Review article

Chee Leong Tan and Hooman Mohseni*

Emerging technologies for high performance infrared detectors

https://doi.org/10.1515/nanoph-2017-0061
Received May 30, 2017; revised August 2, 2017; accepted August 6, 2017

Abstract: Infrared photodetectors (IRPDs) have become important devices in various applications such as night vision, military missile tracking, medical imaging, industry defect imaging, environmental sensing, and exoplanet exploration. Mature semiconductor technologies such as mercury cadmium telluride and III–V material-based photodetectors have been dominating the industry. However, in the last few decades, significant funding and research has been focused to improve the performance of IRPDs such as lowering the fabrication cost, simplifying the fabrication processes, increasing the production yield, and increasing the operating temperature by making use of advances in nanofabrication and nanotechnology. We will first review the nanomaterial with suitable electronic and mechanical properties, such as two-dimensional material, graphene, transition metal dichalcogenides, and metal oxides. We compare these with more traditional low-dimensional material such as quantum well, quantum dot, quantum dot in well, semiconductor superlattice, nanowires, nanotube, and colloid quantum dot. We will also review the nanostructures used for enhanced light-matter interaction to boost the IRPD sensitivity. These include nanostructured antireflection coatings, optical antennas, plasmonic, and metamaterials.

Keywords: infrared detectors; 2D material; optical antenna; plasmonics.

1 Introduction

Infrared (IR) radiation was discovered 217 years ago. However, the first IR photodetectors (IRPDs) were only invented in the late 20th century. Detecting and sensing IR radiation is extremely important in various applications. IR radiation can generally be divided into different categories, namely near-wavelength region from 0.8 to 3 μm, mid-wavelength IR (MWIR) region from 3 to 5 μm, and long wavelength IR (LWIR) region from 8 to 12 μm. Note that there are different conventions for naming these bands. For example, 1.55 μm is considered short-wavelength IR, or SWIR, within the Department of Defense community, but is considered near-IR (NIR) within the astronomy community. The MWIR and LWIR regions are most important for free-space communications and astronomy, since the high transparency of the atmosphere at these wavelengths allows transmission without significant losses. The same regions are also widely used for night vision and missile tracking for military applications. IR detection also enables many non-contact, non-destructive inspection methods that are widely utilized in industry, such as automotive and defect detection in electronic circuits. IR detection has similarly found many medical applications. Most such IR radiations directly come from a thermal source, and hence they carry information on the identity of the source itself. For example, the temperature of the emitting object can be determined from the IR radiation received, allowing for non-contact, high-sensitivity temperature measurements. Moreover, the wavelength spectrum of the detected IR radiation can also provide a great deal of information on the composition of the source or intermediate absorbing medium. The bonding energies of many materials, especially organic compounds, have resonant frequencies in the IR region. Detecting these resonances (i.e. IR spectroscopy) is a very powerful tool for analyzing chemical bonds in molecules and is used in astronomy to analyze stars and exoplanets.

The first semiconductor photodetectors were fabricated using lead sulfide (PbS) in Germany [1, 2] for military application such as night vision as early as 1933. In 1941, Cashman [3] successfully improved the technology using...
thallous sulfide \[\text{[4]}\] and successful produced PbS detectors for the first time in the US at Northwestern University in 1944 \[\text{[5]}\]. After World War II, the IRPDs were widely used in military and astronomy applications such as simple radiometric instruments, passive night vision technology capable of allowing vision under ambient starlight conditions, and anti-air missile seekers \[\text{[5]}\]. Even up to now low-cost versatile PbS and PbSe polycrystalline thin films remain the photoconductive detectors of choice for many applications in the 1–3 \(\mu\)m and 3–5 \(\mu\)m spectral range.

The first extrinsic (photon energies smaller than the bandgap) photoconductive detectors were reported in the early 1950s \[\text{[6, 7]}\] after the discovery of the transistor, which stimulated a considerable improvement in growth and material purification techniques. Extrinsic photoconductive response from copper, mercury, zinc, and gold impurity levels doped in germanium gave rise to devices operating in the 8–14 \(\mu\m\) (LWIR spectral window) and the 14–30 \(\mu\m\) very long wavelength IR (VLWIR) region. They must be operated at very low temperatures to achieve performance similar to that of intrinsic detectors, and they sacrifice quantum efficiency to avoid needing thick detectors. Although Si has several advantages over Ge (namely, a lower dielectric constant giving shorter dielectric relaxation time and lower capacitance, higher dopant solubility and larger photoionization cross-section for higher quantum efficiency, and lower refractive index for lower reflectance), the performance is much lower compared to highly developed Ge detectors. After being dormant for about 10 years, extrinsic Si was reconsidered after the invention of charge-coupled devices (CCDs) by Boyle and Smith \[\text{[8]}\]. In 1973, Shepherd and Yang \[\text{[9]}\] proposed the metal silicide/silicon Schottky-barrier detectors. For the first time it became possible to have much more sophisticated readout schemes – both detection and readout could be implemented on a silicon chip.

At the same time in the early 1950s, rapid advances were being made in narrow bandgap semiconductors that would later prove useful in extending wavelength capabilities and improving sensitivities of devices. The beginning of the 1960s saw the introduction of narrow gap III–V semiconductor alloys \((\text{InAs}_{1-x}\text{Sb}_x)\), IV–VI \((\text{Pb}_{1-x}\text{Sn}_x\text{Te})\), and II–VI \((\text{Hg}_{1-x}\text{Cd}_x\text{Te})\) material systems. These materials have provided an unprecedented degree of freedom in IR detector design. The bandgap energy tunability results in IR detector applications that span all IR bands. The fundamental properties of narrow-gap semiconductors (high optical absorption coefficient, high electron mobility, and low thermal generation rate), together with the capability for bandgap engineering, make these alloy systems almost ideal for a wide range of IR detectors. The difficulties in growing mercury cadmium telluride (HgCdTe) material, significantly due to the high vapor pressure of Hg, encouraged the development of alternative detector technologies over the past 40 years. One of these was PbSnTe, which was vigorously pursued in parallel with HgCdTe in the late 1960s and early 1970s \[\text{[10, 11]}\]. PbSnTe was comparatively easy to grow and good-quality LWIR photodiodes and lasers were readily demonstrated. However, in the late 1970s two factors led to the abandonment of PbSnTe detector work: high dielectric constant and large temperature coefficient of expansion (TCE) mismatch with Si. The scanning-based IR imaging systems of the 1970s required relatively fast response times so that the scanned image was not smeared in the scan direction. A large dielectric constant hinders making low-capacitance devices, which are needed for fast operation. The second drawback, large TCE, can lead to failure of the indium bonds in hybrid structure (between silicon readout and the detector array) after repeated thermal cycling from room temperature to the cryogenic temperature of operation. Eighty years since the development of the first semiconductor IRPD, the development of IRPDs has been matured and well studied. At the end of the last millennium there were two dominant IR systems, InSb- and HgCdTe-based detectors, which were well developed and available in commercial systems. While these two systems saw improvements over the last 12 years, their change has not nearly been as marked as that of the engineered quantum-based detectors [i.e. quantum well infrared photodetectors (QWIPs), quantum dot-based infrared photodetectors (QDIPs), quantum dot in well infrared photodetectors (DWELL-IPs), and strained-layer superlattices (SLS)-based photodetectors]. However, the mature IRPD technology is costly, difficult to fabricate, and requires low temperature operation. In the last two decades, a lot of new nanostructures and technologies have emerged to reduce the cost of IRPDs and improve the IRPD performance especially for high temperature operation.

In this paper, we review the progress made in all of these systems over the last decade. We will also review the new material systems, such as two-dimensional (2D) materials, carbon nanotubes, and metal oxides. In the second section, we review IRPD devices based on the mature material namely II–VI, III–V, and silicon and compare them with emerging new materials such as graphene, transition metal dichalcogenides (TMDCs), and carbon nanotubes. In the third section, we will review the current progress of nanostructures such as graphene, 2D materials, quantum well (QW), quantum dot (QD), DWELL, SLS and colloid quantum dot (CQD) which dramatically improve the electronic properties of IRPDs. Newly emerging nanostructures
such as nanowires and nanotubes are also discussed. In the fourth section, we will focus on nanostructures used to enhance the coupling of light with the IRPD. These include nanostructured antireflection (AR) coatings, optical antennas, plasmonic and metamaterials. In the last section, we will briefly list the pros and cons of all the discussed material systems and future prospect of the structures and materials covered here.

