Chapter 2
Photon Management

2.1 Fundamentals of Photon Management in Photodetectors

2.1.1 Introduction

The optical methods of optimization have a goal to maximize the quantum efficiency of a photodetector for a given geometry (thickness and active area) and a given spatial and frequency dependence of the absorption coefficient. They are described as photon management or light management methods.

Various photon management methods considered in this chapter are mostly generally applicable, not only for infrared semiconductor devices. We consider a general case of a photodetector as a device that converts optical energy into another form of energy. Most often this energy is electrical signal, although other forms may be used like thermal [6], motion (like in some microcantilever-based detectors) [75], optical signal at another frequency (up- or down-converted) [76, 77], etc. A notable example of devices to which a majority of the methods described here can be applied are thin-film solar cells [78–81].

There are different methods of light management in a photodetector that can be divided into four groups, Fig. 2.1: optical concentration, use of antireflection structures, optical path increase, and light localization.

Optical concentration methods actually collect incident radiation from a larger area (denoted as "optical" area) and concentrate it (focus) to the smaller active area of photodetector (the "electrical" area). The concentration efficiency can be then defined as the ratio between the optical and the electrical area, minus absorption and scattering losses. This method of light management is typically done by utilizing structures which are not themselves a part of the detector, but can be integrated with it. The simplest case of a concentrator would be an immersion lens, but there are a number of different other structures to serve the same purpose. Roughly, one could

© Springer International Publishing Switzerland 2014

Z. Jakšić, *Micro and Nanophotonics for Semiconductor Infrared Detectors*,
DOI 10.1007/978-3-319-09674-2_2

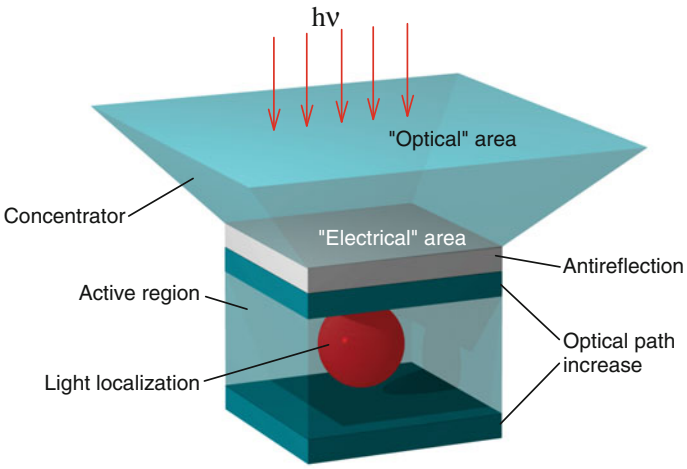


Fig. 2.1 Methods of photon management: use of optical concentrator, antireflection structure, structures for optical path increase (cavity enhancement) and light localization structures

divide them into refractive, reflective, and diffractive elements, although hybrid solutions are also possible.

After arriving to the detector surface, the incident beam should enter the photodetector area with losses as low as possible. Semiconductor materials used for photodetectors have large values of refractive index and thus large values of reflection coefficient at the device surface. This reflection is minimized using antireflection structures. Basically, these structures serve as impedance matching media between the detector environment (most often, but not always free space) and the active region.

Once useful radiation enters the detector-active region, the goal is to keep it there as long as possible and thus obtain maximum interaction between it and the active material. Probably the simplest way to do that is to place a mirror at the back side of the detector, thus doubling the optical path through the active region. However, the methods of optical path increase can be much more sophisticated and may include different cavities with reflective walls, as well as surface structuring. These methods belong to the optical trapping approaches.

There is another method of optical trapping, enabled by the advent of nanophotonics: the subwavelength optical localization utilizing plasmonic nanocomposites. Some metal-dielectric structures ensure the possibility of light localization on a level much smaller than the operating wavelength. To this purpose they utilize the propagating or localized surface plasmons polaritons.

This chapter investigates in some detail all of the mentioned methods of photodetector enhancement.

2.1.2 Fundamental Limits of Photon Management

Basically, all photon management techniques are intended to improve absorption of light in the detector, thus ensuring a higher degree of the conversion of incident optical radiation into a useful signal. Obviously, the efficiency of any conversion is limited by basic physical laws. A question is posed what are the fundamental limits of photodetector enhancement through light management.

A detector system is presented in Fig. 2.2. A background optical flux is incident to the active area of a photodetector with a thickness d. Both in the case of solar cells and night vision photodetectors, the optical flux is blackbody radiation, described by the Planck's law. In a general case, the detector material may incorporate nanostructuring that could localize optical field and create hotspots with high density of states. A perfect mirror is placed at the rear side of the device—i.e., it is assumed that the incident light is unidirectional, while the internal radiation is bidirectional.

The detector surface is corrugated in order to increase the optical path through the detector. The corrugation may be random or ordered, but in both cases its basic purpose is to change the direction of light incident upon the active surface and to make use of total internal reflection to ensure repeated passing of the beams through the active region. Light can escape if the direction of the internal beam falls within the escape cone, for which according to Snell's law $\sin \theta_{cr} = 1/n$ (θ_{cr} is the critical angle of total reflection, n is the refractive index of the active region).

We first consider the case limited by geometrical optics, which has been established by Yablonovitch [82–85]. In the literature it is variably denoted as the conventional limit, the ergodic light trapping limit, the ray-optics limit and the Lambertian limit. It is assumed that the detector-active material can be described by an effective absorption coefficient α isotropic throughout the device and that the detector thickness is much larger than the operating wavelength in free space ($d \gg \lambda/2n$), so that one considers a bulk process. The absorbance within the photodetector for a single pass across the structure (absorption without enhancement) is

$$A(\omega) = 1 - \exp(-\alpha(\omega)d) \approx \alpha(w)d, \tag{2.1}$$

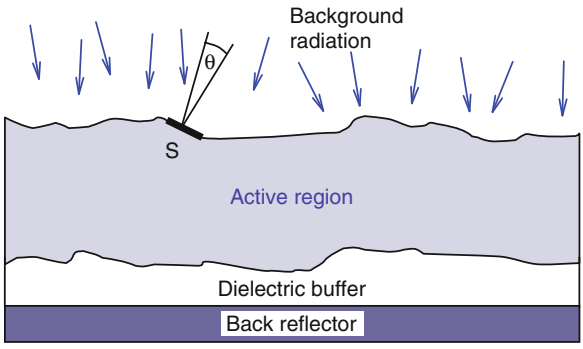


Fig. 2.2 The structure of the active region of a detector with corrugated surface and ideal backside reflector

i.e., the absorbance is equal to the optical thickness of a photodetector, which is defined as the $\alpha(\omega)d$ product.

Since a bulk case is considered, it is further assumed that interference/diffraction effects can be neglected and that the intensity of light within the detector medium is in equilibrium with external blackbody radiation. The density of states within the medium is proportional to n^2. The next assumptions are that the equipartition theorem is valid (the internal occupation of states is equal to the external one, the internal states are ergodic) and that the surface corrugation performs a full randomization of the incident signal over space. This is not always satisfied, but the assumption holds in a vast majority of cases [82]. A sufficient condition for randomization of light by multiple scattering corrugated surfaces is that these surfaces upon averaging behave as Lambertian. The internal distribution of the light within the medium is then isotropic.

According to the statistical ray optics approach [82], the relation between internal and external intensity of light is

$$I_{\text{int}}(\omega, x) = 2n^2(\omega, x) I_{\text{ext}}(\omega), \tag{2.2}$$

The same result is also obtained according to the principle of detailed balancing of the light [86] applied between the light incident to a small surface element of the detector-active area and escaping from that same element through the loss cone and by applying the brightness or radiance theorem (e.g., [87]) stating that the spectral radiance of light cannot be increased by passive optical devices (based on the principle of reversibility).

To determine the enhancement of absorption, one has to consider the loss of light due to various mechanisms. According to Yablonovitch [82, 83], there are three such mechanisms: the escape of light through the light cone, the losses due to imperfect reflection at the surfaces, and the absorption in bulk. The absorbance of a photon is the ratio of the rate at which absorption occurs and the sum of the absorption and the photon loss through the escape cone. For the volume absorption in the limiting case when $\alpha(\omega)d \ll 1$ and taking account the angle of the loss cone θ, this expression is

$$A(\omega) = \frac{\alpha(\omega)}{\alpha(\omega) + \frac{\sin^2\theta}{4n^2 d}}, \tag{2.3}$$

so that the absorption enhancement limit in the bulk case with internal randomization becomes $4n^2$. For $\theta = \pi/2$, this assumes the more often used simple form

$$A(\omega) = \frac{\alpha(\omega)}{\alpha(\omega) + \frac{1}{4n^2 d}}. \tag{2.4}$$

A more complex situation is encountered with the devices using plasmonic localization for absorption enhancement. In this case, many of the above

assumptions introduced for ergodic limit are not valid. The crucial points are that the light distribution is not isotropic anymore (and actually the volumes with a strongly enhanced density of electromagnetic states may be deeply subwavelength) and the thickness of the detector is usually subwavelength.

A number of treatises are dedicated to the situations in which the ray optics limit is exceeded and optical modes are confined at subwavelength scale [88–90]. However, until now no generally valid solution has been given for the extension of the ray optics limit [91].

2.2 Nonimaging and Imaging Optical Concentrators

One of the obvious methods to increase the optical input to the detector is to collect optical signal from a larger area and to concentrate it to the smaller electrically active area. This can be done by some kind of optical concentrator, i.e., focusing optics integrated with the detector unit.

The usual way to concentrate light from a larger area to a smaller one is to utilize optical elements that can create an image of an object. However, besides such structures, there are also concentrators that belong to the wide class of nonimaging optics [92–97]. This is actually the branch of geometrical optics in which the main goal is not to form the best possible object image, but rather to gather as large portion of the incident radiation as possible and to concentrate it to the active area of the detector, regardless of the price to be paid from the point of image fidelity. A single photodetector element in infrared technique is not an imaging device and only the total intensity of the radiation impinging on its active surface is important, not its spatial distribution. A similar situation is met in the case of solar cells. Thus for such photodetectors one can apply both imaging and nonimaging concentrators.

A schematic presentation of an optical concentrator is given in Fig. 2.3. The shape of the system can be arbitrary, as well as its mechanism of focusing. The important factor of the system is its concentration ratio, defined as a ratio between the input (collector) area and the exit area. In an ideal concentration system, this ratio should be equal to the ratio of incident optical flux to the exit flux. In a real system, possible absorption losses (for instance in metal parts) will have to be deducted. Thus the concentration ratio of a light concentrator can be defined either as the geometric concentration ratio C, the ratio of the entry aperture area ($A_{\text{collector}}$) to the exit aperture area (A_{exit}), or as the irradiance gain C_i, the ratio of the irradiance on the collector (Ir_c) to that on the entry aperture (Ir_m). The two concentration terms are related by the fraction of total incident power entering the module that reaches the concentrator exit aperture.

Different strategies can be used to concentrate the incident optical power. There are three main groups of optical concentrators. One of them is based on refractive optics (conventional lenses). The second one is reflective concentrators (mirrors), while the third group is diffractive optical elements. Obviously, a system may simultaneously incorporate two or even all three of the mentioned structures.

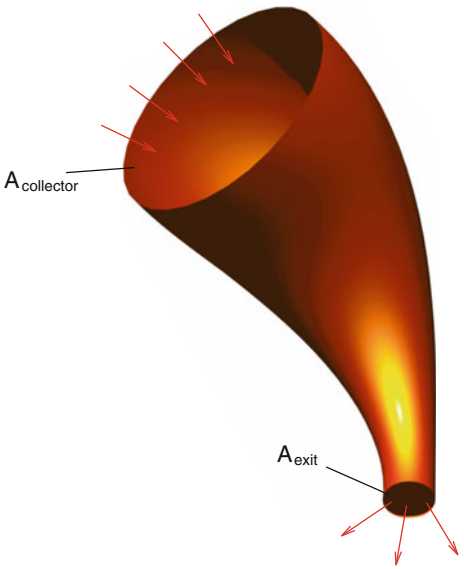


Fig. 2.3 Schematic presentation of an arbitrary light concentration system. The area at the entry of the system is $A_{\text{collector}}$, and at the output A_{exit}

Monolithic or hybrid integration may be used to integrate concentrators with the photodetector (optical immersion of detectors and focusing optics). The concentrators may be partially integrated with the detector, i.e., positioned immediately to the detector surface, but without physical integration. In that case, there is a gap filled with air or some transparent buffer material with a real part of its refractive index comparable to that of the detector-active area or smaller than it and at a distance of an order of the operating wavelength (pseudoimmersion). Finally, concentrators may be integrated with the detector housing, i.e., encapsulated together with the active element, but separated from the detector surface by a distance much larger than the operating wavelength (discrete lenses).

From the point of view of coupling efficiency (minimization of reflection losses), the best solution is monolithic or hybrid integration; however, in practical situations one can encounter all of the mentioned approaches. Two main types of focusing lenses may be used—either refractive lenses fabricated in material with high real part of refractive index and low absorption coefficient at IR wavelengths, or diffractive lenses. Any of them may be either discrete or arrayed. Reflective optical concentrators may be used, also reflective holographic optical elements or any of their combinations.

Besides the maximum increase of incident flux, the main requirements posed to the concentrators in infrared photodetection are the simplicity of fabrication technology, minimal losses, and good reproductivity during fabrication. The dimension tolerances for optics are of secondary importance and allowed aberrations are higher than those posed, e.g., to the microlenses used with optical fibers. Even in imaging optics for the infrared range the tolerances (especially those for MWIR and LWIR) are much less stringent than those for visible light, since larger wavelengths are used. The nonimaging nature of the concentrators further relaxes the tolerances.

2.3 Refractive Concentrators

2.3.1 Immersion Microlenses

Probably the most obvious way to collect illumination from a larger area and to concentrate it to the active area of a photodetector is to use a microlens. Such a lens should have dimensions sufficiently small to be mounted on a detector. Since detector diameters are typically below 1 mm and very often below 0.1 mm, this means that the "micro" in the name of the microlens only means "small," since the dimensions of the lens itself will be of the order of millimeters.

In order to avoid reflective losses between the detector and the lens, typically some kind of refractive index-matching immersion will be used. The lenses for the photodetector improvement can be thus called immersion lenses. Obviously, the material of the lens has to be transparent in the wavelength range of interest and is typically used with some kind of antireflective coating.

There are a number of different geometries convenient for immersion microlenses in photodetection. Probably, the most well known and widely used form is hemisphere. The use of microsystem technologies allows the fabrication of various discrete or arrayed microlenses, with spherical surfaces (calottes, hemispheres and truncated spheres, full spheres), aspheric (ellipsoids, paraboloids, cylinders, cones), toroid, as well as various nonmonotonic surfaces consisting of two or more monotonous segments. Most of the microlenses convenient to increase the incident flux to the detector are plano-convex ones.

Figure 2.4 shows some types of refractive microlenses that can be fabricated utilizing the standard microfabrication procedures in materials convenient for the MWIR and LWIR ranges. Most of them are loosely based on the solutions for microlenses used in fiber optics to improve coupling between laser sources and fibers [98]. These immersion lenses were thus intended for the operation with coherent and monochromatic radiation, while most of the microlenses in the field of IR detector technology are intended for incoherent, mono- or polychromatic Lambertian sources and, of course, they operate in different atmospheric windows.

Among spherical microlenses utilized in optoelectronics literature quotes spherical calottes [99] (Fig. 2.4a), hemispheres [100, 101] (Fig. 2.4b) and truncated spheres [102] (Fig. 2.4e). These lenses may be defined as the refractive surfaces that can be represented as a part of a sphere with a radius r and with its center in the focus of the lens. The immersion lenses are usually formed by bringing into contact the planar surface of such a lens and the incident surface of the detector. Besides the above-mentioned lenses, ball-shaped microlenses are also used [103].

The following microlens shapes were also fabricated by microsystem technologies: elliptic calottes [104], hemiellipsoids [105], ellipsoids [103], hemicylinders [101, 105] (Fig. 2.4g), and cylinders [106] (Fig. 2.4h). These types of microlenses are especially convenient if the beam shape is elongated. Also among aspherics are conic microlenses with straight or curved generatrix [107], (Fig. 2.4i). In some optical systems, paraboloid or hyperboloid surfaces are sometimes used, usually

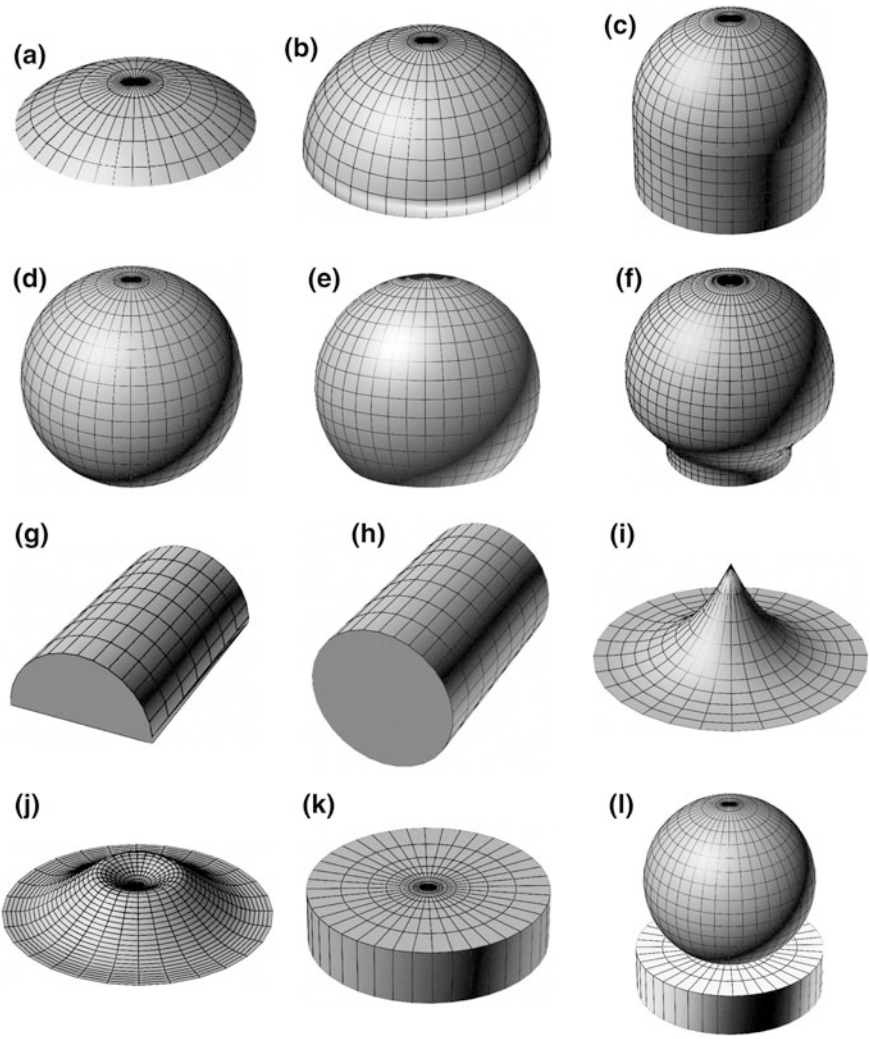


Fig. 2.4 Some spherical and aspheric discrete microlenses for optical concentrators in MWIR and LWIR range that may be fabricated by microsystem technologies. **a** calotte; **b** hemisphere; **c** hyperhemisphere; **d** ball lens; **e** truncated sphere; **f** "bulb"; **g** hemi-cylinder; **h** cylinder; **i** curvilinear cone; **j** concave concentrator; **k** gradient-index lens (GRIN); **l** complex two-element lens (sphere/GRIN). The grid corresponds to the homogeneity of refractive index, i.e., describes its gradient

fabricated by polishing. We should mention here optical concentrators with concave middle area [98] (Fig. 2.4j).

A very important type of optical concentrators, both in optical telecommunications and for MWIR and LWIR photodetectors are hyperhemispheric microlenses (combinations of a hemisphere and a cylinder with joined bases, i.e., the "rod lens")

[103] (Fig. 2.4c). Literature also mentions microlenses shaped as rounded pyramids, fabricated by micromachining [108].

An important class of microlenses are planar lenses with gradient refractive index (GRIN-microlenses, Gradient Index), Selfoc-lenses or Luneberg's lenses [103] (Fig. 2.4k; the grid shows the nonuniformity of refractive index, high in the middle, and decreasing toward the edges.) Within this context one could also mention the "electronic lenses" where the refractive index profile depends on impurity concentration and thus allows an external electrical control of the focus [109].

Confocal systems with two microlenses are often met in telecommunications [110], most often a combination ball lens-GRIN (Fig. 2.4l). Typical procedure is to focus one of the lenses and to align the other subsequently, so that the combination itself improves the allowed position tolerances several times. Also, for some applications (irregular beam shape) it is convenient to have more than one lens to allow a larger degree of freedom when adjusting beam collimation. Dependent on application, systems with three or more microlenses are also used [110].

2.3.2 MEMS Fabrication of Infrared Microlenses

As mentioned before, most of the structures for infrared detection can be fabricated by microsystem technologies (MST) and the microlenses for optical concentration are not an exception, regardless of the type of the optical element in question (refractive, reflective or DOE/HOE).

A point of special importance is the choice of proper material for MWIR and LWIR optical concentrators. A number of materials are customarily used to fabricate infrared optics. Besides these, optical concentrators may be fabricated directly in the photodetector material and indeed be monolithically integrated with it.

The usual requirements posed for infrared optics and regarding mechanical and water resistance, robustness and durability, insensitivity to temperature changes are significantly less stringent here, owing to the fact that microlenses are integrated within the detector housing. Some water-soluble materials are used for IR concentrator microlenses, because they are readily polished with water and cut by a wet string saw [1]. Among the materials used for infrared is germanium, since it ensures relatively simple machining. The most important requirements remain a maximized transmission in a given spectral range and the possibility to machine the microlens by some of the MST procedures.

A number of procedures are used to fabricate microlenses for telecommunications and some of them can be modified and adapted to fabricate optical concentrators for the MWIR and LWIR range. Some of these procedures already became standard.

Chronologically, the oldest approach to microlens fabrication is thermal melting. The action of heat is used to soften or melt convenient material (e.g., germanium in LWIR range) and form it into a desired shape. An alternative is to use material that

polymerizes into a desired shape, or a convenient solver may be used that evaporates and leaves a solid object. Melting is the technology utilized for the least demanding applications, but at the same time it is probably the most often used one.

One of the oldest methods used with molding, but at the same time the least accurate, is cast molding (e.g., injection molding, casting, etc.) [111]. This method is used both for refractive and diffractive lenses. First, a high-accuracy "master" is fabricated. It is subsequently used to produce large series of microlenses. The method is especially convenient for work with plastic materials transparent in IR, like TPX (polymethylpentane, polymerized dimer of propylene.)

The procedures to fabricate masters are handled later in this section. Practically, all higher accuracy methods used to fabricate microlenses can be used to produce masters. In all of the quoted methods, it is important to choose polymers transparent in the IR range.

The methods for inexpensive replication and mass production of microlenses include

- Conventional injection molding (the most often used technique for low-end diffractive elements, for instance for the best-known diffractive optical elements, CD, DVD, and BluRay disks) and epoxy molding (the latter being more expensive because of longer curing cycles).
- Embossing: heat and pressure are applied to transfer surface relief from a mold into a plastic substrate, typically a thermoplastic foil.
- UV replication: probably the most convenient method for DOE mass production and also the most usual for higher quality and low-cost diffractive elements. The substrate is usually glass, and the mold is fabricated by electroforming. UV curable resin is potted between the mold and the glass and replicas are molded by photopolymerization. Alternatively, the mold may be used for e-beam evaporation of various inorganic materials to the substrate.

One of the molding methods often used to fabricate refractive lenses is to deposit a cube of lens material on the substrate using, e.g., some of the planar technology procedures, and then to treat it thermally until it melts. Surface tension turns it then into a hemisphere, calotte, or a flattened hemisphere [112]. An even better method is to use a similar procedure to deposit a "base" for the microlens with a desired diameter, and then to deposit lens material on it [113]. After melting, the lens material flows exactly to the edge, where surface tension prevents it to go farther (melt stop.) In this way, the lens surface is formed with an extreme accuracy.

One of the often used methods is to dip the lens support into the melt and remove it. In the course of this, a "drop" is formed at the end of the support. After cooling and hardening it is mechanically cut away. Finally, one of the used approaches is to drop the melt into cooling liquid. The shape and the size of the cooled drop define the lens.

The next important technology is photolithography. It is used both to fabricate lenses ready for use and masters for further replication, for any of refractive, reflective, or diffractive concentrators.

The most often used lithography procedures are electron-beam writing and optical lithography. Electron-beam writing is actually scanning of the photoresist-covered (spin-coated) surface by a focused electron beam, and thus exposing the resist. This technique enables the fabrication of continuous and curved profiles. It poses problems with positioning, requires computer control, and very high mechanical and vibration stability and must be performed in vacuum. Optical lithography uses light beam for writing and may be performed in the air. Some of its variants are direct laser writing in photoresist (laser beam lithography), holographic lithography (the use of interference pattern to write a DOE/HOE without a mask), and conventional photolithography (where mask can be also a hologram). The latter can use contact-printing (direct contact between the mask and the resist surface, furnishes the best results if performed in vacuum), proximity printing, and projection printing (conventional for volume production, but less accurate).

The next step in all the versions of lithography is removal of the photoresist, either the exposed part (positive photoresist) or unexposed (negative p.) and the final step is etching.

Two dry etching methods are often used: Reactive-Ion Etching (RIE) and ion milling. RIE is actually a simultaneous etching by chemical reaction and by particle bombing. It is anisotropic (vertical component is much more pronounced than the lateral). Ion milling uses a stream of inert gas to strike the surface and remove material by physical sputtering.

Wet etching includes chemical micromachining (etching in a chemically aggressive solution) and, depending on the properties on the substrate and the etching solution, it can be isotropic or anisotropic. Among the available techniques is laser-assisted chemical etching (convenient for very deep vertical holes). Blazing can be done by the melted resin technique and the mass transport method. Masters for simpler kinds of diffractive concentrators can be formed by holography or by mechanical ruling using, for example, a diamond cutter.

The fabrication of hemispheres and hemicylinders was reported in Cohen (1974), Bear (1980), Lee (1985). The microlenses here were fabricated in negative photoresist by point UV illumination to obtain a hemisphere, or by bar-like illumination for hemiellipsoides and hemicylinders. The excess resist is removed by solvent. A negative property of this method is that resist lenses are mechanically sensitive, stronger sources (e.g., industrial lasers) may easily damage them and they are applicable only for certain wavelength ranges.

This method is convenient for fabrication of GRIN lenses, where the required refractive index gradient is obtained by point diffusion [114]. Refractive microlenses with arbitrary geometry may be fabricated by photolithography so that first masks and etching are used to form a mesa or multiple-step structure directly in the substrate, and then additional etching is used (bromine-methanol for HgCdTe or InSb) to round it into a spherical or aspheric surface [115].

