Chapter 22
Light-to-Electricity Conversion

Abstract The use of diodes for the conversion of electromagnetic radiation (light) into electrical signals is discussed with many examples for devices such as photoconductors, pn-, pin-, MSM- and avalanche photodiodes and charge-coupled devices. Energy conversion with solar cells is explained and standard and advanced solar cells concepts are introduced.

22.1 Photocatalysis

The absorption of light in a semiconductor across the band gap creates free electrons and holes. In particular, for small particle size in powders[1] these charge carriers can reach the surface of the semiconductor. At the surface they can react with chemicals. The hole can form •OH radicals from OH^- attached to the bead. The electron can form $O_2\bullet^-$. These radicals can subsequently attack and detoxify, e.g., noxious organic pollutants in the solution surrounding the semiconductor. Such photocatalytic activity has been found, e.g., for TiO_2 and ZnO powders. A review of photocatalysis, in particular with TiO_2 particles and their surface modifications with metals and other semiconductors, is found in [1527].

The efficiency of the photocatalytic activity depends on the efficiency of the charge separation (Fig. 22.1). Any electron–hole pair that recombines within the bulk or the surface of the particle is lost for the catalytic activity. Thus, surfaces must exhibit a small density of recombination centers. Surface traps, however, can be beneficial for charge-carrier separation when they 'store' the charge-carrier rather than letting it recombine. Small particles are expected to exhibit more efficient charge-carrier separation than larger ones. Electrons at the surface can be donated and reduce an electron acceptor, typically oxygen, $A \rightarrow A^-$. A hole at the surface can oxidize a donor species, $D \rightarrow D^+$.

An example of increased photocatalytic activity are TiO_2 powders with deposited metal particles (such as Pt) for H_2 evolution and metal-oxide particles (such as RuO_2) for O_2 evolution. Such a system behaves as a short-circuited microscopic

[1] 'Small' is here in relation to the diffusion length and does not need to be in the range where quantization effects (quantum dots) are present.

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_22

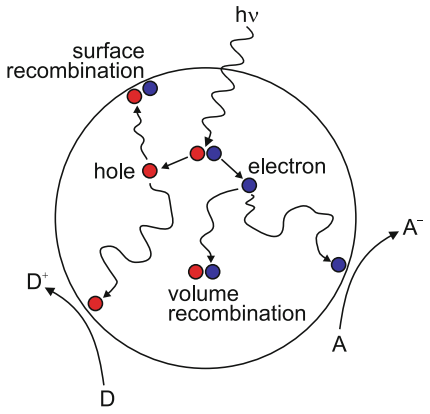

Fig. 22.1 Principle of photocatalytic activity. Light absorption creates an electron–hole pair. The electron and hole diffuse and can recombine in the bulk or at the surface. Free carriers can react at the surface with species from the surrounding solution, reducing an electron acceptor or oxidizing a donor species. Adapted from [1527]

photoelectrochemical cell in which Pt is the cathode and RuO_2 is the anode [1528]. Excitation with light energy above the band gap in the TiO_2 particle (3.2 eV) injects electrons into the Pt particles and holes into the RuO_2 particles. Trapped electrons in Pt reduce water to hydrogen and trapped holes in RuO_2 oxidize water to oxygen.

The photocatalytic activity is also tied to the geometrical shape of the semiconductor. Generally, powders with nanosized grains have much higher activity than those with microsized particles [1569]. In Fig. 22.2 it is shown that nanosized objects with high surface-to-volume ratio are more effective catalysts than rather compact surfaces.

In sun-protection cream only the UV absorption is wanted in UVA (330–420 nm) and UVB (260–330 nm) ranges. Subsequent photocatalysis on the skin and the presence of radicals are unwanted. Thus the semiconductor particles (∼10–200 nm diameter) are encapsulated in microbeads (∼1–10 μm diameter) of silica, PMMA or urethane, also improving ease of dispersion, aggregation, stability and skin feel.

22.2 Photoconductors

22.2.1 Introduction

Charge carriers can be generated in the semiconductor through the absorption of light with a photon energy above or below the band gap (Fig. 22.3). Absorption involving impurities occurs typically in the mid- and far-infrared spectral regimes (cf. Sect. 9.7). The additional charge carriers cause an increase in the conductivity (8.11).

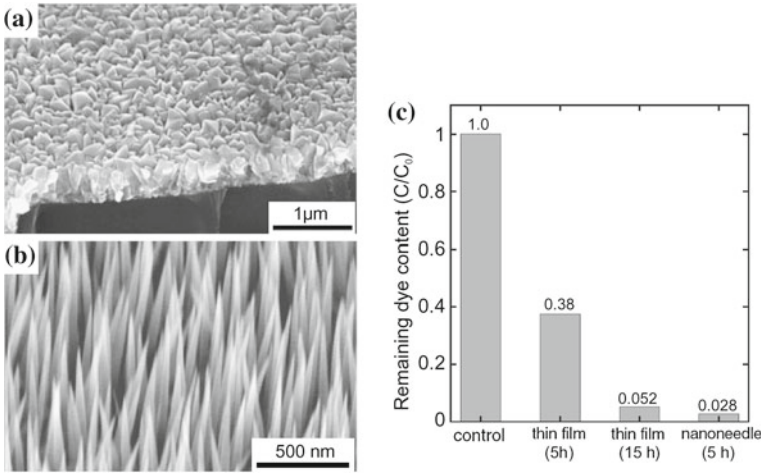

Fig. 22.2 SEM images of MOCVD-grown (**a**) ZnO thin film and (**b**) ZnO nanoneedle layer. (**c**) Comparison of the photocatalytic activity (decoloration of the dye Orange II in aqueous solution) of the ZnO thin film (irradiation with a Hg lamp for 5 h and 15 h) and the ZnO nanoneedles (irradiation 5 h). The sample labeled 'control' (scaled to 100 %) is the start situation (absorption of the dye Orange II) without photocatalytic process. Adapted from [1530], reprinted with permission

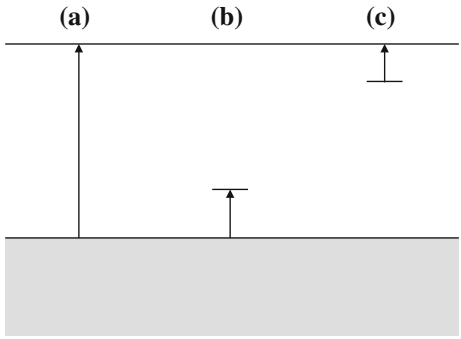

Fig. 22.3 Absorption and charge-carrier generation in a photoconductor: (**a**) band–band transition, (**b**) valence-band to acceptor and (**c**) donor to conduction-band transition

22.2.2 Photoconductivity Detectors

In stationary equilibrium for constant illumination of power P_{opt} and photon energy $E = h\nu$ the generation rate G is given by

$$G = \frac{n}{\tau} = \eta \frac{P_{\text{opt}}/h\nu}{V}, \tag{22.1}$$

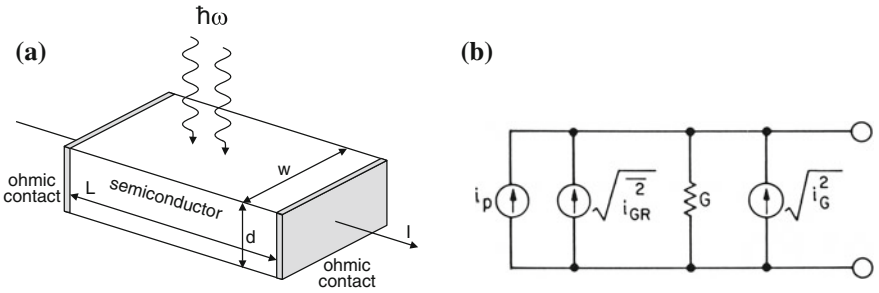

Fig. 22.4 (**a**) Scheme of photoconductor. (**b**) Equivalent circuit of photoconductor

where V is the volume ($V = wdL$, see Fig. 22.4) and τ denotes the charge-carrier lifetime. η is the quantum efficiency, i.e. the average number of electron–hole pairs generated per incoming photon. The photocurrent between the electrodes is

$$I_{\mathrm{ph}} = \sigma E w d \approx e \mu_{\mathrm{n}} n E w d, \tag{22.2}$$

assuming that $\mu_{\mathrm{n}} \gg \mu_{\mathrm{p}}$ and with $E = V/L$ denoting the electric field in the photoconductor, V being the voltage across the photoconductor. We can then also write

$$I_{\mathrm{ph}} = e \left(\eta \frac{P_{\mathrm{opt}}}{h\nu}\right) \left(\frac{\mu_{\mathrm{n}} \tau E}{L}\right) = g I_{\mathrm{p}}. \tag{22.3}$$

With the primary photocurrent $I_{\mathrm{p}} = e \left(\eta \frac{P_{\mathrm{opt}}}{h\nu}\right)$ we deduce a gain

$$g = \frac{I_{\mathrm{ph}}}{I_{\mathrm{p}}} = \frac{\mu_{\mathrm{n}} \tau E}{L} = \frac{\tau}{t_{\mathrm{r}}}, \tag{22.4}$$

where $t_{\mathrm{r}} = L/v_{\mathrm{d}}$ is the transit time through the photoconductor.

Now we consider a modulated light intensity

$$P(\omega) = P_{\mathrm{opt}} \left[1 + m \exp(\mathrm{i}\omega t)\right], , \tag{22.5}$$

where m is between 0 and 1. For $m = 0$ it is a constant light power, for $m = 1$ the intensity is sinusoidally modulated between 0 and $P_{\max} = 2P_{\mathrm{opt}}$. The rms optical power[2] is given by $\sqrt{2} m P_{\mathrm{opt}}$. In the case of $m = 1$ this is equal to $P_{\max}/\sqrt{2}$.

The rms photocurrent ($i^2 = \langle I^2 \rangle - \langle I \rangle^2$) is

$$i_{\mathrm{ph}} \approx \frac{e \eta m P_{\mathrm{opt}}}{\sqrt{2} h\nu} \frac{\tau}{t_{\mathrm{r}}} \frac{1}{\sqrt{1 + \omega^2 \tau^2}}. \tag{22.6}$$

[2]The rms value is the square root of the time average of the square of the power, $\sqrt{\langle P^2 \rangle}$.

Besides the photocurrent which is considered the signal, several sources of noise must be considered. Noise is in this case a fluctuation current i_n with $\langle i_n \rangle = 0$ (see Appendix J).

The thermal noise (Appendix J.3.1) at a conductivity $G = 1/R$ is[3]

$$i_G^2 = 4\,k\,T\,G\,B, \tag{22.7}$$

with B being the bandwidth over which the noise spectrum is integrated. The thermal noise at a resistor was experimentally found by Johnson [1531, 1532] and theoretically explained by Nyquist [1533].[4]

The statistical nature (Poisson statistics) of photon arrival (and absorption) and equally that of recombination of (photo-) excited electrons leads to fluctuations of the carrier density and consequently to fluctuating conductivity and gain. This so-called generation–recombination noise (Appendix J.3.4 is given by [1534]

$$i_{GR}^2 = 4\,e\,I_{ph}\,B\,g\,\frac{1}{1+\omega^2\,\tau^2} \tag{22.8}$$

for the modulation frequency ω, I_{ph} being the photocurrent in steady state (22.3). The equivalent circuit with the ideal photocurrent source and the noise currents is depicted in Fig. 22.4b. A detailed treatment can be found in [1535].

The signal-to-noise ratio of the power is then given by

$$S/N = \frac{i_{ph}^2}{i_G^2 + i_{GR}^2} = \frac{\eta\,m^2\,(P_{opt}/h\nu)}{8\,B}\left[1 + \beta^{-1}\frac{t_r}{\tau}(1+\omega^2\tau^2)\,\frac{G}{I_0}\right]^{-1}. \tag{22.9}$$

An important quantity is the noise equivalent power (NEP). This is the light power $(mP_{opt}/\sqrt{2})$ for which the S/N ratio is equal to 1 (for $B = 1$). The responsivity of a detector to light has been termed 'detectivity' and is the inverse of the noise equivalent power. It typically depends on the square root of the detector area A and the bandwidth B [1536]. Thus the detectivity D^* (D star) has been introduced defined by [1537]

$$D^* = \frac{\sqrt{A\,B}}{\mathrm{NEP}}, \tag{22.10}$$

in order to make various detectors comparable. The unit of D^* is $\mathrm{cm\,Hz^{1/2}\,W^{-1}}$, also known as Jones. The detectivity should be stated together with the modulation frequency. It can be given for monochromatic radiation of a particular wavelength λ or a blackbody spectrum of given temperature T. As refined measure for detectivity, D^{**} (D double star) has been defined to take into account the solid angle Ω from

[3] For frequencies $h\nu \ll kT$; at room temperature kT/h is in the THz regime.

[4] The formula for thermal noise (22.7) is the fluctuation-dissipation theorem in statistical physics, providing a general relation between the response of an equilibrium system to small external perturbations and its spontaneous fluctuations.

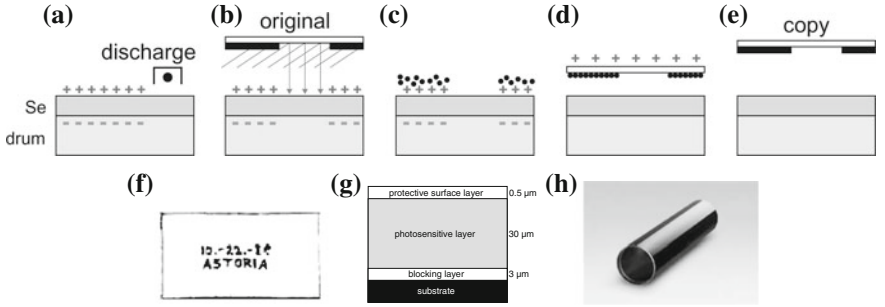

Fig. 22.5 Principle of xerography: (**a**) charging of the selenium-covered drum, (**b**) (reflection) exposure of the Se, exposed areas become uncharged, (**c**) toner addition, (**d**) toner transfer to paper for copy, and (**e**) fixation of the toner on the copy and preparation of drum for the next cycle. (**f**) First xerox copy (Oct. 22nd 1938). (**g**) Schematic cross section of coating of photosensitive drum. The indicated *thicknesses* are approximate. (**h**) Image of drum with photosensitive layer made from amorphous silicon. Part (**h**) from [1539]

which radiation can reach the detector [1538],

$$D^{**} = \sqrt{\Omega/\pi}\, D^{*}\ ; \tag{22.11}$$

for a Lambertian characteristic $D^{**} = D^{*}$.

22.2.3 Electrophotography

The principle of the Xerox copy machine is based on a photoconductive layer (Fig. 22.5). This layer is normally insulating such that both sides of the layer can be oppositely charged. If light hits the layer it becomes photoconductive and neutralizes locally. This requires a small lateral diffusion of charge carriers. Initially amorphous selenium ($E_g = 1.8\,\text{eV}$) was used. The conductivity in the dark of a-Se is $10^{16}\,\Omega$/cm. Se was subsequently replaced by organic material. The highest performance is currently achieved with amorphous silicon.

On the charged areas of the photosensitive layer toner can be attached. The toner pattern is subsequently transferred to the copy sheet and fixated. A copy takes typically more than one rotation of the drum. The principle was invented in 1938 by Chester F. Carlson (1906–1968) with sulfur as the photoconductor.[5]

[5] In 1947 the Haloid company bought the rights to this process, renamed itself XeroX and brought the first copy machine to the market in 1958 based on amorphous selenium. The word 'xerography' stems from the Greek word $\xi\acute{\epsilon}\rho o\varsigma$ (dry). The last 'X' in XeroX was added to mimic the name of the KodaK corporation.

22.2.4 QWIPs

Quantum-well intersubband photodetectors (QWIPs) are based on the absorption of photons between two quantum well subbands (Fig. 22.6). A review can be found in [1540]. Quantized electron or hole states can be used. Besides an oscillator strength for this transition, the lower level must be populated and the upper level must be empty in order to allow this process. The Fermi level is typically chosen by appropriate doping such that the lower subband is populated.

For infinite barrier height the energy separation between the first and second quantized levels (in the effective-mass theory) is (cf. (12.6))

$$E_2 - E_1 = 3\,\frac{\hbar^2}{2m^*}\,\frac{\pi^2}{L_z^2}. \tag{22.12}$$

For real materials the barrier height determines the maximum transition energy. Typical absorption and transmission spectra of a QWIP structure are shown in Fig. 22.7. The spectral response is in the mid- or far-infrared.

The dipole matrix element $\langle z\rangle = \langle \Psi_2|z|\Psi_1\rangle$ can be easily calculated to be

$$\langle z\rangle = \frac{16}{9\pi^2}\,L_z. \tag{22.13}$$

The oscillator strength is about 0.96. The polarization selection rule causes the absorption to vary $\propto \cos^2\phi$, where ϕ is the angle between the electric-field vector and the z direction (Fig. 22.8). This means that for vertical incidence ($\phi = 90°$) the absorption vanishes. Thus schemes have been developed to allow for skew entry of the radiation (Fig. 22.9a). The strict selection rule can be relaxed by using asymmetric potential wells (breaking of mirror symmetry/parity), strained materials (band mixing) or quantum dots (lateral confinement). Also, a grating can be used to create a finite angle of incidence (Fig. 22.9b).

