SEMICONDUCTOR PHYSICS and DEVICES

Prof. S.M. SZE

Introduction

0.1 Semiconductor Devices†

Semiconductor devices are the foundation of the electronics industry, which is the largest industry in the world.

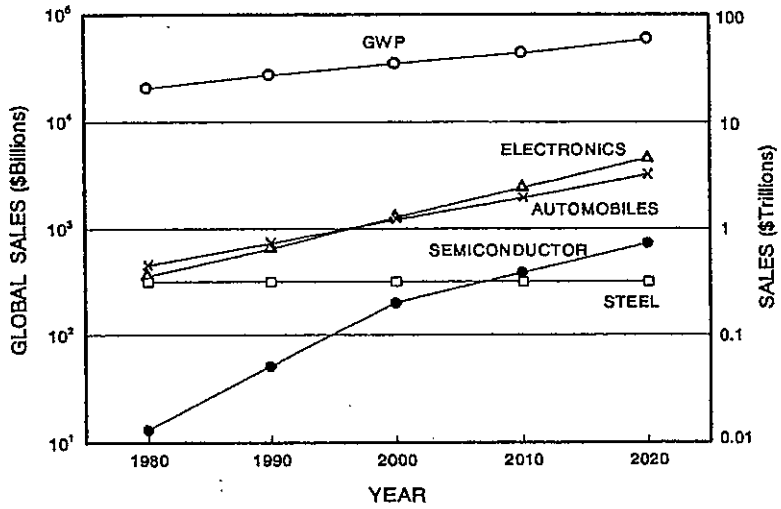


Fig. 1 Gross world product (GWP) and sales volumes of the electronics, automobile, semiconductor, and steel industries from 1980 to 2007 and projected to 2020.[1,2]

0.1.1 Device Building Blocks

We have about 60 major devices, with over 100 device variations related to them. However, all these devices can be constructed from 4 device building blocks:

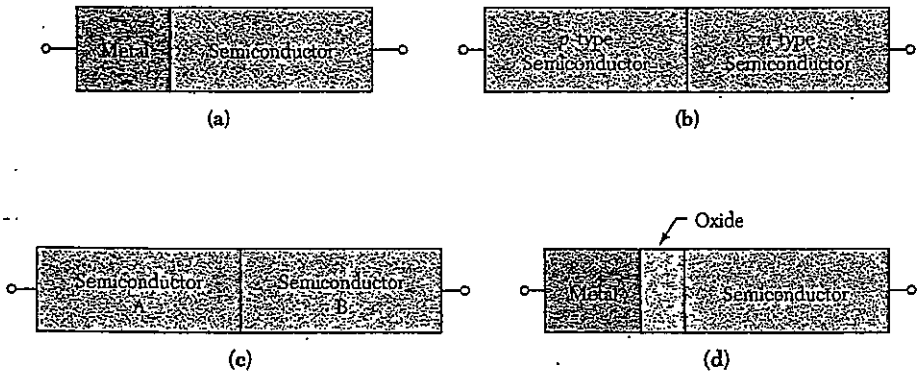


Fig. 2 Basic device building blocks. (*a*) Metal-semiconductor interface; (*b*) *p–n* junction; (*c*) heterojunction interface; and (*d*) metal-oxide-semiconductor structure.

0.1.2 Major Semiconductor Devices

Table 1 lists all major semiconductor devices in chronological order; those with a superscript b are 2-terminal devices, otherwise they are 3-terminal or 4-terminal devices.

TABLE 1 Major Semiconductor Devices

Year	Semiconductor Device[a]	Author(s)/Inventor(s)	Ref.
1874	Metal-semiconductor contact[b]	Braun	5
1907	Light emitting diode[b]	Round	6
1947	Bipolar transistor	Bardeen, Brattain, and Shockley	7
1949	*p–n* junction[b]	Shockley	8
1952	Thyristor	Ebers	9
1954	Solar cell[b]	Chapin, Fuller, and Pearson	10
1957	Heterojunction bipolar transistor	Kroemer	11
1958	Tunnel diode[b]	Esaki	12
1960	MOSFET	Kahng and Atalla	13
1962	Laser[b]	Hall et al	15
1963	Heterostructure laser[b]	Kroemer, Alferov and Kazarinov	16,17
1963	Transferred-electron diode[b]	Gunn	18
1965	IMPATT diode[b]	Johnston, DeLoach, and Cohen	19
1966	MESFET	Mead	20
1967	Nonvolatile semiconductor memory	Kahng and Sze	21
1970	Charge-coupled device	Boyle and Smith	23
1974	Resonant tunneling diode[b]	Chang, Esaki, and Tsu	24
1980	MODFET	Mimura et al.	25
1994	Room-temperature single-electron memory cell	Yano et al.	22
2001	20 nm MOSFET	Chau	14

[a]MOSFET, metal-oxide-semiconductor field-effect transistor; MESFET, metal-semiconductor field-effect transistor; MODFET, modulation-doped field-effect transistor.

We have four important photonic devices: LED, laser, solar cell, and photodetector (e.g., p-n junction). We have six microwave devices: tunnel diode, TED, IMPATT, MESFET, MODFET, and HBT.

The most important logic device for advanced ICs is the MOSFET, and the most important memory device is the nonvolatile semiconductor memory (NVSM) (especially, the Flash Memory). Both devices are covered in Ch. 6.

0.2 Semiconductor Technology

0.2.1 Key Semiconductor Technologies

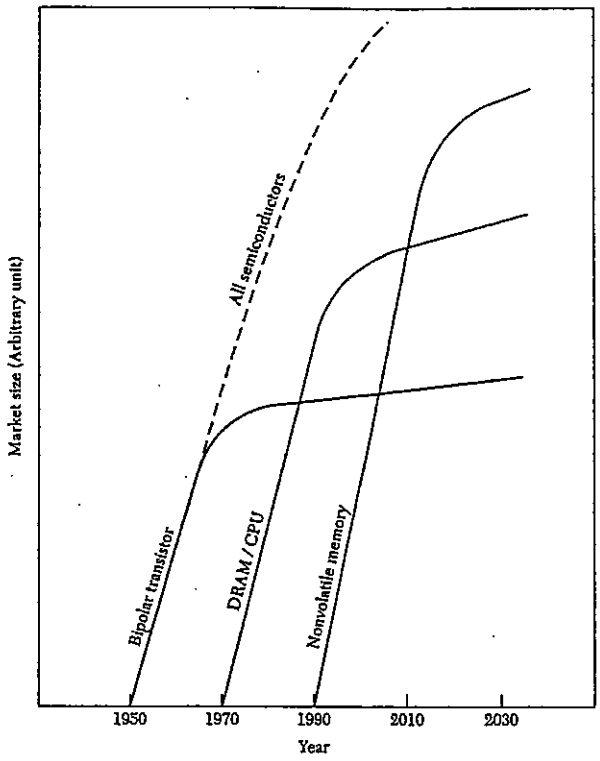
TABLE 2 Key Semiconductor Technologies

Year	Technology[a]	Author(s)/Inventor(s)	Ref.
1918	Czochralski crystal growth	Czochralski	27
1925	Bridgman crystal growth	Bridgman	28
1952	III-V compounds	Welker	29
1952	Diffusion	Pfann	31
1957	Lithographic photoresist	Andrus	32
1957	Oxide masking	Frosch and Derrick	33
1957	Epitaxial CVD growth	Sheftal, Kokorish, and Krasilov	34
1958	Ion implantation	Shockley	35
1959	Hybrid integrated circuit	Kilby	36
1959	Monolithic integrated circuit	Noyce	37
1960	Planar process	Hoerni	38
1963	CMOS	Wanlass and Sah	39
1967	DRAM	Dennard	40
1969	Polysilicon self-aligned gate	Kerwin, Klein, and Sarace	41
1969	MOCVD	Manasevit and Simpson	42
1971	Dry etching	Irving, Lemons, and Bobos	43
1971	Molecular beam epitaxy	Cho	44
1971	Microprocessor (4004)	Hoff et al.	45
1982	Trench isolation	Rung, Momose, and Nagakubo	46
1989	Chemical mechanical polishing	Davari et al.	47
1993	Copper interconnect	Paraszczak et al.	48

[a] CVD, chemical vapor deposition; CMOS, complementary metal-oxide-semiconductor field-effect transistor; DRAM, dynamic random access memory; MOCVD, metalorganic CVD.

0.2.2 Technology Trends

The growth curves for different technology drivers are shown in the figure on p. iv. At the beginning of the modern electronics era (1950 - 1970), the bipolar transistor was the technology driver. From 1970 to 1990, the DRAM and the microprocessor were the technology drivers because of the rapid growth of PCs. Since 1990, NVSM has been the technology driver, because of the rapidly growth of mobile electronic systems (cellular phone, notebook computer, MP3, digital camera, global positioning system, etc).



Growth curves for different technology drivers.[50]

0.3 Summary

Although the *semiconductor-device* field is a relatively new area of study, it has enormous impact on our society and the global economy. This is because semiconductor devices serve as the foundation of the largest industry in the world—the electronic industry.

In this introductory chapter, we have presented a historical review of major semiconductor devices from the first study of metal-semiconductor contact in 1874 to the fabrication of an ultrasmall 20-nm MOSFET in 2001. Of particular importance are the invention of the bipolar transistor in 1947, which ushered in the modern electronic era; the development of the MOSFET in 1960, which is the most important device for integrated circuits; and the invention of the nonvolalite semiconductor memory in 1967, which has been the technology driver of the electronic industry since 1990.

We have also described key semiconductor technologies. The origins of many technologies can be traced back to the late eighteenth and early nineteenth centuries. Of particular importance are the development of the lithographic photoresist in 1957, which established the basic pattern-transfer process for semiconductor devices; the invention of the integrated circuits in 1959, which was seminal to the rapid growth of the microelectronic industry; and the developments of the DRAM in 1967 and the microprocessor in 1971, which constitute the two largest segments of the semiconductor industry.

We have a vast literature on semiconductor-device physics and technology.[51] To date, more than 300,000 papers have been published in this field, and the grand total may reach one million papers in the year 2012. In this book, each chapter deals with a major device or a key technology. Each is presented in a clear and coherent fashion without heavy reliance on the original literature. However, we have selected a few important papers at the end of each chapter for reference and for further reading.

† The references (1–51) can be found in Chapter 1 of S. M. Sze, "Semiconductor Devices: Physics & Technology" 2nd Ed. Wiley, Hoboken, N.J. (2002).

Chapter 1

Physics and Properties of Semiconductors – A Review

1.1 Introduction

Most important semiconductors:

- Silicon (> 96% of all semiconductor devices used) (1.12 eV).
- III–V compounds (~ 3%):

  Binary: GaAs (1.42 eV), InP (1.35), GaN (3.44)

  Ternary: $Al_x Ga_{1-x} As$ (1.42 – 2.51 eV)

  $Ga_{0.47} In_{0.53} As$ (0.75 eV)

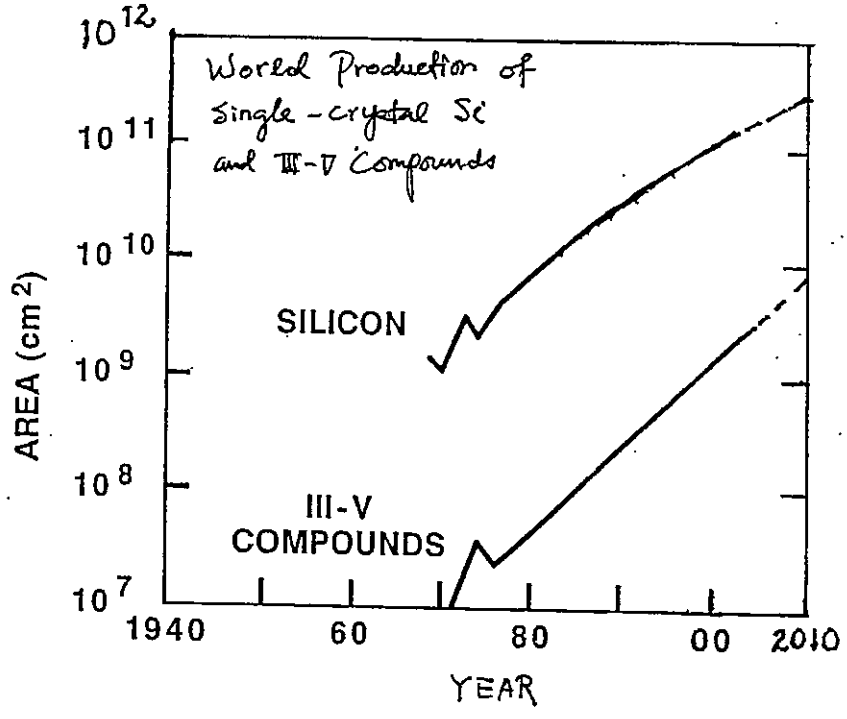
  $Al_{0.48} In_{0.52} As$ (1.45 eV)

  Quaternary: $Ga_x In_{1-x} As_y P_{1-y}$ (0.36 – 2.26 eV)

- II–VI compounds (~ 0.1%): CdSe (1.7 eV)

  ZnS (3.66 eV)

see Fig. 3.2 (p. 56 of textbook)
and App. F (p. 789)



• **Abundance of the Semiconductor Constituents in the Earth's Crust**

Element	Abundance	Element	Abundance	Element	Abundance
Silicon	.283	Zinc	7×10^{-5}	Cadmium	2×10^{-7}
Aluminum	.083	Gallium	1.5×10^{-5}	Indium	1×10^{-7}
Phosphorus	.001	Germanium	5×10^{-6}	Mercury	8×10^{-8}
Sulfur	2.6×10^{-4}	Arsenic	1.8×10^{-6}	Selenium	5×10^{-8}
Carbon	2×10^{-4}	Antimony	2×10^{-7}	Tellurium	1×10^{-9}

• material cost

$\sim \left(\text{atomic abundance}\right)^{-1}$

in weight fraction

• **Mechanical Properties of Common Semiconductors and Metals**

	Melting Point (° C)	Knoop Hardness (kg/mm^2)	Yield Strength (10^{10} dynes/cm^2)	Thermal Conductivity (W/cm-°C)	Thermal Expansion ($10^{-6}\delta L/L$-C°)
Diamond	>3550	7000	53	20	1.0
Tungsten	3410	485	4.0	1.78	4.5
Silicon carbide	2700	2480	21	3.5	3.3
Molybdenum	2610	275	2.1	1.38	5.0
Aluminum arsenide	1740	481		0.8	5.20
Zinc selenide	1520	150		0.13	7
Gallium phosphide	1467	945		0.97	5.8
Silicon	1415	1150	7.0	1.57	2.56
Gallium arsenide	1238	750		0.54	6.8
Stainless steel	1000–1400	600	2.1	0.32	12
Cadmium telluride	1098	100		0.07	5.5
Indium phosphide	1070	535		0.68	4.5
Indium arsenide	943	381		0.26	5.19
Aluminum	660	130	0.17	2.36	25
Indium antimonide	525	223		0.18	5.04

• bonding strength ~ melting point

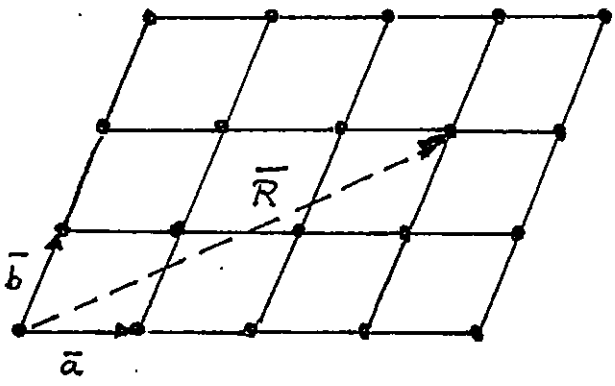
• Mechanical strength

Si > III-V and II-VI Compounds

1.2 Crystal Structures

1.2.1 Primitive Cell and Crystal Plane

2D (two dimensional lattice)



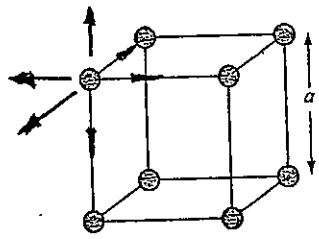
$\bar{R} = 3\bar{a} + 2\bar{b}$

$\bar{a}$, $\bar{b}$: primitive basis vectors

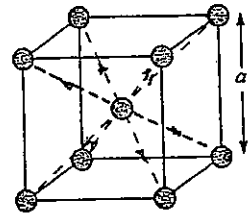
3D lattice $\bar{R} = m\bar{a} + n\bar{b} + p\bar{c}$ (1)

m, n, p integers

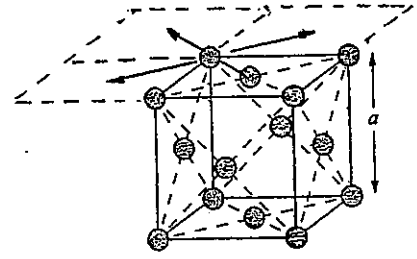
- Simple cubic (Polonium only) – 6 neighbors
- body-centered cubic (Na, W, etc) – 8 neighbors (bcc)
- face-centered cubic (Al, Au, etc) – 12 neighbors (fcc)



sc

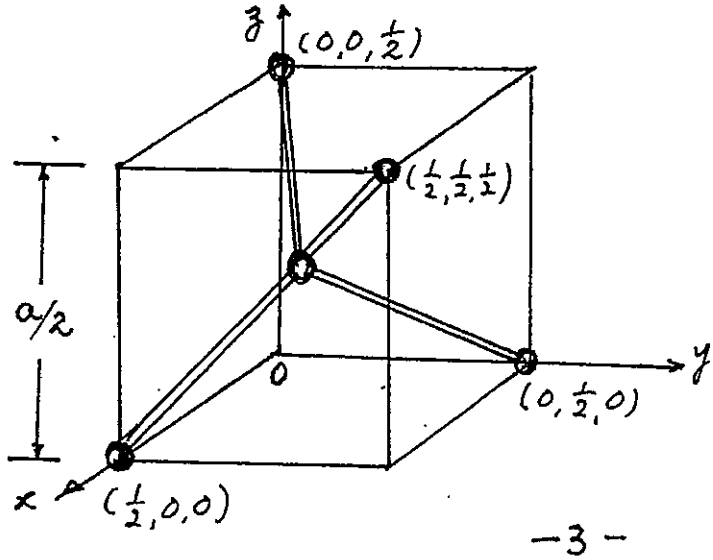


bcc

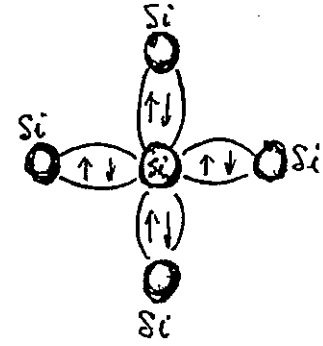


fcc

- Diamond lattice is formed from two interpenetrating fcc lattices with one sublattice displaced from the other by $\frac{1}{4}$ of distance along the body diagonal of the cube (i.e., $a\sqrt{3}/4$).
- Zincblende lattice: same as the diamond lattice except, for example, one sublattice is Ga, other is As.



Tetrahedral Bond

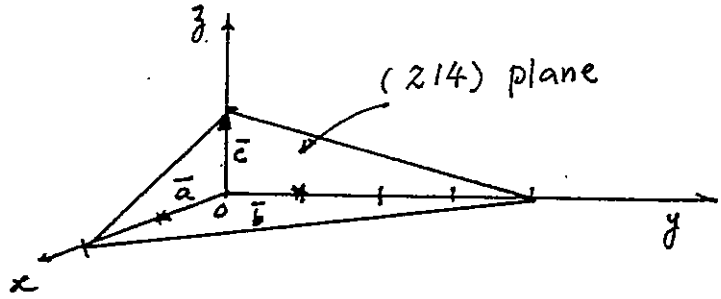


- Miller indices

  (1) intercept of plane with three axes in terms of lattice constant

  (2) reciprocals of the numbers

  (3) reducing to smallest 3 integers having the same ratio



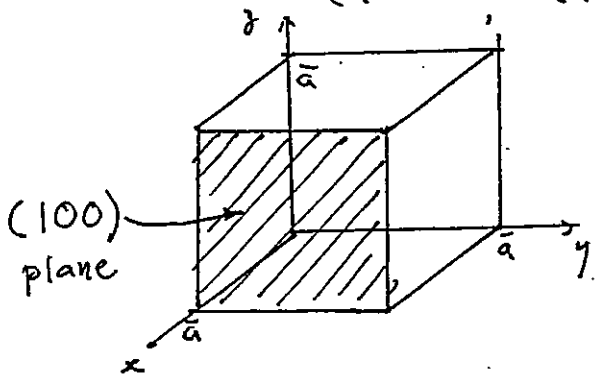
(1) $2\bar{a}$, $4\bar{b}$, $\bar{c}$

(2) $\frac{1}{2}$, $\frac{1}{4}$, 1

(3) (214) plane

- most important planes are shown in Fig. 2

(100) (110), (111)



(1) a, ∞, ∞

(2) $\frac{1}{a}$, 0, 0

(3) (100) plane

- Easiest breakage (cleavage) plane is (111) for Si
  (110) for GaAs

1.2.2 Reciprocal Lattice

$$\bar{G} = h\bar{a}^* + k\bar{b}^* + l\bar{c}^* \quad (h, k, l \text{ are integers})$$

$$\bar{a}^* = 2\pi \frac{\bar{b}\times\bar{c}}{\bar{a}\cdot\bar{b}\times\bar{c}} \qquad \bar{b}^* = 2\pi \frac{\bar{c}\times\bar{a}}{\bar{a}\cdot\bar{b}\times\bar{c}} \qquad \bar{c}^* = 2\pi \frac{\bar{a}\times\bar{b}}{\bar{a}\cdot\bar{b}\times\bar{c}}$$

$$\therefore \bar{a}\cdot\bar{a}^* = 2\pi, \quad \bar{a}\cdot\bar{b}^* = 0 \quad \text{etc}$$

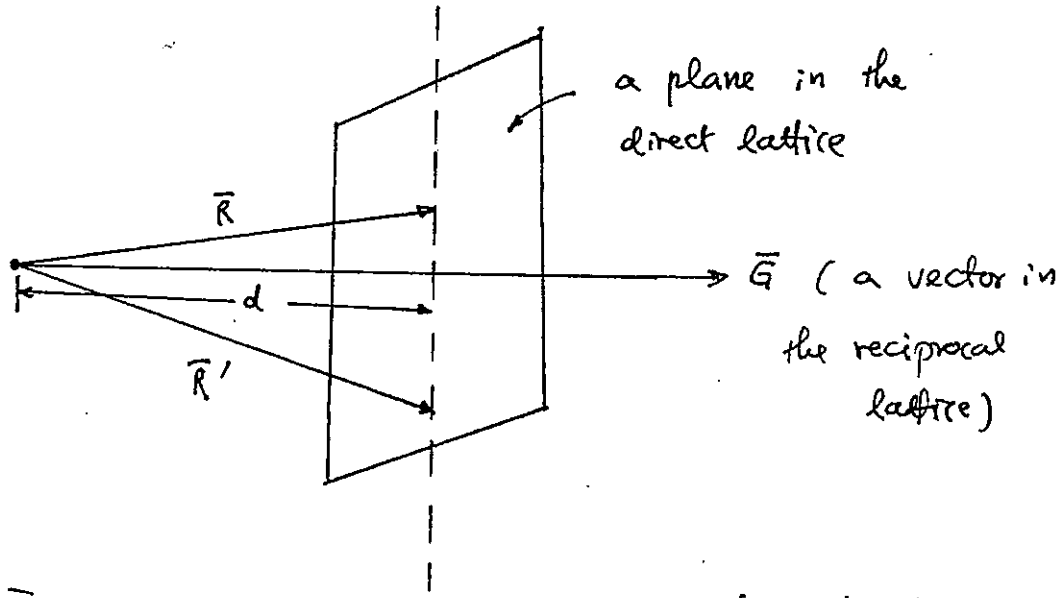
(p.11 of textbook)

Therefore

$$\bar{G} \cdot \bar{R} = (ha^* + kb^* + lc^*) \cdot (m\bar{a} + n\bar{b} + p\bar{c})$$

$$= 2\pi (hm + kn + lp) = \text{integer} \times 2\pi$$

- To Prove: a vector in the reciprocal lattice is normal to a set of planes in the direct lattice.



$$\bar{G} \cdot \bar{R} = 2\pi (hm + kn + lp) = 2\pi N \quad (N = \text{integer})$$

Let $\bar{R}' = (m - Zl)\bar{a} + (n - Zl)\bar{b} + [p + Z(h+k)]\bar{c}$

$$\therefore \bar{G} \cdot \bar{R}' = 2\pi \{ h(m - Zl) + k(n - Zl) + l(p + Z(h+k)) \}$$

$$= 2\pi (hm + kn + lp) = 2\pi N = \bar{G} \cdot \bar{R}$$

$\therefore \bar{R}'$ has the same projection on $\bar{G}$ and must be on the plane normal to $\bar{G}$, at a distance d from the origin. We have constructed one of the lattice planes. Q.E.D.

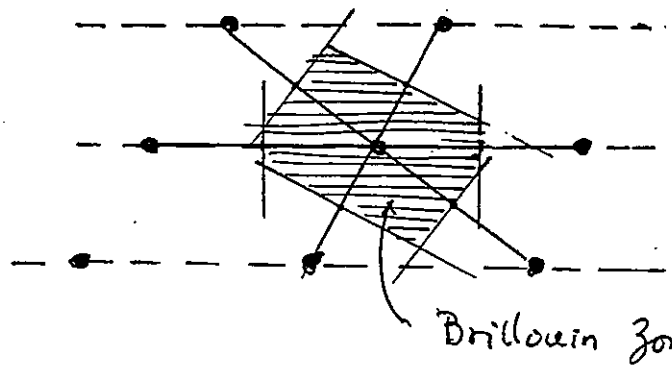
- Volume in the reciprocal lattice

$$a^* \cdot b^* \times c^* = (2\pi)^3 \frac{(\bar{b}\times\bar{c})\cdot(\bar{c}\times\bar{a})\times(\bar{a}\times\bar{b})}{(\bar{a}\cdot\bar{b}\times\bar{c})^3}$$

$$= \frac{(2\pi)^3}{\bar{a}\cdot\bar{b}\times\bar{c}} = \frac{8\pi^3}{V_c}$$

where V_c is the volume in the direct lattice.

- Unit cell in reciprocal lattice for 2D



lines bisecting the midpoint of the reciprocal lattice

For 3D (see Fig. 3 on p. 11)

- A body-centered cubic lattice may have many different sets of primitive basic vectors. A simple set can be:

$$\bar{a} = a\bar{x}\ ,\quad \bar{b} = a\bar{y}\ ,\quad \bar{c} = \frac{a}{2}(\bar{x}+\bar{y}+\bar{z})$$ (Fig. A) P. 7

the point $\bar{P} = -\bar{a} - \bar{b} + 2\bar{c}$

A more symmetric set is:

$$\bar{a} = \frac{a}{2}(\bar{y}+\bar{z}-\bar{x}),\quad \bar{b} = \frac{a}{2}(\bar{z}+\bar{x}-\bar{y}),\quad \bar{c} = \frac{a}{2}(\bar{x}+\bar{y}-\bar{z})$$

the point $\bar{P} = 2\bar{a} + \bar{b} + \bar{c}$ (Fig. B. P. 7)

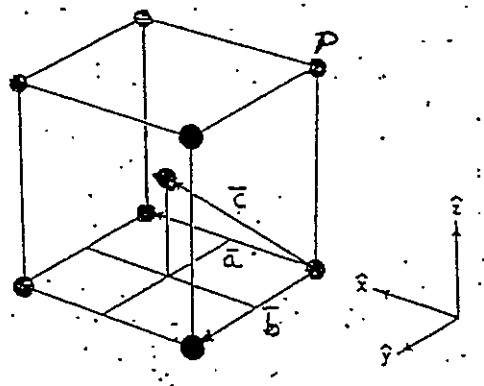


Fig A

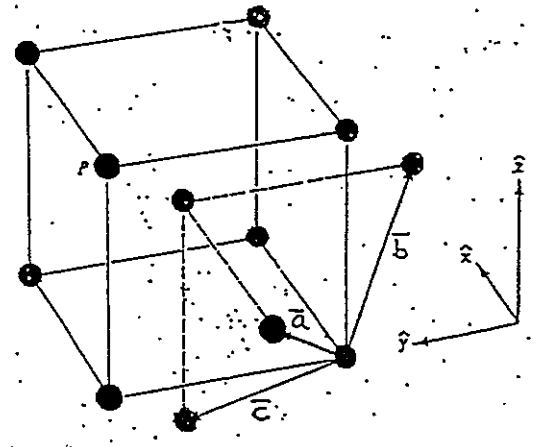


Fig. B

- A face-centered cubic lattice has a symmetric set of primitive basis vectors:

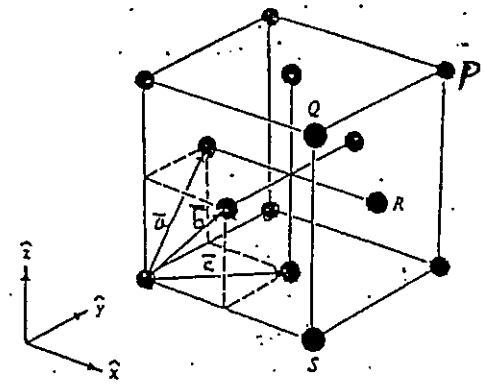
$$\bar{a} = \frac{a}{2}(\bar{y}+\bar{z}), \quad \bar{b} = \frac{a}{2}(\bar{z}+\bar{x}), \quad \bar{c} = \frac{a}{2}(\bar{x}+\bar{y})$$

$\therefore P = \bar{a}+\bar{b}+\bar{c}$

$Q = 2\bar{b}$

$S = -\bar{a}+\bar{b}+\bar{c}$

$R = \bar{b}+\bar{c}$



- 3D Wigner-Seitz cell

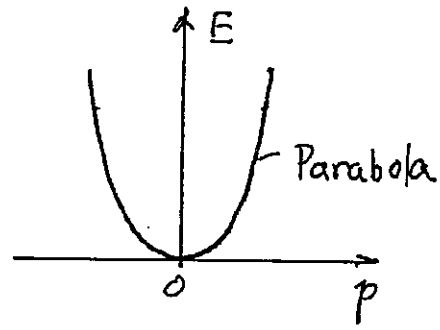
Direct lattice (bcc) ⟶ reciprocal (fcc)

Direct lattice (fcc) ⟶ reciprocal (bcc)

Symmetry points and symmetry lines

Γ : $\frac{2\pi}{a}(000)$	zone center	
L : $\frac{2\pi}{a}(\frac{1}{2}\frac{1}{2}\frac{1}{2})$	zone edge along <111> axis	Λ
X : $\frac{2\pi}{a}(001)$	" " " <100> "	Δ
K : $\frac{2\pi}{a}(\frac{3}{4}\frac{3}{4}0)$	" " " <110) "	Σ

1.3 Energy Bands and Energy Gap



For a free electron

$$E = p^2/2m$$

or $$\frac{d^2E}{dp^2} = \frac{1}{m} \quad \text{and} \quad m = \frac{1}{(d^2E/dp^2)}$$

Therefore, the electron mass is inversely proportional to the second derivative of E vs p.

For electrons in a crystalline solid, we must consider the influence of the lattice potential.

- In classic mechanics, we have Newton's Law:

$$F = ma$$

$$a \equiv \frac{d^2x}{dt^2} \quad \text{and} \quad F = -\frac{dV(x)}{dx}$$ where $V(x)$ is the potential energy

$$\therefore \; -\frac{dV(x)}{dx} = m\frac{d^2x}{dt^2}$$

the above equation with initial conditions can be solved to find $x(t)$, i.e., the position of a particle at a given time.

- In quantum mechanics, we have the Schrödinger equation:

$$i\hbar\frac{\partial \Phi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2 \Phi}{\partial x^2} + V(x)\Phi$$

where $\hbar = \frac{h}{2\pi}$ = reduced Planck's constant $= 1.054 \times 10^{-34}$ J·s.

We can find the wave function $\Phi(x,t)$ from the above

equation, which is the probability of finding the particle at point x at time t.

$$\int_{-\infty}^{\infty} |\Psi(x,t)|^2 dx = 1$$ The particle got to be somewhere.

We can use separation of variables, if

$$\Psi(x,t) = \phi(x) f(t)$$

Then $\frac{d\Psi}{dt} = \phi \frac{df}{dt}$, $\frac{d^2\Psi}{dx^2} = \frac{d^2\phi}{dx^2} \cdot f$

The Schrödinger equation becomes

$$i\hbar \phi \frac{df}{dt} = -\frac{\hbar^2}{2m} \frac{d^2\phi}{dx^2} f + V\phi f$$

Dividing by ϕf on both sides

$$\underbrace{i\hbar \frac{1}{f} \frac{df}{dt}}_{\text{function of } t} = \underbrace{-\frac{\hbar^2}{2m} \frac{1}{\phi} \frac{d^2\phi}{dx^2} + V}_{\text{function of } x} = \text{constant} \equiv E$$

$$\therefore i\hbar \frac{1}{f} \frac{df}{dt} = E \quad \text{or} \quad f(t) = e^{-iEt/\hbar}$$

and

$$-\frac{\hbar^2}{2m} \frac{d^2\phi}{dx^2} + V(x)\phi = E\phi \qquad 1D$$

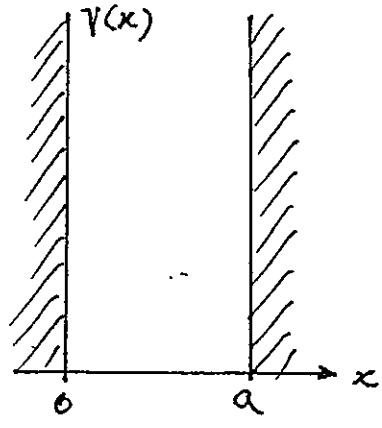
$$-\frac{\hbar^2}{2m} \nabla^2\phi + V(\vec{r})\phi = E\phi \qquad 3D$$

This is the time-independent Schrodinger equation.

- Example 1 Infinite square well

  For $x < 0$, $x > a$ $\phi(x) = 0$

  (i.e., zero probability to find a particle there).

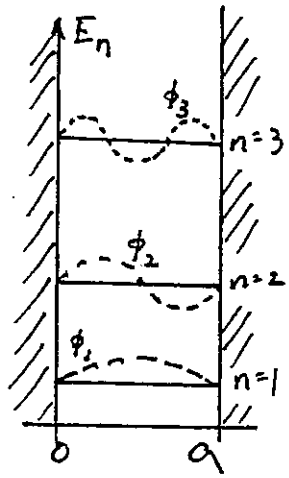


For $0 \le x \le a$ $\quad V(x) = 0$

$$-\frac{\hbar^2}{2m}\frac{d^2\phi}{dx^2} = E\phi$$

wave vector

or $\frac{d^2\phi}{dx^2} = -\frac{2mE}{\hbar^2}\phi \equiv -k^2\phi$ $\qquad k \equiv \frac{\sqrt{2mE}}{\hbar}$ (in 1/cm)



solution : $\phi = A\sin kx + B\cos kx$

Boundary condition : $\phi(0) = 0$, $\phi(a) = 0$

$\therefore$ $A\sin 0 + B\cos 0 = B = 0$ $\qquad \therefore$ $B = 0$

$\phi(x) = A\sin kx$

$\phi(a) = 0 = A\sin ka$

$\therefore ka = \pm\pi, \pm 2\pi, \pm 3\pi \cdots$

or $\frac{\sqrt{2mE}}{\hbar}a = n\pi$ $\quad (n = \pm 1, \pm 2, \pm 3 \cdots)$

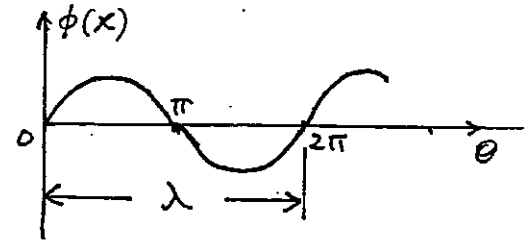
$$\therefore E_n = \frac{\hbar^2 n^2 \pi^2}{2ma^2} \qquad (n = 1, 2, 3 \cdots)$$

a quantum particle in an infinite well can have only discrete energy levels.

$$\int_0^a |\phi(x)|^2 dx = \int_0^a |A|^2 \sin^2 kx\, dx = |A|^2 \frac{a}{2} = 1$$

$$\therefore A = \sqrt{\frac{2}{a}}$$

$$\boxed{\phi(x) = \sqrt{\frac{2}{a}}\sin\left(\frac{n\pi}{a}x\right)}$$



Let $kx \equiv \theta$ where $\theta = 2\pi$

$x = \lambda$ = wavelength

$\therefore k\lambda = 2\pi$ or $\boxed{k = \frac{2\pi}{\lambda}}$ (wave vector, 1/cm)

$\boxed{\lambda = \frac{2\pi}{k}}$ (wavelength, cm)

- Example 2 Free Particle

$$V(x) = 0 \qquad -\frac{\hbar^2}{2m}\frac{d^2\phi}{dx^2} = E\phi$$

$$\therefore \frac{d^2\phi}{dx^2} = -k^2\phi \qquad k \equiv \frac{\sqrt{2mE}}{\hbar}$$

$$\phi(x) = \underbrace{Ae^{jkx}}_{\text{move to right}} + \underbrace{Be^{-jkx}}_{\text{move to left}}$$

Consider the 1st term only. For free electrons, the wavefunction is given by a plane wave. In 3D

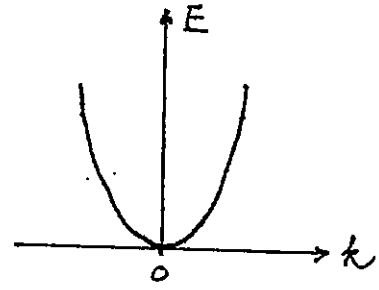
$$\phi_k(\bar{r}) = Ce^{j\bar{k}\cdot\bar{r}}$$

the de Broglie wavelength is $\lambda = \frac{h}{p}$

$$\therefore p = \frac{h}{\lambda} = \frac{h}{\frac{2\pi}{k}} = \frac{hk}{2\pi} = \hbar k$$

Thus, in quantum mechanics, the momentum associated with an excitation of wave vector k is $\hbar k$. The energy of a free particle is then

$$E = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m} \sim k^2$$



- Bloch Function. In a crystal, the electron wave function is given by Bloch function

$$\phi_k(\bar{r}) = e^{j\bar{k}\cdot\bar{r}}\, u_n(\bar{k}, \bar{r})$$

where $u_n(\bar{k}, \bar{r})$ is periodic with periodicity of the direct lattice.

We now <u>prove the Bloch Theorem</u> (p. 12 of textbook)

Crystal symmetry requires that crystal potential has the periodicity of the lattice site, i.e.,

$$V(\bar{r}) = V(\bar{r}+\bar{R}) \quad \text{where } \bar{R} \text{ is a direct lattice site.}$$

Therefore $|\phi(\bar{r})|^2 = |\phi(\bar{r}+\bar{R})|^2$

and $\phi(\bar{r})$ and $\phi(\bar{r}+\bar{R})$ can only differ by a constant A

$$AA^* = 1$$

Let $A = e^{j\bar{k}\cdot\bar{R}}$ Then $AA^* = e^{j\bar{k}\cdot\bar{R}} e^{-j\bar{k}\cdot\bar{R}} = 1$

We have $$\phi(\bar{r}+\bar{R}) = A\phi(\bar{r}) = e^{j\bar{k}\cdot\bar{R}}\phi(\bar{r}) \quad \cdots\cdots (1)$$

$$\phi(\bar{r}) = \frac{1}{e^{j\bar{k}\cdot\bar{R}}}\phi(\bar{r}+\bar{R}) = e^{-j\bar{k}\cdot\bar{R}}\phi(\bar{r}+\bar{R})$$

$$= e^{+j\bar{k}\cdot\bar{r}} \underbrace{\exp[-j\bar{k}(\bar{r}+\bar{R})]\phi(\bar{r}+\bar{R})}_{\equiv u(\bar{r})} \quad \text{---------- (2)}$$

Now we shall prove $u(\bar{r})$ is periodic with the periodicity of the crystal lattice. From Eq. (2) above, for any lattice vector $\bar{b}$:

$$u(\bar{r}+\bar{b}) = \exp[-j\bar{k}(\bar{r}+\bar{b}+\bar{R})]\phi(\bar{r}+\bar{b}+\bar{R})$$

$$\phi(\bar{r}+\bar{b}+\bar{R})\Big|_{(\bar{r}+\bar{R})\equiv\bar{r}'} = \phi(\bar{r}'+\bar{b}) = e^{j\bar{k}\cdot\bar{b}}\phi(\bar{r}') \quad \text{from (1)}$$

$$= e^{j\bar{k}\cdot\bar{b}}\phi(\bar{r}+\bar{R})$$

$$\therefore u(\bar{r}+\bar{b}) = \exp[-j\bar{k}(\bar{r}+\bar{b}+\bar{R})]e^{j\bar{k}\cdot\bar{b}}\phi(\bar{r}+\bar{R})$$

$$= \exp[-j\bar{k}(\bar{r}+\bar{R})]\phi(\bar{r}+\bar{R}) = u(\bar{r}) \quad \text{---------- (3)}$$

Q.E.D.

