Semiconductor Devices

A. Biased *p-n* Junctions
B. Metal/Semiconductor Schottky Contact
C. Bipolar Transistors
D. Field Effect Transistors
E. Dynamic Random Access Memory
F. Charge-Coupled Devices
G. Complementary MOS in Logic Gates
H. Semiconductor lasers

Biased *p-n* Junction

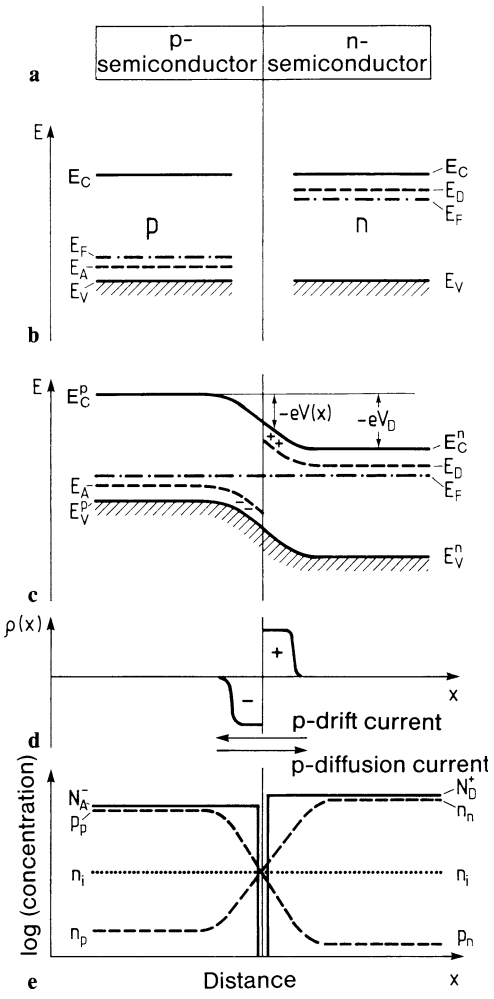
If an external electrical voltage *U* is applied to a *p-n* junction, thermal equilibrium is destroyed, and the situation in the *p-n* junction can be described as a stationary state in the vicinity of thermal equilibrium. Because the space-charge zone between $-d_p$ and d_n has a considerably higher electrical resistance than the region outside the *p-n* junction, the potential drop across the space-charge zone accounts for nearly all of the externally applied voltage *U*. Thus the potential drop across the space-charge zone, instead of being equal to the diffusion voltage V_D, now has the value

$$V_n(\infty) - V_p(-\infty) = V_D - U \, .$$

One thus has

$$d_n(U) = d_n(U=0)(1 - U/V_D)^{1/2} \, ,$$

$$d_p(U) = d_p(U=0)(1 - U/V_D)^{1/2} \, .$$



Let us consider the balance in the electron currents: we are concerned on the one hand with the drift currents of the minority carriers coming from the *p* region (where electrons are the minority carriers), which are drawn across into the *n* region. Because these minority carriers are continually generated in the *p* region by thermal excitation, this current is called the *generation current*, I^{gen} .

$$I^{\text{gen}} = e^{-eE_g/2kT}$$

This current is largely independent of the value of the diffusion voltage and therefore also of the external voltage.

The diffusion current of electrons from the *n* region, where the electrons are majority carriers, into the *p* region (called the *recombination current* I^{rec}). With an externally applied voltage *U*:

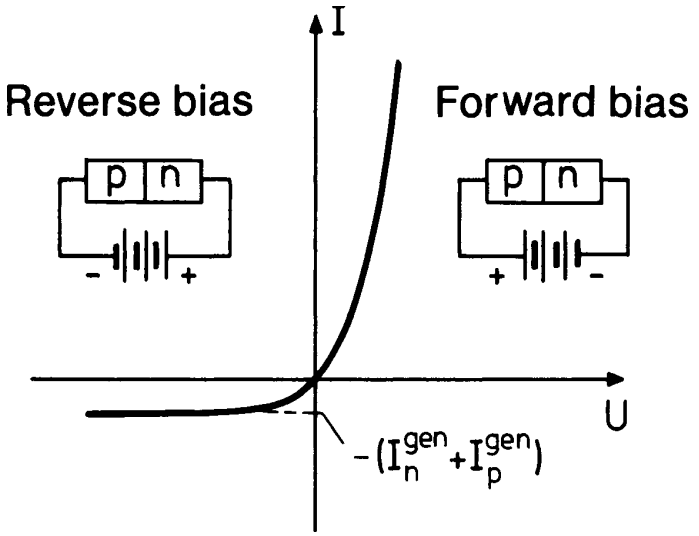
$$I_n^{\text{rec}} \propto \mathrm{e}^{-e(V_{\mathrm{D}}-U)/\not{k}\,T} = I_n^{\text{gen}}\mathrm{e}^{e\,U/\not{k}\,T} ,$$

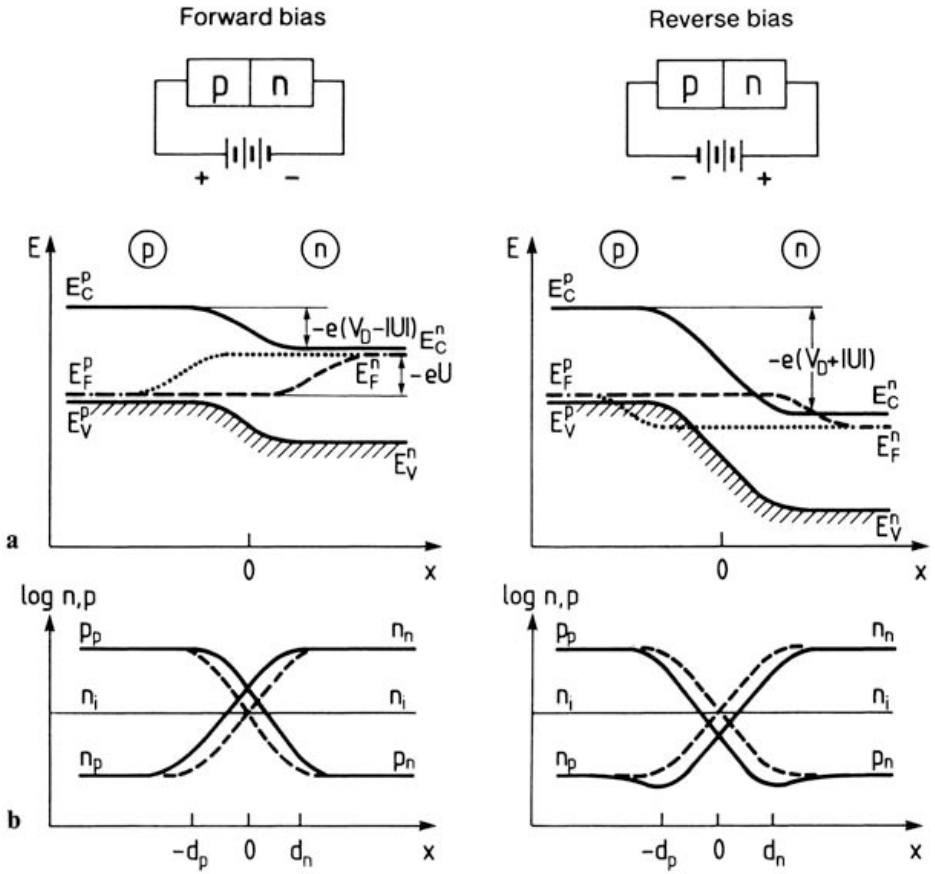
and therefore a total electron current I_n is

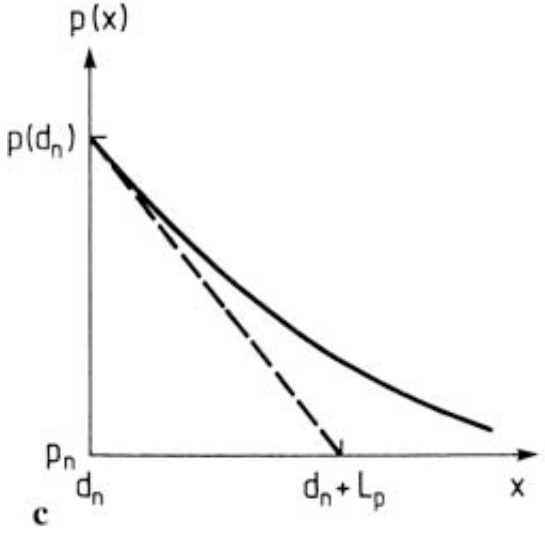
$$I_n = I_n^{\text{rec}} - I_n^{\text{gen}} = I_n^{\text{gen}}(\mathrm{e}^{eU/\not{k}\,T} - 1) .$$

The total current $I(U)$ under applied external U is

$$I(U) = (I_n^{\text{gen}} + I_p^{\text{gen}})\left(\exp\frac{e\,U}{k\,T} - 1\right) .$$