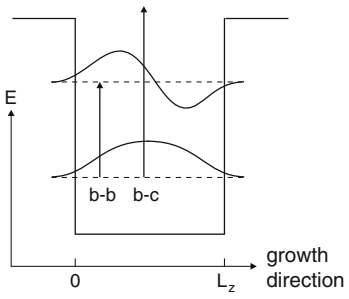

Fig. 22.6 Schematic level diagram of a quantum well. Optical intersubband transitions between the first and second quantized level (b–b) and the ground state and the continuum (b–c)

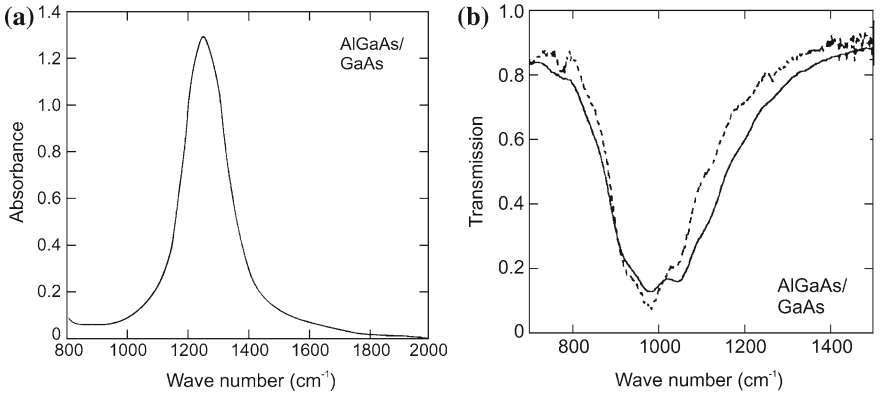

Fig. 22.7 (**a**) AlGaAs/GaAs QWIP absorption spectrum for multiple reflection geometry. Adapted from [1540]. (**b**) Transmission of AlGaAs/GaAs QWIP (100 QWs) in double reflection geometry (45° angle of incidence). The well doping is $1.0 \times 10^{12}\,\mathrm{cm}^{-2}$ (*dashed line*) and $1.5 \times 10^{12}\,\mathrm{cm}^{-2}$ (*solid line*). Adapted from [1541]

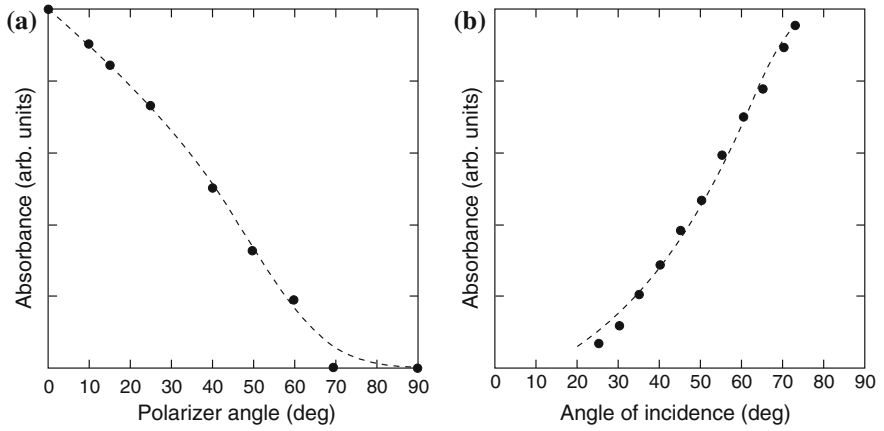

Fig. 22.8 Dependence of the QWIP response on (**a**) polarization and (**b**) angle of incidence. *Dashed lines* are guides to the eye. Adapted from [1540]

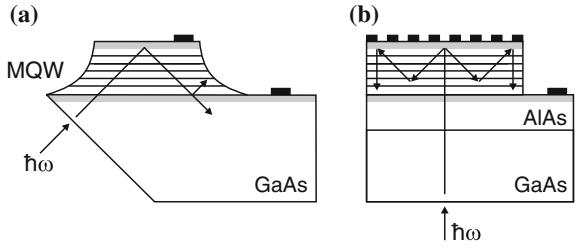

Fig. 22.9 QWIP geometries: (**a**) 45° edge coupled with multiple quantum-well (MQW) absorber and (**b**) grating coupled with GaAs substrate, AlAs reflector and metal grating on top. *Grey areas* are highly n-doped contact layers

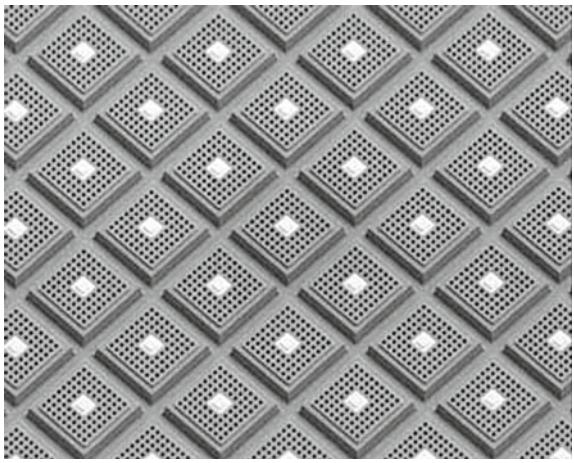

Fig. 22.10 Part of a 256×256 QWIP focal plane array (FPA) with grating coupler (area of one pixel: $37\,\mu m^2$). From [1543]

Besides a useful detectivity (2×10^{10} cm Hz$^{1/2}$/W at 77 K) QWIPs have the advantage, e.g. against HgCdTe interband absorbers, that the highly developed GaAs planar technology is available for the fabrication of focal plane arrays (FPA) as shown in Fig. 22.10. A FPA is an image sensor (in the focal plane of an imaging infrared optics) and is used, e.g., for the detection of heat leaks in buildings or night surveillance. In particular, night vision support in cars may become a major market. A competing technology are bolometric arrays with thermally insulated pixels based on MEMS technology. A review of FPA technologies can be found in [1542].

The carriers that have been optically excited into the upper state leave the QW by tunneling or thermionic emission. Also, a QWIP can be made based on the direct transfer from the (populated) subband into the continuum.

The incoming infrared radiation creates a photocurrent density of

$$i_{\mathrm{ph}} = e\,\eta_{\mathrm{w}}\,\Phi, \tag{22.14}$$

where η_{w} is the quantum efficiency of a single quantum well (including the escape rate) and Φ is the photon flux per time and unit area. During the transport of the charge carriers through the barrier they can be (re-)captured by the QW with the probability p_{c}. The capture probability decreases exponentially with the applied bias. The total photocurrent (including generation and recapture) is

$$I_{\mathrm{ph}} = (1 - p_{\mathrm{c}})\,I_{\mathrm{ph}} + i_{\mathrm{ph}} = \frac{i_{\mathrm{ph}}}{p_{\mathrm{c}}}. \tag{22.15}$$

If the quantum efficiency is small, the efficiency of N_{w} quantum wells $\eta \approx N_{\mathrm{w}} \times \eta_{\mathrm{w}}$. With this approximation the total photocurrent of N_{w} quantum wells is given by

$$I_{\mathrm{ph}} = e\,\eta\,\Phi\,g, \tag{22.16}$$

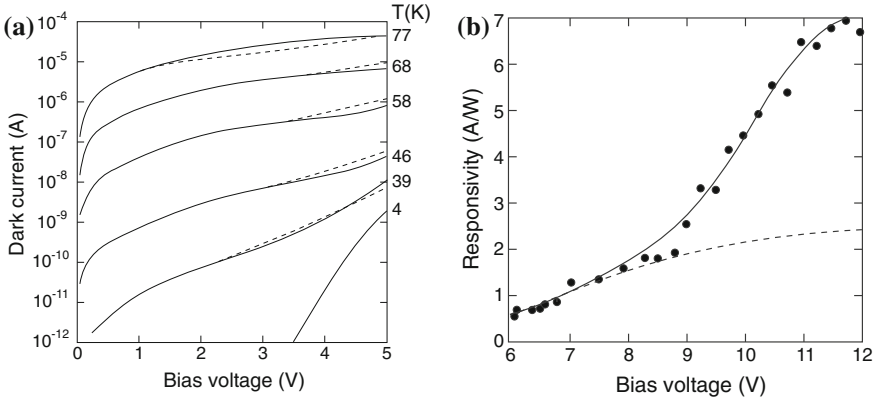

Fig. 22.11 (**a**) Dark current of a QWIP at 10.7 μm, experimental (*solid lines*) and theoretical (*dashed lines*) response. (**b**) QWIP responsivity as a function of the applied voltage. The *solid line* (*dashed line*) is the theoretical dependence with (without) the effect of avalanche multiplication. Adapted from [1540]

where g is termed the gain of the structure and is given by

$$g = \frac{1}{p_c}\,\frac{\eta_w}{\eta} \approx \frac{1}{N_w\, p_c}. \tag{22.17}$$

The dark current can be calculated from thermionic emission and agrees fairly well with experiment (Fig. 22.11a). When the voltage is increased further, avalanche multiplication can occur while the carriers are transported through the barrier(s). This mechanism provides further gain as shown in Fig. 22.11b.

22.2.5 Blocked Impurity-Band Detectors

Impurity absorption allows photoconductivity detectors in the mid- and far-infrared regions to be made. In particular, for THz spectroscopy in medicine and astronomy the extension to longer wavelengths is interesting. For conventional photoconductors the impurity concentration is well below the critical dopant concentration (cf. Sect. 7.5.7). Long-wavelength response can be achieved by going to impurity/host systems with smaller ionization energy, such as Si:B (45 meV) → Ge:As (12.7 meV) → GaAs:Te (5.7 meV). By applying stress to Ge the energy separation between impurity and conduction bands can be lowered and subsequently the detector response is shifted towards longer wavelengths.

For high doping the impurity level broadens to an impurity band and thus allows smaller ionization energy and thus stronger long-wavelength detector response. However, conduction in the impurity band leads to dark current and makes such detectors unfeasible. In a blocked impurity band (BIB) detector [1544–1546] an additional

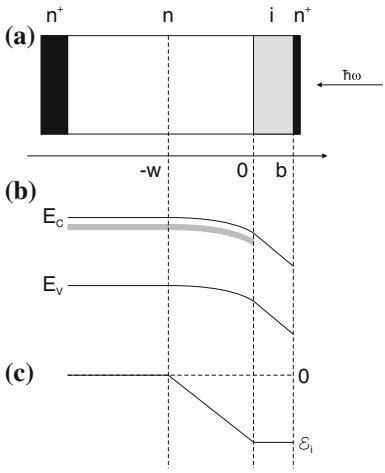

Fig. 22.12 (**a**) Structure of BIB photodetectors. Highly doped contact layers (*black*), doped semiconductor (*white*) and blocking (intrinsic) layer (*grey*). (**b**) Band diagram under small forward bias. *Shaded area* represents the donor impurity band. (**c**) Electric field in the structure

intrinsic blocking layer is sandwiched between the absorption layer and the contact (Fig. 22.12a). Such a structure is similar to a MIS diode, the insulator being the intrinsic semiconductor. We assume in the following an n-type semiconductor, such as Si:As or GaAs:Te, but also p-type BIBs can be made, e.g., from Ge:Ga.

The semiconductor is highly doped (N_D) and partly compensated (N_A). Typically, the acceptor concentration must be small, about $10^{12}\,\mathrm{cm}^{-3}$, and controls the formation of the electric field as shown below. The doping is so high that the impurities form an impurity band. Some of the electrons recombine with the acceptors $N_A^- = N_A$ and leave some donors charged $N_D^+ = N_A$. For GaAs, e.g., the donor concentration in the doped semiconductor is $>10^{16}\,\mathrm{cm}^{-3}$ and $\sim 10^{13}\,\mathrm{cm}^{-3}$ in the i-layer.

Under an external forward bias V, i.e. the positive pole is at the insulator, part of the applied voltage drops over the blocking layer of thickness b. If ideally no charges are present here, the electric field is constant. In the n-doped material electrons move in the impurity band towards the insulator, forming neutral donors in an electron accumulation layer of thickness w in the presence of the charged acceptors N_A^-. This layer is the absorption layer. The mechanism can also be considered as if positive charge (the charged donors, N_D^+) moves (via hopping conduction) towards the back contact. In the literature the layer close to the insulator is thus also termed a 'depletion layer'. The band diagram and the electric field are shown in Fig. 22.12b,c. Due to the blocking layer the carriers on the donors in the n-type material cannot spill via the impurity band into the contact but must be lifted (by photoabsorption) into the conduction band.

From the Poisson equation the electric field is given by

$$E(x) = -\frac{e}{\epsilon_s} N_A\,(w + x)\,, \quad -w \le x \le 0 \tag{22.18a}$$

$$E(x) = -\frac{e}{\epsilon_s} N_A\, w = E_i, \quad 0 \le x \le b. \tag{22.18b}$$

The voltage drops across the blocking layer V_b and the doped semiconductor V_s fulfill

$$V = V_b + V_s. \tag{22.19}$$

Integration of the fields yields

$$V_s = \frac{e}{\epsilon_s} N_A \frac{w^2}{2} \tag{22.20a}$$

$$V_b = \frac{e}{\epsilon_s} N_A \, w \, b. \tag{22.20b}$$

Substituting (22.20a, b) into (22.19) results in the width of the 'depletion layer'

$$w = \sqrt{\frac{2\,\epsilon_s\,V}{e\,N_A} + b^2} - b. \tag{22.21}$$

The high dopant concentration allows for much thinner absorption layers than in a conventional photoconductivity detector, making it less susceptible to background high-energy cosmic radiation. The recombination in the depletion layer is negligible. Detector performance is modeled in [1547].

22.3 Photodiodes

22.3.1 Introduction

The principle of the photodiode is the interband absorption of light in the depletion layer of a diode (or the i-zone of a pin-diode) and the subsequent separation of electrons and holes by the electric field. There are opposite requirements for fast detectors (thin depletion layer) and efficient detectors (complete light absorption, sufficiently thick depletion layer). For this reason generally semiconductors with high absorption coefficient are most suited (Fig. 22.13). In Fig. 22.14 the quantum efficiency and detectivity D^* of various semiconductor detectors are compared.

A diode can be operated without bias (photovoltaic mode) using the built-in field. An improvement in the speed of a pn-diode is achieved with a reverse bias since it increases the field strength in the depletion layer. However, the reverse bias is below the breakdown voltage. Operation near breakdown is exploited in the avalanche photodiode (APD). In the following we will discuss pn-, pin-, MS- (Schottky-), MSM- and heterostructure-diodes and APDs.

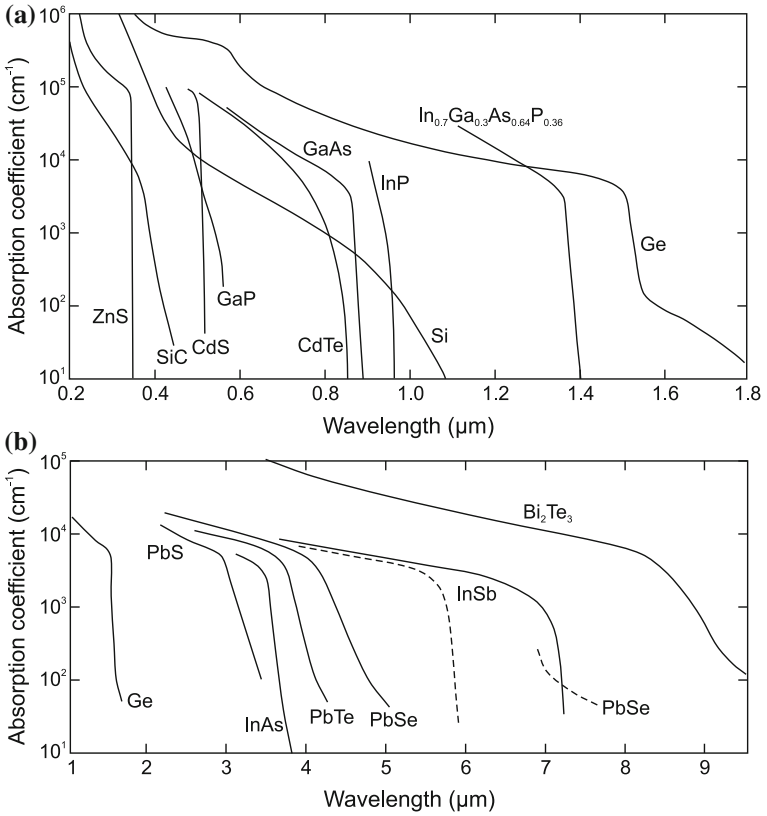

Fig. 22.13 Optical absorption coefficient of various semiconductor materials (as labeled) used for photodetectors (**a**) in the UV, visible and near-infrared range (at room temperature) and (**b**) in the mid-infrared spectral range at room temperature (*solid lines*) and at 77 K (*dashed lines*). Based on [1548]

22.3.2 pn Photodiodes

The most important figures of merit are the quantum efficiency, responsivity, noise equivalent power (NEP) and the response speed.

If the depletion layer is hit by a photon flux with a generation rate G_0 (i.e. electron–hole pairs per unit volume per unit time) the photogenerated current is added to the diffusion current. The photocurrent density j_p (per unit area) is

$$j_\mathrm{p} = -e\, G_0\, L_\mathrm{p} \tag{22.22}$$

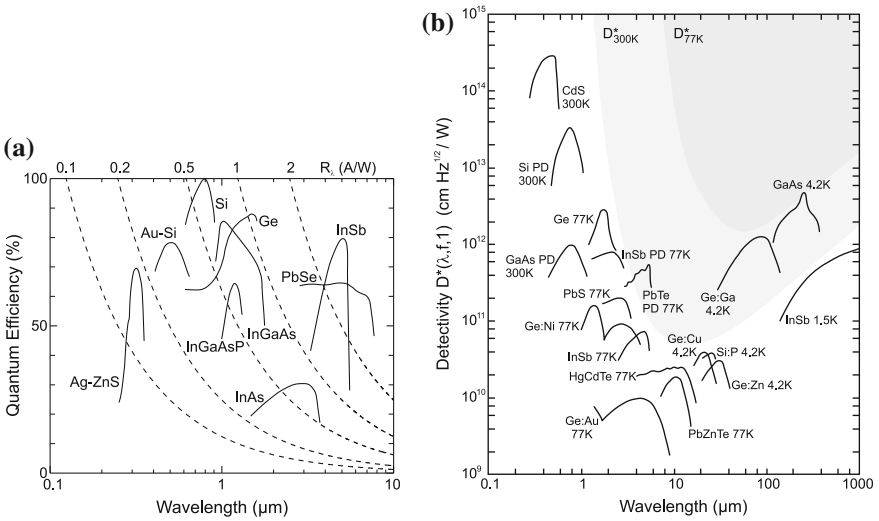

Fig. 22.14 (**a**) Quantum efficiency of various photodetectors. The *dashed lines* depict lines of equal responsivity (R_λ in A/W) as labeled on top of the panel. (**b**) Detectivity D^* of various photoconductors and photodiodes (PD). The *lighter* (*darker*) *shaded area* indicates the range unachievable at 300 K (77 K) due to background radiation. Adapted from [500]

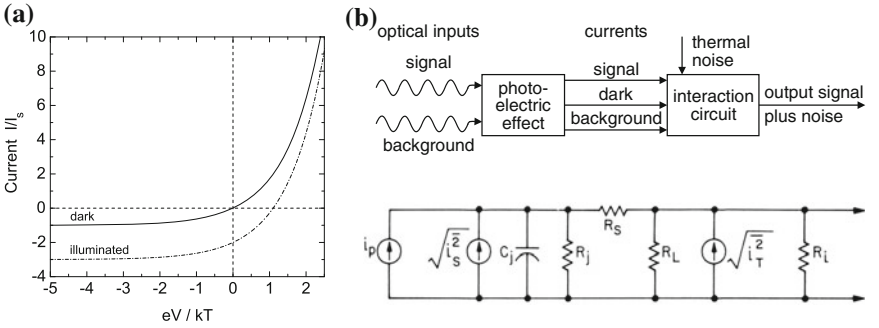

Fig. 22.15 (**a**) Schematic dark and illuminated I–V-characteristics of a photodiode (for the case $j_p = -2j_s$). (**b**) Schematic representation of currents in a photodiode and equivalent circuit. Part (**b**) adapted from [1549]

for a p^+n-diode. In order to obtain this result the diffusion and continuity equations have to be solved for the depletion region.[6] Equation (22.22) means that the dark I–V characteristic is shifted by j_p as shown in Fig. 22.15a. The number of electron–hole pairs that are generated per photon of energy $h\nu$ by the absorption of the (monochromatic) light power P_{opt} is

[6]This derivation is done in Sect. 22.3.3. Equation (22.22) is obtained from (22.33) for vanishing thickness w and $\alpha L_p \ll 1$.

$$\eta = \frac{I_{\mathrm{ph}}/e}{P_{\mathrm{opt}}/(h\nu)}, \tag{22.23}$$

where $I_{\mathrm{ph}} = Aj_{\mathrm{ph}}$ is the photogenerated current over the surface A. The *responsivity* R_λ of the photodiode (for monochromatic radiation) is defined as

$$R_\lambda = \frac{I_{\mathrm{ph}}}{P_{\mathrm{opt}}} = \frac{e}{h\nu}\,\eta \approx \frac{\lambda}{1.24\,\mu\mathrm{m}}\,\eta. \tag{22.24}$$

For a modulated light intensity P_{opt} must be replaced by $mP_{\mathrm{opt}}/\sqrt{2}$. The equivalent circuit including noise sources for a photodiode is shown in Fig. 22.15b.

Random processes lead to shot noise $\langle i_{\mathrm{S}}^2 \rangle$ (Sect. J.3.3). Besides the photocurrent I_{ph} itself, the background radiation (I_{B}, in particular for infrared detectors) and the thermal generation (dark current, I_{D}) of carriers contribute:

$$\langle i_{\mathrm{S}}^2 \rangle = 2\,e\,\left(I_{\mathrm{ph}} + I_{\mathrm{B}} + I_{\mathrm{D}}\right)\,B, \tag{22.25}$$

with B being the bandwidth. Additionally, the parallel resistances cause thermal noise

$$\langle i_{\mathrm{T}}^2 \rangle = 4\,kT\,B/R_{\mathrm{eq}}. \tag{22.26}$$

The resistance R_{eq} is given by the resistance of the depletion layer (junction) R_{j}, the load R_{L} and the input of the amplifier R_{i} as $R_{\mathrm{eq}}^{-1} = R_{\mathrm{j}}^{-1} + R_{\mathrm{L}}^{-1} + R_{\mathrm{i}}^{-1}$. The series resistance R_{s} of the photodiode can be usually ignored in this context.

