Reports on Progress in Physics

REPORT ON PROGRESS

Recent advances in infrared imagers: toward thermodynamic and quantum limits of photon sensitivity

To cite this article: Simone Bianconi and Hooman Mohseni 2020 *Rep. Prog. Phys.* **83** 044101

IOP Publishing Reports on Progress in Physics
Rep. Prog. Phys. **83** (2020) 044101 (27pp) https://doi.org/10.1088/1361-6633/ab72e5

Report on Progress

Recent advances in infrared imagers: toward thermodynamic and quantum limits of photon sensitivity

Simone Bianconi[1] **and Hooman Mohseni**[1,2]

[1] Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208, United States of America
[2] Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208, United States of America

E-mail: hmosheni@northwestern.edu

Received 17 April 2019, revised 17 December 2019
Accepted for publication 4 February 2020
Published 16 March 2020

Corresponding editor Professor Masud Mansuripur

Abstract

Infrared detection and imaging are key enabling technologies for a vast number of applications, ranging from communication, to medicine and astronomy, and have recently attracted interest for their potential application in optical interconnects and quantum computing. Nonetheless, infrared detection still constitutes the performance bottleneck for several of these applications, due to a number of unsolved challenges, such as limited quantum efficiency, yield and scalability of the devices, as well as limited sensitivity and low operating temperatures. The current commercially dominating technologies are based on planar semiconducting PIN or avalanche detectors. However, recent developments in semiconductor technology and nano-scale materials have enabled significant technological advancement, demonstrating the potential for groundbreaking achievements in the field. We review the recent progress in the most prominent novel detection technologies, and evaluate their advantages, limitations, and prospects. We further offer a perspective on the main fundamental limits on the detectors sensitivity, and we discuss the technological challenges that need to be addressed for significative advancement of the field. Finally, we present a set of potential system-wide strategies, including nanoscale and low-dimensional detectors, light coupling enhancement strategies, advanced read-out circuitry, neuromorphic and curved image sensors, aimed at improving the overall imagers performance.

Keywords: infrared imaging, sensitivity limit, photodetector, focal plane array, low light level, short-wave infrared

(Some figures may appear in colour only in the online journal)

1. Background

Observation of infrared radiation has historically been one of the most prominent ways to better understand different phenomena in our Universe. Infrared imaging however, has seen massive growth in number of applications during the past few decades. This rapid growth has fueled exciting technological developments that could spur new paradigm-shifting advancements in infrared detection and the numerous fields related to it. This review intends to focus on the most important recent progresses in the field. However, in light of recent controversies in proper measurement and formulation of photon

detectors (see for example [1]), this review also provides the key background information and formulations required for a proper evaluation of different technologies reported in the literature.

1.1. A short history of infrared detectors and imagers

The discovery and first detection experiments of infrared radiation date back to the early 19th century: the detection technology was initially entirely based on thermal detectors [2, 3]. The first photon detectors were developed in the early 20th century, and the years during World War II saw the birth of modern IR detector technology, initially employing materials such as Tl, Cs and PbS [4, 5]. Thanks to major interest from military funding agencies, the research in infrared photodetector saw a significant impulse throughout the rest of the 20th century: the most significant developments happened during the '60s with the introduction of III–V and II–VI based detectors, in particular HgCdTe [6, 7]. Several technologies were developed, based mainly on these materials systems, to span the whole wavelength range of infrared radiation (~1–30 μm). Most of the research effort was directed since the beginning to the fabrication of arrays of photodetectors (focal plane arrays, FPAs) and cooling systems that could be assembled into imagers, with significant achievement [2]. More recently, cryogenic and superconducting technologies have enabled novel detection processes and devices. Importantly, a few of the superconducting detection technologies could be successfully fabricated into imager arrays [8, 9]. While cooling to temperatures required for supercconducting IR detectors is possible, the limitations produced by the volume, power consumption, lifetime, and cost of the cooling systems have held back wide adoption and utilization of this technology. This limitation might be addressed by novel infrared imagers based on nanoscale and low-dimensional materials, which have been recently proposed and demonstrated for approaching room-temperature highly-sensitive infrared detection.

1.2. Recent attention on technological applications of the SWIR wavelength band

Infrared radiation has been employed in a myriad of different fields and applications, including military tracking and ranging applications, night vision, medical diagnostics, astronomy, spectroscopy, and non-destructive inspection. While attempting to maintain generality when possible, in this work we will focus on the recent developments for the detection of short-wavelength infrared radiation (SWIR), typically corresponding to 1–3 μm wavelength. This wavelength range is unique in that it avoids both the optical absorption and scattering related to electronic transition typical of shorter wavelengths, and that due to free-carrier and molecular vibration absorption, at longer wavelengths. As a result, optical transmission in the SWIR band is fairly high in a wide range of materials, spanning from biological tissues to fiber optics, and the Earth's atmosphere. Such fundamental advantage has attracted great interest from a wide range of different fields, and we are recently witnessing a dramatic increase in number of applications and technologies based on SWIR. Furthermore, as shown in figure 1(a), detectors operating in this band can typically reach higher performances thanks to a lower background noise, compared to longer wavelengths. Furthermore, unlike mid- (MWIR) and long-wavelength (LWIR), this field was not traditionally dominated by military-funded research, since it also represents, for example, one of the key enabling technologies for optical communication. In addition, this wavelength is of particular interest in astronomy, due to both the existence of an atmospheric transparent window around 1.5 μm, as shown in figure 1(b), and because it gives the best contrast (for thermal emission) and superior resolution, for imaging exoplanets [10–14]. Moreover, increasing interest has recently developed towards expanding the domain of optical communication to the chip level, currently dominated by electrical interconnections, in order to address one of the most crucial technological challenges towards decreasing computational power consumption [15–17]. As a result, multiple new devices, processes and systems centered around SWIR detection integrated on chips have been proposed, that have attracted great interest and funding from the whole electronics community [18–24]. Finally, thanks to its relatively low absorption in some human tissues, SWIR radiation has enabled the development of important optical coherent tomography techniques (OCT), for medical diagnostics [25, 26].

In this paper, we review the recent progress in SWIR detection technology, presenting the most mature technologies and their limitations, as well as the prospects for the most promising newly developed technologies, such as nanoscale and low-dimensional materials, and we attempt to present their requirements and future challenges in order to represent a viable alternative.

1.3. Current requirements, guidelines and trends in SWIR detectors and imagers

The current direction of research efforts and technological development for infrared imagers is aimed at achieving megapixel, low-cost, light and efficient imagers with advanced on-chip functionalities. As proposed by Rogalski *et al* [27] (figure 2), these goals characterize the development of the fourth generation of infrared imagers, started in recent years. Of particular relevance for several fields of interest is the concept of size, weight and power consumption (SWaP), as this is crucial to the portability, system complexity and suitability of the imagers [2, 28, 29]. This is intimately connected to the system requirements imposed by the imager of choice, such as optics size, cooling systems and cold shields [29, 30].

Especially for the case of SWIR and MWIR imagers, these development guidelines translate into four major design strategies which have been pursued during the last few decades [31]:

- Small pixels and optics
- High operating temperature
- Wafer-scale and novel process and integration
- Multi-functional pixels and read-out

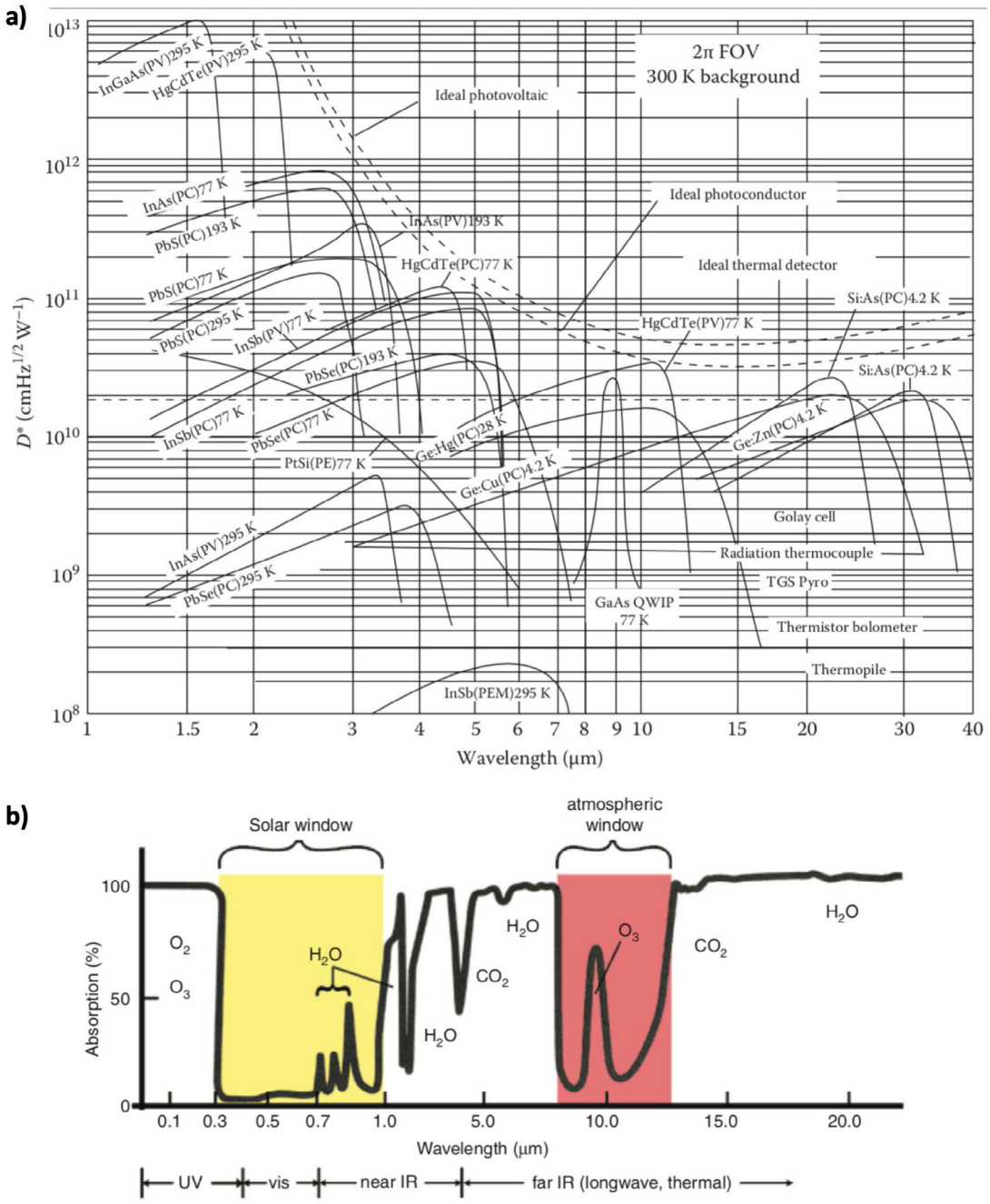

Figure 1. (a) Comparison of detectivity (D^*) of various commercially available infrared detectors operating at the indicated temperature. [2] Taylor & Francis Ltd. http://tandfonline.com. (b) The absorption spectrum of Earth's atmosphere. Reproduced with permission from [13]. © Oxford University Press.

1.3.1. Small pixels and optics. With the consistent improvement of the fabrication capabilities of both the FPAs and the integrated read-out circuitry, the size of the detector pixels has been constantly reduced over time [27, 32]. This enabled the number of pixels in a imager (for a given sensor area) to scale exponentially over time, similar to the Moore's law, allowing third-generation FPAs to reach Megapixel resolution (figure 2) [27, 33]. In addition, small pixel size and pitch can play a crucial role in reducing the SWaP of the system by enabling smaller detector size. For a fixed F-number ($F/\#$), the area, size and weigth of the optics required scale at least quadratically with the detector size [27, 30, 34]. As an example, the miniaturization of visible CMOS imagers, driven in particular by the mobile phone market, has enabled a huge improvement on the cost and portability of such devices, which have become ubiquitous [35, 36]. Most current commercial mobile phone cameras have pixel sizes ranging between 1–2 μm [37].

As the nanofabrication feature size has reached and surpassed the diffraction-limited spot size for all infrared wavelengths, a discussion has developed on the potential benefits and limitations of further reducing the pixel size to and below the diffraction limit [27, 30, 38]. Contrary to the traditional infrared imager design criteria, it has been theoretically proven that both long-range and large-field of view imaging modes can benefit from a diffraction-limited imaging system [30, 31, 34, 39, 40]. As a result, for optimal resolution and sensitivity

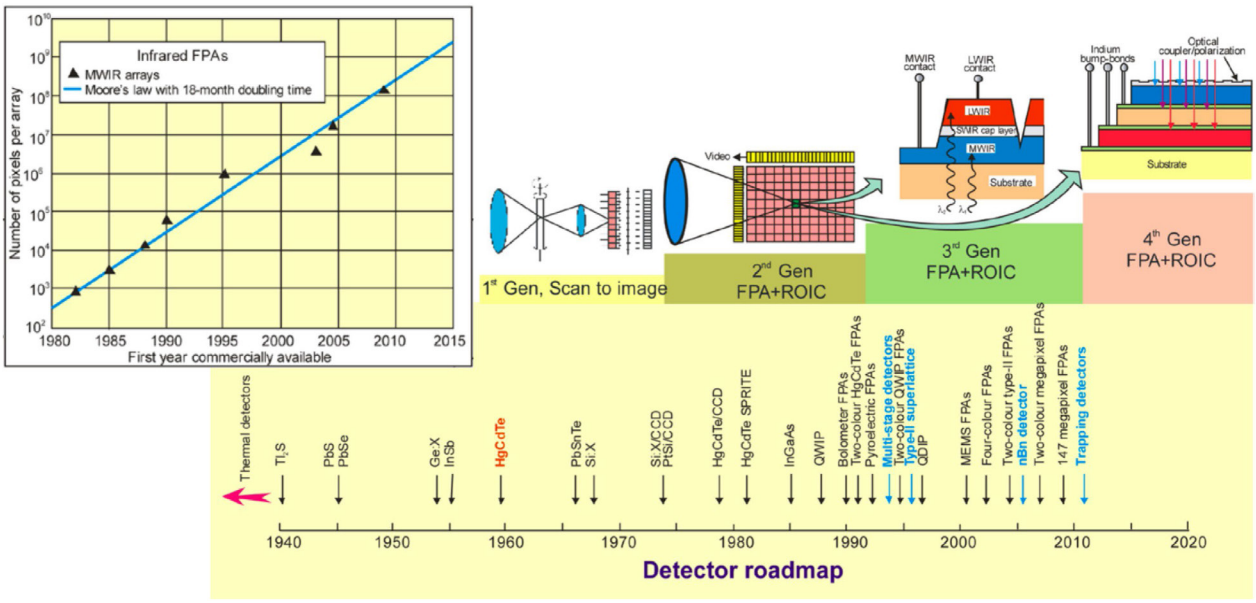

Figure 2. History of the development of infrared detectors and systems, with the four generations of infrared imagers, as proposed by Rogalski *et al* [27]. Highlighted in blue are the major novel detector concepts developed over the last three decades. The inset shows the evolution of the number of pixels in MWIR FPAs over time, which follows Moore's law exponential trend. Reproduced from [27].

in the next generation of infrared imagers, diffraction-limited pixel sizes of 3 and 5 μm have been proposed for the MWIR and LWIR wavelength range respectively. Extending these calculations to the SWIR band, it is evident that detectors sizes in the order of micron and sub-micron are necessary to achieve the advantages of diffraction-limited and sub-wavelength imaging. Finally, there are additional potential benefits to implementing sub-wavelength sized pixel detectors, which we will discuss in section 2.4.

1.3.2. High operating temperature. Typically, infrared detectors need to operate at cryogenic temperatures in order to suppress the thermally-activated noise mechanisms intrinsic to their narrow band gap. Since cooling system are expensive and bulky, increasing the operating temperature is another primary strategy for reducing both the SWaP and the cost of infrared imagers [41, 42]. In addition, higher operating temperature intrinsically enhances the performance of the imagers, such as by reducing the cooling-down dead time [27]. Finally, obviating the need for cryogenic cooling is crucial for extending the application of SWIR imagers to portable devices as well as for operation in remote areas [14, 41, 43]. The goal for high-temperature operation is generally set to 200–220 K, since this temperature range can be achieved by thermoelectric coolers, with the ultimate goal of room-temperature operation. The typical temperature ranges of operation for the most common SWIR detection technologies commercially available or developed in research environments are shown below [22, 44, 45]. Generally, SWIR imagers based on III–V technologies, such as InGaAs PIN and heterojunction phototransistors (HPT) can operate in a range of temperature that is fully covered by thermoelectric cooling, up to room temperature.

Technology	Operating Temperature
Superconducting	<4 K
HgCdTe	30–140 K
InGaAs PIN	210 K—room T
InGaAs HPT	200–240 K

1.3.3. Wafer-scale and novel processing and integration. Infrared imagers are considerably more expensive than their visible CMOS-based counterparts because they require high-quality materials with the appropriate narrow bandgap. These materials typically need to be deposited at high temperature and low pressure on lattice-matched crystalline substrates [46]. As a result, high-quality infrared detectors cannot be directly deposited on standard CMOS read-out circuitry, and therefore require some form of hybridization and integration, which adds up to a good portion of the total imager cost, beside the intrinsic material production cost. Most current commercial imagers are fabricated independently at the individual die level and then integrated with the CMOS read-out circuitry, typically by means of flip-chip bump bonding. This complex process prevents wafer-level parallelization. In addition, the large number of steps, and in particular bump bonding affect the yield of the process. It is for these reasons that the infrared community has increasingly been looking into developing wafer-scale processes that can drastically reduce the cost of production of infrared imagers, thanks to the economy-of-scale. Recently, funding agencies have started several programs dedicated to achieving wafer-scale production, such as DARPA's WIRED program [47]. Research efforts have been devoted to alternatives and improvements to the bump bonding hybridization method. As an example, large-area and

wafer-scale direct bonding of III–V semiconductors to Si and CMOS substrates has been demonstrated [18, 22, 24, 48]. Moreover, a variety of bottom-up fabrication techniques for infrared detectors, such as self-assembly, colloidal dispersion and low-temperature growth on multiple substrates have been reported [19, 49–52].

1.3.4. Multi-functional pixels and read-out. Another fundamental feature of the third and fourth generations of infrared imagers is the enhanced functionality at the individual pixel level. This includes pixel design features, such as multi-color imaging, as well as read-out level functionalities, such as dynamic reconfigurability of the detector operation mode. Multi-color pixels, consist of two or more detectors, each sensitive to a different spectral band, co-located and coinciding with the same pixel. Typically, the shorter wavelength detector layer is placed in front of the longer wavelength one, so as to act as a long-pass filter. Multi-color pixels have been demonstrated across all infrared bandwidths, including both sequential and simultaneous architectures [2, 29]. In the former case, all layers of the pixel share one contact, and the selection of sensing and read-out color is made by bias selection [32]. In simultaneous operation, each layer is contacted and read out individually. In this case, contact density will potentially represent a limitation to the scalability of these devices to sub-diffraction pixels, as discussed above [27].

'Smart' pixels with the ability to dynamically reconfigure their sensing mode can be enabled by novel design of the read-out (ASIC) [53]. These advanced functionality is highly desirable for implementing edge computing as well as enabling advanced imaging techniques such as foveation, i.e. simultaneous passive imaging and dynamic ranging in different regions of the imager [54]. Such smart pixels could control the integration times with the ability to switch between readout modes such as full-frame imaging and zero-suppressed data, modification of the number of pixel data bits and independent programmability for distinct detector regions.

