Optical and Acoustical Methods in Science and Technology

Infrared Detectors for the Future

A. Rogalski
Institute of Applied Physics, Military University of Technology
S. Kaliskiego 2, 00-908 Warsaw, Poland

In the paper, fundamental and technological issues associated with the development and exploitation of the most advanced infrared detector technologies are discussed. In this class of detectors both photon and thermal detectors are considered. Special attention is directed to HgCdTe ternary alloys on silicon, type-II superlattices, uncooled thermal bolometers, and novel uncooled micromechanical cantilever detectors. Despite serious competition from alternative technologies and slower progress than expected, HgCdTe is unlikely to be seriously challenged for high-performance applications, applications requiring multispectral capability and fast response. However, the nonuniformity is a serious problem in the case of LWIR and VLWIR HgCdTe detectors. In this context, it is predicted that type-II superlattice system seems to be an alternative to HgCdTe in long wavelength spectral region. In well established uncooled imaging, VO_x microbolometer arrays are clearly the most used technology. In spite of successful commercialization of uncooled microbolometers, the infrared community is still searching for a platform for thermal imagers that combine affordability, convenience of operation, and excellent performance. Recent advances in microelectromechanical systems have led to the development of uncooled IR detectors operating as micromechanical thermal detectors. Between them the most important are biomaterial microcantilevers.

PACS numbers: 42.79.Pw, 07.57.Kp, 73.21.Cd, 78.67.Pt, 79.60.Jv

1. Introduction

Hitherto, many materials have been investigated in the infrared (IR) field. Fig. 1 gives approximate dates of significant development efforts for the IR materials. The years during World War II saw the origins of modern IR detector technology. Photon IR technology combined with semiconductor material science, photolithography technology developed for integrated circuits, and the impetus of Cold War military preparedness have propelled extraordinary advances in IR capabilities within a short time period during the last century [1].

Two families of multielement detectors can be considered for principal military and civilian IR applications; one used for scanning systems and the other used for staring systems. The scanning system, which does not include multiplexing functions in the focal plane, belongs to the first generation systems.

The second generation systems (full-framing systems) have typically three orders of magnitude more elements ($> 10^6$) on the focal plane than first generation systems and the detectors elements are configured in a two-dimensional (2D) arrays. These staring arrays are scanned electronically by circuits integrated with the arrays. These devices are 2D arrays of photodiodes connected with indium bumps to a readout integrated circuit (ROIC) chip as a hybrid structure, often called a sensor chip assembly (SCA). Development of this technology began in the late 70's last century and took the next decade to reach volume production. In the early 1990's, fully 2D arrays provided a means for staring sensor systems to begin production. Large IR detector arrays are

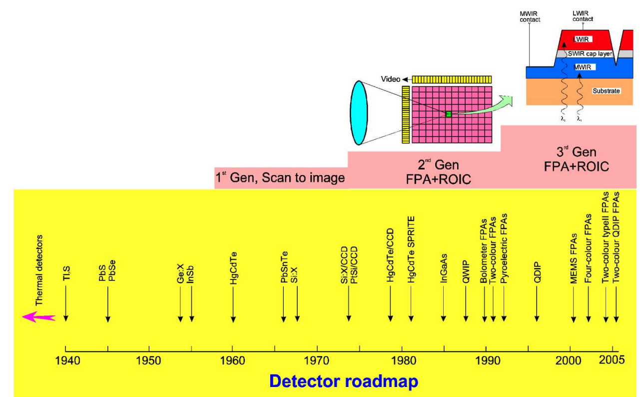

Fig. 1. History of the development of infrared detectors and systems. Three generation systems can be considered for principal military and civilian applications: 1st Gen (scanning systems), 2nd Gen (staring systems — electronically scanned) and 3rd Gen (multicolour functionality and other on-chip functions).

now available that meet the demanding requirements of the astronomy and civil space applications. Astronomers in particular have eagerly waited for the day when electronic arrays could match the size of photographic film. Development of large format, high sensitivity, mosaic IR sensors for ground-based astronomy is the goal of many observatories around the world (large arrays dramatically multiply the data output of a telescope system). This is somewhat surprising given the comparative budgets of the defence market and the astronomical community.

For the last 25 years array size has been increasing at an exponential rate, following a Moore law grow path (see Fig. 2), with the number of pixels doubling every 19 months [2, 3]. The graph shows the log of the number of pixels per SCA as a function of the year first used on astronomy for MWIR SCAs. Arrays exceeded $4\text{k} \times 4\text{k}$ format — 16 million pixels — in 2006, about a year later than the Moore law prediction. A subsequent expansion to $8\text{k} \times 8\text{k}$ with 10 μm pixel array is foreseen for 2009.

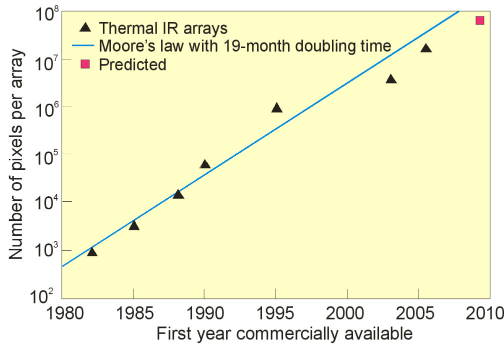

Fig. 2. The number of pixels on an infrared array has been growing exponentially, in accordance with Moore's law for 25 years with a doubling time of approximately 19 months. An $8\text{k} \times 8\text{k}$ array was predicted for 2009 but is likely at least a year later (after Refs. [1] and [2]).

The trend of increasing pixel's number is likely to continue in the area of large format arrays. This increase will be continued using close-butted mosaic of several SCAs as shown in Fig. 2. Raytheon manufactured a 4×4 mosaic of $2\text{k} \times 2\text{k}$ HgCdTe SCAs with 67 million pixels and assisted in assembling into the final focal-plane configuration (see Fig. 3) to survey the entire sky in the Southern Hemisphere at four IR wavelengths [2].

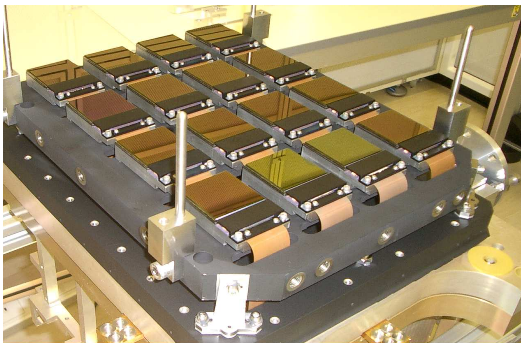

Fig. 3. Sixteen 2048×2048 HgCdTe SCAs were assembled for the VISTA telescope. The SCA are attached to a precision ground plate that ensures that all pixels are within 12 μm of the desired focus. The detectors are ready to be placed in the telescope camera's vacuum chamber and cooled to 72 K (after Ref. [1]).

While the size of individual arrays continues to grow, the very large focal plane arrays (FPAs) required for many space missions by mosaicking a large number of individual arrays. An example of a large mosaic developed by Teledyne Imaging Sensors, is a 147 megapixel FPA that is comprised of 35 arrays, each with 2048×2048 pixels. This is currently the world's largest IR focal plane [3]. Although there are currently limitations to reducing the size of the gaps between active detectors on adjacent SCAs, many of these can be overcome. It is predicted that focal plane of 100 megapixels and larger will be possible, constrained only by budgets, but not technology [4].

Multicolour detector capabilities are highly desirable for advanced IR imaging systems, since they provide enhanced target discrimination and identification, combined with lower false-alarm rates. Systems that collect data in separate IR spectral bands can discriminate both absolute temperature as well as unique signatures of objects in the scene. By providing this new dimension of contrast, multiband detection also offers advanced colour processing algorithms to further improve sensitivity above that of single-colour devices. This is extremely important for identifying temperature differences between missile targets, warheads, and decoys. Multispectral IR focal plane arrays (FPAs) are highly beneficial for a variety of applications such as missile warning and guidance, precision strike, airborne surveillance, target detection, recognition, acquisition and tracking, thermal imaging, navigational aids and night vision, etc. [5, 6]. They also play an important role in Earth and planetary remote sensing, astronomy, etc. [7].

Military surveillance, target detection, and target tracking can be undertaken using single-colour FPAs if the targets are easy to identify. However, in the presence of clutter, or when the target and/or background are uncertain, or in situations where the target and/or background may change during engagement, single-colour system design involves compromises that can degrade overall capability. It is well established that in order to reduce clutter and enhance the desired features/contrast, one will require the use of multispectral focal plane arrays. In such cases, multicolour imaging can greatly improve overall system performance.

Multispectral imaging systems will include very large sensors feeding an enormous amount of data to the digital mission processing subsystem. As these imaging arrays grow in detector number for higher resolution, so will the computing requirements for the embedded digital image processing system. One approach to solving this processing bottleneck problem could be to incorporate a certain amount of pixel-level processing within the detector pixel, similar to the technique implemented in biological sensor information processing systems. Currently, several scientific groups in the world have turned to the biological retina for answers as to how to improve man-made sensors [8, 9].

2. Third generation infrared systems

In the 1990's (see Fig. 1) third generation IR detectors emerged after the tremendous impetus provided by

detector developments. The definition of third generation IR systems is not particularly well established. In the common understanding, third generation IR systems provide enhanced capabilities such as larger number of pixels, higher frame rates, better thermal resolution, as well as multicolour functionality and other on-chip signal processing functions. According to Reago et al. [10], the third generation is defined by the requirement to maintain the current advantage enjoyed by the U.S. and allied armed forces. This class of devices includes both cooled and uncooled FPAs [11, 12]:

- high performance, high resolution cooled imagers having multi-colour bands,
- medium- to high-performance uncooled imagers,
- very low cost, expendable uncooled imagers.

When developing third generation imagers, the IR community is faced with many challenges. Some of them, such as:

- noise equivalent temperature difference (NEDT),
- pixel and chip size issues,
- uniformity, and
- identification and detection ranges

are considered in recently published paper [13].

2.1. Noise equivalent difference temperature

For infrared FPAs the relevant figure of merit is the NEDT. It can be shown that [14]:

$$\mathrm{NEDT} = \left(\tau C \eta_{\mathrm{BLIP}} \sqrt{N_{\mathrm{w}}}\right)^{-1}, \tag{1}$$

where τ is the optics transmission spectrum and C is the thermal contrast and η_{BLIP} is the ratio of photon noise to composite FPA noise. N_{w} is the number of photogenerated carriers integrated for one integration time, t_{int}, and Q_{B} is the photon flux density incident on the detector area A_{d}:

$$N_{\mathrm{w}} = \eta A_{\mathrm{d}} t_{\mathrm{int}} Q_{\mathrm{B}}. \tag{2}$$

The contrast in the MWIR bands at 300 K is 3.5–4% compared to 1.6% for the LWIR band.

From the above formulae, an important result is that the charge handling capacity of the readout, the integration time linked to the frame time, and dark current of the sensitive material becomes the major issues limiting the performance of IR FPAs. The NEDT is inversely proportional to the square root of the integrated charge, and therefore the greater the charge, the higher the performance.

