Infrared (IR) technologies (materials, devices and systems) represent an area of excellence in science and technology and, even if they have been generally confined to a selected scientific community, they have achieved technological and scientific highlights constituting 'innovation drivers' for neighbouring disciplines, especially in the sensors field. The development of IR sensors, initially linked to astronomical observations, since World War II and for many years has been fostered essentially by defence applications, particularly thermo-vision and, later on, smart vision and detection, for surveillance and warning. Only in the last few decades, the impact of silicon technology has changed the development of IR detectors dramatically, with the advent of integrated signal read-outs and the opening of civilian markets (EO communications, biomedical, environmental, transport and energy applications). The history of infrared sensors contains examples of real breakthroughs, particularly true in the case of focal plane arrays that first appeared in the late 1970s, when the superiority of bi-dimensional arrays for most applications pushed the development of technologies providing the highest number of pixels. An impressive impulse was given to the development of FPA arrays by integration with charge coupled devices (CCD), with strong competition from different technologies (high-efficiency photon sensors, Schottky diodes, multi-quantum wells and, later on, room temperature microbolometers/cantilevers). This breakthrough allowed the development of high performance IR systems of small size, light weight and low cost – and therefore suitable for civil applications – thanks to the elimination of the mechanical scanning system and the progressive reduction of cooling requirements (up to the advent of microbolometers, capable of working at room temperature). In particular, the elimination of cryogenic cooling allowed the development and commercialisation of IR Smart Sensors; strategic components for important areas like transport, environment, territory control and security. Infrared history is showing oscillations and variations in raw materials, technology processes and in device design and characteristics. Various technologies oscillating between the two main detection techniques (photon and bolometer effects) have been developed and evaluated as the best ones, depending on the system use as well as expectable performances. Analysis of the ‘waving change’ in the history of IR sensor technologies is given with the fundamental theory of the various approaches. Highlights of the main historical IR developments and their impact and use in civil and military applications is shown and correlated with the leading technology of silicon microelectronics: scientific and economic comparisons are given and emerging technologies and forecasting of future developments are outlined.

Keywords: infrared; smart sensors; thermovision; photon sensors; microbolometers

1. Introduction
Infrared (IR) detectors are devices transforming the radiant energy of the IR region (from 0.7 to 300 μm) of electro-magnetic (e.m.) spectrum incident on the sensor in another form of energy which can be easily measurable, generally in the form of an electric signal [1–7].

This conversion of energy is normally done by:

- Photoelectric effects:
  
  (a) Photoelectric external effects by generating free electrons from the surface of the sensor that have been hit by a photon with sufficient energy.
  
  (b) Photoelectric internal effects, the most relevant phenomena in sensor development, by generating a couple of hole–electron pairs inside a photoelectric material; generally a photoconductor or a photovoltaic structure.

- Bolometric effects: effects based on variations of vibration energy of a crystalline reticle, which are generally measured by the variation of electrical resistance. Recently advanced developments of IR detectors have been achieved
S

S

1.1. Parameters characterising the IR detectors

The main parameters of IR detectors are: spectral response, signal-to-noise ratio per incident unit power, response time and working temperature.

- The spectral response \( R_s (\lambda) \) in \( \text{V W}^{-1} \) is generally represented by a curve showing the responsivity versus wavelength, where \( R_s \) is the rms (root mean square) of the electrical output voltage of the sensor per unit of radiating rms power at the wavelength \( \lambda \).
- The response time, which is the time needed for the signal to achieve 70.7% of the equilibrium value.
- The signal-to-noise per incident unit power in watts, the so-called **detectivity**, is given by:

\[
D = \frac{S}{N(EA)},
\]

where \( S, N \) are, respectively, the rms of the signal voltage and of the noise voltage, \( E \) (irradiance) is the rms of the incident radiation, \( A \) is the sensitive area of the IR sensor. The detectivity is a most important parameter for characterising IR detectors and generally is indicated as \( D(T, f, \Delta f) \), where \( T \) is the black body temperature of the radiating source, \( f \) is the frequency of modulation of the signal and \( \Delta f \) is the normalised bandwidth of the signal output amplifier.

A normalised detectivity is given by:

\[
D' = \frac{S(\Delta f)^{1/2}}{N(EA^{1/2})}.
\]

The assumptions made in the above formula are valid only if noise is proportional to \( A^{1/2} \) and to \( \Delta f \). The first hypothesis is valid only for the photon detectors and for variation of the sensor area only of one to two orders of magnitude, and is not valid for some bolometric sensors; the proportionality to \( \Delta f \) is also valid in a limited frequency variation.

Another parameter often used in thermal analysis is the noise equivalent temperature difference (NETD), that is, the temperature difference generating an rms signal equal to the detector noise.

1.2. Theory of intrinsic photoconductivity

The intrinsic photoconductive sensor is essentially a photoresistance, that is a resistance changing its value when it is radiated by some form of e.m. energy – normally light – as is shown in Figure 1. When the energy of a photon of radiation light is higher than the forbidden energy gap, \( E_g \), in the photoconductive semiconductor, a couple of electron–hole pairs are generated, generating therefore a voltage variation that is measured as a signal output at the load resistance \( R_L \), as shown in Figure 1.

The voltage responsivity is expressed by

\[
R_V = \frac{\eta \lambda \tau V_b (b + 1)}{Lw t \hbar c b n_0 + p_0},
\]

where \( w, t, L \) are the dimensions of the photoconductor (Figure 1) and \( b = \mu_e |\mu_h| \mu_k \) is the electronic mobility, \( \mu_h \) is the hole mobility and \( V_b \) is the polarisation voltage and the lifetime of carriers is given by \( \tau = \Delta t / \eta \Phi_c \).

This means that the basic requirement for achieving a high responsivity in a photoconductor at a certain wavelength \( \lambda \) requires a high quantum efficiency \( (\eta) \) and a long lifetime \( (\tau) \) of the excess carriers, and also a small distance between the electrodes, a low concentration of thermally generated carriers \( n_0, p_0 \) at equilibrium, with the highest applicable polarisation voltage \( V_{Pr} \).

In reality this model ignores important effects deriving from a too short interelectrode distance or by other physical–structural characteristics, which can originate ‘draining’ effects or surface recombination.

1.3. Photovoltaic detectors

The photovoltaic effect is present in sensor structures with an internal voltage difference which is moving the photo-generated carriers (electrons and holes) in opposite directions. The most common example of the photovoltaic effect is based on an abrupt p–n junction in a semiconductor and is known as a ‘photodiode’.

The photons with energy higher than the energy-gap which are incident over the surface of the sensor are creating a ‘electron–hole’ couple on both sides of the p–n junction that, thanks to the induced
field, diffuse within the diffusion path of the junction reaching the region of spatial carrier where the high electric field separates the electron–hole couples in such a way that the minority carriers are accelerated, becoming majority carriers on the other side of the electrical junction.

In this way the generated photocurrent, modifying the current–voltage characteristics, is given by the inverse negative current $I_{ph}$, as is shown in Figure 2.

$$I_{ph} = \frac{\eta q A \Phi}{R}$$

where $A$ is the photodiode area, $\Phi$ is the flux of incident photons and $\eta$ is the quantum efficiency.

Normally the gain in current in a photovoltaic detector with a simple structure (e.g. not of avalanche or tunnel type) is equal to 1.

When the p–n junction is working with an open circuit, the accumulation of electron–hole couples in the junction carriers produces an open circuit voltage which, in the case of a load resistance $R_l$ connected to the diode, generates a current which achieves its highest value when the diode is short-circuited (short-circuit current $I_{sh}$).

The open circuit voltage can be obtained by just multiplying the $I_{sh}$ per the incremental resistance $R = (\delta V / \delta I)$ when $V = V_b$. That is

$$V_{ph} = \eta q A \Phi R,$$

where $V_b$ is the voltage polarisation and $I = f(V)$ is the $I$–$V$ curve. In many cases the photodiode is operating at polarisation voltage equal to zero

$$R_0 = \left( \frac{\partial I}{\partial V} \right)_{V_b=0}^{-1}. \quad (6)$$

A figure of merit normally used is the value of the $R_0 A$ product

$$R_0 A = \left( \frac{\partial I}{\partial V} \right)_{V_b=0}^{-1}, \quad (7)$$

where $I = I/A$ is the current density.

In normal detection the photodiodes are working at zero bias, while the inverse polarisation is used for high frequency applications for reducing the RC constant of the device.

