CHEMICAL PROCESSING OF ELECTRONS AND HOLES

TIMOTHY J. ANDERSON
University of Florida
Gainesville, FL 32611

IN THE EARLY twentieth century the four engineering disciplines of chemical, civil, electrical, and mechanical were founded. In a very short period of time, each discipline evolved along relatively independent paths to produce, in part, quite different curricular contents. These differences are best exemplified by chemical and electrical engineering. Chemical engineering has developed into the most general of the founding engineering disciplines and is characterized by an isolated and rigid curriculum with an emphasis on the engineering sciences, particularly those which involve chemical change. Textbooks in our discipline tend to experience longevity, time as a variable is not emphasized, and mature technologies are often graduated (*e.g.*, nuclear engineering, environmental engineering, petroleum engineering, polymer engineering, metallurgical engineering, and biomedical engineering). In contrast, the electrical engineering curriculum is more option oriented, reflecting a historical retention of developed technologies (*e.g.*, power engineering, solid-state electronics, computer architecture, optical engineering). The increased technological content of the curriculum also translates to short textbook lifetimes.

Mother Nature, however, is totally unaware of our somewhat arbitrary partitioning of her behavioral patterns. As a result, the foundations of chemical engineering prepare the student to understand a variety of topics included in other curricula once the terminology and nomenclature are translated. One such example is the operation of solid state electronic devices in an integrated circuit. Presented below is a synopsis of four lectures given in an elective senior-level electronic materials processing course which introduces the student to solid state electronics. The terminology of chemical engineering is used primarily with the equivalent electrical engineering terminology contained in brackets.

Tim Anderson joined the faculty at the University of Florida in 1978 after receiving degrees in chemical engineering from Iowa State University (BS) and the University of California, Berkeley (MS,PhD). He has been a visiting scientist at RADC Air Force laboratory and a Fullbright Scholar at the University of Grenoble. He has an active research program in bulk crystal growth and epitaxy of compound semiconductors and ceramic superconductors.

© Copyright ChE Division ASEE 1990

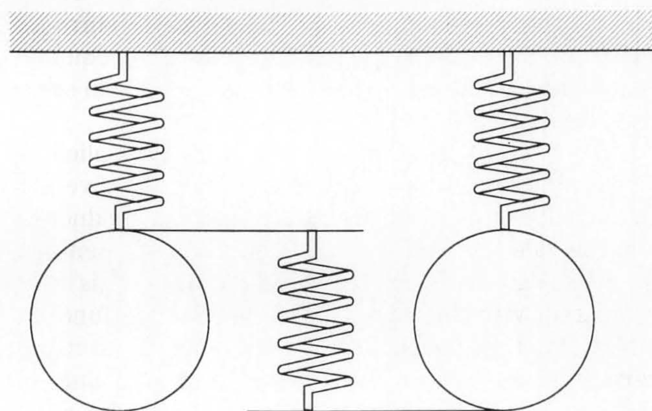


FIGURE 1. Mechanical analogy of electron energy state splitting.

ELECTRONS IN SOLIDS

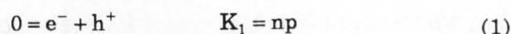
In order to understand the behavior of electrons in the solid state, the student must first appreciate the concept of electron energy states. This is introduced by comparing the free electron, for which all energy values are accessible; the hydrogen atom, for which only discrete states exist; and a large collection of H atoms in the solid state, for which the collection of electron states becomes so closely spaced in energy that we speak of bands of energy states (*e.g.*, 1s band, 2s band). The splitting of electron energy states when atoms are brought together is illustrated by recalling from quantum mechanics the bonding and anti-bonding states between two hydrogen atoms and also by the mechanical analogy shown in Figure 1. This figure illustrates two identical balls suspended from two

By their senior year, students have already received the foundations of chemical engineering. One of the objectives of these lectures is to convince seniors that the digital integrated circuit is nothing more than a chemical processing plant.