- The energy $E_{\bar{k}}$ is periodic in the reciprocal lattice.

i.e., $E_{\bar{k}} = E_{\bar{k}+\bar{G}}$.

Prove: $\phi_{\bar{k}}(\bar{r}) = e^{j\bar{k}\cdot\bar{r}} u_{\bar{k}}(\bar{r})$ Block Function ---(1)'

$$= e^{j(\bar{k}+\bar{G})\cdot\bar{r}} e^{-j\bar{G}\cdot\bar{r}} u_{\bar{k}}(\bar{r})$$

$$\phi_{\bar{k}+\bar{G}}(\bar{r}) = e^{j(\bar{k}+\bar{G})\cdot\bar{r}} u_{\bar{k}+\bar{G}}(\bar{r}) \quad \text{----------(2)'}$$

If $u_{\bar{k}+\bar{G}}(\bar{r}) = e^{-j\bar{G}\cdot\bar{r}} u_{\bar{k}}(\bar{r})$ ----------(3)'

Then $\phi_{\bar{k}+\bar{G}}(\bar{r}) = e^{j(\bar{k}+\bar{G})\cdot\bar{r}} e^{-j\bar{G}\cdot\bar{r}} u_{\bar{k}}(\bar{r})$

$$= e^{j\bar{k}\cdot\bar{r}} u_{\bar{k}}(\bar{r}) \quad \left[= \text{Eq (1)}'\right] \text{-----(4)}'$$

Thus Eq (4)' is a solution of the Schrödinger equation for the same energy because it is the same as Eq (1)'.

We now prove $u_{\bar{k}+\bar{G}}(\bar{r})$ is a periodic function of $\bar{R}$.

$$u_{\bar{k}+\bar{G}}(\bar{r}+\bar{R}) = e^{-j\bar{G}\cdot(\bar{r}+\bar{R})} u_{\bar{k}}(\bar{r}+\bar{R})$$

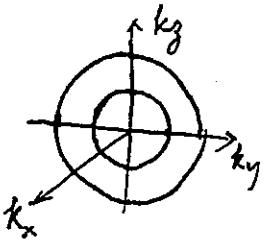
$$= e^{-j\bar{G}\cdot\bar{r}} \underbrace{e^{-j\bar{G}\cdot\bar{R}}}_{=1} \underbrace{u_{\bar{k}}(\bar{r}+\bar{R})}_{=u_{\bar{k}}(\bar{r}) \text{ from Eq.(3)}} = e^{-j\bar{G}\cdot\bar{r}} u_{\bar{k}}(\bar{r})$$

$$= \text{Eq (3)}' = u_{\bar{k}+\bar{G}}(\bar{r})$$

$\therefore \bar{k}$ and $(\bar{k}+\bar{G})$ are equivalent. The energy $E_{\bar{k}}$ is a periodic function of $\bar{k}$ in the reciprocal lattice. We need only to consider the restricted value of k (for 1D case $-\frac{\pi}{a} \le k \le \frac{\pi}{a}$) — the first Brillouin zone.

- Constant energy surfaces

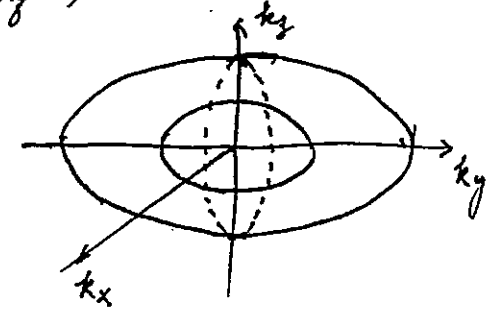
Near the minimum of the conduction band, E vs. k can be approximated by:



$$E(k) = \frac{\hbar^2 k^2}{2m_n^*} \quad \text{for spherical (GaAs)}$$

or

$$E(k) = \frac{\hbar^2}{2}\left(\frac{k_x^2}{m_x^*} + \frac{k_y^2}{m_y^*} + \frac{k_z^2}{m_z^*}\right) \quad \text{ellipsoidal}$$

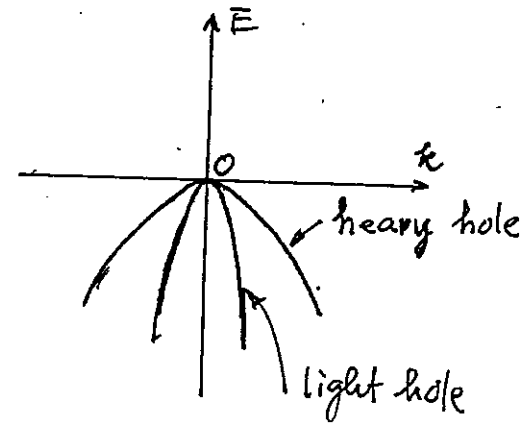


Similary for the valence band

$$E(k) = E_1 - \frac{\hbar^2 k^2}{2m_p^*}$$

The effective mass tensor:

$$m_{ij} = \frac{\hbar^2}{\left(\frac{\partial^2 E}{\partial k_i \partial k_j}\right)}$$



- The temperature dependence of E_g

$$E_g(T) = E_g(0) - \frac{\alpha T^2}{T+\beta}$$ <see p. 16. Fig. 6>

@ 300K $E_g = 1.12$ eV for Si

$= 1.42$ eV for GaAs

- The pressure dependence for Si

$$\frac{dE_g}{dP} = -2.4\times10^{-6}\ \text{eV}/(\text{N/cm}^2) = -2.4\times10^{-10}\ \text{eV}/(\text{N/m}^2)$$
$$= -2.4\times10^{-10}\ \text{eV/Pa}$$

(1 atm = 760 mm Hg = 760 Torr = 1.013×10^5 Pa = 1.013×10^5 N/m^2).

1.4 Carrier Concentration at Thermal Equilibrium

phosphorus (noun)
phosphorous (adjective)
Complementary (Mutually supplying each other's lack) CMOS
Complimentary (given free as a favor)
Dielectric permittivity ϵ_s (F/cm)
Dielectric constant ϵ_s/ϵ_0 (no units)

- Derivation of the density of states in a semiconductor

To calculate the electron and hole concentrations in the conduction and valence bands, respectively, we need to know the density of states, that is, the number of allowed energy states per unit energy per unit volume (i.,e., in the units of number of states /eV/cm^3).

When electrons move back and forth along the x-direction in a semiconductor material, the movements can be described by standing-wave oscillations. The wavelength λ of a standing wave is related to the length of the semiconductor L by

$$\frac{L}{\lambda} = n_x, \tag{1}$$

where n_x is an integer. The wavelength can be expressed by de Broglie hypothesis:

$$\lambda = \frac{h}{p_x}, \tag{2}$$

where h is the Planck constant and p_x is the momentum in the x-direction. Substituting Eq. 2 into Eq. 1 gives

$$Lp_x = hn_x. \tag{3}$$

The incremental momentum dp_x required for a unity increase in n_x is

$$L\,dp_x = h. \tag{4}$$

For a three-dimensional cube of side L, we have

$$L^3\,dp_x\,dp_y\,dp_z = h^3. \tag{5}$$

The volume $dp_x\,dp_y\,dp_z$ in the momentum space for a unit cube (L = 1) is thus equal to h^3. Each incremental change in n corresponds to a unique set of integers (n_x, n_y, n_z), which in turn corresponds to an allowed energy state. Thus, the volume in momentum space for an energy state is h^3. The figure shows the momentum space in spherical coordinates. The volume between two concentric spheres (from p to $p + dp$) is $4\pi p^2 dp$. The number of energy states contained in the volume is then $2(4\pi p^2 dp)/h^3$, where the factor 2 accounts for the electron spins.

The energy E of the electron (here we only consider the kinetic energy) is given by

$$E = \frac{p^2}{2m_n} \tag{6}$$

or

$$p = \sqrt{2m_n E}, \tag{7}$$

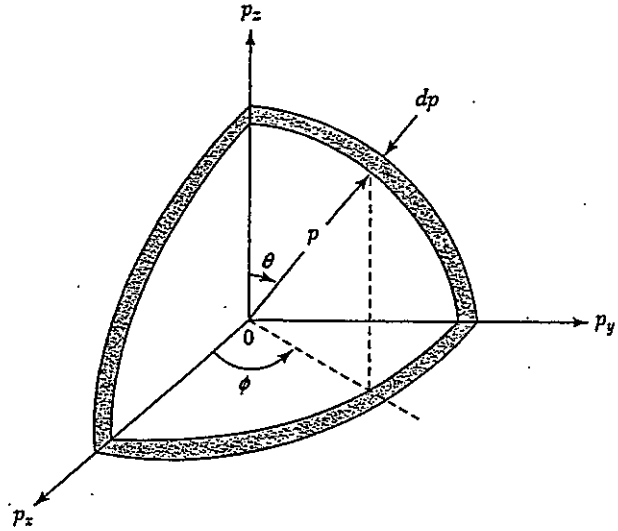
where p is the total momentum (with components p_x, p_y, and p_z in Cartesian Coordinates) and m_n is the effective mass. From Eq. 7, we can substitute E for p and obtain

$$N(E)dE = \frac{8\pi p^2 dp}{h^3} = 4\pi\left(\frac{2m_n}{h^2}\right)^{\frac{3}{2}} E^{\frac{1}{2}}\, dE \tag{8}$$

and

$$N(E) = 4\pi\left(\frac{2m_n}{h^2}\right)^{\frac{3}{2}} E^{\frac{1}{2}}, \tag{9}$$

where $N(E)$ is called the density of states.



- Gamma Function

$$\Gamma(n) \equiv \int_0^\infty x^{n-1} exp(-x)\, dx$$

n	$\Gamma(n)$
1	1.00
1.25	0.906
1.5	0.886
1.75	0.919
2	1.00

$\Gamma(\frac{1}{2}) = \sqrt{\pi}$

For positive integer values of n

$$\Gamma(n) = (n-1)!$$

The values of $\Gamma(n)$ for $n < 1$ and for $n > 2$ may be evaluated using the following formulas:

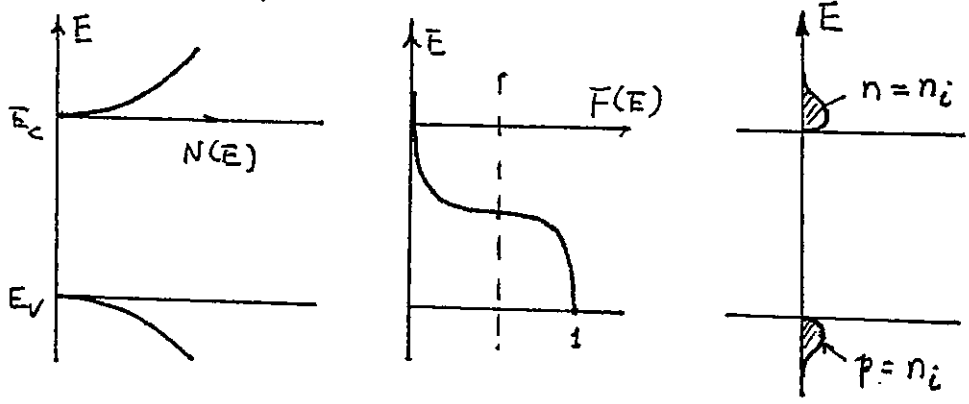
$$\Gamma(n) = \Gamma(n+1)/n$$

$$\Gamma(n) = (n-1)\Gamma(n-1)$$

Example: $\Gamma(2.5) = 1.5\Gamma(1.5) = 1.33$

1.4.1 Carrier Concentration and Fermi Level

- Derivation of Eq. 21 (p. 19)



$$n = \int_{E_c}^{\infty} N(E)\,F(E)\,dE$$

$$= \int_{E_c}^{\infty} \underbrace{M_c \frac{\sqrt{2}}{\pi^2}\frac{(E-E_c)^{1/2}}{\hbar^3}(m_{de})^{3/2}}_{N(E)} \cdot \underbrace{\frac{1}{1+\exp\left(\frac{E-E_F}{kT}\right)}}_{F(E)} dE$$

Let $\eta \equiv (E-E_c)/kT$ $\qquad \eta_f \equiv (E_F-E_c)/kT$

$$N_c \equiv 2\left(\frac{2\pi m_{de} kT}{h^2}\right)^{3/2} M_c \qquad F_{\frac{1}{2}} \equiv \int_0^{\infty} \frac{\eta^{1/2}\,d\eta}{[1+e^{(\eta-\eta_f)}]}$$

$$\therefore\ n = \left(\frac{2}{\sqrt{\pi}}N_c\right)F_{\frac{1}{2}} = \left(\frac{2}{\sqrt{\pi}}N_c\right)\int_0^{\infty}\frac{\eta^{1/2}\,d\eta}{1+e^{\eta-\eta_f}}$$

$n-1=\frac{1}{2},\ n=\frac{3}{2}$

If $(\eta-\eta_f) \gg 1$

$$n = \left(\frac{2}{\sqrt{\pi}}N_c\right)\int_0^{\infty} e^{-(\eta-\eta_f)}\eta^{1/2}\,d\eta = \left(\frac{2}{\sqrt{\pi}}N_c\right)e^{\eta_f}\int_0^{\infty} e^{-\eta}\eta^{1/2}\,d\eta$$

$$= \left(\frac{2}{\sqrt{\pi}}N_c e^{\eta_f}\right)\Gamma\left(\frac{3}{2}\right) = \left(\frac{2}{\sqrt{\pi}}N_c e^{\eta_f}\right)\left(\frac{\sqrt{\pi}}{2}\right) = N_c e^{\eta_f}$$

$$= N_c \exp\left(-\frac{E_c-E_F}{kT}\right) \qquad \text{--------- (21)}$$

[$\Gamma\left(\frac{3}{2}\right) = \left(\frac{3}{2}-1\right)\Gamma\left(\frac{3}{2}-1\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$]

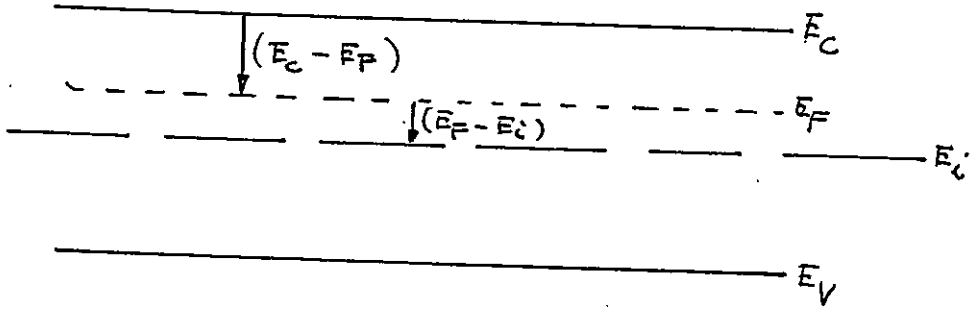
• Mass action law:

$$n = N_C \exp\left(-\frac{E_C - E_F}{kT}\right) \quad \text{---------- (21)}$$

$$p = N_V \exp\left(-\frac{E_F - E_V}{kT}\right) \quad \text{---------- (23)}$$

$$\therefore np = N_C N_V \exp\left(-\frac{E_C - E_V}{kT}\right) = N_C N_V \exp\left(-\frac{E_g}{kT}\right)$$

= independent of E_F.



From Eq. 21, assume full ionization ($n = N_D$)

$$n = N_D = N_C \exp\left(-\frac{E_C - E_F}{kT}\right)$$

$$\therefore E_C - E_F = kT \ln\left(\frac{N_C}{N_D}\right)$$

For intrinsic case $E_F \longrightarrow E_i$, $n \rightarrow n_i$

$$\therefore n = n_i = N_C \exp\left(-\frac{E_C - E_i}{kT}\right)$$

or $$E_C - E_i = kT \ln \frac{N_C}{n_i} \quad \text{---------- (Q)}$$

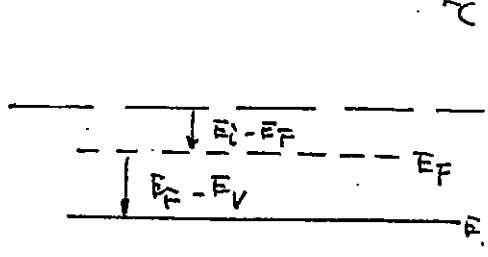
Eq. Q − Eq. 21

$$E_F - E_i = (E_F - E_C) + (E_C - E_i) = kT \ln\left(\frac{N_D}{N_C}\right) + kT \ln\left(\frac{N_C}{n_i}\right)$$

$$= kT \ln\left(\frac{N_D}{n_i}\right) \approx kT \ln\left(\frac{n}{n_i}\right)$$

Similarly $$E_F - E_V = kT \ln\left(\frac{N_V}{N_A}\right)$$

$$E_i - E_F = kT \ln\left(p/n_i\right)$$



1.4.2 Donors and Acceptors

- Derivation of Eq. 34 (p. 22)

The probability of occupancy: $F(E) = \dfrac{1}{1+\frac{h}{g}\exp\left(\frac{E-E_F}{kT}\right)}$

h = number of electrons that can physically occupy the level E

g = " " " " " be accepted by the level E

(1) For free electrons in the conduction band, each level can accept 2 electrons ($\uparrow\downarrow$ spins), and actually 2 electrons can occupy that energy level, thus $h=2$, $g=2$, $\frac{h}{g}=1$.

(2) For donor level, once a donor level is filled by an electron it can no longer accept another electron: $\therefore h=1$, $g=2$

$$F(E) = \frac{1}{1+\frac{1}{2}\exp\left(\frac{E-E_F}{kT}\right)}$$

$$\therefore N_D^+ = N_D\left[1-F(E)\right] = N_D\left[1-\frac{1}{1+\frac{1}{2}\exp\left(\frac{E_D-E_F}{kT}\right)}\right]$$

$$= \frac{N_D}{1+2\exp\left(\frac{E_F-E_D}{kT}\right)} \qquad \text{---------- (34)}$$

(3) For holes, $h=1$, $g=4$ due to two bands at $k=0$

$$\therefore N_A^- = N_A\left[1-\frac{1}{1+\frac{1}{4}\exp\left(\frac{E_F-E_A}{kT}\right)}\right]$$

$$= \frac{N_A}{1+4\exp\left(\frac{E_A-E_F}{kT}\right)} \qquad \text{---------- (35)}$$

• Derivation of Eqs. 39 and 40 (p. 25 of textbook)

From charge neutrality condition

$$n + N_A^- = p + N_D^+$$

or

$$n = N_D^+ - N_A^- + p$$

Assume p is very small and $N_A^- = N_A$.

From Eq. 34

$$n = N_D\left[1 - \frac{1}{1+\frac{1}{2}\exp\left(\frac{E_D - E_F}{kT}\right)}\right] - N_A$$

$$= (N_D - N_A) - \frac{N_D}{1+\frac{1}{2}\frac{N_c}{N_c}\exp\left(\frac{E_c - E_F}{kT}\right)\exp\left(-\frac{E_c - E_D}{kT}\right)}$$

$$= (N_D - N_A) - \left.\frac{N_D}{1+\frac{\frac{1}{2}N_c\exp(-E_d/kT)}{n}}\right| \begin{array}{l} n = N_c\exp\left(-\frac{E_c - E_F}{kT}\right) \\ E_d \equiv E_c - E_D \end{array}$$

$$= (N_D - N_A) - \left.\frac{N_D n}{n + N_c'}\right| N_c' \equiv \frac{1}{2}N_c\exp(-E_d/kT)$$

$$\therefore\ n^2 + n(N_c' + N_A) - (N_D - N_A)N_c' = 0$$

If $N_A = 0$, we have from above equation

$$n^2 + nN_c' - N_D N_c' = 0$$

$$n = \left.\frac{-N_c' + \sqrt{N_c'^2 + 4N_D N_c'}}{2}\right| \approx \frac{\sqrt{4N_D N_c'}}{2} = \sqrt{N_D N_c'} \quad N_D \gg N_c'$$

$$= \frac{1}{\sqrt{2}}(N_D N_c)^{1/2}\exp(-E_d/2kT) \quad \text{------(39)}$$

If $N_A \gg N_c'$

$$n = \frac{-N_A + \sqrt{N_A^2 + 4(N_D - N_A)N_c'}}{2} \approx \frac{-N_A + N_A\left[1 + \frac{4}{2}\frac{(N_D - N_A)N_c'}{N_A^2}\right]}{2}$$

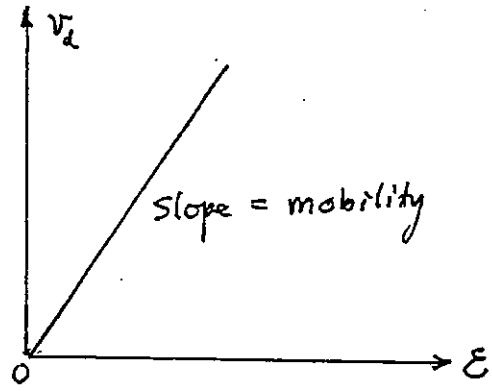
$$= (N_D - N_A)N_c'/N_A = \left(\frac{N_D - N_A}{2N_A}\right)N_c\exp(-E_d/kT) \quad \text{----(40}$$

1.5 Carrier Transport Phenomena (p. 28)

1.5.1 Drift and Mobility

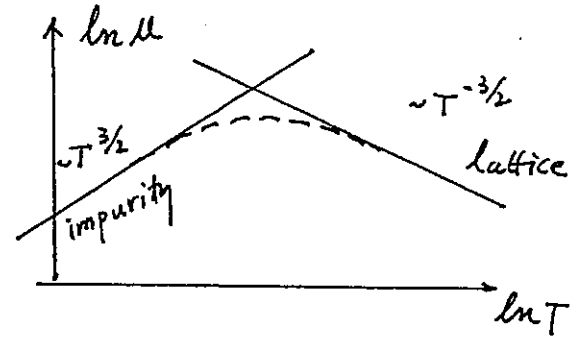
$v_d \sim \mathcal{E}$

$v_d = \mu \mathcal{E}$ ..... (48)

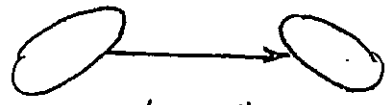


$\mu_l \sim (m^*)^{-5/2} (T)^{-3/2}$

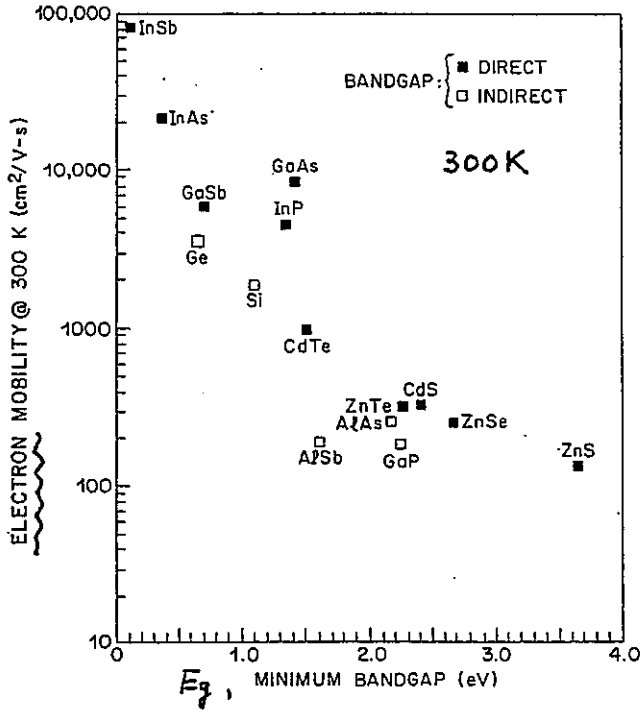
$\mu_i \sim (N_I)^{-1} (m^*)^{-1/2} T^{3/2}$



intravalley
long-wavelength
phonon



intervalley
energetic phonon



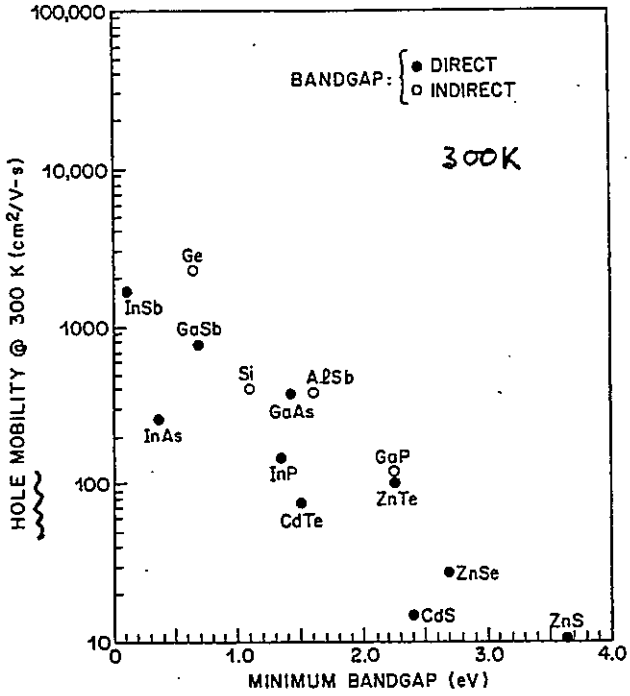
Low-field electron mobilities of the common semiconductors

- $\mu_n \sim \frac{1}{E_g}$.

At a given E_g

μ_n (direct bandgap)

$> \mu_n$ (indirect bandgap)



Low-field hole mobilities of the common semiconductors (at room temperature).

$\mu_p \sim \frac{1}{E_g}$

$\mu_n > \mu_p$ for all semiconductors shown (except AlSb).

1.5.2 Resistivity and Hall Effect

$$\rho\ (\text{resistivity}) = \frac{1}{\sigma\ (\text{conductivity})} = \frac{1}{q(\mu_n n + \mu_p p)}$$

n, p are carrier concentrations which are equal or less than the impurity concentration (i.e., $n \leq N_D$)

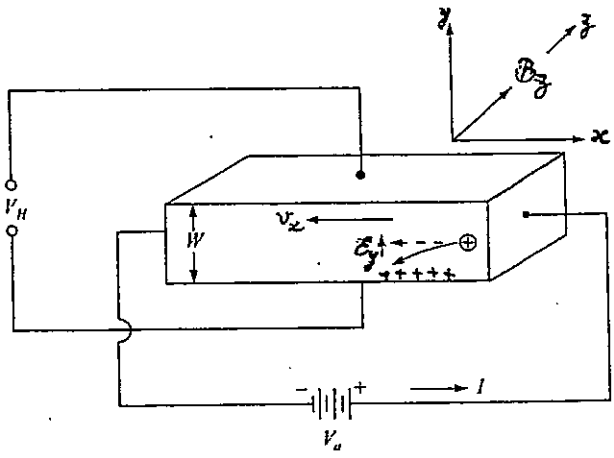
Example: $N_A = 10^{17}\ cm^{-3}$ $\quad (E_A - E_V) = 0.073\ eV$ (for Ga).

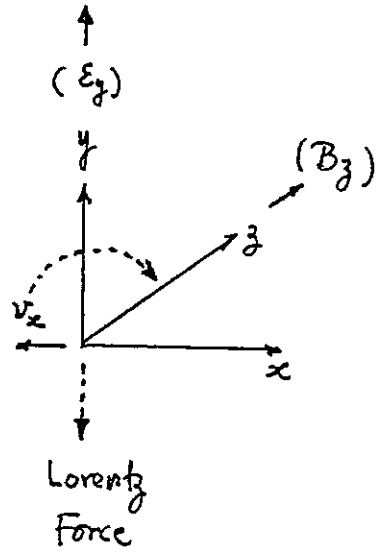
We find $E_F - E_V = 0.14\ eV$

From Eq. 35 $\quad N_A^- = \dfrac{10^{17}}{1 + 4\exp\left(\dfrac{0.073 - 0.14}{0.0259}\right)} = 7.7 \times 10^{16}\ cm^{-3}$ ionized

$(\therefore\ 2.3 \times 10^{16}\ cm^{-3}$ unionized!)

- Hall Effect





$$\underbrace{q\mathcal{E}_y}_{\text{Force due to electric field}} = \underbrace{q v_x B_z}_{\text{Lorentz force due to magnetic field}}$$

$$\therefore \quad \mathcal{E}_y = v_x B_z$$

$$J_p = q v_x p \quad (\text{drift current only})$$

$$v_x = \left(J_p / qp\right)$$

$$\therefore \quad \mathcal{E}_y = \left(\frac{J_p}{qp}\right) B_z = \left(\frac{1}{qp}\right) J_p B_z = R_H J_p B_z$$

$$\therefore \quad R_H \equiv \frac{1}{qp} \ (\text{holes}), \qquad R_H = -\frac{1}{qn} \ (\text{electrons})$$

R_H: Hall coefficient

$$\therefore \quad p = \frac{1}{qR_H} = \frac{1}{q\left(\mathcal{E}_y / J_p B_z\right)} = \frac{J_p B_z}{q\mathcal{E}_y} = \frac{(I/A) B_z}{q(V_H/W)}$$

Example: $N_D = 10^{16}\ cm^{-3}$, $W = 500\ \mu m$, $A = 2.5\times10^{-3}\ cm^2$

$I = 1\ mA$, $B_z = 10^{-6}\ wb/cm^2$

$$\therefore R_H = -\frac{1}{qn} = \frac{-1}{1.6\times10^{-19}\times10^{16}} = -625\ cm^3/C$$

$$\therefore V_H = \mathcal{E}_y W = \left[R_H\left(\frac{I}{A}\right) B_z\right] W = \left[-625\left(\frac{10^{-3}}{2.5\times10^{-3}}\right)\times10^{-6}\right]\times 500\times10^{-6}$$

$$= -1.25\ mV$$

1.5.3 High-Field Properties

From Eq. 49 on (p. 28)

$$\mu_e = \frac{2q\ell}{3}\sqrt{\frac{2}{\pi m_c^* kT}} \quad \text{where} \quad \ell \equiv \frac{\pi\hbar^4 C_e}{3E_{ds}^2 m_c^{*2} kT} = f(T)$$

The average energy of electrons increases with electric field. Let the electrons have a temperature T_e in a crystal at temperature T. The mean free path ℓ is determined by the crystal temperature T, but the random velocity of the electrons is determined by the electron temperature T_e. The mobility is then

$$\mu(T, T_e) = \frac{2q\ell}{3}\sqrt{\frac{2}{\pi m_c^* k T_e}} = \underbrace{\frac{2q\ell}{3}\sqrt{\frac{2}{\pi m_c^* kT}}}_{\mu_\ell(T)}\left(\frac{T}{T_e}\right)^{1/2}$$

$$= \mu_\ell\sqrt{\frac{T}{T_e}}$$

$$\therefore v_d = \mu\mathcal{E} = \sqrt{\frac{T}{T_e}}\,\mu_\ell\mathcal{E}$$

Consider the energy balance equation: the rate of acquisition of energy from the field per unit volume minus the rate of energy loss to phonons:

$$\frac{dE}{dt} = q\mathcal{E}v_d - \frac{dE}{dt}(\text{phonon})$$

$$\frac{dE}{dt}(\text{phonon}) = \frac{\delta m_c^* C_s^2}{\sqrt{\pi}\,\ell}\sqrt{\frac{2kT_e}{m_c^*}}\left(\frac{T_e}{T} - 1\right) \qquad \text{p.206 Moll}$$

At steady state $\frac{dE}{dT} = 0$, or

$$q\mathcal{E} v_d - \frac{8 m_c^* C_s^2}{\sqrt{\pi}\, l}\sqrt{\frac{2kT_e}{m_c^*}}\left(\frac{T_e}{T} - 1\right) = 0$$

Substitution l <Eq. 49 in textbook> into above equation:

$$q\mathcal{E}\underbrace{\sqrt{\frac{T}{T_e}}\mu_e\mathcal{E}}_{v_d} - \frac{8 m_c^* C_s^2}{\sqrt{\pi}}\underbrace{\left(\frac{2q}{3\mu_e}\sqrt{\frac{2}{\pi m_c^* kT}}\right)}_{1/l}\sqrt{\frac{2kT_e}{m_c^*}}\left(\frac{T_e}{T} - 1\right) = 0$$

The above equation can be simplified to

$$\left(\frac{T_e}{T}\right)^2 - \left(\frac{T_e}{T}\right) - \frac{3\pi}{32}\left(\frac{\mu_e\mathcal{E}}{C_s}\right)^2 = 0$$

where C_s is the velocity of sound

$$C_s = \sqrt{\frac{Y}{\rho}} = \sqrt{\frac{\text{Young's modulus}}{\text{density}}} = \sqrt{\frac{1.3\times10^{11}\ N/m^2}{2.33\times10^3\ kg/m^3}}$$

$$= 7\times10^3\ m/s = 7\times10^5\ cm/s$$

$$\therefore \frac{T_e}{T} = \frac{1}{2}\left\{1 + \left[1 + \frac{3\pi}{8}\left(\frac{\mu_e\mathcal{E}}{C_s}\right)^2\right]^{1/2}\right\} \qquad \text{------(78)}$$

(p. 30)

(1) when $\mathcal{E} \to 0$ $\quad T_e = T$ as expected

(2) when $\mathcal{E}$ increases but $\mu_e\mathcal{E} < C_s$

$$\frac{T_e}{T} = \frac{1}{2}\left\{1 + \left[1 + \frac{1}{2}\left(\frac{3\pi}{8}\left(\frac{\mu_e\mathcal{E}}{C_s}\right)^2\right)\right]\right\} = 1 + \frac{3\pi}{32}\left(\frac{\mu_e\mathcal{E}}{C_s}\right)^2$$

(3) when $\mu_e\mathcal{E} = \frac{8C_s}{3} = 2.66\,C_s$

$$\frac{T_e}{T} = \frac{1}{2}\left\{1 + \sqrt{1 + \frac{3\pi}{8}\left(\frac{8}{3}\right)^2}\right\} = 2 \quad , \quad T_e = 2T$$

$$\therefore \mu = \mu_e\sqrt{\frac{T}{T_e}} = \frac{\mu_e}{\sqrt{2}} = 0.7\,\mu_e \quad (\text{drop by } 30\%)$$

(4) when $\mu_e \mathcal{E} \gg c_s$, from Eq. 78

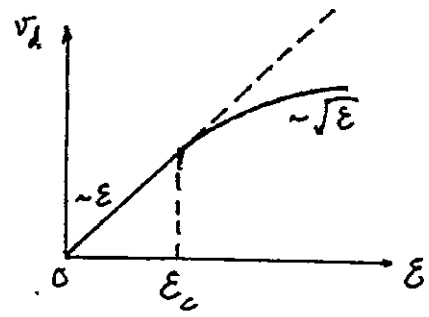
$$T_e \approx T\sqrt{\frac{3\pi}{32}\left(\frac{\mu_e \mathcal{E}}{c_s}\right)}$$ the last term dominates

$$\mu = \mu_e\sqrt{\frac{T}{T_e}} = \mu_e\left(\frac{32}{3\pi}\right)^{1/4}\left(\frac{c_s}{\mu_e \mathcal{E}}\right)^{1/2} \equiv \mu_e\sqrt{\frac{\mathcal{E}_c}{\mathcal{E}}} \sim \frac{1}{\sqrt{\mathcal{E}}}$$

$$\mathcal{E}_c \equiv \sqrt{\frac{32}{3\pi}}\left(\frac{c_s}{\mu_e}\right) = (1.84)\,\frac{7\times10^5\ cm/s}{10^3\ cm^2/Vs} = 1.2\times10^3\ V/cm$$

$\therefore$ For $\mathcal{E} > \mathcal{E}_c$

$$\boxed{v_d = \mu\mathcal{E} = \mu_e\sqrt{\frac{\mathcal{E}_c}{\mathcal{E}}}\,\mathcal{E} = (\mu_e\sqrt{\mathcal{E}_c})\sqrt{\mathcal{E}} \sim \sqrt{\mathcal{E}}}$$



At very high field, only emission of optical phonons will occur. From energy balance equal:

$$\frac{dE}{dt} = q\mathcal{E}v_d - \frac{dE}{dt}\,(\text{phonon})$$

$$\therefore\ 0 = q\mathcal{E}v_d - E_p\frac{\langle v_\theta\rangle}{l_r}$$

where E_p = optical phonon energy

l_r = optical phonon mean free path

$\langle v_\theta\rangle$ = electron average velocity $= \sqrt{\frac{8kT_e}{\pi m_c^*}}$ for Maxwellian distribution

$$\therefore \quad \mathcal{E} = \frac{E_p \langle v_\theta \rangle}{l_r} \frac{1}{\beta v_d}$$

$$v_d = \mu \mathcal{E} = \frac{2\beta l_r}{3}\sqrt{\frac{2}{\pi m_c^* kT_e}} \cdot \mathcal{E} = \frac{2\beta l_r}{3}\sqrt{\frac{2}{\pi m_c^* kT_e}}\left(\frac{E_p \langle v_\theta \rangle}{l_r}\frac{1}{\beta v_d}\right)$$

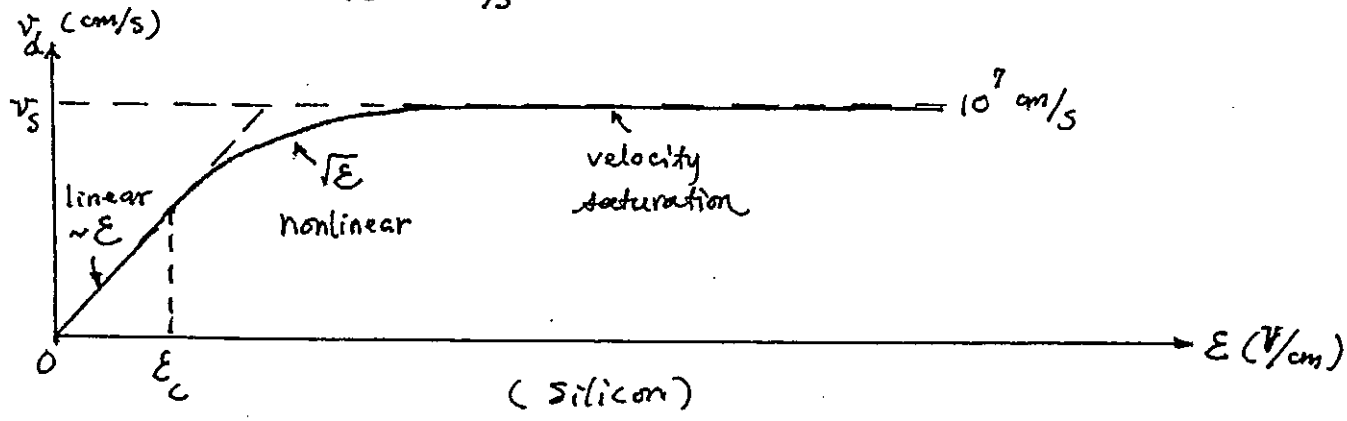
$$= \frac{2\beta l_r}{3}\sqrt{\frac{2}{\pi m_c^* kT_e}}\left(\frac{E_p}{l_r}\sqrt{\frac{8kT_e}{\pi m_c^*}}\frac{1}{\beta v_d}\right)$$

$$= \frac{8E_p}{3\pi m_c^*}\left(\frac{1}{v_d}\right)$$

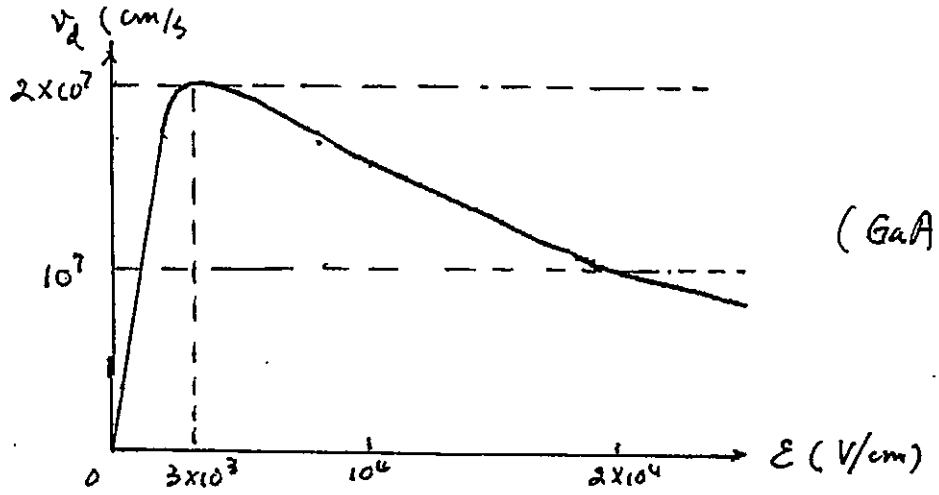
$$\therefore \quad v_d = \sqrt{\frac{8E_p}{3\pi m_c^*}} \neq f(\mathcal{E}) \qquad \cdots\cdots\cdots\cdots (80)$$

$$\approx \sqrt{\frac{8 \times 0.063 \times 1.6 \times 10^{-19}\ J}{3\pi \times 9.1 \times 10^{-31}\ kg}} = 9.7 \times 10^4\ m/s$$

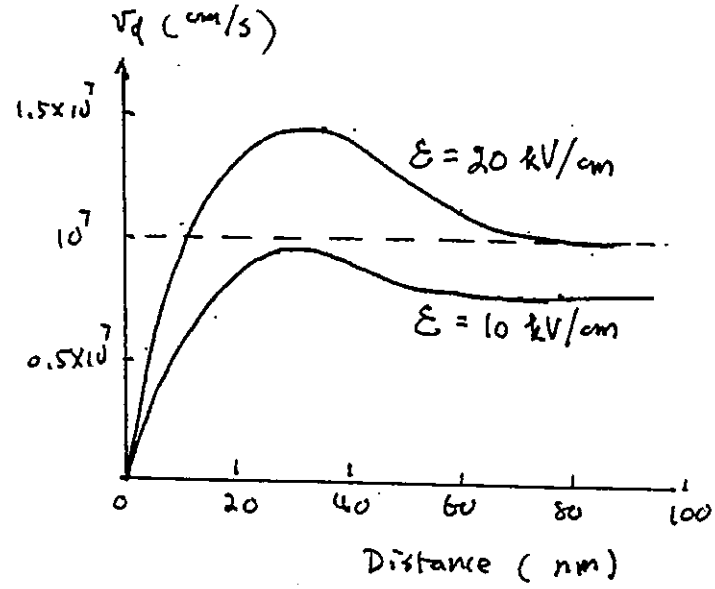
$$\approx 10^7\ cm/s$$



(Silicon)



(GaAs)



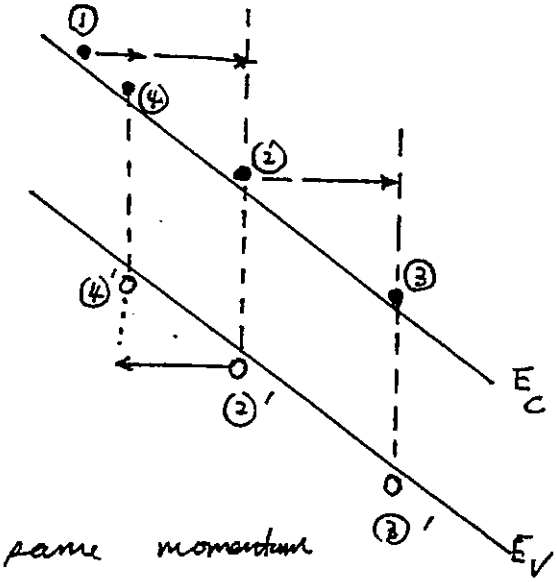
Velocity overshoot

For even higher fields, impact ionization will occur:

No.1 before collision

kinetic energy $= \frac{1}{2} m_1 v_s^2$

Momentum $= m_1 v_s$



After collision, we have

No.1, No.2, No 2'

Assume all have same mass,

same kinetic energy & same momentum

Total energy: $\frac{3}{2} m_1 v_f^2$

Total momentum: $3 m_1 v_f$

$\therefore \quad \frac{1}{2} m_1 v_s^2 = E_g + \frac{3}{2} m_1 v_f^2$ Conservation of energy

$m_1 v_s = 3 m_1 v_f$ " of momentum

$\therefore \quad E_0 = \frac{1}{2} m_1 v_s^2 = 1.5\, E_g$ (kinetic energy for ionization process)

Actually $E_0 \approx 3.6$ eV for electrons in Si

$E_0 \approx 5.0$ eV for holes in Si.