1.3.1. Photodiode currents

Various mechanisms are involved in the current phenomena in photodiodes. The most important are:

(a) Dark current mainly due to thermally generated carriers in the crystal and in the depletion layer of the p–n junction, and surface currents due to surface states and surface leakages.

(b) Diffusion current given by:

$$J_D = J_s [\exp(qV/kT) - 1] \quad (8)$$

where $J_s$ is given by

$$J_s = (kT)^{1/2} \frac{n_i^2 q^{1/2}}{\rho_{po}^{1/2}} \left[ \frac{1}{n_{po}^{1/2}} + \frac{1}{n_{no}^{1/2}} \left( \frac{\mu_n}{\tau_n} \right)^{1/2} \right]^{-1}. \quad (9)$$
where \( n_i \) is the concentration of intrinsic carriers, \( p_o \) and \( n_o \) are the concentrations of the majority carriers and \( \tau_e \) and \( \tau_h \) are the lifetime of electrons and holes respectively in the p and n regions [8–10].

The diffusion current is changing versus temperature as the square of intrinsic density of electrons (\( n_i^2 \)) and is generally the current dominating at high temperatures.

(c) Generation-recombination current:

Such a mechanism can be the dominant one at low temperatures and is given by [8]:

\[
J_{GR} = \frac{q n_i w}{(\tau_e \tau_h)} \frac{2 \sinh(q V/2kT)}{q(V - V)/kT} f(b),
\]

(10)

where \( V_{bi} \) is the induced voltage and \( \tau_e \) and \( \tau_h \) are the lifetimes of the electrical carriers in the depletion layer and \( f(b) \) is a complex function which is normally close to \( \approx 1 \). The generation-recombination (g-r) current can be simplified by:

\[
J_{GR} = \frac{q n_i w}{2\tau_e},
\]

(11)

which, taking into account that the width of the barrier is varying as the square root of the applied voltage (\( w \approx V^{1/2} \)) in the presence of abrupt junctions or as the cubic root (\( w \approx V^{1/3} \)) for linearly graded junctions, is showing the proportional dependence for the g-r current. Moreover, the g-r current is proportional to \( n_i \), while the diffusion current is proportional to \( n_i^2 \). Therefore, there is a temperature \( T_e \) where the two currents are comparable, while below \( T_e \) the g-r current is dominant.

Other current phenomena are related to tunnelling effects and more importantly, to surface leakages. The dark current taking care of all these phenomena is expressed by:

\[
I = I_s \left[ \exp \left( \frac{q(V - IR_s)}{\beta kT} - 1 \right) + \frac{V - IR_s}{R_{sh}} + I_T \right],
\]

(12)

where \( R_s \) is the series resistance and \( R_{sh} \) is the shunt resistance of the photodiode.

If the diffusion current is dominating, the \( b \) coefficient is close to 1, but if the main carrier transport is due to g-r current, \( b \) is equal to \( a \approx 2 \).

1.3.2. \( R_0A \) product

In the case of a classic diode, where \( d \gg L_e \)

\[
(R_0A)_D = \frac{(KT)^{1/2}}{q^{3/2}n_i^2} N_a \left( \frac{\mu_e}{\tau_e} \right)^{1/2}.
\]

(13)
If the thickness of the crystal layer is smaller than the diffusion length of the minority carriers, the \( R_0A \) product is increased and in such a case, we obtain
\[
(R_0A)_{GR} = \frac{V_{b}T_b}{q\mu n L_{ct}}
\]
with an increase in the product \( R_0A \) equal to \( L_c/d \).

1.4. Noise mechanisms

All the IR sensors are limited in their detection performance capabilities by the various forms of noise generated either from the sensor itself or from fluctuations in the radiation environment, either from the electronic amplifier for the signal read-out (with the most recent very low-noise amplifiers this type of noise can be ignored – in comparison to the other two) [8,9].

The noise due to the fluctuation of background radiation is given by
\[
V_{ph} = \frac{2\pi^{1/2}V_b}{(h\nu)^{1/2}bn_0 + p} \int_0^\infty \frac{\eta(\nu)^2 \exp(h\nu/kT_b)}{c^2[\exp(h\nu/kT_b) - 1]^2} \times \frac{n\Delta f}{(1 + \omega^2\tau^2)^{1/2}},
\]
where \( T_b \) is the background temperature, and \( n_0 \) is the frequency corresponding to the cut-off wavelength \( 1/\lambda \).

The most important internal noise types are:

- The thermal noise (Johnson–Nyquist noise) associated with any device with a resistance ‘\( R \)’ (pure capacitors and inductors don’t have this type of noise, although they can have other forms of noise, e.g. capacitive noise due to electronic switching).

This type of noise is related with the random thermal fluctuations of the electrical carriers which are moving within the semiconductor (the total fluctuation of the carriers is generating the other type of noise, the so-called g-r noise described later on).

The thermal noise is present in the absence of external biasing and generates a current fluctuation independently from the method of measurement.

The root-mean square (rms) value is given by
\[
V_f = (4kTR\Delta f)^{1/2},
\]
where \( k \) is the Boltzmann constant, \( T \) is the temperature and \( \Delta f \) is the frequency bandwidth.

This type of noise has a flat distribution and is therefore called ‘white noise’.

- The generation recombination noise (g-r or flicker noise) is due to the random fluctuations of the free electronic carriers due to the lattice vibrations of the semiconductor crystal causing a current fluctuation at the microscopic level.

The g-r noise, for an intrinsic photoconductor, is given by
\[
V_{gr} = \frac{2V_b}{(Lw)^{1/2}bn + p} \left( \frac{np}{n + p} \right)^{1/2} \left( \frac{\tau\Delta f}{1 + \omega^2\tau^2} \right)^{1/2}. \tag{17}
\]

It is interesting to underline that the measurements of the g-r noise is allowing one to obtain the value of the lifetime \( \tau \) just by measuring with a spectrum analyser the knee of the curve at \( \nu = 1/\tau \).

- The so-called 1/f noise characterised by a spectrum where the noise power is inversely proportional to the frequency \( f \), according to
\[
I_{1/f} = \left( \frac{K\mu^2AF}{f^\beta} \right)^{1/2}, \tag{18}
\]
where \( K \) is a proportionality factor, \( I_b \) is the current bias, \( \alpha \) is a constant almost equal to 2 and \( \beta \) is a constant almost equal to 1.

The 1/f noise is normally associated with the presence of potential microbarriers at the boundaries of polycrystalline grains in the semiconductor and the reduction of 1/f noise is almost an art in which one has to take care with the realisation of the electrical contacts and the preparation of photosensitive surfaces.

Normally the IR photodetectors show a 1/f noise at low frequencies, while at higher frequencies the predominant noise is the g-r noise until the 1/\( \tau \) frequency where the Johnson noise starts to prevail (Figure 3).

2. Complex devices: IR focal plane arrays (FPA)

The most important application of IR detectors is in thermovision which is the ability to see the thermal emission of the scene. Thermovision systems have strongly evolved over time since the first thermovision systems developed during World War II, based on a simple opto-mechanical scanning focusing scene onto a single detector [10].

Such type of image reconstruction, based on a serial scanning of the image points (pixels), has been improved, mainly by enhancing the signal-to-noise ratio thanks to the increase in sensor numbers. This was first achieved by using linear detector arrays, with a serial read-out integrated in time (time delay integration [TDI], Figure 4(a)) or by using a parallel structure to read simultaneously more rows of the scanned image and then by integrating them in the
The importance of the development of FPA detectors, working in the so-called ‘staring’ mode, i.e. capable of seeing the image by a simultaneous vision of the scene thanks to a mosaic sensor structure positioned in the focal plane of the image avoiding the use of any opto-mechanical scanning, in a way similar to human vision, is self-evident.

Various types of electronic read-out have been developed since the 1970s, starting with a pseudo-bidimensional read-out based on the sequential

image reconstruction (Figure 4(b)) (both applied especially to satellite remote sensing).
read-out of rows and columns by using multiplexers and a shift-register (Figure 5(a)), passing then to the so-called X–Y addressing by using integrated read-out devices (Figure 5(b)) or external addressing capable of selecting specific areas of the mosaic sensors (Figure 5(c)).

The read-out was continuously evolving until 1970, when the new charge coupled device (CCD) read-out, based on electronic charge transfer, allowed a completely integrated read-out (Figure 5(d)), with the development of FPA IR sensors with more than one million pixels of detectors.

