absorbe tres huecos. La reacción resultante de este proceso sería:



aplicando la ley de acción de masas a la ecuación anterior y suponiendo que la mayor parte de los átomos de zinc están en sitios substitucionales:

$$K_1 = \frac{[Zn_s]p^3}{[Zn_i][V_{Ga}]} = \frac{N_s p^3}{N_i [V_{Ga}]} \quad (4.85)$$

suponiendo que $p = Zn_s$ y $[V_{Ga}] \propto p_{As}^{1/2}$ se tendrá:

$$K_2 = \frac{N_s^4}{N_i p_{As}^{1/2}} \quad (4.86)$$

o sea que la concentración de átomos intersticiales será:

$$N_i = \frac{N_s^4}{K_2 p_{As}^{1/2}} \quad (4.87)$$

La aproximación para impurezas con este mecanismo es (ec. 4.68):

$$D \approx D_i \frac{\partial N_i}{\partial N_s} \quad (4.88)$$

y de (4.87) se puede derivar que :

$$\frac{\partial N_i}{\partial N_s} = \frac{4N_s^3}{K_2 p_{As}^{1/2}} \quad (4.89)$$

entonces el coeficiente de difusión efectivo será:

$$D \approx D_i K_3 [Zn]^3 p_{As}^{-1/2} \quad (4.90)$$

Esto es, el zinc será un difusor muy rápido a altas concentraciones y tenderá a ser lento para material intrínseco. Hay que considerar que dado el mayor ancho de banda del GaAs, este casi siempre será intrínseco a la temperatura de difusión.

4.6 INTRODUCCIÓN DE IMPUREZAS A TRAVÉS DE ÓXIDO DE SILICIO

El óxido de silicio puede ser utilizado en la Tecnología Planar como capa para evitar la difusión de impurezas en el substrato en determinadas regiones, gracias a que el coeficiente de difusión en el SiO₂ de impurezas de boro, fósforo, arsénico y antimonio, es menor que en el silicio. Para conocer el espesor requerido de la capa de óxido para producir este enmascaramiento hay que resolver el problema de difusión en una doble capa:

$$D_l \frac{\partial^2 C_1}{\partial x^2} = \frac{\partial C_1}{\partial t} \quad \text{para } -X_o < x < 0 \quad (4.91)$$

$$D_2 \frac{\partial^2 C_2}{\partial x^2} = \frac{\partial C_2}{\partial t} \quad \text{para } x > 0 \quad (4.92)$$

con las condiciones siguientes:

En $t=0$ $C_1=C_2=0$

$C_2 \rightarrow 0$ para $x \rightarrow \infty$

$C_1=C_s$ para $x=-X_o$;

El flujo de átomos que dejan el óxido es igual al que penetra en el silicio:

$$D_1 \frac{\partial^2 C_1}{\partial x^2} = D_2 \frac{\partial^2 C_2}{\partial x^2} \quad \text{en } x=0 \quad (4.93)$$

debido al proceso de segregación de las impurezas en la frontera entre el SiO_2 y el silicio, debe cumplirse que:

$$C_2(0, t) = m \cdot C_1(0, t) \quad (4.94)$$

donde m es el coeficiente de segregación (Fig. 4.16.).

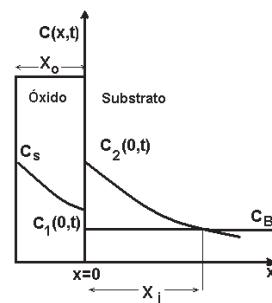


Fig. 4.16 Difusión de una impureza a través de una estructura SiO_2 - Si .

La solución de este problema da una expresión en forma de una serie infinita del tipo de:

$$C_1 = C_s \sum_{j=0}^{\infty} \left[\frac{m-r}{m+r} \right]^j \cdot \left[\operatorname{erfc} \left(\frac{X_o(2j+1)+x}{2\sqrt{D_1 t}} \right) - \left(\frac{m-r}{m+r} \right) \cdot \operatorname{erfc} \frac{X_o(2j+1)-x}{2\sqrt{D_2 t}} \right] \quad (4.95)$$

$$C_2 = \frac{2rmC_s}{m+r} \sum_{j=0}^{\infty} \left[\left(\frac{m-r}{m+r} \right)^j - erfc \left(\frac{X_o(2j+1)}{2\sqrt{D_1 t}} + \frac{x}{2\sqrt{D_2 t}} \right) \right] \quad (4.96)$$

donde

$$r = \sqrt{\frac{D_1}{D_2}} \quad (4.97)$$

Si $X_o > \sqrt{4D_1 t}$, la solución puede aproximarse al primer término de la serie cuando $j=0$, ya que la función error complementario decrece muy rápidamente. Las expresiones (4.95) y (4.96) pueden escribirse en ese caso, de la forma:

$$C_1 = C_s \cdot \left[erfc \left(\frac{X_o + x}{2\sqrt{D_1 t}} \right) - \left(\frac{m-r}{m+r} \right) \cdot erfc \frac{X_o - x}{2\sqrt{D_2 t}} \right] \quad (4.98)$$

$$C_2 = \frac{2rmC_s}{m+r} \cdot erfc \left(\frac{X_o}{2\sqrt{D_1 t}} + \frac{x}{2\sqrt{D_2 t}} \right) \quad (4.99)$$

Si se quiere que no se forme unión en las zonas protegidas por el óxido, se debe sustituir en la expresión (4.100) la condición de que la unión se produzca como caso límite en $x=0$, o sea,

$$C_2(x_j, t) = C_s \quad \text{en } x_j = 0 \quad (4.100)$$

y

$$\frac{2rmC_s}{m+r} \cdot erfc \left(\frac{X_o}{2\sqrt{D_1 t}} \right) = C_s \quad (4.101)$$

En caso general puede calcularse la profundidad de la unión en función del espesor de óxido o la profundidad a que penetra la impureza en el óxido para una temperatura y tiempo de difusión dados. Estos resultados se pueden dar en forma de gráficos.

4.7 DIFUSIÓN HACIA AFUERA O EXODIFUSIÓN

Se llama difusión hacia afuera o exodifusión, al proceso mediante el cual la especie activa, presente en el semiconductor, sale de este hacia el medio que lo rodea. La ecuación (4.41) se resuelve para las condiciones de frontera

$$C(\infty, t) = C_s \quad (4.102)$$

$$gV(x, 0) = -D \frac{\partial C}{\partial x} \Big|_{(0, t)} \quad (4.103)$$

y la condición de inicial

$$C(x, 0) = C_s \quad (4.104)$$

La solución es de la forma:

$$C(x, t) = C_s \operatorname{erf} \left(\frac{x}{2\sqrt{Dt}} \right) + \exp \left[-\frac{ht}{\sqrt{Dt}} \right] \cdot \frac{x}{2\sqrt{Dt}} + \left(\frac{ht}{\sqrt{Dt}} \right)^2 \cdot \operatorname{erfc} \left(\frac{ht}{\sqrt{Dt}} + \frac{x}{2\sqrt{Dt}} \right) \quad (4.105)$$

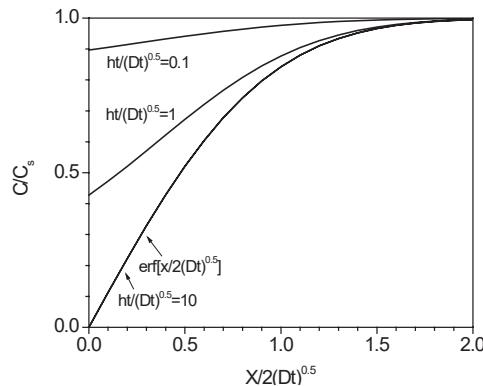


Fig. 4.17 Distribución de impurezas en el semiconductor luego de un proceso de redistribución hacia fuera en función del parámetro $\frac{ht}{\sqrt{Dt}}$.

En la *Fig. 4.17* se muestran perfiles de concentración para diferentes valores de $\frac{ht}{\sqrt{Dt}}$, cuando ocurre este efecto.

4.8 REDISTRIBUCIÓN DE IMPUREZAS

Cuando se crece un óxido de silicio consume una película de silicio de aproximadamente el 45% el espesor del óxido. Esto implica un cambio en el perfil de las impurezas que estaban originalmente en este lugar debido al incremento del volumen por la incorporación del oxígeno, al establecimiento de condiciones de equilibrio en la interfaz y a la misma difusión a que están sujetas las impurezas. Esta situación es común tanto cuando se realiza un proceso de redistribución de impurezas, ya que normalmente se oxida la superficie para evitar una exodifusión, como en los diferentes procesos tecnológicos de enmascaramiento propios de la tecnología del silicio.

La concentración de impurezas en el óxido es inferior a la que originalmente tenía el silicio en:

$$C_{SiO_2} = 0.45C_{Si} \quad (4.106)$$

El equilibrio se establece en la interfaz Si-SiO₂ de acuerdo al coeficiente de distribución o coeficiente de segregación:

$$k = \frac{C_{Si}}{C_{SiO_2}} \quad (4.107)$$

Que relaciona las concentraciones de impurezas en el silicio y en el óxido. Adicionalmente los procesos térmicos a que se someta la oblea darán lugar a difusión entre las mismas impurezas, tanto en el óxido como en el semiconductor, de acuerdo a las leyes de Fick. Esto modifica la distribución de impurezas en la interfaz como se muestra en la figura 4.18.

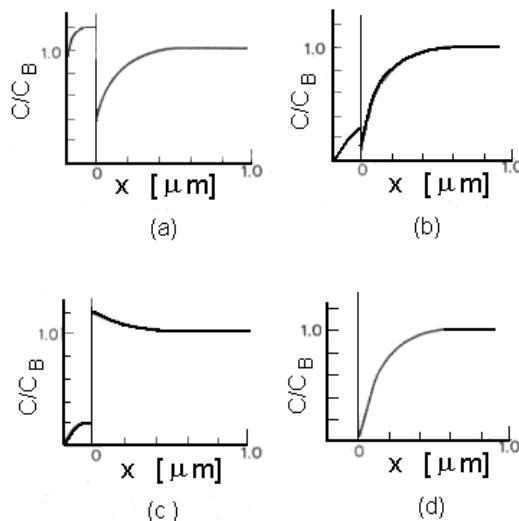


Fig. 4.18 Distribución de impurezas en silicio después del crecimiento de una película de óxido. a) $k < 1$, difusión en óxido lenta, b) $k < 1$, difusión en óxido rápida, c) $k > 1$, difusión de óxido lenta, d) $k > 1$, difusión en óxido rápida.

Esto es, habrá una discontinuidad en la interfaz de acuerdo al valor de k , por otro lado siempre se presentará un empobrecimiento en la superficie del óxido debido a la exodifusión. Cuando el coeficiente de difusión de la impureza en óxido sea pequeño, como es el caso de la mayor parte de las impurezas empleadas (boro, fósforo, arsénico) el efecto de empobrecimiento solo se notará en la región muy cercana a la superficie, mientras que en el caso de que el coeficiente de difusión en el óxido sea rápido (galio), el efecto se extenderá incluso hasta el silicio [8].

Esto es si el coeficiente de segregación es menor a la unidad se tendrá un empobrecimiento en la superficie del semiconductor, mientras que si es mayor a uno se dará una acumulación. Por otro lado, si el coeficiente de

difusión en el óxido es alto, se presentará exodifusión, que eventualmente también puede afectar la distribución en la interfaz.

4.8.1 Efectos de interacción

La presencia de otras impurezas en la estructura modifica los coeficientes de difusión.

Uno de los efectos más conocidos es el llamado “emitter push” o empuje del emisor. Debido a que una región de la base debe llegar hasta la superficie para posteriormente poder poner contactos óhmicos, el área de emisor es menor que la de la base. Al realizar la difusión de emisor la base se difundirá adicionalmente de acuerdo a las leyes de Fick, sin embargo el coeficiente de difusión en la zona bajo el emisor será mayor, por lo que se verá una deformación de la difusión de base como se ve en la figura 4.19.

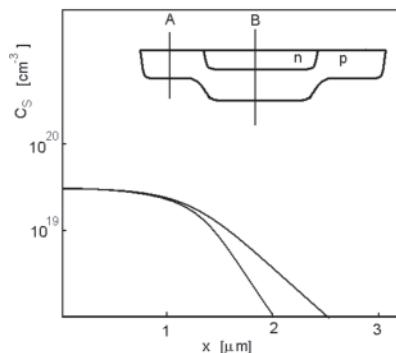


Fig. 4.19. Difusión anormal en la base debido a la interacción de las impurezas de emisor.

Existen dos teorías para la explicación de este fenómeno:

Generación de un campo eléctrico por los portadores libres generados por las impurezas de emisor. Considerando que la concentración de emisor es muy alta con respecto a la base, por lo que la primera región sería extrínseca, mientras que la segunda intrínseca. Dada la polaridad opuesta

entre base y emisor, los portadores libres generados por las impurezas del emisor tenderán a repeler a las impurezas de base, incrementando su difusión.

Generación de defectos en la rede cristalina. Dada la alta concentración de impurezas en el emisor se genera una deformación de la red cristalina, que puede ser un efecto de compresión, en el caso de que las impurezas tengan un menor radio tetraedral, o de expansión en el caso inverso. Esta deformación favorece la formación y difusión de vacancias en la red cristalina, lo que favorece la difusión de impurezas. Este efecto no se percibe en el caso del emisor, ya que son varios los factores que modifican el coeficiente de difusión a altas concentraciones, sin embargo para el caso de la base se notará una diferencia en el perfil de impurezas en la parte cercana al emisor con respecto a la parte alejada al mismo.

4.8.2 Difusión bidimensional

Los modelos expuestos en este capítulo consideran que la difusión se realiza en una sola dimensión. Esta aproximación es válida cuando las dimensiones de las estructuras son muy grandes con respecto a la profundidad de la difusión. Sin embargo en los bordes de las mascarillas también habrá difusión hacia los lados manteniéndose un perfil cilíndrico, donde la profundidad de difusión es alrededor de 0.7 del valor hacia el interior de la oblea. También encontraremos efectos de dos dimensiones en el caso de que las dimensiones de la ventana sean comparables a la profundidad de difusión en este caso el perfil, en coordenadas polares, se puede aproximar como:

$$N = N_s e^{-\rho/\sigma} \quad (4.108)$$

donde ρ es la distancia de la superficie.

BIBLIOGRAFIA

- [1] B.I. Boltaks, Difusión in Semiconductors, Academia , New Cork, 1963
- [2] J.Crank, The Mathematics of Diffusion, Oxford University Press, London, 1070.
- [3] T.H. Yen y W. Armstrong, “Diffusion of Boron in Silicon”, Electrochemical Society Spring Meeting Abstract No. 69, Indianapolis 1961.
- [4] Tabla de magnitudes físicas, editado por K.K. Kikoin, Moscú, Atomisdat, 1976.
- [5] R.B. Fair “Silicon Integrated Circuits”, Part B in Applied Solid Science Series Supplement 2, D. Kahng, ed., Academic Press, NY, 1981.
- [6] B.L. Sharma, “Diffusion in Semiconductors”, Trans. Tech. Publications, Clausthal, Germany, 1970.
- [7] K.H. Nicholas, “Studies of anomalous diffusion of impurities in Silicon”, SSE, 9 (1966) 35.
- [8] A.S. Grove, O. Leistiko y C.T. Sah, “Redistribution of acceptor and donor impurities during thermal oxidation of silicon”, Appl. Physic, 35 (1964) 2665.

CAPÍTULO 5

EPITAXIA

Una gran proporción de los dispositivos de semiconductores se fabrica a partir de obleas cortadas de lingotes monocristalinos. Sin embargo en muchos la oblea sirve solamente como soporte mecánico, y las estructuras se fabrican sobre capas epitaxiales.

El término epitaxia proviene del griego $\varepsilon\pi\iota$ (sobre) y $\tau\alpha\xi\eta$ (arreglo), y se aplica al depósito de películas monocristalinas sobre un sustrato. La estructura y orientación de capa y sustrato no necesariamente tienen que ser idénticas. Un depósito será epitaxial si la película crecida es monocristalina.

Cuando la película crecida es de la misma naturaleza que el sustrato se hablará de “*homoepitaxia*” o “*autoepitaxia*”. Este es el caso de Silicio sobre silicio, GaAs en GaAs, etc. El objetivo de depositar capas de la misma naturaleza del sustrato es tener material de mejor calidad o la fabricación de estructuras especiales.

Cuando los materiales de sustrato y depósito son diferentes se hablará de heteroepitaxia. Como ejemplos de este caso se pueden citar GaAs/Ga_xAl_{1-x}As, GaAs/Si, Ga_{1-x}In_xAs/InP. Es factible que los crecimientos se realicen sobre cristales de diferentes redes cristalinas, por ejemplo se puede crecer Si(100) sobre Zafiro (0112), teniendo el primero estructura de diamante, mientras que el segundo es hexagonal.

Una película epitaxial produce un material de mejor calidad, con menor cantidad de defectos y mayor pureza que una oblea obtenida a partir del fundido debido principalmente a que los procesos se realizan a una

temperatura considerablemente menor que el punto de fusión del semiconductor¹⁵. Adicionalmente, el control de los procesos es mas fino y preciso.

Las estructuras realizadas a partir de procesos epitaxiales se pueden diseñar con mas libertad que los procesos de difusión, ya que no se tienen las limitaciones asociadas por las matemáticas que regulan este último. Así es posible la realización de perfiles hiperabruptos, o tener capas con alta concentración de portadores entre películas de baja impurificación.

La heteroepitaxia ha permitido el desarrollo de dispositivos como los diodos láser, transistores de heterounión o sensores ópticos ajustados a un rango de sensibilidad específico. También ha sido fundamental en el estudio de estructuras cuánticas, mismas que deberán ser la base de muchos dispositivos en el futuro.

En silicio las películas epitaxiales se utilizan para tener material con mayor tiempo de vida de portadores minoritarios y de mayor resistividad. Los circuitos integrados con tecnología bipolar se realizan generalmente sobre material epitaxial. Corrientemente las capas crecidas en este material están en el rango de 2 a 5 μm .

La mayor parte de los dispositivos a base de arseniuro de galio se realizan a partir de capas epitaxiales, como capas activas y de bajo ruido de transistores de efecto de campo, diodos emisores de luz, diodos láser etc. El rango de espesores en películas epitaxiales de GaAs puede ir desde monocapas¹⁶ hasta depósitos de 100 μm .

¹⁵ Recuérdese que la concentración de vacancias, el coeficiente de difusión, así como otros defectos son térmicamente activados y se incrementan exponencialmente con la temperatura.

¹⁶ Hay una técnica denominada en inglés *Atómico Layer Epitaxy* (ALE), donde se depositan las capas atómicas una a una.

5.1 LOS PROCESOS EPITAXIALES DE ACUERDO AL MEDIO DE TRANSPORTE.

Para realizar un depósito epitaxial es necesario un medio de transporte de los elementos que componen la película hasta el sustrato. En la superficie del mismo se realiza una serie de procesos conducentes a la incorporación de los elementos a la red cristalina de una forma controlada. Los procesos epitaxiales son altamente sensibles a la contaminación y deben ser realizados en las máximas condiciones de pureza. De acuerdo al medio de transporte los procesos epitaxiales se pueden clasificar en:

Epitaxia por Haces Moleculares (Molecular Beam Epitaxy o MBE).

Los elementos se depositan por evaporación en un sistema de vacío. Permite el uso de varias técnicas de caracterización. Es caro.

Epitaxia en Fase de Vapor. (Vapor Phase Epitaxy, VPE). Se introducen compuestos químicos gaseosos a un reactor y se hacen reaccionar en la superficie de los sustratos. Es un conjunto de técnicas versátiles que permiten la producción en masa.

Epitaxia en Fase Líquida (Liquid Phase Epitaxy, LPE). El crecimiento se realiza a partir de la sobresaturación de una solución en contacto con un sustrato. El excedente tiende a cristalizarse sobre este último.

Existen varios casos particulares que pueden ser modificaciones a las técnicas anteriores o combinaciones de las mismas. Por ejemplo: se habla de MOVPE, (Metalorganic-VPE) cuando se utilizan compuestos organometálicos como reactivos, LPVPE (Low Pressure-VPE) si el proceso se realiza a presión reducida, CBE (Chemical Beam Epitaxy) si el sistema de haces moleculares utiliza compuestos químicos como precursores en lugar de elementos, etc.

5.2 EPITAXIA POR HACES MOLECULARES.

El proceso de epitaxia por haces moleculares consiste en un depósito por evaporación en alto vacío, calentando los materiales a evaporar y su posterior condensación sobre un sustrato cristalino. En el caso de los semiconductores compuestos se evapora cada elemento por separado: e.g. galio y arsénico para formar GaAs; indio y fósforo para obtener InP; cadmio y telurio para CdTe, etc. Para semiconductores elementales solo se evapora el mismo.

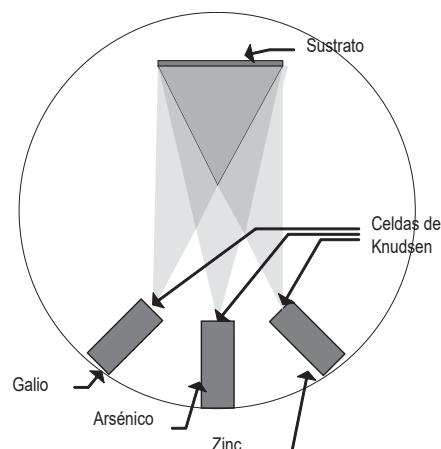


Figura 5.1 Sistema de epitaxia por haces moleculares

Gracias al alto vacío, los elementos tienen una trayectoria libre mucho más grande que las dimensiones del sistema y no hay colisiones entre moléculas. Por ello el transporte se controla por efectos térmicos y no, como es el caso de depósito en fase de vapor, por efectos de difusión, y la razón de depósito será proporcional a la razón de evaporación.

Dado que cada elemento tiene una presión de vapor diferente, cada uno se coloca en una cámara independiente, la cuál se calienta para obtener la presión parcial deseada. Estas se fabrican de cerámica o de algún otro material refractario que no contamine al ser calentado. La salida del vapor se realiza a través de una apertura que está diseñada para que se tenga una distribución uniforme sobre la oblea. A estas cámaras se les llama *Celdas de Efusión de Knudsen*.

Dado, dado que para la incorporación de átomos a la red cristalina son necesarios procesos de difusión superficial y de incorporación a la red, debe haber una temperatura suficientemente alta para favorecerlos. Los rangos típicos para procesos MBE oscilan entre 400 y 800°C.

El proceso se realiza en sistemas de ultra alto vacío (10^{-8} - 10^{-9} Torr) lo que hace que los sistemas sean de muy alto costo con la intención principal de reducir al mínimo la contaminación por los gases residuales (uno de los compuestos mas difíciles de eliminar es el agua debido a que se adsorbe fácilmente en las superficies).

El vacío se realiza por medio de bombas de sublimación de titanio o criogénicas de baja temperatura. Las cámaras de reacción generalmente son de acero inoxidable y se someten a procesos intensos de degasado mediante calentamiento a presiones muy bajas. Una vez puestas en operación no se exponen a la atmósfera sino en casos excepcionales de mantenimiento. La introducción de obleas se realiza por medio de esclusas, o sea, cámaras intermedias.

El alto vacío permite la instalación de varias técnicas de caracterización que se pueden usar simultáneamente al realizar el proceso de depósito, por lo que este tipo de sistemas es muy popular en centros de investigación para el estudio de los mecanismos de crecimiento de cristales.

La técnica de epitaxia por haces moleculares es particularmente útil cuando se requiere el depósito de películas muy delgadas, ya que permite un control de espesores muy preciso. Por ello se utiliza en la fabricación de superredes o estructuras cuánticas. Por otro lado, debido a que se realiza a temperatura relativamente baja no se utiliza para películas gruesas.

Debido a su alto costo y dado que la calidad no es mejor que las otras técnicas, los sistemas MBE no se han integrado en las líneas de producción.

5.3 EPITAXIA EN FASE DE VAPOR (VPE)

En un crecimiento epitaxial a partir de la fase de vapor los elementos se hacen reaccionar compuestos gaseosos de los elementos que se desean depositar. Los diferentes procesos que tienen lugar se resumen en el siguiente esquema:

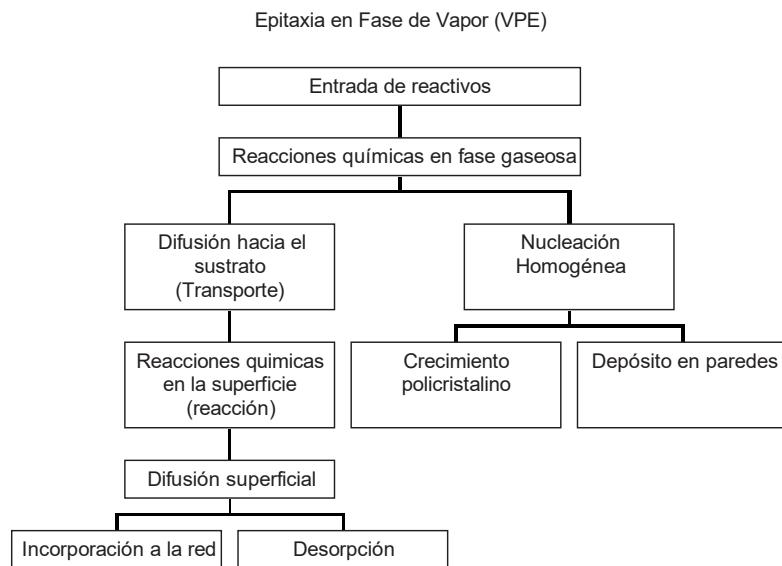


Diagrama a bloques de los procesos en el crecimiento epitaxial en fase de vapor.

Se introducen al reactor los compuestos químicos necesarios. Puede haber una reacción, generalmente por la temperatura, entre ellos antes de llegar al sustrato. En algunos casos el proceso estará dominado por los fenómenos de transporte (arrastre del flujo y difusión en el gas), mientras que también se da el caso de que las reacciones superficiales (reacciones químicas, difusión de átomos en la superficie e incorporación a la red cristalina) son las que controlan el proceso. La nucleación homogénea, o formación espontánea de núcleos es indeseable, ya que puede provocar crecimiento policristalino sobre el sustrato y depósito parásito sobre las paredes del reactor.

5.3.1 Transporte y Reacción.

El proceso epitaxial se realiza bajo un flujo de algún gas portador. Los flujos pueden ser turbulentos o laminares. El primero se caracteriza por la formación de remolinos y es impredecible. Este se da cuando existe una viscosidad alta del fluido, alta velocidad u obstáculos mecánicos. También los gradientes de temperatura pueden provocar turbulencias, debido a las corrientes de convección. En los procesos epitaxiales es importante tener flujos laminares.

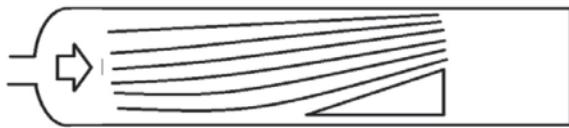


Fig. 5.2 Flujo laminar

En un flujo laminar el fluido se distribuye en capas o *láminas* paralelas entre si, este es predecible y se pueden calcular sus parámetros. En la cercanía de una superficie la velocidad del flujo se reduce a cero, debido a

los efectos de fricción y viscosidad, esta se incrementa conforme se aleja de la superficie. La región de velocidad variable entre la superficie y la corriente principal del flujo se denomina *capa frontera*.

La capa frontera se considera como de velocidad despreciable con respecto a la corriente principal. Cuando esta se inicia en un escalón tendrá una forma parabólica a partir su inicio.

Al introducir en el reactor un gas con una concentración constante de reactivos el transporte de los mismos hacia las superficies se realizará por difusión a través de la capa frontera. Si se considera que en la superficie de las obleas se consumen aquellos, se creará un gradiente de concentración. Aplicando la primera ley de Fick se tendrá:

$$j_t = -D \frac{dn}{dx} = -\frac{D}{\delta} (N_G - N_0) = h(N_G - N_0) \quad (5.1)$$

Donde: j_t : flujo de reactivos (por transporte)

D : Coeficiente de difusión en el gas.

δ : Espesor de la capa frontera.

N_G : Concentración del reactivo en el gas fuera de la capa frontera.

N_0 : Concentración del reactivo en el gas, en contacto con la superficie

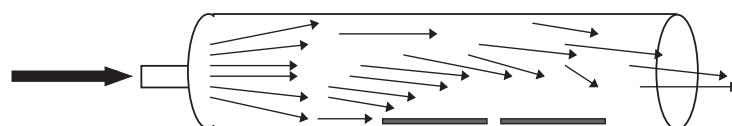


Fig. 5.3. Transporte de reactivos..

Por otro lado en la superficie del semiconductor, debido a las diferentes reacciones químicas, procesos de difusión, desorción y otros eventos , se consumirán los reactivos en proporción a la concentración N_0 . Podemos definir un flujo por reacción j_r como:

$$j_r = kN_0 \quad (5.2)$$

donde:

$$k = k_0 e^{-\Delta E_a / kT} \quad (5.3)$$

Entonces:

$$N_0 = \frac{j_r}{k} \quad (5.4)$$

Considerando estado estacionario, esto es no hay variación en la concentración a lo largo de la capa frontera, esto implica que el flujo por transporte será igual al flujo por reacción:

$$j = j_r = j_t; \quad (5.5)$$

y el flujo de impurezas estará dado por:

$$j = h \left(N_G - \frac{j}{k} \right) \quad (5.6)$$

Reordenando los términos se tendrá:

$$j \left(1 + \frac{h}{k} \right) = hN_G \quad (5.7)$$

$$j = N_G \frac{h}{\left(1 + \frac{h}{k} \right)} = N_G \frac{hk}{(h+k)} \quad (5.8)$$

Si n es el número de átomos en el cristal por volumen unitario y j el flujo de átomos por unidad de área, la razón de depósito de las películas (espesor por unidad de tiempo) estará dada por:

$$\frac{dx}{dt} = \frac{j}{n} = \frac{N_G}{n} \frac{hk}{(h+k)} \quad (5.9)$$

De acuerdo a la ec. 3, k depende exponencialmente de la temperatura. Si el valor de ésta última es bajo, h será mucho mayor que k , y podremos aproximar:

$$\frac{dx}{dt} = \frac{kN_G}{n} \quad (5.10)$$

En este caso se hablará de crecimiento limitado por reacción y serán los procesos en la superficie y las condiciones termodinámicas quienes determinen el espesor de la película.

En el caso inverso la siguiente aproximación será válida:

$$\frac{dx}{dt} = \frac{hN_G}{n} \quad (5.11)$$

En este caso se hablará de crecimiento limitado por transporte. Este es mas deseable, dado que solo habrá que controlar las cantidades de reactivos que se introducen al reactor y debido a que en un proceso limitado por reacción este es sensible a la introducción de otros gases, como puede ser necesario cuando se impurifica el material o se depositan estructuras de varias capas semiconductoras con características diferentes.

5.3.2 Nucleación

La transición de fase gaseosa a sólida requiere de la existencia de núcleos, los cuáles proveerán de la superficie necesaria para atrapar a las moléculas. Cuando la este proceso es espontáneo se llamará *nucleación*

homogénea en contraste con *nucleación heterogénea*, la cual se presenta cuando la condensación se realiza en una superficie existente. Esta última es la que genera la película epitaxial, mientras que la nucleación homogénea es indeseable ya que puede dar lugar a crecimiento policristalino y pérdida de reactivos.

Para comprender la formación espontánea de núcleos es conveniente partir de consideraciones termodinámicas de equilibrio.

La energía libre de Gibbs de una molécula en estado gaseoso está dada por:

$$G_v = kT \int_p^{P_\infty} \frac{dP}{P} = kT \ln \frac{P}{P_\infty} \quad (5.12)$$

donde: P : presión parcial (valor real que presenta el gas).

P_∞ : presión de vapor (valor de equilibrio del sólido con el gas).

Esto es, si la presión parcial del compuesto es inferior a la presión de vapor el valor de la energía libre será negativo, y habrá una tendencia para que las moléculas del sólido pasen a la fase gaseosa, en el caso contrario el gas tenderá a condensarse. Para todas las moléculas en el volumen:

$$\Delta G_v = -nkT \ln \frac{P}{P_\infty} \quad (5.13)$$

Por otro lado, el núcleo tendrá una superficie propia con una energía que estará dada por:

$$G_s = A * \gamma \quad (5.14)$$

donde γ es la energía por unidad de área.

La energía total de un núcleo será la suma de ambas:

$$\Delta G_{\text{hom}} = \left(\frac{4}{3} \pi r^3 \right) \Delta G_v + (4 \pi r^2) \gamma \quad (5.15)$$

De acuerdo a la ec. 5.12, si $P < P_\infty$ al aumentar el volumen del núcleo disminuirá la componente de la energía debido al mismo, por otro lado, el término asociado a la energía superficial siempre se incrementa de forma proporcional al área. De acuerdo a los criterios de equilibrio, la energía libre de un cristal tenderá a un valor mínimo, por ello ΔG tenderá disminuir. El núcleo será estable, o sea que puede subsistir solo si su energía libre se reduce al crecer. Esto será válido si el radio del mismo sobrepasa un valor crítico, que se obtiene al igualar la ec. 5.15 a cero y está dado por:

$$r_{\text{crit}} = -\frac{2\gamma}{\Delta G_v} \quad (5.16)$$

O sea que núcleos con un radio menor tenderán a desaparecer y si el radio es mayor aumentarán en tamaño.

La formación de núcleos que puedan tener el tamaño crítico se puede aproximar aplicando la función estadística de Boltzmann:

$$n_{\text{crit}} = n_0 e^{-\Delta G_{\text{crit}}/kT} \quad (5.17).$$

La nucleación homogénea produce depósitos parásitos sobre paredes y policristalinidad sobre el sustrato, por lo que debe evitarse. La reducción de la probabilidad de formación de núcleos críticos puede eliminarse mediante la reducción de la presión parcial del material (ec. 12), reduciendo la probabilidad de formación de estos disminuyendo el tiempo en que los reactivos se exponen a altas temperaturas, mediante las velocidades de flujo y presión total en el reactor, o mediante la presentación de superficies (sustratos) que propicien nucleación heterogénea en competencia con el primer proceso.

Cuando las moléculas o los núcleos se condensan sobre una superficie existente se presentará nucleación heterogénea. En este caso hay tres tipos de superficies que se pueden formar:

- Nucleo-vacío.
- Nucleo-sustrato.
- Sustrato-vacío.

Cada una de las cuales tendrá asociada una energía superficial γ (γ_{NV} , γ_{NS} y γ_{SV}). De acuerdo a los criterios de equilibrio cuando la energía sea mínima se tendrá estabilidad. Esto es que se tienen tres estados posibles:

- 1.- Que el núcleo no moje al sustrato, lo que implica tener una superficie sustrato-vacio y otra nucleo-vacio.
- 2.- Que el núcleo moje parcialmente al sustrato, donde se tendrían los tres tipos de superficies presentes.
- 3.- Que el núcleo moje totalmente al sustrato, donde no habría superficie sustrato-vacio, ya que esta sería cubierta por el núcleo.

$$\gamma_{SV} + \gamma_{NV} < \gamma_{NS} \quad \gamma_{SV} = \gamma_{NS} + \gamma_{NV} \cos\theta \quad \gamma_{NS} + \gamma_{NV} < \gamma_{SV}$$

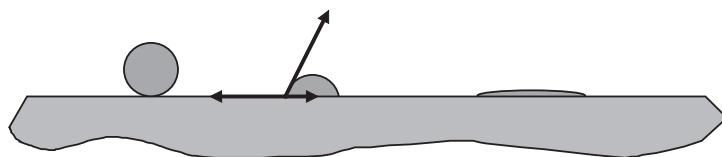


Fig. 5.4. Mojado de superficies.

Si la energía de la superficie del núcleo en contacto con el sustrato es inferior a la suma de las energías asociadas sustrato-vacio mas nucleo-vacio se dará el primer caso.

El tercer caso se presenta si la energía sustrato-vacio es inferior a la suma de las otras dos.

El segundo caso es intermedio de las dos anteriores, y el núcleo presentará un ángulo de mojado determinado por la condición de que la suma vectorial de las tres energías sea igual a cero.

Si hay mojado completo, la probabilidad de nucleación heterogénea será mayor que la de nucleación homogénea¹⁷. O sea que este será el proceso dominante en la cercanía del sustrato. Sin embargo si la sobresaturación en la fase gaseosa (P/P_∞) es muy grande, también habrá nucleación homogénea lejos del sustrato. Los núcleos formados podrán transportarse hasta el sustrato y provocar defectos de morfología superficial o policristalinidad¹⁸. Dado que la desviación del equilibrio depende de las condiciones termodinámicas y de las propiedades de los reactivos, esto deberá ser tomado en consideración para el diseño del sistema de epitaxia.

5.3.3 Diseño de sistemas.

Los procesos epitaxiales para materiales semiconductores demandan condiciones extremas de pureza. Los reactores generalmente se fabrican en cuarzo y las piezas metálicas de acero inoxidable pasivado y de alta calidad. Casi siempre se emplea hidrógeno como gas portador ya que este puede purificarse en la línea a muy bajo costo¹⁹. Los reactores de paredes calientes (calentamiento resistivo) favorecen la difusión de impurezas contaminantes de las paredes hacia las obleas, por lo que se prefieren los de paredes frías (calentamiento por radiofrecuencia o infrarrojo) cuando es posible.

¹⁷ Considerando que las moléculas individuales son núcleos en sí.

¹⁸ Los núcleos muy pequeños pueden orientarse con el cristal generando solamente defectos morfológicos, los que no se orientan provocan crecimiento policristalino.

¹⁹ Los sistemas de purificación a base de paladio son económicos y confiables.

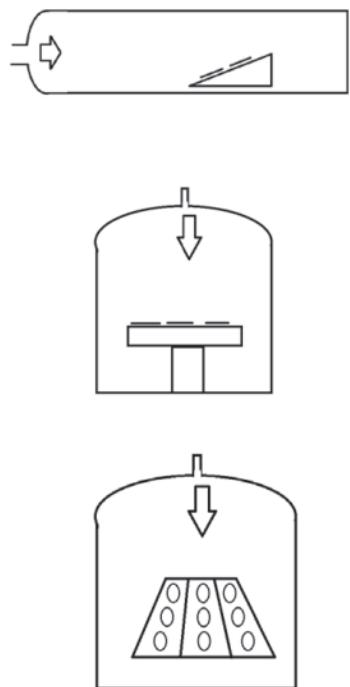


Fig. 5.5 Diferentes tipos de reactor (horizontal, vertical y de barril)

El estado de la superficie antes de iniciar el crecimiento es determinante de la calidad del depósito. Previamente a la introducción las obleas son limpiadas químicamente para eliminar contaminantes, sin embargo es prácticamente imposible evitar la formación de una capa muy delgada de óxido al exponerlas al aire durante el proceso de carga del reactor. El hidrógeno reduce estos óxidos en GaAs a 600°C sin embargo la remoción del SiO₂ es mas difícil y este tiene que ser calentado arriba de 1000°C para lograrlo. En algunos casos se hace decapado “*In situ*”, o sea en el mismo reactor introduciendo algún reactivo como HCl.

El control de parámetros es más crítico que en el caso de difusión, ya que en el caso de epitaxia no se trabaja en regímenes máximos. Los flujos se manejan por medio de controladores de flujo de masa.

El crecimiento a alta temperatura implica mejor calidad cristalina, mayor difusión de impurezas.

La orientación de los reactores puede ser horizontal o vertical. En este último caso se presentan problemas de convección y turbulencias, aunque permiten un mayor número de ondas que los reactores horizontales, los cuales presentan flujos laminares. Reactor de barril tiene las ventajas de flujos laminares y un alto número de ondas (Fig. 5.5.).

Los procesos epitaxiales a presión reducida tienen la ventaja de una mayor uniformidad debido a un incremento de la difusión en la fase gaseosa y mejor calidad cristalina y morfología al reducirse la nucleación homogénea. Por otro lado estos sistemas tienden a tener mayores gradientes de temperatura.

5.3.3.1 Precursosores.

Los precursores son los compuestos químicos a partir de los cuales se van a liberar los elementos para la reacción, deben ser gases a la temperatura del proceso y su función es servir como medio de transporte. Deben ser compatibles con las condiciones de pureza y deben dejar residuos que modifiquen la conductividad del semiconductor. Los compuestos deben ser estables a temperatura ambiental, y descomponerse a las temperaturas del depósito. Generalmente se utiliza un precursor para cada elemento que se quiera incorporar a la película.

Entre los compuestos más usados tenemos:

Haluros: Compuestos de Cl, Br o I. Generalmente son gaseosos a condiciones ambientales, forman ácidos, los cuales pueden, los que pueden

auxiliar en decapado “*in situ*”, aunque también pueden degradar la superficie.

Hidruros: Compuestos a base de hidrógeno. Gaseosos a condiciones ambientales. Altamente tóxicos y/o explosivos. Requieren de manejo en condiciones de seguridad.

Organometálicos: Compuestos orgánicos. Pueden ser líquidos o sólidos. Existe una gran variedad de opciones, dada la variedad de compuestos posibles. Se utilizan las fórmulas mas sencillas, teniendo cuidado especial que los residuos no sean contaminantes. Su costo es alto.

5.3.4 Procesos epitaxiales de silicio

El silicio se deposita a partir de tetracloruro de silicio, triclorosilano, diclorosilano o silano. La tabla 1 compara algunas propiedades de estos compuestos.

Los compuestos de cloro reaccionan con el hidrógeno para formar SiCl₂. El crecimiento se realiza a partir de éste último con la reacción:



ésta se realiza sobre la superficie y es catalizada por la misma, por lo que en este proceso es mas factible la. nucleación heterogénea que la homogénea.

El crecimiento se realiza cerca del equilibrio termodinámico. Si se incrementa la concentración del reactivo, hay depósito policristalino debido a efectos de la superficie. Un incremento posterior resulta en la reducción de la velocidad de crecimiento de las películas pudiendo llegar a ser negativo, o sea decapado, este fenómeno es causado por la generación de HCl en el reactor.

Tabla 1. Cuadro comparativo entre los compuestos usados para epitaxia de silicio.

Tetracloruro de silicio	(SiCl ₄)	Mas estable. Requiere mayor temperatura. (1150-1250°C)
Triclorosilano	(SiHCl ₃)	1100-1200°C
Diclorosilano	(SiH ₂ Cl ₂)	Mayor eficiencia (menos pérdida de silicio) 110-1200°C.
Silano	(SiH ₄)	Reacción irreversible. Nucleación homogénea. Muy sensible al oxígeno. <1000°C

El SiCl₄ es el compuesto mas estable y requiere mayor temperatura de crecimiento, por ello se prefiere el triclorosilano cuando se pretende evitar la difusión de las impurezas entre las películas.

El crecimiento con silano se realiza a menor temperatura, la reacción de descomposición de este compuesto es irreversible, o sea que el proceso se realiza lejos del equilibrio termodinámico. El proceso debe realizarse en reactores de paredes frías para evitar la nucleación homogénea.

5.3.5 Epitaxia de Arseniuro de Galio.

No existe un compuesto que transporte galio y arsénico simultáneamente²⁰, por lo que se empleará un precursor diferente para cada uno de ellos.

²⁰ Existen algunos compuestos orgánicos que han sido investigados para esta función, pero los resultados no han sido satisfactorios.

El arseniuro de galio se descompone con la temperatura, liberando arsénico, por lo que hay que prevenir las fugas del elemento durante el crecimiento epitaxial. Esto se hace con su precursor. Dado que su presión de vapor es alta el exceso quedará en fase gaseosa. Por otro lado el galio en exceso tiende a condensarse y formar gotas de material, mismas que pueden reaccionar después degradando la morfología superficial.

Como precursores de arsénico se utilizan casi siempre el tricloruro de arsénico (AsCl_3) y la arsina (AsH_3). Este último es extremadamente tóxico y se deben tomar precauciones para su manejo. La selección del precursor de galio es más difícil. El cloruro de galio no es estable a temperatura ambiental y debe ser generado en el interior de los mismos hornos. Esto se realiza haciendo reaccionar ácido clohídrico con galio dentro de la cámara de reacción hay dos variantes, el *proceso de los haluros* y el *proceso de los hidruros*.



Fig. 5.6 Técnica de los haluros.

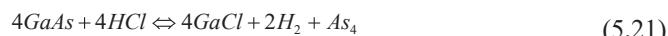
La fig. 5.6. muestra un diagrama del proceso de los haluros este se realiza en un horno de dos zonas de temperatura, en una de las cuales se coloca galio. Al introducir galio en el reactor, este se descompone con la formación de ácido clohídrico y arsénico:



Una porción del arsénico se disuelve en la superficie formando una costra de GaAs:



El ácido clorhídrico a su vez reacciona con la costra, para la formación de HCl



Las reacciones 5.20 y 5.21 son reversibles, esto es el proceso se realiza cerca del equilibrio termodinámico. Las cantidades de cada una de las especies involucradas estarán determinadas por la temperatura, presión y actividad de cada componente.

El gas se transporta hacia la zona del sustrato, el cual se encuentra a una temperatura inferior, con lo que las condiciones termodinámicas varías y existe una sobresaturación del GaCl en la fase gaseosa, en este caso la reacción procede en sentido inverso, realizándose un depósito cristalino.

En esta técnica la razón de crecimiento depende fundamentalmente de la diferencia de temperaturas de fuente de galio y de sustrato. Entre sus inconvenientes se encuentra el hecho de que la fuente requiere de estabilización para la formación de la costra de GaAs, por lo que los transitorios pueden modificar las características del proceso. Para solucionar este problema se desarrolló la técnica de los hidruros.

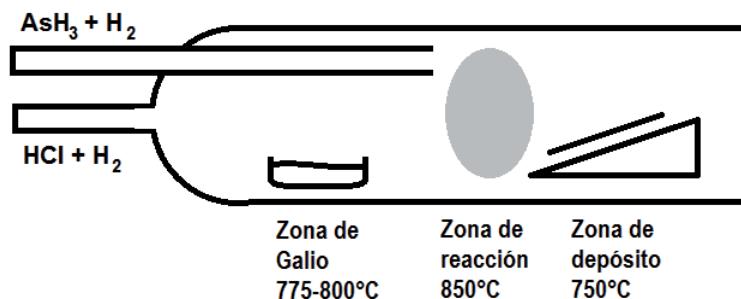


Fig.5. 7. Técnica de los hidruros.

En la figura 5.7 se muestra la técnica de los hidruros. En esta el precursor de arsénico es AsH_3 , y se introduce en un punto posterior a la fuente de galio. La arsina se decomponen por la temperatura:



Para formar el GaCl se utiliza ácido clorhídrico y de esta forma se realiza la reacción:



Tanto el GaCl como el arsénico son transportados por el gas portador a la zona del sustrato realizándose una reacción como la ec. 5.21.

El proceso de los hidruros es menos sensible a las condiciones termodinámicas, sin embargo sigue realizándose cerca del equilibrio, y no es adecuado para el crecimiento de heteroestructuras. Además, la introducción de gases para impurificar modifica las condiciones de equilibrio con alteraciones en la razón de crecimiento. Una solución al respecto sería el *proceso organometálico*

El proceso organometálico se basa en reacciones irreversibles, por lo que se reduce drásticamente la sensibilidad a los parámetros de depósito. Como precursor de arsénico se utiliza arsina. Para el galio se emplea

trimetilgalio $[(CH_3)_3Ga]$. Las impurezas también pueden ser transportadas por compuestos orgánicos. Se seleccionan aquellos que sean volátiles y cuyos residuos sean estables. Generalmente son líquidos a temperatura ambiente, por lo que se usan burbujaeadores para su manejo. La reacción del crecimiento de GaAs sería:



El hidrógeno atómico liberado por la arsina reacciona eficientemente con los radicales metil, con la formación de metano e impidiendo la incorporación de carbono en las películas.

5.4 EPITAXIA EN FASE LÍQUIDA.

De acuerdo al diagrama de fases Ga-As, si se tiene en equilibrio un líquido en contacto con un sustrato de GaAs la composición del primero será fija y dependerá de la temperatura. Si esta varía tendremos un comportamiento para restaurar el equilibrio de acuerdo al siguiente esquema:

La epitaxia en fase líquida se ha empleado muy poco en silicio, pero es popular en GaAs, debido principalmente a la calidad del material que se obtiene, particularmente adecuado para aplicaciones en optoelectrónica. Además de ser una técnica de bajo costo. Entre sus desventajas están la dificultad de obtener una buena morfología debido principalmente a la tensión superficial del líquido; lograr un buen control de espesores, sobre todo en el rango submicrométrico y no ser muy adecuada para la producción en masa.

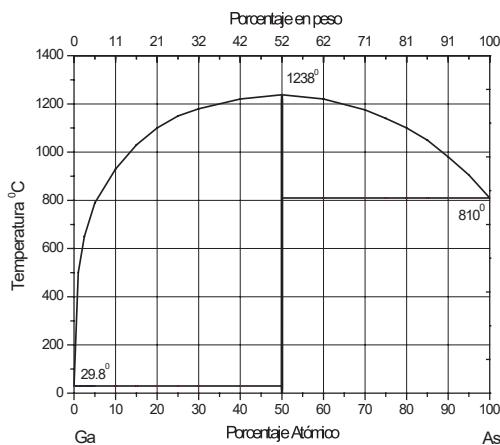


Fig. 5.8 Diagrama de fases Ga-As

En un proceso epitaxial habrá un solvente y un soluto. Los solventes serán activos cuando impurifiquen el material. Este es caso del Galio y el Aluminio para el silicio, y del germanio y el estaño para el GaAs. Por otro lado hablaremos de solventes inactivos los que no afecten la conductividad eléctrica del semiconductor, como el Sn en silicio y Ga y Pb en GaAs. La impurificación se realiza mediante la introducción de pequeñas cantidades de los elementos.

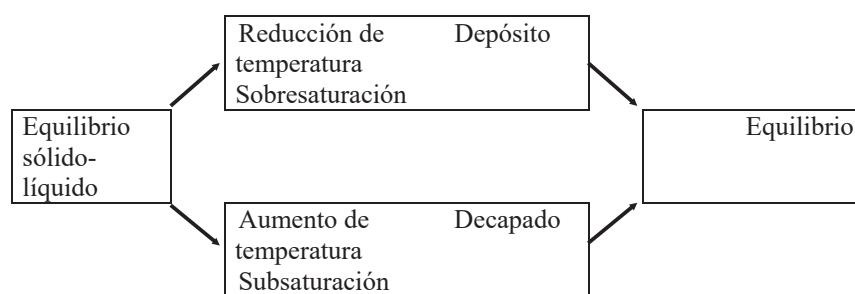


Fig. 5.9 Esquema del proceso de epitaxia en fase líquida

Como en el caso de la epitaxia en fase de vapor los procesos se realizan en atmósferas de gases puros, generalmente en hidrógeno. En este caso la función del gas solo es purgar el sistema y reducir la cantidad de contaminantes, por lo que los flujos son pequeños.

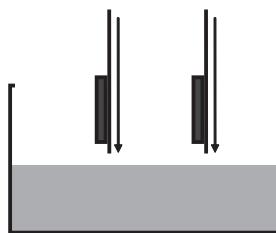


Fig. 5.9. Sistema de inmersión

Los sistemas epitaxiales pueden ser de varios tipos. El mas sencillo es el de inmersión (fig. 5.9), donde se introducen las obleas en un líquido. Para iniciar el proceso se reduce lentamente la temperatura. Para terminarlo se retira.

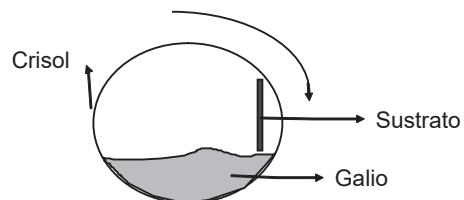


Fig. 5.10 Sistema de rotación

En el de rotación (Fig 5.10) el crisol es cilíndrico, y en su interior se encuentran galio y el sustrato. Para iniciar el crecimiento se gira el cilindro de tal forma que el sustrato se sumerja dentro de la solución saturada.

El sistema más popular es el de bote de rejilla deslizante (fig. 5.11), ya que permite la obtención de estructuras de varias capas. En este se colocan

las diferentes soluciones en depósitos. En la parte inferior se tienen dos sustratos. El que entra en primero con la solución tiene la función de equilibrar la mezcla, o sea que las concentraciones sean exactamente las de equilibrio. En el segundo es donde se depositará la estructura de interés. La realización del crecimiento se hace reduciendo lentamente la temperatura.

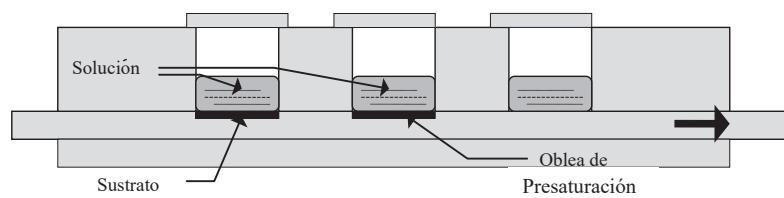


Fig. 511. Esquema de un bote de rejilla deslizante,

CAPÍTULO 6

IMPLANTACIÓN IÓNICA

6.1 INTRODUCCIÓN

En 1954 William Shockley y N. J. Madison de los laboratorios Bell sometieron una patente para la formación de dispositivos de semiconductores por bombardeo iónico. La implantación iónica evolucionó considerablemente en los años 60 y en los años 70 aparecen las primeras máquinas comerciales. Actualmente es un método muy utilizado en la industria de dispositivos semiconductores.

La implantación iónica consiste fundamentalmente en el bombardeo de una superficie con iones a muy alta velocidad, estos se impactan penetrando en el interior del material y modificando sus propiedades.

Entre las ventajas que presenta el método son: pureza, control de dosis, alta reproducibilidad, versatilidad y el hecho de realizarse normalmente a temperatura ambiental, que disminuye la contaminación asociada a fenómenos de alta temperatura y compatibilidad del proceso con un mayor número de materiales, como es la resina fotolitográfica. Además se trata de un proceso fuera del equilibrio, por lo que no depende de, ni está limitado por las condiciones termodinámicas. Entre sus características más importantes está la posibilidad de controlar de manera independiente la concentración y la profundidad del máximo de la impureza que se implanta, así como la posibilidad de introducir muy bajas concentraciones de impurezas, imposible de lograr con el proceso de predepósito o predifusión descrito en el capítulo anterior.

Su principal desventaja es el costo, ya que los equipos se cotizan en los millones de dólares. También produce, en el material que se implanta, daño estructural provocado por el bombardeo de iones, cuya eliminación requiere de procesos de recristalización a temperaturas del orden de 800-1000 °C. En algunos casos el efecto sobre algunos parámetros del material, como es el tiempo de vida de portadores minoritarios, puede ser negativo para las características del dispositivo.

6.2 SISTEMAS DE IMPLANTACIÓN

En la figura 6.1 se muestran los componentes básicos de un sistema de implantación iónica.

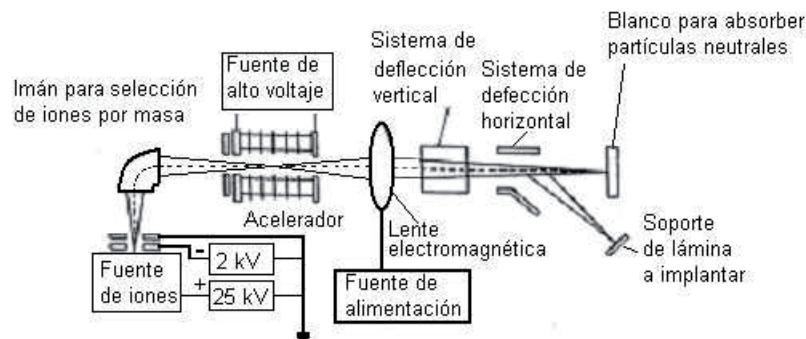


Fig. 6.1 Partes de que consta un implantador iónico.

El funcionamiento de un equipo de implantación iónica aplica en varias de sus partes la fuerza de Lorentz que ejerce un campo electromagnético sobre una partícula cargada,

$$F = q(\vec{E} + \vec{v} \times \vec{B}) \quad (6.1)$$

De acuerdo a la segunda ley de Newton, esta fuerza generará una aceleración:

$$\vec{a} = \frac{\vec{F}}{m} \quad (6.2)$$

dónde:

\vec{a} : aceleración (vector)

\vec{F} : fuerza (vector)

q : Carga del electrón (1.60219×10^{-19} C).

\vec{E} : Campo eléctrico (vector).

\vec{v} : Velocidad de la partícula cargada (vector).

\vec{B} : Campo magnético (vector).

El uso de estas dos ecuaciones permite controlar, comprender y predecir la ionización de moléculas, la aceleración de los iones, la purificación del haz y la reducción de la dispersión de la velocidad de impacto.

La energía de impacto de los iones se reporta en unidades de electrón-volt²¹

6.2.1 Fuente de iones

En primer lugar tenemos la fuente de iones, es aquí donde se generan estos a partir de un material fuente. Para ionizar una molécula es necesario cargarla, ya sea positiva o negativamente. En el caso de implantación iónica se introduce el material a ionizar en un campo electromagnético. Dado que la fuerza del campo depende del signo de la carga, habrá fuerzas opuestas sobre los electrones con carga negativa y el núcleo de carga

²¹ Un electrón-volt es la energía que adquiere una partícula con una carga igual a la del electrón acelerada por un campo eléctrico bajo una diferencia de potencial de un voltio.

positiva. Si estas son lo suficientemente fuertes, los electrones mas débilmente ligados se desprenderán del átomo, o sea que en este momento se ionizarán los átomos.

Un átomo ionizado tenderá a recuperar su electrón. Esto puede suceder a presión atmosférica casi instantáneamente debido a las colisiones, por otro lado, si hay una gran densidad de átomos, cada uno de ellos compensará parcialmente el campo eléctrico, haciéndose mas difícil obtener un número considerable de iones, la eficiencia de ionización se incrementa al reducir la presión a niveles de vacío moderado. En alto vacío la cantidad de átomos disponibles para ionizar es muy reducida.

Un filamento caliente pude proveer electrones que aumenten la ionización. La ionización también se puede incrementar mediante el calentamiento del material a ionizar, de tal manera que los electrones alcancen niveles energéticos más cercanos al nivel de vacío.

En la fuente de iones se introducen pequeñas cantidades de algún compuesto que contenga el elemento que se quiere implantar, se prefieren las fuentes en estado gaseoso, ya que son más fáciles de controlar, aunque también se pueden introducir pequeñas cantidades de algún sólido dentro de la misma cavidad de la fuente.

Los iones generados salen por un pequeño orificio de la misma fuente y son acelerados por un campo eléctrico que los conduce a la siguiente parte del implantador.

6.2.2 Purificador del haz

Las fuentes de iones difícilmente son puras, esto es a la vez de obtener iones del elemento deseado, se generan también iones de otros elementos o compuestos, es por ello que es necesaria la purificación de los iones.

Dependiendo del diseño del implantador, la purificación del haz puede darse antes o después de la aceleración de los iones.

Los iones se pueden identificar por su relación carga masa. Cada isótopo de un elemento tendrá una masa diferente. Generalmente se consideran aquellos que resultan cuando el elemento pierde un solo electrón, así, un ion con relación carga/masa igual a 1 será hidrógeno, por no existir otro que presente este valor. En algunos casos podremos tener duplicidad, tal es el caso del isótopo de masa 28 del silicio y la molécula ionizada del nitrógeno (N_2) también de masa molecular igual a 28. Dado que el nitrógeno se encuentra presente como gas residual, y muy probablemente en alguno de los componentes del implantador, generalmente se trata de implantar el isótopo de masa 30 del silicio.

Combinando la Fuerza de Lorentz (Ec. 1) con la segunda Ley de Newton (Ec. 5.2) tendremos:

$$\bar{a} = \frac{q}{m}(\bar{E} + \bar{v} \times \bar{B}) \quad (6.3)$$

En este caso vemos que la aceleración que recibe una partícula será proporcional a la relación carga/masa. Existen dos filtros que nos permiten la purificación de haz. El primero nos permitirá seleccionar iones con una velocidad determinada, esto se logra cuando exponemos al haz a un campo electrostático opuesto a la dirección $v \times B$, en este caso la aceleración de los iones será cero si se cumple la condición:

$$|E| = |vB| \quad (6.4)$$

En este caso podemos ajustar los valores de los campos para que sólo los iones con una velocidad determinada no sufran ninguna aceleración lateral. Esto es una desviación. Al final del filtro se pondrá una rendija para que solo pasen los iones que no hayan sido desviados.