For a fully modulated signal the signal-to-noise ratio of the photodiode is given by

$$S/N = \frac{i_{\mathrm{ph}}^2}{\langle i_{\mathrm{S}}^2 \rangle + \langle i_{\mathrm{T}}^2 \rangle} = \frac{\left(e\,\eta\,P_{\mathrm{opt}}/h\nu\right)^2/2}{2e\,\left(I_{\mathrm{ph}} + I_{\mathrm{B}} + I_{\mathrm{D}}\right)\,B + 4\,kT\,B/R_{\mathrm{eq}}}. \tag{22.27}$$

Therefore the NEP is given by

$$\mathrm{NEP} = \frac{2\,h\nu\,B}{\eta}\left[1 + \sqrt{1 + \frac{I_{\mathrm{eq}}}{e\,B}}\right]. \tag{22.28}$$

The current I_{eq} is given by $I_{\mathrm{eq}} = I_{\mathrm{B}} + I_{\mathrm{D}} + 2kT/(eR_{\mathrm{eq}})$. If $I_{\mathrm{eq}}/eB \ll 1$, the NEP is determined by the shot noise of the signal itself. In the other limit $I_{\mathrm{eq}}/eB \gg 1$ the detection is limited by the background radiation or thermal noise. In this case, the NEP is (for $B = 1\,\mathrm{Hz}$, in $\mathrm{W\,cm^2\,Hz^{1/2}}$)

$$\mathrm{NEP} = \sqrt{2}\,\frac{h\nu}{\eta}\,\sqrt{\frac{I_{\mathrm{eq}}}{e}}. \tag{22.29}$$

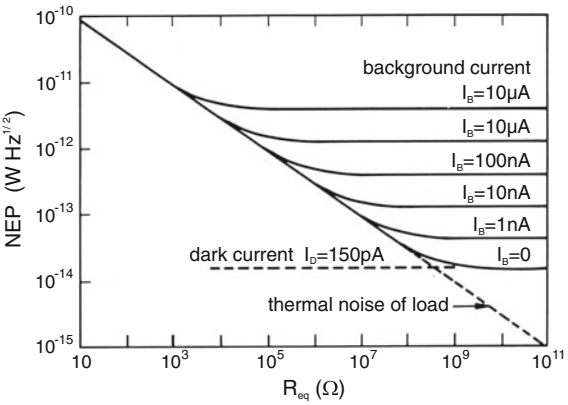

Fig. 22.16 NEP as a function of the resistance R_{eq} for a Si photodiode. From [1549]

In Fig. 22.16 the situation is shown for a silicon photodiode as a function of R_{eq}. The diode has a quantum efficiency of 75 % at $\lambda = 0.77\,\mu\text{m}$. A high value of $R_{\text{eq}} \sim 1\,\text{G}\Omega$ is necessary to ensure detection limited by dark current.

22.3.3 Pin Photodiodes

The depletion layer in pn-diodes is relatively thin such that the incident light is not completely absorbed. An almost complete absorption of light can be achieved by using a thick intrinsic absorption layer. The field in the intrinsic region is constant or slowly varying linearly (Fig. 21.70). The generation rate per unit area decreases exponentially following the Lambert–Beer law (9.16) as shown in Fig. 22.17c:

$$G(x) = G_0 \exp(-\alpha x). \tag{22.30}$$

The initial generation rate $G_0 = \Phi_0 \alpha$ is given by the incident photon flux per unit area Φ_0 and the reflectance of the surface R as $\Phi_0 = P_{\text{opt}}(1 - R)/(Ah\nu)$.

The drift current in the i-region collects all those carriers (if recombination in the depletion layer is neglected). The electron drift current is given by

$$j_{\text{dr}} = -e \int_0^w G(x)\,\text{d}x = e\,\Phi_0 \left[1 - \exp(-\alpha w)\right], \tag{22.31}$$

with w being the thickness of the depletion layer that is approximately the same as the thickness of the i-region. In the bulk (neutral) region ($x > w$) the minority-carrier density is determined by drift and diffusion[7] (10.76). The diffusion current density

[7] At the edge of the depletion layer, $x = w$, all photo-generated carriers are transported away instantly, thus the excess carrier density from photo-generated carriers is zero there and (10.76) applies.

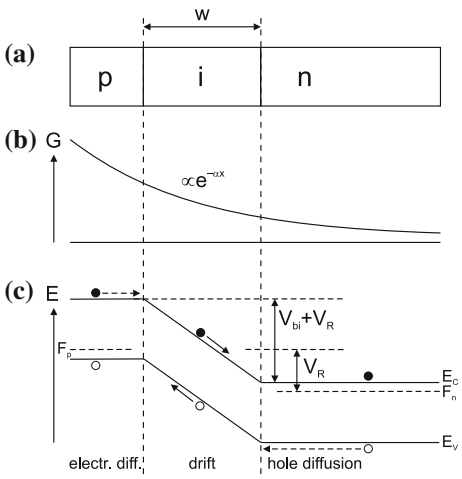

Fig. 22.17 (**a**) Schematic cross section of pin-diode, (**b**) profile of carrier generation due to light absorption and (**c**) schematic band structure under reverse bias. The generation of three electron-hole pairs is shown; subsequent drift (diffusive) transport is indicated with *solid* (*dashed*) *arrows*

at $x = w$ is thus given by

$$j_{\text{diff}} = e\,\Phi_0 \exp(-\alpha\, w)\;\frac{\alpha\, L_\text{p}}{1+\alpha\, L_\text{p}} + e\,p_{\text{n}_0}\,\frac{D_\text{p}}{L_\text{p}}. \tag{22.32}$$

The first term is due to the diffusion current of photo-generated carriers (10.76), the second term due to thermally generated carriers (21.129). The total current $j_{\text{tot}} = j_{\text{diff}} + j_{\text{dr}}$ is given by

$$j_{\text{tot}} = e\,\Phi_0 \left[1 - \frac{\exp(-\alpha\, w)}{1+\alpha\, L_\text{p}}\right] + e\,p_{\text{n}_0}\,\frac{D_\text{p}}{L_\text{p}}. \tag{22.33}$$

The first term is due to the photocurrent, the second term is due to the diffusion current known from the p^+n-diode. In normal operation, the second can be neglected compared to the first. The quantum efficiency is

$$\eta = \frac{j_{\text{tot}}/e}{P_{\text{opt}}/h\nu} = (1-R)\left[1 - \frac{\exp(-\alpha\, w)}{1+\alpha\, L_\text{p}}\right]. \tag{22.34}$$

For a high quantum efficiency, of course low reflectance and high absorption coefficient, i.e. $\alpha w \gg 1$, are necessary.

However, for $w \gg 1/\alpha$ the transit time through the depletion layer $t_\text{r} \approx w/v_\text{s}$ (at sufficiently high field, v_s being the drift-saturation velocity) increases too much. The 3 dB cutoff frequency $f_{3\,\text{dB}}$ (Fig. 22.18) is

$$f_{3\,\text{dB}} \cong \frac{2.4}{2\pi\, t_\text{r}} \cong \frac{0.4\, v_\text{s}}{w}. \tag{22.35}$$

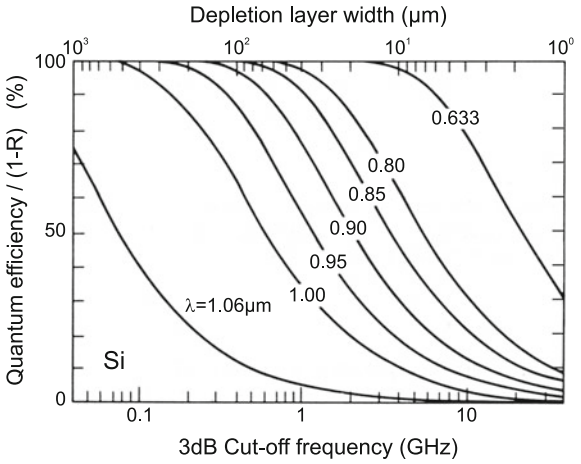

Fig. 22.18 Quantum efficiency and 3 dB cutoff frequency of a Si pin-diode at $T = 300\,\mathrm{K}$ for various wavelengths of input radiation. Adapted from [1549]

Therefore a tradeoff exists between the quantum efficiency and the response speed of the pin-photodiode (Fig. 22.18). Choosing $w \cong 1/\alpha$ is a good compromise.

22.3.4 Position-Sensing Detector

In a position-sensing detector (PSD) two electrodes are placed at opposite edges of a photodetector. The current output depends linearly on the beam position in between the electrodes, similar to a voltage divider. If two pairs of electrodes, one on the front and one on the back of the detector, are fabricated in orthogonal directions (Fig. 22.19a), the beam position can be measured in both x and y directions.

22.3.5 MSM Photodiodes

A MSM photodiode consists of a piece of semiconductor between two Schottky contacts (MS contacts). These are typically arranged laterally (as shown in Fig. 22.24b) but will first be considered at the front and back of the semiconductor [1551]. The band structure in thermodynamic equilibrium is shown in Fig. 22.20.

In the general case two different metals with two different barriers ϕ_{n1}, ϕ_{n2} and built-in voltage V_{D1}, V_{D2} are considered. If a voltage is applied across the MSM diode, one of the junctions is biased in the forward, the other in the reverse direction. We assume in Fig. 22.21 that the voltage biases the first contact in the reverse direction, i.e. the '+' pole is on the left contact. The applied voltage V is split between the two contacts, the larger voltage will drop at the reverse-biased contact (here: $V_1 > V_2$)

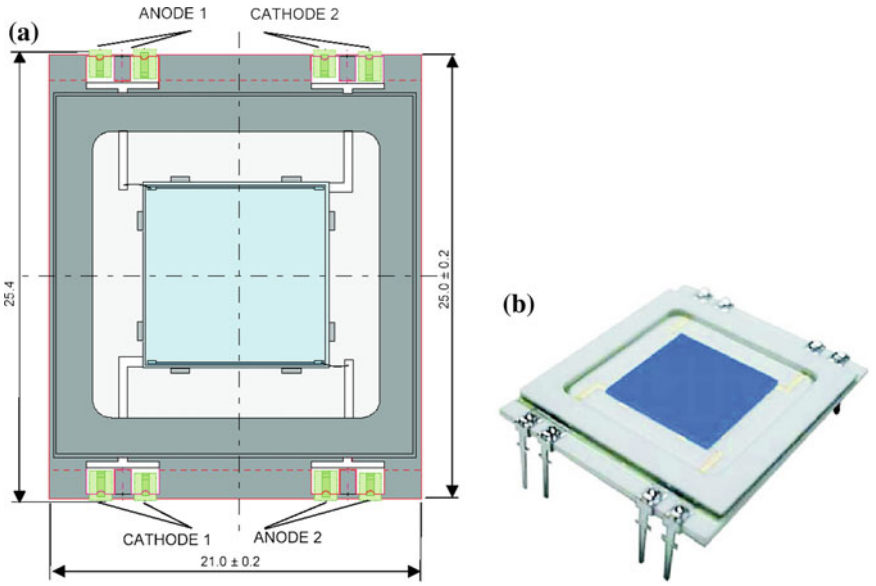

Fig. 22.19 (**a**) Scheme of two-dimensional position-sensing detector (PSD), (**b**) image of PSD. From [1550]

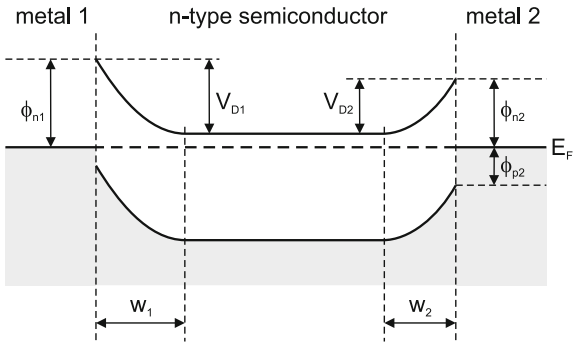

Fig. 22.20 Band diagram of a MSM structure with an n-type semiconductor in thermal equilibrium. In the general case, two different metals cause two different Schottky barrier heights and related depletion layer widths. Adapted from [1551]

$$V = V_1 + V_2. \tag{22.36}$$

The electron current arises from thermionic emission at contact 2. Due to current continuity (without recombination since we inject majority charge carriers) this is also the current through contact 1, i.e.

$$j_{n1} = j_{n2}. \tag{22.37}$$

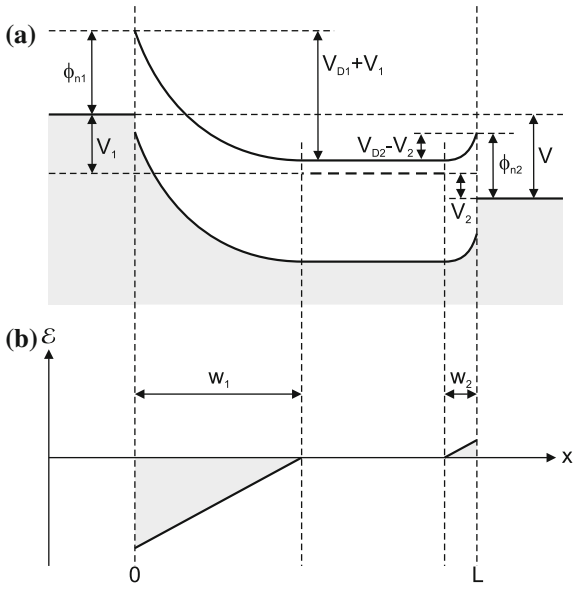

Fig. 22.21 (**a**) Band diagram for a MSM structure under bias ($V < V_{RT}$), (**b**) electric field distribution. Adapted from [1551]

The reverse current at contact 1 is

$$j_{n1} = A_n^* T^2 \exp(-\beta\phi_{n1}) \exp(\beta\Delta\phi_{n1}) \left[1 - \exp(-\beta V_1)\right], \tag{22.38}$$

where $\Delta\phi_{n1}$ is the barrier reduction due to the Schottky effect (Sects. 21.2.3 and (21.27)). The forward current at contact 2 is

$$j_{n2} = -A_n^* T^2 \exp(-\beta\phi_{n2}) \exp(\beta\Delta\phi_{n2}) \left[1 - \exp(\beta V_2)\right]. \tag{22.39}$$

For a symmetric structure, i.e. $\phi_{n1} = \phi_{n2}$ and $V_{D1} = V_{D2} = V_D$, (22.37)–(22.39) yield together with (21.27)

$$\left(\frac{e^3 N_D}{8\pi^2 \epsilon_s^3}\right)^{1/4} \left[(V_D + V_1)^{1/4} - (V_D - V_2)^{1/4}\right] = \frac{1}{\beta} \ln\left[\frac{\exp(\beta V_2) - 1}{1 - \exp(-\beta V_1)}\right]. \tag{22.40}$$

Together with (22.36) a numerical or graphical solution can be found. Initially (for small voltages) the injected hole current (from contact 2) is much smaller than the electron current and diffusion occurs in the neutral region.

The reach-through voltage V_{RT} is reached when the width of the neutral region is reduced to zero (Fig. 22.22a). At the juncture of the two depletion regions inside the semiconductor material the electric field is zero and changes sign. For a larger voltage V_{FB} flat-band conditions are present at contact 2, i.e. the electric field is zero at contact 2 (Fig. 22.22b). At even larger voltage V_B breakdown occurs.

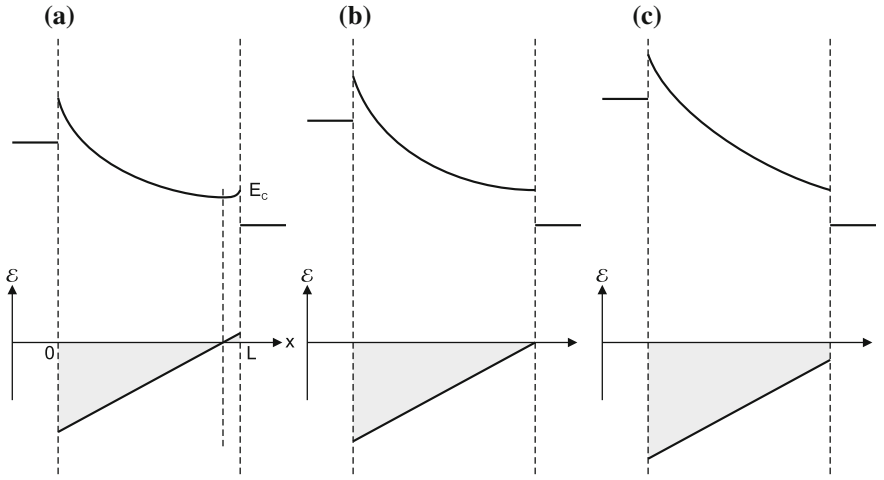

Fig. 22.22 Band diagram (*upper parts*) and electric field distribution (*lower parts*) in a MSM diode for various bias conditions: (**a**) at reach-through voltage V_{RT}, (**b**) at flat-band voltage V_{FB} and (**c**) for $V > V_{FB}$. Adapted from [1551]

At $V = V_{RT}$ we have

$$w_1 = \left[\frac{2\,\epsilon_s}{e\,N_D}\,(V_1 + V_{D1})\right]^{1/2} \tag{22.41a}$$

$$w_2 = \left[\frac{2\,\epsilon_s}{e\,N_D}\,(V_{D2} - V_2)\right]^{1/2} \tag{22.41b}$$

$$L = w_1 + w_2, \tag{22.41c}$$

and therefore (with (22.36))

$$V_{RT} = \frac{e\,N_D}{2\,\epsilon_s}\,L^2 - L\left[\frac{2\,e\,N_D}{\epsilon_s}\,(V_{D2} - V_2)\right] - \Delta V_D, \tag{22.42}$$

with $\Delta V_D = (V_{D1} - V_{D2})$, vanishing for a symmetric MSM structure. At and after reach-through the electric field varies linearly from 0 to L within the semiconductor. The point of zero electric field shifts towards contact 2. At the flat-band voltage this point has reached the contact 2 and the width of the depletion layer at contact 2 is zero. This condition leads (as long as no breakdown occurred) to

$$V_{FB} = \frac{e\,N_D}{2\,\epsilon_s}\,L^2 - \Delta V_D. \tag{22.43}$$

The maximum electric field is at contact 1 and is given (for $V > V_{FB}$) by

$$E_{m1} = \frac{V + V_{FB} + 2\,\Delta V_D}{L}. \tag{22.44}$$

If in a part of the structure the critical field E_B for impact ionization is reached (this will be at contact 1, since the field is highest there), the diode breaks down. Therefore the breakdown voltage is given by

$$V_B \approx E_B\,L - V_{FB} - 2\,\Delta V_D. \tag{22.45}$$

The current–voltage characteristic for a Si-MSM structure is shown in Fig. 22.23. At small voltages only small currents flow since one contact is in reverse bias. The hole current is much smaller than the electron current. Only those holes that diffuse through the neutral region contribute to the hole current. After reach-through the barrier $\phi_{p2} + V_{D2} - V_2$ for hole injection is strongly reduced that leads to strong hole injection. Beyond the flat-band voltage the hole current increases only weakly since a lowering of the barrier occurs only via the Schottky effect. For high fields ($V > V_{FB}$, before breakdown) the hole current is

$$j_{p1} = A_p^*\,T^2\,\exp\left(-\beta\phi_{p2}\right)\,\exp\left(\beta\Delta\phi_{p2}\right) = j_{p,s}\,\exp\left(\beta\Delta\phi_{p2}\right), \tag{22.46}$$

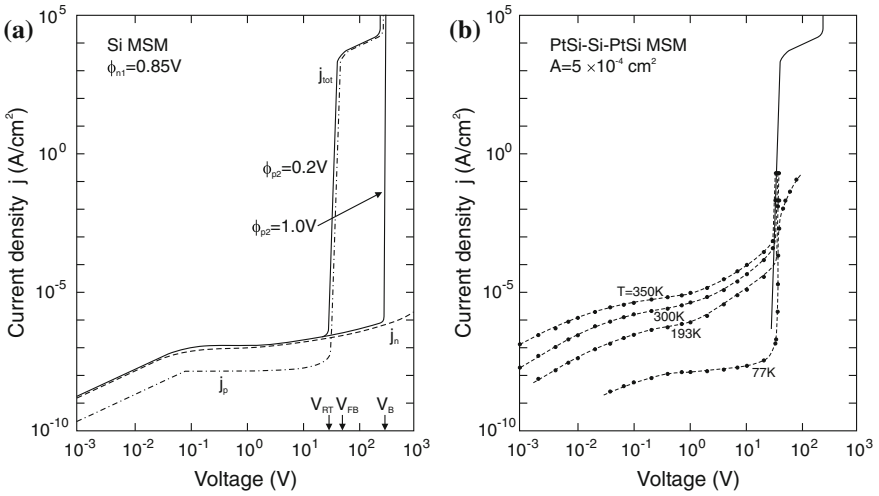

Fig. 22.23 Current–voltage characteristics of a Si MSM structure, $N_D = 4 \times 10^{14}\,cm^{-3}$, $L = 12\,\mu m$ (thin, polished, ⟨111⟩-oriented wafer), $T = 300\,K$. (**a**) Theory for two different values of ϕ_{p_2}, (**b**) experiment (for $\phi_{p_2} = 0.2\,V$). Adapted from [1551]

and the total current is

$$j = j_{\mathrm{n,s}} \exp\left(\beta \Delta\phi_{\mathrm{n1}}\right) + j_{\mathrm{p,s}} \exp\left(\beta \Delta\phi_{\mathrm{p2}}\right), \tag{22.47}$$

with $j_{\mathrm{n,s}} = A_{\mathrm{n}}^{*} T^{2} \exp\left(-\beta\phi_{\mathrm{n1}}\right)$ and $j_{\mathrm{p,s}} = A_{\mathrm{p}}^{*} T^{2} \exp\left(-\beta\phi_{\mathrm{p2}}\right)$.