3. Historical scenario

Infrared as radiation theory and advanced technology is new (only two centuries old), but the concept of heat and temperature measurements is quite old. In fact the ancient Egyptians moved their hands across the surface of the human body to scan and monitor changes in temperature. The concept of temperature was originated by the need to define the differences more precisely than just the adjective hot/hotter and cold/colder. Based on the ideas of Aristotle who had defined four qualities: hot, cold, moist and dry. Galen (130–200 AD) was the first man to describe heat and cold by a number and the word ‘temperature’ was originated from ‘temper’, determining the ‘complexion’ of a person by the proportion in which the above four qualities were ‘tempered’ [11,12].

The instrument by which the temperature of bodies is measured is called ‘thermometer’ (warmth measurer) and the history of thermometry is a recent part of the history of science. In fact, although the ancient Greeks (Philo of Byzantium, second century BC, and Heron of Alexandria, first century BC), knew of the principle of air thermal expansion and Abū Alī ibn Sinā (known as Avicenna) in the early eleventh century already demonstrated that the expansion and contraction of air moved the position of the water-air interface inside a tube with air inside and partially filled with water, it wasn’t until the sixteen/seventeenth centuries that the first air thermometers were developed.
In 1603 Galileo Galilei built a thermoscope made of a glass cylinder filled with a liquid whose density sensibly expands with temperature (Figure 6). Inside the glass cylinder there are small glass ampoules filled with coloured liquids: such ampoules have different densities and they are labelled with a certain temperature value. When the device is in thermal equilibrium with the surrounding ambient, one can read the temperature on the lowest floating ampoule [11–13].

Robert Fludd (1574–1651) was also regarded as an independent inventor of the thermometer, modifying Philo’s apparatus that demonstrated the expansion of air by heat. One limit of the air thermometer was that it also responded to the change in atmospheric pressure, even if originally there was no concept of atmospheric pressure. The sealed liquid-in-glass thermometer invented by Ferdinand II, Grand Duke of Tuscany, in 1654 was to overcome two other problems, the lack of portability and the evaporation of water, just by using the thermal dilation of a liquid instead of air to sense temperature changes (the first liquid in glass was spirits of Chianti wine). In 1663 the Members of the Royal Society of London agreed to use one of several thermometers made by Robert Hooke as the standard for the calibration of other thermometers.

In 1714 Fahrenheit improved the precision of the thermometer using a cylinder rather than a globe as the bulb of the thermometer and substituting alcohol spirits with mercury, with a more linear thermal expansion, and fixing the lower temperature point by using salt with ice water and the higher point to boiling water at 212 degrees. Celsius, in 1742, created a decimal scale, using ‘zero’ as the boiling point of water and 100°C as the freezing point. His scale was reversed by the botanist Linnaeus: increasing heat was indicated by higher temperatures. In 1948 the name of the ‘Celsius’ scale was given to the centigrade scale.

Another branch of thermometry is gas thermometry and thermodynamic thermometry. Before the Fahrenheit thermometer appeared, in 1660 Robert Boyle reported on his study of air trapped in a U tube that he found that volume at constant pressure was a function of temperature. While Fahrenheit studied his liquid-in-glass thermometer, the French scientist Amontons developed the constant volume gas thermometer. He used air as the thermometric medium, and concluded that the lowest temperature which could exist would correspond to a zero gas pressure. This must have been the first step on the way to understanding the concept of temperature. According to Amontons we could define the temperature as being simply proportional to the pressure of a gas, and thus we would need only one fixed point to define a scale [14]. But the new temperature scale didn’t appear at that time, maybe because of the cumbersome operation of the gas thermometer. Jacques-Alexandre Charles studied the phenomena again in 1787. Jaspeh L. Gay-Lussac extended Charles’ work, all his gases – air, oxygen, nitrogen, hydrogen, and carbon dioxide – expanded the same amount when heated from the ice point to the boiling point. The result showed that the mean volumetric coefficient of thermal expansion, at constant pressure, is equal to \( 1/267 \) (degree)–1. In 1847, Victor Regnault obtained a better value of \( 1/273 \). Later experiments revealed that all gases had very slightly different thermal coefficients, however, all were found to approach a common value \( 1/273.15°C \) as the pressure approached zero.
The gases that obey the temperature–volume relationship exactly at constant pressure (and the pressure–volume relationship exactly at constant volume) were defined the perfect (‘ideal’ or ‘simple’) gas. In fact, all gases at extremely low pressure approach perfect gas behaviour: the volume of the perfect gas in the gas thermometer will approach zero as \( t = -273.15 \). We can establish a new zero of temperature at the ice point \(-273.15\), namely the absolute zero degree, and the perfect gas temperature scale \( T = t + 273.16 \) was defined in 1954 [15].

The precise meaning of temperature came after Kelvin’s thermodynamic temperature appeared. In 1824, Carnot had laid down his principle of reversibility with regard to ‘quantity of heat’ in an ideal ‘body’ that ‘undergoes changes’ and in 1845 (after 30 years of experiments) Joule demonstrated mechanical equivalence. Heat now became an engineering measurable ‘quantity’ and the higher the temperature the greater the quantity of heat in Carnot’s ‘body’. A real relation of temperature to heat was achieved by the first and second laws of thermodynamics: it is a standing proof that for a reversible heat engine operating over a Carnot cycle between two temperatures and the ratio of the heat taken in at the higher temperature to that given out at the lower temperature is proportional simply to the ratio of the same function of each of the two temperatures. William Thomson (later Lord Kelvin) realised, in 1848, that this relation could be used to define the ratio of any two temperatures. The values of the temperatures would depend upon the functional form, but by taking the simplest possible form of the function he defined a temperature, which he called thermodynamic temperature, \( T \). Thermodynamic temperature thus has the property that ratios of \( T \) are defined in terms of the properties of reversible heat engines and are independent of the working substance. The definition of the quantity thermodynamic temperature then has to be completed by assigning a particular numerical value to an arbitrary fixed point of temperature, triple point of water. So the real meaning of the temperature is a physical quantity which is fundamental for the field of thermodynamics and is directly related to the basic laws of thermodynamics [14,15]. In principle, any suitable thermodynamic interaction may be used as the basis for a thermometer, such as the acoustic thermometer, the thermal noise thermometer, the gas thermometer, the radiation thermometer, etc. However, thermodynamic thermometers cannot achieve the desired precision, and are complex and time consuming to use. To overcome these difficulties, the above empirical temperature scale was defined as the International Temperature Scale (ITS) by the Comité des Poids et Mesures (CIPM) under the Convention du Metre, the founding treaty for the SI, and was revised with the current version agreed to in 1990, and known as ITS90. The ITS are empirical temperature scales giving a close approximation to the known thermodynamic scale, but are more precise and easier to use [65].

The history of IR radiation sensors starts in 1800 when the astronomer, William Herschel, discovered the existence of infrared radiation, a form of energy in the light beyond the ‘red’ (from the Latin ‘infra’ – below), by trying to measure the heat of the separate colours of the rainbow spectrum cast on a table in a darkened room. After noticing that the temperature of the colours increased from the violet to the red part of the spectrum, Herschel decided to measure the temperature just beyond the red portion of the spectrum in a region apparently devoid of sunlight. To his surprise, he found that this region had the highest temperature of all. In April 1800 he reported it to the Royal Society as Dark Heat and, making further experiments on what he called the ‘calorific rays’ that existed beyond the red part of the spectrum, he found that they were reflected, refracted, absorbed and transmitted just like visible light [16,17].

The basic laws of IR radiation (Kirchhoff’s law, Stefan–Boltzmann’s law, Planck’s law and Wien’s displacement law) were developed many years after the discovery of IR radiation. In 1859 Gustave Kirchhoff found that a material that is a good absorber of radiation is also a good radiator. Kirchhoff’s law states that the ratio of radiated power and the absorption coefficient: (1) is the same for all radiators at that temperature, (2) is dependent on wavelength and temperature, and (3) is independent of the shape or material of the radiator. If a body absorbs all radiation falling upon it, it is said to be ‘black’. For a black body the radiated power is equal to the absorbed power and the emissivity (ratio of emitted power to absorbed power) equals one. In 1884, L.E. Boltzmann, starting from the physical principles of thermodynamics, derived the theoretical formula for the T4 black body radiation law, stated empirically in 1879 by J. Stefan, by developing the Stefan–Boltzmann’s law (19):

\[
W = \sigma T^4. \tag{19}
\]

where \( W \) = radiation power, \( T \) = absolute temperature, and \( \sigma \) = Stefan–Boltzmann’s constant.