Another approach to form a lens directly in the substrate material is to form first a lens in photoresist deposited onto the substrate using some of the above procedures. After that reactive-ion etching (fast, but relatively small choice of materials) or ion milling (slower, more materials) is used to transfer lens into the substrate

[116]. The method is convenient for fabrication of optical elements in the wavelength ranges where resist is not transparent and generally offers a larger degree of freedom in choice of lens material. Finally, one of the options is to immerse the whole system into an electrolyte and to perform photochemical etching through a mask [117].

Most of the above photolithographic procedures can be utilized to fabricate diffractive lenses [118].

We mentioned holographic lithography as one of the types of photolithography used for IR concentrator fabrication [119]. Its main advantage is that the key steps of forming and exposing a mask are omitted. Resist (or some other medium, e.g., bichromated gelatin [120]) can be directly exposed by an interferometric pattern obtained by laser irradiation [107, 121]. A stable optical bench is necessary for that (or, alternatively, a pulse laser with high enough output power). The second alternative is to use computer-generated holography. In that case, a mask is necessary, but very complex systems can be fabricated [122]. Some authors mention the use of bichromated gelatin for the infrared optical elements [3].

Chemical micromachining is one of the methods of choice for fabrication of microlenses [123]. Silicon and germanium were used as lens materials. For spherical and aspheric geometries solutions for anisotropic etching are used, while nonmonotonous surfaces can be fabricated by combining masks and anisotropic etching. This method was used to fabricate various types of optical concentrators, among them cones and inflected surfaces [107].

Single microlenses with relatively larger dimensions can be efficiently produced mechanically, e.g., using abrasion or polishing. To fabricate both refractive and diffractive microlenses diamond turning is used [124]. One of the methods in this group is microgrinding [125].

One of the modifications of this approach is the use of a combination of mechanical and chemomechanical polishing, for instance, by simultaneous action of abrasive and bromine-methanol for lenses in Ge, CdTe, and InP [126].

2.4 Reflective Concentrators

Reflective nonimaging concentrators are not often used with infrared photodetectors, but are very well known and intensively applied in solar cell energy production [93, 123, 127–129]. Typically their structure is a tapering hollow with reflective walls, its wider aperture serving as the collecting surface, while multiple reflections from the walls collect beams toward the exit opening. The cavity may be empty (i.e., filled with air) or filled with some dielectric. The geometry may be two-dimensional (various kinds of trough) or three-dimensional, the latter being either a body of revolution or some other kind of tapering structure. The prototypical 3D structure is the so-called Winston cone, while various other tapering profiles may be used as well. They start from the simple ones like parabolic, hyperbolic, circular, conical shapes, extend to their combination, but may have different complex forms

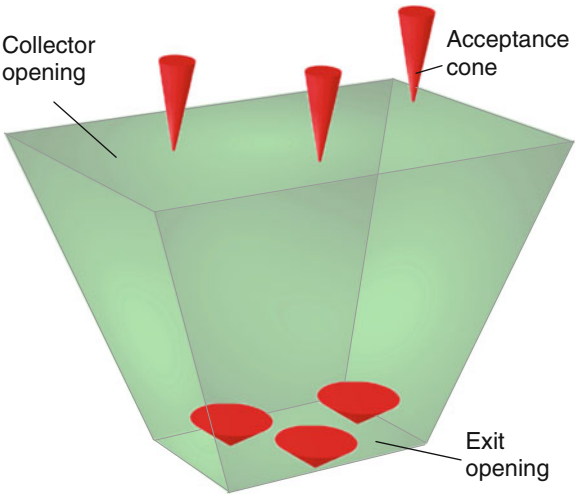


Fig. 2.5 Angular distribution of light showing larger acceptance angle cones at the exit opening and smaller at the collector opening

as determined by a particular application and the optimization procedure implemented to obtain maximum performance. The surface also may be multipart/segmented, with each segment having a different profile. Other solutions include Fresnel or echelon reflectors.

Figure 2.5 shows the connection between the spatial angle from which incident flux may arrive in order to be concentrated (acceptance cone) and the spatial angle under which radiation leaves the concentrator. Any radiation outside the acceptance cone will be rejected, i.e., the system will not bring it to the exit aperture. It can be seen that the acceptance cones at the input cover much smaller spatial angle than the light cones at the exit.

Regarding materials, the structures of microreflectors and their arrays are usually made of metal with good reflection coefficient. There are a wide choice of metals and alloys convenient for the fabrication of reflector-type concentrators.

The shape of a 3D reflector top and basis may be rounded, but also polygonal, the latter being more convenient if a high filling factor is desired when producing 2D arrays of reflectors. Figure 2.6 shows some possible shapes and the manners of their stacking into larger matrices. The rightmost example shows that the possible shapes may include combinations of polygonal entrance with round exit.

2.4.1 CPC (Winston Collector)

Compound parabolic concentrator (CPC, also called Winston collector or Winston cone) [95–97] is a reflective nonimaging light concentrator intended to collect incident light from a spatial angle larger than that of imaging concentrators and to

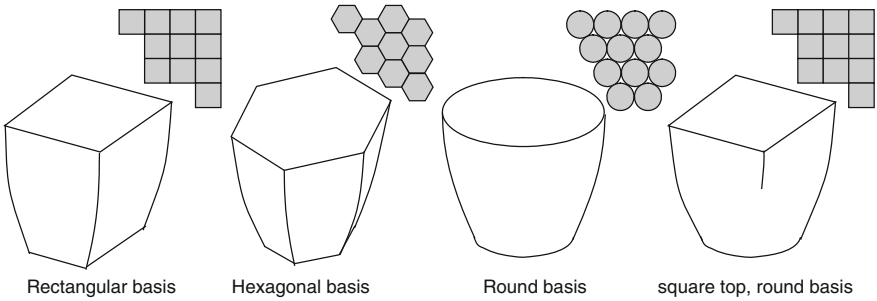


Fig. 2.6 Different entrance and exit geometries shapes for reflective light concentrators. *Gray matrices* in the right upper angle of each shape show the manner of concentrator packaging into arrays

direct it toward its exit opening. It has been the first discovered nonimaging reflective concentrator. Its 2D depiction is schematically shown in Fig. 2.7.

Two parabolas p_1, p_2 can be seen in the figure, with their respective focal points in F_1, F_2, and their axes being a_1, a_2. The parabolas are cut at the line of their focal points and then mutually rotated so that the focal point of one of them coincides with the curve of the other parabola and the very edge of the exit opening. A 3D version of this structure is obtained as a body of revolution whose sides are parabolic segments satisfying the same condition. Thus a Winston cone is an off-axis parabola of revolution.

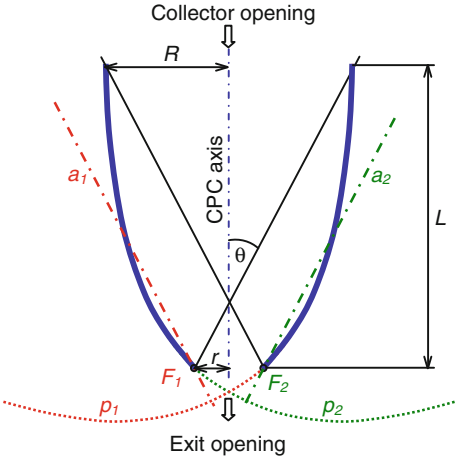


Fig. 2.7 Schematic presentation of CPC collector. R is the entry (collector opening) radius, r is the exit opening radius, a_1 is the axis of the parabola p_1, a_2 is the axis of the parabola p_2, F_1 and F_2 are focal points of the parabolas 1 and 2, respectively. Parabolas are rotated so that the focal point of one of the parabola coincides with the curve of the other parabola at the edge of the exit opening

A CPC concentrator increases the efficiency of concentration by collecting off-axis rays that would else be lost and directing them toward the exit opening. These rays reflect several times within the CPC and after such multiple bounces reach the exit. The rays arriving at an angle larger than the acceptance angle do enter the concentrator, but after bouncing several times off its walls they are lost, i.e., they do not reach the exit opening.

The total length of the CPC concentrator is

$$L = \frac{r(1 + \sin\theta)\cos\theta}{\sin^2\theta} \tag{2.5}$$

Obviously, a CPC can be truncated and in real situations often is. This will result in a lower concentration ratio, but the overall length might be more acceptable for implementations.

There is a relation between the radii of the entry and exit openings

$$R = \frac{r}{\sin\theta} \tag{2.6}$$

The equation defining the surface of a CPC in a radial system in which the CPC axis is in z-direction while the radial distance is described as the coordinate ρ is

$$\begin{aligned} &2r\rho\,(1 + \sin\theta)^2 + (\rho\cos\theta + z\sin\theta)^2 \\ &\quad = 2rz\cos\theta(2 + \sin\,\theta)^2 + r^2(1 + \sin\theta)(3 + \sin\theta) \end{aligned} \tag{2.7}$$

A special case of the rotational Winston collector is a CPC trough. Both 2D and 3D CPC structures are ideal concentrators. 2D CPC is convenient if the target travels along a line since then there is no need for its tracking.

One possible improvement of the CPC is to fill it with dielectric having a refractive index n. This shortens the structure and increases the acceptance angle. At the same time it uses total internal reflection which ensures a higher reflection than from metal surrounded by air. The maximum concentration ratio in this case is larger by a factor of n^2. Another useful modification is to utilize a two-stage CPC where the exit opening of the first stage CPC coincides with the entry opening of the second stage.

2.4.2 Parabolic Reflector

Geometrically, a 3D parabolic reflector represents a truncated paraboloid, as shown in Fig. 2.8. Parabolic reflectors are actually imaging optical elements, used in various optical systems. For instance, a parabolic reflector is often combined with a diverging mirror that directs light beams through the exit aperture. This is the well-known Cassegrain configuration, the standard design for reflector telescopes.

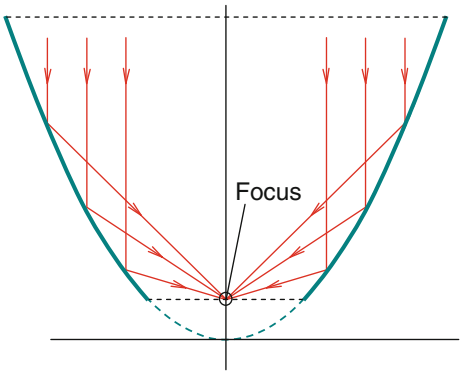


Fig. 2.8 Schematic presentation of a parabolic reflector. The parabola is cut off in the focal point



Fig. 2.9 Parabolic trough

The microfabricated version of parabolic reflector is used in solar energy concentration [130].

Contrary to the CPC, a parabolic concentrator is a high-concentrating device, but is more sensitive to scattering of incident radiation than the CPC. A problem with parabolic concentrators is that their total transmission is much smaller than that of CPC and reaches about 0.60.

A 2D implementation of the full paraboloid geometry is the parabolic trough [131], Fig. 2.9. Although a full three-dimensional paraboloid is more energy efficient than a trough, 2D implementation is less complex and can be useful in situations where there is a linear array of photodetectors.

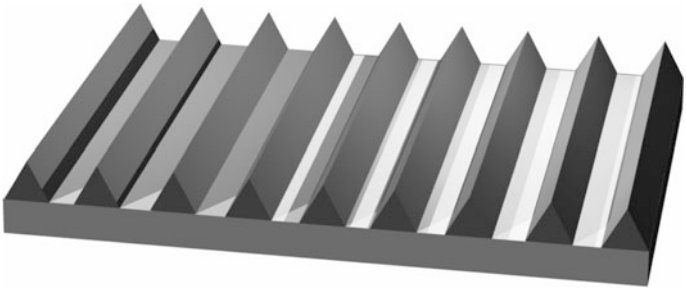


Fig. 2.10 A V-trough for reflective nonimaging concentrators

2.4.3 Conic Concentrator

The conic concentrator is actually a truncated cone with reflective sides. It is not an ideal concentrator, and its transmission reaches about 0.8. Its 2D version (with a triangular profile) is called a V-trough, Fig. 2.10.

2.4.4 Hyperbolic (Trumpet) Geometry

A geometry for reflective concentrators is the trumpet concentrator [128], Fig. 2.11. This is actually a hyperboloid of revolution. Besides the CPC it is denoted as one of the "maximum collection efficiency" nonimaging concentrators. It is characterized by multitudinous reflections near the exit aperture.

Trumpet concentrators are ideal concentrators both in 2D and 3D case. They also offer the advantage of concentrating not only internal but also external rays (those latter are concentrated to an annular shape around the exit aperture). Trumpet concentrators are characterized by a small material expenditure. External rays are not usable for detector matrices.

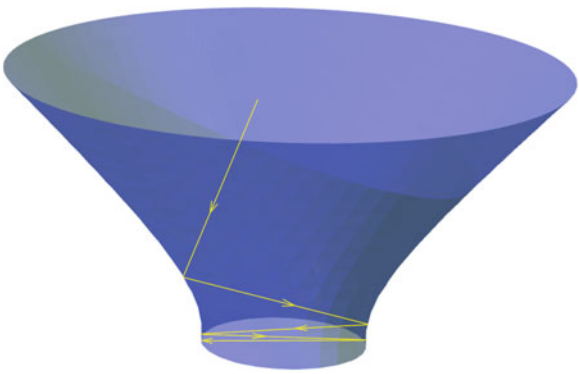


Fig. 2.11 'Trumpet' concentrator

2.4.5 Other Reflector Geometries

There is a plethora of reflector geometries for nonimaging concentrator optics. Among them are elliptic and compound elliptic concentrators (CEC) as well as high-flux two-stage optical designs based on tailored edge-ray concentrators (ordinary and compound). They also include different lens-mirror combinations, utilizing reflection (X), refraction (R), and total internal reflection (I), for instance XR, RX, XX, XRI, etc. [128]. Complex forms for ultra-high concentrations of input radiation are obtained by using different algorithms, for instance global optimization procedures.

2.4.6 Fresnel Reflector

The idea of the Fresnel reflector is to divide a curved reflector into segments and to move these segments to a single joint plane [132, 133]. Thus one is able to effectively reduce a 3D object to a thin (quasi-2D) object while keeping the identical optical path to that of the 3D object. For instance, if one starts a 3D body of revolution like, for instance, a parabolic reflective concentrator or a Winston cone, this body can be divided into a number of annular segments which are subsequently placed concentrically on a flat surface. This process is illustrated in Fig. 2.12a, where a curved body is cut into segments and placed on a surface. An incident beam reflects under the same angles as it would from the full 3D body.

The surface of the original body of revolution can be turned into a Fresnel reflector by being cut into annular segments with the reflective area identical with the original, i.e., with curvilinear surfaces, Fig. 2.12b. It can be also approximated with blazed geometry, Fig. 2.12c, where the cross-section of the annular segments is fabricated in triangular form, with the internal side of the triangle following the original form as accurately as possible.

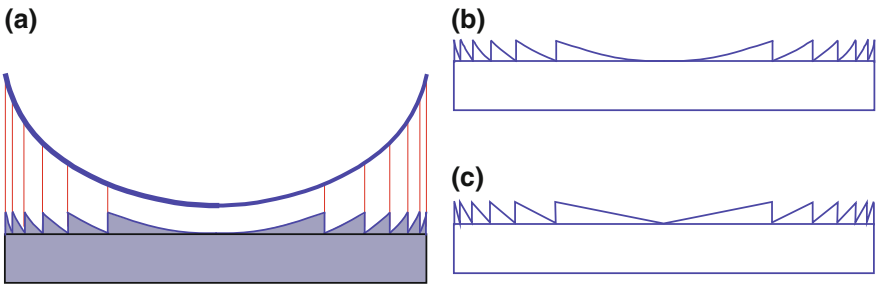


Fig. 2.12 Fresnel reflector. **a** Method of cutting reflective surface. **b** Reflective surface represented by in-plane curvilinear annular segments. **c** Reflective surface approximated by flat (blazed) segments with triangular profile

Fresnel reflectors are much more compact and convenient than full 3D bodies of revolution or 2D troughs. They can be both of imaging and nonimaging type, depending on the body of revolution (or 2D trough) they approximate.

2.5 Diffractive Optical Concentrators

2.5.1 Definitions

We start this section with some definitions. First, we define a diffractive optical element (DOE) as a passive optical element, its properties being either stationary or switchable, with operation based on diffraction or interference of light beams passing through it or being reflected from it. This is a very broad definition and includes a large number of specific subclasses.

Any DOE is basically an optical grating structure, either surface- or volume-type. The grating lattice is defined by some kind of modulation of the complex refractive index of the grating material.

A binary *DOE* is an optical grating element which is defined by a finite number of discrete levels in the values of its optical constants. The smallest value of this number is 2. Due to technological issues, this number is in general case equal to 2^N, where N is the number of technology steps required for its fabrication, typically photolithography steps. Some authors actually define binary optics elements as those passive optical elements which are fabricated using VLSI technology.

A grayscale *DOE* is a generalization of a binary element and represents those cases when N is large enough to be regarded as infinite.

If a DOE is fabricated using holography, it is called a holographic optical element (HOE). The terms HOE and DOE are sometimes used interchangeably.

Any diffractive optical element may be computer-generated (CGH, computer-generated hologram) or designed directly using a physical interference pattern, either from some object or generated in an interferometer.

The family of diffractive optical elements is schematically shown in Fig. 2.13.

In dependence on which part of the complex refractive index was used to modulate the diffractive lattice, a DOE may be of amplitude type (the imaginary or absorptive part of the refractive index was used for modulation, i.e., the grating is

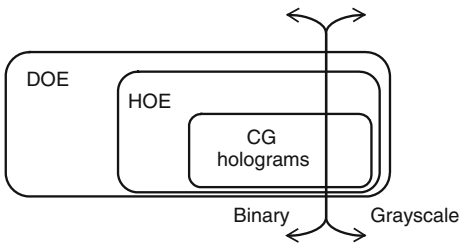


Fig. 2.13 The family of diffractive optical elements

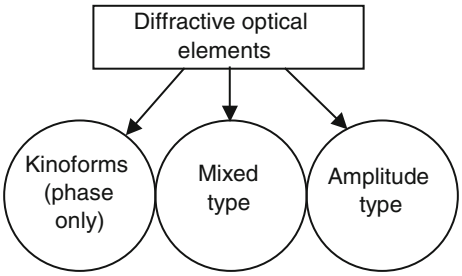


Fig. 2.14 Classification of DOE in dependence on the mechanism of refractive index modulation

built from transmissive and opaque parts, and thus only the amplitude of the incident beam is changed), phase type (only the real (refractive) part of the refractive index is modulated, thus only the beam phase is changed) and mixed (the material used changes both refraction and absorption of the incident beam). The diffractive optical elements influencing only the phase of the incident beam are kinoforms.

Figure 2.14 shows the division of DOE to classes in dependence on refractive index modulation.

Because DOEs function by breaking the incident light wavefront into wavelets and restructuring it, they offer a larger degree of freedom in designing their functions. Besides replacing most of the conventional optical elements, DOEs have enabled the development of some products that could not be produced with previous optical technologies. In addition to that, their area can be larger, even much larger than that of the conventional optical elements, they are much thinner and their cost is lower. They also offer an enormously wider choice of materials for their fabrication, providing the user with a greater flexibility in selecting the material for a particular application.

A diffractive optical element can be either transmissive or reflecting. It may be polarization sensitive or it can be insensitive to it.

One of the most important features of DOEs is their angular and wavelength sensitivity or dispersivity. This means that their focal length and aberrations vary with wavelength and incident angle.

The properties of some DOEs are constant within a narrow wavelength band, e.g., 10–20 nm, which may be a serious shortcoming for some applications. However, there are solutions for multicolor applications, among them being the use of DOEs that widen the operating frequency band by integrating in themselves several separate narrow-band elements, each covering a part of the desired spectrum. Also, some DOEs have wider wavelength ranges.

Since almost any material may be used for DOEs, the available media include electrochromic, liquid crystal and other recording materials. Thus DOEs can be easily switched between two or more desired functions, i.e., recorded and erased and generally controlled in time by the user.

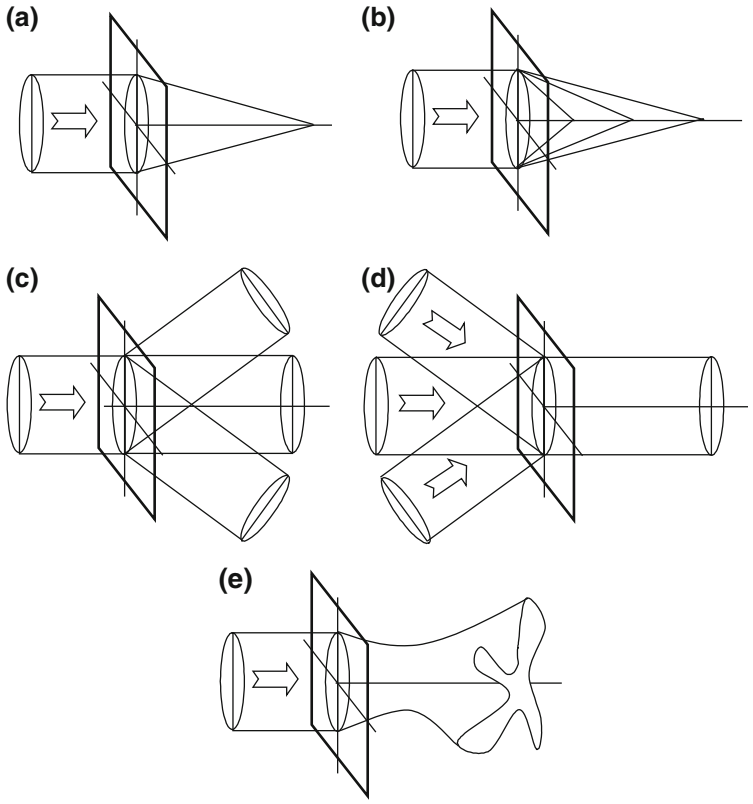


Fig. 2.15 DOE functions of interest in photodetection. **a** Simple focusing; **b** focusing with variable focal length; **c** beam splitting; **d** beam combining; **e** beam shaping

Diffractive elements are very lightweight, as they are typically formed in thin films of a few micrometer thickness only (this is valid even for the so-called "volume structures" where a thickness of only a few wavelengths usually suffices.) They can be completely flat even if they perform functions of conventional elements with curved surfaces. On the other hand, a DOE can be fabricated on any arbitrary shape of the substrate. Their fabrication is cheap in comparison to conventional optics (typically a master element is fabricated and its replicas are created by some mass production methods).

Figure 2.15 shows some functions achievable by the use of diffractive optical elements that are of interest for photodetection. The principal function of interest is focusing or light concentration. This may be done in a way similar to conventional lenses (Fig. 2.15a), but one of the possibilities offered by DOEs is also multiple focusing, i.e., simultaneous focusing of different beams to several foci at once with a single DOE, in dependence on the beam properties or on external controls. These foci may be along the same axis (only focal length is changed) (Fig. 2.15b) or

separated in space (both focal length and focus position are changed)—beam-splitting with simultaneous focusing. Naturally, DOEs may be utilized for beam-splitting without focusing (Fig. 2.15c). The selectivity of a DOE may be designed to depend on any desired property of the incident waves (e.g., their wavelength, polarization, spatial angle of incidence, etc.). DOEs in photodetection may be utilized as beam combiners (Fig. 2.15e).

Practically, it may be said that several different diffractive optical elements can coexist on the same substrate without interfering with one another. They can perform the same function, either selectively, as described above, or implement completely different functions. For example, a single DOE can be simultaneously used as a lens, beam splitter and spectral filter.

A DOE can be used for beam shaping, i.e., it can transform an incident wave from a given initial form to a desired final form (Fig. 2.15e), for example from planar to Gaussian distribution or vice versa. It can serve as a generator of arbitrary aspheric waves, or a most general wavefront transformer. This includes both spatial redistribution and phase correction and may incorporate generation of an interferogram of existing or nonexistent object, sampling of wavefronts, etc.

Finally, DOEs can serve as various optical spectral filters, including low-, high-, or bandpass filters, spectral transformers, notch-filters, etc. All of these features can be used to enhance photodetectors.

2.5.2 Fresnel Lenses

Probably the most important and definitely the simplest DOE for optical concentration in photodetection is the Fresnel Zone Plate (FZP) [134]. An FZP lens represents a set of concentric rings with their width decreasing toward the edge. Light is diffracted from these rings and the diffraction maximum is in the focal point of the lens. The alternating transparent and opaque rings are calculated to result in adding the electromagnetic waves with the same sign, and blocks those with the opposite sign. The resulting function is similar to that of a conventional convex lens and the structure works as a light concentrator. The intensity of the electric field in the focal plane is linearly proportional to the number of the zones, i.e., the number of transparent rings. The maximum diffraction efficiency (defined as the ratio between the incident and the transmitted radiation power) of the FZP does not exceed 10 %, because approximately one half of the radiation is stopped at the opaque rings, while the phase distribution of the other half is sinusoidal.

A FZP is a 2D form and a large number of annular lenslets present in this form are arrayed on a single optical surface, thus replacing a thick refractive optical lens by an essentially planar structure. Figure 2.16 represents a typical FZP. Instead of having sharp edges, the transitions between the opaque and transparent zones are more gradual, which is a situation often met in reality.

The optimum diameter of a Fresnel zone for a case of a planar incident wave is calculated as [135]

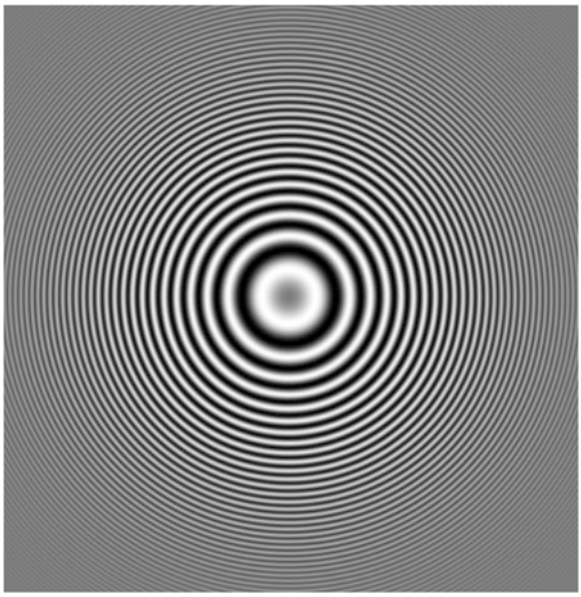


Fig. 2.16 Fresnel zone plate

$$D_m = \sqrt{mf_F\lambda + \frac{m^2\lambda^2}{4}}, \tag{2.8}$$

where λ is the wavelength, f_F focal length, and m is the number of the ring (counted from the center of the FZP.)

The irradiation distribution in the focal plane for a binary amplitude FZP is calculated by the Fresnel-Kirchhoff integral, performing integration across all of the transparent rings $S_1,\ldots, S_n$ [135]

$$E_p = -\frac{E_{0i}}{\lambda} \iint\limits_{S_1,\ldots,S_n} \frac{e^{i\frac{2\pi}{\lambda}r}}{r} \frac{1-\cos\left[\vec{n}\,\vec{r}\right]}{2} ds \tag{2.9}$$

where E_p is the electric field in the observed point of the focal plane, E_0 is the amplitude of the electric field of the plane wave, $\vec{r}$ is the radius vector of the surface element dS FZP, $\cos[\vec{n},\vec{r}]$ is the cosine of the angle between the normal to the FZP and the radius vector of the point P observed from the dS, S_n is the area of the n-th transparent zone.