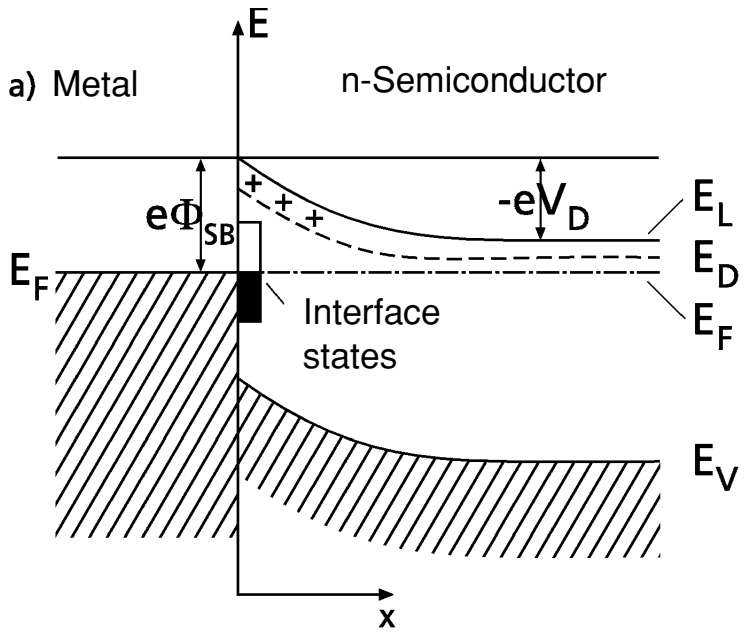






Metal/Semiconductor Schottky Contact

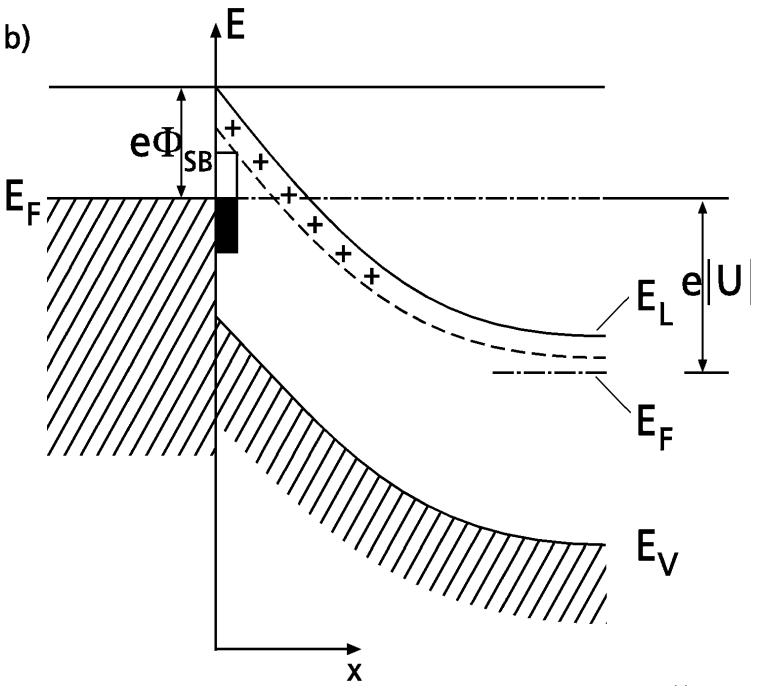
Electronic interface states are formed at the metal/*n*-semiconductor interface. Their spatial extension is limited to a few atomic layers at the interface and their energetic distribution is fixed with respect to the conduction and valence band edges of the semiconductor. These interface states, sometimes called MIGS (metal-induced gap states) originate from the Bloch waves in the metal.



Conduction band states are negatively charged when occupied by electrons and neutral in the unoccupied state; valence band states, however are positive when empty and neutral in the occupied state. The band bending within the semiconductor is determined by a charge balance between negative charge Q_{is} in the interface states and the positive space charge Q_{SC} in the depletion layer.

An external bias produces a voltage drop essentially across the space-charge zone. The mathematical description of the space-charge region below a metal-semiconductor junction is analogous to a *p-n* junction. The thickness of the Schottky contact space charge region in thermal equilibrium, e.g., is obtained as

$$d = \left(\frac{2\,\varepsilon\,\varepsilon_0 V_\mathrm{D}}{e\,N_\mathrm{D}} \right)^{1/2}$$



Similarly, the capacity of a metal-semiconductor junction as a function of external bias can be obtained as

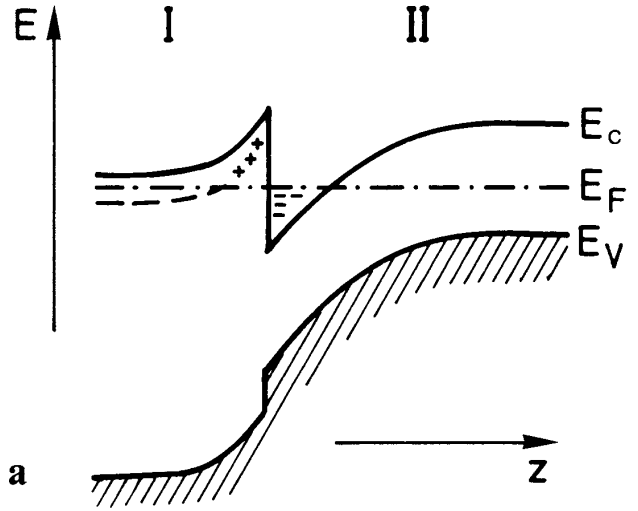
$$C = \frac{A}{2} \left(\frac{2\,e\,\varepsilon\,\varepsilon_0 N_\mathrm{D}}{V_\mathrm{D} - U} \right)^{1/2},$$

where A is the area of the contact.

Electron transport from the metal into the semiconductor requires that the carriers overcome the Schottky barrier $e\Phi_{SB}$.

Isotypic Heterojunctions

Of particular interest are heterojunctions between two different semiconductors with the same doping, so-called isotypic heterojunctions. In this case, because of the continuity conditions for the Fermi level, an accumulation space-charge zone for electrons is created on the side of the semiconductor with a smaller forbidden gap, which leads to an extremely large increase in local electron concentration.

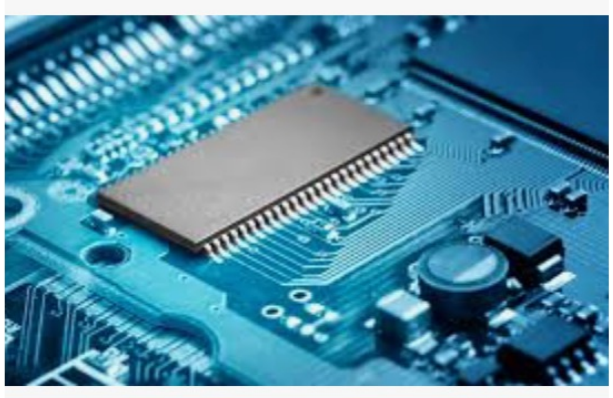


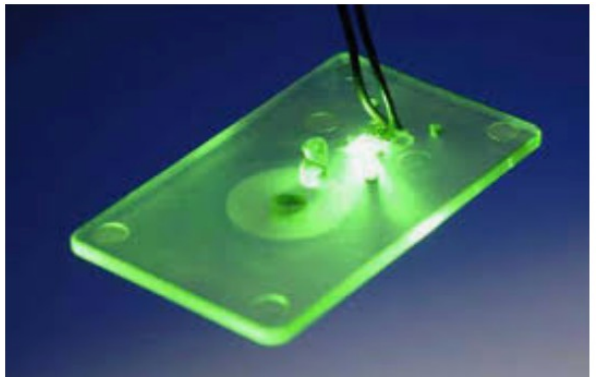
The high concentration of free electrons in this space-charge zone (semiconductor II) is compensated by a depletion space-charge zone in semiconductor I. In this way the high density of free electrons is spatially separated from the ionized impurities from which they originate.

Impurity scattering, which is an important source of electrical resistance at low temperature, is therefore strongly reduced for this free electron gas.

Modern Information Technology

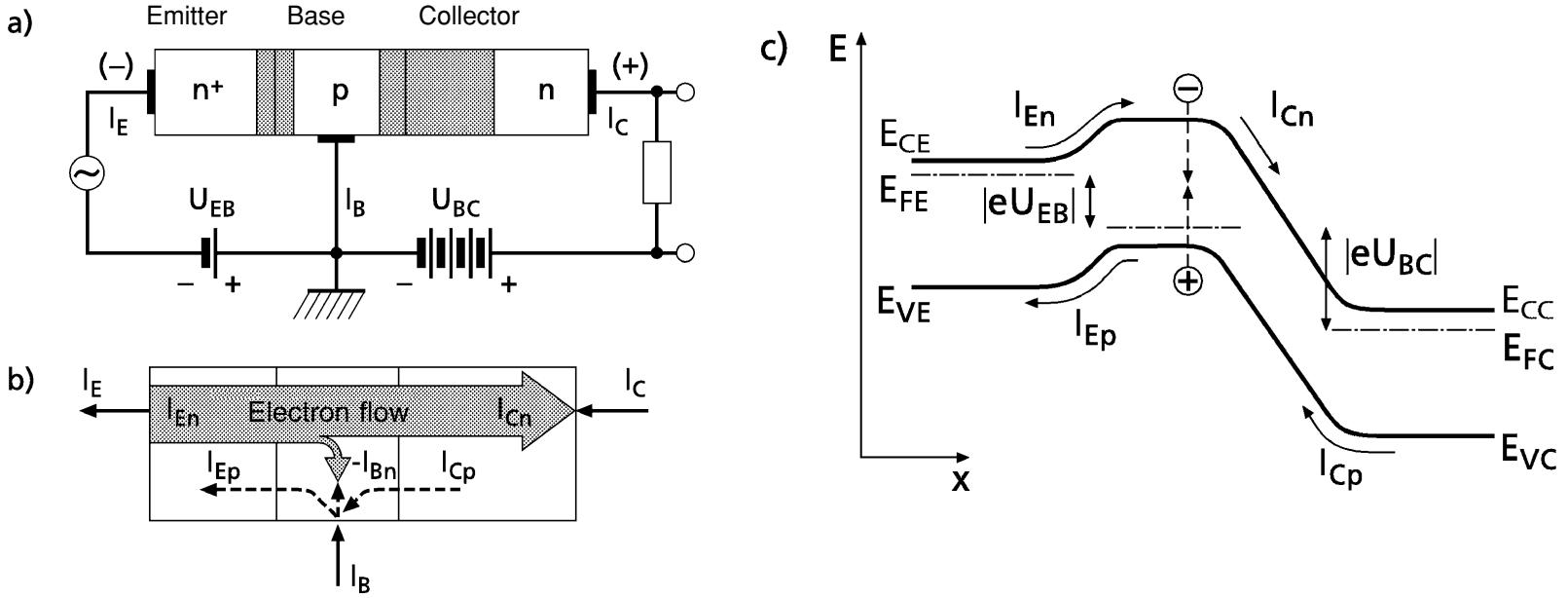
Modern information technology includes data processing (logic), data storage (memory) and data transfer in networks. More than 90% of the integrated circuits (ICs) are fabricated on silicon wafers and are used for data processing and storage. The reason is that the elemental semiconductor silicon permits the highest density of integration. As for data transfer, where optoelectronic devices such as detectors and optical emitters (light emitting diodes, LEDs, and lasers) are the fundamental elements. While the major material base for optoelectronic devices are the III-V semiconductors, and particularly GaAs. The reason is the much stronger coupling of electronic transitions to the electromagnetic field in the case of direct band gap III-V semiconductors is essential for optoelectronic devices.