1.3.5. Novel technologies and trends. In parallel with the research and development of commercially dominating imager technologies, several novel materials, processes and device designs have been proposed in recent years, to try to address or circumvent the challenges described above. Some of these have shown the potential to even surpass the performance of current state-of-the-art technologies, and revolutionize the market and research field of SWIR imagers. Among these, for example, a SWIR imager based on colloidal quantum dots was demonstrated in 2015 [52, 55], and similar devices, including multi-color SWIR-MWIR versions, have since entered the market as affordable alternative to the current commercially dominant HgCdTe and III–V cameras [56–61]. Nanowires and thin-film detectors that can be grown or deposited on CMOS substrates have also shown the potential to disrupt the current state-of-the-art SWIR imager market. Novel methods for the CMOS-integration, as well as the scaling and fabrication of imagers based on conventional semiconductors such as III–V and Ge have also seen a rapid growth of interest and funding. By looking at the research output of the last 20 years in the field of SWIR imagers shown in figure 3, it is apparent that the field has experienced a dramatic expansion, driven in particular by low-dimensional materials and III–V and Ge-based detectors.

2. Key detection parameters and performance limits

2.1. Key infrared detector parameters

A photodetector is a device that transforms a light signal (information) to a detectable electrical signal. One of the main performance metrics of photodetectors is therefore the responsivity, defined as the ratio of the generated electrical signal to the power of the incident radiation. Depending on the class of photodetectors, the signal can be based on either in terms of voltage, V_S, or of current, I_S, such that for each case, responsivity is defined as [2]:

$$R_V = \frac{V_S}{\int_{\lambda_a}^{\lambda_b} \Phi(\lambda)\,d\lambda}\left[\frac{V}{W}\right], R_I = \frac{I_S}{\int_{\lambda_a}^{\lambda_b} \Phi(\lambda)\,d\lambda}\left[\frac{A}{W}\right] \quad (1)$$

where $\int_{\lambda_a}^{\lambda_b} \Phi(\lambda)\,d\lambda$ is the radiant flux within the wavelength range of interest $[\lambda_a, \lambda_b]$, in units of W. The main task that efficient photodetectors need to fulfil is generating a large enough signal to overcome noise: the level of radiation that is sensed by the detector generating a signal output equal to the noise level is called noise-equivalent power (NEP). Since the noise level depends on the bandwidth (BW) of the photodetectors, NEP is typically reported at the standard BW = 1 Hz, so that the NEP is measured in $W\sqrt{Hz}^{-1}$ [2]. The detectivity of a photodetectors is then defined, as $D = \frac{1}{NEP}$; however, since the noise related to a photo-induced amount of charges absorbed over a detector area A_D, and bandwidth BW, is proportional to their square root, $NEP \propto \sqrt{A_D\,BW}$, a normalized detectivity, D^* is usually preferred [2]:

$$D^* = \frac{\sqrt{A_D\,BW}}{NEP} = \frac{\sqrt{A_D\,BW}}{\Phi}SNR \left[\text{Jones} = \text{cm}\sqrt{\text{Hz}}\,\text{W}^{-1}\right] \quad (2)$$

where the definition of signal-to-noise ratio, $SNR = \frac{\Phi}{NEP}$ has been utilized, with Φ radiant flux in W.

All noise sources in a detector can be classified as either electronic or radiation noise. The radiation noise is fundamental: it does not depend on the detector materials or design, as it is related to the quantized nature of light. The radiation noise is either intrinsic to the detected light signal, or can come from the background radiation: in either case, the noise is quantifiable as the square root of the incident photon flux density in the signal or background, respectively. The electronic noise, conversely, has a wide range of sources, including imperfections in the detector, electronic circuitry for readout and amplification, etc. The ultimate limit in performance of an IR photodetector is when all sources of noise are negligible compared to the background noise; condition known as background-limited infrared photodetection (BLIP). In BLIP condition, the

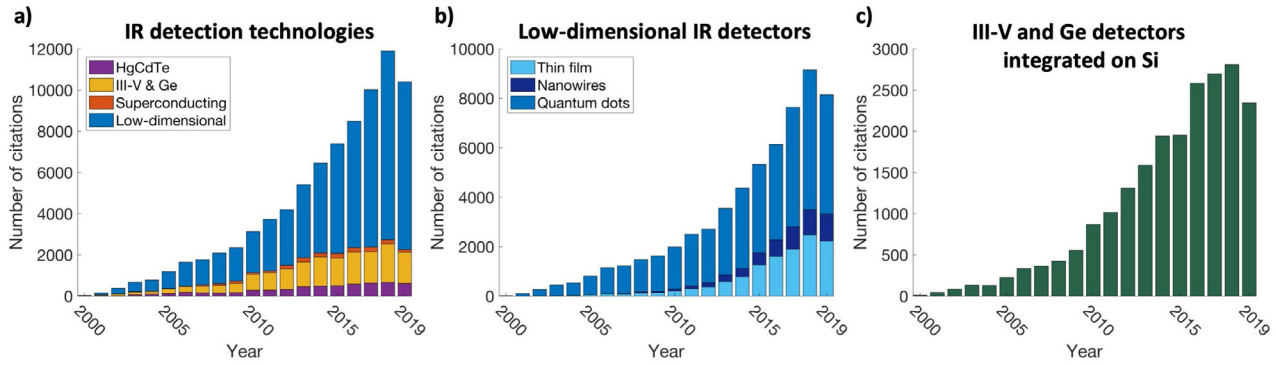

Figure 3. SWIR imager research output over the last 20 years. (a) SWIR detection citation report for the most common and most promising technologies: the citation boost is dominated by low-dimensional materials (b) and III–V and Ge-based detectors, particularly involving Si integration methods (c). Note: 2019 data only include citations from January though October.

SNR will then be defined by the poissonian shot noise of the charges generated by the background radiation, N_B:

$$\mathrm{SNR_{BLIP}} = \sqrt{N_B} = \sqrt{\frac{\eta}{2}\frac{Q_B}{A_D\,BW}} \tag{3}$$

where Q_B is the background photon flux density, in $\#\mathrm{photons}\,(\mathrm{cm^2 s})^{-1}$, and η is the detector quantum efficiency. The background radiant flux is $\Phi_B = Q_B A_D \frac{hc}{\lambda}$, where $\frac{hc}{\lambda}$ is the energy of the photons. The limit BLIP detectivity can therefore be expressed as:

$$D^* = \frac{\sqrt{A_D\,BW}}{\Phi_B}\mathrm{SNR_{BLIP}} = \frac{\lambda}{hc}\sqrt{\frac{\eta}{2Q_B}}. \tag{4}$$

2.2. Key infrared imager parameters

A vast portion of the applications of infrared detectors encompasses imaging techniques: this requires the detectors to be produced in large arrays that could be placed in the focal plane of an optical imaging system. Such an array of detector is called focal plane array (FPA) and is operated globally as an imager. In order to access the large number of pixels at once, and be able to perform imaging at a reasonable frame rate, hybrid integration of an array of photodetectors with a read-out integrated circuit (ROIC) is necessary. The ROIC system constitutes a platform for interfacing the large number of pixels (hundreds of thousand to hundreds of millions), with a reduced number of electrical connection pins (tens to hundreds of pins), to be controlled and processed digitally. The ROIC unavoidably introduces additional sources of noise to the detector system, denominated read noise, which include for instance ADC quantizing noise, reset noise, dark current and thermal noise [62]. Moreover, when considering the imager performance as a system, losses related to the optics need to be taken into account. The most common figure of merit of the performance of an IR imager is the noise equivalent difference temperature (NEDT), representing the minimum temperature difference of an imaged object that can be sensed by the imager with an SNR = 1, defined as [2]:

$$\mathrm{NEDT} = \frac{V_n \frac{\partial T}{\partial Q}}{\frac{\partial V_S}{\partial Q}} = V_n \frac{\Delta T}{\Delta V_S} \tag{5}$$

where V_n is the rms noise of the imager, and ΔV_s is the measured signal as a result of the temperature difference ΔT. Assuming the imager system has a detector area of A_D, and a lens of focal length f and aperture of diameter D, corresponding to $F\# = \frac{f}{D}$, then the radiant flux Φ in W as a function of the irradiance L of the imaged object in W stm^{-2} can be expressed as [2]:

$$\Phi = L\frac{\pi}{4}\frac{A_D}{(F\#)^2}. \tag{6}$$

Using equation (1), the signal voltage can be then expressed by:

$$V_S = \int_{\lambda_a}^{\lambda_b} R_V(\lambda)\Phi(\lambda)\,\mathrm{d}\lambda = \frac{\pi}{4}\frac{A_D}{(F\#)^2}\int_{\lambda_a}^{\lambda_b} R_V(\lambda)L(\lambda)\,\mathrm{d}\lambda \tag{7}$$

from which:

$$\frac{\partial V_S}{\partial T} = \frac{\pi}{4}\frac{A_D}{(F\#)^2}\int_{\lambda_a}^{\lambda_b} R_V(\lambda)\frac{\partial L(\lambda)}{\partial T}\mathrm{d}\lambda \tag{8}$$

and using $D^* = \frac{\sqrt{A_D\,BW}}{NEP} = \frac{\sqrt{A_D\,BW}}{V_n}R_V$:

$$\frac{\partial V_S}{\partial T} = \frac{\pi}{4}\frac{A_D}{(F\#)^2}\frac{V_n}{\sqrt{A_D\,BW}}\int_{\lambda_a}^{\lambda_b} D^*(\lambda)\frac{\partial L(\lambda)}{\partial T}\mathrm{d}\lambda \tag{9}$$

$$\mathrm{NEDT} = V_n\frac{\Delta T}{\Delta V_S} = \frac{4(F\#)^2\sqrt{BW}}{\pi\sqrt{A_D}}\left[\int_{\lambda_a}^{\lambda_b} D^*(\lambda)\frac{\partial L(\lambda)}{\partial T}\mathrm{d}\lambda\right]^{-1}. \tag{10}$$

Thus the best sensitivity, corresponding to smallest NEDT, is achieved for maximum detectivity in the wavelength of interest, and minimum $F\#$ number. Moreover, sensitivity of the imager is inversely proportional to its bandwidth.

2.3. Ultimate limits of photon detection sensitivity:

The current responsivity of a photodetector can be written as [2]:

$$R_i = \frac{\lambda}{hc}\, q\, \eta\, \beta \tag{11}$$

where $\frac{hc}{\lambda} = h\nu$ is the energy of the absorbed photon, q is the charge of the electron, η and β are the detectors quantum efficiency and gain. The electrical current due to noise is:

$$I_{\mathrm{n}} = q\,\beta\sqrt{2\ (G_{\mathrm{op}} + G_{\mathrm{th}} + R)\ A_{\mathrm{e}}\, t\, \mathrm{BW}} \tag{12}$$

where G_{op}, G_{th} and R are the excess carrier optical generation, thermal generation and recombination rates, respectively, and A_{e}, t and BW are the electrically active area, thickness and bandwidth of the detector. Considering $G = G_{\mathrm{op}} + G_{\mathrm{th}}$, and since at equilibrium $G = R$, equations (1) and (2) can be substituted in the expression for D^*, yielding:

$$D^* = \frac{\sqrt{A_{\mathrm{o}}\,\mathrm{BW}}}{\mathrm{NEP}} = \frac{\sqrt{A_{\mathrm{o}}\,\mathrm{BW}}}{I_{\mathrm{n}}} R_i = \frac{\lambda}{hc}\,\eta\,\sqrt{\frac{A_{\mathrm{o}}}{4GA_{\mathrm{e}}\,t}}. \tag{13}$$

Assuming a planar device ($A_{\mathrm{o}} = A_{\mathrm{e}}$), and approximating the quantum efficiency as $\eta \cong 1 - \mathrm{e}^{-\alpha\ t}$, and solving for the optimal device thickness that maximizes D^*, it can be shown that $t_{\mathrm{opt}} = \frac{1.26}{\alpha}$ [2]; so that equation (4) simplifies to:

$$D^* \cong 0.31\,\frac{\lambda}{hc}\,\sqrt{\frac{\alpha}{G}}. \tag{14}$$

The ultimate limit of sensitivity is obtained for an ideal detector at $T_{\mathrm{D}} = 0\,\mathrm{K}$, such that $G_{\mathrm{th}} = 0$, and the detector behaves like a black body with a surrounding cold shield, also assumed at $T_{\mathrm{S}} = 0\,\mathrm{K}$, with an acceptance angle θ. Considering the detector as surrounded by a uniform black body at temperature at $T \neq 0\,\mathrm{K}$, such that the radiation exchange is governed by the Planck distribution:

$$B_\nu\,(\nu, T) = \frac{2\,h\,\nu^3}{c^2\,\left(\mathrm{e}^{\frac{h\nu}{k_{\mathrm{B}}T}} - 1\right)}. \tag{15}$$

The number of modes per unit volume that can couple to the detector of area A_{o} and with acceptance angle θ can by calculated from the k-space volume of a spherical shell of thickness $\mathrm{d}k$ and solid angle $\mathrm{d}\theta$:

$$\mathrm{d}V_k = k^2 \mathrm{d}k\,\mathrm{d}\theta = \frac{8\,\pi^3\nu^2 \mathrm{d}\nu\,\mathrm{d}\theta}{c^3} \tag{16}$$

where $k = \frac{2\pi}{\lambda} = \frac{2\pi\nu}{c}$ has been used. Since in k-space every mode occurs for $k_{x,y,z} = \frac{m\pi}{L_{x,y,z}}$, with $L_{x,y,z}$ characristic lengths of the system, then the k-space volume-per-mode is given by $\left(\frac{\pi}{L}\right)^3$, and dividing $\mathrm{d}V_k$ by this volume, by the real space volume, L^3, and multiplying by 2 to include both polarizations and dividing by 8 to account for the one octant of the k-space volume corresponding to positive k, the number of modes per unit volume is found:

$$\mathrm{d}n = \frac{\mathrm{d}V_k}{\left(\frac{\pi}{L}\right)^3}\frac{2}{8L^3} = \frac{2\nu^2\,\mathrm{d}\nu\,\mathrm{d}\theta}{c^3} \tag{17}$$

where the power transfer between the two black bodies, expressed as $\mathrm{d}P = h\nu\,\mathrm{d}n f(\nu) c\,\mathrm{d}A_{\mathrm{o}}$, with $f(\nu) = \frac{1}{\mathrm{e}^{\frac{h\nu}{k_{\mathrm{B}}T}} - 1}$ the Bose-Einstein distribution function, yields exactly Planck's law (equation (15)). Considering the detector as a gray body with quantum efficiency $\eta(\nu, \theta)$, however:

$$\mathrm{d}P = \frac{2h\nu^3\eta(\nu,\theta)}{c^2\,\left(\mathrm{e}^{\frac{h\nu}{k_{\mathrm{B}}T}} - 1\right)}\mathrm{d}\nu\,\mathrm{d}\theta\,\mathrm{d}A_{\mathrm{o}}. \tag{18}$$

Hence, the total flux of background noise photons absorbed in the detector, $N_{\mathrm{BG}} = \iiint \frac{\mathrm{d}P}{h\nu}$ is the ultimate limit of G:

$$GA_{\mathrm{e}}\,t \geqslant N_{\mathrm{BG}} = \iiint \frac{2\,\nu^2\eta(\nu,\theta)}{c^2\,\left(\mathrm{e}^{\frac{h\nu}{k_{\mathrm{B}}T}} - 1\right)}\mathrm{d}\nu\,\mathrm{d}\theta\,\mathrm{d}A_o. \tag{19}$$

Substituting this value for G in equation (3) yields a new expression for the theoretical limit of detectivity:

$$D^* = \frac{\eta(\nu,\theta_0)}{2\,h\nu_0}\left(\iint \frac{2\nu^2\eta(\nu,\theta)}{c^2\,\left(\mathrm{e}^{\frac{h\nu}{k_{\mathrm{B}}T}} - 1\right)}\mathrm{d}\nu\,\mathrm{d}\theta\right)^{-1/2} \tag{20}$$

where A_{o} cancels out when carrying out the integral. An expression for the NEP can subsequently be derived from the definition of D^* (equation (4)) as:

$$\mathrm{NEP} = 2\frac{h\nu_0}{\eta\,(\nu,\theta_0)}\left(\mathrm{BW}\iint \frac{2\,\nu^2\eta(\nu,\theta)}{c^2\,\left(\mathrm{e}^{\frac{h\nu}{k_{\mathrm{B}}T}} - 1\right)}\,\mathrm{d}\nu\,\mathrm{d}\theta\right)^{1/2}. \tag{21}$$

Using equation (10), the theoretical sensitivity limit can be calculated for different materials, given their absorption coefficient, as a function of wavelength and temperature, as reported in figure 4 for various AlGaSb, InGaAs and HgCdTe-based detectors [63–66].

2.4. Sensitivity limit of sub-wavelength sized detectors:

There are several advantages related to shrinking the size of the detectors, including higher pixel density and increased sensitivity derived from lower capacitance of the device [17, 29, 30, 42]. In particular, as we discussed above, it has been theoretically proven that both long-range and large-field of view imaging modes can benefit from a diffraction-limited imaging system [30, 39, 40]. Furthermore, recent studies have reported record-high responsivity using nanoscale devices and low-dimensional materials [49, 50, 67–69]. In this analysis, we consider an additional benefit to the sensitivity that arises from reducing the number of modes supported by the detector when its size approaches the same order as the wavelength [70]. We do not consider here the limitations on the optics of the system imposed by diffraction-limited operation, and we limit to a theoretical evaluation of the sensitivity limit that can be achieved by decreasing the size of the detector body. Several design strategies can be adopted to facilitate and enhance detection near and beyond the diffraction limit, including optical antennas [71], plasmonics [72] and metamaterials [73], as is discussed in the following sections. When the size of the detector approaches the same order as the wavelength, the density of electromagnetic modes that can be supported by the subwavelength body decreases [70]. As a result,

the black-body radiation formula in equation (15) needs to be adjusted to take into account the size of the detector [74]:

$$B_{\mathrm{bb}}(\nu, T, d) = B_{\nu}(\nu, T)\left(1 - \frac{1}{2\pi^2}\frac{\lambda_{\mathrm{M}}^2}{d^2}\right) \tag{22}$$

where the detector body is modeled as a sphere of diameter d and $\lambda_{\mathrm{M}} = \frac{\lambda}{n(\lambda)}$ is the wavelength in the detector material. The factor in the brackets in equation (5) accounts for the decrease in density of modes as a function of wavelength and detector size: it intercepts zero at $\lambda = \sqrt{2}\pi d$, and all modes corresponding to larger wavelengths are physically unrealizable. This size factor acts as a cutoff filter on the blackbody radiation, as shown in figure 5(a): here the size factors for different detector dimensions are represented by the dashed lines, and the resulting modified blackbody radiances $B_{\mathrm{bb}}(\nu, T, d)$ by solid lines. As a result, shrinking the size of the detector decreases the noise contribution from the long-wavelength tail of the blackbody spectrum. However, the effect on the detector performance can be dual, since the small size can have a detrimental effect on the quantum efficiency $\eta(\nu_0, \theta)$ at the wavelength of interest. Therefore, the sensitivity of the detector can be improved by shrinking its size, up until the size of detector is too small to support the wavelength of interest for detection, i.e. $d \leqslant \frac{\lambda_{\mathrm{M}_0}}{\sqrt{2}\pi} = \frac{\lambda_0}{\sqrt{2}\pi n(\lambda_0)}$, and the detectivity drops to zero. Typically, a large refractive index of the detector materials allows to push this cutoff towards small detector sizes: for an InGaAs-based detector such as shown in figure 5, for instance, the cutoff detector size for supporting 1550 nm wavelength is $d \geqslant \frac{1.55\cdot 10^{-6}\,\mathrm{m}}{\sqrt{2}\pi\ 3.5} \approx 100\,\mathrm{nm}$. The enhancement in detectivity at 1500 nm as a result of shrinking in size for InGaAs-based detectors is shown in figures 5(c) and (d): since the nano-scale size of the detector effectively acts as a cutoff wavelength, the improvement in performance is more significant for detectors based on materials systems with longer cutoff wavelength. By extending these calculations to material systems across the infrared and visible, as shown in figure 6, it is apparent that small sized detectors can prove beneficial across all these wavelengths. Specifically, sub-wavelength detector can act as high-pass filters for the energy of the photons incident on the detectors, cutting the absorption of all longer wavelengths. This is particularly beneficial for the case of materials that do not have a bandgap tailored to the ideal wavelength cutoff of their application. While the bandgap of some materials, such as HgCdTe, can be tailored by composition, many binary and ternary III–V semiconductors do not have this degree of freedom. As a result, for example, InGaAs detectors with a wavelength cutoff around 1.7 μm intrinsic to their bandgap are typically used for detection of SWIR photons at 1.5 μm and 1.3 μm, as could benefit from the reduction of background noise enabled by sub-wavelength detectors cutoff.

