Chapter 2
Physical and Technical Basics

First, in Sect. 2.1 the general basics of infrared (IR) thermography are briefly reviewed, which are also applicable to IR camera based lock-in thermography. In Sect. 2.2, the principles of the lock-in technique itself and of its digital realization are described. In Sect. 2.3, the two principal variants of lock-in thermography, which are serially measuring systems and camera-based systems, are introduced and compared. Different strategies to organize the timing of the lock-in correlation in relation to the frame rate of the IR camera are described in Sect. 2.4. Section 2.5 discusses the influence of non-harmonic (square wave) heat introduction, which is standard in electronic device testing, in contrast to the harmonic (sin-shaped) heat introduction mostly used in NDT. A detailed noise analysis is presented in Sect. 2.6, which relates the noise properties of the temperature measurement system to the noise level of the lock-in thermography result. Here, the "pixel related system noise density" is introduced as a universal parameter describing the figure of merit of different lock-in thermography systems. Section 2.7 deals with the problem of an easy and reliable calibration of lock-in thermography measurement systems by using a resistively heated test structure. In Sect. 2.8 the elementary processes in a solar cell with their heat generation and transport properties are described, which is the physical base for understanding lock-in thermography on solar cells. The detection of free carriers in semiconductors by IR lock-in thermography, which is a novel application of this technique based on non-thermal effects, is introduced in Sect. 2.9.

2.1 IR Thermography Basics

Section 3.1 will show that infrared (IR) thermography is by far not the only way to measure surface temperature distributions. Nevertheless, it is maybe the most elegant one, since it can be applied even to rough surfaces, and it can image the sample from a certain distance without contacting the surface at all. It even may investigate structures at the backside of a silicon wafer by looking through the wafer material, which is essentially transparent to the IR radiation. Most importantly, with the availability of modern focal plane array IR cameras combining high sensitivity with high frame rates, IR camera-based lock-in thermography systems have proven to provide today's best possible performances with respect to the detection sensitivity

within a limited measure time, which will be shown in Sect. 3.1. Thus, IR camera based lock-in thermography systems have already shown to be very useful to test electronic components. Therefore, in the following section the basic principles of IR thermography will briefly be reviewed, which also hold good for IR lock-in thermography.

If electromagnetic radiation (light) falls onto the surface of a specimen, three and only three things possibly may happen with the irradiated light: It may be reflected from the surface, it may be absorbed by the specimen, or it may be transmitted if the specimen is totally or partly transparent to the light. The probabilities of these three processes to happen are described by the reflection coefficient or reflectance ρ, the absorption coefficient or absorbance α, and the transmission coefficient or transmittance τ. These three coefficients are usually wavelength-dependent and may depend on the directional distribution of the irradiation. They are dimensionless, and their sum is always unity. For an ideally reflecting specimen, ρ is unity, and α and τ are zero, for a non-reflecting totally transparent specimen, τ is unity and α and ρ are zero, and for a black body, α is unity, and ρ and τ are zero. Each body at a finite temperature spontaneously emits electromagnetic radiation, which is called thermal radiation. The magnitude M_λ (in units of $\mathrm{W\,m^{-2}\mu m^{-1}}$) is called the spectral specific irradiation. It describes the electromagnetic power, which is irradiated within a differential wavelength range by a plane unit area into one half-space. The specific irradiation of a black body as a function of the wavelength λ is given by Planck's law:

$$M_\lambda(T) = \frac{2\pi \mathrm{hc}^2}{\lambda^5} \left(\mathrm{e}^{\frac{\mathrm{hc}}{\lambda k T}} - 1 \right)^{-1} \tag{2.1}$$

(h = Planck constant, c = velocity of light, k = Boltzmann constant, T = absolute temperature in Kelvin). Figure 2.1 shows the spectral distribution of the specific irradiation of a black body for two temperatures near room temperature. The noticeable room temperature radiation starts at 3 microns, the maximum appears around 10 microns, and the radiation spreads to above 30 microns. For a real (not black) specimen, the thermal emission also depends on optical properties of the specimen. Let us imagine a closed volume with homogeneous optical properties of the walls at a certain temperature in thermal equilibrium. Then the inner surface loses energy by thermal radiation, and it absorbs energy by radiation absorption. Thus at any wavelength, each part of the inner surface has to emit the same amount of radiation as it absorbs, otherwise the system would not be in thermal equilibrium. This means that the probability of a surface to emit radiation (the so-called emissivity ε) has to be equal to the absorption probability α at this wavelength. This identity is known as Kirchhoff's law. For an ideal black body, $\alpha = \varepsilon = 1$ holds, for real bodies ε is more or less smaller than 1. If within a certain wavelength range the emissivity ε is < 1 but wavelength-independent, this specimen is called a "grey emitter". If ε strongly depends on the wavelength, this specimen is called a "selective emitter". In order to obtain the specific irradiation of a real body, (2.1) has to be multiplied by the wavelength-dependent emissivity $\varepsilon(\lambda)$.

