Chapter 6

Uncooled Detection

The ultimate IRFPA will operate at room temperature and exhibit background limited performance (BLIP), regardless of the operating cutoff wavelength. Currently the primary vehicle for room-temperature IR detection is the thermal detector, but it is essentially limited to the detection of LWIR radiation, at relatively slow frame rates, or integration times. These limitations were discussed briefly in Sec. 2.2 for the optimized perfect thermal detector. The successful room-temperature detection of the complete IR spectrum, with BLIP performance, will require the utilization of other detector concepts. The HOT detector of Elliott and Ashley [80] is just such a concept, the potential of which is limited entirely by the quality of the direct bandgap semiconductor utilized to detect the IR radiation, and the contacts and passivation applied to that material. The HgCdTe HOT detector concept is discussed here, together with a consideration of relevant material properties within the active volume of the device for successful room-temperature operation, taking into account the all-important issue of Shockley–Read (S-R) centers in HgCdTe. The HOT concept is not limited to the HgCdTe materials system alone, but can be implemented in any direct bandgap semiconductor material, or its bandgap engineered equivalent.

6.1 Thermal Detection

The principle of the thermal detector is illustrated in Fig. 6.1. The detector element, of heat capacity C_{th}, is thermally coupled to a heat sink by a conductance G_{th}. The thermal detector element is micromachined onto the surface of a silicon readout integrated circuit (ROIC) and the resulting pedestal is connected to the ROIC by thin metal buss leads. The pedestal can be fashioned from any one of a number of temperature-sensitive materials that are process compatible with silicon ROIC processing. We will assume here that the temperature sensitivity of the element is sufficiently large that all noise sources other than temperature fluctuations of the detector element can be ignored. As discussed in Sec. 2.2, the power fluctuation of the element is then given by $\Delta W_{\text{n}} = [4kT^2 G_{\text{th}} \Delta f]^{1/2}$, where Δf is the noise bandwidth, which translates into a noise temperature fluctuation, $\Delta T_{\text{n}} = [4kT^2 R_{\text{th}} \Delta f]^{1/2}$. The detector signal is

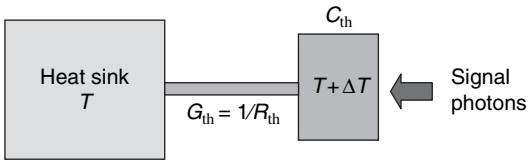

Figure 6.1 IR thermal detector.

given by $\Delta T_s = \eta R_{th}(dP/dT)_\lambda \Delta T A/(1+4F^2)$, where η is the quantum efficiency, $(dP/dT)_\lambda$ the differential change in radiated power per scene temperature change in the spectral region of interest, A the detector area, and F the $F/\#$ of the optical system. Thus the detectivity of the detector, defined as [signal/flux power][$A^{1/2}$/(noise/$\Delta f^{1/2}$)], is given by

$$D^* = [\eta^2 A/(4kT^2 G_{th})]^{1/2}. \tag{6.1}$$

If G_{th} is radiatively limited, then $G_{th} = 4PA/T$, where $P = \sigma_B T^4$. Thus for $\eta = 1$, $D^* = 2 \times 10^{10}$ cm Hz$^{1/2}$/W, and is independent of area.

This value of D^* is comparable to the theoretical limit of a LWIR photon detector, but provides only a limited insight into the suitability of the thermal detector for specific IR systems. As pointed out in Sec. 2.2, signal and spectral bandwidth requirements are also important, and to this end it is meaningful to consider the available noise equivalent temperature difference, NEΔT.

Correlated double sampling [4] of the bandwidth-limited temperature fluctuation, with an integration time τ_{int}, gives $\Delta T_n = [2kT^2(1 - \exp(-\tau_{int}/R_{th}C_{th}))/C_{th}]^{1/2}$, where we have assumed $\Delta f = 1/4R_{th}C_{th}$. Equating this to a signal temperature change given by $\Delta T_s = (dP/dT)_\lambda \Delta T A(1 - \exp(-\tau_{int}/R_{th}C_{th}))(1+4F^2)R_{th}$, gives

$$\text{NE}\Delta T = [2kT^2/C_{th}/(1 - \exp(-\tau_{int}/R_{th}C_{th}))]^{1/2}(1+4F^2)/(dP/dT)_\lambda/AR_{th}. \tag{6.2}$$