2. IRPD based on different bulk material systems

For an IRPD to work, a light-sensitive material capable of absorbing electromagnetic radiation is needed. It converts the absorbed light into a signal that can be collected by a readout device. In order for the absorption process to happen, a semiconductor is used generally. While semiconductor-free methods (such as direct rectification of the electromagnetic radiation, called “rectenna”) have been demonstrated, they have far smaller quantum efficiency compared with the semiconductor-based IRPDs.

For different IR regimes or wavelengths, different semiconductor materials with different effective bandgaps are needed for IR detection. In this section, we review IRPDs based on the mature developed material, namely II–VI, III–V, silicon-based semiconductor, and emerging new materials such as graphene, TMDCs and carbon nanotubes. In the past decade, 2D materials have attracted much attention due to their extraordinary electronic and optical properties, and have a large potential in optoelectronic applications such as photodetection. We will compare the recent progress in IR photodetection based on graphene and graphene-like 2D materials with the commercially mature IRPD materials. Additionally, we will look at the key potentials and the difficulties involved in commercialization and integration of such emerging material into focal plane arrays (FPAs). Figure 1 shows the detectivity of various available detectors with different materials when operated at the indicated temperature.

2.1 II–VI-based IRPD

The II–VI material system, such as HgCdTe, is one of the most important semiconductor alloy systems and the most commonly used material for IR detectors in the spectral range between 1 and 25 μm. The HgCdTe ternary alloy is an ideal IR material system due to these key features:

![Figure 1: Comparison of the detectivity of various available detectors when operated at the indicated temperature. Theoretical curves for the background-limited \( D^* \) (dashed lines) for ideal photovoltaic and photocative detectors and thermal detectors are also shown. HEB, hot electron bolometer; PC, photoconductive detector; PEM, photoelectromagnetic detector; PV, photovoltaic detector [12].](image-url)
composition-dependent tailorable energy bandgap, high optical absorption coefficient, high quantum efficiency and favorable inherent recombination mechanisms leading to long carrier lifetimes and high operating temperatures. The attributes of HgCdTe translate to flexibility and the capability to produce SWIR, MWIR, and LWIR detectors. HgCdTe IRPD can be divided into two base structures, front illuminated and back illuminated. Rogalski has reviewed the HgCdTe IRPD structures in detail [13]. Despite its unparalleled device performance, such as high detectivity, high quantum efficiency (>70%), and fast response time (several ns), HgCdTe IR detector technology has a number of major limitations, including health hazard, high cost, limited array size, low operating temperature and material fragility. Consequently, other competing IR technologies have been regularly proposed to replace HgCdTe to overcome the limitations of HgCdTe-based IR detectors. These include lead salt alloys, InSb, QW IR detectors, QD IR detectors and type II superlattice (T2SL)-based detectors. Although significant effort has been devoted to developing these alternative IR technologies [2, 11, 13–24], at this stage, HgCdTe IR detectors still dominate high performance IR systems. However, HgCdTe IR technology has its own challenges, including: (1) p-type doping control; (2) lower cost and larger array format size for FPAs; (3) higher operating temperatures; (4) multi-band detection; and (5) advanced plasma dry etching control.

2.2 III–V alloy-based IRPD

The first III–V material used for MWIR detection was InSb, which had the smallest energy gap of any III–V binary alloys known at the time. The energy gap of InSb is poorly matched to the 3–5 μm band at higher operating temperatures, and better performance can be obtained from HgCdTe. InAs is a similar compound to InSb, but has a larger energy gap, so that the threshold wavelength is 3–4 μm. However, MWIR HgCdTe photodiodes have significant higher performance in the 30–120 K temperature range. The InSb devices are dominated by generation-recombination currents in the 60–120 K temperature range because of defect centers in the energy gap.

InGaAs ternary alloy was made available after improvements were made in growing techniques. The energy gap of the InGaAs ternary system spans from 0.35 eV for InAs to 1.43 eV for GaAs. In0.53Ga0.47As alloy (Eg = 0.73 eV) lattice matched to the InP substrate has been shown to be a suitable detector material for NIR (1.0–1.7 μm) spectral range compared to HgCdTe and InSb. Having lower dark current and noise than indirect-bandgap germanium, InGaAs is addressing both entrenched applications including lightwave communication systems, low light-level night vision, and new applications such as remote sensing: eyesafe range finding and process control [25–33]. InGaAs photodiodes have shown high device per absorption layer. The photodetector responsivity can be maximized at the desired wavelength of the end-user to enhance the signal-to-noise ratio.

Another material in III–V family which stands out compared to other materials for IRPD is InAsSb. InAsSb ternary alloy is more stable in comparison with HgCdTe and has a fairly weak dependence between the band edge and composition. The stability of this material is conditioned by the stronger chemical bonds available in the lower atomic number III–V family and the larger covalent bonding contribution compared to ionic bonding in HgCdTe. Some other physical properties of InAsSb material are better compared to those of HgCdTe. For instance, its dielectric constant is low and the room temperature self-diffusion coefficient is low. However, the InAsSb ternary alloy does not have a sufficiently small gap at 77 K for operation in the 8–14 μm wavelength range, but work began on SLs [34], which will be discussed in Section 3.1.2.

2.3 Silicon/germanium-based IRPD

Silicon, being an indirect bandgap semiconductor with a centro-symmetric crystalline structure, is not directly suited for optoelectronics. Si/Ge IRPDs are usually divided into two types: intrinsic detectors and extrinsic detectors. Silicon intrinsic detectors are usually designed and fabricated as PIN photodiodes or Schottky-barrier detectors [14].

For intrinsic IRPDs, Si has lower costs and lower leakage current advantage – both stem from the near-perfect processing of this material. Si-based IRPDs are compatible with silicon circuits; however, silicon is not the ideal material for IRPDs due to its bandgap mismatch to many important wavelengths including the telecommunication bands. Si has a lower dielectric constant than Ge, and the related device technology of Si has been extensively developed, including contacting methods, surface passivation, and mature metal oxide semiconductor and CCD technologies. Moreover, Si detectors are characterized by superior hardness for high-energy particles, such as nuclear radiation and space environments. The well-established technology also helps in the manufacturing of uniform detector arrays and the formation of low-noise contacts. Impressive progress in the development of hybrid p-i-n Si-complementary metal oxide semiconductor (CMOS) arrays
for the large visible and NIR imaging market has been presented by Raytheon Vision Systems [35–37]. For the longer wavelengths, the most popular Schottky-barrier detector is the PtSi detector, which can be used for detection in the 3–5 μm spectral range. Radiation is transmitted through the p-type silicon and is absorbed in the metal PtSi, producing hot holes which are then emitted over the potential barrier into the silicon, leaving the silicide charged negatively [38].

For even longer wavelength, extrinsic IRPDs can be used as they can reach up to approximately 300 μm. The spectral range of the extrinsic photodetector is determined by the doping impurity and by the material into which it is introduced. Detectors based on silicon and germanium have found the widest application compared with extrinsic photodetectors on other materials. Increasing interest in extrinsic detectors has been observed due to the creation of multi-element FPAs for application in space and ground-based IR astronomy and on defense space vehicles. Successes in the technology of photodetectors, creation of deep-cooled, low-noise semiconductor preamplifiers and multiplexers, as well as unique designs of photodetector devices and equipment for deep cooling, have ensured the achievement of a record-breaking detectivity close to the radiation limit even under exceedingly low backgrounds in space [39, 40]. A detailed review of a bulk Si/Ge IR detector can be found in [39], which shows the mature application of Si/Ge IR detectors.