In a MSM photodetector the metal contacts are typically formed in an interdigitated structure on the semiconductor surface (Fig. 22.24). These contacts shield some of the active area from photons. An increase in quantum efficiency can be achieved with transparent contacts (e.g. ZnO or ITO) and an antireflection (AR) coating.

The dark current is given by (22.47) and is minimal when electron and hole saturation currents are identical. This conditions leads to the optimal barrier height

$$\phi_{\mathrm{n}} = E_{\mathrm{g}} - \phi_{\mathrm{ph}} = \frac{1}{2}\,\beta^{-1} \ln\left(\frac{m_{\mathrm{e}}}{m_{\mathrm{hh}}}\right) + \frac{1}{2} E_{\mathrm{g}} \tag{22.48}$$

close to middle of the band gap. For InP and optimal barrier $\phi_{\mathrm{n}} = 0.645\,\mathrm{eV}$ a dark current of $0.36\,\mathrm{pA/cm^2}$ is expected for a field of $10\,\mathrm{V/\mu m}$. For deviating barrier height the current increases exponentially. The current–voltage characteristic of an InGaAs:Fe MSM photodetector is shown in Fig. 22.25 for a dark environment and various illumination levels.

The time-dependent response of a MSM photodetector depends on the drift time of the carriers, i.e. the time that a created electron and hole need to arrive at their respective contacts. In Fig. 22.26 a simulation is shown for a MSM detector. The current has two components, a fast one due to the electrons and a slow one due

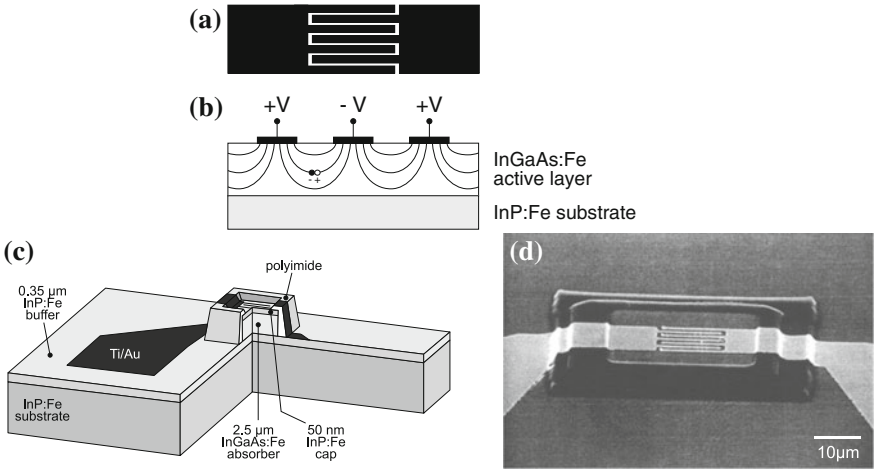

Fig. 22.24 Scheme of MSM photodetector with interdigital contacts in (**a**) plan view and (**b**) cross section. In part (**b**), the electric field lines are shown schematically together with an electron–hole pair ready to be separated. (**c**) Scheme of a MSM mesa structure, (**d**) SEM image of an InGaAs/InP MSM mesa photodetector. Parts (**c**) and (**d**) adapted from [1552]

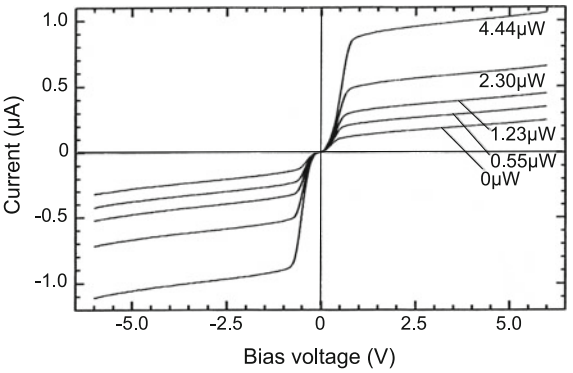

Fig. 22.25 dc I–V characteristic of an InGaAs/InP MSM photodetector (InP:Fe/InGaAs:Fe/InP:Fe, finger separation 1 μm, $\lambda = 1.3\,\mu$m) under illumination for dark environment (0 μW) and various illumination levels as labeled. Adapted from [1552]

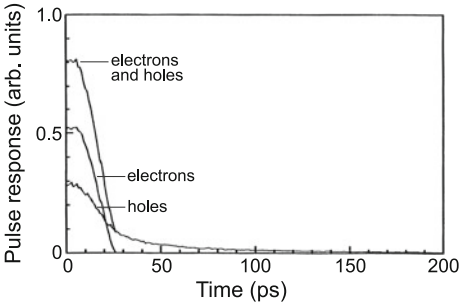

Fig. 22.26 Simulation of the time-dependent response of an InGaAs:Fe MSM photodetector to a short light pulse. Adapted from [1552]

to the holes that have the lower mobility and smaller drift saturation velocity. A similar dependence is found in experiment (Fig. 22.27a). For longer wavelengths the detector is slower since they penetrate deeper into the material and thus the charge carriers have a longer path to the contacts (cf. scheme in Fig. 22.24b). An important role is played by the finger separation; smaller finger separation ensures a more rapid carrier collection (Fig. 22.27b). In [1553] a bandwidth of 300 GHz was demonstrated for 100 nm/100 nm finger width and separation for LT-GaAs[8] and bulk GaAs, limited by the *RC* time constant. For 300 nm/300 nm fingers and a LT-GaAs a bandwidth of 510 GHz (pulsewidth of 0.87 ps) was reported, which is faster than the intrinsic transit time (1.1 ps) and not limited by the *RC* time constant (expected pulse width 0.52 ps), due to the recombination time (estimated to be 0.2 ps).

22.3.6 Avalanche Photodiodes

In an avalanche photodiode (APD) intrinsic amplification due to carrier multiplication (through impact ionization) in a region with high electric field is used to increase the photocurrent. The field is generated by a high reverse bias in the diode. In an ideal

[8] LT: grown at *low temperature*, i.e. containing many defects that reduce the carrier lifetime.

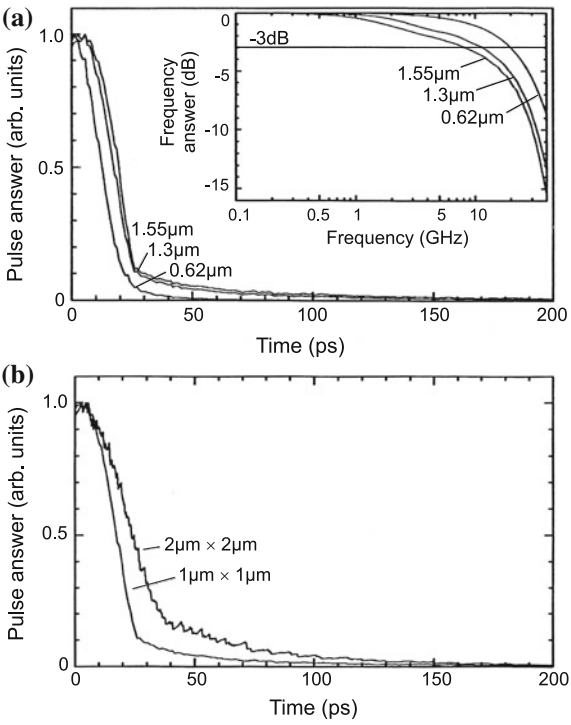

Fig. 22.27 (**a**) Experimental time-dependent response of an InGaAs:Fe MSM photodetector to a short light pulse for three different wavelengths, *inset* shows the frequency response from a Fourier transform. (**b**) Response of the MSM for two different finger widths and separations (both 1 or 2 μm, respectively), InGaAs layer thickness 2 μm, $\lambda = 1.3\,\mu$m and bias voltage 10 V. Adapted from [1552]

APD only one type of carrier is multiplied, resulting in the lowest noise. If electrons are injected into the field region at $x = 0$ (Fig. 22.28a), the multiplication factor for electrons is

$$M_\mathrm{n} = \exp(\alpha_\mathrm{n}\, w), \tag{22.49}$$

for $\alpha_\mathrm{p} = 0$. Typically, both carrier types suffer multiplication. If the electron and hole impact ionization coefficients are the same ($\alpha_\mathrm{n} = \alpha_\mathrm{p} = \alpha$), the multiplication factor for electrons and holes M is given by

$$M = \frac{1}{1 - \alpha\, w}. \tag{22.50}$$

The rms value of the current noise is the same as in the case of the pn-diode (22.25), only that now the gain M is added

$$\langle i_\mathrm{S}^2 \rangle = 2\, e\, \left(I_\mathrm{ph} + I_\mathrm{B} + I_\mathrm{D}\right)\, \langle M^2 \rangle\, B. \tag{22.51}$$

The term $\langle M^2 \rangle$ is written as $\langle M \rangle^2\, F(M)$ with $F(M) = \langle M^2 \rangle / \langle M \rangle^2$ being the excess noise factor that describes the additional noise introduced by the random nature of the impact ionization. For multiplication started with electron injection, it is given

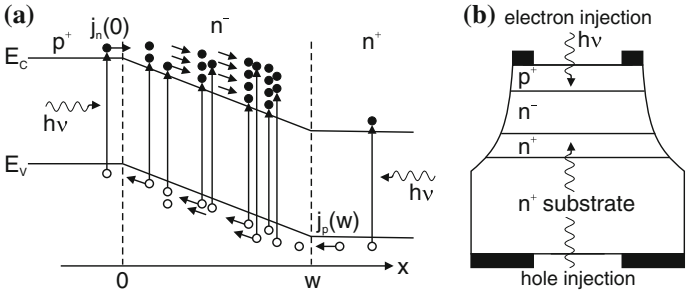

Fig. 22.28 Schematic band structure (**a**) and schematic device setup (**b**) of an avalanche photodiode (APD). From [1047]

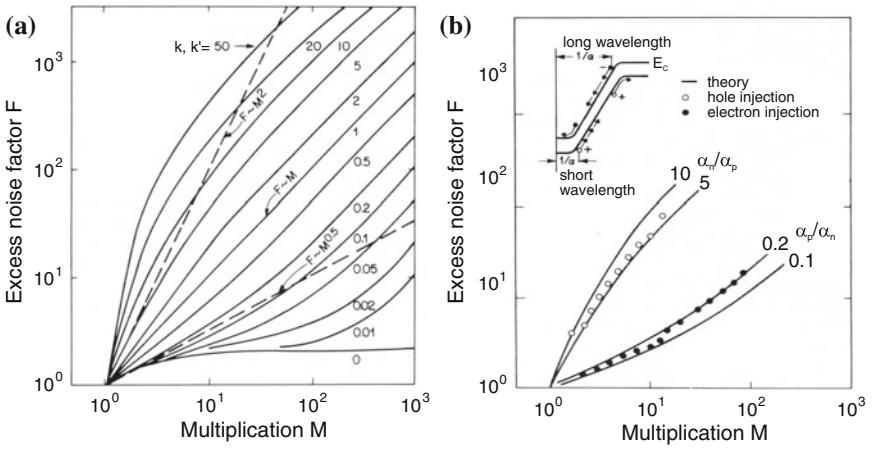

Fig. 22.29 (**a**) Excess noise factor for various values of the ratio of ionization coefficients k or k'. Adapted from [1554]. (**b**) Experimental results for F for a Si APD with 0.1 μA primary current. The *empty* (*full*) symbols are for short (long) wavelengths [primary hole (electron) current]. The *inset* shows the schematic band diagram of the np-diode under reverse bias. Adapted from [1555]

by [1554]

$$F(M) = k\,M + (1-k)\left(2 - \frac{1}{M}\right), \tag{22.52}$$

with $k = \alpha_\mathrm{p}/\alpha_\mathrm{n}$. For hole injection starting the multiplication (22.52) holds with k substituted by $k' = \alpha_\mathrm{n}/\alpha_\mathrm{p}$. In Fig. 22.29a the excess noise factor is shown versus the average multiplication for various values of k and k'.

Experimental data are shown in Fig. 22.29b for a Si APD. For short wavelengths absorption is preferential at the surface (n-region) and we have the case of hole injection. The data for the excess noise factor are fairly well fit with $k' \approx 5$. For longer wavelengths, the data for electron injection are fit by $k \approx 0.2 = 1/k'$.

For a fully modulated signal the signal-to-noise ratio is given by

$$S/N = \frac{(e\,\eta\,P_{\mathrm{opt}}/h\nu)^2/2}{2\,e\,\left(I_{\mathrm{ph}} + I_{\mathrm{B}} + I_{\mathrm{D}}\right)\,F(M)\,B + 4\,kT\,B/(R_{\mathrm{eq}}\,M^2)}. \tag{22.53}$$

If S/N is limited by thermal noise, the APD concept leads to a drastic improvement of noise.

The APD can used for single photon detection; then the variation in pulse height (with reasonable limits) plays no role for the count rate. The number of dark counts can be reduced by cooling the APD. Using constant fraction triggering, also the arrival time of the photon can be determined, allowing time resolution typically in the 100 ps regime.

A particular APD structure is known as a *solid-state multiplier*. It has separate absorption and amplification regions (SAM structure). In the low-field region the light is absorbed. One type of carrier is transported with the drift field E_{d} to the multiplication region in which a large field E_{m} is present and multiplication occurs. In Fig. 22.30a a homo-APD with SAM structure is shown. Regions with different electric field are created by a special doping profile.[9] A π-p-π structure leads to regions with homogeneous low and high field strengths.

The performance of a commercial silicon APD is depicted in Fig. 22.31. With increasing reverse bias, both the dark current and the multiplication factor increase. Before breakdown at about 77 V, the best ratio of M and I_{dark} is found. This particular diode has a typical excess noise factor of $F = 2$ for $M = 100$. Using (22.52), this shows that only one kind of carriers is multiplied.

In the case of a heterostructure-APD with SAM structure (Fig. 22.30b) absorption (of light with sufficiently long wavelength with an energy smaller the InP band gap) takes place only in the InGaAs layer. Since no light is absorbed in the multiplication region, the device functions similarly for front and back illumination. In [1557] a multi-stage InGaAs-based APD on InP is described that is optimized for electron multiplication in ten subsequent gain sections; a total gain of 10^3 is achieved with an excess noise factor F of about 40, belonging to an effective ionization ratio of $k = 0.036$.

22.3.7 Traveling-Wave Photodetectors

In a standard photodetector there was a tradeoff between the thickness of the absorption layer and the speed of the detector. In a traveling-wave photodetector the light absorption occurs in a waveguide such that for sufficient length L all incident light is absorbed. Complete absorption is achieved ('long' waveguide) if $L \gg (\Gamma\alpha)^{-1}$, α being the absorption coefficient and $\Gamma \leq 1$ being the optical confinement factor,

[9] Employing Poisson's equation $\partial(\epsilon_{\mathrm{s}}(x)E(x))/\partial x = \rho(x)$.

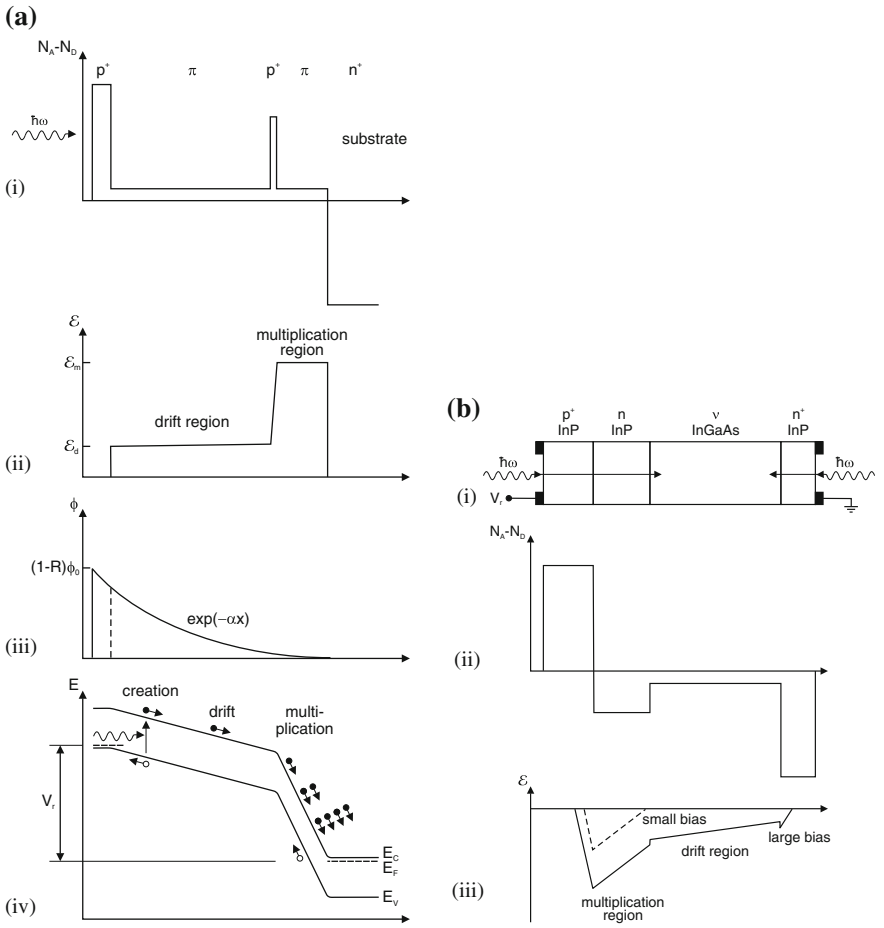

Fig. 22.30 (**a**) Homo-APD with SAM structure. (**i**) doping profile, (**ii**) electric field, (**iii**) photon flux or electron–hole pair generation rate and (**iv**) schematic band diagram under reverse voltage V_r with charge-carrier transport. The multiplication is for $\alpha_n \gg \alpha_p$. Adapted from [500]. (**b**) (**i**) Scheme of an InP/InGaAs hetero-APD with SAM structure, (**ii**) doping profile and (**iii**) electric field for small (*dashed line*) and large (*solid line*) reverse bias V_r. Adapted from [1047]

the geometrical overlap of the optical mode with the cross section of the absorbing medium (cf. also Sect. 23.4.4).