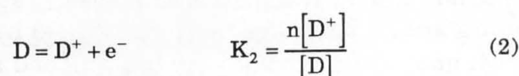
identical springs. Neglecting frictional losses, an initial displacement of each uncoupled ball will result in a single natural vibrational frequency. If the two balls are permitted to interact through the third spring shown in this figure, two natural vibrational states are possible for a given displacement of each ball: a low frequency state when each ball is initially displaced in the same direction and a higher frequency state when displaced by the same amount but in opposite directions.

The concept of energy bands is next applied to semiconductors by showing a plot of energy versus the density of electron states for both the conduction and the valence bands, pointing out the gap in energy for which no intrinsic states exist. This plot is then compared with the Fermi-Dirac distribution function which gives the probability of finding an electron at a certain energy value. Integration of the product of the density of states and the probability of a state being occupied for energies above the conduction band minimum gives the number of electrons in the conduction band. These electrons are essentially free because the vast majority of conduction band states are not occupied. A similar integration over unoccupied states in the valence band gives the concentration of holes (empty electron states). The Fermi-Dirac distribution contains a parameter called the Fermi energy that is a function of temperature, pressure, and concentration. The Fermi energy is equivalent to the electrochemical potential of electrons, a quantity that is understandable to chemical engineers.

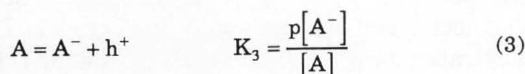
The process of doping a semiconductor is next described by illustrating the incorporation of B and P in Si. The group III dopant B introduces an electron state with energy level located just above the highest valence band energy, while the group V dopant P introduces a state just below the conduction band minimum energy. A doped semiconductor material under equilibrium conditions is a good example of chemical equilibrium. Consider the following three reactions at equilibrium:



$$0 = e^- + h^+ \qquad K_1 = np \tag{1}$$



$$D = D^+ + e^- \qquad K_2 = \frac{n[D^+]}{[D]} \tag{2}$$



$$A = A^- + h^+ \qquad K_3 = \frac{p[A^-]}{[A]} \tag{3}$$

These three chemical reactions involve the chemical species free electron (e^-), free hole (h^+) neutral donor (D), ionized donor (D^+), neutral acceptor (A), and ionized acceptor (A^-) with corresponding equilibrium concentrations of n, p, [D], $[D^+]$, [A], and $[A^-]$. The chemical species 0 represents an electron combined with a hole in the valence band (normal state) with a large, nearly constant concentration.

Equations (1-3) are further constrained by the condition of charge neutrality

$$p + [D^+] = n + [A^-] \tag{4}$$

Since the electron and hole are highly mobile in the semiconductor, even at room temperature, the material reaches equilibrium very rapidly. These four equations contain six concentration variables. The total donor dose, $N_D = [D^+] + [D]$, and acceptor dose, $N_A = [A^-] + [A]$, however, are usually specified. Solution of Equations (1-4) in terms of N_D and N_A involves finding the roots of a 4th order polynomial in the variable n. If the donor and acceptor ionization energies are small compared to the bandgap energy, E_g, and comparable or smaller than RT, then reactions (2) and (3) are shifted to right and

$$N_D \approx [D^+] \qquad N_A \approx [A^-]$$

This simplification leads to a quadratic equation with meaningful root

$$n = \frac{1}{2}\left\{N_D - N_A + \left[(N_D - N_A)^2 + 4K_1\right]^{\frac{1}{2}}\right\} \tag{5}$$

This example is easily understood by the senior chemical engineering student and is translated into electrical engineering terminology according to the relationships

$$K_1 = \exp(\Delta S^\circ / R)\ \exp(-\Delta H^\circ / RT) = N_C N_V\ \exp(-E_g / RT) = n_i^2 \tag{6}$$

In this equation N_c and N_v are conduction and valence band effective density of states

$$N_i = 2\left(2\pi m_i^e kT / h^2\right)^{\frac{3}{2}}$$

and

m_i^e = effective mass of an electron or hole

and are related to the entropy of reaction (1). The quantity n_i is the intrinsic carrier concentration and represents the electron or hole concentration in the undoped semiconductor ($n = p$ in the intrinsic material since a hole is created for every electron promoted to the conduction band).