The distinction between integration time and the FPA frame time must be noted. At high backgrounds it is often impossible to handle the large amount of carriers generated within a frame time compatible with standard video rates. Off-FPA frame integration can be used to attain a level of sensor sensitivity that is commensurate with the detector-limited D^* and not the charge-handling-limited D^*. Even though the detectivity of LWIR detectors is background limited, the ROIC can collect only about 1% of the charge within the unit cell for a flux of 10^{16}–10^{17} photons/(cm^2 s). Unit cell capacitors fill up in about 100 μs, while the frame time is on the order of 10 ms.

The well charge capacity is the maximum amount of charge that can be stored on the storage capacitor of each unit cell. The size of the unit cell is limited to the dimensions of the detector element in the array. Usually, it is assumed that the integration time is such that the readout node capacity is maintained at half full. For a 30×30 μm^2 pixel size, the storage capacities are limited to 1 to 5×10^7 electrons. For LWIR HgCdTe FPAs the integration time is usually below 100 μs. Since the noise power bandwidth $\Delta f = (1/2)t_{\mathrm{int}}$, a short integration time results in extra noise in the integration process.

Current readout technology is based upon CMOS circuitry that has benefited from dramatic and continuing progress in miniaturizing circuit dimensions. Second generation imagers provide NEDT of about 20–30 mK with $f/2$ optics. A goal of third-generation imagers is to achieve sensitivity improvement corresponding to NEDT of about 1 mK. From Eq. (1) it can be determined that in a 300 K scene in the LWIR region with thermal contrast of 0.02, the required charge storage capacity is above 10^9 electrons. This high charge-storage density cannot be obtained within the small pixel dimensions using standard CMOS capacitors. Although the reduced oxide thickness of submicrometer CMOS design rules gives large capacitance per unit area, the reduced bias voltage largely cancels any improvement in charge storage density. Ferroelectric capacitors may provide much greater charge storage densities than the oxide-on-silicon capacitors now used. However, such a technology is not yet incorporated into standard CMOS foundries.

To provide an opportunity to significantly increase both, the charge storage capacity and the dynamic range, the vertically-integrated sensor array (VISA) program has been sponsored by DARPA [15, 16]. The approach being developed builds on the traditional "hybrid" structure of a detector with a 2D array of indium-bump interconnects to the silicon readout. VISA allows additional layers of silicon processing chips to be connected below the readout to provide more complex functionality. It will allow the use of smaller and multicolour detectors without compromising storage capacity. Signal-to-noise ratios will increase for multicolour focal plane arrays. This will permit LWIR focal planes arrays to improve the sensitivity by a factor of ten.

2.2. Uniformity

It is well known that, when the detectivity is approaching a value above 10^{10} cm Hz$^{1/2}$/W, the FPA performance is uniformity limited prior to correction and thus essentially independent of the detectivity (see Fig. 4). An

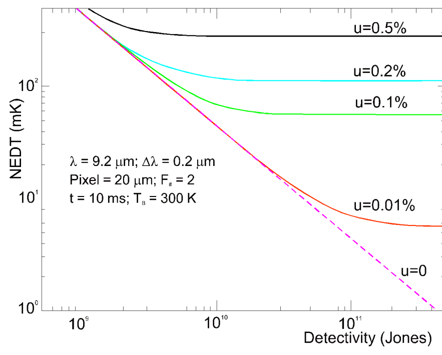

Fig. 4. NEDT as a function of detectivity. The effects of nonuniformity are included for $u = 0.01\%$, 0.1%, 0.2% and 0.5%. Let us note that for $D^* > 10^{10}$ cm $\mathrm{Hz}^{1/2}$/W, detectivity is not the relevant figure of merit.

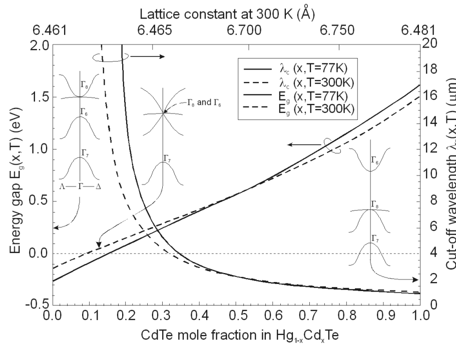

Fig. 5. The band gap structure of $\mathrm{Hg}_{1-x}\mathrm{Cd}_x\mathrm{Te}$ near the Γ-point for three different values of the forbidden energy gap. The energy band gap is defined at the difference between the Γ_6 and Γ_8 band extremes at $\Gamma = 0$.

improvement in nonuniformity from 0.1% to 0.01% after correction could lower the NEDT from 63 to 6.3 mK.

The nonuniformity value is usually calculated using the standard deviation over mean, counting the number of operable pixels in an array. For a system operating in the LWIR band, the scene contrast is about 2%/K of change in scene temperature. Thus, to obtain a pixel to pixel variation in apparent temperature to less than, e.g. 20 mK, the nonuniformity in response must be less than 0.04%. This is nearly impossible to obtain in the uncorrected response of the FPA, so a two-point correction is typically used.

FPA uniformity influences an IR system complexity. The uniformity is important for accurate temperature measurements, background subtraction, and threshold testing. Nonuniformities require elaboration of compensation algorithms to correct the image and by consuming a number of analog-to-digital bits they also reduce the system dynamic range.

Figure 5 plots the energy band gap, $E_g(x,T)$, for $\mathrm{Hg}_{1-x}\mathrm{Cd}_x\mathrm{Te}$ versus alloy composition x for temperature of 77 K. Also plotted is the cut-off wavelength $\lambda_c(x,T)$, defined as that wavelength at which the response has

TABLE I

Cut-off wavelength, λ_c, for x variations of 0.1% and the corresponding λ_c shift for $\mathrm{Hg}_{1-x}\mathrm{Cd}_x\mathrm{Te}$ at 77 K.

Composition x	Cut-off wavelength λ_c [μm]	Uncertainty $\Delta\lambda_c$ [μm]
0.395	3	0.012
0.295	5	0.032
0.210	10	0.131
0.196	14	0.257
0.187	20	0.527

dropped to 50% of its peak value. Table I shows uncertainty in cut-off wavelength for x variations of 0.1%. This variation in x value is typical of a good material. For short wavelength IR ($\approx$ 3 μm) and MWIR ($\approx$ 5 μm) materials, the variation in cut-off wavelength is not large. However, the nonuniformity is a serious problem in the case of VLWIR HgCdTe detectors. The variation of x across the $\mathrm{Hg}_{1-x}\mathrm{Cd}_x\mathrm{Te}$ wafer causes much larger spectral nonuniformity; e.g. at 77 K, a variation of $\Delta x = 0.1\%$ gives a $\Delta\lambda_c = 0.032$ μm at $\lambda_c = 5$ μm, but $\Delta\lambda_c = 0.53$ μm at 20 μm, which cannot be fully corrected by the two or three point corrections [6]. Therefore required composition control is much more stringent for LWIR than for MWIR. For applications that require operation in the LWIR band as well as two-colour LWIR/VLWIR bands most probably HgCdTe will not be the optimal solution.

Alternative candidate for third generation IR detectors is Sb-based III–V material system. This material is mechanically robust and has a fairly weak dependence of band gap on composition (see Fig. 6).

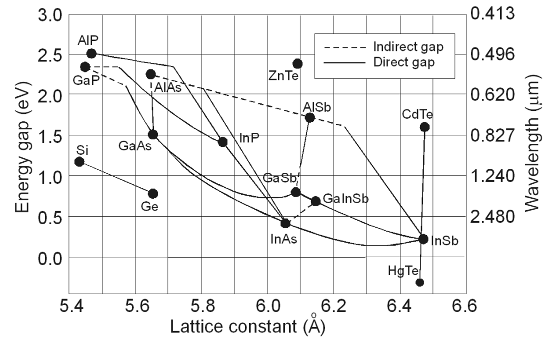

Fig. 6. Composition and wavelength diagram of Sb-based III–V material systems.

3. Material systems for future infrared detectors

In the wavelength regions of interest such as MWIR, LWIR and VLWIR, three detector technologies that are developing multicolour capability are visited here: HgCdTe, quantum well infrared photoconductors (QWIPs) and antimonide based type-II superlattices.

Both HgCdTe photodiodes and QWIPs have demonstrated the multicolour capability in the MWIR and LWIR range. Each of these technologies has its advantages and disadvantages. QWIP technology is based on the well-developed $A^{III}B^{V}$ material system, which has a large industrial base with a number of military and commercial applications. HgCdTe material system is only used for detector applications. Therefore QWIPs are easier to fabricate with high yield, high operability, good uniformity and lower cost. On the other hand, HgCdTe FPAs have higher quantum efficiency, higher operating temperature and potential for the highest performance. So far, however, the HgCdTe process yield in molecular beam epitaxy (MBE) growth of complex multiplayer structures, especially those working in the LWIR and VLWIR is low, pixel outage rates high and the material is consequently very expensive.

3.1. HgCdTe

The HgCdTe ternary alloy is a close to ideal infrared detector material system. Its unique position is dependent on three key features [1, 17]:

- composition-dependent tailorable energy band gap over the entire 1–30 μm range,
- large optical coefficients that enable high quantum efficiency, and
- favourable inherent recombination mechanisms that lead to long carrier lifetime and high operating temperature.

These properties are a direct consequence of the energy band structure of this zinc-blende semiconductor. Moreover, additional specific advantages of HgCdTe are the ability to obtain both low and high carrier concentrations, high mobility of electrons, and low dielectric constant. The extremely small change of lattice constant with composition makes it possible to grow high quality layered and graded gap structures. As a result, HgCdTe can be used for detectors operating in various modes, photoconductor, photodiode or metal–insulator–semiconductor (MIS) detector.

Third-generation HgCdTe IR systems have emerged as a result of technological achievements in the growth of heterostructure devices used in the production of second-generation IR focal plane arrays. Dual band detectors are grown by MBE on lattice matched CdZnTe wafers (see Fig. 7) [18, 14, 19–22]. Recently, Raytheon Vision Systems (RVS) has developed two-colour, large-format infrared FPAs to support the US Army's third generation FLIR systems. RVS has produced 1280×720 MWIR/LWIR FPAs with a 20 μm pixel pitch and cut-offs ranging out to 11 μm at 78 K. These FPAs have demonstrated excellent sensitivity and pixel operabilities exceeding 99.9% in the MW band and greater than 98% in the LW band. Table II provides a summary of the sensitivity and operability data measured for the three best 1280×720 FPAs fabricated to date [23]. Median 300 K NEDT values at $f/3.5$ of approximately 20 mK for the MW and 25 mK for the LW have been measured for dual-band TDMI operation at 60 Hz frame rate with integration times corresponding to roughly 40% (MW) and 60% (LW) of full well charge capacities.

TABLE II

Performance summary of three best 1280×720 MW/LW FPAs fabricated to date (after Ref. [23]).