1.5.4 Recombination, Generation, & Carrier Lifetimes

For indirect-bandgap semiconductors, such as silicon, a direct recombination process is very unlikely, because the electrons at the bottom of the conduction band have nonzero crystal momentum with respect to the holes at the top of the valence band (refer to Fig. 4 in Chapter 1). A direct transition that conserves both energy and momentum is not possible without a simultaneous lattice interaction. Therefore the dominant recombination process in such semiconductors is indirect transition via localized energy states in the forbidden energy gap.[6] These states act as stepping stones between the conduction band and the valence band. Because the transition probability depends on the energy differences between the step and the conduction and valence band edges, these intermediate states can substantially enhance the recombination process.

Figure 13 shows various transitions that occur in the recombination process through intermediate-level states (also called recombination centers).[†] We illustrate the charging state of the center before and after each of the four basic transitions takes place. The arrows in the figure designate the transition of the electron in a particular process. The illustration is for the case of a recombination center with a single energy level that is neutral when not occupied by an electron or negative when it is occupied.

Process (*a*) is for electron capture—an electron in the conduction band is captured by the center. Because only one electron can occupy a given center, a center that is occupied by an electron cannot capture another one. Therefore, the rate of electron capture is proportional to the concentration of centers that are not occupied by electrons (i.e., those centers that remain neutral). If the concentration of centers in the semiconductor is N_t, the concentration of unoccupied centers is given by $N_t(1 - F)$, where F is the Fermi distribution function for the probability that a center is occupied by an electron. In equilibrium,

$$F = \frac{1}{1 + e^{(E_t - E_F)/kT}} \tag{48}$$

where E_t is the energy level of the center and E_F is the Fermi level.

Therefore, the rate of capture of electrons, process (*a*), is given by

$$R_a \sim nN_t(1 - F). \tag{49}$$

We shall designate the proportionality constant by the product $v_{th}\sigma_n$, so that

$$R_a = v_{th}\sigma_n nN_t(1 - F) \tag{50}$$

where v_{th} is the thermal velocity of the carriers given in Eq. 1 and σ_n is the capture cross section. The quantity σ_n describes the effectiveness of the center to capture an electron and is a measure of how close the electron has to come to the center to be captured. We expect that the capture cross section would be

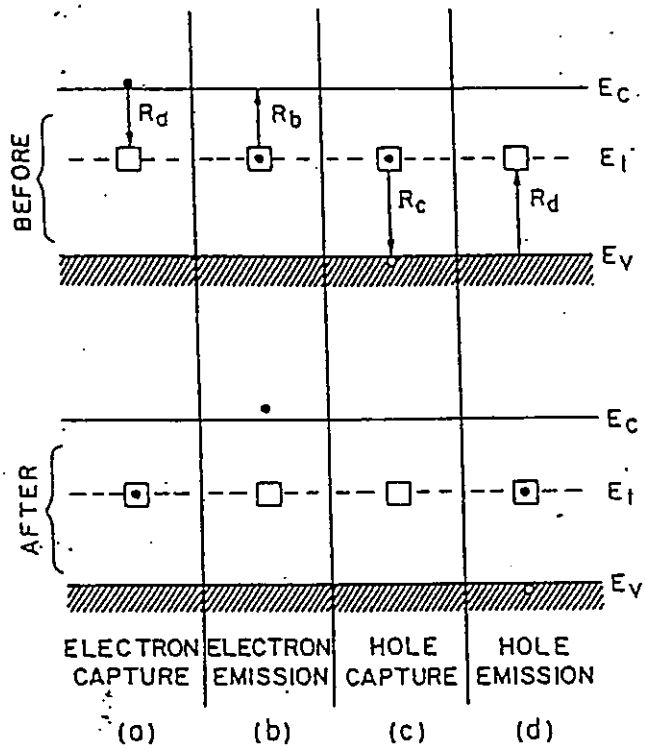


Fig. 13 Indirect generation-recombination processes at thermal equilibrium.[6]

of the order of the atomic dimensions, that is, of the order of 10^{-15} cm^2. The product $v_{th}\sigma_n$ may be visualized as the volume swept out per unit time by an electron with cross section σ_n. If the center lies within this volume, the electron will be captured by it.

The rate of emission of electrons from the center, process (*b*) in Fig. 13, is the inverse of the electron capture process. The rate is proportional to the concentration of centers occupied by electrons, that is, $N_t F$. We have

$$R_b = e_n N_t F . \tag{51}$$

The proportionality constant e_n is called the emission probability. At thermal equilibrium the rates of capture and emission of electrons must be equal ($R_a = R_b$). Thus, the emission probability can be expressed in terms of the quantities already defined in Eq. 50:

$$e_n = \frac{v_{th}\sigma_n n(1 - F)}{F} . \tag{52}$$

Since the electron concentration in thermal equilibrium is given by

$$n = n_i e^{(E_F - E_i)/kT} \tag{53}$$

and

$$\frac{(1 - F)}{F} = e^{(E_i - E_F)/kT} \tag{54}$$

we obtain

$$e_n = v_{th}\sigma_n n_i e^{(E_i - E_i)/kT} . \tag{55}$$

Note that if the center is close to the conduction band edge (i.e., $E_t - E_i$ is larger), the electron emission from the center becomes more probable, since e_n increases exponentially with $E_t - E_i$.

The transitions between the recombination center and the valence band are analogous to those described above. Process (*c*) in Fig. 13 is for hole capture. The rate is proportional to the centers occupied by electrons, $N_t F$, and the hole concentration. The proportionality constant is given by the product of the thermal velocity v_{th} and the capture cross section σ_p of holes. Thus,

$$R_c = v_{th}\sigma_p p N_t F . \tag{56}$$

The fourth process, Fig. 13*d*, is for hole emission, which describes the excitation of an electron from the valence band to the unoccupied center. By arguments similar to those for electron emission, the rate is

$$R_d = e_p N_t (1 - F) . \tag{57}$$

The emission probability e_p of a hole may be expressed in terms of v_{th} and σ_p by considering the thermal-equilibrium condition for which $R_c = R_d$:

$$e_p = v_{th}\sigma_p n_i e^{(E_i - E_t)/kT} . \tag{58}$$

Again, note that the emission probability increases exponentially as the center level E_t approaches the valence band edge. Because we can measure experimentally the capture cross sections σ_n and σ_p, the concentration of the recombination center N_t, and the center's energy level E_t, we can use Eqs. 50, 51, 56, and 57 to determine the kinetics of indirect recombination under nonequilibrium conditions.

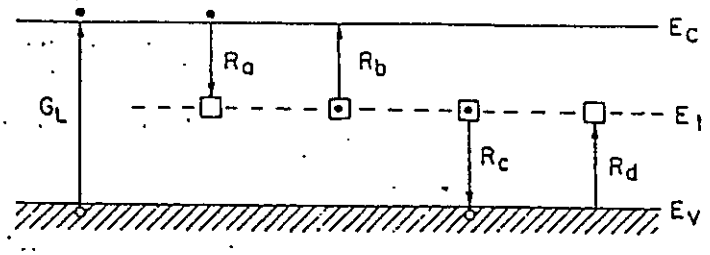


Fig. 14 Generation-recombination processes under illumination.

Figure 14 shows a nonequilibrium case in which an *n*-type semiconductor is illuminated uniformly to give a generation rate G_L. Thus in addition to the processes shown in Fig. 13, electron–hole pairs are generated as a result of light. In steady state the electrons entering and leaving the conduction band must be equal. This is called the *principle of detailed balance,* and it yields

$$\frac{dn_n}{dt} = G_L - (R_a - R_b) = 0 . \tag{59}$$

Similarly, in steady state the detailed balance of holes in the valence band leads to

$$\frac{dp_n}{dt} = G_L - (R_c - R_d) = 0 . \tag{60}$$

Under equilibrium conditions, that is, $G_L = 0$, $R_a = R_b$ and $R_c = R_d$. However, under steady-state nonequilibrium conditions $R_a \neq R_b$ and $R_c \neq R_d$. From Eqs. 59 and 60 we obtain

$$G_L = R_a - R_b = R_c - R_d . \tag{61}$$

Inserting the expressions for R_a through R_d into Eq. 61 gives

$$\begin{aligned} &v_{th}\sigma_n N_t[n_n(1-F) - n_i e^{(E_t - E_i)/kT}F] \\ &\quad = v_{th}\sigma_p N_t[p_n F - n_i e^{(E_i - E_t)/kT}(1-F)] \end{aligned} \tag{62}$$

We can eliminate F from Eq. 62 and solve for the net recombination rate U:

$$U \equiv R_a - R_b = \frac{v_{th}\sigma_n\sigma_p N_t(p_n n_n - n_i^2)}{\sigma_p[p_n + n_i e^{(E_t - E_i)/kT}] + \sigma_n[n_n + n_i e^{(E_i - E_t)/kT}]} \tag{63}$$

= Eq. 92 <p. 43>

• Recombination Lifetime

Assume $\sigma_n = \sigma_p = \sigma_o$

$$\therefore\ U = v_{th}\sigma_o N_t \frac{(p_n n_n - n_i^2)}{p_n + n_n + 2n_i \cosh\left(\frac{E_t - E_i}{kT}\right)}$$

$$\because\ p_n n_n = (p_{no} + \Delta)(n_{no} + \Delta) = p_{no}n_{no} + \Delta(p_{no} + n_{no}) + \Delta^2$$

$$\approx n_i^2 + (p_n - p_{no})(p_{no} + n_{no})$$

$$\therefore\ U = v_{th}\sigma_o N_t \frac{(p_n - p_{no})}{\dfrac{p_n + n_n}{(p_{no} + n_{no})} + \dfrac{2n_i\left(\cosh\left(\frac{E_t - E_i}{kT}\right)\right)}{(p_{no} + n_{no})}}$$

$$\approx v_{th}\sigma_o N_t \frac{(p_n - p_{no})}{1 + \dfrac{2n_i}{(p_{no} + n_{no})}\cosh\left(\frac{E_t - E_i}{kT}\right)} \equiv \frac{p_n - p_{no}}{\tau_r}$$

$$\therefore\ \tau_r \equiv \left[1 + \frac{2n_i}{(p_{no} + n_{no})}\cosh\left(\frac{E_t - E_i}{kT}\right)\right] \Big/ v_{th}\sigma_o N_t$$

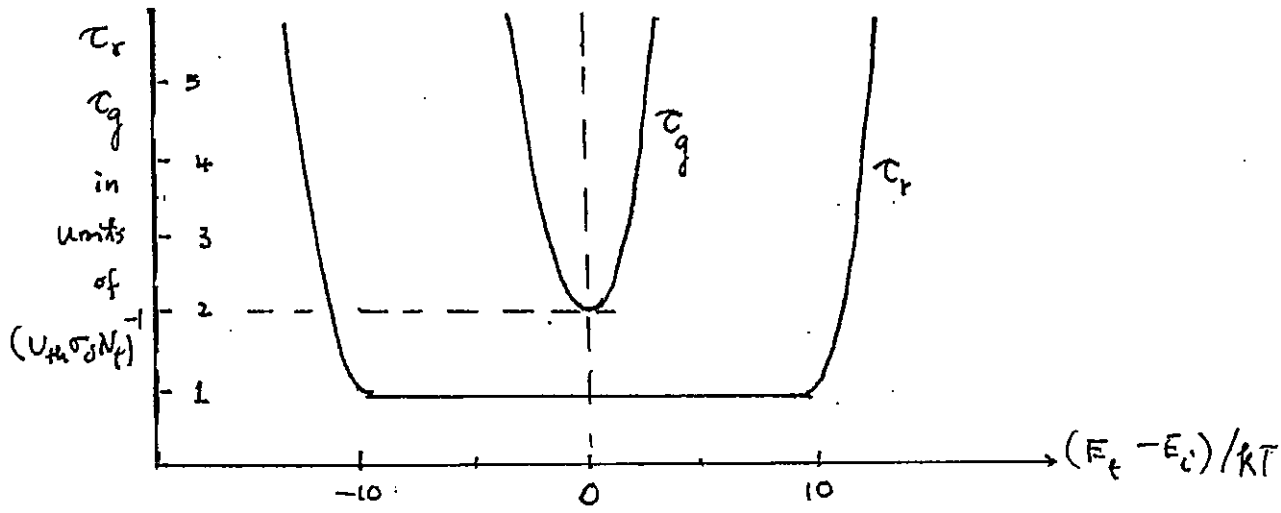
If $\frac{2n_i}{(p_{no}+n_{no})}\cosh\left(\frac{E_t-E_i}{kT}\right) \ll 1$

Then $\tau_r = \frac{1}{v_{th}\sigma_0 N_t}$

- Generation Lifetime ($p_n < n_i$, $n_n < n_i$)

$$G = -U = \frac{v_{th}\sigma_0 N_t n_i^2}{2n_i \cosh\left(\frac{E_t-E_i}{kT}\right)} = \frac{v_{th}\sigma_0 N_t n_i}{2\cosh\left(\frac{E_t-E_i}{kT}\right)} \equiv \frac{n_i}{\tau_g}$$

$$\tau_g \equiv 2\cosh\left(\frac{E_t-E_i}{kT}\right)\Big/ v_{th}\sigma_0 N_t \rightarrow \frac{2}{v_{th}\sigma_0 N_t} \quad \text{as } (E_t-E_i)\rightarrow 0$$



At $(E_t - E_i) = 4kT$, we have

$$\frac{\tau_g}{\tau_r} = 50.$$

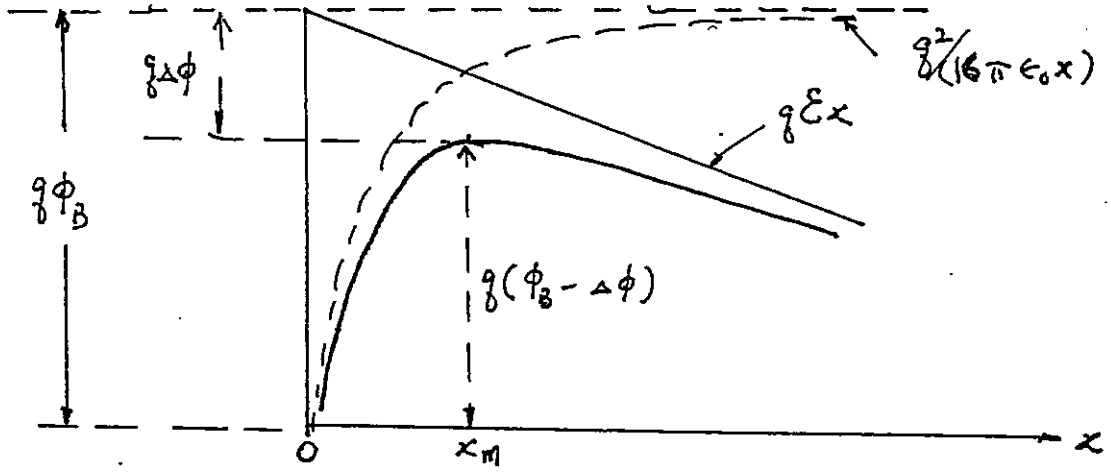
1.5.5 Diffusion

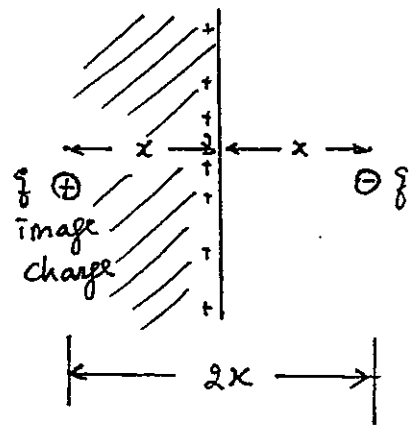
$$\cdot\ D_n = \left(\frac{kT}{q}\right)\mu_n$$
$$D_p = \left(\frac{kT}{q}\right)\mu_p$$
} Einstein Relation

one obtains D_n or D_p from Fig. 15 (p. 29)

1.5.6 Thermionic Emission

- Schottky emission (see Chapter 3)





The attractive force is

$$F = \frac{-q^2}{4\pi(2x)^2\epsilon_0} = \frac{-q^2}{16\pi\epsilon_0 x^2}$$

The work done by an electron from ∞ to x is

$$E(x) = \int_{\infty}^{x} F\,dx = \frac{q^2}{16\pi\epsilon_0 x}$$

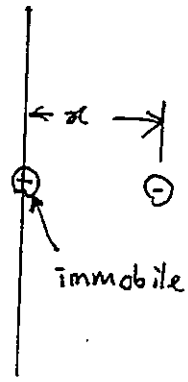
Potential energy $PE = \dfrac{q^2}{16\pi\epsilon_0 x} + q\mathcal{E}x$

$$x_m\left[\frac{dPE}{dx} = 0\right] = \sqrt{q/16\pi\epsilon_0\mathcal{E}}$$

$$\boxed{\Delta\phi = \sqrt{\frac{q\mathcal{E}}{4\pi\epsilon_0}} = 2\mathcal{E}x_m}$$

This is the Schottky lowering.

- Frenkel- Poole Emission



If the positive charge is immoble

$$\text{Force} = \frac{-q^2}{4\pi(x^2)\epsilon_0}$$

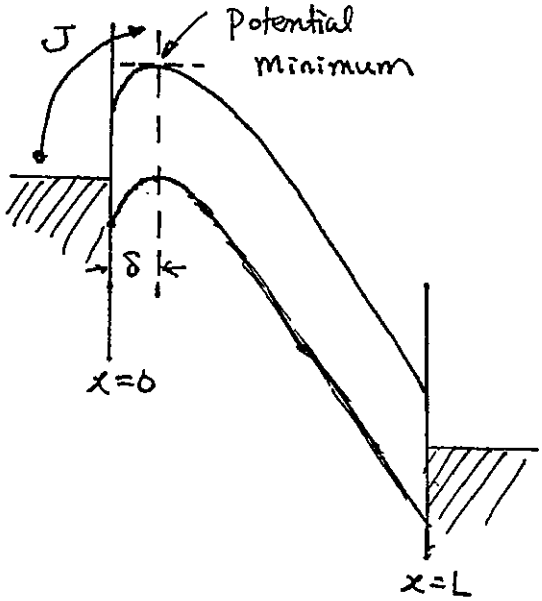
$$E = \frac{q^2}{4\pi\epsilon_0 x}$$

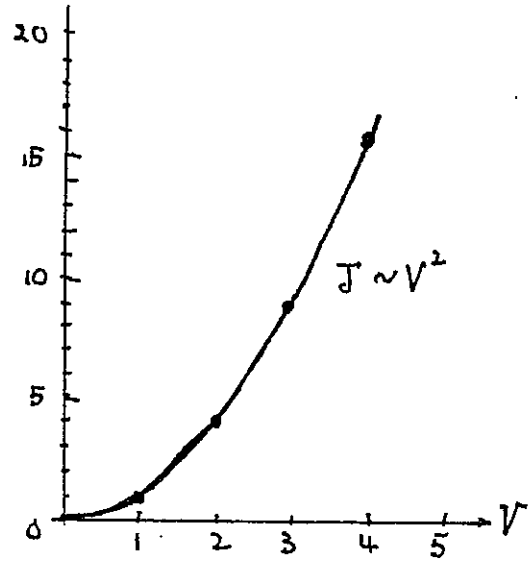
$$PE = \frac{q^2}{4\pi\epsilon_0 x} + qEx$$

$$\therefore \Delta\phi = \sqrt{\frac{qE}{\pi\epsilon_0}}$$ which is twice as large as the Schottky lowering

1.5.7 Tunneling (see chapter 8)

1.5.8 Space-charge Effect





For insulators or intrinsic semiconductors, the space charge (i.e., the injected charge carriers from the metal into the insulator) can not be neutralized by charges of opposite sign (i.e., n ≫ ionized donor density N_D^+).

We have the Poisson Equation and the current density Equation:

$$\frac{d\mathcal{E}}{dx} = \frac{\delta n}{\epsilon_i} \qquad (1)$$

$$J = q n \mu \mathcal{E} \qquad (2) \quad \text{(assume constant mobility)}$$

no diffusion current

The boundary conditions are

$$\mathcal{E} = 0 \quad \text{and} \quad V = 0 \quad \text{at} \quad x = 0 \quad (\text{assume } \delta \to 0)$$

$$V = V \quad \text{at} \quad x = L$$

substituting (2) into (1)

$$\frac{d\mathcal{E}}{dx} = \frac{J}{\epsilon_i \mu \mathcal{E}}$$

$$\therefore \mathcal{E}\, d\mathcal{E} = \frac{J}{\epsilon_i \mu} dx$$

$$\mathcal{E}^2 = \frac{2J}{\epsilon_i \mu} x \quad (\mathcal{E} = 0,\ x = 0) \quad \text{or} \quad \mathcal{E} = \sqrt{\frac{2J}{\epsilon_i \mu}}\sqrt{x}$$

$$\because |\mathcal{E}| = \frac{\partial V}{\partial x} \qquad \therefore dV = \sqrt{\frac{2J}{\epsilon_i \mu}}\sqrt{x}\, dx$$

$$V = \frac{2}{3}\left(\frac{2J}{\epsilon_i \mu}\right)^{1/2} x^{3/2} \quad (V = 0 \text{ at } x = 0)$$

$$V = \frac{2}{3}\left(\frac{2J}{\epsilon_i \mu}\right)^{1/2} L^{3/2}$$

$$\boxed{\therefore \quad J = \frac{9\, \epsilon_i \mu V^2}{8 L^3} \sim V^2}$$ for constant mobility

Similarly $J = \dfrac{2 \epsilon_i v_s V}{L^2} \sim V$ for velocity saturation

$$J = \frac{4 \epsilon_i}{9 L^2}\left(\frac{2q}{m^*}\right)^{1/2} V^{3/2}$$ for ballistic regime

1.6 Phonon, Optical and Thermal Properties

A diatomic lattice such as NaCl will be more complicated than the linear array of like masses that is coupled together by springs. There is a vibration of the sodium chloride lattice in which all of the sodium atoms move together and all of the chlorine atoms move together. A description of the waves associated with such a vibration requires that the vibrational state of each unit cell be specified. In reference to Fig. 9-5, for example, there is a vibrational mode in which the relative positions of the two masses M and m in each of the unit cells move with relation to each other. This vibration is formally the same as the vibrations of identical particles. In addition, there is a vibration in which the two masses in the unit cell move with respect to each other but the centers of gravity of the unit cells do not move.

If we suppose that the masses interact only with nearest neighbors, then, since even-numbered particles have mass M and odd-numbered particles have mass m, the equations of motion are

$$F_{2p} = M\frac{d^2x_{2p}}{dt^2} = K(x_{2p+1} + x_{2p-1} - 2x_{2p})$$
$$F_{2p+1} = m\frac{d^2x_{2p+1}}{dt^2} = K(x_{2p+2} + x_{2p} - 2x_{2p+1}) \tag{9-28}$$

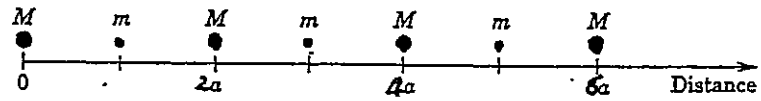
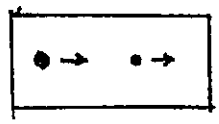
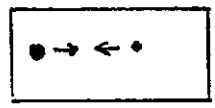


Fig. 9-5 Diatomic lattice. In such a lattice, there can exist mechanical vibrations in which neighboring atoms move together as in the case for the main atomic lattice, as well as vibrations in which adjacent planes move in opposite directions.


Acoustic phonon


Optical phonon

We shall assume wave solutions of the form

$$x_{2p} = A_M \exp\,[i(pa\beta - \omega t)]$$
$$x_{2p+1} = A_m \exp\left\{i\left[(2p+1)\frac{a\beta}{2} - \omega t\right]\right\} \tag{9-29}$$

where the coefficients A_M and A_m may differ in amplitude and phase. The frequency $\omega/2\pi$ and wave numbers $\beta/2\pi$ are identical for the two displacements. Substitution of (9-29) into (9-28) results in a pair of equations for the coefficients A_M and A_m. Thus,

$$A_M(-\omega^2 M + 2K) - A_m 2K\cos\frac{\beta a}{2} = 0$$
$$-A_M 2K\cos\frac{\beta a}{2} + A_m(-\omega^2 m + 2K) = 0 \tag{9-30}$$

Equation (9-30) has a nontrivial solution for the coefficients A_M, A_m only if the determinant of coefficients vanishes or if

$$(-\omega^2 M + 2K)(-\omega^2 m + 2K) - 4K^2 \cos^2 \frac{\beta a}{2} = 0 \tag{9-31}$$

Equation (9-31) is satisfied if

$$\omega^2 = K\frac{M+m}{Mm} \pm \sqrt{K^2\left(\frac{M+m}{Mm}\right)^2 - \frac{4K^2}{Mm}\sin^2\frac{\beta a}{2}} \tag{9-32}$$

= Eq. 123

$K \equiv \alpha_f$
$M \equiv m_1$
$m = m_2$
$\beta = k_{ph}$

There are two frequencies that satisfy the conditions for propagation in the lattice for each β. At $\beta \approx 0$, the two frequencies are

$$\omega_a \approx 0$$
$$\omega_0 \approx \sqrt{2K\left(\frac{1}{M} + \frac{1}{m}\right)} \tag{9-33}$$

In (9-32) and (9-33), there is a branch of the frequency spectrum that starts at zero and, near the edge of the Brillouin zone, reaches a value of $\sqrt{2k/M}$. This branch includes the usual acoustic waves in solids and is called the acoustic branch. The higher-frequency branch that has frequency $\sqrt{2K(1/M + 1/m)}$ at zero wave number can interact with infrared frequencies of the electromagnetic spectrum and was thought at one time to be responsible for the optical behavior of solids.[1] The high-frequency branch is called the optical branch.

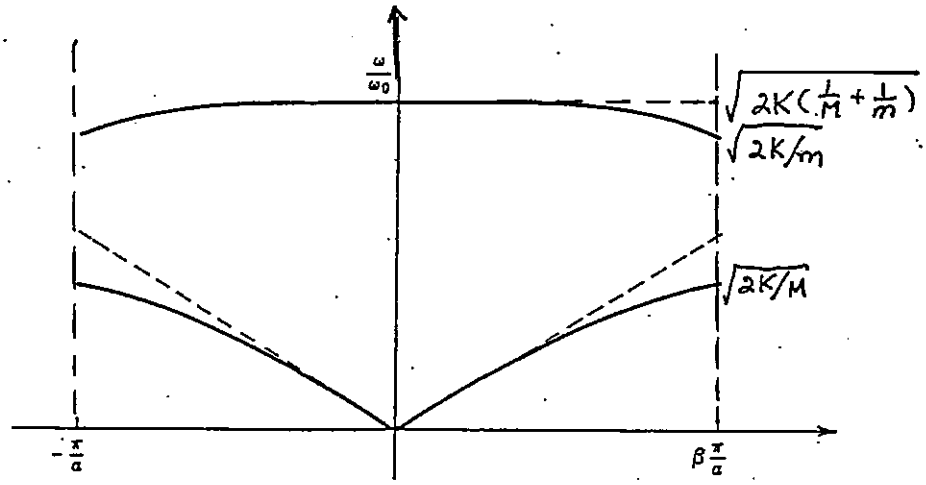


Fig. 9-6 Dispersion diagram (ω-β) for diatomic crystal with only nearest-neighbor interaction, showing acoustic and optical branches.

For the acoustic mode, $\omega \approx 0$, we find from Eq. (9-30) that the ratio

$$\frac{A_M}{A_m} = +\frac{2K\cos(\beta a/2)}{2K - \omega^2 M} \tag{9-34}$$

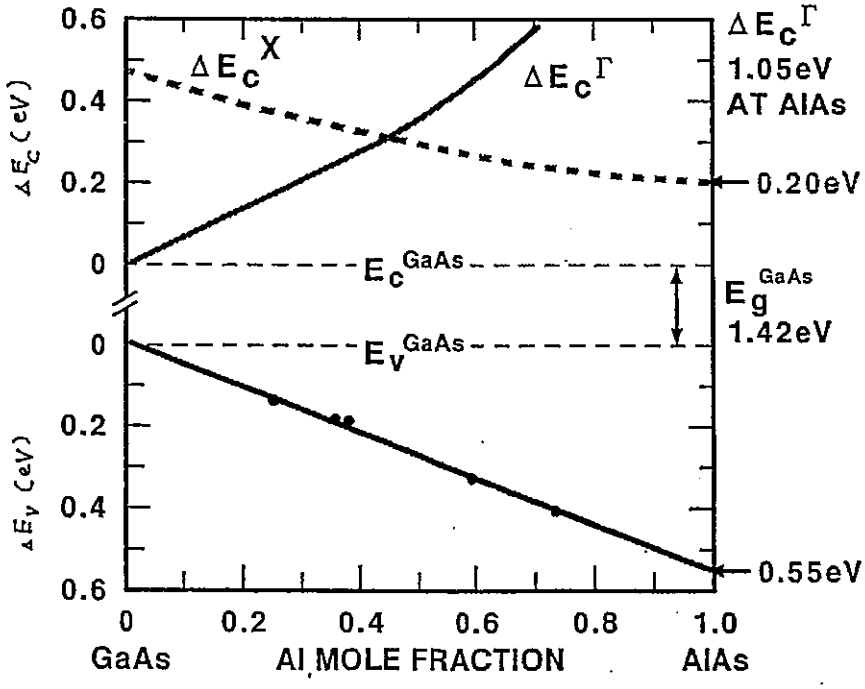
is unity if $\beta a \approx 0$, $\omega \approx 0$. For the optical mode, with $\beta a = 0$ and $\omega = \sqrt{2K(1/M + 1/m)}$, we find that

$$\frac{A_M}{A_m} = -\frac{m}{M} \quad \text{optical mode} \tag{9-35}$$

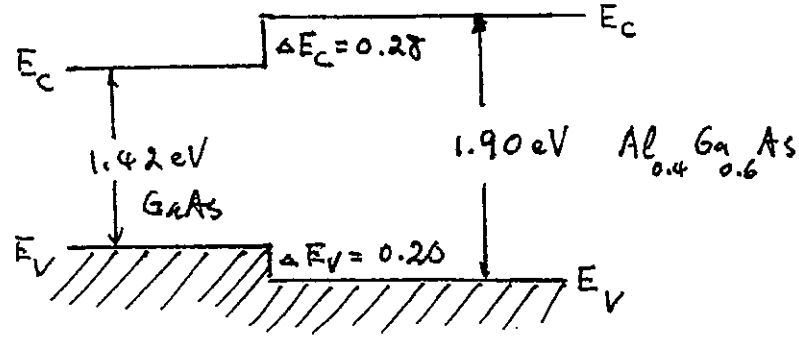
1.7 Heterojunction and Nanostructures

Epitaxy technology is used to grow lattice-matched semiconductor layers and strained layers.

- Lattice-matched epitaxy: best example is $Al_xGa_{1-x}As$ on GaAs with good lattice match for all x, i.e., $0 \le x \le 1$.



For example, for $x = 0.4$, i.e., $Al_{0.4}Ga_{0.6}As$ has a lattice constant very close to that of GaAs ($a = 5.6533\,\text{Å}$).



- Strained-layer epitaxy (also called pseudomorphic):
The critical thickness is the maximum thickness of epitaxial layer, above which misfit dislocation will occur.

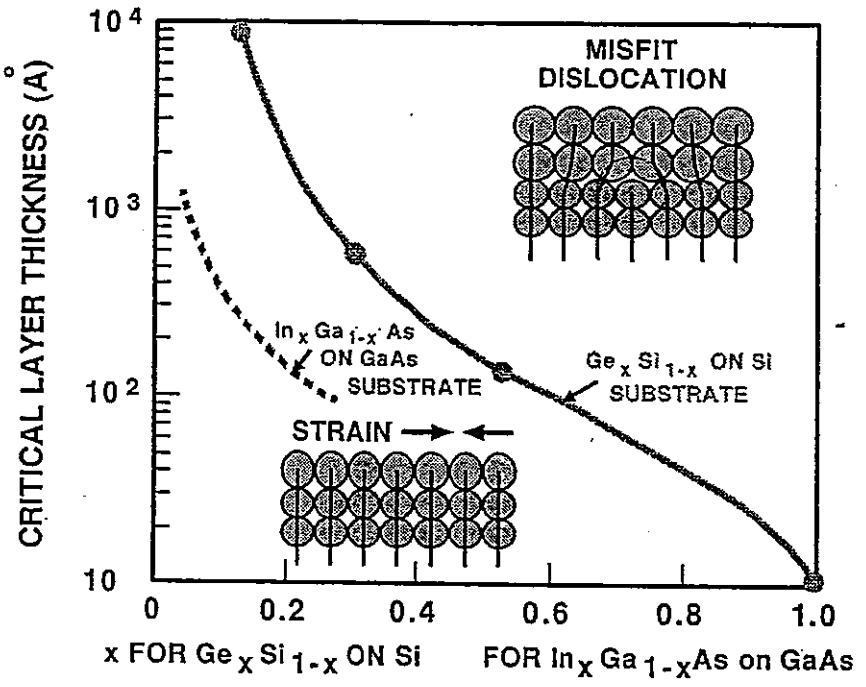
The critical thickness is approximately

$$t_c \approx \frac{a_e^2}{2|a_e - a_s|}$$

where a_e = lattice constant of epitaxial layer

a_s = " " " the substrate

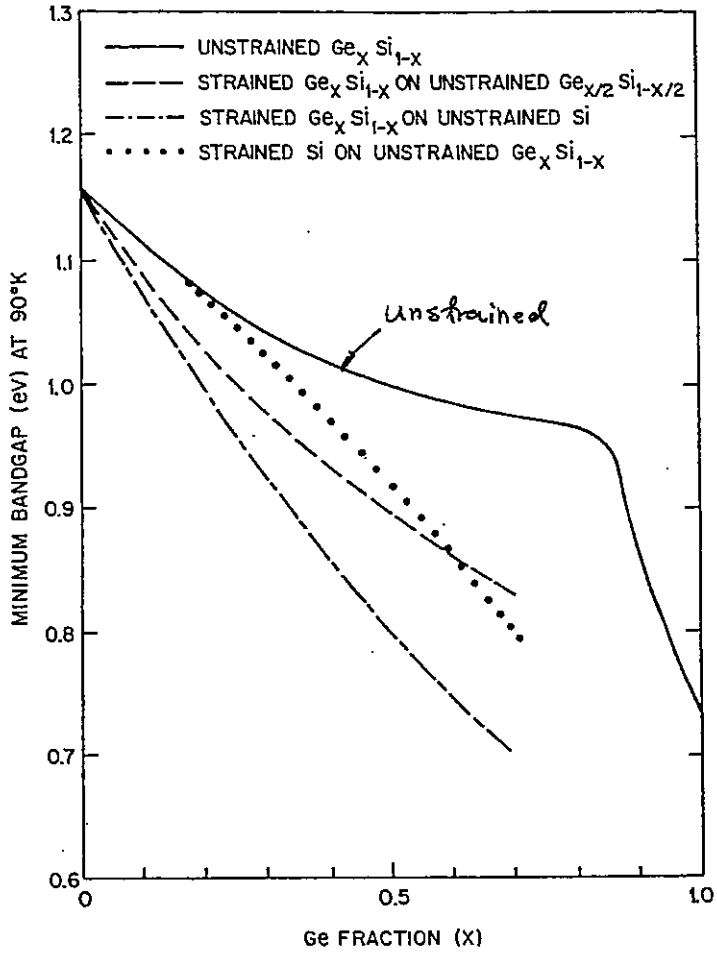
If $a_e = 5\text{Å}$, mismatch is 5%, then $t_c \approx 100\text{Å} = 10\text{ nm}$



For $Ge_{0.2}Si_{0.8}$ epitaxial layer on silicon, the critical layer thickness is over 2000 Å (200 nm).