SEMICONDUCTORS UNDER NON-EQUILIBRIUM CONDITIONS

As in operational chemical plants, functioning semiconductor devices operate under non-equilibrium conditions by the action of external influences (*e.g.*, electric field, magnetic field, optical excitation, electron bombardment). The basic equations that describe transport of electrons and holes include species material balances, a statement of species flux in terms of available potential gradients, and Maxwell's equations since these two species are charged. The principles of basic device operation can be illustrated with a simplified set of these equations. Considering only low electric fields and one-dimensional transport in the absence of magnetic fields for an ideal (dilute) solution of electrons and holes, the species material balances [continuity equations] are:

$$\frac{\partial n}{\partial t} = \frac{1}{q}\frac{\partial J_n}{\partial x} + R_n \quad \text{(electrons)} \tag{7}$$

and

$$\frac{\partial p}{\partial t} = \frac{-1}{q}\frac{\partial J_p}{\partial x} + R_p \quad \text{(holes)} \tag{8}$$

where

q = magnitude of electric charge
t = time
J_i = flux of positive charge [current density](charge/cm^2·s)
R_i = net rate of production of species i (number/cm^3·s) by homogeneous reaction [carrier recombination, optical excitation, impact ionization]

The charge fluxes [current densities] are given by

$$J_n = q\mu_n n\xi + qD_n\frac{\partial n}{\partial x} \tag{9}$$

and

$$J_p = q\mu_p p\xi - qD_p\frac{\partial p}{\partial x} \tag{10}$$

where

μ_i = mobility of species i (cm^2/V·s)
ξ = electric field (V/cm)
D_i = concentration independent diffusion coefficient of species i (cm^2/s)

Electrons and holes can migrate in response to both an electric field [drift] and concentration gradient [diffusion]. The most important Maxwell equation to device physics is the Poisson equation which relates the electric field gradient to the net charge distribution, ρ, according to

$$\frac{\partial \xi}{\partial x} = \frac{\rho}{\varepsilon} \tag{11}$$

where

ε = semiconductor permittivity (F/cm)

The senior chemical engineering student who has completed courses in transport phenomena and introductory physics can easily understand the significance of these equations; the only new term in the transport equations is migration due to an electric field.

A simple and useful example of these equations is illustrated in Figure 2. A uniformly n-type doped semiconductor slab is illuminated on one side with light (Figure 2a). The photon energy is chosen so that

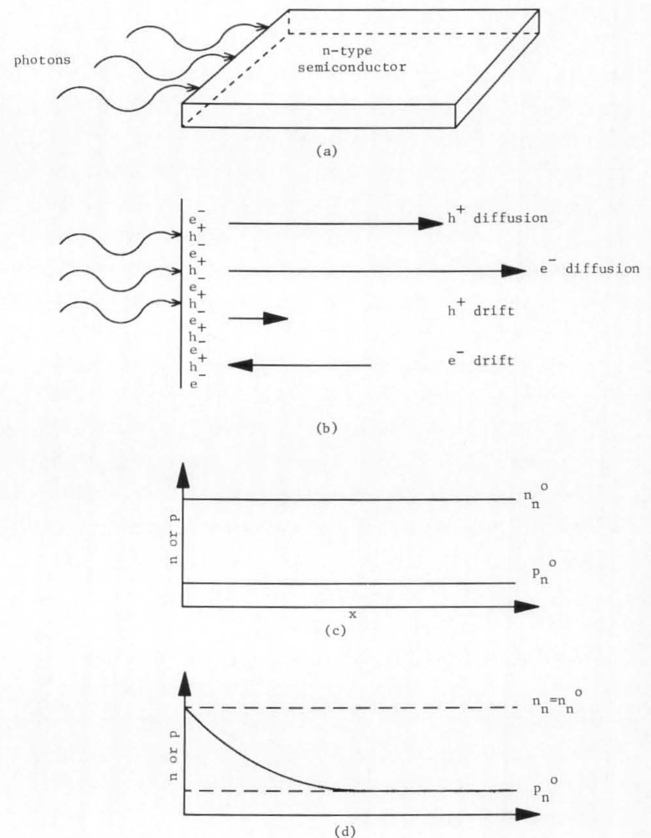


FIGURE 2. ***Surface absorption of photons in an n-type semiconductor; (a) schematic, (b) sketch of relative magnitude of hole and electron currents due to diffusion and drift, (c) carrier concentration at equilibrium (no illumination), (d) steady state carrier concentration with illumination.***