Figure 2.17 shows the distribution of irradiation in the focal plane, calculated numerically using (2.9). The irradiation power drastically decreases and the focus diameter increases if the number of the zones is smaller.

We further give a connection between the dimensions of a binary amplitude FZP and the active area of an IR detector. For the sake of simplicity, we assume that the

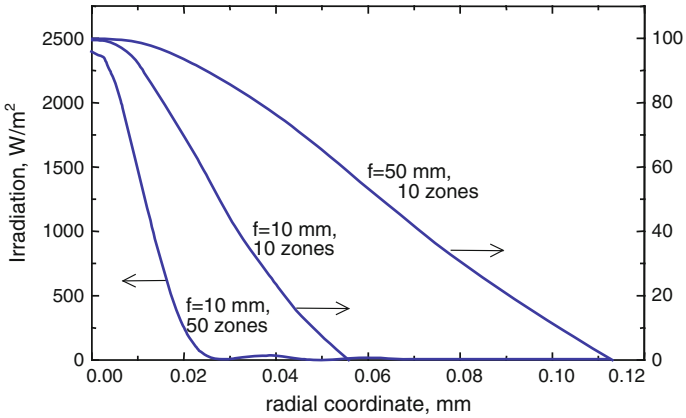


Fig. 2.17 Distribution of irradiation intensity in the focus of a Fresnel zone plate for various numbers of zones and different focal lengths

active area is circular and that a planar monochromatic wave illuminates the FZP. The increase of the electric area of the detector is

$$\frac{A_0}{A_e} = \eta_F \frac{\pi R_F^2}{\pi r_d^2} \tag{2.10}$$

where R_F is the radius of the outer zone of the FZP, FZP, r_d is the radius of our idealized detector, and η_F is the efficiency of the FZP (the ratio between the incident and the transmitted radiation power), here equal to 0.1.

The effective area can be increased either by increasing the FZP radius or by decreasing the photodetector-active area. The maximum diameter of the diffractive lens is limited by the properties of the photomask and by the dimensions of the housing. The minimum diameter of the idealized detector is equal to the diameter of the focus. In the above case, when the focal length is 10 mm, and the FZP diameter calculated according to (2.2) is 5 mm (50 zones), as shown in Fig. 2.18, the optimum detector diameter is about 20 μm. Such a detector coupled with the binary amplitude FZP theoretically receives about 2,000 times more radiation than the same device without a concentrator. For a detector with an active area diameter of 250 μm, the effective improvement achieved by the same diffractive lens is 40 times.

In the case of a detector for the detection of CO_2 laser radiation (10.6 μm) with a FZP diameter of 5 mm, the total number of Fresnel zones 50, and the distance to the active area 10 mm, the focus diameter is 650 μm.

The FZP is the simplest and the least efficient planar diffractive optical concentrator of Fresnel type. Other types are shown in Fig. 2.19.

The simplest and the least efficient binary DOE is the mentioned Fresnel zone plate (Fig. 2.19a), containing only alternating opaque and transparent rings (i.e., an

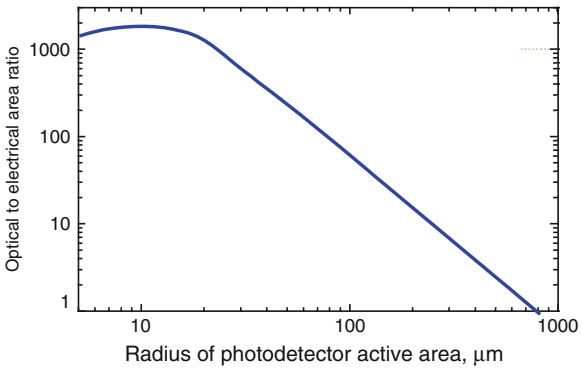


Fig. 2.18 Ratio of optical and electrical area in dependence on the active area radius of a photodetector for a binary amplitude FZP with 50 zones, intended for the detection of IR radiation at 10.6 μm and with a focal length of 10 mm

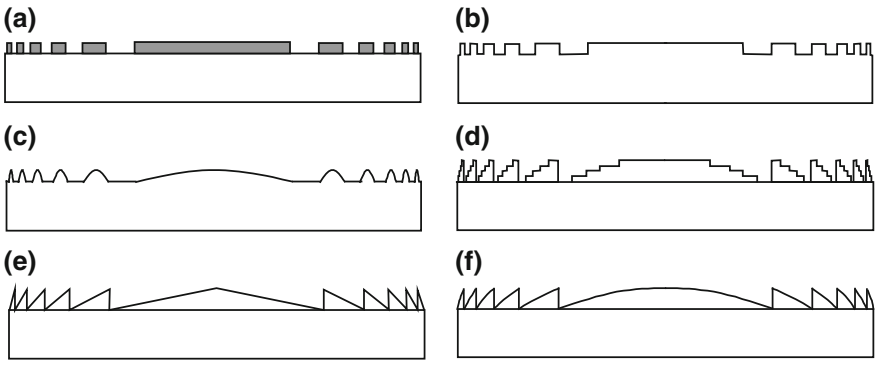


Fig. 2.19 Various types of DOE lenses. **a** Fresnel zone plate; **b** Fresnel phase plate; **c** Fresnel phase plate with "soft" edges; **d** binary multilevel phase FZP; **e** blazed Fresnel lens; **f** Fresnel lens (continual kinoform lens, equivalent to a binary DOE lens with an infinite number of levels)

amplitude-modulated element). Similar but more efficient is the single-level Fresnel phase plate (Fig. 2.19b), where the opaque rings are replaced by phase-shifting rings introducing a phase change of $\pi/2$ (a kinoform-type DOE). Instead of losing 50 % of the incident light intensity, one utilizes all illumination.

In reality, the Fresnel phase plates are often with gradually changing edges, Fig. 2.19c.

The next level of sophistication is the multilevel binary Fresnel lens (Fig. 2.19d). It represents a grayscale-DOE with a finite number of steps (due to the nature of photolithographic process used in its fabrication, it will have 2^N levels, where N is the number of the photolithographic steps.

The diffraction efficiency of a multilevel Fresnel phase plate (the ratio between the incident and the transmitted optical power η_F is calculated as [136]

$$\eta_F = \sin^2\left(\frac{\pi/M}{(\pi/M)^2}\right), \tag{2.11}$$

where $M = 2^m$, while m is the number of levels. For the case shown in Fig. 2.19b ($m = 1$), the efficiency is 41 %, and 99 % for $m = 4$.

The next level is the blazed planar lens where the steps are replaced by sloped planes (Fig. 2.19e).

The most sophisticated is the full Fresnel lens, actually a curved-surface kinoform (Fig. 2.19f), which corresponds to a multilevel binary Fresnel plate where N tends to infinity, but also corresponds to a planarized conventional refractive lens. In other words, this Fresnel lens would be obtained if a conventional lens was cut into rings and if each ring was then thinned to equal thickness by removing from its basis all the excess material that does not contribute to focusing.

2.5.3 Holographic Optical Elements

A holographic optical element (HOE) is a diffractive optical element either designed by computer or used for image forming by holography. In other words, a DOE can be denoted as a holographic optical element if fabricated by holography or intended to be used as a hologram. As an example, a DOE fabricated by ruling using a diamond cutter is not a HOE, but an identical DOE fabricated holographically belongs to the HOE. The design of holograms using numerical methods on a computer is called Computer-generated Holography (CGH).

Holography is defined as a method of recording and reconstructing wavefronts, and optical holography is connected with the light in visible, ultraviolet, and infrared spectrum. A hologram is a material recording of the wavefront phase and intensity distribution of a light beam. Thus it is actually an interferometric pattern created either by an interferometer, a wavefront from a real object interfering with another coherent beam, a wavefront from another holographic recording or a computer-generated interferometric pattern corresponding to an existing or even a nonexistent object. Obviously, holographic fabrication may be used to manufacture any type of diffractive optical element.

Holograms can be classified in a manner similar to DOEs as amplitude (absorption-type) HOE, phase only (kinoform), and mixed ones. Another classification is based on their thickness. They are thin (or planar) if their total thickness is not much larger than their fringe spacing (i.e., the operating wavelength). If it is, they are thick (volume, Bragg-type) holograms.

A volume hologram may be regarded as a superposition of three gratings recorded in the depth of the medium, each satisfying the Bragg law. A volume hologram produces maximum change in refractive index and/or absorption index and offers the largest diffraction efficiencies. It reconstructs the virtual image at the original position of the object if the reconstruction beam exactly coincides with the

reference beam. The conjugate image and higher order diffractions are absent in a volume hologram.

Focal length and wavefront shape of an output beam from a HOE do not depend on the refractive index of the utilized material, although it influences, e.g., the angular dependence of some of its parameters. HOEs are used if it is not practical or is impossible to use refractive optics in a given application [137]. Among the advantages of HOEs are their reconfigurability, the possibility to obtain several different functions in a single element (see Fig. 2.2), the possibility to use beam-splitting to focus radiation on a larger number of spatially separated positions (as in detector arrays), the beam-shaping possibilities to adjust the focused beam to the shape of a detector element or a detector array, the possibility to separate different modes of incident radiation to different detector elements, etc.

Among HOEs an important position belongs to kinoforms, thin-phase holograms with a surface microrelief corresponding to the phase modulation of a given object [138, 139] (i.e., without any data on the radiation amplitude.) Kinoform-based HOEs can operate with incoherent light, i.e., it may be designed for a range of operating wavelengths and allows a relatively wide field of view. Kinoform may be transferred from standard holographic materials and replicated in MWIR and LWIR materials using a number of microfabrication procedures, e.g., reactive-ion etching [140].

HOEs may be regarded as a generalization of Fresnel lenses and their definitions partly overlap—for instance, the kinoform structure shown in Fig. 2.2e is a typical HOE. We may say that Fresnel lenses are a subgroup of holographic optical elements.

The most important parameter of HOEs for infrared detection is η, diffraction efficiency, defined as the ratio between the incident optical flux and the flux forming its image. It is predominantly dependent on absorption losses, i.e., on the absorption coefficient of the HOE material within the given operating wavelength range, but also on the incident angle. Diffraction efficiency of a general HOE is calculated using the coupled-wave theory [134]. The maximum η is obtained for transmission-type phase HOEs (e.g., kinoforms, $\eta \approx 1$), while it is almost two orders of magnitude lower for transmission-type amplitude HOEs ($\eta \approx 0.037$). The second parameter usually applied to describe the behavior of a HOE, signal-to-noise ratio, is of much lesser importance for optical concentrators for photodetectors.

Holographic optical elements with desired characteristics are typically designed using the methods of computer-generated holography (CGH) [141]. The most often used CGHs are the Fourier holograms [142]. Two predominant methods for the calculation of CGHs are Lohman's [143] and Lee's [144]. Holograms are computer generated using the FFT algorithm (Fast Fourier Transform) [145]. The reconstructed image is determined by the inverse FFT. In real optical systems, the Fourier transform of an optical signal is obtained by using refractive lenses.

In Fresnel-type HOEs, the light distribution in the hologram plane is the Fresnel diffractive image of the object. For radially symmetrical kinoforms (rings and circles), 2D Fourier and Fresnel integrals are reduced to 1D Hankel's integrals [146].

During the last years, non-Fourier CGHs are increasingly used [147]. They are used without any additional optics and allow the design of holograms of arbitrary objects, including those not existing in reality, as well as the design of full 3D holograms. Their shortcomings is that they are extremely demanding for CPU resources, especially when calculating reconstructed images. Usually some acceleration methods are used, e.g., the Monte-Carlo method [146] or smaller 2D arrays are generated, e.g., 512 × 512 pixels, and subsequently repeated.

One point is important for the use of HOEs in infrared detection. Standard holographic materials today are photosensitive mostly in the short-wavelength part of the visible spectrum. Since they must be used for infrared radiation, this could cause large aberrations, as well as the breaking of the Bragg condition, especially if the incident wave is not planar. This is the reason why holographic optics for larger wavelengths are typically designed utilizing some of the iterative CGH methods, e.g., Gerchberg–Saxton's algorithm [122]. In this manner, one obtains a holographic optical element diffraction-limited in a desired wavelength range [148]. To this purpose, a Fourier transform is generated or some other method is used to calculate the CGH to be fabricated using standard procedures. Then the image is reconstructed at the operating wavelength (MWIR or LWIR). Finally, the function of the reconstructed image aberration is minimized (a variant of the so-called method of conditional gradient) [122]. Due to a slow convergence and the effect of stagnation, some methods of improvement of these iterative procedures are introduced, e.g., reset of the complex amplitude in each iterative step [149] or the use of adaptive algorithms [122].

DOE lenses for HgCdTe detector arrays were described within the context of HgCdTe detector focal plane arrays [150]. Diffractive (8–16-phase-level) Ge microlenses reducing the detector size in focal plane arrays by 3–6 times were described in [151].

2.6 Antireflection Structures

The increase of incident optical flux into an optical element by decreasing reflection from the incident surface is one of the oldest problems in optics (e.g., [87]). The customary approach is to form some kind of antireflection (AR) structure directly on the surface of the optical element. Antireflection structures may be denoted as thin homogeneous or inhomogeneous layers deposited on the incident optical surface or formed directly in it with the goal to use destructive interference or diffraction to maximize the amount of flux penetrating the interior of the element. An optimum antireflection structure for a given application furnishes the minimum of reflection for given optical properties of the element and the ambient within a given spectral range.

The use of AR structures in the case of IR detectors is even more important than in the visible spectrum, because materials used for the fabrication of both active regions of the detectors and optical concentrators usually have large real parts of

refractive index (typically 3.5–4, for instance in the case of InSb for MWIR, HgCdTe for LWIR and Ge for optical windows and concentrators.) Antireflection structures are usually deposited directly on the active area of the detector, on the collimating optics and on the optical windows.

In our consideration, we assume that the AR structure is nonabsorbing or weakly absorbing, which is valid even for the AR structures incorporating narrow-bandgap semiconductors because of their very small total thickness within a given AR structure. The necessary data for the design of an optimized AR structure are the required spectral range of minimized reflection, the real part of the refractive index of the substrate and the ambient, as well as the minimum and maximum limits of the refractive index for the AR structure. Sometimes the maximum optical thickness of the whole AR structure is also given.

Two main groups of antireflection structures are single- or multilayer AR layers composed of dielectric films (interference films) [152] and single or multiple periodical diffractive structure of either amplitude of phase type. Multiple AR structures may have quarterwave or quarterwave/half-wave periodicity or they may be aperiodic [153].

2.6.1 Antireflection Interference Films

An optically thin film may be defined as a plane-parallel structure with a thickness of the order of the wavelength of light, characterized by its complex refractive index and thickness. The simplest and cheapest, but the least effective approach to the minimization of reflections on the incident optical surface is to deposit one or several optically thin films on it, with a thickness and refractive index value optimized to furnish maximum transmission. The method is based on destructive interference between the strata within the AR thin film. Such interference-based dielectric films [152] or interference coatings were chronologically the oldest structures used to minimize reflection on the incident optical surface. They consist of layers whose optical thickness (the product of their real thickness and refractive index) is equal to the targeted quarter-wavelength multiplied by an odd integer. This method was conceived in the early 1950s [87]. Usually, the refractive index is alternating between neighboring layers, thus forming layer pairs, while the optical thickness is kept constant. During the last few decades, optimized AR films are also produced by varying both the optical thickness of layers and their refractive index.

Antireflection interference thin-film coatings based on homogeneous strata may be grouped as [154]

(a) single layer,
(b) multilayer, periodic,
(c) multiple, quasiperiodic,
(d) multiple, aperiodic (for instance step-down profile), and
(e) multiple, random.

The simplest AR coatings are the single-layer ones. They consist of a single homogeneous layer used for most of the low-end applications. Currently, they still comprise the majority of all existing AR coatings, since their performance is satisfactory if a transmission of the order of 95 % fulfills the posed requirements. Theoretically, there is a single wavelength within a given range for which a single AR layer furnishes 100 % transmission.

For single-layer quarter-wavelength homogeneous antireflection film, the condition of reflection minimum can be calculated as

$$n_1 = \sqrt{n_0 n_s}, \tag{2.12}$$

where the refractive index of the single homogeneous AR film is n_1, that of the substrate n_s and of the surrounding medium n_0. The optical thickness of the layer should be

$$d_1 = \frac{\lambda}{4n_1}. \tag{2.13}$$

Multilayer structures offer much better performance and vastly superior possibilities for optimization. Among their main advantages are the possibility to obtain minimal reflection in a wider spectral range and facilitated choice of convenient materials and technologies.

The condition of reflection minimum for a double-layer quarter-wavelength antireflection film made from homogeneous strata is

$$n_1 n_2 = \sqrt{\frac{n_s}{n_0}} \tag{2.14}$$

where the refractive indices of the first and the second film within the AR coating are n_1, and n_2.

The condition of reflection minimum for a three-layer quarter-wavelength antireflective film with homogeneous strata is

$$\frac{n_1 n_3}{n_2} = \sqrt{\frac{n_s}{n_0}} \tag{2.15}$$

where the refractive indices of the first, second, and third film are n_1, n_2, and n_3, respectively.

With an increase of the number of layers, the spectral bandwidth of the reflectivity minimum becomes wider and the mean deviation from the zero reflectivity lower.

Figure 2.20 shows a schematic presentation of a generalized multilayer AR film with a total of m layers. The refractive index and the thickness of each layer can be different.

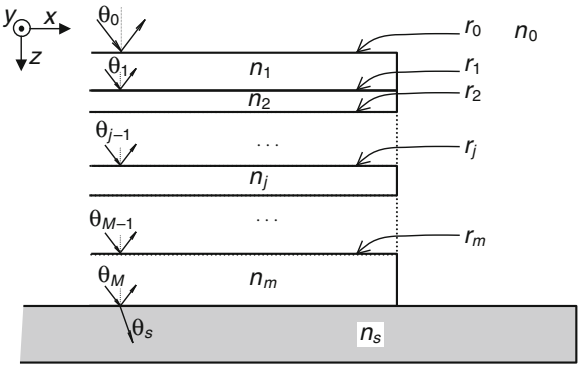


Fig. 2.20 Multilayer antireflection film consisting of discrete and homogeneous all-dielectric strata

Reflection and transmission coefficients for multiple interference antireflection layers may be calculated in the following manner. For each j-th layer, we write the expression for Fresnel amplitude reflection coefficient r_j^* for the next m–j layers located behind it. This expression features the reflection at the surface of this layer r_j, its phase φ_j and the amplitude reflection coefficient of all the remaining m–j–1 layers. By writing successively in the same manner the expressions for the remaining layers, we obtain a recursive system with an infinite number of solutions

$$r_j^* = \frac{r_j e^{i\varphi j} + r_{j+1}^* e^{-i\varphi_{j+1}}}{e^{i\varphi_j} + r_j r_{j+1}^* e^{-i\varphi_j}}, \tag{2.16}$$

$$r_j = \frac{n_{j-1} - n_j}{n_{j-1} + n_j}, \tag{2.17}$$

$$j = 1, 2, \ldots, j, \ldots, m.$$

Here m denotes the total number of layers, n_j is the real part of refractive index for the j-th layer. The layers are all-dielectric, i.e., the absorption of the AR film is neglected. The amplitude reflection coefficient of the whole film is further made equal to zero (the condition for the reflection minimum) and then the parameters of the whole multilayer structure are numerically optimized.

For given spectral limits, the best results are obtained if using discrete layers with homogeneous composition and with a quarter-wavelength optical thickness [155],

$$d_1 n_1 = d_2 n_2 = \cdots = d_j n_j = \cdots = d_m n_m = \frac{\lambda}{4} \tag{2.18}$$

For a substrate with a refractive index larger than that of its ambient it is simultaneously necessary that

$$n_0 < n_1 < \cdots < n_j < \cdots < n_s \tag{2.19}$$

(step-down refractive index) [155].

The thickness of the j-th layer is

$$d_j = \frac{d}{\sum\limits_{\substack{k=0 \\ j \neq k}}^{m} \left(\frac{n_j}{n_k} + 1\right)}. \tag{2.20}$$

For a single AR layer, the above condition reduces to $d = \lambda/4n_1$ and $n_1 = n_S^{1/2}$, assuming that the incident medium is air.

For multilayer quarter-wavelength antireflection films with homogeneous strata (step-down refractive index film) [156, 157], the amplitude reflection coefficient determined for the whole structure in the aforementioned manner is made equal to zero (the condition of reflection minimum) and numerical optimization is further performed to obtain the strata parameters. In that case, the above-mentioned step-down condition is applied that refractive index decreases in steps from the n_s value at the substrate surface to the n_0 value of the surrounding medium [155].

A problem with quarter-wavelength layers is to find materials with convenient refractive index, since material with the required refractive index value may not even exist. There are several possible approaches to this problem [158]. One is to utilize a combination of two material furnishing an effective refractive index near the required value. Another way is to deposit two homogeneous layers whose transfer matrix is equal to the matrix of the desired material (the Herpin equivalent-index concept) [159]. Third method is to use AR layers aperiodic both regarding the layer thickness and their composition. It was shown that a combination of an arbitrary number of different materials can be mathematically reduced to a combination of two materials only, for instance by utilizing the equivalent-index concept. This fact may be used to simplify the design of AR layers.

The problem to find an optimal AR film in the most general case when both thickness and refractive index may be arbitrarily changed reduces to an optimization procedure in a multidimensional parametric space. It is assumed [160] that for a given spectral range and a given interval of values of refractive index there must be a single optimal generalized AR film structure, i.e., there must be a global solution of the problem, giving the lowest reflection coefficient in the whole range.

A method to solve the problem is to determine in the Fourier space the connection between the logarithm of refractive index values and the amplitude reflection and transmission coefficients, represented as complex wavelength-dependent functions. The global minimum of thus obtained dependence is then determined. The solution is an inhomogeneous layer, further transformed into a two-material system and subsequently subjected to a new procedure of fine optimization.

Among a number of approaches to the calculation of the global minimum, we mention the needle optimization and the method of quadratic programming [153]. The needle optimization obtained its name because it is based on a procedure of introducing a Dirac function (“needle”) into the refractive index profile and thus

determining the best position to introduce a new layer; fine optimization is performed afterward. The procedure is repeated until no new layers can be introduced.

The method of quadratic programming starts from a quality function equal to the integral of the ratio $R(\lambda)/[1 - R(\lambda)]$ over wavelength from the beginning to the end of the optimization interval. The solution here is also an inhomogeneous layer transformed subsequently into a two-material one and finely optimized [158].

Different approaches to the calculation of the global reflection minimum of a generalized AR film furnish very similar solutions, independently on the algorithm used and the initial conditions, which points to an existence of a unique solution for the given conditions. The maximums of refractive index in an optimized AR film are grouped. This is a consequence of the fact that complex filters are basically a combination of two interference filters with different periods [153]. A combined AR film profile is a result of the beating of their periods. The long-wavelength filter causes the mentioned grouping, determines its period, and creates a high-reflection peak at the long-wavelength boundary of the AR range. The short-wavelength filter determines the fine structure of groups and defines a reflection peak at the opposite end of the AR range. The reflection peaks at the boundaries of thus designed AR layers are very high, reaching almost 100 %.

The main disadvantage of all-dielectric interference-based AR films is that due to their very nature their performance sharply drops for oblique angles, i.e., only small deviations of the incident angle are permitted from the value for which a coating was designed. A typical AR interference coating is designed for normal incidence of light. This can be somewhat improved if material pairs with a large refractive index ratio are used. Other problems include technological issues, for instance cohesion problems and thermal matching between the substrate and the deposited layers. A design problem is a limited choice of materials with suitable refractive values of refractive index. This can be alleviated by the use of porous materials.

The most often used materials for the fabrication of AR layers in the MWIR and LWIR range are ZnS, MgF_2, and similar ones. Magnetron sputtering is very often used, reactive or nonreactive, either dc or RF. It gives layers with optical quality while at the same time offering the possibility to control their thickness very accurately. Among the methods used to deposit thin single or multiple coatings with antireflective properties is the deposition from solution and further curing, baking, and chemical processing. Various kinds of evaporation are used, both reactive and nonreactive. A method offering less porous coatings is ion-assisted deposition. One of the methods is to deposit first a thin metal layer and then to convert it to oxide or hydroxide AR coating. Chemical vapor deposition may also be used.

As an illustration, Fig. 2.21 shows experimental reflection diagrams for a germanium optical concentrator for uncooled HgCdTe photoconductors with a single-layer zinc sulfide AR coating deposited by RF sputtering.

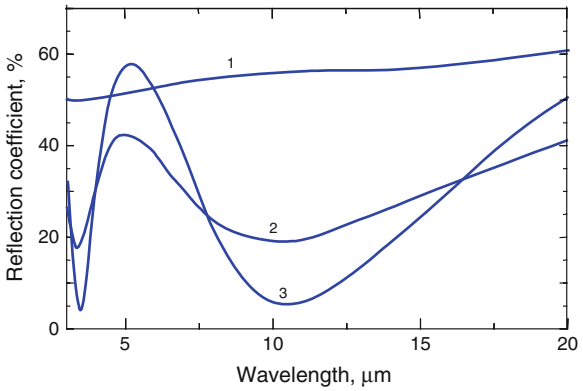


Fig. 2.21 Experimental spectral reflection for single-layer ZnS AR coatings. *1* Ge substrate 2 mm without AR coating. *2* ZnS layer 1.2 μm deposited on the top surface of the Ge substrate; *3* The same layer as in (*2*) deposited both on *top* and the *bottom* surface

2.6.2 Gradient-Index Single-Layer Dielectric Films

If the refractive index of a coating is reduced uniformly and continually within the AR coating from the substrate value to the surrounding medium value, excellent antireflection properties are obtained in a very wide wavelength range (e.g., [161, 162]). This is not surprising, since this situation can be represented as a step-down index profile with an infinite number of infinitesimally thin homogeneous strata. However, contrary to the conventional interference-based AR single- or multilayers, their operation is omnidirectional. They are polarization insensitive, i.e., their reflectivity is low for both transverse electric and transverse magnetic electromagnetic waves. A theory of gradient-index thin films can be found in [163].

Different functional dependences may be used for gradient-index AR layers. The simplest one is linear, but various polynomial dependences are met, for instance cubic or quintic. Sinusoidal dependence is also used. Of these, quintic dependence has been reported as the closest to optimum [164].

The fabrication of gradient-index materials with refractive index values ranging from that of the air to the values found in semiconductors presents significant technological challenge [152]. Optical materials with the values of refractive index close to 1 are difficult to produce. One of the possible approaches is to use porous materials whose effective refractive index nears that of the air. Xi et al. reported material with an effective refractive index of 1.05 [165] based on porous silica utilizing nanorods produced by oblique-angle vapor deposition. Minot used chemical etch-leach process to produce porous silica gradient films [166]. Electrochemical etch procedures for porous silicon structures furnishing controllable refractive index gradient was reported in [167]. Polymer-based porous gradient-index films using polystyrene-block-poly(methyl methacrylate) were described in [168].