The electrical connections are designed along this waveguide on the sides (coplanar layout, Fig. 22.32). The bandwidth limitation due to a *RC* time constant is now replaced by the velocity match of the light wave $v_{opt} = c/n$ and the traveling electric wave in the contact lines $v_{el} \approx 1/\sqrt{LC}$. While the two waves travel along the waveguide, energy is transferred from the light wave to the electric wave. The 3 dB

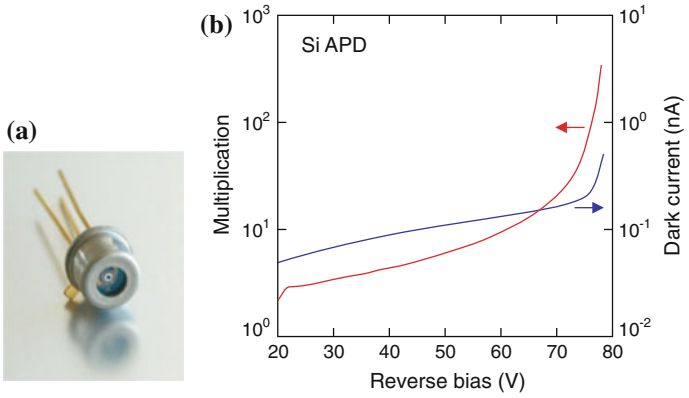

Fig. 22.31 (**a**) TO casing of Si APD. (**b**) Dark current (*blue*) (active area 0.2 mm^2) and multiplication factor (*red*) as a function of reverse bias (at 23 °C). Adapted from [1556]

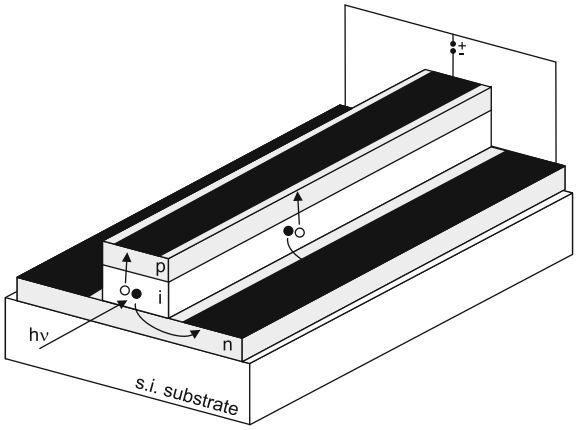

Fig. 22.32 Scheme of a traveling-wave photodetector with pin structure and coplanar contacts

bandwidth due to velocity mismatch B_{vm} (for impedance- matched, long waveguides) is given by [1558]

$$B_{\mathrm{vm}} = \frac{\Gamma\,\alpha}{2\pi}\,\frac{v_{\mathrm{opt}}\,v_{\mathrm{el}}}{v_{\mathrm{opt}} - v_{\mathrm{el}}}. \tag{22.54}$$

For a MSM structure, whose electrode separation has been designed with a self-aligned process (without extensive effort in lateral patterning) by an etch depth of a few 100 nm (Fig. 22.33), 3 dB cutoff frequencies beyond 500 GHz have been achieved (Fig. 22.34). The quantum efficiency of this detector was still 8.1 %.

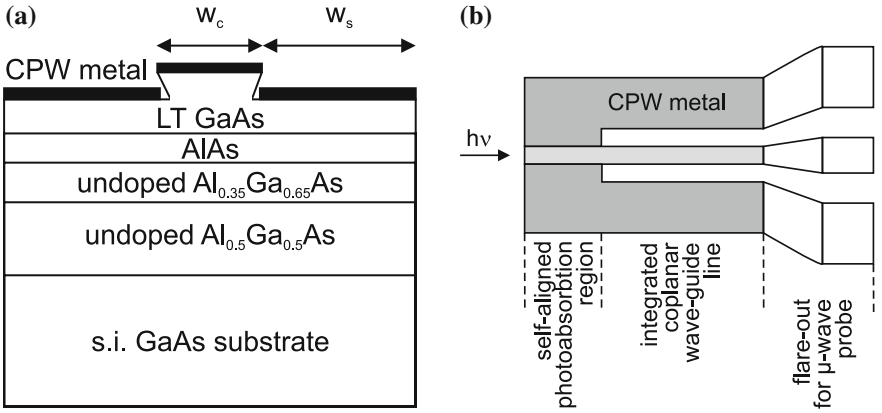

Fig. 22.33 Scheme of a MSM traveling-wave photodetector in (**a**) cross section and (**b**) plan view. Adapted from [1559]

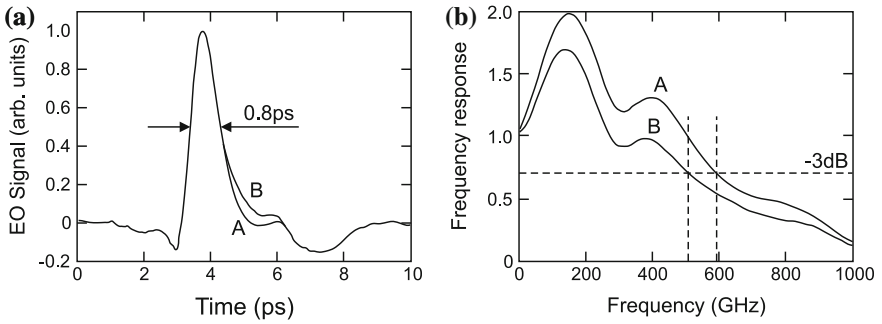

Fig. 22.34 (**a**) Pulse response (FWHM = 0.8 ps) and (**b**) frequency response (Fourier transform of time response) of a MSM traveling-wave photodetector (bias 5 V) for various illumination intensities, A: 1 mW, B: 2.2 mW. Adapted from [1559]

22.3.8 Charge Coupled Devices

The concept of the charge coupled device (CCD), an array of connected photodetectors serving as an image sensor, was devised by W.S. Boyle and G.E. Smith [1560] (Fig. 22.35) and realized [1561]. As textbook for further details [1562, 1563] may serve.

A MIS diode (mostly a silicon-based MOS diode) can be designed as a light detector. The diode is operated in deep depletion. When a large reverse voltage is applied, initially a depletion layer is formed and the bands are strongly bent as shown in Fig. 22.36b. We note that in this situation the semiconductor is not in thermodynamic equilibrium (as it is in Fig. 22.36d) when the quasi-Fermi level is constant throughout the semiconductor. The inversion charge has yet to build up.

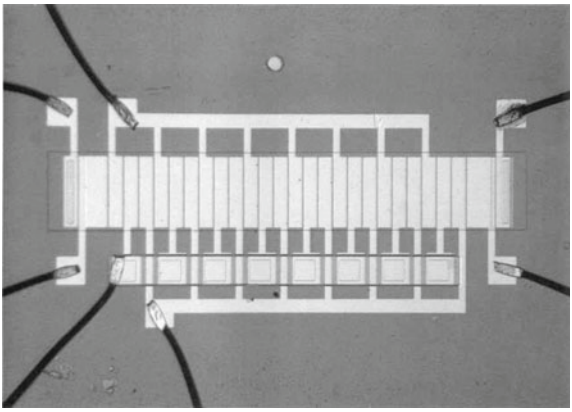

Fig. 22.35 First 8-bit charge coupled device (1970). The chip (size: $1.5 \times 2.5\,\mathrm{mm}^2$) consists of 24 closely packed MOS capacitors (*narrow rectangles* in the center grid). The *thick rectangles* at either end of the grid are input/output terminals

There are three mechanisms to generate the inversion charge. (a) generation–recombination, (b) diffusion from the depletion-layer boundary and (c) carrier generation by light absorption. Mechanisms (a) and (b) represent dark currents for the photodetector. The conductivity due to these two processes is shown in Fig. 22.37 and slowly builds up the inversion charge. Two temperature regimes are obvious; at low temperatures the generation dominates ($\propto n_\mathrm{i} \propto \exp\left(-E_\mathrm{g}/2kT\right)$), at high temperatures the diffusion ($\propto n_\mathrm{i}^2 \propto \exp\left(-E_\mathrm{g}/kT\right)$). The latter process can be strongly suppressed by cooling the device.

The gate voltage V_G and the surface potential Ψ_s are related to each other via

$$V_\mathrm{G} - V_\mathrm{FB} = V_\mathrm{i} + \Psi_\mathrm{s} = \frac{e\,N_\mathrm{A}\,w}{C_\mathrm{i}} + \frac{eN_\mathrm{A}w^2}{2\epsilon_\mathrm{s}}, \tag{22.55}$$

where w is the width of the depletion layer. w will be larger than w_m in thermodynamic equilibrium. The first term in the sum is $|Q_\mathrm{s}|/C_\mathrm{i}$ and the second is obtained by integrating the Poisson equation for the constant charge density $-eN_\mathrm{A}$ across the depletion layer. The elimination of w yields

$$V_\mathrm{G} - V_\mathrm{FB} = \Psi_\mathrm{s} + \frac{1}{C_\mathrm{i}}\sqrt{2e\epsilon_\mathrm{s}N_\mathrm{A}\Psi_\mathrm{s}}. \tag{22.56}$$

If light is absorbed in the depletion layer (process (c)), the hole (for p-Si) drifts towards the bulk material. The electron is stored as part of the signal charge Q_sig close to the oxide semiconductor interface (Fig. 22.36b).

$$V_\mathrm{G} - V_\mathrm{FB} = \frac{Q_\mathrm{sig}}{C_\mathrm{i}} + \frac{eN_\mathrm{A}w}{C_\mathrm{i}} + \Psi_\mathrm{s}. \tag{22.57}$$

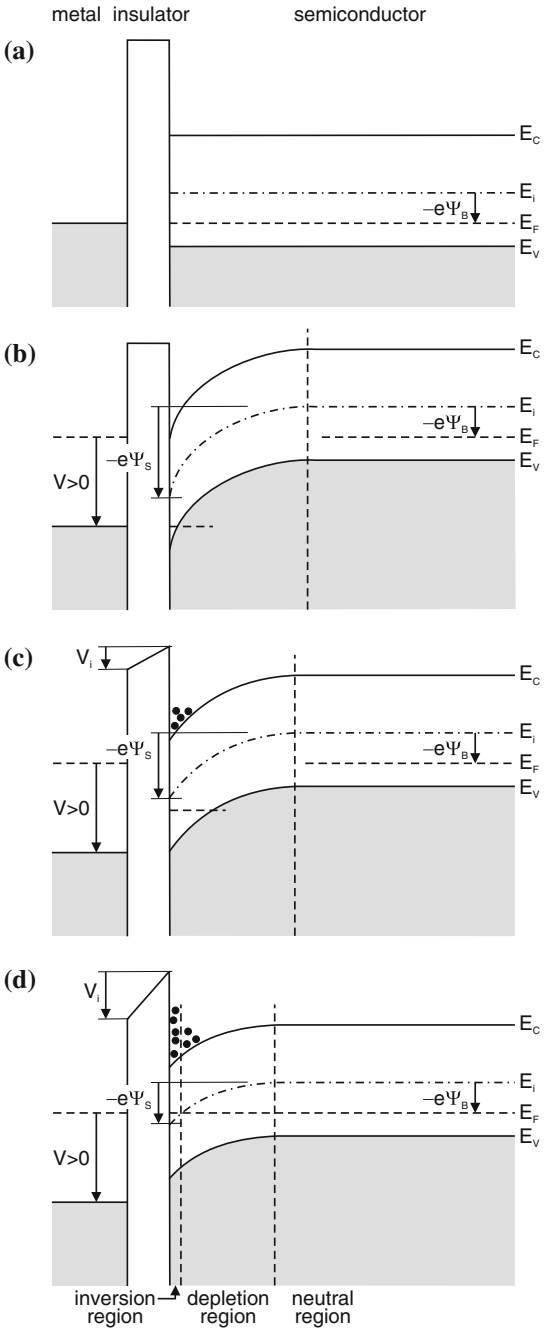

Fig. 22.36 Ideal MIS-diode (with p-type semiconductor) as photodetector (principle of a CCD pixel). (**a**) Without bias (see Fig. 21.29b). (**b**) Immediately after an external (reversely poled) voltage $V > 0$ has been applied, the surface potential is $\Psi_s = V$ and no charges have moved yet. (**c**) Strong depletion (still not in thermodynamic equilibrium) with signal charge and reduced surface potential $\Psi_s < V$. (**d**) The semiconductor in equilibrium (E_F is constant) with depletion and inversion layer (see Fig. 21.33). For all diagrams, $V = V - \mathrm{i} + \Psi_s$

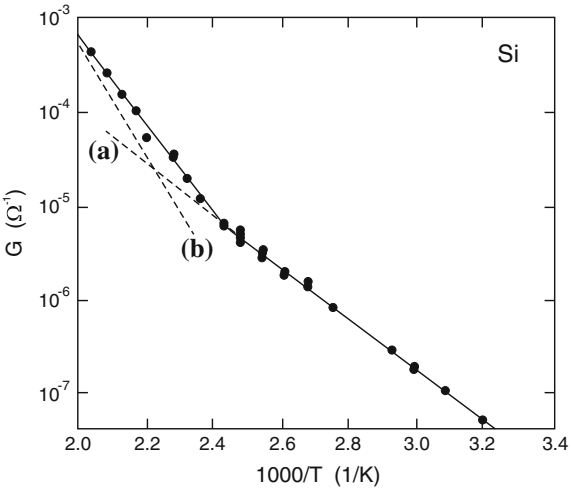

Fig. 22.37 Conductivity of a n-Si/SiO_2-diode as a function of temperature ($1/T$). The slope of the *dashed lines* is (a) ∼0.56 eV ($\approx E_g/2$) and (b) ∼1.17 eV ($\approx E_g$). Adapted from [1564]

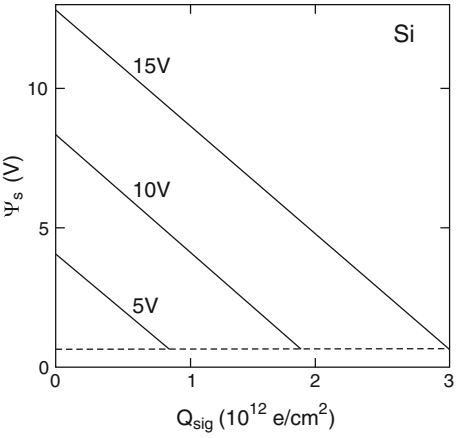

Fig. 22.38 Surface potential as a function of the signal charge Q_{sig} for various values of the bias $V_G - V_{FB}$ as labeled for a SiO_2/p-Si diode with $N_A = 10^{15}\,cm^{-3}$ and an oxide thickness of 100 nm. The *dashed line* represents the limit for inversion given by $\Psi_s = 2\Psi_B \approx 0.6\,V$. Adapted from [1565]

As a consequence of the increase in signal charge the potential well becomes shallower (22.57). For each gate voltage there is a maximum charge (well capacity). The maximum signal charge is reached for $\Psi_s \approx 2\Psi_B$ (Fig. 22.38).

In a charge coupled device (CCD) many light-sensitive MIS diodes, as described above, are fabricated in matrix form to create an image sensor. Upon application of a gate voltage they accumulate charge depending on the local exposure to light. The read out of this charge occurs by shifting the charge through the array to a read-out circuit. Therefore charge is transferred from one pixel to the next. Several schemes have been developed for this task. The three-phase clocking is shown schematically in Fig. 22.39. Other clocking schemes involve four, two or only one electrode per pixel [1566].

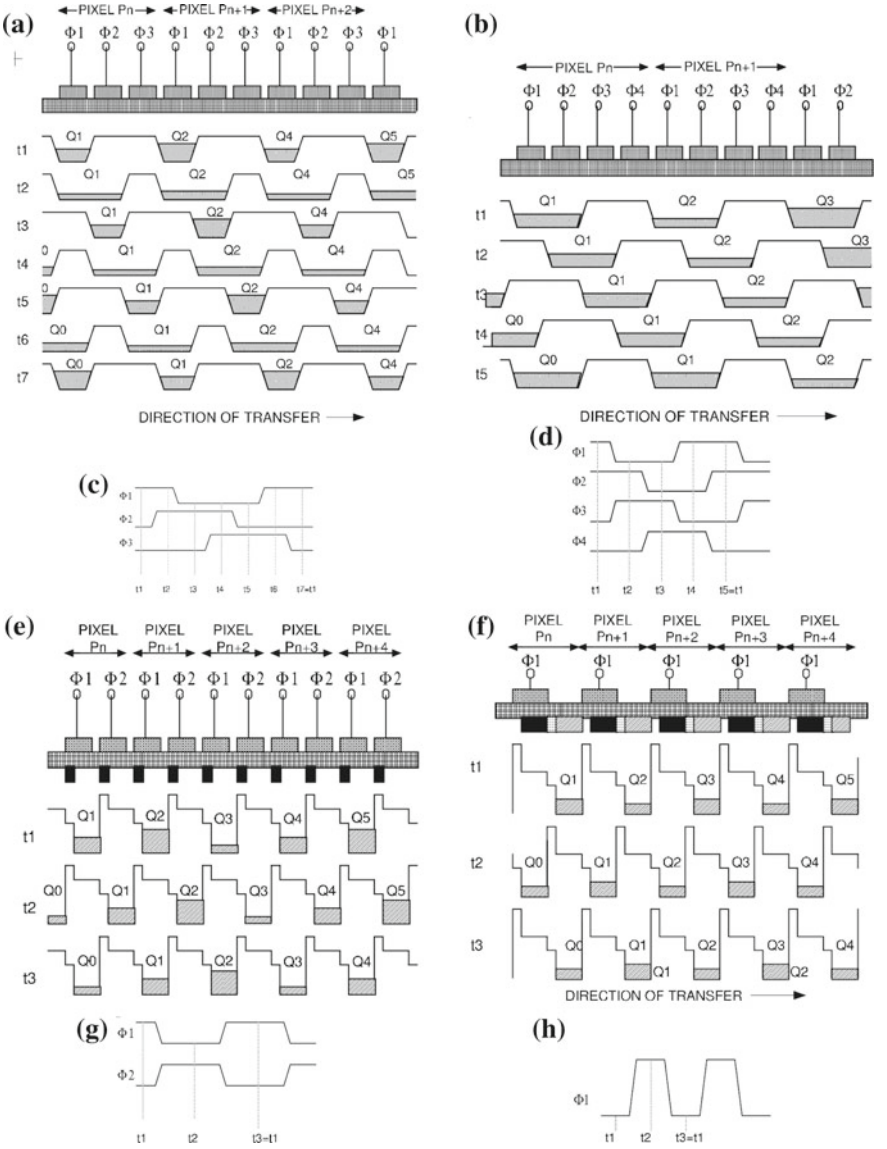

Fig. 22.39 (**a**) Three-phase CCD. Each pixel has three electrodes that can be switched independently (phases 1–3). (**b**, **e**, **f**) Schematic of CCDs with four, two or one phase, respectively. (**c**) (t_1) Charge accumulated after light exposure. A lateral potential well is formed along the row of pixels by the voltages at the three electrodes, e.g. $P_1 = P_3 = 5\,\mathrm{V}$, $P_2 = 10{,}\mathrm{V}$. (t_2–t_7) transfer of charge, (t_7) has the same voltages as (t_1), the charge has been moved one pixel to the right. (**d**, **g**, **h**) Timing schemes for 4-, 2- and 1-phase CCDs, respectively. From [1566]

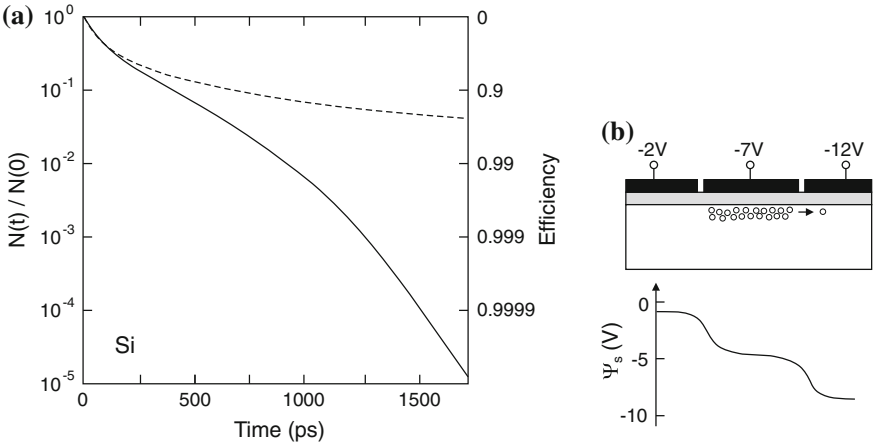

Fig. 22.40 (**a**) Efficiency of charge transfer with (*solid line*) and without (*dashed line*) the effect of the fringing field. (**b**) Schematic plot of the CCD electrodes and bias with the fringing field for a three-electrode CCD, oxide thickness 200 nm, doping $N_D = 10^{15}\,\mathrm{cm}^{-3}$. The electrode pitch is 4 μm and the gap between electrodes is 200 nm wide. Adapted from [1567]

Since the CCD sensor has many pixels (e.g. up to 4096) along a line, the charge transfer must be highly efficient. The transfer of charge carriers occurs via thermal (regular) diffusion, self-induced drift and the effect of the fringing field (inset of Fig. 22.40). The time constant with which the charge carriers move due to diffusion (in a p-type semiconductor) is

$$\tau_{\mathrm{th}} = \frac{4L^2}{\pi^2 D_{\mathrm{n}}}, \tag{22.58}$$

where L is the length of the electrode. For a sufficiently large charge packet the self-induced drift due to Coulomb repulsion is important. The decay of charge is then given by

$$\frac{Q(t)}{Q(0)} = \frac{t_0}{t + t_0}, \tag{22.59}$$

with $t_0 = \pi L^3 W_{\mathrm{e}} C_{\mathrm{i}}/(2\mu_{\mathrm{n}} Q(0))$. W_{e} is the width of the electrode. This dependence is shown as a dashed line in Fig. 22.40a. The last electrons are efficiently transferred by the drift induced by the fringing field of the electrodes (solid line in Fig. 22.40a). The origin of the fringing field is schematically shown in Fig. 22.40b; the minimum fringing field shown is 2×10^3 V/cm. In about 1–2 ns practically all $(1 - 10^{-5})$ charges are transferred. This enables clock rates of several 10 MHz.