absorption creates electron-hole pairs near the surface only (photon energy greater than the bandgap energy and large absorption coefficient). Before illumination the slab is at equilibrium (Figure 2c) and the concentration of electrons [majority carrier] greatly exceeds the concentration of holes [minority carrier]; for example, Si at room temperature and doped at

$$n = 10^{17}\ \text{cm}^{-3} \qquad \text{gives} \qquad p_n^o = 2.1 \times 10^3\ \text{cm}^{-3}$$

Upon illumination electron/hole pairs are generated at the surface and are transported into the slab by diffusion. The diffusion coefficient of an electron, however, is normally greater than that for a hole (by a factor of 3 for Si) and a small electric field is established. The direction of the field is such that the electron flux is reduced and the hole flux is enhanced (Figure 2b). Realizing there is no net current in the slab ($J_n = J_p$) and the carrier concentration gradients are nearly identical at a specified position [electroneutrality approximation], it can be shown that the minority carrier transports almost exclusively by diffusion provided the quantity

$$\left|\left(\mu_p / \mu_n - 1\right) p_n / n_n\right| << 1$$

This condition is satisfied if the photon flux is not too large ($p_n \ll n_n$). The same conclusion is not found for the electron [majority carrier], since there is a large population of electrons to respond to the electric field (Figure 2b).

As the carriers diffuse into the slab, they attempt to return to their equilibrium concentrations through homogeneous reaction [recombination]. The steady state minority carrier concentration profile can be determined by solution of Equation (10), with only the diffusion term, and Equation (8):

$$0 = D_p \frac{\partial^2 p_n}{\partial x^2} + R_p \tag{12}$$

The applicable boundary conditions are

$p_n(x=0)$ equal to a constant (due to steady illumination)

and

$p_n(x=\infty) = p_n^o$ (semi-infinite slab)

An expression for the net rate of generation of holes, R_p, is required. The simplest homogeneous reaction mechanism is direct recombination of a conduction band electron with a valence band hole [band to band recombination]

$$0 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} e^- + h^+ \tag{13}$$

The rate of production of h^+ by this reversible reaction is

$$R_p = k_1 - k_{-1} n_n p_n \tag{14}$$

The rate of the forward reaction is pseudo-zero order since the concentration of electrons in the valence band and holes in the conduction band are not significantly changed by the reaction at low doping levels. The rate constant k_1 can be eliminated in favor of the known equilibrium constant

$$K_1 = \frac{k_1}{k_{-1}} = n_n^o\ p_n^o$$

to give

$$R_p = k_{-1}\left(n_n^o\, p_n^o - n_n p_n\right) \approx k_{-1}\, n_n^o\left(p_n^o - p_n\right) \tag{15}$$

The rate at which holes disappear by this particular mechanism is thus pseudo-first order since the equilibrium majority carrier concentration is barely disturbed at low illumination. In device physics texts the quantity

$$\frac{1}{k_{-1}\, n_n^o}$$

is termed the minority carrier lifetime, τ_p. The solution to Equation (12) with the recombination rate given above is:

$$\frac{p_n(x) - p_n^o}{p_n(x=0) - p_n^o} = e^{-x/(D_p \tau_p)^{1/2}} \tag{16}$$

The term $(D_p\ \tau_p)^{1/2}$ is called the diffusion length for obvious reasons.

In chemical engineering terms, the above problem is simply one of diffusion with first-order homogeneous reaction into a semi-infinite slab from a constant composition source. The simplicity of this example allows students to make the connection between semiconductor physics and their own background in chemical engineering. The problem is also useful since it is the basis for understanding minority carrier injection necessary to describe the operation of p-n junction devices (diode, bipolar transistor). Useful homework problems include analysis of: carrier concentration decay with steady and uniform photoexcitation (batch reactor), transient and steady-state transport of carriers generated by localized illumination with/without an electric field (transient and steady-state dispersion of a line source with plug flow/no flow), recombination rates for materials having mid-gap states (homogeneous catalysis), and surface recombination with uniform and steady illumination (homogeneous and heterogeneous reaction in a semi-infinite stagnant liquid).

p-n JUNCTIONS

The p-n junction is the basic building block of many solid state devices, including the junction diode and bipolar transistor. The lecture material begins with the equilibrium p-n junction and then examines the junction under non-equilibrium conditions with both a positive and a negative applied potential. Finally, the behavior of two p-n junctions (bipolar transistor) under non-equilibrium conditions is described.