In 1901, Nobel Prize laureate Max Karl Ernst Ludwig Planck developed Planck’s law which stated that the radiation from a black body at a specific wavelength can be calculated from

\[
I(\nu) \delta \nu = \frac{2h\nu}{c^2} \frac{1}{\exp(h\nu/kT) - 1} \delta \nu \tag{20}
\]
IR detector developments, even after the discovery of infrared radiation by Sir H. Herschel in 1798, were mainly based on the use of thermometers which dominated IR applications until World War I, although in 1821 J.T. Seebeck had discovered the thermoelectric effect and in 1829 L. Nobili had fabricated the first thermocouple, allowing in 1833 the multi-element thermopile development by Macedonio Melloni with the first detection of a human being at 10 m. Early thermal detectors, mainly thermocouples and bolometers, were sensitive to all infrared wavelengths and operating at room temperature normally, and until a few years ago, they were of relatively low sensitivity and slow response time.

The first photon detectors (based on the photovoltaic effect discovered by Smith in 1873 in selenium and, later on, by Bose in photovoltaic lead sulphide, but not applied for many years) were developed by Case in 1917. In 1933 Kutzscher developed IR PbS detectors (using natural galena found in Sardinia); these sensors were widely used during World War II. These detectors have been extensively developed since the 1940s. Lead sulfide (PbS) was the first practical IR detector, sensitive to infrared wavelengths up to \( \sim 3 \mu \text{m} \). In the meantime Cashman developed TaS, PbSe and PbTe IR detectors with high performances supporting developments in the UK and US.

High level results were achieved in the 1940s, especially in lead salts (good PbSe and PbTe stable cells were developed by the Office of Scientific Research and Development [OSRD] in co-operation with MIT, Harvard and the British Telecommunication Research Establishment [TRE], Great Malvern). The history of IR detector developments therefore has been almost coincident with optoelectronics for military applications for many decades, strongly conditioning the cultural behaviour of the IR industry and in some way of R&D labs.

Thanks to the discovery of transistors in 1948, infrared technology during the 1950s was enjoying a great growth, especially in the development of solid state IR sensors. Almost contemporarily, lead selenide (PbSe), lead telluride (PbTe), and indium antimonide (InSb) cooled detectors extended the spectral range beyond that of PbS, providing sensitivity in the 3–5\( \mu \text{m} \) region, needing very low temperature with the use of liquid helium. But we had to wait until the 1960s to see the first advanced developments coming out thanks to direct gap photon materials based on ternary semiconductor compounds (HgCdTe and PbSnTe) [18]. This was a real breakthrough, because in the meantime microelectronics was offering new advanced manufacturing technologies like photomasking and integrated microsoldering and assembly. So, thanks to these advanced technologies, the first linear arrays of tens of elements were developed at the end of the 1960s with strong competition between HgCdTe and PbSnTe (this compound could offer more stability and reliability in performance, moreover at the beginning of the 1970s Lincoln Labs, MIT researchers were developing the first solid state IR 10\( \mu \text{m} \) lasers applied to environmental control) [19]. On the other hand, PbSnTe was showing higher dielectric constant, limiting high frequency performances, and a high thermal expansion coefficient, with a strong limitation for integration into silicon microelectronics. (The speed limit could be overcome if forecasting the FPA array developments which could allow the use of slower detectors.)

The choice, supported by US industries, of concentrating efforts and resources in a specific technology produced the ‘first generation of linear detector arrays’, which allowed one to obtain BLIP detectors at liquid nitrogen temperature (this first generation of CMT linear arrays was the basis for the ‘common modules’ LWIR FLIR systems with a number of pixels from 60 up to 180, each detector connected with feedthroughs to the room temperature read-out electronics).

The invention of charge coupled devices (CCDs) in 1969 [20] made it possible to start developing the ‘second generation’ FPA detector arrays coupled with on-focal-plane electronic analogue signal readouts which could multiplex the signal from a very large array of detectors. In the mid-1970s, while the first common module IR arrays were produced, the first CCD IR bidimensional arrays [21,22] were appearing in the USA and, the first Smart Sensors based on LTT RF sputtered thin films, using X-addressing read-out, were developed in Italy [23].

In 1975 the first CCD TV camera was realised and this allowed the forecast of the ‘second generation FPAs’ capable of staring vision, although the necessity for very high spatial resolution and high reliability even in complex structures, with an extremely high
number of pixels (up to one million pixels), were pushing towards alternative solutions, with materials less difficult than CMT, in the manufacturing process (e.g. extrinsic silicon detectors). The high quantum yield of CMT and the top performances required by the military allowed one to improve the performances of sensor linear arrays by integrating time delay and integration inside the detector structure itself (SPRITE detector [24]). From the late 1970s to the 1980s, CMT technology efforts focused almost exclusively on PV device development because of the need for low power and high impedance for interfacing to readout input circuits in large arrays (photoconductive CMT was not suitable due to its low impedance). This effort was concretised in the 1990s with the birth of 'second generation IR detectors' which provided large 2D arrays with the number of pixels up to many hundreds of thousands thanks to hybrid integration (indium bumps or loophole soldering) of CMT bidimensional arrays in silicon substrate with CCD and more recently CMOS read-out. At the same time, other significant detector technology developments were taking place. Silicon technology generated novel platinum silicide (PtSi) detector devices which have become standard commercial products for a variety of MWIR high resolution applications. Monolithic extrinsic silicon detectors were demonstrated first in the mid-1970s [25,26]. Thanks to PtSi Schottky barrier IR properties, great attention was dedicated to FPA arrays based on integrated silicon Schottky sensors which were showing reliable monolithic silicon CMOS integrated technology and high uniformity in detectivity, but were operating in the short wavelength region and with the limitation of low working temperatures. Similar considerations can be made for the long wavelength GaAs/GaAlAs multiquantum well IR FPA arrays [27], which, although if with lower quantum efficiency, are close to CMT performances even showing higher homogeneity and stability in sensitivity thanks to a more reliable manufacturing process, but with the strong limitation of working at lower temperatures (<77 K). This requires the use of cryogenic structures with the associated high cost of purchasing and maintenance; therefore, improving the restriction of the main use to military applications, limiting the market size and, as a consequence, product growth.

In all the latest developments the real driving key technology has been the integration of IR technology with silicon microelectronics and more and more emergence of the importance freeing IR from the constraints of the cooling requirements due to its high cost (almost a 1/3 of the total cost) and low reliability and heavy need for maintenance. For the above reasons, work on uncooled infrared detectors has shown impressive growth since the first developments, allowing the real expectation for production of low cost, high performance detector arrays which finally should follow the rules of a real global market, opening a real market for civil applications following the winning rules of silicon microelectronics. For these reasons the emerging room temperature detectors in the 1970s by the use of pyroelectric materials [28], which shows the limitations of not being fully monolithic, but the more innovative room temperature silicon microbolometers [29] appearing on the IR scene in the 1990s, appear to be a real breakthrough for future IR sensors.

Table 1 reports a schematic synthesis of the most important developments in thermal sensing and Table 2 reports the highlights in thermal sensor and IR sensor developments since Herschel’s discovery.

4. State of the art

Electro-optics (EO) technology is continuously growing in complexity and difficulty in science and in technology development: in particular, IR sensors might be considered at the frontier of solid state technology [17–26].

Electro-optics recently has shown an impressive growth in performance thanks to the reliability and cost achievable by the integration with advanced silicon microcircuit technology.

IR detectors can be structured into two main classes: cooled (photon detectors) and uncooled detectors (thermal or bolometer detectors) and each divide further into two classes, hybrid and monolithic. Infrared history is showing oscillations and variations in raw materials, technology processes and in device design and characteristics. In the IR sensor technologies we can perceive a general oscillation between the two main detection techniques (photon and bolometer effects) evaluated as the best ones depending, initially, on the single sensor performances and later, since the 1980s, on system use, especially after the development of the CCD/CMOS bidimensional FPA (Figure 7).