Image smearing considerations require $\tau_{int} \approx 2R_{th}C_{th}$, resulting in

$$\text{NE}\Delta T = [2kT^2/C_{th}]^{1/2}(1+4F^2)/(dP/dT)_\lambda/AR_{th}.$$

For $R_{th} < R_{rad}$

$$\text{NE}\Delta T \approx [2kT^2 C_{th}/\tau_{int}^2]^{1/2}(1+4F^2)/(dP/dT)_\lambda/A, \tag{6.3}$$

and for $R_{th} = R_{rad}$

$$\text{NE}\Delta T \approx [32k/C_{th}]^{1/2}(1+4F^2)/[(dP/dT)_\lambda/P]. \tag{6.4}$$

This predicted dependence of NEΔT on integration time, τ_{int}, for LWIR, and MWIR detection is shown in Fig. 6.2, for a 25-μm pixel, operating with $F/1$ optics, and $\eta = 1$.

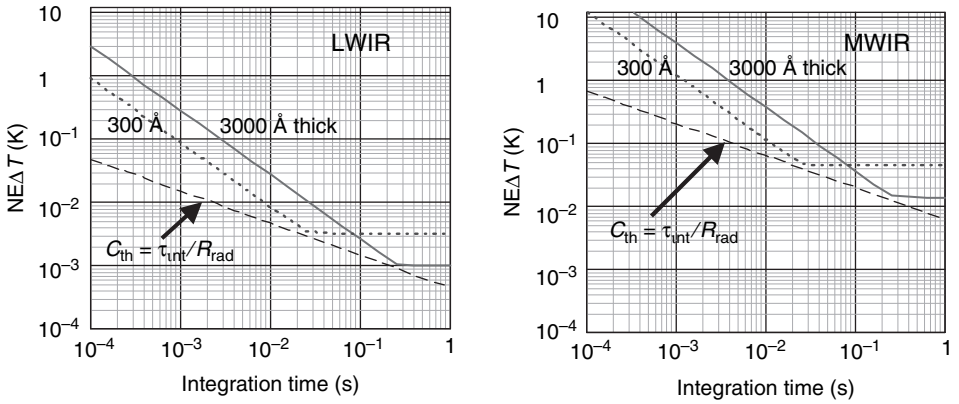

Figure 6.2 NEΔT vs. integration time for perfect LWIR, MWIR thermal detectors, for 25 μm pixels, $F/1$ optics, and $\eta = 1$.

The curves labeled $C_{th} = \tau_{int}/R_{rad}$ represent the optimum mode of detector operation, assuming that (1) the detector thermal conductance is limited by radiation to its surroundings, and (2) the thermal capacitance can always be made small enough to satisfy the requirement of $\tau_{int} \approx R_{rad} C_{th}$. In practice, the available thermal capacitance is limited by the thickness of the bolometer pedestal and material specific heat. Realistic values for bolometer thickness are in the range 300–3000 Å, and the volumetric specific heat is 1–2 J/cm^3/K. Examples of fixed bolometer thickness values of 300 Å, and 3000 Å are shown in Fig. 6.2, encompassing both ranges of R_{th} described by Eqs. (3) and (4). Bolometers today are typically 3000 Å thick, and their performance approaches the values shown in Fig. 6.2 for LWIR at relatively long integration times, such as 15–30 ms. However, it is apparent that high performance in a snapshot mode ($\tau_{int} \approx 2$ ms) will not be realized. At best, NEΔT values can approach 0.1 K for LWIR under these conditions. The NEΔT achieved by thermal detectors in the MWIR is mediocre at best. This is due to the low value of $(dP/dT)_\lambda$ available for signal in this spectral band, relative to the temperature noise, which is determined by all of the black-body flux absorbed by the bolometer. Theoretically, this situation could be improved by resonant tuning of the pedestal to its surroundings for the same IR spectral band as the desired signal; however, this would not only be difficult but, if successful, would also increase R_{rad} significantly, and result in considerably longer thermal time constants. For slower optics the situation is exacerbated, varying as $(F/\#)^2$. Thinner detector elements improve the NEΔT as $t^{1/2}$.