For the clocking of the CCD the lateral variation of potential depth with the applied gate voltage is used. In Fig. 22.41 it is shown how a lateral variation of doping or oxide thickness creates a lateral potential well. Such structures are used to confine the row of pixels against the neighboring rows (channel stops, Fig. 22.42). In order

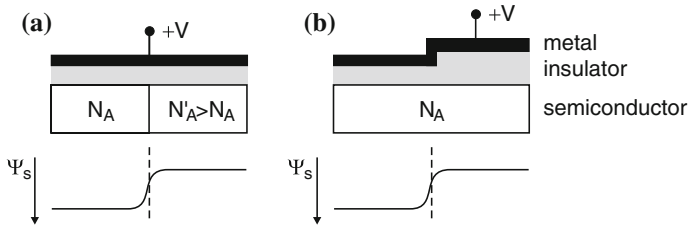

Fig. 22.41 Creation of a lateral potential well (barrier) in a MIS structure with (**a**) varying doping via diffusion or implantation and (**b**) varying (stepped) oxide thickness. *Upper row* shows schematic geometry, *lower row* depicts schematic lateral variation of the surface potential

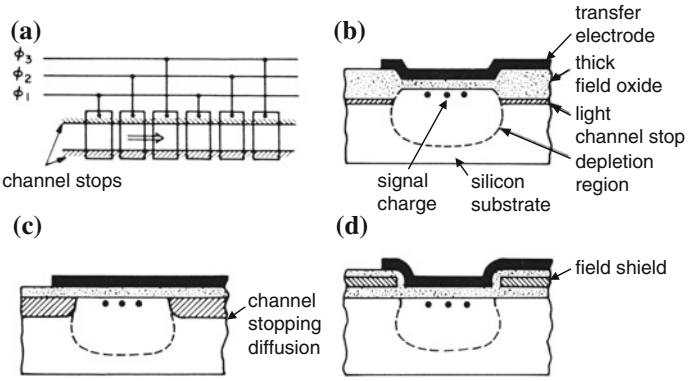

Fig. 22.42 (**a**) Schematic image of channel isolation. Cross section for channel isolation by (**b**) variation of oxide thickness, (**c**) highly doped region and (**d**) field effect. Adapted from [1568]

to avoid carrier loss at the interface between the oxide and the semiconductor a buried-channel structure is used (Fig. 22.43).

For front illumination parts of the contact electrodes shield the active area of the device. Higher sensitivity (in particular in the UV) is achieved for back illumination. For this purpose the chip is thinned (polished). This process is expensive and makes the chip mechanically less stable. For red/infrared wavelengths typically interference fringes occur for such thinned chips due to the small thickness. An increase in efficiency for front illumination can be achieved with an onchip microlens (Fig. 22.44).

For color imaging the CCD is covered with a three-color Bayer mask [1571] (Fig. 22.45a). On average there are one blue and one red pixel and two green pixels since green is the most prominent color in typical lighting situations. Thus each pixel delivers monochromatic information; RGB images are generated using suitable image software. Alternatives in high-end products are the use of a beam splitter, static color filters and three CCD chips, one for each color (Fig. 22.45b), or the time-sequential recording of three monochromatic images using one CCD chip and a rotating color-filter wheel (Fig. 22.45c). Another method is to shift

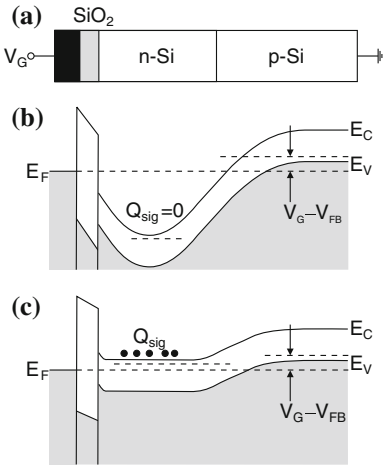

Fig. 22.43 (**a**) Schematic layers of MIS diode with buried-channel structure. Band diagram (**b**) after application of reverse voltage V_G and (**c**) with signal charge Q_{sig}. Adapted from [1569]

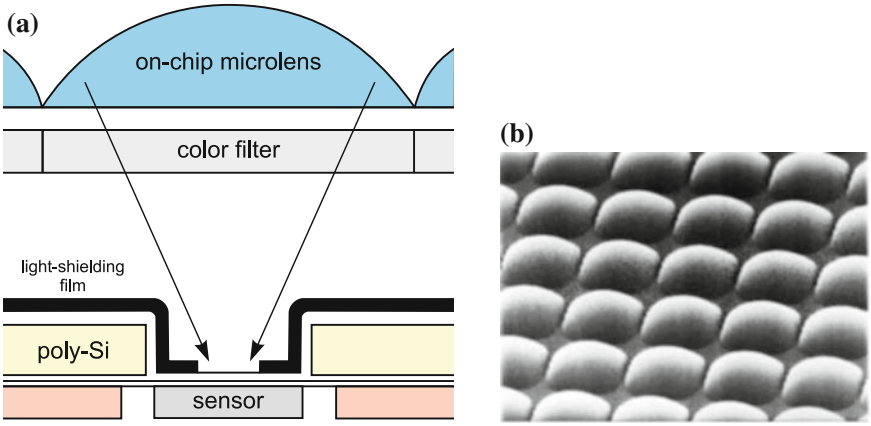

Fig. 22.44 (**a**) Scheme for enhancement of CCD efficiency for front illumination by application of an onchip microlens. (**b**) SEM image of an array of such microlenses. From [1570]

(by piezoelectric actors) the CCD by half or full pixels and take sequential exposures, increasing spatial and color resolution for still objects.

22.3.9 Photodiode Arrays

An array of photodiodes is also suitable to create an image sensor. During illumination each diode charges a capacitor that is read out with suitable electronics. Based on

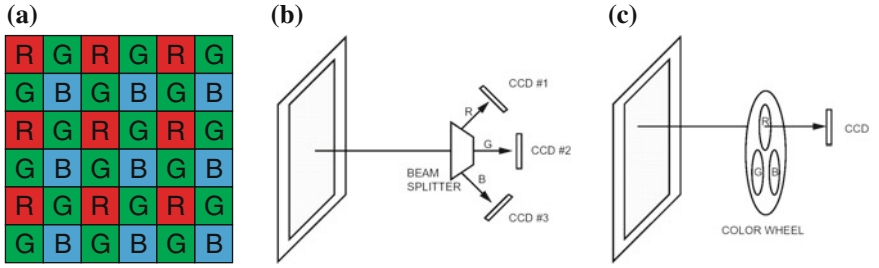

Fig. 22.45 (**a**) Arrangement of colors in a CCD Bayer [1571] color filter ('R': *red*, 'G': *green*, 'B': *blue*). Color splitting with (**b**) static color filters and (**c**) rotating color wheel. Parts (**b**) and (**c**) from [1566]

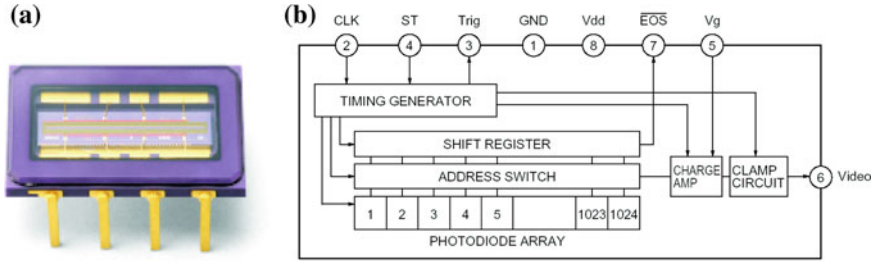

Fig. 22.46 (**a**) CMOS linear array sensor in a 8-pin package. (**b**) Block diagram, the built-in timing generator allows operation only with start and clock pulse inputs. Reprinted with permission from [1572]

CMOS technology (cf. Sect. 24.5.4) cheap image sensors can be made that show currently, however, inferior performance to CCDs. The built-in electronics allows simple outward connections (Fig. 22.46).

The three-color CCD image sensor did not offer RGB color information at each pixel. Therefore the spatial resolution of a color image is not directly given by the pixel distance. This is not a very dramatic drawback since human vision is more sensitive to intensity contrast than color contrast. However, RGB color information for each pixel would be desirable, giving higher resolution, in particular in professional photography. Such a sensor has been fabricated employing the wavelength dependence of the silicon absorption coefficient (Fig. 22.13). Blue light has the shortest and red light the largest penetration depth. By stacking three photodiodes on top of each other (Fig. 22.47) photocurrents at different penetration depth are recorded that can be used to generate a RGB value for each pixel.

In Fig. 22.48a a 16-channel array of silicon avalanche photodiodes is shown. It features a quantum efficiency of $>80\,\%$ between 760 and 910 nm. The pixel size is $648 \times 208\,\mu\text{m}^2$ on a $320\,\mu\text{m}$ pitch. The gain is 100 and the rise time 2 ns.

The InGaAs photodiode array in Fig. 22.48b is hybridized with CMOS read-out integrated circuits. It is useful for detection in the spectral range 0.8–1.7 µm. The

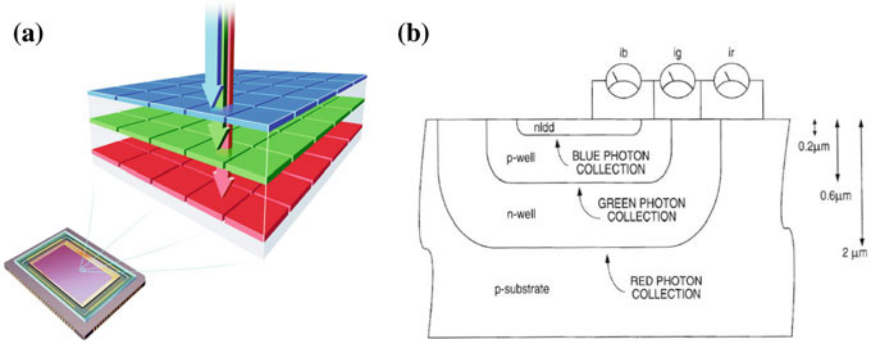

Fig. 22.47 (**a**) Scheme of image sensor with depth-dependent light detection. From [1573]. (**b**) Schematic layer sequence for three-color pixel. i_b, i_g and i_r denote the photocurrents for *blue*, *green* and *red* light, respectively. Adapted from [1574]

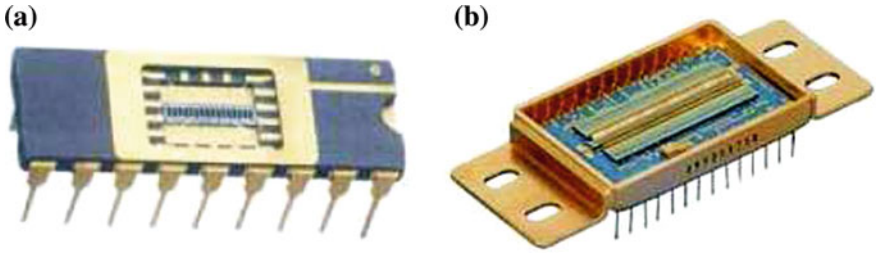

Fig. 22.48 (**a**) Array of 16 silicon APDs. From [1575]. (**b**) 1024-pixel InGaAs photodiode array. From [1576]

asymmetric diode size of $25 \times 500\,\mu\text{m}$ is designed for use in the focal plane of a monochromator.

Another special type of photodiode array is the four-quadrant detector. A light beam generates four photocurrents I_a, I_b, I_c, I_d of the respective parts (Fig. 22.49a). A beam deviation in the horizontal or vertical direction can be detected from the (signed) signals $(I_a + I_d) - (I_b + I_c)$ or $(I_a + I_b) - (I_c + I_d)$, respectively. We note that these signals can also be normalized to the total beam intensity $I_a + I_b + I_c + I_d$.

22.4 Solar Cells

Solar cells are light detectors, mostly photodiodes, that are optimized for the (large-area) conversion of solar radiation (light) into electrical energy. A 1993 review of the historic development of photovoltaics is given in [1578]. The latest data on solar cell efficiencies are compiled in the Solar Cell Efficiency Tables [1579].

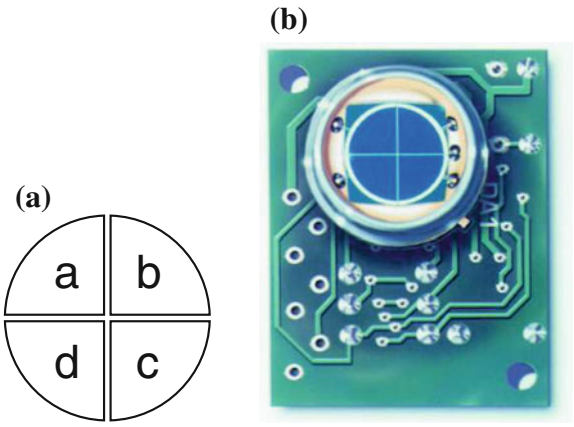

Fig. 22.49 (**a**) Scheme of four-quadrant photodetector with sections 'a', 'b', 'c' and 'd', (**b**) image of four-quadrant silicon photodetector with circuit board. Part (**b**) from [1577]

22.4.1 Solar Radiation

The sun has three major zones, the core with a temperature of 1.56×10^7 K and a density of 100 g/cm^3 in which 40 % of the mass is concentrated and 90 % of the energy is generated, the convective zone with a temperature of 1.3×10^5 K and a density of 0.07 g/cm^3, and the photosphere with a temperature of 5800 K and low density ($\sim 10^{-8}$ g/cm^3). The radius is 6.96×10^8 m and is about 100 times larger than that of the earth (6.38×10^6 m). The distance sun–earth is 1.496×10^{11} m. The angle under which the sun disk appears on earth is 0.54°. An energy density of 1367 ± 7 W/m^2 arrives at the earth in front of its atmosphere.

This value and the according spectrum of the sun's emission, which is similar to a blackbody with temperature 5800 K (Fig. 22.50), is termed air mass zero (AM0). The total energy that reaches the earth from the sun is 1.8×10^{17} W per year. This value is 10^4 times the world's primary energy need.

Air mass zero (AM0) is important for solar cells in satellites. When the solar spectrum passes the earth's atmosphere, it is changed with regard to its shape and the total energy density due to gas absorption (ozone, water, CO_2, ...). Depending on the meridian of the sun γ (Fig. 22.51), the spectrum on the surface of the earth is termed AMx with $x = 1/\sin\gamma$. In spring and fall (March 21st and September 21st), Leipzig (51°42′N latitude) has about AM1.61. At the summer (June 21st) and winter (December 21st) solstices the air mass in Leipzig is AM1.13 ($\gamma = 61.8°$) and AM3.91 ($\gamma = 14.8°$), respectively. Additionally, the duration of sunshine and thus the light power density is regionally different across the earth due to climate and weather (Fig. 22.52). For AM1.5, the incident power density is 844 W/m^2.

The global radiation reaching a photovoltaic cell has three components: (i) the direct radiation, (ii) diffuse radiation and (iii) reflected radiation. The relative amounts

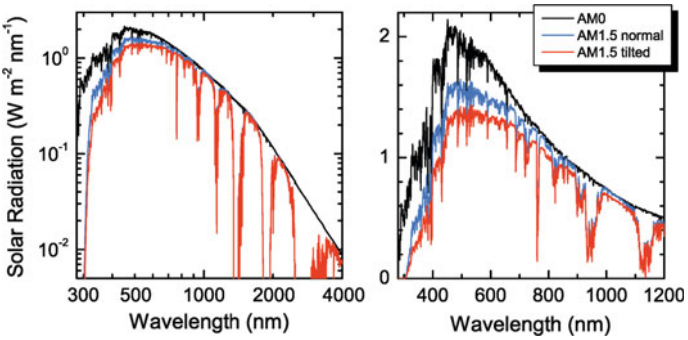

Fig. 22.50 Solar spectra (power per area and wavelength interval) for AM0 (*black line*, extraterrestrial irradiance) and AM1.5 (sun at 41.8° elevation above horizon) for direct normal irradiance (*blue line*) and global total irradiance (*red line*) on a sun facing surface (tilted 37° towards the equator). Left (right) graph in log–log (linear) scales

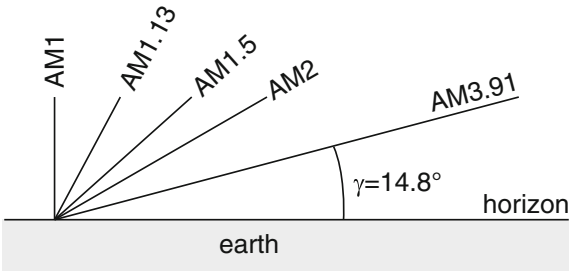

Fig. 22.51 Schematic path of sunlight through the atmosphere and definition of the air mass AM*x*

and their spectra depend on details such as the climate (e.g. humidity) or the environment (e.g. outdoors vs. urban).