2.5. Sensitivity and bandwidth limit of nano-scale phototransistor detectors

A growing number of novel infrared photodetectors utilizes transistor action to produce an internal amplification mechanism, including most low-dimensional detectors that obtain a gain from the modulation of a potential barrier to the majority carrier [75–80]. When designing high-sensitivity infrared detectors, approaching the thermodynamic sensitivity limit, phototransistors represent one of the most attractive device architectures. Their main advantage is generating a large electrical signal—even from a single photon—to be detceted with high fidelity using conventional electronic circuits. At low light levels, a simple equivalent circuit can be used to describe the dynamic response of the detectors, as shown in figure 7(a) [42]. Here, the response is dominated by the charge transfer across the device capacitance, and the time constant can be expressed as $\tau_{\mathrm{RISE}} = r_{\mathrm{d}} C_{\mathrm{TOT}}$, where $r_{\mathrm{d}} = \frac{\frac{k_{\mathrm{B}}T}{q}}{I_{\mathrm{d}}}$ is the dynamic resistance and C_{TOT} is the sum of all parallel capacitances (base-emitter, base-collector, parasitics). Upon receiving a flux of photons, ϕ, the detector current is raised above the dark current, I_{d}, by the signal $I_{\mathrm{ph}} = \eta\phi q\left(1 - e^{-\frac{t}{\tau_{\mathrm{RISE}}}}\right)$, where η is the quantum efficiency and t is the length of the photon pulse, as shown in figure 7(b). The signal-to-noise ratio of the system can then be written as [81]:

$$\mathrm{SNR} = \frac{I_{\mathrm{ph}}}{\sqrt{I_{n_{\mathrm{d}}}^2 + I_{n_{\mathrm{ph}}}^2 + I_{n_{\mathrm{R}}}^1}} \tag{23}$$

where $I_{n_{\mathrm{d}}} = 2qI_{\mathrm{d}}\mathrm{BW}$ and $I_{n_{\mathrm{ph}}} = 2qI_{\mathrm{ph}}\mathrm{BW}$ represent the dark current and photocurrent shot noise respectively, and $I_{n_{\mathrm{R}}}$ the read noise. For the sake of calculating the theoretical sensitivity limit, read noise can be neglected considering an internal gain mechanism that allows to amplify the detector output beyond the read noise. Furthermore, for the circuit equivalent shown in figure 7(a), the bandwidth can be approximated by $\mathrm{BW} \approx \frac{1}{4\tau_{\mathrm{RISE}}}$ [42]. As a result, equation (23) simplifies to:

$$\mathrm{SNR} = \frac{\eta\phi q\left(1 - \mathrm{e}^{-\frac{t}{\tau_{\mathrm{RISE}}}}\right)}{\sqrt{\frac{q}{2\tau_{\mathrm{RISE}}}\left(I_{\mathrm{d}} + \eta\phi q\left(1 - \mathrm{e}^{-\frac{t}{\tau_{\mathrm{RISE}}}}\right)\right)}}. \tag{24}$$

Hence at low light levels the junction capacitance determines the device bandwidth, acting as a low-pass filter on the white noise contribution; as a result, small device capacitance lowers the noise contribution and enhances the SNR.

In the limit of large pulse or fast device, $t \gg \tau_{\mathrm{RISE}}$, represented by the blue response curve in figure 7(b), and defining $g = \frac{t}{\tau_{\mathrm{RISE}}}$ and $N_{\mathrm{ph}} = \phi t$ the number of photons within a pulse:

$$\mathrm{SNR} \approx \frac{\eta\phi q}{\sqrt{\frac{q}{2\tau_{\mathrm{RISE}}}\left(I_{\mathrm{d}} + \eta\phi q\right)}} = \frac{\eta N_{\mathrm{ph}}}{\sqrt{\frac{1}{2}\left(g^2\frac{C_{\mathrm{TOT}}}{C_0} + g\eta N_{\mathrm{ph}}\right)}} \tag{25}$$

where $C_0 = \frac{q^2}{k_{\mathrm{B}}T}$ is defined as thermal fundamental capacitance [42].

Conversely, for a pulse consisting of a small number of photons, the pulse duration will be significantly small compared to the characteristic rise time of the device, $\frac{t}{\tau_{\mathrm{RISE}}} \to 0$, such that $1 - \mathrm{e}^{-\frac{t}{\tau_{\mathrm{RISE}}}}$ can be Taylor-expanded, yielding:

$$\mathrm{SNR} \approx \frac{\eta\frac{N_{\mathrm{ph}}}{t}q\frac{t}{\tau_{\mathrm{Rise}}}}{\sqrt{\frac{q}{2\tau_{\mathrm{RISE}}}\left(I_{\mathrm{d}} + \eta\frac{N_{\mathrm{ph}}}{t}q\frac{t}{\tau_{\mathrm{Rise}}}\right)}} = \frac{\eta N_{\mathrm{ph}}}{\sqrt{\frac{1}{2}\left(\frac{C_{\mathrm{TOT}}}{C_0} + \eta N_{\mathrm{ph}}\right)}}. \tag{26}$$

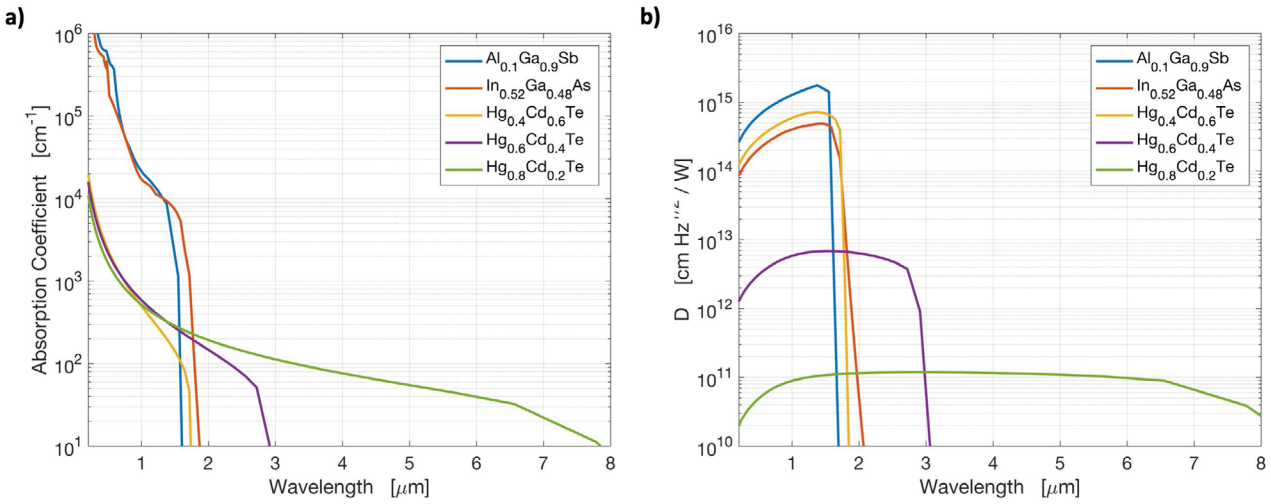

Figure 4. Absorption coefficient (a) and normalized detectivity (b) as a function of wavelength for detectors based on different materials. The normalized detectivity D^* is calculated from equation (10), assuming a background temperature of 300 K and a field of view of 30°.

Solving for the sensitivity in terms of number of absorbed photons detected with a certain SNR, equation (26) yields:

$$\eta N_{\text{ph}} = \frac{1}{4}\text{SNR}^2\left(1+\sqrt{1+\frac{8}{\text{SNR}^2}\frac{C_{\text{TOT}}}{C_0}}\right). \quad (27)$$

The SNR as a function of the size of the detector for different numbers of absorbed photons, ηN_{ph}, is shown figure 7(c), assuming a cylindrical detector, with a depletion width $t = 300$ nm, such that the total capacitance can be calculated as$C_{\text{TOT}} = \varepsilon_0\varepsilon_{\text{R}}\frac{A}{t}$. As the device operating temperature is increased from 4 K to 77 K and 300 K, a smaller detector size is required to achieve a certain SNR at a given photon flux: for instance, in order to detect 10 absorbed photons with an SNR = 3 at room temperature, a 200 nm detector is needed, compared to a 2 μm detector at 4 K. Similarly, the sensitivity as a function of the detector size is shown in figure 7(d) for the same three different operating temperature.

In summary, ultra-low capacitance phototransistors possess the potential for reaching unprecedented levels of sensitivity by shrinking the detector size. Moreover, nanoscale photodetectors that can achieve capacitance in the order of a few attofarad are almost insensitive to operating temperature, as evident from figures 7(c) and (d), shedding the cooling requirements even at high sensitivity levels. This approach suggests that room-temperature single-photon sensitivity approaching the quantum limit of SNR = 1 can theoretically be achieved by photodetectors of a few hundreds of nm in size.

2.6. Sensitivity and bandwidth considerations for thin-film detectors:

A large number of novel detectors design are based on thin-films, including for example quantum dots, perovskites, 2D materials and organics. Due to their intrinsic differences in physics and operation of the devices, it is important that these devices are accurately characterized in order to be compared to state-of-the-art detector technologies in a fair way. In particular, as indicated by Fang *et al* [1], thin film detectors often possess high surface-to-volume ratio, which makes them more susceptible to the physics of trap states. As a result, approximations made for the sake of extrapolation, such as white noise (frequency-independent) and response linearity at low-light level are often not accurate, causing unrealistic estimates of the detectivity (D^*) and sensitivity (NEP) [1, 82]. As such, it is crucial that these parameters are determined by direct measurement methods that do not require extrapolation assumptions, as proposed by Fang *et al*, in order to obtain useful terms of comparison with the current state-of-the-art infrared technologies [1].

2.7. ROIC bandwidth and SNR considerations:

When a photodetector array is integrated to form an imager, the bandwidth of the ROIC also affects the system SNR. The effective noise bandwidth of the system will be limited by the lowest between the detector and the ROIC bandwidth. The bandwidth of the ROIC can be tuned by changing the integration time, as $\text{BW}_{\text{ROIC}} \sim \frac{1}{T_{\text{int}}}$, and because ROICs can typically operate with integration times spanning from μs to seconds, likely both regimes are accessible for most imagers [83]. When $T_{\text{int}} < 2\tau_{\text{RISE}}$, the bandwidth of the system is detector-limited, and equation (24) for the SNR is recovered. Conversely, when $T_{\text{int}} > 2\tau_{\text{RISE}}$, the bandwidth of the system in ROIC-limited, and equation (24) is modified into:

$$\text{SNR}_{\text{ROIC-limited}} = \frac{\eta\phi q\left(1-\text{e}^{-t/\tau_{\text{RISE}}}\right)}{\sqrt{\frac{q}{T_{\text{int}}}\left(I_{\text{d}}+\eta\phi q\left(1-\text{e}^{-t/\tau_{\text{RISE}}}\right)\right)}}. \quad (28)$$

3. Recent advances in infrared imagers

In this section, we offer a review of the most prominent technologies for infrared photon detection, with a particular emphasis on those that can support large array fabrication for imagers. The main state-of-the-art infrared detectors

technologies are based on semiconductors, superconductors, or thermal detectors, which differ by absorption mechanism, physical principle of operation, and device architecture and performance.

In the present manuscript, we only consider photon detection technologies, such as seminconducting and superconducting detectors. Infrared thermal detectors such as microbolometers have consistently improved in recent years, positioning them as an important technology, especially in the LWIR band, mainly thanks to their economic feasibility, enabled by lower cost materials and wafer-scale fabrication [28, 84]. In addition, state-of-the-art microbolometers can instrinsically operate at room temperature, which is extremely attractive for infrared imagers [28, 29]. However, their performance at shorter wavelengths, especially in the SWIR range, is significantly inferior. This is mainly related to the absorption mechanism in such detectors, which does not discriminate on the energy of the incident photons, contrary to what semiconductors do by virtue of their bandgap [28, 84]. Since thermal noise spectrum at 300 K is peaked around 10 μm wavelength, the detection of much shorter wavelength with microbolometers is flawed by low signal-to-noise ratio resulting from high absorption of the thermal noise at longer wavelengths. As a result, the performance metrics of microbolometers typically lay orders of magnitude below those of photon detectors in the SWIR band [84, 85].

Furthermore, in this discussion we also do not consider superconducting detectors in details [86]. While this class of devices has emerged as a leading technology for the detection of long-wave and microwave photons, finding extensive application especially in astronomy [87–90], it is flawed by some intrinsic limitations that currently limit its applicability and prospects for the future development of infrared detection and imaging. One of the main drawbacks of these detectors is that in order to take advantage of the superconducting zero-resistance state, they need to operate below the superconducting transition temperature, which is typically around a few degrees Kelvin [91]. As discussed in the previous section, however, the future of infrared imagers will be dominated by detectors operated in HOT conditions (above 200 K), which hampers the technological interest in superconducting detectors, except for a few specific types of applications. Perhaps more importantly, superconducting detectors typically suffer from low quantum efficiency and poor large-area scalability [92–95]. Some improvements in this merit have recently been reported, for example for the case of microwave kinetic inductance detectors (MKID), which have been fabricated in arrays of 1024, and integrated with microlenses [9, 96–98]. Nonetheless, due to both poor fabrication uniformity and yield, and to the complexity and bulkiness of the read-out circuitry, the current MKID pixel counts are still rather limited. As a result, we opted not to include superconducting detectors in this review which is mostly focused on imagers. However, we note that future technological advancements might enable large 2D arrays of superconducting detectors.

Finally, a brief overview of low-dimensional produced from 'bottom-up' methods is also presented, despite most of its typical implementations not yet representing a mature enough technology for supporting infrared imagers. In conclusion, the main strategies for enhancing light coupling in order to increase detector efficiency are also discussed in this section.

3.1. Semiconductor devices

The vast majority of commercially available infrared detectors and camera are currently based on semiconducting materials: unlike superconductors, semiconductor devices can operate above cryogenic and up to room temperature, and can take advantage of a far larger number of device architectures, materials and fabrication processes. These advantages have led to commercial production of IR imagers with over a million pixels in one FPA. The absorption and detection of IR photons is associated with the transition of carriers across an energy gap, which can correspond to the bandgap in interband absorption devices, inter-subband in superlattices, or to shallow trap energy states of engineered defects [2]. The last two architectures are generally limited to longer wavelengths, where high-quality materials with the required narrow bandgap are not readily available, while they are usually outperformed at shorter wavelength by interband devices, thanks to their higher mobility and carrier lifetime [2]. Here, we will focus only on the interband semiconductors technologies. Figure 8 shows the bandgap and corresponding cutoff wavelength of a wide range of semiconductors employed for IR detection [28]. The most common device architectures for semiconductor detectors are the photodiode and the phototransistor made from this material system [29].

3.1.1. HgCdTe. $Hg_{1-x}Cd_xTe$ is a direct bandgap, ternary semiconductor alloy with a zinc-blende crystal structure [6]. The key feature of this material is that its bandgap can be tuned by varying its x-composition between the negative energy gap of HgTe and the relatively high energy gap (1.5 eV) of CdTe, as represented in figure 9 [28]. As a result, its absorption properties can be engineered to effectively cover the whole IR 1–30 μm wavelength range [31]. Furthermore, the electronic and optical properties of this alloy make it ideal for IR detection application: it exhibits high optical absorption, low dielectric constant, high mobility, low recombination at room temperature [32]. In addition, carrier concentration and type can be effectively controlled by doping processes, allowing for a wide range of device architectures. Finally, the extremely narrow bandgaps and impact ionization properties make HgCdTe an ideal candidate for realizing detectors with avalanche-based gain mechanisms [99]. For all these reasons, HgCdTe is currently the most widely employed material for IR detection applications.

Due to the nature of its main funding sources, the research and development of HgCdTe technology has always been aimed at the demonstration of FPAs, with several successful milestones over the last 50 years, from the demonstration of the first linear array, in the '70, to the recent achievement of megapixel FPA [33].

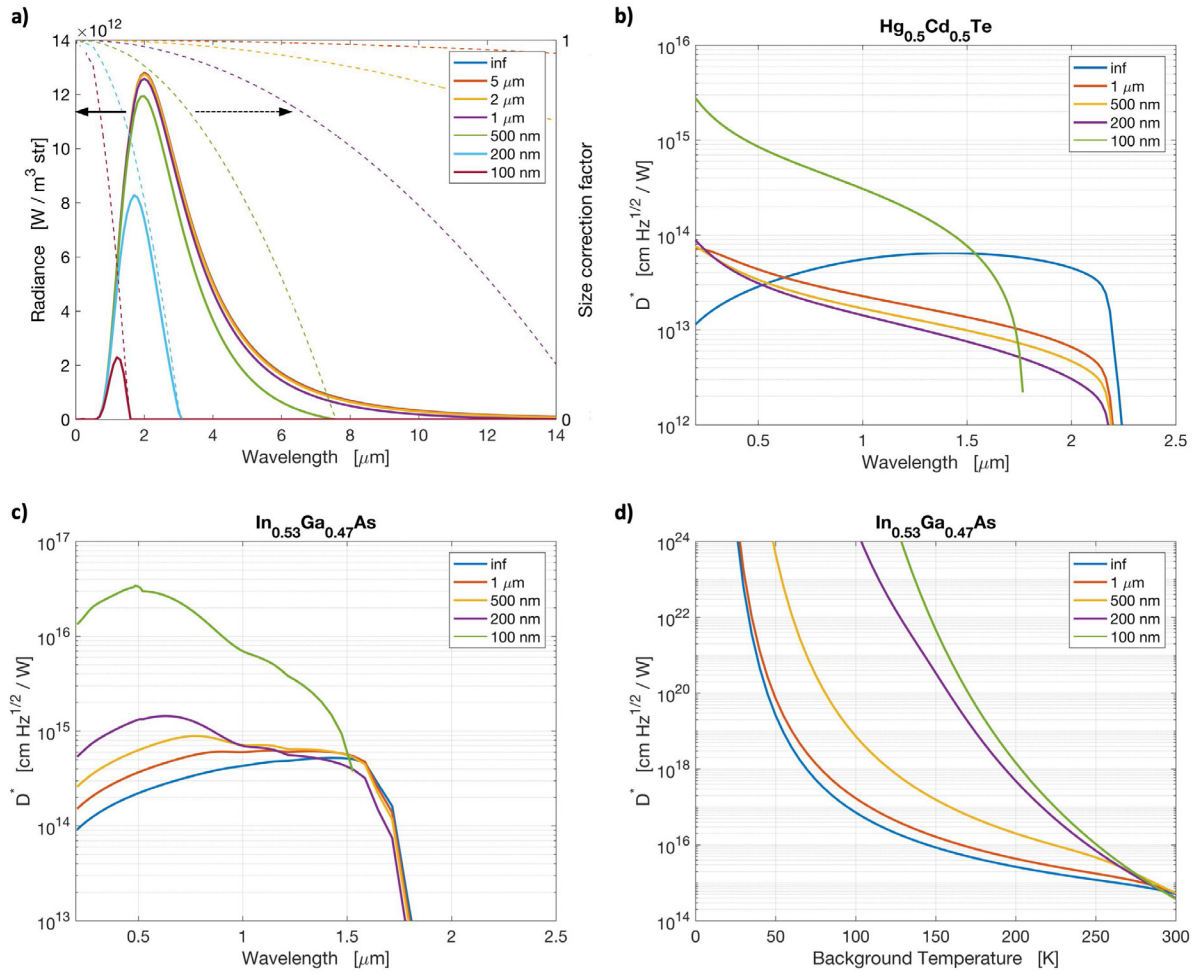

Figure 5. Effect of detector size on the radiance and detectivity. (a) Modified blackbody spectra $B_{bb}(\nu, T, d)$ for different detector sizes (solid lines). The corresponding correction factors are represented by the dashed lines and refer to the right y-axis. The spectra correspond to radiance emitted by a blackbody at 5000 K. (b) and (c) Normalized detectivity D^* as a function of wavelength for different detector sizes, for $Hg_{0.5}Gd_{0.5}Te$-based (b) and $In_{0.53}Ga_{0.47}As$-based (c) detectors. D^* was calculated from equations (10)and (12), assuming a background temperature of 300 K and a field of view of 30°. (d) Normalized detectivity D^* of an $In_{0.53}Ga_{0.47}As$-based detector at 1500 nm as a function of background temperature for different detector sizes, calculated from equations (10) and (12), assuming a field of view of 30°.