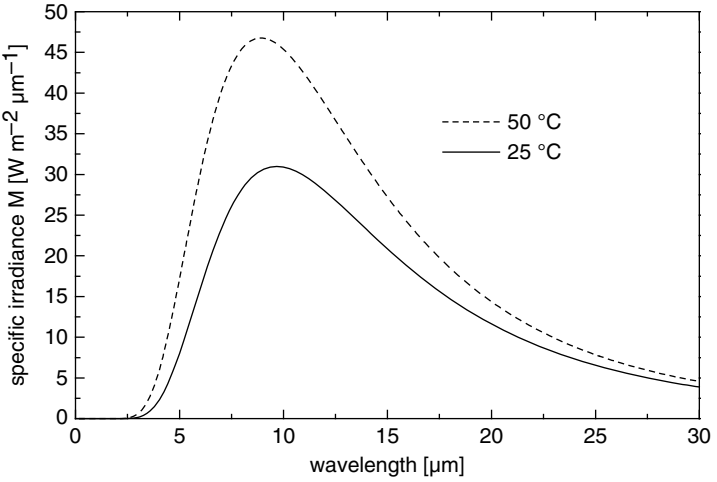

Fig. 2.1 Black body radiation for room temperature (25°C) and for 50°C

The dominant uncertainty in quantitative thermography is due to the fact that the IR emissivity ε of the investigated object is often unknown and may even be position-dependent. In order to determine the temperature of an object from the thermal radiation, the emissivity in the detection range has to be known or measured (see Sect. 5.3). Thus, though thermograms are usually scaled in °C, this scaling is based on the assumption of a certain emissivity, which does not need to be correct. Especially, reflecting objects like metallic surfaces usually show a low emissivity due to their high reflectivity, and the radiation coming from these objects contains a high amount of reflected light from the surrounding. This makes thermography on highly reflecting objects especially difficult. A proven way to overcome this problem is to cover the surface of these objects with a thin layer (e.g. a black paint), which has a high emissivity within the sensitivity range of the camera.

Thermography was invented for imaging temperature distributions from certain distances through the air. If there is any absorption of thermal radiation by the air, this causes errors in thermographic T-measurements. In order to keep these errors small, the wavelength range of a thermocamera has to be chosen so to avoid the dominant absorption regions. Figure 2.2 shows the transmission of a typical wet atmosphere as a function of the wavelength. The absorption is mostly due to water vapor and CO_2. In the interesting IR range between 3 and 30 µm there are two wavelength "windows" where the atmosphere is essentially transparent. The so-called "mid-range" window is between 3 and 5 µm, and the "long-range" window is above 8 µm. Thermocameras are usually supplied to work within one of these windows. The residual absorption within these windows may be regarded in the software-based scaling of modern IR cameras as a function of the distance, the temperature, and the humidity. Note, however, that for Fig. 2.2 a large distance of 6000 ft (1250 m) was chosen. If lock-in thermography has to be performed on single