2.6.3 Antireflective Diffractive Structures

Antireflective properties may be generated by forming a subwavelength surface relief diffractive structures on the detector-active area (ARS, antireflective structures [169]). Such structures are also denoted as high spatial frequency structures or subwavelength surface textures [119]. Antireflection structures generally have a twofold function in light management in photodetectors. Since effectively they represent gradient-index structures, they ensure minimum reflection at the surface. However, they also function as scatterers for light arriving from the inside of the active region and return the beams back, which also increases the optical path and ensures improved absorption.

Most of the ARS function as diffractive gratings of the zero-order. The optical behavior of subwavelength gratings can be satisfactorily approximated in a majority of situations by using the effective medium theory [170]. Figure 2.22 shows two main situations that can be encountered.

One of the presented structures is a monodispersion of subwavelength inclusions *i* (spheres) in dielectric host *h*, Fig. 2.22a. The other is polydispersion, Fig. 2.22b. The first situation can be described by the well-known Maxwell-Garnett model [171], the oldest effective medium model, obtained by the use of Clausius-Mossotti/Lorenz-Lorentz equation. The other case is polydispersion, described by the implicit Bruggeman expression [172, 173].

Diffractive ARS are usually calculated either using the effective medium approximation, or the rigorous coupled-wave analysis [169]. Direct numerical solution of Maxwell equations and the approach using the plane wave theory were also used [174].

The effective medium theory (EMT) or the homogenization method regards a structured diffractive medium as a homogeneous material with an effective refractive index equal to a weighted and spatially averaged value of the refractive index of both structural materials. This is a very old theory and it was considered even in the works of Lord Rayleigh near the end of the nineteenth century.

The simplest is the zeroth-order EMT, that considers the diffractive medium in static approximation (grating transmits radiation as an unstructured medium). The

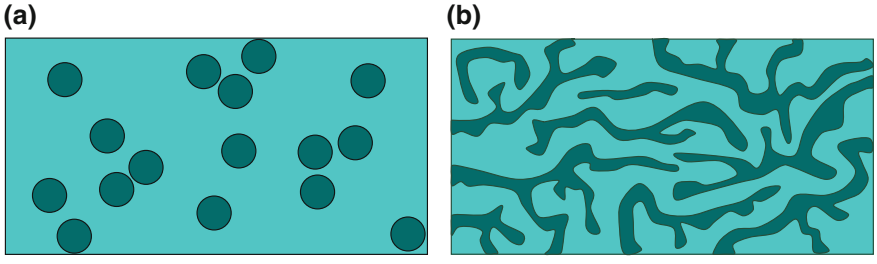


Fig. 2.22 Effective medium approach. **a** Spherical subwavelength inclusions as monodispersion in homogeneous host (Maxwell-Garnett model); **b** polydispersion (Bruggeman model)

condition of validity of the zeroth-order EMT is obtained according to the well-known Bragg equation of optical grating where the diffractive order is made equal to zero

$$\frac{\Lambda}{\lambda} < \frac{1}{\max\left(n_j, n_s\right) + n_1 \sin\,\theta_{\max}} \tag{2.21}$$

where $\theta_{\max}$ denotes the maximum incident angle. The condition of the existence of a zeroth-order transmitted wave is

$$\begin{aligned} &n_j < n_s \rightarrow 0 \le \theta \le \pi/2 \\ &n_j > n_s \rightarrow 0 \le \theta_j \le \arcsin\left(\frac{n_s}{n_j}\right) \end{aligned} \tag{2.22}$$

Thus we can calculate the angles of total reflection for the propagation through a subwavelength grating from an optically denser to an optically thinner medium. For instance, for HgCdTe this angle is 16.5°. Therefore the flux within the active area is additionally increased, because the returning beam is reflected from the incident surface back into the active region.

Based on the above diffraction equation, we introduce the miniaturization factor of an AR structure as $\Lambda/\lambda = [\beta(n_j + n_s)]^{-1}$. This factor should be maximized in order to simplify technological procedures for grating fabrication. In practical situations, the ratio between the grating step and the wavelength is about 10.

The zeroth-order EMT is not applicable for large refractive index modulations. Also, due to technological reasons, large patterns not satisfying the conditions of validity of zeroth-order EMT are used for ARS. In these cases, the second-order theory is used. It was developed in 1956 by Rytov for the case of magnetic and conductive materials [175–177]. In the case of ARS, one can utilize its form for dielectric media [178–180].

First, we define the grating vector (stratification vector) $\vec{K}$ as a normal to the boundary between two periods of a 1D optical grating. We utilize a coordinate system where the *x*-axis is parallel with this vector, while the *y*-axis is normal to it and to the incident surface. We use f_S to denote the fill factor equal to the ratio of the width of the part with the material having a permittivity ε_s to the total grating period Λ.

In the second-order EMT, the effective value of relative dielectric permittivity (square of effective refractive index) can be written as [176]

$$\begin{aligned} \varepsilon_{\vec{E}\|\vec{K}}^{(2)} &= \varepsilon_{\vec{E}\|\vec{K}}^{(0)} \left[1 + \frac{\pi^2}{3}\left(\frac{\Lambda}{\lambda}\right)^2 f_S^2 (1 - f_s) \frac{\left(\varepsilon_s - \varepsilon_j\right)^2}{\varepsilon_0} \varepsilon_{\vec{E}\perp\vec{K}}^{(0)} \frac{\varepsilon_{\vec{E}\|\vec{K}}^{(0)}}{\varepsilon_s \varepsilon_i}\right] \\ \varepsilon_{\vec{E}\perp\vec{K}}^{(2)} &= \varepsilon_{\vec{E}\perp\vec{K}}^{(0)} \left[1 + \frac{\pi^2}{3}\left(\frac{\Lambda}{\lambda}\right)^2 f_S^2 (1 - f_s) \frac{\left(\varepsilon_s - \varepsilon_j\right)^2}{\varepsilon_0} \frac{1}{\varepsilon_{\vec{E}\perp\vec{K}}^{(0)}}\right]. \end{aligned} \tag{2.23}$$

The index (0) denotes the EMT approximation of the zeroth order

$$\begin{aligned} \varepsilon_{\vec{E}\perp\vec{K}}^{(0)} &= f_S\varepsilon_s + (1-f_s)\varepsilon_j \\ \varepsilon_{\vec{E}\parallel\vec{K}}^{(0)} &= \left[\frac{f_S}{\varepsilon_s} + \frac{1-f_S}{\varepsilon_j}\right]^{-1} \end{aligned} \tag{2.24}$$

It can be seen that both zeroth and second-order EMT expressions are polarization dependent, which means that the 1D ARS has a property of structural (built-in) form birefringence.

For multilayer 1D ARS after introducing the effective refractive index and reducing a single layer to an equivalent homogeneous one, we may use the same recurrent relations as for the usual AR dielectric layers. A m-level film may be produced by using 2^m photolithographic steps. It is interesting that multilevel structures do not have to be symmetric, i.e., only the fill factor of a given level is important, not the geometrical position of the filling itself. Structures with a fill factor continually changing from the substrate to the top may be approximately represented as a large enough number of infinitesimally thin discrete and homogeneous rectangular elements.

For two-dimensional AR structures, literature quotes a number of different approximations (e.g., [162]). Lalanne proposed expressions for effective index in the form of infinite series [181]. Raguin [169] and dos Santos [182] showed that there is a direct correspondence between the optical properties of two-dimensional subwavelength ARS and homogeneous materials.

In practical calculations of 2D ARS (both single- and multilevel) and of 1D near cutoff (comparable Λ and λ), one may use the rigorous coupled-wave analysis (RCWA) [180]. This theory was used for the case of rectangular AR structures by Moharam and Gaylord [174, 183, 184]. The electromagnetic field for a planar structure is first expanded over harmonics and together with the Bloch periodicity condition replaced in the wave equation written over TE or TM polarized waves. In this way, an infinite set of coupled-wave equations is obtained [174]. By transforming this system and applying the boundary conditions for tangential components of electric and magnetic field, an infinite set of linear equations is obtained. Finally this set is truncated, and an arbitrary accuracy is obtained by retaining a sufficient number of equations. This is used to calculate the amplitude of certain orders of electromagnetic waves, and based upon this the diffraction efficiency of the structure is determined.

In the case of surface relief with an arbitrary shape, the relief is divided into a large number of homogeneous thin layers parallel with the surface. The above method is then applied to each of the layer and the boundary conditions are applied to the dividing surface between each pair of the planar layers. For a rectangular structure, this is reduced to a single layer.

The function of an ARS can be described by considering its profile, as shown in Fig. 2.23.

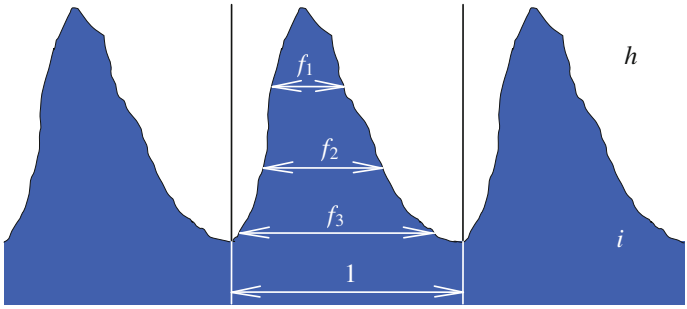


Fig. 2.23 Applying effective medium approach to diffractive antireflection structures. Here f_i denotes the filling factor at different heights, the largest value being 1 and the lowest 0; $f_1 < f_2 < f_3$

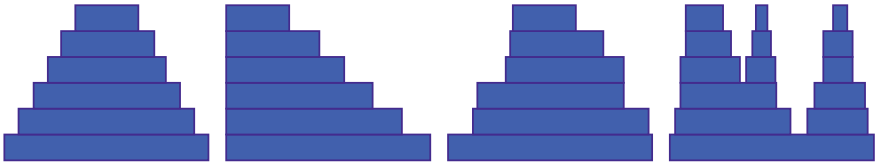


Fig. 2.24 Equivalent ARS structures with different profiles but with identical dependence of filling factor on height

It can be seen that the ratio of the amount of the ARS material (denoted by i) and the surrounding medium (denoted by h) changes with height, f being larger nearer the active surface (bottom). If one divides the ARS in infinitesimally thin horizontal sheets and applies the effective medium theory to each one of them, one sees that the relative dielectric permittivity increases gradually from about 1 on the top (since there $f = 0$) to the value of the active semiconductor permittivity at the bottom—i.e., the ARS shown in Fig. 2.23 effectively behaves as a graded refractive index structure, shown to be the most efficient planar AR structure [162].

Since the structures in question are subwavelength and observed as an effective medium with a given filling factor at a given height from the detector surface, it is then irrelevant which profile exactly it follows, as far as the filling factor does not change. In other words, the profiles shown in Fig. 2.24 are identical from the point of view of their properties as antireflective structures.

The range of acceptable incident angles for an ARS is very large and it can be designed for optimum operation for a wide range of operating wavelengths. Another important advantage is that the ARS can enable fabrication of any desired effective refractive index, including the values that do not exist in nature. Also, they do not pose the problem of material adhesion to the substrate and thermal expansion coefficient matching between the substrate and the ARS.

A relief ARS array may be one-dimensional (Fig. 2.25) or two-dimensional, Fig. 2.26. In both cases it may be fabricated with a continuous profile, but also as binary or multilevel. Most frequently mentioned in the literature are 1D ARS with a

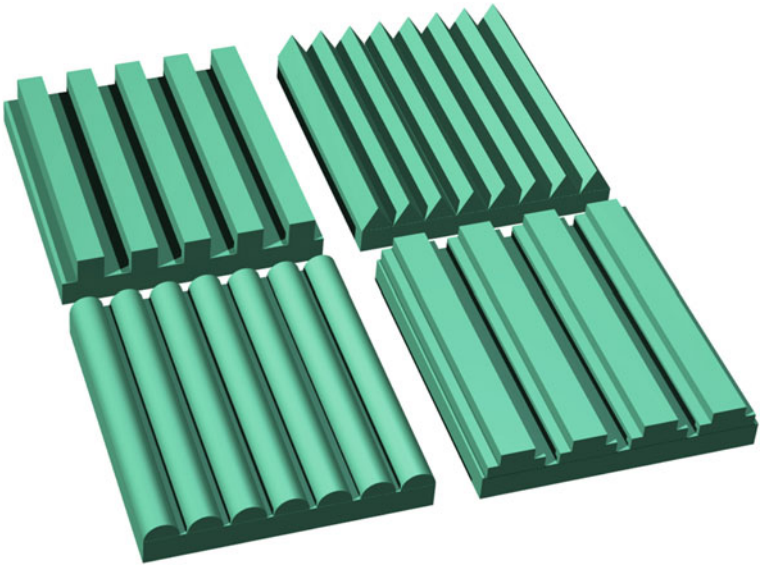


Fig. 2.25 Various 1D subwavelength ARS. *Top left* 1D rectangular structure; *top right* 1D sawtooth (triangular, V-groove) structure; *bottom left* cylindrical; *bottom right* 1D two-level, rectangular

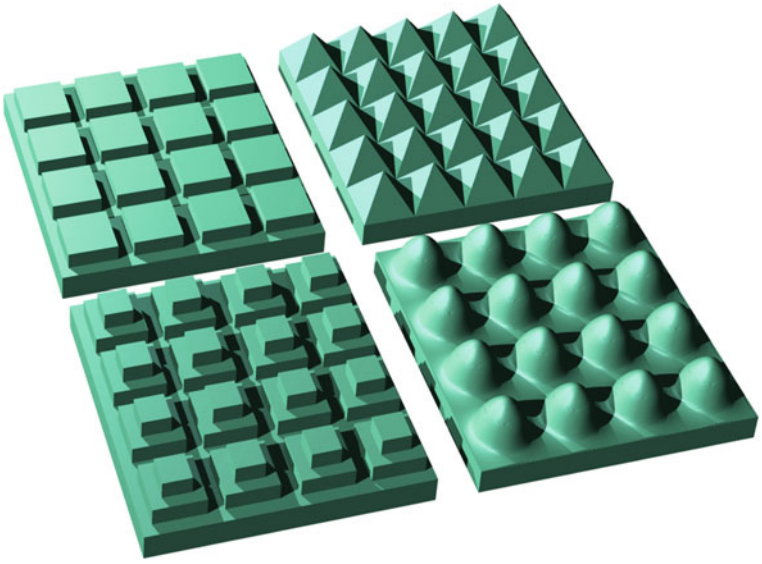


Fig. 2.26 2D subwavelength ARS. *Top left* parallelepiped profile; *top right* pyramidal shape; *bottom left* two-level parallelepiped (binary); *bottom right* 2D sinusoidal

rectangular relief profile [174], triangular ("toothed," "saw," "sawtooth," or "V-shaped") [185] and sinusoidal [119, 182]. Two-dimensional ("crossed") structures may be parallelepiped-shaped (single- or multilevel) [182], pyramidal (truncated or full) [186], sinusoidal [187], etc. Obviously, the structures shown in Figs. 2.25 and 2.26 are only a few of the basic versions and many other geometries are possible. For instance, one may use the "negatives" of each of the presented structures—where protrusions are shown, one can use instead holes of the same shape. Instead of pyramids shown in Fig. 2.26 top right, one could use pyramidal holes, etc.

Different technologies are available for the fabrication of micrometer or submicrometer-sized subwavelength ARS for middle- and long-wavelength infrared radiation. An ARS profile may be fabricated in resist utilizing photolithography, directly in the active area of the photodetector or in an amorphous buffer layer deposited for instance by sputtering. A variant is holographic lithography (being a standard method for the fabrication of Bragg gratings), used for submicron-pattern ARS. Direct-writing electron beam and focused ion beam lithography are also used.

To transfer a photolithographically formed pattern of an ARS into the active area or into a buffer layer techniques of wet and dry etching are used. Among the methods of wet etching, a frequently mentioned one is lift-off technology. In the processes of etching, sometimes additional buffer layers are used as an etch stop to accurately control etching depth. Much more often methods of dry etching are used, especially reactive-ion etching (RIE) and chemically assisted ion etching. Electron-beam lithography, ion milling and ion implantation are also used.

Numerous methods for nanofabrication were applied to fabricate patterned AR 3D profiles. Bataglia et al. used self-assembled nanospheres to perform nanosphere lithography by reactive-ion etching and obtain patterns of nanocavities on glass [188]. Spinelli et al. [189] produced "black" silicon by depositing silicon nanopillars 250 nm in diameter, 150 nm high on the silicon substrate. The nanopillars were ordered in a square 2D lattice with a 450 nm pitch. They acted as subwavelength resonant Mie scatterers, thus producing large scattering cross-sections with large forward scattering. The obtained reflection was 1.3 % in the visible. Like other ARS, the structure was omnidirectional.

One of the possible approaches for the photodetectors in (8–14) μm range (mercury cadmium telluride and similar zincblende crystals) is to fabricate a 2D array of openings by anisotropic etching through windows in photolithographic mask (along crystallographic axes). The method is somewhat similar to the one used to characterize etch-pit dislocations. In dependence on the crystallographic orientation and the etching solution used, the shape of the openings may be tetrahedral (orientation (111)Cd), trapezoidal (100)Cd, or pyramidal $(1\bar{1}\bar{1})$Cd. If preferential etching is performed without the aid of a mask, stochastically distributed tetrahedrons are obtained.

High-resolution (submicrometer) photomasks may be used to fabricate microopenings by isotropic etching of HgCdTe using Br-methanol solution. In this way calottes and hemispheres are obtained that may be separated or overlapping.

Diffractive ARS offer important advantages over interference AR coatings. They can enable fabrication of any desired effective refractive index. They do not pose the problem of material adhesion to the substrate and thermal expansion coefficient matching between the substrate and the ARS. Contrary to homogeneous dielectric films, diffractive ARS can be used for a much larger range of incident angles, exceeding 50°.

2.6.4 Bionic Approach: Moth-Eye Structures

Biomimetics, bionics, or bio-inspired design is a very important direction of fabrication of highly efficient engineering structures generally. Particularly, numerous works have been dedicated to biomimetics in photonics [190, 191]. It has been observed that some insects developed organic formations that act as very efficient antireflection structures. Some species of moths have their compound eyes covered by ordered arrays of submicrometer nipples. These structures, also variously called protuberances or nodules, are rounded at the top and arranged in a hexagonal lattice with a periodicity of about 250 nm. It has been found that they act as very efficient antireflection structures in a way similar to the previously mentioned antireflection diffractive surface relief with gradient effective refractive index profile. Very similar structures have been observed on the transparent wings of some hawkmoths, where they are responsible for their visual appearance of blackness, like a kind of natural "stealth" design. First considerations of such extraordinary properties of the eyes of moths came relatively early [192].

Figure 2.27 shows the appearance of a moth-eye surface profile. The structure exhibits strong antireflection properties for all polarization and is omnidirectional. Its bandwidth is extremely wide and covers the range from UV to the terahertz region.

Some literature sources go so far that according to them all subwavelength antireflection reliefs with a gradient of effective refractive index are called moth-eye structures (e.g., [193]), including 2D pyramidal and nanopillar arrays.

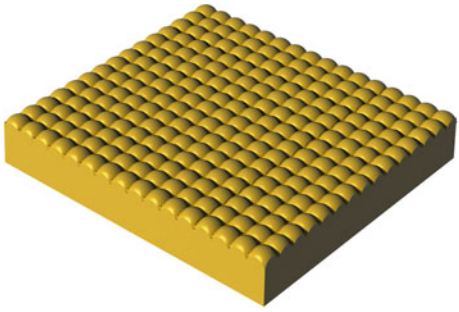


Fig. 2.27 Illustration of an artificial antireflection structure based on the moth-eye principle

Decades passed before it was technologically possible to mimic such nanostructures and fabricate artificial bionic moth eye ARS [191, 194]. Technologically, moth eyes can be produced at a correct scale using lithography, crossing three gratings at 120°. Another approach to fabrication is electron-beam etching.

Harker et al. produced artificial moth eye structures in silicon, germanium, and diamond [193]. Huang et al. fabricated bio-inspired silicon nanotips as antireflection structures [194].

One should mention here the works of Wang who implemented a topology optimization procedure by genetic algorithm to developed a complex optimized surface relief (not a moth-eye structure), as a result claiming efficiencies up to three times over the Yablonovitch limit [195].

2.6.5 Random Surface Corrugations

Besides the regular (patterned) profiles, there are a number of random or "stochastic" antireflective diffractive structures, i.e., structures utilizing arbitrarily distributed elements (whose shapes themselves may be regular or not). Usually their behavior is similar to those with pyramidal or sinusoidal relief, because their effective density gradually increases from the ambient toward the substrate. These stochastic reliefs may consist of randomly distributed smaller or larger geometrical shapes obtained by substrate etching, e.g., tetrahedrons, rounded "pillars" [196], hemispheres, etc. or the geometrical shapes themselves may be fully random (Fig. 2.28).

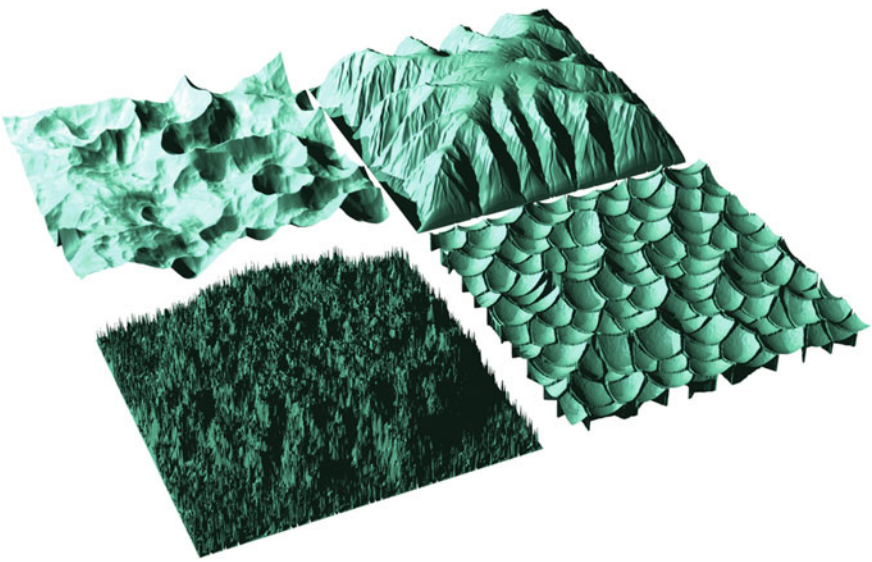


Fig. 2.28 Some stochastic surface reliefs for ARS

The question which kind of antireflection surface relief is the most efficient, regular or random, is still open to debate [188]. Some of the most efficient detectors in the solar cells field actually have random surface profiles—stochastically distributed pyramids made in transparent conductive oxide (zinc oxide) using chemical vapor deposition [197]. However, ordered and random profiles both effectively behave as gradient refractive index profile starting from the effective value nearly equal to that of the environment index and ending with that of the substrate. In this manner, the question reduces to the determination of an optimal gradient-index profile.

The scattering properties of surface profiles can be described by two main parameters, the haze and the angular distribution functions [198]. The haze is defined as the portion of the light that is diffusely (i.e., nonspecularly) scattered into the far field, while angular distribution function describes angularly resolved scattering. Rockstuhl et al. [199] made a connection between these parameters and the absorption properties in the active region and performed a numerical calculation to determine the optimum stochastic surface profile. The conclusion is that a random surface relief should scatter light as strongly as possible and that simultaneously the angular domain of that scattering should be as large as possible in order to ensure maximum transmission.

To create "stochastic" subwavelength antireflection relief anisotropic etching of single crystalline layer was used [200, 201]. Another method is reactive sputtering in plasma [196]. Some authors utilized chemical vapor deposition [197]. Random arrays of core-shell nanowires grown by vapor–liquid–solid method were reported in [202]. For HgCdTe, wet chemical etching and dry etching in plasma were used [203]. The most often used wet etchant is the solution of bromine in methanol. Other solutions of bromine in alcohols (butanol, ethylene glycol) or acids are also utilized. These methods are preferred since they do the least damages to the HgCdTe crystal lattice. Various dry etching solutions are applied as well, using plasma based on methane, hydrogen, argon, and nitrogen, sometimes with added halogens like fluorine or chlorine (convenient for stochastic surface profiles).

A problem with random profiles is their repeatability from sample to sample, since it depends on a number of process parameters and their very nature is stochastic. Some solutions that could be applied to overcome this is to use advanced replication technologies like, e.g., nanoimprint to replicate the optimized random profile. Another is to apply top-down planar fabrication to pattern quasiperiodic structures that behave as pseudorandom, although they are actually deterministic.

2.6.6 Subsurface Scatterers

It is possible to increase the optical path through a photodetector by incorporating different structures within the active region itself so that they act as scatterers of light. In this way, a beam that would else leave the detector remains within the absorptive domain. Structurally, one could use inclusions, fillers, pores, voids, etc.

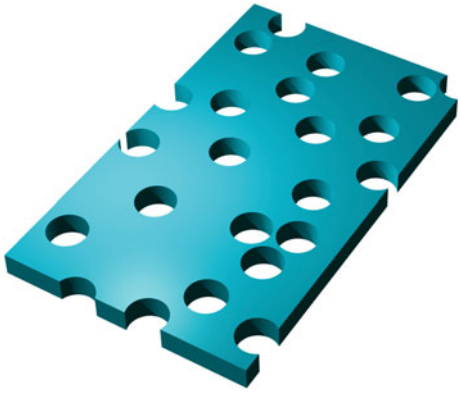


Fig. 2.29 Disordered 2D array of nanoholes for light trapping in thin absorptive media

with subwavelength dimensions. After entering the active region, light beams are diffracted on scatterers which increases their overall optical path and effectively traps them. This mechanism is actually similar to the scattering of the internal beams within the detector from the surface antireflection reliefs. Like in the case of other structural arrays for detector enhancement, the subsurface scatterers may be ordered/periodic, quasiperiodic (with a benefit that one has a deterministic and repeatable structure that behaves as a quasi-random one), aperiodic or fully random.

A method of the use of scatterers incorporated into the absorptive region was described by Vynck et al. who produced a random array of nanoholes in absorptive film [204]. Multiple scattering here is two-dimensional and leads to an appearance of disordered quasi-guided modes. Their structure is schematically presented in Fig. 2.29.

2.7 Metamaterials as Antireflection Structures

Compositing of materials at subwavelength level brought to a novel class of structures known as electromagnetic metamaterials [205, 206]. They could be defined as artificial structures whose effective electromagnetic properties exceed those found in nature. By tailoring their geometry and material properties, one is able to directly influence their frequency and spatial dispersion. In this manner, it is possible to obtain effective relative dielectric permittivity, relative magnetic permeability, and refractive index in a vastly expanded range compared to naturally occurring materials. Group velocities in metamaterials can be designed to be much larger than in free space (the superluminal group velocity—the so-called "fast light," corresponding to very low values of refractive index, including those near zero), much smaller ("slow light"—very high effective refractive index), and even negative ("left-handed light"—effective refractive index below zero) [207].