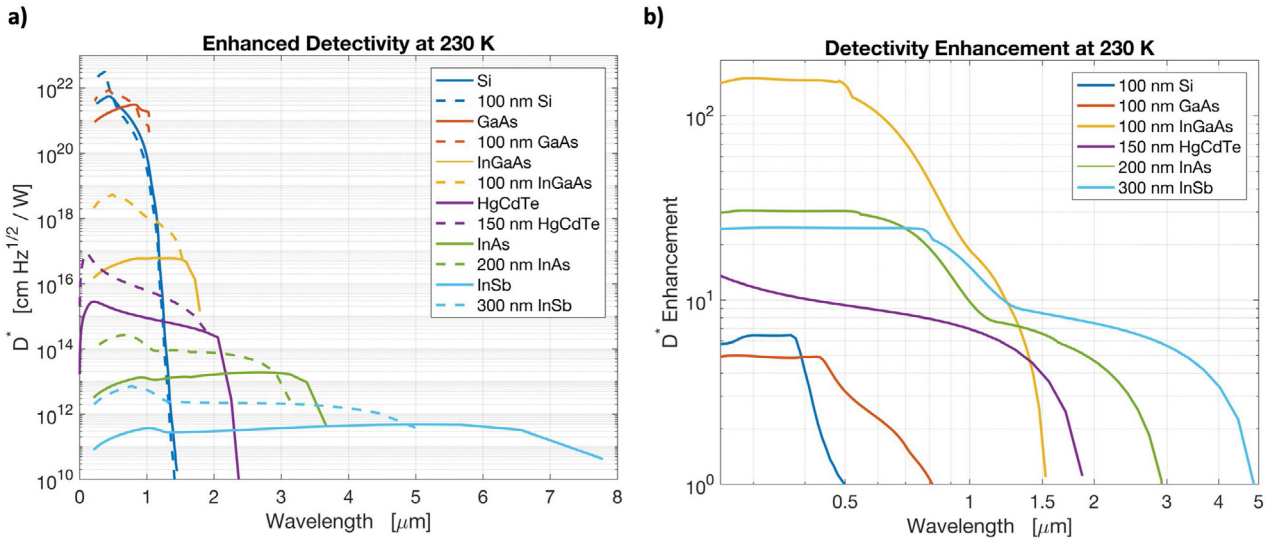

Figure 6. Detectivity enhancement enabled by sub-wavelength sized detectors for common detector materials across the infrared and visible range. (a) Solid lines represent the detectivity D^*, calculated from equations (10) and (12), assuming a background temperature of 300 K and a field of view of 30°, for a bulk detector (size much greater than wavelength). Dashed lines represent the enhanced detectivities enabled by the modified blackbody spectra $B_{bb}(\nu, T, d)$ for subwavelength detector sizes. (b) Relative detectivity enhancement (ratio of modified sub-wavelength detectivity to bulk).

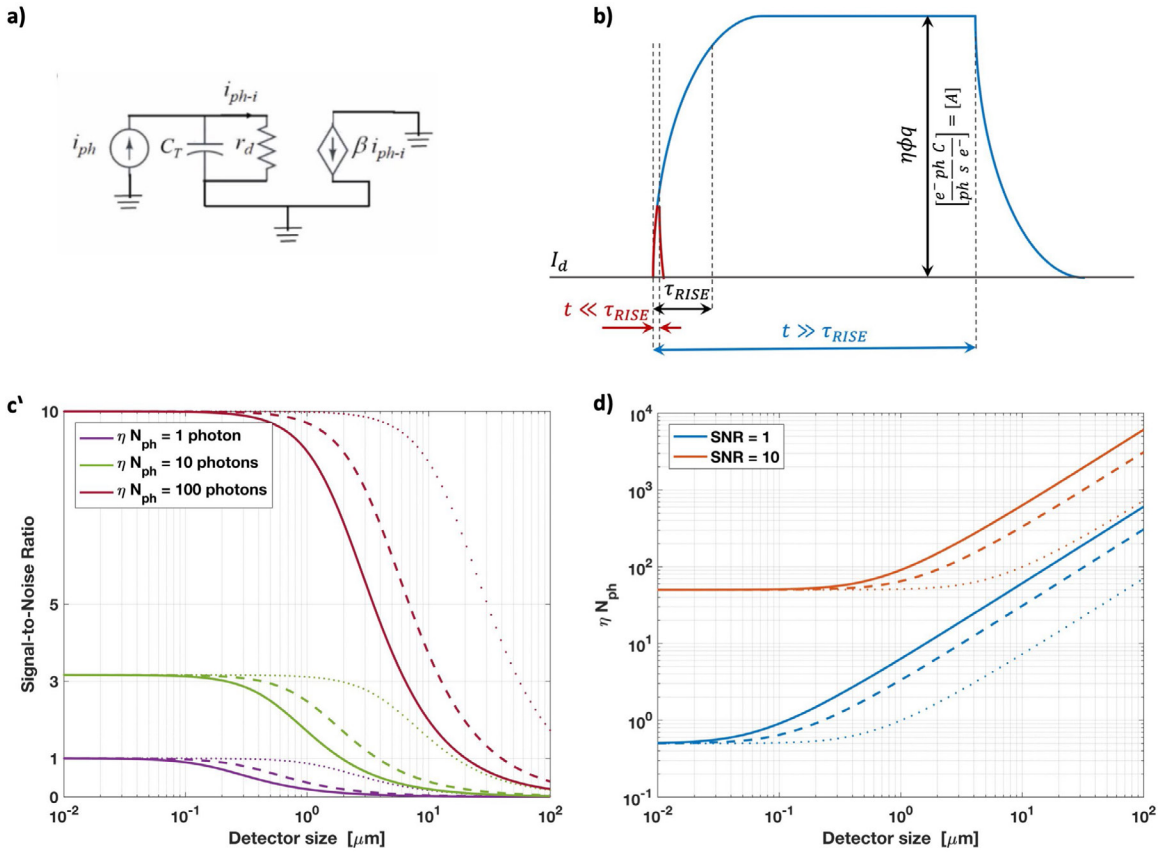

Figure 7. (a) Circuit equivalent of e phototransistor detector, as proposed in Rezaei *et al* [42]. Reproduced with permission from [42]. © IEEE; (b) schematic of a photocurrent pulse in a detector, showing the two different regimes; (c) detector SNR as a function of detector size for three different absorbed photons fluxes; (c) detector sensitivity in number of absorbed photons detected with a certain SNR as a function of detector size. In both (c) and (d), the solid line represents the solutions for $T = 300$ K, the dashed lines for $T = 77$ K and the dotted lines for $T = 4$ K.

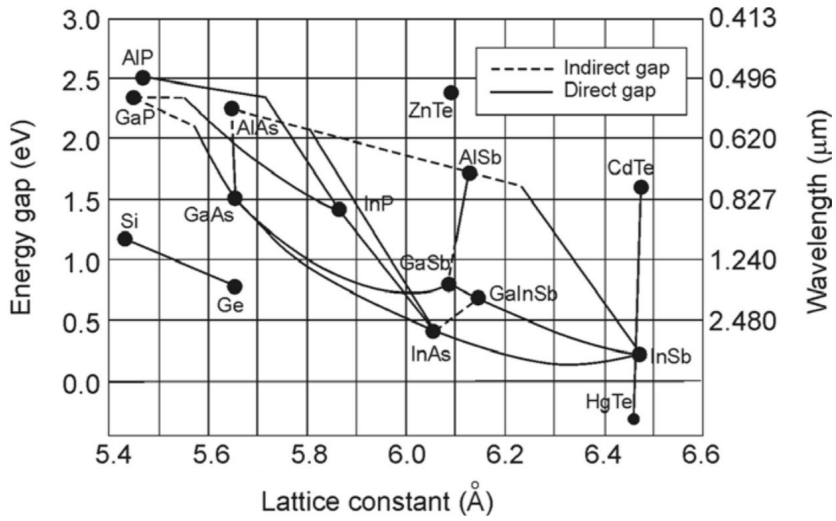

Figure 8. Direct and indirect bandgap and corresponding cutoff wavelength of some of the most common semiconductors in IR detection application. Reprinted from [28], Copyright (2011), with permission from Elsevier.

In recent years, major efforts in the advancement of the technology have been devoted to the demonstration of simultaneous multi-color detection, which can be achieved in a HgCdTe material system with graded composition, thanks to its extremely small change in lattice constant with composition [31, 32]. Multi-color FPAs based on a higher bandgap HgCdTe photodiode placed optically in front of a smaller bandgap photodiode have been demonstrated in all infrared spectral ranges [31]. These devices are typically run sequentially, selecting the read out of either of the photodiodes by means of the polarity of the applied voltage. Simultaneous multi-color detection has also been demonstrated using HgCdTe, in a number of different device architectures [31].

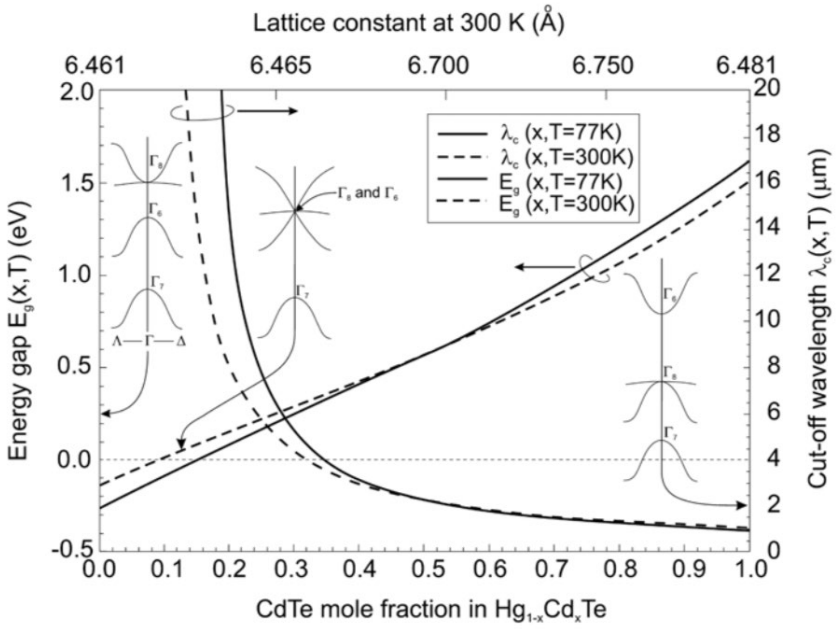

Figure 9. Band gap and corresponding cutoff wavelength $E_G = \frac{hc}{\lambda}$ of $Hg_{1-x}Cd_xTe$ near the Γ-point for varying x-composition of the alloy. Reprinted from [28], Copyright (2011), with permission from Elsevier.

Another crucial step in the development of HgCdTe technology was the implementation of avalanche processes for signal amplification, enabling unprecedented sensitivities to very small number of photons, thanks to this gain mechanism. A typical avalanche-photodetector (APD) architecture is shown in figure 10(a): the photogenerated carriers are multiplied via impact ionization in a high-field, narrow-bandgap region [100, 101]. For avalanche to take place, the kinetic energy of the accelerated carrier before ionization must be $E_0 > \alpha\, E_G$, where $\alpha > 1.5$ (3.2 for electron in Si) [102]. Therefore, the avalanche multiplication gain will be strongly dependent on the bandgap and applied field, as shown in figures 10(b) and (c) [28, 45]. The bandgap of HgCdTe can be engineered for simultaneously achieving the optimal avalanche gain and cutoff wavelength in the multiplication and absorption region respectively. FPAs of APD have been successfully fabricated and extensively characterized, and represent the state of the art for SWIR detector FPA [45, 101, 103, 104].

The main drawbacks of the HgCdTe material system are related to its technological complications, such as weak Hg-Te bond causing surface and interface defects and instabilities [32]. In addition, the high vapor pressure of Hg and the solid-liquid separation and related segregation represent major challenges in the growth of high-quality HgCdTe. Typically, uniformity and yield, are inferior to those of the more well-established III–V semiconductor fabrication processes, especially in the compositions utilized for longer wavelength spectral range [28, 32]. In addition, HgCdTe materials are usually grown on lattice-matched CdZnTe wafers, and cannot be grown on Si or GaAs substrates, due to the large lattice mismatch (19% [28]). This constitutes a major factor in determining the high cost of the material, due to both the high cost and small scale of these substrates, and also precludes direct integration with CMOS read-out, which is instead achieved by bumb bonding hybridization, further affecting the devices cost and yield. Moreover, the high field required for avalanche processes (figure 10(b)) intrinsically limits its application in fields where power consumption and auxiliary circuitry are of relevance: recently, novel avalanche device architectures based on localized field enhancement have been proposed in order to mitigate this limitation [20, 105].

For these reasons, while HgCdTe remains the leading detector material for LWIR and MWIR, it is seriously challenged at shorter wavelengths by other existing semiconducting materials, mostly of the III–V group.

3.1.2. III–V semiconductors. As can be inferred from figure 8, the majority of the semiconductors employed for IR detection applications belongs to the III–V group. The smallest energy gaps are achieved in the InSb and InAs alloys, which can achieve cutoff wavelengths in the MWIR spectral band, although the performance deteriorates above 30 K. The InGaAs ternary alloy, with direct bandgap between 0.35 eV for InAs to 1.43 eV for GaAs has established itself as one of the prominent materials for detection in the SWIR spectral range. In particular, $In_{0.53}Ga_{0.47}As$ alloy ($E_G = 0.73$ eV) lattice matched to InP has found applications in lightwave communication systems, low light-level night vision, etc. Germanium and the Si-Ge alloy is recently establishing as a direct competitor of InGaAs, despite its higher dark current and lower absorption, mainly thanks to its compatibility with the Si-based CMOS substrates and processing. Finally, recent advancements in the Sb growth have granted access to the InAsSb ternary alloy, which is more stable and uniform than HgCdTe and could compete with it for applications in the MWIR and LWIR.

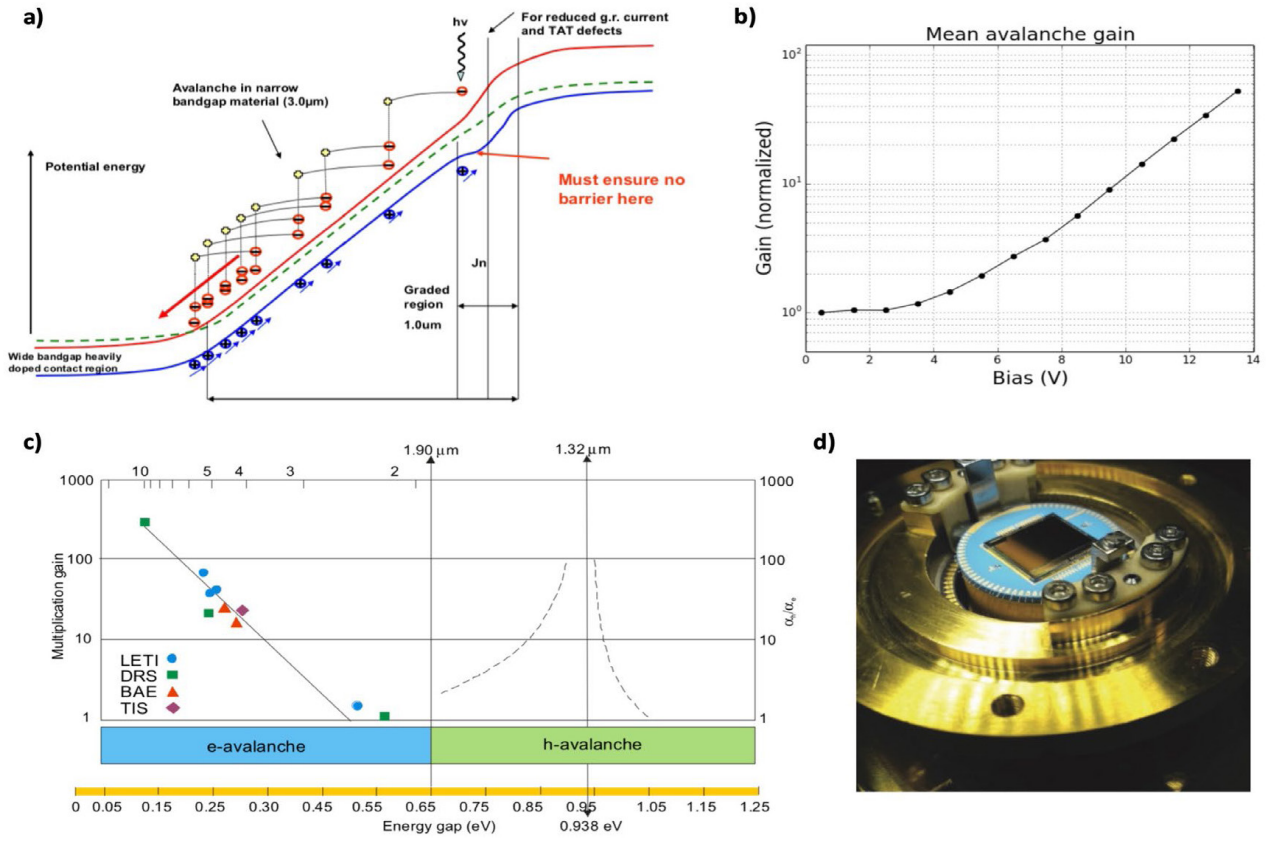

Figure 10. (a) Band diagram schematic of APD detectors [101], (b) voltage dependency of the avalanche gain in SAPHIRA APD FPA, reproduced with permission from [101]. © 2014 SPIE; (c) electron and hole avalanche multiplication gain as a function of HgCdTe bandgap, reprinted from [28], Copyright (2011), with permission from Elsevier, (d) RAPID [103] 320 × 255 IR APD array integrated in cryostat. Reproduced from [103]. CC BY 3.0.

3.1.3. III–V superlattices. One major advantage of III–V semiconductor technology are its mature and well-established growth and fabrication techniques, which ensure a material quality and uniformity over large areas superior to that of HgCdTe [28]. In addition, the versatility of the processing tools developed, as well as the doping capabilities enable complex device architectures, such as phototransistors and three-dimensional structures [44, 106]. An additional advantage of the plurality and compatibility of III–V materials is that they can be combined forming complex periodical structures such as superlattices and quantum wells, which allow to change the material's effective bandgap with ease, covering a range of detection from 1 to 30 μm [46, 107, 108]. While quantum well infrared photodetectors (QWIP) are widely employed in LWIR, their quantum efficiency is typically poor, due to the forbidden optical transition for normal incident light and the consequent need for grating coupling [28]. Superlattices, on the contrary, present quantum efficiencies comparable to those of HgCdTe and other bulk III–V. In addition, the combination and robustness of the available material systems enables band engineering for the suppression mechanisms that are detrimental to the detection, such as Auger and dark current [107]. Finally, superlattices have also attracted increasing attention for their application to implementing multi-color detectors, such as the one reported in figure 11, with both 2- and 3-color infrared detectors having been demonstrated [107, 109, 110].