6.2 Photon Detection

6.2.1 HOT detector theory

The high operating temperature (HOT) photon detector concept, first proposed by Elliott and Ashley, is depicted in Fig. 6.3 for an $n^+/\pi/p^+$ architecture, where π

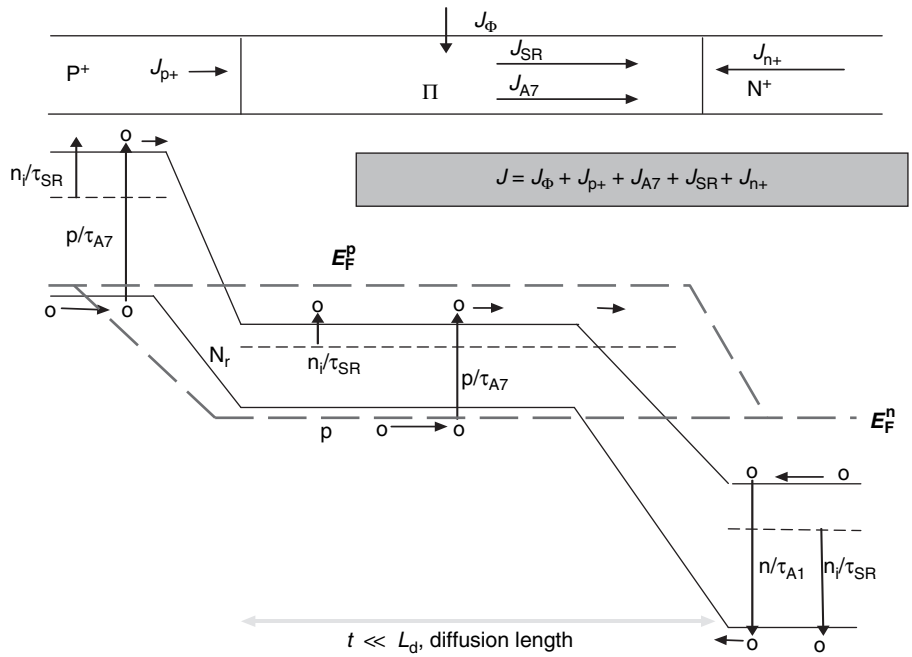

Figure 6.3 HOT detector concept with dark current sources.

designates an intrinsic region containing a p-type background dopant. The intrinsic IR absorbing volume is connected to both a minority carrier contact and a majority carrier contact. The geometry of the active volume is small relative to a minority carrier diffusion length. It is operated in strong non-equilibrium by reverse biasing the minority carrier contact to completely extract all of the intrinsically generated minority carriers. Charge neutrality in the active volume is violated, creating an electric field to sweep out the intrinsically generated majority carriers until the majority carrier concentration equals the background dopant concentration. The active volume thus consists of a depletion region, with a width determined by the doping concentration and applied bias, and a field-free region with a majority carrier concentration determined by background doping. The elimination of minority carriers throughout the device means that minority carrier recombination is a non-issue, and dark current is determined entirely by generation mechanisms within the active volume of the device, plus possible components associated with injection from the contact and surface regions. The relevant dark current mechanisms within the active volume of the HOT detector are (1) Auger generation associated with majority carriers in the field-free region, and (2) thermal generation through S-R centers throughout the whole non-equilibrium volume. As discussed earlier, the arguments of Humphreys suggest that radiative mechanisms are not important in this device.

Auger generation–recombination in direct-gap semiconductors is a well-understood phenomenon involving the interaction of three carriers. The Auger1

lifetime in n-type material is due to the interaction of two electrons and a heavy hole and is given by

$$\tau_{A1} = 2\tau_{Ai1} n_i^2 / n(n+p), \tag{6.5}$$

$$\tau_{Ai1} = 7.6 \times 10^{-18} \varepsilon^2 (1+\mu)^{1/2} (1+2\mu) / ((m_e/m_o)|F_1 F_2|^2 (kT/E_g)^{3/2}) \times [\exp((1+2\mu)E_g/(1+\mu)kT)], \tag{6.6}$$