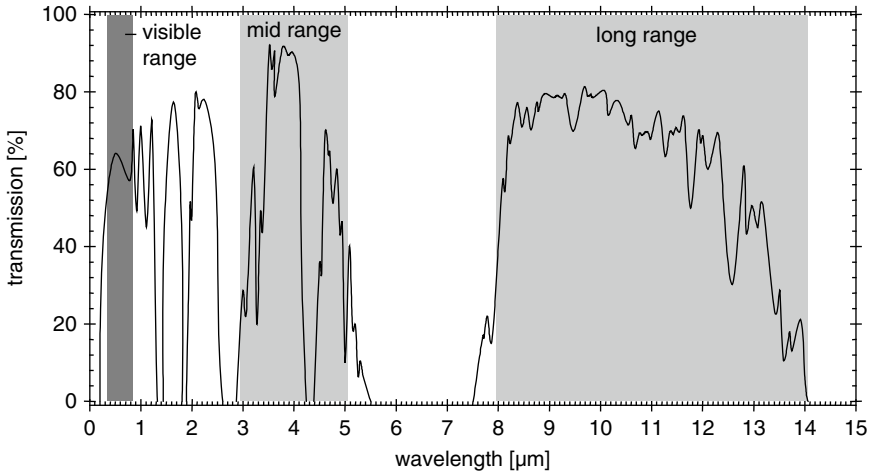

Fig. 2.2 Transmission of the atmosphere in the wavelength range between 0.2 and 14 μm. This measurement was done over a distance of 6000 ft (1250 m) at 20°C and 100% relative humidity [1]

electronic devices, the typical imaging distances are well below 1 m. Even in the range between 5 and 8 μm the transmission of wet air over a distance of 30 cm is still above 80%. Hence, for this special application the sensitivity range of IR cameras might also reach into this range, which could be advantageous in achieving a higher sensitivity. Unfortunately, today's highly sensitive IR cameras generally are still working in only one of these windows and do not use the spectral range between 5 and 8 μm. The spectral range of mid-range and long-range thermocameras is often artificially cut below 3 and 8 μm, respectively, using an optical filter.

The major inaccuracy of radiometric temperature measurements arises from the uncertainty in knowing ε and from the influence of reflected light. Fortunately, reflected light disturbs lock-in thermography only slightly, since this is an a.c. measurement. Though, in principle, also the reflection coefficient may be temperature-dependent, it can be assumed to be constant within the small temperature fluctuations appearing in lock-in thermography. Thus, reflected light contributes a constant additive component to the thermographic images, which is effectively cancelled by the lock-in process if it does not include frequency components near the lock-in frequency (see Sects. 2.2 and 2.3). Therefore lock-in thermography usually does not require any shading of the imaging scene, and persons may move around the measurement set-up, which may be disturbing in standard thermography. Only for low lock-in frequencies around or below 1 Hz, the light reflected from a laboratory surrounding contains frequency components near the lock-in frequency, which may affect also a lock-in thermography measurement.

Some remarks should follow as to the different types of thermocameras. The classical thermocameras, which are no more available, were actually serially working. Hence, they contained only one highly sensitive IR detector and a mechanical mirror scanner system, which successively leads the light pixel by pixel from the imaged

scene to the detector. The actual IR detector was a quantum detector (photodiode type), usually liquid nitrogen cooled and made from InSb (indium-antimonide) or HgCdTe (mercury-cadmium-telluride, MCT). InSb detectors are only sensitive in the mid-range spectrum (3–5 μm), whereas MCT detectors may be designed to work either in the mid-range or in the long-range one (8–10 μm).

All modern thermocameras are so-called "staring" or focal plane array (FPA) cameras [11], usually having quartz-stabilized frame rates. Just like the well-known silicon-based CCD cameras working in the visible range, these FPA cameras employ a 2-dimensional array of IR-detectors, which is positioned in the focal plane of the IR optics. However, this focal plane array has to be made from a material like InSb or MCT, which is sensitive in the infrared range. Each photodiode of this detector array has to be electrically connected with one readout channel of a separate silicon readout-chip attached to the detector chip, which is technologically very demanding. These detectors have to be cooled to about liquid nitrogen temperature (<80 K), which is done today mostly by using Stirling coolers. Also thermocameras based on Quantum Well Infrared Photodetectors (QWIP) on a GaAs wafer are available, which are not quantum detectors but rather photoconductors. Their sensitivity is comparable to that of InSb or MCT detectors, but their frame rate is lower. In the past also, detectors based on platinum-silicide Schottky diodes on a silicon wafer were used, but they also have disappeared from the market.

Today's most popular thermocameras are uncooled cameras, which are based on an array of thin film thermoresistors on free-lying membranes on an Si-chip. These so-called microbolometer cameras are sensitive in the long range. They are about a factor of 2–4 less sensitive than cooled quantum detector cameras and operate at a frame rate between 30 and 60 Hz, but they are much cheaper than any cooled detector camera. In cases where an ultimate detection sensitivity is not needed, they can also be used for lock-in thermography.