3.1.4. Charge compression devices. An emerging strategy for highly sensitive photon detection is the so-called photogenerated charge compression and transfer. This concept leverages a 3D design of the detector band structure, aimed at decoupling the optical and the electronic functionalities of a photodetector. The responsivity of electronic sensors is ultimately related to their junction capacitance, which has driven the continuous shrinking in size of sensor electronics over the last several decades [15, 42]. Conversely, efficient light coupling and absorption typically require a large volume for maximizing the light-matter interaction. This intrinsic tradeoff has traditionally represented the ultimate limit to the photodetector performance, which are currently flawed by either poor quantum efficiency or low sensitivity [42]. By near-lossless transfer of the photogenerated charges from a large area absorbing region to a small-volume collection and multiplication region, charge compression has shown the potential to overcome this intrinsic limitation, enabling a disruptive advancement of the state of the art of photodetectors. Furthermore, this hybrid integrated architecture allows for enhanced area fill factor, high pixel density, increased full-well capacity and high conversion gain [111].

An example of this design concept has recently been demonstrated in a CMOS-based megapixel pump-gate jot device for quanta image sensor in the visible range [112]. A schematic of the pump-gate jot architecture is shown in figure 12:

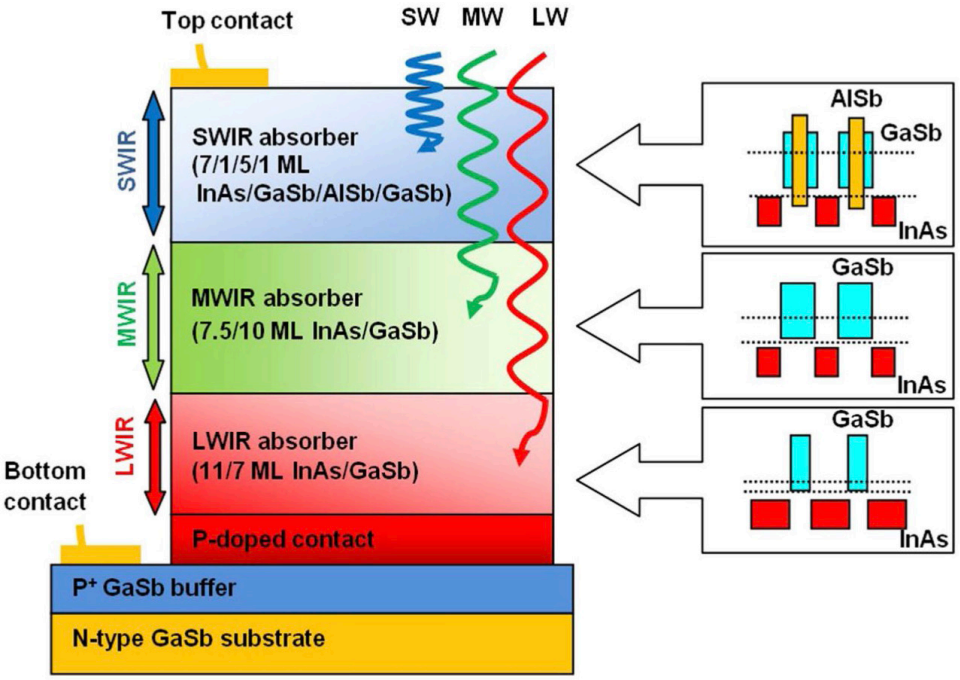

Figure 11. Schematic diagram of the triple-band superlattice SWIR-MWIR-LWIR photodetector reported by Hoang *et al* [110]. The band alignment of the three different superlattices employed are reported on the right (colored squares represent the forbidden bandgaps and the dashed lines the effective bandgaps). Reproduced from [110]. CC BY 4.0

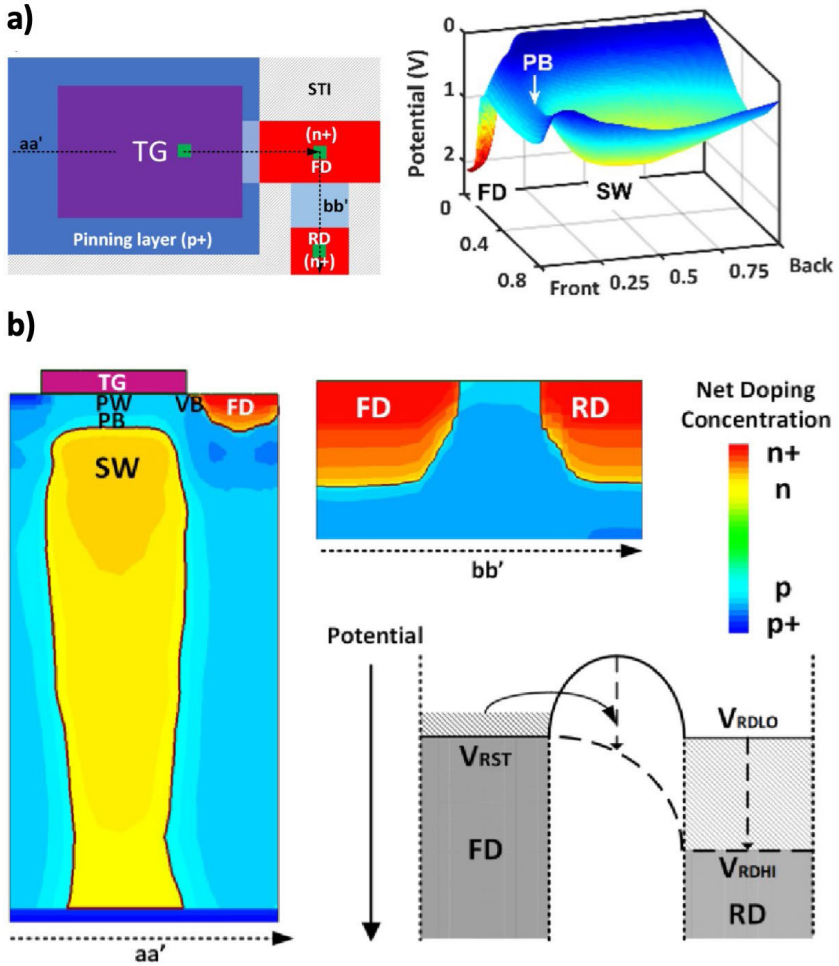

Figure 12. (a) Schematic of the geometric design of the EI detector, combining a large-area absorption region with a small-volume hole-trapping multiplication region (nanoinjector) [115]; cross-sectional 3D (b) and axial (c) band structure showing lateral band bending for charge compression and confinement [116, 119]. Reproduced from [112]. CC BY 3.0.

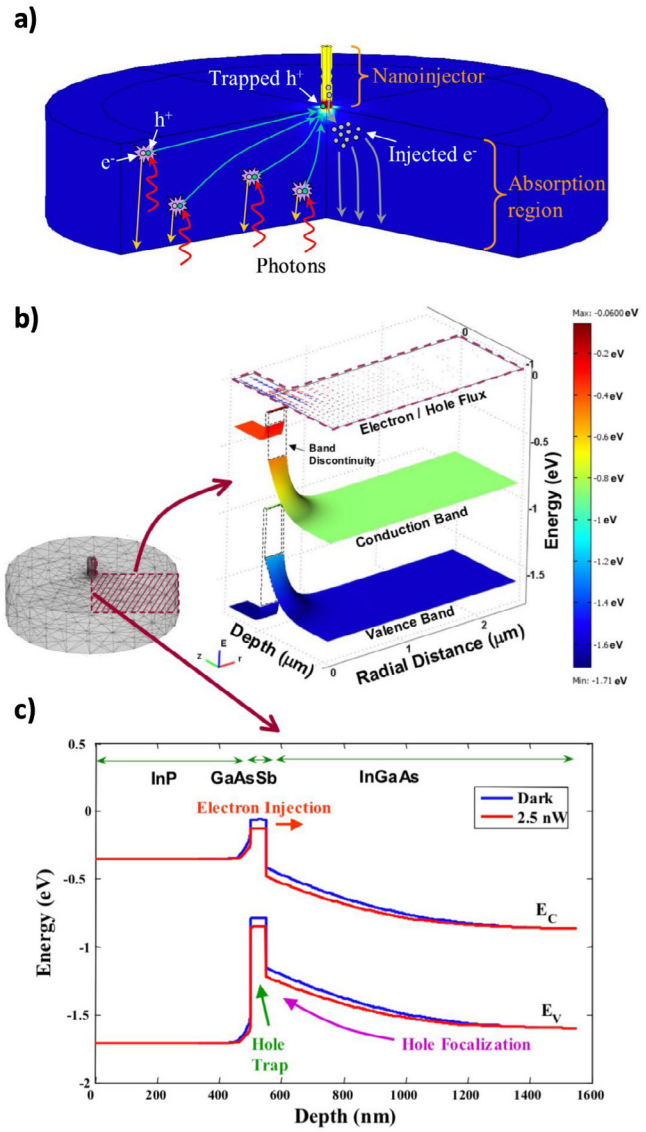

Figure 13. (a) Top-view schematic and 3D potential profile in a pump-gate jot. Reproduced with permission from [115]. © 2007 SPIE. (b) Cross sectional potential profiles along lines aa' and bb', and band diagram schematic of the two steps of pump action, corresponding to transfer gate (TG) on and off along the charge transfer path. Reproduced with permission from [116]. © IEEE: SW indicates the storage well, PB and PW a p-type barrier and well respectively, VB the virtual phase barrier, and FD the floating diffusion. Reprinted from [119], with the permission of AIP Publishing..

a two-step pump action is employed, similar to two-phase CCD and global-shutter CMOS imagers [112, 113]. The photogenerated carriers from a large-area absorber are first collected in a shallow storage well under a transfer gate (TG) during an integration time; subsequently, the potential of the TG is reverted, causing the charges to drift laterally onto the floating diffusion (FD). The compression of photogenerated charges in a small-volume FD allows to drastically reduce the overall sensor capacitance, and hence improved its sensitivity, achieving single-photon sensitivity in the visible spectral range [112].

Another example of implementation of the charge compression design strategy is that of the electron injector (EI) detectors [114]. This design encompasses an epitaxial phototransistor (PT) architecture, whose geometrical dimensions are engineered to maximize the device conversion gain [106]. A typical EI geometry and band structure is shown in figure 13: a large area absorption region corresponding to the PT collector is coupled to a small electronic volume consisting of the emitter and floating base, acting as charge injector for amplification [115]. The photogenerated charge is transported laterally to the multiplication region thanks to diffusion and built-in field, as shown in figure 13(b), and subsequently 'compressed' in a small-area trapping layer in the injector [116–118]. As a result, the floating base potential barrier is modulated, as shown in figure 13(c), similarly to the VB in the pump-gate jot device, enabling the gain of the device [119]. This approach has been successfully implemented for InGaAs-based SWIR detectors, exhibiting large gain and responsivity [120, 121]; moreover, the great uniformity and yield of the III–V material processing have enabled the demonstration of FPAs with performance surpassing that of currently available commercial cameras [114, 122].

3.2. Low-dimensional materials

In recent, a new wave of interesting developments has come from detectors based on nano-scale and low-dimensional (LD) materials, such as quantum dots (QD), nanowires and two-dimensional (2D) materials, shown in figure 14 [49, 123, 124]. These devices have attracted ever-increasing attention thanks to their unmatched responsivity [79] and bandwidth [50], potentially allowing them to reach single-photon sensitivity [125]. Typically, they rely on the same fundamental concept that has driven the development of charge compression devices: shrinking the size of the detectors in order to increase their sensitivity or confinement.

The exceptional charge confinement of LD materials offers unique optical and electronic properties. As an example, QD and 2D materials have been extensively used as efficient photon emitters [126, 127]. The amplification mechanisms, as shown in figure 14(b), is mostly similar to that of a PT, where the photogenerated carriers modulate a potential barrier at the interface or surface of the LD material, which can generate high amplification, thanks to the exceptional electronic properties of LD materials. The main limitation resides in the poor absorption in the intrinsically small volume of LD materials, which drastically affects the quantum efficiency of these devices. A successful implementation of LD materials for IR detection has been achieved through the integration of LD materials in conventional detector architectures: for example, colloidal QD can be employed as photo-sensitive CMOS gate layer of a field-effect transistor, where the absorption in the QD layer modulates the transistor turn-on potential, enabling the detection and amplification [44, 49]. While this approach has enabled the electronic integration of LD materials in IR detectors, it does not address the issue of poor light coupling

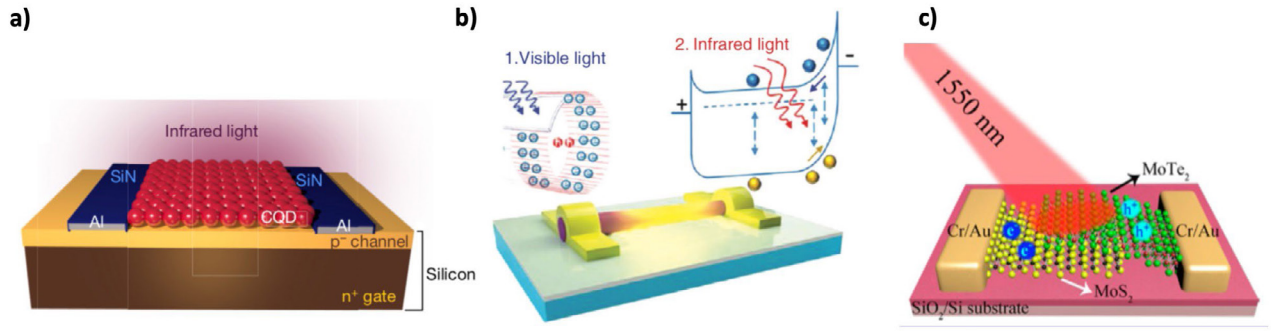

Figure 14. Low-dimensional photodetector architectures based on QD (a) Reprinted by permission from Macmillan Publishers Ltd: Nature [49], Copyright (2017), nanowires (b) Reprinted with permission from [123]. Copyright (2016) American Chemical Society, and 2D material (c) Reprinted with permission from [124]. Copyright (2016) American Chemical Society.

limiting the quantum efficiency. Therefore, a successful demonstration of IR detectors based on LD materials will require the implementation of strategies for enhancing optical coupling, in particular involving light enhancement and confinement in very small volumes. Such strategies include cavities, microlenses, metamaterials, plasmonic enhancement and optical antennas [44, 71]. Furthermore, the uniformity and yield of LD materials processing has not yet matched that of semiconductors and represents a major technological challenge to the implementation of large-scale and large-area LD-based IR detectors and imaging arrays.

4. Strategies for enhancing optical coupling

4.1. Anti-reflection and light trapping

Since most materials utilized for IR detection have a relatively large refractive index (3–4), a significant enhancement of the device quantum efficiency can be offered by minimizing the reflection at the surface [44]. Conventionally, anti-reflection (AR) coating can be implemented exploiting interference effects in a stack of thin film dielectric layers with varying refractive index to cancel out the reflection [128]. This approach however is fundamentally limited to low-angle and narrow spectral response, since its principle of operation is only effective for a small range of wave vectors, all perpendicular to the surface [129]. Recently, a novel approach to AR strategies has been introduced, which takes advantage of engineered surface texture which effectively act as graded-index structures [130]. This approach is capable of producing a broadband impedance matching layer, as well as efficient mode coupling to many modes of different polarization inside the detector volume, especially useful for detector materials with specific polarization, such as QWIP [131]. Moreover, this strategy can be implemented in a rather simple and cost-efficient techniques, based on self-ordering processes, such as reactive ion treatments [132]: an example of self-assembled Si nanostructured cones obtained with reactive ion etching is shown in figure 15 [133].

An even more significant performance enhancement, especially for detectors based on LD materials, can be generated by increasing the path of light through the detector, through effects known as light trapping. The most common such strategy is that of placing the detector active absorbing region between two highly reflective surfaces, as shown in figure 15(b), forming a cavity where light interacts with the active volume for longer time before leaving the detector [134, 135]. Because of reciprocity, however, this approach in incompatible with AR strategies, and the cavity usually comprises a perfect mirror and a distributed Bragg reflector (DBR). Furthermore, the resonant nature of light trapping will increase the spectral selectivity of the detectors, and limit its speed, which becomes limited by carrier transit time [136].

4.2. Enhanced light-matter interaction

This class of strategies encompasses the engineering of optoelectronics materials and systems to enhance the light absorption probability. It includes lensing and microlensing, plasmonic enhancement, photonics crystals, optical antennas, and metamaterials.

Lensing allows focusing the far field into a small area, increasing the fill factor of the detector and hence its quantum efficiency. As discussed in previous sections, small pixel size and pitch has increasingly been pursued in order to reduce the SWaP of the system by enabling smaller imagers size. For a fixed F-number ($F/\#$), the area, size and weigth of the optics required scale at least quadratically with the detector size [27, 30, 34]. In addition, it has been theoretically proven that diffraction-limited optics can benefit imaging systems for both long-range and large-field of view imaging modes [30, 31, 34, 39, 40]. As a result, reducing the size of the pixels, and hence of the imager and optics, has increasingly become a primary strategy for achieving efficient and light infrared imaging systems.

In imaging, two strategies for implementing microlensing arrays for FPA fabrication have been proposed, as shown in figures 16(a) and (b). In the first approach, microlenses of dimensions comparable to the pixel pitch are independently designed and fabricated, and subsequently integrated in the FPA using backside alignment through the detector substrate [137]. The alternative proposed solution is to monolithically fabricate the lensing array into the detector substrate, by etching a parabolic-shape structure, exploiting total internal reflection for light focusing, effectively acting like a parabolic mirror [45]. In recent years, a novel approach to light concentration has emerged, based on subwavelength dielectric and metallic structures, known as electromagnetic metamaterial

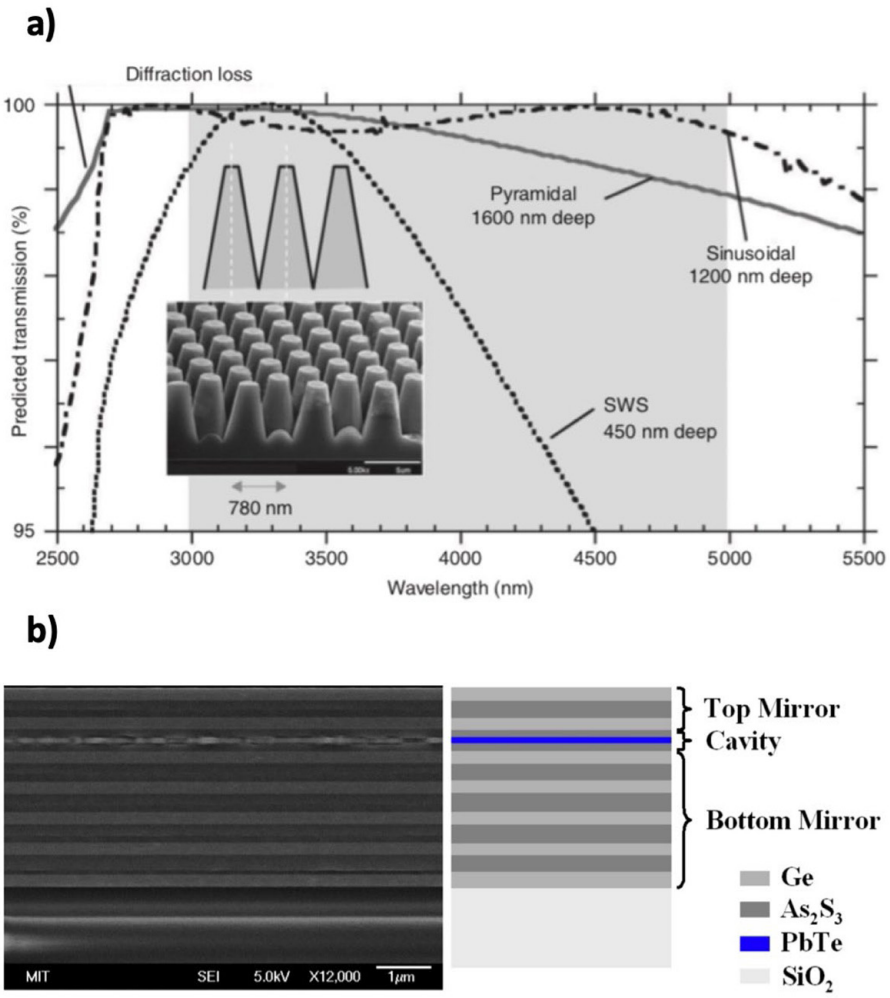

Figure 15. (a) Nano-structured Si cones for MWIR AR application. Reproduced with permission from [133]. © 2005 SPIE, (b) schematic of a cavity-based IR detector architecture for enhanced quantum efficiency, adapted from [134]. CC BY 3.0.