Una vez que se tiene un haz de iones a una velocidad constante es sometido a un campo magnético perpendicular a su dirección de movimiento. Este tendrá una aceleración tangencial igual a:

$$a = \frac{v^2}{r} = \frac{q}{m} v B \quad (6.5)$$

o sea que los iones tomarán una órbita circular con radio:

$$r = \frac{mv}{qB} \quad (6.6)$$

Conociendo la velocidad de los iones y ajustando el campo magnético se puede determinar el radio de curvatura para aquellos iones que tengan una relación q/m específica, ésta curvatura se ajustará a una geometría específica del sistema y se acoplará a una rendija de salida donde sólo podrán salir aquellos iones que tengan una trayectoria predeterminada de acuerdo a los valores de velocidad, campo magnético y sobre todo su relación carga/masa.

La separación de iones mediante su relación carga/masa permite la purificación de los iones a grados mayores que con cualquier otra técnica conocida. Si se toman las precauciones adecuadas es poco probable la contaminación con iones con la misma relación q/m .

6.2.3 Aceleración

Los iones son partículas cargadas, en nuestro caso de interés, negativamente. Si son sujetos a un campo electrostático presentarán una aceleración de acuerdo a:

$$\bar{F} = q\bar{E} = ma \quad (6.7)$$

Generalmente la energía adquirida después de una aceleración de los iones se da en unidades de electronvoltios o sea la diferencia de potencial aplicada en el acelerador.

El potencial de aceleración aplicado en un sistema de implantación iónica es del orden de cientos de miles de voltios, por lo que será necesario aplicar técnicas de alto voltaje al sistema. El acelerador es un tubo (cañón) de algunos metros de longitud que separa la fuente de iones del blanco, esto es las obleas de semiconductor. Asimismo se colocan regularmente electrodos con divisores de tensión, a fin de tener una distribución uniforme del campo eléctrico a lo largo del cañón.

El cañón está al alto vacío. Este es importante a fin de prevenir las colisiones de los iones con moléculas de gas y su desviación o neutralización. Generalmente se utilizan bombas de vacío turbomoleculares, ya que éstas además de alcanzar la presión requerida en un tiempo adecuado son muy limpias.

Teniendo los iones la misma carga eléctrica, estos tenderán a dispersarse, por lo que en el cañón también se colocan lentes electrostáticos para hacer converger el haz de nuevo.

Debido al alto voltaje es necesario aislar eléctricamente la fuente de iones. Por ello su alimentación puede ser por medio de baterías mientras esté funcionando el acelerador. Asimismo su comunicación con la unidad de control será por medio de fibra óptica, a fin de evitar conductores que pudieran provocar un arco eléctrico.

6.2.4 Blanco

Los iones deberán impactarse sobre las obleas de material semiconductor. En este caso tendremos diferentes diseños, dependiendo de la aplicación. En general se tendrá un sistema que permita la carga y

descarga de las obleas por medio de esclusas, para no interrumpir el vacío del acelerador. Existen implantadores de doble cámara, que permiten cargar una, mientras en la otra se está implantando. También existen sistemas de carrusel, donde se introducen varias obleas y se ponen a girar durante el proceso de implantación.

Las obleas quedarán protegidas por un cilindro metálico, que recogerá cualquier carga que pudiera ser emitida de la oblea, como serían electrones o emisión de iones secundarios, éste funcionará como una jaula de Faraday, y será importante para el control y medición de la dosis.

6.2.5 Deflexión del haz

A la salida del acelerador el haz de iones será puntual, pero el propósito será implantar las obleas uniformemente. Esto se hace mediante sistemas de deflexión cuya función es realizar un barrido sobre la superficie de la oblea. La deflexión puede ser electrostática en ambos ejes, aunque también existe la modalidad mecánico-electrostática, donde las obleas se mueven en una dirección, mientras que el haz se mueve perpendicularmente a la misma. Este es el caso particular de los sistemas de carrusel.

6.2.6 Control de dosis

Una de las ventajas de la implantación iónica es el control de la cantidad de impurezas implantadas. Dado que los iones están cargados eléctricamente estos se traducirán en una corriente eléctrica que puede ser medida con gran precisión. Esto permite conocer la carga implantada y con ello el número de iones que se impactan sobre la superficie del blanco. Generalmente se reporta la dosis en número de iones por unidad de área, el perfil de impurezas se calcula en base a la distribución de acuerdo a los modelos que se exponen mas adelante.

6.3 CARACTERÍSTICAS DE PENETRACIÓN DE LOS IONES EN EL BLANCO

6.3.1 Dispersión

Los iones se impactan como proyectiles a muy alta velocidad penetrando en el blanco, éstos van dañando el cristal, desplazando a todo lo que encuentran a su paso antes de llegar al reposo. Si la energía con que llegan no es demasiado alta, en cada colisión los iones se desvían de su trayectoria de forma aleatoria como se muestra en la Fig. 6.2. En cada choque pierden parte de su energía y pueden desplazar a los átomos de la estructura cristalina de sus posiciones destruyendo la red, por lo que el material cristal puede pasar a una fase amorfa. Se llama proyección media del recorrido a la proyección sobre el eje perpendicular al plano de incidencia de la distancia de la superficie al punto en que se detuvo el ión dentro del material.

Existen principalmente dos mecanismos mediante los cuales los iones pierden su energía: frenado nuclear y frenado electrónico. El primero se aproxima mediante colisiones entre esferas, manteniendo las leyes de conservación de energía y momento mientras que el segundo es una interacción con los electrones libres y las corazas electrónicas de los átomos de la red y se aproxima más a pérdida por fricción. Dado el gran número de iones implantados, la distribución final se puede tratar por métodos estadísticos con bastante precisión, [1-4].

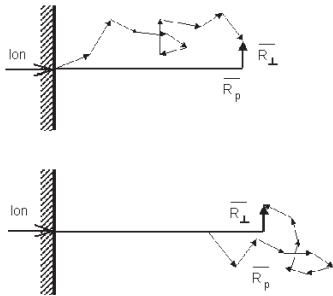


Fig. 6.2 Movimiento de los iones dentro del material para baja y alta energía de implantación.

6.3.2 Frenado electrónico

La interacción entre los iones y los electrones es principalmente por efecto Coulomb. Dado que la masa de un ion es mucho mayor que la de un electrón, las colisiones directas no afectarán la trayectoria del ion, más bien el frenado es efecto del gas electrónico en el semiconductor y se puede aproximar como una viscosidad cuya fuerza será proporcional a la velocidad del ion [1]:

$$F_D \propto v \propto \sqrt{E} \quad (6.8)$$

Entonces se definirá la potencia de frenado electrónico como:

$$S_e = \frac{dE}{dx} \Big|_e = k_e \sqrt{E} \quad (6.9)$$

El modelo de viscosidad deja de ser válido a energías muy altas.

6.3.3 Frenado nuclear

La interacción de los iones con los átomos del semiconductor se puede aproximar mediante colisiones elásticas entre los iones y los núcleos del

blanco. Debido a su velocidad los iones pueden atravesar la nube de electrones de un átomo, por lo que la interacción se puede aproximar entre los núcleos cargados. Las colisiones se suceden de manera aleatoria modificándose la trayectoria del ion en cada choque. La pérdida de energía de un ion por colisiones depende de varios factores como son: las masas del ion y del átomo de la red, la velocidad, y el ángulo de choque. Esta se incrementa de manera lineal con la energía del ion, para llegar a un máximo y luego decrecer, sin embargo se aproxima como un valor constante.

El frenado nuclear produce un desplazamiento de los átomos del blanco, lo que da lugar a que este pierda su cristalinidad. Los iones pueden penetrar debajo de los bordes de la máscara de implantación debido a la dispersión causada por los cambios de dirección. Dadas las condiciones de este proceso es difícil predecir la posición final de un ion en particular, sin embargo, considerando un gran número de iones es posible aplicar una solución estadística,

Dado que el frenado electrónico se incrementa linealmente con la velocidad y el frenado nuclear se puede aproximar por una constante, hay una energía crítica en la que estos se igualan. A valores más bajos el proceso dominante será el nuclear y el electrónico dominará a altas velocidades de los iones. Esta energía crítica también depende de la naturaleza del ion y es más baja para iones ligeros. Generalmente se puede decir que iones ligeros y altas energías favorecen el frenado electrónico, mientras que iones pesados y bajas energías de implantación favorecen el frenado nuclear.

La distribución de iones se puede aproximar mediante una forma gaussiana [1]:

$$N(x) = N_o e^{-\left[\frac{(x-R_p)^2}{2\Delta R_p^2} \right]} \quad (6.10)$$

Donde:

R_p es el Rango proyectado.

ΔR_p es la dispersión del rango proyectado; y

N_o es la dosis o cantidad de iones implantados.

Los valores de R_p y ΔR_p para diferentes impurezas y blancos se muestran en las tablas A-5 a A-8 en el Anexo.

El Rango Proyectado es la posición en que se encuentra un máximo de iones medidos a partir de la superficie. El valor del Rango proyectado se determina experimentalmente y es diferente para cada ion. Generalmente varía en forma proporcional a la energía de implantación.

La distribución real difiere ligeramente de la gaussiana. Generalmente esta no es completamente simétrica, por otro lado es posible encontrar que la campana es más plana que lo ideal. Estas desviaciones han sido estudiadas, sin embargo tienen un efecto pequeño sobre el funcionamiento final de las estructuras semiconductoras.

La dosis total implantada D_o se calcula según:

$$D_o = \int_{-\infty}^{\infty} N(x) dx = \int_{-\infty}^{\infty} \exp - \left[\frac{(x-R_p)^2}{2\Delta R_p^2} \right] dx = N_o \Delta R_p \sqrt{2\pi} \quad (6.11)$$

obteniéndose la relación entre la dosis total y los parámetros de la distribución de impurezas.

$$D(x_l) = \frac{D_o}{\sqrt{\pi}} \int_{-\infty}^{x_l} \exp - \left[\frac{(x-R_p)^2}{2\Delta R_p^2} \right] dx = D_o \operatorname{erf} \left[\frac{-(x_l-R_p)^2}{2\Delta R_p^2} \right] \quad (6.12)$$

Si $x_1 < R_p$,

$$D(x_1) = \frac{D_o}{2} \left[1 - \operatorname{erf} \left[\frac{(R_p - x_1)^2}{\sqrt{2}\Delta R_p} \right] \right] \quad (6.13)$$

Si $x_1 > R_p$

$$D(x_1) = \frac{D_o}{2} \left[1 + \operatorname{erf} \left[\frac{(R_p - x_1)^2}{\sqrt{2}\Delta R_p} \right] \right] \quad (6.14)$$

El perfil de impurezas introducidas por implantación iónica presenta las siguientes características de interés:

Para cualquier energía de implantación, R_p es diferente de cero, por lo tanto, el máximo de impurezas se encuentra siempre a determinada distancia de la superficie, dentro del substrato que se implanta. Para ubicar un máximo de impurezas en la superficie es necesario implantar a través de una capa de otro material, como por ejemplo el SiO₂.

Como R_p es siempre mayor que $2\Delta R_p$, luego de una implantación sobre un substrato, la mayor parte de la dosis implantada se encuentra dentro de este, independientemente de la localización de R_p , o sea, para cualquier energía. Por el contrario, si hay que implantar a través de una capa u oxidar el substrato implantado, parte de la dosis se perderá.

La función gaussiana es una representación de primera aproximación a un perfil de impurezas implantadas. Para impurezas donadoras, como el fósforo y el arsénico, se obtiene una mejor descripción del perfil de implantación utilizando una doble media gaussiana, o sea dos gausseanas con diferente R_p y ΔR_p hacia cada lado del máximo de concentración.

Para impurezas acceptoras, la distribución más utilizada es la de Pearson IV,

Realizando varias implantaciones a diferentes dosis y energías se pueden obtener diferentes perfiles de concentración, cuando la estructura en formación así lo requiere.

6.4 IMPLANTACIÓN A TRAVÉS DE MULTICAPAS

Durante la fabricación de dispositivos semiconductores se realizan procesos de implantación en el silicio a través de una o varias capas que recubren la superficie del silicio. Cada una de estas capas, según el material que sea, tendrán una R_p y ΔR_p diferente. Los simuladores tecnológicos permiten determinar el comportamiento de las impurezas que se implantan en cada capa y dentro del substrato, que permiten determinar con precisión tanto la posición del máximo de implantación según el tipo de impureza, las capas que atraviesa y la energía de implantación. De forma sencilla se puede realizar también un estimado de esta posición, utilizando las siguientes propiedades del proceso de implantación.

Para dos materiales homogéneos independientes en los cuales se implanta a la misma energía una dosis igual, el espesor del material en que se almacena igual cantidad de impurezas debe cumplir que:

$$\frac{R_{p2} - x_2}{\Delta R_{p2}} = \frac{R_{p1} - x_1}{\Delta R_{p1}} \quad (6.15)$$

Esta expresión (6.15) se obtiene de igualar las áreas bajo las curvas gaussianas correspondientes al perfil en cada uno de los materiales; y permite, conocido el espesor de un material dado en que se encuentra una cantidad de impurezas dada, calcular el espesor de otro material donde se

almacenaría igual cantidad de impurezas que en el primero. Esto se obtiene despejando x_2 en la expresión (6.15).

$$x_2 = R_{p2} - \frac{\Delta R_{p2}}{\Delta R_{p1}} \cdot R_{p1} + \frac{\Delta R_{p2}}{\Delta R_{p1}} \cdot x_1 \quad (6.16)$$

En una doble capa, para una misma energía de implantación, la cantidad de impurezas retenidas por la primera capa de espesor X_o será igual a la que retendrá una capa homogénea de igual material que esta primera capa y de igual espesor X_o . La cantidad de impurezas almacenadas en este espesor X_o para una capa homogénea puede expresarse como:

$$D(X_o) = \frac{D_o}{2} \left[1 - \operatorname{erf} \left[\frac{(R_{p1} - X_o)^2}{\sqrt{2} \Delta R_{p1}} \right] \right] \quad (6.17)$$

Para una doble capa, cuando el máximo de impurezas cae dentro de la segunda capa, la cantidad de impurezas almacenada en la primera de espesor luego de una implantación con una energía dada, puede expresarse como:

$$D(X_o) = \frac{D_o}{2} \left[1 - \operatorname{erf} \left[\frac{(R'_p - X_o)^2}{\sqrt{2} \Delta R_{p2}} \right] \right] \quad (6.18)$$

donde R'_p es la distancia desde la superficie de la estructura formada por la doble capa, hasta la coordenada donde se localiza el máximo de concentración.

También se podría considerar que según se definió la distancia R'_p , una capa con espesor $x' = R'_p - X_o$ y rango proyectado R_{p2} , almacenaría la

misma cantidad de impurezas que una de X_o , con rango R_{p1} , según (5.15). Por lo tanto puede escribirse que:

$$D(x_1) = \frac{D_o}{2} \left[1 - \operatorname{erf} \left(\frac{(R_{p2} - X_{eq})^2}{\sqrt{2} \Delta R_{p1}} \right) \right] \quad (6.19)$$

donde X_{eq} sería el espesor de una capa de material 2, que almacenaría la misma dosis de impurezas que una capa de espesor X_o del material 1. El valor de X_{eq} se determina mediante cualquiera de las siguientes expresiones:

$$\frac{R_p - X_o}{\Delta R_{p2}} = \frac{R_{p1} - X_o}{\Delta R_{p1}} = \frac{R_{p2} - X_{eq}}{\Delta R_{p2}} \quad (6.20)$$

Resumiendo, para analizar la distribución de impurezas que resulta de implantar una impureza en un substrato a través de una o varias capas de diferente material que cubren el substrato se puede realizar lo siguiente:

Convertir el espesor de cada capa a un espesor equivalente de silicio X_{eq} a través de (6.15), donde x_1 es el espesor de la capa cuyo espesor se desea convertir a espesor equivalente; R_{p1} es el rango proyectado correspondiente dicha capa y R_{p2} el rango proyectado del substrato.

Determinar el valor de R_{p1} y R_{p2} para la impureza que se implanta y los materiales de la capa en cuestión y del substrato.

La distancia desde la superficie del substrato hasta el máximo se determina como:

$$x_2 = R_{p2} - X_{eq} \quad (6.21)$$

En la Fig. 6.3 se indican las distancias mencionadas.

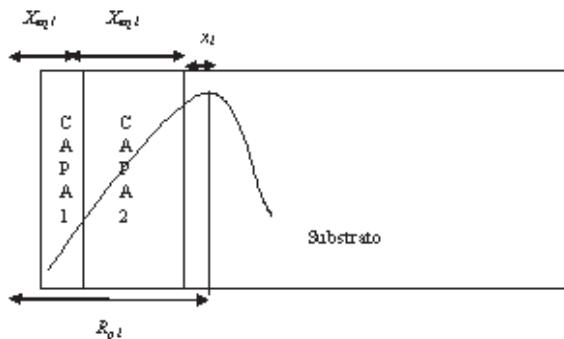


Fig. 6.3 Determinación de la posición del máximo del perfil de implantación en el substrato para una implantación a través de dos capas, utilizando la técnica de convertir el espesor real de cada capa en espesor equivalente de substrato.

6.5 CANALES DE PROPAGACIÓN

Cuando la implantación se realiza en materiales cristalinos, es posible que los iones se propaguen por un arreglo de intersticios de la red. Este sería un canal en el cual no hay colisiones y por lo tanto el frenado de iones es muy lento. Esto es particularmente notable cuando el haz de iones incide paralelamente a una orientación cristalina. A este fenómeno se le denomina en inglés “channeling”, o sea, propagación por canales.

La propagación a través de canales en el cristal es fuertemente dependiente de la orientación del cristal con respecto al haz de iones y es muy difícil de predecir y resulta en una cola adicional al perfil gaussiano, misma que se puede extender a distancias mucho mayores que el rango proyectado. Es por ello que se trata de evitar. La solución más sencilla es inclinar la oblea para evitar la propagación de los iones. Típicamente se usa un ángulo de aproximadamente 7°.

A pequeñas dosis siempre se encuentra un cierto grado de propagación por canales, debido a que después de una colisión el ion puede alterar su dirección y alinearse con un canal. Por otro lado, cuando se implanta a muy altas dosis (más de 10^{14} iones/cm²), el efecto de amorfización que sufre el cristal por los impactos de los iones destruye los canales por los que se pudieran propagar los iones, reduciendo el efecto de canalización. Implantando a temperatura del orden de los 300 °C o realizando tratamientos térmicos después de la implantación a temperaturas superiores a los 600 °C es suficiente para recristalizar las regiones dañadas [5,7].

Si la implantación se realiza a través de una capa amorfa, como por ejemplo el SiO₂, también se reduce el efecto de canalización en el substrato.

6.6 DAÑO DE IMPLANTACIÓN

El impacto de los iones daña fuertemente la estructura cristalina del sustrato, siendo mayor el daño que causan los iones pesados. Algunos átomos de la superficie pueden ser desprendidos del material, pero el principal efecto es el desplazamiento de los iones dentro de la misma red con la pérdida casi total de la estructura cristalina en las primeras capas del semiconductor. Dado que esta es precisamente la región activa de los dispositivos es necesario restablecer la estructura cristalina, por otro lado los iones implantados tampoco ocupan sitios substitucionales de la red, por las impurezas no son eléctricamente activas.

La manera más sencilla de restablecer la estructura cristalina y de activar las impurezas implantadas es mediante un tratamiento térmico, dado el grado de amorfización, la recristalización se presenta a unos pocos de

cientos de grados centígrados, la temperatura de activación eléctrica es diferente para cada impureza [8-10]. Dado que el proceso también afecta el tiempo de vida de portadores minoritarios, cuando se requiera restablecer este se requerirán mayores valores de tiempo y temperatura del tratamiento térmico. Más recientemente se están utilizando sistemas de RTP (Rapid Thermal Processing) como alternativa [11].

La activación de impurezas en arseniuro de galio requiere de prevenir la evaporación de arsénico, por lo que generalmente se recubre el semiconductor con una película protectora, como pueden ser óxido o nitruro de silicio. También es posible realizar el recocido en una atmósfera con sobrepresión de arsénico.

6.7 BIBLIOGRAFÍA

- [1] J. Lindhard, N. Scharff y H.E. Schiott, “Range Concepts and Hedavy ion ranges”, Mat. Fys. Med. Dan. Vidensk. Selk, 33 (1963) 14.
- [2] J.F. Gibson y W. S. Jonson, “Projected Range Statistics Semiconductor and Related Materials”, Standford University Bookstore, USA, 1969.
- [3] W.S. Johnson y J. K. Gibson, “Statistical Range distribution of ions in single and multiple elements substrates” Appl. Phys. Letters 9 (1966) 321.
- [4] I. D:E: Davies “Range of implanted boron, phosphorous and arsenic in silicon”, J. App. Phys 47 (16) (1972) 134.
- [5] D.S. Gemmell, “Channeling and related effects in the motion of charged particles through crystals, Rev. Mod. Phys. 46 (1974) 129.

[6] J.F. Ziegler, "Channeling of ions near the silicon <100> axis", Pll. Phys Lett. 46 (1985) 358.

[7] N.L. Turner, Effects of planar channeling using modern ion implant equipment", Solid State Tech., 2 (1985) 163.

[8] B.L. Crowder and R.S. Title, "The distribution of damage produced by ion implantation of silicon at room temperature" Radiation effects, 6 (1981) 163.

[9] Formation of P⁺-n junctions by B implantation in silicon substrate with amorphous layers" Jap. J. Appl. Phys 24 (1985) 568.

[10] G. Carter y W.A. Grant, "Ion implantation in semiconductors" ediciones John Wiley and Sons, NY, 1976.

[11] T. E. Seidel, D. J. Linscher, C. S. Pai, R. V. Knoell, D. M. Mather y D.C. Jonson, "A review of rapid thermal annealing (RTA) of B, BF₂, and As Implanted into silicon", Nucl Instrum. Methods B 7/8 (1985) 251.

PARTE II

CAPÍTULO 1

INTRODUCCIÓN A LA TECNOLOGÍA PLANAR

Durante la segunda mitad de la década del 50, comienza a utilizarse el silicio como semiconductor para fabricar dispositivos semiconductores, así como el proceso de difusión de boro y fósforo para formar zonas p y n en el substrato de silicio; sin embargo, no es hasta 1959 que Atalla y Tannenbaum descubren las propiedades pasivadoras del óxido de silicio (SiO_2) que comienza una nueva época (en lo que a tecnología se refiere) con el surgimiento de la Tecnología Planar.

1.1 CARACTERÍSTICAS DE LA TECNOLOGÍA PLANAR

La Tecnología Planar combina, en la fabricación de dispositivos semiconductores, las propiedades del SiO_2 obtenido por oxidación térmica del silicio, las posibilidades de la fotolitografía y la introducción de impurezas localizadas. Esta tecnología permitió un notable aumento en la productividad y fiabilidad de los dispositivos; al mismo tiempo hizo posible la realización de circuitos integrados al permitir aislar entre sí transistores construidos en una misma oblea de silicio.

Entre las propiedades del óxido de silicio obtenido por oxidación térmica del silicio se pueden mencionar las siguientes:

Pasivación. Es la propiedad del óxido de silicio que permite estabilizar las características de la superficie del silicio, evitando su interacción con el medio ambiente. Esta propiedad permite que durante su procesamiento, una vez oxidada, la oblea pueda estar en contacto con el aire sin que se alteren las características eléctricas finales de los dispositivos. Además, aumenta la

repetibilidad de las características de los dispositivos realizados en diferentes corridas.

Enmascaramiento. El coeficiente de difusión de elementos como B, P, Sb y As, que se utilizan para formar las zonas *p* o *n* de las uniones, es un orden menor en el SiO₂ que en el Si. Esto permite proteger con óxido las zonas de la oblea donde no se desea introducir las impurezas por difusión, ya que en el período de tiempo en que la impureza penetra, por ejemplo, 1 μm en el silicio, en el óxido de silicio penetrará menos de 0.1 μm. De esta forma, en el ejemplo señalado, se puede evitar que en las zonas protegidas con óxido de un espesor mayor de 0.1 μm, la impureza llegue al silicio (Fig. 1.1).

Introducción localizada de impurezas en el silicio. En las zonas donde se dejó al descubierto el silicio por el proceso de fotolitografía pueden introducirse impurezas de forma localizada, que no penetran en las zonas cubiertas de óxido por la propiedad de enmascaramiento.

Ataque selectivo. Fue posible encontrar soluciones químicas que atacan el óxido de silicio pero que no reaccionan con el silicio. En la actualidad, el ataque selectivo se logra también mediante las técnicas de ataque en seco (en plasma).

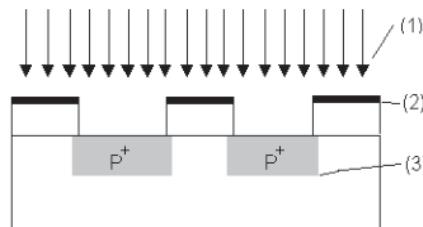


Fig. 1.1 Difusión selectiva en el silicio: 1. impurezas; 2. Zona hasta dónde penetró la impureza en el SiO₂; 3. Zona hasta dónde penetró la impureza en el Si.

Características del SiO₂ y de la interfaz SiO₂-Si. El SiO₂ presenta relativa baja permeabilidad a los iones móviles y la interfaz SiO₂-Si presenta baja densidad de cargas. Estas propiedades dependen del proceso tecnológico utilizado en la obtención del óxido y su perfeccionamiento permitió pasar de la tecnología bipolar a la tecnología MOS.

Fotolitografía. Se llama fotolitografía al proceso de formación de un relieve sobre la superficie de un substrato, que permite conformar los elementos del circuito que se desea fabricar. Para esto, el relieve se imprime inicialmente sobre un material fotosensible que se deposita previamente sobre la oblea de silicio cubierta de SiO₂ y/o otras capas, y se traslada posteriormente a la oblea mediante los ataques selectivos de las capas que cubren el substrato (Fig. 1.2).

La impresión de la fotorresina se realiza exponiéndola a la luz ultravioleta, generalmente a través de placas de vidrio donde previamente, por un proceso fotográfico, se han delineado los motivos cuya geometría se desea imprimir en la fotorresina, y que reciben el nombre de máscaras o máscaras.

Para confeccionar las máscaras hay que contar primero con un dibujo que determine la posición relativa, la geometría y las dimensiones de los elementos del circuito en escala natural. A este dibujo se le llama topología o dibujo geométrico del circuito y generalmente se realiza con un sistema de diseño interactivo con ayuda de la computadora (CAD).

En la variante mas generalizada de confeccionar máscaras para circuitos integrados, el sistema de diseño interactivo prepara como salida un soporte magnético que controla un equipo que recibe el nombre de generador de patrones. Este equipo va trazando la topología o dibujo geométrico del circuito sobre un material fotosensible que recubre una placa de vidrio denominada *retículo* mediante un haz luminoso (generalmente con un aumento de 10x) o mediante un haz de electrones.

En un paso posterior, también por un proceso fotográfico, se reduce esta imagen dibujada en el retículo a su tamaño natural y se imprime muchas veces, en el sentido de los ejes x e y , sobre una placa de vidrio. Esta placa de vidrio recibe el nombre de máscara original y se muestra esquemáticamente en la figura 1.3. Cada cuadrado oscuro representa la topología del circuito integrado a realizar.

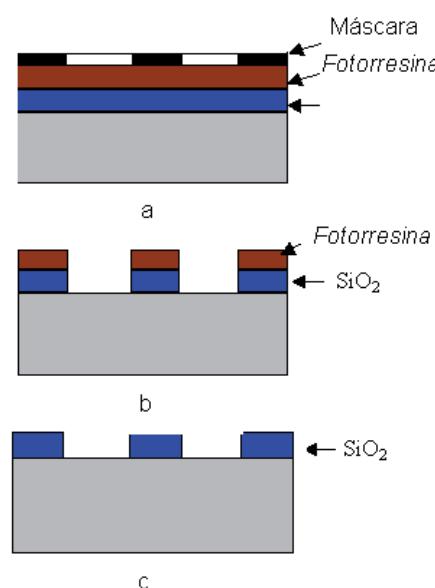


Fig. 1.2 Proceso de fotolitografía: a) impresión de la fotorresina exponiéndola a la luz ultravioleta a través de placas de vidrio donde previamente, por un proceso fotográfico se definieron las regiones oscuras y claras; b) Ataque selectivo de la fotorresina y del SiO₂; c) Eliminación de la fotorresina

Según la capa con que esté recubierta la placa de vidrio, las máscaras pueden ser de emulsión o metálicas, opacas o semitransparentes. En todos los casos en que se usa el método de alineación por contacto o por aproximación, las máscaras se dañan con el uso y pueden utilizarse un

número limitado de veces. El proceso de fabricación de máscaras es un proceso costoso, por lo que la necesidad de sustituir los juegos de máscaras, afecta significativamente el costo final del procesamiento de un dispositivo o circuito.

Alineación y exposición. La traslación del motivo de la máscara a alguna de las capas previamente depositadas sobre la oblea de silicio, se realiza utilizando la propiedad que tienen las fotorresinas de polimerizarse bajo la acción de la luz.

Por ejemplo, la oblea de silicio se oxida previamente para formar una capa de SiO₂ sobre la que se deposita la fotorresina. Las zonas transparentes de la máscara hacen que esta resina se polimerice. Si la fotorresina utilizada es positiva, el revelado hace que se disuelvan las partes expuestas a la luz, dejando al descubierto en esas zonas el SiO₂. El SiO₂ en estas regiones en las que se eliminó la fotorresina, puede eliminarse por métodos de ataque químico húmedo o seco (por plasma), hasta dejar al descubierto la superficie del silicio.

Por otro lado, como para la fabricación de dispositivos es necesario utilizar varias máscaras, cada una de ellas debe alinearse con respecto a la anterior, para lo cual se utilizan marcas o motivos definidos en la mascarilla en uso, los cuales se alinean con marcas o motivos definidos en la oblea por la máscara anterior. Este proceso se realiza en equipos llamados alineadoras. Un proceso de fabricación de circuitos integrados utiliza frecuentemente más de 10 máscaras.

Estas alineadoras pueden ser de contacto, de aproximación o de proyección, según el método que utilicen para trasladar el dibujo de la mascarilla a la superficie de la oblea. En el caso de este último tipo de alineadora, la máscara sufre menos deterioro durante el proceso de alineación, pero los equipos son más complejos y costosos que los de contacto o aproximación.

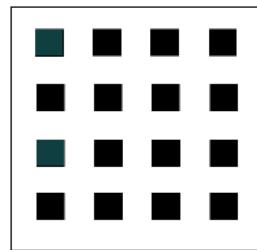


Fig. 1.3 Mascarilla indicando la repetición de los motivos de cada chip a lo largo de los ejes X e Y. a) Imagen reducida del retículo a su dimensión final

La tecnología planar utiliza los siguientes procesos tecnológicos:

- Crecimiento de cristales semiconductores y preparación de obleas;
- Procesos de limpieza de obleas;
- Crecimiento epitaxial;
- Introducción de impurezas;
- Oxidación térmica del silicio;
- Depósito de capas dieléctricas, semiconductoras y metálicas;
- Procesos de fotolitografía;
- Ataques selectivos de capas;

CAPÍTULO 2

CARACTERÍSTICAS DEL SiO₂ POR OXIDACIÓN TÉRMICA DEL SILICIO

El primer método utilizado para la formación de capas de óxido de silicio para la Tecnología Planar, fue la oxidación térmica del silicio. Las propiedades del SiO₂ dependen de su método de obtención, siendo el óxido térmico el que mejores características presenta en su conjunto.

Entre ellas se pueden citar las siguientes:

1. Conductividad eléctrica baja.
2. Campo eléctrico mayor de 4×10^6 V/cm, lo que permite aplicar altos voltajes a través de él para espesores de capas relativamente pequeños.
3. Reacciona con soluciones químicas o mezclas gaseosas que no atacan al silicio. Una solución muy frecuentemente usada para estos ataques selectivos es el llamado buffer de ácido fluorídrico, que contiene ácido fluorídrico y fluoruro de amonio en una mezcla de 7 partes a 1, respectivamente. Esta solución disuelve al SiO₂ a una velocidad del orden de los 70 nm/min a la temperatura de 25 °C,
4. Buen acoplamiento en la interfaz con el silicio, desde el punto de vista de dilatación térmica.
5. Baja densidad de carga en la superficie producto de los enlaces no saturados, centros cargados, estados superficiales o impurezas adsorvidas.
6. Buena adherencia a las fotorresinas utilizadas en los procesos de fotolitografía.
7. Baja permeabilidad a iones móviles.

2.1 CINÉTICA DE CRECIMIENTO

Las capas de SiO₂ obtenidas por oxidación térmica se obtienen en hornos de alta temperatura (hasta 1200 °C), en presencia de oxígeno puro, vapor de agua y sus mezclas en diferentes porcentajes. Las reacciones químicas que tienen lugar son:



Aunque todavía se trabaja en esta temática, y existen aspectos divergentes en las conclusiones de diferentes investigadores, el tratamiento clásico para el estudio de la cinética de crecimiento de los óxidos térmicos fue planteada por primera vez por B. E. Deal y A. S. Grove [1,2], siendo la mayoría de sus planteamientos aplicables en la actualidad.

El modelo propuesto, y que explica en gran parte los procesos que ocurren, establece que la oxidación del silicio se produce por el movimiento de la especie oxidante hacia la frontera Si–SiO₂, a través de la capa de óxido que va creciendo. Para que la especie oxidante alcance la superficie del silicio, donde tiene lugar la reacción, debe ocurrir lo siguiente:

1. Que la especie oxidante llegue desde el volumen del gas hasta la superficie de la oblea, (flujo F₁ figura 2.1).
2. Que se difunda la especie oxidante a través del SiO₂ que se está formando, hasta alcanzar la frontera SiO₂ -Si (flujo F₂, figura 2.1).

3. La reacción de la especie oxidante con la superficie del silicio consumirá cierta cantidad de la especie activa que se representa por el flujo F_3 en la figura 2.1.

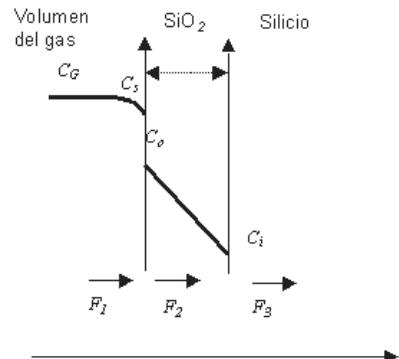


Fig. 2.1 Flujos de especie oxidante durante el proceso de oxidación del silicio en un horno de alta temperatura: F_1 - flujo del volumen del gas a la superficie del SiO_2 que va

Para encontrar a qué es igual la corriente del gas hacia la superficie, se analizará lo que ocurre en esa zona en un modelo muy simplificado que llamaremos modelo de la capa límitrofe (MCL).

2.1.1 **Modelo de la capa límitrofe o fronteriza**

Dentro del tubo del horno donde ocurre el proceso de oxidación, el gas penetra por un extremo y se mueve hacia el otro. Si toda la masa del gas se mueve como un todo, se dice que tiene lugar una corriente o flujo laminar. El gas se mueve según se muestra en la figura 2.2.

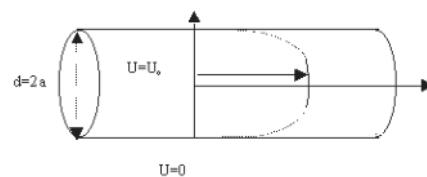


Fig. 2.2 Distribución de la velocidad del gas U en el interior de un tubo de diámetro d .

En las paredes del tubo la velocidad del gas es cero, pero si el diámetro del tubo d es suficientemente grande, habrá una región cercana al eje central del tubo donde la velocidad del gas puede considerarse uniforme e igual a:

$$U = U_o \cdot \left(1 - \frac{(2r)^2}{d^2}\right) \quad (2.3)$$

Si en esa zona se introduce una oblea de silicio, la velocidad del gas en la superficie de ésta será también igual a cero, ya que la oblea representa un obstáculo al paso del gas. Lo que ocurre en esa zona puede representarse según se muestra en la figura 2.3, donde la región dentro de la frontera marcada por $\delta(x)$ limita la zona a partir de la superficie de la oblea, en la que la velocidad del gas varía entre 0 y $0.99 U_o$.

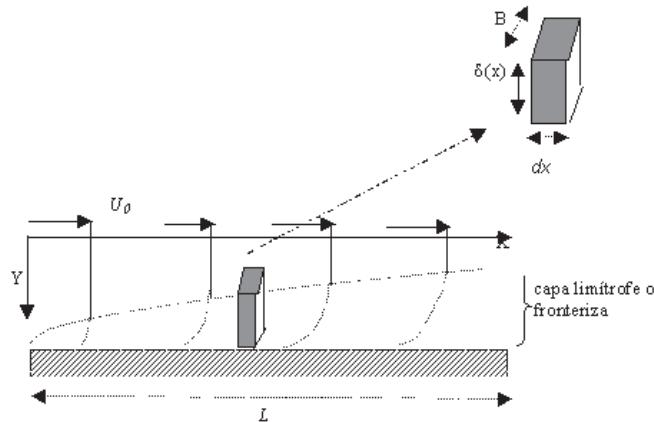


Fig. 2.3 Formación de una capa fronteriza debida al paso del gas bordeando un obstáculo fijo de longitud L .

En esta misma zona de altura $\delta(x)$, la concentración de la especie activa en el gas también varía a lo largo del eje “y”, por lo que tendrá lugar un

proceso de difusión de la misma, desde el volumen del gas hacia la superficie de la oblea, que en la Fig. 2.3 está colocada a lo largo del eje “x”.

La relación entre el flujo y el gradiente de concentración se expresa a través de la primera ley de Fick [2]

$$F_1 = -D \nabla C, \quad (2.4)$$

la cual se simplifica para el caso unidimensional a lo largo del eje “y”, quedando de la forma:

$$F_1 = -D \frac{\partial C}{\partial y}. \quad (2.5)$$

Si $\delta(x)$ es pequeña, el gradiente de concentración puede expresarse como la variación de concentración a través de la capa con un espesor promedio $\bar{\delta}$, o sea:

$$-\frac{\partial C}{\partial y} \approx \frac{C_G - C_s}{\bar{\delta}} \quad (2.6)$$

y $D=D_G$, donde D_G es la difusividad de la especie activa en el gas.

Sustituyendo (2.6) en (2.5) queda la expresión

$$F_1 = -h_G (C_G - C_s) \quad (2.7)$$

donde $h_G = \frac{D_G}{\bar{\delta}}$ es el coeficiente de transferencia de masa de la fase gaseosa.

Para determinar el valor de $\bar{\delta}$ se considera que la desaceleración del fluido es producida por la fricción de este sobre la pared del obstáculo.

Si se analiza un volumen unitario de gas formado por la zona rayada en la figura 2.3, el área de fricción será Bdx y la fuerza de fricción por unidad de área será

$$F'_{fric} = \eta \frac{\partial U}{\partial y}, \quad (2.8)$$

donde η es el coeficiente de viscosidad del gas (Fig. 2.4 a) y la fuerza que actúa sobre el elemento dx será $F_{fric} = F'_{fric} \cdot dx$.

La aceleración del elemento de gas será

$$\frac{\partial u}{\partial t} \approx \frac{\partial u}{\partial x} \cdot \frac{\partial x}{\partial t} = \frac{\partial u}{\partial x} \cdot U \quad (2.9)$$

y la masa de este volumen unitario será: $\rho \delta(x) dx \cdot B$, donde ρ es la densidad del gas (Fig. 2.4 b).

Aplicando la segunda ley de Newton se tiene que:

$$\eta \frac{\partial u}{\partial y} \cdot Bdx = \rho \cdot \delta(x)dx \cdot B \cdot \frac{du}{dx} U. \quad (2.10)$$

Esta expresión se simplifica y queda como

$$\eta \frac{\partial u}{\partial y} \cdot dx = \rho \cdot \delta(x)dx \cdot \frac{du}{dx} U. \quad (2.11)$$

En adelante, el modelo toma en cuenta para el cálculo de δ las consideraciones siguientes:

- 1 La velocidad del gas en la frontera de la zona limítrofe U es igual a 0,99 U_0 .

$$2 \quad \frac{du}{dx} \approx \frac{\Delta U}{\Delta x} \approx \frac{U_o}{x}, \quad (2.12)$$

o sea, la velocidad del gas varía linealmente con la coordenada x en la zona limítrofe, siendo las condiciones de frontera las siguientes:

$$\begin{aligned} U &= 0 \text{ en } y=0, \\ U &= U_o \text{ en } y \rightarrow \infty, \\ U &= U_o, \text{ en } x=0 \text{ para cualquier } y. \end{aligned}$$

$$3) \quad \frac{\partial U}{\partial y} \approx \frac{\Delta U}{\Delta y} \approx \frac{U_o}{\delta(x)}. \quad (2.13)$$

Con este valor promedio y definiendo una magnitud adimensional que se conoce como el número de Reynolds

$$\text{Re} = \frac{UL\rho}{\eta}, \quad (2.14)$$

que relaciona los efectos iniciales con los efectos de viscosidad, se puede calcular el parámetro h_G según la expresión

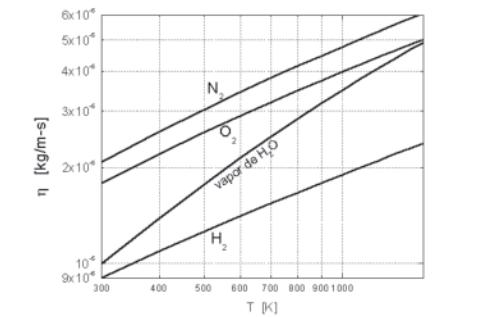
$$h_G = \frac{D_G}{L} \sqrt{\text{Re}}. \quad (2.15)$$

Puede demostrarse que cálculos más exactos permiten llegar a una expresión similar

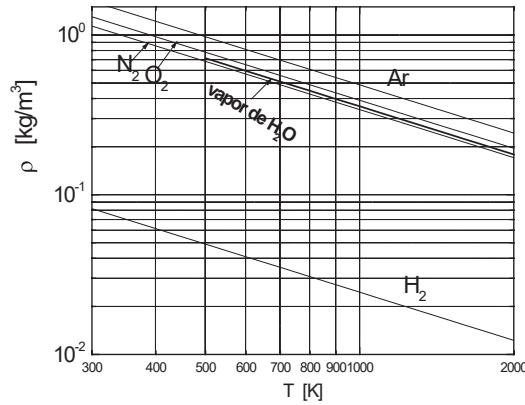
$$h_G = \frac{2D_G}{3L} \sqrt{\text{Re}} \cdot \sqrt{\frac{\eta}{\rho D_G}}, \quad (2.16)$$

donde $\frac{\eta}{\rho D_G}$ es el número de Schmidt, que oscila entre 0,6 y 0,8 para los diferentes gases.

Las expresiones (2.15) y (2.16) se diferencian en un factor de aproximadamente tres.



(a)



(b)

Fig. 2.4 a) Viscosidad del H_2 , O_2 , vapor de agua y N_2 en función de la temperatura; b) densidad del H_2 , O_2 , Ar, vapor de agua y N_2 en función de la temperatura.

2.1.2 Cálculo del flujo de especie activa durante la oxidación

Para calcular este flujo F_2 en la figura 2.1 se considera que el espesor del óxido X_0 es lo suficientemente pequeño como para que el gradiente de concentración a través de él pueda expresarse según:

$$\frac{dC}{dx} = \frac{C_o - C_i}{X_0}, \quad (2.17)$$

donde C_o y C_i son las concentraciones de la especie activa en el óxido junto a la superficie externa y en la interfaz SiO₂-Si respectivamente.

De esta forma, según la ley de Fick:

$$F_2 = D \frac{C_o - C_i}{X_0}, \quad (2.18)$$

donde D es el coeficiente de difusión de la especie activa en el óxido.

Además, si se supone que:

- durante el proceso de difusión de la especie activa se cumple la ley de Henry, por lo que la concentración de la especie dentro del sólido es proporcional a la presión parcial de la especie activa en el gas que la rodea, o sea:

$$C_o = H \cdot P_s, \quad (2.19)$$

- que se tiene un gas ideal, para el cual

$$C_S = P_S / kT. \quad (2.20)$$

Sustituyendo (2.19) en (2.20) queda:

$$C_S = C_0 / HkT. \quad (2.21)$$

Si de forma análoga se define una concentración en el sólido C^* correspondiente a una presión parcial P_G , que es la presión parcial de la especie activa en el volumen del gas, según la ley de Henry y la ley de los gases ideales

$$C^* = HP_G \quad (2.22)$$

y

$$C_G = P_G / kT. \quad (2.23)$$

Sustituyendo (2.21), (2.22) y (2.23) en (2.7) se obtiene el flujo de la especie activa, al difundirse ésta desde el volumen del gas hasta la superficie de la oblea en función de las concentraciones en el sólido

$$F_l = \frac{h_G}{HkT} (C^* - C_0). \quad (2.24)$$

En la frontera $\text{SiO}_2\text{-Si}$ tiene lugar la reacción de la especie activa con el silicio, dando lugar al óxido de silicio. Esta reacción, que como ya se señaló puede ser según (2.1) y (2.2) para los casos de oxidación en oxígeno y vapor de agua, respectivamente, va a considerarse que es de primer grado. De esta manera el flujo F_3 (Fig. 2.1) puede determinarse según

$$F_3 = k_s C_i, \quad (2.25)$$

donde k_s , es la constante de reacción.

En estado estacionario los tres flujos deben ser iguales entre si, por lo que utilizando esta condición, y a partir de (2.18), (2.24) y (2.25), se obtienen las concentraciones C_i y C_0 .

Igualando (2.18) y (2.25) se obtiene

$$C_0 = \left(1 + \frac{k_s X_0}{D}\right) \cdot C_i. \quad (2.26)$$

Igualando (2.24) y (2.25) y haciendo $h=h_G/HkT$, se obtiene

$$C_i = \frac{C^*}{1 + \frac{k_s}{h} + \frac{k_s X_0}{D}}. \quad (2.27)$$

Por último, sustituyendo (2.27) en (2.26)

$$C_0 = \left(1 + \frac{K_s X_0}{D}\right) \cdot \frac{C^*}{\left[1 + \frac{k_s}{h} + \frac{k_s X_0}{D}\right]}. \quad (2.28)$$

En las expresiones anteriores se puede analizar dos casos límites:

Proceso de oxidación controlado por difusión. En este caso el coeficiente de difusión es pequeño, predominando el proceso de transporte de masa sobre el proceso de difusión. La concentración de la especie activa en la frontera SiO₂-Si tiende a cero y en la frontera gas-SiO₂, dentro del sólido, tiende a su valor máximo.

$$D \rightarrow 0$$

$$C_i \rightarrow 0$$

$$C_0 \rightarrow C^*$$

Proceso de oxidación controlado por reacción. En este caso el coeficiente de difusión es relativamente grande por lo cual se cumple que

$$\frac{k_s X_0}{D} \ll 1 \quad (2.29)$$

y las expresiones (2.26) y (2.27) se transforman en

$$C_i = C_0 = \frac{C^*}{1 + \frac{k_s}{h}} \quad (2.30)$$

La concentración de la especie activa en las fronteras gas-SiO₂ y SiO₂-Si tenderá al mismo valor, el que dependerá de la presión parcial de la especie en el gas junto a la superficie del sólido. En este caso el transporte de la especie por difusión ocurre más rápido que su consumo en la reacción.

2.1.3 Velocidad de crecimiento

Conocidos los procesos que ocurren en las fronteras gas-SiO₂ y del SiO₂-Si durante la oxidación térmica, se puede determinar la velocidad de crecimiento de la capa de óxido.

El número de moléculas de SiO₂ en un cm³ de óxido es de 2.2x10²². Si la oxidación de silicio ocurre en presencia de oxígeno, tendrá lugar la reacción (2.1) y se incorporará una molécula de oxígeno a cada molécula de SiO₂. Si la oxidación ocurre en presencia de agua, tendrá lugar la reacción (2.2) y se incorporarán dos moléculas de H₂O a cada una de SiO₂.

Por lo tanto se necesitará incorporar 2.2×10^{22} moléculas por cm^3 de oxígeno o 4.4×10^{22} moléculas de H_2O por cm^3 durante la oxidación, según corresponda.

El flujo de oxidante hacia la superficie del silicio está relacionado con la velocidad de crecimiento de la capa dX_o/dt a través del número de moléculas a incorporar por unidad de volumen N_1 . Si se sustituye (2.27) en (2.25) resulta

$$F_1 = N_1 \cdot \frac{dX_0}{dt} = \frac{k_S C^*}{1 + \frac{k_S}{h} + \frac{k_S X_0}{D}}. \quad (2.31)$$

Resolviendo (2.31) con respecto a X_0 se obtiene

$$X_0 = \frac{A}{2} \cdot \left[\sqrt{1 + \frac{(t + \tau)}{A^2/4B}} - 1 \right], \quad (2.32)$$

donde:

$$A = 2 \cdot D \cdot \left(\frac{1}{k_s} + \frac{1}{h} \right), \quad (2.33)$$

$$B = 2D \frac{C^*}{N_1}, \quad (2.34)$$

$$\tau = \frac{X_1^2 + AX_1}{B}, \quad (2.35)$$

y $X_1 = X_0(0)$, que representa el espesor de una capa de óxido que se supone existe previamente sobre el silicio.

El análisis de (2.32) revela de nuevo la existencia de dos casos límites:

$$1. \text{ Si } t \gg A^2/4B, \quad (2.36)$$

o sea, para tiempos largos de oxidación,

$$X_o = \sqrt{B(t + \tau)} \quad (2.37)$$

y el crecimiento de óxido sigue un régimen parabólico con respecto al tiempo.

$$2. \text{ Si } (t + \tau) \ll A^2/4B, \quad (2.38)$$

es decir, para tiempos cortos de oxidación,

$$X_0 = \frac{B}{A}(t + \tau). \quad (2.39)$$

Estos dos regímenes de crecimiento se observan en las curvas de espesor de óxido contra tiempo de crecimiento, para diferentes agentes oxidantes y temperaturas.

En la figura 2.5 y 2.6 se observa en un gráfico logarítmico, la presencia de dos pendientes que precisamente definen las regiones en que predomina cada uno de estos regímenes para oxidación húmeda y seca respectivamente. A mayor temperatura, el régimen parabólico predomina para tiempos menores de oxidación. En la Fig. 2.6 también se observa la dependencia de la velocidad de crecimiento con la orientación de la oblea, para la oxidación en O₂. Aunque no se muestra en la Fig. 2.5 lo mismo ocurre para la oxidación en vapor de agua. Esta dependencia viene dada por el coeficiente lineal y por ello la diferencia entre las curvas es mayor a menores tiempos o menores temperaturas.

Una vez obtenidas experimentalmente, a partir de ellas se determina el coeficiente parabólico B y el coeficiente lineal B/A.

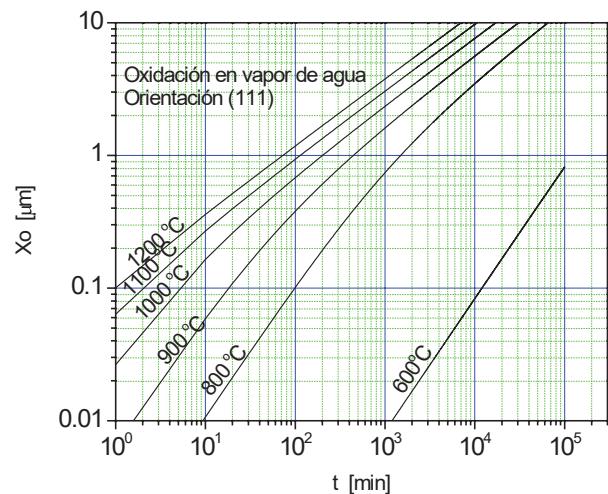


Fig. 2. 5 Espesor del óxido en función del tiempo de oxidación en vapor de agua a presión atmosférica para diferentes temperaturas de oxidación.

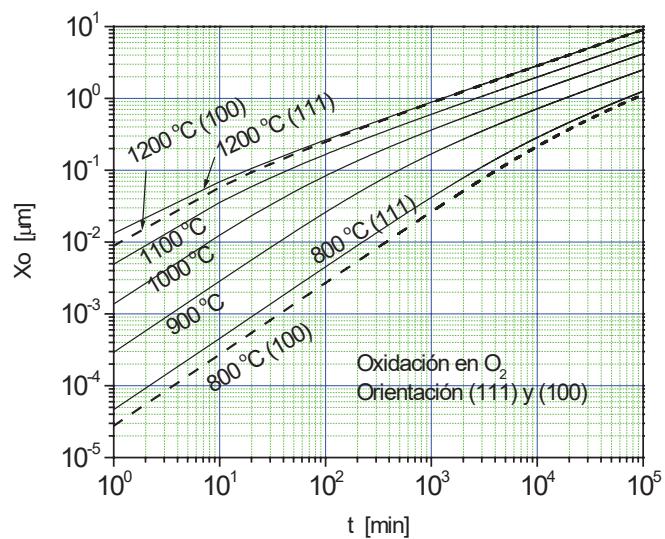


Fig. 2.6 Espesor del óxido en función del tiempo de oxígeno seco a presión atmosférica para diferentes temperaturas de oxidación y orientación de substrato.

También debido a la dependencia de C^* , y por tanto del coeficiente parabólico con la presión de la especie activa según las expresiones (2.21) y (2.33), a mayor presión de oxidación, la velocidad de oxidación será mayor. Por ello, aumentando la presión del agente oxidante se puede obtener el mismo espesor de óxido en igual tiempo pero a menor temperatura, o crecer óxidos más gruesos en tiempos iguales o menores.

2.2 DATOS EXPERIMENTALES QUE CONFIRMAN EL MODELO DE OXIDACIÓN DE GROVE

El análisis de los coeficientes B y B/A demuestra que según (2.34), el coeficiente B depende de los parámetros siguientes:

$$B(T) = (2C^*/N_A) D(T), \quad (2.40)$$

por lo que la dependencia térmica de B debiera estar dada por la dependencia térmica del coeficiente de difusión de la especie activa. Experimentos realizados [2] demuestran que la dependencia térmica de B es exponencial (Fig. 2.7) y que la energía de activación es de 1.24 eV para el oxígeno y 0.7 eV para el agua, valores bastante cercanos a los reportados para energías de activación de la difusividad respectiva del oxígeno y del vapor de agua a través de sílice fundida.

Una vez determinado B y conocido el valor de D, el cálculo de C^* [2] da el valor de $5 \times 10^6 \text{ cm}^3$ para la oxidación en oxígeno seco y $3 \times 10^{19} \text{ cm}^{-3}$ para la oxidación en vapor de agua, los cuales concuerdan satisfactoriamente con los datos de solubilidad del oxígeno y agua en sílice fundida a una atmósfera de presión.

Según (2.31) y (2.32) el coeficiente lineal B/A depende de los parámetros k_s , h y C^* de la forma siguiente:

$$B/A = (2C^*/N_I) (1/k_s + 1/h). \quad (2.41)$$

Experimentos realizados indican que la dependencia térmica observada en B/A es también exponencial, con una energía de activación cercana a los 2 eV para ambos agentes oxidantes (Fig. 2.8). Esta energía es del orden de la energía de ruptura de los enlaces en el silicio, lo que indica que el coeficiente lineal está relacionado con la reacción que ocurre en la frontera $\text{SiO}_2\text{-Si}$.

De acuerdo con esta hipótesis, este coeficiente debía tener dependencia con la orientación del substrato de silicio, cuestión que se observa también experimentalmente. A un plano más denso, por ejemplo el (111), le corresponde un coeficiente B/A mayor y también mayor velocidad de crecimiento que a uno menos denso, por ejemplo el (100). En la región parabólica esta dependencia de la velocidad de crecimiento con la orientación no se observa. Experimentalmente la relación del valor de B/A para la orientación (111), con el valor para la orientación (100) es como promedio de 2.68.

El estimado del coeficiente h por el modelo de capa limítrofe indica que para el proceso de oxidación, la constante de reacción k_s es menor que el coeficiente de transporte de masa h . El hecho de que la variación en el flujo de oxidante afecta poco al coeficiente lineal ratifica también esta conclusión.

Un aspecto que no queda claro en el modelo planteado, es el hecho de que para el proceso de oxidación en oxígeno seco, las expresiones propuestas concuerdan con los datos experimentales solo si se supone, previo al primer proceso de oxidación, que ya existe un óxido inicial. Para

explicar esta situación, el autor del modelo, [2], propone que el crecimiento inicial del óxido ocurre a una mayor velocidad que la que se desprende de las expresiones, debido a que en ellas no se ha considerado la posibilidad de difusión en presencia de campo eléctrico.

Según su hipótesis, cuando comienza el proceso de oxidación, la molécula de oxígeno se adsorbe en la interfaz y puede disociarse dando lugar a un ión negativo y a una carga positiva. Ambos comienzan a moverse, pero si uno tiene mayor movilidad tiende a arrastrar al otro, sin que por ello aparezca corriente neta. Este fenómeno fue analizado por Cabrera y Mott [3], quienes en su teoría predijeron que esto se debía observar fundamentalmente para espesores de óxido inferiores a un valor crítico que puede calcularse para cada caso.

La determinación del valor de espesor de óxido inicial, necesario para que el modelo de Deal y Grove sea válido, puede calcularse extrapolando la curva X_0 vs. t [2] hasta el eje de tiempo igual a cero. Para el caso de oxidación en oxígeno seco el valor que se obtiene es de aproximadamente 20 nm, mientras que para oxidación en vapor de agua es de solo 0.5 nm y se considera despreciable.

Posteriormente este problema ha sido objeto de innumerables trabajos, sobre todo a medida que los espesores de óxido fueron siendo reducidos. Entre las hipótesis y modelos de comportamiento para explicar este fenómeno de crecimiento muy rápido al inicio de la oxidación, se encuentra la hipótesis del efecto de aceleración debido a cargas en el óxido e interfaz que establecen un campo eléctrico entre 15 y 20 nm a partir de la interfaz, [4,5].

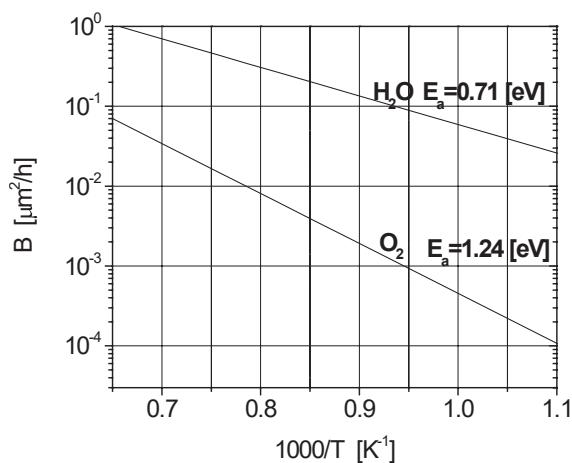


Fig. 2.7 Dependencia del coeficiente parabólico B con la temperatura de oxidación.

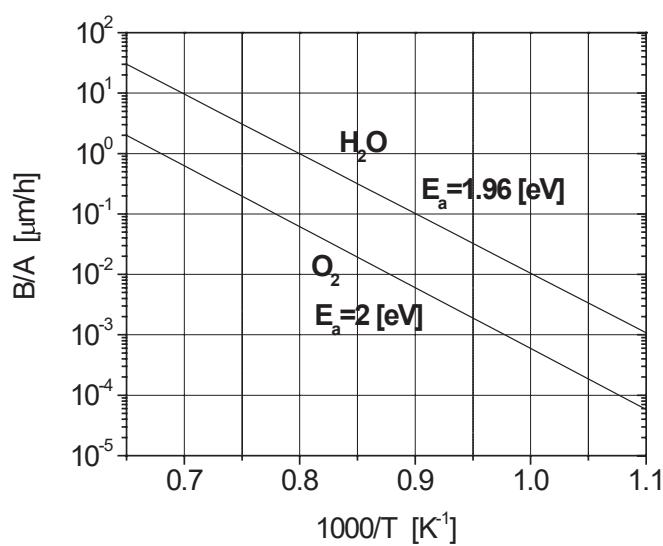


Fig. 2.8 Dependencia del coeficiente lineal B/A con la temperatura de oxidación.