Another important aspect is the frame rate of the camera. As the noise analysis in Sect. 2.6 will show, a high frame rate is desirable for attaining a high detection sensitivity. In this respect, the long range MCT FPAs are optimum of all, since they reach their sensitivity of about 20–30 mK at a frame integration time of only about 200 μs. The frame integration time is the time where "the shutter of the camera is opened". Of course, modern FPA cameras do not have a mechanical shutter anymore but an electronic one, which governs the exposure time when photons are captured. Depending on the object temperature and on the properties of the objective, there is an optimum integration time for each measurement. The signal-to-noise ratio of the camera increases with the square root of the integration time. Hence, if the integration time is chosen too low, the camera noise increases. On the other hand, if the integration time is chosen too large, the storage capacitors of the readout circuit get saturated, hence the pixel readout values do not depend on the object temperature anymore. This has to be avoided, of course. In older FPA cameras the moment of exposure was dependent on the image position, hence different pixels were not only read-out sequentially but also exposed to the light sequentially. This operation mode was called "rolling frame mode". Modern FPA cameras are working mostly in the "snapshot mode", hence all pixels are exposed to the light at the same time. In

the past, only one capacitor per image pixel was used in the read-out circuit to store the collected charges of this pixel. Here the readout process started after the exposure period ("integrate-then-read", ITR-mode), hence the minimum frame period was the sum of the frame integration time and the readout time. Newer IR-cameras contain two storage capacitors per pixel and may charge-up one of them while the other is read-out ("integrate-while-read", IWR-mode). In these cameras the maximum possible frame rate may be close to the inverse of the frame integration time. Thus, long range cameras, needing an integration time of only 200 µs for achieving optimum signal-to-noise ratio at room temperature, could reach a frame rate of about 5 kHz. In practice, however, the readout circuit still limits the maximum full frame rate to values between 100 and 800 Hz. The readout speed is given in units of megapixel per second. Hence, the larger the array, the harder is it to reach high frame rates. However, if only some part of the detector is used (sub-frame modus), depending on the size of the sub-frame, the attainable frame rates may reach several kilohertz. It can be hoped that the speed of readout circuits will further increase in future. The InSb and MCT mid range FPAs need a frame integration time of about 1–2 ms to reach their best sensitivity at room temperature of 10–20 mK. Thus, they can be used up to a frame rate of about 500 Hz, which is within the possibilities of modern readout circuits. Uncooled IR cameras show a frame rate of only 30–60 Hz, hence also for this reason they are less appropriate to achieve highest detection sensitivity (see Sect. 2.6). Regarding the spectral distribution of the thermal radiation given in Fig. 2.1, one could assume that the sensitivity of cameras in the long-range should generally exceed that in the mid-range. However, the practically obtained signal-to-noise ratio depends not only on the flow rate of received photons but also on the noise properties, on the quantum yield, and on the sensitive area of one pixel (pitch size) of the detector elements as well as on the noise parameters of the readout electronics. Also, the relative change of the photon flux with temperature plays a role, which is larger in the mid-range than in the long-range. Altogether, mid-range cameras usually show an even somewhat higher nominal sensitivity at room temperature than long-range ones [11]. If microscopic investigations have to be made, like failure analysis on ICs, mid-range cameras have to be preferred, since they provide a better diffraction-limited spatial resolution, see Sect. 3.4. Thus, the choice of the optimum thermocamera for a certain application depends on many factors, which here could be discussed only briefly.

It should be mentioned that compared to the old scanner cameras the highly sensitive Focal Plane Array cameras may have special problems. Thus, as a rule, not all pixels of an array are operative. Some of them show strongly outrunning properties, some of them are totally insensitive, and some of them are "flickering". Usually, also the scatter of the light conversion parameters (gain and offset) of the nominally operative pixels is so large that the raw image from such a camera looks very poor. Note that monocrystals of the III-V compounds InSb and HgCdTe (MCT), which the most sensitive FPAs are made from, cannot be grown as perfectly as monocrystalline silicon crystals can be made. Fortunately, these problems can be solved using digital technology. All modern FPA cameras are digital ones, hence their originally analog video signal is digitized and further processed as a stream of data within the