[73, 138]. Engineering the geometrical and materials properties of these nanoscale periodic structures allows access to domains of the electromagnetic properties that were previously forbidden, enabling new technological applications such as subwavelength cavities [139], detection directivity [140] and planar lensing [73]. As a result, lensing effect from quasi-flat metamaterials, such as shown in figure 16(c), have recently attracted increasing interest, for their wide range of applicability, and can also represent a viable solution for enhancing light coupling in FPAs [138, 141].

The second class of light enhancing strategies encompasses the use of effect that can create localized electromagnetic fields within subwavelength areas. Such effects include plasmonic and photonics crystals, and optical nanoantennas, which inherently exhibit broad spectral band. Plasmonic effects, collective charge oscillation at optical frequencies near metallic-dielectric boundaries and interfaces, allow to locally enhance the field coming from propagating light, into regions much smaller than its wavelength. Such effects can be used in periodic nanostructures, called plasmonic and photonic crystals (PC), to tailor the properties of light propagation and trapping to favor absorption. For example, PC allow to confine the incident light in a very narrow (subwavelength) mode inside the detector [142], especially useful for LD-based detectors with very thin absorbing layer (e.g. graphene [143]), as shown in figures 17(a) and (b). Moreover, PC can be employed to guide light in subwavelength volumes, and to produce cavity effects, as shown in figure 17(c) [144]. Optical antennas are structures that couples the propagating light into evanescent near field, favoring absorption and detection. These structures are particularly useful when utilizing detector materials with polarization-dependent absorption, such as quantum well superlattices (QWIP), since they allow compact conversion of polarization for a maximum coupling [71, 145]. Among common optical antennas are nanospheres [146], bowtie [147], and nanohole arrays [131]. Further engineering of these antennas have been performed by combining several of these approaches. For example, a novel metallo-dielectric hybrid antenna has been proposed, based on coupling a metallic cavity antenna coupled with a photonic jet produced by a dielectric microsphere, as shown in figure 17(d). This structure offers significant improvements in quantum efficiency and directivity, compared to bowtie optical antenna [148].

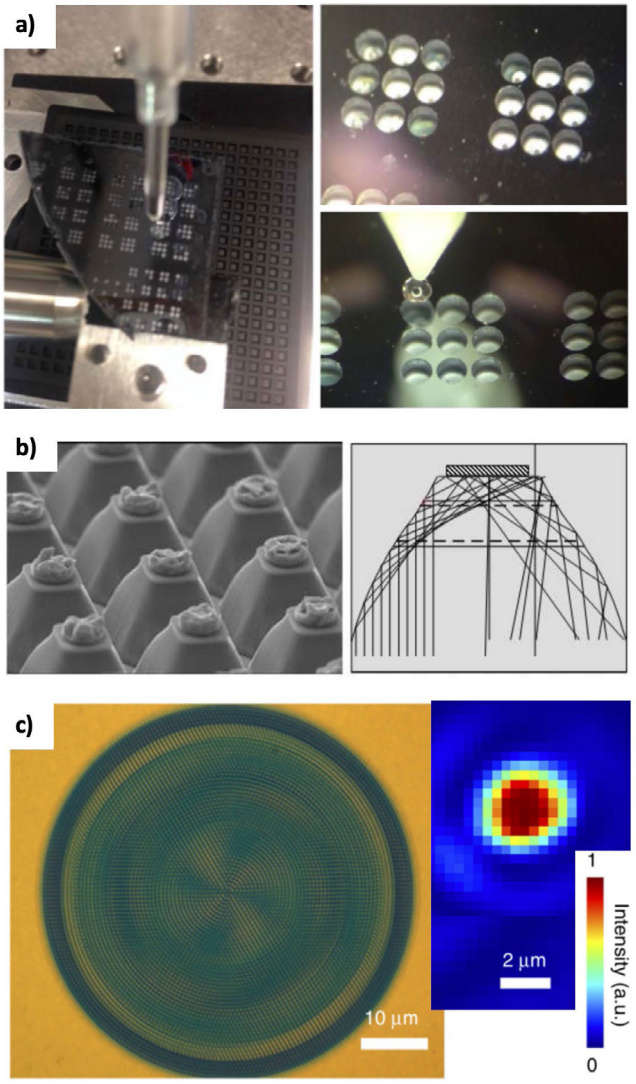

Figure 16. (a) Microscope image of the fabrication steps for microlens arrays. Reproduced with permission from [137]. © 2015 SPIE, (b) scanned electron image and schematic of light concentrators array for HgCdTe FPA [45]; (c) image of a quasi-flat achromatic lens based on space-varying dielectric gradient metastructure, with ~2 μm measured spot size from 1200 nm to 1600 nm wavelengths. Adapted from [141]. CC BY 4.0.

5. Challenges, strategies and potential future directions for SWIR imagers

The future success and impact of SWIR imaging technology will require three major technological advancements: reduced manufacturing costs, improved performance and advanced functionality. We here summarize some of the most promising strategies to address the fundamental hurdles in these three areas. We present an overview of the prospects for large area scalability and low-cost manufacturing, from a commercial and technological standpoint, with an attention to the most promising novel and future designs and concepts. We describe a promising strategy for enhancing the performance of the imagers, both in terms of low-light sensitivity and bandwidth, based on nanoscale and low-dimensional photodetectors with ultra-low capacitance. We review some of the proposed novel ROIC architectures that could enable advanced functionalities of SWIR imagers. Finally, we present some of the most promising and impactful potential applications of SWIR imagers, such as neuromorphic, stereoscopic and foveated vision, that can be enabled by leveraging the above mentioned advanced functionalities.

5.1. Manufacturing considerations: large-area scalability and yield

Scalability of the manufacturing processes to both large-area and highly parallelized processing are of fundamental importance for the future development of infrared imagers. As an example, the large-scale fabrication of CMOS imagers has enabled a huge cost reduction, which have significantly contributed to the success of these devices, that have today become ubiquitous [35, 36]. Similarly, high fabrication and integration yield is essential both for high-resolution imaging applications, and because of the high relative materials cost.

As discussed in a previous chapter, large part of the limitations on cost and large-area scalability of SWIR imagers is related to the need for an integration with a CMOS-based ROIC. This process introduces additional steps and risks to the fabrication, and currently prevents wafer-level processing. As such, most research efforts in both academia and industry are currently devoted to exploring alternative means of CMOS integration of the SWIR sensors. Recently, Ge-based SWIR detectors have attracted increasing atttention thanks to their compatibility with the Si material system, which allows epitaxial growth on CMOS. However, Ge detectors typically present higher dark current and lower absorption than its III–V competitors, such as InGaAs [20]. Wafer-level III–V to CMOS hybridization is very attractive since it can drastically reduce the cost of production of infrared imagers thanks to economy-of-scale [47]. Alternatively, large-area direct wafer-bonding of III–V semiconductors to Si and CMOS substrates has been demonstrated [18, 22, 24, 48]. This method is particularly attractive if combined with CMOS-compatible device processing, since all the fabrication can be performed on large CMOS wafer, with drastic reduction of the cost [18, 22]. Finally, bottom-up fabrication techniques for infrared detectors, such as self-assembly, colloidal dispersion and low-temperature growth on multiple substrates have recently attracted great intested due to their potential drastic reduction of the manufacturing costs and ease of CMOS integration [19, 49–52]. Typically, most bottom-up techniques do not guarantee large-area uniformity and scalability, however, some have been optimized over years of research and development and have come to compete in uniformity and cost with current state-of-the-art semiconductor fabrication. As an example, a new class of SWIR imager based on colloidal quantum dots has recently entered the market as affordable alternative to the current commercially dominant HgCdTe and III–V cameras [56–61]. These imagers can easily be implemented into multi-color pixel architectures by simply varying the size of the quantum dots [52]. Nanowires and thin-film SWIR detectors that have shown record responsivity can also be easily grown or deposited on CMOS substrates. Although large-area

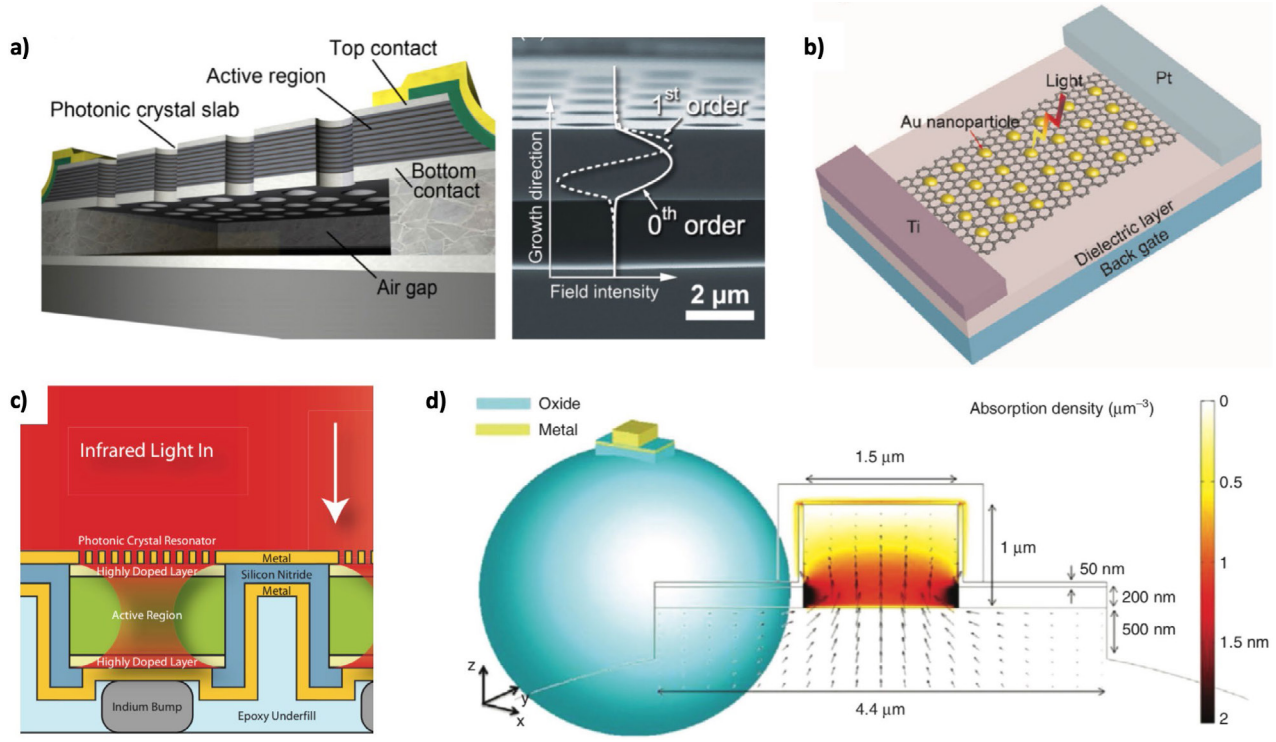

Figure 17. (a) Schematic of a nanoholes-array photonic crystal slab and modes of field distribution. Adapted from [142]. CC BY 4.0. (b) schematic of plasmonics-enhanced graphene photodetector architecture. Reprinted with permission from [143]. Copyright (2012) American Chemical Society, (c) schematic of PC-based resonator implementation in an FPA. Reproduced with permission from [144]. © 2010 SPIE, (d) schematic and FDTD field enhancement simulation for the hybrid metallo-dielectric photonic jet nanoantenna. Reproduced with permission from [148]. © 2013 SPIE.

uniformity has not yet been demonstrated for these materials, recent progress has shown promising results towards high uniformity large-area nanowire growth [105].

In conclusion, while the most reliable direction for near term commercial development is that of expanding the large-area capabilities of the fabrication and integration of conventional semiconductor-based technologies, it is undoubted that novel bottom-up fabrication techniques are already starting to play a significant role in the SWIR imagers market.

5.2. Ultra-low capacitance phototransistor arrays

Phototransistor detectors, including low-dimensional (LD) detectors, represent the most promising candidate for next-generation infrared imagers. They can amplify the signal to allow shot-noise limited performance, have superior performance at elevated teperatures, and have the potential for fabrication in large arrays. As shown in Ch.2, the sensitivity of these devices is directly related to their capacitance and can therefore be improved by shrinking the size of the detectors, with the potential for reaching photon-counting sensitivity even at room temperature [42]. The technological challenges posed by this approach are multiple. Currently, LD detectors cannot be fabricated into large arrays with good yield and uniformity, and conventional top-down fabrication methods based on epitaxial substrates incur limitations at very small dimensions. These issues are mainly related to the mechanical stability, alignment, and surface effects. As the size of the devices decreases, their surface-to-volume ratio increases, resulting in dominant contribution from surface effects processes such as trap-assisted recombination [80, 149]. Nevertheless, these do not represent fundamental limitations but rather technological, and recent advancements in this fronts have demonstrated the feasibility of the approach [105, 150, 151].

The key technological issues that need to be addressed for the successful fabrication of ultra-low capacitance infrared phototransistor arrays are:

- development of reliable, high-yield nanoscale fabrication methods
- alignment and electronic integration of nanoscale detectors
- development and characterization of appropriate passivation techniques

On the other hand, this sensitivity improvement will likely come at the price of a reduced quantum efficiency, due to poor coupling of light into such a small volume. Charge compression and transfer, allowing the integration of a large-volume photon absorber with a small amplification region, offers a viable solution to mitigate this drawback [85]. However, its effects are limited by parasitic capacitance and surface effects. A more compelling approach is that of enhancing the light coupling into the small sub-wavelength detector volume through strategies such as those presented in Ch.3, including optical antennas, photonic jets, integrated plasmonic. As an example, enhanced light detection and emission from quantum dots have been demonstrated by means of photonic jets [152] and surface plasmonic effect [153], suggesting that

a viable integration of nanoscale and low-dimensional detectors in large scale and camera format can be achieved thanks to sub-wavelength light enhancement techniques.

Finally, another proposed solution is that of decreasing the typical pixel pitch size of the FPA to match that of the small-volume detectors, allowing to increase the fill factor [154]: this can be achieved for instance by implementing a full three-dimensional CMOS ROIC, which is part of the following section.

5.3. Advanced ROIC

Generally, the ROIC electronics introduce additional noise in the imager system, which is called read noise. This noise is typically far exceeds the signal generated by a few photons, and hence prevented photon-counting sensitivities in many integrated imaging system [62]. However, photodetectors with internal gain mechanisms, such as the aforementioned phototransistors, overcome this limitation thanks to the amplification of the small photon signal to a larger electrical signal, which makes the read noise of modern ROIC negligible compared to the internal noise of the detectors.

Nevertheless, ROICs introduce other major limitations to the imaging system, which have not yet been addressed in full. Here we review some of the key challenges, and present the most promising ROIC designs, proposed to address each of these challenges.

5.3.1. Dynamic range limitations. In conventional ROIC designs, the photocurrent generated at each detector pixel is accumulated locally in a capacitor acting like a potential well (C_W). Since the maximum dynamic range of the camera is determined by the amount of charge that can be stored at the pixel, expressed as $Q = C_W V$, increasing the dynamic range requires either a larger capacitor or higher operating voltages. Both parameters are strictly limited in a conventional ROIC, because of pixel pitch size, leakage, and electrical discharges. Furthermore, the capacitance of the well and the SNR readout requirements also limit the speed at which the data output can be read, and hence require a large number of output taps for fast and efficient readout.

A solution that has been proposed and implemented is in-pixel signal digitization by means of analog-to-digital conversion (ADC) combined with a counter [155]. Beside increasing the effective well size (i.e. the dynamic range) by orders of magnitude and allowing for fast data rates, this approach enables enhanced functionality at the pixel level. These include real-time filtering and feature extraction, background suppression, and non-uniformity compensation [155]. An alternative proposed solution is based on differential pixel read for increasing the dynamic range, which also allows compensation of common mode noise [156].

5.3.2. 3D edgeless and deadzone-less tileable ROICs for large-area FPAs. Another limitation of current ROIC technology stems from the architecture design of the ROIC, which typically separates the in-pixel analog and digital circuitry

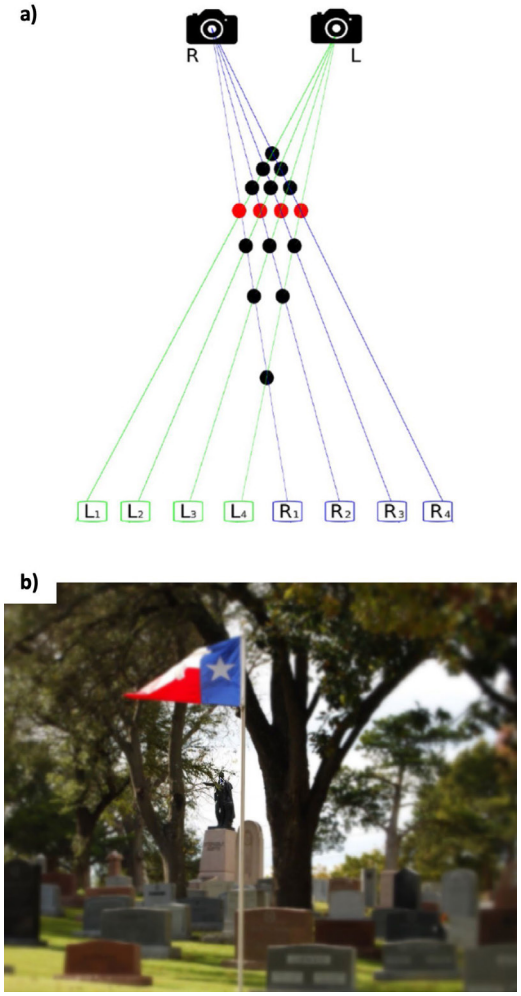

Figure 18. (a) Schematic of artificial stereo vision employing two cameras (left, L and right, R), with representation of a typical correspondence problem (the four red dots representing the solution). Reproduced from [163]. CC BY 3.0. (b) Example of foveated image with point of fixation on the Stephen F Austin statue in the background, adapted from en.wikipedia.org/wiki/Foveated_imaging#/media/File: Texas_state_cemetery_foveated1.png.

from the peripheral auxiliary circuitry, input-output and insulation layers. This ROIC design limits the scalability and integration of large-area detector arrays, as tiling multiple ROIC chips will result in deadzones within the detection area, which are highly undesirable, especially for high-sensitivity applications [157]. Edgeless, 3D-integrated advanced ROICs with negligible amount of edge inactive area, that can be tiled without causing detection deadzone can enable fully modular scalability of the detector size, with the potential for reaching beyond Megapixel array resolution [157].