Table 1 shows some of the best existing IRPD imaging arrays based on the aforementioned material systems.

3 Nanostructures to enhance the electronic properties of the IRPD

Based on the considerations of the previous section, we can further improve the photodetector performance

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Size</th>
<th>Pixel size (μm)</th>
<th>Detector material</th>
<th>Spectral range (μm)</th>
<th>Operating temp (K)</th>
<th>$D^* (\text{cm Hz}^{1/2}/\text{W})$ and NETD (mK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goodrich corporation</td>
<td>320 × 240</td>
<td>25 × 25</td>
<td>InGaAs</td>
<td>0.9–1.7</td>
<td>300</td>
<td>1 × 10^{13}</td>
</tr>
<tr>
<td></td>
<td>640 × 512</td>
<td>25 × 25</td>
<td>InGaAs</td>
<td>0.4–1.7</td>
<td>300</td>
<td>&gt;6 × 10^{13}</td>
</tr>
<tr>
<td>Raytheon vision systems</td>
<td>1024 × 1024</td>
<td>30 × 30</td>
<td>InSb</td>
<td>0.6–5.0</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2048 × 2048</td>
<td>25 × 25</td>
<td>HgCdTe</td>
<td>0.6–5.0</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2048 × 2048</td>
<td>20 × 20</td>
<td>HgCdTe</td>
<td>0.8–2.5</td>
<td>4–10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2048 × 2048</td>
<td>15 × 15</td>
<td>HgCdTe/Ge</td>
<td>3.0–5.0</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1024 × 1024</td>
<td>25 × 25</td>
<td>Si:As</td>
<td>5–28</td>
<td>6.7</td>
<td></td>
</tr>
<tr>
<td>Teledyne imaging sensors</td>
<td>2047 × 1024</td>
<td>25 × 25</td>
<td>Si:As</td>
<td>5–28</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4096 × 4096</td>
<td>10 × 10 or 15 × 15</td>
<td>HgCdTe</td>
<td>1.0–1.7</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4096 × 4096</td>
<td>10 × 10 or 15 × 15</td>
<td>HgCdTe</td>
<td>1.0–2.5</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4096 × 4096</td>
<td>10 × 10 or 15 × 15</td>
<td>HgCdTe</td>
<td>1.0–5.4</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2048 × 2048</td>
<td>18 × 18</td>
<td>HgCdTe</td>
<td>1.0–1.7</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2048 × 2048</td>
<td>18 × 18</td>
<td>HgCdTe</td>
<td>1.0–2.5</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2048 × 2048</td>
<td>18 × 18</td>
<td>HgCdTe</td>
<td>1.0–5.4</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Teledyne imaging sensors</td>
<td>2047 × 1024</td>
<td>25 × 25</td>
<td>HgCdTe</td>
<td>0.8–2.5</td>
<td>&lt;200</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1000 × 256</td>
<td>30 × 30</td>
<td>HgCdTe</td>
<td>3.7–4.8</td>
<td>77–100</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>1280 × 1024</td>
<td>15 × 15</td>
<td>HgCdTe</td>
<td>7.7–9.5</td>
<td>77–80</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>384 × 288</td>
<td>25 × 25</td>
<td>HgCdTe</td>
<td>8.0–9.0</td>
<td>73</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>640 × 512</td>
<td>20 × 20</td>
<td>HgCdTe</td>
<td>MIR</td>
<td>77–80</td>
<td>15–20</td>
</tr>
<tr>
<td></td>
<td>640 × 512</td>
<td>24 × 24</td>
<td>HgCdTe</td>
<td>LWIR/MWIR</td>
<td>77–80</td>
<td>20–25</td>
</tr>
<tr>
<td>Selex</td>
<td>384 × 288</td>
<td>40 × 40</td>
<td>Type II SL</td>
<td>MW</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1024 × 768</td>
<td>16 × 16</td>
<td>HgCdTe</td>
<td>3–5</td>
<td>140</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>640 × 512</td>
<td>24 × 24</td>
<td>HgCdTe</td>
<td>8–10</td>
<td>90</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>640 × 512</td>
<td>24 × 24</td>
<td>HgCdTe</td>
<td>MW/LW</td>
<td>80</td>
<td>28</td>
</tr>
<tr>
<td>AIM</td>
<td>384 × 288</td>
<td>40 × 40</td>
<td>Type II SL</td>
<td>MW</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1024 × 1024</td>
<td>15 × 15</td>
<td>Si:As</td>
<td>3–5</td>
<td>77</td>
<td>20</td>
</tr>
<tr>
<td>SCD</td>
<td>2048 × 2048</td>
<td>18 × 18</td>
<td>Si:As</td>
<td>5–28</td>
<td>7.8</td>
<td></td>
</tr>
<tr>
<td>DRS Technologies</td>
<td>2048 × 2048</td>
<td>18 × 18</td>
<td>Si:Sb</td>
<td>5–40</td>
<td>7.8</td>
<td></td>
</tr>
</tbody>
</table>
(detectivity bandwidth product) by two methods. First method is based on engineering the electronic properties of the absorbing material for improved figures of merits such as detectivity and quantum efficiency. The second method is based on increasing the coupling of light to the electronic states of the absorbing material. Interestingly, both methods can be implemented by utilizing different micro- and nanofabrication techniques. In the following we will discuss the nanostructures used to enhance the electronic properties of the IRPD, while in Section 4 we will cover the nanostructures used to enhance the optical coupling of the IRPD to free-space photons.

We will first discuss QW and semiconductor super-lattice (SL) structures, which are amongst the oldest and most mature nanotechnologies used to enhance the efficiency of the IRPD. SL and QW structures have continued to dominate the field due to the possibility of creating new classes of materials and heterojunctions with unique electronic and optical properties with high uniformity of epitaxial growth over large areas.

3.1 QW-based IRPD

3.1.1 Intersubband QWIP

Figure 2 shows the low-temperature energy bandgap diagram of different semiconductors with diamond and zinc-blende structure versus their lattice constant matching. Thin layers of semiconductors with close lattice constants – within each of the gray zones – can be grown on top of each other, using strain compensation. Such thin QWs and barriers can be designed as shown in Figure 3 to produce localized quantum states within each conduction or valence band, with a specific energy difference to resonate with the targeted IR radiation. Such “intersubband” photodetectors are called QWIPs.

Among the different types of QWIPs, GaAs/AlGaAs QW detectors are the most mature due to the almost perfect natural lattice match between GaAs and AlGaAs [42–54]. GaAs/AlGaAs QW devices have many advantages, including their ability to be used with standard GaAs-based manufacturing techniques and processing technologies, highly uniform and well-controlled molecular beam epitaxy growth on six-inch GaAs wafers, high yield and thus low cost, more thermal stability, and extrinsic radiation hardness. However, the optical cross-section absorption is limited by the dopant concentrations, leading to a low quantum efficiency and relatively poor performance at temperatures >40 K due to short intersubband lifetimes (dominated by fast phonon relaxation). While some complex quantum structures have been explored [51], the emerging ideas to address this performance limitation are based on optical antenna and plasmonic structures, which will be discussed in Section 4.

3.1.2 Interband superlattice IRPD

A rapidly maturing III–V IRPD structure is InAs/InGa_{1-x}Sb (InAs/InGaSb) strained-layer T2SL detectors [52–55]. The optical absorption of T2SL occurs in interband transition, which makes it an intrinsic device with high quantum

![Figure 2](image)

**Figure 2**: The low-temperature energy bandgap diagram of different semiconductors with diamond and zinc-blende structure versus their lattice constants by Nobel Laureate Esaki [41].

![Figure 3](image)

**Figure 3**: Schematic of the conduction energy band profile for a GaAs/AlGaAs QWIP under zero (above) and finite (below) bias.
efficiency. T2SLs can produce a bandgap that is smaller than constituting semiconductors, due to the staggered (type II) energy band lineup of these layers. However, this band lineup causes low overlap integral between the electron and hole wavefunctions, and hence a low IR absorption. In recent years, T2SL has been further improved by a current blocking layer which has larger energy bandgap than the surrounding SL layers [56, 57]. An example of this new band diagram can be seen in Figure 4.