When an epitaxial layer grown in strained-layer form, the compression or dilation of bonds is equivalent to that produced by pressure of order of 100,000 atm!

Therefore, the effect will change the bandgap from that of the unstrained alloy.



Calculations showing the dramatic effect of strain upon semiconductor bandgaps. Solid line: bulk (unstrained) GeSi. Dotted line: silicon strained to match the lattice constant of Ge_xSi_{1-x}/Si. Dashed line: Ge_xSi_{1-x} strained to lattice match $Ge_{x/2}Si_{1-x/2}$/Si. Dash-dot line: Ge_xSi_{1-x} strained to lattice match silicon.

- Quantum Well and Superlattice

A quantum well is formed by two heterojunctions or 3 layers of materials such that the middle layer has the lowest E_c. When the barrier layers between quantum wells become very thin, a superlattice is formed.

- Quantization in quantum wells :

SCHRÖDINGER EQUATION IN EFFECTIVE MASS APPROXIMATION

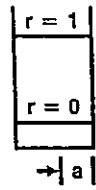
$$\left(\frac{P^2}{2m^*} + U\right)\psi = E\psi \qquad (1)$$

FOR ONE-DIMENSIONAL POTENTIAL ENERGY U(z), EQ.1 REDUCES TO

$$\left(\frac{P_z}{2m^*} + U\right)\phi = E_r\phi \qquad E_r \equiv E - \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2)$$

FOR INFINITE SQUARE WELL

$$E_r = \frac{\hbar^2\pi^2}{8m^*a^2}(r+1)^2 \quad r = 0, 1, 2, \ldots$$

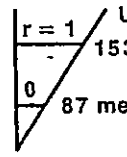


224 meV (For $m^* = 0.067\ m_0$, $a = 100\text{Å}$)

56 meV

FOR TRIANGULAR WELL

$$E_r = \left(\frac{\hbar^2 q^2 \mathcal{E}^2}{2m^*}\right)^{1/3}\left(\frac{3}{2}\pi\left(r+\frac{3}{4}\right)\right)^{2/3} \quad r = 0, 1, 2, \ldots$$

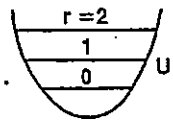


153 meV (For $\mathcal{E} = 10^5\ \text{V/cm}$)

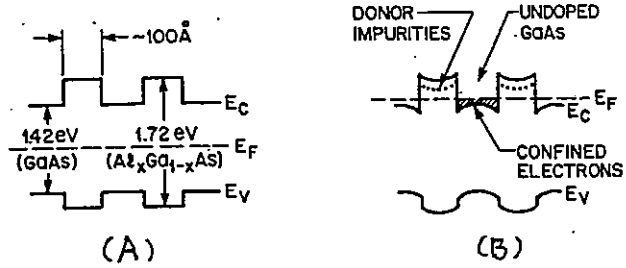
87 meV

FOR PARABOLIC WELL

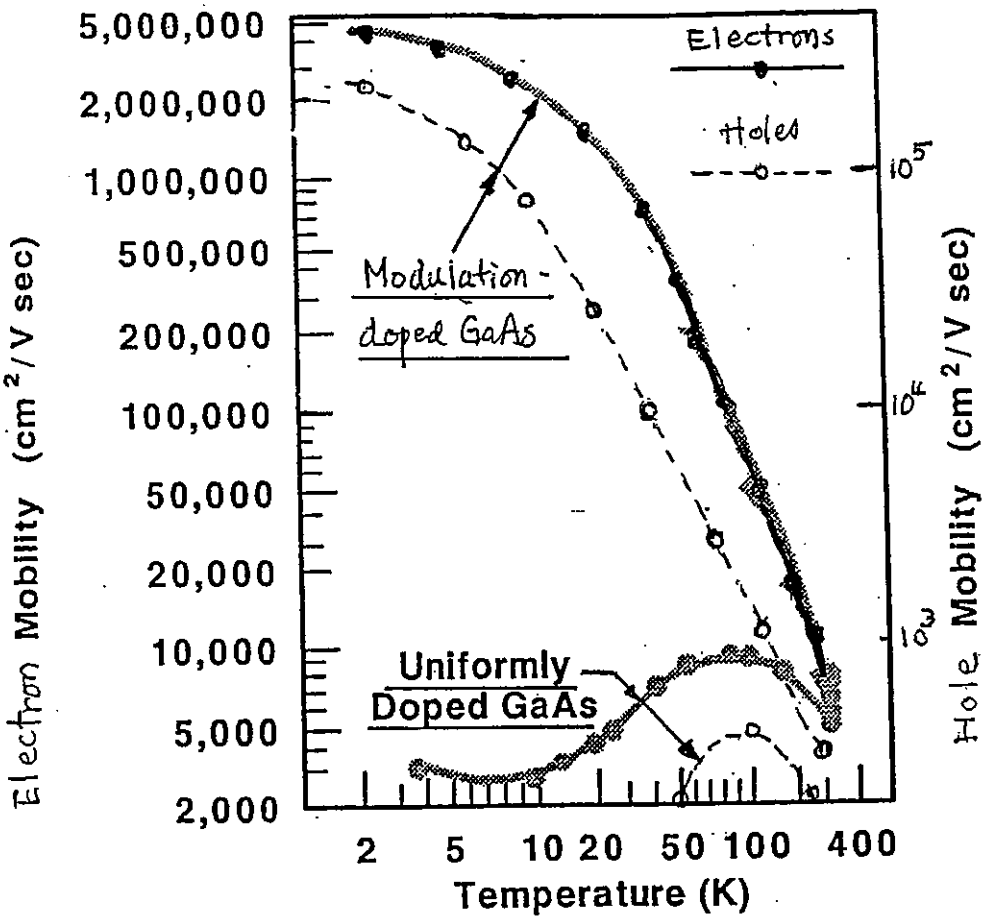
$$E_r = \hbar\omega\left(r+\frac{1}{2}\right) \quad r = 0, 1, 2, \ldots$$



For an undoped-superlattice, the band diagram is shown in Fig A. For a modulation-superlattice (Fig. B), only the large-bandgap layers are doped. The electrons from ionized donors will drop into the undoped GaAs quantum wells. Since there are no impurity scatterings in the undoped layers, the mobility can be substantially increased at lower temperatures.

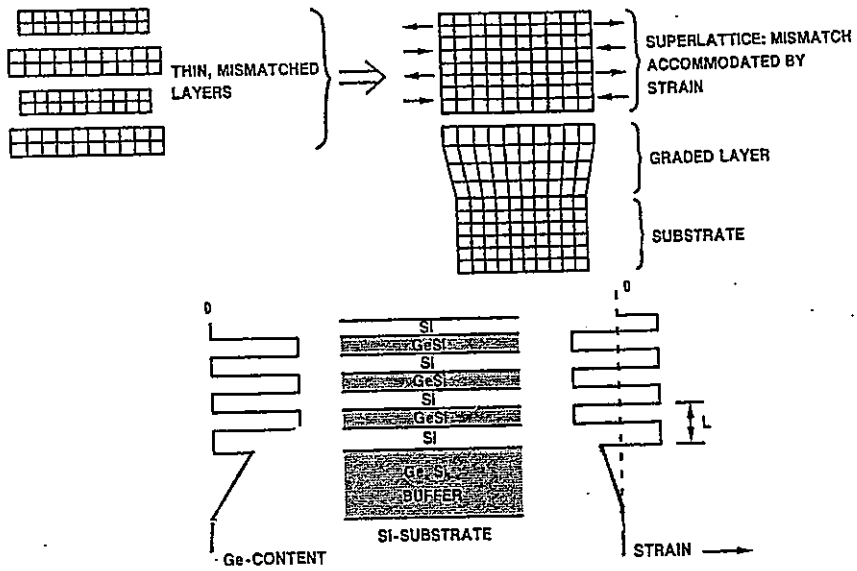


(A) Energy-band diagram for undoped superlattice structure of GaAs (1.42 eV) and $Al_{0.3}Ga_{0.7}As$. (B) Energy-band diagram for modulation-doped superlattice structure.



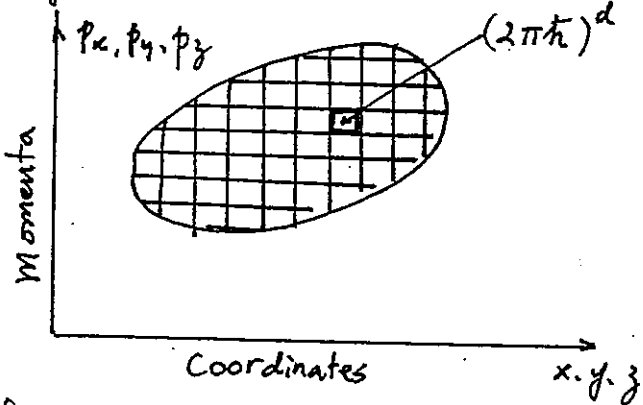
For example at 10K, $\mu_n \approx 2\times10^6$ cm²/V-s for modulation-doped GaAs, while the bulk μ_n for uniformly doped GaAs is only 3×10^3 cm²/V-s.

- **STRAINED-LAYER SUPERLATTICE (SLS)**



- Quantum Wire and Quantum Dot

Consider the phase space of a single particle in d dimensions ($d=0$ for quantum dot, $d=1$ for quantum wire, $d=2$ for quantum well, $d=3$ for bulk material). It has $2d$ axes, corresponding to d coordinates and d momenta of the particle.



Following the same derivation as in <Sec. 1.4>, the hypercube $V^{(d)} = L^d p^d$ in the phase space contains N distinct states. The volume for a state is $h^d \equiv (2\pi\hbar)^d$.

$$\therefore \quad N = \frac{2V^{(d)}}{(2\pi\hbar)^d} \qquad (2 \text{ for spin-}\tfrac{1}{2}\text{ particles})$$

For free electron gas, the energy dispersion relation is isotropic, $E = p^2/2m^*$, which allows a simple counting of states in spherical energy shells, $dE = (2E/m^*)^{1/2} dp$.

$$\therefore \quad dN = \frac{2L\,dp}{2\pi\hbar} = \left(\frac{\sqrt{m^*}}{\pi\hbar\sqrt{2E}}\right) L\,dE \qquad \text{for 1D.}$$

$$dN = \frac{2L^2 2\pi p\,dp}{(2\pi\hbar)^2} = \left(\frac{m^*}{\pi\hbar^2}\right) L^2\,dE \qquad \text{for 2D.}$$

$$dN = \frac{2L^3 4\pi p^2 dp}{(2\pi\hbar)^3} = \left(\frac{m^*\sqrt{2m^*E}}{\pi^2\hbar^3}\right) L^3\,dE \qquad \text{for 3D.}$$

The quantity in the bracket is the density of states $N(E) = dN/dE$ per unit length, area, or volume of the electron gas.

1.8 Basic Equations & Examples (p.63 of textbook)

1.8.1 Basic Equations

- Poisson's Equation $\nabla \cdot D = \rho$ $\qquad \nabla \cdot \mathcal{E} = \frac{\rho}{\epsilon_s}$
- Current density equation

$$J_n = q\mu_n n \mathcal{E} + q D_n \frac{dn}{dx}$$

$$J_p = q\mu_p p \mathcal{E} - q D_p \frac{dp}{dx}$$

- Continuity equation

$$\frac{\partial n}{\partial t} = G_n - U_n + \frac{1}{q}\frac{dJ_n}{dx}$$

$$\frac{\partial p}{\partial t} = G_p - U_p - \frac{1}{q}\frac{dJ_p}{dx}$$

1.8.2 Examples

- Stevenson-Keyes method — minority life time
- Injection ratio — transistor action
- Haynes-Shockley experiment — drift velocity
- Surface recombination — surface recombination velocity:

assume $J \sim q[p_n(0) - p_{no}]$

$\therefore J = qS[p_n(0) - p_{no}]$ $\qquad S$ = proportionality constant

$\therefore S = \dfrac{J}{q[p_n(0) - p_{no}]}$ = unit of velocity (cm/s)

$S = 1 - 10$ cm/s for Si

$\sim 10^3$ cm/s for InP

$\sim 10^6$ cm/s for GaAs.

Chapter 4

Metal-Insulator-Semiconductor Capacitors

4.1 Introduction

MIS capacitor, especially the MOS capacitor, is

(1) the most useful device to study reliability and stability of semiconductor devices

(2) related to most silicon planar devices and integrated circuits (MOS capacitor forms the central part of the MOSFET)

4.2 Ideal MIS Capacitor

An ideal MIS capacitor is defined as:

(1) no interface traps at no oxide charges

(2) resistivity of insulator (oxide) is infinite

(3) flat band when there is no applied voltage

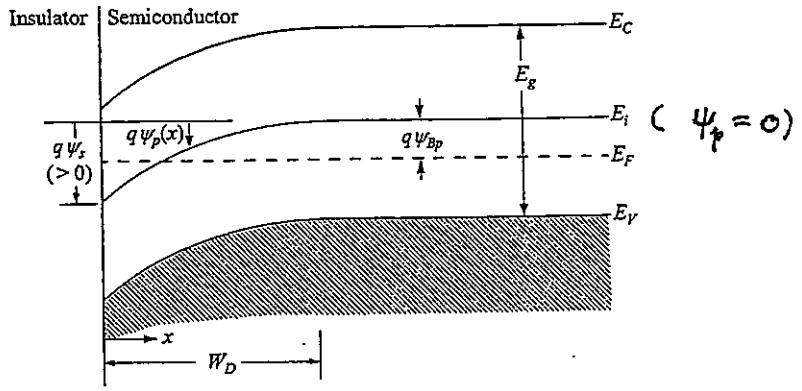
Three conditions at semiconductor surface:

(1) accumulation of majority carriers (holes for p type)

(2) depletion of majority carriers

(3) inversion: minority carriers (electrons for p type) is larger than the majority carriers. Inversion is most important for MOSFET operation.

4.2.1 Surface Space-Charge Region



From Eq. 30b < p. 21 of textbook >

$$p = n_i \exp\left(\frac{E_i - E_F}{kT}\right)$$

Under dc bias, there is no current flow, therefore E_F is constant. Referring to the above figure, we have

(1) At far right :

$$p_{p0} = n_i \exp\left(\frac{E_i - E_F}{kT}\right) = n_i \exp\left(\frac{q\psi_{Bp}}{kT}\right) = n_i \exp(\beta\psi_{Bp}) \qquad \beta \equiv q/kT$$

(2) At x near the surface :

$$p_p(x) = n_i \exp\left[\frac{E_i(x) - E_F}{kT}\right] = n_i \exp\left(\frac{q\Delta}{kT}\right)$$

$$= n_i \exp\left[\frac{q(\psi_{Bp} - \psi_p)}{kT}\right] = \underbrace{n_i \exp\left[\beta(\psi_{Bp}}_{p_{p0}} - \psi_p)\right]$$

$$= p_{p0} \exp(-\beta\psi_p) \qquad \cdots\cdots\cdots\cdots\cdots (3b)$$

$$n_p(x) = n_i^2 / p_p(x) = \frac{n_i^2}{p_{p0}\exp(-\beta\psi_p)} = n_{p0}\exp(\beta\psi_p) \cdots (3a)$$

(3) At $x = 0$:

$$n_p(0) = n_{p0}\exp(+\beta\psi_s)$$
$$p_p(0) = p_{p0}\exp(-\beta\psi_s)$$

} where ψ_s is the surface potential.

< p. 201 > Multiply Eq. 8 by $d\psi_p$ on both sides, then integrate

$$\int d\psi_p \left(\frac{d^2\psi_p}{dx^2}\right) = \int -\frac{q}{\epsilon_s}\left\{ p_{po}\left[\exp(-\beta\psi_p) - 1\right] - n_{po}\left[\exp(\beta\psi_p) - 1\right]\right\} d\psi_p$$

$$\downarrow$$

$$\int d\psi_p \frac{d}{dx}\left(\frac{d\psi_p}{dx}\right) = \int -\frac{q}{\epsilon_s}\left\{ \cdots\cdots\cdots\cdots\cdots\cdots \right\} d\psi_p$$

$$\downarrow$$

$$\int \left(\frac{d\psi_p}{dx}\right) d\left(\frac{d\psi_p}{dx}\right) = \int -\frac{q}{\epsilon_s}\left\{ \cdots\cdots\cdots\cdots\cdots\cdots \right\} d\psi_p$$

$$\downarrow$$

$$\frac{1}{2}\left(\frac{d\psi_p}{dx}\right)^2 = \int_0^{\psi_p} -\frac{q}{\epsilon_s}\left\{ \cdots\cdots\cdots\cdots\cdots\cdots \right\} d\psi_p$$

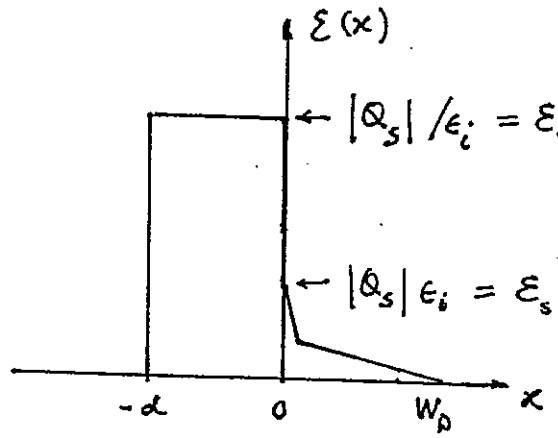
$$\downarrow$$

$$\frac{1}{2}\mathcal{E}^2 = \frac{1}{2}\left(\frac{\sqrt{2}kT}{q}\right)^2 \underbrace{\left(\frac{q p_{po}\beta}{\epsilon_s}\right)}_{\text{III}} \underbrace{\left\{\left[e^{-\beta\psi_p} + \beta\psi_p - 1\right] + \frac{n_{po}}{p_{po}}\left[e^{\beta\psi_p} - \beta\psi_p - 1\right]\right\}}_{\text{III}}$$

$$(L_D)^{-2} \qquad \left[F\left(\beta\psi_p, \frac{n_{po}}{p_{po}}\right)\right]^2$$

$$\therefore \mathcal{E} = \pm \frac{\sqrt{2}kT}{qL_D} F\left(\beta\psi_p, \frac{n_{po}}{p_{po}}\right) \cdots\cdots\cdots\cdots\cdots\cdots (13)$$

(p. 204) Fig. 6C



From Eqs. 15 $\quad |Q_s| = \epsilon_s \mathcal{E}_s$

From Gauss Law $\quad \epsilon_s \mathcal{E}_s = \epsilon_i \mathcal{E}_i$

$$\therefore \mathcal{E}_s = \mathcal{E}_i\left(\frac{\epsilon_i}{\epsilon_s}\right) = \mathcal{E}_i\left(\frac{3.9}{11.9}\right) \approx 0.32\,\mathcal{E}_i$$

(p.205)

$$C_D = \frac{dQ_s}{d\psi_s} = \frac{1}{d\psi_s}\left[\frac{\sqrt{2}\epsilon_s kT}{qL_D} F\left(\beta\psi_s, \frac{n_{p0}}{p_{p0}}\right)\right] = \frac{\sqrt{2}\epsilon_s kT}{qL_D} \frac{\frac{1}{2}\left\{-\beta e^{-\beta\psi_s} + \beta + \frac{n_{p0}}{p_{p0}}(\beta e^{\beta\psi_s} - \beta)\right\}}{F(\beta\psi_s, n_{p0}/p_{p0})}$$

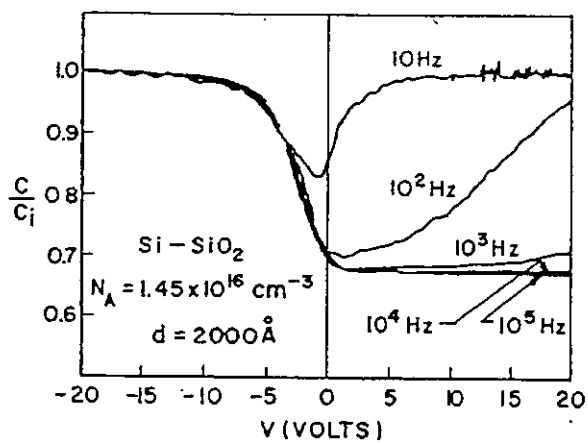
$$= \left(\frac{\epsilon_s}{\sqrt{2}L_D}\right) \frac{\left[1 - e^{-\beta\psi_s} + \frac{n_{p0}}{p_{p0}}(e^{\beta\psi_s} - 1)\right]}{F(\beta\psi_s, n_{p0}/p_{p0})}$$

$$C_D(\text{flat-band}) = \left(\frac{\epsilon_s}{\sqrt{2}L_D}\right) \frac{\left[1 - (1 - \beta\psi_s) + \cdots\cdots\right]}{\left[1 - \beta\psi_s + \frac{\beta^2\psi_s^2}{2} + \beta\psi - 1\right]^{1/2}} = \frac{\epsilon_s}{\sqrt{2}L_D} \frac{\beta\psi_s}{\left(\frac{\beta^2\psi_s^2}{2}\right)^{1/2}}$$

$$= \frac{\epsilon_s}{L_D} \qquad \text{----------(23)}$$

$$C_{FB}(\psi_s = 0) = \frac{C_i C_D(FB)}{C_i + C_D(FB)} = \frac{\frac{\epsilon_i}{d} \times \frac{\epsilon_s}{L_D}}{\epsilon_i/d + \epsilon_s/L_D} = \frac{\epsilon_i}{d + \left(\frac{\epsilon_i}{\epsilon_s}\right) L_D}$$

<p.207> The frequency effect on MIS C-V curve. The onset of low-frequency curves occurs at $f \lesssim 100$ Hz.



<p.209> From Eq. 17 at onset of heavy inversion

$$Q_n \to 0$$

$$\therefore Q_s = Q_n + qN_A W \approx qN_A W$$

$$\therefore V_T = \frac{Q_s}{C_i} + 2\psi_{Bp} = \frac{\sqrt{2\epsilon_s q N_A (2\psi_{Bp})}}{C_i} + 2\psi_{Bp} \qquad \text{-----(33)}$$

4.3 Silicon MOS Capacitor

4.3.1 Interface Traps

From Fig. 14a and b we have

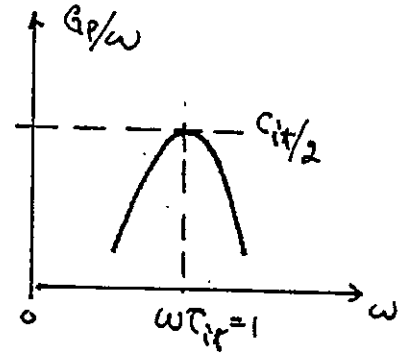
$$\frac{G_p}{\omega} = \frac{C_{it}\,\omega\tau_{it}}{1+\omega^2\tau_{it}^2}$$

$$\frac{d}{d\omega}\left(\frac{G_p}{\omega}\right) = (C_{it})\,\frac{(1+\omega^2\tau_{it}^2)\,\tau_{it} - \omega\tau_{it}(2\omega\tau_{it}^2)}{(1+\omega^2\tau_{it}^2)^2} = 0$$

$$\therefore\ 1+\omega^2\tau_{it}^2 - 2\omega^2\tau_{it}^2 = 0$$

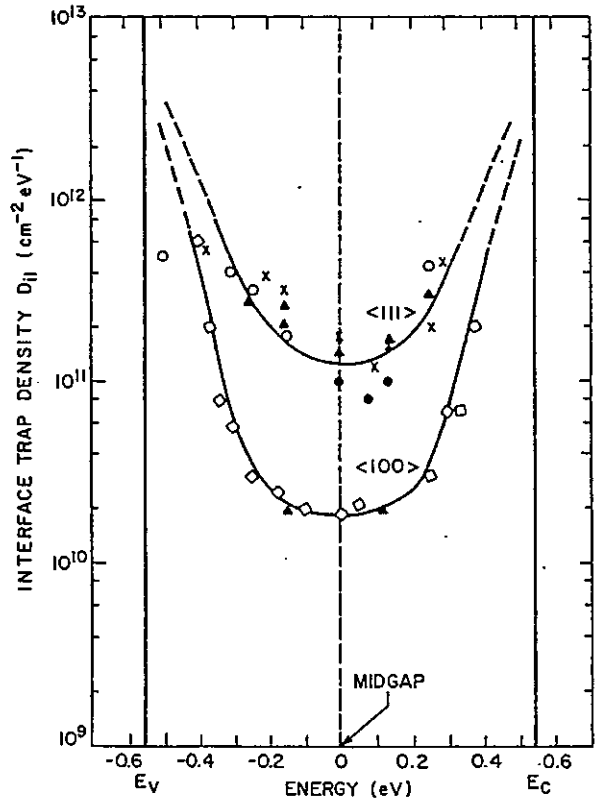
$$\text{or}\quad \omega\tau_{it} = 1 \qquad \text{so}\quad \tau_{it} = 1/\omega$$

$$\left.\frac{G_p}{\omega}\right|_{max} = \frac{C_{it}}{2}$$



(p. 220) $\because\ D_{it} = \frac{1}{g}\frac{dQ_{it}}{dE}$, $C_{it} = \frac{dQ_{it}}{dV}$

$$\therefore\ D_{it} = \frac{1}{g^2}\left(\frac{dQ_{it}}{dV}\right) = \frac{C_{it}}{g^2}$$



Left figure shows interface-trap density D_{it} in thermally oxidized silicon. The <100> surface has much lower D_{it} than <111> surface.

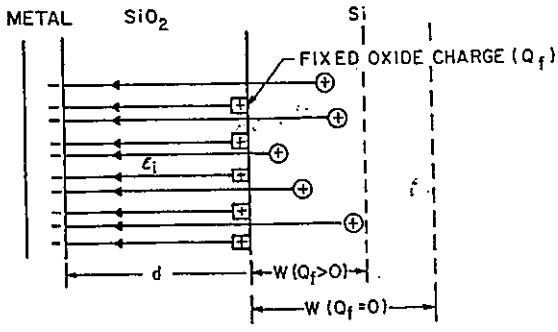
4.3.3 Oxide Charges and Work-Function Difference

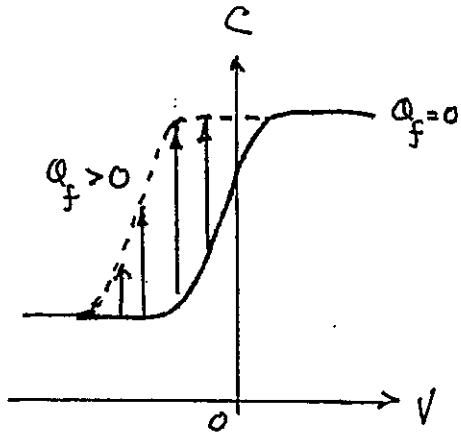
There are 3 kinds of oxide charges.

Q_f (fixed oxide charge)

Q_m (mobile ionic charge)

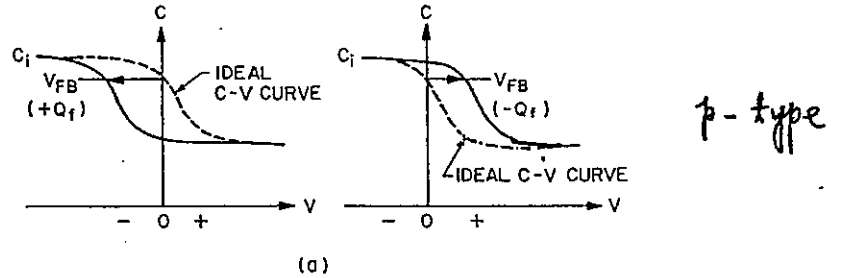
Q_{ot} (oxide trapped charge)



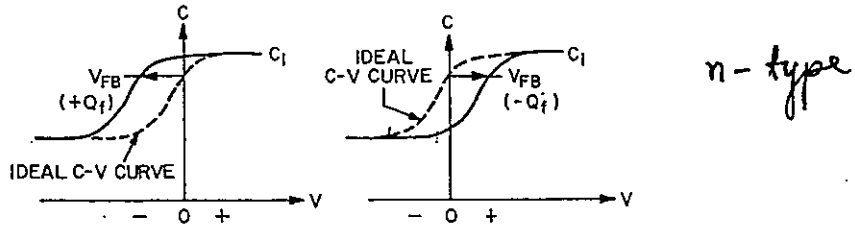


W (for $Q_f > 0$) is smaller than W (for $Q_f = 0$)

Thus C (for $Q_f > 0$) increases, causing the C-V curve to shift left. If $Q_f < 0$, C-V curve will shift right.



(a)



(b)

C–V curve shift along the voltage axis due to positive or negative fixed oxide charge. (a) For *p*-type semiconductor. (*b*) For *n*-type semiconductor.

The fixed charge Q_f is located within approximately 3 nm of the SiO_2-Si interface. This charge is fixed and cannot be charged or discharged over a wide variation of surface potential ψ_s. Generally, Q_f is positive and depends on oxidation and annealing conditions and on silicon orientation. It has been suggested that when the oxidation is stopped, some ionic silicon is left near the interface. These ions, along with uncompleted silicon bonds (e.g., Si-Si or Si-O bonds) at the surface, may result in the positive interface charge Q_f · Q_f can be regarded as a charge sheet located at the SiO_2-Si interface. Typical fixed-oxide charge densities for a carefully treated SiO_2-Si interface system are about 10^{10} cm^{-2} for a ⟨100⟩ surface and about 5×10^{10} cm^{-2} for a ⟨111⟩ surface. Because of the lower values of Q_{it} and Q_f, the ⟨100⟩ orientation is preferred for silicon MOSFETs.

The oxide-trapped charges Q_{ot} are associated with defects in the silicon dioxide. These charges can be created, for example, by X-ray radiation or high-energy electron bombardment. The traps are distributed inside the oxide layer. Most of process-related Q_{ot} can be removed by low-temperature annealing.

The mobile ionic charges Q_m, such as sodium or other alkali ions, are mobile within the oxide under raised-temperature (e.g., >100°C) and high-electric field operations. Trace contamination by alkali metal ions may cause stability problems in semiconductor devices operated under high-bias and high-temperature conditions. Under these conditions mobile ionic charges can move back and forth through the oxide layer and cause shifts of the C-V curves along the voltage axis. Therefore, special attention must be paid to the elimination of mobile ions in device fabrication.

The charges are the effective net charges per unit area in C/cm^2. We now evaluate the influence of these charges on the flat-band voltage. Consider a positive sheet charge per unit area, Q_o, within the oxide, as shown in Fig. 11. This positive sheet charge will

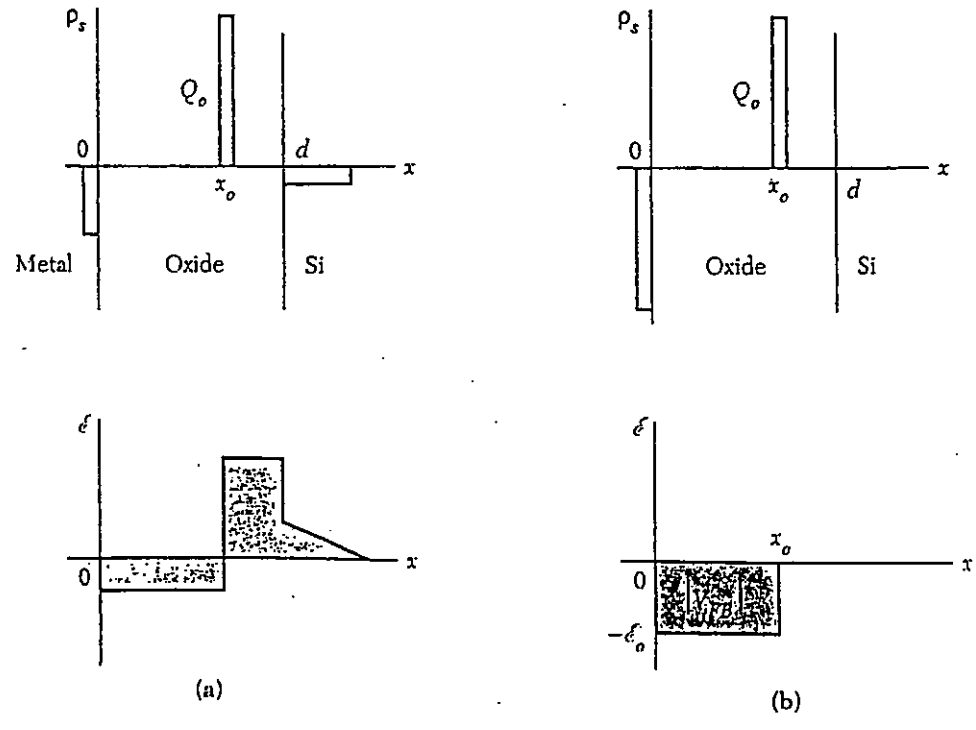


Fig. 11 Effect of a sheet charge within the oxide.[2] (*a*) Condition for $V_G = 0$. (*b*) Flat-band condition.

induce negative charges partly in the metal and partly in the semiconductor as shown in the upper part of Fig. 11*a*. The resulting field distribution, obtained from integrating Poisson's equation once, is shown in the lower part of Fig. 11*a* where we have assumed that there is no work function difference, or $q\phi_{ms} = 0$.

To reach the flat-band condition (i.e., no charge induced in the semiconductor), we must apply a negative voltage to the metal, as shown in Fig. 11*b*. As the negative voltage increases, more negative charges are put on the metal and thereby the electric-field distribution shift downward until the electric field at the semiconductor surface is zero. Under this condition the area contained under the electric-field distribution corresponds to the flat-band voltage V_{FB}:

$$V_{FB} = -\mathscr{E}_o x_o = -\frac{Q_o}{\varepsilon_{ox}} x_o = -\frac{Q_o}{C_o}\frac{x_o}{d}.$$

The flat-band voltage is, thus, dependent on both the density of the sheet charge Q_o and its location x_o within the oxide. When the sheet charge is located very close to the metal—that is , if $x_o = 0$—it will induce no charges in the silicon and therefore have no effect on the flat-band voltage. On the other hand, when Q_o is located very close to the semiconductor—$x_o = d$, such as the fixed-oxide charge Q_f, it will exert its maximum influence and give rise to a flat-band voltage

$$V_{FB} = -\frac{Q_o}{C_o}\frac{d}{d} = -\frac{Q_o}{C_o}.$$

For the more general case of an arbitrary space charge distribution within the oxide, the flat-band voltage is given by

$$\boxed{V_{FB} = \frac{-1}{C_o}\left[\frac{1}{d}\int_o^d x\rho(x)dx\right],}$$

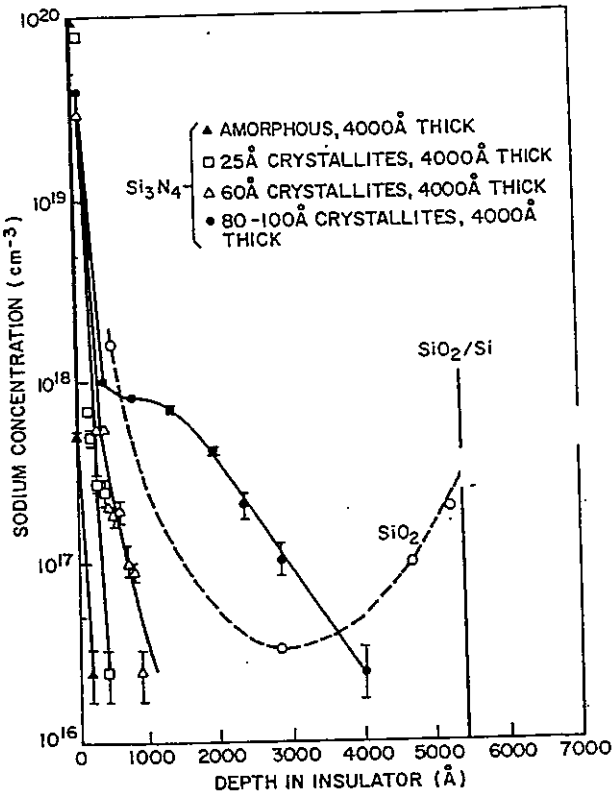
where $\rho(x)$ is the volume charge density in the oxide. Once we know $\rho_{ot}(x)$, the volume charge density for oxide-trapped charge, and $\rho_m(x)$, the volume charge density for mobile ionic charges, we can obtain Q_{ot} and Q_m and their corresponding contribution to the flat-band voltage:

$$Q_{ot} \equiv \frac{1}{d}\int_o^d x\rho_{ot}(x)dx,$$

$$Q_m \equiv \frac{1}{d}\int_o^d x\rho_m(x)dx.$$

If the value of the work function difference $q\phi_{ms}$ is not zero and if the value of the interface-trapped charges is negligible, the experimental capacitance-voltage curve will be shifted from the ideal theoretical curve by an amount

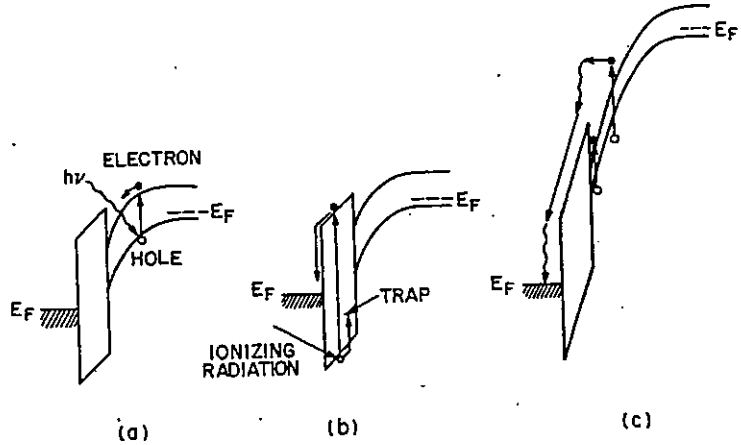
$$\boxed{V_{FB} = \phi_{ms} - \frac{(Q_f + Q_m + Q_{ot})}{C_o}.}$$



Sodium concentration versus depth in silicon dioxide and silicon nitride with different crystallite sizes.

To prevent Q_m contamination of the oxide during device life we can protect it with a film impervious to Q_m, such as using amorphous or small crystallite Si_3N_4 (shown above).

4.3.5 Nonequilibrium at Avalanche



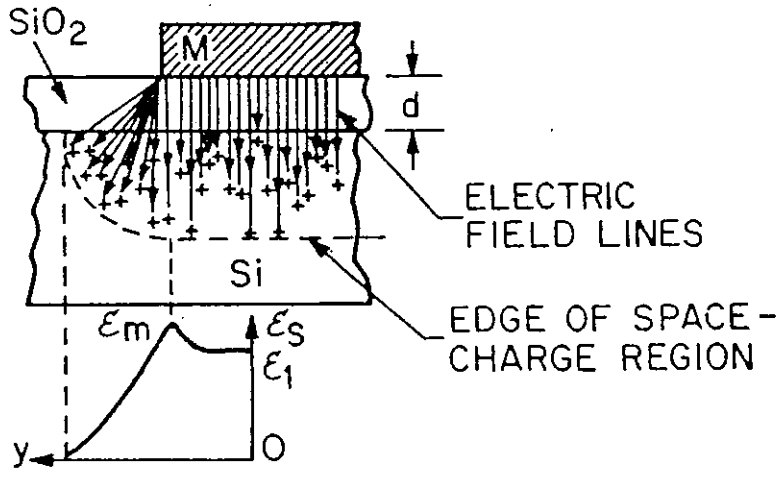
Band diagrams of MIS diode under the following conditions: (a) illumination; (b) radiation; (c) avalanche injection. (After Nicollian and Brews, Ref. 7.)