Structurally, metamaterials are usually metal-dielectric composites ordered in 1D, 2D, or 3D arrays, where a certain metallic pattern (denoted as electromagnetic

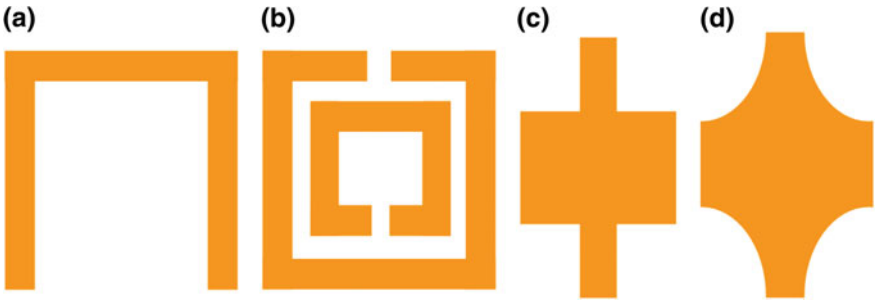


Fig. 2.30 Some metamaterial building blocks ("particles"). **a** Single split ring, **b** double split ring, **c** square fishnet, **d** elliptical fishnet

"atom", "molecule" or "particle" because of the analogy with natural particles and their electromagnetic response) is designed to furnish a desired response and then repeated to ensure the designed effective properties, Fig. 2.30. This is why such artificial "atoms" (which obviously comprise a very large number of natural atoms) must be much smaller than the operating wavelength in free space. Engineering of group velocity is obtained by adjusting their capacitive or inductive behavior, i.e., their resonant/antiresonant response to electromagnetic radiation.

The most often met class of metamaterials are those with negative effective refractive index. These are structures in which effective permittivity and permeability are negative in the same frequency range (the "double-negative" materials). Envisioned as a curiosity in 1967 by Veselago [208], practically introduced by Pendry [209] and experimentally fabricated by Shelby et al. [210], they have a plethora of unusual properties. The direction of the Poynting vector in negative index metamaterials is opposite to that of the wavevector, i.e., electric, magnetic, and wave vectors form a left-oriented triplet. Hence they are often called "left-handed materials".

It has been suggested that negative index metamaterials can be used for the fabrication of near-field lenses that exceed the diffraction limit and operate in subwavelength range (perfect lenses and superlenses) [211]. A whole new subfield of electromagnetic optics has been derived from the paradigm of metamaterials, the transformation optics [212, 213]. One being able to tailor optical response at will, it is in principle possible to perform a mapping of the optical space from one state to another, thus obtaining a desired electromagnetic behavior. Some of the proposed applications are invisibility cloaks [214, 215], superconcentrators [216], hyperlenses [217], etc.

Since the properties of metamaterials and metamaterial-containing structures can be engineered almost at will, an obvious question that occurs is if it is possible to use them as superior impedance matching structures. The problem reduces to the use of metamaterials as impedance matching structures between the surrounding medium and the material of the detector-active region.

2.7.1 Metamaterial-Containing Antireflection Multilayers

The simplest way to use metamaterials as antireflection structures is to make a 1D multilayer in which isotropic layers of negative index metamaterial and lossless dielectric are alternating. Such structures were extensively researched [218–222]. Metamaterial-containing multilayers are characterized by many peculiar properties with both theoretical and practical interest. Using such structures it is possible to fabricate resonant cavities with subwavelength dimensions [223].

Compared to conventional Bragg-type quarterwave-stack interference multilayers that exhibit a single stop-band (1D photonic band gap) where propagation of electromagnetic waves is forbidden, metamaterial-containing multilayers have three additional kinds of stop bands besides such Bragg stop-band. One of the novel stop bands appearing in alternating metamaterial-dielectric stacks is related to zero-value of refractive index averaged over volume of the multilayer and is denoted as zero-index band gap [220]. Another one is in a similar manner related with zero permittivity and the third one with zero permeability bandgap.

The new band gaps appearing in metamaterial-containing stacks only are insensitive to change of scale. They are also insensitive to disorder, and remain even in quasiperiodic structures [224, 225]. Even more important, it has been shown that metamaterial-containing multilayers do not change their properties with a variation of incident angles, i.e., they exhibit omnidirectional behavior. Most of such metamaterial properties are consequences of phase compensation [226]. Namely, the metamaterial slab fully cancels the phase of a wave arriving from the dielectric part, so that each pair of strata can be made completely transparent. Due to the same effect, higher and wider bandgaps appear in such structures than in conventional quarterwave stacks.

It has been suggested that multilayer structures comprising of alternating dielectric and negative index metamaterial could be used as a novel kind of antireflection coatings [227]. Compared with conventional Bragg-type all-dielectric interference films, these structures are much less sensitive to the incident angle, which could improve the characteristics of antireflective coatings [227].

A problem with such use of metamaterials are large absorption losses in the range of negative refractive index and relatively narrow bandwidths due to their resonant nature. Technologically, the fabrication of metamaterial structures is a nontrivial task. For the visible and infrared range it is typically done by top-down approach (planar technologies and dry etching), but the necessary resolutions are deeply subwavelength. A multilayer metamaterial based on the so-called double-fishnet structure has been reported by Valentine et al. [228].

2.7.2 Metamaterial Superabsorbers

Plasmonic metamaterials can be utilized to match the impedance of environment (air) to the substrate (semiconductor material) in the optical range (or actually any other range) directly by optimizing their "particles" design. Thus one can obtain "perfect black" surfaces denoted as superabsorbers [229, 230]. Such nanostructured surfaces are able to absorb light in a wide bandwidth regardless of its polarization and for literally any incident angle.

Basically, all superabsorbers proposed until now were metal-dielectric–metal sandwiches where the top metal layer of the sandwich is nanostructured to obtain a desired wave impedance.

Figure 2.31 shows two typical structures for superabsorbers. The bottom one represents an ordered array of plasmonic nanoparticles on dielectric. The shape of the nanoparticle array may be rectangular, circular, etc. [230] The top structure represent a 2D array of nanoholes in the top metal sheet [231]. In both cases dielectric is located on a continuous metal sheet.

Another structure represents 1D trapezoidal or crossed 2D arrays (Fig. 2.32). The basic form of this super absorber is again metal–dielectric–metal [232].

As far as photodetectors are concerned, superabsorber structures were described within the context of photovoltaic solar cell enhancement [232]. However, other wavelength ranges can be covered by changing the dimensions of the elements of the superabsorber.

Technologically, the metal-dielectric-metal sheet with satisfactory properties can be fabricated utilizing conventional radiofrequent sputtering. Nanostructuring of the

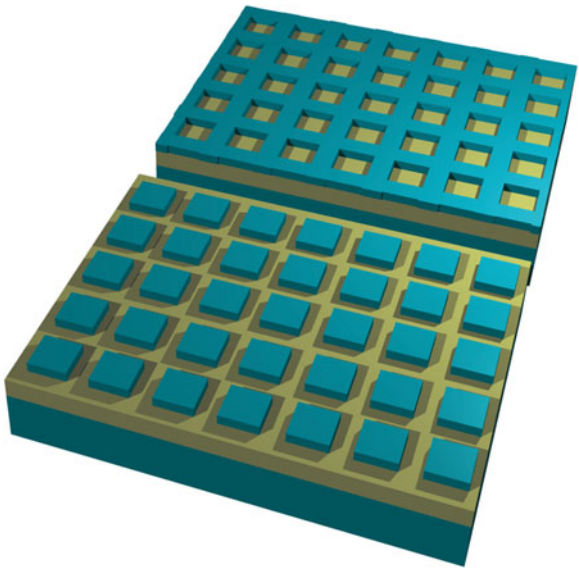


Fig. 2.31 Plasmonic metamaterial-based super absorbers. *Bottom* Metal nanoparticles on dielectric; *top* Nanohole array in metal layer on dielectric. *Bottom* layer is metal

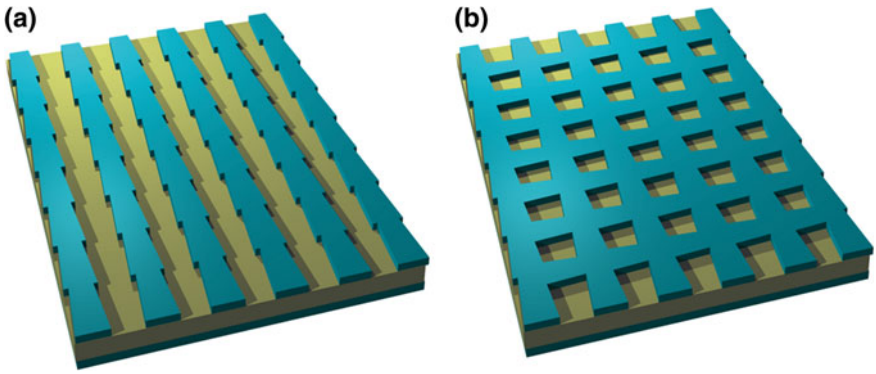


Fig. 2.32 Plasmonic super absorber based on trapezoidal plasmonic elements. **a** 1D array; **b** 2D array. Metal layers are on the *top* and the *bottom*, divided by dielectric

top surface can be done utilizing photolithography and electron-beam or focused ion beam methods [230]. In addition to that, structures of the type shown in Fig. 2.31 bottom may be fabricated utilizing functionalization by plasmonic nanoparticles [233]. In that case disordered nanoparticle arrays are obtained, but the effective properties of the structures remain similar to the ordered ones for similar values of the filling factor *f*.

Alici et al. described a metamaterial absorber based on a 2D array of single split ring resonators [234, 235]. The structure was designed for the infrared range. Sun et al. designed a metamaterial absorber with wide operating range by using several layers, each with split ring resonators with different dimensions [236].

2.8 Light Trapping by Optical Path Increase

A possible way to increase the absorption probability while at the same time keeping the detector optical thickness αd constant is to ensure a longer path of light beam through the detector. The idea is to trap the light that has already entered the active region and keep it as long as possible. The most general way to implement such light trapping is to influence the path of the beam, i.e., modify the optical space in some manner so that the radiation continues to propagate within the detector. This could be done in several ways. The beam could be returned into the detector from its surface, immediately before leaving the structure. It could be made to scatter and to repeatedly change its path within the structure. It could be also concentrated within the structure itself, by making some kind of internal hotspots through the influence to the local density of the optical space.

The simplest and the most obvious approach is to utilize some kind of high-reflection structure to change the path of the beam and return it back to the active

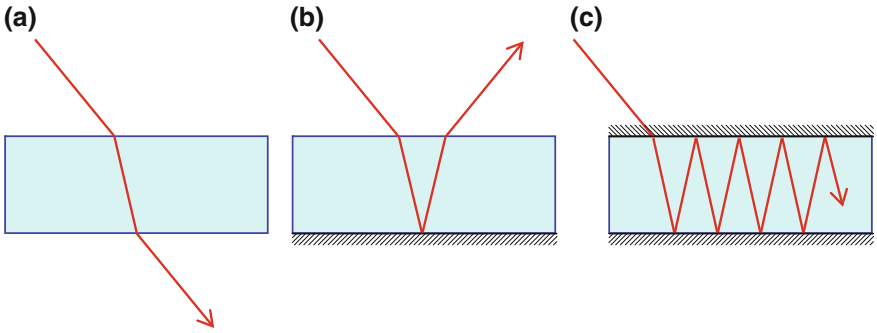


Fig. 2.33 Idealized representation of optical path increase in photodetector using high-reflection surfaces. **a** No high-reflection surfaces are used: light refracts and leaves the detector after a single pass; **b** backside reflector used: light passes the active region twice; **c** cavity enhancement is used: light continues to reflect multiply

region from the exit path. Figure 2.33 shows an idealized representation of optical path through a thin semiconductor detector without light trapping and with two different schemes utilizing high-reflection surfaces.

Figure 2.33a shows a photodetector before any light trapping scheme is implemented. It is assumed that an antireflection coating has been applied to the detector surfaces, ensuring almost complete transmission of the incident beam into the active region of the photodetector. In a real structure, a part of the radiation would be reflected both on the front and on the back surface due to the refractive index mismatch between the surrounding medium (low index) and semiconductor material (high index), and no AR coating would ensure 100 % transmission. Obviously, a part of the beam intensity would be lost due to the absorption within the structure, the effect we desire to maximize. The beam refracts at the front and at the back surface and is immediately lost after a single pass.

A different situation is met if a high-reflection layer is deposited on the detector backside, as shown in Fig. 2.33b. The signal is completely reflected from it and passes a twice longer path through the active region before leaving the structure. This solution is very often implemented and represents almost a rule for practically all conventional thin photodetectors. In real situations, there are again multiple reflections due to refractive index mismatch. A calculation that takes into account these reflections has been presented in [11].

The efficiency can be further improved if there are high-reflection surface both on the front and the backside of the detector, Fig. 2.33c. In other words, the detector is placed into a cavity with highly reflective walls. Light will enter into such structure resonantly or through some kind of light pipe, as described later in this text. Upon entering, the beam will reflect multiply and will be kept within the active region much longer. Large enhancements ensue as a result. This is the basis of the cavity enhancement of detectors. The cavity itself may support one or a larger number of modes and may be implemented in different ways. Examples of cavity

enhancement include resonant cavity enhancement (RCE) and photonic crystal enhancement (PCE) [237].

The next section analyzes different high-reflection structures convenient for cavity enhancement. It also considers strategies and conditions of their use.

2.9 High-Reflection Structures

2.9.1 Introduction

A reflective surface can be specular (mirror-like, glossy) or diffuse, Fig. 2.34. The retroreflective surfaces return the beam in the direction the incident light came from. Regardless of the type of reflectivity, most of the radiation is returned from the illuminated surface toward the radiation source. This radiation can be practically unchanged or can be shifted in frequency spectrum, either toward higher or toward lower frequencies ("hot" and "cold" mirrors).

The basic parameter of a reflective surface is its reflection coefficient, R, describing the percent of the incident radiation returned from the reflective surface. The largest values of R are offered by multilayer dielectric films. Reflection coefficient is usually dependent on light wavelength, and there is a spectral range of reflection maximum for a given type of material. Typically, metals have a very wide spectral range of high R, while for dielectric mirrors this range depends on the properties of a particular structure and is typically much narrower. On the other hand, dielectric mirrors offer much higher values of the reflection coefficient.

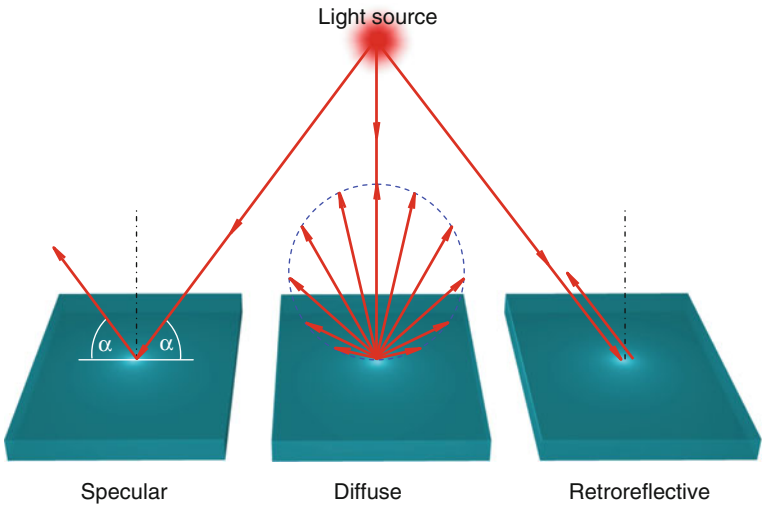


Fig. 2.34 Different types of reflectivity: specular (mirror-like), diffuse, and retroreflective

Another issue of interest is the angular dependence of reflection coefficient. For some structures, this dependence is very strong and large values of R are obtained only for near-normal incidence, thus furnishing surfaces with very high directivity of reflected light. The other pole are the surfaces offering near-omnidirectional or even fully omnidirectional reflectivity.

A surface can have significant diffuse reflectance, dependent on the scattering properties of the material. There is a range of possible intermediate situations between a fully specular and fully diffused optical reflector surface.

In this text, we handle different types of high-reflectance structures that could be of use for photodetector enhancement through optical path increase, starting from the simplest and well-known ones like metal mirrors and continuing to more complex ones like interference mirrors and ending with structures like photonic crystals. The function of dielectric mirrors and photonic crystals is based on light wave interaction effects (interference and diffraction). Both utilize nanoscale-patterned structures and the effects of multiple reflection on structures with periodic changes of refractive index. They are part of a larger field of the so-called "designer" nanostructured materials that represent the current limits of materials engineering.

2.9.2 Specular Reflectivity

The specular type of reflectivity means that scattering on a given optical surface is low or zero. The angle of incidence with respect to the target surface is equal to the angle of reflection (Fig. 2.34). The incident beam and the reflected beam lie in the plane defined by the incident beam and the normal to the surface. In other words, the surface is mirror-like. Only the macroscopic curvature of the surface may influence the shape of a reflected optical image.

Specularly reflective surface can be used in image forming. It should be mentioned that this is not of prime importance for photodetectors, since no image is formed in a single photodetector element, and the only important factor is the value of reflection coefficient for different wavelengths and incidence angles.

The conventional solution to achieve specular reflectance is to use flat metal surfaces. Other solutions are interference-based multilayer dielectric reflectors (Bragg mirrors) and, as their generalization, photonic bandgap structures (photonic crystals) of all-dielectric and metal-dielectric type, etc. Nanoscale interferometric and diffractive structures offer extremely large values of reflection coefficient (in excess of 99.99 %).

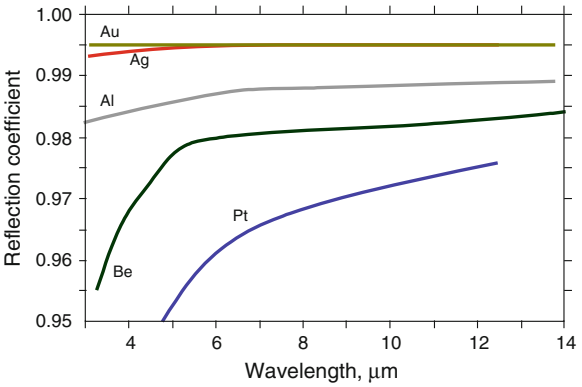


Fig. 2.35 Spectral dependence of reflection coefficient of some metals used for reflective layers in IR detectors

2.9.3 Metal Mirrors

The spectral range of thin metal mirrors is typically very wide. Also, they are relatively insensitive to the changes of the incident angle. Because of their low cost and relatively high-reflection coefficient they are often used for the deposition to the backside of the detector. Since thin metal mirrors quickly become covered by an oxide layer spoiling their reflection, in some applications additional dielectric protective coatings are deposited on them, for instance magnesium fluoride or silicon monoxide. This is usually not necessary when they are used for the improvement of photodetector quantum efficiency.

The most often used materials for the infrared range are aluminum and gold [5]. There is a wealth of literature data on these materials in the IR range. As an example, Fig. 2.35 shows spectral characteristics of the reflection coefficient of some polished metals in the (3–14) μm range, according to [1, 238–240]. All dependences are valid for normal incidence and at room temperature.

One of the important parameters for the use of metal reflectors in MOEMS (microoptoelectromechanical systems) is the penetration depth (skin depth) of electromagnetic radiation at a given wavelength. In the (3–14) μm range, its value is typically between below 10 nm (aluminum) and 20 nm (nickel) [240].

2.9.4 Dielectric Interference Mirrors

Interference mirrors are dielectric thin film coatings where low- and high-refractive index layers alternate. The optical thickness of each of the layers is equal to quarter-wavelength ($\lambda/4n$). They are denoted as distributed Bragg mirrors or distributed Bragg reflectors (DBR), sometimes simply as Bragg mirrors. Other names include quarter-wave mirrors (QWM), quarterwave stacks (QWS) and highly reflective (HR) layers.

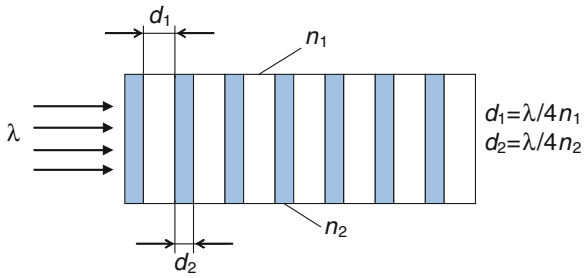


Fig. 2.36 Schematic presentation of a high-reflection quarterwave stack

Bragg mirrors are routinely fabricated by alternatively depositing quarter-wavelength stacks with high (n_h) and low (n_l) refractive index (approximately lossless dielectrics are assumed). Their principle of operation is similar to that of multilayer antireflection films, but the incident beam arrives at the layer with high-refractive index, and in this case the interference on the quarter-wavelength stacks is constructive.

The theory of dielectric mirrors belongs to the general theory of optical interference coatings [152, 163, 180]. The simplest type of specular multilayer reflectors are high-reflection (HR) dielectric quarterwave stacks. These are structures consisting of a number of micrometer- to nanometer-thick dielectric layers. The optical thickness of each separate layer of such a structure (the ratio of real thickness and refractive index) is equal to a quarter of the operating wavelength (thus the name "quarterwave stacks") and the value of refractive index is periodically alternating between high (n_2) and low (n_1) values (See Fig. 2.36).

After a light beam enters the stack, each interface between two adjacent layers will reflect a part of the incident radiation. All the emerging beams will be in phase and thus their intensities will add. In that way constructive interference will greatly enhance the reflection of the whole stack.

An approach that can be generally applied to various multilayer structures is the well-known transfer matrix technique [87, 180]. It can be used regardless of the geometry of the layers or the intended application of the multilayer. Besides being usable for the calculation of quarterwave Bragg mirrors, it can be used without modification to accurately determine the properties of antireflection coatings, step-down structures, various quasiperiodic, aperiodic and random structures, but also 2D and 3D photonic crystals as well [241]

In comparison to an arbitrary multilayer coating, a quarterwave 1D dielectric stack has additional symmetry which allows one to determine its transfer matrix analytically, thus significantly simplifying the procedure of calculation of its properties and optimization of the final structures.

To apply the transfer matrix method to a quarterwave stack the electric field in each separate homogeneous layer of the stratified Bragg mirror structure is written as a superposition of the incident ($E_j^{(i,l)}$) and reflected ($E_j^{(r,l)}$) plane wave, where j denotes the number of the pair (period) of the quarterwave mirror, $l = 1, 2$ is the number of the stratum within the given pair, r denotes the reflected wave, and

i incident. Assuming that a plane wave has Bloch (Floquet) periodicity and applying the condition of continuity of electric and magnetic fields and their first derivatives on the interfaces between strata, the connection between the j-th and (j–1)-th layer is [178–180]

$$\begin{bmatrix} E_j^{(i,l)} \\ E_j^{(r,l)} \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} E_{j-1}^{(i,l)} \\ E_{j-1}^{(r,l)} \end{bmatrix} = [P] \begin{bmatrix} E_{j-1}^{(i,l)} \\ E_{j-1}^{(r,l)} \end{bmatrix}, \tag{2.25}$$

where $[P]$ denotes the transfer matrix with the elements

$$\left.\begin{aligned} P_{11} &= e^{e\pm ik_1 d_1}\left[\cos k_2 d_2 \pm \frac{i}{2}\left(\frac{n_2}{n_1}+\frac{n_1}{n_2}\right)\sin\, k_2 d_2\right] \\ P_{12} &= e^{\pm ik_1 d_1}\left[\pm\frac{i}{2}\left(\frac{n_2}{n_1}-\frac{n_1}{n_2}\right)\sin\, k_2 d_2\right] \\ P_{21} &\\ P_{12} &\end{aligned}\right\} k_1 = \frac{2\pi}{\lambda} n_l, \quad l = 1, 2\ . \tag{2.26}$$

For an M layer structure the Chebyshev equality can be applied [87] to obtain

$$\begin{bmatrix} E_0^{(i,l)} \\ E_0^{(r,l)} \end{bmatrix} = \begin{bmatrix} P_{22}U_{M-1} - U_{M-2} & -P_{12}U_{M-1} \\ -P_{21}U_{M-1} & P_{11}U_{M-1} - U_{M-2} \end{bmatrix} \begin{bmatrix} E_M^{(i,l)} \\ E_M^{(r,l)} \end{bmatrix}, \tag{2.27}$$

where

$$U_M \frac{\sin[(M+1)\kappa d]}{\sin(\kappa d)} \tag{2.28}$$

is the Chebyshev polynomial of the second kind, and κ is the Bloch wavenumber

$$\kappa = \frac{1}{d}\arccos\left(\frac{P_{11}+P_{12}}{2}\right). \tag{2.29}$$

The Fresnel reflectance of the structure is [180]

$$r_M = \frac{P_{21}U_{M-1}}{P_{11}U_{M-1} - U_{M-2}}. \tag{2.30}$$

In the case when the refractive index of the incident medium n_a is not equal to n_1, we utilize [242]

$$r = \frac{\frac{|n_a - n_1|}{n_a + n_1} + r_M}{1 + \frac{|n_a - n_1|}{n_a + n_1}}. \tag{2.31}$$

The total reflection coefficient is calculated as the square of the absolute value of the Fresnel reflectance (2.31).

The reflection range of a quarterwave mirror is determined as

$$\frac{\Delta\omega}{\omega} = \frac{4}{\pi} \arcsin \frac{|n_2 - n_1|}{n_2 + n_1}. \tag{2.32}$$

The procedure of transfer matrix calculation of dielectric mirrors is described in detail in, e.g., [243]. An approach similar to the above is used for 2D and 3D structures, but the obtained expressions are much more involved and must be solved numerically. In [244] a detailed approach is described and the diagrams of photonic bandgap structures for 2D and 3D photonic crystals are given. An alternative method is presented in [241].

As an example of the application of transfer matrix method to the calculation of spectral reflection of quarterwave stacks, Fig. 2.37 shows the spectral reflection of a silicon-silica Bragg mirror for different numbers of layer pairs under normal incidence of light. A stop band is readily recognizable already for a small number of layer pairs. It is important that variations of the layer thickness and the materials used.

Reflectivity is greatly increased in a dielectric stack with a larger number of layers. A typical spectral dependence of a dielectric mirror is shown in Fig. 2.37. The wavelength range of maximum reflectivity is called the stop band (sometimes also called one-dimensional photonic band gap). Increasing the ratio between the low and high-refractive index parts extends the width of the stop band.