22.4.2 Ideal Solar Cells

When a solar cell made from a semiconductor with a band gap E_g is irradiated by the sun, only photons with $h\nu > E_g$ contribute to the photocurrent and the output power. The I–V characteristic under illumination (Fig. 22.53) is given by

$$I = I_s \left[\exp(\beta V) - 1\right] - I_L, \tag{22.60}$$

with I_L being the current due to generation of excess carriers by the absorption of the sunlight. Assuming a simple n^+p-diode solar cell model, the current consists of two components: the depletion layer current j_{DL} from carriers absorbed in the depletion

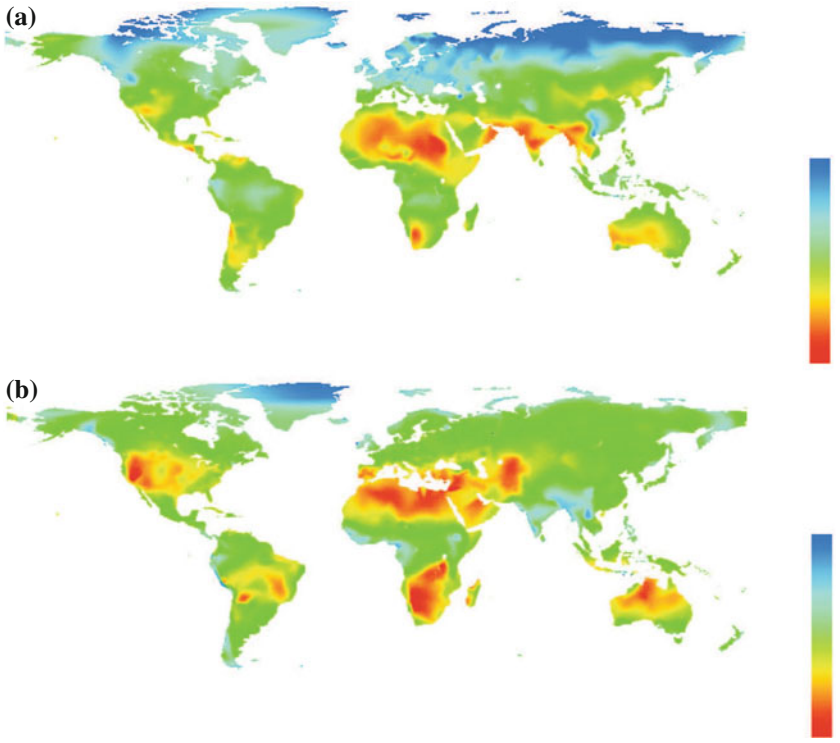

Fig. 22.52 Global sunshine distribution in (**a**) January and (**b**) July. The sunshine fraction is the actual number of bright sunshine hours over the potential number, and is thus expressed as a percentage figure. The color scale reaches from 0 to 100 %. The sunshine data are in a 0.5° grid based on data from [1580]

layer (field region) and the diffusion current j_D from absorption in the neutral region ($j = I/A$).

For the drift current out of the depletion layer of width w it can be assumed that it is collected fast and recombination plays no role. Thus (cmp. (22.31))

$$j_{DL}(\lambda) = e\, n_{ph}(\lambda)\,[1 - R(\lambda)]\,[1 - \exp(-\alpha(\lambda)\, w)], \tag{22.61}$$

where λ is the wavelength of the incident radiation, R the reflectance of the surface, α the absorption coefficient and $n_{ph}(\lambda)$ the photon flux (number of photons per area and time) at the given wavelength. For a solar spectrum, an integral needs to be performed over the spectral distribution:

$$j_{DL} = \int j_{DL}(\lambda)\, d\lambda, \tag{22.62}$$

the total photon flux being $n_{ph} = \int n_{ph}(\lambda)\, d\lambda$.

The diffusion current collected at the back contact is obtained from solving (10.71) (now for electrons in p-type material) with the appropriate boundary conditions (reversely bias depletion layer, $n_p(w) = 0$, $\Delta n_p(\infty) = n_p(\infty) - n_{p_0} = 0$ [1581]):

$$j_D(\lambda) = e\, n_{ph}(\lambda)\,[1 - R(\lambda)]\,\frac{\alpha\, L_n}{1 + \alpha\, L_n}\,\exp(-\alpha\, w) + e\, n_0\,\frac{D_n}{L_n}, \tag{22.63}$$

Dropping the wavelength dependence and neglecting the dark term, the usual formula is obtained,

$$j_L = e\, n_{ph}\,(1 - R)\left[1 - \frac{\exp(-\alpha\, w)}{1 + \alpha\, L_n}\right], \tag{22.64}$$

The last bracket represents the quantum efficiency [1582]. The model can be extended for taking into account a non-zero surface recombination velocity at the back contact at finite distance [1583].

Here a voltage *independent* photo-generated current I_L is assumed. If the diffusion length is small compared to the transport path, the carrier collection efficiency η_c becomes voltage dependent [1582]. The reduction of the diffusion potential for forward voltage decreases the carrier collection efficiency [1584], possibly to zero close to the built-in voltage.

The saturation current density is given by (21.128) and (21.129)

$$j_s = \frac{I_s}{A} = e\, N_C\, N_V\left(\frac{1}{N_A}\sqrt{\frac{D_n}{\tau_n}} + \frac{1}{N_D}\sqrt{\frac{D_p}{\tau_p}}\right)\exp\left(-\frac{E_g}{kT}\right), \tag{22.65}$$

with A being the cell area.

The voltage at $I = 0$ is termed the *open-circuit* voltage V_{oc}, the current at $V = 0$ is termed the *short-circuit* current $I_{sc} = I_L$ (Fig. 22.53). Only a part of the rectangle $I_{sc} \times V_{oc}$ can be used for power conversion. By choice of the load resistance R_L, the work point is set. At I_m and V_m, the generated power $P_m = I_m V_m$ is maximal. The filling factor F is defined as the ratio

$$F = \frac{I_m\, V_m}{I_{sc}\, V_{oc}}. \tag{22.66}$$

The open-circuit voltage is given by

$$V_{oc} = \frac{1}{\beta}\,\ln\left(\frac{I_L}{I_s} + 1\right) \cong \frac{1}{\beta}\,\ln\left(\frac{I_L}{I_s}\right) \tag{22.67}$$

and increases with increasing light power and decreasing dark current. The output power is

$$P = I\, V = I_s\, V\ (\exp(\beta V) - 1) - I_L\, V. \tag{22.68}$$

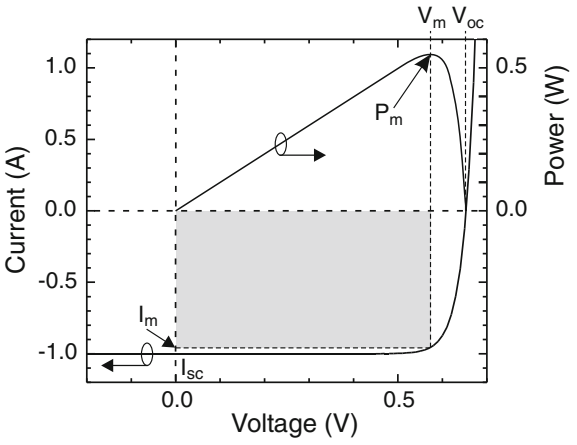

Fig. 22.53 Schematic I–V characteristics of a solar cell under illumination (*left scale*) and extracted power (*right scale*). The *grey area* is the maximum power rectangle with $P_\mathrm{m} = I_\mathrm{m} V_\mathrm{m}$

The condition $\mathrm{d}P/\mathrm{d}V = 0$ yields the optimal voltage at which the solar cell has to be operated and is given by the implicit equation

$$V_\mathrm{m} = \frac{1}{\beta} \ln\left(\frac{I_\mathrm{L}/I_\mathrm{s} + 1}{1 + \beta V_\mathrm{m}}\right) = V_\mathrm{oc} - \frac{1}{\beta} \ln\left(1 + \beta V_\mathrm{m}\right). \tag{22.69}$$

The current at maximum power is

$$I_\mathrm{m} = I_\mathrm{L}\left(1 - \frac{1 - \beta V_\mathrm{m}\, I_\mathrm{s}/I_\mathrm{L}}{1 + \beta V_\mathrm{m}}\right) \cong I_\mathrm{L}\left(1 - \frac{1}{\beta V_\mathrm{m}}\right). \tag{22.70}$$

E_m is the energy that is delivered per photon at the load resistor at the power maximum. The maximum power is $P_\mathrm{m} = I_\mathrm{L} E_\mathrm{m}/e$ and E_m is given by

$$E_\mathrm{m} \cong e\left[V_\mathrm{oc} - \frac{1}{\beta} \ln\left(1 + \beta V_\mathrm{m}\right) - \frac{1}{\beta}\right]. \tag{22.71}$$

The ideal solar cell has a (power) conversion efficiency $\eta = P_\mathrm{m}/P_\mathrm{in}$ that can be determined from Fig. 22.54a.

The right curve (1) in Fig. 22.54a shows the integral number n_ph of photons in the solar spectrum (per area and time) with an energy larger than a given one (E_g). For a given value of n_ph, the left curve (2) represents the value of E_m. The efficiency is the ratio of $E_\mathrm{m} n_\mathrm{ph}$ and the area under curve (1). The efficiency as a function of the band gap is shown in Fig. 22.55a. It has a fairly broad maximum such that many semiconductors can be used for solar cells, in principle. The maximum theoretical efficiency for a single junction is 31 % for nonconcentrated sunlight (AM1.5). This limit corresponds to the classic Shockley–Queisser limit [1586–1588], assuming radiative recombination as the only charge-carrier recombination mechanism. In [1589], the limit for a single material is found to be 43 % for an optimally tailored

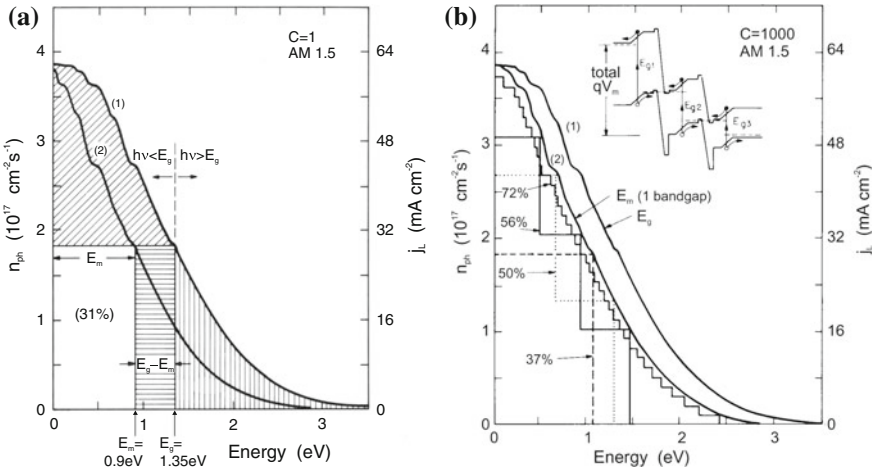

Fig. 22.54 (**a**) Number of photons n_{ph} per area and time in the sun spectrum (AM1.5, $C = 1$ sun) with an energy larger than a cutoff energy (*curve* 1) and graphical method to determine the quantum efficiency (from *curve* 2). Adapted from [1585]. (**b**) Number of photons in concentrated solar spectrum (AM1.5, $C = 1000$ sun) with an energy larger than a given energy and graphical method to determine the quantum efficiency of multi-junction solar cells. Adapted from [500] after [1585]

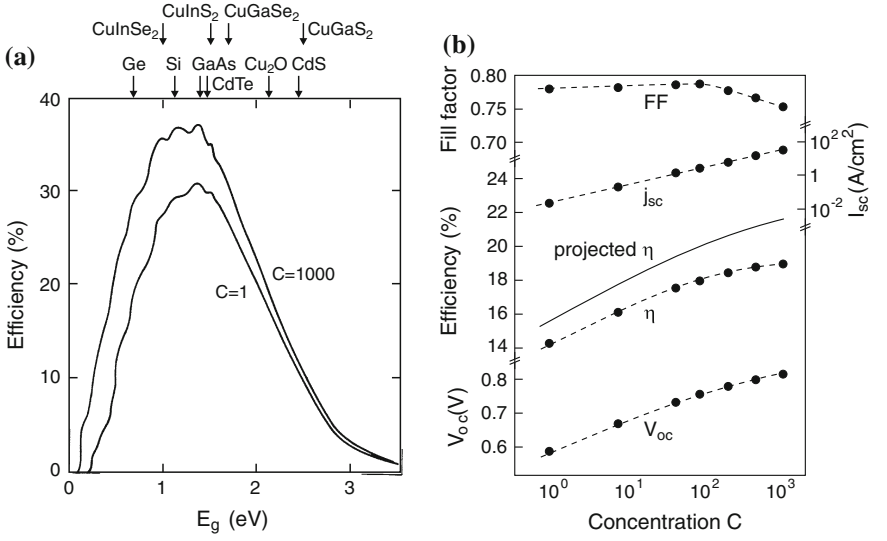

Fig. 22.55 (**a**) Ideal quantum efficiency of solar cells (single junction) as a function of band gap and light concentration C. The band gaps of some important semiconductors are denoted by *arrows*. Adapted from [1585]. (**b**) Properties of a silicon solar cell (water cooled) as a function of light concentration. *Solid line* is theoretically projected efficiency, *dashed lines* are guides to the eye. Adapted from [500]

band structure that allows carrier multiplication by optically excited hot carriers. The solar cell as a heat engine is dicussed in [1590].

When the sunlight is concentrated, e.g. by a lens, the efficiency increases (Fig. 22.55b). The short-circuit current increases linearly. The effect is mostly due to the increase of the open-circuit voltage. For $C = 1000$, the maximum theoretical efficiency for a single-junction solar cell is 38 %.

A further increase of efficiency can be achieved with multiple junctions using various materials for absorption. In a tandem cell (two junctions), the upper layer absorbs the higher-energy photons in a wide band gap material. The material with the lower band gap makes use of the low-energy photons. Thus, the cell works with two different values of E_m (Fig. 22.54b). With band gaps of 1.56 and 0.84 eV, an efficiency of 50 % can be reached theoretically. With three materials 56 %, and for a large number of materials 72 % is the limit. Between the junctions, tunneling diodes (Sect. 21.5.9) must be used to allow carrier transport through the entire structure. It is a nontrivial task to fabricate multiple heterojunctions due to incompatibilities of the lattice constants. Besides heteroepitaxy, wafer bonding can also be used for fabrication. A lattice-matched InGaP/GaAs/InGaAsN cell seems a viable solution for high-efficiency solar cells.

22.4.3 Real Solar Cells

For a real solar cell, the effect of parallel resistance R_{sh} (shunt resistance due to leakage current, e.g. by local shorts of the solar cell) and serial resistance R_s (due to ohmic loss) must be considered. The I–V characteristic is then (cf. (21.154))

$$\ln\left(\frac{I + I_L}{I_s} - \frac{V - IR_s}{I_s R_{sh}} + 1\right) = \beta\,(V - I\,R_s)\,. \tag{22.72}$$

The serial resistance affects the efficiency more strongly than the shunt resistance (Fig. 22.56). Therefore, it is frequently enough to consider R_s only and use (cf. Fig. 21.153)

$$I = I_s\,\exp\left(\beta(V - I\,R_s)\right) - I_L. \tag{22.73}$$

In the example of Fig. 22.56, a serial resistance of 5 Ω reduces the filling factor by a factor of about four.

At open circuit voltage the photo-generated carriers have nowhere to go; in an ideal solar cell, the only process is the radiative recombination and the photon escape. Certainly, the internal quantum efficiency should be high and the non-radiative recombination rate small compared to the radiative one (cf. Sect. 10.10). The open circuit voltage and thus the energy conversion efficiency depends also on the light extraction efficiency χ_{ex} which will be discussed in greater detail in the context of LEDs (Sect. 23.3.3). Based on [1592, 1593], the open circuit voltage V'_{oc} of a real solar cell is *reduced* ($\chi_{ex} \leq 1$) from the ideal value V_{oc} given in (22.67),

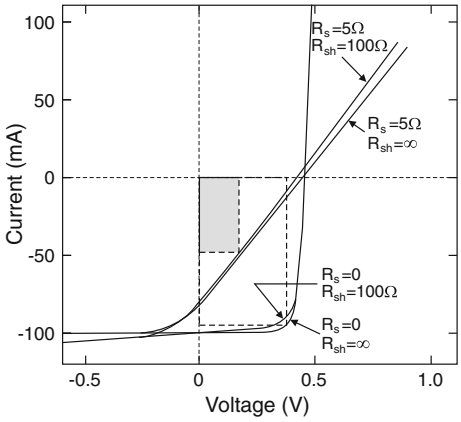

Fig. 22.56 *I–V* characteristics of a solar cell considering shunt and series resistances R_s and R_{sh}, respectively. The efficiency of the real cell (*shaded* power *rectangle*) is less than 30 % of that of the ideal cell. Adapted from [1591]

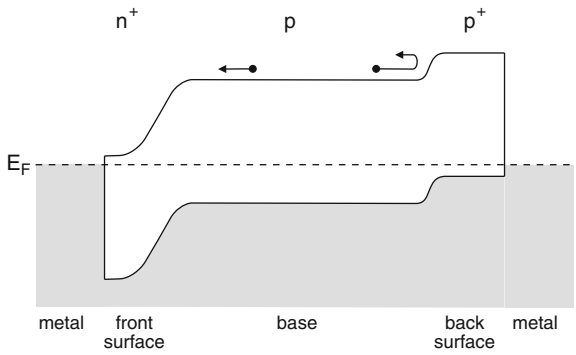

Fig. 22.57 Increase of the carrier collection efficiency by a back surface field. Adapted from [1594]

$$V'_{oc} = V_{oc} + \beta^{-1} \ln \chi_{ex}. \tag{22.74}$$

22.4.4 Design Refinements

In order to collect electrons most efficiently, a back surface field is used (Fig. 22.57). A higher-doped region at the back contact creates a potential barrier and reflects electrons back to the front contact.

An important point for optimization is the management of the reflection at the solar cell surface. First, a dielectric antireflection (AR) layer (or multi-layers) can be used. These layers should have a broad AR spectrum. Additionally, a textured surface reduces reflection (Fig. 22.58d), giving reflected photons a second chance for penetration (Fig. 22.58c). The reflectance of bare Si, 35 %, can be reduced to 2 %. An AM0 efficiency of over 15 % was reached using textured multi-crystalline cells. Alkaline KOH-based etches attack Si (001) anisotropically and yield pyramidal structures (Fig. 22.58b) with {111} facets. Recently an acidic HF/HNO_3-based

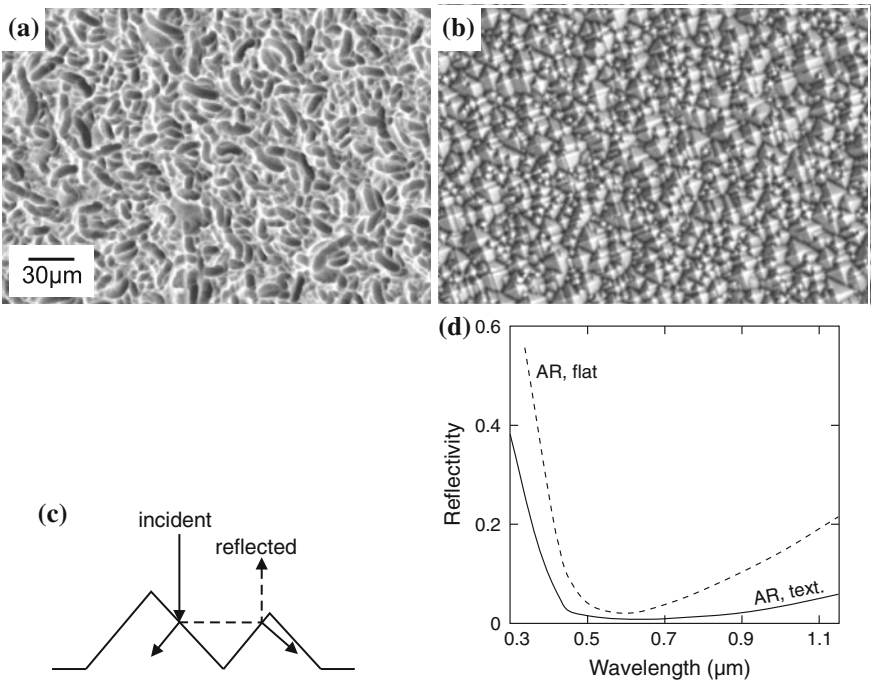

Fig. 22.58 (**a**) SEM image of topology of acidically etched multi-crystalline silicon wafer. (**b**) SEM image of alkaline etched mono-crystalline silicon wafer. (**c**) Exemplary light path. (**d**) Reflectance of antireflection-coated flat (*dashed line*) and textured (*solid line*) surface). Parts (**a**) and (**b**) adapted from [1596], part (**d**) adapted from [1597]

process has been established [1595], resulting in a worm-like surface pattern on multi-crystalline silicon wafers (Fig. 22.58a) with superior anti-reflection properties.