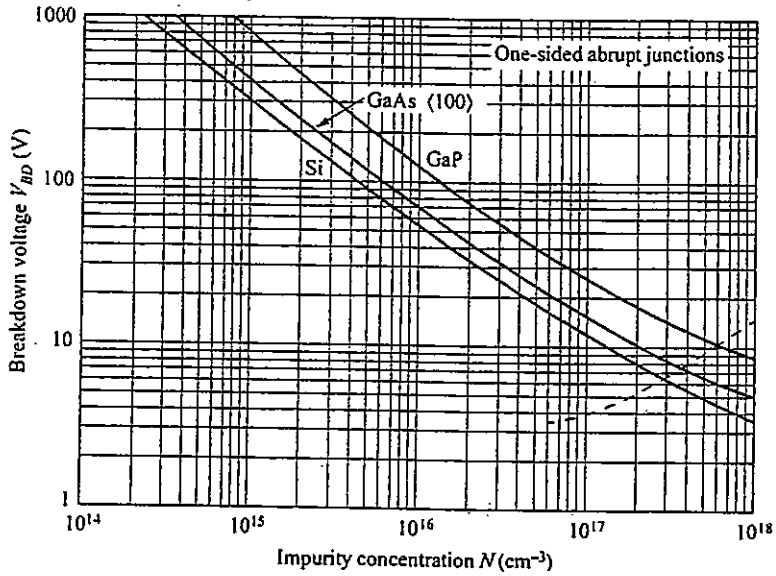
(2) Illumination: The capacitance (on the C-V curve) in the inversion region will approach the low-frequency value as the intensity of illumination is increased. This is due to the generation of electron-hole pairs by photons, which causes a decrease of ψ_s under constant applied voltage $\rightarrow W \downarrow \rightarrow C \uparrow$.

(2) Radiation (x-ray or γ-ray): can create electron-hole pairs. electrons will drift rapidly & flow out into the external circuit. holes may be trapped in the oxide to form Q_{ot} — causing C-V curve to shift.

(3) Avalanche

With reference to Fig. 27 (p. 232), For a given oxide thickness, a doping exists which yields minimum breakdown voltage; edge breakdown ($\mathcal{E}_m > \mathcal{E}_1$ shown before) dominates to the left of the minimum, & uniform breakdown ($\mathcal{E}_m = \mathcal{E}_1$) dominates to the left. To avoid edge breakdown, the ratio $d/W_{max} > 0.3$ where W_{max} is the maximum depletion-layer width at breakdown.

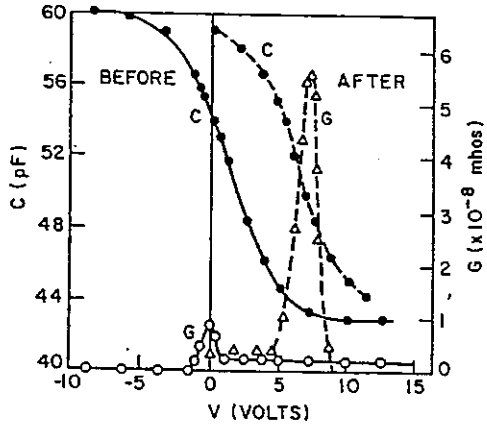




Avalanche breakdown voltage in Si, <100>-oriented GaAs and GaP one-sided abrupt junctions. The dashed line indicates the maximum doping beyond which tunneling will dominate the breakdown characteristics.

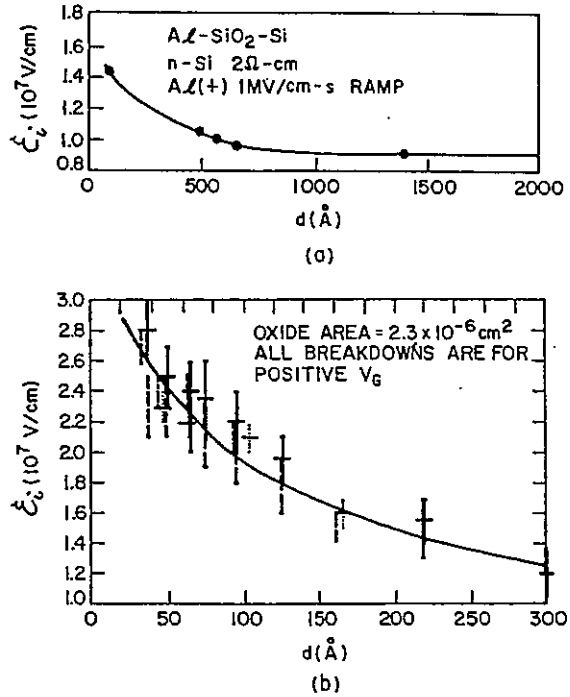
If we compare the above curve with Fig. 27 <p. 232> we notice that at lower dopings the MOS capacitor has a lower breakdown voltage than that of the p-n with the same doping. This is due to edge breakdown. At higher dopings, the MOS capacitor has much higher breakdown voltage, because

V_B (MOS capacitor) = V_i (voltage across the oxide = $\mathcal{E}_i d$)
+ V_B (semiconductor breakdown voltage with a maximum field $\mathcal{E}_m$ at oxide-semiconductor interface).



Capacitance and conductance versus voltage for a MIS diode before and after avalanche injection of electrons.

The hot electron injection cause a flat-band shift of C-V curve. ($-Q_f$). The increase of G is due to an increase of the interface-trap density ($1.2\times10^{11} \rightarrow 7.9\times10^{11}$ / cm^2/eV) (p.236). As the oxide thickness becomes thinner, $\mathcal{E}_i$ ↑. However for $d \leq 4nm$, $\mathcal{E}_i$ will drop due to an increase of tunneling current.



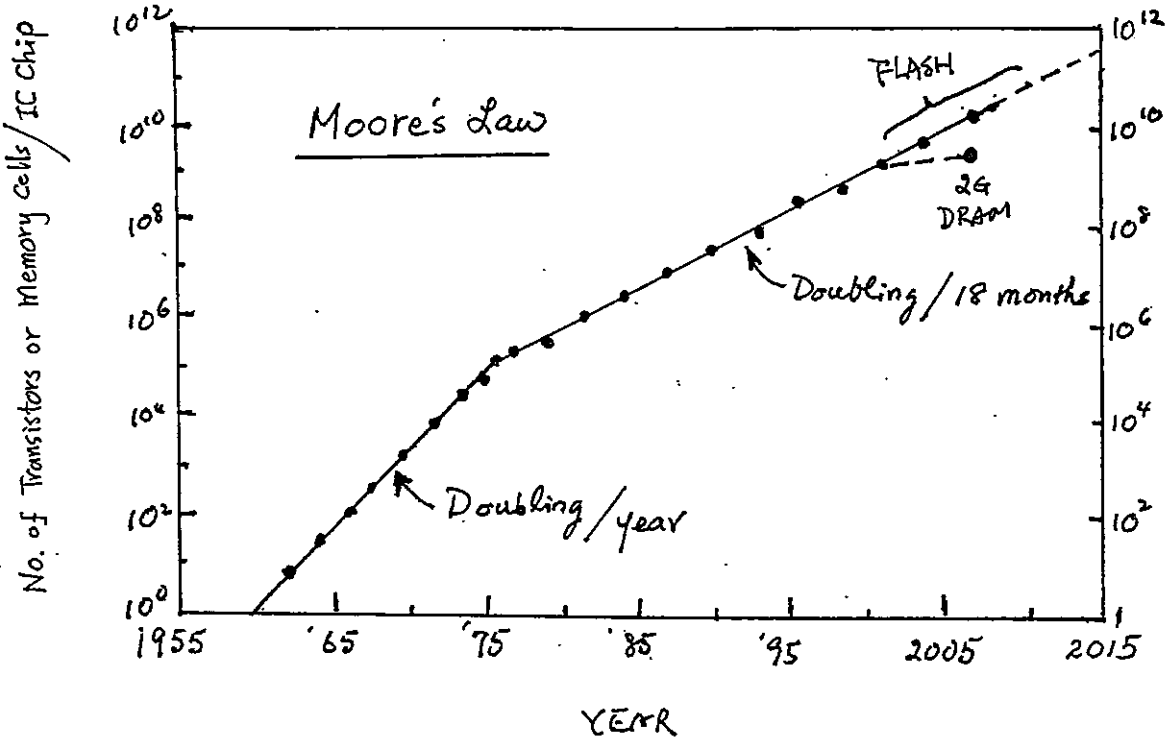
Breakdown field versus oxide thickness in SiO_2. (a) From 100 to 2000 Å. (b) From 40 to 300 Å.

Chapter 6

MOSFETs

6.1 Introduction

MOSFET is the most important device for advanced ICs. From 1959 to 1975, the complexity doubled every year. Since 1975, the complexity has been doubling every 18 months (i.e., increasing by a factor of 4 every 3 years).



6.2 Basic Device Characteristics

(p.302) $$Q_B = -qN_A W_m = -qN_A\sqrt{\frac{2\epsilon_s \psi_s}{qN_A}} = -\sqrt{2qN_A\epsilon_s\psi_s}$$

(p. 208 Eq. 32)

$$= -\sqrt{2qN_A\epsilon_s(V_D + 2\psi_B)} \quad \text{---- (12)}$$

For strong inversion the F function (Eq. 7) is dominated by the 2nd and the 4th terms:

$$F(\beta\psi_p, V_D, \frac{n_{po}}{p_{po}}) \cong \sqrt{\beta\psi_p + \left(\frac{n_{po}}{p_{po}}\right)\exp(-\beta V_D)\exp(\beta\psi_p)}$$

Therefore

$$|Q_n| = Q_s - Q_B = \frac{\sqrt{2}\,\epsilon_s kT}{qL_D} F(\beta\psi_s, V_D, \frac{n_{po}}{p_{po}}) - \sqrt{2qN_A\epsilon_s\psi_s}$$

$$= \sqrt{2}qN_A L_D\left[\sqrt{\beta\psi_s + \left(\frac{n_{po}}{p_{po}}\right)\exp(\beta\psi_s - \beta V_D)} - \sqrt{\beta\psi_s}\right] \quad \text{----(13)}$$

$$\left(L_D \equiv \sqrt{kT\epsilon_s/N_A q^2}\right)$$

At onset of strong inversion, the electrical field $\mathcal{E}_s$ at the oxide-semiconductor interface reaches its maximum value, and we only consider the second term in the F function:

$$\mathcal{E}_s = \frac{\sqrt{2}kT}{qL_D}\sqrt{\exp(-\beta\psi_s) + \beta\psi_s - 1 + \frac{n_{po}}{p_{po}}\exp(-\beta V_D)\left[\exp(\beta\psi_s) - \beta\psi_s\exp(\beta V_D) - 1\right]}$$

$$\approx \frac{\sqrt{2}kT}{qL_D}\sqrt{\beta\psi_s} = \sqrt{\frac{2qN_A\psi_s}{\epsilon_s}} = \sqrt{\frac{2qN_A(\Delta\psi_i + 2\psi_B)}{\epsilon_s}} \quad \text{----(18)}$$

To derive Eq. 19 (↓ Eq. 19 p. 203)

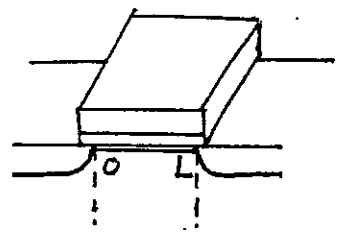
$$|Q_n| = Q_s - Q_B = V_i C_{ox} - Q_B = (\mathcal{E}_{ox} d) C_{ox} - Q_B$$

$$= \left[V_G - \psi_s(y)\right] C_{ox} - Q_B$$

$$= \left[V_G - \Delta\psi_i(y) - 2\psi_B\right] C_{ox} - \sqrt{2\epsilon_s qN_A\left[\Delta\psi_i(y) + 2\psi_B\right]} \quad \text{----(19)}$$

Equation 19 above will be used as the channel charge to calculate the conduction current in the channel.

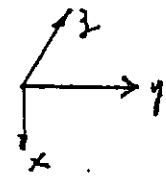
6.2.2 Current - Voltage Characteristics



From Eq. 21, we have

$$I_D(y)\,dy = Z\,|Q_n(y)|\,v(y)\,dy$$

$$\int_0^L I_D(y)\,dy = \int_0^L Z\,|Q_n(y)|\,v(y)\,dy$$



$\because$ $I_D(y)$ = constant, independent of y

$$\therefore I_D = \frac{Z}{L}\int_0^L |Q_n(y)|\,v(y)\,dy \quad \cdots\cdots (22)$$

(p.305) If $V_D \ll (V_G - V_T)$, we can expand Eq. 23 in power series around V_D:

$$I_D \approx \frac{Z}{L}\mu_n C_{ox}\left\{(V_G - V_{FB} - 2\psi_B - \frac{V_D}{2})V_D - \frac{2}{3}\frac{\sqrt{2\epsilon_s qN_A}}{C_{ox}}\left[(2\psi_B)^{3/2}\left(1+\frac{3}{2}\frac{V_D}{2\psi_B} + \frac{\frac{3}{2}\cdot\frac{1}{2}}{2}\left(\frac{V_D}{2\psi_B}\right)^2 + \cdots\cdots\right] - (2\psi_B)^{3/2}\right\}$$

$$= \frac{Z}{L}\mu_n C_{ox}\left\{(V_G - V_{FB} - 2\psi_B - \frac{V_D}{2})V_D - \frac{\sqrt{2\epsilon_s qN_A(2\psi_B)}}{C_{ox}}V_D \cdots\right\}$$

$$= \frac{Z}{L}\mu_n C_{ox}\left\{V_G - \underbrace{\left[V_{FB} + 2\psi_B + \frac{\sqrt{2\epsilon_s qN_A(2\psi_B)}}{C_{ox}}\right]}_{V_T} - \frac{V_D}{2}\right\}V_D$$

$$= \frac{Z}{L}\mu_n C_{ox}\left(V_G - V_T - \frac{V_D}{2}\right)V_D \quad \cdots\cdots (24)$$

$\therefore$ Drain conductance:

$$g_D \equiv \left.\frac{\partial I_D}{\partial V_D}\right|_{V_G = \text{constant}} = \frac{Z}{L}\mu_n C_{ox}(V_G - V_T - V_D)$$

Transconductance:

$$g_m \equiv \left.\frac{\partial I_D}{\partial V_G}\right|_{V_D = \text{constant}} = \frac{Z}{L}\mu_n C_{ox} V_D$$

(p. 306) Assume $|Q_n(y)| = C_{ox}\left[V_G - V_T - M\Delta\psi_i(y)\right]$

$$I_D = \frac{Z}{L}\int_0^L C_{ox}\left\{V_G - V_T - M\Delta\psi_i(y)\right] v(y)\,dy \Big|_{v(y)=\mu_n\mathcal{E}_y=\mu_n\frac{d\Delta\psi_i}{dy}}$$

$$= \frac{Z}{L}\int_0^{V_D} C_{ox}\left[V_G - V_T - M\Delta\psi_i(y)\right]\mu_n\, d\Delta\psi_i$$

$$= \frac{Z}{L}\mu_n C_{ox}\left(V_G - V_T - \frac{MV_D}{2}\right)V_D \quad \cdots\cdots\cdots (31)$$

(p 308) $I_D = Z\left|Q_n(y)\right|\dfrac{\mu_n\mathcal{E}}{1+\mathcal{E}/\mathcal{E}_c}$

$$\therefore I_D\left(1+\frac{d\Delta\psi_i/dy}{\mathcal{E}_c}\right) = Z\left|C_{ox}(V_G - V_T - M\Delta\psi_i)\mu_n\frac{d\Delta\psi_i}{dy}\right.$$

$$\int_0^L I_D\left(\mathcal{E}_c + \frac{d\Delta\psi_i}{dy}\right)dy = \int_0^{V_D} Z\mu_n C_{ox}\left(V_G - V_T - M\Delta\psi_i\right)d\Delta\psi_i$$

$$I_D\left(\mathcal{E}_c L + V_D\right) = Z\mu_n C_{ox}\left(V_G - V_T - \frac{MV_D}{2}\right)V_D$$

$$\therefore I_D = \frac{Z\mu_n C_{ox}\mathcal{E}_c}{L\mathcal{E}_c + V_D}\left(V_G - V_T - \frac{MV_D}{2}\right)V_D \quad \cdots\cdots (39)$$

Under velocity saturation condition, the inversion charge at the source is given by Eq. 19 (p. 302)

$$|Q_n| = \left(V_G - 0 - 2\psi_B\right)C_{ox} + \sqrt{2\varepsilon_s q N_A\left(0 + 2\psi_B\right)}$$

$$= \left(V_G - V_T\right)C_{ox}$$

$$\therefore I_{Dsat} = Z\left|Q_n\right|v_s = Z v_s\left(V_G - V_T\right)C_{ox} \quad \cdots (41)$$

The transconductance is

$$g_m \equiv \frac{\partial I_{Dsat}}{\partial V_G} = Z C_{ox} v_s \neq f(V_G) \quad \cdots\cdots (42)$$

(p. 311) Fig. 14 For ballistic transport, if the inversion carrier density is $10 \times 10^{12}\ cm^{-2}$, v_{inj} is $2 \times 10^{7}\ cm/s$. The maximum current is then

$$I_{max} = |Q_n|\, v_{inj} = (10 \times 10^{12})(1.6 \times 10^{-19}) \times 2 \times 10^{7}$$

$$= 3.2 \times 10\ A/cm = 3.2\ A/mm.$$

(p. 312)

$$I_{D\,sat} \sim (V_G - V_T)^2 \quad \text{for constant mobility regime}$$

$$\sim (V_G - V_T) \quad \text{for saturation-velocity regime}$$

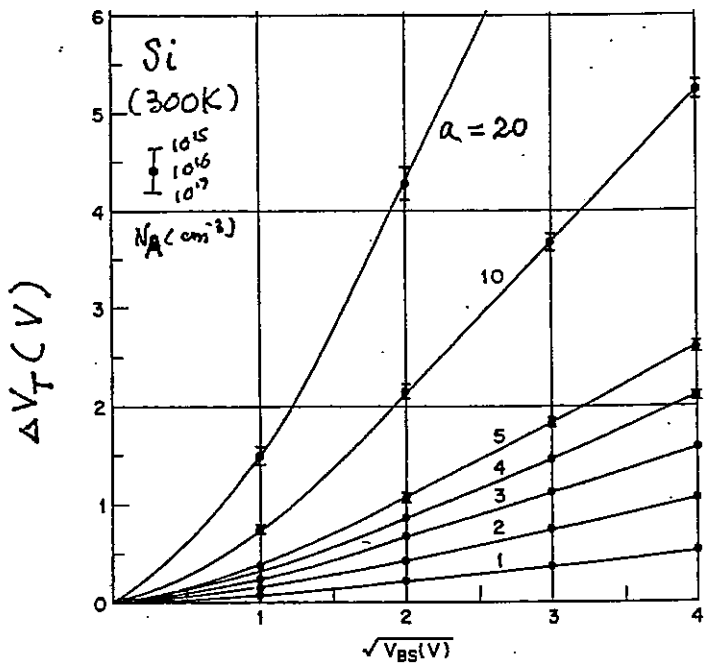
$$\sim (V_G - V_T)^{3/2} \quad \text{for ballistic regime.}$$

6.2.3 Threshold Voltage

The threshold voltage shift ΔV_T as a function of the substrate bias can be expressed as

$$\Delta V_T = \frac{a}{\beta}\left(\sqrt{2\beta\psi_B - \beta V_{BS}} - \sqrt{2\beta\psi_B}\right) \quad \cdots\cdots (51)$$

where $\beta \equiv q/kT$ $\quad a \equiv \sqrt{2}(\epsilon_s/\epsilon_{ox})(d/L_D)$

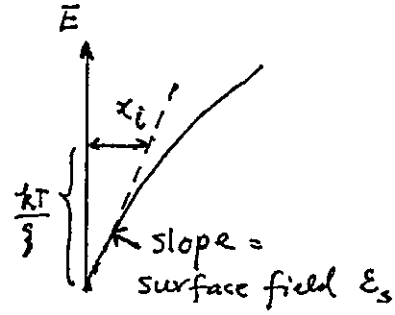


$a \sim d$,

$a \sim 1/L_D \sim \sqrt{N_A}$

As L_D decreases (i.e., $N_A \uparrow$) and/or d increases, then "a" increases causing a large threshold voltage shift with V_{BS}. To minimize ΔV_T due to V_{BS}, small "a" should be used — low substrate doping and thin oxide thickness.

6.2.4 Subthreshold Region



$$\mathcal{E}_s = \frac{1}{q}\frac{dE}{dx} = \frac{kT/q}{x_i}$$

$$\therefore x_i = \frac{kT}{q\mathcal{E}_s}$$

$$I_D = -qAD_n\frac{dn}{dy} = +qAD_n\frac{n(0)-n(L)}{L}$$

$$D_n = \frac{kT}{q}\mu_n$$

A = area $= x_i Z = \left(\frac{kT}{q\mathcal{E}_s}\right)Z$

$$\mathcal{E}_s = \frac{|Q_B|}{\epsilon_s} = \sqrt{2qN_A\psi_s/\epsilon_s}$$

$$n(0) = n_{p0}e^{\beta\psi_s}$$

$$n(L) = n_{p0}e^{\beta\psi_s - \beta V_D}$$

$$\therefore I_D = q\left(\frac{kTZ}{q\mathcal{E}_s}\right)\left(\frac{kT}{q}\mu_n\right)\frac{n_{p0}\left[e^{\beta\psi_s} - e^{\beta\psi_s - \beta V_D}\right]}{L}$$

$$= \frac{Z\mu_n}{L\beta^2}\sqrt{\frac{q\epsilon_s N_A}{2\psi_s}}\left(\frac{n_i}{N_A}\right)^2 \exp(\beta\psi_s)\left[1 - \exp(-\beta V_D)\right]$$

$$\approx \frac{Z\mu_n}{L\beta^2}\sqrt{\frac{q\epsilon_s N_A}{2\psi_s}}\left(\frac{n_i}{N_A}\right)^2 \exp(\beta\psi_s) \quad \text{for } V_D \gg kT/q \quad \cdots(57)$$

To derive the subthreshold swing S: from Eq. 57

$$I_D = \frac{\text{Constant}}{\sqrt{\psi_s}}\exp(\beta\psi_s) \sim \exp(\beta\psi_s)$$

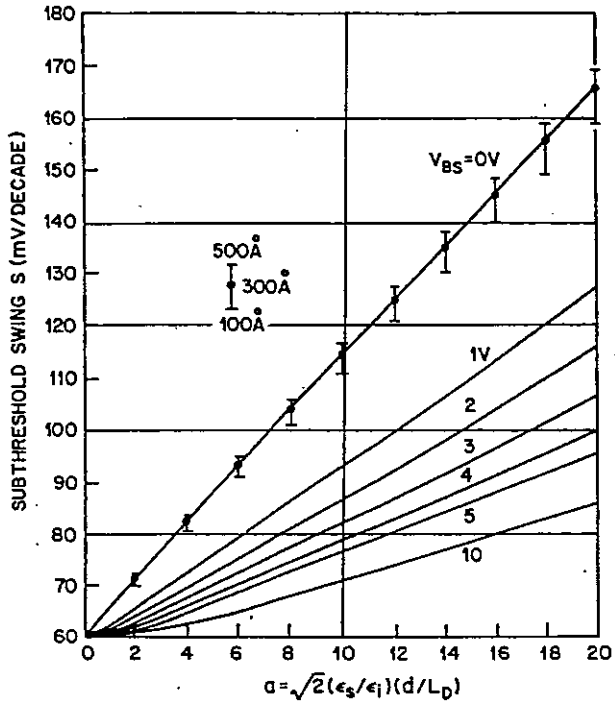
$$\ln I_D \sim \beta\psi_s = \frac{q\psi_s}{kT} \qquad d(\ln I_D) = \frac{q}{kT}d\psi_s$$

$$S \equiv (\ln 10)\frac{dV_G}{d(\ln I_D)} = (\ln 10)\frac{dV_G}{\frac{q}{kT}d\psi_s} = (\ln 10)\left(\frac{kT}{q}\right)\frac{dV_G}{d\psi_s}$$

From Eq. 58, 59

$$= (\ln 10)\left(\frac{kT}{q}\right)\left(\frac{C_{ox}+C_D}{C_{ox}}\right) \quad \cdots\cdots(60)$$

When a substrate bias is applied, it increases the value of ψ_s, thus C_D is reduced and therefore S is reduced



Subthreshold swing versus a for various substrate reverse bias.

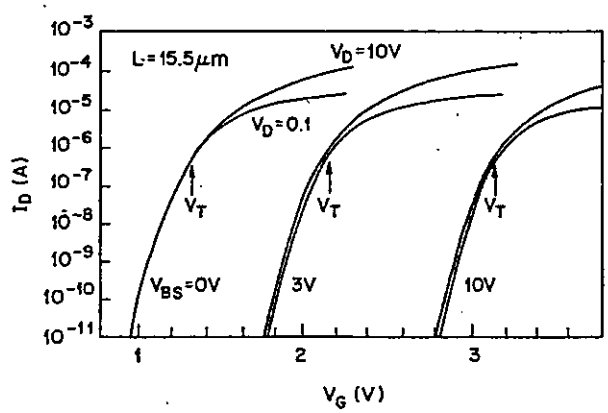
$|V_{BS}|$ ↑

C_D ↓

S ↓

For a given V_{BS}

S ↑ with "a"



Experimental subthreshold characteristics for a long-channel device

$d = 57$ nm

$N = 5.6 \times 10^{15}$ cm^{-3}

$L = 15.5\,\mu$m

$a = 4$

$V_{BS}=0$ $S = 83$ mV/dec

$V_{BS}=3$ $S = 67$ "

$\Delta V_T = 0.75$ V

$V_{BS}=10$ $S = 63$

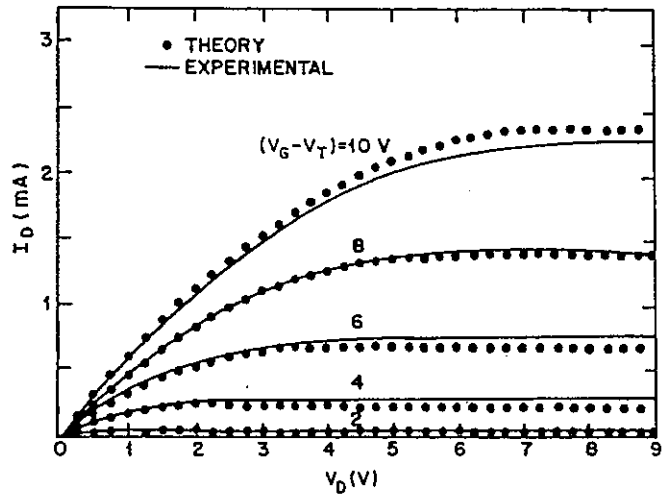
$\Delta V_T = 1.7$ V

(p. 316)

The total drain current density including both drift and diffusion components is given by Eq. 62. The drain current is given by Eq. 63.

$$V_G - V_{FB} = \frac{\sqrt{2}\,\epsilon_s kT}{C_{ox} q L_D} F\left(\beta\psi_s, \Delta\psi_i, \frac{n_{po}}{p_{po}}\right) + \psi_s \quad \text{-----(64)}$$

For a particular device with known parameters, Eq. 63 can be calculated numerically to give accurate results for the entire range of drain voltage, from the linear region to the saturation region:



Theoretical (dots) and experimental (solid lines) drain characteristics of a p-channel MOSFET having $d = 2000$ Å, $N_D = 4.6 \times 10^{14}$ cm^{-3}, and $\mu_p = 256$ cm^2/V-s.

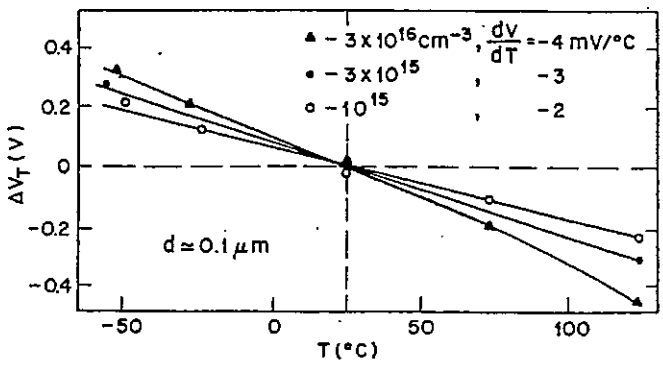
6.2.6 Temperature Dependence

$$n_i \sim T^{3/2} \exp(-E_{g0}/2kT)$$

$$\ln n_i \approx \ln T^{3/2} - E_{g0}/2kT$$

$$\frac{d\psi_B}{dT} = \frac{d}{dT}\left[\frac{kT}{q}\ln\left(\frac{N_A}{n_i}\right)\right] \approx \frac{k}{q}\ln\left(\frac{N_A}{n_i}\right) + \frac{kT}{q}\left[\frac{-E_{g0}(2k)}{(2kT)^2}\right]$$

$$= \frac{1}{T}\left(\psi_B - \frac{E_{g0}}{2q}\right) \quad \text{-- -- -- -- (70)}$$

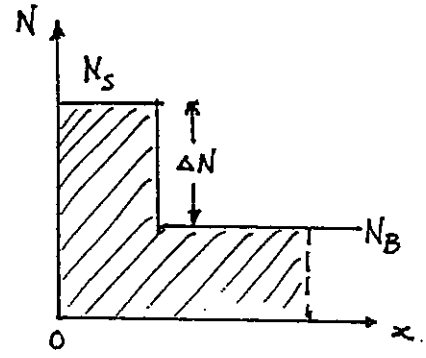


Experimental measurement of V_T versus temperature for a given N_A and d

$dV_T/dT \approx$ constant (straight line)

6.3 Nonuniform Doping and Buried-Channel Device

6.3.1 High-Low Profile



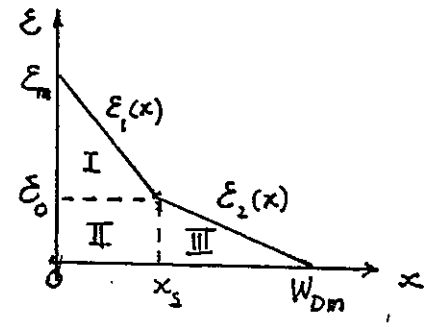
From Poisson Equation

$$\mathcal{E} = \int_0^x \frac{qN(x)\,dx}{\epsilon_s}$$

$$\mathcal{E}_1 = \mathcal{E}_m - \frac{qN_S x}{\epsilon_s} \qquad (1)$$

$$\mathcal{E}_0 = \mathcal{E}_m - \frac{qN_S x_s}{\epsilon_s} \qquad (2)$$

$$\mathcal{E}_2 = \mathcal{E}_0 - \frac{qN_B(x-x_s)}{\epsilon_s} \qquad (3)$$



At $x = W_{Dm}$, $\mathcal{E}_2(W_{Dm}) = 0$

From Eq. 3

$$\mathcal{E}_0 - \frac{qN_B(W_{Dm}-x_s)}{\epsilon_s} = \left(\mathcal{E}_m - \frac{qN_S x_s}{\epsilon_s}\right) - \frac{qN_B(W_{Dm}-x_s)}{\epsilon_s} = 0$$

(Eq. 2)

$$\therefore \quad \mathcal{E}_m = \frac{qx_s(N_S-N_B)}{\epsilon_s} + \frac{qN_B W_{Dm}}{\epsilon_s} \qquad (4)$$

From Eq. 2 & 4

$$\mathcal{E}_0 = \mathcal{E}_m - \frac{qN_S x_s}{\epsilon_s} = \frac{qx_s(N_S-N_B)}{\epsilon_s} + \frac{qN_B W_{Dm}}{\epsilon_s} - \frac{qN_S x_s}{\epsilon_s} = \frac{qN_B(W_{Dm}-x_s)}{\epsilon_s}$$

The area under $\mathcal{E}$-x plot is the voltage $(\psi_s + V_{BS})$:

$$\therefore \psi_s + V_{BS} = \text{I} + \text{II} + \text{III} = (\mathcal{E}_m - \mathcal{E}_0)\frac{x_s}{2} + \mathcal{E}_0 x_s + \mathcal{E}_0(W_{Dm}-x_s)/2$$

$$= \frac{qN_S x_s}{\epsilon_s}\left(\frac{x_s}{2}\right) + \frac{qN_B}{\epsilon_s}(W_{Dm}-x_s)\left(x_s + \frac{W_{Dm}-x_s}{2}\right)$$

$$= qN_S x_s^2/2\epsilon_s + \frac{qN_B}{2\epsilon_s}(W_{Dm}^2 - x_s^2)$$

$$\therefore \frac{qN_B W_{Dm}^2}{2\epsilon_s} = \psi_s + V_{BS} - \frac{qx_s^2}{2\epsilon_s}(N_S-N_B) = \psi_s + V_{BS} - \frac{qx_s^2\Delta N}{2\epsilon_s}$$

$$W_{Dm} = \sqrt{\frac{2\epsilon_s}{qN_B}\left(\psi_s + V_{BS} - \frac{q\Delta N x_s^2}{2\epsilon_s}\right)} \qquad \text{-------(74)}$$

$$V_T = V_{FB} + \underbrace{2\psi_B}_{\psi_s} + \frac{Q_B}{C_{ox}}$$ ← Cross hatched area on p. 66

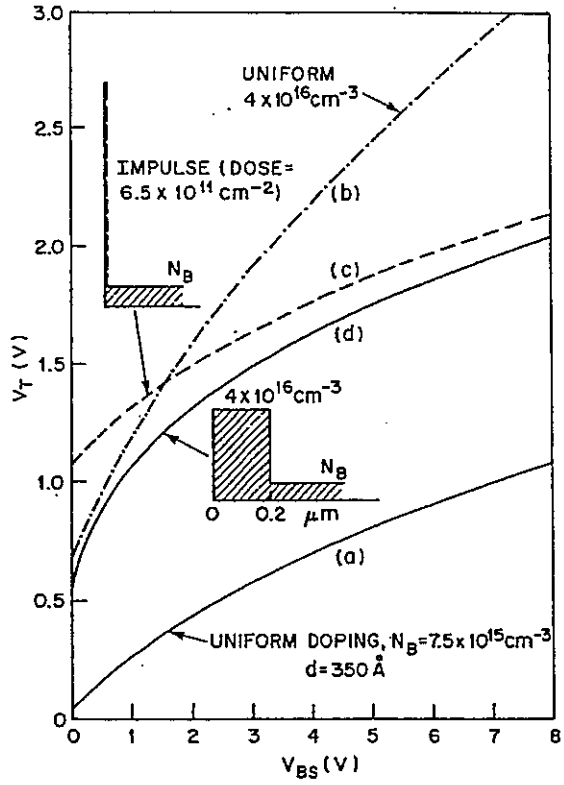
$$= V_{FB} + 2\psi_B + \frac{\gamma N_B W_{dm} + \gamma \Delta N x_s}{C_{ox}}$$

$$= V_{FB} + 2\psi_B + \frac{1}{C_{ox}}\sqrt{2\gamma\epsilon_s N_B \left(2\psi_B - \frac{\gamma \Delta N x_s^2}{2\epsilon_s}\right)} + \frac{\gamma \Delta N x_s}{C_{ox}} \quad \text{.....(73)}$$

If $x_s \to 0$ and $\Delta N x_s \to D_I$ (total dose) delta-function dose.

$$\therefore \quad \Delta V_T = \gamma D_I / C_{ox} \quad \text{- - - - - - - - (75)}$$

This is one way to do threshold adjust.



Calculated substrate sensitivity for various doping profiles.

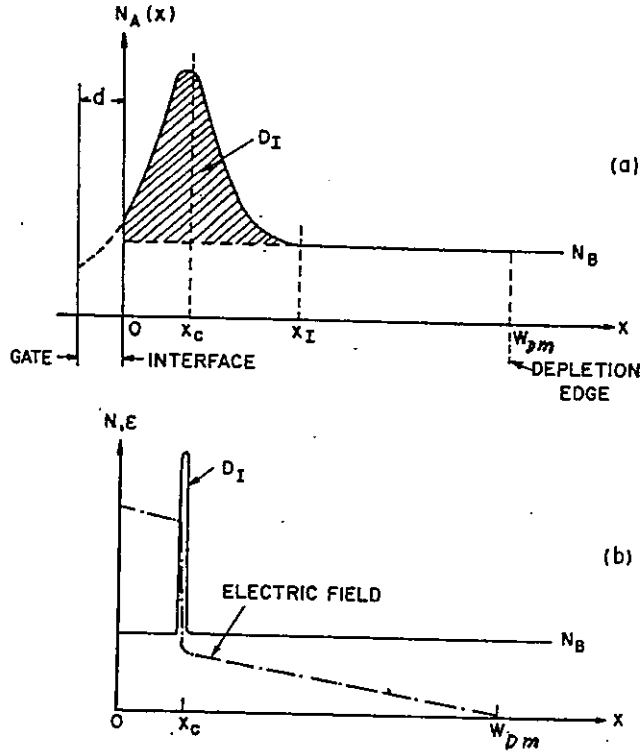
Case (a) Uniformly doped $7.5 \times 10^{15}\ cm^{-3}$ ($a = 3.2$)

Case (b) " " $4 \times 10^{16}\ cm^{-3}$ ($a = 7.4$)

Case (c) delta-function dose of $6.5 \times 10^{11}\ cm^{-2}$
($\gamma D_I / C_{ox} = 1.1\ V$) Parallel to case (a)

Case (d) High-low doping: for $W_{dm} < 0.2\mu m$, follows (b)
for $W_{dm} > 0.2\mu m$, follows (c)

To obtain an accurate V_T we have to consider the actual doping profile, which can be replaced by D_I and the centroid x_c:

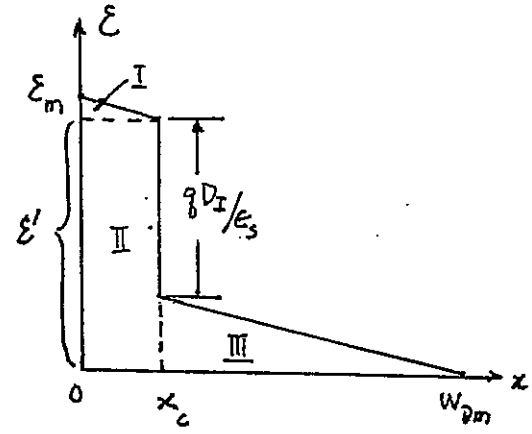


(a) Implanted doping profile. (b) Delta function equivalent to (a).

$$D_I \equiv \int_0^{W_{Dm}} \Delta N(x)\,dx$$

$$x_c \equiv \frac{1}{D_I}\int_0^{W_{Dm}} x\,\Delta N(x)\,dx$$

From the right figure, we have



$$\mathcal{E}_m - \frac{qN_B x_c}{\epsilon_s} - \frac{qD_I}{\epsilon_s} - \frac{qN_B(W_{Dm}-x_c)}{\epsilon_s} = 0$$

$$\therefore \mathcal{E}_m = \frac{q}{\epsilon_s}(N_B W_{Dm} + D_I)$$

$$\mathcal{E}' = \mathcal{E}_m - \frac{qN_B x_c}{\epsilon_s} = \frac{q}{\epsilon_s}\left(N_B W_{Dm} + D_I - N_B x_c\right)$$

ψ_s = area under $\mathcal{E}$-x curve

$$= \text{I} + \text{II} + \text{III}$$

$$= \frac{qN_B x_c}{\epsilon_s}\left(\frac{x_c}{2}\right) + \frac{q}{\epsilon_s}\left(N_B W_{Dm} + D_I - N_B x_c\right)x_c + \frac{qN_B}{2\epsilon_s}(W_{Dm} - x_c)^2$$

$$= \frac{qN_B}{\epsilon_s}\left[\frac{x_c^2}{2} + W_{Dm}x_c + \frac{D_I x_c}{N_B} - x_c^2 + \frac{W_{Dm}^2 - 2W_{Dm}x_c + x_c^2}{2}\right]$$

$$= \frac{qN_B}{\epsilon_s}\left(\frac{D_I x_c}{N_B} + \frac{W_{Dm}^2}{2}\right)$$

$$W_{Dm} = \sqrt{\frac{2\epsilon_s}{qN_B}\left(\psi_s - \frac{qD_I x_c}{\epsilon_s}\right)} \quad \text{---- (79)}$$

$$V_T = V_{FB} + \psi_s + \frac{Q_B}{C_{ox}} = V_{FB} + 2\psi_B + \frac{qN_B W_{Dm} + qD_I}{C_{ox}}$$

$$= V_{FB} + 2\psi_B + \frac{1}{C_{ox}}\sqrt{2q\epsilon_s N_B\left(2\psi_B - \frac{qD_I x_c}{\epsilon_s}\right)} + \frac{qD_I}{C_{ox}} \quad \text{---- (78)}$$

- When $x_c = 0$, the implant is a delta function at the Si-SiO_2 interface

$$\Delta V_T = qD_I/C_{ox}$$

- As x_c increases, $\Delta V_T \downarrow$. Eventually, x_c reaches the depletion-layer width W_{Dm}, and the depletion edge becomes clamped to the implant

. For a uniform N_B doping, the depletion-layer width is

$$W_{DmO}^2 = \frac{2\epsilon_s}{qN_B}\psi_s$$

From Eq. 79

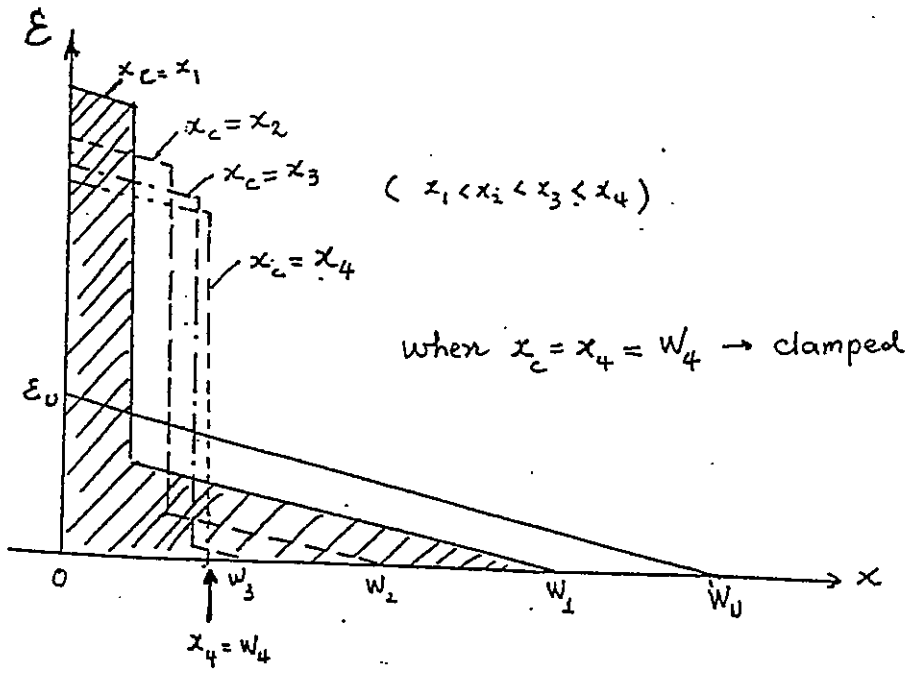
$$W_{Dm}^2 = \frac{2\epsilon_s}{qN_B}\left(\psi_s - \frac{qD_I x_c}{\epsilon_s}\right) = W_{DmO}^2 - \frac{2D_I x_c}{N_B}$$

when x_c clamped with W_{Dm}, i.e., $x_c = W_{Dm}$ then

$$x_c^2 = W_{DmO}^2 - \frac{2D_I x_c}{N_B}$$

$$\therefore D_I(x_c = W_{Dm}) = \frac{N_B(W_{DmO}^2 - x_c^2)}{2x_c}$$

- When $x_c > W_{Dm}$, it no long has any effect on threshold voltage & depletion-layer width.