and $\mu = (m_e/m_h)$. τ_{Ai1} is defined as the intrinsic Auger1 lifetime. The largest uncertainty in the Auger model lies in the calculation of the overlap integral $F_1 F_2$, and empirical values, provided by a comparison to experimental data [81], are typically used. For HgCdTe, the empirical value for $F_1 F_2$ is ~0.22. In p-type material Auger recombination involves two holes and an electron and is referred to as Auger7. Calculations by Casselman [82] for HgCdTe give $\tau_{Ai7} \approx 6\tau_{Ai1}$, although experimental data would indicate that it may be somewhat higher, namely $\sim 10\tau_{Ai1}$. This longer lifetime suggests that p-HgCdTe is the material of choice for the HOT detector active volume with regard to ultimate dark current performance. The Auger generation rate/unit volume in the field-free region of the HOT detector, under non-equilibrium conditions, is proportional to the majority carrier concentration, and is given by

$$G_{A7} = n_a / 2\tau_{Ai7}. \tag{6.7}$$

In p-HgCdTe under non-equilibrium conditions, the minority carrier generation rate/unit volume associated with an S-R center of density N_r, located E_r from the conduction band, is given by

$$\begin{aligned} G_{SR} &= n_i^2 / [\tau_{pr} n_1 + \tau_{nr}(p + p_1)], \\ n_{1r} &= N_c \exp(-E_r q/kT), \\ p_{1r} &= N_v \exp(-(E_g - E_r)q/kT), \end{aligned} \tag{6.8}$$

where $\tau_{nr} = 1/\gamma_n N_r$, $\tau_{pr} = 1/\gamma_p N_r$. γ_n and γ_v are the capture coefficients for electrons and holes into N_r, and N_c, and N_v are the conduction-band and valence-band density of states respectively. In the field-free region of the detector $p = n_a$, whereas in the depletion region $p = 0$.

The dark current density in a simple planar geometry HOT detector of thickness t, with a depletion region width W, is thus given by

$$J = q n_a (t - W)/2\tau_{Ai7} + \sum^{r} q n_i^2 \{W/(\tau_{pr} n_{1r} + \tau_{nr} p_{1r}) + (t - W) / [\tau_{pr} n_{1r} + \tau_{nr}(n_a + p_{1r})]\}, \tag{6.9}$$

where the summation is carried out over all the S-R centers in the device. The materials technology development required for HOT detector optimization is clear:

- Reduce the majority carrier concentration in the active volume of the device, and
- Reduce the density of S-R centers in the active volume to a minimum.

Reducing the majority carrier concentration is beneficial in two respects. It not only results in a reduction in Auger generation rate, but also results in a larger depletion region for the diode at a constant bias, and hence a smaller required field-free volume. For an applied bias of 1 V, a depletion region width of 5 μm is achieved for doping concentrations of $10^{14}\,\mathrm{cm}^{-3}$. Such a depletion region width will provide ample absorption of the IR radiation, with good quantum efficiency, and completely eliminates Auger generation due to the total absence of majority carriers. The device is then essentially a P-I-N diode, and either p-type or n-type material can be employed for the intrinsic region. Interestingly enough, doping concentrations of $1–2 \times 10^{14}\,\mathrm{cm}^{-3}$ are already available using indium background doping in LPE layers grown in Te-rich melts. A P-I-N diode is shown in Fig. 6.4, where the p^+ contact is shown as wide-bandgap HgCdTe, to avoid the possibility of minority carrier injection into the intrinsic active volume of the HOT device. The band-filling in the n^+ contact is assumed deep enough to inhibit Auger transitions in that region, and hence minority carrier generation and injection. A wide-bandgap n^+ region could also be utilized for this purpose. The dark current in such a device is given by

$$J_\mathrm{d} = \sum_\mathrm{r} q n_\mathrm{i}^2 [W/(\tau_\mathrm{pr} n_\mathrm{1r} + \tau_\mathrm{nr} p_\mathrm{1r}], \tag{6.10}$$

where the summation is over all the S-R centers, and the depletion width is equal to the device width. The most efficient S-R centers for thermal generation are normally located close to the intrinsic energy level in the bandgap, for $\tau_\mathrm{pr} \approx \tau_\mathrm{nr}$. In HgCdTe, the intrinsic level is located far above mid-gap due to the fact that the conduction band effective mass is much less than the heavy hole mass.

Thus, for the well-designed P-I-N device, operating under reverse bias, the dark current is determined solely by the density and distribution of S-R centers in the bandgap of the active volume of the diode, and the success or failure of the HOT detector depends entirely upon the management of these bandgap states.