camera via digital signal processors (DSPs). In the so-called two-point calibration procedure, both the gain and the offset of each pixel are software corrected on-line to yield the correct temperature values. The user can gain the data necessary for the two-point calibration procedure by imaging a homogeneous black body at two different temperatures in the interesting temperature range. Independently, he can easily make an offset correction procedure by imaging any homogeneous object at any temperature. Instead of using a black body, the simplest way to perform the calibration procedure is the following: First cover the objective with a homogeneous object at room temperature, like the plastic lens cap or a sheet of paper, for creating a homogeneous room temperature image and make the first image, which corresponds to room temperature (typically 25°C). Then cover the objective with the heel of your hand (without touching the lens!) and make the second image, which corresponds to the temperature of your hand (approximately 35°C). If not high accuracy measurements have to be performed, the accuracy of this procedure is sufficient. Note that the IR emissivity of the human skin is close to 1, and objects directly in front of the lens, lying far outside of the lens focus, appear very blurred. Hence, even if your hand shows a somewhat inhomogeneous temperature distribution, this will average out across the whole area. If even after this calibration procedure the image quality is not good enough, this is often an indication that the frame integration time was chosen too large, so that the storage capacitors were already saturated at the higher temperature. Note that the data for the image correction depend on both frame rate and frame integration time of the camera. The DSP of the camera can also replace missing pixels by the content of one or several neighboring pixels on-line, so that the final image looks defect-free. The data necessary for a bad-pixel correction are usually given by the manufacturer of the camera. More sophisticated digital cameras used for standard thermography also allow one to correct the image on-line for different emissivities and atmospheric absorption (depending on distance, humidity, and temperature) as well as to regard the non-linear dependence of the IR irradiation of the temperature according to Planck's law (2.1). They may even be equipped with a thermostatted internal black body to provide automatic temperature calibration. For lock-in thermography on electronic devices, however, only the emissivity correction might be interesting, since the absolute temperature reading does not influence the result, the temperature modulations hardly exceed some K, and the objects are usually imaged from a short distance. Details of the local emissivity correction procedure will be described in Sect. 5.3.

Let us finally mention the so-called "Narcissus effect", which is well known to every experienced thermographer. Narcissus was the man in Greek mythology who fell in love with his own mirror image. In thermography the Narcissus effect describes the black spot which is visible whenever a cooled thermocamera is directed to a flat reflective surface where it "sees itself". Since the interior of the camera is cooled and the objective is transparent to the detected IR radiation, the image of the objective of an IR camera in operation detects a cool object. The farer away the mirror is the smaller becomes the dark spot and vice versa. In a microscope objective where the object is very close, the dark spot usually occupies the whole image field. Then the influence of reflected light from the surrounding is

very small, which simplifies the emissivity correction of microscopic IR images (see Sect. 5.3). Since the Narcissus effect is a steady-state phenomenon, it only affects the topography image but not any lock-in thermography result. It can be avoided by slightly inclining a plane object from the normal orientation. Then the dark spot is lying outside of the image field.

2.2 The Lock-in Principle and its Digital Realization

The lock-in principle is the technique of choice, if signals have to be extracted from statistical noise. Prerequisite to using this technique is that the primary signal (before the detection and the first amplification stage!), can be periodically pulsed or anyhow else amplitude-modulated with a certain frequency called "lock-in frequency" $f_{\text{lock-in}}$. In some cases, this modulation is part of the experiment, e.g. in cyclical mechanical loading experiments. If any kind of light causes the primary signal as, e.g., in light absorption experiments, mechanical chopping of the light beam or simply switching the light on and off are convenient means to produce an amplitude-modulated signal. The most elegant method to produce an amplitude-modulated signal is given if the signal generation itself can be controlled electronically. This usually occurs in the functional diagnostics of electronic components if the generated heat is our primary signal. In the simplest case, the heat generation can be modulated by simply pulsing the supply voltage of the tested component. Other more sophisticated triggering modes will be discussed at the end of Sect. 2.3.

The aim of the lock-in principle is to evaluate only the oscillating "alternating current" (a.c.) part of the detected signal. The classical way of converting an a.c. signal into a "direct current" (d.c.) one is to rectify it via some diodes. In a full-wave rectifier, the negative part of a signal is converted into a positive one and then passed by some diodes. The positive part is directly passed by other diodes. Hence, the momentary sign of the a.c. signal itself controls whether the signal is passed by directly, or inverted. If this procedure is performed with a noisy signal, also the noise is rectified, therefore the noise share contributes to the output signal. In order to suppress the noise, the signal before rectification can be fed through an appropriate small-band amplifier, with its center frequency matched to the signal frequency. This measure already improves the signal-to-noise ratio at the output. Nevertheless, also the small-band noise at the modulation frequency makes a positive contribution after conventional rectification, if its intensity is strong enough.