When a piece of n-type semiconductor is metallurgically joined to a piece of p-type semiconductor, the hole and highly mobile electron species diffuse in directions of lower chemical potential. Electrons, in excess in the n-type material, will diffuse into the p-type material, where their concentration is extremely small, and holes will diffuse in the opposite direction. If immobile donor and acceptor ions were not present, this process would continue until the electron and hole chemical potentials were the same in both materials. As diffusion occurs, however, a positive space charge density in the n-type material and a negative charge density in the p-type material are "uncovered." The resulting charge distribution produces a diffusion potential [built-in voltage] that opposes further diffusion. When the spatial variation of the chemical potential is just balanced by the variation in electric potential, the joined semiconductors are in equilibrium and the electrochemical potentials [Fermi level] are constant. The discussion of the equilibrium p-n junction continues with a numerical illustration of a Si abrupt junction in which the charge distribution is approximated as a step function [depletion approximation]. The details of this example are given in most device textbooks [1-6] and it invokes our first use of Poisson's equation.

The steady state operation of a p-n junction with an applied external voltage is treated next. The equilibrium junction described above is dynamic, representing a balance between drift and diffusion currents. To understand the operation of the junction with a positive voltage applied to the n-type material [reverse bias], different sources of carriers are examined. If I place myself at the metallurgical junction and count the electrons and holes which cross the junction, I will see three sources of carriers. The first source is homogeneous reaction in the depletion region [generation]. The rate of electron production equals that of hole production as given by Eq. (15) $pn \ll K_1$ since carriers are assumed to be depleted in this region. The electric field in the depletion zone will sweep an equal number of generated electrons and holes in opposite directions towards material of the same type. Therefore, at the metallurgical plane I can count the holes coming from the n-type material side of the depletion region and the electrons originating from the p-type material side. The total charge crossing the plane is equal to the rate of hole production in the entire depletion region, or equivalent electron production, times the total width of the depletion region, W

$$J = q k_{-1} n_i^2 W \tag{17}$$

The second source of carriers crossing the junction plane are produced by diffusion of minority carriers to the boundary between the neutral and depletion regions, where they are swept across the depletion zone by the electric field. This problem is similar to the surface illumination problem treated earlier, except that the minority carrier concentration is reduced [depleted] at the boundary instead of elevated to a constant value by the photon absorption. The flux of minority carriers at the edges of the depletion region can be determined from Fick's first law of diffusion and Equation (16) with $p_n(x=0) = 0$. The currents arising from extraction of minority carriers are given by

$$J_p = q p_n^o \left(D_p / \tau_p\right)^{1/2} \tag{18}$$

and

$$J_n = q n_p^o \left(D_n / \tau_n\right)^{1/2} \tag{19}$$

The currents given by Equations (17-19) have the same sign and at equilibrium are just balanced by the third source of carriers; majority carriers from the neutral regions capable of overcoming the built-in potential. Application of a positive voltage, V_R, to the n-type material increases the potential which majority carriers must overcome, thus decreasing the majority carrier diffusion current (proportional to $\exp[qV_R/kT]$). The minority carrier extraction currents given by Equations (18) and (19), however, are independent of voltage while the current due to homogeneous reaction (Equation (17)) actually increases since the depletion region widens with increasing V_R, enlarging the reactor volume. Therefore, with increasing V_R, the majority current rapidly becomes small and the reverse bias current is given by the sum of currents in Equations (16-18). This current is small since the quantities

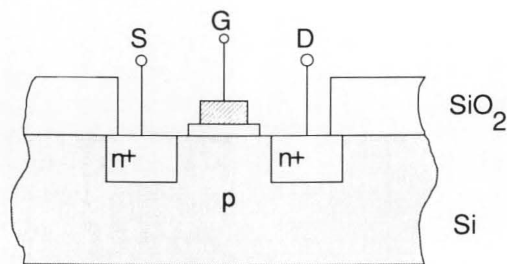
$$W, \quad p_n^o, \quad \text{and} \quad n_p^o$$

are small.