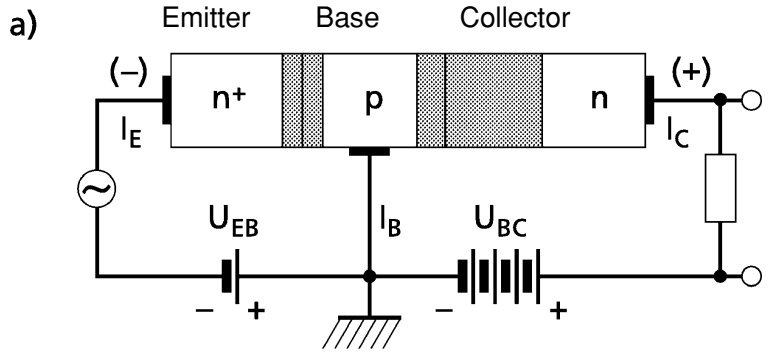


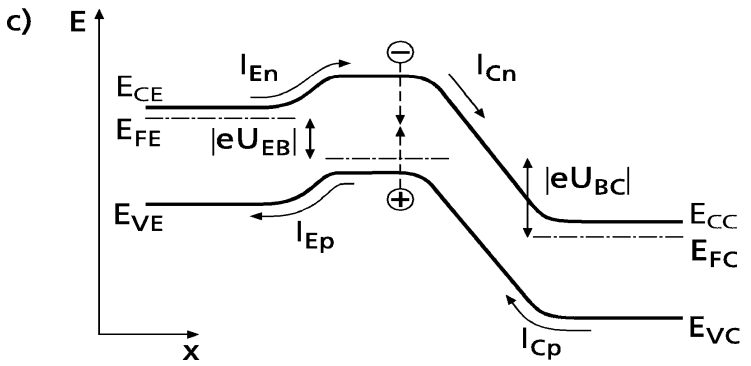
Bipolar Transistors

The classical bipolar transistor, invented by Bardeen, Brattain and Shockley in 1947, consists of two oppositely biased *pn*-junctions. Accordingly there are both, *npn*- and *pnp*-transistors; in *npn* devices the current is essentially carried by electrons, while in the *pnp* by holes.



A scheme of an *npn*-transistor together with its external circuitry is shown: the first n^+p-junction (emitter-base) is biased in the forward direction and the base-collector junction is biased in the reverse direction.





$$I_{\mathrm{E}} = I_{\mathrm{E}n} + I_{\mathrm{E}p} \ ,$$

$$I_{\mathrm{C}} = I_{\mathrm{C}n} + I_{\mathrm{C}p} \ ,$$

$$I_{\mathrm{B}} = I_{\mathrm{E}} - I_{\mathrm{C}} = I_{\mathrm{E}p} + I_{\mathrm{B}n} - I_{\mathrm{C}p} \ .$$

Emitter current I_E: the electrons emitted to the base I_{En} + the holes diffused from the base I_{Ep}.

Collector current I_C: the electrons reached the base-collector junction I_{Cn} + the thermally generated holes I_{Cp} .

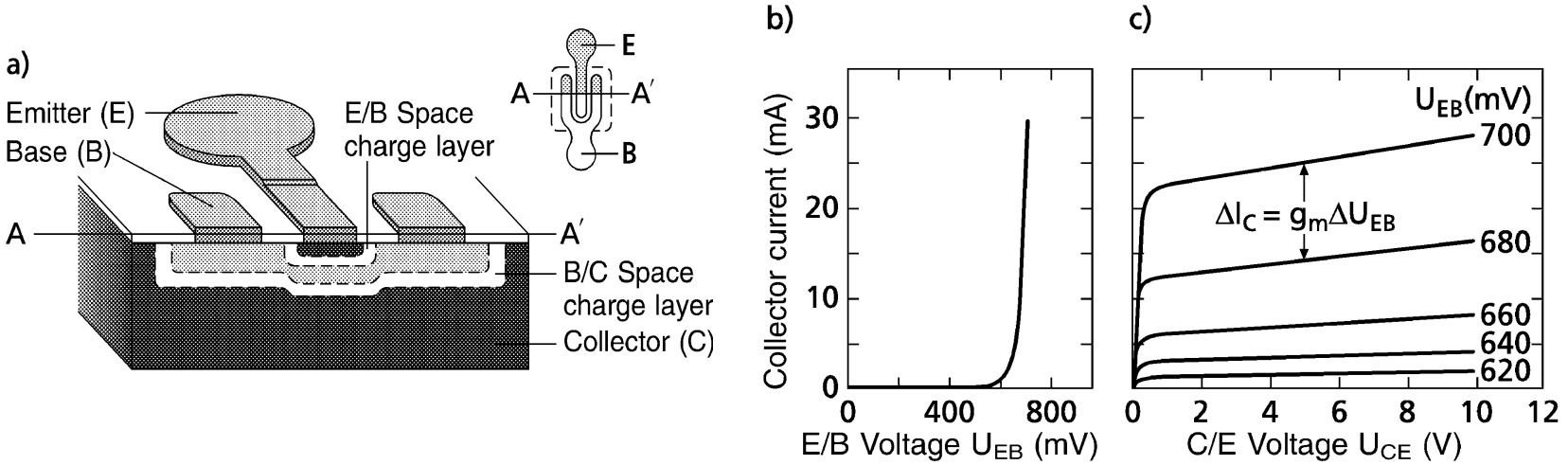
Base current I_B: the holes diffused into the emitter I_{Ep} + the portion I_{Bn} due to the recombination of holes with electrons + the thermally generated holes I_{Cp} from the collector region.

The collector current I_C is thus essentially controlled by the base current I_B .

Since the base current I_B is typically below 1% of the collector current I_C the collector current, which follows

$$I_C \approx I_E \approx A \frac{eD_n n_B}{L_n} [\exp(eU_{EB}/kT) - 1] .$$

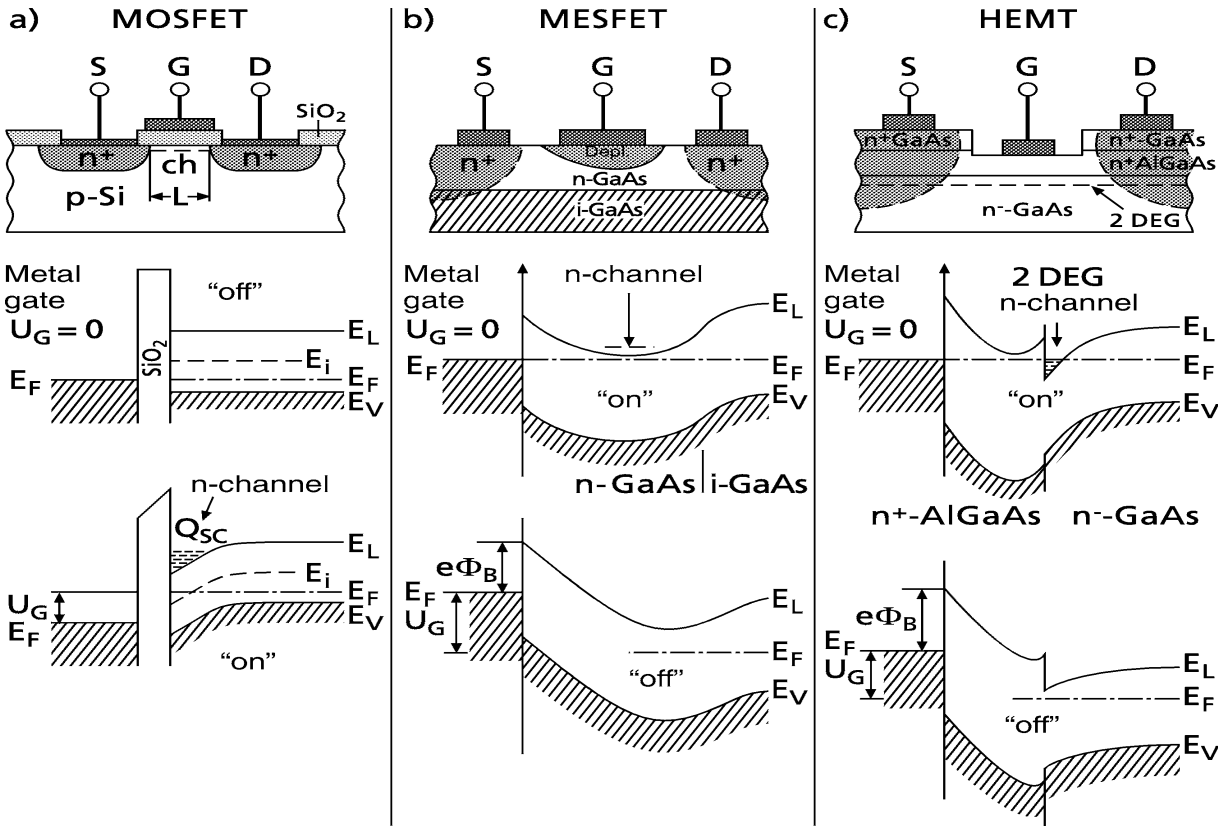
Here, A is the current-carrying area of the transistor, D_n the diffusion constant of electrons in the base region, n_B their equilibrium concentration therein, and L_n the electronic diffusion length in the *p*-doped base.