In the lock-in process, the condition whether the signal is passed by directly or inverted, is not controlled by the sign of the a.c. signal itself but rather by a noise-free reference a.c. signal, which is derived from the signal generation process. Figure 2.3 illustrates this procedure. Here, the signal generation process is symbolized by a black box delivering the (noisy) signal of interest and a (noise-free) reference signal. It is irrelevant whether a free-running oscillator controls both the primary signal and the reference signal, or whether the periodic signal generation process is free-running and the reference signal comes from the signal generation process. The dashed circles symbolize different typical signal shapes in different

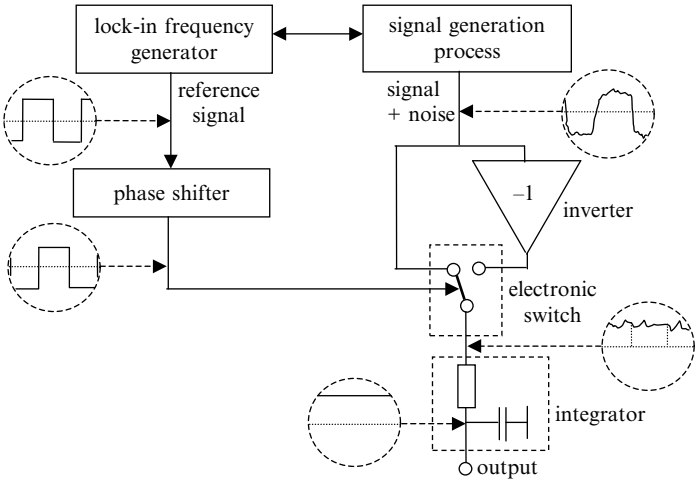

Fig. 2.3 Schematic of an analog wide-band lock-in correlation procedure

positions of the circuit. The actual rectification is performed here via an electronically controlled switch, which connects alternatively the inverted or not inverted signals to the output. Thus, if the reference signal is symmetric (positive period equals the negative one), any d.c. component of the signal is perfectly suppressed by the lock-in process. If the phase of the detected (beneficial) signal coincides with the reference phase, the lock-in procedure behaves to the beneficial signal like an ordinary signal rectification process. Hence, the beneficial a.c. signal is correctly rectified by the lock-in process, leading to a positive d.c. signal at the output. If the phase of the beneficial signal does not coincide with that of the reference signal, both phase positions have to be matched by inserting a phase shifter in the reference path to get the correct signal retrieval. This phase shifter provides a controlled time delay of the reference signal.

The crucial point of the lock-in process is that statistical noise remains statistical even after the rectification process, since it is not correlated with the reference signal. If there are noise components near the reference frequency, after the lock-in procedure they produce another noise signal which, however, has the average value of zero. Therefore, at the final integration stage, which averages the output signal over a certain integration time, this noise averages out and only the beneficial signal passes as a d.c. signal. The longer the integration time, the more efficient is the noise suppression effect, but the longer is the response time for the beneficial signal. In this way it is possible to separate periodic signals, which are embedded in noise being orders of magnitude above the signal level, from the noise. Effectively, the lock-in procedure presents a frequency conversion of the noisy measured signal with the noise-free reference signal towards d.c., where the integration time is inversely proportional to the output bandwidth and thus to the effective detection bandwidth. If the integration stage is a simple RC element, which shows an exponential response characteristic, one speaks of exponential averaging. Averaging the

output over a well-defined integration time t_{int} is called linear averaging. Linear averaging, which can be performed most easily in digital lock-in correlation (see below), has the advantage that there is no cross talk between successive measurements. Choosing the integration time always represents a compromise between the degree of noise suppression and the bandwidth of the output signal.