FPA	Wafer	MW t_{int} [ms]	MW median NETD [mK]	MW response operability	LW t_{int} [ms]	LW median NETD [mK]	LW response operability
7607780	3827	3.14	23.3	99.7%	0.13	30.2	98.5%
7616474	3852	3.40	18.0	99.8%	0.12	27.0	97.0%
7616475	3848	3.40	18.0	99.9%	0.12	26.8	98.7%

However, near lattice matched CdZnTe substrates have severe drawbacks such as lack of large area, high production cost and, more importantly, a difference in thermal expansion coefficient (TEC) between the CdZnTe substrates and the silicon readout integrated circuit. Furthermore, interest in large area two-dimensional IR FPAs (1024×1024 and larger) have resulted in limited applications of CdZnTe substrates.

Currently, readily producible CdZnTe substrates are limited to areas of approximately 50 cm^2. At this size, the wafers are unable to accommodate more than two 1024×1024 FPAs. Not even a single dye can be accommodated for very large FPA formats (2048×2048 and larger) on substrates of this size.

The use of Si substrates is very attractive in IR FPA technology not only because it is less expensive and available in large area wafers, but also because the coupling of the Si substrates with Si readout circuitry in an FPA structure allows fabrication of very large arrays exhibiting long-term thermal cycle reliability. The 7×7 cm^2 bulk CdZnTe substrate is the largest commercially available, and it is unlikely to increase much larger than its present size.

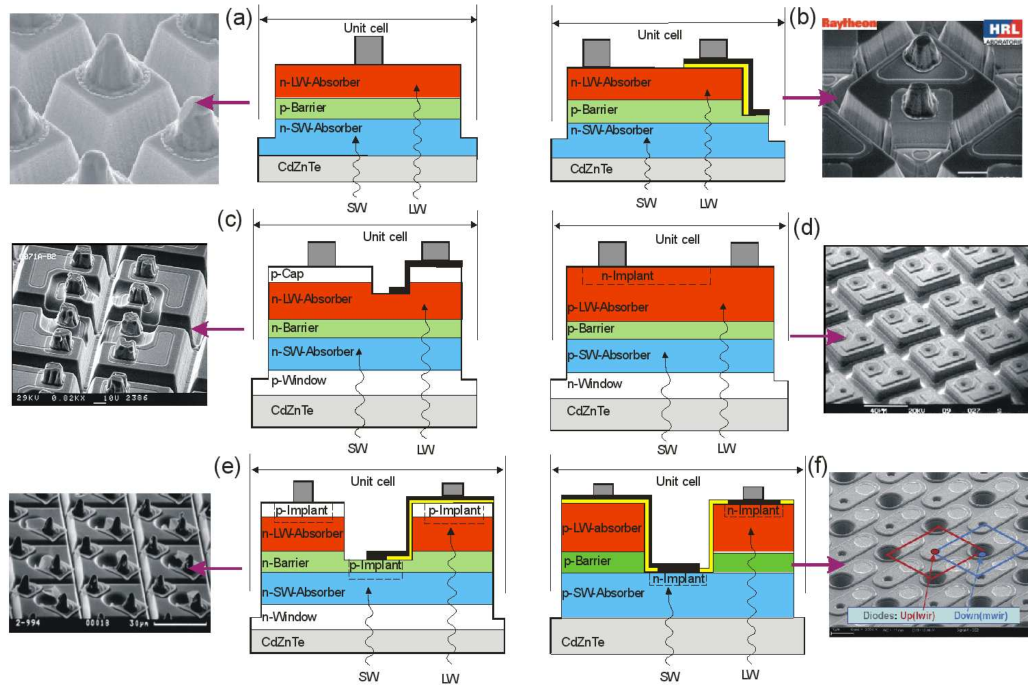

Fig. 7. Cross-section views of unit cells for various back-illuminated dual-band HgCdTe detector approaches: (a) bias--selectable n–p–n structure reported by Raytheon [17], (b) simultaneous n–p–n design reported by Raytheon [18], (c) simultaneous p–n–n–p reported by BAE Systems [19], (d) simultaneous n–p–p–p–n design reported by Leti [20], (e) simultaneous structure based on p-on-n junctions reported by Rockwell [21], and (f) simultaneous structure based on n-on-p junctions reported by Leti [22].

With the cost of 6 inch Si substrates being ≈ \$100 vs. \$10 000 for the 7×7 cm^2 CdZnTe, significant advantages of HgCdTe/Si are evident [24]. Despite the large lattice mismatch (≈ 19%) between CdTe and Si, MBE has been successfully used for the heteroepitaxial growth of CdTe on Si. Using optimized growth condition for Si(211)B substrates and a CdTe/ZnTe buffer system, epitaxial layers with etch pit density (EPD) in the 10^6 cm^{-2} range have been obtained. This value of EPD has little effect on both MWIR and LWIR HgCdTe/Si detectors [18, 14]. By comparison, HgCdTe epitaxial layers grown by MBE or LPE on bulk CdZnTe have typical EPD values in the 10^4 to mid-10^5 cm^{-2} range where there is a negligible effect of dislocation density on detector performance. At 77 K, diode performance with cut-off wavelength in LWIR region for HgCdTe on Si is comparable to that on bulk CdZnTe substrates [25].

3.2. QWIPs

An alternative hybrid detector for the middle and long wavelength IR region are the QWIPs. Despite large research and development efforts, large photovoltaic LWIR HgCdTe FPAs remain expensive, primarily because of the low yield of operable arrays. The low yield is due to sensitivity of LWIR HgCdTe devices to defects and surface leakage, which is a consequence of basic material properties. With respect to HgCdTe detectors, GaAs/AlGaAs quantum well devices have a number of potential advantages, including the use of standard manufacturing techniques based on mature GaAs growth and processing technologies, highly uniform and well-controlled MBE growth on greater than 6 in GaAs wafers, high yield and thus low cost, more thermal stability, and extrinsic radiation hardness.

LWIR QWIP cannot compete with HgCdTe photodiode as the single device, especially at higher temperature operation (> 70 K) due to fundamental limitations associated with intersubband transitions. QWIP detectors have relatively low quantum efficiencies, typically less than 10%. The spectral response band is also narrow for this detector, with a full-width, half-maximum of about 15%. All the QWIP data with cut-off wavelength about 9 μm is clustered between 10^{10} and 10^{11} cm Hz$^{1/2}$/W at about 77 K operating temperature. However, the advantage of HgCdTe is less distinct in temperature range below 50 K due to the problems involved in an HgCdTe material (p-type doping, Shockley–Read recombination, trap-assisted tunnelling, surface and interface instabilities). Table III compares the essential properties of three types of devices at 77 K.

TABLE III

Essential properties of LWIR HgCdTe and type-II SL photodiodes, and QWIPs at 77 K.

Parameter	HgCdTe	QWIP (n-type)	InAs/GaInSb SL
IR absorption	normal incidence	$E_{\text{optical}} \perp$ plane of well required normal incidence: no absorption	normal incidence
quantum efficiency	$\geq 70\%$	$\leq 10\%$	≈ 30–40%
spectral sensitivity	wide-band	narrow-band (FWHM $\approx 1 \div 2$ μm)	wide-band
optical gain	1	0.2–0.4 (30–50 wells)	1
thermal generation lifetime	≈ 1 μs	≈ 10 ps	≈ 0.1 μs
R_0A product ($\lambda_c = 10$ μm)	300 Ω cm^2	10^4 Ω cm^2	100 Ω cm^2
detectivity ($\lambda_c = 10$ μm, FOV $= 0$)	2×10^{12} cm Hz$^{1/2}$ W^{-1}	2×10^{10} cm Hz$^{1/2}$ W^{-1}	5×10^{11} cm Hz$^{1/2}$ W^{-1}

Even though QWIPs are photoconductors, several its properties such as high impedance, fast response time, and low power consumption comply well with the requirements for large FPAs fabrication. The main drawbacks of LWIR QWIP FPA technology are the performance limitation for low integration time applications and low operating temperature. Their main advantages are linked to performance uniformity and to availability of large size arrays. The large industrial infrastructure in III–V materials/device growth, processing, and packaging brought about by the utility of GaAs-based devices in the telecommunications industry gives QWIPs a potential advantage in producibility and cost. The only known use of HgCdTe, to this date, is for IR detectors.

State of the art QWIP and HgCdTe FPAs provide similar performance figure of merit, because they are predominantly limited by the readout circuits. It can be shown that NEDT value for charge-limited QWIP detectors is even better than that of HgCdTe photodiodes by a factor of $(2g)^{1/2}$, where g is the photoconductive gain, since a reasonable value of g is 0.4 [26]. The above deduction was confirmed experimentally by a research group at Fraunhofer IAF. Based on the photovoltaic low--noise four-zone QWIP structure, the Fraunhofer group [27, 28] has manufactured a 256×256 FPA camera operating at 77 K with the 9 μm cut-off wavelength. The camera exhibits record-low NEDT values of 7.4 mK with 20 ms integration time and 5.2 mK with 40 ms. It is the best temperature resolution ever obtained in the LWIR regime.

The very short integration time of LWIR HgCdTe devices of typically below 300 μs is very useful to freeze a scene with rapidly moving objects. QWIP devices achieve, due to excellent homogeneity and low photoelectrical gain, an even better NEDT, however, the integration time must be 10 to 100 times longer for that, and typically it is 5–20 ms. Decision of the best technology is therefore driven by the specific needs of a system. Even HgCdTe photodiodes intrinsically exhibit higher performance than the QWIP detectors, QWIP detectors are used for large formats (e.g. 1024×1024 and larger) with low frame rates and large integration time. Recently, 1 megapixel hybrid MWIR and LWIR QWIPs with 18 μm pixel size have been demonstrated with excellent imaging performance [29, 30]. The MWIR detector array has demonstrated a NEDT of 17 mK at a 95 K operating temperature with $f/2.5$ optics at 300 background and the LWIR detector array has demonstrated a NEDT of 13 mK at a 70 K operating temperature with the same optical and background conditions as the MWIR detector array [30]. This technology can be extended to a 2k $\times$ 2k array, but at present the limitation is the readout availability and cost.

QWIPs are ideal detectors for the fabrication of pixel co-registered simultaneously readable two-colour IR FPAs because a QWIP absorbs IR radiation only in a narrow spectral band and is transparent outside of that absorption band [31, 32]. Thus it provides zero spectral cross-talk when two spectral bands are more than a few μm apart. Devices capable of simultaneously detecting two separate wavelengths can be fabricated by vertical stacking of the different QWIP layers during epitaxial growth. Separate bias voltages can be applied to each QWIP simultaneously via doped contact layers that separate the MQW detector heterostructures.

Powerful possibilities of QWIP technology are connected with multicolour detection. A four-band FPAs has been demonstrated by stacking different multi--quantum well structures, which are sensitive in 4–6, 8.5–10, 10–12, and 13–15 μm bands [33]. The 640×512 format FPA consists of four 640×128 pixel areas which are capable of acquiring images in these bands. Four separate detector bands were defined by a deep trench etch process and the unwanted spectral bands were eliminated by a detector short-circuiting process. The unwanted top detectors were electrically shorted by gold-coated reflective 2D etched gratings.