Most of the blocking barrier consists of aluminum incorporated into either the InAs or GaSb layer to form the band structure period of the SL either like a W or an M shape. However, the precise position of the barrier in the middle of the well is needed to increase the number of energy states available at the desired energy level [57, 59]. Razeghi’s group [58, 60–62] showed a blocking barrier T2SL with detectivity as high as $1.05 \times 10^{12}$ cm Hz$^{1/2}$/W at 150 K in the MWIR regime and then a $320 \times 256$ pixel FPA was constructed using this design, with imaging possible up to 170 K.

### 3.2 QD-based IRPD

The desires for higher operating temperatures, higher efficiency, and a lower cost IRPD has given birth to even lower quantum dimensionality than QWs. QDs as show in Figure 5, have zero dimensionality, have been incorporated into different material systems to enhance the efficiency and higher operating temperatures. Here we reviewed the two popular types of QDIPs based on III–V material (InGaAs/GaAs and InAs/InP) [64–66] and silicon-based material (Ge/Si) [67–69].

#### 3.2.1 QDIP based on III–V material

In 1998, the first QDIP based on intersub-level transitions was demonstrated along with the first QD laser [70]. QDIPs are similar to QWIPs but with extra advantages due to the three-dimensional (3D) confinement in the QDs compared to the 2D confinement in QW. There are three major advantages of QDIPs compared to QWIPs: (1) QDIPs are intrinsically sensitive to IR irradiation at normal incidence due to the breaking of the polarization selection rule [71], (2) QDIPs have lower dark current than QWIPs because of weaker thermionic emission from the QDs with 3D quantum confinement of carriers [72], and (3) the discrete energy levels in QDs have no dispersion, which reduces the phonon scattering and can lead to longer carrier lifetime (>100 ps), and so higher operating temperatures [73,

---

**Figure 4:** Schematic diagram of W- and M-T2SL LWIR photodiodes. (A) Band profiles of W-T2SL [58] and (B) p-doped M-T2SL structure [59].

**Figure 5:** Schematic view of (A) the quantum dot structure and (B) the conduction band of the QD [63].
The detection mechanism in QDIPs is also based on intersubband transitions between the quantized energy levels of the dots and continuum states [63, 75]. However, QDIPs have lower absorption quantum efficiency due to the small fill factor area of QDs and large inhomogeneous broadening of the self-assembled QDs [63, 75]. This inhomogeneous broadening also makes fine tuning to specific wavelengths difficult, and leads to issues with strain and dislocations with increasing number of absorption layers, diminishing the absorption coefficient.

In the last few decades, Lin et al. [76] improved the performance of QDIP responsivity, detectivity, as well as demonstrating higher operating temperatures and lower dark currents by using sandwiched QD layers with Al$_{0.3}$Ga$_{0.7}$As current blocking barriers. The background-limited infrared photodetector (BLIP) of the improved QDIP was demonstrated to be as high as 250 K [77]. Lin et al. [78] showed that the QDIP can be further improved in regard to the responsivity and lower dark current by direct doping the QDs. Later on QDIPs with QW active regions, known as DWELL structures, were proposed to improve the detection wavelength tunability with InGaAs strain relief layers [79, 80]. Surprisingly, DWELL IRPDs turned out to have advantages such as enhancement of absorption, low dark current, and high responsivity [81–83]. DWELL IRPDs have been extensively explored by Krishna’s research group, and some of the main achievements are presented in their review articles [84]. They have demonstrated some impressive results including: (1) multiwavelength operation in the MWIR, LWIR, and VLWIR, (2) fine tuning operating wavelength, (3) tunable bias, and (4) higher operating temperatures, especially for the VLWIR response. A 320 × 256 LWIR FPA was developed and images were obtained at 78 K [84]; however, this FPA has low responsivity and quantum efficiency.

### 3.2.2 QDIP based on Ge/Si heterostructures

Germanium QDs can be grown on Ge/Si and Ge/SiGe/Si multilayer heterostructures using the effect of the self-assembly of semiconductor nanostructures during the heteroepitaxial growth of materials with a large lattice mismatch [85]. Similar to other materials, Ge QDIP improves the performance of the heterostructures in multiple ways, such as by allowing absorption of normally incident electromagnetic radiation, high (up to 1000) photoelectric gain, and lower dark currents in Ge/Si QDIP [86]. High temperature operation QDIP at mid-IR (MIR) has also been achieved, even up to room temperature [87]. However, the quantum efficiency and photocurrent of Ge/Si QDIP are far lower than those of the bulk material. Despite the low quantum efficiency, Ge/Si QDIP is still one of the most popular QDIPs because of its lower cost. This is due to well-established silicon technology for fabricating MIR devices with low defect content and a high spatial homogeneity of photovoltaic characteristics. Ge/Si QDIPs also have better thermal expansion coefficients, which can easily be matched to a silicon readout circuit so that the array size is not limited by the cooling-induced mechanical stresses of the entire structure.

### 3.3 Nanowires and nanopillar for IRPD

Semiconductor nanowires (NWs) exhibit unique photonic and electrical properties such as high sensitivity and AR due to their unique anisotropic geometry, and high surface-to-volume ratio. NW architectures can be bandgap engineered using compound semiconductor materials and current advanced fabrication technologies. Moreover, NWs of almost all the existing semiconductor materials can now be realized by either top-down or bottom-up approaches. Although a lot of work has been devoted to investigating NW IRPDs in the past few years [88–94], most of the reported NW IRPDs are limited in the visible and ultraviolet spectral regions [89–91], and little work is conducted in the IR region [92, 93]. III–V semiconductor NWs with narrow bandgaps are considered as promising candidates for constructing IRPDs. To date, IRPDs based on InAs, InP, InGaAs, InPAs, and InGaSb nanowires have been reported [88, 94–96]. The fabrication of NW IRPDs can be categorized into three types, namely, photoconductor, phototransistor (where photosensitivity is greatly enhanced by an additional gate bias), and heterostructure NW devices including homo-/heterojunction and Schottky photodiodes.

#### 3.3.1 Photoconductor NW IRPD

NW photoconductors are NW-based photodetectors where the material is sandwiched between two electrodes [97]. NW photodetectors based on III–V compound semiconductors are one of the most studied due to their excellent transport properties, easy fabrication, and wide tunability wavelength by alloy bandgap engineering. Also, their low capacitance leads to high operating speeds. For example, an intersecting array of InP NWs can have a fast photoconductive response (14 ps FWHM at
In recent years, InGaAs and InAs have been studied intensely. InGaAs has potential bandgap tunability from the NIR to MIR, as well as high electron mobility and small leakage current. Single-crystal In_{0.65}Ga_{0.35}As nanowires are reported to have a large responsivity of $6.5 \times 10^3$ A/W over a broad range from 1100 to 2000 nm [99]. The photoresponsivity of $5.3 \times 10^3$ A/W was reported in Schottky-Ohmic contacted InAs NW photodetectors as shown in Figure 6 at the wavelength around 1.5 μm [100].

II–VI material NWs, such as CdS, CdSe, ZnSe, CdTe and ZnTe, are typically used for visible light detection. CdS NWs and nanobelts are among the most studied photoconductors in group II–VI for visible and ultraviolet light detection [89]. In the visible-NIR range (400–800 nm), CdTe nanoribbons with p-type conductivity present significant photoresponse to irradiation with a high responsivity of $7.8 \times 10^4$ A/W [101].

3.3.2 Phototransistor NW IRPD

A phototransistor is a bipolar or unipolar transistor in which light is converted to carriers and the current can be amplified through transistor action, leading to much greater photosensitivity. However, most of the NW phototransistors today have surface recombination loss, as carriers recombine before they can be collected. This problem can be solved by a core/shell design. A good example is an InAs NW phototransistor with a self-assembly photogating layer containing randomly distributed lattice defects as a trapping layer to capture photoexcited electrons [102]. This core/shell-like n-InAs NW phototransistor has a very large photoconductive gain of $10^5$ and a fast response time of 12 ms at the wavelength of 1.2 μm [102]. Single GaAsSb NWs have also shown a decent responsivity of 2.37 A/W with a low operating bias voltage of 0.15 V achieved at the wavelength of 1.3 μm [103]. These GaAsSb NWs were grown in a self-assembly fashion using a horizontal flow metalorganic vapour phase epitaxy reactor at 100 mbar.