4.1. Photon cooled detectors

4.1.1. Hybrid CMT detectors

In the form of photoconductive (PC) detectors, the hybrid photon detectors have been the most fully developed detectors which utilise the photon-induced electronic change in bulk conductance as a detection mechanism. PC detectors must be biased with an external voltage. This technology has been used primarily in first generation scanned sensors. It was not a viable option for second generation focal plane arrays because of their difficulties to be integrated in a
complex structure; moreover, the power dissipation due to the continuous bias voltages produces considerable cooling power requirements. Photovoltaic (PV) detectors do not require a bias voltage and can hence be operated in the low power dissipation mode required for large second generation infrared arrays. PV technology is actually the primary detector technology being developed for large array applications; although PV detectors have a theoretical photo-current which is twice that of PC devices, because both holes and electrons contribute due to the charge separation caused by the internal junction field. Unfortunately, photovoltaic technology is intrinsically more non-uniform than PC detectors because of the variations due to intrinsic junction formation and consequent large variation in diode impedance. PV detectors are normally operated at a slightly negative bias voltage in scanned systems which are not affected by low frequency 1/f noise; staring PV detector arrays are normally operated at zero bias in order to minimise the 1/f noise which is present in the low frequency read-out noise bandwidth. HgCdTe is the most attractive option as a photon material because of the availability of high quantum efficiency PV detector arrays with the appropriate cut-off wavelength; moreover, HgCdTe diodes have quantum efficiencies in the 60%–70% range and array non-uniformities in the 10%–30% range. Current PV arrays are fabricated using liquid phase epitaxy. MOCVD growth technology is expected to improve the yield and non-uniformity of these arrays. FPA HgCdTe detectors with a million pixels have been developed by integrating CMT on a silicon substrate by MBE GaAs buffer layers for matching the thermal expansion coefficient. The major disadvantage of HgCdTe is its non-uniformity. Each of the above-described materials is normally associated with a particular multiplexer interconnection. HgCdTe focal plane arrays are hybridly integrated with silicon-based readout electronics. The detector readout connection can utilise either indium solder bumps, in a flip-chip configuration, or over the edge connections, if the HgCdTe strips are bonded directly to the silicon readout chip.

Table 1. History of thermal sensors.

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>3000 BC</td>
<td>Egyptian medical wizards</td>
</tr>
<tr>
<td>400 BC</td>
<td>Hippocrates</td>
</tr>
<tr>
<td>340 BC</td>
<td>Aristotle</td>
</tr>
<tr>
<td>II Century BC</td>
<td>Galeno</td>
</tr>
<tr>
<td>II–I Century BC</td>
<td>Philo of Byzantium – Heron of Alexandria</td>
</tr>
<tr>
<td>1605</td>
<td>Galileo</td>
</tr>
<tr>
<td>1654</td>
<td>Ferdinand II Duke of Tuscany</td>
</tr>
<tr>
<td>1724</td>
<td>Fahrenheit</td>
</tr>
<tr>
<td>1742</td>
<td>Celsius/Linneus (0–100)</td>
</tr>
<tr>
<td>1868</td>
<td>C. Wunderlich (thermometry in medicine)</td>
</tr>
<tr>
<td>1877</td>
<td>Lehman (liquid crystal thermometer)</td>
</tr>
<tr>
<td>1860</td>
<td>J.L. Gay-Lussac 1/267 (degree)–1</td>
</tr>
<tr>
<td>1847</td>
<td>Regnault 1/273 (degree)–1</td>
</tr>
<tr>
<td>1824</td>
<td>Carnot (ciclo)</td>
</tr>
<tr>
<td>1845</td>
<td>J.P. Joule (heat/work)</td>
</tr>
<tr>
<td>1860</td>
<td>W. Thomson (Lord Kelvin)</td>
</tr>
<tr>
<td>1900</td>
<td>ITS (International Temperature Scale) by CIPM (Comité des Poids et Mesures)</td>
</tr>
</tbody>
</table>

Table 2. History of IR detectors.

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1800</td>
<td>IR radiation</td>
</tr>
<tr>
<td>1821</td>
<td>Thermoelectric effect</td>
</tr>
<tr>
<td>1829</td>
<td>Thermocouple</td>
</tr>
<tr>
<td>1833</td>
<td>Thermopile</td>
</tr>
<tr>
<td>1836</td>
<td>Optical pyrometer</td>
</tr>
<tr>
<td>1873</td>
<td>Photodetection (selenium)</td>
</tr>
<tr>
<td>1884</td>
<td>IR radiation law</td>
</tr>
<tr>
<td>1902</td>
<td>Photoconductivity effect</td>
</tr>
<tr>
<td>1917</td>
<td>Lead sulphide</td>
</tr>
<tr>
<td>1933</td>
<td>Lead sulphide (galena)</td>
</tr>
<tr>
<td>1940</td>
<td>TI-S</td>
</tr>
<tr>
<td>1942</td>
<td>Golay cell</td>
</tr>
<tr>
<td>1948</td>
<td>transistor</td>
</tr>
<tr>
<td>1950s</td>
<td>PbS, PbSe, PbTe</td>
</tr>
<tr>
<td>1959</td>
<td>HgCdTe</td>
</tr>
<tr>
<td>1960s</td>
<td>Ge:Ga, InSb</td>
</tr>
<tr>
<td>1969</td>
<td>CCD</td>
</tr>
<tr>
<td>1970s</td>
<td>PbSnTe/HgCdTe,Si:Ga</td>
</tr>
<tr>
<td>1973</td>
<td>Common modules</td>
</tr>
<tr>
<td>1975</td>
<td>IR smart sensors</td>
</tr>
<tr>
<td>1978</td>
<td>Si:Ga/CCD/PSi/CCD</td>
</tr>
<tr>
<td>1980s</td>
<td>HgCdTe SPRITE</td>
</tr>
<tr>
<td>1990s</td>
<td>Pyroelectric FPAs</td>
</tr>
<tr>
<td>2000</td>
<td>MEMS FPAs – cantilever IR nanotubes/ nanowires</td>
</tr>
</tbody>
</table>
4.1.2. *InSb/LTT detectors*

The importance of integrating infrared detectors within a read-out electronics was the main reason for pushing the development of lead salts (PbS, PbTe, PbSnTe and PbSnSe) in the form of layers deposited on silicon substrate by the use of buffer layers of CaF$_2$ and BaF$_2$ or by a special thin film technology. It has to be underlined that lead salts are excellent IR detectors either in the form of polycrystalline layers as well as single crystals and are particularly suitable for complex pattern shapes. Moreover, well acquainted technologies such as MBE are particularly suitable for lead salt growth. A well-established material technology is based on using InSb with a CID readout structure. At 77 K working temperature, InSb cut-off wavelength is near 5 $\mu$m. Array non-uniformities in the range of 10% have been achieved; this technology has the limitation of short cut-off wavelength. InSb material is more mature than HgCdTe and good quality few inch diameter bulk substrates are commercially available: InSb devices are usually made with both p-n junction and MIS capacitors. Fabrication techniques for InSb photodiodes use gaseous diffusion, and a subsequent etch results in a p-type mesa on an n-type substrate. To maximise the available resolution and response of photodiodes, the bulk material is thinned to about 10 $\mu$m. Backside illuminated staring arrays with a format up 1024 x 1024 have been fabricated. The cryogenically cooled InSb and HgCdTe arrays have comparable array size and pixel yield at the MWIR spectral band. However, the shorter response time and high quantum efficiency have made HgCdTe the preferred material, although if crystal growth and process fabrication need higher complexity, it shows lower reliability.

4.2. *Monolithic photon detectors*

4.2.1. *Silicon Schottky diodes*

The main developments are based on Schottky barrier silicon (PtSi) detectors, with a rapid growth in recent years in the USA and Japan which allowed achievement of the highest number of pixels in FPA (>10$^6$ pixels). PtSi has a relatively low cut-off coupled to the need for high cooling; the only operational monolithic development working at 77 K is the PtSi arrays in the 3–5 $\mu$m range, but it has better uniformity than other detector technologies and can be integrated with silicon CCDs in monolithic structures. The negative charge of silicide is transferred to a CCD by the direct charge injection method. The effective quantum efficiency in the 3–5 $\mu$m atmospheric window is very low, of the order of a few percent, but useful sensitivity is obtained by means of near full frame integration in
area arrays. Moreover, for relatively long integration time (>10 ms), PtSi responsivity can be adequate for detection and hence PtSi appears to be one of the more promising technologies. Schottky photoemission is independent of such factors as semiconductor doping, minority carrier lifetime, and alloy composition, and as a result of this, has spatial uniformity characteristics that are far superior to those of other detector technologies. Uniformity is only limited by the geometric definition of the detectors. At present FPAs up to millions of pixels have been developed, but slow progress in the future is expected for this technology.