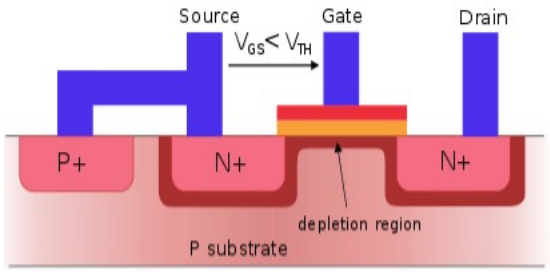
Significant voltage and power amplification can be achieved since the input circuit of the transistor contains the low-resistance emitter-base *pn* junction biased in the forward direction, while the output current I_C flows through the high resistance of the base-collector *pn* junction biased in reverse direction.

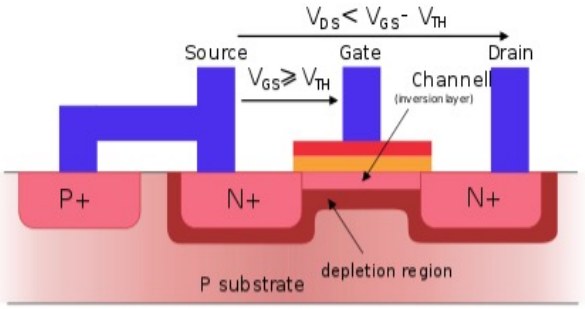
Field Effect Transistor (FET)

A field effect transistor (FET) is essentially a resistor that is controlled by an external bias voltage. It consists of a current channel with two contacts, the source and the drain, and a third contact, the gate, which is separated from the current channel by an insulating barrier.

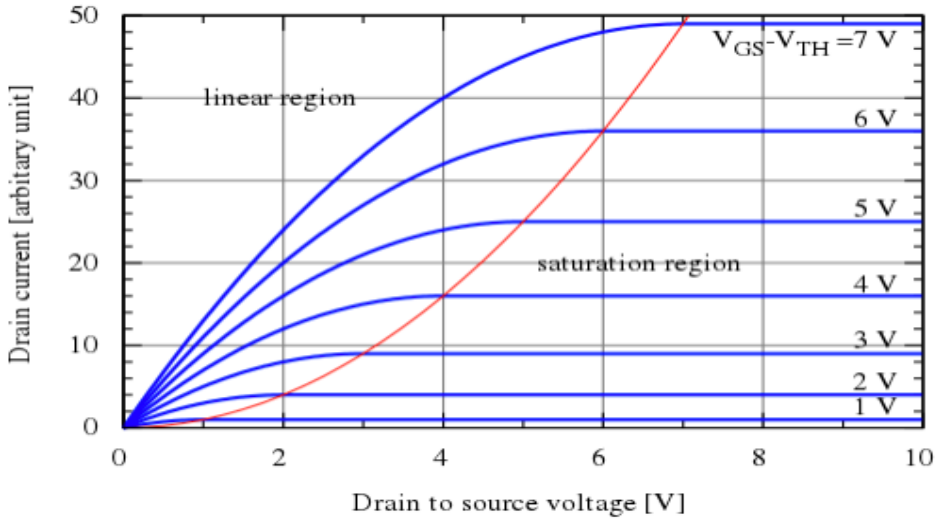


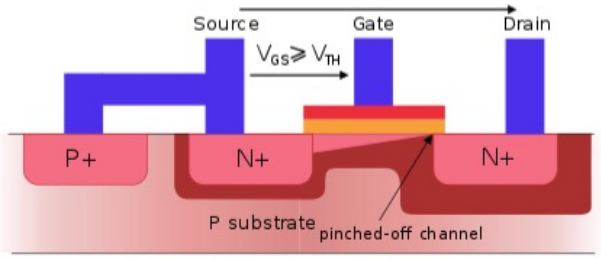
Operation of nMOSFET



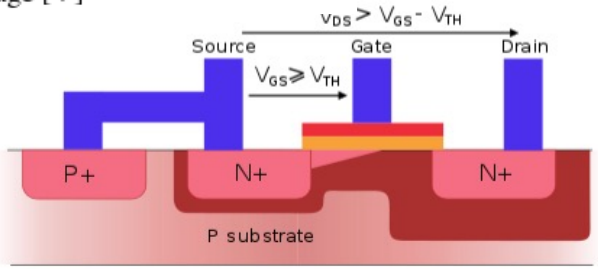


Linear operating region (ohmic mode)



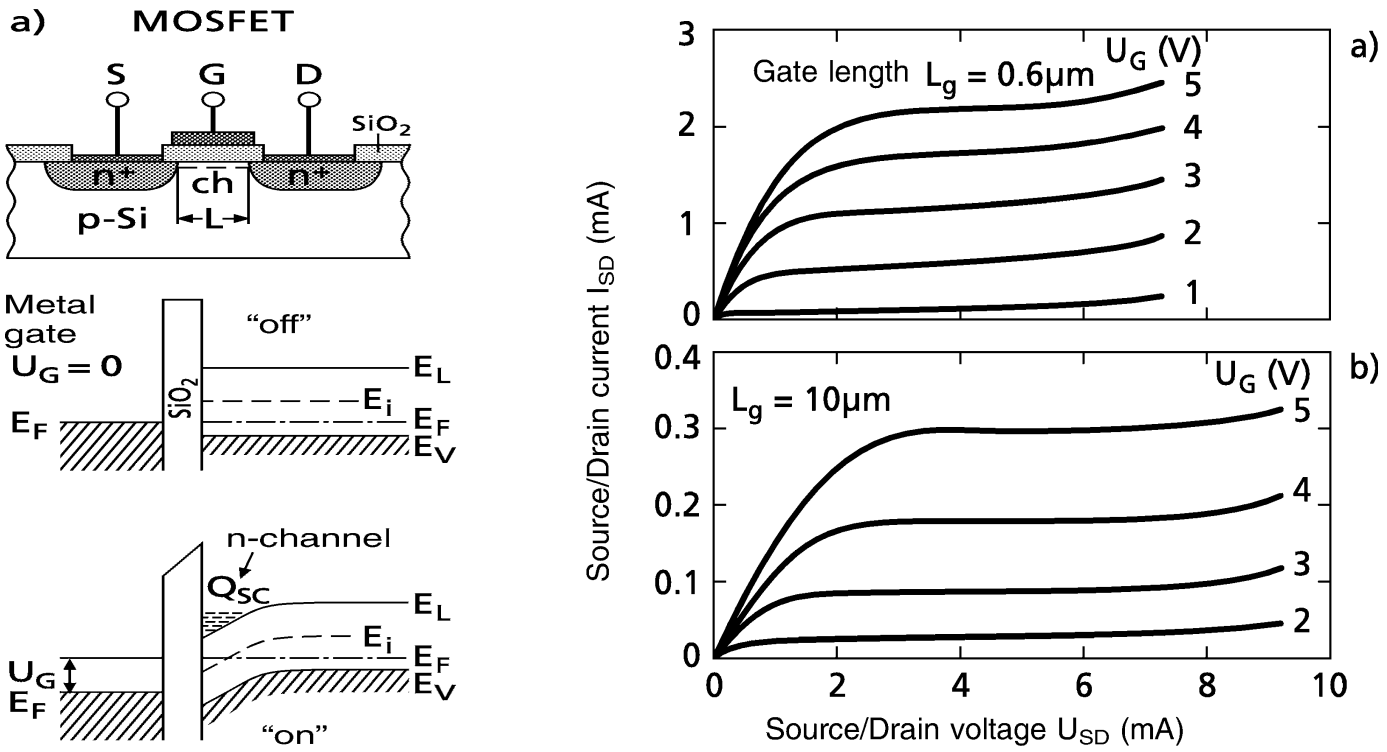


Saturation mode at point of pinch-off



Saturation mode

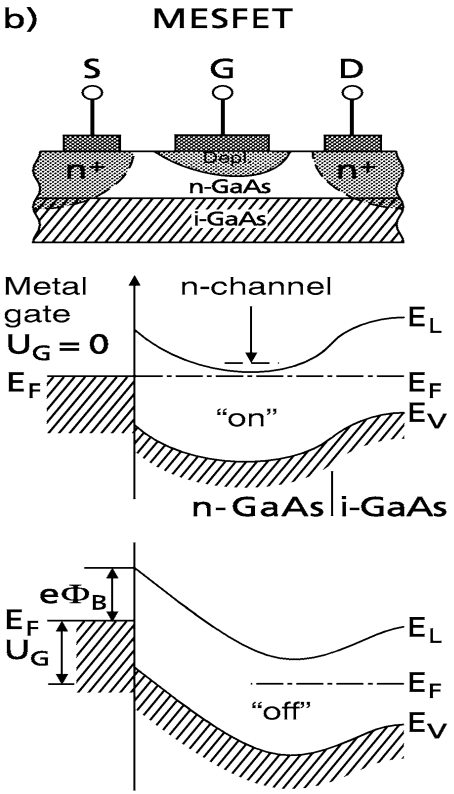
Metal-Oxide-Semiconductor FET (MOSFET)



A bias voltage applied to the gate gives rise to a large electric field across the gate oxide layer that in turn causes a band bending in the underlying semiconductor. For a sufficiently positive gate voltage U_G the initial *p*-region (at U_G=0) below the gate is inverted ($E_i < E_F$) to become an *n*-accumulation layer. The two source and drain n^+-wells are thence connected by a conductive *n*-accumulation channel. The transistor has switched from ``off'' to ``on''.