Mathematically, the lock-in signal treatment (usually called lock-in correlation procedure) can be described as a multiplication of the detected signal $F(t)$ by a correlation function $K(t)$, which is a symmetric square wave function for the case shown in Fig. 2.3. Here, for linear averaging over a certain integration time t_{int}, the output signal S can be written as

$$S = \frac{1}{t_{\mathrm{int}}} \int_0^{t_{\mathrm{int}}} F(t)K(t)\,dt \quad \text{with } K(t) = \begin{cases} +1 \text{ (first half period)} \\ -1 \text{ (second half period)} \end{cases} \tag{2.2}$$

If $K(t)$ is symmetric and if a complete number of periods falls into the integration time, the averaged value of $K(t)$ itself is exactly zero, hence (2.2) provides a perfect suppression of any d.c. part of the signal $F(t)$.

Digital lock-in correlation uses an Analog-Digital-Converter (ADC) to digitize the input signal $F(t)$ leading to a set of numbers F_k. Then the whole lock-in correlation procedure is performed numerically. Hence, also the correlation function $K(t)$ has to be replaced by a set of numbers K_k, and (2.2) has to be replaced by a sum. These numbers are also called "weighting factors". They may be positive and negative, and for obtaining an efficient d.c.-suppression their total sum has to be zero. In most lock-in thermography applications the weighting factors follow not a square function as shown in Fig. 2.3 but rather a harmonic function (see below). The digital lock-in correlation procedure consists in averaging the product of the measured values F_k and a set of weighting factors K_k up to the total number of measured values M:

$$S = \frac{1}{M} \sum_{k=1}^{M} F_k K_k \tag{2.3}$$

There are two strategies of performing a digital lock-in correlation: on-line correlation (evaluating the data during the measurement) and off-line correlation (all data are stored and correlated after the measurement). The weighted sum in (2.3) can be summed up during the measurement, but this "standard lock-in correlation" is not the only possible digital correlation technique. Krapez [29] has reviewed different digital signal correlation techniques. Of these the Least Squares Method and also the Fast Fourier Method used by Kaminski et al. [30] have to be performed off-line, since for them all data have to be available prior to the calculation. It is without question that for long integration times and high sampling rates, the storing of all data needs a lot of storage capacity, and the off-line evaluation may become a serious time factor of the whole measurement procedure. This is especially true for lock-in thermography, where whole images have to be stored. Since within the scope of this book we are most interested in highly sensitive measurements requiring

a large amount of averaging, in the following we will consider only standard lock-in correlation, which may be performed on-line.

Note that the (usually equidistant) digitizing or sampling events and the periodic lock-in correlation do not necessarily have to be synchronized. In some cases, a synchronization of the digitization with the experiment is complicated or even impossible to manage. This may hold, e.g., for thermo-elastic investigations, where the sample is treated by a free-running machine independently of the digitizing events. In the following, we will speak of "asynchronous correlation" whenever the lock-in frequency and the sampling rate are not synchronized to each other. It is even possible that the sampling rate is lower than the lock-in frequency. This so-called "undersampling" technique as well as different strategies to organize the timing of lock-in thermography systems will be discussed in detail in Sect. 2.4. Krapez [29] has evaluated some problems connected with a non-synchronized lock-in frequency. For example, in the case of undersampling, there are some "forbidden" lock-in frequencies, which would interfere with the sampling frequency. Under non-synchroneous conditions, due to the usually incomplete number of lock-in periods in one measurement, the d.c. rejection of the lock-in process may not be perfect. Moreover, any frequency drift and phase jitter of both experiment and camera may cause additional noise. Generally, synchronizing the digitizing rate and the lock-in frequency avoids these problems. Synchroneous correlation should therefore be preferred, if technically possible. For the functional diagnostics of electronic components this should always be possible, as electronic components can be triggered via an external signal deduced from the frame trigger of the camera. The main results of Krapez [29] were that more sophisticated correlation methods are advantageous over the standard lock-in method only for a small number of lock-in periods, and if the noise level is below the resolution of the ADC.