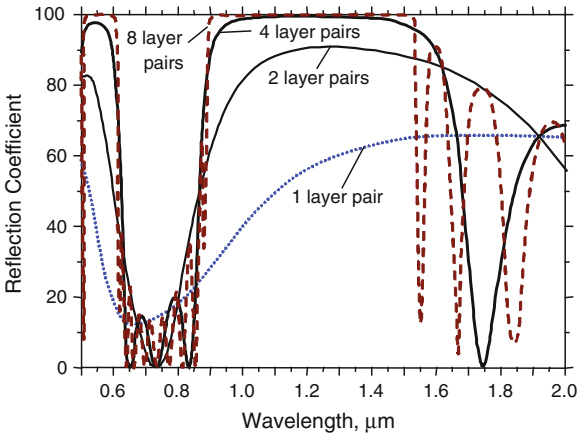


Fig. 2.37 Spectral reflectance of a HR dielectric stack for different numbers of layer pairs under normal incidence of light. Materials are Si ($\varepsilon = 11.958$, thickness 174 nm) and SiO_2 ($\varepsilon = 2.1316$, thickness 75 nm)

For some applications it is of interest to extend the high-reflectance zone (the stop band) and cover a larger portion of spectrum than in the case of quarter-wave stacks. There are several strategies to that purpose that deviate from the quarterwave stack concept. One of them is to stagger the thickness of each successive layer in some kind of progression (e.g., harmonic, arithmetic, geometric, etc.) [152]. Theoretically this could be also done by step-grading the refractive index from layer pair to layer pair, but the problem is the choice of materials. Chirped mirrors with Bragg wavelength varying between layer pairs in, e.g., linear fashion offer extended high-reflection range and at the same time enable control of the group delay [245, 246]. The stop bands in dielectric mirrors can be also extended by superimposing two or more quarterwave stacks designed for maximums in different wavelength ranges and to choose their ranges to be continuous. The proposed solutions include the use of dielectric stacks with gradient change of layer thickness and the use of layers with stochastically distributed thickness.

2.9.5 Omnidirectional Reflectors

The concept of omnidirectional mirrors, also denoted as perfect mirrors, was introduced in 1998 [247, 248]. These are one-dimensional, all-dielectric mirrors (1D photonic crystals) that keep their reflective properties even for large oblique incident angles in a wide and fully tailorable range of frequencies.

Basically, omnidirectional mirrors are dielectric quarterwave stacks in which TIR and Bragg mirroring wavelength ranges overlap, i.e., an overlapping band gap regime exists that extends above the light cone.

For a normally incident beam the stop bands for transverse magnetic (TM) and transverse electric (TE) polarization is identical. For increasingly oblique incidence the TM gap decreases, and the TE gap increases.

The width of the omnidirectional frequency range can be approximated for quarter-wave stacks as [249]

$$\frac{\Delta\omega}{2c} = \frac{(d_1 + d_2)\cos\left(-\sqrt{\frac{K_1-2}{K_1+2}}\right)}{d_1 n_1 + d_2 n_2} - \frac{(d_1 + d_2)\cos\left(-\sqrt{\frac{K_2-2}{K_2+2}}\right)}{d_1\sqrt{n_1^2 - 1} + d_2\sqrt{n_2^2 - 1}} \tag{2.33}$$

where c is the speed of light in the ambient, d_1 and n_1 are the thickness and the refractive index of the first layer in the pair, while d_2 and n_2 are the corresponding values for the second pair.

Figure 2.38 shows the spectral dependence of reflection coefficient for various incident angles for an Si/SiO_2 omnidirectional stack (dielectric permittivity 11.958/2.1316, thickness 174 and 75 nm, respectively, 8 layer pairs). The stop band changes with incident angle, becoming narrower and shifting toward shorter wavelengths with increasing incident angle. However, it does exist for all

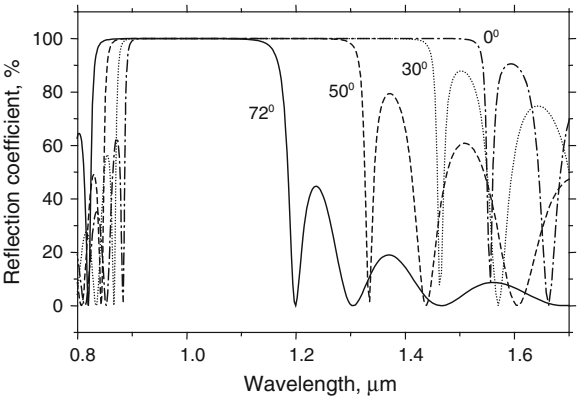


Fig. 2.38 Spectral reflectance of omnidirectional stack with 8 layer pairs for various incident angles, TM mode. Material pair is Si/SiO_2 (dielectric permittivity 11.958/2.1316), thickness 174 and 75 nm, respectively

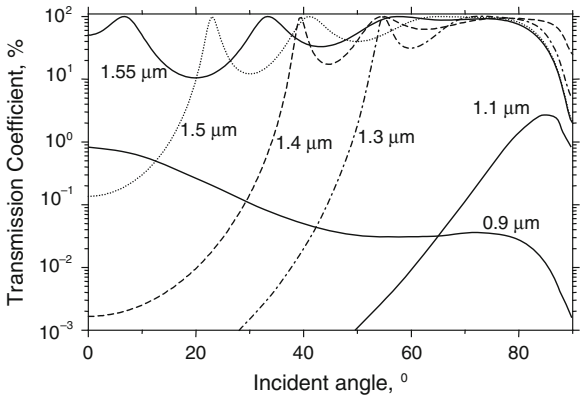


Fig. 2.39 Transmission coefficient of omnidirectional stacks with 8 layer pairs versus incident angle for different wavelengths, TM mode. Material pair is Si/SiO_2 (dielectric permittivity 11.958/2.1316), thickness 174 and 75 nm, respectively

considered incident angles, thus ensuring full functionality of the mirror for a wide range of oblique incidences.

Figure 2.39 shows the dependence of the transmission coefficient of omnidirectional 1D stacks on the incident angle for various wavelengths within the photonic band gap. The allowed angles decrease with the operating wavelength increase. For wavelengths outside the photonic band gap (1.55 μm) the transmission is high for all angles, remaining between 10 and 100 %. The lowest transmissions and highest reflections are obtained for the shortest operating wavelengths within the PBG.

The stop band (1D PBG) is fully tailorable to different wavelengths through changes of the materials used as high and low permittivity parts and through changes of the thickness of layers. Fink et al., who first described omnidirectional mirrors, performed experiments with polystyrene and tellurium and showed omnidirectional reflection in the LWIR wavelength range 10–15 μm [247]. Silicon and silica were used for the 3–5 μm MWIR range [250]. Numerous other material pairs are available for different wavelength ranges.

2.9.6 Metal-Dielectric Mirrors

Pure metal mirrors are omnidirectional reflectors, but the price to pay are their non-negligible absorption losses. On the other hand, all-dielectric multilayers exhibit very high reflectivity, but their reflectivity drops for oblique incident angles. Thus the idea occurs to combine both into a single structure.

A very old and obvious solution is to deposit a dielectric layer onto metal, thus reducing the absorptance and increasing reflectance. The transmittance remains very low or zero in such structures, but this fact is of no consequence for light trapping structures for photodetectors.

A generalization of this approach is to deposit a quarterwave dielectric mirror onto a metal reflector. In this way, very high reflectance is obtained for near-normal incidence. For more oblique incident angles the all-dielectric part will be transmissive, but in that case the metal part will reflect these beams, albeit with a lower reflectance. In this way the benefits of both types of the used structures are simultaneously employed. In [251] the authors mention the use of hybrid reflectors consisting of a Bragg-type dielectric mirror with an additional thin metal reflector deposited on its backside.

Another type of structures for metal-dielectric mirrors are metallodielectric multilayers. In this case alternating quarterwave or subwavelength stacks of metal and dielectric are deposited. Typical for such multilayers is a low reflection in visible, but large in infrared wavelength range. Thus they basically behave as low-pass optical filters. Such structures were denoted in literature as heat mirrors. First heat mirrors were fabricated as early as in 1950s [252]. The simplest heat mirrors consist of three layers only, dielectric-metal-dielectric or, alternatively, dielectric-transparent conductive oxide-dielectric [253]. Full multilayer metal-dielectric reflectors with binary but also ternary layers were also considered [254]. Because of their high reflectance in infrared, but also because of their plasmonic properties [255] metal-dielectric multilayer mirrors are of interest for cavity enhancement of infrared detectors.

A generalization of metal-dielectric photonic bandgap materials are metamaterial-containing multilayers. Due to the effect of phase compensation [226], higher and wider bandgaps appear in metamaterial-containing multilayer compared to the conventional all-dielectric and metal-dielectric quarterwave stacks. Such layers exhibit omnidirectional reflection [256, 257]. Like in the case of AR structures based on such multilayers, a problem are large absorption losses.

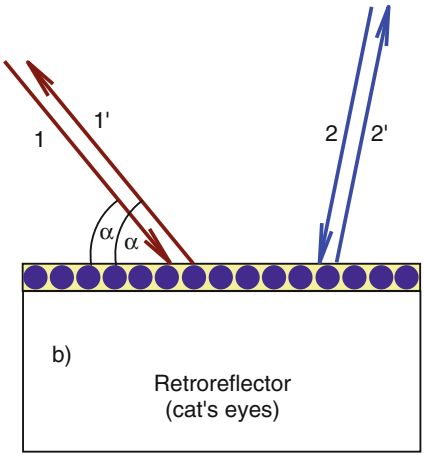


Fig. 2.40 Representation of a retroreflector

2.9.7 Retroreflectors

In certain types of mirror surfaces the reflected beam is returned in the same direction from which the incident beam arrived. This is maintained over a wide range of directions of the incident beam (Fig. 2.40). The phenomenon is called retroreflection [258]. Retroreflectors are various kinds of catadioptric structures ("cat's eyes"). Basically, a retroreflector may be a surface-diffractive structure with triangular, pyramidal, spherical, etc. shapes of the grating. Incident light typically reflects twice before returning in the direction it came from.

The patterns may also be comparable to the operating wavelength or larger, even much larger, and actually many retroreflector structures in everyday life are macroscopic. However, retroreflectors can also have subwavelength dimensions [259].

Instead of the conventional geometrical optics approach, one can use instead the transformation optics [243] which allows one to tailor the beam path at will. In this way one use the mapping of the optical space to arrive at optical components with vastly improved characteristics, in this case a retroreflector with omnidirectional operation [260].

2.9.8 Diffuse Mirrors

A light beam reflected from a diffuse surface does not remain in the same plane, but may be instead reflected omnidirectionally, i.e., in any spatial direction (e.g., [261]). Such behavior is ideally described by Lambert's cosine law.

Diffuse reflectivity is caused by scattering centers which may be either on the reflecting surface or within the volume of the reflecting body and are usually subwavelength or mesoscopic. Multiple scattering tends to produce a diffusely

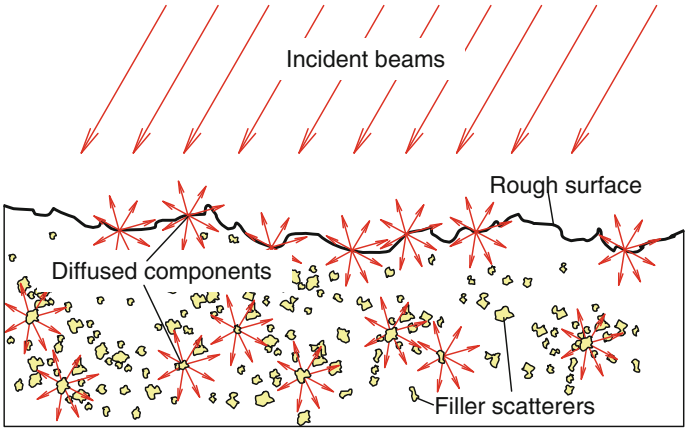


Fig. 2.41 Generalized 2-layer scattering surface

reflected beam emanating in many directions. Two main types of scattering centers are shown in Fig. 2.41.

The total reflectance of a surface is the result of a combination of multiple scattering from the volume inclusions and from the reflector surface. These two components may be in any ratio. Multiple scattering by pigment particles within the coating volume readily achieve diffuse reflectances of 0.80–0.90. The values are much higher if the structure is optimized for diffuse reflectance [262]. Such optimization can be obtained for example by numerical techniques.

One type of diffusing surfaces is a stochastically corrugated reflecting surface. Such structures can range from simple etched metal surfaces to more complex metal-multilayer dielectric stack systems. Surface corrugation is relatively the simplest approach to the fabrication of diffuse reflectors. Typically an optical surface is etched to obtain convenient roughness and then a high-reflection coating is deposited, or, alternatively, a mirror surface is directly roughened.

The other types of diffusing surfaces are various random distribution structures with enhanced scattering in the coating volume. Such structures are macroporous media with pores of the order of hundreds of nanometers (a high refractive index material with air spheres or other kinds of porosity, the "scattering sponges"), colloidal structures (particles of high-refractive index material or dye dispersed in a transparent medium), various powder complexes (sintered or processed otherwise) and generally disordered optical media. A property of all of the volume structures is that their refractive index is strongly—and stochastically or pseudostochastically—dependent on spatial coordinate. In all cases of strongly diffusing surfaces, regardless of their type, it is desirable that the dimensions of the disorder are comparable to the wavelength of light or smaller.

Enhanced backscattering can lead to near-ideal Lambertian illumination and "super-white" coatings, or "superdiffusers". Diffuse reflectors are described by the theory of disordered media. Generally, this theory is very complex and the methods

for its solution include Mie theory, but also various methods like Monte Carlo approach [263].

In structures where light is randomly scattered the scattering strength is characterized by the mean free path, defined as the distance light travels in average between two scatterings. If the mean free path is comparable to the operating wavelength or even smaller than it, coherent backscattering (weak localization) occurs, described as a precursor to Anderson localization (strong localization) [264]. This is similar to the behavior in a photonic crystal, where within its photonic band gap light propagation is also inhibited, but in this case it is caused by the structure disorder, not by its periodicity.

Light propagating through disordered media is multiply scattered, thus its exact description is very difficult. If all interference effects are neglected, which is an extremely rough approximation, then the light transport may be described by diffusion. Light propagates by randomly changing its direction and the characteristics of such transport is the mean free path and thus the diffusion constant. Internal reflections at the boundaries of such medium are also described by diffusion. The details of the calculations of transmission and reflection based on the diffusion approach can be found for example in [265].

Real reflectance can be described analytically or empirically [261]. There are several analytic reflectance models that can be used to describe various types of surfaces. The simplest diffuse source is Lambertian. The Mie theory can compute light scattering by spherical particles, as well as some other simple shapes like elongated ellipsoids [266].

2.10 TIR Structures

One of the approaches to the increase of the optical path through the active area is to use total internal reflection (TIR) and thus ensure multiple reflections from within the detector. To this purpose the active area surface is roughened (textured) in some way, for instance by etching. The obtained roughness may be stochastic, or alternatively some kind of regular pattern may be formed. The pattern shape ensures that the photon flux, after entering the detector and reflecting from the highly reflective backside, arrives from the inside to the corrugation surface at an angle larger than Brewster's. The goal is to obtain total internal reflection of as large percentage of the returning flux as possible. The result is that the absorption path within the active region may be increased several times.

The concept of TIR structures and that of antireflection DOEs are connected. In both of those similar or even identical surface reliefs are used. The difference is in their dimensions. While in the case of AR reliefs the texture constant remains subwavelength ($\Lambda < \lambda$ or in the worst case $\Lambda \approx \lambda$—in the case when $\Lambda \ll \lambda$ the effective medium approximation is valid), in the case of TIR reliefs $\Lambda > \lambda$ or $\Lambda \gg \lambda$—in this text we assume that TIR structures are those where interference effects

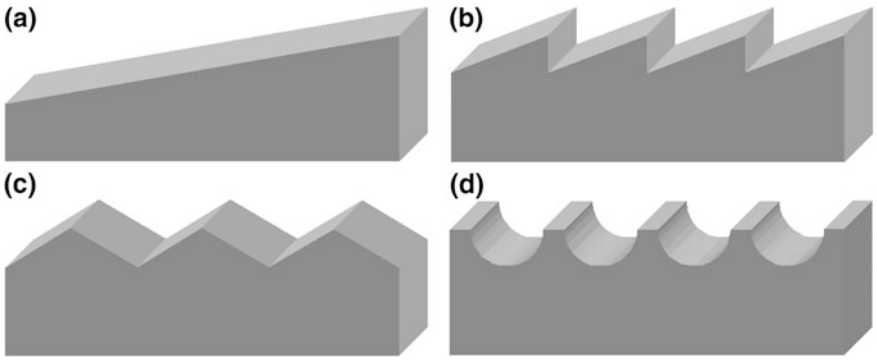


Fig. 2.42 1-D profiles of TIR structures for IR detectors; **a** wedge detector; **b** sawtooth structure; **c** V-groove (triangular structure); **d** hemispherical holes

can be neglected, or alternatively, those with features much larger than mesoscopic and subwavelength ones.

All nonbinary geometries listed in the text on AR subwavelength DOEs can be utilized for TIR reliefs as well: sawtooth geometries (triangles with one vertical side) [267], triangular (and their special case, V-grooves [268]), sinusoidal, hemispherical [269] and irregular corrugations (this includes stochastically distributed tetrahedrons with different dimensions [201, 270], etc. All of these geometries may be 1D or 2D. Figure 2.42 shows examples for some of the mentioned TIR geometries.

In 1970 Lee and Sze described a wedge-shaped TIR structure [271] to improve performance of Si diodes. This was a planar structure polished to obtain a sloped active area (i.e., a sawtooth detector with its period $\Lambda/2$ equal to the full length of detector) and thus receiving incident radiation at an oblique angle. One of the solutions to follow was sawtooth geometry, and another subdividing the structure to segments to reduce it to planar geometry [272] (similar to the connection between a conventional refractive lens and the corresponding Fresnel lens).

V-groove TIR structures (or, as a variation, a truncated—trapezoidal shape) and tetrahedral ones were successfully utilized for solar cells [200, 267].

Several different microfabrication techniques are used to form TIR structures. One of these is photolithography combined with bulk micromachining [196], also ion etching and mechanical polishing [271]. An application of isotropic etching is presented in [269]. TIR structures may be fabricated on one side of the detector only or on both of its sides and may be combined with, e.g., AR films/structures.

The calculation of quantum efficiency improvement by TIR is done by standard methods of geometrical optics (ray tracing) [268]. The wedge-shaped detector was calculated in [271], while triangular and sawtooth structures were analyzed in [268]. A typical approach is to divide the optical surface into a network of points where reflection and refraction of different beams is determined.

2.11 Resonant Cavity Enhancement (RCE)

Absorption within detector can be significantly increased by placing the active area into a resonant cavity, e.g., in a Fabry-Perot resonator. If the distance between resonator mirrors is equal to a small number of half wavelengths and if the product of the effective absorption coefficient and the active area thickness is small enough, the cavity will maintain a single allowed optical mode, while the other modes will be suppressed. In that manner interference causes a large change of optical mode distribution in comparison to the situation in conventional detector structures, including a change of spontaneous emission [273]. Detectors in a 1D optical resonant cavity are denoted as Resonant Cavity Enhanced (RCE) [251] or REsonant CAvity Photodetector (RECAP) [274].

The amplification of the radiation at the resonant wavelength in an RCE device and rejection of others results in a increased spectral selectivity of resonant detectors. A logical consequence is an increase of the detector speed [275] which becomes limited by carrier transit time. Another consequence is an increase of the allowed operating temperature because of the shift of the BLIP limit [276]. The high spectral selectivity of RCE detectors may prove advantageous or disadvantageous, depending on the concrete application.

Interference structures with multiple reflection applied in photodetection were analyzed as early as in 1970s [277]. At the end of 1980s and at the beginning of 1990s the interest for these detectors suddenly increased. It was the result of the requirements for faster photodetectors with narrow spectral range in optical telecommunication, i.e., with minimal crosstalk in wavelength division demultiplexers (WDM). This coincided with the availability of the convenient fabrication technologies [243].

RCE structures were used to improve performance of different optoelectronic devices—Schottky diodes [278, 279], p-i-n devices [243], avalanche photodiodes [280], MSM detectors [281], phototransistors [282, 283], LED diodes [284], etc. In the infrared range RCE was used for p-i-n diodes [285], photoconductors [286] and LED diodes [287].

Figure 2.43 shows a typical RCE photodetector structure.

The absorptive active region (for MWIR and LWIR devices typically fabricated from a narrow-bandgap material, its thickness d and the absorption coefficient α) is positioned between two layers of wider-bandgap material (thickness L_1 and L_2 and the absorption coefficient of both layers α_{ex}). DBR mirrors are on the top and the bottom side. In some cases the top Bragg mirror does not exist, and its role is assumed by the interface between semiconductor and the incident medium (air), which furnishes a top surface reflection coefficient of about 30 % [251].

The consideration of an RCE photodetector reduces to an analysis of a 1D microcavity with losses (absorption used for the detection). The interfaces between the DBR mirrors and the wide-bandgap layers are described by the Fresnel complex reflectances $r_1 = |r_1|\exp(i\psi_1)$ and $r_2 = |r_2|\exp(i\psi_2)$.

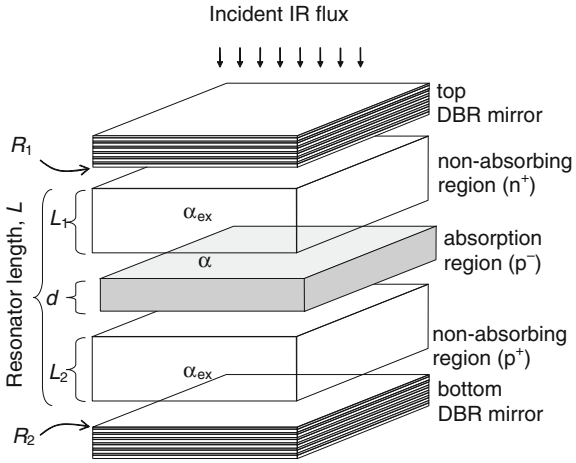


Fig. 2.43 Schematic presentation of an RCE photodetector

If the refractive index of the active region is denoted by n and the wave propagation constant is $\beta = 2n\pi/\lambda$, the quantum efficiency of the photodetector can be obtained as the ratio of the absorbed optical power to the incident power [251], taking into account the interference effects and neglecting the spatial distribution of the optical field within the RCE structure

$$\eta = \left\{ \frac{e^{-\alpha_{ex}L_1} + e^{-\alpha_{ex}L_2}|r_2|^2 e^{-\alpha_{sve}L}}{1 - 2r_1 r_2 e^{-\alpha_{sve}L}\cos(2\beta L + \psi_1 + \psi_2) + |r_1|^2 r_2|^2 e^{-2\alpha_{sve}L}} \right\} \left(1 - |r_1|^2\right)\left(1 - e^{-\alpha d}\right), \tag{2.34}$$

i.e.,

$$\eta = A\left(1 - |r_1|^2\right)\left(1 - e^{-\alpha d}\right). \tag{2.35}$$

Here $\alpha_{sve} = [\alpha_{ex}(L_1 + L_2) + \alpha d]/L$. The factor A describes the enhancement of quantum efficiency by interference effects. The peaks of quantum efficiency are obtained when the term in the cosine is equal to an integer number of 2π. The wavelength interval between two peaks is denoted as free spectral range.

The spatial distribution within a resonant cavity can be calculated by introducing the standing wave (SW) factor with a position-dependent value. In that case the absorption coefficient within the active area can be replaced by its effective value [251]

$$\alpha_{eff} = SW\alpha = \frac{\frac{1}{d}\int_0^d \alpha(Z)|E(z,\lambda)|^2 dz}{\frac{2}{\lambda}\int_0^{\lambda/2} |E(z)|^2 dz}. \tag{2.36}$$

If an assumption is made that the product of the absorption and the layer thickness is small, that the absorption coefficient and the refractive index are constant and that the reflection at the boundary of the active layer can be neglected, the following approximation is valid

$$|E|^2 = \frac{\left|1 - r_1^2\right|}{\left|1 - r_1 r_2 e^{i(2\beta L + \psi_1 + \psi_2)}\right|^2} \left[1 + r_2^2 + 2r_2 \cos(2\beta(L - z) + \psi_2)\right] |E_{\text{dir}}|^2 \quad (2.37)$$

where E_{dir} is the electric field corresponding to the incident wave. Then the standing wave coefficient is

$$\text{SW} = 1 + \frac{2r_2}{\beta d\left(1 + r_2^2\right)} [\sin\ \beta d\ \cos(2\beta L_2 + \beta d + \psi_2)] \quad (2.38)$$

As an illustration, we consider here resonant cavity enhancement of a HgCdTe detector at 10.6 μm. If we choose Ge–PbF_2 material pair for the RCE Bragg mirrors (Ge thickness 0.66 μm, PbF_2 thickness 1.63 μm), the structure shown in Fig. 2.12 can be designed in the following manner: the top mirror is PbF_2–Ge–PbF_2–Ge–PbF_2, followed by a Ge defect with a thickness of 2.12 μm. The nonabsorbing region is CdTe 0.38 μm, followed by the $Hg_{1-x}Cd_xTe$ active region. The bottom side nonabsorbing region is CdTe 0.76 μm, and the bottom mirror is four Ge–PbF_2 pairs followed by a single Ge layer.

The cadmium molar fraction of the photodetector was $x = 0.165$, which is the optimum for the operation at 300 K. The detector-active area was assumed to be 1 mm x 1 mm, one of the standard dimensions for uncooled HgCdTe photoconductors for the detection of CO_2 laser radiation. The bias voltage was 1.5 V, and the detector resistance 50 Ω.

Figure 2.44 shows the calculated reflection coefficient from the incident medium to the active area. R_1 denotes the top (front) surface, and R_2 bottom, as seen from the active area.

Figure 2.45 shows the dependence of quantum efficiency on the active region thickness. The same diagram also shows this dependence for a conventional HgCdTe photodetector ($R_1 = 0$). It was assumed that the detector had a backside Bragg mirror with its parameters identical to the one used in the RCE case. The front surface was protected with a PbF_2 antireflection coating. The diagram also shows the dependence of the similar structure with taking into account interference effects, i.e., an unoptimized RCE structure.

The first maximum in the RCE dependence in Fig. 2.45 is found for a thickness of 0.184 μm, and its value is 98 %. The spectral dependence of this quantum efficiency is shown in Fig. 2.46.

Figure 2.47 shows the changes of the quantum efficiency with varying thickness of the active region. This "sensitivity" of detector parameters to the variations of the thickness of the active region is important from the technological point of view, since a more "spread" curve means more relaxed conditions for the fabrication of

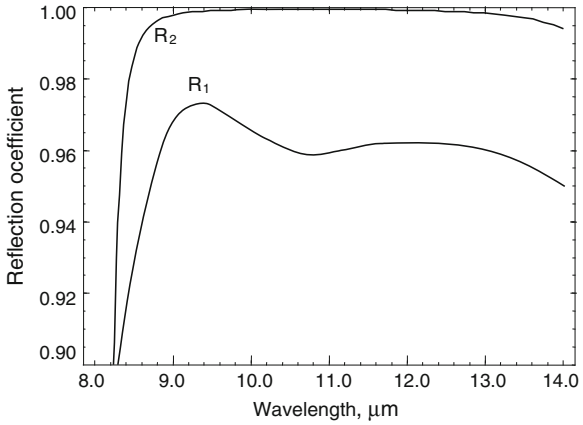


Fig. 2.44 Reflection coefficients of top and bottom DBR layers of an RCE structure (as seen from the active region)

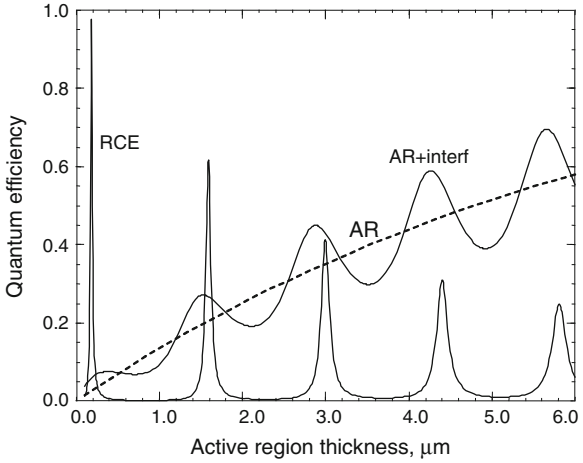


Fig. 2.45 Quantum efficiency of HgCdTe RCE photodetector designed for the LWIR wavelength of 10.6 μm. *RCE* resonant structure; *AR* (*dashed*) conventional detector; *AR + interf* detector with an antireflection layer taking into account interference (unoptimized RCE)

the RCE structure. It can be seen that the thickness changes of about 3 nm result in a quantum efficiency drop of near 10 %, while an error of 6 nm results in a 20 % drop.