Metal-Semiconductor FET (MESFET)

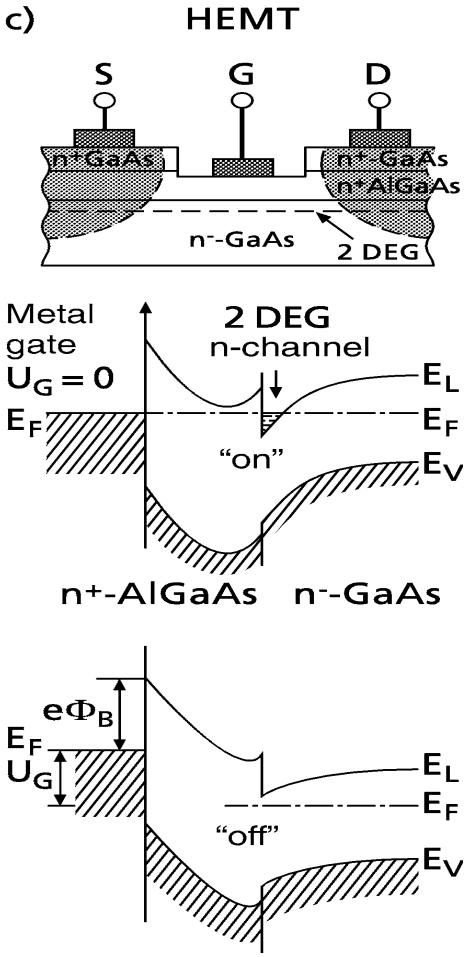
The performance of a MOSFET depends essentially on the perfection of the gate SiO_2 layer and its interface to the underlying Si channel. III-V semiconductors such as GaAs do not form perfect oxide overlayers. The MESFET (metal semiconductor field effect transistor) is therefore the most common FET on III-V compounds.



A highly resistive, intrinsic (i) or semi-insulating GaAs (achieved by doping with Cr) with the Fermi level E_F located near the mid-gap position serves as the substrate on which a thin, crystalline *n*-doped GaAs layer is epitaxially deposited or generated by ion implantation. Since in the metal/GaAs junction the Fermi level is pinned near mid-gap at the interface, the electronic bands are bent upwards with respect to the conducting *n*-channel. The metallic gate is thus separated from the *n*-channel by a depletion region of high resistance.

High Electron Mobility Transistor (HEMT)

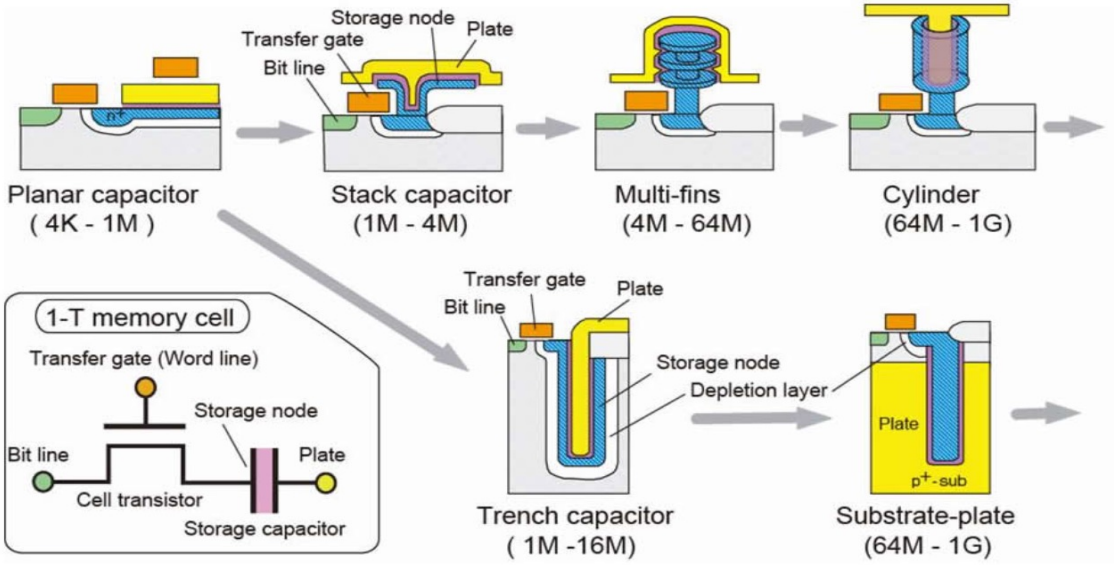
The HEMT is derived from the MESFET concept, which is extremely fast (20-300GHz) and exhibits very low noise. Its major fields of application are therefore radar and satellite communication.



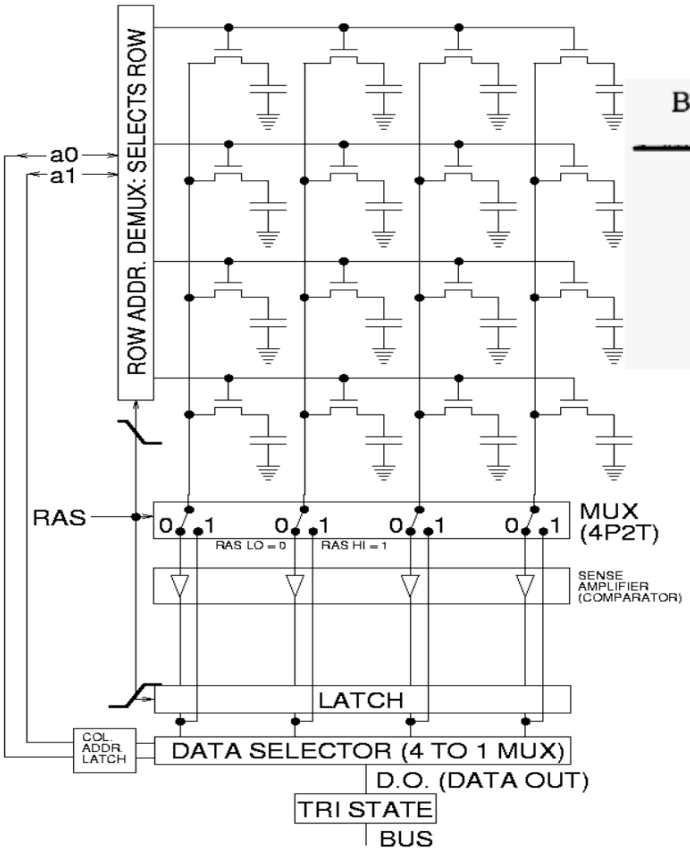
As a ``heterostructure MESFET'' the HEMT is switched by a Schottky barrier, separating the metal gate contact electrically from the conducting channel. The conduction channel, however, is a modulation-doped 2DEG at the interface. Characteristic figures of merit for a FET are its transconductance $g = \partial I_D / \partial U_G$ and the gate capacity C_g. The ratio between transconductance and gate capacity determines the cut-off frequency $f_{max} = g/C_g$. Fast, high-performance HEMTs in the material system GaInAs/InP reach transconductances g of about 600 mS/mm and upper cut-off frequencies f_{max}~300 GHz at channel lengths of 0.1 µm.

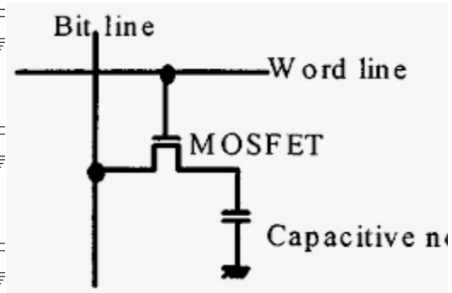
Dynamic Random-Access Memory (DRAM)

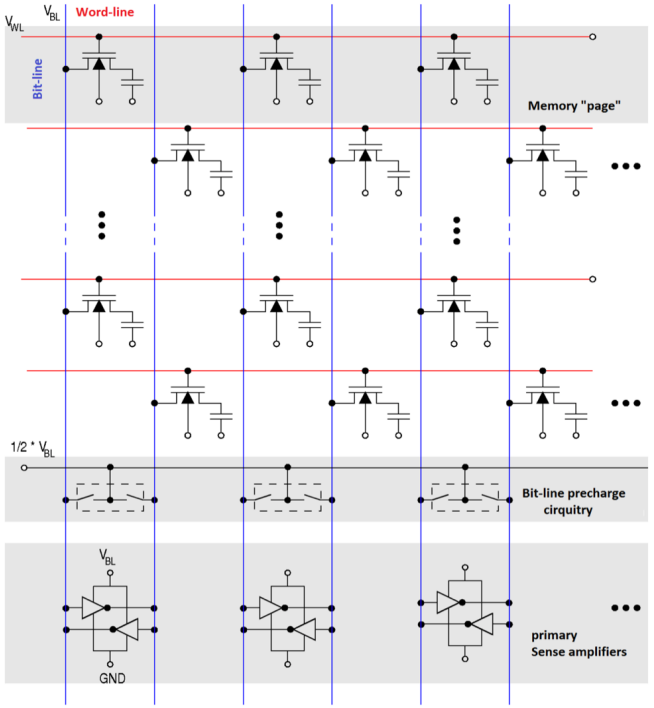
Dynamic random-access memory (dynamic RAM or DRAM) is a type of random-access semiconductor memory that stores each bit of data in a memory cell, usually consisting of a tiny capacitor and a transistor, both typically based on metal-oxide-semiconductor (MOS) technology. The capacitor can either be charged or discharged; these two states are taken to represent the two values of a bit, conventionally called 0 and 1. To prevent the leaking of electric charge on the capacitors, DRAM requires an external memory refresh circuit which periodically (every 64 ms) rewrites the data in the capacitors, restoring them to their original charge.