If the sign of the applied potential is reversed [forward bias], the potential barrier decreases and the number of majority carriers capable of overcoming the decreased potential barrier dramatically increases

(proportional to $\exp[qV_F/kT]$). Development of the current equations for the forward bias condition is similar to the reverse bias case, requiring only the use of Equations (15) and (16). Instead of extracting minority carriers from each side of the depletion region, they are injected, and instead of carrier generation in the depletion region, they recombine ($pn > p^o n^o$). The p-n junction device is thus shown to operate as a leaky check valve, permitting a large current to flow under forward bias and only a very small current to flow under reverse bias. An interesting homework problem is the analysis of a p-n junction under uniform illumination (solar cell, photodetector).

With a background in p-n junction behavior the operation of a p^+-n-p bipolar (both electrons and holes participate) transistor (*tran*sfer re*sistor*) is next discussed. This transistor consists of three slabs of semiconductors joined in the series p^+-n-p and electrical contacts made to each slab. The transistor is biased such that the p^+ (heavily doped) -n junction is forward biased and the second n-p junction is reverse biased.



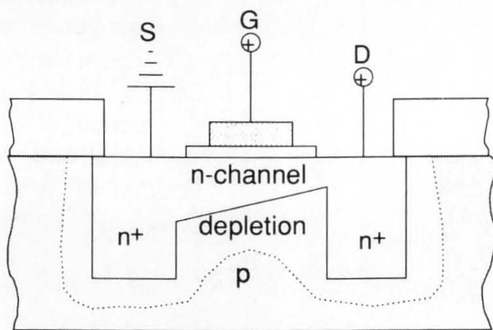


FIGURE 3. Metal oxide semiconductor (MOS) transistor; (a) cross section view of the device; (b) schematic of the gate voltage induced n-channel.

The diffusion current at the forward biased p^+-n junction is largely due to holes because of the heavy doping in the p^+ slab [hole emitter junction]. These holes then diffuse as a minority species across the n-type slab [base]. If the width of this region is kept small compared to the diffusion length, $(D_p/\tau_p)^{1/2}$, most of the holes reach the depletion region of the reverse biased junction and are swept across this junction [hole collector junction] by the favorable electric field. These holes are now a majority carrier in the p-type slab and appear as the collector current. The electrons that enter the base contact and are extracted from the reverse biased junction either participate in a small diffusion current at the forward biased junction or react homogeneously with holes in the base region. With proper transistor design the collector current can be significantly greater than the base current [amplification], thanks to the "pumping" action of the emitter. The bipolar transistor can act as either a pump [amplification] or an on-off valve [switch].

METAL-OXIDE-SEMICONDUCTOR (MOS) JUNCTIONS

Though the bipolar transistor can be made to act as a fast switch, the power requirements can be fairly high. In order to decrease the base current, an insulating oxide layer is sandwiched between the base and the metal contact as illustrated in Figure 3. The MOS transistor is a three-terminal device with a source (S), a gate (G), and a drain (D). In this particular configuration, the source and drain lead wires are connected to "pockets" of n-type material which are isolated by a p-type region. Application of a potential between the drain and source contacts will not produce any significant current since one of the junctions is reverse biased. The application of a positive voltage to the gate contact attracts electrons and repels holes in the semiconductor near the oxide interface, uncovering immobile ions. For a sufficiently large applied gate voltage [threshold voltage], the population of electrons near the interface will exceed that of the holes [inversion], and a continuous n-type channel forms between the source and the gate that permits a current to flow. Of course, the n-type channel and p-type material are separated by a depletion region. A further increase in the applied voltage will increase the n-type channel "pipe" diameter to produce a larger "flow rate." The device can be operated as either an "on-off valve" or "gate valve."