Figure 2.48 shows spectral responsivity and detectivity of the HgCdTe RCE detector with an active area thickness of 0.184 μm.

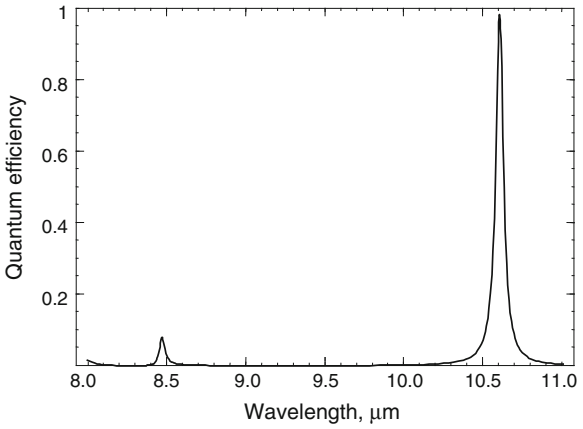


Fig. 2.46 Quantum efficiency for an HgCdTe RCE detector 0.184 μm thick

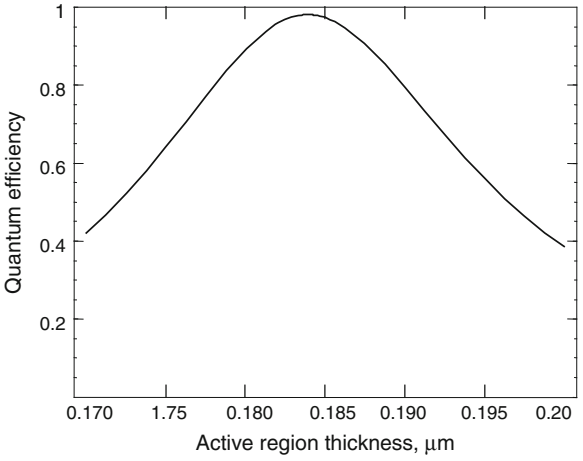


Fig. 2.47 Quantum efficiency of a HgCdTe RCE photodetector versus active region thickness

The maximum current responsivity (at 10.6 μm) of the 0.184 μm thick structure is 0.129 A/W, and the corresponding specific detectivity 1.94×10^9 cmHz$^{1/2}$/W. For a comparison, a conventional structure reaches the same responsivity for an active region thickness of 27.4 μm, and an unoptimized RCE for 14 μm. Also for a comparison, the best conventional uncooled detectors with d = 5 μm [8] reach $\eta \approx 0.60$.

A conventional uncooled detector (no interference, but the same Cd molar fraction, the same back mirror and an AR coating) with a thickness identical to the RCE (i.e., 0.184 μm) reaches $R_I = 0.034$ A/W and thus $D^* = 5.2 \times 10^7$ cmHz$^{1/2}$/W, i.e., an almost two orders of magnitude lower value. A conventional detector with a current responsivity identical to that of the RCE (d = 27.4 μm) reaches $D^* = 1.59 \times 10^8$ cmHz$^{1/2}$/W. This means that the specific detectivity is more than 12 times higher if all other parameters are identical.

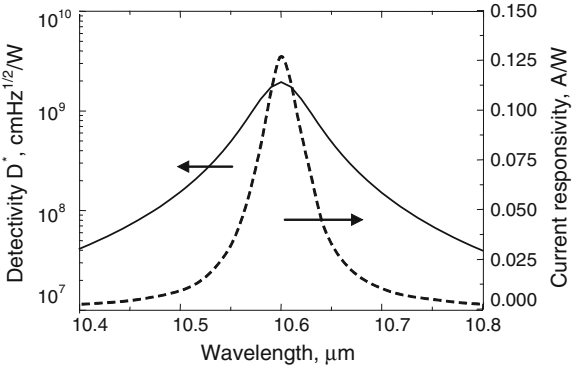


Fig. 2.48 Spectral current responsivity and specific detectivity of a HgCdTe RCE detector optimized for 10.6 μm

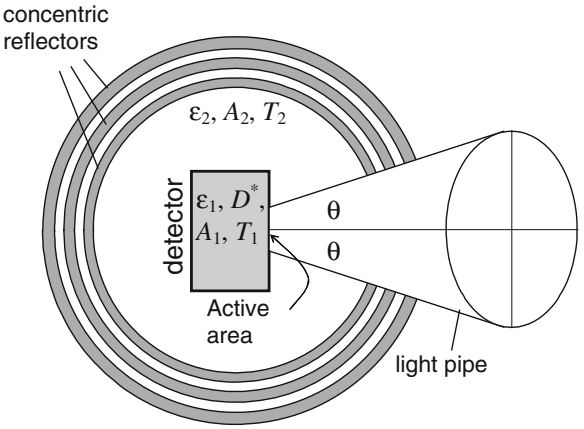


Fig. 2.49 Performance improvement by radiative shield

2.12 Radiative Shields

Radiative shields are screens with low emissivity (high-reflection coefficient) partially or completely enclosing the detector. In 1981 Putley proposed this method to improve BLIP detectivity of an ideal uncooled thermal detector [6]. We consider here the use of a radiative shield as a macro- or microcavity that may be resonant or nonresonant. In the most general case it would be a full 3D photonic bandgap structure (the PBG coinciding or being wider than the electronic bandgap of the semiconductor material used). A defect mode/light pipe may be introduced into the radiative shield on the incident side.

Figure 2.49 shows a schematic presentation of a photodetector surrounded by a radiative shield. A light pipe is shown on the right side as a cone-shaped opening with a solid angle θ. Similar to the radiative shield itself, its walls are made of highly reflective material. The detector radiates as a blackbody source (temperature T_1) and at the same time its radiation must be in equilibrium with its surroundings.

If the emissivity of the radiative shield ε_2 is much lower than unity, the level of equilibrium radiation emitted by the detector must drop (suppression of spontaneous radiation) and thus its BLIP detectivity will increase. The ideal result would be obtained for an aperture of the light pipe close to zero, but in that case diffractive effect completely change the behavior of the whole system.

Putley proposed to use the so-called superisolators as the material of the radiative shield. These materials were developed for thermal isolation in cryogenic technique and their thermal impedance is larger than that of vacuum. These are actually concentric layers of isolators and highly reflective metal layers.

In our consideration we assume that the detector and the radiative shield are at the same temperature $T_1 = T_2 = T$), which is near or equal to the room temperature. The detectivity of a photoconductive BLIP detector is given by [5]

$$D^*_{\mathrm{PC}} = \frac{\lambda}{2\mathrm{hc}}\sqrt{\frac{1}{\Phi_\mathrm{b}}}. \tag{2.39}$$

while for a photovoltaic detector it is $\sqrt{2}$ times larger. Here Φ_b denotes the background flux which can be absorbed in a detector with a cutoff wavelength of λ_{co} and which is given by the Planck integral

$$\Phi_\mathrm{b} = \frac{1}{\varepsilon_1}\int_0^{\lambda_{\mathrm{co}}} \frac{2\pi c}{\lambda^4}\frac{d\lambda}{e^{\mathrm{hc}/\lambda k_b T} - 1}. \tag{2.40}$$

The factor $1/\varepsilon_1$ takes into account the detector emissivity.

The following expression is valid for the radiation flux of a detector fully surrounded by a radiative shield

$$\Phi_{\mathrm{shield}} = \Phi_\mathrm{b}\sqrt{\frac{\varepsilon'}{\varepsilon_1}}, \tag{2.41}$$

While deriving the above equation the angular dependence of emissivity was not taken into account. ε' denotes the equivalent emissivity of the shield. Its value is obtain by a thermodynamical consideration of heat exchange between the detector and the shield. For a diffusively reflective external surface

$$\varepsilon' = \left[\varepsilon_1^{-1} + (A_1/A_2)\left(\varepsilon_2^{-1} - 1\right)\right]^{-1} \tag{2.42}$$

and for a specular one

$$\varepsilon' = \left[\varepsilon_1^{-1} + \left(\varepsilon_2^{-1} - 1\right)\right]^{-1} \tag{2.43}$$

In both cases for a low ε_1 and a high ε_2 it will be $\varepsilon' \approx \varepsilon_2$ (we assume that $A_1 \approx A_2$, i.e., that the shield is directly deposited to the active area).

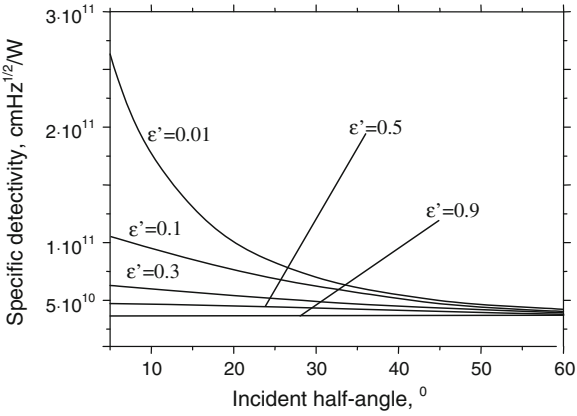


Fig. 2.50 Specific detectivity of photoconductor versus incident half-angle of light pipe for different shield emissivities. T = 297 K, ε_1 = 0.9, λ_{co} = 10.6 μm

If there is a light pipe, the above expression for the Φ_{shield} is modified my multiplying it with $\sin^2(\pi - \theta)$, while the flux arriving through the light pipe is given by $\Phi_{pipe} = \Phi_b \sin^2\theta$. The total flux is obtained by summing these two fluxes, which is equivalent to summing the noise equivalent powers for each flux separately

$$\Phi_{tot} = \left[\frac{\varepsilon'}{\varepsilon_1^2}\cos^2\theta - \frac{\sin^2\theta}{\varepsilon_1}\right]\int_0^{\lambda_{co}} \frac{2\pi c}{\lambda_1^4}\frac{d\lambda_1}{e^{hc/\lambda_1 k_b T} - 1}. \tag{2.44}$$

We assume that ε_1 and ε_2 do not depend on wavelength in the spectral range of interest.

Figure 2.50 shows the influence of the decrease of the emissivity of the radiative shield and of the decrease of the solid angle of the light pipe to the increase of the maximum specific detectivity of the detector. Figure 2.51 gives a comparison of a detector without a radiative shield, the one with the shield having a light pipe with a 15° angle (i.e., relatively large) and the one with a near-ideal case of radiative shield (light pipe angle tends to zero). Finally, Fig. 2.52 shows the specific detectivity of a photoconductor versus wavelength for a total enclosure within a radiative shield for various shield emissivities.

2.13 Photonic Crystal Enhancement (PCE)

Photonic crystals or photonic bandgap materials may be defined as artificial 1D, 2D, or 3D structures where the dielectric constant periodically changes, with a result that photons within it behave similarly to electrons in semiconductor crystal lattice [288, 289]. A photonic bandgap (PBG) appears in photonic crystals, an energy

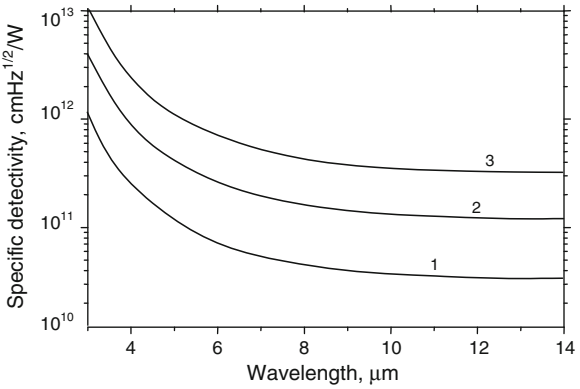


Fig. 2.51 Spectral dependence of specific detectivity of photoconductor for *1* conventional detector (no radiative shield); *2* radiative shield with a light pipe incident angle of $\theta = 15°$; *3* the case of ideal total radiative shield, $\theta = 0°$. $T = 297$ K, $\varepsilon_1 = 0.9$, $\varepsilon_2 = 0.01$, $\lambda_{co} = 10.6$ μm

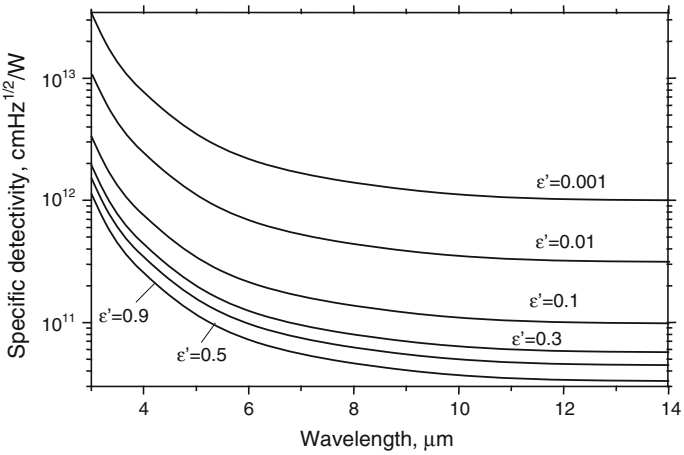


Fig. 2.52 Specific detectivity of a photoconductor versus wavelength for a total enclosure within a radiative shield for various shield emissivities. $T = 297$ K, $\varepsilon_1 = 0.9$

range in which electromagnetic waves cannot propagate through the structure. A photonic crystal behaves as a high reflection coefficient mirror for all wavelengths within the photonic bandgap, and the waves with such wavelengths are evanescent within it. In 3D structures this behavior is independent on the propagation direction (i.e., the radiation incident angle), while in 2D structures it has a strong angular dependence in a plane orthogonal to the propagation direction, and in 1D structures in two planes, which are at the same time orthogonal to the propagation direction and one to each other.

The fabrication of 3D photonic crystals for optical wavelengths is still a formidable technological problem, since their structure must have mesoscopic features (comparable to the operating wavelength radiation), and their fabrication accuracy must be at least an order of magnitude better. A number of methods have been proposed for their fabrication [289]. The problem is much simpler with 2D structures where it requires submicrometer photolithography, while in the case of 1D crystals it reduces to conventional thin film technologies.

An important property of photonic crystals are defect modes, i.e., the allowed energy modes within their photonic bandgap which appear when the symmetry of the photonic crystal is broken [290, 291]. They are analogous to the impurity levels in semiconductor lattices. The symmetry can be perturbed by increasing the dimensions of the high-refractive index material in the photonic unit cell (donor, introducing a mode originating from the bottom of the conduction band) or by removing a part of it (equivalent to adding the low-index material; this is acceptor, originating from the top of the photonic valence band). The result in both cases is a sharp peak at the defect wavelength. The defect mode wavelength can be adjusted by changing the properties of the defect (its thickness or the refractive index).

The use of photonic crystals for the enhancement of solar cells by utilizing PBG structure instead of the high-reflection Bragg mirrors has been reported by Bermel et al. [292]. They proved that a six-period triangular two-dimensional PBG structure made of air holes in silicon increases power generation more than 2 %, even more if a combination of a Bragg mirror and 2D PBG is used. A similar improvement is obtained if an eight-period inverted opal photonic crystal [293] is utilized.

The use of photonic crystal cavity to improve characteristics of infrared photodetectors (PCE, Photonic Crystal Enhancement) was described in [237]. The photodetector is enclosed within a PBG structure (3D in a general case) with a photonic bandgap either equivalent to the electronic bandgap of the utilized semiconductor material or larger than it. A defect may be introduced in the photonic crystal on the incident side of the photodetector, and its transmission peak adjusted to a desired wavelength. Such an approach improves the performance of the detector in two ways at the same time. As other methods of photodetector enhancement through optical path increase, it increases the probability of absorption of useful radiation in the active region and thus improves quantum efficiency and the responsivity. At the same time, reabsorption (photon recycling) increases the radiative lifetime and thus shifts the BLIP detectivity limit. The basic concept of the method is illustrated in Fig. 2.53. Although the presented photonic crystal is an inverted opal, any other type of PBG structure may be actually used instead.

Both RCE and PCE represent cavity enhancement methods. The main difference between them is that in the latter the cavity maintains more than a single mode, thus ensuring a range of operating wavelengths within the photonic bandgap.

The influence of a PBG structure to the detectivity of a photonic detector may be considered in a manner analogous to that presented in Sect. 2.12. Actually a photonic crystal may be considered the ideal case of a radiative shields, describing

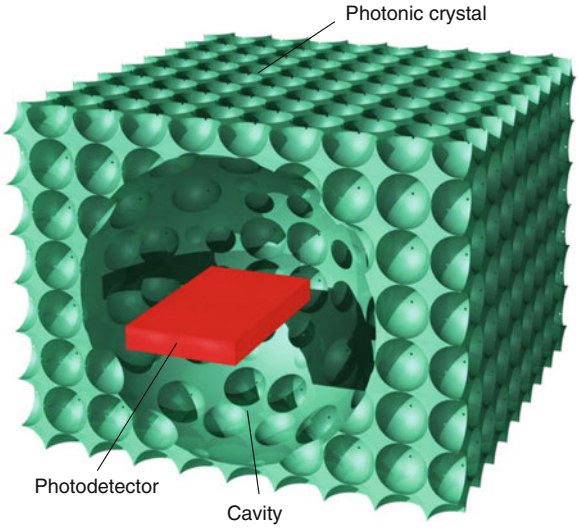


Fig. 2.53 Schematic presentation of cavity enhancement in a photodetector using photonic crystal. The *front half* of the PBG structure is removed to show the photodetector slab. The photonic crystal is an inverted opal, but this is for illustration purpose only since it can be any other type of PBG structure

the situation when the spatial angle of the light pipe is zero. The BLIP detectivity of an uncooled photoconductor calculated utilizing the approach from 9.3 is shown in Fig. 2.53.

In the further text, we shortly consider PCE utilizing a 1D photonic crystal with a defect. This situation is chosen for its simplicity. It can be generalized to 2D and 3D structures, but qualitatively the results remain similar.

In 1D case a defectless photonic crystal reduces to a Bragg distributed mirror (DBR). The defect modifies its spectral characteristics by introducing a transmission maximum dependent on the defect properties and deforms the 1D photonic bands to some extent.

Figure 2.54 shows a photodetector enclosed within a 1D PBG structure.

Using the Maxwell equations for the general case of propagation of electromagnetic waves through a one-dimensional medium, a wave equation can be written [179] with a form equivalent to the Schrödinger equation for electrons in semiconductor crystal in 1D case

$$\frac{\partial^2 E(x,t)}{\partial x^2} = \frac{\varepsilon(x)\partial^2 E(x,t)}{c^2 \quad \partial t^2}, \tag{2.45}$$

and if we assume that the electric field has the form of a plane wave $E(x,t) = E(x)\, e^{-i\omega t}$:

$$\frac{\partial^2 E(x)}{\partial x^2} = -\frac{\omega^2}{c^2}\varepsilon(x)E(x). \tag{2.46}$$

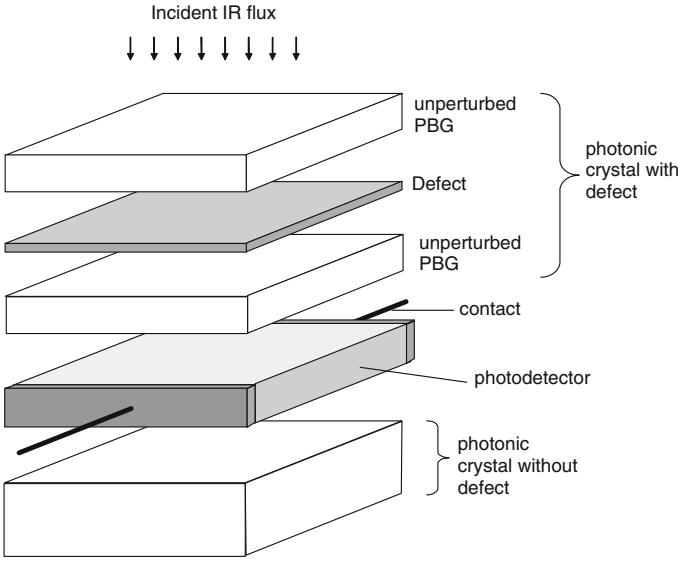


Fig. 2.54 Idealized presentation of a photodetector surrounded by 1D photonic crystals. The photonic bandgap of both of the crystals coincides with the spectral sensitivity range of the detector. Additionally, the photonic crystal at the incident side has a built-in defect at the desired operational wavelength

Here the dielectric permittivity is spatially periodic, $\varepsilon(x + d) = \varepsilon(x)$, which means that the Kronig-Penney model is applicable. If we introduce the Bloch vector, i.e., the crystal momentum κ, the periodicity of the electric field will be described by $E(x + d) = e^{i\mu d}E(x)$—Bloch or Floquet condition. The solution of the wave equation for an infinite 1D lattice with periodically changing dielectric permittivity should have the form of a sum of a direct and a reflected wave

$$E_i(x) = A_i e^{ik_i x} + B_i e^{-ik_i x}, \tag{2.47}$$

where $k_i = \omega\varepsilon_i^{1/2}$. While solving the above equation, we assume the continuity of the electric and the magnetic field on the boundary between layers. For the photonic bandgap of a 1D photonic crystal, we obtain the following transcendent equation

$$\cos(\kappa d) = \cos[k_1(d-a)]\cos(k_2 a) - \frac{1}{2}\left(\frac{k_1}{k_2} - \frac{k_2}{k_1}\right)\sin[k_1(d-a)]\sin(k_2 a), \tag{2.48}$$

This approach for the 1D case reduces to the consideration of a Bragg dielectric mirror.

In the case when there is a defect, the wavelength of its localized peak is obtained by solving the equation

$$tg[k_1(d-a)] = -\frac{tg(k_2 a)}{2}\left\{\left(\frac{k_1}{k_2}+\frac{k_2}{k_1}\right) \pm \left(\frac{k_2}{k_1}+\frac{k_1}{k_2}\right)\frac{\cos(k_1 d)}{\cos[k_1(d-a)]}\right\}. \quad (2.49)$$

The equation

$$\sigma\cos(\gamma d) = \cos[k_1(d-a)]\cos(k_2 a) - \frac{1}{2}\left(\frac{k_1}{k_2}+\frac{k_2}{k_1}\right)\sin[k_1(d-a)]\sin(k_2 a) \quad (2.50)$$

determines the symmetry (the sign factor $\sigma = \pm 1$) and the position ($1/\gamma$) of the defect modes.

Using (2.38) a photonic bandgap map can be obtained for 1D crystals [244]. Figure 2.55 shows two maps for 1D PBG structures. One of them is designed for the atmospheric window (3–5) μm an consists of a pair silicon–silica, while the other is intended for the (8–12) μm window and its material pair is lead fluoride and germanium.

The reflection coefficient for an unperturbed PBG structure chosen according to the map of photonic bandgaps is calculated by the transfer matrix method.

A defect is taken into account by writing the transfer matrix for the part of the PBG structure starting from the incident surface and ending with the layer next to the defect. According to the notation from Sect. 2.9.4, the degree of this transfer matrix is M/2, because usually the middle layer is replaced with a defect. Thus

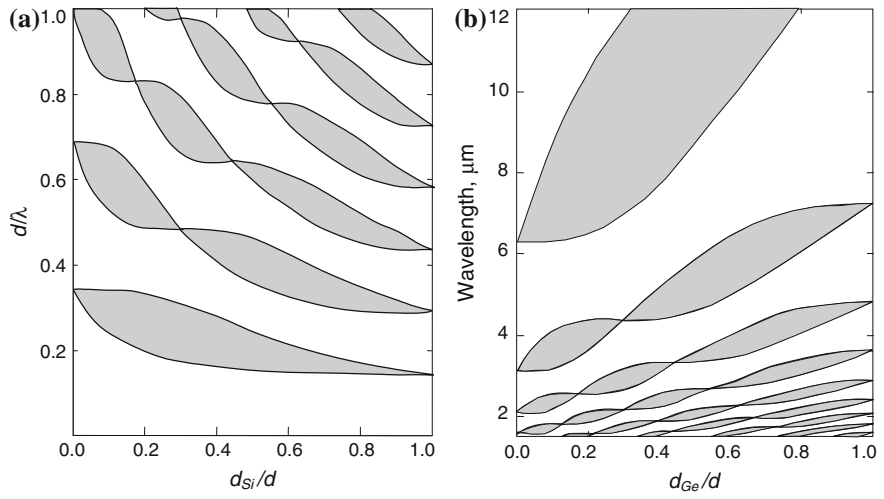


Fig. 2.55 Photonic bandgap maps for 1D photonic crystals for different material pairs. **a** silicon/silica; **b** germanium/lead fluoride. *Shaded areas* correspond to photonic bandgaps

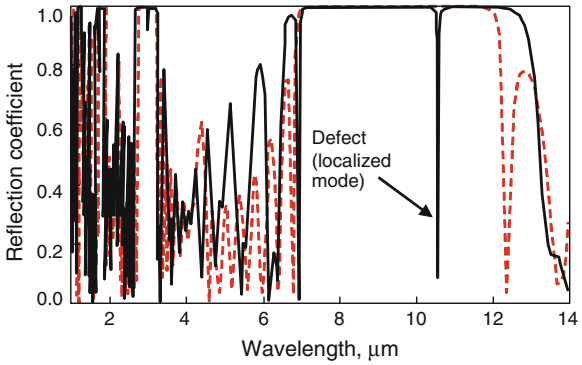


Fig. 2.56 Spectral dependence of the reflection coefficient for 1D Ge/PbF_2 photonic crystal with seven periods. *Dashed line* denotes a defectless structure, *solid line* is valid for a structure with a defect at 10.6 μm

obtained matrix is multiplied from the right with the matrix describing the defect layer and then again with a M/2 degree matrix which is identical to the first one. The reflection coefficient is further calculated in the previously described manner.