3.3.3 Heterostructure NW IRPD

A lot of promising heterostructured NW architectures as shown in Figure 7 adopted methods from thin-film compound semiconductor photodetector technologies such as homogeneous, inhomogeneous, Schottky junction photodiodes, and avalanche photodetector. The p-n/p-i-n junction is normally used for homo/heterojunction NW photodiodes which operate under reverse bias. A single NW of InAs/InAsP axial heterojunctions is reported to have very low dark current, strong polarization dependence, and high photoconductivity when operated at 77 K [108]. Another example of NW heterostructures is core-shell NW junctions which attract a lot of interest due to high efficient photosensing and high light absorption enhancement [109]. For IR detection, GaAs systems are the most widely investigated. Examples include GaAs/AlGaAs core-shell [104, 110–113] and GaAs/InGaP/GaAs core-multishell NWs [114]. Among them, an NIR photodetector based on a core-shell GaAs/AlGaAs NW exhibited a peak photoresponsivity of 0.57 A/W at room temperature, comparable to large-area planar commercial GaAs photodetectors, and a high detectivity of $7.2 \times 10^{10}$ cm Hz^{1/2} W^{-1} at $\lambda = 855$ nm [115].

Another example of heterostructure NWs is Schottky photodiode structures [usually the construction is a metal semiconductor metal (MSM) photodetector], which exhibit fast response speed [105]. Arguably, the most promising NW photodetectors are the ones based on avalanche...
multiplication. NW avalanche photodiodes (APDs) with diverse configurations have been demonstrated. These include the “crossed” n-CdS/p-Si NW heterojunction [116], axial p-i-n single Si NW homojunction [116], InAsP QD in the axial InP p-n junction [117], radial GaAs nanoneedle junction [107], and radial GaAs p-n junction [118]. GaAs-based APD GaAs nanoneedles exhibit an excellent multiplication factor as large as 29 at −2 V and 263 at −8 V [118].

3.3.4 Si/Ge NW IRPD

Si/Ge NW IRPDs naturally have indirect bandgap which limits it from the high optical absorption application. However, by reducing their dimension, one can increase the quantum confinement and change the band structure. Also, impurity doping can lead to better performance photodetectors. A detailed review of Si/Ge NWs and their applications has been published recently by Ray et al. [119]. However, most of the works on Si/Ge NWs are focused at visible wavelength and very limited applications on NIR and MIR. The most notable work on Si NW in the NIR region with a high responsivity of $2.5 \times 10^4$ A/W in NIR wavelength (900 nm) at a zero bias condition has been reported recently [120]. The Si NWs with 80–100 nm diameter were fabricated by metal (Ag)-assisted electroless chemical etching techniques. Further enhancement of the Au NP-decorated Si NW has also been reported by Jee et al. [121]. The Au NP-decorated Si NWs achieved 63 times enhancement at the wavelength of 1000 nm, but the reported responsivity is around 0.1 A/W.

3.4 New emerging nanomaterials for detection

Graphene was the first 2D material which attracted the attention of many scientists due to its appealing features for nanophotonic applications originating from its zero-bandgap nature [122]. With linear dispersion near the Dirac point [123] and various forms of light-matter interaction [124], graphene offers a high optical responsivity in a broad spectral range. However, graphene has low absorption as an IRPD, which is due to its short carrier lifetime.
and its zero-bandgap nature. This gives rise to interest in another family of 2D materials – the single-layer TMDCs [126], such as molybdenum disulfide (MoS₂) and tungsten diselenide (WSe₂). Unlike graphene, the TMDC family has indirect-to-direct bandgap transitions that occur when the material thickness decreases from multilayer to monolayer. Moreover, TMDCs can easily be incorporated into a wide range of heterostructures with atomic-level control to achieve a higher performance IRPD.

### 3.4.1 Graphene-based photodetector

Graphene is a promising material for photodetectors compared to conventional semiconductors due to ultrahigh mobility, making it suitable for high-speed communications [127–130]. Single-atomic-layer graphene is stable, low cost, easy to fabricate, and has high internal quantum efficiency [131] when used in the NIR and MIR regions for optoelectronic applications. However, single-layer graphene is an inherently weak light absorber, which originates from its short interaction length [132–134], and the narrow effective area of lateral graphene p-n junctions also limits the efficiency of photocarrier extraction [135–137]. In other words, the applications of graphene-based photodetectors are limited by the lower external quantum efficiency and photoresponsivity in comparison to traditional photodetectors [130].

Various approaches have been proposed to enhance the sensitivity, including the introduction of a bandgap or an electron trap layer, such as building graphene QD-like structures [138] or cutting graphene into nanoribbons [139]. However, these bandgap opening methods degraded the electronic performances of graphene, especially the high mobility, thus leading to a decrease in the photoresponse speed [138]. Unfortunately, the overall quantum efficiency of graphene is quite low compared to conventional materials such as III–V semiconductors, due to the nature of zero bandgap structures. Figure 8A summarized the responsivity versus the response time for detectors based on difference 2D material, and comparison with commercial silicon and InGaAs photodiodes. Figure 8B also showed the wavelength dependency of different 2D material use in photodetector.

### 3.4.2 TMDC-based photodetector

Recently, another family of 2D materials with far better optical properties has emerged. Single-layer TMDCs [145], such as MoS₂ and WSe₂, have been intensely researched. The direct bandgap of monolayer TMDCs leads to efficient light emission covering the energy range from below 1 eV to well above 2.5 eV [153] and beyond, making them promising for a wide range of optoelectronic devices. Bandgap tunability with the layer thickness is another important property of this material. The bandgaps of trilayer, bilayer, and monolayer MoS₂, determined by photoluminescence, change from 1.35 to 1.65 and 1.8 eV [154], respectively. The bandgap of other typical TMDCs, such as WSe₂, MoSe₂ [146], WS₂ [147], and GaTe [148], also increases with decreasing layer thickness, due to the quantum confinement of carriers in the direction normal to the 2D plane. 2D TMDC photodetectors exhibit decent responsivity from the IR to the near UV range. The first reported monolayer MoS₂ photodetector [149] exhibited a responsivity of 75 mA/W in the visible range. Choi et al. [150] demonstrated a responsivity of ~100 mA/W by using a multilayer MoS₂ photodetector. However, the speed of photoresponse is relatively slow (ranging from microseconds to seconds) owing to the trapping of photocarriers. It is also noticeable that most of the high-sensitivity TMDC-based photodetectors reported above are for visible and NIR application. However, the recently re-discovered few-layer black phosphorus (bP) is an interesting material for photodetection for IR and far infrared (FIR), especially due to its intermediate bandgap between graphene (i.e. zero bandgap) and TMDCs. The study by Youngblood et al. [151] shows an intrinsic responsivity of up to 135 mA/W and 657 mA/W in 11.5-nm- and 100-nm-thick devices, respectively, at room temperature and at a frequency of 3 GHz. The few-layer bP is embedded within an on-chip waveguide structure with a few-layer graphene top gate, allowing for optimal interaction with light and tunability of the carrier density. The bP-based device presented in [151] stands out from the rest by showing comparable performance to both a commercial silicon photodiode and a graphene-based detector. Despite the large amount of funding and research invested on 2D materials, there is currently very limited set of 2D material covering the IR region of the spectrum [148]. This might represent an important future research area, considering the strong need for conformal and thin IR detectors.