Hoy en día el estudio del comportamiento del crecimiento de una capa de óxido por oxidación térmica del silicio se realiza utilizando simuladores tecnológicos que contienen diferentes modelos para describir los procesos tecnológicos que se simulan, aunque el modelo de Deal y Grove sigue siendo el modelo básico al que se le han introducido varias modificaciones para que tenga mejor concordancia con el experimento. Por ejemplo, para las oxidaciones en vapor de agua el coeficiente lineal y el parabólico se describen en dos rangos de temperatura, uno por encima de 950 °C y otro por debajo. En cada región, la energía de activación y el valor del coeficiente varían para representar el efecto de las tensiones durante el crecimiento, el cual resulta más importante en el régimen de temperaturas inferiores a los 950 °C. A temperaturas mayores el óxido puede comportarse con cierto grado de viscosidad que ayuda a disminuir los esfuerzos que debieran aparecer durante la oxidación, pues el volumen de silicio que se consume es inferior al volumen del óxido que crece, (se consume 0.45 de silicio por cada unidad de SiO_2 que se crece).

La tendencia de que las temperaturas de procesamiento de los dispositivos sean lo más bajas posible incluye también a los procesos de oxidación térmica a presión atmosférica. En ese caso, el régimen de crecimiento del óxido térmico es generalmente lineal.

Importancia especial reviste la adecuación de este modelo para que describa el comportamiento de óxidos muy finos, cuyo espesor se hace comparable o incluso menor a los 20 nm requeridos para el ajuste del modelo al comportamiento real, en el caso de la oxidación seca. Por ejemplo en el modelo del simulador tecnológico ATHENA, basado en el simulador SSUPREM desarrollado en la universidad de Stanford, para simular el crecimiento de óxidos menores de 50 nm en ambiente seco se incluye un término semi-empírico con 3 constantes de ajuste que deben ser calibradas para el proceso que se desea modelar [6].

La expresión para la velocidad de crecimiento se expresa como:

$$\frac{dX_o}{dt} = \frac{B}{A + 2X_o} + R_{th}, \quad (2.42)$$

donde R_{th} se expresa como:

$$R_{th} = R_{th0} \cdot e^{\frac{-E_{Aa}}{kT}} \cdot e^{-\frac{X_o}{L}} \cdot P^{a_p}. \quad (2.43)$$

R_{th0} es el valor intrínseco del parámetro R_{th} del orden de $5.87 \times 10^6 \mu\text{m}/\text{min}$; E_{Aa} es su energía de activación, del orden de 2.32 eV; L es una dimensión característica de ajuste del orden de 0.0078 μm y a_p es otro parámetro de ajuste que depende de la presión.

Este simulador requiere para el cálculo la presencia de un óxido inicial que generalmente existe debido al óxido nativo en la oblea.

2.3 ESTRUCTURA DEL ÓXIDO DE SILICIO Y CARACTERÍSTICAS DE LA FRONTERA $\text{SiO}_2\text{-Si}$

En este epígrafe se estudiará la estructura del óxido de silicio, la formación de defectos y las características de la interfaz $\text{SiO}_2\text{-Si}$.

2.3.1 Estructura del óxido de silicio

El óxido de silicio térmico es una estructura amorfa en la que se observa una disposición atómica semejante a la del cristal de cuarzo. La estructura de este último fue estudiada por primera vez por Sachariasen,

según el cual cada átomo de silicio está rodeado por cuatro átomos de oxígeno, formando un tetraedro; cada átomo de oxígeno, a su vez, está enlazado con dos átomos de silicio, jugando el papel de puente entre ellos [7].

Para describir la estructura (Fig. 2.9) se utiliza la notación $\text{Si}^{4+}(\text{O}^{2-}/2)$ que indica que cada catión de Si^{4+} está rodeado por 4 iones O^{2-} y cada uno de ellos está compartido por dos iones de silicio. El radio del ión O^{2-} es de 0.138 nm y el del Si^{4+} es de 0.039 nm. Se dice que en la estructura de SiO_2 se observa solo orden cercano, porque cada tetraedro está unido por sus vértices con otros, pero la orientación relativa entre ellos varía. Además se conoce que la estructura no es compacta y ocupa solo el 43% de su volumen total. El diámetro de las zonas no ocupadas es del mismo orden que el diámetro del ión oxígeno.

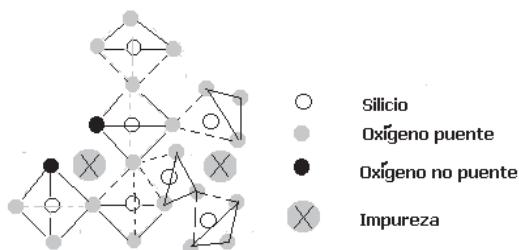


Fig. 2.9 Estructura de orden cercano del SiO_2 , indicando que los átomos de oxígeno se disponen alrededor del átomo de silicio en las esquinas de un tetraedro. Cada tetraedro puede tomar una orientación arbitraria con respecto a los otros.

La posibilidad teórica de que en la frontera entre dos cristales reales no existan estados superficiales fue analizada en un trabajo de F. Forstmann [8], donde se demuestra que la presencia de estos estados localizados es posible, solo si se cumplen las condiciones para la existencia de estados del

tipo de Shockley en uno de los dos materiales. En caso contrario, no deben observarse estos estados en la interfaz entre los dos cristales.

Según los parámetros de la red del silicio y del SiO₂, la frontera de SiO₂-Si corresponde precisamente a aquella donde no deben existir estados superficiales localizados.

En la práctica la situación real es bastante más compleja, ya que el silicio no es un cristal ideal, durante el crecimiento de la capa de SiO₂ surgen defectos y además la estructura de la capa crecida es incluso amorfa. Por esto, en una frontera real de SiO₂-Si existen estados superficiales; pero su densidad puede ser hasta cinco órdenes menor que la observada en la superficie limpia del silicio, además de ser repetible y no depender del medio ambiente luego de crecida la capa. También se observa una fuerte dependencia de la densidad de los estados con el tratamiento térmico que sufre la estructura.

2.3.2 Características de la interfaz SiO₂-Si: estados superficiales, carga fija y carga móvil, su origen y naturaleza física

El estudio de la estructura SiO₂-Si ha permitido determinar experimentalmente la existencia de tres tipos de cargas:

1. Cargas en estados localizados en la interfaz Si-SiO₂ o dentro del dieléctrico, pero cercanas a la interfaz.

Sus niveles energéticos se encuentran dentro de la banda prohibida del semiconductor y pueden intercambiar carga con la banda de conducción y la de valencia al variar el potencial externo

aplicado. A estas cargas se les denomina cargas en estados superficiales Q_{ss} .

2. Centros cargados positivamente, localizados en la interfaz, o en el dieléctrico cercano a ésta.

Sus niveles energéticos están cercanos a la banda de valencia o de conducción y por ello su población de electrones prácticamente no varía al variar el voltaje externo aplicado a la estructura. La densidad de carga originada por estos niveles se denomina carga fija, Q_{cf} .

3. Cargas positivas móviles Q_p que pueden desplazarse en el interior del dieléctrico.

La naturaleza física de estas cargas en el óxido de silicio y en la interfaz SiO₂-Si puede deberse a:

- a) Defectos estructurales;
- b) Átomos de impurezas.

La formación y comportamiento de las diferentes cargas en esta estructura, debidas a defectos estructurales fue explicada por Kooi [9], (Fig.2.10).

En el semiconductor, el átomo de silicio debe estar rodeado de 4 átomos de silicio. Si por algún motivo este átomo está unido solo a 3 átomos de silicio, uno de sus electrones estará no pareado y tenderá a alejarse del átomo (quedando cargado positivamente), o a atraer otro electrón para cargar negativamente al átomo. Este tipo de centro de defecto fue llamado Si⁽³⁾ y puede manifestar con igual probabilidad una acción tanto donora como aceptora.





De igual forma pueden existir en la frontera entre el dieléctrico y el semiconductor, o dentro del dieléctrico, defectos formados por átomos de silicio unidos a uno, a dos o a tres átomos de oxígeno; los que fueron denominados $\text{Si}^{(2)}$, $\text{Si}^{(1)}$, $\text{Si}^{(0)}$, respectivamente. Cada uno de estos defectos puede comportarse de forma análoga al $\text{Si}^{(3)}$, tomando o donando un electrón, aunque la presencia en ellos del oxígeno, hace más electronegativo al centro, predominando la acción donora y por lo tanto, el centro con carga positiva.

Los centros del tipo $\text{Si}^{(3)}$, pueden describir la carga en centros localizados que varían su población de electrones con el voltaje aplicado $Q_{ss}(V)$ y que pueden tener su origen en defectos creados durante el crecimiento del óxido.

Los centros del tipo $\text{Si}^{(2)}$ y $\text{Si}^{(1)}$, pueden describir la carga positiva fija que prácticamente no varía con el voltaje externo y que está localizada cercana a la interfaz Q_{fc} . Este tipo de defecto ha sido asociado a la presencia del oxígeno que se difundió a través del óxido y que llega a la frontera $\text{SiO}_2\text{-Si}$, donde queda atrapado sin llegar a reaccionar con el silicio [10].

Los centros del tipo $\text{Si}^{(0)}$ pueden describir un tipo de carga positiva móvil, capaz de desplazarse dentro del dieléctrico bajo al acción de la temperatura y polarización. Éstas son las llamadas vacancias de oxígeno [11], que pueden producirse cuando se deposita aluminio a temperaturas superiores a los 300 °C. En este caso, el aluminio al depositarse sobre el SiO_2 reacciona con éste y lo reduce. El oxígeno perdido por el óxido da lugar a un defecto tipo $\text{Si}^{(0)}$ con valencia +2, capaz de desplazarse bajo la acción de una polarización a temperaturas superiores a los 200 °C, dando

lugar a inestabilidad en algunas características eléctricas de los dispositivos. Esta reacción se acelera con la temperatura, siendo este tipo de carga móvil inherente a la estructura aluminio-silicio.

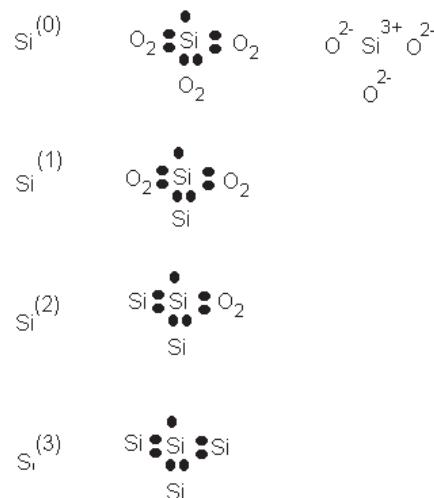


Fig. 2.10 Posibles centros de defecto en la estructura de SiO_2 - Si según Kooi [9].

Este fenómeno resultó muy importante cuando se hacían transistores MOS con compuerta de aluminio ya que disminuir las impurezas no evitaba el desplazamiento del voltaje de umbral. Hoy en día los transistores con compuerta de polisilicio ya no presentan este problema.

La carga positiva móvil dentro del dieléctrico puede estar también originada por iones de sodio (Na^+) que penetran en el óxido en alguna de las etapas del procesamiento [12].

2.3.3 Influencia de las cargas en la interfaz y en el dieléctrico sobre los parámetros de los dispositivos MOS

Para detectar la presencia de cargas en la interfaz dieléctrico-semiconductor, se utilizan estructuras metal-óxido-semiconductor (MOS) para las cuales se define un parámetro que recibe el nombre de voltaje de banda plana (V_{FB})

$$V_{FB} = \Phi_{GS} - \frac{Q_{ss} + Q_{cf} + Q_i}{C_i}, \quad (2.46)$$

donde:

Φ_{GS} - diferencia de potencial de contacto entre la compuerta y el semiconductor;

Q_{ss} - densidad de carga debida a estados superficiales localizados en la interfaz;

Q_{cf} - densidad de carga debida a carga fija localizada en la interfaz o en el dieléctrico cercano a ella;

Q_i - densidad de carga debida a carga móvil que se encuentra en el interior del dieléctrico;

C_i - capacitancia por unidad de área de la estructura debida a la capa de óxido. Este parámetro expresa el voltaje externo que hay que aplicar al metal para que las bandas del semiconductor homogéneo bajo el óxido, estén planas, o sea, que el potencial superficial en el semiconductor sea cero.

La presencia de $Q_{cf} + Q_i$ tiende a desplazar el valor del voltaje de banda plana hacia valores más negativos.

La presencia de Q_i tiende a su vez a provocar una inestabilidad en el voltaje de banda plana, ya que al desplazarse a través del óxido, la carga

equivalente que se manifiesta aplicada en la frontera SiO₂-Si varía. Esta inestabilidad se observa al polarizar el dispositivo cuando éste se encuentra a temperaturas del orden de los 150 °C.

El proceso que ocurre se explica de forma muy sencilla si se supone que se tiene un plano de carga Q localizado dentro del dieléctrico a la distancia x de la superficie de la estructura (Fig. 2.11). El aporte de esta carga a V_{FB} será:

$$V_{FB} = \frac{x}{k_i \epsilon_0} Q, \quad (2.47)$$

ya que para lograr que el potencial en la superficie se haga igual a cero, hace falta aplicar al metal un voltaje igual a V_{FB}.

El efecto de una redistribución arbitraria de carga en el dieléctrico, por ejemplo, se puede calcular suponiendo que cada plano de carga localizado en la coordenada x aportará al V_{FB} un incremento de:

$$\Delta V_{FB} = \left[\frac{x}{x_i} \cdot \rho(x) \cdot x \right] \cdot \frac{1}{C_i} \quad (2.48)$$

donde p(x) es la densidad de carga por unidad de volumen (Fig. 2.12).

La superposición del efecto de cada p(x) dará:

$$V_{FB} = -\frac{1}{C_i} \int_0^{x_i} \frac{x}{x_i} \cdot \rho(x) dx \quad (2.49)$$

Resumiendo:

1. De la expresión (2.49), si $x=0$; $V_{FB} = 0$, o sea, el efecto de carga de volumen dentro del dieléctrico sobre V_{FB} es cero cuando ésta se encuentra junto al metal.
2. V_{FB} será máximo cuando $x=x_i$, es decir, cuando la carga se encuentra toda junto al semiconductor.
3. Este desplazamiento empieza a hacerse notable a temperaturas del orden o superiores a los 120 °C y su comportamiento depende del tipo de carga móvil, es decir, si se deben a impurezas o a vacancias de oxígeno.

Las cargas móviles producen inestabilidad en el valor V_{FB} , ya que pueden alejarse o se acercarse a la interfaz según el signo y la magnitud del voltaje aplicado entre la compuerta y el semiconductor

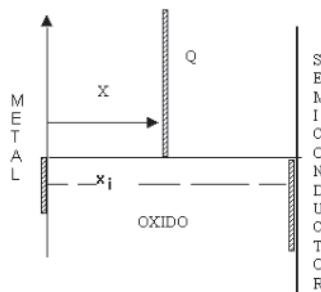


Fig. 2.11 El efecto de un plano de carga dentro del dieléctrico en ausencia de campo externo.

El método más utilizado para la evaluación del tipo y magnitud de las cargas presentes en las estructuras es el método capacitancia-voltaje conocido como C-V. La carga fija produce desplazamiento de la curva C-V a lo largo del eje de las x , o sea, del eje del voltaje. Las cargas en estados superficiales Q_{SS} producen deformación de la curva C-V.

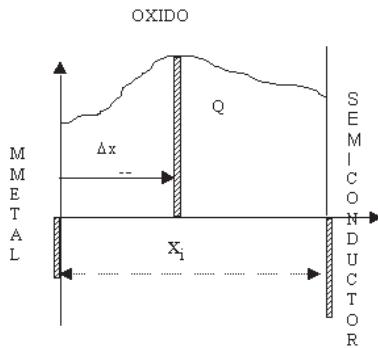


Fig. 2.12 Efecto de una distribución arbitraria de carga dentro del dieléctrico

En los dispositivos MOS, el voltaje para el cual el dispositivo pasa de conducción a no conducción recibe el nombre de Voltaje Umbral (V_T) y se encuentra relacionado con el voltaje de banda plana V_{FB} a través de la expresión:

$$V_T = V_{FB} + \frac{Q_B}{C_i} + 2\varphi_F, \quad (2.50)$$

donde Q_B es la carga en la zona de carga espacial del semiconductor y φ_F es el potencial de Fermi del semiconductor.

Por lo tanto, la presencia de los diferentes tipos de carga asociados al óxido implicará un incremento indeseado y no controlable en el valor del V_T , así como inestabilidad durante el funcionamiento del dispositivo. El control de la densidad de cargas en la interfaz es fundamental para el trabajo normal de los dispositivos y circuitos integrados MOS, siendo un problema crítico para la obtención de altos rendimientos de fabricación en los circuitos integrados.

En el caso de los dispositivos bipolares, el efecto mas dañino producido por las cargas Q_{SS} , Q_{CF} y Q_i lo producen las cargas en estados superficiales Q_{SS} , ya que su presencia incrementa la recombinación superficial, lo que se manifiesta también en un incremento de la corriente inversa y en una disminución de la ganancia de corriente o beta de los transistores.

2.4 FORMAS DE DISMINUIR LA DENSIDAD DE CARGAS EN LA INTERFAZ SIO₂-SI Y EN EL DIELÉCTRICO

Si las cargas Q_{SS} y Q_{cf} se deben a defectos durante el crecimiento del óxido, es de suponer que su densidad pueda controlarse mediante el régimen de oxidación y con tratamientos térmicos posteriores a la oxidación.

En [9] se mostró y explicó por primera vez una dependencia entre la densidad de carga fija y el régimen térmico del proceso de oxidación, así como con el tratamiento térmico posterior a la oxidación. El gráfico que se muestra en la figura 2.13 se conoce con el nombre de triángulo de oxígeno seco y describe cualitativamente, de forma satisfactoria, el comportamiento de la densidad de carga fija y de la carga en estados superficiales con la temperatura de oxidación y luego de un tratamiento térmico a alta temperatura (TTAT) en gas inerte posterior a la oxidación.

Con oxidación en oxígeno seco a una temperatura superior a los 1100 °C, se obtiene la menor densidad de carga. La densidad de carga en la interfaz también puede ser disminuida sometiendo la estructura a un proceso de TTAT en gas inerte por determinado tiempo en un rango amplio de temperatura.

Las flechas indican el sentido de disminuir o aumentar la densidad de carga según el proceso utilizado. Por ejemplo, si se oxida una oblea a 1100

$^{\circ}\text{C}$ en oxígeno seco, se puede obtener una densidad de carga menor de $3 \times 10^{11} \text{ cm}^{-2}$. Si se oxida a 900 $^{\circ}\text{C}$ se obtiene una mayor densidad de carga (línea hipotenusa del triángulo). Este efecto puede explicarse ya que según el modelo de oxidación, a menor temperatura, la velocidad de la reacción de oxidación es menor y parte del agente oxidante que se difundió hasta la interfaz no ha reaccionado al terminar el proceso de oxidación, quedándose ionizado dentro de la estructura. Si la temperatura de oxidación es mayor, la mayor parte de lo que se difunde hasta la interfaz será consumido y la densidad de carga en la interfaz debida a átomos de oxígeno que no llegaron a reaccionar, será menor.

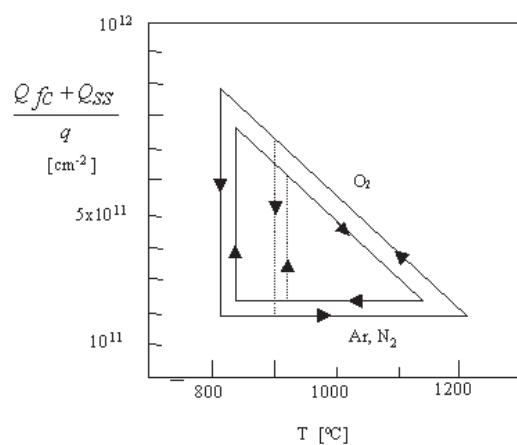


Fig. 2.13 Triángulo de oxígeno seco [13]. Muestra cómo varía la densidad de carga en la interfaz con la temperatura de oxidación y con el TTAT en gas inerte.

Si en lugar de sacar la oblea al terminar el proceso de oxidación, se deja un rato más a la misma temperatura de oxidación, pero en gas inerte, el oxígeno que aun no reaccionó, o termina de reaccionar o se difunde de nuevo hacia afuera, por lo que es de esperar que la densidad de carga en la

interfaz disminuya, tal y como se observa en la figura 2.13. Esto ocurrirá siempre independientemente de la temperatura a que se haya realizado la oxidación. Si el tiempo que se deja la oblea en este TTAT es suficiente, se logrará disminuir la densidad de carga hasta un valor similar, tal como se muestra en la Fig. 2.13 para dos procesos, uno realizado alrededor de 900 °C y otro alrededor de 800 °C. En todos los casos se puede en principio llegar a la misma densidad de carga obtenida cuando se oxida a mayor temperatura, que en el gráfico se indica como menor de $2 \times 10^{11} \text{ cm}^{-2}$.

En el caso de oxidación en vapor de agua, la densidad de carga en la interfaz puede ser mayor que para el oxígeno, aunque el efecto del TTAT también se observa.

La densidad de cargas móviles puede reducirse de dos formas, según su origen.

Para reducir la presencia de vacancias de oxígeno se utiliza un tratamiento térmico a baja temperatura (TTBT) en presencia de hidrógeno, que al mismo tiempo reduce también la densidad de carga debida a Q_{ss} y Q_{cf} que haya quedado luego del TTAT [16,18]. Esto se explica debido a que el hidrógeno se introduce en el dieléctrico y puede formar enlaces saturados del tipo Si--H, los que resultan estables a temperaturas inferiores a los 500 °C, reduciendo la densidad de carga en los centros de defecto debida a enlaces no saturados.

Debe señalarse que este proceso no puede ser utilizado para disminuir altas densidades de carga, por cuanto implica una alteración de la estructura del SiO_2 debido a la formación de los enlaces Si-H. Una alta densidad de estos enlaces tiende a disminuir la densidad del óxido y a empeorar considerablemente sus propiedades.

Para reducir la presencia de átomos de impurezas en el dieléctrico, fundamentalmente sodio (Na^+), hay que recurrir a la **tecnología limpia**, que no es otra cosa que evitar su penetración en el óxido en cualquiera de las

etapas del procesamiento. Para esto deben utilizarse productos químicos de pureza electrónica con bajo contenido de sodio, agua deionizada de más de 20 MΩ-cm para temperatura menor de 20 °C. También se requieren gases de alta pureza en los procesos de alta temperatura, así como tubos especiales de cuarzo, carburo de silicio o polisilicio con bajo contenido de sodio y con estructura resistente a la difusión del sodio proveniente del ambiente exterior a través de sus paredes.

A esto último debe añadirse que los hornos de alta temperatura con tubos de cuarzo deben mantenerse siempre a una temperatura superior a los 500 °C. Si el tubo se enfria por debajo de esa temperatura, recristaliza en otra estructura cristalina que ocupa un volumen mayor que la anterior, y tiende a rajarse o a hacerse poroso, incrementando la probabilidad de contaminación de las estructuras con impurezas que atraviesan sus paredes. Hoy en día estos tubos pueden sustituirse por tubos de carburo de silicio o Si, que no presentan la dificultad antes señalada, aunque resultan bastante más costosos.

Por último, además de los indicados, existen otros métodos que se utilizan para disminuir la densidad de cargas en la interfaz y en el dieléctrico; por ejemplo la oxidación en presencia de HCl o después de purgar el horno durante cierto tiempo con este producto, con vistas a eliminar las impurezas de sodio; la utilización de vidrio de fosfósilicato (VFS) para atrapar las cargas móviles que ya existen en el dieléctrico y el recubrimiento de las estructuras con nitruro de silicio (Si_3N_4) para evitar que penetren impurezas luego de crecidos los óxidos.

2.5 MEDIDAS DE CONTROL DE LA CONTAMINACIÓN DURANTE EL PROCESAMIENTO DE LA OBLEA

Como se señaló en el párrafo anterior, el estricto control de la contaminación debida al sodio en el SiO₂ de la compuerta de un dispositivo MOS resulta fundamental para el correcto funcionamiento del mismo. Este hecho hizo que aunque el principio de funcionamiento del dispositivo de efecto de campo se conociera antes del principio de funcionamiento del transistor bipolar, tecnológicamente no fuera factible su realización hasta varios años después de establecida la tecnología de fabricación de los dispositivos y circuitos integrados bipolares. Fue el concepto de Tecnología Limpia el que permitió la fabricación de dispositivos MOS.

2.5.1 *Tecnología limpia*

El concepto de tecnología limpia, como ya se indicó, se utiliza para incluir todo un conjunto de medidas para evitar al máximo la contaminación por impurezas indeseadas de los dispositivos durante los procesos de fabricación. Uno de los pasos más críticos, resultó el control de la pureza del óxido de compuerta y de la interfaz SiO₂-Si. Para ello, resulta necesario disminuir el nivel residual de Na en cualquier producto químico o gas que se utilice en la fabricación de los dispositivos.

Además del Na, otros metales, como el K, Fe, Au, Cu, Ag y otros resultan igualmente dañinos ya que son capaces de crear estados energéticos permitidos cercanos al medio de la banda prohibida. Estos metales originan trampas que actúan como centros de generación-recombinación, afectando sensiblemente las características de los dispositivos.

Durante la obtención de un producto químico o gas, en él quedan residuos de otros productos intermedios, iones y partículas.

Por estos motivos, el desarrollo de la Tecnología Planar y en especial de la tecnología MOS requirió del desarrollo y producción de productos químicos, gases y materiales que cumpliesen normas muy estrictas en cuanto al nivel de contaminación con estos productos. Surgió así un grado de pureza denominado pureza electrónica o grado electrónico, el cual a su vez, con el desarrollo de nuevas tecnologías fue incrementando sus requerimientos, dando lugar a diferentes niveles de pureza y de tamaño de partículas que contienen, dentro del propio grado electrónico. Los valores límites de impurezas establecidos para estos productos está hoy en día por debajo de 1 ppb de los elementos indicados en el producto final

La presencia de partículas en el proceso de fabricación de C.I. no solo resulta importante en los productos químicos y gases que se utilizan, sino también en las áreas de trabajo donde se fabrican los dispositivos. Esto resulta especialmente importante en las áreas donde se realizan las fotolitografías, ya que una partícula del tamaño de un micrómetro que se deposita sobre la máscara o sobre la oblea durante el proceso de alineación y exposición, puede dar lugar a la alteración de las formas de las figuras que se desea transferir a la oblea.

En este caso, las áreas de trabajo también se clasifican de acuerdo al tamaño y concentración de partículas suspendidas en el aire. Según el estándar Federal 209D de los Estados Unidos, de 1989, las áreas se clasifican, según su contenido de partículas, ver Tabla 2.1.

Las clases 0.1 y 0.01 son muy difíciles de implementar, ya que una persona en un área de 10 m^2 genera lo suficiente para dar lugar a un área clase 1.

Tabla 2.1 Clasificación de áreas limpias

Tipo de área	Número de partículas por pie cúbico con tamaño mayor de:				
	0.1 μm	0.2μm	0.3μm	0.5μm	5μm
CLASE 1000	--	--	--	<1000	<7
CLASE 100	--	<750	<300	<100	--
CLASE 10	<350	<75	<30	<10	--
CLASE 1	<35	<7.5	<3	<1	--

Por lo señalado, algunas medidas generales se toman en consideración tanto en la manipulación de productos como en la higiene personal y organización de los locales. Entre ellas podemos mencionar:

- Limpieza diaria;
- Limpieza personal, nada de cosméticos, ni cigarro;
- Eliminar papeles, cosas que guarden polvo o que lo desprendan;
- Entrada restringida; En cuartos de clase 10 y 1 entrar solo lo imprescindible.
- Batas de tela sintética que no suelte polvo y cobertura al máximo de la piel del la persona, como se muestra en la Fig. 2. 14;
- Filtros de agua, gases y fotorresina locales;
- Sistemas locales de agua deionizada;
- Cuarzos de alta pureza, tubos de silicio o de carburo de silicio para los hornos de alta temperatura
- Los productos químicos deben manipularse solo en cuartos limpios;
- No se pueden almacenar en recipientes intermedios;
- En las especificaciones del proceso en que se usan, se establece el tiempo máximo de utilización o el número máximo de obleas a procesar;

- Las líneas de gases deben ser de acero 304, 306 o de níquel, pulidas por su pared interior; en algunos casos se usa tubería de teflón o PVC;
- Los reguladores y conexiones son especiales para este tipo de gases. Las interconexiones no deben soldarse, se usan conexiones Swagelok.

En la Fig. 2.15 se muestra una estructura típica de un cuarto clase 10 con una cabina de flujo laminar clase 1, donde se observa los grados de anidamiento de un área de clase menor dentro de otra de clase mayor, así como el uso de filtros de alta eficiencia para atrapar las partículas mayores de un micrómetro, los llamados filtros HEPA. En la Fig. 2.16 se muestra la organización típica de un área limpia incluyendo las áreas de servicio.



Fig. 2.14 Ropa para trabajar en un área limpia.

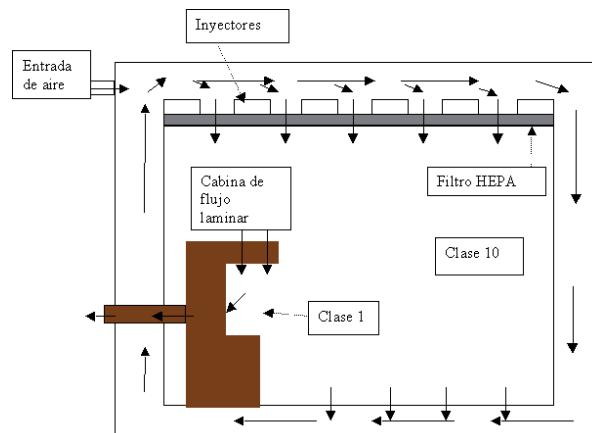


Fig. 2.15 Estructura típica de un cuarto clase 10 con una cabina de flujo laminar clase 1.

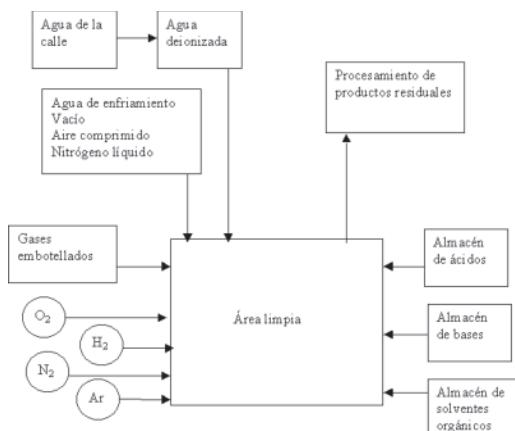


Fig. 2.16 Organización típica de un área limpia con sus áreas de servicio.

2.5.2 Proceso de limpieza de obleas

El proceso de la oblea antes de la oxidación resulta fundamental para lograr una buena calidad del óxido crecido, así como de la interfaz SiO₂-Si.

Ya se señaló que los productos utilizados tienen que ser de grado electrónico, pero además, la limpieza busca un conjunto de objetivos en los diferentes pasos que la componen.

Desde finales de los años 60, los procesos de limpieza se basan en el llamado proceso RCA [13]. Este proceso incluye los siguientes pasos y objetivos:

Tabla 2.2 Pasos del proceso de limpieza de la oblea y objetivos de cada uno.

Proceso	Objetivo
1. Tricloroetileno US 15 min;	Eliminar residuos orgánicos
2. Acetona US 15 min;	Eliminar residuos orgánicos
3. SO ₄ H ₂ +H ₂ O ₂ 80 °C 15 min;	Eliminar residuos orgánicos e iónicos
4. FH (puro o diluido);	Eliminar el óxido nativo
5. NH ₄ OH+ H ₂ O ₂ ;	Eliminar residuos iónicos y crear una pequeña película de óxido formado durante la limpieza.
6. ClH+ H ₂ O ₂	Eliminar residuos iónicos y sustituir el óxido nativo, por una pequeña película de óxido formado durante la limpieza.

La secuencia utilizada parece lograr que los residuos que deja cada etapa de limpieza vayan siendo cada vez menores. Por ejemplo, la acetona se usa después del tricloroetileno, por ser más volátil y entre el paso 5 y 6, el 6 deja menos residuos que el 4.

Entre los pasos e al 6 de cada limpieza se realiza un enjuague en agua deionizada.

Este proceso se utiliza con ligeras modificaciones en diferentes etapas del proceso. Por ejemplo, si ya se oxidó la oblea y no se desea atacar el óxido, se elimina el paso 4.

Si la oblea tiene ya depositada alguna capa que se ataque en alguno de estos productos, ese paso se elimina o sustituye por otro.

2.6 BIBLIOGRAFÍA

- [1]. B.E. Deal, A.S. Grove, "General Relationship for the Thermal Oxidation of Silicon", Journal of Applied Physics, 36 (1965) 3770.
- [2]. A,S, Grove, "Physics and Technology of Semiconductor Devices", John Wiles and Sons, Inc., New York, 1967.
- [3]. N. Cabrera, N.F. Mott, "Theory of the Oxidation of Metals," en Reports of Progress in Physics, 12 (1948) 163.
- [4]. S.A. Schafer, S.A. Lyon, "New model of the rapid initial oxidation of silicon", Appl. Phys. Lett. 47 (1985) 154.
- [5]. Y. Z. Lu, C. Cheng, "A new model for the growth of silicon oxide", J. Appl. Phys. 56 (1984) 1608.
- [6]. H.Z. Massoud, J.D. Plummer, E.A. Irena, "Thermal oxidation of Silicon in Dry Oxigen: Growth rate enhancement in the Thin Regime I, Experimental Results", J. Electrochem. Soc. 132 (1985) 2685-2693.
- [7]. W .A Weyl, S.C Marboe, "The Constitution of Glasses", John Willey and Sons, New York, 1962.
- [8]. F. Forstmann, "Theory of Imperfect Crystalline Solids", Trieste Lectures,1970. pp. 511-533.
- [9]. E. Kooi, "Characteristics of the Surface-State Charge of Thermally Oxidized Silicon". Philips, Research Report, 21 (1966) 477.

- [10]. B.E. Deal, M. Sklar, A.S. Grove, E.H. Snow, "Characteristics of the surface-state charge (Q_{SS}) of thermally oxidized silicon", Journal of the Electrochemical Society, 1441 (3) (1967) 226-273.
- [11]. E. Thomas, D.R. Young, " Space Charge model for surface potential shifts in silicon passivated with thin insulating layers, IBM Journal, 9 (1964) 368.
- [12]. E.H. Snow, A.S. grove, B.E. Deal, C.T. Sah, "Ion Transport Phenomena in Insulating Films", Journal of Applied Physics, 36 (5) (1965) 1664.
- [13]. W. Kern, "Handbook of Semiconductor wafer Cleaning technology: Science, Technology and applications", Noyes, Park Ridge, N.J. 1993.

CAPÍTULO 3

CURVAS CAPACITANCIA-VOLTAJE (C-V) DE ESTRUCTURAS MDS: FUNDAMENTO TEÓRICO Y APLICACIONES FUNDAMENTALES

3.1 FUNDAMENTO TEÓRICO DE LAS CURVAS C-V

Las curvas de capacitancia-voltaje (C-V) de estructuras metal-dieléctrico-semiconductor (MDS) permiten caracterizar procesos tecnológicos así como determinar algunos parámetros de dichas estructuras.

Es conocido que cuando se aplica un potencial externo a una estructura MDS, tienen lugar los efectos que se muestran en la Fig. 3.1, donde según el tipo de substrato y el potencial aplicado, se pueden obtener regiones de acumulación, de empobrecimiento o de inversión en la región del semiconductor adyacente al dieléctrico.

Los fenómenos que tienen lugar en esa región superficial pueden describirse resolviendo la ecuación de Poisson:

$$\frac{d^2\Psi}{dx^2} = -\frac{\rho(x)}{k_s \epsilon_0} . \quad (3.1)$$

La concentración de electrones y huecos, por ejemplo, para el caso de un substrato homogéneo de tipo P, viene dada por:

$$n_p = n_{p0} e^{\frac{q\Psi}{kT}} = n_{p0} e^{\beta\Psi}, \quad (3.2)$$

$$p_p = p_{p0} e^{-\frac{q\Psi}{kT}} = p_{p0} e^{-\beta\Psi}. \quad (3.3)$$

La concentración de electrones y huecos en la superficie será:

$$n_s = n_{p0} e^{\beta\Psi_s}, \quad (3.4)$$

$$p_s = n_{p0} e^{-\beta\Psi_s}. \quad (3.5)$$

La densidad de carga en la región de carga espacial será:

$$\rho(x) = q(N_d^+ - N_a^- + p_p - n_p). \quad (3.6)$$

En el volumen del semiconductor alejado de la superficie hay neutralidad eléctrica, o sea, $\psi(x)=0$.

Sustituyendo (3.1), (3.2) y (3.3) en (3.6) resulta:

$$\frac{d^2\Psi}{dx^2} = -\frac{q}{k_s \epsilon_0} \left[p_{p0} (e^{-\beta\Psi} - 1) - n_{p0} (e^{\beta\Psi} - 1) \right] \quad (3.7)$$

Integrando desde el volumen hacia la superficie:

$$\int_0^x \frac{d\Psi}{dx} d\left(\frac{d\Psi}{dx}\right) = -\frac{q}{k_s \epsilon_0} \int_0^\Psi \left[p_{p0} (e^{-\beta\Psi} - 1) - n_{p0} (e^{\beta\Psi} - 1) \right] d\Psi, \quad (3.8)$$

luego el campo eléctrico F será:

$$F^2 = \left(\frac{2kT}{q} \right)^2 \frac{q\beta p}{2k_s \epsilon_0} \left[\left(e^{-\beta\psi} + \beta\psi - 1 \right) + \frac{n_p p_0}{p_0} \left(e^{\beta\psi} - \beta\psi - 1 \right) \right]. \quad (3.9)$$

Si expresamos la longitud extrínseca de Debye en un semiconductor tipo P a través de:

$$L_D = \sqrt{\frac{kT k_s \epsilon_0}{q^2 p_0}}, \quad (3.10)$$

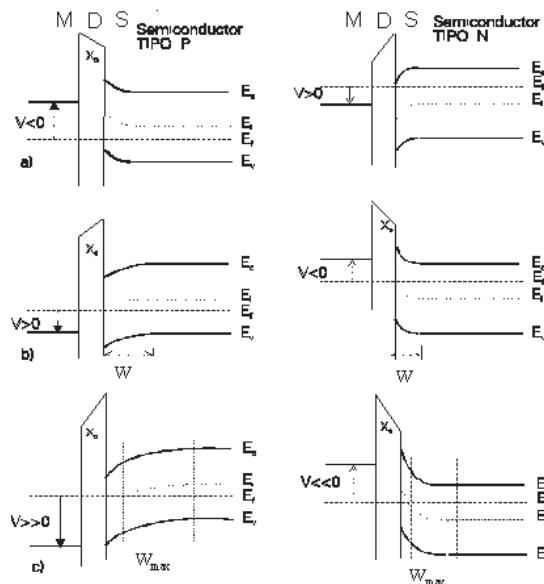


Fig. 3.1 Diagrama de bandas de energía de una estructura MDS ideal con una polarización externa V aplicada. Caso de: a) acumulación; b) empobrecimiento y c) inversión de la superficie del semiconductor, para substratos tipo P (izquierda) y tipo N (derecha).

y definimos una función F_o como:

$$F_o(\beta\psi, \frac{n_p 0}{p_p 0}) = \left\{ \left[e^{-\beta\psi} + \beta\psi - 1 \right] + \frac{n_p 0}{p_p 0} \left[e^{\beta\psi} - \beta\psi - 1 \right] \right\}^{\frac{1}{2}} \geq 0, \quad (3.11)$$

el campo eléctrico se puede expresar finalmente como:

$$F = \pm \frac{\sqrt{2kT}}{qL_D} F_o(\beta\psi, \frac{n_p 0}{p_p 0}). \quad (3.12)$$

En la superficie, el campo será igual a:

$$F_s = \pm \frac{\sqrt{2kT}}{qL_D} F_o(\beta\psi_s, \frac{n_p 0}{p_p 0}) \quad (3.13)$$

y la carga por unidad de área para producir este campo será igual a:

$$Q_s = k_s \epsilon_0 \frac{\sqrt{2kT}}{qL_D} F_o(\beta\psi_s, \frac{n_p 0}{p_p 0}). \quad (3.14)$$

Si el potencial en la superficie es $\psi_s=0$, para semiconductores homogéneos, la carga Q_s será cero y se dice que hay condición de banda plana, ver Fig. 3.2.

Si $\psi_B > \psi_s > 0$, donde $\psi_B = (E_f - E_i)/q$, entonces $Q_s < 0$ y existirá empobrecimiento bajo la superficie. La densidad superficial de carga tenderá a:

$$Q_s \rightarrow \sqrt{\psi_s}. \quad (3.15)$$

Si $\psi_s > 2\psi_B$ se tiene el caso de inversión fuerte, donde

$$Q_s \rightarrow -e^{\frac{q\psi_s}{2kT}}. \quad (3.16)$$

La capacitancia diferencial en la zona de empobrecimiento dentro del semiconductor será:

$$C_D = \frac{dQ_S}{d\psi_S} = \frac{k_S \epsilon_0}{\sqrt{2} L_D} \frac{\left[\left(1 - e^{-\beta\psi} \right) + \frac{n_{p0}}{p_{p0}} \left(e^{\beta\psi} - 1 \right) \right]}{F_0 \left(\beta\psi, \frac{n_{p0}}{p_{p0}} \right)}. \quad (3.17)$$

La capacitancia del semiconductor cuando hay banda plana será:

$$C_D = \frac{k_S \epsilon_0}{L_D}. \quad (3.18)$$

La curva C-V para estas estructuras, en el caso ideal, se obtiene si consideramos que parte del voltaje aplicado V cae a través del dieléctrico V_i y parte en la región de carga espacial del semiconductor ψ_S .

$$V = V_i + \Psi_s. \quad (3.19)$$

El voltaje que cae a través del dieléctrico es:

$$V_i = F_i X_o = \frac{k_S}{k_i} X_o F_S = \frac{Q_S X_o}{\epsilon_0 k_S} = \frac{Q_S}{C_i}. \quad (3.20)$$

La capacidad total $C = \frac{dQ_s}{dV}$ será la resultante de la capacidad del dieléctrico $C_i = k_i \epsilon_0 / X_o$ en serie con la capacidad del semiconductor C_D , o sea:

$$\frac{1}{C} = \frac{1}{C_i} + \frac{1}{C_D}. \quad (3.21)$$

Importancia especial tiene el valor de la capacitancia de esta estructura correspondiente a la condición de banda plana. Para los substratos homogéneos, la expresión de la capacitancia de banda plana será:

$$C_{FB}(\psi_s = 0) = \frac{k_i \epsilon_0}{X_o + \frac{k_i}{k_s} L_D}. \quad (3.22)$$

En la medición de la curva C-V, generalmente están presentes una señal de polarización y una señal alterna de muy pequeña amplitud para medir la capacitancia diferencial de la estructura MDS a cada voltaje de polarización dado. Según la frecuencia con que varían estas dos señales pueden ocurrir diferentes comportamientos de la estructura.

Caso 1: La frecuencia de la señal de voltaje de corriente alterna (CA), aplicada a la compuerta para la medición de la capacitancia diferencial es lo suficientemente alta para que la concentración de portadores minoritarios que se generan cuando ocurre la condición de inversión, no puedan seguir las variaciones de dicha señal en el tiempo, pero sí las de la señal de polarización cuya frecuencia de variación se fija suficientemente baja.

En este caso, como la variación en el tiempo de la señal de polarización es lenta, cuando la magnitud y sentido de la polarización es el adecuado según el tipo de substrato, se forma la región de inversión inmediatamente debajo de la superficie. Cuando hay región de inversión, todo nuevo incremento de voltaje (o sea de carga en el electrodo externo) se compensa por un incremento de carga en la región de inversión.

Hay que recalcar que como estas cargas en la región de inversión se deben a portadores minoritarios que se generan en el semiconductor, su velocidad de reacción (dada por el tiempo requerido para generarse o recombinarse los portadores y por lo tanto aumentar o disminuir la carga en la zona de inversión al variar la carga en el electrodo exterior), es relativamente lenta, ya que ésta depende del tiempo de generación y de recombinación del material.

Cuando se aplica a la compuerta una señal de medición de CA para medir la capacitancia diferencial de la estructura MOS, la frecuencia de esta señal puede escogerse de manera que sea lo suficientemente alta, para que en lo que ella varía no de tiempo a que la concentración de portadores minoritarios de la región de inversión cambie.

La variación de carga en el electrodo externo producto de la variación de potencial aplicado a la compuerta tiene que ser compensado por una variación de carga en el semiconductor. Si los portadores minoritarios en la región de inversión no pueden compensar dicha variación de carga en la compuerta, la región de carga espacial debe aumentar o disminuir ligeramente su ancho para compensar la carga externa. Esto es posible porque el ancho de la zona de empobrecimiento crece o disminuye al variar la concentración de portadores mayoritarios en el semiconductor y esta concentración de portadores mayoritarios puede variar con el tiempo de relajación de Maxwell correspondiente al material semiconductor en cuestión. Esto permite perfectamente seguir señales externas con

frecuencias en el orden de 1 MHz, que es la frecuencia típica utilizada en la señal de medición de C-V de alta frecuencia.

En este caso la **curva C-V de alta frecuencia** que se obtiene, variará según se muestra en la Fig. 3.2. Por otro lado, los nuevos incrementos del voltaje de polarización, una vez que se forma la región de inversión no incrementan más el ancho de la zona de empobrecimiento. La capacidad de empobrecimiento que se mide a partir de ese voltaje se mantendrá constante e igual que su valor máximo al momento de formarse la región de inversión. La capacitancia total también se mantendrá constante.

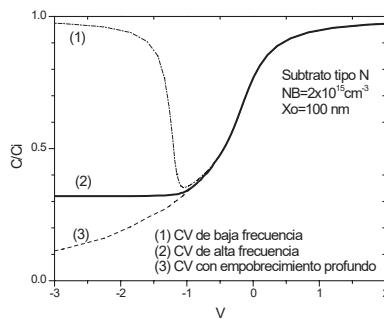


Fig. 3-2 Curvas C-V de una estructura MDS con substrato tipo n, para diferente rango de frecuencias de la señal de medición y del voltaje de polarización.

Caso 2: La frecuencia de la señal de medición aplicada a la compuerta es lo suficientemente baja para que la concentración de portadores minoritarios en la región de inversión pueda seguir a las variaciones de dicha señal, una vez que se forma la región de inversión.

En este caso, a medida que se incrementa el voltaje, la estructura se comporta cada vez mas como un condensador de placas paralelas con las cargas acumuladas a ambos lados del dieléctrico. La curva C-V vuelve a

subir hasta alcanzar el valor de la capacitancia en acumulación. Este régimen de medición se conoce como régimen de baja frecuencia y la curva que de él se obtiene se conoce como **curva C-V de baja frecuencia**.

Caso 3: La variación de la señal que se utiliza para polarizar el dispositivo es suficientemente rápida, para que varíe sin dar tiempo a que se generen los portadores minoritarios necesarios para formar la región de inversión.

En este caso la región de inversión no se forma y la curva C-V no se mantiene constante en la región de inversión, ya que cada nuevo incremento de voltaje produce un aumento en el ancho de la región de carga espacial para poder compensar el incremento de carga en el electrodo exterior y por lo tanto la capacitancia disminuye todo el tiempo. Esto ocurre hasta que se produce la ruptura del semiconductor por efecto de campo. A esta condición de medición se le llama **empobrecimiento profundo**, ver Fig. 3.2.

Para estas estructuras ideales, descritas hasta el momento, se han calculado un conjunto de curvas que se muestran en las Fig. 3.3 y 3.4, cuya utilidad se verá a continuación. En la Fig. 3.3 se muestra la dependencia de la capacitancia de banda plana (3.22) relativa a la capacitancia máxima, (C_{FB}/C_i) en función del espesor del dieléctrico y de la concentración de impurezas de un substrato homogéneo. En la Fig. 3.4 se muestran los valores de capacitancia mínima relativa para la curva de alta frecuencia, calculados a partir de la expresión:

$$C_{min}' = \frac{k_i \epsilon_0}{X_o + \frac{k_i}{k_s} W_{max}}, \quad (3.23)$$

donde W_{max} es el ancho máximo de la región de carga espacial. W_{max} se calcula a través de la expresión para el ancho de la zona de carga espacial de una unión abrupta, sustituyendo el voltaje interconstruido por el potencial superficial cuando se alcanza la inversión en la superficie del semiconductor $\psi_s(inv)=2\psi_F$:

$$W_{max} = \sqrt{\frac{2K_S \epsilon_0 (2\psi_F)}{qN_B}}. \quad (3.24)$$

3.2 APLICACIÓN DE LAS CURVAS C-V DE ALTA FRECUENCIA A LA DETERMINACIÓN DE LAS PROPIEDADES DE LA INTERFAZ SiO₂-Si

En las estructuras metal-dieléctrico-semiconductor, en la realidad, tiene lugar todo un conjunto de fenómenos complementarios que afectan el comportamiento de las curvas C-V ideales descritas hasta ahora. Es precisamente este hecho, el que se utiliza para, a través del análisis de las diferencias entre las curvas reales y las ideales, determinar todo un conjunto de parámetros de las estructuras, que sirven para su caracterización e incluso para el control tecnológico de los procesos que se utilizan en su fabricación. Es este último, uno de los usos más generalizados y más importantes de estas curvas C-V.

Como ejemplo de utilización analizaremos una estructura formada por un metal, óxido de silicio como dieléctrico y silicio. Esta estructura resulta muy importante en la Tecnología de fabricación de dispositivos Semiconductores. Como el SiO₂ se usa para proteger la superficie del silicio y sobre él se depositan otras capas de metales con los cuales se

realizan las interconexiones entre los diferentes elementos del circuito integrado, donde quiera que se tenga una capa de metal sobre el SiO₂, el cual a su vez está sobre el silicio, hay una estructura MDS o MOS.

Además el SiO₂ se utiliza como dieléctrico de compuerta en la tecnología MOS. En este caso los requisitos que se imponen a las características del óxido y a la interfaz SiO₂-Si son muy estrictos. Las curvas C-V se usan para caracterizar y controlar la calidad de estas estructuras.

En estructuras MDS reales existen diferentes cargas en el óxido y en la interfaz SiO₂-Si, que influyen en la condición de banda plana. En caso general, éste se define como:

$$V_{FB} = \phi_{MS} - \frac{Q_{ss}(0) + Q_{cf} + Q_i}{C_i} - \frac{Q_s(0)}{C_i}, \quad (3.25)$$

donde:

ϕ_{MS} es la diferencia de potencial de contacto entre el metal y el semiconductor usados en la estructura;

Q_{ss} es la densidad de carga debida a estados superficiales localizados en la interfaz; su valor depende del potencial aplicado;

Q_{cf} es la densidad de carga debida a la carga fija localizada en la interfaz con el dieléctrico o en el dieléctrico cercano a ella; no depende del potencial aplicado.

Q_i es la densidad de carga debida a cargas móviles que se encuentran dentro del dieléctrico; su valor depende de la posición de estas cargas en el interior del dieléctrico.

Estas cargas, presentes en toda estructura real, hacen que incluso cuando no haya voltaje externo aplicado, las bandas del semiconductor no estén planas. Para hacerlas planas es necesario aplicarle un potencial externo igual a V_{FB} .

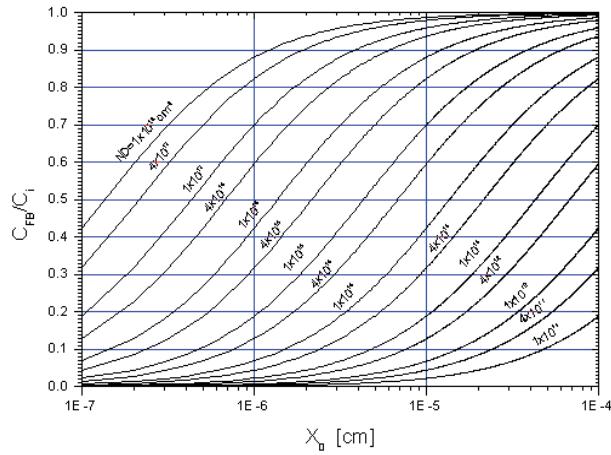


Fig. 3-3 Curvas C_{FB}/C_i de en función del espesor del óxido de silicio X_o para diferentes concentraciones de impurezas N_D en el substrato.

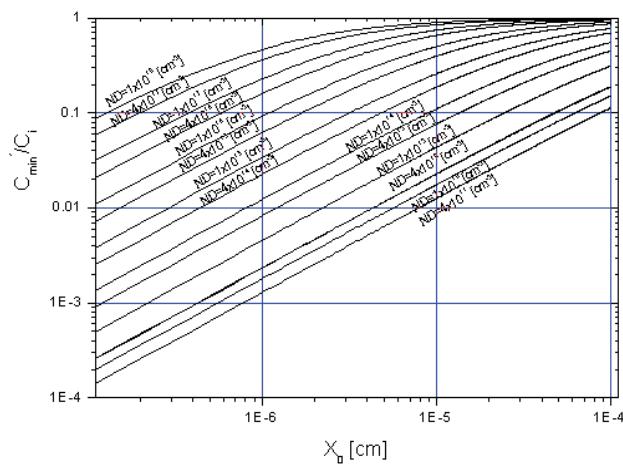


Fig. 3-4 Curvas de C'_{min}/C_i en función del espesor del óxido de silicio X_o para diferentes concentraciones de impurezas N_D en el substrato.

Esto hace que una curva C-V real no tenga la forma descrita en la Fig. 3.2, sino que se encuentre desplazada generalmente hacia valores negativos de voltaje y que además pueda tener ligeras deformaciones en la región en que varía.

El hecho de que generalmente la curva C-V se encuentre desplazada hacia valores negativos de voltaje se explica debido a que la carga fija y la carga móvil son cargas positivas, luego para compensarlas hace falta aplicar un voltaje negativo a la estructura. Solamente en el caso en que estas cargas fueran muy pequeñas y que la Q_{ss} predominara y fuera negativa, se observaría un corrimiento hacia voltajes positivos de la curva. La carga total que aportan estos estados Q_{ss} puede tener cualquier signo y su valor depende del voltaje aplicado y de la frecuencia de medición, ya que se deben al intercambio de estos centros con portadores. Cada captura o cesión se caracteriza por una energía de activación y un tiempo de relajación. Actualmente con el perfeccionamiento de las tecnologías el efecto de Q_{ss} puede predominar dado lo pequeñas que se han logrado hacer tanto la carga móvil como la fija.

3.2.1 Determinación del voltaje de banda plana

En la Fig.3.5 se muestra la curva C-V real de una estructura MOS. La relación C'_{min}/C_i es aproximadamente 0.2. La concentración homogénea de impurezas en el substrato de partida se calcula buscando en las curvas C'_{min}/C_i vs. X_o de la Fig. 3.4, para qué concentración del substrato el valor de C'_{min}/C_i es igual a 0,2, cuando el espesor del óxido es de 98 nm. El valor de concentración de impurezas que se obtiene en este ejemplo es de $6 \times 10^{14} \text{ cm}^{-3}$. Seguidamente, en la Fig.3.3, se selecciona la curva correspondiente a $N_B = 6 \times 10^{14} \text{ cm}^{-3}$ y para un óxido de 98 nm se busca el valor de C_{FB}/C_i . El

valor obtenido es de 0.64. Si determinamos en la Fig. 3.5, qué valor de voltaje corresponde a ese valor de la capacitancia relativa, habremos determinado el valor del V_{FB} para la estructura en análisis, que en este caso es de -0.83 V.

Si conocemos el tipo de material utilizado en la compuerta se puede calcular el valor de Φ_{MS} , determinando previamente los valores de Φ_M y Φ_{OS} (trabajo de extracción del metal y del semiconductor intrínseco), así como el de $q\Phi_{FS}=[E_f-E_i]$.

En varios libros o manuales pueden encontrarse valores de Φ_M y Φ_S para diferentes materiales. Debe tenerse en cuenta que los valores que aparecen dependen de la técnica usada para su determinación. En la Fig.3.6 se muestra el gráfico de E_f-E_i vs. T , para diferentes concentraciones de substrato N_B .

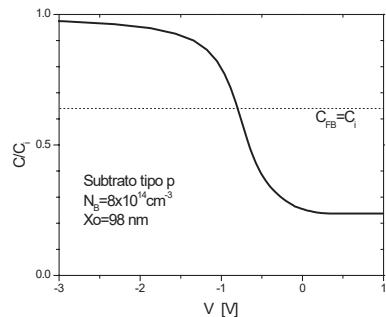


Fig. 3.5 Curva C-V de alta frecuencia para un capacitor MOS cuyo espesor es de 98 nm y la

Por ejemplo, si tomamos $q\Phi_M= 4.1$ eV para el aluminio y $\Phi_{os}=4.7$ eV para el silicio intrínseco, y tenemos un substrato de silicio tipo N cuyo $E_f=0.4$ eV, el trabajo de extracción del semiconductor será:

$$q\Phi_S = 4.7 - 0.4 = 4.3 \text{ eV} \quad (3.26)$$

y la diferencia de potencial de contacto entre el metal y el semiconductor será:

$$q\Phi_{MS} = 4.1 - 4.3 = -0.2 \text{ eV.} \quad (3.27)$$

El valor de Φ_S puede calcularse también, tomando en consideración que la energía desde la banda de valencia hasta el nivel de vacío es de 5.25 eV, y por lo tanto:

$$q\Phi_{OS} = 5.25 - Eg/2 - 0.4 = 4.3 \text{ eV} \quad (3.28)$$

y

$$q\Phi_{MS} = 4.1 - 4.3 = -0.2 \text{ eV} \quad (3.29)$$

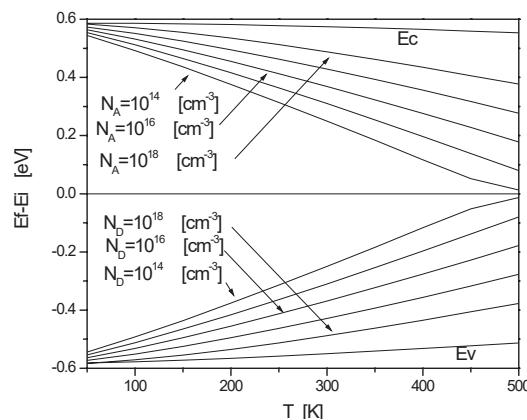


Fig. 3.6 Gráfico de la variación de $E_f - E_i$ en función de la temperatura para diferentes concentraciones de impurezas en un substrato de silicio P y N.

La diferencia de potencial de contacto caerá parte en el dieléctrico y parte en el semiconductor. La parte que cae en cada uno puede calcularse mediante las expresiones (3.20) y (3.13). Para determinar el valor de $Q_{ss} + Q_{fc} + Q_i$, se calcula Φ_{MS} y $C_i = k_s \epsilon_0 / X_o$, donde $k_s = 11.7$, $k_i = 3.9$, $\epsilon_0 = 8.86 \cdot 10^{-14}$ F/cm y X_o es el espesor del óxido. Estos los valores se sustituyen en (3.22), junto con el valor medido de V_{FB} .

Como ya se señaló, los valores de Φ_M dependen del método utilizado para su determinación. Así por ejemplo para el aluminio, además del valor de 4.1 eV se puede encontrar en la literatura un valor de 4,25 eV, si se utiliza otra técnica de medición. Además, si se utiliza polisilicio como material de compuerta, la diferencia de potencial de contacto vendrá dada por la diferencia entre los niveles de Fermi del polisilicio y del substrato. Como resulta bastante difícil poder determinar con exactitud la concentración de impurezas en el polisilicio, también será difícil el poder determinar con exactitud su nivel de Fermi correspondiente.

Todo esto hace que, aunque en principio es posible la determinación de la densidad de carga en la interfaz mediante este método, en la práctica solo se puede obtener un valor aproximado. Para comprobar la repetibilidad de un proceso tecnológico, generalmente se acostumbra a determinar una carga equivalente que se define como:

$$Q_{eqv} = V_{FB} \cdot C_i . \quad (3.30)$$

Esta carga no tiene sentido físico, pero sí permite verificar la repetibilidad de los parámetros de los procesos tecnológicos, así como que los mismos se encuentren dentro de determinado rango. En este caso no es necesario conocer el valor de Φ_M ni de Φ_{MS} . Esta técnica de caracterización es ampliamente utilizada en el control de procesos en la industria.