The success of quantum well structures for infrared detector applications has stimulated the development of QDIPs. In general, QDIPs are similar to QWIPs but with the quantum wells replaced by quantum dots, which have size confinement in all spatial directions. QDIPs theoretically have several advantages compared with QWIPs including the normal incidence response,

lower dark current, higher operating temperature, higher responsivity and detectivity. Theoretical predictions indicate even that QDIPs are expected to compete with HgCdTe photodiodes [34]. However, comparison of theoretically predicted and experimental data indicates that, as so far, the QDIP devices have not demonstrated their potential advantages [35]. The main disadvantage of QDIPs is the large inhomogeneous linewidth of the quantum-dot ensemble variation of dot size in the Stranski–Krastanow growth mode. Poor QDIP performance is generally linked to nonoptimal band structure and controlling the QDs size and density (nonuniformity in QD size).

3.3. Type-II superlattices

InAs/$Ga_{1-x}In_xSb$ (InAs/GaInSb) strained layer superlattices (SLSs) can be considered as an alternative to the HgCdTe and GaAs/AlGaAs IR material systems and as a candidate for third generation IR detectors. The low quantum efficiency of QWIPs is largely due to fact that the optical transition is forbidden for normal incidence of light. Straylight generated by reflecting gratings is required to achieve reasonable quantum efficiency. On the other hand, this straylight degrades the modulation transfer function of QWIPs since some light intensity is guided by the residual substrate into neighbours. In the case of InAs/GaInSb SLS structures the absorption is strong for normal incidence of light. Consequently, the SLS structures provide high responsivity, as already reached with HgCdTe, without any need for gratings. Further advantages are a photovoltaic operation mode, operation at elevated temperatures and well established III–V process technology.

The InAs/GaInSb material system is however in a very early stage of development. Problems exist in material growth, processing, substrate preparation, and device passivation. Optimization of SL growth is a trade-off between interface roughness, with smoother interfaces at higher temperature, and residual background carrier concentrations, which are minimized on the low end of this range. The thin nature of InAs and GaInSb layers (< 8 nm) necessitate low growth rates for control of each layer thickness to within 1 (or 1/2) monolayer (ML). Monolayer fluctuations of the InAs layer thickness can shift the cut-off wavelength by about ± 2 μm for a 20 μm designed cut-off. Typical growth rates are less than 1 ML/s for each layer.

The type-II superlattice has staggered band alignment such that the conduction band of the InAs layer is lower than the valence band of InGaSb layer, as shown in Fig. 8. This creates a situation in which the energy band gap of the superlattice can be adjusted to form either a semimetal (for wide InAs and GaInSb layers) or a narrow band gap (for narrow layers) semiconductor material. In the SL, the electrons are mainly located in the InAs layers, whereas holes are confined to the GaInSb layers. This suppresses Auger recombination mechanisms and thereby enhances carrier lifetime. Optical transitions occur spatially indirectly and, thus, the optical matrix element for such transitions is relatively small. The band gap of the SL is determined by the energy difference between the electron miniband E_1 and the first heavy hole state HH_1 at the Brillouin zone centre and can be varied continuously in a range between 0 and about 250 meV. An example of the wide tunability of the SL is shown in Fig. 8b.

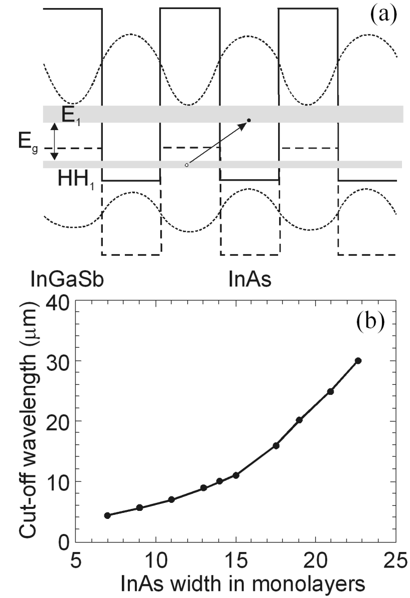

Fig. 8. InAs/GaInSb strained layer superlattice: (a) band edge diagram illustrating the confined electron and hole minibands which form the energy band gap; (b) change in cut-off wavelength with change in one superlattice parameter — InAs layer width (after Ref. [36]).

It has been suggested that InAs/$Ga_{1-x}In_xSb$ SLSs material system can have some advantages over bulk HgCdTe, including lower leakage currents and greater uniformity [37]. Electronic properties of SLSs may be superior to those of the HgCdTe alloy. The effective masses are not directly dependent on the band gap energy, as it is the case in a bulk semiconductor. The electron effective mass of InAs/GaInSb SLS is larger ($m^*/m_0 \approx 0.02$–0.03, compared to $m^*/m_0 = 0.009$ in HgCdTe alloy with the same band gap $E_g \approx 0.1$ eV). Thus, diode tunnelling currents in the SL can be reduced compared to the HgCdTe alloy. Although in-plane mobilities drop precipitously for thin wells, electron mobilities approaching 10^4 cm^2/(V s) have been observed in InAs/GaInSb superlattices with the layers less than 40 Å thick. While mobilities in these SLs are found to be limited by the same interface roughness scattering mechanism, detailed band structure calculations reveal a much weaker dependence on layer thickness, in reasonable agreement with experiment [38].

Theoretical analysis of band-to-band Auger and radiative recombination lifetimes for InAs/GaInSb SLSs showed that in these objects the p-type Auger recombi-

nation rates are suppressed by several orders, compared to those of bulk HgCdTe with similar band-gap [39, 40], but n-type materials are less advantageous. In p-type superlattice, the Auger rates are suppressed due to lattice--mismatch-induced strain that splits the highest two valence bands (the highest light band lies significantly below the heavy hole band and thus limits available phase space for the Auger transitions). In n-type superlattice, the Auger rates are suppressed by increasing the InGaSb layer widths, thereby flattening the lowest conduction band and thus limiting available phase space for the Auger transition. However, the promise of the Auger suppression has not yet to be observed in practical device material.

Comparison of theoretically calculated and experimentally observed lifetimes at 77 K for 10 μm InAs/GaInSb SLS and 10 μm HgCdTe indicates on good agreement for carrier densities above 2×10^{17} cm^{-3}. The discrepancy between both types of results for lower carrier densities is due to the Shockley–Read (SR) recombination processes having a $\tau \approx 6 \times 10^{-9}$ s which has been not taken into account in the calculations. For higher carrier densities, the superlattice (SL) carrier lifetime is two orders of magnitude longer than in HgCdTe, however in low doping region (below 10^{15} cm^{-3}, necessary in fabrication high performance p-on-n HgCdTe photodiodes) experimentally measured carrier lifetime in HgCdTe is more than two orders of magnitude longer than in SL. More recently published upper experimental data [41–43] coincide well with HgCdTe trend-line in the range of lower carrier concentration. In general however, the SL carrier lifetime is limited by SR mechanism with influence of trap centres located at an energy level of $\approx 1/3$ band gap below the effective conduction band edge [41].

Narrow band gap materials require the doping to be controlled to at least 1×10^{15} cm^{-3} or below to avoid deleterious high-field tunnelling currents across reduced depletion widths at temperature below 77 K. Lifetimes must be increased to enhance carrier diffusion and reduce related dark currents. At present stage of development, the residual doping concentration (both n-type as well as p-type) is typically about 5×10^{15} cm^{-3} in superlattices grown at substrate temperature ranging from 360°C to 440°C [44]. Low to mid 10^{15} cm^{-3} residual carrier concentrations are the best that have been achieved so far.

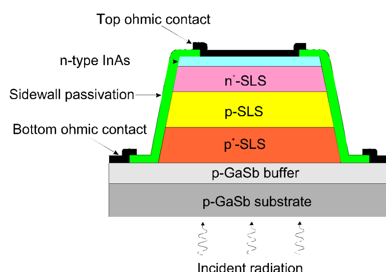

Fig. 9. Cross-section schematic of p–i–n InAs/GaInSb superlattice photodiode.

InAs/GaInSb SL photodiodes are typically based on p–i–n structures with an unintentionally doped, intrinsic region between the heavily doped contact portions of the device. A cross-section scheme of a completely processed mesa detector is presented in Fig. 9. The layers are usually grown by MBE at substrate temperatures around 400°C on undoped (001) oriented two inch GaSb substrates. With the addition of cracker cells for the group V sources, the superlattice quality becomes significantly improved. Despite the relatively low absorption coefficients, GaSb substrates require thinning the thickness below 25 μm in order to transmit appreciable IR radiation. Since the GaSb substrates and buffer layers are intrinsically p-type, the p-type contact layer, intentionally doped with beryllium at an acceptor concentration of 1×10^{18} atoms/cm^3, is grown first (see Fig. 8).

Sensors for the MWIR and LWIR spectral ranges are based on binary InAs/GaSb short-period superlattices [36, 45]. The layers needed are already so thin that there is no benefit to using GaInSb alloys. For the formation of p–i–n photodiodes the lower periods of the SL are p-doped with 1×10^{17} cm^{-3} Be in the GaSb layers. These acceptor doped SL layers are followed by a 1 to 2 μm thick, nominally undoped, superlattice region. The width of the intrinsic region does vary in the designs. The width used should be correlated to the carrier diffusion lengths for improved performance. The upper of the SL stack is doped with silicon (1×10^{17} to 1×10^{18} cm^{-3}) in the InAs layers and is typically 0.5 μm thick. The top of the SL stack is then capped with an InAs:Si ($n \approx 10^{18}$ cm^{-3}) layer to provide good ohmic contact.

To approach cut-off wavelengths in the 8-to-12 μm wavelength range also the InAs/GaInSb superlattice p–i–n photodiodes, with the indium molar fraction in the ternary GaInSb layers close to 20%, are fabricated [44].

The main technological challenge for the fabrication of small area size photodiodes is the occurrence of surface leakage currents mainly due to tunnelling electrons. The mesa side walls are a source of excess currents. Several materials and passivation processes have been explored. Some of the more prominent thin films studied have been silicon nitride, silicon oxides, ammonium sulfide and most recently, aluminium gallium antimonide alloys [36]. Rehm et al. [46] have chosen and demonstrated the good results achieved with lattice matched AlGaAsSb overgrowth by MBE on etched mesas. It is expected that the exposed side walls are larger band gap material, which tends to generate less excess currents.

In general, the inversion potentials are bigger for higher band gap materials, and therefore SiO_2 can passivate high band gap materials (MWIR photodiodes) but not low band gap material (LWIR photodiodes). Using this property, a double heterostructure that prevents the inversion of the high band gap p-type and n-type superlattice contact regions has been proposed with low--temperature ion-sputtered SiO_2 passivation [47]. Passivation with polyimide has also proved very effective [48].

An alternate method of eliminating excess currents due to side walls is shallow-etch mesa isolation with band-graded junction [49]. The primary effect of the grading is to suppress tunnelling and generation-recombination currents in the depletion region at low temperatures. Since both processes depend exponentially on band gap, it is highly advantageous to substitute a wide gap into depletion region. In this approach, the mesa etch terminates at just past the junction and exposes only a very thin (300 nm), wider band gap region of the diode. Subsequent passivation is therefore in wider gap material. As a result, it reduces electrical junction area, increases optical fill factor, and eliminates deep trenches within detector array.