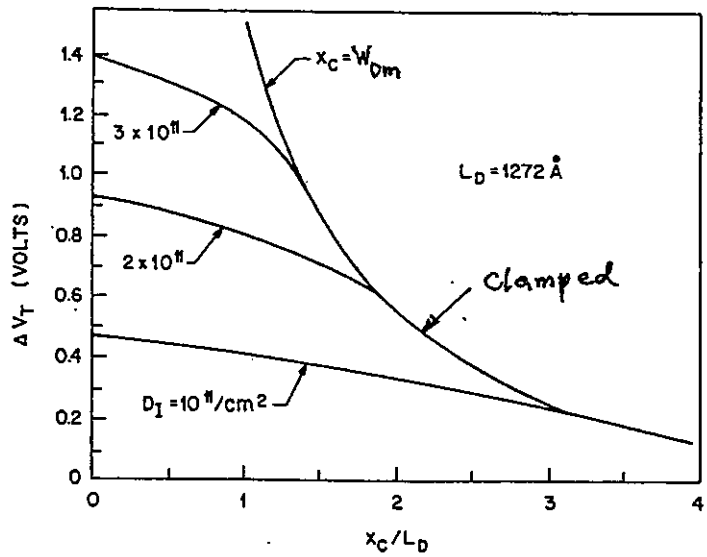
ψ_{SU} = Area of the triangle for uniformly doped N_B

$= \mathcal{E}_U - W_U - 0$

ψ_{SI} = Area of implanted case = cross hatched area for $x_c = x_1$

As $x_c \uparrow$, $W \downarrow$

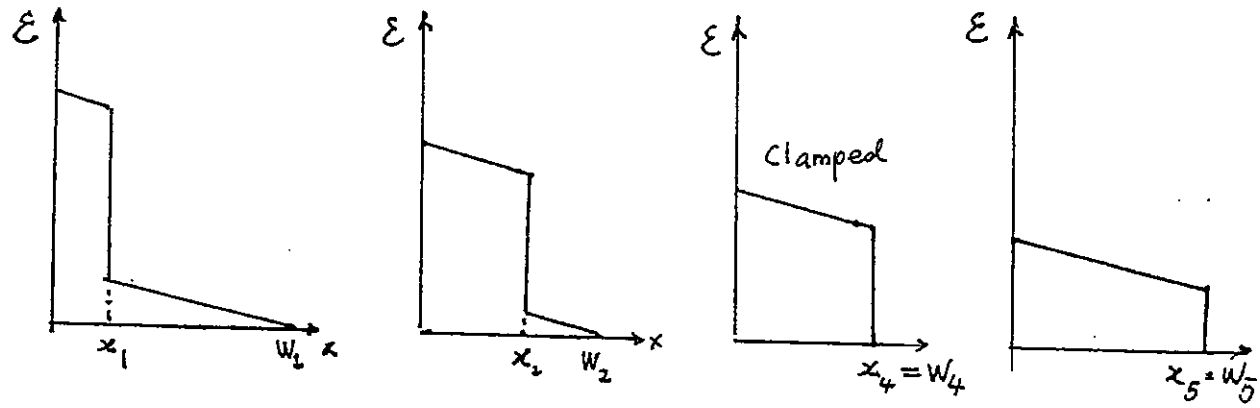
eventually $x_c = x_4 = W_4$ the W_{DM} becomes clamped to and increases with the implant centroid x_c.



Threshold voltage shift versus centroid for different ion doses.

- Subthreshold swing.

$$S \sim (\ln 10)\left(\frac{kT}{q}\right)\left(\frac{C_D + C_{ox}}{C_{ox}}\right)$$

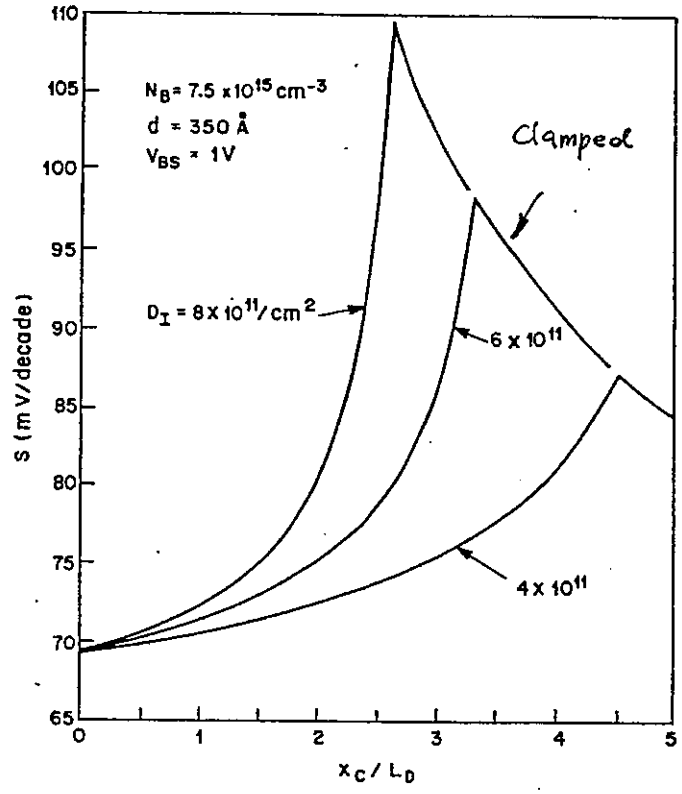


$W_2 < W_1$
$C_D \uparrow$
$S \uparrow$

$W_4 < W_2$
$C_D \uparrow$
$S \uparrow$

After clamp
$W_5 > W_4$
$C_D \downarrow$
$S \downarrow$

Therefore S increases with x_c,
once clamped $x_c = W_{Dm}$ as W_{Dm} increases with x_c, the depletion capacitance C_D will decrease, and S decreases

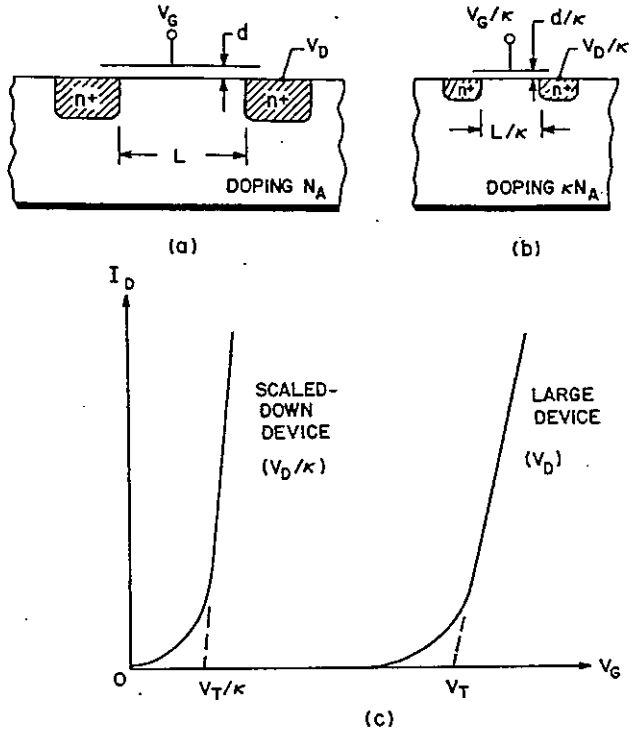


Subthreshold swing versus centroid for different ion doses.

6.4 Device Scaling and Short-Channel Effects

6.4.1 Device Scaling

- Constant Field Scaling (see Table 2. p 329)



Scaling approach for device miniaturization. (a) Long-channel device. (b) Scaled device. (c) Drain characteristics of these devices.

$$L' = L/\kappa \qquad (\kappa > 1)$$

$$d' = d/\kappa$$

$$\mathcal{E}' = \mathcal{E}$$

$$V' = \mathcal{E}'L' = \mathcal{E}L/\kappa = V/\kappa$$

$$C_{ox} = \frac{\epsilon_{ox}}{d'} = \frac{\epsilon_{ox}}{d/\kappa} = \kappa C_{ox}$$

$$W_{Dm}' = \frac{W_D}{\kappa} = \sqrt{\frac{2\epsilon_s}{qN_A}(V_{bi} + V)}/\kappa$$

$$\sim \sqrt{V/N_A}/\kappa = \sqrt{V/\kappa N_A \kappa} = \sqrt{V'/N_A'}$$

$$\therefore \; N_A' = N_A \kappa$$

Therefore

$$\text{Density}' = (\text{Density})\,K^2$$

$$\tau' = L'/v = \tau/K$$

$$\text{dc Power}' = \left(\frac{I}{K}\right)\left(\frac{V}{K}\right) = P_{dc}/K^2$$

$$\text{ac power}' = (C_{ox}' A')\,V'^2 f_T' = \frac{C_{ox}A}{K}\left(\frac{V}{K}\right)^2 K f = P_{ac}/K^2$$

$$\text{Energy}' = \frac{1}{2}(C_{ox}A)'\,V'^2 = \frac{1}{2}\frac{C_{ox}A}{K}\left(\frac{V}{K}\right)^2 = \frac{E}{K^3}$$

- Constant Voltage Scaling

$$L' = L/K$$

$$V' = V$$

$$\mathcal{E}' = V/L = K\mathcal{E}$$

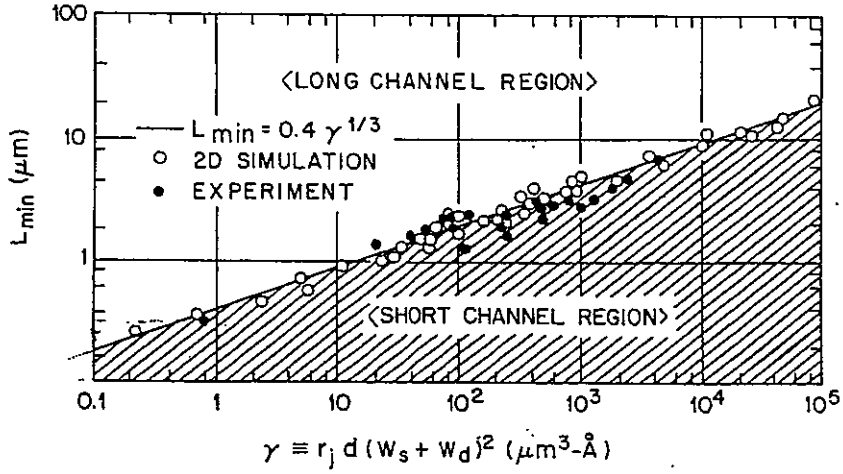
$$C_{ox}' = C_{ox}K$$

$$V_T' = V_T \simeq 2\psi_B + \frac{\sqrt{2\epsilon_s q N_A (2\psi_B)}}{C_{ox}} \sim \frac{\sqrt{N_A}}{C_{ox}} = \frac{\sqrt{N_A}}{C_{ox}'/K}$$

$$= \frac{\sqrt{N_A K^2}}{C_{ox}'} = \frac{\sqrt{N_A'}}{C_{ox}'}$$

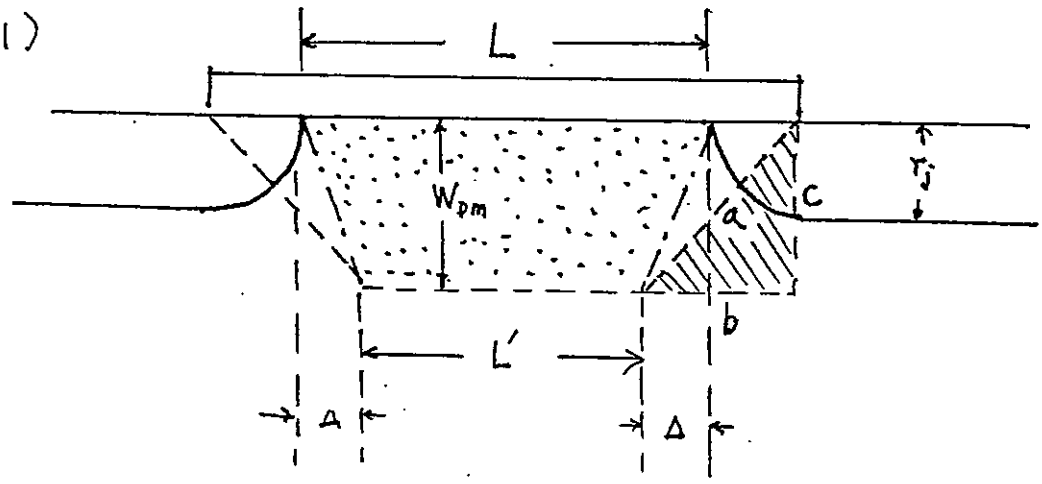
$$\therefore \boxed{N_A' = K^2 N_A}$$

- Flexible Scaling



Minimum channel length versus $r_j \cdot d(W_S + W_D)^2$.

(p.331)



Refer to the triangle at right

$$a \cong r_j + W_{Dm}$$

$$b = r_j + \Delta$$

$$c = W_{Dm}$$

$$\therefore (r_j + W_{Dm})^2 = (r_j + \Delta)^2 + W_{Dm}^2$$

$$r_j^2 + 2r_j W_{Dm} + W_{Dm}^2 = r_j^2 + 2r_j\Delta + \Delta^2 + W_{Dm}^2$$

$$\Delta^2 + 2r_j\Delta - 2r_j W_{Dm} = 0$$

$$\therefore \Delta = -r_j + \sqrt{r_j^2 + 2r_j W_{Dm}}$$

$$L' = L - 2\Delta$$

$$\frac{L+L'}{2L} = \frac{2L-2\Delta}{2L} = 1 - \frac{\Delta}{L} = 1 - \frac{r_j}{L}\left(\sqrt{1 + \frac{2W_{Dm}}{r_j}} - 1\right)$$

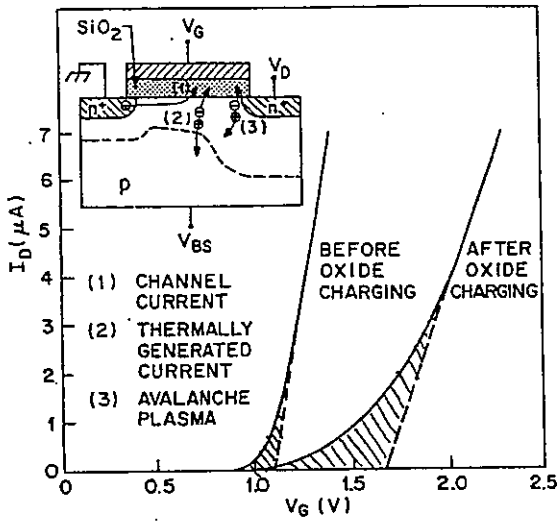
$$\Delta V_T = \frac{1}{C_{ox}}\left(\frac{Q_B'}{2L} - qN_A W_{Dm}\right) = \frac{qN_A W_{Dm}}{C_{ox}}\left(\frac{L+L'}{2L} - 1\right)$$

$$= -\frac{qN_A W_{Dm} r_j}{C_{ox} L}\left(\sqrt{1 + \frac{2W_{Dm}}{r_j}} - 1\right) \quad \text{------ (99)}$$

As L becomes shorter, $|\Delta V_T|$ increases. The transistor has lower V_T and it is easier to turn on <see Fig. 27 on p.333>

6.4.5 Multiplication and Oxide Reliability

- Oxide charging



V_T : 1.1 V → 1.7 V

$g_m \left(=\frac{dI_D}{dV_G}\right)$: reduced

Subthreshold current (= Area under curve): increased

The oxide charging causes threshold voltage shift, reduction of transconductance (due to reduced mobility), and higher subthreshold current (due to increased interface trap density).

6.5 MOSFET Structures

6.5.1 Gate Stack :

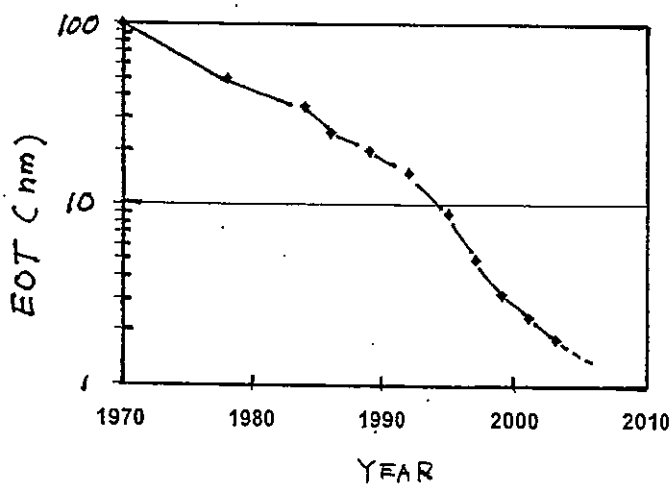
$$C_i = \frac{\epsilon_i}{d_i} = \frac{\epsilon_{ox}}{d_i\left(\frac{\epsilon_{ox}}{\epsilon_i}\right)} = \frac{\epsilon_{ox}}{EOT}.$$

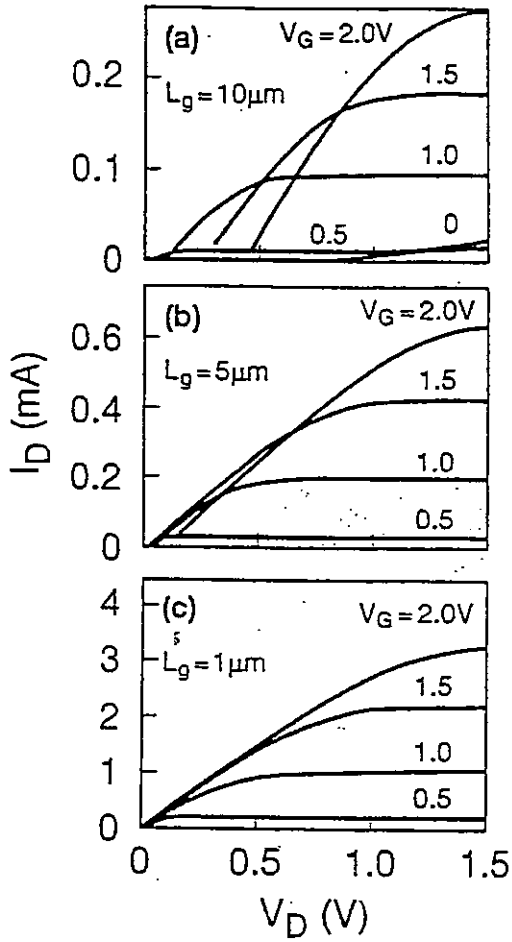
∴ EOT = effective oxide thickness $= d_i\left(\frac{\epsilon_{ox}}{\epsilon_i}\right)$

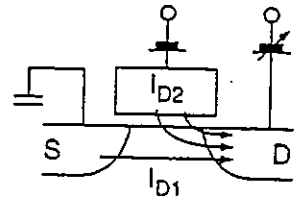
Assume $K \equiv \left(\frac{\epsilon_i}{\epsilon_0}\right) = 3.9$, ϵ_0 = Permittivity in Vacuum.

$K(SiO_2) = 3.9$

An insulator with a thickness d_i of 20 nm has an EOT of 2 nm $\left(20 \times \frac{3.9}{39}\right)$.







$I_{D1} \sim \frac{1}{L}$ channel current

$I_{D2} \sim L$ tunnel current

$\therefore \frac{I_{D2}}{I_{D1}} \sim L^2$

If L is reduced by 10 times, I_{D2} will be lowered by 100 times. and becomes insignificant.

6.5.4 SOI and TFT

Structure	Conventional	Gate-All-Around	Ground Plane
Schematic	N_A L_{eff} t_{ox} t_{Si} t_{box}	n+ t_{ox} t_{Si} t_{ox}	t_{ox} t_{Si} Ground
Scaling	$\lambda = \sqrt{\frac{\varepsilon_{Si}}{\varepsilon_{ox}} t_{Si} t_{ox}}$	$\lambda = \sqrt{\frac{\varepsilon_{Si}}{2\varepsilon_{ox}} t_{Si} t_{ox}}$	$\lambda = \sqrt{\frac{\varepsilon_{Si}}{2\varepsilon_{ox}} \frac{t_{Si} t_{ox}}{(1 + \frac{\varepsilon_{Si} t_{ox}}{\varepsilon_{ox} t_{Si}})}}$
	(a)	(b)	(c)

Scaling relations for the conventional SOI MOSFET, the gate-all-around configurations and the ground-plane structure.

λ is a natural length scale. To minimize short-channel effect, the ratio L/λ should be larger than 5. (see classnotes on "MOSFET and Related Devices", 2nd paper.)

6.6 Circuit Applications

6.6.1 Microwave Performance

The cutoff frequency is obtained for

$$i_{out}/i_{in} = 1$$

$$|i_{in}| = \omega C_G' \tilde{v}_g = \omega(C_{GS}' + C_{GD}')\tilde{v}_g \quad \text{assuming no parasitics}$$

$$i_{out} = g_m \tilde{v}_g$$

$$\therefore \quad \omega(C_{GS}' + C_{GD}')\tilde{v}_g = g_m \tilde{v}_g$$

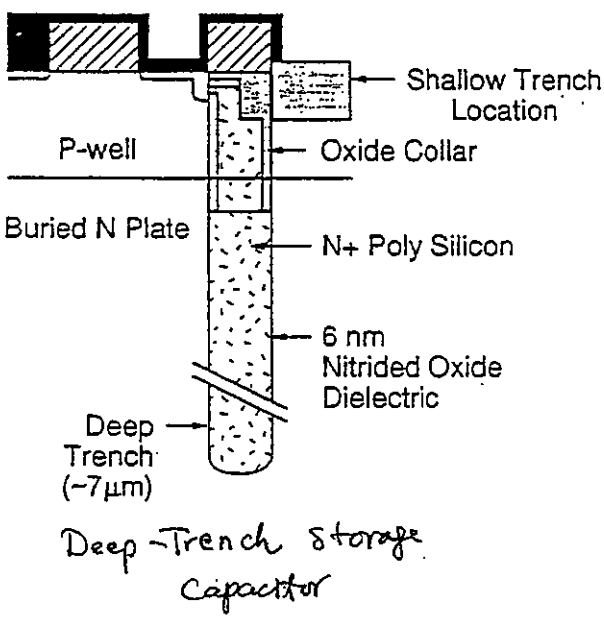
$$f_T = \frac{\omega}{2\pi} = \frac{g_m}{2\pi(C_{GS}' + C_{GD}')}$$

For the ideal case $(C_{GS}' + C_{GD}') = C_{ox} ZL$, under velocity-saturation condition $g_m = Z C_{ox} v_s$

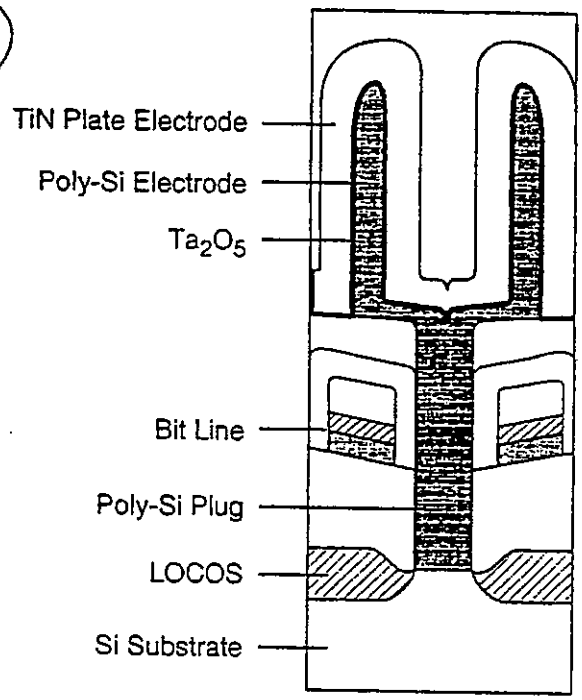
$$\therefore f_T = \frac{Z C_{ox} v_s}{2\pi(C_{ox} ZL)} = \frac{v_s}{2\pi L} = \frac{1}{2\pi \tau_t}$$

where τ_t is the transit time of carriers from source to drain.

6.6.2 DRAM $\left(C = \frac{A\epsilon_{ox}}{d}\right)$



Deep-Trench storage capacitor



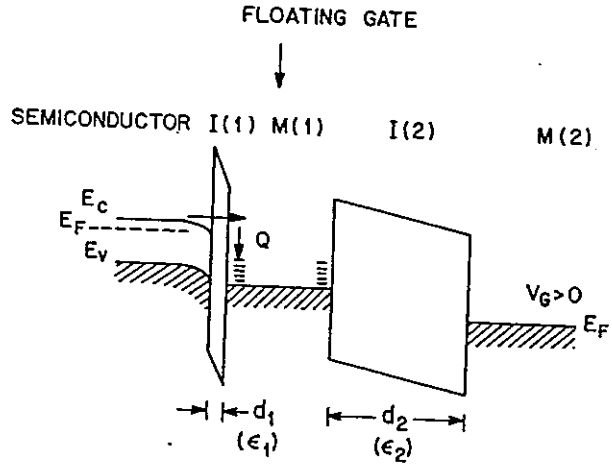
Crown-type storage capacitor

DRAM Capacitor is given by

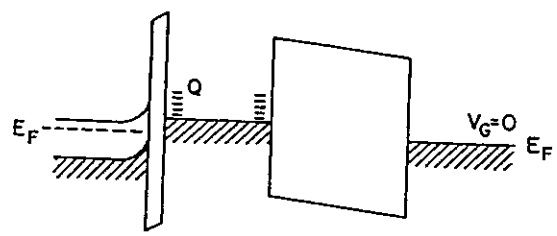
$$C = \epsilon_i A / d_i$$

For a minimum d_i, only two parameters are adjustable: ϵ_i and A. We can use Ta_2O_5, $SrTiO_3$ etc to increase ϵ_i, and use trench or crown structure to increase A.

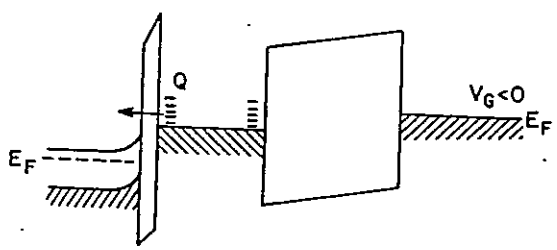
6.7 Nonvolatile Memory Devices



(a) Programming Mode



(b) Storage Mode



(c) Erasing Mode

From Gauss Law

$$\epsilon_1 \mathcal{E}_1 = \epsilon_2 \mathcal{E}_2 + Q$$

$$V_G = V_1 + V_2 = d_1 \mathcal{E}_1 + d_2 \mathcal{E}_2 \quad \text{--- (115)}$$

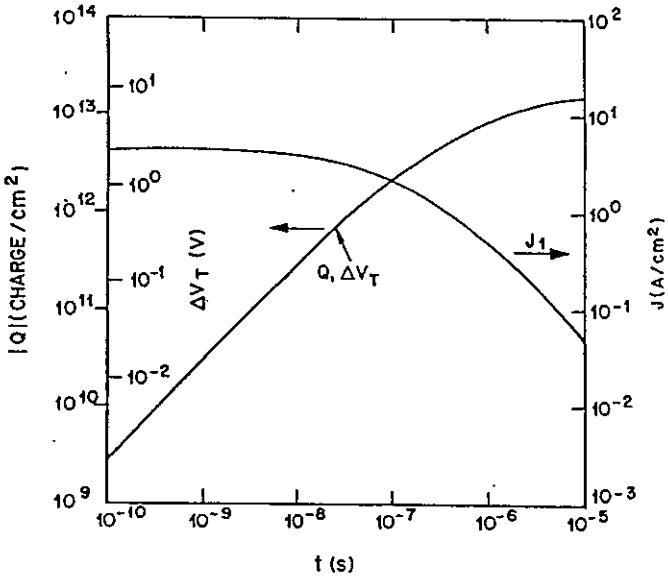
$$\therefore \epsilon_1 \mathcal{E}_1 = \epsilon_2 \left(\frac{V_G - d_1 \mathcal{E}_1}{d_2} \right) + Q = \frac{\epsilon_2 V_G}{d_2} - \frac{d_1}{d_2} \epsilon_2 \mathcal{E}_1 + Q$$

$$\therefore \mathcal{E}_1 \left(\epsilon_1 + \frac{d_1}{d_2} \epsilon_2 \right) = \frac{\epsilon_2}{d_2} V_G + Q$$

$$\mathcal{E}_1 = \frac{\frac{\epsilon_2}{d_2} V_G + Q}{\epsilon_1 + \frac{d_1}{d_2} \epsilon_2} = \frac{V_G}{d_1 + \left(\frac{\epsilon_1}{\epsilon_2}\right) d_2} - \frac{|Q|}{\epsilon_1 + \epsilon_2 \left(\frac{d_1}{d_2}\right)} \quad \cdots (116)$$

$$\Delta V_T = - \frac{d_2 Q}{\epsilon_2} \quad \cdots\cdots (118)$$

The charge in the floating gate will cause a threshold voltage shift, and the charge is like a fixed oxide charge at the interface of M(1) and I(2).



Calculated charging current and stored charge as a function of charging time.

$$J_1 = C_4 \mathcal{E}_1^2 \exp\left(-\frac{\mathcal{E}_0}{\mathcal{E}_1}\right)$$

$$Q = \int_0^t J_1 \, dt$$

When Q is small, $\mathcal{E}_1$ is given by the first term of Eq. 116. As |Q| increases $\mathcal{E}_1$ will decrease, causing a reduction of J_1. For sufficiently long time, J_1 drops to zero, both Q and ΔT approach steady values. For example, at $t = 10^{-7}$ s, $J_1 = 2 A/cm^2$, $Q = 2 \times 10^{12}/cm^2$, $\Delta V_T = 1 V$.

• Retension Time (p. 356 - 357)

$$Q(t) = Q_0 e^{-t/t_0} \qquad t_0 = \frac{1}{\nu} \exp(q\phi_B/kT)$$

$$\frac{Q(t)}{Q_0} = \frac{1}{2} = e^{-\tau/t_0}$$

$$\therefore \tau = (\ln 2) t_0 = \frac{\ln 2}{\nu} \exp(q\phi_B/kT).$$

As $T \uparrow$ $\tau \downarrow$

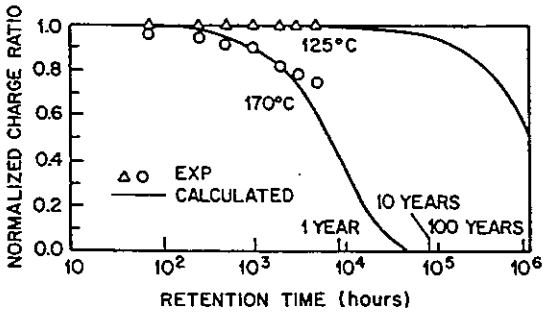
$q\phi_B \approx 1.7$ eV

at 125°C

$\tau = 100$ years

at 170°C

$\tau = 1$ year



Normalized charge storage versus time for two ambient temperatures.

6.8 Single-Electron Transistor

• The minimum energy of a single electron is

$$E = \frac{CV^2}{2} = \frac{C}{2}\left(\frac{q}{C}\right)^2 = \frac{q^2}{2C}$$

Let $E = 1$ eV, a capacitor has an oxide layer of 2nm and an area of $l \times l$, then

$$C = \frac{q^2}{2E} = \frac{q}{2} = \frac{1.6\times10^{-19}}{2} = 3.45\times10^{-13} \times \frac{l^2}{2\times10^{-7}}$$

$$\therefore l \doteq 2.1\times10^{-7}\ \text{cm} = 2.1\ \text{nm}.$$

• Tunneling resistances

power of a resistor : $P = I^2R$

energy of a resistor : $\Delta E = I^2R\,\Delta t = \left(\frac{q}{\Delta t}\right)^2 R\,\Delta t$

$= q^2R/\Delta t$

Uncertainty Principle:

$$\Delta E \, \Delta t > \frac{\hbar}{2}$$

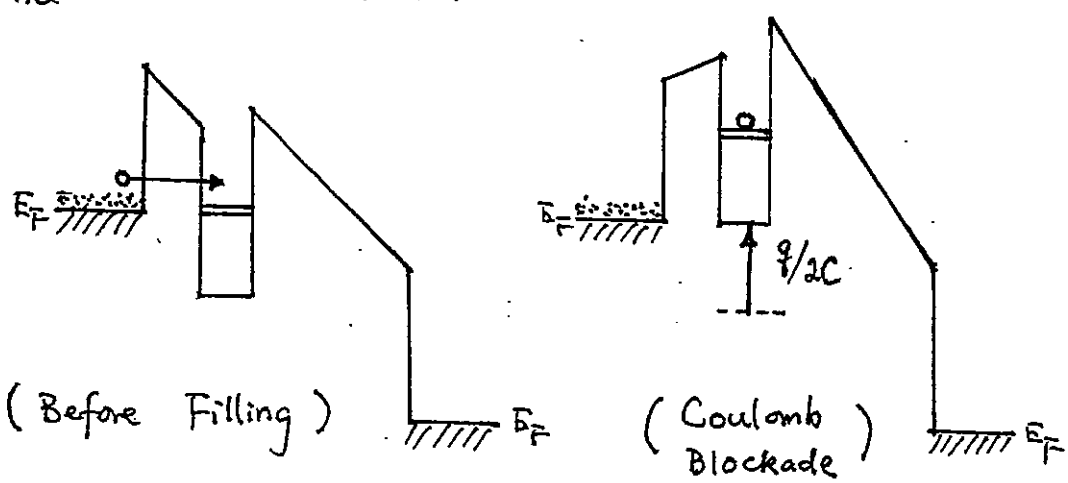
$$\therefore \quad q^2 R > \frac{\hbar}{4\pi}$$

$$R > \frac{\hbar}{q^2 4\pi} = \frac{6.62 \times 10^{-34}}{(1.6 \times 10^{-19})^2 4\pi} = 2.05 \text{ k}\Omega$$

$$\therefore \quad R > 100 \text{ k}\Omega$$

- Single-electron memory cell (SEMC) is a limiting case of the floating-gate structure. By reducing the length of the floating gate to nm regime, we obtain the SEMC. Because of its small size, the capacitance is very small ($\sim 10^{-19}$ F).

When an electron tunnels into the quantum well (the floating gate), the potential on the left side is reduced and the transfer of another electron is blocked — the Coulomb Blockade:



(Before Filling) (Coulomb Blockade)

SEMC is an ultimate floating-gate memory cell. Note that the above band diagrams are similar to those shown in the figure on p. 78 (classnotes).

A Floating Gate and Its Application to Memory Devices

By D. KAHNG and S. M. SZE

(Manuscript received May 16, 1967)

A structure has been proposed and fabricated in which semi-permanent charge storage is possible. A floating gate is placed a small distance from an electron source. When an appropriately high field is applied through an outer gate, the floating gate charges up. The charges are stored even after the removal of the charging field due to much lower back transport probability. Stored-charge density of the order of $10^{12}/cm^2$ has been achieved and detected by a structure similar to an metal-insulator-semiconductor (MIS) field effect transistor. Such a device functions as a bistable memory with nondestructive read-out features. The memory holding time observed was longer than one hour. These preliminary results are in fair agreement with a simple analysis.

It has been recognized for some time that a field-effect device, such as that described by Shockley and Pearson,[1] can be made bistable utilizing switchable permanent displacement charges on ferroelectric material.[2] Subsequent studies of ferroelectric material have revealed,[3] however, that the inherent speed capability of a device incorporating a ferroelectric material is limited by domain motion, whose highest speed is limited by the acoustic velocity. In the absence of highly ordered, near-ideal thin film ferroelectric material, the speed capability of a bistable device, therefore, is in the microsecond range at best.[4] In addition, many ferroelectric materials suffer from irreversible mechanical disorder after many cycles of polarization switching,[2] rendering some uncertainty on the long term device reliability aspect.