During its course over the sky during the day, the sun changes its angle towards a fixed solar cell.[10] A tracking mechanism can optimize the angle of incidence during the day and increase the overall efficiency of the solar cell (Fig. 22.59).

22.4.5 Modules

In order to cover a large area and supply certain values of output voltage and current, several solar cells are connected into modules. Arrays are built up from several modules (Fig. 22.60). If solar cells are connected in parallel, the total current increases; if they are connected in series, the output voltage increases. We note that in partially shadowed modules the reverse characteristics of solar cells are important [1600, 1601]; local breakdown can lead to hot spots and irreversible degradation.

[10] We are of course aware that the earth rather rotates around the sun.

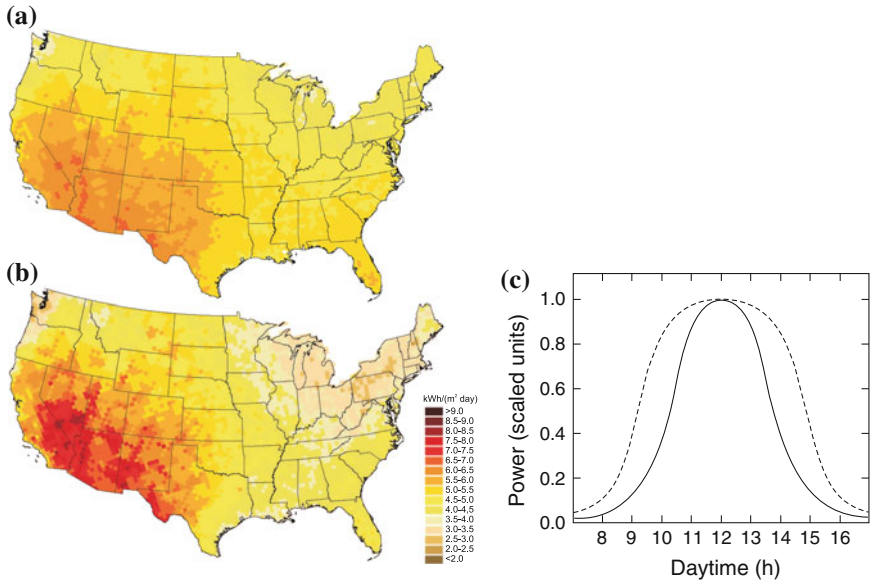

Fig. 22.59 Annual average solar energy (in kWh/(m^2 day)) for (**a**) an optimally tilted south-facing fixed panel and (**b**) an optimally two-axis tracked panel (in mainland US). Adapted from [1598]. (**c**) Power generation of a solar cell versus time (in daytime hours) for a stationary setup facing the sun at constant angle (*solid line*) and mounting with tracking (*dashed line*) to optimize the angle towards the sun. Adapted from [1599]

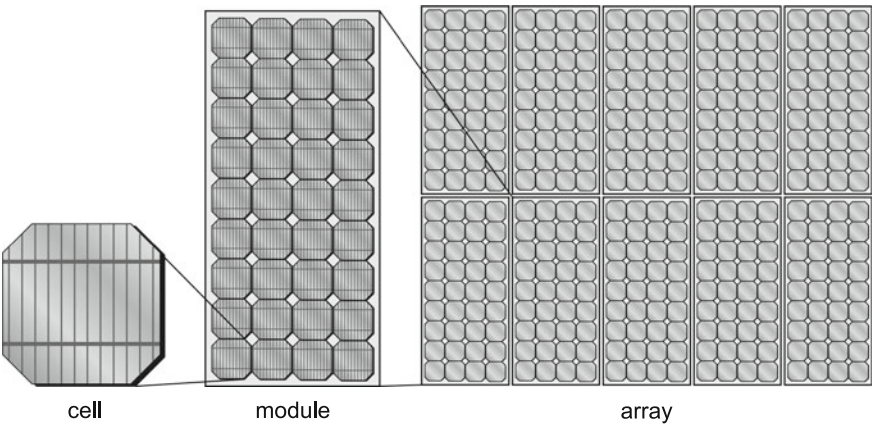

Fig. 22.60 Schematic drawing of a solar cell (with contact grid), a module (36 cells) and an array of ten modules

22.4.6 Solar-Cell Types

First Generation Photovoltaics

Silicon is the most frequently used material for solar cells. Cells based on single-crystalline silicon (wafers) have the highest efficiency but are the most expensive (Fig. 22.62a). The efficiencies of various solar cells are compiled in Table 22.1. Polycrystalline (for large grains also called multi-crystalline) silicon (Fig. 22.62b) is cheaper but offers less performance. Material design is oriented towards increasing the grain size and/or reducing their electrical activity. Grain boundaries act as recombination centers with a surface (i.e. interface) recombination velocity of 10^2 cm/s [1602] for particular, electrically fairly inactive grain boundaries, several 10^3 cm/s [1603], several 10^4 cm/s [1604] or even 10^5–10^7 cm/s [1605]. The grain boundaries reduce the effective diffusion length and thus carriers recombine before they can reach the contacts. A detailed theory of solar cell performance for polycrystalline material has been worked out in [1604] and explains the reduction of efficiency with decreasing grain size as shown in Fig. 22.61.

These solar cells are also called 'first-generation' photovoltaics. Thin sheets of crystalline silicon drawn from a melt between two seed crystals in a modified CZ growth (sheet silicon or ribbon silicon) allow cheaper production compared to cells based on 'traditional' polished wafers cut from a large silicon rod. Silicon made particularly for solar cell use is called 'solar-grade' silicon.

Second Generation Photovoltaics

Even cheaper are solar cells from amorphous silicon (Fig. 22.62c). Since silicon is an indirect semiconductor, a fairly thick layer is needed for light absorption. If direct band gap semiconductors are used, a thin layer ($d \approx 1\,\mu$m) is sufficient for complete light absorption. Such cells are called thin-film solar cells. A typical material class used in this type of cell are chalcopyrites, such as $CuInSe_2$ (CIS). The band gap is around 1 eV, which is not optimal. An improvement can be achieved by adding

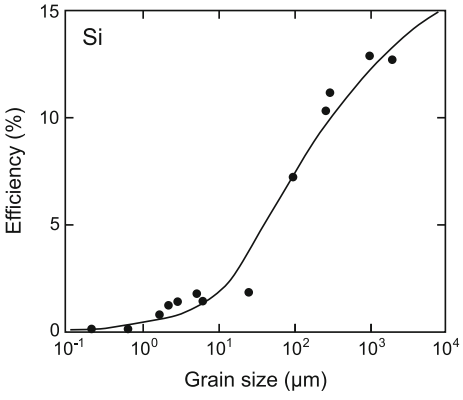

Fig. 22.61 Theoretical dependence of the effect of grain size on efficiency of polycrystalline solar cells (*solid line*) with experimental data points (*circles*). Adapted from [1604]

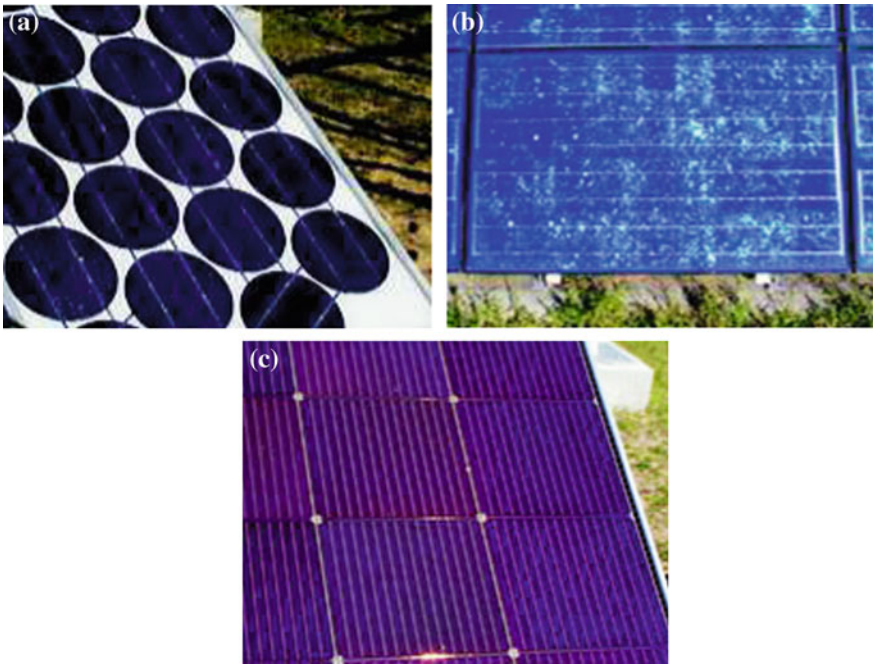

Fig. 22.62 Various types of solar cells: (**a**) monocrystalline silicon solar cell, (**b**) polycrystalline solar cell, (**c**) amorphous silicon solar cell. From [1700]

Table 22.1 Record efficiency of various solar cells (AM1.5, 1000 W/cm^2, 25° unless noted otherwise)

Cell material/type	Efficiency (%)	V_{oc} (V)	j_{sc} (mA/cm^2)	FF (%)	Date m/y
Si (crystalline)	25.6 ± 0.5	0.740	41.8	82.7	2/2014
Si (polycrystalline)	20.8 ± 0.6	0.6626	39.03	80.3	11/2014
Si (amorphous)	10.2 ± 0.3	0.896	16.36	69.8	7/2014
GaAs (crystalline)	25.9 ± 0.8	1.038	29.4	84.7	12/2007
GaAs (thin film)	28.8 ± 0.9	1.122	29.68	86.5	5/2012
GaAs (polycrystalline)	18.4 ± 0.5	0.994	23.2	79.7	11/1995
2J (GaInP/GaAs)	30.3	2.488	14.22	85.6	4/1996
3J (GaInP/GaAs/Ge)	32.0 ± 1.5	2.622	14.4	85.0	1/2003
3J (GaInP/GaAs/InGaAs)	37.9 ± 1.2	3.065	14.27	86.7	2/2013
3J (conc., 240 suns)	40.7 ± 2.4	2.911	3832	87.5	9/2006
4J (conc., 508 suns)	46.0 ± 2.2	4.227	6498	85.1	10/2014
CIGS	20.5 ± 0.6	0.752	35.3	77.2	3/2014
CdTe	21.0 ± 0.4	0.876	30.25	79.4	8/2014
Dye sensitized	11.9 ± 0.4	0.793	19.4	71.4	9/2014
Organic (thin film)	11.0 ± 0.3	0.793	19.4	71.4	9/2014

Most data from [1579], additional data for 3J (conc.) [1606]

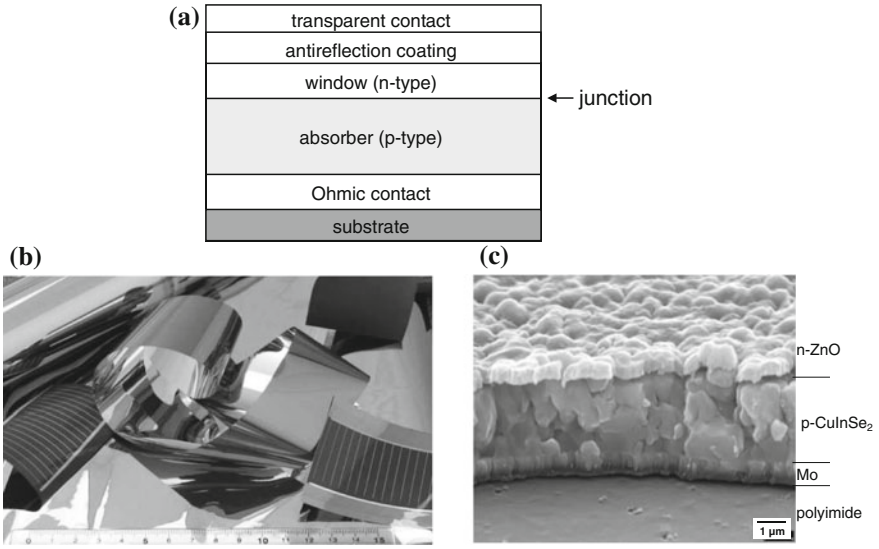

Fig. 22.63 (**a**) Schematic cross section of a polycrystalline thin-film solar cell. (**b**) Rolled sheets of CIS thin film solar cell on flexible Kapton foil. (**c**) SEM cross section of CIS thin-film solar cell. Parts (**b**) and (**c**) reprinted with permission from [1611]

Ga and/or S which increases the band gap, $Cu(In,Ga)(Se,S)_2$ (CIGS), to 1.2–1.6 eV. Using CIGS, an efficiency of over 19 % has been reported in laboratory samples; 12–13 % seeming realistic for production [1607]. Also CdTe is a viable absorber mostly sputtered on glass with over 16 % efficiency demonstrated and 9–10 % realistic in production. Thin-film solar cells can be fabricated on glass substrate or on flexible polymer substrate such as Kapton[11] (Fig. 22.63a, b). Also here, optimization of the grain size is important (Fig. 22.63b). As the front contact, a transparent conductive oxide (TCO), such as ITO ($InSnO_2$) or ZnO:Al, is used. If the front surface is given by the glass substrate, as can be the case for CdTe/glass solar cells, the glass is actually termed 'superstrate'. Thin-film and amorphous silicon solar cells are also termed 'second-generation' photovoltaics. Also organic materials can be used for solar cells [1608], promising low-cost production at acceptable performance, in 2004 2.5 % [1609] and in 2006 about 4 % maximum efficiency [1610].

Third Generation Photovoltaics

'Third-generation' photovoltaics attempt to go beyond the 30 % limit and comprise of multi-junction solar cells, concentration of sunlight, use of hot-carrier excess energy as discussed above and possibly other concepts including photon conversion [1612, 1613], intermediate band absorption [1614–1616], multi-exciton generation [1617] or the use of quantum dots [1618].

[11] Kapton® is a polyimide and a product and registered trademark of DuPont.

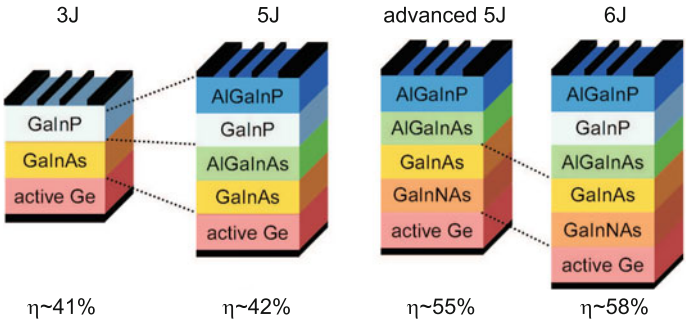

Fig. 22.64 Schematic layer stacking of multi-junction solar cells and expected efficiency. Adapted from [1619]

In multi-junction cells the different absorber layers are stacked on the substrate with increasing band gap and connected via (highly doped) tunneling junctions (Sect. 21.5.9). Under 500-fold AM1.5 illumination a three junction (3J) cell (GaInP/GaInAs/Ge) is expected to exhibit up to 41 % efficiency, for 5 J 42 % or up to 55 % using GaInNAs and with 6 J up to 59 % [1619, 1620] (Fig. 22.64). For a 3 J cell the record efficiency is 40.7 % (240 suns) using the layer structure as shown in Fig. 22.65 [1606]. Details on modeling of III–V multi-junction solar cells can be found in [1621]. In a monolithic cell the absorbers must be tuned such that the same current (Kirchhoff's law) can pass through all layers. Multi-junction solar cells are heteroepitaxial devices and thus expensive; the use of concentration is economically mandatory.

A recent novel route is a tandem cell comprising of a bottom silicon cell and a top perovskite cell. Organo-metal halide peroskites, e.g. of the ammonium trihalogen plumbates type, R-N_3PbI_3, have exhibited quite high conversion efficiencies [1624, 1625] and can be tuned with regard to their absorption range.

22.4.7 Commercial Issues

The cost[12] of producing photovoltaic (PV) modules, in constant dollars, has fallen from as much as $50 per peak watt in 1980 to as little as $3 per peak watt in 2004. A projected cost of 0.2 €/kWh in 2020, a third of the current cost, is realistic and competitive in many applications. In 2002, photovoltaic power of 560 MW was installed worldwide. By the end of 2003, a total photovoltaic power of about 350 MW was installed in Germany. Thus, the PV power is 0.33 % of the total installed electric power of 106 GW in Germany. The current market growth of 30 % (worldwide) is driven by crystalline silicon cells (95 % in 2002).

[12] The following information is taken from [1626, 1627].

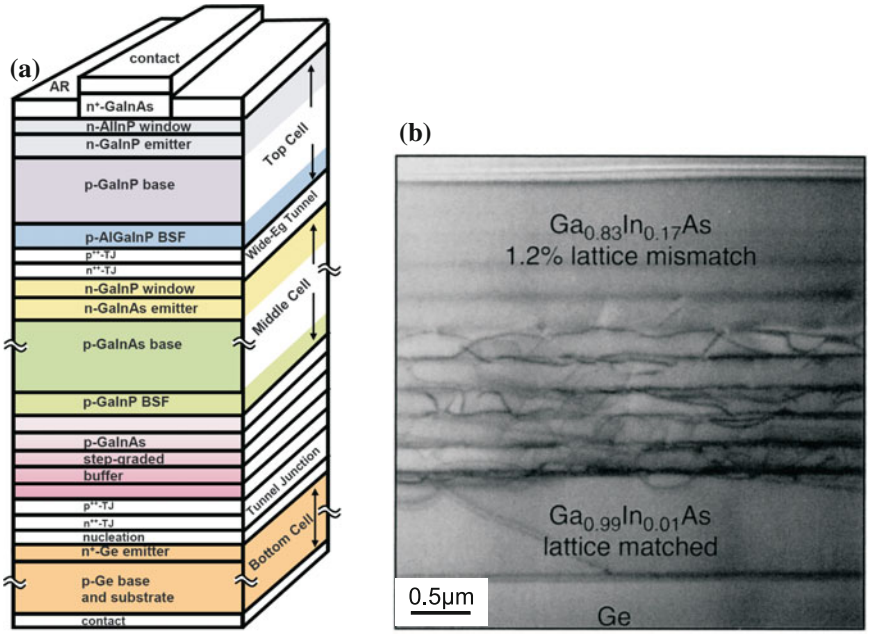

Fig. 22.65 (**a**) Schematic layer stacking of three-junction (3J) solar cell. The step-graded buffer (metamorphic buffer) changes the in-plane lattice constant for the following layers. Adapted from [1622]. (**b**) Cross-section TEM image of metamorphic InGaAs buffer on Ge. Adapted from [1623]

The energy payback period is also dropping rapidly. For example, it takes today's typical crystalline silicon module about 4 years to generate more energy than went into making the module in the first place. The next generation of silicon modules, which will employ a different grade of silicon and use thinner layers of semiconductor material, will have an energy payback of about 2 years. And thin-film modules will soon bring the payback down to one year or less. However, market growth of thin-film modules is currently slow. This means that these modules will produce 'free' and clean energy for the remaining 29 years of their expected life.

PV technology can meet electricity demand on any scale. The solar energy resource in a 100 mile-square area of Nevada could supply the United States with all its electricity (about 800 GW), using modestly efficient (10 %) commercial PV modules. A more realistic scenario involves distributing these same PV systems throughout the 50 states. Currently available sites, such as vacant land, parking lots, and rooftops, could be used. The land requirement to produce 800 GW would average out to be about 17 × 17 miles per state. Alternatively, PV systems built in the 'brownfields', the estimated 5 million acres of abandoned industrial sites in the US, could supply 90 % of America's current electricity. Solar power is expected to contribute 10 % of the US energy need in 2030. For Germany, more than 2 % in 2020 is probable.

In 2001, PV module shipments in the US approached the 400 MW mark, representing a $2.5 to $3 billion market. The US-based industry itself is now approaching $1 billion per year and provides 25,000 jobs. It is expected to grow to the $10–$15 billion level in the next 20 years, providing 300,000 jobs by 2025.