The performance of LWIR photodiodes in the high temperature range is limited by diffusion process. Figure 10 shows the experimental data and theoretical prediction of the R_0A product as a function of temperature for InAs/GaInSb photodiode with 11 μm cut-off wavelength. The photodiodes are depletion region (generation-recombination) limited in temperature range between 80 K and 50 K. The trap-assisted tunnelling is dominant only at low temperature (< 50 K) with almost constant activation trap density (1×10^{12} cm^{-3}) [41]. Similar results have been recently published [42, 43]. The estimated effective carrier lifetime in the depletion layer is several tens of nanoseconds.

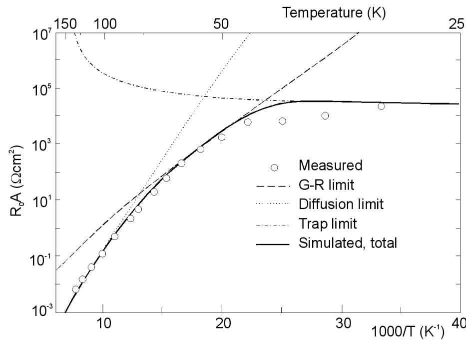

Fig. 10. Experimental data and theoretical prediction of the R_0A product as a function of temperature for InAs/GaInSb photodiode with 11 μm cut-off wavelength. The activated trap density is taken as a constant (1×10^{12} cm^{-3}) in the simulation over the whole temperature range (after Ref. [41]).

Optimization of the SL photodiode architectures is still an open area. Since some of the device design parameters depend on material properties, like carrier lifetime and diffusion lengths, these properties are still being improved. Also additional design modifications dramatically improve the photodiode performance. For example, Aifer et al. have reported W-structured type-II superlattice (WSL) LWIR photodiodes with R_0A values comparable for state-of-the-art HgCdTe [49]. In this design illustrated in Fig. 11a, the AlSb barriers are replaced with shallower $Al_{0.40}Ga_{0.49}In_{0.11}Sb$ quaternary barrier layers. In such structure two InAs "electron-wells" are located on either side of an InGaSb "hole-well" and are bound on either side by AlGaInSb "barrier" layers. The barriers confine the electron wave functions symmetrically about the hole-well, increasing the electron-hole overlap while nearly localizing the wave functions. The resulting quasi-dimensional densities of states give the WSL its characteristically strong absorption near band-edge. The new design W-structured type-II SL photodiodes employ a graded band-gap p–i–n design. The grading of the band gap in the depletion region suppresses tunnelling and generation-recombination currents in the depletion region which have resulted in an order of magnitude improvement in dark current performance, with $R_0A = 216\ \Omega$ cm^2 at 78 K for devices with a 10.5 μm cut-off wavelength.

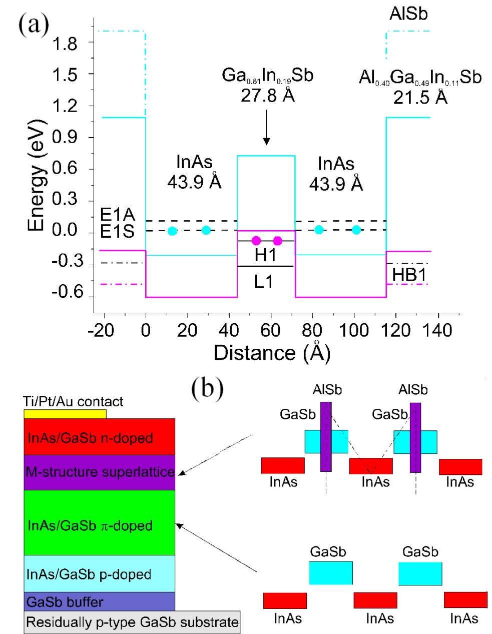

Fig. 11. Schematic diagrams of modified type-II LWIR photodiodes: (a) band profiles at $k = 0$ of enhanced WSL (after Ref. [49]), (b) p–π–M–n SL (band alignment of standard and M shape superlattices are shown) (after Ref. [50]).

Another type-II superlattice photodiode design with the M-structure barrier is shown in Fig. 11b. This structure significantly reduces the dark current, and on the other hand, does not show a strong effect on the optical properties of the devices [50]. The AlSb layer in one period of the M structure, having a wider energy gap, blocks the interaction between electrons in two adjacent InAs wells, thus, reducing the tunnelling probability and increasing the electron effective mass. The AlSb layer also acts as a barrier for holes in the valence band and converts the GaSb hole-quantum well into a double quantum well. As a result, the effective well width is reduced, and the hole's energy level becomes sensitive to the well

dimension. Device with a cut-off wavelength of 10.5 μm exhibits a R_0A product of 200 Ω cm^2 when a 500 nm thick M structure was used.

Figure 12 compares the R_0A values of InAs/GaInSb SL and HgCdTe photodiodes in the long wavelength spectral range. The solid line denotes the theoretical diffusion limited performance of p-type HgCdTe material. As it can be seen in the figure, the most recent photodiode results for SL devices rival that of practical HgCdTe devices, indicating substantial improvement has been achieved in SL detector development.

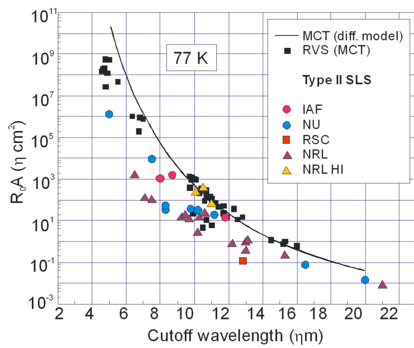

Fig. 12. Dependence of the R_0A product of InAs/GaInSb SLS photodiodes on cut-off wavelength compared to theoretical and experimental trendlines for comparable HgCdTe photodiodes at 77 K (after Ref. [51]).

Next figure (Fig. 13) compares the calculated detectivity of type-II and p-on-n HgCdTe photodiodes as a function of wavelength and temperature of operation with the experimental data of type-II detectors operating at 78 K. The solid lines are theoretical thermal limited detectivities for HgCdTe photodiodes, calculated using a 1D model that assumes diffusion current from narrower band gap n-side is dominant, and minority carrier recombination via Auger and radiative process. In calculations typical values for the n-side donor concentration ($N_\mathrm{d} = 1 \times 10^{15}$ cm^{-3}), the narrow band gap active layer thickness (10 μm), and quantum efficiency (60%) have been used. The predicted thermally limited detectivities of the type-II SLS are larger than those for HgCdTe [53].

>From Fig. 13 it results that the measured thermally limited detectivities of type-II SLS photodiodes are as yet inferior to current HgCdTe photodiode performance. Their performance has not achieved theoretical values. This limitation appears to be due to main two factors: relatively high background concentrations (about 5×10^{15} cm^{-3}, although values below 10^{15} cm^{-3} have been reported) and a short minority carrier lifetime (typically tens of nanoseconds in lightly doped p-type material). Up till now non-optimized carrier lifetimes have been observed and at desirably low carrier concentrations is limited by the Shockley–Read recombination mechanism. The minority carrier diffusion length is in the range of several micrometers. Improving these fundamental parameters is essential to realize the predicted performance of type-II photodiodes.

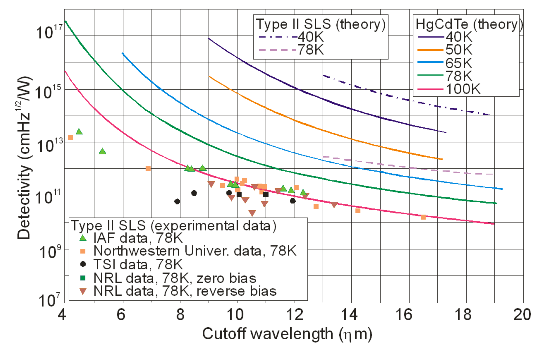

Fig. 13. The predicted detectivity of type-II and p-on-n HgCdTe photodiodes as a function of wavelength and temperature (after Ref. [52]).

Type-II based InAs/GaInSb based detectors have passed rapid progress over the past few years. The presented results indicate that fundamental material issues of InAs/GaInSb SLs fulfil practical realization of high performance FPAs.

First 256×256 SL MWIR [54] and LWIR [55] FPA detectors have been hybridized. The cut-off wavelength of MWIR detector is 5.3 μm. Excellent NEDT value of approximately 10 mK measured with $f/2$ optics and integration time $\tau_\mathrm{int} = 5$ ms has been presented.

Recently, the demonstration of a high performance type-II FPA with cut-off wavelength of 10 μm has been reported [55]. The R_0A product of diodes passivated with SiO_2 was 23 Ω cm^2. Using this photodiode design, a FPA demonstrated NEDT of 33 mK with an integration time of 0.23 ms comparable to HgCdTe.

Also first dual band LWIR/VLWIR photodiodes [56] and dual band MWIR FPAs [57] have been demonstrated. These very promising results confirm that the antimonide SL technology is now competing with MBE HgCdTe dual colour technology.

3.4. Outlook on uncooled detectors

In the beginning of the 1970's in US research programmers started to develop uncooled infrared detectors for practical military applications, mainly to bring thermal imaging to every soldier. In order to do so, thermal imaging cameras needed to become more compact, portable and definitely a lot less expensive than cooled cameras. In 1978, Texas Instruments patented ferroelectric infrared detectors using barium strontium titanate (BST). This technology was demonstrated to the military for the first time in 1979 [58]. At the same time, another technology, micromachining bolometer technology, was developed by Honeywell. The development of both technologies was founding by US military through next fifteen years. However, about 10 years ago, convinced of the advantages VO_x has over BST, the US military decided not to provide any more funding for research into

BST. The loss of government founding for BST meant that research in this technology slowed down drastically. Thus, research efforts supported by DARPA have been moved into development of thin-film ferroelectric (TFFE) detector arrays in Raytheon Commercial Infrared (former Texas Instruments). Further research supported by Missile Defence Agency and L-3 Communication Infrared Products to push technology into marketable products has failed [59]. The research into BST stagnated, while VO_x technology is still continuing to do so with more funding. The size of BST pixels, 50 μm, is still at the same point where it was more than 10 years ago. Moreover, the rotating chopper blocks the detector 50% of the time, which does not benefit the sensitivity of the camera. Mechanical chopper is susceptible to breakdown and sensitive to shock and vibration. As a result, the mean time between failures (MTBF) for BST camera is shorter than for microbolometer camera. Also using of thermoelectric cooling to stabilize electrical polarization is disadvantage of BST detectors in comparison with VO_x and α-Si detectors.

In the mid 1990's a third technology, amorphous silicon, was developed. During this time, the big advantage of using α-Si was their fabrication in a silicon foundry. Further, the VO_x technology was controlled by the US military and export license was required for microbolometer cameras that were sold outside the US. Today, VO_x bolometers can be also produced in a silicon foundry and above both reasons disappeared.