4.2.2. Multi-quantum well detectors

A new class of photon detectors has been developed by creating multilayer structures which allow a quantum well superlattice [27]. These structures are based on two mono-atomic layers of different semiconductor materials (GaAs and GaAlAs) with the creation of modulated electronic bands (normally referred to as NIPI structures). Detectors with \( D^* > 10^{10} \text{cm Hz}^{1/2} \text{W}^{-1} \) have been obtained at working temperatures \( \lesssim 70 \text{K} \). The possibility of this technology to integrate GaAs electronic devices can foreseeably allow important future developments although limitations in operative uses are expected due to the low working temperature needed for limiting the band intermodulation noise. Among the different types of quantum well infrared photodetectors (QWIPs), the GaAs/AlGaAs technology is the most mature with a number of military and commercial applications. QWIP cannot compete with the HgCdTe photodiode as a single device, especially at higher temperature operation (>77 K) due to fundamental limitations associated with inter-sub-band transitions. Even though QWIP is a photoconductor, several of its properties, such as high impedance, fast response time, long integration time, and low power consumption, comply well with the requirements of large FPA fabrication. At low temperature, QWIP has potential advantages over HgCdTe for VLWIR FPA applications in terms of array size, uniformity, yield and cost.

4.3. Room temperature thermal detectors

4.3.1. Hybrid: ferroelectric/pyroelectric detectors

The main technology has been based on ferroelectric materials which, thanks to the dielectric constant dependence versus temperature, can generate a voltage variation with capacitive structure detection and therefore a quick response, although measurable as a thermal variation [28]. These detectors are mainly coupled in a hybrid model (normally by InSn bumps) to an underlying stage of a silicon microcircuit. The performance is nowadays of acceptable level depending on the quality of the thermal insulation and therefore is normally operating in a vacuum encapsulation and thermoelectric thermal regulation. The imaging systems, based on pyroelectric arrays, usually need a reference obtained by optical modulators which chop the incoming radiation. Barium strontium titanate (BST) ceramic is a reliable material with a very high permittivity and usually has improved the performance of pyroelectric FPAs using a bias voltage applied to optimise the pyroelectric effect near the phase transition at the Curie temperature.

Although many applications for this hybrid array technology have been identified, and the imagers employing these arrays have been in mass production for some years, no hybrid technology advances are foreseen. The reason is that up to now the thermal conductance of the bump bonds is limiting the NETD (f/1 optics) to about 50 mK. Pyroelectric array technology therefore is moving toward monolithic silicon microstructure technology with a less complex and more reliable process with the problem that most ferroelectrics lose their properties as the thickness is reduced, even if lead titanate (PbTiO3) and related materials hold their properties well, even in thin film form. Various techniques for the deposition of thin ferroelectric films have been investigated, including radio frequency magnetron sputtering, dual ion beam sputtering, sol-gel processing, and laser ablation.

4.3.2. Monolithic: silicon microbolometers

This completely new approach based on microsystems technology (MST) is showing excellent performances and seems to offer very promising future developments [29]. Monolithic silicon FPAs have been fabricated by realising silicon-based bolometers in microminiature form by etching a thin bridge bolometer layer connected by two thin legs to the underlying silicon substrate [30]. The whole bi-dimensional array with a high number of pixels (>105), with a 50 \( \mu \text{m} \) pixel, is maintained under vacuum and thermally regulated by a single thermo-element cooler in order to obtain outstanding performances (NETD<50 mK). This technology is moreover showing the advance, being integratable with silicon technology, of having the detector read-out circuits enabling a maximum in the filling factor. The most popular thermistor material used in fabrication of the micromachined silicon bolometers is vanadium dioxide, VO2. From the point of view of an IR imaging application, probably the most important property of VO2 is its high negative temperature coefficient of resistance (TCR) at ambient
temperature, which exceeds 4% per degree for a single element bolometer and about 2% for a FPA, with monolithic readout circuits integrated into underlying silicon. At present, several research programmes are focused towards larger array size and enhancement of performances with $D^* > 10^6 \text{cm Hz}^{1/2} \text{W}^{-1}$ and some of them are based on an all-silicon version of the microbolometer. It is anticipated that some new materials will form the basis of the next generation of semiconductor film bolometers, although the most promising material appears to be amorphous silicon which is completely integratable with IC manufacturing processes. This new original solution showing acceptable performances coupled to a low cost approach (full silicon technology integrability and room temperature operation), can create a real breakthrough in infrared sensor technology.

5. Future infrared detectors

Electro-optics (EO) is continuously growing in complexity and difficulty in science and technology developments: in particular, IR sensors might be considered at the frontier of solid state technology [30–39].

Electro-optics has shown winning performances thanks to the reliability and cost achievable by integration with advanced silicon microcircuits. The exploding market of optical telecommunications has, in some way, concentrated the main interest in the near infrared bands, pushing the infrared technology (especially the long wavelengths) towards thermovision and surveillance applications. This gave an impressive impulse to the development of FPA arrays with ever growing performances and tasks for these advanced high-tech components and systems. The main effort nowadays in IR detector developments is oriented towards the highest number of pixels (of the order of $10^6$) with a FPA structure, with integrated electronics for signal read-out and elaboration, with working temperatures close to room temperature and with high uniformity. Bi-dimensional FPAs coupled to $X–Y$ readout integrated microelectronics (CCD and CMOS devices) are allowing great improvements in performances, but also in reliability and cost, improved by the fact that, thanks to the increased integration time possible in staring arrays, NETD values close to the BLIP limit can be more easily achieved. This, with the main task of achieving high optoelectronic FPA performances with smaller and lighter structures, gives possibilities for applications in civil areas thanks to cost reduction by eliminating opto-mechanical scanning and cryogenic cooling. The possibility of an advanced IR FPA (with $>10^6$ pixels in 8–12 μm) working at room temperature, allows one to forecast an incredible growth of uses and applications besides the evident growth of advanced applications of IR science based on thermal measurements.

There is a large amount research activity directed towards 2D staring array detectors consisting of more than $10^6$ elements. IR FPAs have a similar growth rate as dynamic random access memory (RAM) integrated circuits (ICs) [40,41] (it is a consequence of Moore’s Law [42], which predicts the ability to double transistor integration on each IC about every 18 months), but with a lag of about 10 years and with some breakpoints (linear arrays and FPAs) and with a saturation trend due to the physical constraints of the diffraction law, as explained soon after. Figure 8 illustrates the trend in array sizes for various technologies over the past 40 years and some forecasting for the future. (Actually the
largest HgCdTe FPA is a short wavelength IR (SWIR) hybrid 2048 × 2048 with a unit cell size of 18 μm × 18 μm for astronomy and low background applications [43].

The pixel size is conditioning the achievable number of pixels achievable in a feasible FPA chip size conditioned by the size and the cost of the optics, and overall by the fundamental limit on the pixel size determined by the diffraction law [34,35]. In fact the size of the diffraction-limited optical spot, or Airy disk, is given by:

\[ d = \frac{2.44 \lambda}{F} \]  

where \( d \) is the diameter of the spot, \( \lambda \) is the wavelength, and \( F \) is the \( f \) number of the focusing lens (for high luminosity \( F/1 \) optics at 10 μm wavelength, the diffraction limited spot size is \( \sim 25 \mu m \)).

Therefore, in the future a reduction for pixel size mainly thanks to oversampling (up to a factor of 4) could be achieved for applications requiring high spatial resolution: the LWIR pixels could reach a limit of \( \sim 5 \mu m \) and SWIR pixel sizes could shrink to correspondingly smaller dimensions. Anyway the pixel size, unlike the size of a DRAM memory cell, cannot be significantly reduced and therefore to increase the number of pixels in IR FPAs, the chip size should grow with a high cost and limited handling (this a further reason for which room temperature IR sensor technologies allowing monolithic integration with silicon microelectronics are favourite for future developments) [66]. New requirements in most advanced applications in military surveillance, astronomy, medical and environment fields are pushing the request, as a strategic parameter, of the multi-spectral operability (for some applications even hyperspectral resolving detection are required) [66]. This can allow one to foresee either high sensitivity tunable bandgap photon detectors or wide-band microbolometers with integrated coating multispectral filters. The cost of an FPA depends on the type and maturity of the technology with different reduction factors in the case of mass quantity production (again production based on silicon technology is more competitive) [37,67]. The cost of an IR sensors array can be shared in the following main blocks: (a) IR FPA physical detectors, including read-out electronics, (b) assembly and testing, (c) optics with mechanical structure, and (d) cryogenics in the case of cooled detectors. Up to a few years ago there was the thumb’s evaluation of almost 25% for each part, although for high performance photon-cooled hybrid detectors, more than 50% was due to the sensor chip and a significant cost improvement in the cryogenic cooler was to be taken into account in the case of working temperatures, lower than 77 K. This explains why PtSi and QWIP detectors requiring, respectively, wider luminosity optics and lower operating temperatures, have a cost comparable to other photon detectors, even though the raw material (Si or GaAs) and fabrication technologies are much cheaper than for HgCdTe.