An alternative to a ferroelectric gate is a floating gate chargeable by field emission, which hopefully circumvents the above mentioned difficulties. Consider a sandwich structure, metal $M(1)$, insulator $I(1)$, metal $M(2)$, insulator $I(2)$, and finally metal $M(3)$. (See Fig. 1). If the thickness of $I(1)$ is small enough so that a field-controlled electron transport mechanism such as tunneling or internal tunnel-hopping are possible, a positive bias on $M(3)$ with respect to $M(1)$ with $M(2)$ floating [$M(2)$ is called the floating gate henceforth], would cause electron accumulation in the floating gate, provided electron transport

* Bell System Tech. J. 46, 1288 (1967).
(First paper on nonvolatile semiconductor memory)

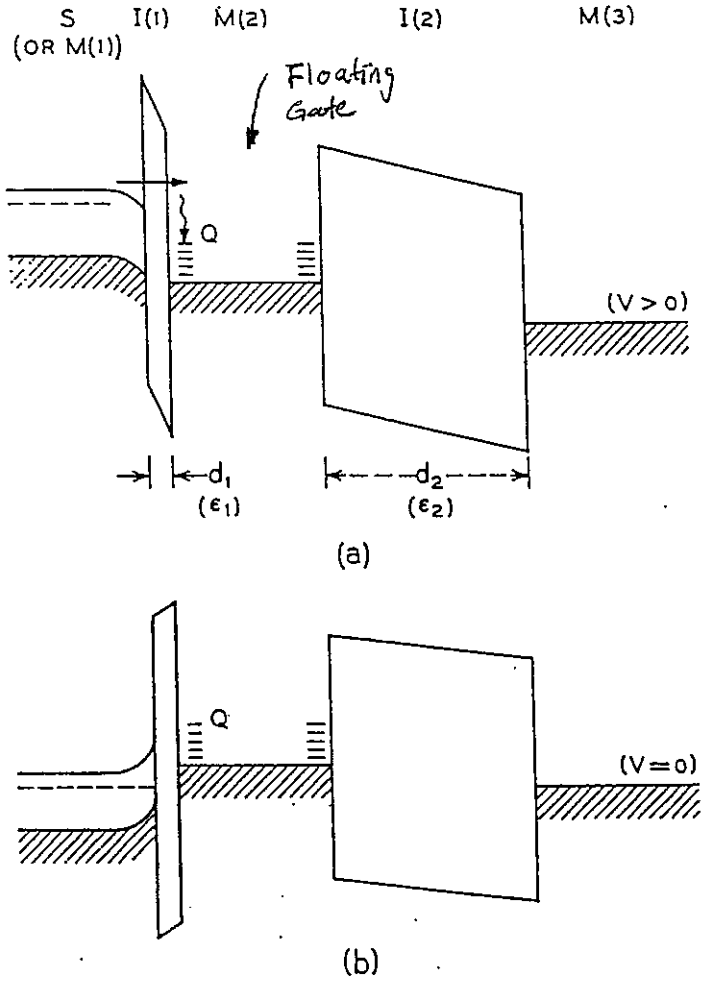


Fig. 1 — Energy band diagram of a floating gate structure with a semiconductor-insulator-metal-insulator-metal sandwich. For calculation of the stored charge, the semiconductor is replaced by a metal $M(1)$. (a) When a positive voltage step is applied to the outer gate. (b) When the voltage is removed. The stored charge Q causes an inversion of the semiconductor surface.

across $I(2)$ is small. These conditions can be met by choosing $I(1)$ and $I(2)$ such that the ratio of dielectric permittivity ϵ_1/ϵ_2 is small and/or the barrier height into $I(1)$ is smaller than that into $I(2)$. The sandwich structure is somewhat similar to the tunnel emitter metal-base transistor proposed by Mead[5] in its structure but with the following essential differences.

(*i*) $M(2)$ is much thicker than the hot electron range, so that emitted electrons are close to the Fermi-level of $M(2)$ before reaching $I(2)$.
(*ii*) No carrier transport is allowed across $I(2)$.
(*iii*) $M(2)$ is floating.

The stored charge Q, as a function of time when a step voltage function with amplitude V is applied across the sandwich, is given by

$$Q(t) = \int_0^t j\,dt' \qquad \text{coul/cm}^2. \tag{1}$$

When the emission is of Fowler-Nordheim tunneling type, then the current density, j, has the form

$$j = C_1E^2 \exp(-E_0/E), \tag{2a}$$

where C_1 and E_0 are constants in terms of effective mass and the barrier height. (We have neglected the effects due to the image force lowering[6] of the barrier, etc., but the essential feature is expected to be retained even after detailed corrections are made). This type of current transport occurs in SiO_2 and Al_2O_3.

When the field emission is of the internal Schottky or Frankel-Poole type, as occurs in Si_3N_4,[7] then j follows the form

$$j = C_2E \exp[-q(\Phi_1 - \sqrt{qE/\pi\epsilon_1})/kT], \tag{2b}$$

where c_2 is a constant in terms of trapping density in the insulator, Φ_1 the barrier height in volts, ϵ_1 the dynamic permittivity.

The electric field in $I(1)$ at all times is a function of the applied voltage V and $Q(t)$, and is obtainable from the displacement continuity requirement as

$$E = \frac{V}{d_1 + d_2(\epsilon_1/\epsilon_2)} - \frac{Q}{\epsilon_1 + \epsilon_2(d_1/d_2)}, \tag{3}$$

where d_1 and d_2 are the thickness of $I(1)$ and $I(2)$, respectively.

Fig. 2(a) shows the results of a theoretical computation using (1), (2a), and (3) with the following parameters: $d_1 = 50$ Å, $\epsilon_1 = 3.8\,\epsilon_0$ (for SiO_2), $d_2 = 1000$ Å, $\epsilon_2 = 30\,\epsilon_0$ (for ZrO_2), and V = 50 volts. One notes that the stored charge initially increases linearly with time and then saturates. The current is almost constant for a short time and then decreases rapidly. The field in $I(1)$ decreases slightly as the time increases. The above results can be explained as follows: When a voltage pulse is applied at $t = 0$, the initial charge Q is zero, and the initial electric field across $I(1)$ has its maximum value, $E_{max} = V/[d_1 + (\epsilon_1/\epsilon_2)d_2]$. As t increases, Q will first increase linearly with time. This is because of the fact that for small Q such that E remains essentially the same, the current will in turn remain the same, so $Q = j(E_{max}) \cdot t$. Eventually, when Q is large enough to reduce the

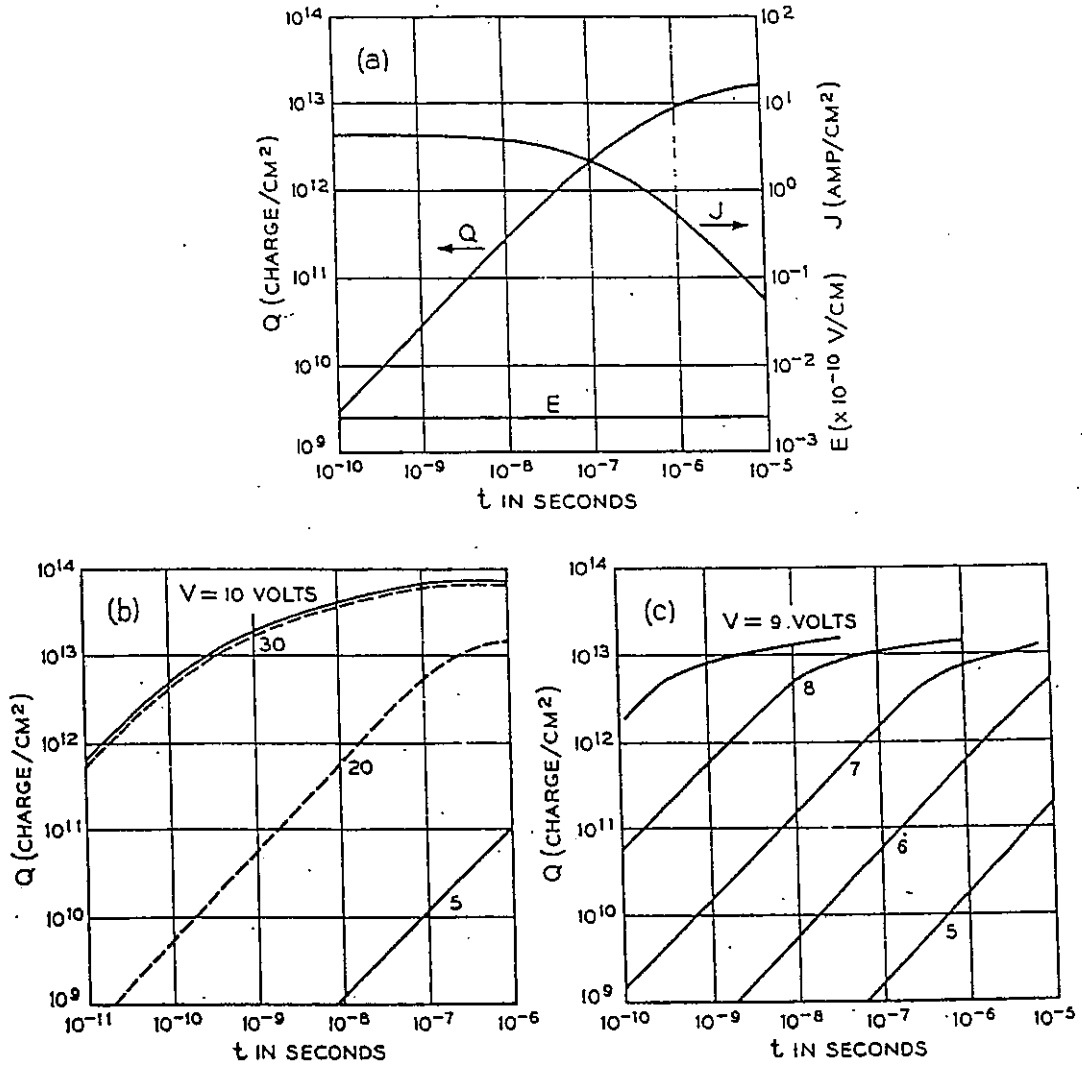


Fig. 2(a) — Theoretical results of the stored-charge density (Q), the current density (J), and the electric field across $I(1)$ as a function of time. $V = 50$ volts, $d_1 = 50$ Å, $\epsilon_1/\epsilon_0 = 3.8$ (for SiO_2), $d_2 = 1000$ Å, $\epsilon_2/\epsilon_0 = 30$ (for ZrO_2). (b) Theoretical results of the stored charge density as a function of time with the same ϵ_1 and ϵ_2 as in (a), and $d_1 = 10$ Å, $d_2 = 100$ Å (solid lines), $d_1 = 30$ Å, $d_2 = 300$ Å (dotted lines). (c) Theoretical results of the stored density as a function of time with $d_1 = 20$ Å, $\epsilon_1/\epsilon_0 = 60$ (for Si_3N_4), $d_2 = 200$ Å, $\epsilon_2/\epsilon_0 = 30$ (for ZrO_2) and various applied voltages.

value of E substantially, then the current will decrease rapidly with time and Q increases slowly.

Fig. 2(b) shows the stored charge as a function of time for the time ϵ_1 and ϵ_2 but different d_1, d_2, and V. It is clear that for a given structure, in order to store a given amount of charge, one can either increase the applied voltage or increase the charging time (pulse width) or both. Fig. 2(c) shows the calculated stored charge for the current transport described by (2b). Here $I(1)$ is a 20 Å thick Si_3N_4 film. There are

marked decreases in the gate voltages required for a given charge compared to Al_2O_3, SiO_2. This is largely due to the much lower barrier height (1.3 volts)[7] compared to SiO_2 ($\simeq$ 4.0 volts).[8]

It is noted that the field in $I(1)$ for appreciable charge storage is in the 10^7 V/cm range. When the outer gate voltage is removed, the field in $I(1)$ due to the stored charge on the inner gate is only 10^6 V/cm or so corresponding to 5×10^{12} charges/cm^2, a large enough charge to detect easily. Since the transport across $I(1)$ is highly sensitive to the field, (2a) and (2b), no charges flow back. The charge loss is actually controlled by the dielectric relaxation time of the sandwich structure,[9] which is very long. When it is desired to discharge the floating gate quickly, it is necessary to apply to the outer gate a voltage about equal in magnitude but opposite in polarity to the voltage which was used for charging. It is evident that net positive charges (loss of electrons) can also be stored in the floating gate if the discharging gate voltages are appropriately chosen in magnitude and duration.

It was mentioned that the stored-charge density of 5×10^{12}/cm^2 was sufficient for easy detection. One of the detection or read-out schemes is to use the surface field effect transistor (MOSFET or IGFET) first fabricated and described by Kahng and Atalla[10] in 1960. For inversion at a silicon surface, the charge required is only about 2×10^{11}/cm^2 for 1 ohm-cm n-type silicon. However, surface-state charges at the silicon-silicon dioxide interface may be as high as 10^{12}/cm^2, depending on the fabrication techniques used. For this reason we have chosen 5×10^{12}/cm^2 as the stored charge required for easy detection. When the Insulated Gate Field Effect Transistor (IGFET) principle is used for read-out, $M(1)$ is now replaced by silicon. This requires a slight correction in the calculation of charge flow through the insulator, but the major features of the results are not expected to be altered significantly. It is to be noted that about one half of the stored charge can be active in creation of the inversion layer since the other half resides near the $M(2)$-$I(2)$ interface due to Colombic repulsion.

To check the feasibility, a floating gate device has been fabricated using an IGFET as shown in Fig. 3(a). The substrate is an n-type silicon, 1 ohm-cm, and (111) oriented. $I(1)$ is a 50 Å SiO_2 thermally grown in a dry oxygen furnace. $M(2)$ and $I(2)$ are Zr (1000 Å) and ZrO_2 (1000 Å), respectively. $M(3)$ and the ohmic contact metals are aluminum deposited in a vacuum system. Fig. 3(b) is another version of the floating gate device using a thin film transistor (TFT) structure.[11]

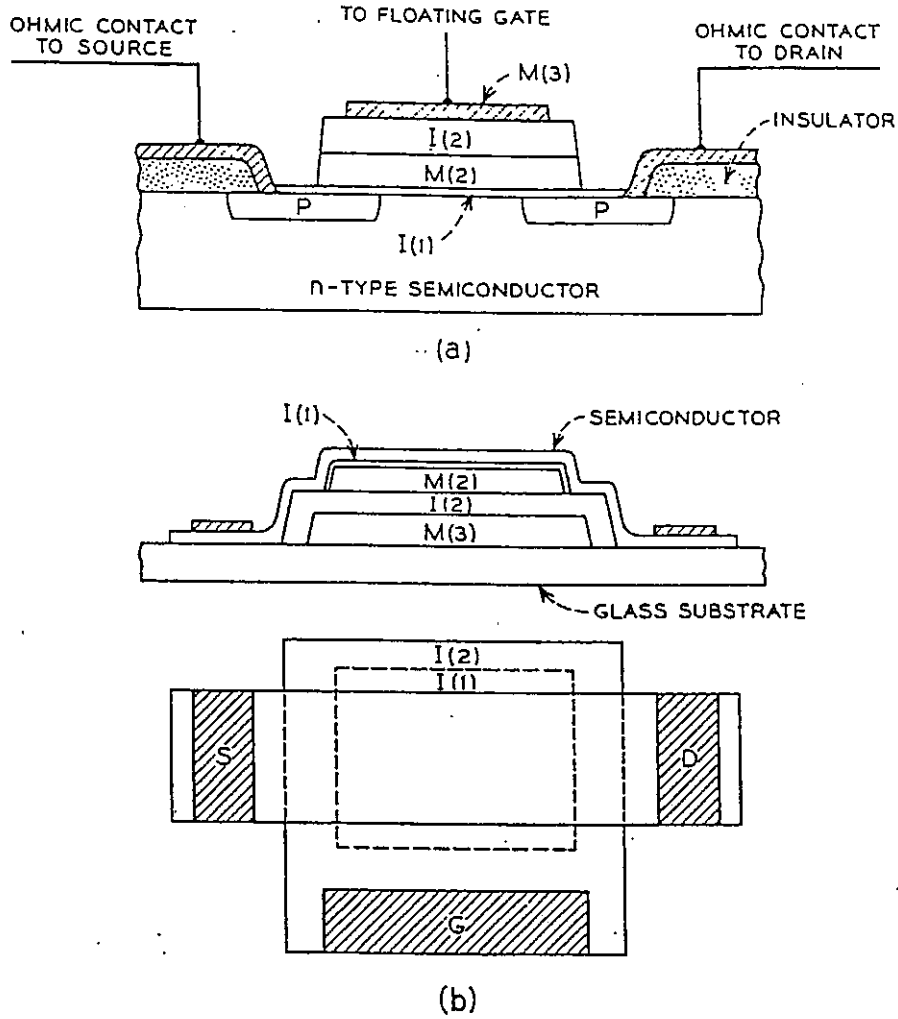
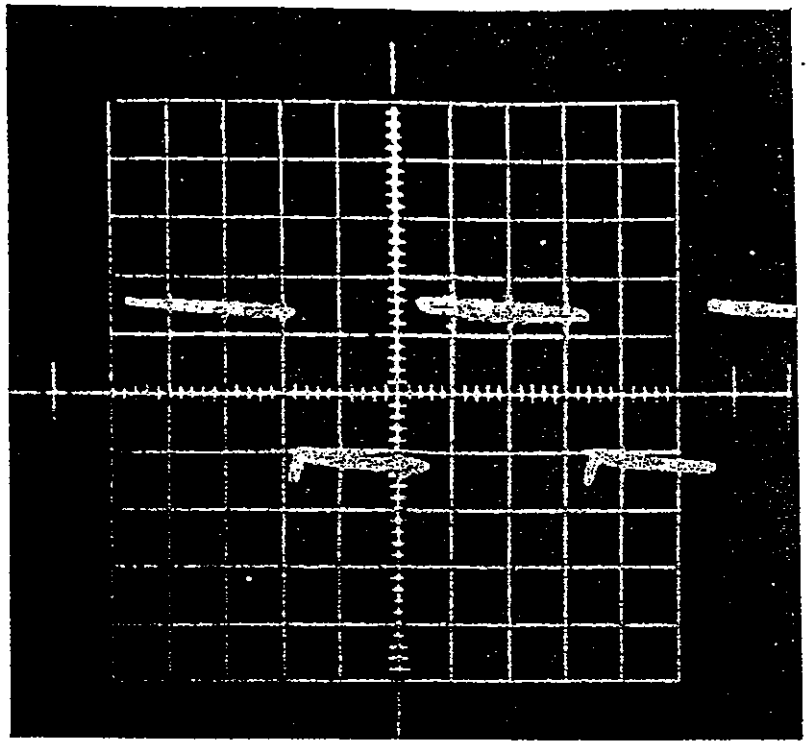
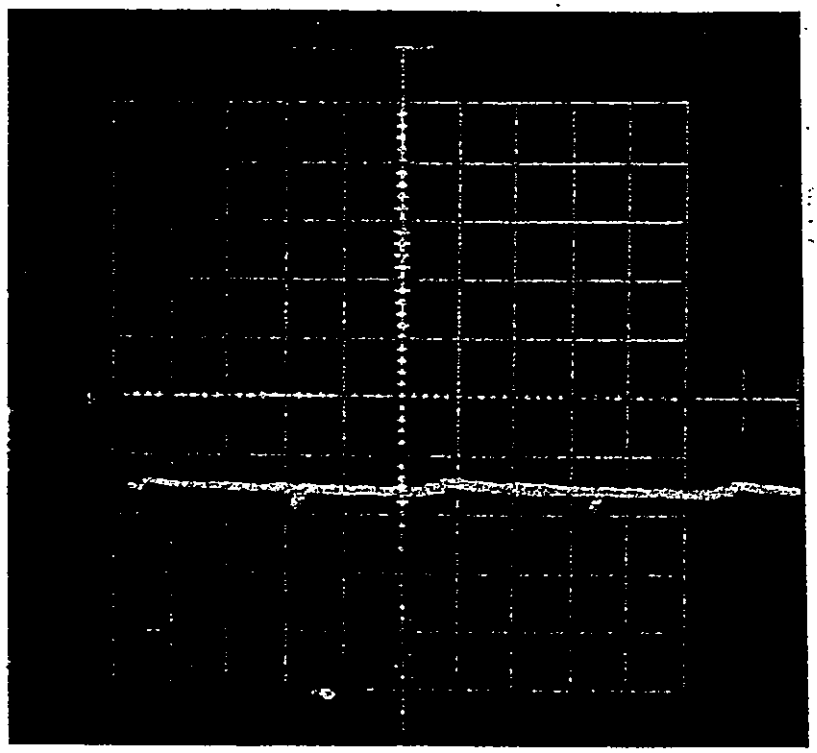


Fig. 3 — (a) Schematic diagram of a floating gate device using an IGFET. The numbers indicated correspond to those shown in Fig. 1. (b) Schematic diagram of a floating gate device using thin film transistor structure.

The IGFET-type floating gate devices have been tested in a pulsing circuit. Because of the relatively thick insulator layers, large voltages ($\approx$ 50 V) and long pulse width ($\approx$ 0.5 μs) have to be applied in order to store $\approx 5 \times 10^{12}$ charges/cm^2. Fig. 4 shows the experimental results. A positive pulse of 50 volts is first applied to the gate electrode, and 60 ms later a negative pulse of 50 volts is applied. Then the pulsing cycle repeats. In Fig. 4(a) the pulse widths are 0.5 μs. One notes that when the positive pulse is applied, a sufficient amount of charge is stored in the floating gate so that the silicon surface is inverted; a conducting channel is thus formed, and the channel current is "on." It can be seen that the channel current decreases only slightly at the end of 60 ms. When the negative pulse is applied, the stored charge is eliminated, and also the channel. The channel current reduces to its



(a)



(b)

Fig. 4 — Experimental results of the channel current of a IGFET-type floating gate device. A positive voltage pulse, V_1, with pulse width W_1, is first applied to the gate, and 60 ms later a negative pulse V_2 with pulse width W_2 is applied. Then the pulsing cycle repeats. Horizontal scale: 20 ms/div. Vertical scale: 0.1 ma/div. (a) $V_1 = V_2 = 50$ volts, $W_1 = W_2 = 0.5$ μs. (b) $V_1 = V_2 = 40$ volts, $W_1 = W_2 = 0.5$ μs.

"off" state. Fig. 4(b) shows results for pulses with the same widths but smaller amplitude (40 V). Since the stored charge is a strong function of the pulse amplitude, only a very small amount of charge is stored, too small to cause inversion. For non-leaky units, the memory holding time of longer than one hour has been observed.

It is clear that a modified IGFET such as a TFT can be used for read-out, as shown in Fig. 3(b). For an academic study of device operation, the floating gate can be partially exposed and a potential probe can be placed nearby.

In conclusion, it has been demonstrated that the controlled field emission to the buried "floating" gate may be capacitively induced by pulsing the outer gate electrode. This combination can therefore, be used as a memory device, with holding time as long as the dielectric relaxation time of the gate structure and with continuous nondestructive read-out capability. There seems to be no inherent reason why read-in read-out cannot be performed in a very short time, say in the nanosecond range or even shorter.

We wish to acknowledge helpful discussions and technical assistance rendered by M. P. Lepselter and P. A. Byrnes in connection with ZrO_2, and skillful technical assistance given to us by G. P. Carey and A. Loya, and J. F. Grandner and his group.

REFERENCES

1. Shockley, W., and Pearson, G. L., Phys. Rev., *74*, 1948, p. 232.
2. Ross, I. M., U. S. Patent 2,791,760, issued May 7, 1957.
3. Jona, F. and Shirane, G., *Ferroelectric Crystals*, MacMillan Co., 1962.
4. Heyman, P. M. and Heilmeier, G. H., Proc. IEEE, *54*, 1966, p. 842.
5. Mead, C. A., Proc. IRE, *48*, 1960, p. 359.
6. Simmons, J. G., J. Appl. Phys., *34*, 1963, p. 1793.
7. Sze, S. M., Current Transport and Maximum Dielectric Strength of Silicon Nitride Films, J. Appl. Phys., June, 1967.
8. Williams, R., Phys. Rev., *140*, 1965, A569.
9. Kahng, D., Semipermanent Memory Using Capacitor Charge Storage and IGFET Read-out, B.S.T.J., this issue, pp. 1296–1300.
10. Kahng, D. and Atalla, M. M., Silicon-Silicon Dioxide Field Induced Surface Devices, presented at the IRE-AIEE Device Res. Conf., Carnegie Inst. Tech., Pittsburgh, Pa., June 1960.
11. Weimer, P. K., An Evaporated Thin Film Triode, presented at the IRE-AIEE Device Res. Conf., Stanford University, Stanford, Calif., June 1961.

A NEW FLASH E²PROM CELL USING TRIPLE POLYSILICON TECHNOLOGY

FUJIO MASUOKA, MASAMICHI ASANO, HIROSHI IWAHASHI, TEISUKE KOMURO
and SHINICHI TANAKA

INTEGRATED CIRCUIT DIV. TOSHIBA CORP.
KOMUKAI 1, SAIWAIKU, KAWASAKI 210 JAPAN

ABSTRACT

A new Flash Electrically Erasable-PROM cell with single transistor per bit as same as conventional UV-EPROM(1)(2) and suitable for 256K bit F-E²PROM with rather conservative 2.0μm design rule is described.

The cell is programmed by a channel hot carrier injection mechanism similar to EPROM. The contents of all memory cells are simultaneously erased by using field emission of electrons from a floating gate to an erase gate in a flash. The F-E²PROM cell with single transistor per bit consists of three layers of polysilicon with select transistor.(3)(4)(5)

Programming is 10msec per bit as same as UV-EPROM. Good erasing characteristics is obtained with 550Å of oxide thickness between floating gate and erase gate.

INTRODUCTION

Conventional UV-EPROM is disadvantageous in that a package with window is very expensive as compared with plastic package. One time EPROM with plastic package can not test reliability after packaging. Conventional E²PROM with two transistors per cell is difficult to get high packaging density over UV-EPROM.(6)

This new F-E²PROM overcomes their disadvantages. Using this new cell, the F-E²PROM can get high reliability because of rejecting failed devices of stored charge with burn-in test after plastic packaging.

The chip size of F-E²PROM also is similar to UV-EPROM by using the new small cell and a simple peripheral circuit for erasing.

CELL STRUCTURE

The cell structure is shown in Fig.1-(a)(b)(c). This new memory cell consists of these layers of polysilicon.

The first polysilicon layer is used as an erase gate which is located between the field oxide and the floating gate. The second polysilicon is used as a floating gate similar to an UV-EPROM. The third polysilicon is used as a control gate which is used as a bit select line at programming and reading.

The equivalent circuit of the structure of Fig.1 is shown in Fig.2.

The memory cell area is $8.0 \times 8.0 \mu m^2$ with conservative 2.0μm design rule and the chip size is $5.7 \times 5.8\ mm^2$.

The gate oxide thickness between floating gate and channel region is 500Å. The thickness of interoxide between floating gate and control gate is 700Å. The thickness of interoxide between floating gate and erase gate is 550Å.

The cell is programmed by a channel hot carrier injection mechanism similar to EPROM. The erasure is accomplished with the first level of polysilicon which serves as an erase electrode causing field emission of electrons from the bottom of the floating gate. In F-E²PROM cell, as shown in Fig.1-(c) and Fig.2, one part of channel is controled by the control gate. The third polysilicon layer is used both a gate of the selection transistor and a control gate of the F-E²PROM cell. This selection transistor enables F-E²PROM cell to keep the enhancement mode even if the floating gate transistor becomes depletion mode after erasure.

OPERATION AND RESULT

Programing: The memory is programmed by a channel hot carrier injection mechanism similar to EPROM. The condition which enables this new memory cell to program is shown in Fig.3-(a), as same as the UV-EPROM with double polysilicon gate.

Erasing: Erasing is accomplished by the field emission of electrons from a floating gate to an erase gate. The erase gate of memory cell is supplied with the boosted voltage (VEG) which enables field emission from the floating gate. This boosted voltage(VEG) is produced from the program voltage(VPP). All of the memory cells are erased simultaneously by the erase gate

(First paper on Flash Memory)

CH2099-0/84/0000-0464 $1.00 © 1984 IEDM

which is commonly connected to each other. Reading: The read data depends on the stored electric charges in the floating gate similar to EPROM. The erased memory cell, which is not charged electrons in the floating gate, conducts by applying voltage to the control gate. The programmed memory cell, which is charged electrons in the floating gate, does not conduct by applying voltage, as shown in Fig.4. The virgin memory cell, which is not programmed yet, has same threshold voltage as the erased memory cell, however their gm is different. The threshold voltage of F-E^{2}PROM cell is controled by the selection transistor, however the gm depends on the value of the stored electric charges in the floating gate, even if the memory cell is erased.

Fig.5 shows the experimental results of erasing characteristics. The threshold voltage shift of the memory cell versus erasing voltage applied to the erase gate with oxide thickness between the erase gate and the floating gate as a parameter are measured. In order to measure the negative threshold voltage of the F-E^{2}PROM, the experimental F-E^{2}PROM cells without the selection transistor are fabricated as shown in Fig.5-(b) and (c).

Good erasing characteristics is obtained with 550Å oxide thickness between the erase gate and the floating gate as shown in Fig.5(a).

Fig.6 shows a SEM photograph of the cross sectional view of the F-E^{2}PROM cell.

CONCLUSION

A new 256K bit F-E^{2}PROM with conservative 2.0μm design rule is reported. A simple cell structure caused small cell area that is $8.0 \times 8.0 \mu m^2$. A chip size of $5.7 \times 5.8 mm^2$ 256K F-E^{2}PROM has been developed successfully using the new memory cell. It is very likely to realize 1M bit F-E^{2}PROM by $5.4 \times 5.4 \mu m^2$ cell area with 1.2μm design rule.

ACKNOWLEDGMENT

The authors are grateful for the cooperative effort of H. Minagawa.

The authors would like to thank Mr. K. Suzuki and Mr. M. Ueno for their continuous encouragement throughout this work.

REFERENCES

(1) F. Masuoka, et al "An electrically programmable and erasable 2048-bit stacked-gate MOSROM" in 1973 IEDM Late News 7.8, Dec. 1973.

(2) F. Masuoka, "Field effect semiconductor memory apparatus with a floating gate" USP 3,825,945, Feb. 1972.

(3) F. Masuoka, "Semiconductor memory device"; USP 4,437,172; USP 4,437,174; Dec. 1980.

(4) J. Kupec, et al "Triple level polysilicon E^{2}PROM with single Transistor per bit" 1980 IEDM 23.7 p.602 ∿ 606.

(5) F. Masuoka, "Bit by bit erasable E^{2}PROM with single Transistor per bit" 1981 IEDM 2.1.

(6) Sindey T.K. Nieh, et al "High-voltage Regulation and Process Consideration for High-Density 5V-only E^{2}PROM's" in IEEE Journal of Solid-State Circuits, Vol. SC-18, No. 5, Oct. 1983.

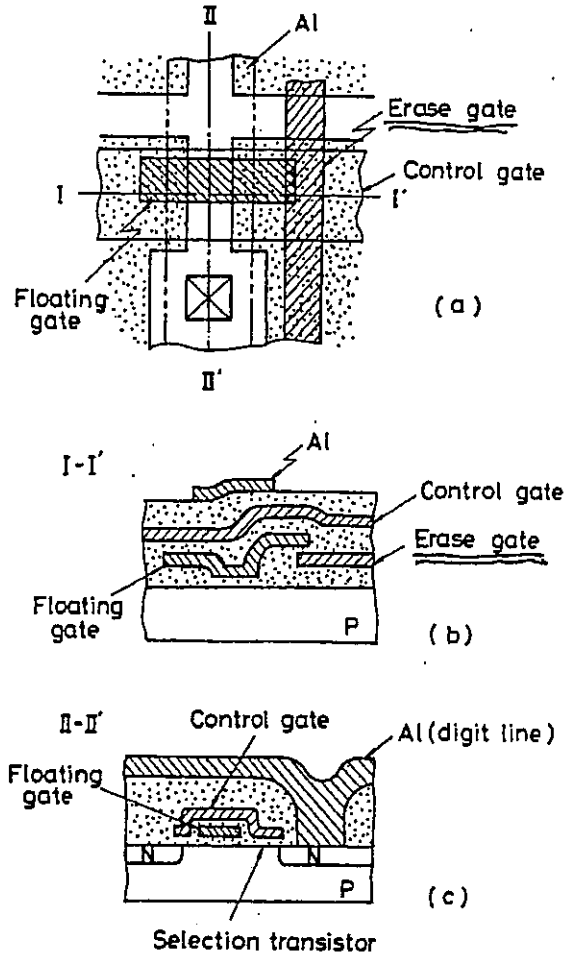


Fig.1 (a) Top view of F-E^{2}PROM cell
(b) Cross sectional view along I-I' line in Fig.1-(a)
(c) Cross sectional view along II-II' line in Fig.1-(a)

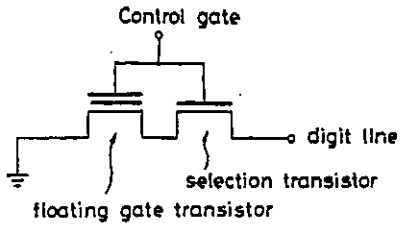


Fig.2 Equivalent circuit of Fig.1

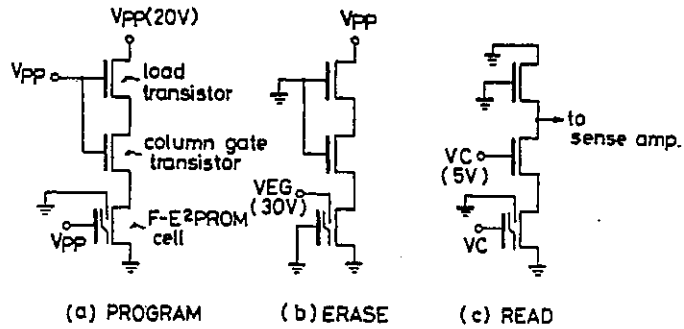


Fig.3 Operating conditions

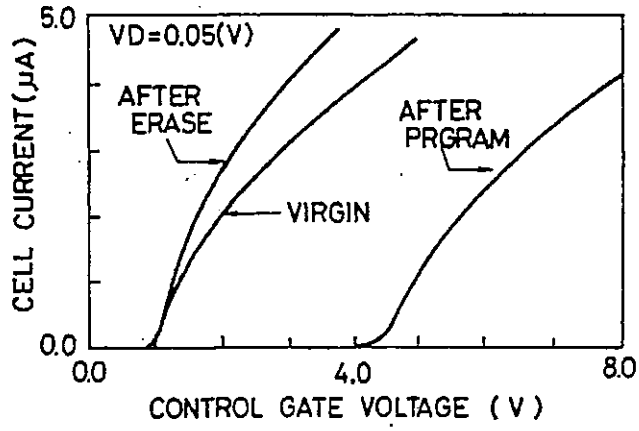


Fig.4 Erase characteristics

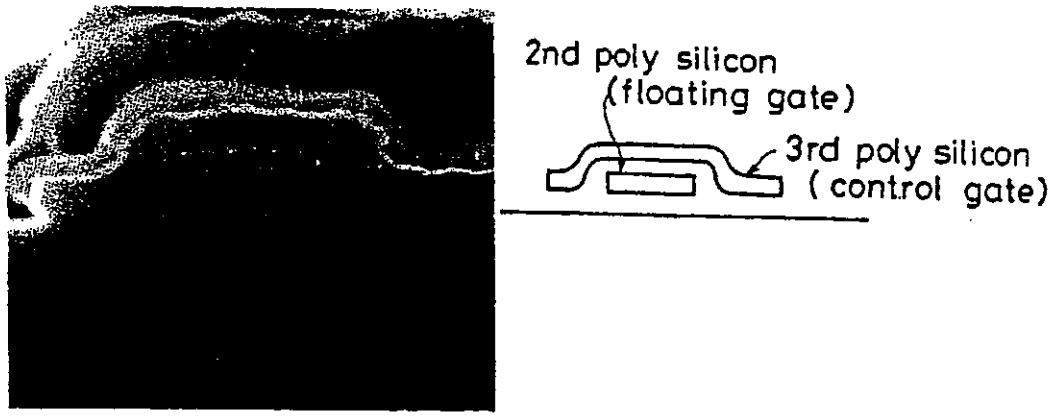
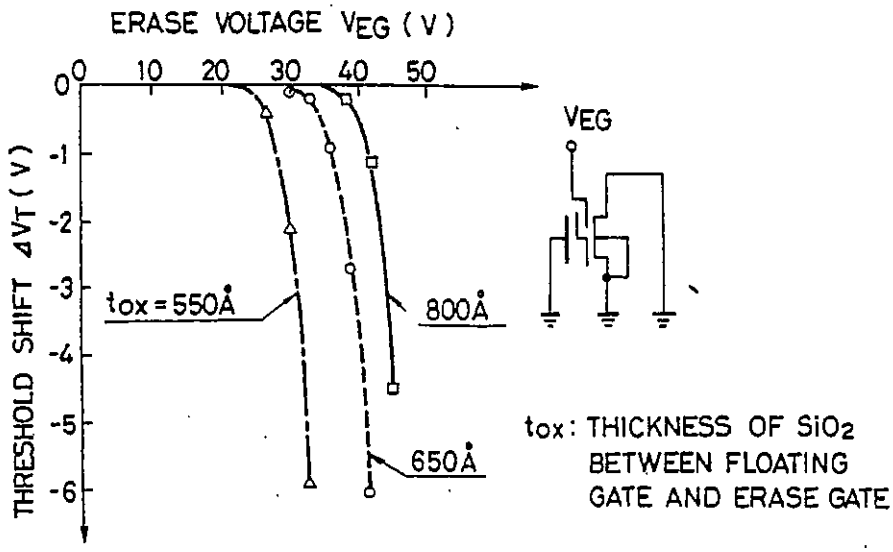


Fig.6 SEM photograph of the cross sectional view of F-E²PROM cell.



(a) ERASE CHARACTERISTICS

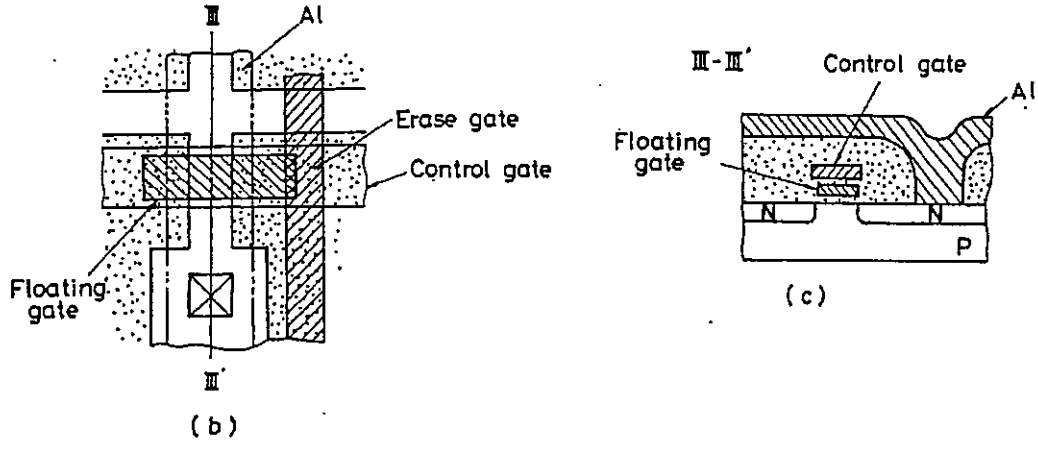


Fig.5 (a) The threshold voltage shift of the memory cell versus erasing voltage applied to the erase gate with oxide thickness between floating gate and erase gate as a parameter.

(b) Top view of this experimental F-E^{2}PROM cell without the selection transistor.

(c) Cross sectional view along Ⅲ-Ⅲ' line in Fig.5-(b)

EVOLUTION OF NONVOLATILE SEMICONDUCTOR MEMORY

From Floating-Gate Concept to Single-Electron Memory Cell

S. M. Sze
UMC Chair Professor
National Chiao Tung University
National Nano Device Laboratories
Hsinchu, Taiwan, R.O.C.

1. INTRODUCTION

Semiconductor memories constitute about 30% of the world integrated circuit (IC) market[1]. In ten years, they will capture the largest market share with over 50% of the total IC sales. Therefore, semiconductor memories will stop being looked at as an auxiliary to the microprocessors (or other logic circuits) and microprocessors will start being looked at as aids to the optimized and predominant memories on IC chips.

Semiconductor memories can be broadly classified into two groups: volatile and nonvolatile[2]. Volatile memories such as DRAM (dynamic random access memory) and SRAM (static random access memory) lose their stored information once the power supply is switched off. Nonvolatile memories, on the other hand, can retain their stored information.

The nonvolatile semiconductor memory (NVSM) group includes the following members[3] :(1) ROM (ready-only memory) with data permanently written during manufacturing, (2) PROM (programmable ROM) with data which can be written only once using fuse or antifuse process, (3) floating-gate memory, (4) silicon-nitride memory, (5) FRAM (ferroelectric RAM), and (6) MRAM (magnetoresistive RAM).

In order to compare the characteristics and performances of these memories, we first define twelve attributes of an ideal memory: (1) nonvolatility with long retention (i.e., the ability to retain data without power over long period of time, typically over 10 years), (2) high density (i.e., small cell area), (3) low-power consumption, (4) in-system rewritability, (5) bit alterability, (6) fast read / write, (7) high endurance (i.e., the ability to maintain the stored information after erase/program/read cycling), (8) low cost, (9) single-power supply, (10) highly scalable(i.e., cell size can be reduced with the minimum feature length), (11)

ruggedness, and (12) highly integrable with silicon IC technology.