The lecture presentation includes a graphical representation of the band diagrams in the equilibrium, accumulation (negative gate voltage), depletion, and inversion regimes. The operation of a functional

capacitor is also analyzed. The p-n junction has a small capacitance since the depletion width changes with applied voltage. Similarly, the MOS structure can be used as a capacitor. Actually, this structure has two capacitors connected in series; a parallel-plate-like capacitor with the oxide as the dielectric material and the accumulation/depletion regions of the semiconductor. The equations which describe the operation of an ideal MOS capacitor are relatively straightforward to develop [1-6] and are presented in the course.

CONCLUDING REMARKS

By their senior year, students have already received the foundations of chemical engineering. One of the objectives of these lectures is to convince seniors that the digital integrated circuit is nothing more than a chemical processing plant. Integrated circuits contain only a few types of devices (resistors, capacitors, transistors, and diodes). These four lectures demonstrate their operation in terms that, for the most part, are understandable by chemical engineers.

A comparison between a large scale chemical processing plant and an integrated circuit is given in Table 1. A typical chemical plant processes hundreds of species, while the integrated circuit processes only two charged species, the electron and the hole. Powerful pumps move fluids through large diameter pipes at high flowrates in a chemical plant, while power requirements, dimensions, and flowrates are orders of magnitude lower in an integrated circuit. A high percentage of the land area at a plant site can be devoted to storage of raw materials and products. In contrast, charge storage in a p-n junction or MOS structure is very limited in an integrated circuit. As discussed above, control valves have analogs in an integrated circuit. Indeed, a diode is used to protect the circuit against excessive voltages, just as a check valve protects against excessive pressures. Relief in the integrated circuit is accomplished by simply dumping the current to ground. The E.P.A., however, does not permit this luxury at a chemical plant site. One of the major difficulties in the simulation of a chemical process is the large number of chemical reactions, often coupled and with unknown rate constants. In the electron-hole plant the reactions involve only recombination and generation, often of reduced order. Due to the size difference in the basic unit operations, the densities are dramatically different, though the costs per unit area are similar. As in chemical plants, the rates of most processes are limited by either reaction or diffusion. The diffusion coefficient and reaction rate constants for electrons and holes are very high. Combining these properties with the small dimensions found in an integrated circuit gives an extremely rapid response time to input parameter changes in the circuit.

TABLE 1
Comparison Between a Chemical Processing Plant and an Integrated Circuit

	Typical Chemical Plant	Typical Integrated Circuit
Raw material source	many but depleting	electrical ground
Number of species	10^2	2 (electron, hole)
Transport	pipe (10 inch O.D.)	Metal interconnect (10^{-5} inch O.D.)
Flow rates	10^3 moles/s	10^{-11} moles/s
Pump	10 hp	10^{-9} hp (bipolar transistor)
Storage	tanks (10^6 moles)	capacitor (10^{-10} moles)
Control	check valve on-off valve gate valve	diode transistor field effect transistor
Reactions	many	recombination/generation
Unit operation density	10^4/mi^2	10^{16}/mi^2
Cost	\$$10^8$ (\$$10^9$/mi^2)	\$10 (\$$10^9$/mi^2)
Waste disposal	problem	electrical ground
Diffusion coefficient	10^{-2} - 10^{-5} cm^2/s	10 - 10^3 cm^2/s
Reaction rate constant (1st order)	10^6 1/moles s	10^{16} 1/moles s

REFERENCES

1. Adler, R.B., A.C. Smith, and R.L. Longini, *Introduction to Semiconductor Physics,* Wiley, New York (1964)
2. Gray, P.E., D. DeWitt, A.R. Boothroyd, and J.F. Gibbons, *Physical Electronics and Circuit Models of Transistors,* Wiley, New York (1964)
3. Grove, A.S., *Physics and Technology of Semiconductor Devices,* Wiley, New York (1967)
4. Streetman, B.G., *Solid State Electronic Devices,* 2nd ed., Prentice Hall, Englewood Cliffs, NJ (1980)
5. Sze, S.M., *Physics of Semiconductor Devices,* 2nd ed., Wiley, New York (1981)
6. Sze, S.M., *Semiconducting Devices: Physics and Technology,* Wiley, New York (1985) ❐