IR imagers mainly for military applications, actually using cryogenic or thermoelectric coolers, complex IR optics, and expensive sensor materials have typical costs of some tens of thousands of Euros, while IR cameras using emerging micromachined silicon bolometer arrays with NEDT close to 10 mK with costs of a few thousands of Euros are expected in the civil market in the near future. This cost in the case of mass production, as is expected for collision avoidance and guidance

Figure 8. Number of pixels in infrared detector arrays versus time (Moore’s law for comparison) [66].
assistance in automobile production, could be reduced down to less than one thousand Euros.

The actual explosions of microsystem technologies (sensors, control and actuators), especially in automotive applications, is, moreover, increasing this positive growth of the IR market with a real possibility of high level products at contained cost. New markets (automotive, law enforcement, intelligent building, environmental control) and old markets (biomedical and medical, industrial process monitoring, energy control, surveillance and warning) will get strong benefits from the new technology feeding also new military market applications (especially portable equipment, soldier of 2000). In Figure 9 is shown the single pixel detector cost in the last 50 years (at actual value of the year 2000) with a reduction of more than $10^6$, that is comparable only to the cost reduction for 1Mb memory chip for computers obtained by silicon microelectronics, which is the most advanced technology developed up to now (this is another factor demonstrating the strong liaison between IR FPAs and silicon microelectronics).

In conclusion, the cost of IR sensors strongly reduced for a single pixel cost is however still high because of the improved performances of the new high quality thermal camera (similar analysis can be done for PC where the cost of a memory bit has been reduced by a factor of $10^8$, but the cost of the top computer is reduced of a factor of 10). A strong reduction of the cost, even for high number pixel IR FPAs is expected thanks to silicon technology, which in the case of the silicon microbolometer is avoiding the high cost of cryogenic components. For future consistent cost reduction, particular efforts should be dedicated to low cost, high luminosity IR optics based on organic materials highly transparent in the long wavelength IR region.

6. Smart sensors

Associated with the push towards the highest number of pixels ($>10^6$) and the highest working temperatures (close to room temperature), the general trends of future detectors will show more and more an increase of the ‘intelligence’ of the sensors which will integrate the sensing function with the signal extraction, processing/’understanding’ (Smart Sensors) [44–47].

The term ‘Smart Sensors’ originated to indicate sensing structures capable of gathering in an ‘intelligent’ way and of pre-processing the acquired signal to give aimed and selected information. The Smart Sensor technology, based on the use of a smarter sensor architecture, allows one to integrate technical design and development (from optics, detector materials, electronics, and algorithms) into the sensor’s functions rather than trying to get the required performance by just relying on one aspect of the technology, for instance the number of sensor pixels.

One of the most advantageous application areas for Smart Sensors is the infrared field where the information to be extracted is generally based on very small signals buried in highly intensive and diffused background noise and often high intensity ‘unwanted signals’. In general, background clutter consists of extended objects that are more slowly varying spatially than the target, therefore temporal filtering as well as spatial filtering, further complemented by multispectral filtering, are required for target signal detection and extraction. This implies that infrared imaging devices require some processing of detector output signals to correct non-uniformity and remove the background effect and to avoid this, without on-focal-plane processing, most of the data would be useless clutter or unwanted data. Because of the whole acquired pattern only a few pixels contain the target information of the selected targets.

Figure 9. Single sensor pixel cost evolution [66].
Therefore, conventional approaches need to process these confused data through all the chain of the readout electronics, the analogue to digital converters and the digital signal processor before, finally, separating and rejecting the clutter. In contrast, the Smart Sensor rejects this clutter before it is read off the focal plane sensors so that most of the useless data is not processed. The Smart Sensor design concept is based on the processing capabilities, at least at some stage of the threshold, inside the sensor structure itself. This means that the Smart Sensor in some way emulates a living eye in the simplest stage, at least at a primordial level, such as an ‘insect eye’ (Figures 10 and 11), and, in future perspective, could reach performances close to the ‘human eye’, thanks to neuronal network development, which can allow pattern recognition and object discrimination [45,48,68].

One of the simplest feature extractions and at the same time most appealing for the numerous applications, is the discrimination of point sources from extended background emissions and/or of fast events (moving targets or changeable emissions) for static or slow moving scenario just as is operating in the fly eye.

In this case a reticule structured detector, which is electronically modulated to obtain a spatial–temporal correlation of the focused spot target, normally buried within the diffused background emission, can allow the detection of a point source or a well-defined shaped target improving the signal-to-clutter ratio: this correlation associated with appropriate temporal signatures, can allow one to discriminate and identify the targets, like is performed by an insect eye thanks to a spatial–temporal correlation [44,45]. IR Smart Sensors have been developed by integrating a columnar three-electrodes sub-structure designed to capture a fast moving point target by a read out processing for clutter rejection performed in an analogue way within the multielectrode sensors on the focal plane. A scheme for the three-electrode electronics read-out with a reticule structure or implementing this analogue processing within the sensor on the focal plane is shown in Figure 12. In Figure 13 the first IR smart FPAs with 1024 sensors, each one structured in 21 sub-pixels [44,45] is shown.

The basic objectives of new IR smart sensors are much more demanding because significant improvements in the performance of VLSI processors and infrared mosaic detector arrays are being achieved, especially in the pre-threshold stage. Target signals are expected to be deeply buried in background clutter noise which can be much higher than the target intensity. Therefore, imaginative pattern recognition processing techniques using all spatial, temporal and/or spectral information of both targets and background clutter should be developed for suppressing background clutter and unwanted signal, but maintaining or even enhancing the target signal. Finally, it is important to underline that the important recent developments of neural networks for advanced computing allow foreseeably an impressive growth of the Smart Sensor concept especially for those detector technologies which will take advantage of the possibility of integrating processing devices.
7. Applications highlights

7.1. Surveillance, detection and warning

The main application for IR technology was in the past and will be in the future surveillance and warning and more specifically for military applications, that conditioned for a long time, the development of IR devices and systems. The highlights of this application field are quite well known and have allowed one to develop an impressive know-how in system performances. The first historical lesson, missed for lack of knowledge of the users, was the underestimation of the strategic value of IR surveillance systems (at that time RADAR was not yet operating although in 1904 Christian Hülsmeyer had used radio waves for detecting ‘the presence of metallic distant objects’, and only in 1922 did Guglielmo Marconi propose the idea of a RadioTelemeter for localising metallic objects at distance and therefore the remote sensing was mainly just optical). In fact in 1910, Bellingham had presented a method to detect the presence of icebergs and steamships by using a mirror and the original thermopile. He later patented this device in 1913. His infrared radiometer’s primary advantage over the disappearing-filament optical pyrometer was that it was able to detect temperatures substantially lower than ambient. If this device had been installed on the Titanic ship, avoiding that grave tragedy, efforts in developing IR surveillance systems would probably have been much greater. During World War II great efforts were dedicated to the development of IR surveillance systems especially in the Army with both...

Figure 11. Smart sensors – foveated sensor [40,45].

Figure 12. Reticule structured smart sensor [40,45].
parties capable of IR detection of enemy tanks and support to night moving.

After the 1970s R&D developments of IR surveillance systems, especially for navy applications, were carried out in the US, UK, France and Italy especially (where the first modular FPA staring omni-directional surveillance system prototype was designed and realised) [44].

In the 1980s the SDI Programme for Ballistic Missile Defence in the US originated a highly advanced EO surveillance system with performances close to BLIP limits. Nowadays, the main efforts are dedicated to multispectral detection capability.

Lastly, wide application of IR warning is expected in automotive applications for smart collision avoidance systems in poor visibility conditions: room-temperature smart IR receivers could be installed with high reliability and a simple, immediate man–machine interface in any type of motor vehicle [49].

7.2. Medical applications

Medical applications were among the first and the most important civil applications, although they have not been successfully recognised for their applicability value because of some specific, marginal limits which, at the beginning caused misunderstandings in their interpretation, especially among medical users who were not expert enough in IR knowledge.