Operation of DRAM Reading



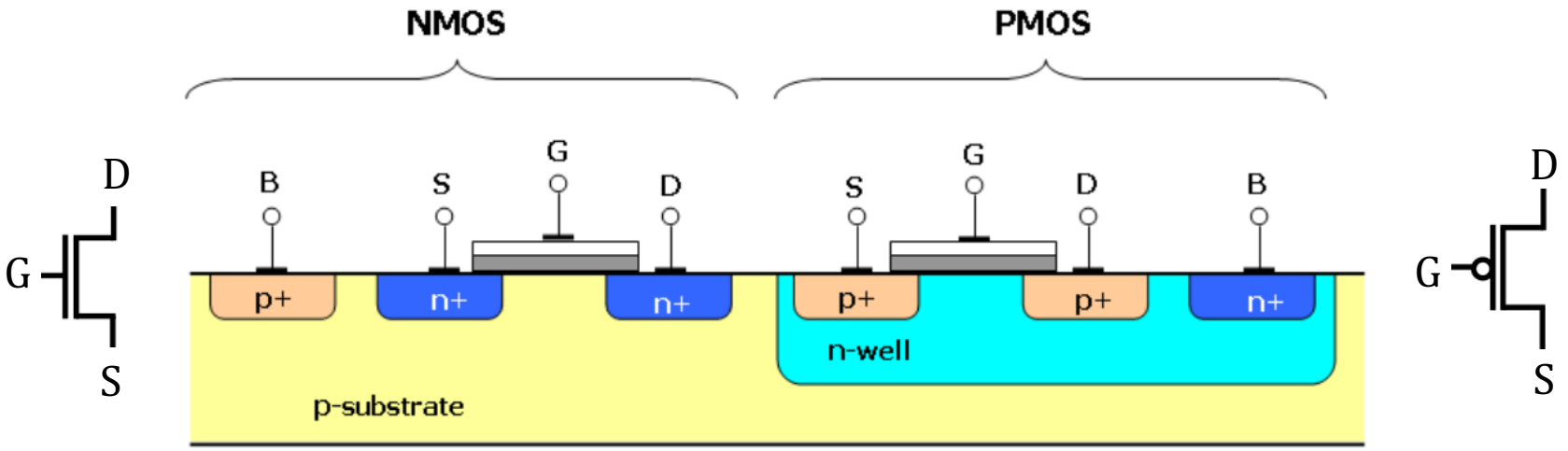




1. The sense amplifiers are disconnected.
2. The bit-lines are pre-charged to exactly in-between voltages.
3. The pre-charge circuit is switched off.
4. The desired row's word-line is driven to connect a cell's capacitor to its bit-line.
5. The sense amplifiers are now connected to the bit-lines pairs.
6. All storage cells in the open row are sensed simultaneously.
7. While reading is occurring, current is flowing back to recharge the cells.
8. When done with reading, the word-line is switched off to disconnect from the bit-lines.

Complementary MOS (CMOS)

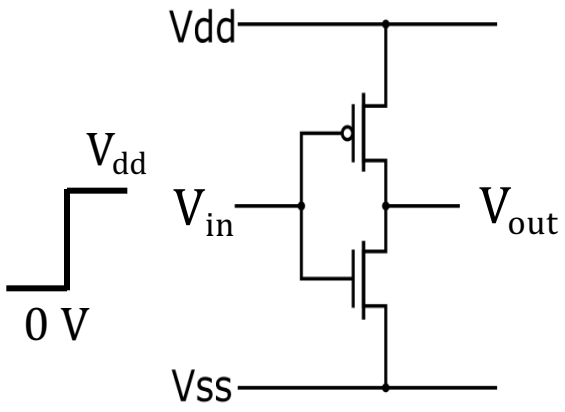
Nearly all advanced silicon ICs consist of combinations of *p*- and *n*-channel MOSFETs (CMOS = complementary MOS). Since one transistor of the MOSFET pair is always off, the combination of both transistor types allows the realization of logic gates that carry current only within the switching period between two bit-operations. CMOS circuits, therefore, consume less power, which is of significant advantage in the ever-increasing integration density.



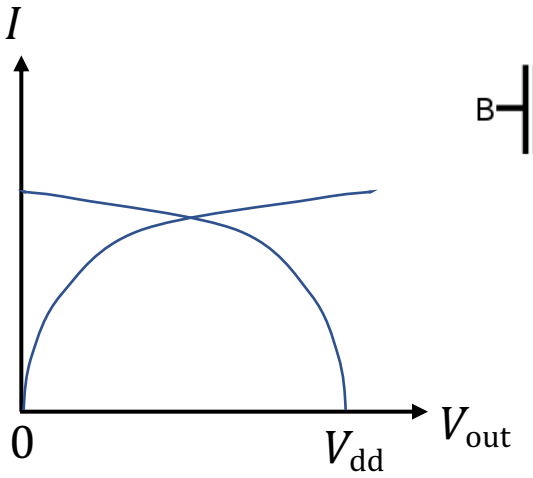
Gate Voltage	*n* channel	*p* channel
High	On	Off
Low	Off	On