### 3.4.3 2D material-based heterostructure IRPD

As discussed in the previous section, conventional graphene and TMDC-based phototransistors have relatively poor responsivity (~10⁻² A/W) owing to their weak optical absorption. Recently this limitation has been addressed by 2D material synthesized with other semiconductor
material, nanostructured or different 2D materials. For example, MoS$_2$ nanocrystal on silicon on insulator structure [155, 156], MoS$_2$/bp heterojunctions [157], graphene/MoS$_2$ heterostructures [158], graphene/MoTe$_2$ heterojunctions, and WSe$_2$/MoS$_2$ heterojunction photodetectors [157] show significant improvement in responsivity to the visible and NIR regions. A recent review paper on different dimensional material heterostructures [159] summarizes the works on different dimensional material for increasing the responsibility of the 2D material in the IRPD. One of the most notable works is QD on the graphene photodetector, where the QD optical absorption leads to a large photoexcited charge carrier in the graphene channel, and added with the high mobility of graphene, the QD/graphene heterojunction photodetector achieves a massive responsivity exceeding 10$^7$ A/W in the NIR region [160]. However, most of the mixed-dimensional heterojunction photodetectors discussed above have a relatively slow response (10$^5$–10$^{-6}$ s) due to their enhanced capacitance and charge traps. Nonetheless, there are a growing number of active research studies on faster IRPD devices with lower responsivity based on 2D heterojunctions. For example, single walled carbon nanotubes/MoS$_2$ p-n heterojunction photodetectors show responsivity in the visible and NIR with a response time of 15 μs [161]. Similarly, field-effect transistors based on the heterojunction between a monolayer of WS$_2$ and PbS CQDs demonstrated a photoresponsivity of ~14 A/W, while exhibiting a fast photoresponse time of ~153 μs. Most of such 2D heterojunction-based devices have the potential to become low-cost and high-performance IRPDs, while significant future research is needed to the commercialization stage.

3.4.4 CQD

In the past decade, CQD as shown in Figure 9 has emerged as one of the most appealing potential nanomaterials for...
IRPDs. CQDs show quantum confinement and are commonly made of II–VI, III–V, and IV–VI semiconductors by means of inexpensive and scalable wet-chemical synthetic procedures [163, 164]. Photodetection by CQD is based on the conductivity of CQD solid films. The first generation of CQD light detectors focused on photoconduction [165], where photogeneration of an electron-hole pair is followed by capture of one type of carrier (can be the electrons or holes), while the other remains substantially free to travel. Most of the CQD photoconductors reported until now are for the NIR to MIR regime, and based on the lead chalcogenide nanocrystals [166, 167]. The speed of these CQD photoconductor devices depends on the carrier lifetime. Very large photoconductive gain, in the range of $10^2$–$10^6$, can be achieved [166]. Photodetectors based on CQDs exploit tuning of the band edge to suitably chosen spectral cutoff energy, enabling dark current minimization for a given desired spectral response. The dark current is amplified via the same gain mechanism as photocurrent. However, it could be minimized by using photo-field-effect transistors [168]. The transistor introduces a threshold, so that turn-on occurs only after a certain photocarrier density (hence light level) is exceeded. PbSe CQD field effect phototransistors based on metallic (Au/Ag) nanowire transparent electrodes with a high responsivity of $2 \times 10^4$ A/W and a high specific detectivity of $7 \times 10^{10}$ Jones at low operation voltage (~1 V) in the NIR have been reported [169].

The speed of CQD detectors has been historically an issue. The extended lifetimes of trapped photocarriers, which are the source of slow response, can be controlled via the CQD surface chemistry and surface passivation [167]. Extracting photocarriers by using the built-in field in a photodiode provides a ready path to achieving efficient photocurrent generation; when the excited-state lifetime is typically ~100 ns, a built-in or applied potential on the order of 1 V can achieve a drift-based extraction time comparable with the excited lifetime when the lower of the electron or hole mobility exceeds $2 \times 10^{-3}$ cm$^2$/V s in the MIR regime [170]. Similarly, in the MIR regime, the CQD photodiode based on monodispersed HgSe CQD photodetectors that are illuminated with radiation resonant with intraband electronic absorption between 3 and 5 μm is reported. The HgSe CQD photodetector achieved the lowest dark current of 80 μA, but with a relatively small detectivity of $8.5 \times 10^8$ Jones at 80 K [171].

The maturity of the CQD in imaging has been much faster compared to other nanostructures. PbS CQD-based FPAs and camera prototype devices have been demonstrated, with a wide range of array and pixel sizes. The imaging arrays display a spectral response from 400 to

Figure 9: An image from the PbS CQD-based focal plane array (FPA) camera operating in the SWIR band. The CQD-based camera can see through a silicon wafer. The first inset shows the image taken using a normal visible camera. Second inset image is the PbS CQD-based focal plane array (FPA) structure [162].
bump-bonding of the sensor chip to the ROIC chip. This approach could in principle produce a high-yield and low-cost alternative to the current indium (ROIC) chips. However, they offer a very attractive low-cost route, since they can be spin-coated on top of silicon readout integrated circuit (ROIC) chips. These FPAs display multispectral sensitivity with detectivity comparable to InGaAs devices that currently dominate the commercial market. However, they have significant limitations due to the change of reflectivity with the incident angle. This is a rather fundamental limitation in thin-film interferometric filters, since the wavenumber (k-vector) parallel to the films is preserved throughout all layers. Another related issue is the narrow spectral response band of this approach, which renders such AR coating layers ineffective in modern “multi-color” and “multi-band” IRPDs. Antireflective condition can be generated by forming SWSs in front of the detector [174]. Such structures are also denoted as high spatial frequency structures or subwavelength surface textures [175]. SWSs generally have a two-fold function in light management in photodetectors: since they effectively represent gradient-index structures, they produce a broadband impedance matching that leads to a broadband AR performance. SWSs can also be designed to function as efficient mode couplers that couple the incident mode to many modes inside the high-index detector volume, and with a diverse polarization. This property can be used to couple to material with specific polarization, such as QWIP. The optical behavior of SWSs can be satisfactorily approximated in a majority of situations by using the effective medium theory [176].

SWS structures can be formed randomly across the surface or as periodic arrays. Random SWS structures give broadband antireflective properties, but repeatability is an issue. On the other hand, periodic SWSs can be designed based on required wavelengths, but usually the fabrication is more complex and costly. Random SWSs can be formed by self-assembled nanoparticles to perform nanosphere lithography by reactive-ion etching and obtain patterns of nanocavities on interface material [177]. For example, a cone-shaped Si nanostructure can reduce the reflectance (a planar Si) from 30% to 3% by using inductively coupled plasma – reactive ion etching at the NIR wavelength (800–1000 nm) [178]. This irregular structure is also referred to as random surface corrugations. Similarly for the periodic SWS, periodic structure lithography (holographic lithography, e-beam lithography) is performed followed by chemical or physical etching. SWSs offer important advantages over interference AR coatings. Examples of period SWSs include moth eye-like structures, 2D pyramidal, and also nanopillars. Figure 10 shows that SWSs can effectively enable fabrication of any desired effective refractive index profile [179]. These SWS nanostructures can efficiently reduce the
reflection of the surface leading to higher absorption in a wide range of wavelengths including the NIR and MIR regions [180].

4.1.2 Resonant cavity structure

One the most commonly used light-trapping structures in IRPDs is the resonant cavity structure (Figure 11).

The absorption within the detector can be significantly increased by placing the active area into a resonant cavity between resonator mirrors (usually consisting of a highly reflective mirror and a distributed Bragg mirror). 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 [182]. The resonant cavity enhanced (RCE) photodetector is usually known as RCEPD.

Figure 10: Predicted transmission of MWIR light propagating from air through an AR texture into silicon for various profile structures [179].

Figure 11: 3D schematic illustration, top view microscope image and cross-sectional scanning electron microscope image of the fabricated RCEPDs [181].
The amplification of the radiation at the resonant wavelength in an RCE device and rejection of others results in an increased spectral selectivity of resonant detectors. However, the RC structure also decreases the speed of the detector [183] which becomes limited by carrier transit time. Another disadvantage is an increase of the allowed operating temperature because of the shift of the BLIP limit [184]. RCE structures were used to improve performance in almost all IRPD systems such as p-i-n diodes [185], QW [181], T2SL [186], QD [187], quantum dot in quantum well [188], and light emitting diodes [189].