A comparison[3] of various memories based on the aforementioned attributes is shown in Table 1. The ROM and PROM are not included, since they can not be re-programmed. Neither is silicon-nitride memory because of its low endurance and low retention. The floating-gate memory can be subdivided into three categories: Flash memory (see Sec.2), EEPROM (electrically erasable programmable ROM), and UV-EPROM (ultraviolet erasable programmable ROM).

As can be seen from Table 1, no memory has all the attributes of an ideal memory. For example, DRAM suffers from relatively high power consumption and a non-scalable storage capacitor in addition to its volatility. SRAM suffers from low density and volatility. FRAM and MRAM are still in their initial development stage and the main concerns are their process uniformity and compatibility with silicon IC processes. We have also included the hard and floppy disks, which suffer from high power consumption and low ruggedness. The memory that has the most desirable attributes is the Flash memory in the floating-gate memory family. In the subsequent discussion, we will concentrate on the dominant NVSM—the floating-gate memory.

2. HISTORICAL DEVELOPMENTS

2.1 Floating-Gate Concept

The NVSM was proposed[4] by Kahng and Sze in 1967. The proposal introduced the floating gate concept for charge storage and nonlinear transport processes for programming and erasing. From a historical perspective, the 1967 proposal recognized for the first time the possibility of rewritable NVSM devices.

Figure 1 shows the cross-sectional view of the first NVSM with a floating gate M(1). The basic operations of programming (e.g., Fowler-Nordheim tunneling), storage, and erase (e.g., reverse FN tunneling) are shown in Fig.2. If the

Table 1. Comparison of Memory Attributes[3].

Attribute \ Device	DRAM	SRAM	Flash	EEPROM	UV-EPROM	FRAM	Hard Disk	Floppy Disk
1. Nonvolatility	—	—	✓	✓	✓	✓	✓	✓
2. High density	✓	—	✓	—	✓	✓	✓	—
3. Low power	—	—	✓	✓	✓	✓	—	—
4. In-system rewritability	✓	✓	✓	✓	—	✓	✓	✓
5. Bit alterability	✓	✓	—	✓	—	✓	✓	✓
6. Fast read/write	✓	✓	✓	✓	✓	✓	—	—
7. High endurance	✓	✓	✓	✓	—	✓	✓	✓
8. Low cost	✓	—	✓	—	✓	—	✓	✓
9. Single-power supply	✓	✓	✓	✓	—	✓	✓	✓
10. Highly-scalable	—	—	✓	—	✓	✓	—	—
11. Ruggedness	✓	✓	✓	✓	✓	✓	—	—
12. Highly integrable	✓	✓	✓	✓	✓	—	—	—
Total Desirable Attributes	9	7	11	9	8	10	7	6

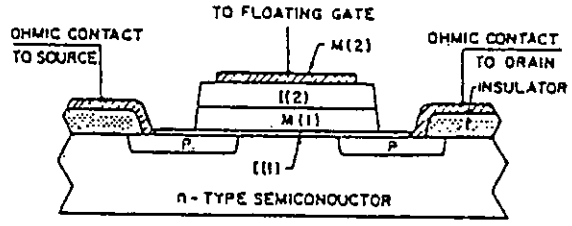


Fig. 1 Cross-sectional view of the first proposed nonvolatile semiconductor memory with a floating gate in 1967. (Ref. 4).

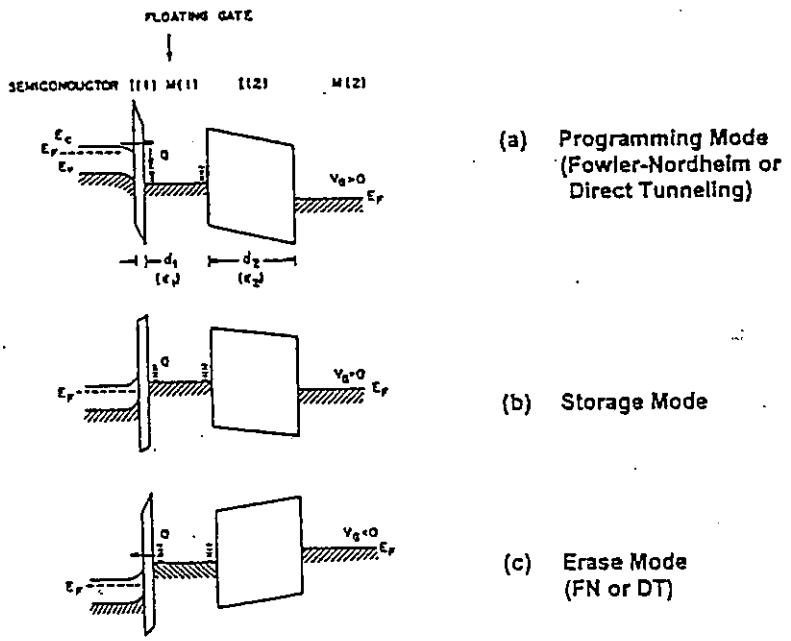


Fig. 2 Programming, storage, and erase modes for the floating-gate memory. (Ref. 4).

insulators I(1) and I(2) are sufficiently thick, the charge in the floating gate can be stored for a long time.

An experimental device was also made using SiO_2 as I(1), Zr as M(1), and ZrO_2 as I(2). When a voltage pulse of 50V with a pulse duration of 0.5 μs was applied to the control gate, M(2), about 10^{12} electrons /cm^2 were transported to and stored in the floating gate. The stored charge caused a large threshold voltage shift, and the device was "on" with a channel current of 0.25 mA. When a large negative pulse was applied to the control gate, the stored charge was depleted and the device was " off " as shown in Fig.3. From the slope of the "on" state, this NVSM can store information (i.e., charge) for a few hundred milliseconds. This result can be considered as the first demonstration of the EEPROM operation.

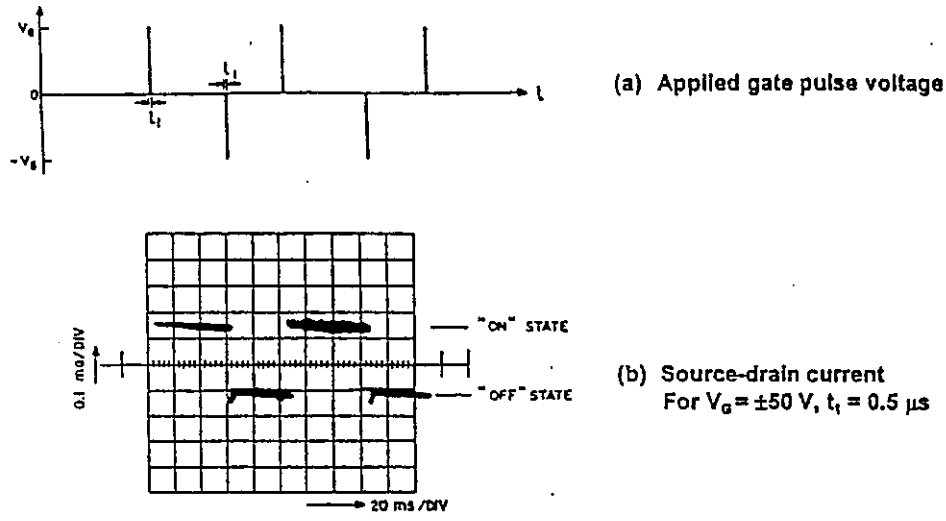


Fig. 3 First demonstration of the operation of an electrically erasable programmable read only memory (EEPROM). (Ref. 4).

2.2 Early Device Structures

In 1971 Frohman-Bentchkowsky developed the FAMOS[5] (floating –gate avalanche-injection MOS) shown in Fig. 4a. This structure is an EPROM and has a floating gate but no control gate. The stored charge in the floating gate came from injection of hot electrons generated at the avalanche region near the drain. Since FAMOS has no control gate, the stored charge can not be erased electrically. One has to use ultraviolet light to do the erasing.

In 1976, Iizuka and his coworkers studied the SAMOS[6] (stacked gate avalanche injection MOS) shown in Fig. 4b. This structure is an EEPROM, and it looks essentially the same as that in Fig.1. However, the injection mechanism is due

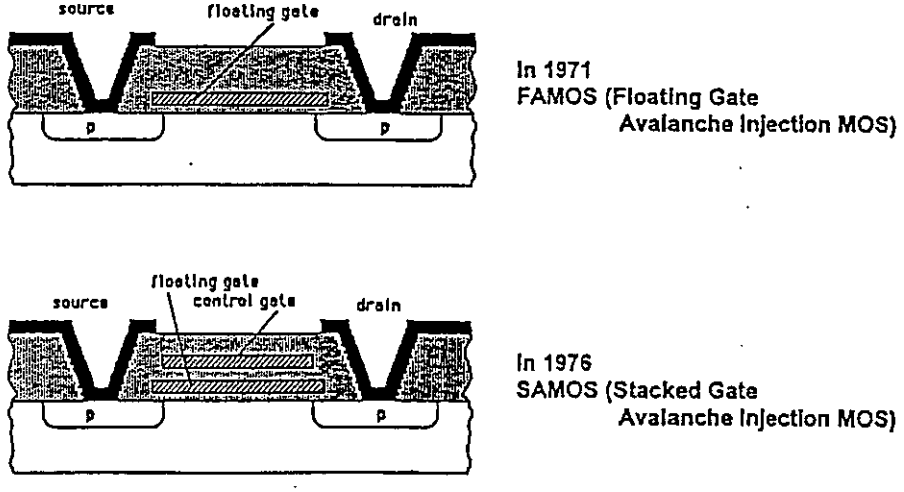


Fig. 4 Two early floating-gate memories. (Ref. 5 and 6).

to avalanche instead of Fowler-Nordheim tunneling. Because of the relatively thick oxide between the floating gate and the channel, a substantial improvement of the retention time was obtained. However, for a typical EEPROM operation, we need two devices per cell, i.e., an EEPROM and a selection MOSFET. Therefore, the cell size is relatively large.

In 1984, Masuoka and his coworkers developed the Flash memory[7] In its erase operation, a whole block is erased, thus the name "Flash". The top and cross-sectional view of a Flash memory are shown in Fig.5. Since there is only one device per cell, Flash memory has the advantages of higher density, lower cost, and higher scalability compared to the EEPROM.

All early NVSMs were not very reliable. They all suffered from low endurance and poor retention. This is expected, since the injection (or reverse injection to remove the stored charge) mechanisms of the avalanche process and Fowler-Nordheim tunneling generally cause reliability problems in conventional MOSFETs. For example, hot-election trapping in oxide or the time-dependent dielectric breakdown (TDDB) in oxide may induce leakage currents; this, in turn, may result in poor endurance and low retention.

To improve the reliability, the programming voltage for Fowler-Nordheim tunneling can be reduced by using thinner oxide between the floating gate and the channel. The programming voltage for hot–electron injection can also be reduced by using the channel-induced substrate electron injection (CISEI) [8] process shown in Fig.6. The injected current I_G is determined by the multiplication M_1, creating holes that ionize in the substrate with multiplication M_2, and then reach the floating gate with probability T. For the CISEI process, the programming voltage can be as low as 2.5 V. Figure 6 also shows a structure with a p-Halo to increase the generation of channel-induced substrate electrons.

2.3 Flash Memory Structures

Figure 7 shows the world NVSM market since 1980. Notice that the market share of the Flash memory has increased rapidly since 1990. At present, it has eclipsed both EPROM and EEPROM to become the dominant NVSM with a market share near 70%.

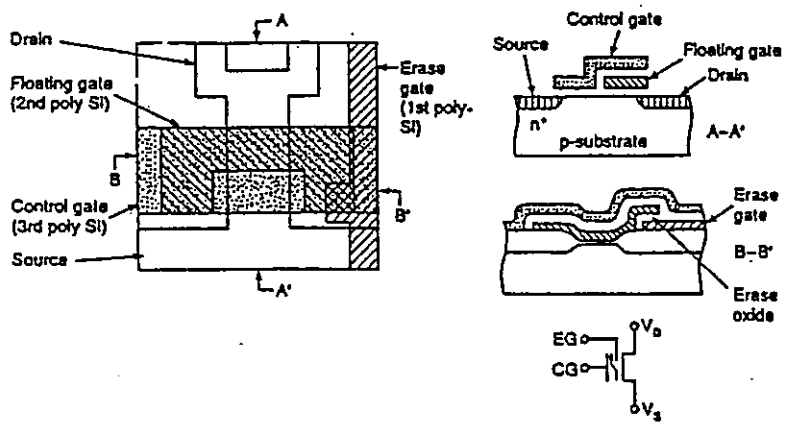


Fig. 5 The Flash memory proposed in 1984. In the erase operation, a whole block is erased, thus the name "Flash". (Ref. 7).

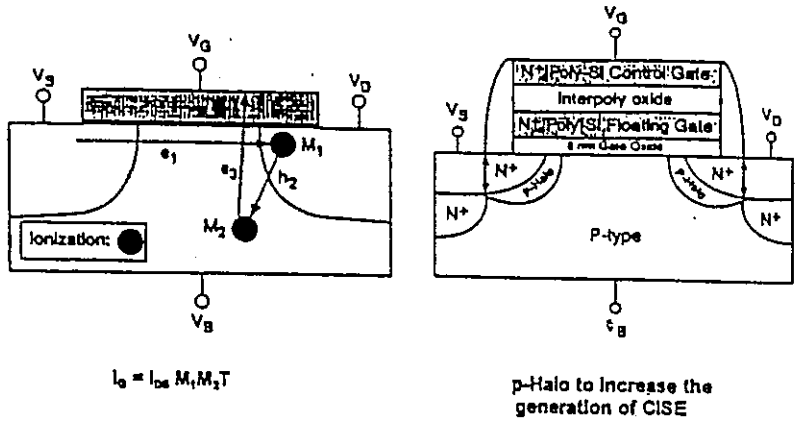


Fig. 6 The gate current I_G is determined by the multiplication M_1, creating holes that ionize in the substrate with multiplication M_2, and then reach the gate with probability T. (Ref. 8).

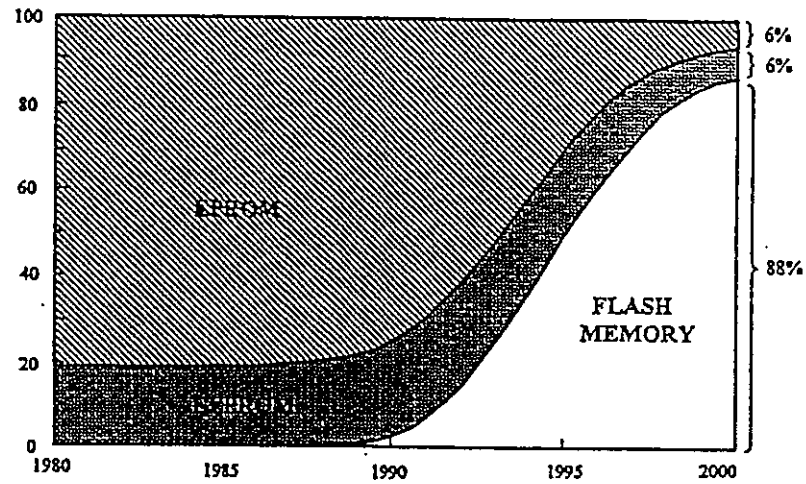


Fig. 7 World NVSM market. By Year 2000, Flash memory will have nearly 90% of the market share.

The key reason for this dominance is its small cell size. Figure 8 shows a comparison of cell sizes of various memory devices versus minimum feature length[9]. SRAM with six transistors (or 4 transistors and 2 load resistors) has, of course, the largest cell size. EEPROM with a selection transistor is also quite large. DRAM is approximately 50% larger than the Flash cell became of its storage capacitor. By using a shallow-trench isolation (STI) approach, a Flash memory can have very small cell size[10]. For example, at a 0.25 μm design rule, the cell area is 0.31 μm^2, and at a 0.18 μm design rule, the cell size is only 0.16 μm^2. A process flow of the STI-cell is shown in Fig.9.

Excellent endurance characteristics have been obtained for the STI cell with 9 nm oxide between the floating gate and the channel. The upper threshold voltage remains at 1.5V and shows no change at all after 10^6 program/erase cycles, while the lower threshold voltage varies from –2.0V to –1.6V. This memory is expected to maintain an acceptable threshold voltage window of 2V at least to 10^8 program/erase cycles.

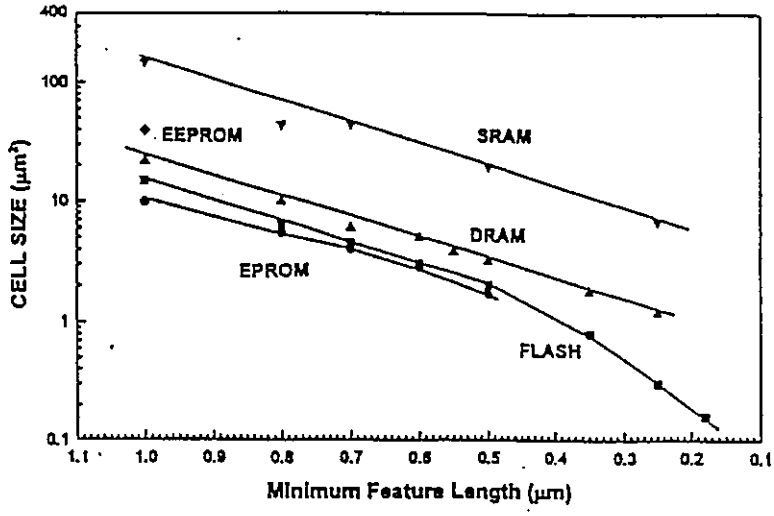


Fig. 8 Memory cell size versus minimun feature length for various memory devices. (Ref. 9).

The Flash memory has many different cell designs[11]. These designs can be divided according to data-access and data-write organization. Arrays that have random access and random program (parallel) are consistent with embedded applications (see Sec.3), while page read and page program (serial) are consistent with mass-storage applications. For programming operations, the different designs use either hot-electron injection or Fowler-Nordheim tunneling. For erasing, Fowler-Nordheim tunneling is used for all designs. In the parallel architectures, we have the asymmetrical contactless transistor, divided bit-line NOR, and common ground standard NOR. In the serial architectures, we have the high-capacitance-coupling-ratio cell, triple-poly virtual-ground cell, and split-gate cell.

2.4 Single-Electron Memory Cell

The single-electron memory cell (SEMC) is actually a limiting case of the floating-gate structure. By reducing the length of the floating gate, M(1), in Fig. 1, to ultra small dimensions, e.g., 10 nm, we obtain the SEMC. A cross-sectional view of a SEMC is shown in Fig.10a. The storage dot corresponds to M(1) in Fig.1. It is located between the control gate and the channel. Because of its small size the capacitance is also very small ($\sim 10^{-18}$ F).

The band diagrams of a SEMC are shown in Fig.10 b. The quantum well (i.e., the storage dot) can only accommodate one electron[12]. When an electron tunnels into the quantum well, the potential on the left side is reduced to a level that the transfer of another electron is blocked. This is called the "Coulomb Blockade." Note that these band diagrams are similar to those shown in Fig. 2a and 2b. The SEMC is an ultimate floating-gate memory cell, since we need at least one electron for information storage. The operation of a SEMC at room temperature was first demonstrated by Yano et al[12] in 1994 using a floating gate with a gate length of 10 nm. Recently a 128 megabit memory using a vertically unified cell structure of SEMCs has been reported[13].

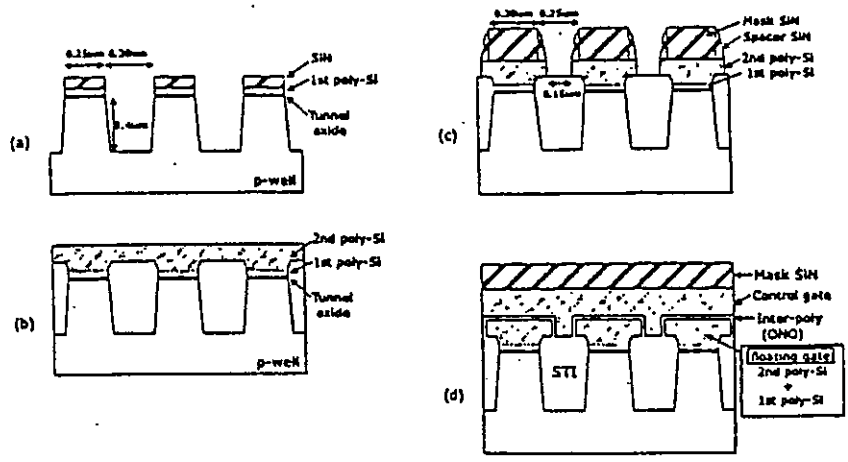


Fig. 9 Process flow of the shallow-trench isolation cell. This cell offers very small cell area. (Ref. 10).

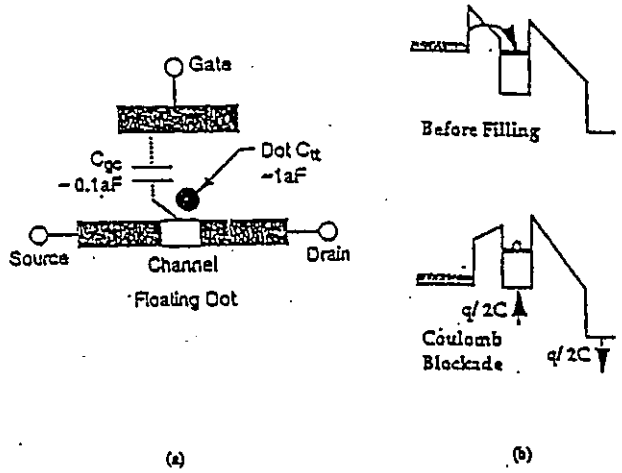


Fig. 10 Single-electron memory cell. (a) Cross-sectional view of the device. The floating gate is reduced to about 10nm. (b) Due to the Coulomb Blockade, the information is stored in the form of a single electron. (Ref. 12)

3. MAJOR APPLICATIONS

3.1 Embedded Memory

NVSM can be integrated with logic systems (e.g., microprocessors) to form systems-on-chip. Such IC chips can have a wide range of applications including[14]:

(1) Automation: laser printer, ink jet printer, hard disk drive, copier.
(2) Automobile: power train control, automatic breaking system,

instrumentation, navigation.

(3) Communication : cellular phone, cordless phone, paging, cable TV-set top box.

(4) Consumer: camera, camcorder, audio recorder, smart card.

(5) Industry: servo control, motor control, bar code reader.

An example of the embedded memory system is the cellular phone. A block diagram for such system is shown in Fig.11. A Flash memory is embedded in the microcontroller unit along with a microprocessor, a RAM, and a timer[14]. Another example is the single-chip multi-media personal computer which has a Flash memory BIOS (basic input and output system) and a Flash disk controller[15].

3.2 Mass Storage

The hierarchy of memories in a computer system has five levels: the register in the central processing unit (CPU), cache memory, main memory, hard and floppy disks, and magnetic tape[16]. The register in the CPU, cache memory, and main memory are generally volatile memories. On the other hand, the hard disk, floppy disk, and magnetic tape are nonvolatile memories. As we move up the hierarchy (i.e., from magnetic tape to the register in CPU), the access time becomes faster; and as we move down the hierarchy, the memory capacity increases. As a result, the market for hard and floppy disks is about five times larger than that for DRAM.

Table 2 shows a comparison of NVSM with hard disk for portable-computing storage[17]. We note that the Flash solid state disk is 70 times faster in access time, 200 times lower in power consumption, 26 times lighter in weight, 29 times smaller in size, and 100 times better in ruggedness than a hard disk. At the present time, the only advantage of a hard disk is its price. However, the price per Mbytes of a NVSM is being reduced at a faster rate than that of a hard disk. It is, therefore, expected that in a few years NVSM will replace most of the hard and floppy disks.

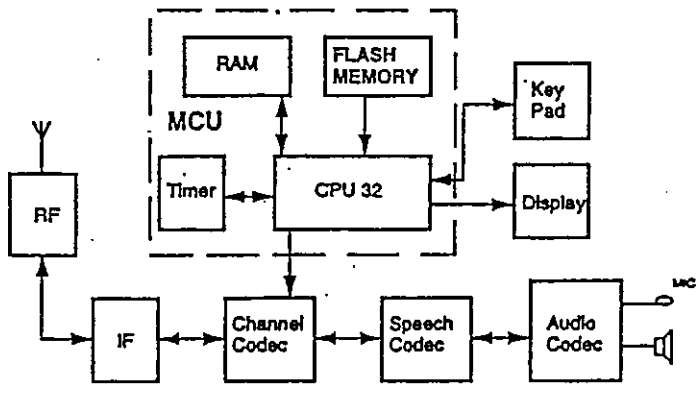


Fig.11 A block diagram of a cellular phone with Flash memory. (Ref. 14).

Table 2. NVSM versus hard disk for portable-computing mass storage[17].

Feature \ Memory	Nonvolatile Memory (Flash Solid State Disk)	Magnetic Disk (2.5" HDD)	Ratio HDD/FSSD
Access Time (ms)	0.2	14	70
Power Consumption (Watt-hours /hour)	0.002	0.4	200
Weight (g/Mbytes)	0.15	4	26
Volume (cm^3/Mbytes)	0.1	2.9	29
(Operating Shock)$^{-1}$ (1/Gs)	0.001	0.1	100
Price ($/Mbytes)	~1	~0.2	0.2

3.3 Technology Drivers

The electronic industry is the largest industry in the world with global sales over US$1,000 billions. The foundation of the electronic industry is the semiconductor industry. Figure 12 shows the sales volumes of these two industries in the past ten years and projects the sales to year 2010. If the current trends continue, the sales volumes of the electronic industry and the semiconductor industry will reach $3000 billions and $500 billions, respectively, in year 2010.

Also shown in Fig.12 are the sales volumes for DRAM, SRAM and NVSM. The global sales of NVSM is about $6 billions with an impact on the system level well over $100 billions. NVSM has surpassed SRAM in 1996. If it maintains a 35% to 45% annual growth rate, NVSM will surpass DRAM in about 5 years to become the dominant memory for the electronic industry.

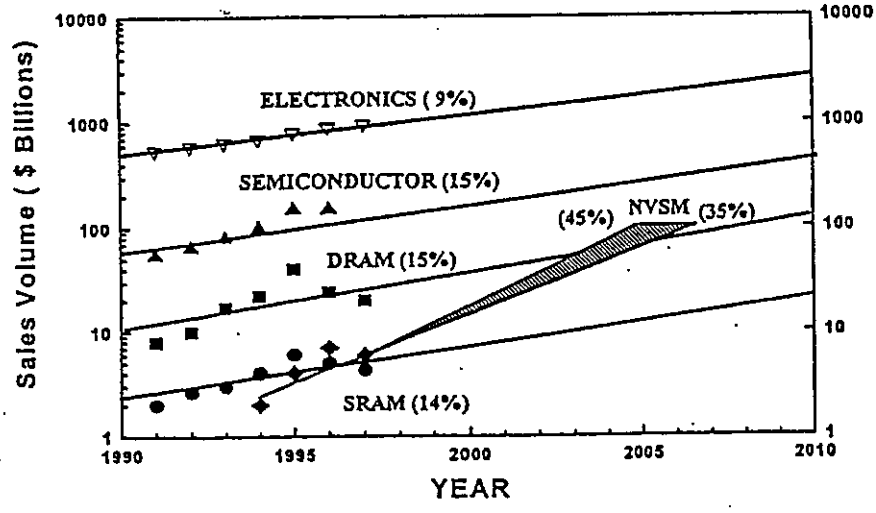


Fig. 12 Sales volume of electronic industry, semiconductor industry, DRAM, SRAM and NVSM. Sales of the NVSM have surpassed SRAM since 1996, and will surpass DRAM in 5 to 10 years.

The growth curves for different technology drivers are shown[16] in Fig. 13. At the beginning of the modern electronic era (1950 to 1970) the bipolar transistor was the technology driver. From 1970 to 1990, the DRAM and microprocessor were the technology drivers due to the rapid growth of personal computers and advanced electronic systems. Since 1990, NVSM has been the technology driver. This is because of the rapid growth of portable electronic systems such as the cellular phone, notebook computer, smart card, etc, which need memories having the most desirable attributes.

4. CONCLUSION

We have presented a brief review of the historical developments of nonvolatile semiconductor memory (NVSM) and projected its trends to year 2010. In the past 30 years, NVSM has emerged from a simple floating-gate concept to a multi-billion-dollar industry. A key member of NVSM family is the Flash memory which has captured about 70% of the NVSM market. The ultimate Flash memory is the single-electron memory cell, which may serve as a building block for 1 Tb (10^{12} bits) nonvolatile memory.

Because of its attributes of high density, low-power consumption, nonvolatility, and electrical rewritability, NVSM has become the dominant memory for portable electronic systems such as cellular phone, notebook computer, smart IC card, etc. NVSM has surpassed SRAM in sales volume. It is now poised to eclipse DRAM's market share and to replace hard and floppy disks in the near future.

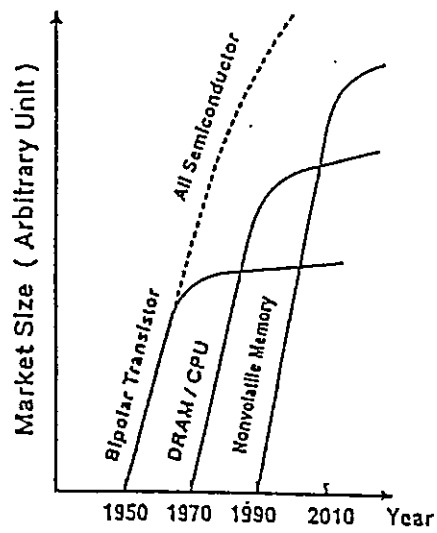


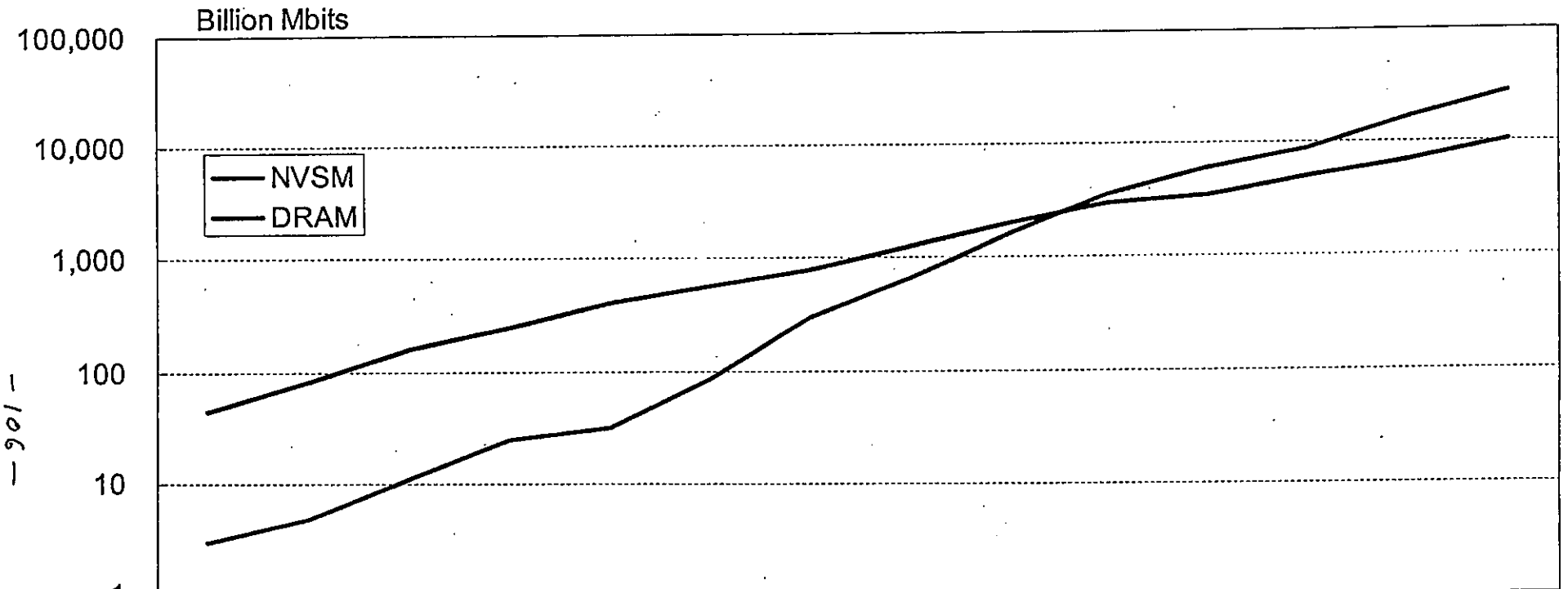
Fig.13 NVSM has been the technology driver since 1990. (Ref. 16).

References

1. *1997 Electronic Market Data Book*, Electronic Industries Association, Washington, D. C., 1997.
2. A. K. Sharma, *Semiconductor Memories-Technology, Testing, and Reliability*, IEEE Press, Piscataway, 1997.
3. W. D. Brown and J. E. Brewer, Eds, *Nonvolatile Semiconductor Memory Technology*, IEEE Press, New York, 1998.
4. D. Kahng and S. M. Sze, "A Floating Gate And Its Application To Memory Devices", *Bell Syst. Tech. J.*, **46**, 1288(1967).
5. D. Frohman-Bentchkowsky, " Memory Behavior In A Floating-gate Avalanche-Injection MOS (FAMOS) Structure", *Appl., Phys, Lett.*, **18**, 332 (1971).
6. H. Iizuka, F. Masuoko, T. Sato, M. Ishikawa, "Electrically Alterable AvalancheInjection Type MOS Read-Only Memory with Stacked Gate Structure", *IEEE Trans. Elect. Dev.*, **ED-23**, 379 (1976).
7. F. Masuoko, M. Asano, H. Iwahashi, T. Komuro, and S. Tanaka. "A New Flash E^2PROM cell using triple polysilicon technology", *IEDM Tech.Dig.*, 464(1984)
8. J. Bude, A. Frommer, M. Pinto, and G. Weber, "EEPROM/ Flash Sub 3.0V Drain-Source Bias Hot Carrier Writing", *IEDM Tech. Dig.*, 990(1995).
9. S. Lai, "The Outlook of Flash EPROM Technology", *Proc. VLSI Technology Systems, and Application*, 147, May 12, 1993.
10. K. Shimizu, K. Narita, H. Watanabe, E. Kamiya, Y. Takeuchi, T. Yaegashi, S. Aritome, and T. Watanabe, "A Novel High Density $5F^2$ NAND STI Cell Technology Suitable for 256 Mbit and 1 Gbit Flash Memories", *Tech. Dig. IEEE IEDM*, 271(1997).
11. P. Pavan, R. Bez, P. Olive, and E. Zanoni, "Flash Memory Cells – An Overview", *Proc. IEEE*, **85**, 1248 (1997).
12. K. Yano, T. Ishii, T. Hashimoto, T. Kabayashi, F. Murai, and K. Seki, "Room Temperature Single-Electron Memory", *IEEE Trans, Elect, Dev.*, **41**, 1628 (1994).
13. K. Yano, T. Jshii. T. Mine, F. Murai, T. Kure, and K. Seki. "128 Mb Early Prototype for Gigascale Single-Electron Memories", *Digest of IEEE Solid State Circuit Conference*, 344 (1998).
14. C. Kuo, " Embedded Flash Memory, Application, Technology, and Design", IEDM Short Course on NVRAM Technology and Application (1995).
15. H. Sasaki, "Multimedia Future and Impact for Semiconductor Technology", *Tech. Digest, IEEE Int. Elec. Dev. Meeting*., 3 (1997).
16. F. Masuoka, "Flash Memory Technology", *Int. Elec. Device and Materials Symposium*, 83(1996).
17. A. G. Barre, " Flash Memory, Magnetic Disk Replacement?" *IEEE Trans. Magnetics*, **29**, 4104(1993).

(A chapter in "Future Trends in Microelectronics", Ed. S. Luryi et al Wiley Interscience, New York, 1999.)

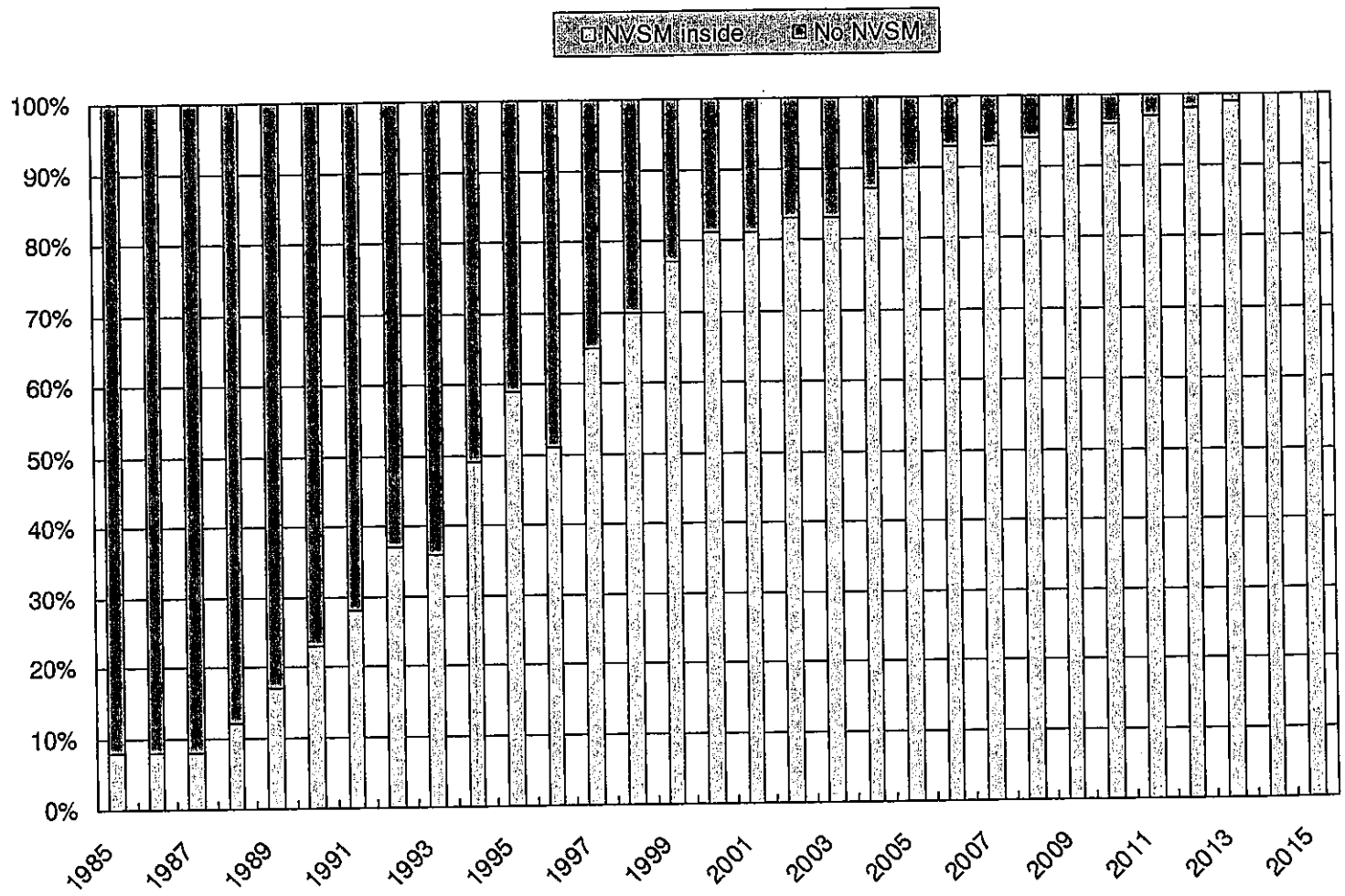
Bits Shipment Trend(NVSM vs DRAM)



	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
NVSM	3	5	11	24	31	85	291	647	1,609	3,565	6,127	8,911	16,594	27,914
DRAM	44	83	155	246	400	563	785	1,260	2,003	2,978	3,499	5,086	7,014	10,496

DRAM Source:1997~2005 is from WSTS actual data and 2006~2010 is from Infineon,San Jose,2005/10
NVSM(NAND+NOR) Source: 1997~2005 is from WSTS actual data and 2006~2010 based on Web Feet Research number * (85%)

NVSM Penetration Trend in Electronics Industry, 1985 - 2015



Source: The Information Network,
adjusted by MXIC