3.2.2 Determinación de la magnitud de la carga móvil

Para determinar la carga móvil, se utiliza el hecho de que la carga móvil que se ha encontrado siempre dentro del dieléctrico es positiva. Por lo tanto, si se aplica un voltaje negativo a la compuerta, toda la carga tenderá a desplazarse hacia la frontera dieléctrico-compuerta. Del análisis del comportamiento de esta estructura descrito en [2,3] se puede ver que el aporte de las cargas móviles al V_{FB} será cero si se logra que, subiendo la temperatura de la estructura en presencia de un campo eléctrico adecuado, todas las cargas se desplacen hacia esa frontera.

En la práctica se ha demostrado que cuando las cargas móviles en el óxido se deben a iones sodio, para lograr el efecto señalado basta aplicar un campo superior a 10^5 V/cm y una temperatura superior a los 150 °C, durante un tiempo superior a los 10 minutos. En el caso de que existan cargas móviles debidas a la presencia de vacancias de oxígeno, se requiere una temperatura superior a los 200 °C, debido a la menor movilidad de estos iones.

De esta forma, se pueden seguir los siguientes pasos para separar el valor de la densidad de carga móvil Q_i , de la carga equivalente presente en una estructura MDS.

1. Aplicar un voltaje de -10 V a la compuerta de dicha estructura con respecto al semiconductor (también referido como substrato);
2. Calentar la muestra a la temperatura de 200 °C y mantenerla durante 10 ó 15 minutos;
3. Enfriar la muestra sin quitar la polarización aplicada;
4. Una vez fría, registrar la curva C-V de alta frecuencia. La curva obtenida corresponderá al caso de una muestra con $Q_i=0$; si se

conoce el valor de Φ_{MS} se puede determinar la suma de la densidad de cargas fijas y de estados superficiales $Q_c + Q_{ss}$.

5. Una vez realizada esta medición, se procede a aplicar a la compuerta un voltaje de +10 V.
6. Se sube la temperatura de nuevo a 200 °C y se mantiene otros 10 a 15 minutos.
7. Se enfriá de nuevo sin quitar el voltaje aplicado.
8. La nueva curva registrada corresponderá al caso en que todas las cargas móviles se encuentran cercanas a la frontera dieléctrico-semiconductor, o sea, incorporando el mayor aporte de ellas al V_{FB} .

La diferencia entre los dos valores de V_{FB} obtenidos, permite calcular la densidad de carga móvil en el dieléctrico mediante la relación:

$$Q_i = \Delta V_{FB} \cdot C_i. \quad (3.31)$$

Esta técnica permite determinar la presencia de cargas móviles cuya densidad es superior a 10^{10} cm^{-2} . Generalmente para el control tecnológico esta cota es suficiente, y constituye la técnica más utilizada. Si por alguna razón se requiere determinar la densidad de cargas móviles cuyo valor es inferior a 10^{10} cm^{-2} , se utilizan otras técnicas, como por ejemplo, las curvas I-V cuasiestáticas, (basadas en la C-V de baja frecuencia) a las cuales nos referiremos posteriormente.

También efectos como el de polarización, muy frecuentemente observados en dieléctricos que contienen iones de fósforo o boro, y que producen desplazamiento en el V_{FB} , pueden ser diferenciados de los efectos producidos por las cargas móviles debido, por ejemplo, al hecho de que se observan a temperatura ambiente, y a que se observa histéresis en las curvas C-V al polarizar la estructura con voltajes positivos y negativos.

3.3 DETERMINACIÓN DE LA CONCENTRACIÓN DE IMPUREZAS EN UN SEMICONDUCTOR

Si se tiene un substrato de silicio con concentración uniforme de impurezas, una forma de determinar dicha concentración es construir una estructura MDS, cuyo espesor de óxido sea conocido. Se mide la curva C-V de alta frecuencia y utilizando la curva de la Fig. 3.4, para el caso de óxido de silicio como dieléctrico, se determina para el espesor de óxido conocido, qué concentración de substrato corresponde al valor de C_{\min} que se obtiene en la curva C-V medida. El grado de precisión en la determinación de la concentración, depende de la precisión con que se determina la curva C-V y del conjunto de curvas para las cuales se tenga el gráfico.

Es importante que la curva C-V medida no presente alteraciones precisamente en la región donde se observa la capacitancia mínima. En esta región puede afectar el hecho de que la señal de polarización varíe tan rápido que se entre en el régimen de inversión profunda o que la frecuencia de la señal de medición no garantice que se esté trabajando en alta frecuencia.

Como se señaló con anterioridad, la condición de trabajar en régimen de alta frecuencia viene dado por el hecho de que la señal de medición varíe lo suficientemente rápido para que los portadores minoritarios que se generan no puedan seguir sus variaciones. Por otro lado, la velocidad de generación-recombinación de los portadores minoritarios varía con la temperatura de la muestra y los efectos de iluminación sobre la misma. Un ejemplo de la influencia de estos dos efectos en las curvas C-V se observa en las Fig. 3.7 y Fig. 3.8 respectivamente para una misma frecuencia de medición.

Otro factor que puede afectar la medición experimental de la curva C-V es la presencia de alguna resistencia en serie con la muestra que se mide. Las curvas C-V que se obtienen pueden mostrar una C_{min}' superior a la real y por lo tanto la concentración de impurezas que se determina también estará alterada.

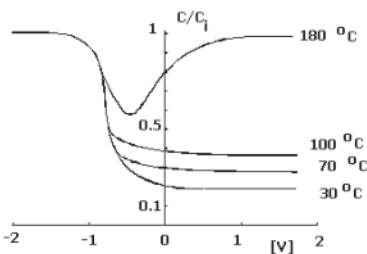


Fig. 3.7 Efecto de la temperatura sobre el comportamiento de las curvas C-V. Al aumentar la temperatura, el comportamiento característico de la curva, medida a 1 MHz empieza a parecerse a una curva medida a baja frecuencia.

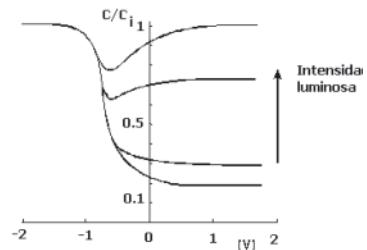


Fig. 3.8 Efecto de la iluminación sobre el comportamiento de las curvas C-V.

Cuando la estructura contiene trampas en la interfaz, el valor de C_D se puede ver afectado por efecto de éstas. Esta alteración se hace más significativa a bajas frecuencias de medición, ya que en este caso las

trampas pueden intercambiar portadores durante el tiempo en que varía la señal. Por este motivo C_D puede descomponerse en 2 términos:

$$C_D = C_{Do} + C_t(w), \quad (3.32)$$

donde $C_t(w)$ es la capacitancia debida a las trampas para bajas frecuencias de medición.

Tanto en el caso de una unión, como en el caso de una estructura MDS, la concentración de impurezas depende de la pendiente a la curva C-V, por lo tanto, cualquier efecto, como por ejemplo, la presencia de trampas en la interfaz de las estructuras MDS capaces de intercambiar portadores a la velocidad de la señal de medición, puede dar lugar al cálculo de un perfil erróneo.

La medición de la curva C-V a alta frecuencia, además de lo ya señalado con respecto al portador minoritario, permite eliminar el efecto de trampas con tiempos de relajación menores que el inverso de la frecuencia de medición.

3.4 TÉCNICAS EXPERIMENTALES DE MEDICIÓN

Las técnicas experimentales utilizadas para la medición de curvas C-V se dividen en dos grandes grupos: las que miden las curvas de alta frecuencia y las que miden las curvas de baja frecuencia.

3.4.1 Técnica de medición de curvas C-V de alta frecuencia

Esta técnica se basa en la utilización de una señal de CA de pequeña amplitud, del orden o menor de los 25 mV y de una frecuencia superior a los 500 kHz, generalmente incluso superior a 1MHz.

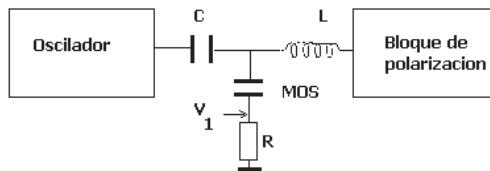


Fig. 3.9 Circuito de medición de curvas de alta frecuencia

En la Fig.3.9 se muestra un circuito de medición que se basa en la detección de la corriente que pasa por una resistencia de pequeño valor conectada en serie con la estructura. Esta corriente es proporcional a la capacitancia de la estructura y por lo tanto el voltaje que se mide en la resistencia es:

$$V_1 = R \cdot I = R \cdot C \cdot \frac{dV}{dt}, \quad (3.33)$$

de donde el valor de la capacitancia será:

$$C = \frac{V}{[R \cdot V_o \cdot \omega]}, \quad (3.34)$$

donde V_0 es la amplitud de la señal de medición y ω la frecuencia.

El bloque de polarización puede ser un generador de barrido o simplemente una fuente de CD de amplitud variable.

Tanto el bloque del oscilador como el de polarización tienen que estar separados de la muestra mediante filtros que impidan el paso de la señal de alta frecuencia hacia el bloque de polarización y de la señal de polarización hacia el oscilador.

Se mide V_1 y el voltaje de polarización V_2 aplicado a la muestra, entre los nodos A y B del circuito y utilizando (3.34) se obtiene la curva C vs. V o bien C/C_i vs. V.

Este método aunque muy sencillo introduce una resistencia en serie con la estructura a medir, que puede alterar los resultados.

Una de las técnicas más utilizadas de medición consiste en un capacitímetro con una señal de medición de 1 MHz; que permita también polarizar la muestra durante la medición y aplicar amplitudes de voltaje de alta frecuencia del orden de los 25 mV. También debe excluir los efectos de capacitancias parásitas asociadas al circuito de medición. Existen pocos fabricantes en el mundo que ofrecen equipos con estas características, entre ellos están los de las marcas Boonton, Keithley y Hewlett-Packard.

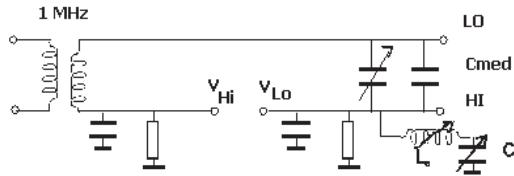


Fig. 3.10 Diagrama de conexión de la estructura a medir en el capacitímetro Boonton.

En la Fig.3.10 se muestra el diagrama de conexión de la muestra a medir en un capacitímetro Boonton. La estructura MDS se conecta entre el HI y el LO de C_{med} . Entre V_{Hi} y V_{Lo} se conecta la fuente de polarización. Es muy importante recalcar que la fuente utilizada para polarizar el dispositivo tiene que estar levantada de tierra, o sea, la salida de LO de la fuente no puede estar conectada a tierra. Si la fuente de voltaje tuviera el LO conectado a tierra y no se pudiera desconectar, es necesario trabajar con

dos fuentes, una para aplicar en el nodo V_{HI} el voltaje de polarización con respecto a tierra y otra ajustada a cero V con respecto a tierra conectada al nodo V_{LO} .

La impedancia del circuito en resonancia $Z=R\rightarrow 0$, se lleva inicialmente a cero con el condensador C1. El circuito consta de un detector de fase que mide la corriente desfasada 90 grados respecto al voltaje aplicado y que será proporcional a la capacidad entre los terminales LO y HI.

En el caso de existir una capacidad a tierra, ésta no se mide, pues la corriente está en paralelo con el circuito resonante, y $Z=R\rightarrow 0$.

La capacidad a tierra del terminal Lo, que pudiera influir en la medición se balancea inicialmente mediante el ajuste de cero del instrumento (condensador C1).

3.4.2 Técnicas de medición de curvas de baja frecuencia

Aunque existen diferentes técnicas de medición de curvas de C-V de baja frecuencia nos detendremos solamente en el método fue propuesto por Kuhn [3], quien propuso la utilización de la curva I-V cuasiestática.

En la Fig.3.11 se muestra esquemáticamente su instrumentación. El mismo se basa en que la corriente que pasa por la muestra analizada será:

$$I = C \cdot \left[\frac{dV}{dt} \right], \quad (3.35)$$

donde la señal de medición aplicada es una rampa lenta de voltaje

$$V=V_o \pm at. \quad (3.36)$$

Por lo tanto la capacitancia que se mide será:

$$C = \frac{I}{a} = \pm \frac{V_{sal}}{R \cdot a}. \quad (3.37)$$

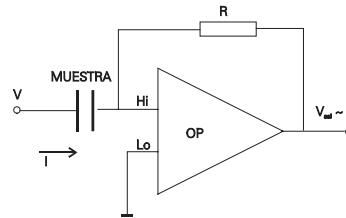


Fig. 3.11 Circuito de medición de curvas I-V cuasiestáticas.

Como la señal de voltaje varía lentamente en el tiempo, la capacitancia que se mide será la de baja frecuencia. Esta técnica aunque aparentemente sencilla, presenta como dificultad fundamental que junto con la corriente de desplazamiento (3.35) se mide cualquier corriente de fuga presente en la estructura, por lo que las condiciones de medición se vuelven muy críticas.

Normalmente se hace que la rampa tenga una variación del orden de los 10 mV/s o menos, para lograr realmente trabajar en el régimen de baja frecuencia. Esto conlleva a que las corrientes de desplazamiento deben estar en el orden de los fA (para valores típicos de capacitancia del orden de décimas de pF), y por lo tanto el efecto de las corrientes de fuga se hace sustancial. A menos que se tomen medidas muy estrictas, o se realicen mediciones para compensar la presencia de estas corrientes, las curvas con línea continua que se obtienen son como las que se muestran en la Fig. 3.12.

Es posible la corrección de dichas curvas midiendo por separado una curva I-V estática y restando la corriente de fuga medida de la total obtenida durante la medición de la curva cuasiestática. Los resultados se

mejoran, cuando ambas curvas son obtenidas al mismo tiempo como lo permiten algunos equipos Keythley.

Estas curvas también permiten la obtención de la densidad de cargas móviles existentes en el dieléctrico con valores por debajo de 10^{10} cm^{-2} . En este caso se polariza la muestra a un voltaje del orden de los -10V aplicados a la compuerta y se mantiene durante unos minutos a esa temperatura. Sin enfriar, se realiza la medición con el barrido lento y se lleva hasta un voltaje de +10 V. Se espera unos minutos y se registra la curva ahora en el sentido de +10V a -10V.

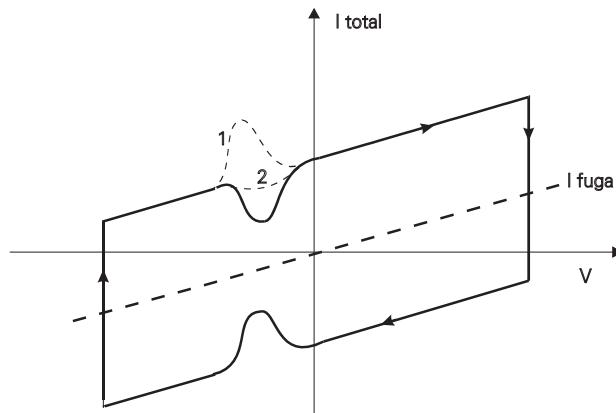


Fig. 3.12 Curva C-V cuasiestática de un MOS mostrando el efecto de la corriente de fuga. La curva (1) corresponde a una muestra con iones móviles en el óxido. La curva (2) a una sin iones móviles.

La Curva I-V que se obtiene se muestra en las curvas (1) y (2) la Fig. 3.12. Como se ve en la curva (2), el efecto de C-V de baja frecuencia ha desaparecido, pues se está midiendo a alta temperatura y como ya se vio al aumentar la temperatura aumenta la velocidad de generación-recombinación de los portadores minoritarios. Llegará el momento en que

no se observa el mínimo de la curva (ver Fig. 3.8). Si hay carga móvil, se observará un pico de corriente definido, en el sentido cuando se va de -10V hacia +10V. El área bajo la curva (1) se corresponde con la densidad de carga presente en el dieléctrico.

3.5 BIBLIOGRAFÍA

- [1]. S.M. Sze, "Physics of Semiconductor Devices", Willes-Interscience, John Wiley and Sons, Inc, (1969).
- [2]. A.F. Grove, "Physics and Technology of Semiconductor Devices", Wiley, New York, (1967).
- [3]. M. Kuhn, " A quasi-state Technique for MOS C-V and Surface Sytate Measurements", Solid State Electronics, 13 (1970) 873.

CAPÍTULO 4

DEPÓSITO DE CAPAS DIELÉCTRICAS Y SEMICONDUCTORAS: CARACTERÍSTICAS Y MÉTODOS DE DEPÓSITO.

Las capas dieléctricas en la Tecnología Planar se usan como aislantes entre capas conductoras de interconexión, como dieléctricos para formar capacitores, como material donde realizar los procesos de fotolitografía, como máscaras para difusión localizada de impurezas, como protección mecánica y contra la penetración de iones móviles, como dieléctrico para la formación de la compuerta en los transistores MOS, como pasivación de la superficie del silicio, para suavizar o planarizar el relieve que se forma sobre la oblea durante los diferentes procesos de fabricación de los dispositivos y como fuentes dopadas para introducir impurezas.

4.1 CAPAS DIELÉCTRICAS UTILIZADAS EN LA TECNOLOGÍA PLANAR

Entre las capas dieléctricas más utilizadas se encuentran el óxido de silicio, el vidrio de fosfossilicato (VFS o PSG del inglés Phosphosilicate glass), el vidrio de boro-fosfossilicato (VBFS o BPSG), el nitruro de silicio Si_3N_4 y la polimida. Esta última es una capa orgánica en forma líquida que se aplica sobre la oblea de silicio como si fuera fotorresina, obteniéndose buenos resultados cuando se deposita como aislante entre capas de interconexión.

4.1.1 Capas de óxido de silicio

El óxido de silicio depositado se comenzó a utilizar en la década del 70 para cubrir el polisilicio de la compuerta, como aislamiento entre capas metálicas y como ayuda en algunos procesos de fotolitografía. Cuando se usa como aislamiento entre líneas de metalización presenta la dificultad de reaccionar con el aluminio, por lo cual es necesario depositar capas de espesores mayores de 500 nm. Esto conlleva a que, al aumentar los espesores de las capas depositadas, aumente también la altura del relieve en la superficie de la oblea.

En todos los casos, todas las capas depositadas, unas más y otras menos según la técnica de depósito utilizada, pueden presentar problemas de cubrimiento de los escalones que conforman el relieve de la oblea. Esto puede verse en la Fig. 3.1, lo que puede dar lugar, por ejemplo, a que queden regiones desprotegidas o se interrumpan las líneas definidas en las capas depositadas.

En este último caso el SiO_2 depositado se usa como capa auxiliar para realizar el ataque con productos químicos que tienen buena selectividad para estructuras como el polisilicio-óxido de silicio o el Si_3N_4 , pero que levantan la resina durante el proceso de ataque. El fotograbado se realiza primeramente sobre la capa de óxido depositado que cubre al polisilicio o al Si_3N_4 , se elimina la fotorresina y se procede al ataque del polisilicio o del nitruro, usando la capa de SiO_2 como máscara.



Fig. 4.1 Formas en que las capas depositadas pueden cubrir los escalones previamente existentes en la superficie de la oblea.

El atacante húmedo, que se utiliza con el nitruro de silicio y que presenta buena selectividad con respecto al óxido, es el ácido ortofosfórico a 180 °C, que también levanta la resina. Debe señalarse que con el surgimiento del ataque selectivo en seco (en presencia de plasma) esta aplicación de las capas de SiO₂ sobre el nitruro de silicio y sobre el polisilicio deja de ser imprescindible.

La adherencia de las fotorresinas sobre los óxidos depositados no es tan buena como sobre los óxidos térmicos. Para que la fotolitografía quede con buena resolución es necesario, o bien someter a las estructuras a un proceso de alta temperatura del orden de los 1000 °C durante unos minutos, preferiblemente en atmósfera oxidante, o utilizar productos especiales que mejoren la adherencia de las fotorresinas sobre las capas depositadas.

El método de depósito mas utilizado para las capas de óxido de silicio, es el depósito químico a partir de la fase de vapor (CVD) a presión atmosférica (APCVD), a baja presión (LPCVD); o con ayuda de plasma (PECVD). También pueden usarse métodos físicos de depósito como la pulverización catódica reactiva. En la Tabla 4.1 se comparan algunos parámetros de estas capas depositadas por los tres métodos de CVD mencionados.

Como puede observarse en dicha Tabla, el depósito a baja presión (LPCVD) tiene una velocidad de crecimiento mucho menor que a la

presión atmosférica, por lo que se requieren tiempos relativamente largos para depositar capas del orden de los 1000 nm, pero se puede controlar fácilmente el depósito de capas de varias decenas de nm de espesor. Las capas que presentan mejores características son las depositadas por PECVD, siguiendo luego las depositadas por LPCVD y finalmente las depositadas por APCVD.

Tabla 4.1
CARACTERÍSTICAS DE ÓXIDOS DE SILICIO DEPOSITADOS
POR DIFERENTES MÉTODOS

Parámetros	CVD	LPCVD	PECVD
Permitividad	4.2	4.2	4.6
Huecos / cm ²	10	5 a 10	3
Cubrimiento de escalones	normal	bueno	excelente
Esfuerzo	tensión	tensión	compresión
Adherencia al metal	mala	buena	excelente
Índice de refracción	1.45	1.45	1.45
Velocidad de depósito [nm/min]	100	10 a 15	60

4.1.2 Capas de vidrio de fosfossilicato (VFS)

Las capas de óxido de silicio dopadas con fósforo dan lugar a la formación de un vidrio de fosfossilicato que se utiliza, fundamentalmente, como protección de la superficie de los circuitos contra la penetración de impurezas y como recubrimiento a los escalones que se producen entre diferentes capas en la superficie de la oblea, producto de los sucesivos procesos de fotolitografía. A diferencia del SiO₂ es posible formar un vidrio VFS con una concentración adecuada de fósforo para que a una

temperatura inferior a los 1000 °C sea capaz de fluir y llenar o planarizar los escalones existentes en la superficie de la oblea.

A inicio de los años 70, las capas de VFS fueron ampliamente utilizadas en los circuitos con dispositivos MOS canal P para estabilizar los parámetros de los transistores, fundamentalmente el voltaje umbral (V_T). También se usaron, aunque en menor grado, como fuentes de impurezas de fósforo.

La estructura de este vidrio de fosfossilicato (VFS) tiene la composición química $P_2O_5 \cdot SiO_2$ con una estructura que se muestra en la figura 4.2, donde cada dos átomos de fósforo comparten 5 de oxígeno. La estructura de orden cercano consta de un átomo de fósforo cargado negativamente (δ^-) y otro cargado positivamente (δ^+). El primer fósforo está unido con 4 oxígenos puente y un oxígeno intersticial. El segundo fósforo está unido con 4 oxígenos puente. Los dos átomos de fósforo cargados están situados a una distancia tal entre sí, que se comportan como un dipolo neutral.

Cada átomo de fósforo con carga negativa (δ^-) puede atrapar a un ión positivo neutralizándolo. En el dipolo queda una carga positiva fija debida al átomo de fósforo ionizado, (δ^+). Mediante este mecanismo, la capa de VFS puede **inmovilizar iones** positivos, aunque **conserva igual número de cargas positivas** fijas. Si hay iones de sodio en la capa, estos se atraparán por el fósforo, (δ^-). Si hay vacancias de oxígeno, éstas también se aniquilarán ya que el oxígeno puente del fósforo (δ^-), se dona a la capa de aluminio. Al final quedan las dos cargas positivas de los dos iones fósforo de la estructura del VFS. Una de estas cargas se compensa con la carga negativa del ión AlO_2^- , y la otra da lugar a la carga positiva resultante en la capa.

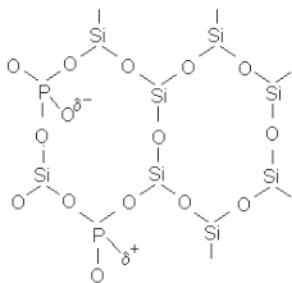


Fig. 4.2 Esquema en dos dimensiones de la estructura de una capa de VFS.

Cada átomo de fósforo debiera estar unidos a 5 oxígenos puente, pero para seguir el comportamiento del silicio, se une a 4 oxígenos puente. Uno de los átomos tiene cerca un oxígeno intersticial con un enlace no saturado, (cargado negativamente) y al otro le falta un enlace por saturar, quedando con carga positiva.

En la práctica se requiere obtener capas con concentraciones de fósforo inferiores al 7 %, ya que de lo contrario estas se hacen muy hidratables y, además, aumenta considerablemente su velocidad de ataque en la solución *buffer* de ácido fluorhídrico (BHF), resultando poco controlable el proceso de fotolitografía.

En la actualidad el método más utilizado para obtener capas de VFS es mediante el depósito químico a presión atmosférica o en presencia de plasma, pudiendo obtenerse un porcentaje de fósforo en las capas inferior al 7 % molar.

Las capas obtenidas por este método pueden ser utilizadas para suavizar el talud de los escalones en la superficie de la oblea. Para esto, luego de depositada la capa, ésta se somete a un proceso de alta temperatura para que el material tienda a fluir, ocupando los espacios vacíos en la superficie

y aplanando los escalones. La temperatura a la que se produce este efecto depende de la concentración de fósforo, resultando menor mientras mayor es la concentración de fósforo.

El porcentaje de fósforo de la capa de VFS se ajusta de manera que se pueda obtener la menor temperatura posible de fluidez, sin que se afecten demasiado las otras propiedades de la capa. Este valor se escoge generalmente alrededor del 7%.

Resumiendo las principales características de estas capas:

- 1) La cantidad de cargas móviles que puede atrapar la capa viene dada por [1] $N_e = (1.33 \times 10^{20}) X_o \text{ m\%}$, donde: m% es la fracción molar de P_2O_5 en el SiO_2 expresada en % y X_o es el espesor de la capa expresada en cm, [1].
- 2) Las capas de VFS no impiden la penetración del ión sodio, sino que solo lo inmovilizan en una zona cercana a la superficie del silicio. Al inmovilizar los iones, se genera igual número de cargas positivas fijas que las móviles atrapadas; pero como estas están localizadas en una zona cercana a la superficie del dieléctrico, producen poco incremento en el valor de V_T .
- 3) Cuando la capa de VFS se forma por difusión de fósforo en SiO_2 , la profundidad hasta la que este penetra es proporcional a la raíz del valor del tiempo de difusión, para una temperatura y concentración de la especie activa en el gas constante; y es también proporcional a la raíz del valor de la presión parcial fósforo para iguales tiempos de difusión e igual temperatura.
- 4) La velocidad de ataque de estas capas en soluciones de HF, aumenta con el porcentaje molar de P_2O_5 , en la capa, pudiendo ser muy superior a la velocidad de ataque de las capas de SiO_2 puro.

- 5) Las capas presentan efecto de polarización a temperaturas del orden de los 300 °C. Los dipolos formados por las cargas unitarias de signo contrario (δ^+) y (δ^-), en ausencia de campo eléctrico, tienen una orientación arbitraria. Por lo tanto, el momento neto de los dipolos es cero. Para temperaturas del orden o superiores a los 100 °C, en presencia de campo eléctrico, estos dipolos se orientan dando lugar a un momento dipolar resultante que depende del espesor de la capa, del potencial superficial y del porcentaje molar de fósforo, pudiendo observarse incrementos en el voltaje de banda plana del orden y superiores a los 0,25 V.
- 6) La temperatura de fluidez del VFS es inversamente proporcional a la concentración de fósforo en la capa.

4.1.3 Capas de vidrio de borofosfosilicato

Si a estas capas de VFS se le añade B_2O_5 introduciendo una especie activa de boro en la corriente de gas, durante el proceso de depósito se formarán capas de vidrio de borofosfatosilicato (VBFS). Estas capas tienen también las propiedades de inmovilizar los iones de sodio y de suavizar los escalones, luego de un proceso a alta temperatura para aumentar su fluidez. A diferencia de las capas de VFS, las capas de VBFS se hacen fluidas a temperaturas bastante por debajo de los 1000 °C, por lo que resultan muy convenientes en los circuitos en los que se requiere bajar la temperatura de procesamiento.

Las capas de VBFS tienen también como ventaja con respecto a las de VFS, que su velocidad de ataque en una solución buffer de ácido

fluorhídrico es mucho menor y del orden de la velocidad de ataque en esta solución de las capas de SiO_2 .

De esta manera se pueden obtener capas de VBFS con concentraciones de fósforo menores que el 6% en peso, cuyas temperaturas de fusión son inferiores a los 1000 °C. Los porcentajes en peso de boro utilizados son del orden de 4 a 6%.

La constante dieléctrica de las capas de VBFS es 3.8 a 1 MHz, acabadas de depositar y 4,5 luego de fundidas a 950 °C. Si la fusión se realiza a 800 °C la constante dieléctrica no varía.

La tensión mecánica de estas capas es mucho menor que la de las capas de VFS. Si se depositan por CVD a presión atmosférica se encuentran tensionadas y, luego de fundidas, pasan a ser compresibles con un esfuerzo mecánico en el orden de los $7\text{-}8 \times 10^7 \text{ N/m}^2$ [4].

4.1.4 Capas de nitruro de silicio (Si_3N_4)

El depósito de nitruro de silicio tiene su uso más generalizado en la pasivación final de los circuitos integrados, para lo cual se hace uso de sus propiedades de ser poco permeable a los iones de impurezas como el sodio y el boro, y de poseer una altísima resistencia mecánica. Como en este caso el depósito se realiza luego de depositado el aluminio, la temperatura de depósito debe ser inferior a los 500 °C, lo que se logra mediante depósito químico con ayuda de plasma o por depósito catódica reactiva con enfriamiento del cátodo.

El nitruro se usa también en la tecnología MOS para realizar las oxidaciones localizadas que permiten el aislamiento entre zonas activas, o sea, aquellas donde se realizan los dispositivos de que consta el circuito. Para esto se utiliza la propiedad de que bajo las zonas protegidas por

nitruro no ocurre oxidación del substrato. El procedimiento utilizado se muestra en la Fig. 4.3.

Se realiza una oxidación sobre toda la oblea, del orden de los 50 a 150 nm; sobre ella se deposita una capa de nitruro del orden de los 150 nm y sobre esta última se deposita una capa de óxido de unos 100 nm, la que se usará como máscara del nitruro. Mediante un proceso de fotolitografía se abren las ventanas hasta el silicio en las zonas donde se quiere realizar la oxidación localizada (Fig. 4.3 a y b). Posteriormente se somete toda la oblea al proceso de oxidación localizada (Fig. 4.3c) y se elimina la multicapa $\text{SiO}_2\text{-Si}_3\text{N}_4\text{-SiO}_2$, dejando descubierta la superficie del silicio en las áreas sobre las que posteriormente se harán los dispositivos activos del circuito (Fig. 4.3 d).

La velocidad de ataque del nitruro depende de los procesos térmicos a que ha sido sometido. En la actualidad el proceso de fotolitografía de estas capas en forma selectiva se realiza por plasma en atmósfera de freón 14 y oxígeno, no siendo necesaria la capa de SiO_2 , sobre el Si_3N_4 .

El proceso de fotolitografía se realiza atacando selectivamente el SiO_2 , y el Si_3N_4 . También se puede utilizar ataque húmedo utilizando BHF y ortofosfórico al 85% y a 180 °C, respectivamente. Este último producto debe ser secado previamente, para lo cual se hiere a 190 °C durante más de 1 hora antes de utilizarlo [5]. Se obtienen velocidades de ataque del orden de 10 nm/min para el nitruro y despreciables para el óxido.

Estos atacantes son severos con la fotorresina, como se indicó anteriormente. El nitruro puede atacarse también en HF diluido o en solución BHF; pero entonces no se obtiene un ataque selectivo con respecto al óxido.

En la zona del borde, donde termina la triple capa y comienza la ventana que deja al descubierto la superficie del silicio, durante la oxidación, se produce difusión de oxígeno a través del pequeño espesor de óxido que hay

debajo del nitruro. Por ello, en esa zona y bajo el nitruro, se producirá también una oxidación del silicio hasta una profundidad lateral B_2 , (Fig. 4.3c). La estructura de la multicapa se levantará una altura H_3 y se formará una región de transición entre el óxido grueso y las zonas activas que recibe el nombre de **pico de pájaro** por la forma que se observa en la Fig. 4.3d.

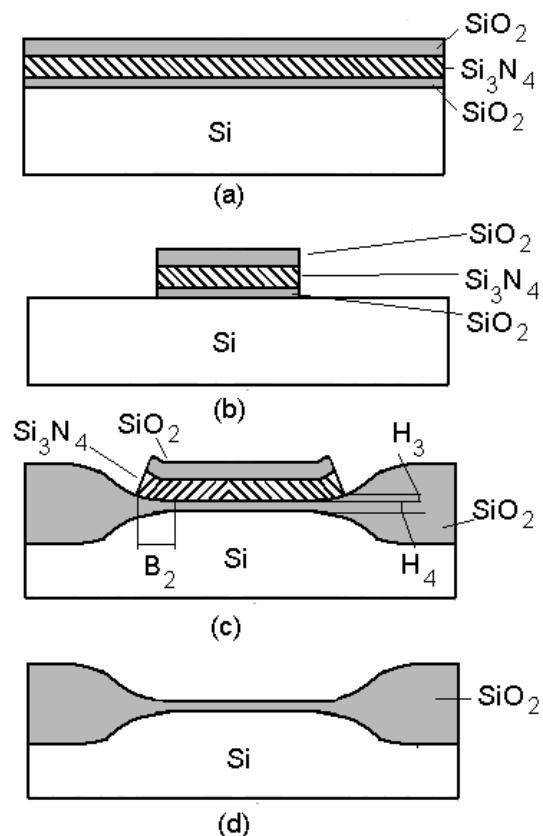


Fig. 4.3 Corte transversal de una oxidación localizada mostrando el "pico de pájaro"

Para el cálculo de los parámetros B_2 y H_3 que se utilizan en los programas de simulación tecnológica se hace uso de expresiones empíricas, como por ejemplo las que reportan Lee y Dutton [6].

Conocer y evaluar el fenómeno del "pico de pájaro" resulta importante, pues su presencia disminuye el área activa de los circuitos integrados.

4.2 DEPÓSITO QUÍMICO A PARTIR DE LA FASE, DE VAPOR (CVD)

El depósito químico a partir de la fase vapor consiste en la formación de capas sólidas estables, gracias a una reacción heterogénea del tipo gas-sólido o gas-líquido en la superficie del substrato. Las reacciones químicas pueden ocurrir de varias formas:

Pirólisis: El substrato se calienta hasta una temperatura lo suficientemente alta para provocar la descomposición de los compuestos gaseosos. Sobre el substrato se depositan los compuestos no volátiles. Los compuestos químicos que se usan para depositar por este mecanismo son generalmente halogenuros metálicos o hidruros metálicos, como por ejemplo el silano (SiH_4).

Reducción del hidrógeno: La reacción tiene lugar a alta temperatura por reducción en presencia de hidrógeno de un gas. Los compuestos más frecuentemente, utilizados son los halogenuros como Cl_4Si y Cl_2SiH_2 .

Para ambos tipos de reacción, la forma de suministrar la energía incluye calentamiento y plasma.

4.2.1 Cinética de crecimiento

Al igual que en otros procesos analizados, los fenómenos físicos que determinan el régimen de depósito son:

- 1) El transporte de los gases reaccionantes hacia la superficie del substrato calentado;
- 2) La velocidad de reacción de estos gases en la superficie del substrato; la concentración de las especies reaccionantes.

En la figura 4.4 se muestra el proceso de depósito donde C_G y C_S son las concentraciones del gas reaccionante en el volumen del gas portador y en la superficie de la oblea, respectivamente. Como en este caso la reacción ocurre en la superficie del óxido, el proceso de crecimiento se define solamente por dos flujos:

El de transporte de masa

$$F_1 = h_G \cdot (C_G - C_S) \quad (4.1)$$

y el que se consume durante la reacción

$$F_2 = k_S \cdot C_i. \quad (4.2)$$

La velocidad de crecimiento de la capa se puede expresar como

$$\frac{dX_o}{dt} = \frac{k_S C_i}{N_1}. \quad (4.3)$$

donde N_1 - cantidad de moléculas por centímetro cúbico en el óxido de silicio.

Considerando que durante el crecimiento se cumplen la ley de Henry y la ley de los gases ideales:

$$C_i = H \cdot P_s \quad (4.4)$$

$$C_s = \frac{P_s}{kT} \quad (4.5)$$

$$C_G = H \cdot P_G \quad (4.6)$$

$$C^* = \frac{P_G}{kT} \quad (4.7)$$

Igualando los flujos y según (4.5) y (4.7)

$$h(C^* - C_i) = k_s C_i \quad (4.8)$$

donde.

$$h = \frac{h_G}{HkT} \quad (4.9)$$

$$C_i = \frac{h \cdot C^*}{(k_s + h)} \quad (4.10)$$

y

$$\frac{dX_o}{dt} = \frac{1}{N_1} \cdot \frac{k_s C^*}{\left(1 + \frac{k_s}{h}\right)} \quad (4.11)$$

Se observan dos casos límite:

1) Si $h \gg k_s$

$$\frac{dX_o}{dt} = \frac{1}{N_1} \cdot k_s C^* \quad (4.12)$$

y tiene lugar el control por reacción. A baja temperatura es de esperar que la velocidad de crecimiento varíe aproximadamente de forma exponencial con la temperatura y lineal con la presión parcial de la especie activa.

2) Si por el contrario $h \ll k_s$

$$\frac{dX_o}{dt} = \frac{1}{N_1} \cdot h \cdot C^* \quad (4.13)$$

y tiene lugar el control por transporte de masa. Esto da lugar a que para altas temperaturas, la velocidad de crecimiento de la capa no depende prácticamente de la temperatura y solo se observa dependencia con la concentración de la especie activa en el gas.

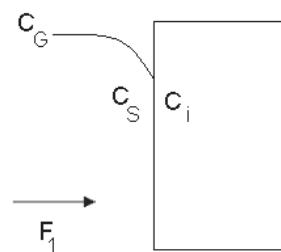


Fig. 4.4 Modelo para el proceso de depósito química de capas a partir de la fase vapor

El proceso de reacción depende de la concentración de los gases reaccionantes de la temperatura de la superficie de la oblea y de la presión a la que ocurre la reacción,

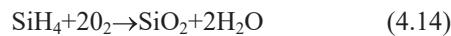
Cuando el depósito a partir de la fase vapor se realiza a presiones entre 0,2 Torr (26,7 Pa) y 1 Torr (1.33 Pa), la difusividad de los gases inyectados se incrementa, por lo que el coeficiente de transferencia de masa de los gases reaccionantes y de los productos residuales se incrementa con respecto a su valor para el régimen a presión atmosférica. Esto hace que los

procesos a baja presión sean controlados fundamentalmente por reacción, la dependencia térmica se acentúa, disminuye la velocidad de depósito y hay un mayor control del proceso de crecimiento para capas de espesores pequeños.

4.3 DEPÓSITO DE SiO₂

4.3.1 Depósito de SiO₂ por APCVD

El depósito de SiO₂ a presión atmosférica se realiza a partir de la reacción química:



donde el SiH₄ se arrastra mezclado en una corriente de gas inerte. La velocidad de crecimiento depende de:

- a) la temperatura de depósito;
- b) la relación O₂/SiH₄
- c) la geometría del reactor.

La reacción (4.14) comienza a ocurrir temperaturas del orden de los 240 °C. Para temperaturas entre 240 °C y 310 °C la velocidad de crecimiento aumenta rápidamente con la temperatura, mientras que a temperaturas entre 310 y 450 °C el incremento de la velocidad con respecto a la temperatura es menor (Fig. 4.5).

Si la relación entre los gastos de oxígeno y silano (O₂/SiH₄) se mantiene constante, aumentar el gasto de silano hace que se incremente la velocidad de depósito linealmente (Fig. 4.6) lo que era de esperar según (4.11) ya que C* depende de la cantidad de SiH₄ que entra a la cámara.

Manteniendo constante la temperatura de depósito, la velocidad de crecimiento depende de la relación de los flujos de O_2/SiH_4 , presentando un máximo cuyo valor se incrementa con la temperatura.

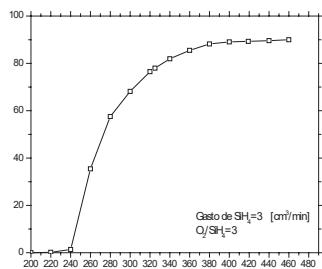


Fig. 4.5 Dependencia de la velocidad de depósito con la temperatura para una relación O_2/SiH_4 constante.

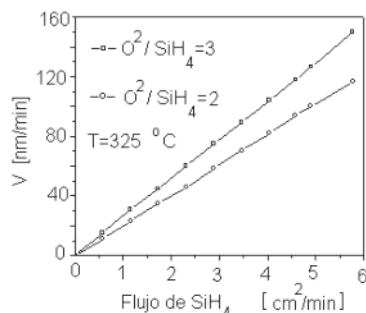


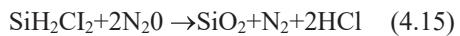
Fig. 4.6 Velocidad de crecimiento de SiO_2 depositado por CVD en función del gasto de SiH_4 para dos valores de O_2/SiH_4 a temperatura constante.

También, al aumentar la temperatura, el máximo se desplaza hacia valores mayores de la relación O_2/SiH_4 , a la vez que se hace más ancho. Por ejemplo, a 475 °C la velocidad de crecimiento máxima se observa para una relación entre O_2/SiH_4 de 14, pero el máximo es tan ancho que incluso hasta para una relación de O_2/SiH_4 de 33 se puede considerar que la

velocidad de crecimiento prácticamente se mantiene constante. Para relaciones mayores de 33 se observa una disminución de la velocidad de crecimiento que se atribuye a un proceso de inhibición de la reacción por exceso de oxígeno, el que en este caso, se puede adsorber en la superficie de la capa que se va depositando. Mientras no se observa este fenómeno, la reacción que tiene lugar puede considerarse de primer grado [7].

4.3.2 Depósito de SiO₂ por LPCVD

El depósito de SiO₂ a baja presión (LPCVD) se realiza a presiones del rango de 26,66 a 133,3 Pa (0,2 a 1 Torr), a partir de la reacción



En ella no se produce hidrógeno y se reporta [9] que puede lograrse una uniformidad en el espesor de las capas del orden de un 2% dentro de la misma oblea y para una temperatura de depósito de 925 °C en el centro del tubo, 930 °C en la entrada y 910 °C en la salida; una presión de 0,4 Torr (53,32 Pa); 56 cm³/min de SiH₂Cl₂ y 113 cm³/min de N₂O.

El óxido de silicio también puede obtenerse a baja presión a partir de la reacción (4.14), variando la temperatura a lo largo del horno entre 400 °C y 450 °C; pero se obtienen diferencias de espesores hasta de un 20% dentro de una misma oblea. Algunas correcciones adicionales a la geometría de los botes y formas de introducir los gases [10] han permitido reducir esta falta de uniformidad hasta un 5 %, aunque siempre resulta mayor que la que se obtiene a partir de la reacción (4.15).

4.4 DEPÓSITO DE CAPAS DE VIDRIO DE FOSFOSILICATO (VFS)

El depósito de capas de vidrio de fosfositato (VFS) por CVD a presión normal, se realiza a partir de la reacción química



Generalmente los hidruros están disueltos en nitrógeno o argón. La velocidad de crecimiento de la capa depende de:

- a) la temperatura de depósito;
- b) la relación de los gastos de oxígeno con respecto a los hidruros $\text{O}_2 / (\text{SiH}_4 + \text{PH}_3)$
- c) la relación del gasto de fosfamina con respecto al gasto de hidruros;
- d) la geometría del reactor.

La concentración de fósforo en la capa dependerá a su vez de:

- a) la temperatura de depósito;
- b) la relación del gasto de oxígeno con respecto al gasto de hidruros;
- c) la relación del gasto de fosfamina con respecto al gasto de hidruros.

En la figura 4.7 se muestra como ejemplo, un gráfico que muestra el comportamiento de la velocidad de crecimiento, al variar la relación de los gastos oxígeno/hidruros para una temperatura dada.

A medida que la temperatura aumenta, el máximo de velocidad de crecimiento aumenta en su valor y su posición se desplaza hacia mayores valores de la relación entre los gastos de oxígeno/hidruros (Tabla 4.2); mientras el ancho del máximo el cual la velocidad de crecimiento se mantiene casi constante) se hace mayor, de forma similar a lo descrito para

las capas de óxido. Este comportamiento es similar al observado para el depósito de SiO_2 , aunque los valores obtenidos dependen del equipo y otras condiciones de depósito.

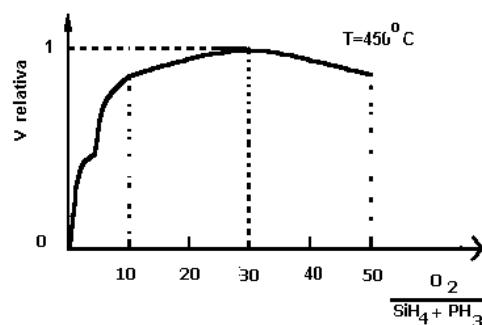


Fig. 4.7 Variación de la velocidad de depósito de una capa de VFS en función de la relación $\text{O}_2/(\text{SiH}_4+\text{PH}_3)$ para una temperatura, gasto total de SiH_4 y relación de SiH_4/PH_3 constantes.

En la figura 4.8 se muestra la dependencia de la concentración de fósforo en la capa de VFS con respecto a la relación de los gastos de oxígeno /hidruros (SiH_4/PH_3) para la misma temperatura y relación de SiH_4/PH_3 . La concentración de fósforo aumenta con la relación entre los gastos de oxígeno y de hidruros $\text{O}_2/(\text{SiH}_4+\text{PH}_3)$ y es inversamente proporcional a la temperatura. Al aumentar la temperatura, la dependencia con la relación oxígeno/hidruros se debilita. Por ejemplo, a 450 °C, el % molar de fósforo en la capa varía entre el 2% y el 6%; mientras que para 350 °C lo hace entre 2% y el 10%. Desde este punto de vista resulta conveniente depositar a temperaturas del orden de los 450 °C.

El fenómeno descrito en el párrafo anterior se explica porque la fosfamina reacciona a temperaturas menores que el silicio, por lo que a bajas temperaturas ya puede haber reaccionado completamente, mientras el

silano se encuentra aún en etapa de incrementar su velocidad de reacción con el oxígeno. De esta forma, aumenta la relación fosfamina/silano que ha reaccionado.

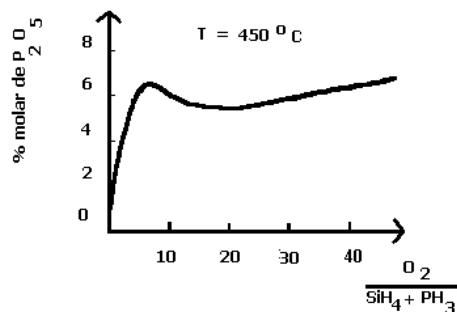


Fig. 4.8 Variación del % de fósforo y del % molar de fósforo de una capa de VFS en función de la relación $O_2/(SiH_4+PH_3)$ para una temperatura de 450 °C, y una relación de $SiH_4/PH_3=20$.

La composición de fósforo de la capa de VFS depende también de la relación fosfamina/silano. Si se mantienen constantes el resto de los parámetros, se observa una dependencia lineal entre el porcentaje molar de P_2O_5 , en SiO_2 y la fracción molar SiH_4/PH_3 hasta alcanzar un 8% molar de P_2O_5 . En la figura 4.9 se muestra esta dependencia para datos calculados y experimentales.

Para el cálculo de la curva estequiométrica se parte del porcentaje molar de PH_3 en SiH_4+PH_3 se calcula el peso total del fósforo disponible y, por lo tanto, el posible porcentaje molar de P_2O_5 en la capa, si las relaciones molares de SiH_4/PH_3 y $O_2/(SiH_4+PH_3)$ garantizan un peso de silicio y de oxígeno suficientes para que reaccione todo el fósforo presente

Tabla 4.2
 RELACIÓN DE GASTOS Y TEMPERATURAS PARA LOS
 CUALES SE PRODUCE EL MÁXIMO EN LA VELOCIDAD DE
 CRECIMIENTO

Relación de gastos $O_2/(SiH_4+PH_3)$	T [°C]
5	300
5-10	350
18-22	400
20-28	450

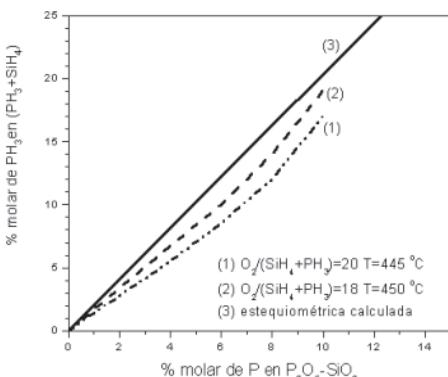


Fig. 4.9 Variación de la concentración de fósforo en la capa de VFS respecto al cálculo estequiométrico para 2 valores de la relación PH₃/SiH₄, en capas depositadas a 450 °C.

El porcentaje molar de P₂O₅ puede expresarse también como porcentaje en peso de fósforo en la capa, si se calcula el peso de un mol de P₂O₅·SiO₂, para un % molar de P₂O₅, dado.

Por ejemplo, para un 10 % de P₂O₅ en la capa, el peso de fósforo se calcula de la manera siguiente:

El peso de un mol de $P_2O_5 \cdot SiO_2$, es 10% de 142+90% de 60 lo que es igual a 68,2g. El peso molecular de P_2O_5 es de 142 g y el de SiO_2 es de 60 g. Del peso total de 69,2 g, habrá un peso de fósforo igual a 2 (P.A. del fósforo) (% molar de P_2O_5), o sea, $2(30,9) / (0,1) = 6,18$ g. De esta manera el porcentaje de fósforo en la capa será de $6,18/68,2 = 9\%$.

En la Tabla 4.3 se muestra la conversión ce algunos valores de % molar de P_2O_5 expresados en % en peso de fósforo en la capa,

Tabla 4.3
CONVERSIÓN DE PORCENTAJE MOLAR DE P_2O_5 A

% molar de P_2O_5	% en peso de P
2	2
4	3.9
5	4.8
6	5.7
7	7.6
8	7.4
10	9

Otro de los parámetros que se hace necesario controlar en las capas de VFS es la tensión mecánica, encontrándose que esta depende fuertemente de la relación oxígeno/hidruros. Para espesores de capas entre 1500 nm y 1900 nm, depositadas a 400 °C y para una relación de $O_2/(SiH_4 + PH_3)$ de 10, no se observaban rajaduras por tensión [7]. Si se incrementa esta relación, aparecen rajaduras, y ya para una relación de 40 se observa una alta densidad de ellas. Si se incrementa la temperatura de depósito hasta 450 °C, para relaciones oxígeno/hidruros entre 3 y 36 y concentraciones de fósforo del orden de 10 %, no se manifiestan tensiones mecánicas ni

rajaduras. La tensión de la capa depende también de la concentración de fósforo.

En principio, las capas de VFS pueden depositarse a baja presión de forma muy similar a la utilizada para el SiO₂ sin embargo, por causas semejantes también se produce la mayor uniformidad cuando el depósito se hace a partir de diclorosilano a temperaturas del orden de los 900 °C, por lo que no puede usarse como pasivación final; además, la velocidad de depósito resulta bastante más pequeña, por lo que se requiere tiempos demasiado largos para crecer los espesores acostumbrados. En general, este método de depósito no se encuentra generalizado para estas capas.

4.5 DEPÓSITO DE CAPAS DE VIDRIO DE BOROFOSFOSILICATO (VBFS)

4.5.1 Depósito de capas de vidrio de borofosfосilicato por CVD

Las capas de vidrio de borofosfосilicato se depositan por CVD a presión atmosférica [12] a partir de la simultánea oxidación de los hidruros SiH₄, PH₃ y BH₃ con el oxígeno, en un gas de arrastre constituido por nitrógeno.

La velocidad de crecimiento de las capas depende de:

- a) la temperatura de depósito;
- b) la relación de los gastos de oxígeno con respecto a los hidruros.

Para temperaturas entre 410 y 450 °C y relaciones oxígeno hidruros de 20 a 1 se obtienen valores de velocidad de crecimiento del orden de los 100 nm/min.

Las capas crecidas por CVD a presión atmosférica presentan variaciones en su espesor inferiores al 1,5%.

4.5.2 Depósito de capas de vidrio de borofosfosilicato por LPCVD

El depósito de capas de vidrio de borofosfosilicato por CVD a baja presión [12] se realiza a partir del SiH₄ PH₃ y CIB₃, con oxígeno y en ausencia de gas portador. Las relaciones de oxígeno/hidruro son solo del orden de 1,5; pero se requieren altas concentraciones de hidruros, por ejemplo 30 % de BH₃ y 15 % de PH₃ en SiH₄ puro. Las temperaturas de depósito son del orden de los 400 °C.

Para las cámaras verticales con flujo vertical de gases puede utilizarse el diborano (BH₃). Sin embargo, para los tubos horizontales de paredes calientes el diborano produce una descomposición pirolítica en los inyectores [13], por lo que se recomienda sustituirlo por el CIB₃. Las velocidades típicas de crecimiento están en el orden de 1,9 a 2,5 nm/min para presiones entre 23 y 44 Pa.

4.5.3 Depósito de capas de vidrio de borofosfosilicato por PECVD

El depósito de capas de vidrio de borofosfosilicato por LPCVD [15] se realiza tanto en sistemas horizontales con pared caliente como en sistemas con electrodos paralelos y de pared fría. En el primer caso es necesario utilizar un gradiente de temperatura a lo largo del horno, para garantizar una uniformidad en los espesores del 2%. La presión de trabajo es del orden de 313 Pa.

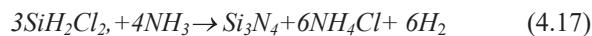
Las ventajas del método de PECVD son: mayor velocidad de depósito que el LPCVD y menores temperaturas de depósito, así como menores

concentraciones de hidruros. Entre las desventajas se encuentra la inclusión de gases, como H₂, H₂O, PH₃ y N₂ en las capas crecidas.

4.6 DEPÓSITO DE CAPAS DE NITRURO DE SILICIO (Si₃N₄)

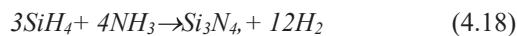
4.6.1 Depósito de nitruro de silicio por LPCVD

Las capas de Si₃N₄ se depositan por LPCVD a partir de la reacción.



a una temperatura del orden de los 750 °C con una uniformidad en los espesores entre el centro y borde de una oblea de 100 mm en el orden de 3,2 % y de 1,5 % entre una oblea y otra.

Por motivos análogos a los ya señalados durante el estudio de otras capas, las capas de nitruro de silicio que se obtienen a partir de la reacción



A la temperatura del orden de los 800 °C, presentan una mayor no uniformidad en el espesor de las capas del orden de un 20% en una misma oblea. Con algunas modificaciones en la configuración de los botes este valor puede reducirse a menos del 5%.

La velocidad de crecimiento de las capas depende de la temperatura y de la presión parcial de diclorosilano; así como de la relación molar entre el NH₃ y el SiH₂Cl₂ y de la presión a que se realiza el proceso. Esta velocidad aumenta a medida que decrece la relación molar, lo que se explica porque

la presencia del NH_3 altera la disociación térmica del diclorosilano, que es el proceso predominante temperaturas del orden de los 800 °C. La uniformidad de la capa en la oblea mejora ligeramente.

Para compensar el empobrecimiento del gas a lo largo del horno se utiliza la presencia de un gradiente de temperatura. El perfil de temperatura que se requiere para mejorar la uniformidad de la capa, depende de la presión de depósito y del porcentaje molar de NH_3 en SiH_2Cl_2 . En la Fig. 4.10 se muestra un perfil de temperatura típico para compensar el empobrecimiento del gas a lo largo del horno de LPCVD, para dos valores de presión de trabajo. Como puede observarse el efecto de falta de uniformidad en la velocidad de depósito se incrementa al aumentar la presión promedio en la cámara.

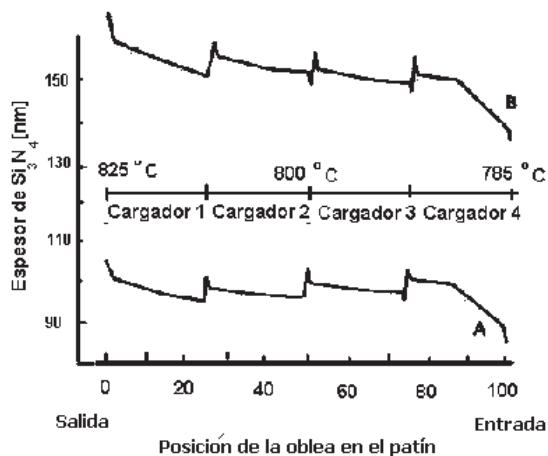


Fig. 4.10 Variación del espesor de la capa con la posición de la oblea en el horno para $\text{NH}_3/\text{SiH}_2\text{Cl}_2=3.4$ y diferentes presiones de trabajo. Curva A) $P=34.6 \text{ Pa}$ (0.26 Torr); Curva B) $P=49.3 \text{ Pa}$ (0.37 Torr).

4.6.2 Depósito de nitruro de silicio por CVD con ayuda de plasma (PECVD).

Como se señaló, la temperatura de depósito del nitruro por LPCVD es del orden de los 800 °C, por lo que estas capas no pueden usarse para la pasivación final, ya que en este caso deben depositarse sobre el aluminio que funde a 550 °C. El nitruro para esta aplicación puede depositarse por PECVD, es decir, depósito química a partir de fase vapor, con ayuda de plasma. En la Tabla 4.4 se muestran, algunos parámetros de capas de nitruro crecidas por LPCVD y PECVD.

Durante el crecimiento de Si_3N_4 por PECVD tiene lugar la reacción siguiente:



donde el NH_3 puede o no estar presente. Cuando lo está, el esfuerzo mecánico de la capa es tensionado y las capas contienen menor concentración de hidrógeno; además, su presencia permite obtener índices de refracción entre 2 y 2,1, ya que como la molécula de NH , tiene menor energía de disociación que la de nitrógeno, las capas no están enriquecidas en silicio, no tienen alterado el índice de refracción ni la velocidad de ataque.

La velocidad de depósito de la capa depende:

- a) del gasto de silano;
- b) de la relación nitrógeno/amoníaco;
- c) de la geometría del reactor;
- d) de la potencia de radiofrecuencia para crear el plasma.

La uniformidad de la capa que se crece depende también de la geometría del reactor, del gasto de silano y de la relación N₂/NH₃ así como de la temperatura y de la potencia de radiofrecuencia.

Tabla 4.4
COMPARACIÓN DE PARÁMETROS DE CAPAS DE NITRURO DE
SILICIO CRECIDAS POR PECVD Y LPCVD

Parámetros	PECVD	LPCVD
Gases reaccionantes	SiH ₄ NH ₃	SiH ₂ Cl ₂ ; NH ₃
Temperatura, °C	380	850
Presión, Torr	2	0,5
Pa	267	67
Velocidad de depósito, nm/min	30	3,5
% atómico de hidrógeno en la capa	17	6
Esfuerzo	compresión	tensión
Recubrimiento de escalones	bueno	bueno
índice de refracción	2	2
Relación Si/N ₂	0,78	0,75

4.6.1 Depósito de nitruro de silicio por pulverización catódica

El nitruro de silicio puede depositarse también mediante pulverización catódica reactiva, en presencia de nitrógeno. Se han reportado capas crecidas en plasma de nitrógeno a una presión de 10⁻² Torr (2,27 Pa) y 8 W /cm² de potencia en el blanco [16]. La velocidad de depósito es de 15

nm/min, por lo que resulta un poco alta para obtener capas del orden de los 30 nm de forma repetible. Esta velocidad puede ajustarse variando la presión de la cámara y la potencia del blanco. La constante dieléctrica, la resistividad y el espectro de absorción de las capas de nitruro de silicio obtenidas por pulverización catódica reactiva son semejantes a los de capas obtenidas por PECVD y LPCVD.