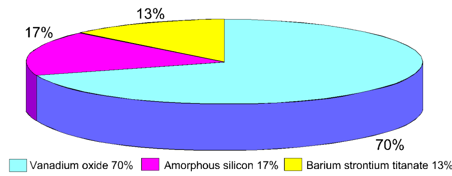

Fig. 14. Estimated market shares for VO_x, α-Si and BST detectors (after Ref. [58]).

At present, VO_x microbolometer arrays are clearly the most used technology for uncooled detectors (see Fig. 14). VO_x is winner of the battle between the technologies and vanadium oxides detectors which are being produced at a much lower cost than either of the two other technologies [58].

The key trade-off with respect to uncooled thermal imaging systems is between sensitivity and response time. The thermal conductance is an extremely important parameter, since NEDT is proportional to $G_{\rm th}^{1/2}$ (where $G_{\rm th}$ is the thermal conductance), but the thermal response time of the detector, $\tau_{\rm th}$, is inversely proportional to $G_{\rm th}$. Therefore, a change in thermal conductance due to improvements in material processing technique improves sensitivity at the expense of time response. Typical calculations of the trade-off between NEDT and time response carried out in Ref. [60] are shown in Fig. 15 [61].

If the NEDT is dominated by a noise source that is proportional to $G_{\rm th}$, which has place when Johnson and $1/f$ noises are dominated, and since $\tau_{\rm th} = C_{\rm th}/G_{\rm th}$ ($C_{\rm th}$ is the thermal capacity of the detector), then the figure of merit (FOM) given by

$$\mathrm{FOM} = \mathrm{NEDT} \times \tau_{\rm th} \tag{3}$$

can be introduced [62]. Users are interested not only in the sensitivity, but also in their thermal time constants and the FOM described by Eq. (3) recognizes the trade-off between thermal time constant and sensitivity. Figure 16 shows the dependence of NEDT on thermal time constant for two NEDT$\times\tau_{\rm th}$ products. Table IV contains an overview of the main suppliers and specifications for existing products and for bolometer arrays that are in the R&D stage whereas Fig. 17 presents development efforts for the VO_x FPAs.

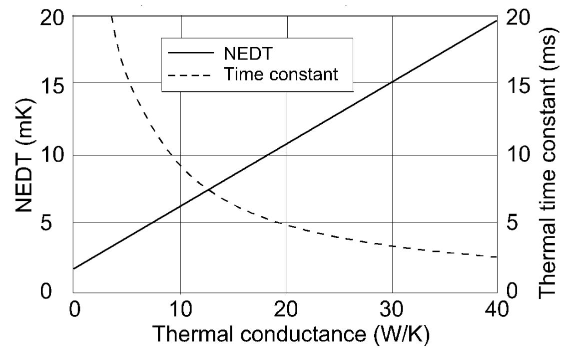

Fig. 15. Trade-off between sensitivity and response time of uncooled thermal imaging systems. 8–14.5 μm, $f/1$, pitch = 25 μm (low mass) (after Ref. [61]).

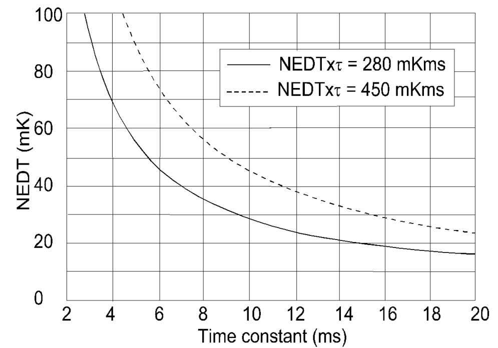

Fig. 16. Calculated microbolometer NEDT and thermal time constant, $\tau_{\rm th}$, for two $NEDT \times \tau_{\rm th}$ products (after Ref. [62]).

TABLE IV

Commercial and state-of-the-art R&D uncooled infrared bolometer array.

Company	Bolometer type	Array format	Pixel pitch [μm]	Detector NEDT [mK] ($f/1$, 20–60 Hz)
FLIR (USA)	VO_x bolometer	160×120–640×480	25	35
L-3 (USA)	VO_x bolometer	320×240	37.5	50
	α-Si bolometer	160×120–640×480	30	50
	α-Si/α-SiGe	320×240–1024×768	R&D:17	30–50
BAE (USA)	VO_x bolometer	320×240–640×480	28	30–50
	VO_x bolometer (standard design)	160×120–640×480	17	50
	VO_x bolometer (standard design)	1024×768	R&D:17	
DRS (USA)	VO_x bolometer (umbrella design)	320×240	25	35
	VO_x bolometer (standard design)	320×240	17	50
	VO_x bolometer (umbrella design)	640×480	R&D:17	
Raytheon (USA)	VO_x bolometer	320×240–640×480	25	30–40
	VO_x bolometer (umbrella design)	320×240–640×480	17	50
	VO_x bolometer (umbrella design)	640×480, 1024×768	R&D:17	
ULIS (France)	α-Si bolometer	160×120, 640×480	25–50	35–80
	α-Si bolometer	1024×768	R&D:17	
Mitsubishi (Japan)	Si diode bolometer	320×240, 640×480	25	50
SCD (Israel)	VO_x bolometer	384×288	25	50
	VO_x bolometer	640×480	25	50
NEC (Japan)	VO_x bolometer	320×240	23.5	75

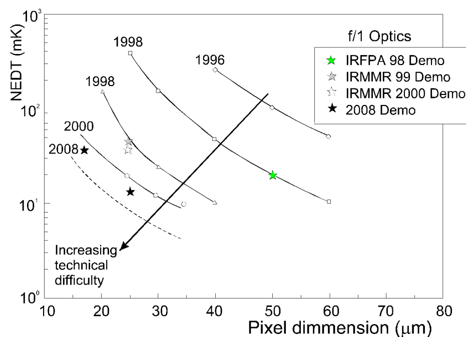

Fig. 17. VO_x focal plane array development (after Refs. [62] and [63]).

The microbolometer detectors are now produced in larger volumes than all other IR array technologies together. The demonstrated performance is getting closer to the theoretical limit with the advantages regarding weight, power consumption and cost. The 240×320 arrays of 50 μm microbolometers are fabricated on industry-standard wafer (4 in diameter) complete with monolithic readout circuits integrated into underlying silicon. However, there is a strong system need to reduce the pixel size to achieve several potential benefits. The detection range of many uncooled IR imaging systems is limited by pixel resolution rather than sensitivity. Because the cost of the optics made of Ge, the standard material, depends approximately upon the square of the diameter, so reducing of the pixel size causes reducing cost of the optics. These reductions in optics size would have a major benefit in reducing the overall size, weight and cost of manportable IR systems. In addition the reduction in pixel size allows a significantly larger number of FPAs to be fabricated on each wafer which allows a significantly larger of FPAs to be fabricated on each wafer. However, the NEDT is inversely proportional to the pixel area, thus, if the pixel size is reducing from 50×50 μm^2 to 17×17 μm^2, and everything else remained the same, the NEDT would increase by the factor of nine. Improvements in the readout electronics are needed to compensate for this. For future arrays, the $f/1$ NEDT performance of 17 μm pitch microbolometer FPAs is projected to be below 20 mK (see Table IV and Fig. 17) [62, 63]. The development of highly sensitive 17 μm microbolometer pixels, however, presents significant challenges in both fabrication process improvements and in pixel design. Microbolometr pixels fabricated with conventional single-level micromachining processes suffer severe performance degradation as the unit cell is reduced below 40 μm. This problem can be mitigated to some degree if the microbolometer process capability (design rules) is improved dramatically.

TABLE V

Performance and cost comparison (after Ref. [65]).

Features	Microbolometers	Pyrometers	Microcantilever
ultimate sensitivity [mK]	20	40	3
response time [ms]	15–20	15–20	5–10
dynamic range	10^4	10^3	$>10^5$
optics	large, expensive	large, expensive	small, cheap
power requirements	low	low	low
ease of fabrication	difficult	difficult	standard IC fabrication
size	moderately small	moderately small	small
cost of camera	\$20–50 k	\$7–25 k	\$5–15 k

Despite successful commercialization of uncooled microbolometers suitable for thermal imaging, the community is still searching for a platform for thermal imagers that combine affordability, convenience of operation, and excellent performance [64]. At present, thermal-imaging modules for less than \$5000 are produced [65]. It means a tenfold reduction in costs, compared with the approximate price for current IR imaging systems (see Table V). Recent advances in microelectromechanical systems (MEMS) have led to the development of uncooled IR detectors operating as micromechanical thermal detectors. Between them the most important are biomaterial microcantilevers that mechanically respond to the absorption of the radiation [66, 67].

In dependence on readout techniques, the novel uncooled detectors can be devoted on: capacitative, optical, piezoresistive, and electron tunnelling. Many research groups are involved in development low-cost optical-readable imaging arrays in which the infrared radiation detection and subsequent reconstruction of an image is based on the deflection of individual microcantilever pixels. This approach was adapted from standard atomic force microscopy (AFM) imaging systems [68]. With this approach the array does not require metallization to individually address each pixel. In comparison electrically-coupled cantilevers, the optical readout has a number of important advantages [69]:

- the array is simpler to fabricate enabling reduced cost,
- the need for an integrated ROIC is eliminated,
- the layout complexity of matrix addressing is not required,
- parasitic heat from ROIC is eliminated, and
- absence of electrical contacts between pixels and substrate eliminates a thermal leakage path.

The most important practical implication of the above approach is, however, related to their straightforward scalability to much larger ($>2000\times2000$) arrays [70].

Figure 18 demonstrates a schematic diagram and components of the optomechanical IR imaging system. It consists of an IR imaging lens, a microcantilever FPA, and an optical readout. Visible light that comes from the LED becomes parallel via collimating lens. Subsequently the parallel light is reflected by the pixels of the FPA and then passes through a transforming lens. The reflected diffracting rays synthesize the spectra of the cantilever array on the rear focal plane of the transforming lens. When the incident IR flux is absorbed by the pixels, their temperature rises, and then causes a small deflection of the cantilevers. Consequently, the changes in the reflected distribution of visible light are collected and analyzed by a conventional CCD or CMOS camera. The small aperture of the lens mounted on camera makes it possible to achieve the required angle-to-intensity conversion. This simple optical readout uses 1 mW power of the light beam while the power per FPA pixel is a few nanowatts. The dynamic range, intrinsic noise, and resolution of the camera largely determine the performance of the system.

Recently, Agiltron Inc. produced a 280×240 photomechanical IR sensors with an optical readout for both MWIR and LWIR imaging at a speed of up to 1000 frames per second [72]. Results on the detection of rapid occurrence events, such as gunfire and rocket travel, were reported. At present stage of development, the imager has a NEDT of approximately 120 mK at $f/1$ optics.

4. Adaptive focal plane arrays

Research beyond third generation detector arrays is focused on adaptive multi/hyperspectral imaging. A number of recent developments in the area of MEMS-based tuneable IR detectors have the potential to deliver voltage-tuneable multi-band infrared FPAs. These technologies have been developed as part of the DARPA-funded adaptive focal plane array (AFPA) program, and have demonstrated multi-spectral tuneable IR detector structures [73–76].