A proposed approach encompasses multi-layered CMOS design that integrates the peripheral circuitry within a small area of each pixel, guaranteeing the full functionality at no cost in term of floor area occupation. The ROIC chip is hybridized with the detector via bump bonding, and with the electronic board by taking advantage of low-capacitance, low-noise through silicon vias (TSVs), achieving truly edgeless ROIC integration [53, 157]. More advanced 3D architectures,

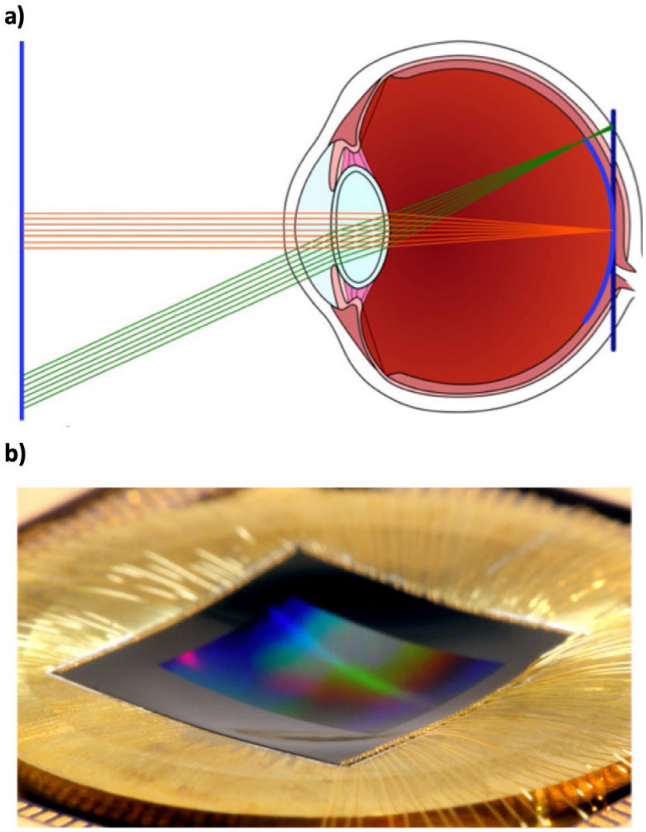

Figure 19. (a) Schematic of the eye's aberration-free curved imaging sensor, compared to the aberration from a single thick lens related with a planar imager. (b) 8 megapixel CMOS curved imager bonded to a precisely curved mold surface. Reproduced with permission from [164]. © The Optical Society.

including TSVs and bonding of a digital and analog chips together before integration into the camera and board assembly are also being pursued [157].

5.3.3. Redundancy in transmitted data. In many cases of interest, especially in high-sensitivity applications, the set of pixels containing useful information is very sparse. In such cases, the conventional ROIC logic implies a very inefficient readout scheme, that wastes power and bandwidth by reading out the 'empty' pixels signal.

As a solution, ROIC systems employing novel logic architecture have been proposed, to enable dynamic zero-suppression readout mode, enabling techniques such as image foveation [157]. In particular, reconfigurable systems that can evaluate and select the most appropriate imaging mode in real-time, to ensure efficient usage of the data bandwidth, are highly desirable in order to achieve a highly versatile infrared imaging system.

5.3.4. Timing accuracy. Recently, advanced pixel circuitry and functionality have been demonstrated, enabling photon counting and time-stamping at a pixel level. However, for efficient continuous operation, the readout process should be faster than the integration frame window for avoiding signal corruption or the need to introduce deadtime through a buffer. Conventional daisy-chaining of the pixel output in conjunction with prioritization algorithms are intrinsically limited in speed, hindering the accurate detection of fast events for time-resolved observations.

A proposed solution encompasses smart binary tree logics for prioritizing the readout and transmission, where the pixels are synchronized and entirely controlled by the output data transmitter through a series of switching pulses [158]. Furthermore, advanced re-synchronization techniques that allow for synchronous operation and resetting without corruption of the data that is being read out can represent an additional tool for tackling this issue [157].

5.4. Neuromorphic, stereoscopic and foveated SWIR vision systems, and bio-inspired curved imagers

State-of-the-art imaging systems are still significantly outperformed by biological systems when it come to live processing of large amount of information. Artificial vision, as an example, still struggles to recognize patterns and motion at the speed and efficiency of biological systems, despite its net superiority in terms of floating-point operation [159]. This is inherent to the image acquisition and processing architectures common to all digitally-integrated imagers and sensors. Specifically, the continuous analog data acquired at the sensors are typically digitized at the pixel level and subsequently transferred and processed in a sequential fashion. For this reason, even when imagers are combined with dedicated machine vision or pattern recognition algorithms, they cannot match the image virtuosity and efficiency of biological systems due to data transfer and memory access latencies, digitization artifacts and the lack of real-time feedback. Conversely, biological vision massively employs parallel, non-clocked, real-time processing, developing what is commonly referred to as early vision, to infer distance, surface and movements in the observed objects [159]. This is enabled by real-time data computation, such as filtering, edge-detection and object tracking, taking place at the level of single individual sensing element, thereby enabling real-time adjustment of adaptive photoreceptors. Neuromorphic vision is the attempt to apply the same image processing architecture to the state-of-the-art imaging technologies [160]. This is achieved relying on advanced ROIC systems, such as the ones discussed in the previous section, that are capable of performing complex tasks and computation at the pixel level, as well as to operate the detectors in different imaging modes, for example by changing the detector bias voltage, optical gain, pixel bin size or amplification mechanisms [161]. In addition, such advanced read-out circuitry can enable event-based detection, where a pixel is read out only when triggered by a stimulus above a certain threshold, particularly useful both for lowering data transfer density and for efficient sparse imaging, quite common, for example, in astronomy and particle physics. An example of this concept was applied in the implementation of so called retinomorphic vision sensor, where analog circuits can trigger based on the spatial or temporal gradients in the image, generating 'spiking' signals resembling those of neurons [162]. Bio-inpsired three-dimensional stereo vision, such as shown in

figure 18(a), can also take advantage of event-based imaging enabled by spiking networks, by adding time-resolved sensing to aid the efficient and accurate recognition of the matching points of the two or more stereo images (correspondence problem) [163]. Finally, another imaging processing technique inspired by the retina is foveation: varying resolution across the image, mimicking the resolution of the retina, based on one or more fixation points, as shown in figure 18(b) [53]. This is typically employed in post-processing for image compression, but if implemented at the pixel read-out level, it could help decrease the image memory requirements and data transfer density, allowing for considerable speedups in the image acquisition and processing.

Beside early vision and image processing that is inspired by nature, inspiration from the biological imaging sensors can also benefit the architecture and optics of the imagers. One feature of major technological interest that is currently being pursued is that of curved imaging sensors. These imagers instrinsically allow to avoid aberration from the lenses and optics, as shown in figure 19(a), and are therefore ubiquitous in biological vision systems [164]. This seemingly simple feature, however, posed a formidable challenge to the conventional FPA fabrication, which is traditionally based on planar substrates and processes. Several methods to overcome these limitations have been proposed with relative success, including bonding the detectors to flexible or curved molds, as shown in figure 19(b), as well as flexible detectors and mosaic tiling [165].

Acknowledgments

Authors would like to acknowldge the partial support by the W.M. Keck Foundation Engineering Award, ARO award #W911NF1810429, and NIH award #1R21EY029516-01. SB gratefully acknowledges support from the Ryan Fellowship and the International Institute for Nanotechnology at Northwestern University.

ORCID iDs

Simone Bianconi https://orcid.org/0000-0002-3828-6513
Hooman Mohseni https://orcid.org/0000-0002-0183-4213

References

[1] Fang Y, Armin A, Meredith P and Huang J 2019 Accurate characterization of next-generation thin-film photodetectors *Nat. Photon.* **13** 1–4
[2] Rogalski A 2010 *Infrared Detectors* (Boca Raton, FL: CRC Press)
[3] Herschel W 1800 XIV. Experiments on the refrangibility of the invisible rays of the sun *Phil. Trans. R. Soc.* **90** 284–92
[4] Case T W 1920 The thalofide cell: a new photoelectric substance *Phys. Rev.* **15** 1920
[5] Hudson R D and Hudson J W 1975 *Infrared Detectors* (Stroudsburg, PA: Dowden, Hutchinson & Ross)
[6] Lawson W D, Nielson S, Putley E H and Young A S 1959 Preparation and properties of HgTe and mixed crystals of HgTe–CdTe *J. Phys. Chem. Solids* **9** 325–9
[7] Kruse P W, Blue M D, Garfunkel J H and Saur W D 1962 Long wavelength photoeffects in mercury selenide, mercury telluride and mercury telluride-cadmium telluride *Infrared Phys.* **2** 53–60
[8] Guo W *et al* 2017 Counting near infrared photons with microwave kinetic inductance detectors *Appl. Phys. Lett.* **110** 212601
[9] Mazin B A *et al* 2013 ARCONS: a 2024 pixel optical through near-IR cryogenic imaging spectrometer *Publ. Astron. Soc. Pac.* **125** 1348–61
[10] Guyon O, Martinache F, Clergeon C, Russell R, Groff T and Garrel V 2011 Wavefront control with the subaru coronagraphic extreme adaptive optics (scexao) system *Astronomical Adaptive Optics Systems and Applications IV*, vol 8149 (International Society for Optics and Photonics) p 814908
[11] Guyon O, Martinache F, Cady E J, Belikov R, Balasubramanian K, Wilson D, Clergeon C S and Mateen M 2012 How ELTs will acquire the first spectra of rocky habitable planets *Adaptive Optics Systems III* vol 8447 (International Society for Optics and Photonics) p 84471X
[12] Hinz P *et al* 2012 TIGER: a high contrast infrared imager for the Giant Magellan Telescope *Ground-based and Airborne Instrumentation for Astronomy IV* vol 8446 p 84461P (International Society for Optics and Photonics)
[13] Turco R P 2002 *Earth Under Siege: from Air Pollution to Global Change* (New York: Oxford University Press)
[14] Strausbaugh R, Jackson R and Butler N 2018 Night vision for small telescopes *Publ. Astron. Soc. Pac.* **130** 2018
[15] Miller D A B 2017 Attojoule optoelectronics for low-energy information processing and communications *J. Lightwave Technol.* **35** 2017
[16] Miller D A B 2009 Device requirements for optical interconnects to silicon chips *Proc. IEEE* **97** 2009
[17] Miller D A B 1997 Physical reasons for optical interconnections *Int. J. Optoelectron.* **11** 155–68
[18] Roelkens G *et al* 2007 III–V/Si photonics by die-to-wafer bonding *Mater. Today* **10** 2007
[19] Takenaka M, Kim Y, Han J, Kang J, Ikku Y, Cheng Y, Park Y, Yoshida M, Takashima S and Takagi S 2017 Heterogeneous CMOS photonics based on SiGe/Ge and III–V semiconductors integrated on Si platform *IEEE J. Sel. Top. Quantum Electron.* **23** 2017
[20] Assefa S, Xia F and Vlasov Y A 2010 Reinventing germanium avalanche photodetectors for nanophotonic on-chip optical interconnects *Nat. Lett.* **464** 80–4
[21] Chen H T *et al* 2015 Hihg sensitivity 10Gb/s Si photonic receiver based on a low-voltage waveguide-coupled Ge avalanche photodetector *Opt. Express* **23** 815–22
[22] Settaluri K T *et al* 2015 Demonstration of an optical chip-to-chip link in a 3D integrated electronic-photonic platform *ESSCIRC*
[23] Kang Y *et al* 2008 Monolithic germanium/silicon avalanche photodiodes with 340 GHz gain-bandwidth product *Nat. Photon.* **247** 59
[24] Park M-S, Rezaei M, Nia I, Brown R, Bianconi S, Tan C L and Mohseni H 2018 InGaAs/InP quantum well infrared photodetector integrated on Si substrate by Mo/Au metal-assisted wafer bonding *Opt. Mater. Express* **8** 413–9
[25] Ginner L, Schmoll T, Kumar A, Salas M, Pricoupenko N, Wurster L M and Leitgeb R A 2018 Holographic line field en-face OCT with digital adaptive optics in the retina *in vivo Biomed. Opt. Express* **9** 472–85
[26] Fathipour V, Schmoll T, Bonakdar A, Wheaton S and Mohseni H 2017 Demonstration of shot-noise-limited swept source OCT without balanced detection *Sci. Rep.* **7** 1–9
[27] Rogalski A, Martyniuk P and Kopytko M 2016 Challenges of small-pixel infrared detectors: a review *Rep. Prog. Phys.* **79** 42

[28] Rogalski A 2011 Recent progress in infrared detector technologies *Infrared Phys. Technol.* **54** 136–54

[29] Kinch M A 2007 *Infrared Detector Materials* (Bellingham, WA: SPIE Press)

[30] Driggers R G, Vollmerhausen R, Reynolds J P, Fanning J and Colst H G 2012 Infrared detector size: how low should you go? *Opt. Eng.* **51** 2012

[31] Kinch M A 2015 The future of infrared; III–Vs or HgCdTe? *J. Electron. Mater.* **4** 2969–76

[32] Rogalski A 2005 HgCdTe infrared detector material: history, status and outlook *Rep. Prog. Phys.* **68** 2267–336

[33] Castelein P, Marion F, Martin J-L, Baylet J P, Moussy N, Gravrand O, Durand A, Chamonal J-P and Destefanis G L 2003 Megapixel HgCdTe MWIR focal plane array with a 15 μm pitch *Infrared Technology and Applications XXIX* vol 5074 (International Society for Optics and Photonics) pp 52–9

[34] Kinch M A 2014 *State-of-the-Art Infrared Detector Technology* (Bellingham, WA: SPIE Press)

[35] Clark R N 2005 Digital cameras: does pixel size matter? factors in choosing a digital camera (does sensor size matter?) (www.ClarkVision.com)

[36] Chen T 2000 How small should pixel size be? *Proc. SPIE* **3965** 451–9

[37] Waller J Complete guide to image sensor pixel size ePhotozine, 2016 (www.ephotozine.com/article/complete-guide-to-image-sensor-pixel-size-29652#Smartphone)

[38] Farrel J, Xiao F and Kavusi S 2006 Resolution and light sensitivity trade-off with pixel size *Proc. SPIE* **6069** 60690N

[39] Holts G 2007 Imaging system performance based upon Fλ/d *Opt. Eng.* **46** 103204

[40] Fiete R D 1999 Image quality FN/P and for remote sensing systems *Opt. Eng.* **38** 1229–40

[41] Kinch M A 2018 Room-temperature infrared focal plane array performance *J. Electron. Mater.* **47** 5879–88

[42] Rezaei M, Park M-S, Tan C L and Mohseni H 2017 Sensitivity limit of nanoscale phototransistors *IEEE Electron Device Lett.* **38** 1051–4

[43] Tokunaga A T, Simons D A and Vacca W D 2002 The Mauna Kea observatories near-infrared filter set. II. Specifications for a new JHKL′ M′ filter set for infrared astronomy *Publ. Astron. Soc. Pac.* **114** 180

[44] Tan C L and Mohseni H 2017 Emerging technologies for high performance infrared detectors *Nanophotonics* **7** 169–97

[45] Atkinson D, Hall D, Baranec C, Baker I, Jacobson S and Riddle S 2014 Observatory deployment and characterization of SAPHIRA HgCdTe APD arrays *High Energy, Optical, and Infrared Detectors for Astronomy VI* vol 9154

[46] Razeghi M 2010 *The MOCVD Challenge: A Survey of GaInAsP-InP and GaInAsP-GaAs for Photonic and Electronic Device Applications* (Boca Raton, FL: CRC Press)

[47] Mason W Wafer scale infrared detectors (WIRED) DARPA (www.darpa.mil/program/wafer-scale-infrared-detectors)

[48] Kang C-M, Kang S-J, Mun S-H, Choi S-Y, Min J-H, Kim S, Shim J-P and Lee D-S 2017 Monolithic integration of AlGaInP-based red and InGaN-based green LEDs via adhesive bonding for multicolor emission *Sci. Rep.* **7** 1–9

[49] Adinolfi V and Sargent E H 2017 Photovoltage field-effect transistors *Nature* **542** 324–7

[50] Ko W S, Bhattacharya I, Tran T-T D, Ng K W, Gerke S A and Chang-Hasnain C 2016 Ultrahigh responsivity-bandwidth product in a compact InP nanopillar phototransistor directly grown on silicon *Sci. Rep.* **6** 1–11

[51] Kim J-H, Aghaeimeibodi S, Richardson C J, Leavit R P, Englund D and Wak Es 2017 Hybrid integration of solid-state quantum emitters on a Silicon photonic chip *Nano Lett.* **17** 7394–400

[52] Temple D S, Hilton A and Klem E D 2016 Towards low-cost infrared imagers: how to leverage Si IC ecosystem *Proc. SPIE* **9989** 99890E

[53] Fahim F, Deptuch G W, Hoff J R and Mohseni H 2015 Design methodology: edgeless 3D ASICs with complex in-pixel processing for pixel detectors *Proc. SPIE* **9555** 95550M

[54] Wang Z, Lu L and Bovik A C 2003 Foveation scalable video coding with automatic fixation selection *IEEE Trans. Image Process.* **12** 243–54

[55] Klem E J, Gregory C, Temple D and Lewis J 2015 PbS colloidal quantum dot photodiodes for low-cost SWIR sensing *Proc. SPIE* **9451** 945104

[56] Buurma C, Pimpinella R E, Ciani A J, Feldman J S, Grein C H and Guyot-Sionnest P 2016 MWIR imaging with low cost colloidal quantum dot films *Proc. SPIE* **9933** 993303

[57] Hafiz S B, Scimeca M, Sahu A and Ko D-K 2019 Colloidal quantum dots for thermal infrared sensing and imaging *Nano Convergence* **6** 7

[58] SWIR Vision Systems 3021 E Cornwallis Road, Research Triangle Park, NC 27709 (www.swirvisionsystems.com)

[59] SPIE 2019 SWIR vision systems aims quantum-dot cameras at industrial imaging *SPIE News* (https://optics.org/news/9/3/44)

[60] Quantum dot SWIR to compete with InGaAs *Imaging Mach. Vis. Europe* March 21st 2019 (https://www.imveurope.com/news/quantum-dot-swir-compete-ingaas)

[61] Wallace J 2019 Quantum dots could make IR cameras cheaper, more widely available *Laser Focus World* (https://www.laserfocusworld.com/detectors-imaging/article/16566818/quantum-dots-could-make-ir-cameras-cheaper-more-widely-available)

[62] Janesick J R 2007 *Photon Transfer* (Bellingham, WA: SPIE Press)

[63] Ferrini R, Patrini M and Franchi S 1998 Optical functions from 0.02 to 6 eV of $Al_xGa_{1-x}Sb$/GaAs epitaxial layers *J. Appl. Phys.* **84** 1998

[64] Liu K, Chu J H and Tang D Y 1994 Composition and temperature dependence of the refractive index in $Hg_{1-x}Cd_xTe$ *J. Appl. Phys.* **75** 1994

[65] Jaksic Z and Jaksic O 1995 Simple approximation for absorption coefficient in degenerate HgCdTe in *Proc. MIEL '95 (Nis, Serbia)*

[66] Adachi S 1989 Optical dispersion relations for GaP, GaAs, GaSb, InP, InAs, InSb, $Al_xGa_{1-x}As$ and $In_{1-x}Ga_xAs_yP_{1-y}$ *J. Appl. Phys.* **66** 6030–40

[67] Konstantatos G, Badioli M, Gaudreau L, Osmond J, Bernechea M, De Arquer F P G, Gatti F and Koppens F H 2012 Hybrid graphene-quantum dot phototransistors with ultrahigh gain *Nat. Nanotechnol.* **7** 363–8

[68] Ackerman M M, Tang X and Guyot-Sionnest P 2018 Fast and sensitive colloidal quantum dot mid-wave infrared photodetectors *ACS Nano* **12** 7264–71