4.7 CAPAS DE POLISILICIO. CARACTERÍSTICAS Y APLICACIONES

Uno de los materiales depositados de más importancia en la fabricación de dispositivos semiconductores es el silicio. Según las características del proceso utilizado y del substrato sobre el que se deposita, la capa crecida puede ser epitaxial o policristalina.

Estas capas se usan como nivel de interconexión, como una de las placas de un condensador y en la tecnología MOS se usan ampliamente como material para la compuerta de los transistores. Pueden depositarse por CVD, LPCVD y por pulverización catódica.

Según la temperatura a la que se depositen, las capas pueden ser amorfas, si la temperatura de depósito es inferior a los 600 °C, o policristalinas, si se depositan a temperaturas superiores. Como estas capas son sometidas después de su depósito a temperaturas del orden de los 1000 °C en los tratamientos térmicos necesarios para el procesamiento de los dispositivos, su estructura final es policristalina con un tamaño de grano que depende de la temperatura y del tiempo de los procesos térmicos. Se recomienda depositar las capas a temperaturas superiores a los 600 °C para obtener una estructura policristalina desde que se crecen, de lo contrario, al pasar la capa de amorfa a policristalina, pueden formarse cristales

demasiado grandes que hacen la capa porosa e irregular. Un análisis del mecanismo de crecimiento de los granos fue realizado en [17-18].

Las capas de polisilicio pueden crecerse intrínsecas o dopadas. En el primer caso se dopan en un proceso de difusión o implantación posterior a su depósito. La resistividad depende de la concentración de impurezas y del tamaño y número de los granos; por este último motivo, la resistividad de las capas depende del régimen de depósito (Fig. 4.15), así como de los procesos de TTAT que sufre la oblea luego de depositan las capas. Cuando las capas se depositan intrínsecas, para obtener una resistividad inferior a los $80\Omega/\square$ es necesario que durante el proceso de difusión o con posterioridad a este, la oblea sea sometida a un proceso térmico a temperaturas del orden de los 1000 °C durante algunos minutos.

Las capas de polisilicio se atacan en soluciones de nítrico con fluorhídrico, como el CP-8 que consta de 1 parte de HF al 50%, 50 partes de NO_3H al 70% y 40 partes de H_2O cuya velocidad de ataque es unos 30 nm/s para capas tipo n, y algo menor para las de tipo p. Estos productos para la fotolitografía levantan la fotorresina durante el ataque, por lo que las capas se recubren generalmente de óxido de silicio que se usa como máscara al igual que en caso ya visto del nitruro de silicio. El método de ataque en seco es el más utilizado actualmente.

El espesor de las capas de polisilicio es generalmente inferior a 300 nm. En el caso en que el polisilicio se use como interconexión, el parámetro más importante a controlar es su resistividad por cuadrado, ya que para los valores que generalmente se obtienen entre 25 y $80\Omega/\square$, la caída de voltaje en líneas largas no siempre es despreciable. Además, si por estas interconexiones se propagan señales de alta frecuencia, hay que representarlas como líneas de transmisión con parámetros RC concentrados.

Cuando el polisilicio se utiliza como compuerta de transistores MOS, debe tomarse en cuenta el valor de su diferencia de potencial de contacto con el substrato. Esta diferencia de potencial de contacto depende del tipo y concentración del dopaje de la compuerta y del substrato.

La propiedad del VFS y del BVFS de planarizar los escalones es ampliamente utilizada luego de depositar el polisilicio de la compuerta. Un problema importante desde el punto de vista tecnológico resulta el poder evitar que las líneas de metalización que cruzan por sobre las líneas de polisilicio se abran producto de la presencia de taludes abruptos en los bordes de polisilicio. Esta dificultad, conocida como el problema de los escalones tiene diversas soluciones, entre las que se encuentra el recubrimiento de las estructuras con una o dos capas de VFS. En la Fig. 4.11 se muestra el efecto de planarización utilizando el depósito de una capa de VFS del 7 % que se somete a un tratamiento térmico alrededor de 100 °C por 20 minutos para que fluya y planarice el relieve de la oblea. La Fig. 4.11 a muestra un talud abrupto que es eficientemente suavizado luego del proceso de planarización. Finalmente la estructura queda como se muestra en la Fig. 11 b.

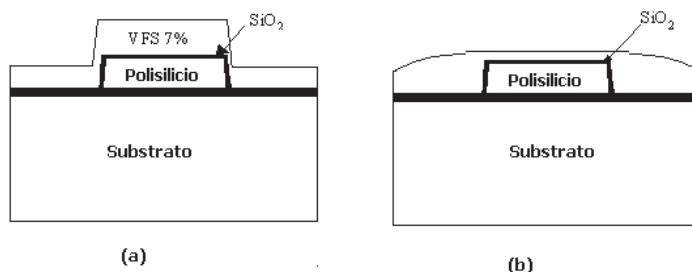


Fig. 4.11 Corte transversal de una estructura de polisilicio cubierta de óxido depositado sobre el óxido fino de compuerta, recubierta por una capa de VFS al 7 %: (a) antes y (b) después del tratamiento térmico para planarizarla.

El uso de estas capas introduce, sin embargo, un nuevo problema durante el fotograbado para hacer los huecos de contacto entre el aluminio de interconexión y las zonas de difusión cuando se usan ataques húmedos. Este surge, como ya se señaló, debido a que las capas de VFS tienen una velocidad de ataque en BHF o en soluciones diluidas de FH, mucho mayor que la del óxido de silicio. Como el VFS se deposita sobre toda la oblea, en los lugares en que se van a abrir los huecos de contacto ya hay una fina capa de óxido de silicio, por ejemplo, del orden de 100 nm.

Cuando se realiza el fotograbado, si no se tornan medidas especiales, durante el tiempo en que se ataca la capa de óxido, se produce un ataque lateral del motivo en la capa de VFS que distorsiona por completo la geometría del hueco que se desea obtener. Para evitar esto se usan diversas medidas: una de ellas implica tener dos máscaras para la fotolitografía del hueco de contacto: la primera, de dimensiones mayores que la que se desea obtener, abre el hueco en la capa de VFS, como se muestra en la Fig. 4.12a, descubriendo la capa de óxido de silicio. El tiempo de ataque será el correspondiente al ataque de la capa de VFS, que es del orden de un micrómetro.

Luego se realiza un nuevo proceso de fotolitografía con una máscara con huecos de menores dimensiones que, por lo tanto, caigan dentro de la zona de óxido que se dejó descubierta con el primer fotograbado (Fig. 4.12 b). De esta manera en este segundo fotograbado solo se ataca el óxido de silicio y no se produce el comido lateral del VFS.

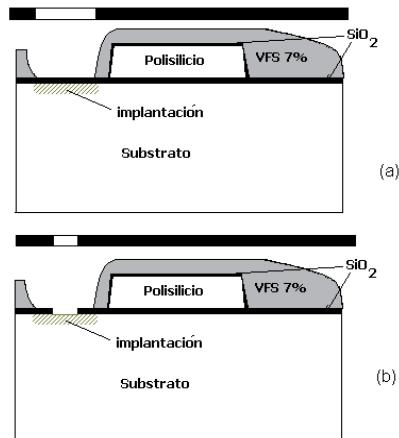


Fig. 4.12 Apertura de un hueco de contacto en una región implantada, mediante el uso de 2 máscaras con diferentes dimensiones de hueco.

(a) fotolitografía con una máscara con una hueco

El segundo método utiliza dos veces una máscara con las dimensiones correspondientes al tamaño del hueco que se desea abrir. En la primera fotolitografía se abre un hueco hasta el silicio que como es de esperar define una figura en la capa de VFS con dimensiones bastante mayores a la que se tiene en la máscara, (Fig. 4.13 a). Seguidamente se deposita una capa de VFS al 4%, la cual cubre toda la oblea y por supuesto también la región sobre el silicio implantado donde se realizó la fotolitografía y se quitó el VFS y el SiO_2 . Seguidamente se realiza de nuevo una fotolitografía con la misma máscara. Como ahora solo se tiene una capa cubriendo la región implantada, que en este caso es un VFS al 4 %, se puede definir bien la dimensión del hueco deseado, Fig. 4.13 b. El hecho que la velocidad de ataque de la capa de VFS al 4 % sea menor que la del 7% también ayuda a obtener un buen control del tiempo de ataque.

El uso del VBFS elimina en parte este problema, ya que puede lograrse la planarización a menor temperatura y con un porcentaje de fósforo menor, por lo que la diferencia en la velocidad de ataque de la capa de VBFS y la de SiO_2 no es tan significativa. Finalmente, la utilización de ataques en seco anisotrópicos, reducción de dimensiones y nuevas y mas

complejas estructuras utilizadas actualmente en la tecnología planar se han desarrollado otras muchas formas de abrir los huecos de contacto.

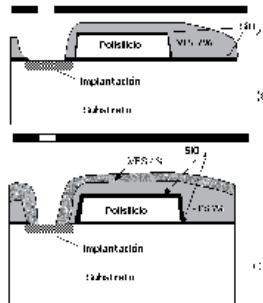


Fig. 4.13 Fotolitografía de hueco de contacto utilizando una máscara y dos procesos fotolitográficos

4.7.1 Depósito de polisilicio por LPCVD

El polisilicio se deposita por LPCVD a partir de la descomposición pirolítica del SiH₄ o la reacción del diclorosilano (SiCl₂H), (ver Tabla 4.2), aunque la mayor uniformidad en el espesor de las capas se logra mediante el diclorosilano. La temperatura de depósito debe ser superior a los 625 °C. El método de LPCVD permite obtener una velocidad de depósito suficientemente baja (ver Tabla 4.2) para poder controlar sin dificultad el depósito de capas del orden de los 100 a 300 nm.

Si la temperatura de depósito es inferior, las capas se depositan amorfas. Por eso motivo las técnicas de APCVD y PECVD no se utilizan en este caso.

Las capas de polisilicio se pueden depositar dopadas con fósforo o con boro. Para ello, se añade a la cámara donde se depositan una entrada con fosfeno (PH₃) o díborano (B₂H₆) respectivamente, diluido en nitrógeno, argón o hidrógeno.

El polisilicio puede depositarse también mediante pulverización catódica utilizando un cátodo de silicio de alta pureza. Las capas quedan de buena calidad y resulta fácil obtener control y repetibilidad en los espesores, aunque desde el punto de vista industrial se dificulta su aplicación porque requiere hacer cátodos de material de alta pureza y de diámetros lo suficientemente grandes para permitir uniformidad en la capa depositada en toda la superficie de la oblea. Las ventajas en cuanto al ahorro de materiales (gases reactivos) son considerables.

4.8 CARACTERÍSTICAS DEL EQUIPAMIENTO UTILIZADO EN EL DEPÓSITO QUÍMICO A PARTIR DE LA FASE VAPOR (CVD)

El equipamiento utilizado para depositar capas dieléctricas y semiconductoras por CVD se puede clasificar según la temperatura y presión utilizadas, y según el tipo de calentamiento; tal como se muestra en la Tabla 4.7.

Tabla 4.5
CLASIFICACIÓN DE LOS EQUIPOS DE DEPÓSITO

T [°C]	Presión atmosférica		Baja presión		Baja presión con plasma	
	Pared caliente	Pared fría	Pared caliente	Pared fría	Pared caliente	Pared fría
200-500	0	1	1	0	0	2
500-1000	0	1	1	0	0	0

0- No se usa comercialmente; 1-Uso limitado; 2-Uso dominante.

Como ya se señaló, la geometría del equipo de depósito influye fuertemente en los parámetros de las capas depositadas. Los primeros

equipos comerciales que se utilizaron fueron verticales y de pared fría. Se buscaba que cargaran el mayor número posible de obleas, por lo que el diámetro de su cámara era grande. Como ejemplo de reactor comercial de los años 70 se puede señalar uno que cargaba 10 obleas de 50 mm de diámetro, con una uniformidad en el espesor de las capas obtenidas de un 7,5% dentro de una oblea y 10% entre obleas [7], ver Fig. 4.14. Estos equipos podían trabajar a presión atmosférica, a baja presión y con ayuda de plasma. El aumento continuo de los diámetros de las obleas, conllevó con el tiempo a buscar nuevas soluciones entre ellas los equipos de pared caliente y baja presión, que utilizan hornos similares a los utilizados en la oxidación del silicio, donde se añade dispositivos de hermeticidad y de bombeo para establecer la presión requerida, ver Fig. 4.15. Sin embargo, debe señalarse que debido a las características constructivas de los equipos utilizados se presentan, entre otros, los problemas siguientes

1. Cuando se usan tubos con paredes calientes, se produce un depósito indeseado del producto sobre ellas.
2. Se produce una caída de presión a lo largo del tubo y entre las paredes y el centro del tubo que da origen a una variación en la velocidad de crecimiento de las capas a lo largo del horno y a lo largo de cada oblea.

Para lograr una mayor uniformidad en este espesor de oblea a oblea, según la posición en el tubo, generalmente se requiere establecer un gradiente de temperatura (a lo largo del tubo) que tienda a compensar la variación de presión señalada. También puede usarse un sistema de inyección lateral del gas a lo largo de la zona donde se colocan las obleas con lo que también se compensa en algo la caída de presión.

Además de lo señalado, hay un mayor consumo de especie activa al principio del tubo debido a que hay una mayor velocidad de crecimiento; esto hará que aún quede menos producto para consumir por las obleas

posteriores, pudiendo darse el caso de que no alcance para todas o, al menos, acentúe la variación en la velocidad de crecimiento. Para disminuir este efecto se puede incrementar la corriente de gas reactivo que se inyecta en el horno, manteniendo constante la temperatura y presión parcial de los gases reaccionantes, o sea, el *producto de la presión total por la fracción molar* de los gases. De esta forma, la velocidad de crecimiento en la entrada del tubo no varía, o sea, se mantiene el mismo consumo de gases reaccionantes; pero como se aumentó su cantidad, llegará más especie activa al final del tubo, haciendo menor el gradiente en la velocidad de crecimiento al disminuir la diferencia entre la presión parcial a la entrada y a la salida.

Resultados experimentales indican que cuando las capas se crecen a partir de hidruros, se observa una falta de uniformidad relativamente alta en sus espesores, tanto de oblea a oblea, como dentro de la misma oblea; además, se aprecia que el efecto de transferencia de masa no resulta totalmente despreciable. Este efecto se considera asociado a la absorción del hidrógeno en la superficie de la oblea, lo cual altera la concentración del gas reaccionante a lo largo de ésta. Para lograr mejor uniformidad en el espesor de las capas, dónde la temperatura de reacción lo permita, se recomienda usar halogenuros como producto reaccionante.

En la Tabla 4.6 se resumen los principales regímenes de depósito por LPCVD y algunos de sus parámetros.

Cuando el proceso de depósito se realiza en presencia de plasma, la descomposición térmica del gas reaccionante puede ocurrir a temperaturas considerablemente más bajas [7]. El plasma utilizado se caracteriza por una densidad de electrones entre 10^9 y 10^{10} cm^{-3} , y energías entre 1 y 10 eV. En la descarga gaseosa en forma de plasma, la temperatura promedio del electrón puede ser unas 10 a 100 veces mayor que la temperatura promedio de la molécula de gas; por lo que a temperaturas cercanas al ambiente

puede haber una energía del electrón suficiente para disociar las moléculas, creando especies químicamente activas.

También se utiliza para depositar capas dieléctricas, semiconductoras y metálicas el método de depósito fotoquímico a partir de fase vapor a baja temperatura y en presencia de luz ultravioleta. La luz ultravioleta descompone uno o varios de los gases presentes en la cámara, con lo que se incrementa su reactividad. A diferencia del plasma, no ha daño sobre la superficie del substrato donde se crece la capa.

Por último debe señalarse que la evaluación de la dependencia de un proceso de LPCVD con los parámetros tecnológicos no resulta fácil, ya que por ejemplo, para evaluar la dependencia del espesor de la capa con la temperatura hace mantener la relación molar y la presión total constantes. Sin embargo, en la práctica, al variar la temperatura varía la relación molar y la presión; además, como los equipos tienen una capacidad de bombeo constante, la variación de la presión de trabajo se obtiene variando la corriente total de gas.

Ya en la década de los 80, los diámetros de las obleas exigieron el paso a equipos de pared fija donde el procesamiento de obleas se hace de manera independiente, o sea, una a una. En este último caso, la productividad del proceso requerida por la industria se garantiza mediante cámaras de vacío previo y cargadores automáticos de obleas, ver Fig. 4.17. Este sistema de procesamiento individual es el más utilizado actualmente ya que permite una mayor repetibilidad y uniformidad en los espesores con el diámetro actual de las obleas.

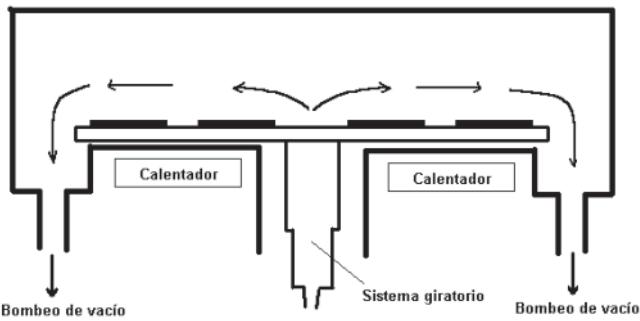


Fig. 4.14 Esquema con las partes principales de una cámara de depósito de un equipo de pared fría para presión atmosférica o baja presión.

Además de estos tipos de equipos, se utilizan equipos de paso continuo, el cual aún se utiliza para depositar capas gruesas de VFS, del orden de un micrómetro de espesor, para pasivación final. Tiene una productividad de cientos de obleas por hora, alcanzando una uniformidad mejor del 5% dentro de la misma oblea. Un esquema de este tipo de reactor se muestra en la figura 4.17.

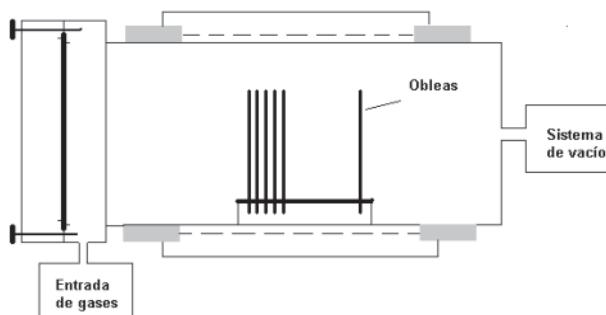


Fig. 4.15 Esquema con las partes principales de un equipo de LPCVD con pared caliente.

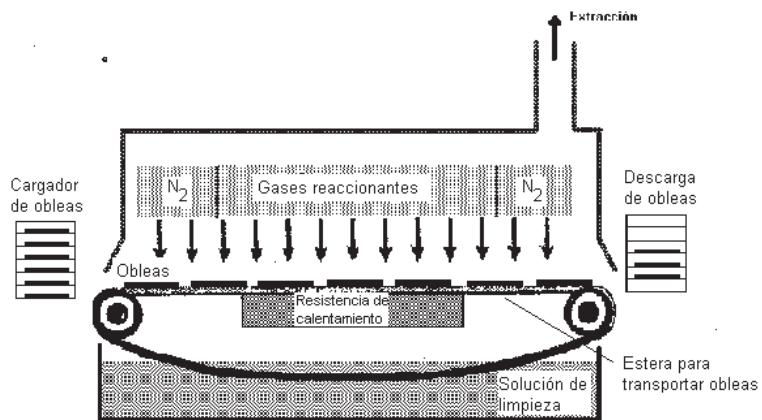


Fig. 4.16 Esquema de un equipo de paso continuo.

El equipo tiene que trabajarse dentro de una cabina de flujo laminar con muy buena extracción, ya que gases altamente tóxicos como el fosgeno y el diborano y altamente reactivos como el silano se encuentran prácticamente al aire, protegidos solamente por cortinas de gas inerte y sistema de evacuación inherente al propio equipo. Una de las ventajas de este tipo de reactor es abaratar el costo de producción.

En este reactor de paso continuo el gas se inyecta a presión directamente sobre la oblea en la que se desea crecer la capa. Frente al inyector se encuentra el sistema de extracción de gases residuales. Las obleas avanzan hacia la cámara de reacción y el calentamiento es resistivo. El espesor de la capa depende de la velocidad de desplazamiento, o sea, del tiempo que está en la cámara de reacción. Por ejemplo, para 6,5 nm/min de velocidad se crece una capa de alrededor de un micrómetro de espesor,

Tabla 4.6
REGÍMENES TÍPICOS DE DEPÓSITO POR LPCVD

Proceso	Gases	Gasto [Torr], [Pa]	Presión Ent-med-Sal [°C]	Temperatura Ent-med-Sal [nm/min]	Vel. epo. espesor %
Polisilicio $\text{SiH}_4 \rightarrow \text{Si} + 2\text{H}_2$	SiH_2	20-50	0,2-1 26,6-133,3	600-615-630 10-15	+3
Si_3N_4 $2\text{SiH}_2\text{Cl}_2 + 10\text{NH}_3 \rightarrow \text{Si}_3\text{N}_4 + 6\text{NH}_4\text{Cl} + 6\text{H}_2$	SiH_4Cl_2	40	0,3-1 39,99-133,3	770-789-808 3-5	±3
SiO_2 $\text{SiH}_2\text{Cl}_2 + 2\text{N}_2\text{O} \rightarrow \text{SiO}_2 + 2\text{N}_2 + 2\text{HCl}$	SiH_4Cl	60	0,4+0,5 120 53,3-66,6	910-925-930 8-10	±5
VFS $\text{SiH}_2\text{Cl}_2 + 2\text{N}_2\text{O} + \text{PH}_3 \rightarrow \text{SiO}_2 \cdot \text{P}_2\text{O}_5 + 2\text{N}_2 + 2\text{HCl}$ 2-10% de P	SiHCl	60	0,5-0,6 120 66,6-79,9	910-925-930 8-10	±5
SiO_2 , $\text{SiH}_4 + \text{O}_2 \rightarrow \text{SiO}_2 + 2\text{H}_2$	SiH_4	60	0,6-0,8 0 ₂ 79,9-106,6	440	12,5-15
VFS $\text{SiH}_2 + \text{PH}_3 + \text{O}_2 \rightarrow \text{SiO}_2 \cdot \text{P}_2\text{O}_5 + 2\text{H}_2$ 2-10% de P	SiH_4	60	0,6-1 0 ₂ 79,9-133,3	440	10-15
VBFS $\text{SiH}_2 + \text{PH}_3 + \text{BCl}_3 + \text{O}_2 \rightarrow \text{SiO}_2 \cdot \text{B}_2\text{O}_3\text{P}_2\text{O}_5 + 2\text{H}_2$	SiH_4 PH_3 BCl_3	2-15	0,318-0,33 42,4-44	420	2,2-2,5

El primer equipo comercial de CVD con ayuda de plasma (PECVD) data del año 1977. Actualmente presentan la forma que se muestra en la figura 4.18. Los principales factores que influyen en la velocidad de depósito son:

1. la potencia de radiofrecuencia (RF);
2. la distribución del plasma;
3. la presión de trabajo;
4. la composición y distribución de los gases.

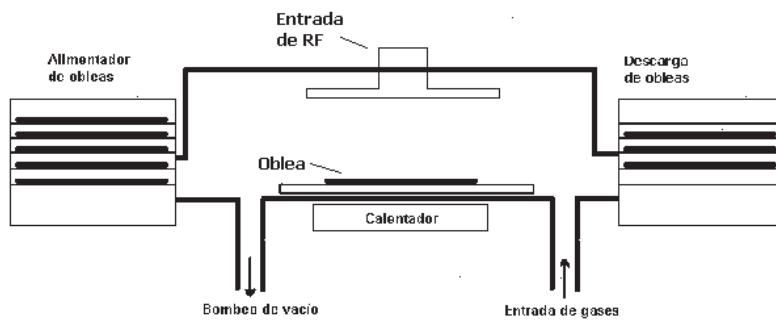


Fig. 4.17 esquema mostrando las partes principales de un equipo de PECVD actual con pared fría y tratamiento independiente de obleas

Los equipos permiten el calentamiento de las obleas hasta alrededor de 400 °C. Los gases utilizados dependen de las capas a depositar. La presión de trabajo se encuentra en el rango entre 266 a 27 Pa (2 a 0.1 Torr). Las obleas deben colocarse en el electrodo no activo, el cual generalmente se busca que tenga un área mayor que el área del electrodo activo, a donde se conecta la señal de RF que produce el plasma por las causas que se verán en el siguiente párrafo.

4.9 DESCARGA ELÉCTRICA AUTOMANTENIDA: REGIÓN DEL PLASMA

Si entre 2 electrodos se aplica un voltaje constante de DC de suficiente intensidad, el espacio entre los electrodos se divide en un conjunto de zonas oscuras e iluminadas. En la Fig. 4.18 se muestran las principales zonas en que se divide el espacio interelectródico cuando se aplica un voltaje de corriente directa entre ellos. En el espacio oscuro de Crook cae gran parte del campo eléctrico aplicado. En él se aceleran los electrones en su camino hacia una región de plasma de laboratorio llamada región de descarga negativa. En esta región, la densidad de iones es igual a la densidad de electrones, pudiendo estar en el rango entre 10^9 y 10^{13} cm^{-3} , (10^{15} - 10^{19} m^{-3}). Los electrones en esta región pueden alcanzar una temperatura entre 1 y 10 eV, que pueden también perder por choques. En el espacio oscuro de Faraday pueden volver a acelerarse hasta alcanzar energías para crear una segunda región de plasma o columna positiva.

El plasma utilizado en el depósito o ataque de capas delgadas es generalmente el correspondiente a la región de la descarga negativa, dentro de la cual se coloca el ánodo.

En la región del plasma existe igual cantidad de iones y de electrones para garantizar neutralidad eléctrica. Para que esta descarga sea automantenida se requiere:

1. Generar electrones primarios y acelerarlos hasta que alcancen la energía suficiente para ionizar el gas;
2. Dirigir los iones positivos que se crean hacia el cátodo negativo, donde dan lugar a electrones secundarios. Estos electrones secundarios se alejan rápidamente del cátodo, dejando una zona de carga espacial cercana a éste, formada por iones positivos, cuya

movilidad es mucho menor que la de los electrones. Acelerados en el campo, los electrones secundarios deben dar lugar a nuevos iones, que comenzarán también a moverse hacia el cátodo. La cantidad de electrones secundarios que se generan debe ser tal que se garantice con sus choques, la ionización del número necesario de átomos del gas para mantener la descarga.

En la región del plasma producido por una descarga gaseosa, el comportamiento de los iones y de los electrones puede ser descrito a través de la distribución de Maxwell, ya que la velocidad de desplazamiento resulta mucho menor que la aleatoria. Por ello se puede considerar que el campo aplicado entre los electrodos no altera la distribución de partículas, sino solo su temperatura.

La energía cinética de los electrones resulta mucho mayor que la de los iones, debido a la mayor movilidad de los primeros con respecto a los segundos. La temperatura de los electrones puede ser alta, pero como la masa de estos es muy pequeña, al chocar con las partículas neutrales del gas, entregan poca energía y no son capaces de subir la temperatura de todo el gas. Los iones positivos tienen mucha mayor masa y pueden entregar mucha más energía en cada choque, pero como adquieren mucha menos energía que los electrones, tampoco son capaces de subir la temperatura del gas.

Resumiendo:

- 1) El plasma es un gas parcialmente ionizado que contiene igual número de electrones e iones positivos además de las moléculas neutrales.
- 2) El grado de ionización que se alcanza, o sea, la relación de partículas ionizadas a neutras es generalmente del orden de 10^{-4} .
- 3) Se llama densidad del plasma a la densidad de iones o electrones (son iguales).

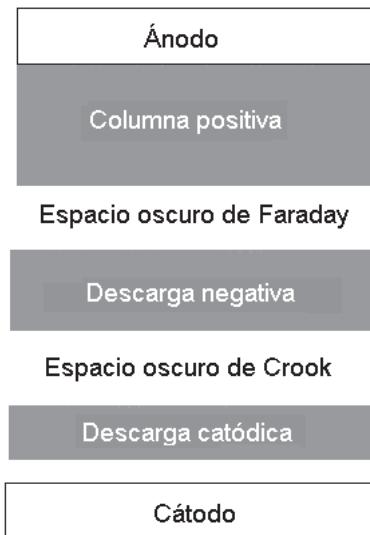


Fig. 4.18 Principales zonas en que se divide el espacio interelectrónico de una descarga luminiscente.

Los iones y electrones que el plasma se comportan de forma similar a como se distribuyen en un conductor metálico, tendiendo a prevenir la penetración del campo eléctrico hacia el interior de dicha región. La profundidad de penetración del campo en el plasma viene dada por la longitud de apantallamiento de Debye D y se expresa como:

$$D = \sqrt{\frac{k \cdot T_e \cdot \varepsilon_0}{n_e \cdot q^2}}, \quad (4.20)$$

donde T_e es la temperatura de los electrones en el plasma expresada en [K]; n_e es la densidad de electrones en el plasma en [m^{-3}]; k es la constante de Boltzman igual a 1.28×10^{-23} [J/K]; q es la carga del electrón igual a 2.9×10^{-19} [C]; ε_0 es la constante dieléctrica del vacío igual a 8.84×10^{-12} [F/m].

La temperatura T_e está asociada a la velocidad de movimiento de los electrones y se expresa a través de:

$$E_C = \frac{m_e v_e^2}{2} = \frac{3}{2} k \cdot T_e, \quad (4.21)$$

donde E_C -es la energía cinética promedio del electrón; m_e la masa del electrón y v_e la velocidad media de los electrones en el gas que para la distribución de Boltzman se expresa como:

$$\bar{v}_e = \sqrt{\frac{8kT_e}{\pi m_e}}. \quad (4.22)$$

Para una energía promedio de los electrones del orden de 2 eV, T_e es del orden de 1.5×10^4 K . La densidad de partículas en el plasma n_i , puede estimarse mediante la ley de gases ideales:

$$P = n_i \cdot k \cdot T, \quad (4.23)$$

donde T es la temperatura del gas y P la presión.

Para $P=13$ Pa (0.1 Torr), $n_i=3 \cdot 10^{21}$ m⁻³. Para un porcentaje de ionización de 10^{-4} se puede estimar una densidad de electrones en el orden de 3×10^{17} m⁻³ y D será del orden de 1.5×10^{-5} m, (1.5×10^{-3} cm), o sea mucho menor que las dimensiones de la cámara. Esto significa que los campos eléctricos estarán localizados muy cerca de las paredes de la cámara o de los electrodos que se encuentren dentro de ella.

Como ya se indicó, dentro de la región del plasma hay neutralidad eléctrica y su potencial, V_p , resulta el más alto de todo el sistema. Si esto no fuera así, tan pronto surgiera una región con potencial más positivo, los electrones se desplazarían hacia allí, empobreciendo de electrones a la

región del plasma. Para alcanzar de nuevo la condición de neutralidad, el potencial del plasma tiene que incrementarse y vuelve por lo tanto a ser el mayor del sistema.

La Tabla 4.7 muestra algunos valores típicos de temperatura y velocidad media de moléculas neutras, iones y electrones en un plasma de argón.

Tabla 4.7

Partículas	Masa [kg]	T [K]	Velocidad media [m/s]
Neutrales	7.6×10^{-26}	293	400
Iones	7.6×10^{-26}	500	520
Electrones	9.1×10^{-31}	1.5×10^4	9.5×10^5

4.9.1 Fenómenos físicos asociados a la introducción de un electrodo en una descarga gaseosa de CD.

Si se introduce un electrodo aislado en un plasma de corriente directa (CD), moléculas neutras, iones y electrones incidirán sobre su superficie, dando lugar los dos últimos a un flujo de carga que será proporcional a la velocidad media de cada una de las partículas cargadas. Como la velocidad media de una partícula es inversamente proporcional a su masa, la velocidad media de los iones será mucho menor que la de los electrones, por lo cual junto al electrodo introducido se acumulará carga negativa que a su vez dará lugar a un potencial más negativo que el presente en la región neutral del plasma, V_p . Las cargas negativas repelerán la llegada de más electrones y atraerán más iones positivos. Este proceso continuará hasta que el flujo de iones se iguale al flujo de electrones. El potencial V_f que

adquiere el electrodo aislado introducido corresponderá al potencial flotante y alrededor del electrodo se habrán acumulado iones positivos que dan lugar a una región no compensada de carga, o sea a una región de carga espacial. Esta región tendrá una densidad de carga espacial y una extensión determinada por la longitud de Debye (4.20).

En la Fig. 4.19 se muestra la característica volt-ampérica de un electrodo introducido en un plasma, donde V_p es el potencial en la región donde se introduce la sonda antes de que esto ocurra, o sea, el potencial del plasma. El eje V_s representa el potencial aplicado a la sonda. La corriente neta que llega a la sonda será la resultante de la corriente de electrones menos la corriente de iones.

Si a la sonda se le aplica un potencial externo V_s igual al potencial del plasma V_p , las cargas (tanto iones como electrones) que llegan a la sonda, serán solo aquellas que llegan por movimiento aleatorio ya que no habrá campo eléctrico actuando sobre ellas:

$$i_{s1} = i_+ - i_e = -i_e \left(1 - \frac{i_+}{i_e} \right). \quad (4.24)$$

La corriente de iones y de electrones será proporcional a la movilidad respectiva de cada uno, μ_+ y μ_e , luego (4.24) se puede escribir:

$$i_{s1} = i_+ - i_e = -i_e \left(1 - \frac{\mu_+}{\mu_e} \right) = -i_e. \quad (4.25)$$

Como $\mu_+ \ll \mu_e$, se ve que la corriente total será fundamentalmente debido a la corriente de electrones. Este caso se ha representado por el punto B de la curva I-V en la Fig. 4.19.

Si se aplica un potencial positivo a la sonda con respecto al plasma ($V_s > V_p$), los iones serán repelidos por éste y frente a ella, se creará una zona de carga espacial (z.c.e.) donde caerá la diferencia de potencial que existe entre la sonda y la región del plasma, ($V_s - V_p$). Las cargas negativas apantallan el potencial negativo con respecto al plasma que adquirió la sonda. La corriente que llega a la sonda se deberá solo a los electrones que llegaron por movimiento aleatorio a la frontera entre el plasma neutral y la región de carga espacial, (z.c.e.) y que luego son atraídos por el campo. La corriente en la sonda (región A Fig. 4.19), $i_{s2} = i_e = i_{esat}$, dependerá poco con el voltaje.

Si se aplica a la sonda un potencial menor que el potencial del plasma, o sea, se hace negativa respecto al plasma, la corriente de electrones caerá bruscamente, ya que para llegar a ella, los electrones han de tener la energía suficiente para vencer la resistencia del campo que los frena. Mientras más se disminuya el voltaje del electrodo con respecto al plasma, habrá menos electrones que cumplen con esta condición (Región C de la curva I-V).

Para un potencial dado, $V_s = V_f$, (punto D de la curva I-V en la Fig. 4.19, la corriente neta se hace cero cuando la pequeña cantidad de electrones que logra atravesar el campo, compensa la corriente de iones. Este potencial corresponde al potencial flotante que ya se mencionó y es el que adquiriere cualquier electrodo aislado que se introduce en el plasma. El mismo siempre será más negativo que el potencial del plasma.

Si el voltaje de la sonda se hace aún más negativo, todos los electrones son repelidos y a ella solo podrán llegar cargas positivas. La sonda se cubre de una capa de iones que dan lugar a una z.c.e. que apantalla su potencial negativo. La corriente que llega a la sonda es solo la debida a los iones que llegaron por el movimiento aleatorio a la frontera entre la z.c.e y el plasma y luego son atraídos por el campo en la z.c.e. Por este motivo la corriente

i_{+sat} , ahora debido a iones positivos, de nuevo no depende del voltaje aplicado a la sonda.

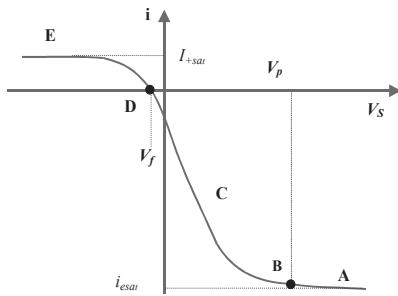


Fig. 4.19 Característica volt-ampérica de un electrodo en un plasma de CD.

Utilizando la distribución de Maxwell-Boltzman para describir el comportamiento de los electrones e iones en la región del plasma, se puede calcular la magnitud de la corriente de electrones y de iones para los casos descritos anteriormente. Si el voltaje aplicado a la sonda es negativo con respecto al plasma, (región C de la curva I-V, Fig. 4.19), los electrones son frenados por el campo en la z.c.e., pudiendo llegar a la sonda solo aquellos cuya velocidad v_x cumpla con la condición:

$$v_x \geq \left[\frac{2q(V_p - V_S)}{m_e} \right]^{\frac{1}{2}}. \quad (4.26)$$

El número de electrones que cumple con la condición (4.26) se calcula mediante la distribución de Maxwell y la corriente debida a ellos será:

$$i_e = qn_e s \bar{v}_e = qn_e s \int_{-\infty}^{\infty} dv_y \int_{-\infty}^{\infty} dv_z \int_{-v_x}^{\infty} f(v_x, v_y, v_z) \bar{v}_x dv_x = s \frac{qn_e \bar{v}_e}{4} e^{-\frac{q(V_p - V_s)}{kT_e}} \quad (4.27)$$

donde n_e es la densidad de electrones en el plasma y s el área del electrodo.

Si $V_s > V_p$, el campo del electrodo atrae todos los electrones que llegan a la frontera, y la corriente se satura al valor:

$$i = i_{esat} \cdot n_e \cdot \frac{\bar{v}_e}{4}. \quad (4.28)$$

Para calcular la corriente de iones que llega a la sonda cuando $V_s < V_f$, una primera aproximación sería sustituir \bar{v}_e por \bar{v}^+ en la ecuación (4.28) donde:

$$\bar{v}^+ = \sqrt{\frac{\bar{v}_e \cdot T_i}{T_e}} \cdot \sqrt{\frac{m}{M}}. \quad (4.29)$$

Sin embargo, un análisis más preciso considera que el potencial que frena a los electrones en la frontera entre el plasma y la z.c.e. es del orden de la energía media de estos, o sea, $V_G \approx kT_e / q \approx 1$ eV, y que la densidad de los electrones en esa frontera será:

$$n_G = n_e e^{-\frac{|V_G|}{kT_e}} \approx n_e \cdot \left(\frac{1}{e}\right). \quad (4.30)$$

Como fuera de la región de carga espacial debe cumplirse la condición de neutralidad eléctrica en la frontera entre el plasma y la z.c.e., ésta solo

será posible si los iones se mueven más rápido que lo que correspondería a su velocidad por movimiento aleatorio. Por ello se considera que los electrones arrastran a los iones en la región neutral cercana a la frontera con la z.c.e. y la velocidad que alcanzan se determina por:

$$\bar{v}^+ \approx \sqrt{2q \cdot \left| \frac{V_G}{M} \right|} \approx \sqrt{\frac{1kT_e}{M}}. \quad (4.31)$$

La corriente de iones que llega a la sonda cuando no hay campo que los repela será:

$$i_{+sat} \approx s \cdot q \cdot n_G \cdot \bar{v}^+ \approx \frac{s \cdot q \cdot n_e}{e} \cdot \sqrt{\frac{2kT_e}{M}}. \quad (4.32)$$

En el punto D, (Fig. 4.19), cuando $V_S = V_f$ la corriente de electrones será igual a la corriente de iones.

$$i_e = s \cdot \left(\frac{q \cdot n_0 \cdot \bar{v}_e}{4} \right) \cdot \exp \left[-\frac{q(V_p - V_f)}{kT_e} \right] = i_{+sat} = \frac{s \cdot q \cdot n_e}{e} \left(\frac{2kT_e}{M} \right)^{\frac{1}{2}}. \quad (4.33)$$

Sustituyendo \bar{v}_e en (4.33), el valor de la diferencia de potencial correspondiente será:

$$V_p - V_f = \frac{kT_e}{2q} \cdot \ln \left[\frac{e^2 M}{4\pi m_e} \right] = \frac{kT_e}{q} \cdot \ln \left[e \cdot \sqrt{\frac{M}{4\pi m_e}} \right], \quad (4.34)$$

luego la diferencia entre el potencial del plasma de CD y el potencial flotante de una superficie dentro del plasma es proporcional a kT_e .

Esta expresión sirve para estimar la máxima energía con la cual los iones positivos pueden bombardear las paredes aisladas de una cámara. Como la mayoría de las energías de umbral de ionización son superiores a los 20 eV, es conveniente que $V_p - V_f < 20$ V para evitar que ocurra pulverización catódica en las paredes del reactor.

4.9.2 Fenómenos asociados a una descarga producida por un campo alterno o plasma de R.F.

El plasma entre dos electrodos que tienen aplicado voltaje de directa se usa prácticamente solo para el ataque o depósito por pulverización catódica de obleas conductoras, colocadas en el cátodo o ánodo respectivamente.

Si el substrato estuviese cubierto por un material aislante, su superficie se comportaría de forma semejante al electrodo flotante que se introduce en el plasma. La superficie se cargará hasta el potencial flotante correspondiente al momento en que el flujo de iones y electrones hacia la superficie se iguala. Para resolver este problema y poder depositar y atacar substratos aislantes, hace falta utilizar descargas o plasmas de corriente alterna, (CA), de frecuencia superior a 1 MHz. En este caso ya no hay un cátodo y un ánodo, sino dos electrodos. Al electrodo al que se aplica la señal de R.F. se le llama electrodo activo. El comportamiento del potencial en cada electrodo respecto al plasma es bien diferente al caso de CD. El valor del potencial flotante ya no será único, sino que depende de la geometría del reactor.

Si aplicamos un voltaje variable V_a a un electrodo de área A_1 con superficie aislante, o a un electrodo activo que tiene en serie un capacitor y que se encuentra enfrente de otro electrodo conectado a tierra y

generalmente de mayor área A_2 , Fig. 4.20, tendrá lugar el efecto de autopolarización que se muestra en la Fig. 4.21.

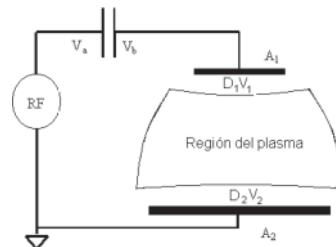


Fig. 4.20 Sistema de PECVD con electrodos asimétricos y condensador en serie.

Si el voltaje aplicado entre los electrodos es una onda cuadrada de 2 kV de amplitud pico a pico, (Fig. 4.21a), en la superficie aislante del electrodo, en el momento inicial, el voltaje V_b variará igual que V_a , ya que el condensador no puede cargarse bruscamente, (Fig. 4.21b). V_b comienza a disminuir hasta $V_b - \Delta V_1$ producto de la carga del condensador. La corriente de carga será la debida a los iones. Si antes de que este se cargue significativamente, en t_1 , el voltaje cambia de polaridad y se incrementa en 2 kV, V_b también se incrementará en 2 kV, a partir del valor de voltaje al que había llegado en t_1 . De nuevo comienza a cargarse el condensador ahora con polaridad contraria, pero, como ahora la corriente es debida a electrones, y ya se vio que esta corriente es mayor que la debida a los iones, al llegar al tiempo t_2 en que vuelve a cambiar la polaridad del voltaje V_a , el voltaje V_b habrá caído en ΔV_2 , siendo $\Delta V_1 < \Delta V_2$. Al variar V_b en 2 kV de nuevo, alcanzará un valor más negativo que su valor inicial. Este proceso se repite en los primeros ciclos, hasta que al final se alcanza un desplazamiento significativo de la señal V_b hacia valores negativos.

Si en lugar de un pulso cuadrado se aplica una señal sinusoidal, se observará una autopolarización del electrodo hacia valores negativos como se muestra en la Fig. 4.21c. El potencial de R.F. V_b en el electrodo, luego de transcurridos los primeros ciclos, prácticamente se encontrará por encima de cero solo en cortos intervalos de tiempo y por lo tanto, el electrodo activo se mantendrá casi todo el tiempo polarizado negativamente. En ese tiempo estará sometido al bombardeo de los iones, cuya energía promedio al llegar al electrodo activo puede estimarse a través de la diferencia entre el potencial del plasma V_p y el potencial de autopolarización negativa V_{DC} adquirido por el electrodo. El valor de V_{DC} , con signo negativo, es aproximadamente la mitad del valor del voltaje de R.F. de pico a pico, V_a , aplicado externamente. La energía de los iones que llegan al electrodo conectado a tierra es menor, pues V_p tiene un valor cercano a cero todo el tiempo. Como el potencial del plasma no puede ser negativo, y además ya se vio que tiene que ser algo mayor que el mayor de los voltajes aplicado a los electrodos, el potencial del plasma variará según se indica en la línea discontinua de la Fig. 4.21c, y su valor será cercano a cero.

Si los electrodos fueran de igual área, (simétricos), la autopolarización no ocurre pues uno no se puede cargar más que el otro. El potencial V_b en el electrodo activo es similar a V_a y V_p alcanza valores mayores y varía más en el tiempo, según se indica en la línea discontinua de la Fig. 4.21d. La diferencia entre los valores promedios de V_p y del voltaje en cada electrodo es similar por lo que hay mas probabilidad de pulverización catódica en ambos electrodos y en las paredes de la cámara.

Esta diferencia en comportamiento es utilizada en la práctica para diferenciar la colocación de los substratos, según se quiera depositar o eliminar material. El menor bombardeo iónico lo sufrirán los substratos colocados en el electrodo de tierra en cámaras con electrodos asimétricos.

Un análisis más detallado de los procesos básicos que tienen lugar en el plasma puede encontrarse en [19].

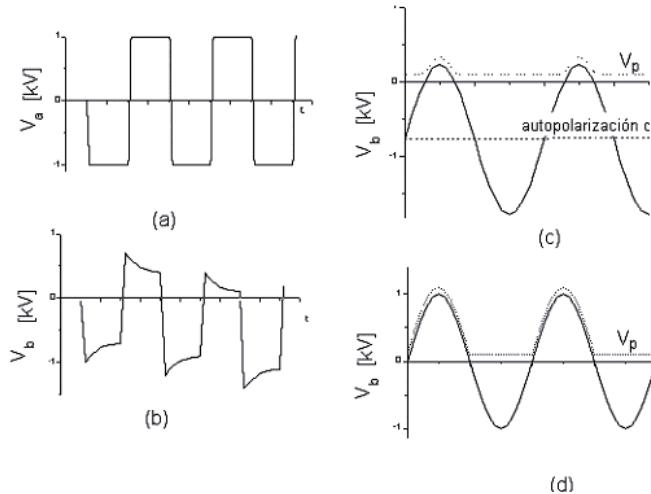


Fig. 4.21 Efecto de autopolarización de los electrodos del plasma.

Koenig y Maissel [20], desarrollaron el siguiente modelo para explicar el comportamiento indicado en electrodos con diferente área.

V_1 y V_2 son los voltajes que caen en la z.c.e adyacente a cada electrodo;

D_1 y D_2 el ancho de cada zona de carga espacial;

A_1 y A_2 las áreas de los electrodos;

El flujo de iones de masa M que provenientes del plasma pueden atravesar la z.c.e. sin colisiones será:

$$j_i \propto \frac{V^{\frac{3}{2}}}{M^{\frac{1}{2}} \cdot D^2}. \quad (4.35)$$

Como la densidad de la corriente de iones tiene ser igual en ambos electrodos,

$$\frac{V_1^{\frac{3}{2}}}{D_1^2} = \frac{V_2^{\frac{3}{2}}}{D_2^2}. \quad (4.37)$$

La capacitancia de la z.c.e. será proporcional al área de cada electrodo e inversamente proporcional al ancho de la z.c.e.

$$C \propto \frac{A}{D} \quad (4.38)$$

El voltaje de R.F. se dividirá entre las dos z.c.e. de manera que:

$$\frac{V_1}{V_2} = \frac{C_2}{C_1}. \quad (4.39)$$

Luego:

$$\frac{V_1}{V_2} = \left(\frac{A_2}{A_1} \right)^4. \quad (4.40)$$

De donde se desprende que:

- La mayor caída de voltaje ocurrirá junto al electrodo menor.
- La campana generalmente se conecta al electrodo de tierra para aumentar el área total de este electrodo.

Como ya se indicó, esta característica es utilizada para determinar la colocación de los substratos en los electrodos, según el proceso a realizar.

En la descarga de R.F., el proceso de ionización se produce por el impacto de los electrones que oscilan a la frecuencia ω de variación del campo externo $F = F_m \cos \omega t$, pudiendo alcanzar una velocidad máxima de:

$$v_o = \frac{qF_o}{m\omega} \quad (4.41)$$

donde: F_o - amplitud del campo eléctrico alterno.

El desplazamiento máximo será:

$$x_m = v_o / \omega \quad (4.42)$$

Para $F_o = 10$ V/cm y $f = 13.56$ MHz, que es la frecuencia de trabajo común de los equipos de depósito por PECVD o de ataque en seco, el desplazamiento máximo será de 2.42 cm. Esto es válido mientras no se consideran las colisiones entre partículas. El recorrido libre medio es proporcional al inverso de la presión y directamente proporcional a la sección eficaz de choque, lo cual a su vez depende del diámetro de la partícula, o sea del gas. En cualquier caso, el desplazamiento medio es generalmente menor que los valores frecuentes de distancia interelectródica.

La combinación de este movimiento oscilatorio y del choque de los electrones con átomos no ionizados del gas hace que durante las oscilaciones, los electrones puedan incrementar su energía hasta ser capaces de ionizar nuevos átomos. Los electrones secundarios no son imprescindibles para mantener el plasma, ni hacen falta campos tan altos.

El comportamiento de los iones en el campo de R.F., es similar al de los electrones pero al ser su masa mucho mayor, la energía que estos alcanzan en sus oscilaciones es mucho menor. Los potenciales que aparecen en

diferentes puntos de la descarga de R.F., y que dependen de la geometría y potencial al que están conectados los electrodos, son importantes para determinar con qué energía inciden los iones sobre las superficies que se atacan, definiendo si predomina la erosión física (energías mayores de varios eV) o la activación química de los gases reaccionantes (energías menores de varios eV).

La energía máxima posible con que los iones bombardean una superficie vendrá dada por el potencial de ésta con respecto al plasma.

La oblea a depositar siempre se coloca en el electrodo a tierra, ya que la energía con que lleguen los iones será menor y se producirá menos daño en ella producto del bombardeo iónico. Si el plasma se utiliza para ataques en seco, la oblea puede colocarse en cualquiera de los electrodos. Se coloca en el electrodo activo, si se desea una mayor influencia de la erosión física y en el electrodo de tierra si se desea que predomine el ataque químico donde el plasma solo favorece el que ocurran las reacciones químicas entre los gases.

Este último también ocurre para obleas colocadas (a potencial flotante), entre los electrodos.

Si durante la erosión catódica en lugar de conformar el plasma en base a argón, se utilizan otros gases reactivos, similares a los utilizados en el ataque por plasma, se obtiene la llamada erosión catódica reactiva. Entre las diferencias fundamentales de la erosión catódica reactiva (Reactive Sputter Etching RSE) y el ataque en plasma se encuentra una menor presión y una mayor energía de los iones en el primer caso en comparación con el segundo. También el substrato se coloca en el electrodo activo (conectado a la R.F.) en el primer caso, mientras que en el segundo se coloca en el electrodo a tierra.

4.10 CARACTERÍSTICAS DEL EQUIPAMIENTO UTILIZADO EN EL DEPÓSITO DE CAPAS POR PULVERIZACIÓN CATÓDICA

El bombardeo de la superficie del electrodo activo por los iones positivos que se producen en la región del plasma, explicado en el epígrafe anterior, da lugar a un desgaste gradual de este, que se conoce como pulverización catódica o *sputtering*.

Entre las primeras hipótesis estudiadas para explicar cualitativamente este fenómeno se pueden mencionar las siguientes:

1. La pulverización catódica es producto de un calentamiento localizado del cátodo por el choque de los iones, lo que da lugar a la evaporación de átomos del material;
2. Los iones al chocar con el cátodo, luego de acelerarse en la región catódica, no tienen energía suficiente para liberar un átomo del cátodo, pero si pueden aumentar su frecuencia de oscilación. Al producirse un incremento de la amplitud de las oscilaciones del átomo superficial, debido a un efecto de resonancia, este puede ser capaz de vencer la fuerza del enlace.

De ellas, la segunda explica mejor lo que se observa, aunque detalles más precisos pueden encontrarse en [19].

El efecto de pulverización catódica permite depositar metales y dieléctricos siendo necesario en este último caso, el uso de un plasma de R.F.

Entre las ventajas de este método de depósito se encuentran:

1. No requiere gases tóxicos.
2. Permite el control de espesores muy finos y repetibilidad en los procesos.
3. Permite depositar capas a baja temperatura.
4. Pueden depositarse tanto capas del material del cátodo como de compuestos químicos del material del cátodo, cuando se usa la deposición reactiva.
5. Permite depositar sin variar la composición, capas de componentes químicos.

Un equipo de pulverización catódica tiene una cámara similar a la mostrada en la figura 3.21 para un equipo de PECVD. Las diferencias básicas se encuentran en que el cátodo es del material que se desea depositar. La oblea sigue siendo colocada en el electrodo de tierra.

Si se introducen gases reactivos, como el oxígeno o el nitrógeno, con un cátodo de silicio pueden depositarse nitruro de silicio u óxido de silicio. En este caso se dice que tiene lugar una pulverización catódica reactiva, siendo un proceso de depósito donde predomina el efecto físico de erosión catódica, aunque combinado con la reacción del material que se desprende del cátodo con el gas que se introduce en la cámara.

Entre las desventajas se encuentra la necesidad de preparar cátodos del material a depositar con la suficiente pureza y el diámetro requerido para garantizar la uniformidad de la capa en toda la oblea.

4.11 BIBLIOGRAFÍA

- [1]. L.H. Kaplan, M.E. Lowe, "Phosphosilicate Glass Stabilization for MOS structures", Journal of the Electrochemical Society, 118 (10) (1972) 1649.

- [2]. J.C. Tombs, F.A. Sewel, “Silicon Oxide as an Etching Mask for Si₃N₄”, Journal of the Electrochemical Society, 115 (1) (1968) 101.
- [3]. W. Van Golder, V.E. Hauser, “Etching of Si₃N₄ in H₃PO₄”, Journal of the Electrochemical Society, 144 (8) (1976) 869.
- [4]. H.G. Lee, E.W. Dutton, “Two Dimensional Low Concentration B Profiles Modeling and Measurements”, IEEE Transaction ED, 28 (10) (1981) 1137.
- [5]. W. Korn, “Advantages in Deposition Process for Passivation Films”, Journal Vac, Science and Technology, 14 (1977) 1082.
- [6]. N. Goldsimth, W. Kern, “The Deposicion of vitreous SiO₂ films from SiH₄, RCA Review, 28 (7) (1967) 153
- [7]. W.A Brown, T.I. Kamins, “An analysis of LPCVD System Parameters for Polysilicon, Silicon Nitride and Silicon Dioxide Deposition”, Solid State Technology, 22 (1979) 51.
- [8]. P.J. Tobin, J.B. Price, L.M. Campbell, “Gas Phase Composition in the Low Pressure Chemical Vapor Deposition of Silicon Dioxide”, Journal of the Electrochemical Society, 127 10 (1980) 2222.
- [9]. L.W. Winkle, C.W. Nelson, “Improved atmospheric-Pressure Chemical Vapor Deposition System for Deposition Silica and Phosphosilicate Glass Thin Films”, Solid state Technology, 84 (9) (1981) 123.
- [10]. W. Kern, G.L. Salmoble, “Chemically Vapor. Deposited Borofosfositicate Glasses for Silicon Device Applications”, RCA Review, 43 (2) (1982) 423.
- [11]. T. Foster, G. Hoeye, J. Goldman, “A Low Pressure BPSG Deposition Process”,Journal of the Electrochemical Society, 132 (2) (1985) 505.

- [12]. W. Kern, R.K. Smeltzer, “Borophosphosilicate Glasses for Integrated Circuits”, Solid State Technology, 28 (6) (1985) 171.
- [13]. S. Shanfield, S. Bay: “Process Characterization of BPSG Plasma Deposition”, Journal of the Electrochemical Society, 131 (5) (1984) 2202.
- [14]. Martell, A., M. Estrada, “Obtención de Capas de SiO₂ y Si₃N₄ Para la Fabricación de Dispositivos MOS”, en Seminario Científico, Facultad de Ciencias, Dic. 1977.
- [15]. L. Mei, M. Rivier, Y. Kwark, R.W. Dutton, “Grain Growth Mechanism in Polysilicon”, Journal of the Electrochemical Society, 129 (8) (1982) 1791.
- [16]. J.P. Colinge, E. Demoulin, F. Delannay, M. Lobet, J.M. Temerson, “Grain Size and Resistivity of LPCVD Polycrystalline Silicon Films”, Journal of the Electrochemical Society, 128 (9) (1981) 2009.
- [17]. B. Chapman, “Glow Discharge Processes”, John Wiley Sons, 1980.
- [18]. H.K. Koenig, L. I. Maissel, “Applications of RF Discharges to sputtering IBM J. Res. Develop. 14 (1970) 168.

CAPÍTULO 5

ATAQUE EN SECO DE CAPAS DIELÉCTRICAS, SEMICONDUCTORAS Y METÁLICAS: MÉTODOS DE ATAQUE Y CARACTERÍSTICAS DE CADA UNO

Los procesos físico-químicos de oxidación térmica, introducción localizada de impurezas, depósito de capas semiconductoras, dieléctricas y metálicas epitaxiales y no epitaxiales, hacen posible la formación de diferentes regiones y capas en y sobre el substrato semiconductor, que deben tener una ubicación previamente determinada para poder conformar un circuito integrado. Mediante la fotolitografía, ya sea antes o después del proceso en cuestión, se logra definir la localización de estas regiones y capas.

El proceso de fotolitografía consta de dos etapas:

a) Inicialmente deben ser trasladadas a la superficie de una capa de resina fotosensible, las figuras geométricas que definirán las regiones que se delimitarán finalmente en cada capa o substrato. Estas figuras a su vez, se definen previamente durante la etapa de diseño y se imprimen fotográficamente en cada una de las máscaras requeridas durante el proceso de realización del juego de máscaras.

b) Seguidamente estas figuras grabadas en la fotorresina, deben ser trasladadas a alguna capa dieléctrica, semiconductor o metálica existente sobre el substrato, o a varias de ellas a la vez. Durante muchos años, esta segunda etapa del proceso de fotolitografía se realizó mediante las técnicas de ataque húmedo, donde los productos utilizados son soluciones que

atacan de forma selectiva a los diferentes materiales. El ataque del material es el resultado de una reacción química entre la solución de ataque y el material. Las reacciones químicas también pueden ocurrir entre un gas o una mezcla gaseosa y el material, dando lugar a una reacción o ataque en seco. Este tipo de ataque permitió obtener líneas con dimensiones menores de micrómetros.

La primera utilización que tuvo el ataque en seco en la tecnología planar fue la eliminación de la fotorresina revelada y atacada. En muchas ocasiones, eliminarlas mediante ataque húmedo resultaba difícil y poco confiable. Este problema se agudizó, al buscarse precisamente una mejor adherencia de la fotorresina y una mayor estabilidad ante los diferentes atacantes, con lo cual se aumentaba la dificultad para su eliminación final. También las resinas expuestas a la implantación iónica resultaban extremadamente difíciles de eliminar.

Al mismo tiempo, a medida que se redujeron las dimensiones mínimas de las figuras utilizadas en los circuitos integrados, la utilización del ataque húmedo comenzó a presentar dificultades debido a su isotropía, o sea, la propiedad de que al atacar en sentido vertical hacia el substrato, también se produce un ataque lateral que afecta las dimensiones finales de la capa que se está atacando. Esto impide que las figuras en la capa en cuestión, queden con las dimensiones determinadas durante el diseño.

Para luchar contra este problema, se utilizó el concepto de correcciones tecnológicas, el cual consiste en determinar un factor por el cual hay que aumentar o disminuir las dimensiones de la topología diseñada para cada capa en cuestión. De esta manera, si se definía una isla cuadrada de lado L_1 y se conocía que el ataque producía un comido lateral de ΔL_1 , en la máscara se preparaba una isla cuadrada de $(L_1 + \Delta L_1)$ de lado. De igual

manera, para obtener un hueco con la dimensión L2, se prepara en la máscara un hueco con dimensiones de $L2-\Delta L2$.