By use of MEMS fabrication techniques arrays of devices, such as etalons, can be fabricated on an IR detector array that permits tuning of the incident radiation

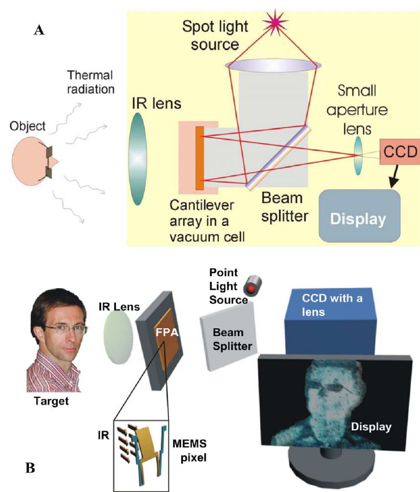

Fig. 18. Uncooled optical-readable IR imaging system: (a) schematic diagram, and (b) components of the thermal imager (after Ref. [71]).

on the detector. If the etalons can be programmed to change distance from the detector surface by the order of IR wavelengths, the detector responds to all wavelengths in a wave band sequentially.

Figure 19 presents general concept of MEMS-based tuneable IR detector. The MEMS filters are individual electrostatically actuated Fabry–Perot tuneable filters. In the actual implementation, the MEMS filter array is mounted so that the filters are facing towards the detector to minimize spectral crosstalk.

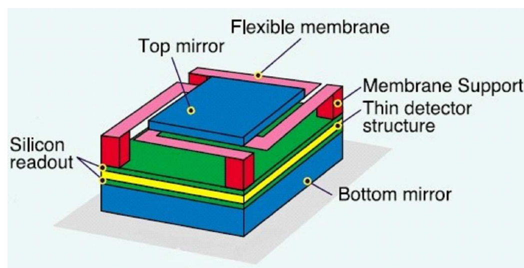

Fig. 19. General concept of MEMS-based tuneable IR detector.

The integration of various component technologies into an AFPA involves a complex interplay across a broad range of disciplines, involving MEMS device processing, optical coating technology, microlenses, optical system modelling, and FPA devices. The goal of this integration is to produce an image-sensor array in which the wavelength sensitivity of each pixel can be independently tuned. In effect, the device would constitute a large-format array of electronically programmable microspectrometers.

Rockwell Scientific Co. has demonstrated simultaneous spectral tuning in the LWIR region while providing broad-band imagery in MWIR band using dual band AFPA (see Fig. 20). The filter characteristics, including LWIR passband bandwidth and tuning range, are determined by the integral thin film reflector and antireflection coatings. The nominal dimension of each MEMS filter is between 100 μm and 200 μm on a side and each filter covers a small subarray of the detector pixels. Employing dual band FPA with 20 μm pixel pitch results in each MEMS filter covering a detector subarray ranging from 5×5 to 10×10 pixels. The MEMS filter array will then evolve to tuneable individual pixels. The device will undoubtedly require a new ROIC to accommodate the additional control functions at each pixel.

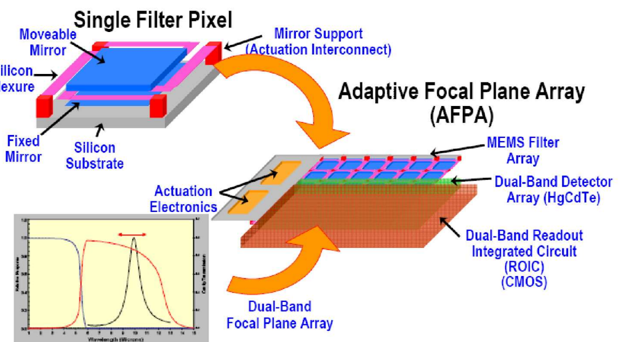

Fig. 20. Dual band adaptive focal plane array (after Ref. [75]).

Figure 21 shows the room temperature spectral transmission of a filter in dual band AFPA illustrating filter tuning for various actuation voltages. The LWIR passbands exhibit low transmission and have measured bandwidths of 200–300 nm.

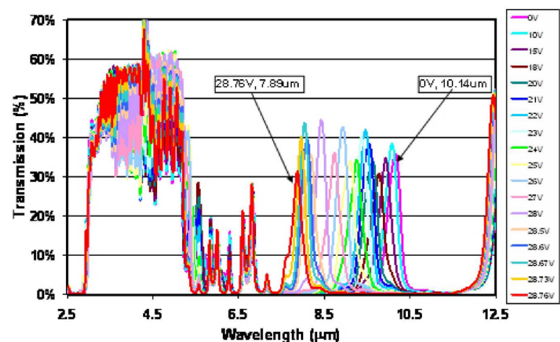

Fig. 21. Measured room temperature spectral transmission of a MEMS tuneable filter over a range of actuation voltages demonstrating tuning in the LWIR with broadband transmission in the MWIR (after Ref. [75]).

The realization of the AFPA concepts offers the potential for dramatic improvements in critical military missions involving reconnaissance, battlefield surveillance, and precision targeting [76].

5. Conclusions

The future applications of IR detector systems require:

- higher pixel sensitivity,
- further increase in pixel density,
- cost reduction in IR imaging array systems due to less cooling sensor technology combined with integration of detectors and signal processing functions (with much more on-chip signal processing),
- improvement in functionality of IR imaging arrays through development of multispectral sensors.

Array sizes will continue to increase but perhaps at a rate that falls below the Moore law curve. An increase in array size is already technically feasible. However, the market forces that have demanded larger arrays are not as strong now that the megapixel barrier has been broken.

There are many critical challenges for future civilian and military infrared detector applications. For many systems, such as night-vision goggles, the IR image is viewed by the human eye, which can discern resolution improvements only up to about one megapixel, roughly the same resolution as high-definition television. Most high-volume applications can be completely satisfied with a format of 1280×1024. Although wide-area surveillance and astronomy applications could make use of larger formats, funding limits may prevent the exponential growth that was seen in past decades.

Third generation infrared imagers are beginning a challenging road to development. For multiband sensors, boosting the sensitivity in order to maximize identification range is the primary objective. The goal for dual-band MW/LW IR FPAs are 1920×1080 pixels, which due to lower cost should be fabricated on silicon wafers. The challenges to attaining those specifications are material uniformity and defects, heterogeneous integration with silicon and ultra well capacity (on the order of a billion in the LWIR).

It is predicted that HgCdTe technology will continue in the future to expand the envelope of its capabilities because of its excellent properties. Despite serious competition from alternative technologies and slower progress than expected, HgCdTe is unlikely to be seriously challenged for high-performance applications, applications requiring multispectral capability and fast response. However, the nonuniformity is a serious problem in the case of LWIR and VLWIR HgCdTe detectors. For applications that require operation in the LWIR band as well as two-colour MWIR/LWIR/VLWIR bands most probably HgCdTe will not be the optimal solution. Type-II InAs/GaInSb superlattice structure is a relatively new alternative IR material system and has great potential for LWIR/VLWIR spectral ranges with performance comparable to HgCdTe with the same cut-off wavelength.

Based on the breakthrough of Sb-based type-II SLS technology it is obvious that this material system is in position to provide high thermal resolution for short integration times comparable to HgCdTe. The fact that Sb-based superlattices are processed close to standard III–V technology raises the potential to be more competitive due to lower costs in series production. The potential low cost compared to HgCdTe is that it can leverage investments in lasers and transistors in the Sb-based industry, and has potential commercial market applications in the future.

Near future high performance uncooled thermal imaging will be dominated by VO_x bolometers. However, their sensitivity limitations and the still significant prices will encourage many research teams to explore other IR sensing techniques with the potential for improved performance with reduced detector costs. Recent advances in MEMS systems have led to the development of uncooled IR detectors operating as micromechanical thermal detectors. One of such attractive approach is optically-coupled cantilevers.

Although in an early stage of development, the potential to deliver FPAs that can adapt their spectral response to match the sensor requirements in real-time, presents a compelling case for future multi-spectral IR imaging systems. Such systems have the potential to deliver much-improved threat and target recognition capabilities for future defence combat systems.

References

[1] A. Rogalski, *Infrared Detectors*, Gordon and Breach, Amsterdam 2000.

[2] A. Hoffman, *Laser Focus World* (February 2006), 81.

[3] A.W. Hoffman, P.L. Love, J.P. Rosbeck, *Proc. SPIE* **5167**, 194 (2004).

[4] J.W. Beletic, R. Blank, D. Gulbransen, D. Lee, M. Loose, E.C. Piquette, T. Sprafke, W.E. Tennant, M. Zandian, J. Zino, *Proc. SPIE* **7021**, 70210H (2008).

[5] P. Norton, J. Campbell, S. Horn, D. Reago, *Proc. SPIE* **4130**, 226 (2000).

[6] M.Z. Tidrow, W.A. Beck, W.W. Clark, H.K. Pollehn, J.W. Little, N.K. Dhar, P.R. Leavitt, S.W. Kennedy, D.W. Beekman, A.C. Goldberg, W.R. Dyer, *Opto-Electron. Rev.* **7**, 283 (1999).

[7] M.N. Abedin, T.F. Refaat, I. Bhat, Y. Xiao, S. Bandara, S.D. Gunapala, *Proc. SPIE* **5543**, 239 (2004).

[8] P. McCarley, *Proc. SPIE* **4288**, 1 (2001).

[9] J.T. Caulfield, in: *Proc. 32nd Applied Imagery Pattern Recognition Workshop*, IEEE 2003, p. 7.

[10] D. Reago, S. Horn, J. Campbell, R. Vollmerhausen, *Proc. SPIE* **3701**, 108 (1999).

[11] P. Norton, J. Campbell, S. Horn, D. Reago, *Proc. SPIE* **4130**, 226 (2000).

[12] P.R. Norton, *Opto-Electron. Rev.* **14**, 283 (2006).

[13] A. Rogalski, J. Antoszewski, L. Faraone, *J. Appl. Phys.* **105**, 091101 (2009).

[14] R.D. Rajavel, D.M. Jamba, J.E. Jensen, O.K. Wu, J.A. Wilson, J.L. Johnson, E.A. Patten, K. Kasai, P.M. Goetz, S.M. Johnson, *J. Electron. Mater.* **27**, 747 (1998).

[15] S. Horn, P. Norton, K. Carson, R. Eden, R. Clement, *Proc. SPIE* **5406**, 332 (2004).

[16] R. Balcerak, S. Horn, *Proc. SPIE* **5783**, 384 (2005).

[17] P. Norton, *Opto-Electron. Rev.* **10**, 159 (2002).

[18] J.A. Wilson, E.A. Patten, G.R. Chapman, K. Kosai, B. Baumgratz, P. Goetz, S. Tighe, R. Risser, R. Herald, W.A. Radford, T. Tung, W.A. Terre, *Proc. SPIE* **2274**, 117 (1994).

[19] M.B. Reine, A. Hairston, P. O'Dette, S.P. Tobin, F.T.J. Smith, B.L. Musicant, P. Mitra, F.C. Case, *Proc. SPIE* **3379**, 200 (1998).