CMOS in Logic Gates

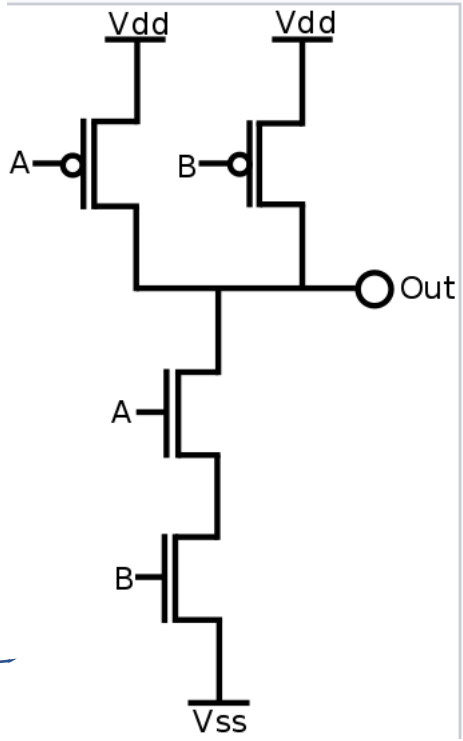
Inverter



In	Out
1	0
0	1



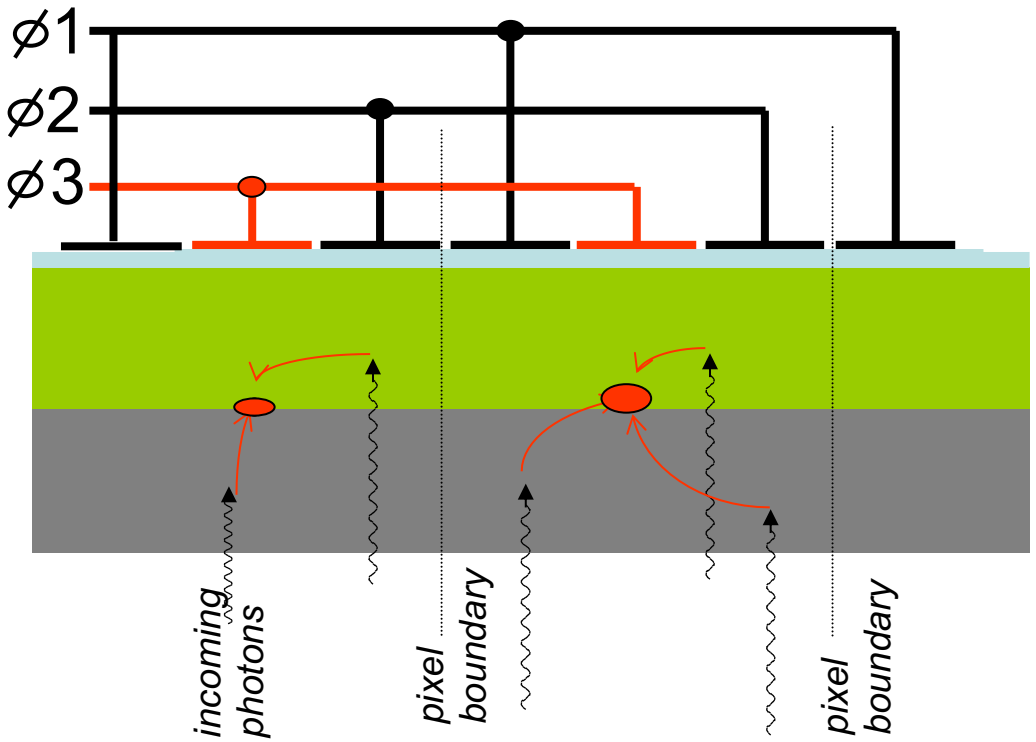
NAND gate

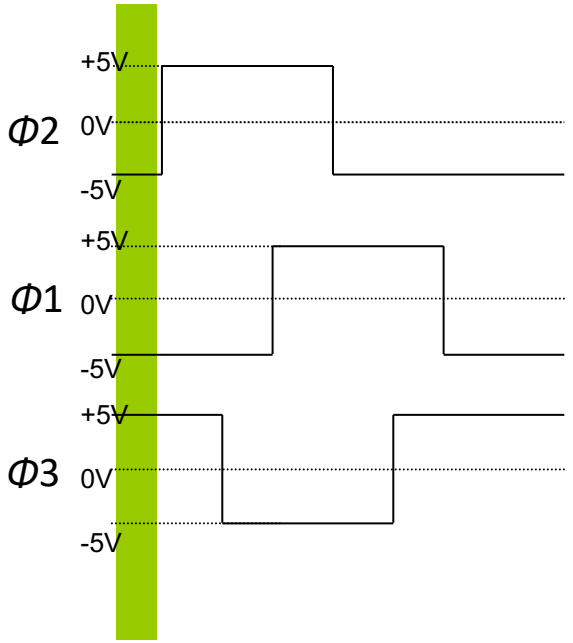


A	B	AND	NAND
1	1	1	0
1	0	0	1
0	1	0	1
0	0	0	1

Charge-Coupled Devices

A charge-coupled device (CCD) is an integrated circuit containing an array of linked, or coupled, capacitors. Under the control of an external circuit, each capacitor can transfer its electric charge to a neighboring capacitor. CCD sensors are a major technology used in digital imaging.





Charge packet

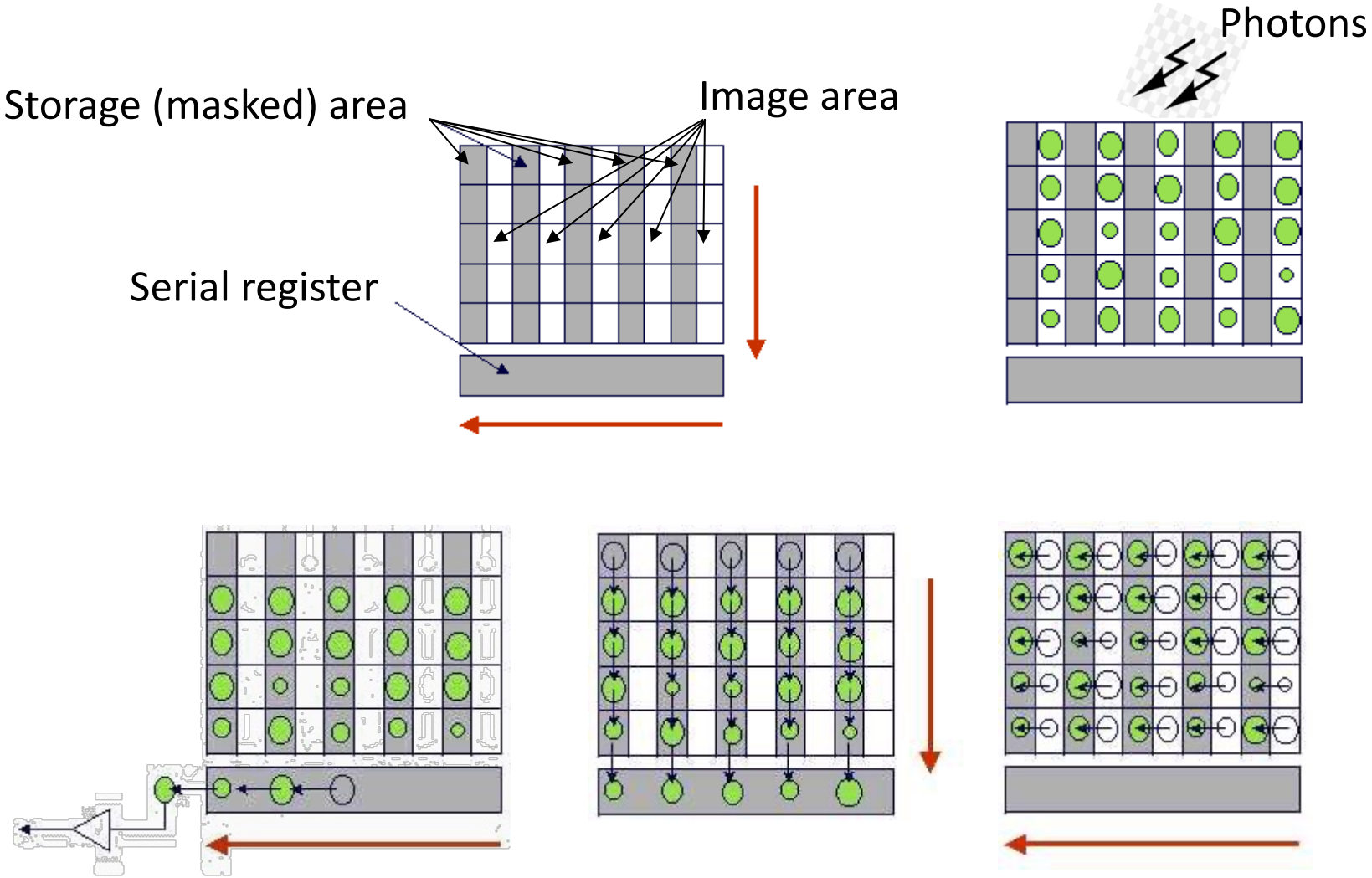
n-type silicon

p-type silicon

Electrode Structure

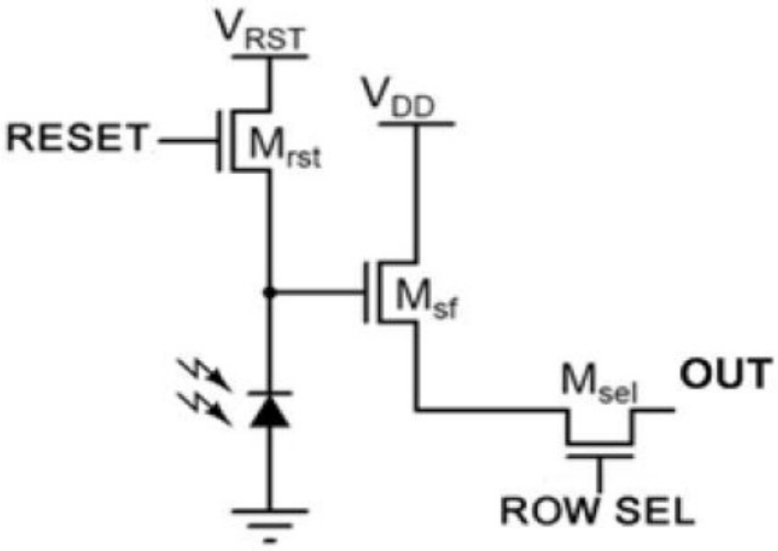
SiO2 Insulating layer

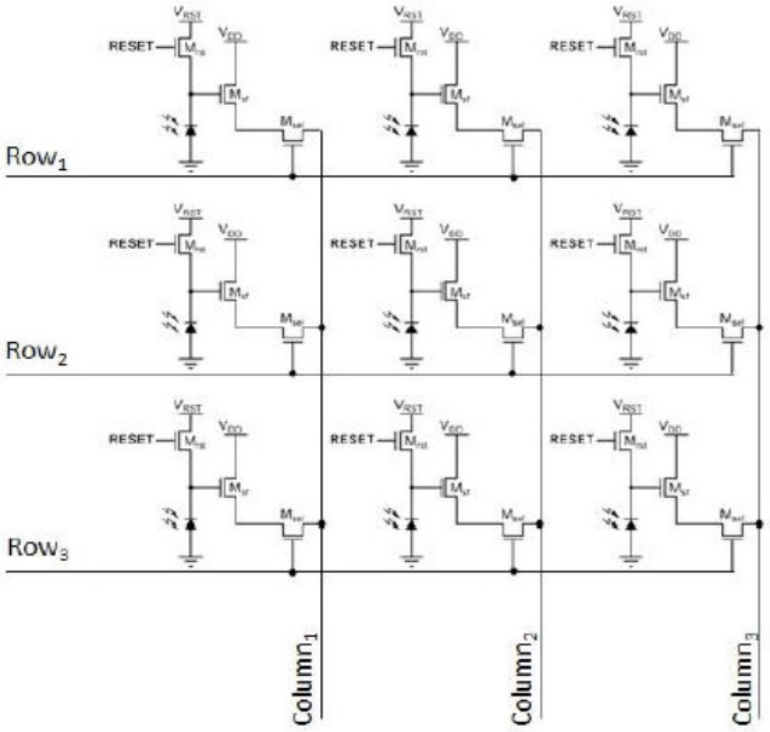
Interline-Transfer CCD



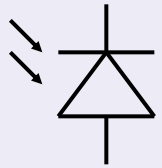
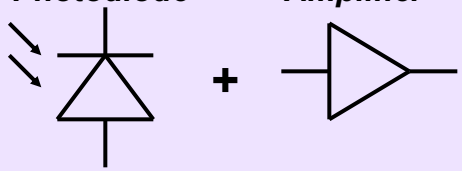
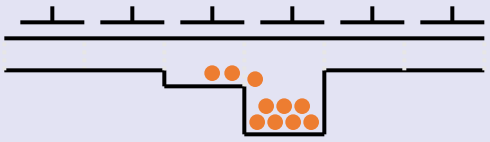
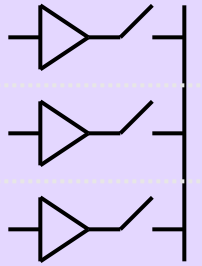
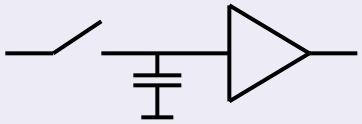
CMOS Image Sensors

A CMOS image sensor (CIS) is an image sensor where each pixel sensor unit cell has a photodetector (typically a pinned photodiode) and one or more active CMOS transistors. CMOS sensors emerged as an alternative to charge-coupled device (CCD) image sensors and eventually outsold them by the mid-2000s.