4.1.3 PC structure

Research on PC structures with a periodic refractive index modulation has opened up several ways to control light. Most existing devices are realized as 2D PC structures, as they are compatible with standard semiconductor processing. PC behaves as a high reflection coefficient mirror for all wavelengths within the photonic bandgap (PBG), and the waves with such wavelengths are evanescent within it. An important property of PC is defect modes, i.e. the allowed energy modes within their PBG which appear when the symmetry of the PC is broken [190, 191].

The use of a PC cavity to improve characteristics of IRPDs was described in [192]. The photodetector is enclosed within a PBG structure with a PBG either equivalent to the electronic bandgap of the utilized semiconductor material or larger than it. A defect may be introduced in the PC 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 photoresponse of the PCS-QWIP (Figure 12) shows a wider response peak but additionally displays several pronounced resonance peaks [193].

4.2 Nanostructures for enhanced light-matter interaction

A field of research that is currently exponentially growing is light-matter interaction structures such as plasmonic and nanoantenna [194]. Dielectric or metallic nanostructures can enhance the absorption cross-sections of the...
photosensitive material, typically by electromagnetic fields localized within subwavelength areas. Unlike cavities discussed in the previous section, plasmonic and metallic nanoantennas as shown in Figure 13 provide extreme field concentration and strong light-matter interaction across a broad spectral band. They have been intensively investigated due to the large enhancement on the absorption of thin layer IRPDs such as QWIP, graphene, and other 2D material.

4.2.1 Plasmonic enhancement in IRPD

Surface plasmon-mediated light-trapping schemes for photodetectors [199, 200] may be divided into two distinct groups; namely, (1) localized surface plasmon consists of plasmonic nanoparticles or nanovoids, and (2) surface plasmon polariton consists of different diffractive structures [197]. The former includes field localization and generation of hotspots [195], plasmon-based singular optics [199], and metamaterial-based transformation [196]. The latter includes structures such as gratings [198] for field coupling into guided modes and subwavelength plasmonic crystals that may be periodic [201] or quasi periodic [194].

Among subwavelength PCs described within the context of IR photodetection, an important position belongs to 2D arrays of nanoapertures in opaque metal film. Such structures first drew attention for their ability to transmit light in spite of the dimensions of nanoapertures being much smaller than the operating wavelength. This effect is commonly known as the extraordinary optical transmission (EOT) [198]. EOT is a result of the excitation of surface plasmon polaritons at the metal surface. It forces the electromagnetic waves incident to the surface to pass through the apertures. The transmission couples all incident electromagnetic energy on the surface to the mode impinging through the aperture. Thus, large field localizations appear in the apertures. This behavior effectively corresponds to impedance matching between propagating waves and the perforated metal film.

Interesting structures of these types similar to Figure 14 have been developed, to enhance the weak optical absorption of graphene. Compared to traditional structures including cavities [205, 206] and waveguides [202, 207], these metal plasmonic structures [203, 204] can

![Figure 13](image.png)

**Figure 13:** Different types of plasmonic structure: (A) bull’s eyes structure [195], (B) plasmonic arrays [196], (C) nanoparticles-based plasmonic photodetector [197], and (D) single slit EOT structure [198].
enhance the light-matter interaction over a broad spectral band. A planar microcavity based on graphene sandwiched between two highly reflecting mirrors achieved a 26-fold enhancement in the absorption of graphene with a responsivity of 21 mA/W [205]. These are also used for other 2D materials [202, 207]. The third approach is to enhance the local electrical field, which is guided into the region of the graphene photoactive area using plasmonic metal structures, hence increasing the light absorption [203, 204]. This field enhancement is proven to dramatically improve the performance of graphene-based photodetectors. The graphene photodetectors can couple with metal nanostructures at different resonance frequencies, and the photocurrent and external quantum efficiency could be enhanced by 1500% [204].

**4.2.2 Optical antenna**

Nanoantenna or optical antenna [194] is a structure that couples the propagating light waves and the evanescent (near) fields. The large change of the photon momentum is the unique property of such devices. Depending on the nanoantenna type and its design, the mode confinement in the evanescent field can be deeply subwavelength. The simplest and the most basic nanoantenna is a nanosphere. This structure actually behaves like a dipole and its scattering properties can be calculated using the Mie theory [208]. Much larger field localizations are obtained using two identical nanospheres (known as dimer) or in an array (nanoparticle chain) [209]. Many other shapes can be used as nanoantennas such as nanorods, two-wire [210] bowtie [211] diablo-type nanoantenna [212], the Yagi-Uda, group of director nanorods [213], spiral nanoantennas [214], and those with fractal geometries [215]. Many other shapes can be used.

In 2010, our group reported an ultra-thin QWIP integrated with a nanohole-array optical antenna (Figure 15) [217]. The optical antenna captures far-field light and enhances its concentration at the active region of QWIP. The polarization selection rule in intersubband QWs prevents absorption of normal-incident light. The nanohole-array antenna can efficiently convert the polarization of normal-incident light to a perpendicular polarization by
exciting the surface plasmons on the surface of the perforated metal. The peaks of responsivity and detectivity both occur at 8 μm, reaching 7 A/W and $7 \times 10^{10}$ cm Hz$^{1/2}$/W, respectively. Such a high responsivity and detectivity is due to a high quantum efficiency (QE) of 40%, which is an order of magnitude larger than the QE of a plain QWIP structure with a similar thickness of ~λe QE.

Several similar structures were reported since then, including a plasmonic optical antenna coupled with a LWIR QDIP [218]. The antenna consists of a pair of half-wavelength dipole antennas with a 20 nm gap in between. E-field enhancement of more than 30 times is obtained in the gap, corresponding to over 900 times light intensity enhancement. This LWIR QDIP shows strong incident-angle-dependent photocurrent enhancement with a clear antenna-induced detection pattern (i.e., directivity).

In the last decade, high-gain and high-directivity IRPDs have been intensely studied. Optical antennas coupled with metal-oxide-metal (MOM) diodes have been used to demonstrate beam steering in the MIR (10 μm) [219]. The devices show narrow beam widths of ~35° FWHM in power, and reception angles of ±50°. Two dipole antennas coupled to coplanar strip waveguides, inspired by similar systems in microwave electronics [220], are coupled to a rectifying MOM diode. Due to the interference between the two antennas, the diode can be placed off-center and “see” off-normal angles. Recently, a double-metal nanoantenna as microcavity has been fabricated on a QWIP detector to confine light in 3D [221]. Strong subwavelength light confinement of this microcavity antenna leads to three times reduction in dark current due to the reduced detection area. The structure achieved

**Figure 15:** An ultra-thin QWIP integrated with a nanohole-array optical antenna. (A and B) SEM images of nanohole array optical antenna integrated on a QWIP photodetector. (C) The measured responsivity of the device at bias 0.7 V at 78 K [216].
a peak responsivity of 0.074 A/W photoconductive gain $g = 0.25$ at a wavelength of 9 μm at all incident angles.

Recently, our group proposed a novel metallo-dielectric hybrid antenna, with a metallic cavity antenna coupled to a photonic jet produced by a dielectric microsphere. This structure as shown in Figure 16 has a high QE of about 50% and a high directivity of about 20 dB. The quantum efficiency and the directivity are significantly better than the bowtie optical antenna with ~1% QE and poor directivity gain (a few dBs) [222]. When integrated with QWIP, the detectivity of the hybrid antennas is more than 30 times higher than the limit of conventional QWIP detectors at all temperatures due to the strong directivity gain. However, the optical antennas cannot increase the detectivity of the interband detector beyond the limit of conventional devices, because of their comparable radiative and non-radiative lifetime. Another similar optical antenna integrates a split-ring bull’s eye antenna on a germanium detector to focus light from a relatively large area into a subwavelength germanium slab. The measured photocurrent is improved by seven times, with a responsivity of 3.93 mA/W at 1310 nm [223].