[69] Hu Y, Zhou J, Yeh P-H, Li Z, Wei T-Y and Wang Z L 2010 Supersensitive, fast-response nanowire sensors by using Schottky contacts *Adv. Mater.* **22** 3327–32

[70] Martynenko Y V and Ognev L I 2005 Thermal radiation from nanoparticles *Tech. Phys.* **11** 1522–4

[71] Bonakdar A and Mohseni H 2014 Impact of optical antennas on active optoelectronic devices *Nanoscale* 10961–74

[72] Stockman M *et al* 2018 Roadmap on plasmonics *J. Opt.* **20** 2018

[73] Jahani S and Jacob Z 2016 All-dielectric metamaterials *Nat. Nanotechnol.* **11** 2016

[74] McGregor W K 1978 On the radiation from small particles *J. Quantum Spectrosc. Radiat. Transfer* **19** 659–64

[75] Zhang H, Babichev A, Jacopin G, Lavenus P, Julien F, Egorov A Y, Zhang J, Pauporte T and Tchernycheva M 2014 Characterization and modeling of a ZnO nanowire ultraviolet photodetector with graphene transparent contact *J. Appl. Phys.* **114** 2014

[76] Kim C-J, Lee H-S, Cho Y-J, Kang K and Jo M-H 2010 Diameter-dependent internal gain in ohmic Ge nanowire photodetectors *Nano Lett.* **10** 2043–8

[77] Liu E, Long M, Zeng J, Luo W, Wang Y, Pan Y, Zhou W, Wang B, Hu W and Ni Z 2016 High responsivity phototransistors based on few-layer ReS_2 for weak signal detection *Adv. Funct. Mater.* **26** 1938–44

[78] Ryzhii V and Ryzhii M 2009 Graphene bilayer field-effect phototransistor for terahertz and infrared detection *Phys. Rev.* B **79** 2009

[79] Zhang A, Kim H, Cheng J and Lo Y-H 2010 Ultrahigh responsivity visible and infrared detection using silicon nanowire phototransistors *Nano Lett.* **10** 2117–20

[80] Konstantatos G and Sargent E H 2007 PbS colloidal quantum dot photoconductive photodetectors: transport, traps and gain *Appl. Phys. Lett.* **91** 173505

[81] Helme J P and Houston P A 2007 Analytical modeling of speed response of heterojunction bipolar phototransistors *J. Lightwave Technol.* **25** 2007

[82] Bianconi S, Rezaei M, Park M-S, Huang W, Tan C L and Mohseni H 2019 Engineering the gain-bandwidth product of phototransistor diodes *Appl. Phys. Lett.* **115** 051104

[83] Tavares S E 1966 A comparison of integration and low-pass filtering *IEEE Trans. Instrum. Meas.* **15** 1966

[84] Talghader J J, Gawarikar A S and Shea R P 2012 Spectral selectivity in infrared thermal detection *Light Sci. Appl.* **1** 2012

[85] Tu X, Xiao P, Kang L, Jiang C, Guo X, Jiang Z, Su R, Jia X, Chen J and We P 2018 Nb_5N_6 microbolometer for sensitive, fast-response, 2 μm detection *Opt. Express* **26** 2018

[86] Van Duzer T and Turner C W 1999 *Principles of Superconductive Devices and Circuits* (Englewood Cliffs, NJ: Prentice Hall)

[87] Gerrits T, Lum D J, Verma V, Howell J, Mirin R P and Nam S W 2018 Short-wave infrared compressive imaging of single photons *Opt. Express* **26** 2018

[88] Gerrits T, Lum D J, Verma V, Howell J, Mirin R P and Nam S W 2017 A short-wave infrared single photon camera *Computational Optical Sensing and Imaging* (Optical Society of America) pp CTu4B-5

[89] Annett J F 2004 *Superconductivity, Superfluids and Condensates* (Oxford: Oxford University Press)

[90] Lita A E, Miller A J and Nam S W 2008 Counting near-infrared single-photons with 95% efficiency *Opt. Express* **16** 2008

[91] Irwin K D 1995 An application of electrothermal feedback for high resolution cryogenic particle detection *Appl. Phys. Lett.* **66** 1998–2000

[92] Calkins B *et al* 2013 High quantum-efficiency photon-number-resolving detector for photonic on-chip information processing *Opt. Express* **21** 2013

[93] Gol'tsman G *et al* 2007 Middle-infrared to visible-light ultrafast superconducting single-photon detectors *IEEE Trans. Appl. Supercond.* **17** 2007

[94] Marsili F, Najafi F, Dauler E, Bellei F, Hu X, Csete M, Molnar R J and Berggren K K 2011 Single-photon detectors based on ultranarrow superconducting nanowires *Nano Lett.* **11** 2048–53

[95] Zmuidzinas J 2012 Superconducting microresonators: physics and applications *Annu. Rev. Condens. Matter Phys.* **3** 169–214

[96] Mazin B A, Bumble B, Meeker S R, O'Brien K, McHugh S and Langman E 2012 A superconducting focal plane array for ultraviolet, optical, and near-infrared astrophysics *Opt. Express* **20** 2012

[97] Mazin B A 2009 Microwave kinetic inductance detectors: the first decade *AIP Conf. Proc.* **1185** 135–42

[98] Day P K, LeDuc H G, Mazin B A, Vayonakis A and Zmuidzinas J 2003 A broadband superconducting detector suitable for use in large arrays *Nature* **425** 817–21

[99] Renker D 2006 Geiger-mode avalanche photodiodes: history, properties and problems *Nucl. Instrum. Methods Phys. Res.* A **567** 2006

[100] Finger G, Baker I, Downing M, Alvarez D, Ives D, Mehrgan L, Meyer M, Stegmeier J and Weller H J 2014 Development of HgCdTe large format MBE arrays and noise-free high speed MOVPE EAPD arrays for ground based NIR astronomy *Int. Conf. on Space Optics (Tenerife, Canary Islands, Spain)*

[101] Finger G, Baker I, Alvarez D, Ives D, Mehrgan L, Meyer M, Stegmeier J and Weller H J 2014 SAPHIRA detector for infrared wavefront sensing *Adaptive Optics Systems IV* vol 9148 (International Society for Optics and Photonics) p 914817

[102] Sze S 2008 *Semiconductor Devices: Physics and Technology* (New York: Wiley)

[103] Feautrier P and Gach J-L 2015 State of the art IR cameras for wavefront sensing using e-APD MCT arrays *Adaptive Optics for Extremely Large Telescopes 4–Conf. Proc.* vol 1

[104] Goebel S B, Hall D N B, Guyon O, Warmbier E and Jacobson S M 2018 Overview of the SAPHIRA detector for adaptive optics applications *J. Astron. Telesc. Instrum. Syst.* **4** 026001

[105] Farrell A C, Meng X, Ren D, Kim H, Senanayake P, Hsieh N Y, Rong Z, Chang T-Y, Azizur-Rahman K M and Huffaker D L 2019 InGaAs–GaAs nanowire avalanche photodiodes toward single-photon detection in free-running mode *Nano Lett.* **19** 582–90

[106] Fathipour V, Jang S J, Hassani Nia I and Mohseni H 2015 Impact of three-dimensional geometry on the performance of isolated electron-injection infrared detectors *Appl. Phys. Lett.* **106** 021116

[107] Razeghi M 2012 Superlattice sees colder objects in two colors and high resolution *SPIE Newsroom* (https://spie.org/news/4122-superlattice-sees-colder-objects-in-two-colors-and-high-resolution?SSO=1)

[108] Wei Y, Gin A, Razeghi M and Brown G J 2002 Type-II InAs/GaSb superlattice photovoltaic detectors with cutoff wavelength approaching 32 μm *Appl. Phys. Lett.* **81** 3675

[109] Huang E K-W, Hoang M-A, Chen G, Ramezani-Darvish S, Haddadi A and Razeghi M 2012 Highly selective two-color mid-wave and long-wave infrared detector hybrid based on type-II superlattices *Opt. Lett.* **37** 4744–6

[110] Hoang M-A, Dehzangi A, Adhikary S and Razeghi M 2016 High performance bias-selectable three-color short-wave/mid-wave/long-wave infrared photodetectors based on type-II InAs/GaSb/AlSb superlattices *Sci. Rep.* **6** 24144

[111] Ma J, Hondongwa D and Fossum E R 2014 Jot devices and the quanta image sensor *IEEE IEDMI*

[112] Ma J, Masoodian S, Starkey D A and Fossum E R 2017 Photon-number-resolving megapixel image sensor at room temperature without avalanche gain *Optica* **4** 1474–81

[113] Ma J and Fossum E R 2015 A pump-gate jot device with high conversion gain for a quanta image sensor *J. Electron Device Soc.* **3** 2015

[114] Memis O G, Kohoutek J, Wu W, Gelfand R M and Mohseni H 2010 Signal-to-noise performance of a short-wave infrared nanoinjector imager *Opt. Lett.* **35** 2010

[115] Mohseni H, Memis O G, Kong S C, Katsnelson A and Wu W 2007 A Novel SWIR detector with an ultra-high internal gain and negligible excess noise *Electro-Optical and Infrared Systems: Technology and Applications IV* vol 6737 (International Society for Optics and Photonics) p 67370W

[116] Memis O G, Wu W, Dey D, Katnelson A and Mohseni H 2007 Detailed numerical modeling of a novel infrared single photon detector for lambda >1 μm *IEEE NUSOD*

[117] Movassaghi Y, Fathipour V, Fathipour M and Mohseni H 2017 Analytical and numerical evaluation of electron-injection detector optimized for SWIR photon detection *J. Appl. Phys.* **121** 2017

[118] Movassaghi Y, Fathipour V, Fathipour M and Mohseni H 2016 Analytical modeling and numerical simulation of the short-wave infrared electron-injection detectors *Appl. Phys. Lett.* **108** 2016

[119] Memis O G, Katnelson A, Kong S-C and Mohseni H 2007 A photon detector with very high gain at low bias and at room temperature *Appl. Phys. Lett.* **91** 171112

[120] Fathipour V, Hassani Nia I, Bonakdar A and Mohseni H 2016 On the sensitivity of electron-injection detectors at low light level *IEEE Photonics J.* **8** 2016

[121] Fathipour V, Memis O G, Jang S J, Brown R L, Hassani Nia I and Mohseni H 2014 Isolated electron injection detectors with high gain and record low dark current at telecom wavelength *IEEE J. Sel. Top. Quantum Electron.* **20** 2014

[122] Memis O G, Kohoutek J, Wu W, Gelfand R M and Mohseni H 2010 A short-wave infrared nanoinjection imager with 2500 A/W responsivity and low excess noise *IEEE Photonics J.* **2** 2010

[123] Fang H *et al* 2016 Visible light-assisted high-performance mid-infrared photodetectors based on single InAs nanowire *Nano Lett.* **16** 2016

[124] Zhang K *et al* 2016 Interlayer transition and infrared photodetection in atomically thin type-II $MoTe_2/MoS_2$ van der Waals heterostructures *ACS Nano* **10** 3852–8

[125] Hadfield R H 2009 Single-photon detectors for optical quantum information applications *Nat. Photon.* **3** 696

[126] Mashford B S *et al* 2013 High-efficiency quantum-dot light-emitting devices with enhanced charge injection *Nat. Photon.* **7** 407–12

[127] Reed J C, Zhu A Y, Zhu H, Yi F and Cubukcu E 2015 Wavelength tunable microdisk cavity light source with a chemically enhanced MoS_2 emitter *Nano Lett.* **15** 2015

[128] Macleod H A 2001 *Thin-Film Optical Filters* (Bristol: Institute of Physics Publishing)

[129] Raguin D H and Morris G M 1993 Antireflection structured surfaces for the infrared spectral region *Appl. Opt.* **32** 1154–67

[130] Ono Y, Kimura Y, Ohta Y and Nishida N 1987 Antireflection effect in ultrahigh spatial-frequency holographic relief gratings *Appl. Opt.* **26** 1142–6

[131] Bonakdar A and Mohseni H 2013 Impact of optical antenna on infrared imagers *Infrared Phys. Technol.* **59** 142–5

[132] Zhu J, Yu Z and Burkhard G F 2009 Optical absorption enhancement in amorphous silicon nanowire and nanocone array *Nano Lett.* **9** 279–82

[133] Hobbs D S and MacLeod B D 2005 Design, fabrication, and measured performance of anti-reflecting surface textures in infrared transmitting materials *Window and Dome Technologies and Materials IX* vol 5786 (International Society for Optics and Photonics) pp 349–64

[134] Wang J, Hu J, Becla P, Agarwal A M and Kimerling L C 2010 Resonant-cavity-enhanced mid-infrared photodetector on a silicon platform *Opt. Express* **18** 12890–6

[135] Ju G W, Na B H, Choi H J, Park K W, Song Y M and Lee Y T 2015 RCEPD with enhanced light absorption by crown-shaped quantum well *IEEE Photonics Technol. Lett.* **27** 2015

[136] Ünlü M S, Onat B M and Leblebici Y 1995 Transient simulation of heterojunction photodiodes part II: analysis of resonant cavity enhanced photodetectors *J. Lightwave Technol.* **13** 406–15

[137] Rezaei M, Nia I H, Bonakdar A and Mohseni H 2015 Ultra-compact hourglass lens for integrated cameras *Current Developments in Lens Design and Optical Engineering XVI* vol 9578 (International Society for Optics and Photonics) p 957809

[138] Lin D, Fan P, Hasman E and Brongersma M L 2014 Dielectric gradient metasurfaces optical elements *Science* **345** 298–302

[139] Engheta N 2002 An idea for thin subwavelength cavity resonators using metamaterials with negative permittivity and permeability *IEEE Antennas Wirel. Propag. Lett.* **1** 10–3

[140] Valentine J, Zhang S, Zentgraf T, Ulin-Avila E, Genov D A, Bartal G and Zhang X 2008 Three-dimensional optical metamaterial with a negative refractive index *Nature* **455** 2008

[141] Wang S *et al* 2017 Broadband achromatic optical metasurfaces devices *Nat. Commun.* **8** 187

[142] Kalchmair S *et al* 2012 Detectivity enhancement in quantum well infrared photodetectors utilizing a photonic crystal slab resonator *Opt. Express* **20** 5622–8

[143] Bao Q and Loh K P 2012 Graphene photonics, plasmonics, and broadband optoelectronic devices *ACS Nano* **6** 3677–94

[144] Rosenberg J, Shenoi R V, Krishna S and Painter O 2010 Design of plasmonic photonic crystal resonant cavities for polarization sensitive infrared photodetectors *Opt. Express* **18** 3672–86

[145] Wu W, Bonakdar A and Mohseni H 2010 Plasmonic enhanced quantum well infrared photodetector with high detectivity *Appl. Phys. Lett.* **96** 161107

[146] Koederink A F 2009 Plasmon nanoparticle array waveguides for single photon and single plasmon sources *Nano Lett.* **9** 4228–33

[147] Schuck P J, Fromm D P, Sundaramurthy A, Kino G S and Moerner W E 2005 Improving the mismatch between light and nanoscale objects with gold bowtie nanoantennas *Phys. Rev. Lett.* **94** 017402

[148] Bonakdar A and Mohseni H 2013 Hybrid optical antenna with high directivity gain *Opt. Lett.* **38** 2726–8

[149] Callewaert F, Hoang A M and Razeghi M 2014 Generation-recombination and trap-assisted tunneling in long wavelength infrared minority electron unipolar photodetectors based on InAs/GaSb superlattice *Appl. Phys. Lett.* **104** 053508

[150] Park M-S, Rezaei M, Barnhart K, Tan C L and Mohseni H 2017 Surface passivation and aging of InGaAs/InP heterojunction phototransistors *J. Appl. Phys.* **121** 2017

[151] Gudisen M S, Lauhon L J, Wang J, Smith D C and Lieber C M 2002 Growth of nanowire superlattice structures for nanoscale photonics and electronics *Nature* **415** 617–20

[152] Bonakdar A, Rezaei M, Brown R L, Fathipour V, Dexheimer E, Jang S J and Mohseni H 2015 Deep-UV microsphere projection lithography *Opt. Lett.* **40** 2537–40

[153] Kim N-Y, Hong S-H, Kang J-W, Myoung N-S, Yim S-Y, Jung S, Lee K, Tu C W and Park S-J 2015 Localized surface plasmon-enhanced green quantum dot light-emitting diodes using gold nanoparticles *RSC Adv.* **5** 19624–9

[154] Bakulin A, D'Souza A I, Masterjohn C, Mei E, Li C, Klem E and Temple D 2018 ROIC for 3 μm Pixel Pitch Colloidal Quantum Dot Detectors *Image Sensing Technologies: Materials, Devices, Systems, and Applications V* vol 10656 (International Society for Optics and Photonics) p 1065614
[155] Schultz K I, Kelly M W, Baker J J, Blackwell M H, Brown M G, Colonero C B, David C L, Tyrrell B M and Wey J R 2014 Digital-pixel focal plan array technology *Lincoln Lab. J.* **20** 2014
[156] Ay S U, Barna S and Fossum E R 2001 Differential mode CMOS active pixel sensor (APS) for optically programmable gate array (OPGA) *Proc. 2001 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensor*
[157] Kmon P, Deptuch G, Fahim F, Grybos P, Maj P, Szczygiel R and Zimmerman T 2019 Active feedback with leakage current compensation for charge sensitive amplifier used in hybrid pixel detector *IEEE Trans. Nucl. Sci.* **66** 664–73
[158] Yang P *et al* 2015 Low-power priority address-encoder and reset-decoder data-driven readout for monolithic active pixel sensors for tracker systems *Nucl. Instrum. Methods Phys. Res.* A **785** 61–9
[159] Koch C and Mathur B 1996 Neuromorphic vision chips *IEEE Spectrum* **33** 38–46
[160] Mead C 1990 Neuromorphic electronic systems *Proc. IEEE* **78** 1629–36
[161] Mahowald M 1994 *An Analog VLSI System for Stereoscopic Vision* (New York: Springer)
[162] Posch C, Serrano-Gotarredona T, Linares-Barranco B and Delbruck T 2014 Retinomorphic event-based vision sensors: bioinspired cameras with spiking output *Proc. IEEE* **102** 1470–84
[163] Steffen L, Reichard D, Weinland J, Kaiser J, Roennau A and Dillmann R 2019 Neuromorphic stereo vision: a survey of bio-inspired sensors and algorithms *Frontiers Neurobotics* **13** 28
[164] Guenter B *et al* 2017 Highly curved image sensors: a practical approach for improved optical performance *Opt. Express* **25** 13010–23
[165] Iwert O and Delabre B 2010 The challenge of highly curved monolithic imaging detectors *High Energy, Optical, and Infrared Detectors for Astronomy IV* vol 7742 (International Society for Optics and Photonics) p 774227

Simone Bianconi received his B.S. and M.S. degrees in energy and nuclear engineering from the University of Bologna, Italy, in 2013 and 2015, respectively. He is currently working toward the Ph.D. degree in electrical engineering, solid-state and photonics at Northwestern University, Evanston, IL, where he has been a research assistant with the Bio-inspired Sensors and Optoelectronics Lab since 2016. His research interests include short-wave infrared single photon detectors and imagers, optical modulators, and novel materials for optoelectronics. He was awarded the Collegio Superiore fellowship by the University of Bologna, and he is a Ryan fellow of the International Institute of Nanotechnology since 2018.

Hooman Mohseni is the AT&T Chair Professor of Electrical and Computer Engineering, and Professor of Physics and Astronomy, at Northwestern University. He is the recipient of several research and teaching award including W.M. Keck Foundation Award, NSF CAREER Award, DARPA Young Faculty Award, and Northwestern Faculty Honor Roll. Mohseni has served at many international conference committees, scientific review panels, and editorial boards. He has published over 250 articles in major journals including Nature, Nano Letters, and ACS Nano. Mohseni has been involved in several startups as technology leader and CTO, including StarSight and Kernel. He holds 14 issued US and International patents. He is a Fellow of SPIE and OSA.