[20] J.P. Zanatta, P. Ferret, R. Loyer, G. Petroz, S. Cremer, J.P. Chamonal, P. Bouchut, A. Million, G. Destefanis, *Proc. SPIE* **4130**, 441 (2000).

[21] W.E. Tennant, M. Thomas, L.J. Kozlowski, W.V. McLevige, D.D. Edwall, M. Zandian, K. Spariosu, G. Hildebrandt, V. Gil, P. Ely, M. Muzilla, A. Stoltz, J.H. Dinan, *J. Electron. Mater.* **30**, 590 (2001).

[22] G. Destefanis, J. Baylet, P. Ballet, P. Castelein, F. Rothan, O. Gravrand, J. Rothman, J.P. Chamonal, A. Million, *J. Electron. Mater.* **36**, 1031 (2007).

[23] D.F. King, W.A. Radford, E.A. Patten, R.W. Graham, T.F. McEwan, J.G. Vodicka, R.F. Bornfreund, P.M. Goetz, G.M. Venzor, S.M. Johnson, *Proc. SPIE* **6206**, 62060W (2006).

[24] J.M. Peterson, J.A. Franklin, M. Readdy, S.M. Johnson, E. Smith, W.A. Radford, I. Kasai, *J. Electron. Mater.* **36**, 1283 (2006).

[25] R. Bornfreund, J.P. Rosbeck, Y.N. Thai, E.P. Smith, D.D. Lofgreen, M.F. Vilela, A.A. Buell, M.D. Newton, K. Kosai, S.M. Johnson, T.J. DeLyon, J.J. Jensen, M.Z. Tidrow, *J. Electron. Mater.* **37**, 1085 (2007).

[26] A.C. Goldberger, S.W. Kennerly, J.W. Little, H.K. Pollehn, T.A. Shafer, C.L. Mears, H.F. Schaake, M. Winn, M. Taylor, P.N. Uppal, *Proc. SPIE* **4369**, 532 (2001).

[27] H. Schneider, M. Walther, J. Fleissner, R. Rehm, E. Diwo, K. Schwarz, P. Koidl, G. Weimann, J. Ziegler, R. Breiter, W. Cabanski, *Proc. SPIE* **4130**, 353 (2000).

[28] H. Schneider, P. Koidl, M. Walther, J. Fleissner, R. Rehm, E. Diwo, K. Schwarz, G. Weimann, *Infrared Phys. Technol.* **42**, 283 (2001).

[29] M. Jhabvala, K. Choi, A. Goldberg, A. La, S. Gunapala, *Proc. SPIE* **5167**, 175 (2004).

[30] S.D. Gunapala, S.V. Bandara, J.K. Liu, C.J. Hill, B. Rafol, J.M. Mumolo, J.T. Trinh, M.Z. Tidrow, P.D. LeVan, *Semicond. Sci. Technol.* **20**, 473 (2005).

[31] S.D. Gunapala, S.V. Bandara, J.K. Liu, J.M. Mumolo, C.J. Hill, D.Z. Ting, E. Kurth, J. Woolaway, P.D. LeVan, M.Z. Tidrow, *Proc. SPIE* **6660**, 66600E (2007).

[32] S.D. Gunapala, S.V. Bandara, J.K. Liu, J.M. Mumolo, C.J. Hill, S.B. Rafol, D. Salazar, J. Woollaway, P.D. LeVan, M.Z. Tidrow, *Infrared Phys. Technol.* **50**, 217 (2007).

[33] S.D. Gunapala, S.V. Bandara, J.K. Liu, C.J. Hill, B. Rafol, J.M. Mumolo, *IEEE Trans. Electron Dev.* **50**, 2353 (2004).

[34] J. Phillips, *J. Appl. Phys.* **91**, 4590 (2002).

[35] P. Martyniuk, A. Rogalski, *Prog. Quantum Electron* **32**, 89 (2008).

[36] G.J. Brown, *Proc. SPIE* **5783**, 65 (2005).

[37] C. Mailhiot, D.L. Smith, *J. Vac. Sci. Technol. A* **7**, 445 (1989).

[38] C.A. Hoffman, J.R. Meyer, E.R. Youngdale, F.J. Bartoli, R.H. Miles, L.R. Ram-Mohan, *Solid State Electron.* **37**, 1203 (1994).

[39] C.H. Grein, P.M. Young, H. Ehrenreich, *Appl. Phys. Lett.* **61**, 2905 (1992).

[40] C.H. Grein, P.M. Young, M.E. Flatté, H. Ehrenreich, *J. Appl. Phys.* **78**, 7143 (1995).

[41] O.K. Yang, C. Pfahler, J. Schmitz, W. Pletschen, F. Fuchs, *Proc. SPIE* **4999**, 448 (2003).

[42] J. Pellegrino, R. DeWames, C. Billman, S. Bandera, "Minority carrier lifetime characteristics in type II InAs/GaSb superlattice $n^+\pi p^+$ photodiodes," *2008 U.S. Workshop on II–VI Materials*, November 11–13, 2008 (presented only).

[43] D.R. Rhiger, A. Gerrish, "Estimation of carrier lifetimes from I–V curve fitting in InAs/GaSb and HgCdTe LWIR diodes," *2008 U.S. Workshop on II–VI Materials*, November 11–13, 2008 (presented only).

[44] L. Bürkle, F. Fuchs, in: *Handbook of Infrared Detection and Technologies*, Eds. M. Henini, M. Razeghi, Elsevier, Oxford 2002, p. 159.

[45] M. Razeghi, Y. Wei, A. Gin, A. Hood, V. Yazdanpanah, M.Z. Tidrow, V. Nathan, *Proc. SPIE* **5783**, 86 (2005).

[46] R. Rehm, M. Walther, J. Schmitz, J. Fleißner, F. Fuchs, W. Cabanski, J. Ziegler, *Proc. SPIE* **5783**, 123 (2005).

[47] P.-Y. Delaunay, A. Hood, B.-M. Nguyen, D. Hoffman, Y. Wei, M. Razeghi, *Appl. Phys. Lett.* **91**, 091112 (2007).

[48] A. Hood, P.-Y. Delaunay, D. Hoffman, B.-M. Nguyen, Y. Wei, M. Razeghi, V. Nathan, *Appl. Phys. Lett.* **90**, 233513 (2007).

[49] E.H. Aifer, J.G. Tischler, J.H. Warner, I. Vurgaftman, W.W. Bewley, J.R. Meyer, C.L. Canedy, E.M. Jackson, *Appl. Phys. Lett.* **89**, 053510 (2006).

[50] B.-M. Nguyen, D. Hoffman, P.-Y. Delaunay, M. Razeghi, *Appl. Phys. Lett.* **91**, 163511 (2007).

[51] C.L. Canedy, H. Aifer, I. Vurgaftman, J.G. Tischler, J.R. Meyer, J.H. Warner, E.M. Jackson, *J. Electron. Mater.* **36**, 852 (2007).

[52] J. Bajaj, G. Sullivan, D. Lee, E. Aifer, M. Razeghi, *Proc. SPIE* **6542**, 65420B (2007).

[53] C.H. Grein, H. Cruz, M.E. Flatte, H. Ehrenreich, *Appl. Phys. Lett.* **65**, 2530 (1994).

[54] W. Cabanski, K. Eberhardt, W. Rode, J. Wendler, J. Ziegler, J. Fleißner, F. Fuchs, R. Rehm, J. Schmitz, H. Schneider, M. Walther, *Proc. SPIE* **5406**, 184 (2005).

[55] P.-Y. Delaunay, B.M. Nguyen, D. Hoffman, M. Razeghi, *IEEE J. Quant. Electron.* **44**, 462 (2008).

[56] E.H. Aifer, J.G. Tischler, J.H. Warner, I. Vurgaftman, J.R. Meyer, *Proc. SPIE* **5783**, 112 (2005).

[57] M. Münzberg, R. Breiter, W. Cabanski, H. Lutz, J. Wendler, J. Ziegler, R. Rehm, M. Walther, *Proc. SPIE* **6206**, 620627 (2006).

[58] http://www.flir.com/uploadedFiles/Eurasia/Cores_and_Components/Technical_Notes/uncooled%20detectors%20BST.pdf .

[59] C.M. Hanson, H.R. Beratan, D.L. Arbuthnot, *Proc. SPIE* **6940**, 694025 (2008).

[60] S. Horn, D. Lohrmann, P. Norton, K. McCormack, A. Hutchinson, *Proc. SPIE* **5783**, 401 (2005).

[61] J.A. Ratches, *Ferroelectrics* **342**, 183 (2006).

[62] M. Kohin, N. Butler, *Proc. SPIE* **5406**, 447 (2004).

[63] J. Anderson, D. Bradley, D.C. Chen, R. Chin, K. Jurgelewicz, W. Radford, A. Kennedy, D. Murphy, M. Ray, R. Wyles, J. Brown, G. Newsome, *Proc. SPIE* **4369**, 559 (2001).

[64] *UKTA News*, Issue 27, December 2006.

[65] "MEMS transform infrared imaging, " *Opto&Laser Europe*, June 2003.

[66] R. Amantea, C.M. Knoedler, F.P. Pantuso, V.K. Patel, D.J. Sauer, J.R. Tower, *Proc. SPIE* **3061**, 210 (1997).

[67] S.R. Hunter, G. Maurer, G. Simelgor, S. Radhakrishnan, J. Gray, K. Bachir, T. Pennell, M. Bauer, U. Jagadish, *Proc. SPIE* **6940**, 13 (2008).

[68] E.A. Wachter, T. Thundat, P.I. Oden, R.J. Warmack, P.D. Datskos, S.L. Sharp, *Rev. Sci. Instrum.* **67**, 3434 (1996).

[69] J. Zhao, *Proc. SPIE* **5783**, 506 (2005).

[70] N. Lavrik, R. Archibald, D. Grbovic, S. Rajic, P. Datskos, *Proc. SPIE* **6542**, 1E-1 (2007).

[71] P. Datskos, N. Lavrik, "Simple thermal imagers use scalable micromechanical arrays," *SPIE Newsroom* 10.1117/2.1200608.036 (2006).

[72] J.P. Salerno, *Proc. SPIE* **6542**, 65421D (2007).

[73] W.J. Gunning, J. DeNatale, P. Stupar, R. Borwick, R. Dannenberg, R. Sczupak, P.O. Pettersson, *Proc. SPIE* **5783**, 336 (2005).

[74] C.A. Musca, J. Antoszewski, K.J. Winchester, A.J. Keating, T. Nguyen, K.K.M.B.D. Silva, J.M. Dell, L.F. Faraone, P. Mitra, J.D. Beck, M.R. Skokan, J.E. Robinson, *IEEE Electron. Device Lett.* **26**, 888 (2005).

[75] W.I. Gunning, J. DeNatale, P. Stupar, R. Borwick, S. Lauxterman, P. Kobrin, J. Auyeung, *Proc. SPIE* **6232**, 62320F (2006).

[76] J. Carrano, J. Brown, P. Perconti, K. Barnard, *SPIE's OEmagazine*, 20, April 2004.