Sin embargo, las dimensiones críticas siguieron disminuyendo y llegó el momento en que la única solución al problema consistía en reducir los ataques laterales, cuestión que solo se puede lograr con el ataque selectivo anisotrópico en seco. Otro de los factores que favorecieron la introducción de los ataques en seco fue el que la fotorresina no era capaz de soportar sin levantarse, determinadas soluciones húmedas como es el caso de la solución de ataque para el polisilicio y para el nitruro. Este problema se combatió durante bastante tiempo, utilizando sobre el polisilicio y sobre el nitruro una capa de óxido de silicio, sobre la cual se deposita la fotorresina. Una vez expuesta y revelada la fotorresina, el óxido se ataca con el tradicional "buffer de HF" y a partir de allí, esta capa es utilizada como máscara, eliminándose incluso la fotorresina inicialmente depositada sobre ella. Esta capa de SiO_2 sí es capaz de soportar a los atacantes del polisilicio y del nitruro, pero constituye un proceso adicional a realizar.

Por todo lo señalado, las técnicas de ataque en seco tienen una importancia fundamental en la realización exitosa de los procesos de fabricación de C.I. de alto nivel de integración.

5.1 MÉTODOS DE ATAQUE EN SECO Y PRINCIPIOS FÍSICOS EN LOS QUE SE BASAN

El ataque húmedo del SiO_2 se realiza en la solución llamada "buffer" teniendo una velocidad de ataque mucho mayor que la que presenta el silicio. Este hecho, ampliamente conocido durante años en la tecnología planar, encontró un inesperado contraste al determinarse que en presencia

de plasmas de freón 14 (CF_4), la velocidad de ataque del SiO_2 puede ser mucho menor que la del silicio.

La explicación a este fenómeno se encontró en el cambio del mecanismo de reacción. En el ataque húmedo a presión normal, tienen lugar reacciones iónicas, donde la probabilidad de que ocurra la rotura de un enlace por interacción entre átomos de diferentes materiales es mayor que entre dos átomos del mismo material. Esto fundamenta el que se observe una mayor velocidad de ataque de compuestos del Si, como el SiO_2 , en comparación con el Si puro.

Las reacciones moleculares fueron descubiertas bastante después y están asociadas precisamente a fenómenos en el plasma, donde además de la ionización de los gases debido al impacto de los electrones, tiene lugar la **disociación molecular que produce radicales, átomos e iones negativos**, como por ejemplo:



Las reacciones de disociación molecular pueden predominar sobre las de ionización, dando lugar a que el ataque de las capas presente características bien diferentes a las que se observan en los ataques húmedos. Un análisis teórico de los fenómenos que allí ocurren puede encontrarse en [1-3].

Por otro lado, los iones del plasma pueden alcanzar en la región de carga espacial interelectródica la suficiente energía para que, al incidir sobre el substrato, extraigan átomos del mismo, dando lugar al fenómeno conocido como erosión catódica.

Los ataques en seco se basan en la acción de gases parcialmente ionizados, (plasma de laboratorio), que contienen iones, electrones y

átomos neutros. Este plasma se produce por descargas eléctricas a baja presión en el rango de 0.133 Pa a 1.33 kPa (10^{-3} Torr a 10 Torr). Entre los ataques en seco se incluyen el ataque en plasma; el ataque por erosión catódica o "sputtering"; el ataque por erosión catódica reactiva (RSE) o "sputtering" reactivo y el ataque iónico reactivo (RIE). Estas técnicas, aunque tienen un origen común, se diferencian por las condiciones de la descarga gaseosa, el tipo de gas, y las formas de los equipos utilizados [1-3].

Mientras la técnica de ataque por erosión catódica se basa fundamentalmente en la erosión física, el ataque en plasma es el que representa el caso más cercano a una reacción química, donde los gases se encuentran en estado de plasma y predominan las reacciones del tipo radical.

El RSE y el RIE añaden a la erosión física las reacciones químicas que dan lugar a compuestos volátiles o semi-volátiles.

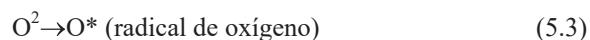
Resumiendo, los procesos de ataque en seco incluyen los procesos basados en erosión física como la erosión catódica; los basados en plasma como activador de reacciones químicas, (ataque en plasma); y los que constituyen una mezcla de ambos procesos, (erosión catódica reactiva y ataque iónico reactivo). Las técnicas más utilizadas en microelectrónica son el ataque en plasma y el ataque iónico reactivo.

5.2 EQUIPOS MÁS UTILIZADOS EN LOS PROCESOS DE ATAQUE EN SECO

5.2.1 *Equipos del tipo de barril y de placas paralelas*

Como ya se indicó, una de las primeras aplicaciones que tuvo el ataque en seco fue la eliminación de las fotorresinas expuestas. Para esto se utilizó

un equipo del tipo de barril y cuyo diagrama se muestra en la Fig. 5.1. Las muestras son colocadas en un soporte de cuarzo o aluminio que se introduce en la cámara. La cámara se evaca hasta alcanzar un vacío primario mediante bomba mecánica y seguidamente se introduce oxígeno, hasta alcanzar la presión de trabajo en el rango de 13.3 Pa a 1.33 kPa, (10^{-1} Torr a 10 Torr). Seguidamente se aplica la radio frecuencia, con una frecuencia típica de 13.56 MHz. y se genera el plasma. Radicales de oxígeno que se forman en el plasma reaccionan con la resina según la reacción:



Los gases residuales se extraen de la cámara.

Posteriormente, estos equipos fueron utilizados también para atacar capas de polisilicio, nitruro de silicio y óxido de silicio entre otros, sufriendo algunas modificaciones con vistas a perfeccionar sus parámetros. Así surgieron los llamados equipos de tipo condensador y de bobina. Como se observa en la Fig. 5.1, en este tipo de equipos, las obleas se colocan de forma simultánea dentro de la cámara.

A medida que creció el tamaño de las obleas y se hicieron más fuertes los requerimientos para obtener líneas más finas, (por debajo de 3 micrómetros), fue necesario dar paso a otra generación de equipos constituida por los equipos de placas paralelas y de tratamiento individual.

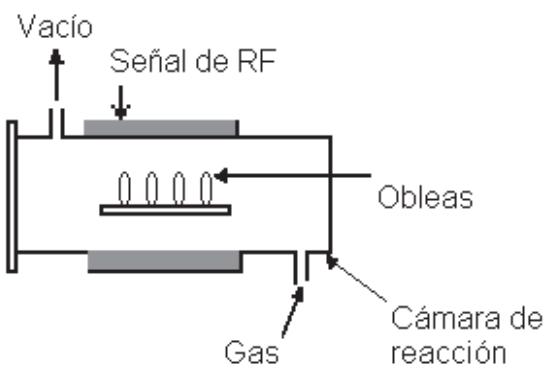


Fig. 5.1 Vista lateral de un equipo de ataque en seco, asistido por plasma, tipo barril.

En la Fig. 5.2 se muestra esquemáticamente un equipo de este último tipo. Contiene un alimentador de obleas y un descargador de obleas. Cada oblea se transporta desde el alimentador hasta la cámara de reacción una a una. Con este tipo de tratamiento individual, el volumen de la cámara se redujo considerablemente y se pudo obtener un ataque más uniforme y más preciso. La inclusión de detectores de fin de proceso permitió ganar además en repetibilidad y control. El esquema del interior de la cámara de reacción de placas paralelas es muy similar al interior de la cámara de reacción utilizada en los equipos de depósito de capas por PECVD. Si el ataque se realiza en plasma, donde predomina el ataque químico, la oblea se coloca en el electrodo de tierra. Si el ataque se realiza por erosión iónica reactiva (RIE), la oblea se coloca en el electrodo activo y se observa una combinación de reacción química y de erosión física.

La secuencia de trabajo en ellos consta de los siguientes pasos:

- La oblea se envía desde el cargador mediante un transportador hasta el soporte. Este soporte se mantiene a una temperatura dada, (del

orden de 100-120 °C para el polisilicio y el nitruro de silicio). El soporte se levanta para permitir el cierre de la cámara, la cual es evacuada mediante una bomba primaria. El gas se introduce a una velocidad controlada mediante un flujómetro; se conecta la radiofrecuencia y se genera el plasma.

- Durante el ataque, se sensa de alguna manera el momento en que termina totalmente el ataque de la capa. Cuando el proceso se acaba, el plasma se apaga y se evacuan los gases residuales. Se da entrada de aire a la cámara, el soporte se baja y la oblea en proceso se envía mediante otro transportador al cargador receptor.

Si el proceso que se realiza es isótropo, a medida que se realiza el ataque vertical, tiene lugar un comido lateral, por lo que es conveniente no excederse en el tiempo de ataque del estrictamente requerido. Incluso si el ataque es anisótropo, al llegar a la capa subyacente a la que se está atacando, comenzará el ataque no deseado de esta. A pesar de que haya una buena selectividad y su velocidad de ataque sea menor que la de la capa que se está procesando, siempre algo se atacará. De allí que sea muy importante el poder determinar de forma precisa el momento en que se termina el proceso de ataque de la capa seleccionada.

Existen diferentes métodos para detectar el fin de proceso entre ellos:

1.- Detección óptica: utiliza la elipsometría para medir el espesor de la capa que se va atacando.

2.- Espectroscopía de masa: determina los productos presentes en la cámara y por lo tanto permite conocer cuando termina la reacción de ataque.

3.- Detección por láser: El haz de fotones de un láser de He-Ne se hace incidir sobre la oblea y se observa el patrón de interferencia resultante, luego se va siguiendo la variación del espesor de la capa.

5.- Espectroscopía de emisión: Se monitorea la radiación producto de las transiciones entre los niveles energéticos de los átomos, moléculas e iones en el plasma, luego también permite determinar cuándo terminó la reacción de ataque.

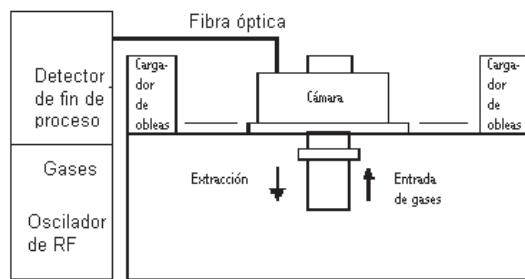


Fig. 5.2 Partes fundamentales de un equipo de ataque asistido por plasma.

Si se está realizando, por ejemplo, el ataque de una capa de polisilicio o del nitruro de silicio en atmósfera de freón 14 (CF_4), la cantidad de flúor en la cámara disminuye durante el proceso debido a que ocurre una de las siguientes reacciones:



Si se controla el espectro de emisión del flúor, se puede obtener un gráfico como el que se muestra en las Fig. 5.3.

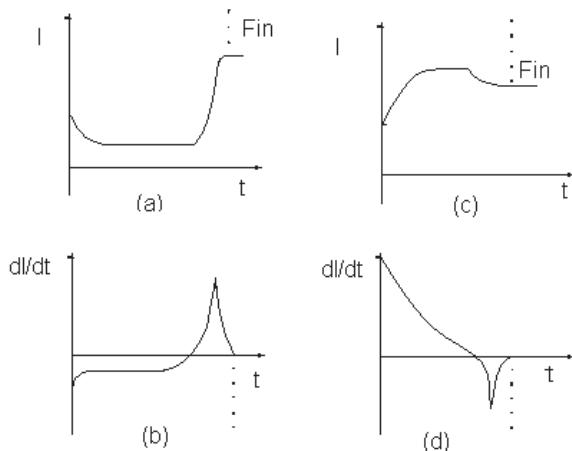


Fig. 5.3 Variación de la intensidad luminosa I y su derivada correspondiente a la línea de emisión del flúor en función del tiempo de ataque para: a) una capa de polisilicio ($\lambda=704$ nm); b) una capa de Si_3N_4 ($\lambda=674$ nm).

El agente atacante fue una mezcla de CF_4 con O_2 al 5%, con una presión total de 47.6 Pa (0.35 Torr). La potencia del plasma fue de 150 W para el polisilicio y de 100 W para el nitruro. En la Fig. 5.3 se muestra el comportamiento de la línea correspondiente a 704 nm para el ataque de polisilicio y de 674 nm para el nitruro respectivamente. En el caso del polisilicio se observa que a medida que termina el proceso, la intensidad aumenta hasta saturar. En el caso del nitruro, el efecto es contrario. Parte de un valor alto y decrece ligeramente. Si en lugar de observar la intensidad de la línea, se observa su derivada en el tiempo, (Fig. 5.3 b y d), es posible detectar de modo más preciso el momento en que finaliza el proceso de ataque, cuando se observa ya sea un máximo o un mínimo en la derivada.

Para observar la variación en el tiempo de la intensidad de la longitud de onda deseada, se utiliza un monocromador o un filtro óptico.

La intensidad de la longitud de onda monitoreada debe cambiar bruscamente al terminar de atacarse la capa que se controla; no debe ser afectada por reacciones con la resina o con el substrato, ni estar cercana a otras longitudes de ondas que puedan aparecer por otras causas en el espectro que se observa. La longitud de onda medida no debe ser afectada por la influencia de otros factores, como pueden ser:

- La variación de los flujos de gas;
- La variación de la descarga del plasma.

Tampoco deben ocurrir cambios en la luz transmitida al detector, debido al depósito de productos de la reacción sobre la superficie de la cámara que impide el paso de la luz a través de ella.

Un esquema simplificado de los bloques fundamentales del detector de fin de proceso se muestra en la Fig. 5.4.

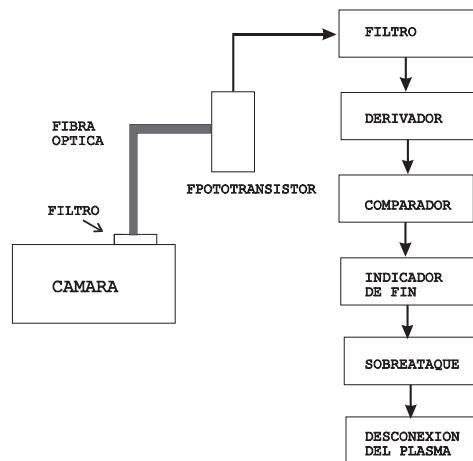


Fig. 5.4 Bloques fundamentales de un detector de fin de proceso por filtro óptico.

5.2.2 Características comunes a los equipos de ataque en seco

Los equipos de ataque en seco deben garantizar los siguientes requerimientos:

- 1.- Densidad uniforme del plasma dentro de la cámara.
- 2.- Estabilidad del plasma
- 3.- Uniformidad en el flujo de gas.
- 4.- Que se mantenga constante la velocidad de ataque y de eliminación para el caso de las resinas, dentro de una misma oblea y entre oblea y oblea.
- 5.- Mínimo daño a la oblea por el plasma.
- 6.- Que no se produzca contaminación de la oblea.
- 7.- Posibilidad de procesar un conjunto grande de obleas por unidad de tiempo.
- 8.- Posibilidad de control y automatización del proceso.
- 9.- Confiabilidad y facilidad en el mantenimiento.
- 10.- Selectividad del ataque mediante los gases utilizados.

Además es conveniente mencionar los siguientes aspectos de interés:

- La energía de los iones que inciden sobre la muestra define el que predomine la erosión física o las reacciones químicas. Si la erosión física es muy alta, puede dar lugar a defectos en la oblea que pueden afectar la uniformidad del proceso e incluso producir inestabilidad en estructuras MOS y/o deterioro de sus características eléctricas.
- El flujo del gas reactivo y la temperatura del proceso influyen en la activación de las reacciones químicas. La selectividad de la reacción también es afectada por la temperatura.
- La presencia del plasma puede dar lugar al calentamiento de la oblea, al cual se añade el calentamiento por reacciones exotérmicas.

- El proceso de ataque depende del tamaño la superficie a atacar. A este efecto se le conoce como efecto de carga. La velocidad de generación de especie activa en la cámara se fija mediante los parámetros de control del proceso, (presión, potencia, frecuencia, etc.) y es independiente del área a atacar. A mayor área, más especie activa se consume, disminuyendo su concentración en el plasma. Como la velocidad de ataque depende de esta concentración, esta velocidad puede variar con el área a atacar.
- En cualquier equipo con placas paralelas, la forma del ataque depende de la presión. A alta presión, el ataque es controlado por difusión con comido lateral significativo. Al bajar la presión, el movimiento de partículas es más perpendicular a la superficie y la direccionalidad del ataque se mejora.
- La eliminación de los productos de la reacción sobre las superficies resulta más efectiva si se trabaja a baja presión debido a que se tiene un camino libre medio mayor para las especies.

5.3 RESISTENCIA DE LA FOTORRESINA AL ATAQUE EN SECO

Hasta el momento se ha descrito de forma independiente la utilización del proceso de ataque en seco por un lado, para realizar ataques en diferentes capas semiconductoras, dieléctricas y metálicas previamente cubiertas de una capa de fotorresina sobre la que se realizó un proceso de fotolitografía y por otro para eliminar la fotorresina después de realizado los ataques correspondientes. El proceso de ataque en seco, a diferencia del húmedo, atacará la capa deseada y la fotorresina.

Esta aparente contradicción se puede comprender analizando la Fig. 5.5, donde se muestra una comparación entre las velocidades de ataque en seco de una capa de SiO_2 térmico y de una capa de fotorresina positiva.

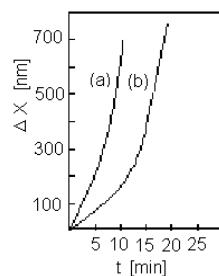


Fig. 5.5 Comparación de: (a) la velocidad de ataque del SiO_2 y (b) la velocidad de ataque de una fotorresina positiva para un ataque en plasma con $\text{CF}_4\text{-O}_2$.

Aunque se observa que la fotorresina se ataca, la velocidad de ataque es menor que la del SiO_2 , por lo cual puede ser utilizada como máscara siempre que no haya una sobre-exposición excesiva en el tiempo de ataque. Esto es perfectamente alcanzable utilizando cualquier medio de detección de fin de proceso de los anteriormente señalados.

5.4 ATAQUES ISÓTROPOS Y ANISÓTROPOS; PRINCIPALES CARACTERÍSTICAS Y APLICACIONES

Con la disminución de las dimensiones mínimas en los C.I. fue necesario introducir los ataques anisótropos, debido a los siguientes problemas inherentes al ataque isótropo.

Durante el ataque isótropo de una capa, al mismo tiempo que se realiza el ataque vertical hasta una profundidad d , se produce un ataque lateral, que en la superficie alcanza la misma magnitud d . La Fig. 5.6 a y b, muestra el caso ideal de un ataque isótropo cuando se abre un hueco y cuando se define una isla respectivamente. Si por el contrario se realizara un ataque totalmente anisótropo, o sea, sin ningún ataque lateral, se obtendrían estructuras como las que se observan en la Fig. 5.7 a y b.

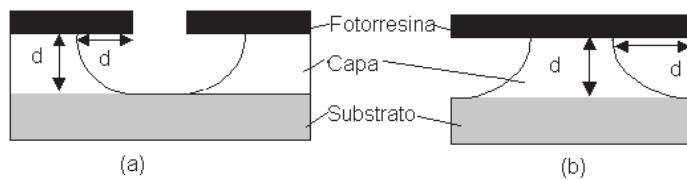


Fig. 5.6 Esquema de un ataque isótropo para: a) abrir un hueco; (b) definir una isla. Se considera que la fotorresina no se ataca durante el proceso.

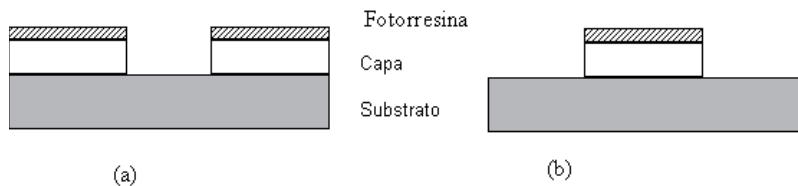


Fig. 5.7 Esquema de un ataque anisótropo ideal: (a) para abrir un hueco; (b) para definir una isla. Se considera que la fotorresina no se ataca durante el proceso.

En la práctica los ataques nunca son totalmente anisótropos, siendo conveniente definir el concepto de grado de anisotropía como:

$$A = I - v_l/v_v, \quad (5.7)$$

donde: v_l y v_v son las velocidades de ataque lateral y perpendicular respectivamente.

De acuerdo a la definición de grado de anisotropía, si $A=0$, También se atacará el substrato. se está en presencia de un proceso isotrópico. Si $A=1$, el proceso es totalmente anisótropo. En los casos intermedios, A varía entre 0 y 1.

Cuando se haya completado el proceso de ataque de una capa, el comido lateral en la superficie de la capa será:

$$l_s = v_l \cdot 2t_t, \quad (5.8)$$

donde t_t es el tiempo total de ataque.

Por otro lado, el tiempo mínimo requerido para atacar una capa de espesor d será:

$$t_c = d/v_v. \quad (5.9)$$

Si $t_t > t_c$, la dimensión de la capa en contacto con el substrato también disminuye lateralmente en un valor igual a $2v_l(t_t - t_c)$ respecto a la dimensión de la figura en la fotorresina.

En la Fig. 5.8 se muestra el comido vertical y lateral de una capa sometida a un ataque en seco si durante el proceso se produce también el comido de la fotorresina. W_{min} es la dimensión inicial de la figura en la fotorresina; W_{fin} es la dimensión superficial de la figura en la capa después del ataque lateral y $t_t = t_c$ es el tiempo de ataque.

Sustituyendo (5.8) y (5.9), en (5.7) se obtiene:

$$A = (l - l_s/2d). \quad (5.10)$$

La otra magnitud importante que se observa en esta figura es W_{sub} , que es el ancho de la estructura en la superficie inferior del motivo en el material que se está atacando. Si la dimensión de la figura en la fotorresina no varía durante el ataque de la capa y $t_i = t_c$:

$$W_{fin} = W_{min} - l_s \quad (5.11)$$

$$\text{y} \quad W_{sub} = W_{min} \quad (5.12)$$

La diferencia $W_{min} - W_{fin}$ indica el ataque lateral total en la superficie de la capa, si se considera que las dimensiones de la fotorresina no se alteran durante el proceso de ataque, $a_1 = a_2 = 0$.

Para el ataque en seco, esta condición no se cumple, produciéndose el efecto que se muestra en la Fig. 5.8, donde $a = a_1 + a_2$ representa la variación de las dimensiones de la figura en la fotorresina debido a la acción del ataque seco sobre ella, durante el proceso de ataque de la capa seleccionada. En caso general:

$$W_{fin} = W_{min} - (l_s + a) \quad (5.13)$$

$$\text{y} \quad W_{sub} = W_{min} - a - l_s \cdot \left(1 - \frac{t_c}{t_t}\right) \quad (5.14)$$

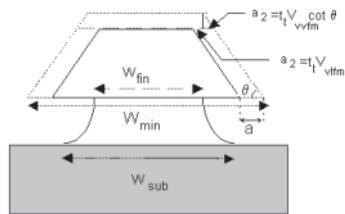


Fig. 5.8 Ataque en seco de una estructura para definir una isla considerando el ataque de la fotorresina y el sobreataque de la capa. v_{vfim} y v_{lfim} son las velocidades máximas de ataque de la fotorresina en sentido vertical y lateral respectivamente. t_i es el tiempo total que se atacó la estructura. W_{sub} y W_{fin} es la dimensión final de la isla junto al substrato y en la superficie respectivamente. W_{min} es la dimensión del motivo en la máscara.

Para obtener una dimensión final de la línea en la superficie de la oblea de W_{fin} es necesario introducir las correcciones tecnológicas en las dimensiones de la máscara, de manera que en la fotorresina, luego de revelada, la figura tenga una dimensión de W_{min} , donde:

$$W_{min} = W_{fin} + l_s + a. \quad (5.15)$$

Si se desea que la fotolitografía en la oblea consista de líneas de dimensión W_{fin} separadas también una distancia $W_s = W_{fin}$, debido a las correcciones tecnológicas, en la fotorresina estas líneas deberán tener una separación entre sí menor a su ancho e igual a:

$$W_s = W_{fin} - l_s - a. \quad (5.16)$$

A la magnitud $W_{min} + W_s$ se conoce como paso o "pitch".

Sustituyendo (5.10) en (5.16) y despejando W_{fin} se obtiene:

$$W_{fin} = W_s + 2d(1 - A) - a . \quad (5.17)$$

Esta expresión indica que la menor figura posible a obtener en la oblea W_{fin} , viene limitada por el menor valor de separación entre líneas W_s que es posible obtener con la técnica de fotolitografía utilizada, por el grado de anisotropía y depende también del espesor d de la capa a atacar.

Es evidente que a medida que se quiera que W_{fin} tienda al límite de la resolución de la fotolitografía utilizada A debe tender a la unidad.

5.5 SELECTIVIDAD Y CONTROL DE LAS DIMENSIONES

Durante el proceso de ataque en seco, además de controlar la velocidad, la repetibilidad del ataque y el comido lateral, es necesario tomar en consideración la selectividad del ataque, o lo que es lo mismo cuán rápido se ataca la capa seleccionada con respecto a la subyacente, o con respecto a la fotorresina utilizada.

Como ya se indicó, en los procesos reales, durante el ataque en seco, la fotorresina se ataca. Además todas las capas depositadas tienen una variación de su espesor de un extremo a otro de la oblea, por lo que para eliminarlas totalmente en el lugar más grueso se usaría un tiempo mayor que el requerido para eliminarla en la región más delgada y $t_l > t_c$ en ese punto. Para complicar más las cosas, las velocidades de ataque vertical y horizontal también varían en diferentes puntos de la oblea.

5.5.1 Aspectos a considerar en el ataque en seco de una capa cubierta con fotorresina

Al atacar una capa cuyo espesor promedio es d , $d(1+\delta)$ corresponderá al mayor espesor de la capa en la oblea y $d(1-\delta)$, al menor espesor, siendo $0 \leq \delta \leq 1$. Supongamos que la velocidad de ataque promedio de la capa en sentido vertical es v_{vco} y que ésta varía en el rango de $v_{vco}(1 \pm \Phi_c)$, donde Φ_c es un parámetro adimensional en el rango de $0 \leq \Phi_c \leq 1$. Considerar la condición de peor caso constituye la aproximación más conservadora para obtener la selectividad necesaria para que el ancho final de la línea en la superficie W_{fin} se encuentre dentro de los límites deseados.

La condición de peor caso corresponde a considerar que se está atacando la parte más gruesa de la capa, con la mínima velocidad de ataque, cuando la velocidad de ataque de la fotorresina es la máxima posible.

El tiempo para completar el ataque será:

$$t_c = d(1+\delta)/[v_{vco}(1-\Phi_c)]. \quad (5.18)$$

El tiempo total de ataque se puede expresar también a través del parámetro $\Delta = \frac{t_c}{t_t}$ como:

$$t_t = [d(1+\delta)(1+\Delta)]/[v_{vco}(1-\Phi_c)]. \quad (5.19)$$

La fotorresina se atacará durante todo el tiempo t_t . Si la velocidad máxima de ataque de la fotorresina en sentido vertical y lateral es v_{vfmax} y v_{lfmax}

respectivamente, los bordes de la fotorresina disminuirán en una dimensión no mayor de:

$$a/2 = [v_{vf_m} \cot \theta + v_{lf_m}] t_b \quad (5.20)$$

donde el ángulo θ se muestra en la misma figura. Este ángulo depende del método fotolitográfico utilizado. Para las alineadoras de proyección θ es del orden de los 60° , mientras que para las de contacto se acerca a los 90° .

Substituyendo (5.19) en (5.20), se puede obtener finalmente:

$$a = \frac{2v_{fm} \cdot d \cdot \left(\cot \theta + \frac{v_{lfm}}{v_{vf_m}} \right) \cdot [(1+\delta) \cdot (1+\Delta)]}{v_{vco} (1 - \Phi_c)} \quad (5.21)$$

Si se define la velocidad de ataque de la fotorresina también a través de un parámetro Φ_f similar al anteriormente definido para la capa, de manera que $v_{vf_m} = v_{vf_0}(1 + \Phi_f)$, donde v_{vf_0} es la velocidad media de ataque de la fotorresina, entonces:

$$S_{ef} = \frac{v_{vco}}{v_{vf_0}} \quad (5.22)$$

será la selectividad promedio deseada de ataque de la capa con respecto a la fotorresina que se requiere tener.

Como también

$$v_{lf_0}/v_{vf_0} = (1 - A_{fo}), \quad (5.23)$$

donde A_{fo} es el grado de anisotropía promedio del ataque de la fotorresina, la ecuación (5.21) puede ser transformada de la siguiente manera:

$$a = \frac{2d}{S_{cf}(1-\Phi_c)} \left[\cot \theta + (1 - A_f) \right] \cdot \left[(1 + \Phi_f)(1 + \delta)(1 + \Delta) \right], \quad (5.24)$$

donde

$$S_{cf} = 2dU_{cf}/a \quad (5.25)$$

y

$$U_{cf} = (1 + \Phi_f)(1 + \delta)(1 + \Delta)/(1 - \Phi_c) \quad (5.26)$$

es el factor de uniformidad que responde a la condición de peor caso.

Si se considera el caso de un ataque completamente anisotrópico en la capa, ($A=1$), donde el comido lateral se produce solamente debido al ataque de la fotorresina; si se supone un porcentaje de uniformidad en la velocidad de ataque tanto para la capa como para la fotorresina del 10%; un porcentaje de uniformidad en el espesor de ambas capas del 5% y una relación $\frac{t_c}{t_t}$ del 20%, se tendrá que:

$$\Phi_c = \Phi_f = 0.1; \delta = 0.05 \text{ y } \Delta = 0.2.$$

Substituyendo estos valores en la ecuación (5.25) se obtiene:

$$S_{cf} = 1.54[\cot \theta + (1 - A_{fo})] \cdot 2d/(W_{min} - W_{fin}). \quad (5.27)$$

La Fig. 5.9 muestra un gráfico de S_{cf} vs. $d/(W_{min} - W_{fin})$, para valores de θ de 60° y 90° en el caso hipotético de ataque totalmente isótropo y anisotropo de la fotorresina.

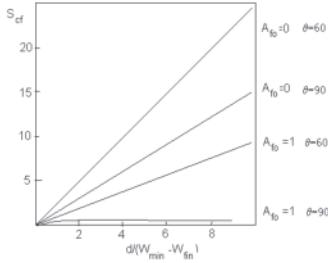


Fig. 5.9 Selectividad del ataque de la capa respecto a la fotorresina en función de $d/[W_{min}-W_{fin}]$ para diferentes θ y grado de anisotropía del ataque de la fotorresina

5.5.2 Selectividad requerida respecto a una capa subyacente a la que se está atacando.

La selectividad requerida con respecto al ataque de la capa subyacente puede obtenerse de forma análoga. De nuevo se analiza el peor caso, que corresponde a la mayor velocidad de ataque en la capa y al menor espesor de ésta a atacar.

El tiempo requerido para atacar la región de esta capa con menor espesor, considerando la mayor velocidad del proceso, será:

$$tI = d(1-\delta)/[v_{vco}(1+\Phi_c)]. \quad (5.28)$$

Si el tiempo total de ataque considerando el sobreataque se expresa por:

$$t_t = t_c(1+\Delta) \quad (5.29)$$

El mayor tiempo en que se está atacando el substrato será:

$$t_o = t_t - tI \quad (5.30)$$

Substituyendo (5.18), (5.25) y (5.26) en (5.27) se obtiene:

$$t_t - tI = \frac{d}{v_{vco}} \left[\frac{(2 + \Delta + \Delta \cdot \delta) \cdot \Phi_c + \delta(2 + \Delta) + \Delta}{(1 - \Phi_c^2)} \right] \quad (5.31)$$

Al mismo tiempo, la mayor profundidad que se come el substrato durante este ataque será

$$d_{smax} = (t_t - tI) \cdot v_{smax}, \quad (5.32)$$

donde v_{smax} es la velocidad máxima a que se ataca el substrato que puede representarse también de la forma:

$$v_{vsmax} = v_{vso} (1 + \Phi_s). \quad (5.33)$$

Si definimos como selectividad del ataque de la capa con respecto al substrato a la magnitud:

$$S_{cs} = v_{vco} / v_{vso}. \quad (5.34)$$

Sustituyendo (5.31), y (5.34) en (5.32) se obtiene:

$$S_{cs} = (d/d_{smax}) \cdot U_{cs}, \quad (5.35).$$

donde d_{smax} es la mayor penetración del ataque permisible en el substrato y

$$U_{cf} = \{ (I + \Phi_c) (I + \delta)(I + \Delta) \} / (I - \Phi_f) \quad (5.36)$$

Si el espesor y la velocidad de ataque de la capa fueran completamente uniformes, ($\delta = \Phi_c = \Phi_s = 0$), y $\Delta = 0$, la selectividad respecto al substrato no tendría importancia. En la práctica siempre resulta imprescindible que $t_r > t_c$ debido a que nunca la uniformidad en el espesor de una capa es perfecta.

En el diseño de las máscaras para la determinación de sus dimensiones es necesario conocer tanto la dimensión final que se requiere como los parámetros del ataque que se vaya utilizar durante la fotolitografía. Si el ataque es en seco, el conocimiento de los parámetros anteriormente señalados resulta fundamental para un diseño correcto de las máscaras.

Tabla 5.1 Mezclas gaseosas mas utilizadas en el ataque seco.

<i>Material de la capa a atacar</i>	<i>Mezclas gaseosas utilizadas</i>
Si	CF ₄
	CF ₄ +O ₂
	SF ₆
	SF ₆ +O ₂
	NF ₃
	Cl ₂
	CCl ₄
	CCl ₃ F
	CCl ₂ F ₂
	CClF ₃
	C ₂ F ₆
	C ₃ F ₆
	Cl ₂ F ₂
	PCl ₃
	CBrF ₃
SO ₂	CF ₄ , CF ₄ +O ₂
Si ₃ N ₄	CF ₄ +H ₂
	C ₂ F ₆
	C ₃ F ₈
	CHF ₃
Al	CCl ₄
Al-Si	CCl ₄ +Cl ₂
	SiCl ₄
	BCl ₃
Al-Cu	BCl ₃ +Cl ₂
Mo	CF ₄
W	CF ₄ +O ₂
Ti	CCl ₄
	CF ₆
GaAs	C ₂ Cl ₂ F ₂
	CHClF ₂
Au	C ₂ Cl ₂ F ₄
	Cl ₂
Pt	CF4
Cr	Cl ₂ ; CCl ₄
Cr ₂ O ₃	Cl ₂ +Ar; CCl ₄ +Ar

Tabla 5.2. Nombres comerciales de las mezclas gaseosas más utilizadas en el ataque en seco.

Mezcla gaseosa	Nombre comercial
CHClF ₂	Freón 22
CClF ₃	Freón 13
CCl ₂ F ₂	Freón 12
Cl+Cl ₂ F	Freón 21
C ₂ Cl ₂ F ₄	Freón 114
CHF ₃	Freón 223
CF ₄	Freón 14
CCl ₃ F	Freón 11
C ₂ Cl ₃ F ₃	Freón 113
C ₂ F ₆	Freón 116

5.6 BIBLIOGRAFÍA

- [1].Takuo Sugano, Applications of Plasma Processes to VLSI Technology, a Wiley-Interscience Publication, John Wiley and Sons, N.Y., 1985.
- [2].J.L.Vossen, W. Kern, Thin Film Processes II, Academic Press INC, 1991.
- [3].VLSI Electronics: Microstructure Science, Vol 8, Series, edited by N.G. Einsbruck y D.M. Brown, Academic Press, Inc. 1984.
- [4].G.S. Oehrlein, T.E.F.M. Standaertm, P.J. Matsuo, “Pattern transfer into low dielectric constant materials by high-density plasma etching, Solid State Tech, 43 (6) 125 (2000).

CAPÍTULO 6

MEDICIÓN DE RESISTIVIDAD

Una de las técnicas utilizadas para la determinación de la concentración de impurezas en substratos semiconductores o en capas epitaxiales, capas predifundidas o difundidas, es la medición de su resistividad. La técnica de medición de resistividad más utilizada es la técnica de medición por cuatro puntas. Esta técnica utiliza una cabeza con 4 puntas separadas entre sí generalmente un milímetro, soportadas sobre muelas independientes, de manera que puedan apoyar con la misma presión sobre la superficie de la oblea. Fig. 5.1. Por las dos puntas exteriores se hace pasar una corriente constante, generalmente de 1 mA, a menos que la resistividad del substrato sea demasiado alta, en cuyo caso esta corriente se puede bajar hasta 1 μ A o menos. Es conveniente suministrar esta corriente mediante un generador de corriente, para no tener que estar constantemente ajustando su valor cada vez que cambie la resistencia que ofrece la oblea al apoyarle las puntas. Esta resistencia se debe a la resistencia propia de la oblea, más la resistencia del contacto de la punta con la oblea, el cual puede ser significativo. Esta resistencia de contacto será mayor, a medida que la superficie sea más pulida. Por este motivo, en muchos casos la medición de las obleas vírgenes se realiza midiendo por su lado posterior, el cual generalmente no está pulido a espejo.

Para evitar el efecto del contacto sobre la medición del voltaje, se utilizan las dos puntas interiores, a través de las cuales no fluye corriente y solo se mide la diferencia de potencial entre ellas.

El fundamento teórico de esta medición fue estudiado en [1], e inicialmente se considera que la medición se realiza sobre un material semi-infinito cuyas fronteras exteriores pueden ser conductoras o no conductoras, y que además se cumplen las siguientes condiciones:

1. La resistividad del material es uniforme.
2. Los portadores injectados por los electrodos de corriente se recombinan cerca de ellos y no llegan a los electrodos en que se mide el voltaje.
3. Las 4 puntas se apoyan sobre una línea recta.
4. El diámetro del contacto debe ser mucho menor que la distancia entre puntas.
5. El contacto es esférico.

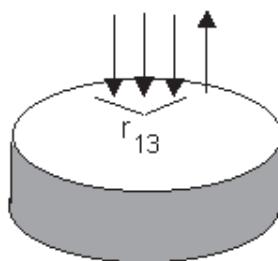


Fig. 7.1. Disposición de las puntas sobre la oblea a medir.

En caso general, si se aplica la ley de Ohms diferencial a una corriente que entra en un punto situado a la distancia r_{12} de otro, que a su vez se encuentra a la distancia r_{24} del punto por donde sale la corriente que entra, Fig.7.2, tendremos que:

$$j = \frac{1}{\rho} \cdot F, \quad (7.1)$$

donde ρ es la resistividad del material, y F es el campo eléctrico.

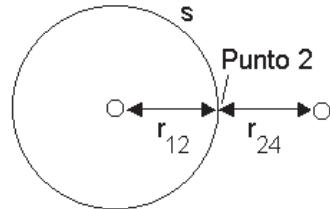


Fig. 7.2. Disposición de las puntas de corriente por donde entra y sale la corriente.

El potencial en el punto 2 producido por la corriente I que entra, será:

$$U_{2\infty} = \frac{1}{4\pi\epsilon\epsilon_0} \cdot \frac{q}{r_{12}} . \quad (7.2)$$

El campo eléctrico en la región será:

$$F = \frac{U_{2\infty}}{r_{12}} = \frac{1}{4\pi\epsilon\epsilon_0} \cdot \frac{q}{r_{12}^2} . \quad (7.3)$$

La relación entre la corriente que entra y el potencial viene dado por:

$$I = \int_S j dS = \frac{1}{\rho} \int_S F dS = \frac{U_{2\infty}}{\rho \cdot r_{12}} \cdot S , \quad (7.4)$$

donde S es la superficie de integración en la que se encuentra el punto 2, Fig.7.2, y que es igual a una semi-esfera en el substrato:

$$S = 2\pi r_{12}^2. \quad (7.5)$$

Luego

$$I = 2\pi r_{12} \cdot U_{2\infty} / \rho. \quad (7.6)$$

El potencial en el punto debido a un solo electrodo de corriente que entra, será:

$$U_{2\infty} = \rho \cdot I / [2\pi \cdot r_{12}]. \quad (7.7)$$

De igual forma, el potencial en el punto 2, producido por la corriente que sale, situada a la distancia r_{24} , de dicho punto, será:

$$U_{2\infty} = \rho \cdot I / [2\pi \cdot r_{24}]. \quad (7.8)$$

El potencial total en el punto será la superposición de ambos, por lo que será igual a:

$$U_2 = U_{2\infty} + U_{2\infty} = [\rho \cdot I / 2\pi] \cdot [1/r_{12} - 1/r_{24}]. \quad (7.9)$$

Si se llama s a distancia entre dos puntas en la cabeza del medidor de 4 puntas, $r_{12} = r_{24} = 2s$, luego:

$$U_2 = [\rho \cdot I / 2\pi] \cdot [1/2s]. \quad (7.10)$$

Análogamente, el potencial en el tercer electrodo, situado a la distancia $2s$ de la corriente que entra y a la distancia s de la que sale será:

$$U_3 = [\rho \cdot I / 2\pi] \cdot [l/2s - s] = [\rho \cdot I / 2\pi] \cdot [-1/2s] \quad (7.11)$$

y la diferencia de potencial entre los electrodos 2 y 3, será:

$$V = U_2 - U_3 = [\rho \cdot I / 2\pi] \cdot [l/s - l/2s] - [l/2s - l/s] = [\rho \cdot I / 2\pi] \cdot [l/s] \quad (7.12)$$

luego:

$$\rho = [V/I] \cdot 2\pi \cdot s \quad (7.13)$$

En la práctica pueden encontrarse diferentes casos en lo que la muestra a medir no es semi-infinita; estos fueron analizados en detalle en [1].

En el caso en que las puntas están colocadas perpendicularmente a una frontera no conductora, a una distancia que no puede considerarse suficientemente grande con respecto a la distancia entre puntas s , la afectación de esta frontera sobre el valor del potencial en el electrodo 2 o 3, puede analizarse según el método de cargas (o corrientes) imágenes según se muestra en la Fig. 7.3. Junto a la frontera, se observará el efecto de cómo si existiese otra corriente que sale, dispuesta a la misma distancia l de la frontera que la corriente real que sale. También aparece otro efecto debido a la imagen de la corriente que entra, la cual está situada a la misma distancia de la frontera que la corriente real que entra, ($l+3s$).

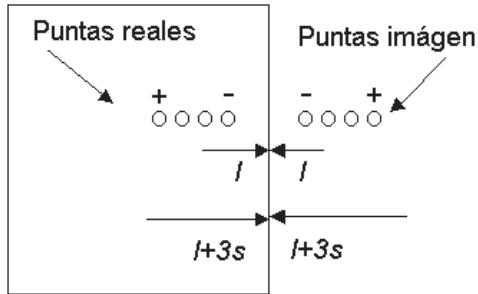


Fig. 7.3 Disposición de las 4 puntas, perpendicularmente y cercanas a una frontera no conductora.

Los potenciales U_2 y U_3 se verán modificados de la siguiente manera:

$$U_2 = [\rho \cdot I / 2\pi] \cdot [1/s - 1/(2s + 1/(2l + 5s)) - 1/(2l + 2s)]. \quad (7.14)$$

$$U_3 = [\rho \cdot I / 2\pi] \cdot [1/2s - 1/s - 1/(s + 2l) + 1/(2l + 4s)]. \quad (7.15)$$

La diferencia de potencial entre el electrodo 2 y el 3 será en este caso:

$$\begin{aligned} V &= U_2 - U_3 = \\ &= [\rho \cdot I / 2\pi] \cdot [1/s - 1/(2s + 1/(2l + 5s)) - 1/(2l + 2s) - 1/2s - 1/s - 1/(s + 2l) + 1/(2l + 4s)] = \\ &= [\rho \cdot I / 2\pi] \cdot F_2(l/s) \end{aligned} \quad (7.16)$$

$$\text{y } \rho = [V/I] \cdot 2\pi \cdot s \cdot F_2(l/s), \quad (7.17)$$

donde $F_2 (l/s)$ es una función de corrección cuyo valor se puede encontrar en [1].

De forma similar, si las puntas se colocan de forma paralela a una frontera no conductora, a una distancia no muy grande con respecto a la distancia entre puntas, Fig.7.4, la expresión para la resistividad, vuelve a quedar de la forma:

$$\rho = [V/I] \cdot 2\pi \cdot s \cdot F_3(l/s), \quad (7.18)$$

donde $F_3(l/s)$, es la función de corrección debido a la cercanía de la frontera no conductora.

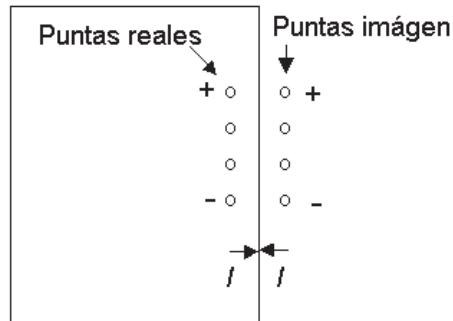


Fig. 7.4 Disposición de las puntas de forma paralela y cercanas a una frontera no conductora.

Si las puntas están colocadas perpendicularmente a la frontera, pero ésta es conductora, la función de corrección será otra, lo mismo que si están paralelas a una frontera conductora. Estas funciones también fueron calculadas en [1]. La resistividad queda en ambos casos de forma similar, respectivamente:

$$\rho = [V/I] \cdot 2\pi \cdot s \cdot F_4(l/s). \quad (7.19)$$

$$\rho = [V/I] \cdot 2\pi \cdot s \cdot F_5(l/s). \quad (7.20)$$

Si la muestra que se mide es suficientemente grande por los lados, pero su espesor no lo es, de igual forma se puede encontrar una función de corrección para el caso de fronteras no conductoras y conductoras. En ambos casos las corrientes imágenes se producirán tanto del lado de debajo de la oblea, como por encima de las fuentes reales, según se muestra en la Fig. 7.5

La expresión para la resistividad en este caso queda de la forma:

$$\rho = [V/I] \cdot 2\pi \cdot s \cdot G_7(w/s), \quad (7.21)$$

donde:

$$G_7\left(\frac{w}{s}\right) = 1 + 4 \cdot \frac{s}{w} \cdot \sum_{n=1}^{\infty} \left\{ \frac{1}{\sqrt{\left(\frac{s}{w}\right)^2 + 4n^2}} - \frac{1}{\sqrt{\left(\frac{2s}{w}\right)^2 + 4n^2}} \right\}. \quad (7.22)$$

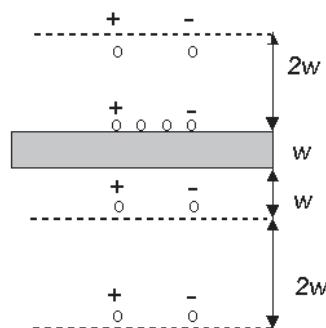


Fig. 7.5 Disposición de las puntas sobre una muestra con espesor finito con superficie inferior no conductora.

Cuando $w \ll s$, la función G_7 se aproxima a:

$$G_7(w/s) \approx (2s/w) \ln 2. \quad (7.23)$$

Sustituyendo (7.23) en (7.21):

$$\rho = 4.53 \cdot [V/I] \cdot w. \quad (7.24)$$

Este es el caso más utilizado para obleas de silicio vírgenes, difundidas o epitaxiales, donde los espesores pueden ser del orden de los 0.2 a 0.5 mm en el primer caso y entre décimas hasta decenas de μm en los otros dos casos.

Para el caso de obleas con frontera conductora, la resistividad se expresa a través de:

$$\rho = [V/I] \cdot 2\pi \cdot s \cdot G_6(w/s), \quad (7.25)$$

donde los valores para $G_6(w/s)$ se encuentran también en [1] donde las corrientes imágenes de cada frontera conductora tiene el signo contrario a la de origen.

La medición de resistividad permite controlar los procesos de predifusión y difusión, así como de implantación iónica para lo cual se utilizan las curvas de Irving [2], actualizadas por Bulucea [3-4], al tomar las aproximaciones analíticas para la resistividad de volumen en substratos dopados con boro y fósforo, según las recomendaciones del ASTM. La conductividad promedio, $\bar{\sigma}(C_s, C_B)$, de una capa superficial con concentración de impurezas C_s , en un substrato con concentración de

impurezas constante C_B , cuyo tipo de conductividad es contrario al de la capa superficial, se expresa como [2]:

$$\bar{\sigma}(C_s, C_B) = \frac{1}{\rho} = \frac{q}{x_j} \int_0^{x_j} \mu_n n(x) dx \quad (7.26)$$

A su vez, la concentración de portadores $n(x)$, puede considerarse aproximadamente igual a la concentración de impurezas activas, $C(x)-C_B$, donde $C(x)$ se describe por $C_s f(x)$ y $f(x)$ es la función que describe el perfil de impurezas. La movilidad hay que expresarla en función de la distribución de portadores, o en su lugar, en función de la totalidad de impurezas en el material, o sea, como función de $[C_s f(x)+C_B]$:

$$\mu = (1/q) \cdot \sigma [C_s \cdot f(x) - C_B] / [C_s \cdot f(x) + C_B], \quad (7.27)$$

donde $\sigma[C_s \cdot f(x) - C_B]$ describe de forma aproximada la conductividad de volumen en función de la concentración neta de impurezas.

Finalmente, la conductividad promedio de las capas superficiales, predifundidas, difundidas o implantadas, se puede calcular mediante la expresión:

$$\bar{\sigma}(C_s, C_B) = \frac{1}{x_j} \int_0^{x_j} \frac{C_s \cdot f(x) - C_B}{C_s \cdot f(x) + C_B} \cdot \sigma[C_s \cdot f(x) - C_B] dx. \quad (7.28)$$

Las curvas que relacionan la resistividad de las capas predifundidas se calculan considerando una distribución de impurezas del tipo función error complementaria. Las curvas para capas implantadas o difundidas se calculan considerando una distribución del tipo gaussiana. La posición de

la unión se determina calculando la coordenada para la cual la concentración de la impureza en la capa predifundida, difundida o implantada se iguala con la del substrato.

El procedimiento para determinar la concentración de impurezas en un substrato homogéneo de espesor w , consiste en medir el valor de resistividad mediante el medidor de 4 puntas y luego ir al gráfico que relaciona la resistividad con la concentración para el tipo de impureza, ya sea boro o fósforo.

El procedimiento para determinar la concentración superficial de impurezas de una capa predifundida, difundida o implantada en un substrato con impurezas de tipo contrario, consiste en determinar el valor de resistividad mediante el medidor de 4 puntas, para lo cual hace falta determinar la profundidad de la unión mediante otro procedimiento, por ejemplo, mediante coloración de biselos. La fórmula que se utiliza para calcular la resistividad si se cumple la condición de que $w \ll s$ es la (7.24). La concentración de impurezas en el substrato, deber haber sido determinada previamente, antes de realizar el proceso de introducción de impurezas. Luego se procede a escoger el gráfico ρ vs. C_s para la C_B dada correspondiente a la distribución de impurezas y al tipo de impurezas en cuestión. En caso de que parte de la superficie haya sido eliminada luego de la introducción de impurezas, pueden utilizarse los gráficos que se muestran en [4].

Además del concepto de resistividad, que constituye una propiedad del material y que se expresa en $\Omega\text{-cm}$, se utiliza bastante en el control de los procesos de introducción de impurezas, el término resistencia por cuadrado, o por hoja, que se expresa como Ω/\square . Este parámetro se define en una tira rectangular de material tomando en consideración que la resistencia entre los extremos de un rectángulo de área transversal con w_1 de ancho y w_2 de profundidad y de largo L se define como:

$$R = \bar{\rho} \cdot \frac{L}{w_1 \cdot w_2} \quad (7.29)$$

Si $w_1=L$, o sea para un cuadrado, la resistencia $R = \bar{\rho} \cdot \frac{L}{w_2}$, será igual a

$$4.53 \cdot \frac{V}{I},$$

o sea, al valor medido V/I , multiplicado por el factor de corrección para el caso de capas de espesor mucho menor que la separación entre puntas.

Este parámetro es cómodo para determinar el valor de resistencias en circuitos integrados producidas por líneas de difusión, ya que conocido este parámetro tecnológico puede calcularse el largo de la línea para obtener el valor de resistencia deseado.

6.1 BIBLIOGRAFÍA

- [1]. L.B. Valdés Proc."Resistivity measurements on Germanium for transistors", IRE, 42, (1954) 420.
- [2]. J.C. Irvin, " Resistivity of Bulk Silicon of Diffused Layers in Silicon", Bell System Tech. J., 41, (1962) 387.
- [3]. C. Bulucea, "" Solid State Electronics, Resistivity Curves for Diffused Layer in Silicon, using Updated Bulk Resistivity Data, Solid State Electroncis, 36 (4) (1993) 489.
- [4]. C. Bulucea, "Resistivity curves for subsurface diffused layers in silicon", Solid State Electronic, 38 (2) (1995) 367.

CAPÍTULO 7

NUEVOS DIELÉCTRICOS PARA DISPOSITIVOS MOS

7.1 LIMITACIONES DEL SiO₂ CON EL ESCALAMIENTO DE LOS DISPOSITIVOS MOS

Como ya se indicó, el óxido de silicio ha sido determinante en el desarrollo de la Tecnología Planar. Por otro lado, la disminución de la longitud del canal de los dispositivos MOS conlleva a la reducción de otras dimensiones como se indica en la Tabla 7.1, para que los dispositivos mantengan sus propiedades [1].

Esta regla de escalamiento indica que por ejemplo, al pasar de dispositivos con longitud de canal de 6 micrómetros que había en los años 70, a los 0.18 micrómetros actuales, el óxido de compuerta debe reducir su espesor de 100 nm a menos de 3 nm.

Las proyecciones hechas en 1994 [2] sobre cómo se comportaría este escalamiento del espesor del óxido se pueden ver en el gráfico de la Fig. 7.1. En ella cabe recalcar que la proyección del comportamiento del espesor del óxido al reducir la longitud del canal, presentaba una saturación hacia el valor de 3 nm.

Esta tendencia tenía su origen en el fenómeno físico que se observa en estos dieléctricos cuando su espesor se hace menor de 3 nm, y que está relacionado con la aparición del efecto de tuneldeo directo.

Tabla 7.1

Reglas para el escalamiento de los dispositivos MOS [1]

Parámetro	Factor de escalado
<i>Longitud de compuerta</i>	$1/\alpha$
<i>Ancho de compuerta</i>	$1/\alpha$
<i>Espesor de óxido de compuerta</i>	$1/\alpha$
<i>Profundidad de unión</i>	$1/\alpha$
<i>Concentración de substrato</i>	$\alpha^2/\alpha' \quad (1 < \alpha' < \alpha)$
<i>Voltaje de operación</i>	$1/\alpha$
<i>Voltaje de umbral</i>	$1/\alpha$
<i>Tiempo de propagación</i>	$1/\alpha$

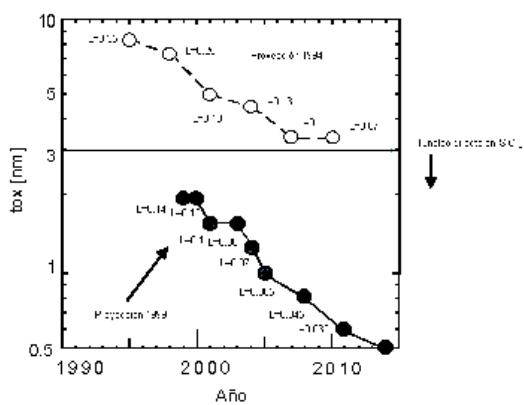


Fig. 7.1 Proyección del escalamiento del espesor del SiO₂ en los transistores MOS submicrométricos según [2].

El efecto de tuneleo directo (TD) incrementa la corriente a través de la compuerta hasta valores prohibitivos para el funcionamiento de los circuitos, incluso en áreas tan pequeñas como las correspondientes a los dispositivos submicrométricos. El valor máximo de densidad de corriente que se ha utilizado hasta el momento para fabricar dispositivos que funcionen aceptablemente es de 1 A/cm^2 , aunque la mayoría utiliza una densidad de corriente un orden menor.

Unos años después, la misma proyección del escalamiento del espesor del dieléctrico de la compuerta, ya pasaba la barrera de los 3 nm, como se puede observar en la misma Fig. 7.1. La respuesta se encontró en el uso de oxinitruros en lugar de óxido de silicio. Estos nuevos dieléctricos, con una constante dieléctrica superior al SiO_2 , permitieron introducir el concepto de espesor equivalente X_{eq} , expresado a través de:

$$X_{eq} = X_i \cdot \frac{k_{\text{SiO}_2}}{k_i}, \quad (7.1)$$

donde X_i y k_i son el espesor real del dieléctrico y su constante dieléctrica relativa al aire y k_{SiO_2} la constante dieléctrica relativa del SiO_2 .

Como se desprende de (7.1), el espesor real de la capa de oxinitruro resulta mayor que el espesor equivalente el cual sería el requerido para que el espesor del dieléctrico cumpliese con las normas de escalamiento, a medida que se disminuye el largo del canal.

Al tener realmente la capa un espesor mayor, la densidad de corriente a través de ella, resulta menor que el de una capa de SiO_2 con el espesor requerido por las reglas de escalamiento, al mismo tiempo que se observaban otras características convenientes. Sin embargo, el límite inferior de espesor utilizable volvía a aparecer cuando X_{eq} se acercaba al valor de 2 nm. Surgió entonces la idea de buscar otros dieléctricos que tuvieran mayor constante dieléctrica y que permitieran seguir

disminuyendo el espesor equivalente. A continuación mencionaremos algunos resultados que se han alcanzado con nuevos materiales dieléctricos, así como técnicas de obtención que se investigan.

7.1.1 Fenómenos físicos que limitan el uso del SiO₂ con espesor menor de 3 nm

Los mecanismos de conducción están bien establecidos para óxidos con espesor mayor de 3 nm. El efecto túnel tipo Fowler-Nordheim (FN) a través de una barrera triangular es el mecanismo responsable por el incremento de corriente, que ocurre a campos eléctricos superiores a 6×10^6 V/cm, [3].

La densidad de corriente por efecto túnel tipo FN se expresa a través de:

$$J_{FN} = C_1 \cdot F_{ox} \cdot e^{\frac{-B}{F_{ox}}}, \quad (7.2)$$

donde $C_1 = \frac{q^3 m_{Si}}{8\pi h m_{ox} \phi_b^3}$,

$$B = \frac{4\sqrt{2m_{ox}\phi_b^3}}{3\hbar q}$$

F_{ox} es el campo a través del óxido

m_{Si} y m_{ox} son la masa efectiva del electrón en silicio y en el dieléctrico respectivamente;

ϕ_b es la altura de la barrera que ve el electrón.

Al reducir el espesor del óxido, se observó un incremento sustancial de la densidad de corriente a través del óxido, respecto a los valores observados para espesores mayores. Este nuevo mecanismo de conducción se debe al tuneldeo directo, que ocurre a bajos campos eléctricos a través de una barrera rectangular suficientemente estrecha. En la Fig. 7.2 se muestran los dos mecanismos de tuneldeo mencionados.

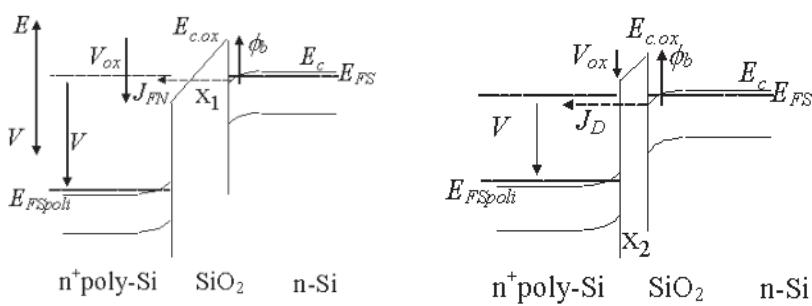


Fig. 7.2 Mecanismo de conducción en el SiO_2 a) por FN y b) por tuneldeo directo.