As an example we consider the design of a 1D photonic crystal for the improvement of the photodetector performance in the LWIR range. According to the map shown in Fig. 2.55b a convenient ratio of the thickness values of the constituents is $d_{Ge}/d = 0.31$, where $d = d_{PbF2} + d_{Ge} = 1.81$ μm, i.e., $d_{Ge} = 0.57$ μm and $d_{PbF2} = 1.24$ μm. For thus chosen PBG structure, the reflection and transmission coefficient are calculated. Then the thickness of the middle Ge layer is varied to determine the transmission peak wavelength (reflection minimum) versus the defect thickness. Using thus obtained value the thickness of the defect is determined which furnishes maximum transmission at the desired wavelength, in this case at 10.6 μm.

Figure 2.56 shows the calculated transmission of a photonic crystal with 7 Ge/PbF_2 layers. The defect at 10.6 μm is obtained by increasing the thickness of the middle layer from 0.57 to 1.545 μm. The transmission peak of the defect mode is narrower if the number of the layers around the defect is larger.

2.14 Radiative Lifetime Increase

It may be generally said that each technique for detector performance enhancement by way of the increase of optical path through the active region simultaneously increases radiative lifetime and thus increases the BLIP detectivity limit. The benefit for detector performance, however, can be seen only in those cases when radiative lifetime is the dominant one. This case is analyzed in this section.

Even the presence of a conventional high-reflection layer (metal, DBR or hybrid mirror) at the photodetector backside influences the reabsorption radiative lifetime

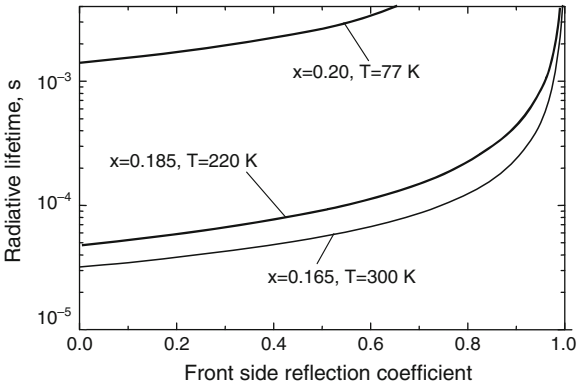


Fig. 2.57 Radiative lifetime versus front-side reflection coefficient for a $Hg_{1-x}Cd_xTe$ detector, thickness 5 μm, at 77 K (x = 0.20), 220 K (x = 0.185), and 300 K (x = 0.165)

[28, 29]. Any strict division between the methods for radiative lifetime increase and those for the optical path increase is impossible.

As an illustration, we consider the calculation of radiative lifetime of bulk $Hg_{1-x}Cd_xTe$ detector structures surrounded by Ge/PbF_2 photonic crystal. HgCdTe bandgap was determined according to [25], Fermi level according to [294], and HgCdTe refractive index according to [295].

Figure 2.57 shows radiative lifetime versus front-side reflection coefficient for an idealized $Hg_{1-x}Cd_xTe$ structure mounted on a highly reflective back surface $R_2 \approx 1$. Operating temperatures are 77 K (standard for BLIP photoconductors at 10.6 μm), 220 K (temperature achievable by multistage thermoelectric coolers) and, for comparison, 300 K. Cd molar fraction values are 0.205, 0.185, and 0.165, respectively, which corresponds to sensitivity maxima at 10.6 μm at each of these temperatures.

Figure 2.58 shows the dependence of radiative lifetime on front-side reflection coefficient for $x = 0.185$ and at $T = 220$ K for detector thickness 3, 5, 10, and 20 μm.

Radiative lifetime increases with a larger number of periods of photonic crystals and decreases with operating temperature. From the point of view of radiative

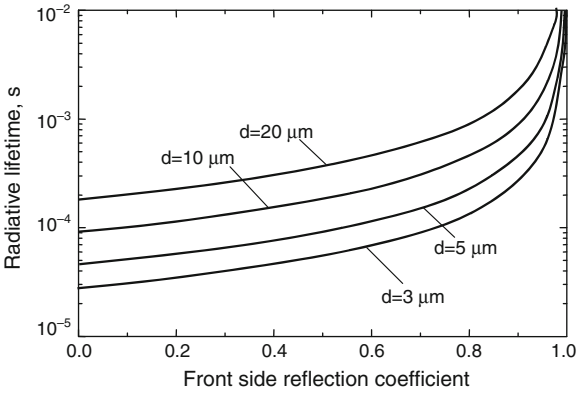


Fig. 2.58 Radiative lifetime versus front-side reflection coefficient for a $Hg_{1-x}Cd_xTe$ detector, thickness 3, 5, 10, and 20 μm at 220 K, composition x = 0.185

lifetime the immersion of the photodetector within a PBG structure is equivalent to a decrease of the operating temperature. The performance improvement is larger for a higher refractive index contrast within the photonic crystal and can be obtained in a larger spectral range.

2.15 Plasmonics for Light Trapping

A field of research that is currently exponentially growing is plasmonics [243]. It deals with the propagation of surface plasmons polaritons (SPP) through metal-dielectric structures (or generally, through composite materials consisting of materials with positive and with negative relative dielectric permittivity in the wavelength range of interest).

Surface plasmons polaritons are oscillations of free electrons in conductive material (the negative permittivity part of the composite) coherently coupled with electromagnetic radiation at an interface with dielectric. The ε-negative material is denoted as plasmonic material. Most often the ε-negative material is metal (typical examples are gold and silver). Other plasmonic metals include chromium, copper, alkali metals, various alloys, etc.). Alternative plasmonic materials are transparent conductive oxides (usable in infrared range), semiconductor materials (gallium arsenide is used in mid infrared), graphene, etc. [243].

A SPP is confined to the interface between the positive and the negative permittivity part and is evanescent away from the interface. SPPs can be propagating along the interface, or they can be nonpropagating, i.e., spatially confined to, e.g., a plasmonic nanoparticle (localized surface plasmons polaritons) [243].

Plasmonic nanocomposites can be ordered on 1D, 2D, or 3D level. They can be periodic, quasiperiodic [296], aperiodic [224], or random [204]. The building blocks of these functions themselves may have different shapes, from simple to complex and from regular to irregular [297], Fig. 2.59.

An important disadvantage of SPPs is their resonant nature, which causes a narrow bandwidth of operation. Another one is related to large absorption losses.

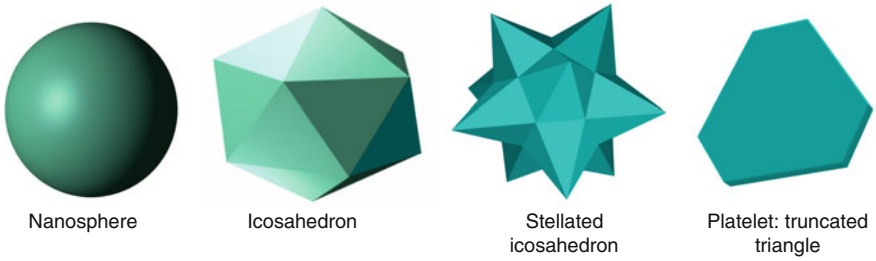


Fig. 2.59 Building blocks of plasmonic nanocomposites: nanoparticles fabricated by bottom-up approach

The relative dielectric permittivity of plasmonic materials is negative below plasma frequency, and its dispersion is by Drude electron resonance model [298]

$$\varepsilon = \varepsilon_\infty - \frac{\omega_p^2}{\omega(\omega + i\Gamma)}, \tag{2.51}$$

where ω_p is the plasma frequency, Γ is the damping factor describing losses, and ε_∞ is the asymptotic relative dielectric permittivity. The plasma frequency is determined by the properties of free carriers as

$$\omega_p^2 = \frac{n_e e^2}{m^* \varepsilon_0}, \tag{2.52}$$

where n_e is electron concentration, e is the free electron charge (1.6×10^{-19} C), ε_0 is the free space (vacuum) permittivity (8.854×10^{-12} F/m), and m^* is the electron effective mass. The damping factor is

$$\Gamma = \frac{e}{\mu m^*} \tag{2.53}$$

where μ is mobility of free carriers.

Dispersion relation of plasmonic material can be tuned by design, leading to an optical behavior surpassing natural materials. These are plasmonic metamaterials [205]. One can use them to obtain superluminal group velocities ("fast light"), near-zero ("slow light"), or negative ("left-handed light") [207]. This is of crucial importance for photodetector enhancement, since the design of arbitrary frequency dispersion gives a possibility to convert propagating far field modes into spatially localized near-field modes. This ensures strongly increased density of states and thus light trapping in subwavelength volumes.

Elimination of absorption losses that are obviously undesired in photodetection is a field of active investigation, and different schemes were proposed for its avoidance [299]. One of the approaches is the use of alternative plasmonic materials [300] including graphene [301].

The applicability of plasmonics for photodetector enhancement including infrared devices has been recognized in 1970s. The first proposed applications of SPP were in photodetection [243]. Surface plasmon polariton-mediated light trapping schemes for photodetectors [299, 302] may be divided into several distinct groups. Probably the most important is enhanced Mie scattering on plasmonic nanoparticles or nanovoids [303], coupling into guided modes [302], field localization and generation of hotspots near the surface of plasmonic material [304], use of plasmon-based singular optics (optical vortices) [299], use of metamaterial-based transformation optics to map the optical space into a desired shape and with an increased density of states [305] and plasmon-enhanced up-conversion media [306].

Besides nanoparticles and nanovoids as scatterers and nanoantennas, different diffractive structures can be used (gratings, lattices) [129] for field coupling into

guided modes; they may be ordered or disordered, as well as subwavelength plasmonic crystals (SPC) which may be periodic [307] or quasiperiodic [243]. Plasmonic nanocomposites with nonresonant operation were also reported [308].

Among subwavelength plasmonic crystals described within the context of infrared photodetection, an important position belongs to 2D arrays of nanoapertures in opaque metal films. Such structures first drew attention for their ability to transmit light in spite of the dimensions of nanoapertures that are much smaller than the operating wavelength. This effect was denoted as extraordinary optical transmission (EOT) [309].

Extraordinary optical transmission is a result of resonant excitation of surface plasmons polaritons at the metal surface. It forces the electromagnetic waves incident to the surface to pass through the apertures. The transmission occurs for all incident electromagnetic energy, not only the one impinging to the EOT aperture. However, the transmitted energy passes only through the apertures. Thus large field localizations appear in the apertures. This behavior effectively corresponds to impedance matching between propagating waves and the perforated metal film. Thus the extraordinary optical transmission arrays act as antireflective structures.

Another application of 2D arrays of subwavelength holes in metal has found an important use in light trapping for infrared detectors. It is based on the so-called designer plasmons.

Pendry et al. [310] have shown that a surface wave propagating along a 2D array of apertures in metal film one can be described by an effective permittivity as

$$\varepsilon_{\text{in-plane}} = \frac{\varepsilon_{\text{hole}}\pi^2 d^2}{8a^2}\left(1 - \frac{\pi^2 c^2}{a^2\omega^2\varepsilon_{\text{hole}}\mu_{\text{hole}}}\right) \tag{2.54}$$

where $\varepsilon_{\text{hole}}$ is the permittivity of the material within the aperture, a is the characteristic dimension of the aperture (side length in the case of square holes), Fig. 2.60, while k_0 is the wavevector in vacuum. The effective plasma frequency of such material is

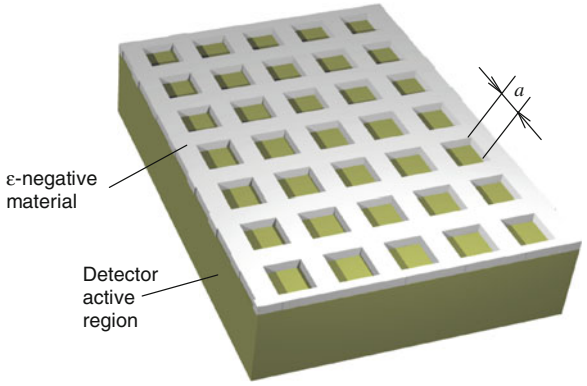


Fig. 2.60 Metallodielectric EOT structure introducing "designer" plasmons with structurally tunable plasma frequency

$$\omega_p = \frac{\pi c}{a\sqrt{\varepsilon_{\text{hole}}\mu_{\text{hole}}}}. \tag{2.55}$$

In other words, the effective dielectric permittivity of an EOT array has a form identical to that of plasmonic materials. At the same time, its plasma frequency and thus its operating range can be freely tuned to the wavelength range of interest. Such surface waves actually mimic real SPP, so that one can fabricate an effective plasmonic material with a desired spectrum. Pendry denoted these surface waves as "designer" plasmons and they are also known as "spoof" plasmons.

The existence of designer SPP is extremely important for infrared photodetectors. One is able to fabricate any desired plasmonic structure and to tune it for the targeted wavelength range. As an example, infrared detectors enhanced by designer plasmon structures tuned to the range of 8–10 μm have been reported [311].

2.16 Optical Antennas

Nanoantenna or optical antenna [243] is a structure that couples with propagating light waves and redirects them into evanescent field. Thus it ensures a highly efficient linking between the two. Depending on the nanoantenna type and its design, the mode confinement in the evanescent field can be tailored and it can be deeply subwavelength. This is of large interest for photodetectors since optical space is basically squeezed into a small region where the interaction with photodetector-active region can be vastly enhanced.

Contrary to the conventional antennas for radio waves, nanoantennas are not connected to any feeding circuitry. They are standalone structures and are used in large arrays. They enhance scattering of electromagnetic waves, the more so the nearer they are to resonance. The literature describes various types of nanoantennas that were experimentally produced. Figure 2.61 shows some of the basic geometries.

The simplest and the most basic nanoantenna is the nanosphere. This structure actually behaves like a dipole. Its scattering properties can be calculated using the Mie theory [266]. Noble metal nanoparticles are often fabricated in spherical form. This makes them the simplest nanoantennas.

If two spherical nanoantennas are brought near to each other so that there is a gap with subwavelength width between them, they form a nanodimer. These nanospheres are electromagnetically coupled, and the coupling gap between them is denoted as the feed gap. A field hotspot appears in the feed gap, where localization is deeply subwavelength and field enhancement is very strong. Much larger field localizations are obtained using such coupled structures than by independent nanospheres. Nanoparticles can be also ordered in an array (nanoparticle chain) [312]. Thus obtained optical antenna effectively behaves as linear nanorod antenna.

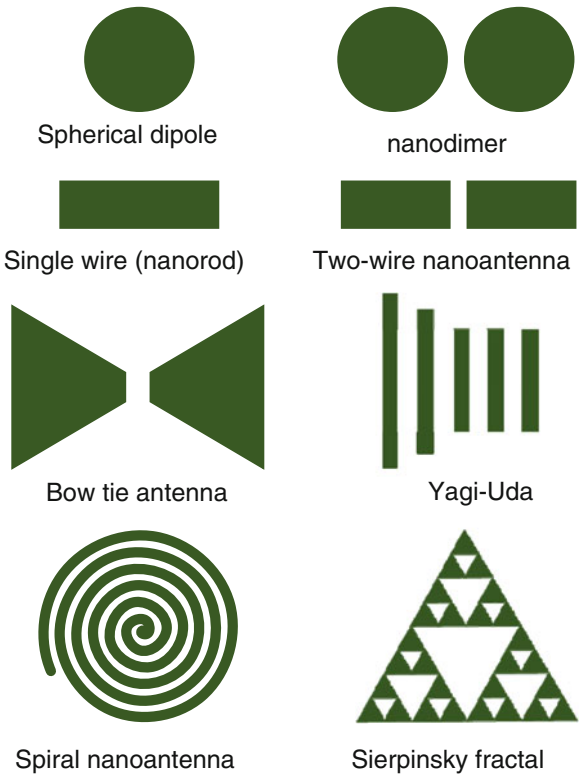


Fig. 2.61 Some different types of experimental plasmonic nanoantennas

A generalization of this case is a nanorod antenna where the dipole length can be tailored. These antennas represent linear dipoles. If two such linear dipoles are aligned and brought together to a subwavelength distance, end-to-end coupling occurs. This is the a two-wire nanoantenna [308]. This can be further generalized by adding two more dipoles perpendicularly to the first ones, all four having a joint feed gap. This configuration is called the cross-antenna.

Another very often met nanoantenna structure is the bowtie nanoantenna [313]. It consists of two triangles aligned along their symmetry axes. A feed gap is formed between their tips. Bowtie antennas have a broader bandwidth together and at the same time ensure large field localizations in the feed gap. A similar type of antenna is the diabolo-type nanoantenna [314], where the triangles overlap.

There is a plasmonic analogy of the well-known Yagi-Uda antenna. It can be fabricated by placing a resonant nanorod antenna between a reflector nanorod and a group of director nanorods [315]. Similar to such antennas used in radiofrequent domain, it ensures a good directivity.

Among the nanoantennas mentioned in the literature are spiral nanoantennas [316] and those with fractal geometries [317]. Many other shapes can be used. Basically, all "particles" used to fabricate metamaterials can also function as nanoantennas. This includes single and double split rings, crescent antennas, etc.

An important group of nanoantennas are those based on the Babinet principle. A metal shape surrounded by dielectric and a dielectric-filled hole in metal with identical shape and size have identical diffraction patterns. Thus bowtie shaped holes in metal can be used, two holes are a Babinet equivalent of a nanodimer, arrays of nanoholes are equivalent of arrays of nanoparticles, crossed arrays of nanoholes correspond to crossed arrays of nanoparticles, etc. [318].

Nanoantennas are used in photodetection to couple propagating and localized modes and to localize fields. Especially, strong localizations are those in the feed gaps between coupled nanostructures. Photodetector enhancement by nanoantennas is a hot topic and numerous works have been published on it [243]. Earliest publications date as far as 1970s [319].

One of the approaches to the use of nanoantennas in photodetection is to use a Schottky metal-semiconductor junction. An optical antenna forms the metal part of the metal-dielectric contact at the semiconductor detector surface [320]. Photoexcitation generates hot electron-hole pairs by plasmon decay and the electrons are injected over the Schottky barrier, thus directly generating photocurrent. A problem with this approach is its low efficiency.

2.17 Plasmonic Nanoparticles and Redshifting

As the simplest nanoantennas, plasmonic nanoparticles can be utilized to enhance the absorption within thin-film solar cells [243]. They couple incoming waves with the localized SPP field, have increased scattering cross-section and strongly localize electromagnetic field just in the thin active region of the detector, Fig. 2.62. The same principle is applicable for infrared detection [321]. This cannot be done with pure noble metal nanoparticles since their surface plasmon resonance is in ultraviolet or visible part of the spectrum. Because of that their response must be redshifted. In this part, two approaches to such redshifting are described.

As Fig. 2.62 shows, the scattering cross-section of a plasmonic nanoparticle is greatly enhanced by plasma resonance. It can readily reach an order of magnitude value larger than the geometrical cross-section. The scattering cross-section for plasmonic nanoparticles at a wavelength λ can be calculated as [322]

$$C_{\text{scat}} = \frac{8}{3\pi}\alpha^2\left[\frac{\pi}{\lambda}\right]^4, \tag{2.56}$$

where

$$\alpha = 3V\frac{\frac{\varepsilon_{\text{np}}}{\varepsilon_{\text{d}}} - 1}{\frac{\varepsilon_{\text{np}}}{\varepsilon_{\text{d}}} + 2}. \tag{2.57}$$

where ε_{np} is the complex relative dielectric permittivity of the plasmonic nanoparticles, determined by the Drude formula, ε_{d} is the permittivity of the dielectric

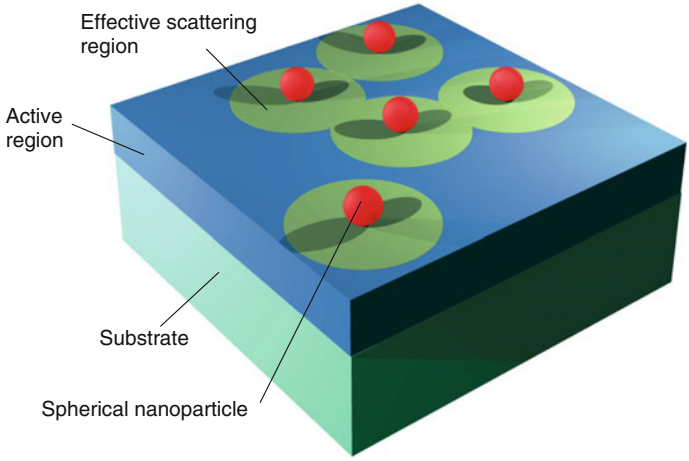


Fig. 2.62 Increase of scattering by nanoparticles on detector surface

medium, and V is the nanoparticle. The plasmon resonance and thus the maximum value of the scattering cross-section are reached at $\varepsilon_{np} = -2\ \varepsilon_d$.

The absorption cross-section is

$$C_{abs} = \frac{2\pi}{\lambda} \mathrm{Im}(\alpha). \tag{2.58}$$

Figure 2.63 shows several schemes for the enhancement of the performance of an IR detector by using nanoparticles and redshifting strategies. Figure 2.62a shows the case when alternative material (transparent conductive oxide like ZnO, ITO, and similar) [243] is used instead of metal nanoparticles. The plasma frequency of TCO nanoparticles is tunable by doping and by the fabrication procedure and in principle reaches longer wavelength (in infrared), while at the same time losses are smaller [243]. Another approach is to use alternative geometries of nanoparticles. Different metal-dielectric combinations and different geometries result in widely varying values of plasma frequency. For instance, nanoshells can have plasma frequency from visible to 9 μm, nanotriangles reach 5 μm, and nanorice 2 μm [323]. A core-shell nanoparticle is shown on the top of the active region (Fig. 2.62b). Instead of nanoparticles with alternative geometry and composition, one can use nanoparticles embedded in high-permittivity dielectric (requires additional antireflection layer at the surface), Fig. 2.62c. Plasmonic nanoparticles can be also placed at the back surface of the detector or within its active region, and not at the top of the detector, Fig. 2.1d. Further redshifting can be obtained by adjusting the distance between the nanoparticles. It has been shown that smaller interparticle distance results in a larger redshift [324]. Obviously, two or more of the quoted strategies can be used simultaneously.

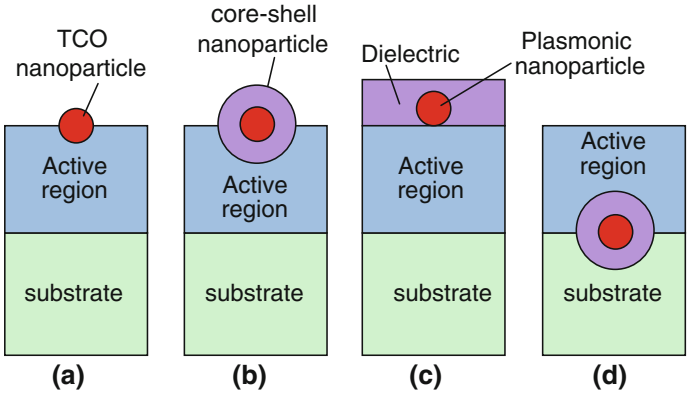


Fig. 2.63 Some schemes for infrared IR detector enhancement by plasmonic nanoparticles. **a** Use of transparent conductive oxide at the surface; **b** core-shell plasmonic nanoparticle embedded in high-permittivity dielectric at the surface; **c** high-permittivity dielectric layer with embedded nanoparticle; **d** core-shell plasmonic nanoparticle embedded in high-permittivity dielectric at the back surface

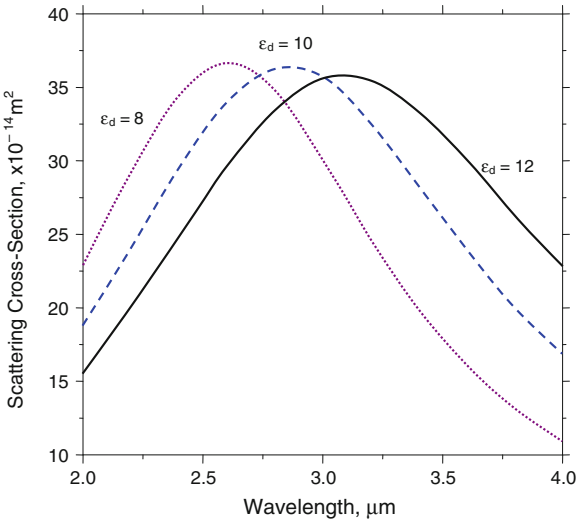


Fig. 2.64 Spectral dependence of scattering S_{scat} cross-section for an embedded ITO particle, r = 60 nm, λ_p = 625 nm, substrate permittivity 10

Figure 2.64 shows the calculated scattering cross-section for a spherical dipole indium tin oxide nanoparticle with a radius of 60 nm at a substrate with a relative dielectric permittivity of 10. The doping concentration was 1.2×10^{21} cm^{-3}, with an effective mass of m* = 0.4 m_0, thus plasma frequency was 4.8×10^{14} Hz. The nanoparticle is located at the top of the active surface of the detector and is embedded in dielectric (the values are shown near each respective curve). Finite

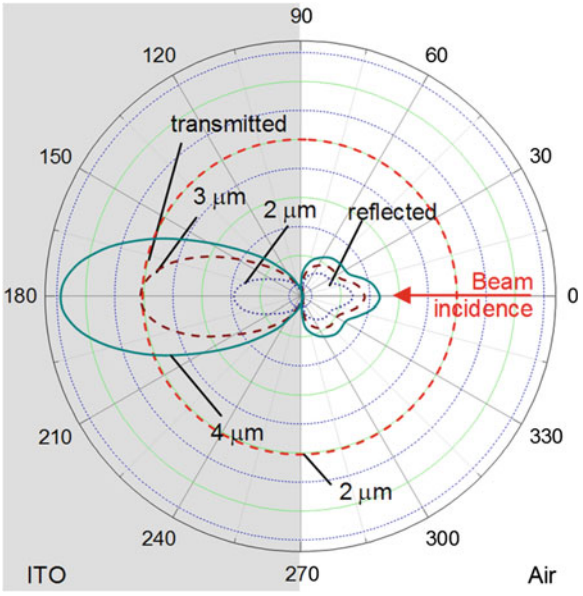


Fig. 2.65 Polar diagram of far field distribution of electric field due to scattering around a spherical ITO nanoparticle obtained by finite element method. r = 60 nm, λ_p = 625 nm, operating wavelengths 2, 3, and 4 μm. Permittivity of the dielectric layer is 10

element method was utilized for calculation. A redshift of the peak of the scattering cross-section increases with the permittivity of the embedding dielectric.

Figure 2.65 shows the radial distribution of the scattered electric field. Forward and back scattering are readily seen. It is visible that scattering increases with the operating wavelength. An increase of the permittivity of the embedding dielectric layer also causes an increase of scattering.

Nanoparticles can be ordered at the detector surface or elsewhere using self-assembly techniques or pattering by top-down approach. They can be distributed in regular patterns, thus accurately controlling the interparticle distance or they can be randomly scattered. A possible way to implement nanoparticles for plasmonic enhancement is to arrange them in a quasicrystal pattern (for instance Penrose tiling), which ensures an isotropic photonic response of the structure [325].



http://www.springer.com/978-3-319-09673-5

Micro and Nanophotonics for Semiconductor Infrared Detectors
Towards an Ultimate Uncooled Device
Jakšić, Z.
2014, XXI, 258 p. 152 illus., 41 illus. in color., Hardcover
ISBN: 978-3-319-09673-5