General CMOS Detector Concept

	CCD Approach	CMOS Approach
Pixel	*Photodiode*  **Charge generation & charge integration**	*Photodiode* *Amplifier*  **Charge generation, charge integration & charge-to-voltage conversion**
Array Readout	 **Charge transfer from pixel to pixel**	 **Multiplexing of pixel voltages: Successively connect amplifiers to common bus**
Sensor Output	 **Output amplifier performs charge-to-voltage conversion**	**Various options possible:** **- no further circuitry (analog out)** **- add. amplifiers (analog output)** **- A/D conversion (digital output)**

Semiconductor Laser

The interaction of the electromagnetic light field with electronic excitations in the semiconductor involves three different processes: (i) an absorption of light quanta by an excitation of electrons from the valence band into the conduction band, or between electronic defect levels, (ii) an inverse stimulated emission, in which the incident photon stimulates an electron to be de-excited from a state of higher energy, and (iii) the spontaneous emission of light. All three types of interactions are used in photoelectric devices: For photodetectors and solar cells the absorption process is essential. Light emitting diodes (LED) utilize spontaneous emission, while lasers are based on stimulated coherent emission. In all three processes energy conservation is obeyed $\hbar\omega = E_2 - E_1$

Under stationary illumination with light the transition rates must compensate each other. With A_{21} as the probability for spontaneous emission and B_{21} and B_{12} the probabilities for stimulated emission and absorption, respectively, then

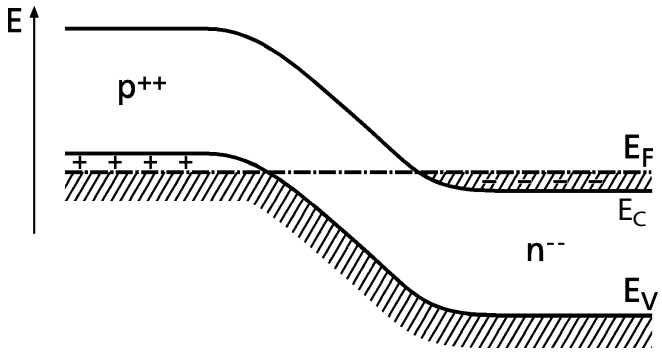
$$\dot{n}_2 = -B_{21} n_2 |\mathscr{E}(\omega)|^2 - A_{21} n_2 + B_{12} n_1 |\mathscr{E}(\omega)|^2 = -\dot{n}_1 = 0$$

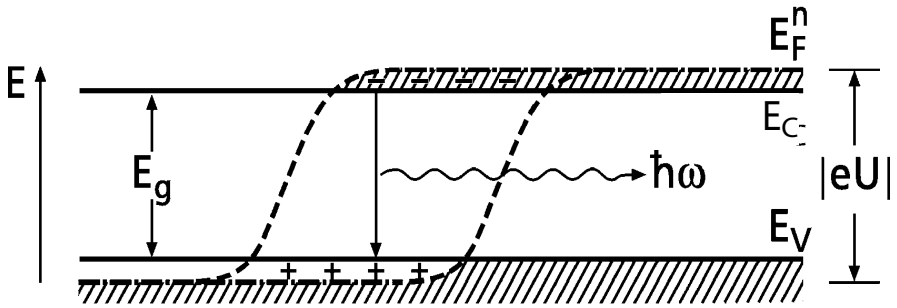
Neglecting spontaneous emission and using the equality $B_{21} = B_{12}$ one obtains the following laser condition:

$$\dot{n}_2 = -(n_2 - n_1) B_{21} |\mathscr{E}(\omega)|^2 < 0 \quad \text{i.e.} \quad n_2 > n_1 \ .$$

This condition describes the inversion of the occupation statistics.

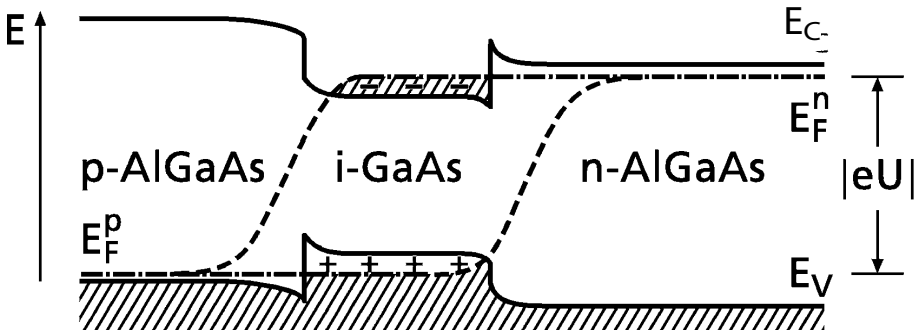
Population inversion is achieved by ``pumping'' into the excited state. In a semiconductor laser a convenient method for pumping is to bias a *p-n* junction in the forward direction and to flood the space-charge region with non-equilibrium electrons and holes.





For laser operation, the quasi-Fermi levels in the n- and p-doped region must be separated from each other by more than the band gap energy E_g.

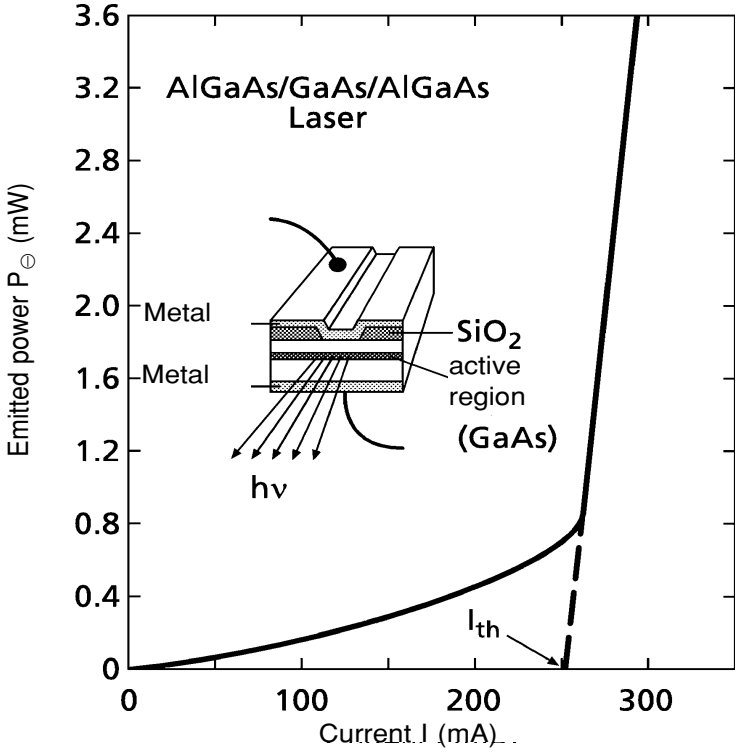
Double Heterostructure Laser



A lowly doped or intrinsic i-GaAs layer is inserted between two low *p*- and *n*-doped wide band gap AlGaAs regions.

In spite of the low doping in the AlGaAs regions, the quasi-Fermi levels are located within the conduction and valence band of the active i-GaAs layer under strong forward bias. Population inversion is thus easily achieved. The active i-GaAs region is flooded with non-equilibrium electrons and holes, which are confined to the active region by the conduction and valence band discontinuities, respectively. This effect, which causes enhanced light emission is called ``electrical confinement''. Additionally, this structure provides an ``optical confinement'', since the refractive index of the active GaAs region exceeds that of the adjacent AlGaAs layers. The light originating from stimulated emission in the GaAs is therefore totally reflected at the AlGaAs layers.

An important figure of merit of a semiconductor laser is a threshold current as low as possible arising from low losses in the resonator. Population inversion is easily achieved in the double heterostructure laser comparing to the highly doped, degenerate semiconductors of the high current densities.



A further figure of merit of a semiconductor laser is the critical temperature T_0, which should be as high as possible because the threshold current I_{th} depends exponentially on the working temperature T as $I_{th} \propto \exp(T/T_0)$, so that the laser shows a slow increase of the threshold current with working temperature T. Typical values of T_0 are close to 100°C for double hetero-structure lasers.

Further improvement of the laser performance is achieved by incorporating into the active zone even thinner layers of a semiconductor with a lower band gap than that of GaAs, which can render the *quantum-well lasers*.

Problems

1. (a) Sketch the energy band diagram for a *p*-*n*-*p* transistor at equilibrium and under the normal active mode of operation.

(b) Sketch a schematic diagram to represent the transistor and indicate all current components.

(c) Write down the relative equations for the currents of emitter, base and collector.

2. (a) Derive expressions of the bulk charge, surface potential, and surface field to show their dependence on the substrate doping density N_a at strong inversion.

(b) Plot the bulk charge, surface potential, and surface field as function of N_a from 10^{14} to 10^{18} cm^{-3}.

3. Try to realize the NOR logic gate in CMOS circuits and use the truth table to confirm the operations.