La probabilidad de tuneldeo para un electrón se expresa como:

$$T_t(E) = \exp\left[-\frac{2\sqrt{2m_{ox}}}{\hbar} \int_0^{x_1} \sqrt{[E_{c,ox}(x)-E]} dx\right] \quad (7.3)$$

donde $E_{c,ox}$ es la energía de la banda de conducción del dieléctrico y m_{ox} la masa efectiva del electrón en el óxido. $E_{c,ox}$ puede expresarse como:

$$E_{c,ox}(x) = E_{c,ox} - qF_{ox}x = E_c + q(\chi_S - \chi_{ox}) = E_c - \phi_b \quad (7.4)$$

χ_s y χ_{ox} son la afinidad del electrón en el semiconductor y en el dieléctrico respectivamente;

Para electrones en una región de acumulación degenerada, $E \approx E_c$ y X_1 , (Fig. 7.2a) será:

$$X_1 = \frac{(E_{c,ox} - E)}{qF_{ox}} \approx \frac{\phi_b}{qF_{ox}} \quad (7.5)$$

Si el espesor del dieléctrico es menor que X_1 , el mecanismo de tuneldeo directo será el predominante, de lo contrario predominará el mecanismo FN.

Luego de integrar entre los límites correspondientes, la probabilidad de tuneldeo para FN y TD respectivamente será:

$$T_{FN}(E) = \exp\left[-\frac{4}{3} \cdot \frac{\sqrt{2m_{ox}}}{q\hbar} \cdot \frac{(\phi_b - E)^{3/2}}{F_{ox}}\right] \quad (7.6)$$

$$T_D(E) = \exp\left[-\frac{4}{3} \cdot \frac{\sqrt{2m_{ox}}}{q\hbar} \cdot \frac{(\phi_b - E)^{3/2} - (\phi_b - E - qV_{ox})^{3/2}}{F_{ox}}\right] \quad (7.7)$$

Donde $\phi_b = q \cdot (\chi_s - \chi_{ox}) = E_{c,ox} - E_c$ (7.8)

La corriente de tuneldeo directo de electrones desde un semiconductor en acumulación y degenerado hacia la compuerta se puede expresar como:

$$J_D = \frac{AF_{ox}^2}{\left[1 - \sqrt{\left(\frac{\phi_b - qV_{ox}}{\phi_b}\right)}\right]^2} \exp\left[-\frac{B}{F_{ox}} \frac{\phi_b^{3/2} - (\phi_b - qV_{ox})^{3/2}}{\phi_b^{3/2}}\right] \cdot F_2 \quad (7.9)$$

$$\text{donde } F_2 = \left\{ 1 - \exp\left[-\frac{3}{2} \cdot \frac{B}{F_{ox}} \frac{\phi_b^{1/2} - (\phi_b - qV_{ox})^{1/2}}{\phi_b^{3/2}}\right] \cdot E_{FS} \right\} \quad (7.10)$$

En la Fig. 7.3 se muestra el comportamiento de la densidad de corriente con el voltaje aplicado al dieléctrico, para diferentes espesores de éste. Se compara el comportamiento de la corriente calculada cuando tiene lugar el mecanismo de TD y el de FN con datos experimentales. Puede observarse que a medida que el espesor de SiO_2 se acerca a los 3 nm, el mecanismo de conducción predominante pasa a ser el TD.

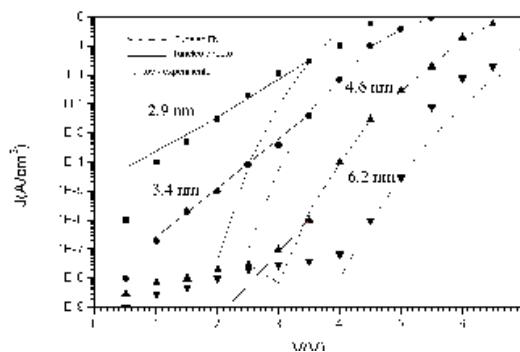


Fig. 7.3 Comportamiento de la corriente a través del dieléctrico cuando predomina cada uno de los dos mecanismos de tuneldeo [4].

En la Fig. 7.4 se muestra la densidad de corriente esperada a través de una capa de oxinitruro con espesor equivalente X_{eq} comparada con la densidad de corriente a través de una capa de SiO_2 con igual espesor, donde se observa que utilizando el oxinitruro se puede disminuir el espesor hasta $X_{eq}=2.3$ nm y obtener la misma densidad de corriente que con 3 nm de SiO_2 , del orden de 0.1 mA/cm^2 .

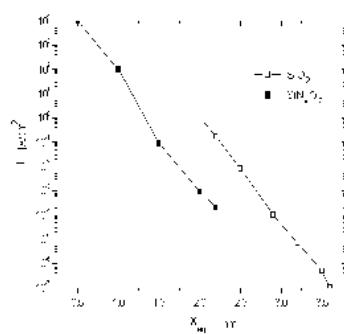


Fig. 7.4 Densidad de corriente para un óxido de espesor X_o y para un oxinitruro de espesor equivalente X_{eq} , con $V=2 \text{ V}$.

Si el tuneleo ocurre desde una región de inversión, los portadores requeridos deben ser suministrados de alguna manera. Si la estructura es un capacitor MOS, el mecanismo será la generación-recombinación, que a su vez será el mecanismo limitante de la corriente, (Fig. 7.5). En transistores, los portadores son suministrados por las regiones de drenaje y fuente.

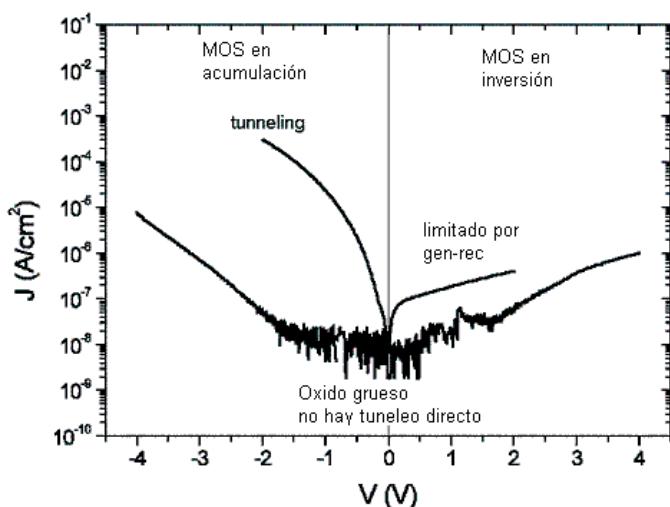


Fig. 7.5 Efecto de limitación de la corriente de tuneleo directo a través del dieléctrico en un capacitor MOS por falta de portadores suficientes, cuando estos son suministrados por generación recombinación en un substrato tipo P con concentración de impurezas del orden de 10^{15} cm^{-3} .

7.1.2 Nuevos dieléctricos substitutos del SiO_2

Por lo dicho anteriormente, la primera intención fue buscar nuevos dieléctricos que tuviesen una constante dieléctrica, k , lo mayor posible y que no reaccionaran con el silicio. De los primeros trabajos teóricos se estableció que los mejores óxidos desde estos dos puntos de vista, parecían ser los óxidos de metales no marcados en la Fig. 7.7.

Estudios complementarios indicaron también que incrementar la k implicaba una reducción del alto de la barrera compuerta-dieléctrico, por lo que había que buscar un compromiso que implicaba que el potencial de la barrera no se redujera por debajo del máximo voltaje de trabajo del

dispositivo. En la Tabla 7.2 [5] se muestran valores de altura de la barrera y permitividad, ancho de banda prohibida y afinidad del electrón para diferentes dieléctricos respecto al silicio. El valor de permitividad relativa aquí mostrado, ϵ_{∞} , es el correspondiente a alta frecuencia, por ello el valor es diferente a la permitividad a baja frecuencia, mas conocido como constante dieléctrica, k. Por ejemplo para el SiO₂ $\epsilon_{\infty}=2.25$ y k=3.9.

The figure shows a periodic table where elements are categorized based on their interaction with silicon (Si). Elements in white boxes are theoretically stable with Si, while elements in shaded boxes react with Si. A legend at the top right provides the key:

- En blanco los que son teóricamente estables con el Si** (White boxes)
- Reaccionan con el Si** (Shaded boxes)
- Estudiados anteriormente** (White boxes)

Specifically, the table highlights the following groups:

- Elements that react with Si (Shaded):** H, Li, Be, Na, Mg, K, Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, Ge, As, Se, Br, Kr, Rh, Sr, Y, Zr, Nb, Mo, Tc, Ru, Rb, Pd, Ag, Cd, In, Sn, Sb, Te, I, Xe, Cs, Ba, † Hf, Ta, W, Re, Os, Ir, Pt, Au, Hg, Tl, Pb, Bi, Po, At, Rn, Fr, Ra, ‡ Rf, Ha, Sg, Ns, Hs, Mt.
- Elements studied previously (White):** He, B, C, N, O, F, Ne, Al, Si, P, S, Cl, Ar.
- Elements theoretically stable with Si (White):** La, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Ac, Th, Pa, U, Np, Pu, Am, Cm, Bk, Cf, Es, Fm, Md, No, Lr.

Fig. 7.6 Tabla de elementos químicos indicando los que no reaccionan con el silicio y los ya estudiados [6].

Según las reglas de escalamiento indicadas en la Tabla 7.1, para una longitud de canal de 0.1 μm , el voltaje máximo de compuerta V_G debería ser menor de 1 V. Esto no es del todo aplicable en la práctica, pues el valor del voltaje de umbral, V_T , no puede ser reducido ilimitadamente y debe garantizarse que V_G-V_T sea mayor de 0.5 V, por lo que resulta difícil trabajar con valores de V_G menores de 1 V. Por ello buscar dieléctricos sustitutos del SiO₂, con k mayor de 30, no resulta conveniente.

Tabla 7.2

Disminución del alto de la barrera con el incremento de k [5]

Material	Gap [eV]	ϵ_m	Altura de Barrera [eV]	Afinidad Electrónica [eV]
SiO ₂	9	2.25	3.5	0.9
Si ₃ N ₄	5.3	4.1	2.4	2.1
Ta ₂ O ₃	.44	4.48	0.3	3.3
BaTiO ₃	3.3	7.1	-0.1	3.9
BaZrO ₃	5.3	4	0.8	2.6
TiO ₂	3.05	7.8	0.005	3.9
HfO ₂	6	3.4	1.5	2.5
Al ₂ O ₃	8.8	4.4	2.8	1
Y ₂ O ₃	6	4	2.3	2
La ₂ O ₃	6	3.8	2.3	2
ZrSiO ₄	7.5	5.3	1.5	2.4
SrBi ₂ Ta ₂	4.1	4.8	0	3.5
ZrO ₂	5.8		1.4	2.5

Entre los óxidos más prometedores se estudian los indicados en la Tabla 7.3, que no reaccionan con el silicio y su k es menor de 30.

Tabla 7.3

Óxidos metálicos que no reaccionan con el silicio y tiene una k entre 7 y 30

Si ₃ N ₄	Al ₂ O ₃	k: 7 - 10
ZrO ₂ -SiO ₂	HfO ₂ -SiO ₂	La ₂ O ₃ -SiO ₂ k: 10 - 20
ZrO ₂	HfO ₂	La ₂ O ₃ k: 20 -30

. Las características de las estructuras con dieléctricos de alta k se investigan tanto cuando el dieléctrico se deposita directamente sobre el silicio como cuando se deposita sobre una capa previa de SiO₂. Esta capa

previa puede ser debida, por ejemplo, al óxido nativo o al formado químicamente luego de la limpieza o por alguna otra técnica.

7.2 OTROS PROBLEMAS ASOCIADOS A LOS ÓXIDOS DE COMPUERTA ULTRAFINOS

La utilización de óxidos de silicio ultra finos (menores de 3 nm) o de dieléctricos alternativos como compuerta de dispositivos MOS presentan otros problemas además de los planteados. Entre ellos se pueden mencionar los siguientes:

1. Penetración de impurezas a través del dieléctrico de compuerta.
2. Empeoramiento de las características de la interfaz.
3. Incremento del valor del voltaje de umbral por efectos cuánticos.
4. Generación de trampas que provocan ruptura dependiente del tiempo (TDB).

7.2.1 Penetración de impurezas a través del dieléctrico

Los dispositivos submicrométricos requieren de varios procesos de implantación para ajustar la concentración de impurezas tanto en el canal como en la compuerta, con vistas a lograr voltajes de umbral pequeños y lo más simétricos posible.

Debido al pequeño espesor del óxido de compuerta, no es extraño esperar que las impurezas implantadas puedan atravesar el óxido, cuando no se desea y alterar la concentración de impurezas en el canal.

En la práctica se ha observado que el problema más crítico se presenta con el boro. Normalmente los procesos de implantación se hacen con muy bajas energías de implantación pero se requiere del proceso posterior de activación de las impurezas que se realiza mediante el tratamiento térmico

rápido (RTA) por pocos segundos. Aún así, la penetración del boro se produce y es un efecto indeseable que requiere un control adicional.

Si en lugar de utilizar SiO_2 se usa oxinitruro, este problema se reduce, pues el oxinitruro detiene mucho más los iones de boro. Por lo tanto esta es una ventaja adicional del uso de oxinitruros. Otros óxidos metálicos pueden presentar un efecto similar con coeficiente de difusión del boro en el óxido.

7.2.2 Empeoramiento de las características de la interfaz

Como es conocido, las características de la interfaz $\text{SiO}_2\text{-Si}$ pueden ser excelentes, cosa que no ocurre con otros dieléctricos como el propio oxinitruro. La interfaz del oxinitruro-silicio presenta una densidad de cargas mayor que la que se puede lograr con el $\text{SiO}_2\text{-Si}$. Esto podría ser un problema importante, sin embargo la magnitud del voltaje de banda plana V_{FB} se expresa como (ver 1.48):

$$V_{FB} = \Phi_{GS} - \frac{Q_{ss} + Q_{cf} + Q_i}{C_i} \quad (7.11)$$

donde $C_i = \frac{\epsilon_0 k_2}{X_i}$ es la capacitancia por unidad de área del dieléctrico de compuerta y X_i su espesor real. Como X_i es muy pequeño, el V_{FB} estará definido fundamentalmente por Φ_{GS} aunque la densidad total de carga $Q_{ss} + Q_{cf}$ sea mayor dentro de determinados límites. Por ejemplo, una carga total de 10^{12} cm^{-2} con un espesor de 3 nm aumentaría el valor de V_{FB} en tan solo 0.05 V. Por ello el oxinitruro y posiblemente otros óxidos metálicos con no tan buenas propiedades de interfaz con el silicio podrían ser utilizados como óxidos de compuerta para espesores inferiores a los 3 nm.

7.2.3 Incremento del valor del voltaje de umbral por efectos cuánticos

Las reglas de escalamiento (Tabla 7.1) indican que al disminuir la longitud del canal hay que incrementar la concentración del substrato. El incremento de la concentración de impurezas del substrato hace que el ancho de la zona de carga espacial sea menor y que la densidad de carga en la zona de inversión mayor. Por ello el campo en la superficie puede ser lo suficientemente alto para producir confinamiento de los portadores en la región de inversión. Si esto ocurre, también se produce una cuantificación de los niveles de energía perpendicularmente a la superficie del semiconductor. La banda de conducción se divide en sub-bandas y el valor de la banda prohibida en la región de inversión aumenta según la expresión [7]:

$$E_g^{QM} = E_g + \frac{13}{9} \Delta E \quad (7.12)$$

donde E_g es el valor de la banda prohibida sin alteración y ΔE el incremento en la posición del fondo de la banda de conducción por este efecto. Este incremento de la banda prohibida modifica la concentración intrínseca de portadores cuyo nuevo valor puede expresarse como:

$$n_i^{QM} = N_i \cdot \exp\left[-\frac{E_g^{QM} - E_g}{2kT}\right], \quad (7.13)$$

y para obtener la misma concentración de portadores se requiere un campo eléctrico mas intenso en la superficie y la región de inversión penetra más.

Por ejemplo en [8] se reporta que el incremento en la penetración puede ser de 1.2 nm cuando ocurre este efecto cuántico. A su vez, esta penetración implica que se produce un efecto como si el espesor del dieléctrico fuera mayor, donde el espesor equivalente se puede expresar ahora como:

$$X_{eq}^{QM} = X_{eq} + \frac{k_{SiO2}}{k_{si}} \cdot \Delta Z \quad (7.14).$$

Para un espesor de óxido de 2 nm, un incremento de $\frac{3.9}{11.7}1.2 = 0.4 \text{ nm}$ representa el 20% del espesor total.

6.2.4. Generación de trampas que provocan ruptura dependiente del tiempo (TDB)

La inyección de electrones hacia el dieléctrico produce la generación de trampas en el mismo. Como el espesor es tan delgado, la probabilidad de que estas trampas se sitúen a lo largo de una dirección formando caminos por donde pueden pasar los electrones a lo largo del dieléctrico mediante el mecanismo de saltos entre trampas se hace factible. Estas trampas dan lugar a la aparición de un efecto de ruptura antes de lo debido, el cual a su vez depende del tiempo que fluye la corriente a través de estas trampas. Es por ello que al este mecanismo de ruptura se le llama ruptura dependiente del tiempo (TDB) y es otro efecto indeseable en las estructuras MOS con dieléctricos ultra finos.

7.3 METODOS DE OBTENCIÓN DE DIELÉCTRICOS DE ALTA K

Los dieléctricos de alta k se han obtenido utilizando, tanto métodos de depósito físico, como químico. Entre ellos se pueden mencionar los que se muestran en la Tabla 7.4. Cualquiera de las técnicas que se utilicen debe permitir el depósito de los espesores deseados de manera controlada.

Dentro de las técnicas de CVD se han reportado variantes como CVD, CVD en plasma remoto, CVD de capas atómicas, (ALCVD) CVD con tratamiento rápido (RTCVD), MOCVD y fotoCVD. Además se han reportado técnicas de depósito físicas, como pulverización catódica, epitaxia de haces molecular y ablación por laser. También se puede utilizar la oxidación en plasma.

Por fotoCVD se han obtenido capas de SiO₂ con buenas características a una presión de 1.5 Torr en presencia de SiH₄ y N₂O en relación de 7.5 a 30 sccm y 200 °C, excitando a través de una ventana de zafiro de 1.5 mm de espesor con una lámpara de deuterio de 30 W [10]. También se han obtenido capas ultra finas con buena calidad, de manera repetible, por oxidación en plasma de O₂, a temperatura ambiente, 1 Torr y 100 W en un plasma de RF de barril, tipo inductivo [10].

Para la obtención de las capas ultra finas de oxinitruro el método de oxinitridación se realiza a temperatura del orden de 900 °C en atmósfera de N₂O o NO, en hornos convencionales o por RTP. La nitridación del SiO₂ se realiza en ambiente de NH₃, luego de formada la capa de SiO₂. Se incorpora H₂ y presenta una alta densidad de carga en la interfaz [11]. También puede realizarse dando un tratamiento térmico en presencia de NO o N₂O a temperatura del orden de los 1000 °C. En este caso no se observa presencia de H₂ y la densidad de carga es menor.

El depósito por CVD se hace en presencia de N₂O o NO. Las características de las capas dependen fuertemente de las condiciones de depósito.

Tabla 7.4
Dieléctricos ultra finos de compuerta y métodos de obtención

Material	Proceso utilizado	Reportado en
SiO ₂	Oxidación térmica a T<900 °C y por RTO	
SiO ₂	fotoCVD	[9]
SiO ₂	Oxidación en plasma a T<200 °C	[10]
SiO _x N _y	Depósito por CVD	[11,12]
SiO _x N _y	Nitridación de SiO ₂	[11,12,13]
SiO _x N _y	Oxinitridación de Si	[11,14]
HfO ₂	Pulverización catódica y RTO	[15,16]
HfO ₂	Pulverización catódica RF	[17]
HfO ₂	CVD	[18]
HfO ₂	PECVD	[19]
TiO ₂	MOCVD	[20]
TiO ₂	ALCVD	[21]
TiO ₂	Crecimiento epitaxial	[22]
TiO ₂	Pulverización catódica y oxidación en plasma	[23]
ZrO ₂	MOCVD	[24]

7.4 NUEVOS DIELÉCTRICOS PARA AISLAMIENTO ENTRE LÍNEAS DE METALIZACIÓN

7.4.1 Características de la Metalización Multiniveles

La densidad de integración de los C.I. está definida no solo por las dimensiones de los transistores, sino por su **interconexión**. Hay dos caminos para acomodar la cantidad de líneas de interconexión requeridas, una de ellas es reducir el paso (pitch), o sea, una magnitud formada por la dimensión de la línea mas la dimensión del espacio entre ellas; la otra es aumentar el número de niveles por encima de la superficie del circuito, por donde pueden ir las líneas.

Hay dos regímenes de interconexión de los elementos de un circuito, que deben ser considerados: la interconexión local y la global. La interconexión local es aquella utilizada dentro de una misma celda o bloque que une transistores vecinos. Para este fin es muy utilizado el polisilicio y los siliciuros. La interconexión global es aquella utilizada para la conexión entre los diferentes bloques, y en ella se busca la menor resistividad posible del metal, así como la menor capacidad entre líneas de interconexión situadas en diferentes niveles y separadas entre sí por algún dieléctrico. Estos multiniveles de interconexión se requieren para lograr la densidad de integración requerida para los CI VLSI y ULSI.

Actualmente se usan hasta 6 ó 7 niveles de interconexión para satisfacer las necesidades de interconexión en C.I. con tecnologías submicrométricas [24]. Un aspecto muy importante de la interconexión es el material que sirve como aislante entre las líneas y el material utilizado para las interconexiones. Tradicionalmente han sido utilizados el SiO₂ y el aluminio y sus eutéticos, aunque para dispositivos submicrométricos con unión superficial, el aluminio no puede ser utilizado debido a la difusión en forma

de pinchos (spikes) que puede penetrar decenas de micras y cortocircuitar la unión como se muestra en la Fig. 7.7.

Al utilizar varias líneas de interconexión se establece el esquema “metal dieléctrico metal”, en el que puede observarse la existencia de capacitancias parásitas entre los dos niveles de metales y entre líneas adyacentes. Además cada metal ofrece una resistencia al paso de la corriente que crece con la longitud de la línea. Estos efectos combinados producen un retardo en la transmisión de las señales debido a la existencia de un circuito RC distribuido, ver Fig. 7.8.



Fig. 7.7. Contacto de aluminio con silicio, mostrando la penetración del

Si se utiliza la aproximación del capacitor de placas paralelas, al reducir las dimensiones de las líneas se esperaría un reducción en la capacitancia; sin embargo, esta aproximación desprecia los efectos de borde que se hacen más importantes a medida que se reducen las dimensiones y con ello la reducción real de la capacitancia es menor que lo esperado. Además del efecto en la capacitancia, la reducción de las dimensiones de las líneas aumenta su resistencia. Por ello al reducir las dimensiones no se reducen los tiempos de retardo como podría esperarse.

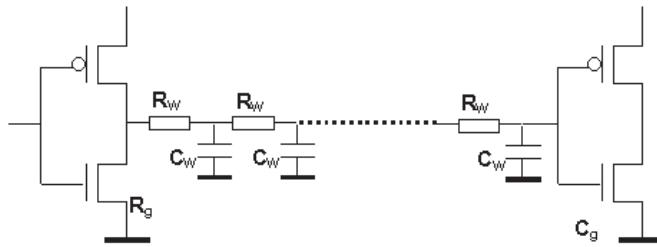


Fig. 7.8. Elementos que proporcionan retardo en la interconexión.

Para las interconexiones globales hay dos parámetros que deben ser tomados en cuenta, los tiempos de retardo y la electromigración. El retardo puede ser calculado como la suma de los retardos de interconexión y los de compuerta, a través de la fórmula de Sakurai [24].

$$T_{pd} = CIR_wC_w + R_wC_g + R_gC_w + R_gC_g \quad (7.15)$$

donde C_1 es una constante de proporcionalidad.

La capacitancia total de la línea esta dada como la suma de las capacitancias entre líneas cercanas (C_L) y la capacitancia entre niveles (C_V), Fig. 7.9.

Se ha determinado que al incrementar la razón de aspecto (espesor/ancho de la línea) se reduce el tiempo de retardo, de forma que si se tiene definido el valor del paso $L+S$, se puede obtener la relación de aspecto óptima que produzca el menor retardo, Fig. 7.10.

Cuando las líneas de interconexión son cortas, el factor que limita al retardo es la capacitancia parásita. Para reducirla y así mejorar la respuesta se debe reducir la razón de aspecto. Para estos casos se ha encontrado que la razón de 1 es la más adecuada. Por otro lado, cuando las líneas de

interconexión son largas, el factor que limita es su resistencia y por ello se debe buscar aumentar la razón de aspecto a fin de reducir la resistencia total de la línea. La razón más adecuada para estas condiciones es de 2; el usar una razón mayor no mejora en mucho el retardo.

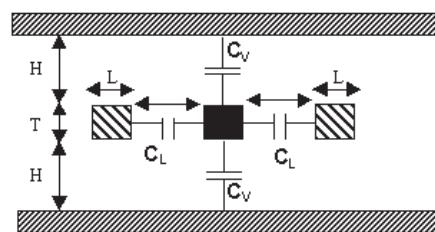


Fig. 7.9 Capacitancias parásitas en las líneas de interconexión: C_V capacitancia entre niveles, C_L capacitancia entre líneas cercanas; L , ancho de la línea; T espesor y H distancia entre una línea en un nivel y otra en el nivel superior o inferior.

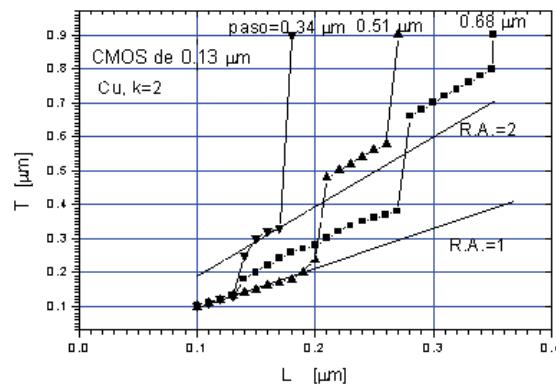


Fig. 7.10 Relación entre la razón de aspecto (R.A.) y el paso [24].

Otro problema presente en las interconexiones y que es acentuado a medida que se reducen las dimensiones de las líneas, es la electromigración. La electromigración es el movimiento de los átomos del metal como resultado de la transferencia de energía de los electrones en conducción. Esto produce un flujo neto de átomos. Como resultado, se producen acumulaciones y estrechamientos a lo largo de la línea. Los estrechamientos pueden producir la ruptura de la línea, mientras que las acumulaciones producen que la distancia efectiva de separación se vea reducida y en un momento dado pueden producir un cortocircuito. Estas características reducen la fiabilidad del circuito.

Una forma de reducir el problema de la electromigración en el aluminio fue mediante la adición de 1 a 4% de cobre al aluminio. Esta aleación puede permitir mayores densidades de corriente (del orden de 10^6 A/cm²) dando fiabilidad a la línea. Un problema que se presenta con estas aleaciones es la formación de AlCu₂ en el metal, el cual es responsable de la corrosión que se ve frecuentemente en este tipo de aleaciones. Otro problema presente es el ataque en seco de la capa mediante plasma. El aluminio generalmente es atacado mediante un plasma de cloro, pero la adición de cobre dificulta dicho ataque, quedando residuos de cobre después del mismo.

Otro material muy atractivo para el uso en interconexiones es el cobre debido a su baja resistividad (menor que la del aluminio) y también debido a que presenta una mayor resistencia a la electromigración [25]. Sin embargo un problema muy delicado que presenta el cobre, es su muy alta difusión, aún a relativamente bajas temperaturas, en el SiO₂ que es el dieléctrico más utilizado entre los niveles de metalización. Debido a esta difusión tan rápida, pueden llegar restos de cobre a la región activa de los transistores produciendo una contaminación muy fuerte que deteriora sus características. Para poder utilizar el cobre como metal de interconexión es

indispensable la formación de barreras contra la difusión del cobre en el SiO_2 .

Una forma de evitar la difusión de cobre, es mediante el uso de una fina capa de nitruro de titanio (TiN) [25] la cual es expuesta a aire para reforzar su capacidad de frenar la difusión pues el oxígeno bloquea las rutas de difusión en el TiN. Cuando no se utiliza la capa de TiN, el SiO_2 presenta una concentración superficial de cobre cercana a 10^{13} cm^{-2} , pero con la utilización del TiN prácticamente no se observa que el cobre penetre. También se usan capas de TaN con igual objetivo.

El proceso básico para la utilización de niveles de metalización con cobre comienza depositando una película de SiO_2 . En las técnicas mas actuales el SiO_2 se sustituye por un dieléctrico de baja constante dieléctrica o por multicapas de SiO_2 , SiN y dieléctrico de bajo k, como se muestra en la Fig. 7.11. En estas se abren, por fotolitografía, los huecos de contacto en las regiones en el silicio o niveles de metalización inferiores a las que se quiere hacer contacto con la línea de metalización. Posteriormente se deposita una fina capa de TiN y se expone a aire. Seguidamente se deposita una o varias capas de Cu lo suficientemente gruesa para que llene los pozos. El exceso de Cu, se elimina mediante un pulido mecánico- químico (CMP). Si se requiere otro nivel de metalización se deposita otra película de TiN y sobre esta otra capa de SiO_2 . Al depositar el último nivel se deposita una película de nitruro de silicio que actúa como pasivación.

Aunado al uso de cobre, los sistemas de interconexión están buscando la utilización de nuevos dieléctricos cuyas constantes dieléctricas (k) sean menores que el SiO_2 . Estas nuevas combinaciones (cobre-dieléctrico de bajo k) logran reducir considerablemente los tiempos de retardo obtenidos en las metalizaciones por multiniveles, con esto se pueden lograr circuitos que operen a una mayor velocidad En la Fig. 7.12 se muestra este efecto.

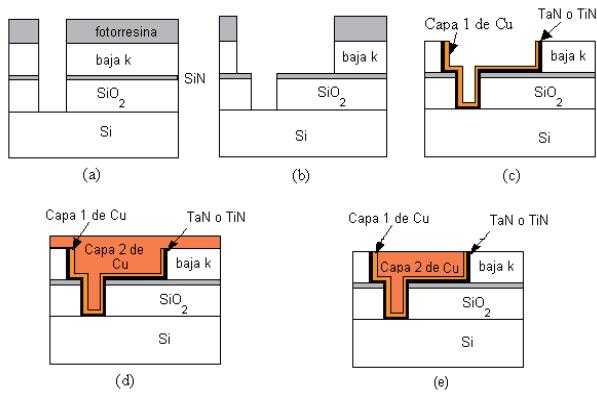


Fig. 7.11 Proceso de metalización utilizando cobre: a) Fotolitografía para abrir hueco de contacto hasta el silicio; b) Fotolitografía para definir ancho de línea de interconexión; c) Depósito de TiN o de TaN para aumentar la resistencia a la difusión del cobre hacia los dieléctricos; d) Depósito de Cu; e) La estructura después del pulido mecánico-químico [25].

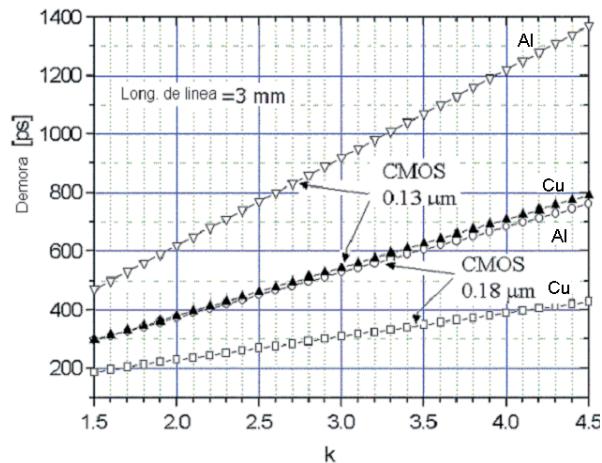


Fig. 7.12. Efecto de nuevos materiales respecto al tiempo de retardo [25].

En la Tabla 7.5 se muestran algunos materiales con baja k, que pueden ser utilizados perspectivamente como dieléctricos entre niveles de metalización.

Tabla 7.5
Dieléctricos de baja k y métodos de obtención

Material	Óxido de Silicio	Vidrio Silicato fluorinado	Polimida	Polimida fluorinada	Carbón amorfo fluorinado	Xerogele (Óxido de Silicio poroso)
Fórmula	SiO_2	$(\text{SiO}_2)_x$ $(\text{SiO}_3\text{F}_2)_{1-x}$	-	-	a-CF	SiO_2
k	3.9-4.5	3.2-4	3.1-3.4	2.5-2.9	2.3	1.1-2.2
Método de depósito	PECVD	PECVD	Spin on	PECVD	PECVD	Spin on

7.5 BIBLIOGRAFÍA

- [1]. G. Baccarani, M. R. Woredeman, R. H. Dennard, “Generalized scaling theory and its application to a $\frac{1}{4}$ micron MOSFET design”, IEEE Tran on ED 31 (1984) 452-462.
- [2]. International Technoloy Roadmap for Semiconductors, <http://public.itrs.net>
- [3]. M. Lenzlinger, E. H. Snow “Fowler-Nordheim tunneling into thermally grown SiO_2 .” J. of Applied Physics. 40 (1) (1969) 278-83.

- [4]. M. Depas, B. V., P. W. Mertens, R. L. Van Meirhaeghe, M. M. Heyns, “Determination of tunnelling parameters in ultra-thin oxide layer poly-Si/SiO₂/Si structures.” Solid-State Electronics **38** (1995) 1465-71.
- [5]. J. Robertson, “ Band structure and offsets of high k dielectrics on Si”, Appl. Surf Science 190 (2002) p2-10.
- [6]. H. Iwai, S. Ohmi, Future COMS technology below 0.1 μm, Proceedings of the XV International Conference on Microelectronics and Packaging SBMicro 2000, p- 2, Manaus, Amazona, Brazil, Sept. 18-24 2000.
- [7]. M. J. Van Dort, P. H. Woerlee, A. J. Walker. “A simple model for quantisation effects in heavily-doped silicon MOSFETs at inversion conditions.” Solid-State Electronics 37 (3) (1994) 411-4.
- [8]. Y. Ohkura, “Quantum effects in Si n-MOS inversion layer at high substrate concentration.” Solid-State Electronics 33 (12) 1581-5 1990.
- [9]. V.M. Sanchez, J. Munguía, M. Estrada, Very Low Photo-CVD deposition rate for high quality thin SiO₂ layers, Proceedings of the Fourth IEEE International Caracas Conference on devices Circuits and Systems, p. D044-1to 4, April 17-19 2002, Aruba, Dutch Caribbean.
- [10].J. C. Tinoco, M. Estrada, “Very low plasma oxidation rate for high quality ultrathin sio₂ layers, accepted for presentation at Sbmicro 2002.
- [11].E. P. Gusev, H. C. Lu, E. L. Garfunkel, T. Gustafsson, M. L. Green, “Growth and characterization of ultrathin nitrided silicon oxide films.” IBM J. Res. Develop. 43 (3) (1999) 265-85.
- [12].D. M. Diatezua, P. A. Tirry, R. Caudano. “Control of radio-frequency-plasma process to improve the reproducibility of silicon oxynitride thin film preparation.” J. Vac. Sci. Technol. A 13 (4) (1995) 2099-104.

- [13].R.I. Hedge, B. Maiti, P.J, Tobin, “Growth and film characteristics of N₂O and NO oxynitride gate and tunnel dielectrics”, J. Electrochem. Soc. 144 (3) (1997) 1981.
- [14].K. A. Ellis, R. A. Buhrman, “Nitrous oxide (N₂O) processing for silicon oxynitride gate dielectrics.” IBM J. Res. Develop. 43 (3) (1999) 287-300.
- [15].L. Kang, K. Onishih, Y. Jeon, B.H. Lee, C. Kang, W. Qi, R. Nieh, S. Gopalan, R. Choi, J.C. Lee, “MOSFET devices with polysilicon on single-layer HfO₂ high-K dielectrics”, Tech. Digest Inter. Electron Dev. Meeting p. 35, San Francisco, USA, Dec. 2000.
- [16].B.H. Lee, R. Choi, L. Kang, S. Gopalan, R. Nieh, K. Onishi, Y. Jeon, W. Qi, C. Kang, J.C. Lee, “ Characteristics of TaN gate MOSFE with ultrathin hafnium oxide (8A-12A)”, Tech. Digest Inter. Electron Dev. Meeting p. 39, San Francisco, USA, Dec. 2000.
- [17].C.T. Hsu, Y. K. Su, M. Yokoyama, “High dielectric constant of RF sputtered HfO₂ thin films, Jap. J. Appl. Phys. 31, (1992) 2501.
- [18].S.J. Lee, H.F. Luan, W. P. Bai, C.H. Lee, T.S. Jeon, Y. Senzaki, D. Roberts, D. L. Kwong, High quality ultra thin CVD HfO₂ gate stack with poly-Si gatge electrode”, Tech. Digest Inter. Electron Dev. Meeting p. 31, San Francisco, USA, Dec. 2000.
- [19].H.S. Kim, D. C. Gilmer, D.L. Polla “Leakage current and electrical breakdown in metalo-organic chemical vapor deposited TiO₂ dielectrics on silicon structures”, Appl. Phys. Lett. 60, (1996) 3860.
- [20].El.L. Lakomaa, S. Haukka, T. Suntola, “ Atomic layer growth of TiO₂ on silica”, Appl. Surf. Scie. 60/61 (1992) 742.
- [21].M.B. Lee, M. Kawasaki, H. Koinuma, “ Formation and characterization of epitaxial TiO₂ and BaTiO₃/TiO₂ films on silicon substrate, “Jpn. J. Appl. Phys. 34 (1995) 808.

- [22].J.C. Tinoco, M. Estrada, G. Romero, "Formation and characterization of SiO₂ and TiO₂ by room temperature plasma oxidation", to be published.
- [23]. T. S. Jeon, J. M. White. a. D. L. Kwong "Thermal stability of ultrathin ZrO₂ films prepared by chemical vapor deposition on Si(100)." *Appl. Phys. Lett.* 78 (3) (2001) 368-70.
- [24]. K. Yamashita and S.Odanaka. "Interconnect Scaling Scenario Using a Chip Level Interconnect Model" *IEEE Trans. Electron Devices*, 47 (1) (2000). 90-97.
- [25]. N. Awaya, H. Inokawa, E. Yamamoto, Y. Okazaki, M. Miyake and T. Kobayashi. "Evaluation of a copper metalization process and the electrical characteristics of copper-interconneted quarter-micron CMOS" *IEEE Trans. Electron Devices*, vol. 43, no. 8 p. 1206-1211 (1996).

CAPÍTULO 8

EL PROCESAMIENTO TÉRMICO RÁPIDO: SU APLICACIÓN EN LA MICROELECTRÓNICA

8.1 INTRODUCCIÓN

En los últimos 25 años la microelectrónica ha tenido un crecimiento impetuoso siguiendo la ley de Moore, o sea duplicando la densidad de componentes en los circuitos integrados de memorias cada 2 años. El paso a los equipos de procesamiento individual de las obleas en un número grande de procesos tecnológicos, la inclusión de dispositivos para el control de fin de proceso, la automatización de los procesos, entre otros factores, ha ayudado en gran medida a alcanzar estos resultados. Sin embargo, un grupo de procesos claves en la Tecnología Planar, entre ellos las oxidaciones y los tratamientos térmicos a alta temperatura (TTAT), se han mantenido utilizando los hornos horizontales o verticales y el procesamiento de obleas por grupos, (batchs), a pesar de ser conocidos desde principios del 1970 las ventajas de lo que se llamó el procesamiento térmico rápido (Rapid Thermal Processing o RTP). El RTP significó la reducción de los tiempos de procesamiento e incluso la aparición de nuevos efectos beneficiosos para el procesamiento de los dispositivos. En ese momento el RTP se investigaba para realizar los TTAT y se denominó tratamiento térmico rápido (Rapid Thermal Annealing, RTA). Los desarrollos posteriores en los equipos de RTP han permitido también la realización de oxidaciones rápidas (RTO) y el depósito de capas (RTCVD).

El proyecto MMST (Microelectronic Manufacturing Science and Technology) desarrollado por la Texas Instruments con ayuda de la Universidad de Stanford [1], las proyecciones hasta el año 2010 en desarrollo tecnológico de la Industria de Semiconductores de los EE.UU., [2], y otros proyectos europeos [3], se propusieron los siguientes objetivos:

- Reducir el área con condiciones ambientales de clase 1 a un mínimo;
- Realizar el procesamiento de las obleas de forma individual en todas las etapas tecnológicas, lo cual resulta casi imprescindible dado las dimensiones que las mismas ya alcanzan [4].
- Reducir el tiempo total de procesamiento, que dura en la actualidad hasta 45-60 días;
- Tener en cuenta los requerimientos de la tecnología submicrométrica, en cuanto a obtener capas dieléctricas y semiconductoras con espesores menores a los 3 nm; mantener las profundidades de penetración de impurezas por debajo de 1 μm .
- Reducir los costos tanto de las áreas productivas como de las de desarrollo.
- Permitir la automatización total de los procesos de que consta la secuencia tecnológica.

Para lograr estos objetivos, el uso del RTP en todos los pasos tecnológicos posibles resulta fundamental [5].

8.2 CARACTERÍSTICAS DE LOS EQUIPOS DE RTP

Los equipos para RTP deben cumplimentar los siguientes requisitos [5]:

- Lograr un calentamiento uniforme de toda la muestra, tanto durante los procesos de calentamiento y enfriamiento, (cuya

duración es de decenas de segundos), como durante el régimen de temperatura constante, cuya duración máxima es generalmente de unos 10 minutos;

- Poder medir la temperatura en forma continua y precisa en distintos puntos de la muestra, mediante un sensor que pueda ser utilizado en el control y automatización del proceso[6];
- Mantener frías las paredes de la campana donde está la muestra;
- Lograr un control dinámico de todo el proceso [7];
- Trabajar con los diferentes gases que se requieren en los procesos;
- Permitir la automatización de los procesos;
- Ser factibles de integrar en estructuras de procesamiento modular flexible [7].

En términos generales, un equipo de RPT pertenece a la categoría de equipos de pared fría, lo que permite reducir la introducción de impurezas indeseables durante los procesos de altas temperaturas. Los elementos para el control de los flujos de gases, así como los utilizados en detectores de fin de proceso, cargadores automáticos y otros dispositivos de control y automatización no difieren mucho de los utilizados en otros sistemas convencionales.

El calentamiento de las obleas se realiza utilizando lámparas de luz, del tipo de tungsteno o de halógeno colocadas por fuera de la cámara donde se coloca la oblea.

En la Fig. 8.1 se muestra el esquema típico de un equipo de RTP, con sus principales elementos: 1) Las fuentes de luz,(lámparas de tungsteno-halógeno de 1 y 2 kW dispuestas en dos a cuatro círculos concéntricos); 2) pantalla reflectora concentradora de la luz sobre la muestra; 3) sistema de

enfriamiento de las lámparas; 4) ventana de cuarzo para separar la zona donde están las lámparas de la cámara donde se coloca la muestra; 5) sensores de temperatura (pirómetros); 6) soporte de la muestra; 7) entrada y salida de gases utilizados en el procesamiento; 8) sistema de vacío para la campana donde está la muestra 9) esclusas para entrada y salida de obleas.

Para alcanzar la temperatura deseada, en el tiempo deseado y con la uniformidad requerida en toda la muestra, ha sido necesario encontrar solución a cómo disponer las lámparas, cuales deben ser sus características de radiación y su potencia, y muy en especial diseñar los bloques para el control de esta última [8].

La medición de la temperatura de la oblea con la precisión requerida ha resultado una tarea difícil pero necesaria, ya que resulta fundamental para poder controlar la potencia de las lámparas y por lo tanto controlar la temperatura de la oblea. Se determinó que las técnicas pirométricas de medición de temperatura permiten un mejor control del proceso cuando varía la geometría de las muestras, que el uso de otros sensores [6].

Además de los sensores, se requieren mecanismos de control que utilizan algoritmos basados en modelos del sistema a controlar. Estos modelos, deben tomar en cuenta la dinámica de las lámparas y de los sensores; las leyes físicas del calentamiento; las reflexiones en las paredes, los flujos de gases y la dinámica de la oblea que se procesa. Otro elemento a tener en cuenta es la forma y tamaño de la muestra, ya que la absorción y reflexión varían. La cantidad de trabajos dedicados a la modelación del comportamiento de la temperatura de la oblea en función de los parámetros antes mencionados es muy numerosa, [9], indicando la importancia de este problema. Este bloque de control constituye el corazón del equipo de RTP.

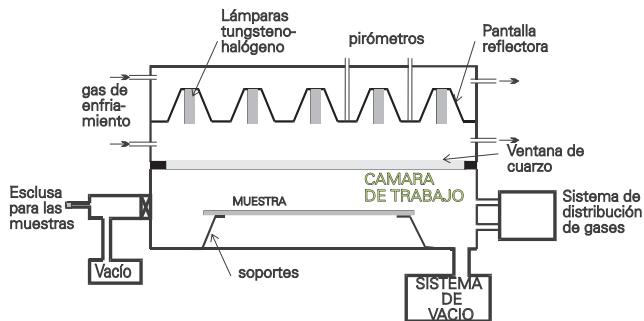


Fig. 8.1 Esquema de una cámara para el Procesamiento Térmico Rápido (RTP), con iluminación por un solo lado y los sistemas de entrada/salida de muestras, de vacío y de distribución de gases. La geometría es circular, con una lámpara central y el resto en círculos con control independiente.

La distribución uniforme de temperatura en la muestra es importante, para generar el menor número posible de imperfecciones de la red y dislocaciones [10], así como para garantizar una velocidad uniforme en los diferentes procesos con dependencia térmica. Se requiere un control que garantice una uniformidad en la temperatura del orden o mejor que los 2 °C, así como los regímenes adecuados de variación de temperatura durante los transientes.

La *Fig. 8.2* muestra un diagrama de temperatura-tiempo típico para los hornos de RTP. En la actualidad, se alcanzan los siguientes parámetros: velocidades de subida y caída de la temperatura de 0-200 °C/s; rango de temperatura de trabajo desde 250 a 1250 °C, controlada con una precisión mejor de ± 2 °C en toda la muestra; tiempo máximo de procesamiento de 10 min.; presiones de trabajo desde 5 mTorr a 760 Torr.

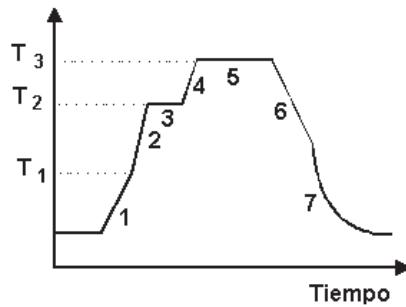


Fig. 8.2 Diagrama típico temperatura-tiempo: 1. Primera rampa (de calentamiento); 2. Segunda rampa (de calentamiento); 3. Platón de precalentamiento; 4. Tercera rampa de calentamiento; 5. Plato de procesamiento, entre 350 - 1000 °C; 7. Cuarta rampa (de enfriamiento), Enfriamiento libre.

8.3 OXIDACIÓN TÉRMICA RÁPIDA (RTO)

La oxidación es uno de los procesos tecnológicos más importantes en la fabricación de dispositivos semiconductores, tanto por las propiedades de las capas de óxido, como de su interfaz con el semiconductor. Una de las aplicaciones mas extendidas del RTO es el crecimiento de óxidos finos menores de 15 nm [11].

Se ha demostrado que por medio de la RTO se pueden obtener óxidos de mejor calidad que los crecidos en hornos convencionales [12]. Algunos estudios, [12], consideran que poder subir a mayor temperatura, sin gran redistribución de las impurezas por lo corto del tiempo de oxidación, permite obtener mejores resultados en el crecimiento de óxidos finos, tanto en la interfaz, como en las características de los mismos. Se ha observado que las tensiones de la estructura pueden ser menores que las obtenidas por técnicas convencionales.

La oxidación puede realizarse con oxígeno seco y húmedo a temperaturas entre 950 y 1000 °C. Como detector de fin de proceso se utiliza un elipsómetro, que mide *in situ* el espesor de la capa que se crece. En ambos casos el tiempo total de procesamiento es de menos de 5 min. Por ejemplo en [13] se crece por RTO el óxido de compuerta de 7.5 nm para circuitos integrados VLSI obteniendo bajas corrientes de fuga y densidad de defectos, así como voltajes de ruptura del óxido para campos eléctricos mayores de 10 MV/cm (1 V/nm). La oxidación húmeda por RTO se utiliza también para los óxidos de enmascaramiento y los de sacrificio.

Se reportan también resultados muy favorables en la oxidación de silicio poroso para fotodetectores [15], así como oxidaciones realizadas en ambiente de N₂O [16].

8.4 TRATAMIENTOS TÉRMICOS RÁPIDOS (RTA)

El tratamiento térmico rápido, RTA fue el primer tipo de procesamiento rápido realizado, pues los equipos tienen menos requerimientos, ya que solamente utilizan gases inertes. El RTA se usa en las mismas aplicaciones que el TTAT e incluso han aparecido otras nuevas, como por ejemplo su utilización en:

- Formación y tratamiento de los contactos con siliciuros (por ejemplo Ti_iN/TiSi₂);
- Tratamientos de los óxidos de compuerta, obteniéndose mejores resultados [12];
- Activación de los iones implantados, sin variar prácticamente los perfiles de distribución de impurezas;
- Sinterización de los contactos Al-Si eliminando los efectos de interdifusión Al-Si (“spiking”) [17];

- Pasivación de la interfaz Si-SiO₂; disminución de las cargas positivas presentes en el óxido e interfaz SiO₂-Si.
- Realizar interconexiones con Cu en sistemas multiniveles, evitando que este llegue al substrato en los tratamientos térmicos posteriores al depósito [18].
- Obtención de perfiles muy abruptos y poco profundos, así como resistencias de emisor bajas, necesarias a la tecnología bipolar avanzada [19].

8.5 CRECIMIENTO DE CAPAS POR RTP (RTCVD)

Las técnicas de RTP se aplican también en la actualidad a los procesos de depósito químico, surgiendo así el RTCVD (Rapid Thermal CVD) [20]. Ya resulta convencional el uso del RTCVD para el depósito de: nitruro de silicio; silicio amorfo; polisilicio; óxido de silicio; metales refractarios, como W, [7,14]. El único tipo de CVD que no ha sido posible realizar por RTCVD es el PECVD, debido a la presencia del plasma. De todas maneras las características de los equipos convencionales de PECVD (pared fría y procesamiento individual), son fácilmente integrables a la estructura modular flexible de esta nueva concepción. Algo similar ocurre con los equipos de ataque en seco.

Las técnicas de RTCVD también han permitido obtener las capas tradicionalmente crecidas por LPCVD y APCVD, e incluso mejorar sus parámetros, así como dar lugar a capas nuevas o con nuevas características [21].

La utilización de estos procesos de RTP ha permitido mejorar las características de los circuitos integrados, celdas solares y dispositivos discretos de todo tipo, tanto en Si, como en otros materiales

semiconductores [22]. Desde el punto de vista industrial y económico el RTP introduce también cambios conceptuales y disminución de costos.

Una fábrica de dispositivos microelectrónicos que usa RTP se diferencia de una convencional en los parámetros que se indican en la Tabla 8.1.

Tabla 8.1 Diferencias entre fábricas que utilizan procesos convencionales y procesos de RTP.

Parámetro	Fábrica convencional	Fábrica con RTP
Equipamiento	grupo+ individual	todo individual
Procesos térmicos	hornos horizontales	todo RTP
Depósito de capas	ATCVD, LPCVD, PECVD	RTCVD y PECVD
Ataque químico	húmedo y seco en grupo e individual	seco individual
Limpieza de obleas	húmeda por grupos	húmeda y seca individual
Transporte de obleas	por cargadores a presión atmosférica	por cargadores a baja presión
Clase de áreas limpias	100 a 10000	<1 a 10000
Área limpia	50000 a 100000 pie ³	3000 a 5000 pie ²
Mínimo ciclo cerrado	4 a 16 semanas	3 días
Costo por fábrica	mas de 700MUS\$	30MUS\$ a 50MUS\$

Esta nueva concepción resulta también muy importante para las áreas de desarrollo. La introducción del RTP permite concebir laboratorios mucho menos costosos, bastante más flexibles y con muchos menos requerimientos en lo que a grado de limpieza en las áreas, se refiere.

Por último, el RTP es aplicable no sólo a la tecnología del silicio, sino a cualquier otro material y en la actualidad resulta un terreno fértil para trabajar en muy variados campos.

8.6 BIBLIOGRAFÍA

- [1]. R.R.Doering, “The MMST program: An Overview”, Sol. State Tech., 37 (1)
- [2]. International Technoloy Roadmap for Semiconductors, editions 1995, 1999 and 2001, <http://public.itrs.net>
- [3]. M. Van Rossum, “Key technologies for Future IC Manufacturing: A View from Europe”, IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop. ASMC ‘92 Proceedings, p.211-215, 1992
- [4]. M.M.Moslehi, C.J. Davis, A. Paranjpe, L.A. Velo, H.N. Najm, C. Schaper, T. Breedijk, Y. J. Lee, D. Anderson, “Single-wafer processing tools for agile semiconductor production”, Solid State Technology, .37 (1) (1994) 35-44.
- [5]. M. Moslehi, L. Velo, A. Paranjpe, R. Chapman, J. Kuehne, J.L. Yong, S. Huang, C. Schaper, T. Breedijk, D. Yin, D. Amderson, C. Davis, “Microelectronics Manufacturing Science and Technology (MMST): Single-wafer RTP-based 0,35 μm CMOS IC Fabrication”, International Electron Device Meeting 1993, Technical Digest, p.649-652, 1993.
- [6]. S. Yu, Y. Sorrel, W.J. Keither, “Temperature Control Strategy for RTP Systems”, Proceeding of the tenth Biennal University/Government/Industry Microelectronics Symposium, p. 229-234, 1993.
- [7]. C. Sharper, M.Moslehi, K.Sarasnat, T.Kailath, “Control of MMST RTP: Repeatability, Uniformity and Integration for Flexible

- Manufacturing”, IEEE Trans. on Semicond. Manufacturing, 7 (1994) 202-219.
- [8]. Y.M. Cho, A. Paulraj, T. Kailath, X. Guanghan, “A Contribution to Optimal Lamp Design in RTP”, IEEE Tran on Semic. Manuf., 7 (1) (1994) 34-41.
 - [9]. R.S. Gyurcsik, R.K. Cavin, F.Y. Sorrell, “Model-based Control of Semiconductor Processing Equipment: RTP Example”, Conference Proceedings of 1993 International Conference on Syystem, Man and Cybernetics. Systems Engineering in the Service of Humans, 2 (1993) 13-18.
 - [10]. M.M.Moslehi, “Process Uniformity and Slip Dislocation Patterns in Linearly Ramped Temperature Transient Rapid Thermal Processing of Silicon”, IEEE Trans. on Semicond. Manufacturing, 2 (4) (1989)130-140.
 - [11]. J.Nulman, J.P.Krusius, A.Gat, “Rapid Thermal processing of thin gate dielectrics. Oxidation of Silicon”, IEEE Electron Devices Lett., 6 (5) (1985) 205-207.
 - [12]. L. Fonseca, F. Campadabal, “Effect of RTP/furnace Processing on the Minority Carrier Lifetime in Very Thin Oxide MOS Capacitors”, Sol. State Elect., 37, 1, pp 115-117, 1994.
 - [13]. L. Fonseca, F.Campadabal, “Breakdown Characteristics of RTO 10 nm SiO₂ Films Grown at Different Temperatures”, IEEE Electron Device Letters, 15 (11) (1994) 449-451.
 - [14]. M.M.Moslehi, A.Paranjpe, L.A.Velo, J.Kuehne, “RTP: Key to future semiconductor fabrication”, Solid State Technology, 37 (5) (1994) 37-45.
 - [15]. C.Tsai, K.H.Li, J.C.Campbell, “Rapid Thermal Oxidised Porous Si Photodetectors”, Electronics Letters, 29 (1) (1993) 134-137.

- [16]. G.Eftekhari, “Effects of oxide thickness and oxidation parameters on the electrical characteristics of thin oxide grown by Rapid Thermal Oxidation of Si in N₂O”, J.Electrochem. Soc., 141 (11) (1994) 3222-3225.
- [17]. W.Chen, Z.Li, H.W.Kraner, “Application of the Rapid Thermal Process: Sintering the Sputtered Aluminum/Silicon Contact in Silicon Detector Fabrication”, IEEE Trans. on Nuclear Science, 39 (4) (1992).
- [18]. H.Miyazaki, H. Kojima, A. Hiaiwa, Y. Homma, “Improved Phosphosilicate Glass Passivation Against Cu Contamination Using the Rapid Thermal Annealing Process”, J. Electroch. Soc., 141 (3) (1994) 734-737.
- [19]. A.G. Oneill, R.D. Barlow, R.G. Biswas, P.J. Phillips, S. Taylor, A. Gundlach, “Diffusion of Delta Doped Boron in Silicon Following Oxidation”, Electronics Letters, 29 (3) (1993) 263-264.
- [20]. M.C.Ozturk, Y.Sorrell, J.J.Wortman, F. S. Johnson, D.T. Grider, “Manufacturability Issues in Rapid Thermal Chemical Vapor Deposition”, IEEE Trans. on Semicond. Manuf., 4 (2) (1991) 155-165.
- [21]. Z.Q. Yao, H.B. Harrison, S. Dimitijev, Y.T. Yeow, “The Electrical Properties of Sub-5-nm Oxynitride Dielectrics Prepared in a Nitric oxide Ambient Using RTP”, IEEE Elect. Dev. Lett., 15 (12) (1994) 516-518.
- [22]. Katz, S.N.G. Chu, B.E. Weir, C.R. Abernathy, W.S. Hobson, S.J. Pearton, W. Savin, “Rapid Isothermal Procesisng of Pt/ti Contacts to P-Type III-V Binary and Related Ternary Materials, IEEE tran on ED, 39, (1) (1992) 184-192.

CAPÍTULO 9

PROCESOS FOTOLITOGRÁFICOS APLICADOS A LA MICROELECTRÓNICA

9.1 INTRODUCCIÓN

El término **litografía** proviene de lithos (piedra), y literalmente significa grabado en piedra. Desde 1978, se utiliza para nombrar a algún proceso que permita la transferencia de un grabado original a otro lugar.

En la tecnología de fabricación de los dispositivos semiconductores se requiere la transferencia de figuras geométricas previamente definidas en una placa de vidrio o cuarzo fundido, a la superficie de una oblea de material semiconductor, lo cual se realiza por métodos fotográficos, y se conoce como **fotolitografía**.

Los procesos fotolitográficos definen las dimensiones mínimas de los dispositivos semiconductores y por lo tanto la densidad de componentes en un chip, así como la velocidad de operación de los mismos. Desde el comienzo de la **tecnología planar** a principios de los años 60, el desarrollo tecnológico en esta rama ha sido impresionante, pasando de decenas de micrómetros en la anchura mínima de líneas, a las actuales dimensiones submicrométricas.

Los procesos y elementos fundamentales que intervienen en la fotolitografía son:

1. El diseño, fabricación y reproducción de las máscaras que contienen las figuras geométricas a transferir.
2. La transferencia de figuras de las máscaras al silicio, y la superposición de las nuevas figuras a transferir con las previamente transferidas a la oblea, (alineación).

3. Los procesos de definición de las figuras en la superficie del semiconductor.

Para la fabricación de un circuito se requiere la repetición de estos procesos de fotolitografía muchas veces en la oblea, unas 15-20 veces como promedio. Por su grado de precisión y por el número de veces que se repiten, los procesos fotolitográficos definen el rendimiento por oblea en la fabricación de los circuitos integrados, o sea, el porcentaje de dispositivos en cada oblea, que cumplen con las especificaciones.

La secuencia de pasos necesarios para hacer un fotograbado incluye:

1. Diseño y fabricación de las máscaras.

Fabricación por alguno de los métodos que se explican a continuación, del conjunto de máscaras, cada una de las cuales contiene una o varias veces repetidas, las figuras geométricas que es necesario transferir a la oblea, antes de realizar el proceso tecnológico al cual están asociadas. El conjunto de figuras geométricas, también llamado topología, correspondiente a cada uno de los procesos fotolitográficos requeridos para la realización del circuito integrado, se definen durante el proceso de diseño de dicho circuito, y dependen de la secuencia tecnológica que se usará en su fabricación, así como de las funciones que realizará el circuito.

2. Recubrimiento de la superficie de la oblea de silicio con una resina fotosensible o **fotorresina**.
3. Superposición de las figuras contenidas en cada máscara con las grabadas en el silicio anteriormente en la oblea. Este proceso se conoce como **alineación**.
4. Sensibilización de la fotorresina.
5. Apertura de ventanas en la fotorresina mediante el proceso de revelado en el cual, las partes sensibilizadas (si la fotorresina es positiva) o de las partes no sensibilizadas (si es negativa), se eliminan.