It is of interest to ascertain the magnitude of N_r required to enable background-limited performance at room temperature for a particular cutoff wavelength.

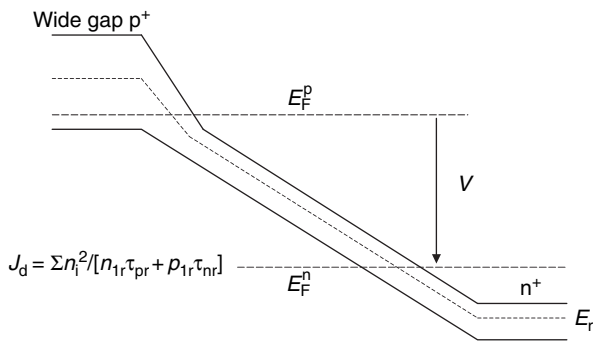

Figure 6.4 A reverse-biased P-I-N Diode.

For neutral S-R centers this requires that

$$G = W n_i \gamma N_r < \Phi_B,$$

where we have assumed unit quantum efficiency. Further assuming that $\gamma \approx 10^{-9}\,\mathrm{cm}^3/\mathrm{s}$, which is typical of neutral S-R centers, $W = 5 \times 10^{-4}\,\mathrm{cm}$, then, for LWIR HgCdTe at 295 K, $n_i \approx 5 \times 10^{16}\,\mathrm{cm}^{-3}$, $\Phi_B = 4 \times 10^{17}/\mathrm{cm}^2/\mathrm{s}$, and the requirement on N_r is $<1.6 \times 10^{13}\,\mathrm{cm}^{-3}$. This refers to the sum total of S-R centers in the material. The maximum allowed density of S-R centers for BLIP operation of a P-I-N diode at 295 K, as a function of cutoff wavelength, is shown in Fig. 6.5. Under these conditions, in the important MWIR/LWIR spectral regions, it is apparent that N_r needs to be $<10^{12}\,\mathrm{cm}^{-3}$, which represents a significant materials challenge, but with enormous potential payoff. For a recombination coefficient of $10^{-10}\,\mathrm{cm}^3/\mathrm{s}$ the requirement on N_r is relaxed by an order of magnitude. The presence of S-R centers away from the intrinsic energy level will modify the predictions of Fig. 6.5, but not too significantly, due to the magnitude of kT relative to the bandgaps involved. This requirement on N_r is at the limit of sensitivity for most impurity detection techniques such as glow discharge mass spectroscopy (GDMS), and secondary ion mass spectroscopy (SIMS).

Knowledge of S-R centers in HgCdTe is limited, but improving. As discussed in Sec. 4.4.1.2, metal vacancies and extrinsic p-type dopants have been characterized by Hall and minority carrier lifetime data over a limited range of $Hg_{1-x}Cd_xTe$ x-values, primarily in the range of $0.2 < x < 0.3$ (LWIR to MWIR).

Metal vacancies are characterized by a donor-like S-R center located ~30mV from the conduction band for compositions in the $0.2 < x < 0.3$ range. Lifetime measurements indicate an electron lifetime for the vacancy given by

$$\tau_{vac} = 5 \times 10^9 [(n + n_1)/p]/N_{vac}. \tag{6.11}$$

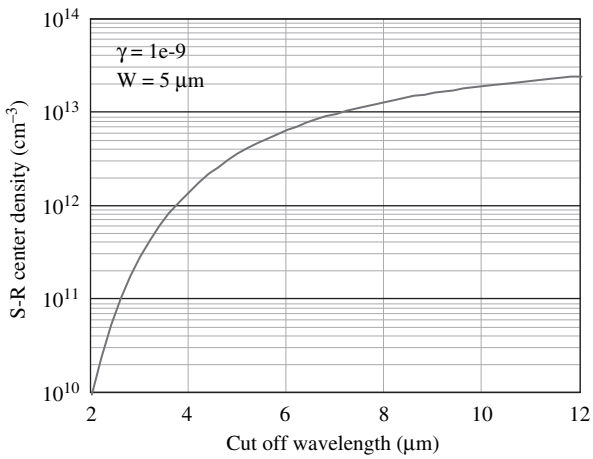

Figure 6.5 Maximum density of allowed S-R centers in a P-I-N diode for BLIP at 295 K.