4.2.3 Metamaterial structures

Subwavelength dielectric and metallic structures made in 1D, 2D, and 3D arrays have created a novel class of structures known as electromagnetic metamaterials [224, 225]. They could be defined as artificial structures whose optical electromagnetic properties cover regions beyond what is found in nature. It is possible to obtain effective relative dielectric permittivity, relative magnetic permeability, and refractive index in a vastly expanded range by tailoring their geometry and material properties (example as shown in Figure 17).

For example, negative index metamaterials are combined with dielectrics to produce a broad range of new device properties [216, 226–229]. Metamaterial-containing multilayers exhibit many peculiar properties, with both theoretical and practical interests. For example, it is possible to fabricate resonant cavities with subwavelength dimensions [230]. Such structures might be used to reduce the volume of the absorber, and hence the dark current/noise, while maintaining a high QE. Another application that is related to IRPDs is a multilayer metamaterial based on the so-called double fishnet structure with enhanced IR absorption and a relatively large directivity [231]. Compared with the aforementioned photonic jet approach, this method is all planar, and uses standard micro/nano-fabrication methods. A persisting problem with such metamaterials is their large absorption losses – specifically in their negative refractive index range, as well as relatively narrow bandwidths due to their resonant nature.

5 Advance application of IRPD

In the past, the main application for IRPDs was for surveillance and warning for military or Department of Defense. Currently, the applications of IRPDs are widely extended to different industries such as automotive, medical, semiconductor, food, and also astronomy. In the past, night vision and surveillance systems were the main application of IRPDs [232]. However, IR cameras have been adopted in various industries such as automation, automotive, civil engineering, and food production in the modern society due to their non-contact

![Figure 16: Schematic of the hybrid antenna and FDTD simulation cross-sectional view of the metallo-dielectric micro cavity. The right figure shows the optical power flow (arrows) and normalized power consumption density (optical absorption density) throughout the cavity at the operating wavelength $\lambda \approx 8 \mu m$ [222].](image-url)
precision temperature measurement and non-destructive monitoring. IR cameras have been used for food quality control [233], semiconductor chip and device control [234], paved road condition control, and various other monitoring activities (http://www.flir.com/home/). As for medical application, various medical tools such as blood vein monitor [235], brain trauma detection [236], diabetic neuropathy [237], thermography, and cancer cell detection [238] are invented using IR technology due to the non-contact and non-destructive monitoring (http://medcitynews.com/2013/02/5-innovative-uses-for-infrared-technology-from-early-bedsore-detection-to-preventing-surgical-errors-video/). However, a lot of the medical tools are in the testing phase and many more...
will be available in the industry within the next decade. One of the most interesting applications for the IRPD camera is astronomy. A review of Figure 1B shows that a lot of molecules can be detected using their unique IR signature. Molecules such as $\text{H}_2$, $\text{O}_2$, $\text{CO}_2$, and water are essential for life [240]. Based on the IR detection from the light from exoplanets, we can potentially determine whether they are habitable or not. This is one of the most important missions in astronomy. However, up to now most highly sensitive IRPD cameras required cryogenic cooling, which makes the camera bulky and increases the maintenance cost. The key emerging factors for future IRPD technologies are, namely, (1) the possibility of making large imaging FPAs, since most applications require arrays of detectors, (2) room-temperature or TE-cooled imaging systems, since most of the current sensitivity IRPD cameras require cryogenic cooling, (3) the ability to integrate with silicon microcircuit technologies to make more compact and low-cost systems, and (4) possibility of mass production, so that the cost of the IRPD devices or cameras can be lowered.

### 6 Conclusion and future prospect

In conclusion, we have reviewed novel nanotechnology approaches for IRPDs, and compared them with the traditional solutions. The rapid progress of the IR photodetection in the recent history is partly due to the advances in micro-nano fabrication and technology. We anticipate that IRPDs will see increasing use in a number of fields, ranging from the military target detection to environmental sensing. Imaging technology has become one of the most important directions in IRPDs. Bulk detectors such as AlGaAs, HgCdTe, and SiGe have been widely commercialized due to ease of large-scale fabrication, flexibility in the absorption wavelengths covering almost all IR bands, and high responsivity at cryogenic temperatures. However, low dark current and high performance IRPDs at high temperatures is highly needed in many emerging application areas. Advances in nanofabrication enable QWIP and QDIP detectors to perform in the longer wavelength bands, and with a higher operating temperature than their bulk counterparts.

<table>
<thead>
<tr>
<th>Nanostructure IRPD</th>
<th>Advantages</th>
<th>Disadvantage</th>
<th>Possibility for FPA fabrication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantum well structure</td>
<td>– Mature GaAs growth and fabrication process&lt;br&gt;– High uniformity&lt;br&gt;– Low cost&lt;br&gt;– Covers MWIR to WLWIR and THz</td>
<td>– Low quantum efficiency&lt;br&gt;– Cannot absorb normal incident angles&lt;br&gt;– Poor performance at elevated temperatures</td>
<td>– Mature technology available for IR FPA production</td>
</tr>
<tr>
<td>Type II superlattice structure</td>
<td>– High operating temperature&lt;br&gt;– Covers SWIR to WLWIR&lt;br&gt;– High absorption coefficient</td>
<td>– High fabrication and processing cost&lt;br&gt;– Low yield in large array fabrication</td>
<td>– Mature technology available for IR FPA production</td>
</tr>
<tr>
<td>Quantum dot</td>
<td>– Sensitive to infrared irradiation at normal incidence&lt;br&gt;– Lower dark current than QWIPs&lt;br&gt;– Higher operating temperature&lt;br&gt;– Longer carrier lifetime</td>
<td>– Low absorption quantum efficiency&lt;br&gt;– Large inhomogeneous broadening of the self-assembled QDs</td>
<td>– Research on going for IR FPA production</td>
</tr>
<tr>
<td>Nanowires and nanopillar</td>
<td>– Higher light sensitivity&lt;br&gt;– Antireflection and light trapping properties&lt;br&gt;– Can be integrated with CMOS technology</td>
<td>– Non-uniformity over large area fabrication&lt;br&gt;– Difficult to form large detector arrays&lt;br&gt;– Difficult to fabricate&lt;br&gt;– Repeatability issues</td>
<td>– Possible for IR FPA production</td>
</tr>
<tr>
<td>Graphene-based photodetector</td>
<td>– High internal quantum efficiency&lt;br&gt;– Low cost&lt;br&gt;– Ease of processing&lt;br&gt;– Ultrafast process</td>
<td>– Weak light absorption in a single layer&lt;br&gt;– Zero bandgap</td>
<td>– Possible for IR FPA production</td>
</tr>
<tr>
<td>Transition metal dichalcogenides</td>
<td>– Low cost&lt;br&gt;– High optical absorption coefficient&lt;br&gt;– Ease of processing</td>
<td>– Low speed&lt;br&gt;– Do not easily achieve LWIR and MWIR&lt;br&gt;– No proven technology for large area fabrication</td>
<td>– Not possible for IR FPA production</td>
</tr>
</tbody>
</table>
Moving forward, progress in the field of IRPDs may come from dramatically different sources. In recent years, graphene and other 2D materials have been intensely researched for higher operating temperature and lower cost IRPD production. However, these new materials suffer from a low absorption coefficient, a high sensitivity to the ambient, and a large area fabrication limitation.

One immediate direction for the 2D material might be to follow the CQD technology development path, which has settled on producing optically thick CQD layers capped with sealing layers. In the long run, however, these low-dimensional materials have the potential to surpass the performance of bulk semiconductors, since they allow intimate integration of new SWs such as plasmonic structures and metamaterial. As such, they hold the potential to achieve higher detectivity in the IR bands and at elevated temperatures. Table 2 summarized the advantages and disadvantages of different type of IRPDs discussed in this review paper, as well as the current status for the applicability to large area FPA fabrication.

References


Annual Conference of IEEE Engineering in Medicine and Biology, Shanghai, China, 2005.