6. Ataque de la superficie de la oblea a través de las ventanas dejadas en la fotorresina, para el grabado de forma definitiva en alguna capa presente sobre la oblea de silicio, de las figuras que se desea transferir.
7. Eliminación de la fotorresina y limpieza del semiconductor.

Existen tres principios físicos para sensibilizar la fotorresina:

- Fotolitografía óptica.
- Fotolitografía con haz de electrones;
- Fotolitografía con rayos X.

9.2 FOTOLITOGRAFÍA ÓPTICA

La fotolitografía óptica es aquella en que la sensibilización de la fotorresina se hace mediante luz ultravioleta. Los sistemas ópticos y el material de la máscara, pueden ser de vidrio o de cuarzo, según la longitud de onda utilizada.

9.2.1 Tipos de máscaras

Para fabricar un circuito integrado (C.I.), se requiere realizar diferentes procesos tecnológicos en partes específicas de la superficie del semiconductor. La selección de estas partes o zonas se realiza mediante la fotolitografía con ayuda de las máscaras. El número de máscaras a utilizar puede ser mayor de 20, según la tecnología de fabricación. El proceso de diseño de la máscara es en la actualidad totalmente computarizado (CAD-Computer Aided Design), y el resultado final del mismo se presenta en forma de algún soporte magnético que contiene la información necesaria

para controlar automáticamente al primero de los equipos a utilizar durante la fabricación de las máscaras. Este paso hace uso del CAM o fabricación con ayuda de la computadora (Computer Aided Manufacturing).

Las máscaras para fotolitografía óptica, son generalmente un soporte de vidrio plano o de cuarzo fundido, de un alto grado de planitud, recubierta de una o varias capas para grabar las figuras de la topología que posteriormente serán transferidas a la oblea. Estas capas pueden ser:

- **Emulsión fotográfica;** similar a una película fotográfica de alta resolución, que se sensibiliza y revela en la forma convencional. Permite alcanzar una resolución de hasta $5 \mu\text{m}$; presentan problemas con la definición de los bordes.
- **Metálica;** la capa metálica puede ser generalmente de cromo, de óxido de hierro o de polisilicio, y está recubierta con una capa de fotorresina sensible a la luz ultravioleta que se expone y revela en las regiones donde se desea eliminar para conformar las figuras de la topología, dejando abiertas ventanas. A través de estas ventanas se ataca la capa metálica, transfiriendo así las figuras de la fotorresina a la capa metálica en cuestión. Estas capas son opacas a la luz ultravioleta y en general también al visible, con la excepción de las capas de óxido de hierro que son semitransparentes a la luz visible. Con cualquiera de ellas se pueden obtener resoluciones submicrométricas.

Para fabricar las máscaras utilizadas en la fotolitografía óptica hay que pasar por varias de las siguientes etapas según el proceso de fabricación de máscaras que se utilice:

- Fabricación de retículos o de máscaras de primera reducción;
- Repetición y reducción de los motivos;
- Fabricación de máscaras patrones;

- Fabricación de máscaras de trabajo;

9.2.2 Métodos de fabricación de máscaras de primera y segunda reducción

Hay dos métodos básicos de preparación de los retículos, el de **reducción fotográfica** y el que se basa en el uso de un equipo llamado **generador de patrones**.

Método de reducción fotográfica. Este método surgió en los años 70, y actualmente no tiene utilización en la industria. Una vez diseñado el circuito que se desea fabricar, la topología final se prepara en un formato con las coordenadas de cada figura que se desea transferir al retículo, de manera que pueda ser entendido por un coordinatógrafo, generalmente utilizando un soporte magnético. El coordinatógrafo en una mesa semejante a una mesa de dibujo, que controla el desplazamiento mediante motores de paso de una cuchilla de corte, en las direcciones “x” y “y”, así como el giro a determinado ángulo. El coordinatógrafo realiza el corte sobre un acetato especial llamado “**mylar**”, el cual está recubierto de una película delgada de color rojo, opaca a la luz ultravioleta (UV). Como resultado final, la cuchilla define mediante el corte de la película roja superficial del “mylar”, las figuras geométricas de que consta la topología final del diseño. Al terminar los cortes, la película roja se despega del acetato en las regiones donde deben quedar ventanas. Las dimensiones de las figuras que se transfieren al “mylar” se aumentan generalmente unas 250 veces con respecto a las dimensiones finales que se desea que tengan en el retículo.

Una vez preparado el mylar, se pasa al proceso denominado de **primera reducción**, que consiste en obtener una fotografía reducida, por ejemplo,

10 veces (10x) del mylar. A esta fotografía se le llama **máscara de primera reducción**. Para esto se requieren equipos fotográficos de mucha resolución y precisión, así como de una pantalla de dimensiones de más de un metro por lado, con una iluminación muy potente y uniforme en toda su área.

Segunda reducción y repetición.

Una vez obtenida la máscara de primera reducción, se pasa a un nuevo proceso de fotografía, para lograr esta vez, una reducción que en total sea igual al aumento con que se definieron las figuras en el “mylar”. En el ejemplo puesto, esta reducción sería de 25x.

La imagen reducida de la máscara de primera reducción se proyecta sobre una nueva placa fotográfica, que una vez terminada se llamará **máscara patrón**. Se repite tantas veces como quepa a lo largo de los ejes “x” y “y”, dejando un espacio fijo entre cada proyección que se denomina calle y que sirve para realizar el corte de los circuitos en la oblea durante el proceso de su independización antes del corte y soldadura. Este proceso de reducción fotográfica, exposición y repetición del motivo para generar la máscara patrón, se realiza en un equipo llamado fotorrepetidor o “step and repeat”.

9.2.3 Método de fabricación de retículos y máscara patrón a través de generador de patrones

Este método también consta de 2 etapas: la preparación del retículo y la preparación de la máscara patrón. En este caso, el soporte magnético controla a un equipo llamado generador de patrones, el cual consta de una rendija con ancho y largo, así como de su posición angular, ajustables. La

luz que atraviesa por la rendija, se proyecta y expone ciertas partes de la superficie de una placa fotográfica, que recibe el nombre de retículo, donde las figuras tienen dimensiones mayores, por ejemplo de 10x, 5x o 4x, respecto a las dimensiones finales que se desean en la máscara patrón. Cada figura se define mediante la exposición de rectángulos yuxtapuestos, según se muestra en la Fig. 9.1

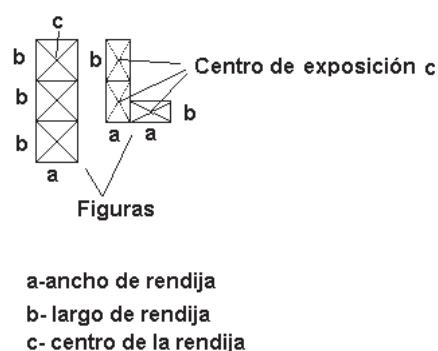


Fig. 9.1: Conformación de una figura por la exposición de un generador de patrones, mediante la yuxtaposición de rectángulos expuestos a través de una rendija de ancho a, largo b y centro situado en la posición c.

Para fabricar el retículo, la placa de vidrio cubierta por una emulsión fotográfica o por una capa metálica cubierta a su vez por fotorresina, se va moviendo con respecto al haz de luz que va sensibilizando la fotorresina que la recubre. Según la complejidad del circuito, la placa de vidrio puede quedar dividida en millones de rectángulos que forman las figuras, por lo que el tiempo que se demora en posicionar el centro de la rendija, definir su ancho y largo, y efectuar la exposición, puede tomar en suma varias horas de trabajo. En el caso en que se utilicen placas cubiertas por una emulsión fotográfica, el tiempo de sensibilización es del orden de milésimas de

segundo, mientras que en el caso de placas con recubrimiento metálico, cubierto a su vez de fotorresina, se requieren décimas de segundo. Para una línea cuya dimensión final en la máscara se desea que sea de 1 μm , en el retículo de 10X se trazará una línea de 10 μm , y en el de 5X es de 5 μm .

Estos retículos se utilizan para fabricar las máscaras patrón usando la reducción fotográfica y repetición del motivo por medio del fotorrepetidor, que se describió en el párrafo anterior.

Las dimensiones de las máscaras patrón fueron creciendo desde unos 50 mm por lado, hasta unos 150 mm por lado. Sin embargo, dada la complejidad que surgió al tener que realizar estas reducciones fotográficas en áreas tan grandes con un mínimo de distorsión, se regresó de nuevo a que la región útil de trabajo de las máscaras fuera de unos 75 mm de lado, donde se realiza la reducción fotográfica del retículo hasta alcanzar las dimensiones finales. Estas máscaras, con los motivos localizados en un área pequeña al centro, se utilizan para exponer obleas de mucho mayor diámetro, mediante las **alineadoras de exposición y repetición**, donde los motivos contenidos en la máscara se exponen y repiten, y a veces también se reducen una vez mas, directamente sobre la oblea (step on the wafer).

Máscaras de Trabajo. Las máscaras de trabajo, se producen mediante copias por contacto de las máscaras patrón. Esto es necesario ya que las máscaras de trabajo se dañan con el uso, por lo que tienen que ser sustituidas frecuentemente en el proceso productivo. La frecuencia de cambio e inspección depende del material de las máscaras y del tipo de alineadora utilizada. Cuando se hace la fotolitografía por contacto se deben cambiar cada 25 alineaciones las de emulsión y cada 50 a 100 las máscaras metálicas. Si se usan alineadoras de aproximación el número posible de veces a utilizar aumenta. Si se usan alineadoras de proyección, prácticamente las máscaras no se dañan durante el proceso de fotolitografía.

En la fabricación de las máscaras patrón y las de trabajo, los procesos se hacen en forma individual y no en grupos. Para el comido del cromo, por ejemplo, se puede utilizar el ataque húmedo o el seco. El proceso de fabricación de un juego de máscaras patrón constituye una de las etapas más costosas en la realización de un C.I.

9.3 FOTORRESINAS

Las fotorresinas para la fotolitografía óptica, son polímeros fotosensibles a la luz ultravioleta en el rango de 450 nm a menos de 150 nm. Una característica general de las mismas es que deben permitir grabar con alta resolución las figuras geométricas que posteriormente se transferirán a la oblea. Deben tener buena adherencia a la superficie de la capa donde se hará la fotolitografía y ser estables durante el ataque de dicha capa.

9.3.1 Tipos de fotorresinas

Hay dos tipos diferentes de fotorresinas, las negativas y las positivas:

Negativas. En este caso la luz incidente activa un proceso que hace al polímero más resistente en la parte iluminada que en la no iluminada. El revelado se realiza con un producto especial que elimina la parte de la capa que no se iluminó. Como las figuras que se graban en el circuito son el complemento de las figuras oscuras en la máscara, el proceso produce un negativo de la imagen original Fig. 9.2 a.

Positivas. En este tipo de fotorresinas las regiones iluminadas se hacen solubles en soluciones acuosas de compuestos alcalinos, por ejemplo de KOH al 1%. Estas fotorresinas pueden requerir mayor tiempo de

exposición, pero la resolución final es generalmente mejor, además de que los productos químicos de revelado son menos nocivos para la salud del operador. Como la parte iluminada es la que se elimina, las figuras no transparentes de la máscara quedan grabadas en forma de islas en la fotorresina, por lo cual el proceso se llama positivo. Además, como la parte de fotorresina que se queda no se transforma, ya que no se iluminó, es más fácil su eliminación.

La respuesta espectral de las fotorresinas debe tener su máxima sensibilidad en la longitud de onda que se utilice en la fuente de luz de la alineadora.

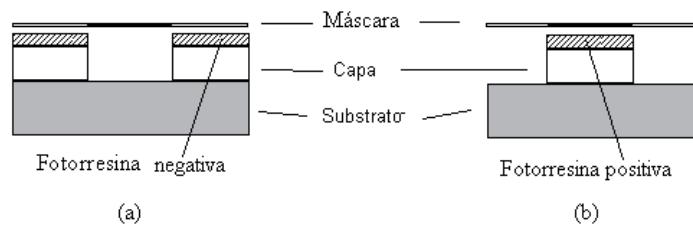


Fig. 9.2 Exposición a través de una máscara de (a) una resina negativa; (b) una resina negativa y figura que se traslada a la capa después del ataque.

9.4 TIPOS DE ALINEADORAS PARA FOTOLITOGRAFÍA ÓPTICA

El primer proceso de fotolitografía sobre la oblea sólo requiere de una orientación de los motivos en la máscara respecto a la orientación cristalina

de la oblea, para que posteriormente se pueda realizar el corte de los dados a lo largo de las direcciones preferenciales de corte de la oblea. A esta orientación de la oblea respecto a la máscara se le llama pre-alineación y consiste en colocar el borde cortado de la oblea en una posición fija respecto a la máscara, lo cual se indica de alguna forma en el equipo que se esté utilizando para exponer y alinear.

Una vez realizado el primer proceso de fotolitografía, los siguientes deben ser alineados respecto al anterior, para que cada nueva figura a trasladar a la oblea, coincida correctamente con la ya existente. Este proceso de alineación se realiza en equipos especiales llamados alineadoras, que también realizan la exposición de la fotorresina.

Existen fundamentalmente tres métodos de alineación y exposición: por contacto, por aproximación y por proyección, este último a su vez incluye la repetición de la exposición sobre la oblea y puede o no incluir una reducción adicional de los motivos de la máscara durante la proyección de los mismos sobre la oblea.

9.4.1 Alineadoras de contacto

En este caso oblea de silicio cubierta de fotorresina, se pone en contacto directo con la placa de vidrio que constituye la máscara, la cual se colocó previamente con la superficie donde se definieron los motivos mirando hacia la superficie de la oblea cubierta de fotorresina. Antes de comenzar el proceso de alineación, la oblea se separa unas cuantos micrómetros de la máscara para realizar visual o automáticamente la alineación, o sea para posicionar correctamente las figuras en la máscara con respecto a las ya existentes en la oblea. Si la alineación es manual, esto se hace con la observación de dos campos visuales diferentes al mismo tiempo

(observación con campo compartido), de manera de ver que figuras, localizadas en los extremos del campo visual, se alinean correctamente. Para posicionar correctamente las figuras en la máscara respecto a las ya existentes en la oblea se utilizan unas figuras especiales llamadas marcas de alineación.

Un ejemplo de marcas de alineación se muestra en la Fig. 9.3. En ella se muestra la alineación de una isla en forma de cruz sobre otra isla, también en forma de cruz, pero de mayor tamaño previamente existente en la oblea. Mediante técnicas visuales, con un aumento adecuado del microscopio a través del cual se observa, es posible diferenciar bien el posicionamiento de dos figuras cuyas dimensiones se diferencien en un micrómetro, por lo que se puede garantizar una precisión en la alineación de $\pm 0.5 \mu\text{m}$. A este parámetro lo llamaremos **precisión en la alineación**.

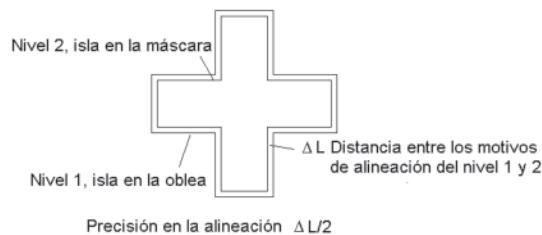


Fig. 9.3 Precisión en la alineación entre dos motivos de alineación correspondientes a diferentes niveles de fotolitografía.

Después del proceso de alineación, la máscara y la oblea se ponen de nuevo en contacto directo y se pasa al proceso de iluminación o exposición de la fotorresina, Fig. 9.4.a. Estos procesos se realizan en forma manual, semiautomática o automática, en dependencia del equipo. En las alineadoras de contacto, en la transferencia de la imagen se produce un menor efecto de difracción en los bordes, por lo que la alteración de las

dimensiones de las figuras en la máscara al ser transferidas a la oblea es menor. Sin embargo, la máscara se ve grandemente afectada por el contacto directo entre las dos superficies, así como por el daño producido por partículas que puedan quedar entre las dos superficies. En cualquier caso, se daña la superficie de la máscara, reduciendo su tiempo de utilización e introduciendo defectos en las figuras.

9.4.2 Alineadoras de proximidad

En este caso la máscara se mantiene paralela a la oblea, a una distancia de 10 a 100 μm durante la alineación y posterior exposición, Fig. 9.4b. La figura transferida presenta un mayor efecto de difracción en los bordes, (se iluminan zonas que debían ser oscuras). En este caso, como en ningún momento hay un contacto íntimo entre la oblea y la máscara, esta última puede utilizarse un número mayor de veces.

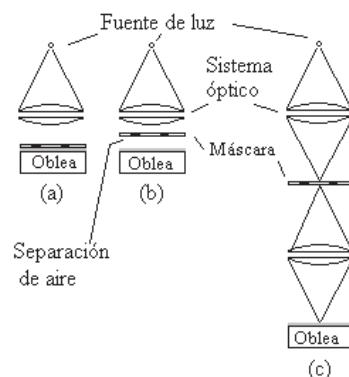


Fig. 9.4 Forma de proyectar la figura de la máscara sobre la superficie de la oblea en los 3 tipos de alineadoras ópticas: a) de contacto; b) de proximidad; c) de proyección.

9.4.3 Alineadoras de proyección

a) **Proyección de toda la máscara una sola vez.** Como se ve en la Fig. 9.4c, en este caso la máscara 1:1 se coloca en una posición intermedia entre la fuente de luz y la oblea, la imagen se proyecta por puntos lográndose una alta precisión y una alta durabilidad de la máscara. Se transfieren todos los chips a un mismo tiempo a la oblea.

b) **Proyección de la máscara y repetición de la misma sobre la oblea.**

A partir de 1980 se comenzó la utilización de una nueva variante de fotolitografía de proyección. En este caso se prepara una máscara con un área útil de unos 75 mm de lado que resulta de la reducción del retículo, por ejemplo en 5 veces, con pocas repeticiones del motivo (chip) a lo largo de los ejes “x” y “y”. La alineación y exposición se hace por pasos, repitiendo el proceso a lo largo de toda la oblea. A este proceso de alineación y exposición de forma repetida hasta cubrir toda la oblea, se le ha dado el nombre de repetidor sobre la oblea o (step on the wafer), uniendo las funciones de los repetidores antes mencionados y de las alineadoras. La reducción en esta última etapa puede ser de 10X, 5X, 4x o 1X, según los equipos que actualmente se ofertan en el mercado. En la Fig. 9.5 se da una idea del procedimiento utilizado.

Entre las ventajas se encuentra la utilización de máscaras con un área útil relativamente pequeña, lo que permite disminuir los efectos de aberración óptica durante la proyección y al mismo tiempo permitir la exposición de obleas que

La calidad de la proyección de las imágenes de la máscara a la oblea viene dada por las características del sistema de lentes ópticos utilizado. En

la actualidad los lentes tienen un alto grado de perfección y se puede considerar que no tienen aberración, estando determinada la perfección de la imagen por los fenómenos de difracción.

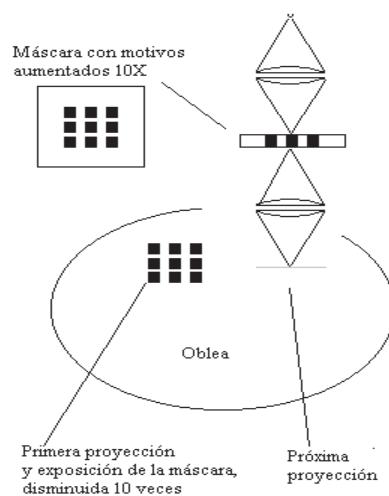


Fig. 9.5 Proceso de fotolitografía por proyección con repetición de la máscara en la oblea y disminución de la imagen durante la proyección.

9.4.4 SITUACION ACTUAL Y TENDENCIAS.

El motor propulsor de la tecnología de la microelectrónica y en especial de los procesos fotolitográficos, son las memorias semiconductoras.

Las exigencias de las memorias para utilizar la fotolitografía óptica hasta sus límites posibles han logrado ampliar su rango de aplicación hasta anchuras de líneas por debajo del micrómetro. En los próximos años, no se prevé un límite para la utilización de la fotolitografía óptica [1-3]

En la tabla 9. 1 se muestran algunos parámetros típicos de memorias que se fabrican actualmente.

Tabla 9.1 Parámetros de memorias semiconductores en producción.

Tipo de memoria	Tamaño aproximado del dato [mm ²]	Dimensión mínima de línea [μm]	Precisión de alineación [μm]
4 Mb	90	0.7	0.35
16 Mb	140	0.5	0.25
64 Mb	200	0.3	0.15
256 Mb	300	0.25	0.1
1 Gb	400	0.18	0.07

Los elementos claves en el desarrollo de la fotolitografía óptica son:

- La resolución,
- La profundidad de foco,
- El traslape
- El control de las dimensiones críticas
- La productividad.

Resolución

Se define como resolución del proceso fotolitográfico a la menor dimensión que puede ser obtenida en la fotorresina, donde aún se mantengan los rasgos característicos de la figura que se transfiere. Generalmente se toma como criterio, que la forma no varíe más de un 10 %. La resolución depende de la alineadora y de la fotorresina utilizada, así como de la longitud de onda utilizada para exponer la fotorresina. Es directamente proporcional a la longitud de onda utilizada, por ello mientras

más hacia el ultravioleta profundo se trabaje, podrán resolverse líneas más pequeñas.

Aunque la resolución puede verse afectada por imperfecciones del sistema óptico como aberración o distorsión, o la separación entre la máscara y la oblea durante la exposición, se busca que el límite venga dado por los fenómenos de difracción y por lo tanto por el criterio de Rayleigh que puede expresarse como:

$$R = k \lambda / NA \quad (9.1)$$

donde k es una constante que depende de la fotorresina y NA es la apertura numérica. A menor R , se podrán resolver dimensiones menores, lo cual se logra disminuyendo λ o aumentando NA .

Apertura numérica

La apertura numérica NA se define a través del ángulo entre el centro del plano donde se obtiene la imagen y la máxima dimensión del objeto que se ilumina, o de la lente con que se proyecta, según sea el caso:

$$NA = n \cdot \sin 2\alpha \quad (9.2)$$

donde n es el índice de refracción en la región entre la lente y el plano de enfoque y α el ángulo de apertura (ver Fig. 9.6). La apertura numérica aumenta al aumentar el índice de refracción o el ángulo de apertura. El ángulo de apertura puede incrementarse aumentando el diámetro de la lente y disminuyendo la distancia focal. La NA se ha incrementado constantemente desde de 0.3 en 1982 a 0.7 en la actualidad. Se trabaja para llegar en los próximos años a 0.8.

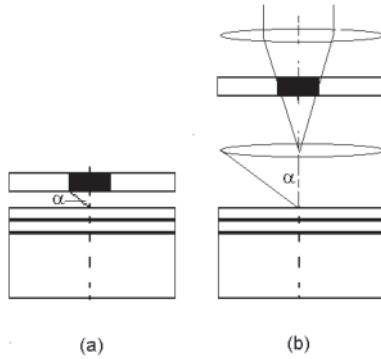


Fig. 9.6 Ángulo de apertura sin (a) y con (b) lente proyectiva

Profundidad de foco

Es un parámetro del sistema óptico que define en cuánto puede ser desplazada la oblea en dirección perpendicular al eje del sistema óptico y aún observar la imagen en foco.

Para un sistema de proyección se expresa a través de:

$$Pf = \frac{\lambda}{NA^2} \quad (9.3)$$

Por lo tanto al lograr una mejor resolución, se disminuye la profundidad de foco.

Precisión en la alineación

Cada vez que se alinea una máscara sobre una figura ya existente en la oblea, habrá un cierto error en el posicionamiento que hará que las dos figuras se encuentren desplazadas una respecto a la otra. La precisión en la alineación depende de la alineadora, del microscopio con que se observa y de la experiencia del técnico que lo realiza. Si la desalineación en cada

proceso de fotolitografía es aleatoria, al final de todos los procesos fotolitográficos que requiere la fabricación del dispositivo, el error total puede ser relativamente pequeño.

Traslape

Define la distancia mínima que puede haber entre figuras de diferentes niveles del diseño topológico de un circuito integrado. Los factores que más influyen son: la precisión en la alineación; la distorsión de las imágenes en la máscara y en la oblea y el corrimiento de los contornos de las figuras durante el procesamiento. La magnitud del traslape ha venido disminuyendo desde 0.5 μm en 1982, hasta menos de 0.05 μm actualmente para que esté en correspondencia con la resolución y utilización de líneas de 0,18 μm .

Productividad

Se define como productividad al número total de obleas que se procesan por unidad de tiempo en un proceso de fotolitografía dado. El aumento en la resolución generalmente implica una disminución en la productividad del proceso, como resulta en el caso de los generadores de patrones por haz de electrones.

Longitudes de onda de exposición de las fotorresinas

Línea g: Corresponde a una longitud de onda de 436 nm. Fue la primera longitud de onda utilizada en la fotolitografía. Con las expresiones antes mostradas, utilizando el factor k de 0.8 y una apertura de 0.45 se obtuvo una resolución de 0.65 nm que permitió la fabricación de las memorias de 4 Mbit.

Línea i: Corresponde a una longitud de onda de 365 nm y permite mejor resolución que la línea g.

Ultravioleta Profundo y ultra profundo (DUV-Deep y VDUV Very Deep Ultra-Violet): Para alcanzar mayores resoluciones se desarrollaron alineadoras que trabajan con la longitud de onda de 248 nm, en el ultravioleta profundo, producida por un láser de excimer de fluoruro de kriptón, (KrF). Otros láseres de excimer tienen frecuencias mas profundas como el ArF con 193 nm y de F₂ con 157 nm. Un elemento a tener en cuenta es que las placas de las máscaras y los lentes ópticos deben de ser de cuarzo para que permitan pasar el DUV y el VDUV. También la alta energía de los fotones del ultravioleta profundo implica tener un nuevo procesamiento de la fotorresina.

La evolución de la fotolitografía óptica es tan intensa que no es posible predecir con mucha antelación los nuevos avances que tendrán lugar. Lo que si parece un hecho es que se seguirán encontrando soluciones que permitan seguir avanzando en mejorar la resolución trabajando con longitudes de onda cada vez menores. Se trabaja en disminuir el valor de k desarrollando nuevas fotorresinas, seguir aumentando la apertura numérica y bajando la longitud de onda hacia el ultravioleta extremo [1-3].

9.5 FOTOLITOGRAFIA CON HAZ DE ELECTRONES

La posibilidad de sensibilizar un material polimérico con un haz de electrones data de la década de los años 70, dando lugar a una nueva técnica de fotolitografía. A partir de 1975 los grandes laboratorios comenzaron a utilizar generadores de patrones que utilizan un haz de electrones para generar el retículo o la máscara de trabajo, directamente con las dimensiones finales o para exponer directamente la oblea. Estos equipos son complejos y costosos.

9.5.1 Generador de patrones utilizando haz de electrones

En la Fig. 9.7 se muestran las partes fundamentales de un equipo de fotolitografía por haz electrónico. En lo esencial son los mismos elementos que un microscopio electrónico de barrido, donde se controla la dimensión y posición del haz electrónico. Por su forma de operación, este equipo es relativamente mas lento que las técnicas de fotolitografía óptica con máscaras de trabajo, ya que para obtener la máxima resolución, la figura se expone píxel a píxel mediante un haz luminoso que se desplaza. Pueden usarse otras opciones donde se exponen varios píxeles a la vez, pero se pierde en resolución. Otra característica importante es el enfoque del haz de electrones, el cual debe tener un diámetro menor de $0.1 \mu\text{m}$ y una alta densidad de corriente enfocada para disminuir el tiempo de exposición. Los cañones electrónicos conforman un haz que tiene una dimensión de 10 a $100 \mu\text{m}$, por lo cual se requiere una óptica-electrónica que reduzca este haz a las dimensiones requeridas. Se necesita controlar por computadora el barrido de este haz en todo el campo de barrido, y el borrado, con una velocidad del orden de MHz.

En estos equipos un elemento fundamental es el movimiento X-Y del soporte de retículo o de la oblea que se sensibiliza. Generalmente se utiliza la interferometría láser para la medición de las distancias recorridas. Este movimiento define la precisión de posicionamiento, la repetibilidad y la alineación entre capas diferentes. La forma de posicionamiento, y/o alineación, es barriendo una marca de referencia hecha previamente en el substrato y detectando los electrones secundarios y la retrodispersión. Estas señales sirven para realizar el posicionamiento automático con una precisión del orden de $\pm 0.2 \mu\text{m}$.

Se usa como regla general que el diámetro del haz sea 4 veces menor que la dimensión mínima, y el campo de barrido sea del orden de 2000 veces la dimensión mínima. De esta manera, para obtener una línea de 1 μm de ancho, se requiere un haz de 0,25 μm y un campo de barrido de 2x2 μm . Como se ve, mientras menores son las dimensiones de las líneas a sensibilizar, mayor será el desplazamiento que se requiere del soporte X-Y, y por lo tanto tomará mas tiempo para hacer una máscara patrón o una exposición directa sobre la oblea. En la información comercial pueden encontrarse numerosos equipos para estas aplicaciones. Por ejemplo la JEOL anuncia un modelo para líneas de 0.1 μm , donde se tiene un diámetro de haz entre 0.005 y 0.2 μm .

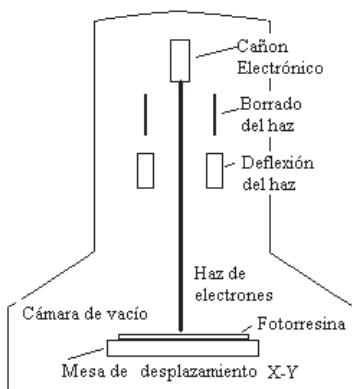


Fig. 9.7 Esquema de las partes principales del generador de patrones por haz de electrones.

9.5.2 Fotorresina electrónica

Fotorresina electrónica se utiliza cuando el elemento de sensibilización de la resina polimérica es un electrón en lugar de un fotón. Esta radiación

provoca cambios físicos y químicos en los polímeros, lo que permite también definir una figura en la fotorresina. Existen dos tipos de resinas:

9.5.2.1 Fotorresinas electrónicas positivas

La radiación conlleva a la ruptura de enlaces de las cadenas, reduciendo el peso molecular de la zona irradiada. El revelador puede diluir esta zona dejando grabada una imagen positiva. Se usan, entre otros el llamado PMMM, que es una solución acuosa de poli-metil metacrilato o el PBS formado por poli-butano-1 sulfona, diluido en un solvente;

Como revelador se puede usar, por ejemplo, una mezcla de 1 en 1 de metil-isobutil-cetona en alcohol isopropílico.

9.5.2.2 Fotorresinas electrónicas negativas

En las resinas negativas, la radiación conlleva al aumento del peso molecular de las moléculas y el revelador disuelve la zona de más baja densidad molecular, por lo cual se forma una imagen negativa. El hecho que la capa se hinche durante el revelado, limita la resolución de esta resina. La más común es la llamada COP, formada por poli-glicidilmetacrilato-co-etil acrilato.

Las resinas positivas requieren de una dosis de electrones por cm^2 (sensibilidad) tres órdenes superior que las negativas, con lo cual el tiempo de exposición es de unas 1000 veces mayor. El área menor en el que se puede detectar si la resina está o no sensibilizada es el píxel, o sea, es la menor dimensión (ancho) de línea que se puede resolver. Para una sensibilidad dada S , se requiere una cantidad de carga eléctrica qNm , donde Nm es el número de electrones que llega al área del píxel, L_p^2 . Esta cantidad de electrones mínima se puede calcular por:

$$Nm = \frac{S \cdot L_p^2}{q}. \quad (9.4)$$

Para tener una probabilidad alta (1×10^{-20}) de que el píxel se sensibilice, se considera que Nm sea mayor de 200. De esta forma, si se conoce la máxima cantidad de carga eléctrica que permite incidir el equipo, se puede conocer también, para una sensibilidad dada de fotorresina, cual es la dimensión mínima que es posible resolver con dicho equipo.

9.5.3 La dispersión y el efecto de proximidad

Los electrones que penetran en la fotorresina y los que llegan al substrato son susceptibles del fenómeno de dispersión elástica e inelástica. Como resultado de estas interacciones se produce la dispersión del haz de electrones y el cambio de trayectoria, y se mueven hasta que pierden toda la energía o salen del material por retrodispersión (backscattering). Los electrones que retornan del substrato penetran en la proximidad de la zona del píxel irradiado, sensibilizando sus alrededores. Esto se ve en las Fig. 9.8 y 9.9 donde se muestran, en diferentes figuras, los dos efectos de proximidad: el efecto de intraproximidad en las esquinas, y el efecto de interproximidad entre figuras. Todos son producidos por la acción de la retrodispersión de un píxel sobre los puntos que lo rodean. También pueden producirse diferencias en el revelado entre diferentes anchuras de líneas, todo lo cual trata de evitarse subdividiendo la figura que se expone en secciones pequeñas y ajustando la dosis de radiación al mínimo valor posible.

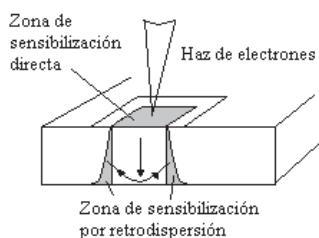


Fig. 9.8 Efecto de aumento del ancho de la figura por efecto de la retrodispersión del haz de electrones

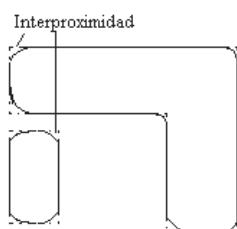


Fig. 9.9 Alteración de los bordes de la figura definida por un generador de patrones de haz de electrones debido a los efectos de proximidad.

9.5.4 Aplicaciones

El uso del haz de electrones se ha introducido en toda la actividad de fotolitografía y es una herramienta más en la actualidad. Como se ve en la Fig. 9.10 esta técnica participa de una forma u otra en los métodos de fotolitografía utilizados para dimensiones submicrométricas.

9.5.4.1 Preparación de retículos y máscaras por exposición directa

Desde el año 1975 hasta el 1980 el trabajo principal del generador de patrones por haz electrónico fue la fabricación de retículos y máscaras por exposición directa. Con la aparición de las alineadoras ópticas de

proyección con repetición en la oblea y reducción de la imagen proyectada, las dimensiones finales en los retículos que se utilizaban en estas alineadoras pudo aumentarse, según la reducción de las dimensiones que permitiera la alineadora, por ejemplo 2:1, 5:1 o 10:1. Por ello, durante un tiempo, el generador de patrones por haz de electrones fue menos utilizado para este fin. Sin embargo, los nuevos requerimientos de dimensiones de líneas por debajo de 0.2 μm han llevado de nuevo a la utilización de los sistemas de haz de electrones, cada vez mas mejorados en precisión de posicionamiento, resolución y definición de contornos.

La preparación de máscaras también es requerida por las alineadoras que utilizan rayos X en lugar de fuentes de luz ultravioleta. Las alineadoras de rayos X permiten definir líneas con dimensiones por debajo de 0.2 μm con alta productividad debido al uso de máscaras en lugar de la exposición directa. Ello requiere que el generador de patrones por haz de electrones fabrique las máscaras 1:1, con las dimensiones requeridas, que pueden ser por debajo de 0.2 μm (ver Tabla 9.2). Los requerimientos fundamentales para anchura de línea mínima 0.25 μm son:

Tabla 9.2

Especificaciones para que un generador de patrones resuelva una línea de 0.25 μm de ancho.

ESPECIFICACIONES	$L_{\min}=0.25 \mu\text{m}$
Control de la anchura	5% de $L_{\min}=0.0125 \mu\text{m}$
Ancho del píxel	$h=4$ píxel por $L_{\min}=0.0625 \mu\text{m}$
Precisión de posicionamiento	$P_p \leq 10\%$ de $L_{\min}=0.025 \mu\text{m}$

9.5.4.2 Escritura Directa

El generador de patrones por haz de electrones permite la escritura directa sobre la oblea. Existen algunos casos en que esto es insustituible.

Como primer ejemplo se tiene la fabricación de circuitos integrados a la orden (ASIC). Para esto se requiere la fabricación de 1 o 2 máscaras finales de interconexión, en el caso de los arreglos de compuertas, o de todas las máscaras en el caso de las Celdas Normalizadas. Cuando la necesidad de circuitos integrados es pequeña, o sea de algunas unidades a unos pocos miles, como es el caso de los nuevos desarrollos experimentales, la producción de juegos de máscaras es costoso y poco ágil. Con la utilización de la escritura directa en la oblea en este caso se obtienen dos grandes ventajas:

- No es necesario fabricar las máscaras, ahorrando en materiales y tiempo, y ganando en precisión;
- Si es necesario introducir cambios en el diseño, solo se requiere introducirlos en el sistema CAD cuyo resultado se guardará en un soporte magnético que controlará el generador de patrones con haz de electrones.

En el caso de las celdas normalizadas que requieren de la fabricación de todas las máscaras, la utilización de la escritura sobre la oblea se considera rentable, cuando la cantidad de obleas a procesar es menor de 500 unidades.

El otro ejemplo de utilización del haz electrónico en escritura directa lo tenemos en el desarrollo de nuevas estructuras submicrométricas de transistores y dispositivos optoelectrónicos, siendo len algunos casos la única forma de obtener estas dimensiones durante la fotolitografía.

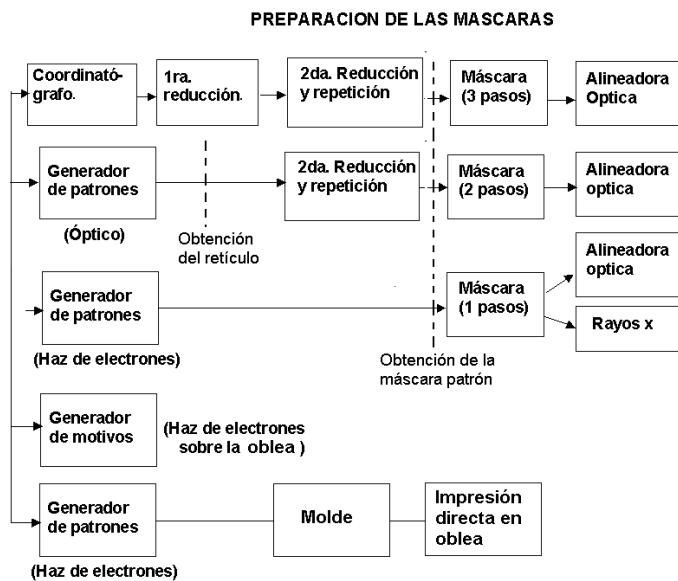


Fig. 9.10 Resumen de las técnicas de fotolitografía, incluyendo las técnicas de preparación de máscaras.

9.6 FOTOLITOGRAFIA DE RAYOS X

9.6.1 Características generales

La fotolitografía de rayos X es en principio lo mismo que la fotolitografía óptica con la técnica de aproximación. En este caso la longitud de onda utilizada es de 0.4 a 5 nm, lo que permite reducir los efectos de difracción de bordes.

Este método combina las ventajas de la alta resolución con una alta productividad. Tiene otras ventajas como son que no hay efectos de dispersión por la baja energía de los rayos, con lo cual se eliminan los efectos de proximidad; los puntos de polvo de baja densidad son transparentes, no ocasionando defectos en las figuras y por la baja absorción de la resina atravesía toda la capa sensibilizándola, produciendo

bordes muy abruptos. Las imágenes se pueden afectar por los llamados efectos geométricos.

El constante desarrollo de la fotolitografía óptica, pasando al campo del ultravioleta profundo, y los altos costos y complejidades de las fuentes del tipo sincrotrón para los rayos X, han hecho que se la fotolitografía óptica, al entrar en la región del ultravioleta muy profundo, haya podido alcanzar la resolución de líneas de dimensiones menores de $0.18 \mu\text{m}$ con facilidad, y se trabaja para llegar a límites inferiores. Esto, que hace unos años parecía imposible ha disminuido un poco el desarrollo de la fotolitografía de rayos X.

9.6.2 Fotorresina para rayos X

El rayo x incidente se desplaza prácticamente sin dispersión hasta que tiene una colisión con un átomo y se excita generando un fotoelectrón. El átomo excitado pasa a su estado estable generando un rayo X fluorescente, que repite el proceso, o un electrón de Auger. En todas las variantes se generan electrones que sensibilizan la fotorresina utilizada. De esta forma, en principio, las fotorresinas que se usan para haz de electrones pueden usarse para rayos x. El problema está en la baja densidad de la radiación de rayos x, que si es del orden de 1 a $10 \text{ mJ}/[\text{cm}^2 \text{ min}]$, requiere de un tiempo muy largo para sensibilizarla.

Se ha tratado de disminuir el umbral de sensibilidad de la fotorresina aumentando la absorción de los rayos x. Esto se logra incorporando una sustancia que tenga una línea de absorción cercana (algo mayor) a la frecuencia de radiación. Para la fotorresina negativa esto se logra con la introducción del Cloro. Además, para mejorar la resolución una técnica utilizada es evitar que se hinche la fotorresina durante el revelado, lo que se logra con el ataque en seco, en plasma de O_2 .

9.6.3 Máscaras

Las máscaras de rayos x deben tener como elemento fundamental un material que absorba los rayos x, que se pueda definir en forma de figuras, sobre otro que deje pasar con transparencia los mismos. El proceso de fabricación es relativamente complejo [4].

9.7 MÉTODOS NO CONVENCIONALES DE FOTOLITOGRAFÍA

Los altos costos del equipamiento para fabricar dimensiones submicrométricas incentivó el desarrollo de técnicas alternativas para obtener estas dimensiones de líneas. Una de las líneas es la litografía por estampado [5]. En este caso no se utilizan lentes de proyección y por lo tanto el valor de la resolución no resulta fundamental. Existen diferentes técnicas para preparar un molde que se pone en contacto con materiales plásticos o poliméricos los que a su vez toman la forma del molde cuando sobre ellos se actúa con alta presión, temperatura y luz ultravioleta, dependiendo del método utilizado. La preparación del molde requiere en todos los casos del uso de generador de patrones por haz de electrones. Ésta técnica incrementa bastante el número de obleas posibles de procesar por unidad de tiempo, además de disminuir los costos del equipamiento relacionado con las alineadoras. Para la alineación entre sí de cada estampado con el anterior se requiere también de un equipo de alineación del molde con la oblea, por lo cual se busca que el molde sea de un material transparente para poder alinearlos sobre la oblea antes de realizar el estampado. El hecho de que sea transparente tiene también otras ventajas que se indican en [5].

Otra alternativa para mejorar la resolución es el uso de capas para desplazar la fase, propuesta por primera vez en [6]. Las máscaras que

utilizan este efecto reciben el nombre de **máscaras con desplazamiento de fase**. El principio de trabajo se muestra en la Fig. 9.11. Cuando dos rendijas cercanas entre sí se iluminan, tiene lugar el fenómeno de difracción de Franhofer en los bordes y el patrón luminoso que se observa hace que entre ellas haya cierta intensidad de luz, en la región donde debía haber oscuridad. Fig. 9.11a. La línea discontinua representa la luz a la salida de una rendija aislada, mientras la continua representa el patrón luminoso de dos rendijas paralelas cercanas entre sí. La interferencia constructiva de los rayos que se difractan de las rendijas cercanas es la causa del incremento de intensidad luminosa en la región que las separa. Si se pudiera lograr que la fase de la luz que sale de las aperturas adyacentes estuviera desfasada en 180 grados, tendría lugar el efecto que se muestra en la Fig. 9.11b, donde la interferencia entre los rayos difractados sería destructiva y no constructiva. Esto tendría lugar si se cubre una de las rendijas con una capa transparente de espesor $d=\lambda/2(n-1)$. Se dice que las rendijas se resuelven, cuando la intensidad de luz en la región oscura entre ellas está por debajo de un nivel dado, que como ya se indicó se conoce como criterio de Raleigh. Por lo tanto, con este mecanismo, es posible que dos rendijas de igual dimensión y situadas a la misma distancia puedan ser resueltas, sin modificar el sistema óptico que las proyecte.

En la práctica el procedimiento de fabricación es algo más complejo y puede ser consultado en [6].

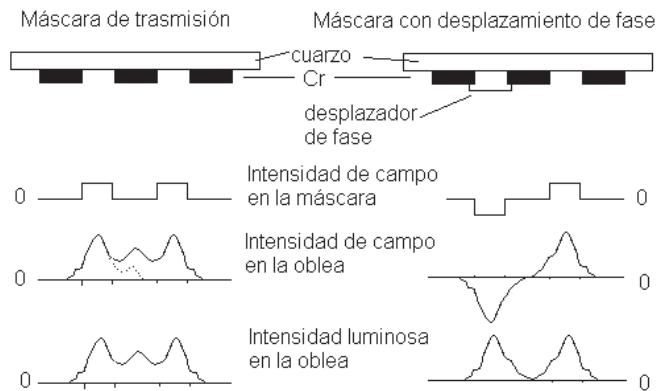


Fig. 9.11 Comparación de los efectos de difracción en una máscara normal y en una con corrección en los bordes por desplazamiento de fase.

9.8 BIBLIOGRAFÍA

- [1]. International Technology Roadmap for Semiconductors, editions 1995, 1999 and 2001, <http://public.itrs.net>
- [2]. P.M. Ware, “Why optics is forever in lithography”, Solid State Tech. 43 (5) (2000) 192.
- [3]. G. Resor, “Charting different roadmaps for lithography at Japanese litho firms”, Solid State Tech. 44 (11) (2001) 59.
- [4]. S. A. Campbell, “The Science and Engineering of Microelectronic Fabrication”, Oxford University Press, 1997.
- [5]. M. Colbyrn, T. Bailey, B.J. Choi, J.G. Ekerdt, S.V. Sreenivasan, C.G. Willson, “Development and advantages of step-and-flash lithography”, Solid State Tech. 44 (7) (2001) 67.
- [6]. M.D. Levenson, N.S. Viswanathan, R. A. Simpson, “Improving Resolution in Photolithography with a phase shifting mask” IEEE Tran on ED 29 (12) (1982) 1828-1837.

CAPÍTULO 10

ANEXOS

10.1 TABLA A-1

Tabla de colores para SiO_2 y Si_3N_4 ,

(Observando perpendicularmente bajo luz fluorescente).

<i>Espesor(μm)</i>	<i>Color para SiO_2</i>	<i>Color para Si_3N_4</i>
0.05	bronceado	amarillo- café
0.075	café	rojo
0.100	violeta oscuro a violeta rojizo	azul claro
0.125	azul royal	plateado
0.150	azul claro a metálico	amarillo
0.175	gris metálico-amarillo verdoso	naranja
0.2	oro pálido a amarillo metálico	rojo
0.225	dorado-naranja amarillento	azul
0.25	naranja a melón	azul verdoso
0.275	violeta rojizo	verde claro
0.3	azul a violeta azuloso	naranja-amarillento
0.310	azul	
0.325	azul a azul verdoso	rojo
0.345	verde claro	
0.350	verde a amarillo verdoso	
0.365	amarillo verdoso	
0.375	verde amarillento	
0.390	amarillo	
0.412	naranja claro	
0.426	rosado color carne	
0.440	violeta rojizo	
0.465	rojo tirando a violeta	
0.470	violeta	
0.480	azul tirando a violeta	
0.490	azul	
0.500	azul verdoso	
0.520	verde (amplio)	
0.540	amarillo verdoso	
0.550	verde amarillento	

Tabla de colores para SiO_2 y Si_3N_4 , (cont.)

<i>Espesor (μm)</i>	<i>Color para SiO_2</i>	<i>Color para Si_3N_4</i>
0.574	amarillo tirando a amarillento	
0.585	naranja claro a amarillento	
0.600	rosado color carne	
0.630	rojo violeta	
0.680	azuloso	
0.720	azul verdoso a verde intenso	

10.2 TABLA A-2

Parámetros para la determinación de la difusividad
de gases utilizados en tecnología planar.

Gas	$T_c(k)$	$P_c(Pa)$	PM	$\eta_0 \left[\frac{kg}{m^3} \right]$	$\rho_0 \left[\frac{kg}{m^3} \right]$	$D_0 [cm^2/s]$
O ₂	154.7	$5.08 \cdot 10^6$	32	1910	1.429	0.18
N ₂	127.25	$3.398 \cdot 10^6$	28	1665	1.251	0.17
H ₂	33	$1.297 \cdot 10^6$	2	840	0.08988	1.28
CIH	324.55	$8.258 \cdot 10^6$	37.45	1320	0.125	
H ₂ O	647.3	$2.21 \cdot 10^7$	18		0.144	
SiCl ₄	507.8	$3.76 \cdot 10^6$	169.89		0.04	0.019
SiH ₄	269.6	$4.84 \cdot 10^6$	32.09	1076	1.44	0.096
PH ₃	324.4	$7.54 \cdot 10^6$	33.975	1072	1.53	0.096
B ₂ H ₆	289.65	4.10^6	17.91			0.108
A ₆ H ₃			77.91	1468	3.48	0.065

η_0 - viscosidad a 0°C;

ρ_0 - densidad a 0°C;

T_c - temperatura crítica;

P_c - presión crítica;

D_0 - difusividad del gas en sí mismo a 0°C;

PM - peso molecular.

10.3 TABLA A-3

TRABAJO DE EXTRACCIÓN PARA DIFERENTES MATERIALES

<i>Elemento J</i>	$q\phi_0 [eV]$	<i>Elemento</i>	$q\phi_0 [eV]$
Li	2.38	Mo	4.60
Be	3.92	3Ru	Ru
B	4.5	Rh	4.60
C	4.7	Pd	4.75
Na	2.35	Ag	4.8
Mg	3.64	Cd	4.3
Al	4.25	In	4.1
Si	4.8	Sn	3.8
K	2.22	Sb	4.38
Ca	2.80	Te	4.08
Sc	3.3	Cs	4.73
Ti	3.95	Ba	1.81
V	4.12	La	2.49
Cr	4.58	Ce	3.3
Mn	3.83	Hf	2.7
Fe	4.31	Ta	3.53
Co	4.41	W	4.12
Ni	4.50	Re	4.54
Cu	4.40	Os	5.0
Zn	4.21	Ir	4.7
Ga	3.96	Pt	4.7
Ge	4.76	Au	5.32
As	4.72	Hg	4.30
Se	4.72	Tl	4.52
Rb	2.16	Pb	3.7
Sr	2.35	Bi	4.0
Y	3.3	Th	4.4
Zr	3.9	U	3.30
Nb	3.99		3.3

10.4 TABLA A-4

TEMPERATURAS A LAS QUE SE ALCANZAN LOS DIFERENTES VALORES DE TENSIÓN DE VAPOR
PARA LOS COMPUESTOS MÁS UTILIZADOS COMO FUENTES DE IMPUREZAS EN SILICIO [1]

		Temperatura a la que se alcanzan las tensiones de vapor [°C]									
Tensión de vapor [kPa] [Torr]	1	0.133	0.267	1.33	2.67	5.33	8	13.3	27.7	3	101
Fuente de impurezas											760
P ₂ O ₅	384	424	442	462	481	493	510	532	556	591	
POCl ₃	-	2	13.6	27.3	35.8	47.4	65	84.3	105.1		
PBr ₃	7.8	34.4	47.8	62.4	799	89.8	103.6	125.2	149.7	175.3	
PCl ₃	-51.6	-31.5	-21.3	-10.2	2.3	10.2	21	37.6	57.9	74.2	
PH ₃	-	-	-	-129.4	-125	-118.8	-109.4	-98.3	-87.5		
AsO ₃	212	242.6	259.7	279.2	299.2	310.3	332.5	370	412.2	457.2	
Sb ₂ O ₃	574	626	666	729	812	873	957	1085	1242	1425	
SbCl ₃	49.2	71.4	85.2	100.6	117.8	128.3	143.3	165.9	192.2	219	
SbBr ₃	93.9	126	142.7	158.3	177.4	188.1	203.5	225.7	250.2	275	
B ₂ H ₆	-159.7	-49.5-	-144.3	-138.5	-131.6	-127.2	-120.9	-111.2	-99.6	-87.5	
SiH ₄	-179.3	168.6	-163	-157.9	-150.3	-147.3	-140.5	-131.6	-122	-111.5	
H ₂ O	-17.3	1.2	11.3	22.2	34.1	41.6	51.6	67.5	83	100	
SiCl ₄	-63.4	-44	-34.4	-24	-12.1	-4.8	5.4	21	38.4	57.8	

TABLA A-4.2

TEMPERATURAS A LAS QUE SE ALCANZAN LOS DIFERENTES VALORES DE TENSIÓN DE VAPOR PARA B_2O_3 EN SILICIO [1]

Temperatura a la que se alcanzan las tensiones de vapor [$^{\circ}\text{C}$]			
Tensión de vapor [kPa] [Torr]	1.33×10^{-5} 10^{-4}	1.33×10^{-4} 10^{-3}	1.33×10^{-2} 10^{-2}
Fuente de impurezas			
B_2O_3	977	1078	1195
			1336

TABLA A-4.3

TEMPERATURA DE FUSIÓN DE FUENTES DE IMPUREZAS EN SILICIO [1]

Fuente	P_2O_5	POCl_3	PBr_3	PCl_3	PH_3	SiH_4	As_2O_3	B_2H_6	Sb_2O_3	SbCl_3	SbBr_3
T de fusión [$^{\circ}\text{C}$]	569	2	-40	-111.8	-132.8	-185	-312.8	-169	656	73.4	97.6

[1] Tabla de magnitudes físicas, editado por K.K. Kikoin, Moscú, Atomisdat, 1977.

10.5 TABLA A-5

**VALORES DE PROYECCIÓN DEL RANGO (RP) Y
DESVIACIÓN MEDIA (ΔR_p) DE VARIAS IMPUREZAS EN SI,
CALCULADOS EN [1]**

Energía [keV]	Antimonio		Arsénico		Boro		Fósforo	
	R_p [nm]	ΔR_p [nm]						
10	8.8	2.6	9.7	3.6	33.3	17.1	13.9	7.9
20	14.1	4.3	15.9	5.98	67.2	28.3	25.3	11.9
30	28.7	5.8	21.5	8.0	98.7	37.1	37.8	17.6
40	23.0	7.1	27.9	9.9	130.2	44.3	48.6	21.2
50	27.1	8.4	32.2	11.8	160.8	50.4	60.7	25.6
60	31.0	9.6	37.4	13.6	190.3	55.6	73.0	29.8
70	34.7	10.7	42.6	15.4	218.8	60.1	85.5	43.0
80	38.5	11.8	47.8	17.2	247.5	64.1	98.1	38.0
90	42.1	13.0	53.0	18.9	273.3	67.7	110.9	41.8
100	45.7	14.0	58.2	29.7	299.4	71.0	123.8	45.6
110	49.3	15.1	63.4	22.4	324.8	73.9	137.7	49.2
120	52.9	17.2	68.6	24.1	349.6	77.6	149.7	52.8
130	57.4	17.2	73.9	25.8	373.7	79.0	162.7	57.2
140	59.9	18.3	79.1	27.5	397.4	81.3	175.7	59.5
150	63.4	19.3	84.5	29.2	420.5	83.4	188.8	62.8
160	67.9	20.3	89.8	30.8	443.2	85.4	201.9	65.9
170	70.4	21.3	95.2	32.5	465.4	87.2	214.9	68.9
180	73.9	22.4	100.5	34.1	487.2	89.0	227.9	71.9
190	77.3	23.4	107.0	35.8	508.6	90.6	240.9	74.7
200	80.8	24.4	111.4	37.4	529.7	92.1	253.9	77.5
220	87.8	27.4	122.3	40.7	570.8	95.0	279.8	82.9
240	94.7	28.3	133.4	43.9	610.8	97.5	305.4	88.0
260	101.7	30.3	144.5	47.0	649.6	99.9	330.9	92.8
280	108.6	32.2	155.8	50.2	687.5	102	357.2	97.4
300	115.6	34.2	167.1	53.3	724.5	104	-	101.7

10.6 TABLA A-6

VALORES DE PROYECCIÓN DEL RANGO (RP) Y DESVIACIÓN MEDIA (ΔR_p) PARA VARIAS IMPUREZAS EN SI, QUE INCLUYE LA PENETRACIÓN LATERAL EN UN MATERIAL AMORFO, CALCULADOSEN [2]

Energía [keV]	Arsénico			Boro			Fósforo		
	R_p [nm]	ΔR_p [nm]	$\Delta X = \Delta Y$ [nm]	R_p [nm]	ΔR_p [nm]	$\Delta X = \Delta Y$ [nm]	R_p [nm]	ΔR_p [nm]	$\Delta X = \Delta Y$ [nm]
20	15	5.6	4.1	77.5	34.6	37.6	25.3	11.4	9.4
40	27.2	9.6	7.9	157.6	58	65.3	48.8	20.1	17.5
60	37.8	13.3	9.6	233.7	75.2	89.9	72.9	28.8	24.9
80	47.3	17.9	12.1	308.1	88.9	110.3	97.4	37.7	32.3
100	57.7	20.4	14.6	381	110.7	141.7	122.6	44.5	39.3
120	68.2	23.7	17.1	450.2	112.6	155.1	147.9	51.6	47.2
140	78.8	27.9	19.6	517.3	119	167.1	173.2	58.1	53.1
160	89.4	30.2	22	580	127.2	177.4	198.8	64.2	59.6
180	100.1	33.4	24.3	641.8	132.7	187.9	224.7	70.2	65.9

10.7 TABLA A-7
VALORES DE PROYECCIÓN DEL RANGO (RP) Y
DESVIACIÓN MEDIA (ΔR_p) PARA VARIAS IMPUREZAS EN SiO_2
CALCULADOS EN [1]

Energía [keV]	Arsénico		Boro		Fósforo	
	R_p [nm]	ΔR_p [nm]	R_p [nm]	ΔR_p [nm]	R_p [nm]	ΔR_p [nm]
20	12.7	4.3	62.2	25.5	19.9	8.4
40	21.7	7.2	128.3	41.8	38.8	15.2
60	30.3	9.9	192.1	54	58.6	21.6
80	38.8	12.5	252.8	63.4	79.2	27.6
100	47.3	15.1	310.4	71	100.2	33.3
120	55.9	17.6	365.3	77.4	121.5	38.7
140	64.6	20.1	417.9	82.7	142.9	43.7
160	73.4	22.6	468.5	87.4	164.4	48.5
180	82.3	25.1	517.2	91.4	185.9	52.9

10.8 TABLA A-8

**VALORES DE PROYECCIÓN DEL RANGO (RP) Y
DESVIACIÓN MEDIA (ΔR_p) PARA VARIAS IMPUREZAS EN Si_3N_4
CALCULADOS EN [1]**

Energía [keV]	Arsénico		Boro		Fósforo	
	R_p [nm]	ΔR_p [nm]	R_p [nm]	ΔR_p [nm]	R_p [nm]	ΔR_p [nm]
20	9.9	3.3	48	19.6	15.4	7.5
40	17.9	5.6	99	32.6	30	11.8
60	23.5	7.7	148.2	42.2	45.3	17.8
80	30.1	9.7	195	49.6	61.2	21.5
100	37.7	11.8	239	55.5	77.4	25.9
120	43.3	13.7	282	60.5	93.9	30.1
140	50	15.7	322.6	64.7	110.5	34
160	57.8	17.6	361.7	68.4	127.1	37.7
180	63.7	19.5	399.4	71.6	143.7	41.1

[1].J. F. Gibbons, W.S. Johnson, Projected Range Statistics Semiconductor and Related Materials, Standford University Booksotre, USA, 1969.

[2].S. Furukawa, et al., “Range of Implanted Boron, Phosporous and Arsenic in Silicon”, Jour. App. Phys., 47 (16) (1969) 1750.



mb!

MoreBooks!
publishing



yes i want morebooks!

Buy your books fast and straightforward online - at one of world's fastest growing online book stores! Environmentally sound due to Print-on-Demand technologies.

Buy your books online at
www.get-morebooks.com

¡Compre sus libros rápido y directo en internet, en una de las librerías en línea con mayor crecimiento en el mundo! Producción que protege el medio ambiente a través de las tecnologías de impresión bajo demanda.

Compre sus libros online en
www.morebooks.es



VDM Verlagsservicegesellschaft mbH

Heinrich-Böcking-Str. 6-8
D - 66121 Saarbrücken

Telefon: +49 681 3720 174
Telefax: +49 681 3720 1749

info@vdm-vsg.de
www.vdm-vsg.de