The first use of thermography in medicine was by the Greek physician, Hippocrates, in about 400 BC. He was able to gain information about diseases by measuring the temperature distribution on immersing the body in wet mud: the areas that dried more quickly indicated a warmer region and were considered to be
diseased zones therefore indicating pathology of the underlying organs. He wrote, 'In whatever part of the body excess of heat or cold is felt, disease is there to be discovered' [12].

Real IR thermography was first used to investigate breast cancer in 1957 by Raymond Law. Since then, there have been thousands of peer-reviewed studies in the medical literature, studying hundreds of thousands of women [50–56]. After approval by the FDA in 1982 of thermography as a breast screening procedure, there have been an impressive number of studies supporting its effectiveness, as well as the development of strict, standardised interpretation protocols, strongly supported by great technological advancements in imaging hardware and computer processing with the introduction of digital image processing, later on better known as digital infrared imaging (DII) [55]. DII emerged in the 1980s, thanks to the use of highly sensitive cameras to accurately measure heat from the surface of the breast, detecting temperature differences of close to 0.1°C and allowing automatic computation of the statistical thermal distribution [57].

DII, a completely non-invasive [69], non-ionising, passive diagnostic technique, demonstrated the capability of detecting a pre-cancerous state and/or the early signs of breast cancer impossible to detect by physical examination or mammography, and even benign tumours, simple cysts, fibro cysts, infections, and other benign conditions could be detected. DII can also be the first indicator that a cancerous situation is developing [51–54], even in the early stages of tumour formation, and can supply invaluable information on prognostic evolution. Finally, each patient has a unique infrared map of their breasts (defined IR pattern signature) which should remain constant in its statistical distribution unless there is a change in the
underlying vascular conditions. The DII has an invaluable diagnostic performance that unfortunately has not yet been recognised worldwide. The main reason for this disappointment is due to its high sensitivity not coupled to a specific cause: that is IR imaging is highly sensitive to thermal variations (down to better than 0.01°C), but it is not able to identify the specific causes of these variations (which moreover are originated only by vascular links close to the skin) and therefore can be easily misinterpreted in the absence of a deep knowledge in thermal physiology. Thus, being a functional and not a morphological diagnostics tool, while the high sensitivity supplies invaluable information for early breast cancer detection especially for young women, deep necrotic breast cancer might not be detected in the case of a weak knowledge in thermography. New positive perspectives have emerged recently thanks to new medical applications in highly specialised fields, where the knowledge of specific diseases are linked to functional behaviour by IR thermography which is showing the whole diagnostic potentiality, especially when applied to specific, well pre-defined diseases related with surface vascular problems (detection of malignant moles, vascular problems, etc.) and/or rheumatism and/or control in high blood flood and dermatology surgery. Lastly DII, thanks to the expected cost reduction for high sensitivity room-temperature thermal cameras, is allowing one to forecast a wide market use in sport-related trauma and veterinary diagnostics.

7.3. Civil industrial applications: energy control and design

With the help of new IR systems the spectrum of civil uses can be applied everywhere where processes inside of an object lead to a change of temperature of its surface. Bearing in mind the current discussion on climatic changes and environment control, applied thermography is gaining importance, as for example, it can quickly discover heat losses from building envelopes.

7.3.1. Quantitative thermography

An overview of the current R&D works about the use of IR thermography for non-destructive testing is a large task due to the continuously increasing popularity of such a method. It is a matter of fact that ASNT included thermography in the main standardised method [58].

The quantitative use of IR thermography covers a large number of different disciplines. Actually, some applications are on the boundary with photo-thermal science, especially when applied to material characterisation [58].

Thermography is a very powerful tool, because with a calibrated device it is possible to measure radiance, temporal and surface properties at the same time. Temperature is not always simple to recover, but in favourable cases, accuracy in the order of millikelvins is possible today. Because the lowest degradation step of any physical phenomena is heat, uses of temperature maps, so accurate and fast as 1 Mpixel in a few milliseconds, are far from being fully exploited.

7.3.2. Thermal non-destructive evaluation of materials

A particularly important and well-documented area is building inspection, conducting infrared analysis of buildings, analysing thermal patterns/thermovision with an in-depth understanding of physical underlying phenomena. Dealing with historical buildings and works of art, different purposes could be achieved, for instance, the non-destructive evaluation of materials, the discovery of the hidden history of a monument, for anti-seismic purposes or very recently, the monitoring of environment conditions [59,60].

A new approach in monitoring the indoor environment of a building has been recently developed using a thermodynamic basis. It uses temperature as the driving parameter and is especially suited for comfort analysis or moisture evaluation. IR thermography is the method more suited to cope with such a task. It allows measuring, at the same time, of surface temperature and all the basic moist air parameters, such as, air temperature, relative humidity and air speed. In such a way a very detailed heat and mass balance could be accomplished close to the wall or in any other part of the room.

Even the energy saving issue could be deeply investigated using an optical measurement of temperature and heat flux. Basically, the non-contact nature of such a measurement is much less prone to error than any contact measurement. In fact, the time constant needed for the thermal equilibrium, which is a great source of errors in traditional measurement techniques, is practically equal to zero. Last, but not least, the temperature recorded by any contact probe is a mixture of surface and air temperature. This is not the case using radiometric measurements. Furthermore, the imaging of a large surface is the only way to achieve a correct mean value and measure local anomalies. The literature provides standard practices for thermographic inspections of buildings [61].
7.3.3. Thermal diagnostics in fluid dynamics

An important application of high resolution, high frame rate thermography, especially for the high scientific knowledge involved, is to visualise and analyse flow phenomena. Infrared technology has been employed in wind tunnels and in flight at subsonic to hypersonic conditions with both local (that is, when the camera and subject surface are mounted on aircraft) and remote camera installations [61–64]. With state-of-the-art equipment and advanced methods, greater success and improved resolution of details of various flow phenomena have become possible. Infrared thermography has many benefits over other methods. It is global in nature and non-intrusive. The basic principle behind IR thermography is the measurement of surface emissions in the IR radiation band, which are directly related to surface temperature [62–64]. Surface shear stress and, thereby, convective heat transfer with the free stream varies with the boundary layer state where the boundary layer state changes, such as at transition. With state-of-the-art equipment the temperature difference changes can be visualised. Shock waves and possibly flow can be measured to an accuracy of nearly 0.1°C over a small area [63].

These characteristics make IR thermography a very powerful tool to visualise certain flow phenomena. In addition to transition, any flow phenomena that creates a measurable temperature change can be visualised.

The deep scientific knowledge required in this application guarantees the reliability of the IR measurements, although as a counterpart, it is limiting the market size to a few R&D centres.

8. Conclusions

Future trends for infrared detectors are linked to various and complex parameters like the emerging technologies of sensor fabrication for mass production especially for civil applications, with a strong two-way synergy between civil and military applications. New markets (automotive, intelligent building, environmental control) and consolidated markets (biomedical and medical, energy control, surveillance and warning) will get strong benefits from microsystem technologies (sensors, control and actuators) especially in automotive applications, with a real possibility of high level products at contained cost feeding a double-wave of technology transfer between civil and military applications (in the future more and more towards new military market applications, especially portable equipment). The competition among various technologies and ‘technical schools’ has been strong with unforeseeable emerging new actors in the last few years (overall, room-temperature microbolometers for future and extremely valid applications in the civil field). Operational requirements (mainly of maintenance and reliability) are pushing IR science to look for new advanced sensors which could avoid the need for cryogenics. New microbolometer technologies, because of the micro size of thin film bolometers, completely integratable with silicon technology and therefore often named silicon microbolometers, have been emerging in the last few years with a very high promise for future IR sensor market growth. For extremely high sensitivity sensors, especially for military and space applications, the technologies with actual major possibilities for future development are mainly based on photon sensors (intrinsic and quantum well). Multispectral and hyperspectral capabilities with spatial–temporal filtering capabilities are also emerging, especially in military and space applications for target identification. The key emerging factor for future IR FPA technologies is room-temperature working for uncooled imaging systems and the complete integrability with silicon microcircuit technology, especially because integrated signal processing (smart sensors) will play a fundamental role in future applications where mass production could allow consistent cost reduction (huge markets are expected especially for automotive applications for the unit cost of a few hundred Euros).

So, for the first time, thanks also to the elimination of cryogenic cooling, wide use of IR Smart Sensors are emerging on the international market, becoming strategic components for major important areas, such as transport (especially cars, aircrafts and helicopters), security, environment and territory control, biomedicine and in helping ‘human beings to a better life’ (intelligent building, energetic control structuring, auxiliaries for disabled people, etc.).

References