In this Section, we always assume synchronization of the lock-in frequency and the digitizing events. Hence, we assume a fixed number of digitizing events n (samples) per lock-in period (which has to be $n \geq 4$, see below), and the weighting factors K_k in (2.3) are the same in each lock-in period. If the measurement is averaged over N lock-in periods, the digital lock-in correlation for a synchronous correlation is given by the sum:

$$S = \frac{1}{nN} \sum_{i=1}^{N} \sum_{j=1}^{n} K_j F_{i,j} \tag{2.4}$$

For function-theoretical reasons, the correlation function optimum to achieve the best signal-to-noise ratio would be the shape of the expected signal. An especially advantageous correlation function, however, is the harmonic (sine or cosine) function. This kind of lock-in correlation is called sin/cos or narrow-band correlation, whereas the square-wave correlation shown in Fig. 2.3 is called wide-band correlation. Typical of the narrow-band correlation is that it evaluates only to the basic harmonic of the signal, which usually carries the dominant information, whereas higher harmonics are suppressed. Electronically, the narrow-band correlation can be

realized either by narrowing the bandwidth of the detected signal prior to the correlation, or by using a sine signal as the reference and an analog multiplier instead of the switches in Fig. 2.3. For the digital lock-in correlation, narrow-band correlation can be realized also either by narrowing the bandwidth of the detected signal or by using the values of a harmonic function for K_j in (2.4). The decisive advantage of the sin/cos correlation is that it allows one to exactly consider the phase of the signal after the measurement (off-line), if the two-channel correlation is used. Two-channel correlation means that there are two sets of weighting factors K_j, one approximating the sine function and the other one approximating the cosine one. The correlation is performed twice in parallel with these two sets of weighting factors. Then the first channel measures the component of the signal in-phase with the sin-function, and the other channel measures the component in-phase with the cos-function, which is 90° phase-shifted to the sin-function. Let us assume that the amplitude of the detected signal is A and its phase (referring to the sin-function) is Φ. Using the addition theorem for the sin-function we get:

$$\begin{aligned} F(t) &= A\sin(2\pi f_{\text{lock-in}}\, t + \Phi) \\ &= A\sin(2\pi f_{\text{lock-in}}\, t)\cos\Phi + A\cos(2\pi f_{\text{lock-in}} t)\sin\Phi \end{aligned} \tag{2.5}$$

If this is inserted into (2.2) with $K^{0^\circ}(t) = 2\sin(2\pi f_{\text{lock-in}} t)$ and $K^{90^\circ}(t) = 2\cos(2\pi f_{\text{lock-in}} t)$ (the factor 2 has to be added to get the correct amplitude), the result of the two correlations over a complete number of periods is:

$$S^{0^\circ} = A\cos(\Phi) \qquad S^{90^\circ} = A\sin(\Phi) \tag{2.6}$$

Exactly the same result is obtained for any even $n \geq 4$, if instead of the integral representation (2.2) the sum representation (2.4) is used for the correlation with K_j being the (doubled) values of the sin and the cos-functions.

$$K_j^{0^\circ} = 2\sin\left(\frac{2\pi(j-1)}{n}\right) \qquad K_j^{90^\circ} = 2\cos\left(\frac{2\pi(j-1)}{n}\right) \tag{2.7}$$

Since the summation in (2.4) goes from 1 to n, j has been lowered by 1 in (2.7) to ensure that the first weighting factor corresponds exactly to a phase position of 0°. S^{0° and S^{90° are usually called the in-phase signal and the quadrature signal. Both signals may be either positive or negative. A negative 0°-signal points to the presence of a signal component 180° phase shifted to the reference, and a negative 90°-signal indicates a 270° signal component. The phase-independent amplitude A (which is always positive) and the signal phase Φ can easily be retrieved from the two results S^{0° and S^{90°:

$$\begin{aligned} A &= \sqrt{(S^{0^\circ})^2 + (S^{90^\circ})^2} \\ \Phi &= \arctan\left(\frac{S^{90^\circ}}{S^{0^\circ}}\right) \left(-180^\circ \text{ if } S^{0^\circ} \text{ is negative}\right) \end{aligned} \tag{2.8}$$

The second term in formula (2.8) for Φ means that an amount of 180° has to be subtracted only if $S^{0°}$ is negative. This measure makes the procedure "quadrant correct", since the arctan-function repeats twice within a whole period of 360°. Knowing A and Φ allows one not only to retrieve the in-phase and the quadrature component using (2.6) but also to construct the component of the signal modulation in any phase position Φ':

$$S^{\Phi'} = A\cos(\Phi' - \Phi) \tag{2.9}$$

This situation is illustrated in the complex vector representation of the phase relations occurring in the lock-in process shown in Fig. 2.4. Here, the phase of the reference signal (0°) is the X-axis representing the real part *Re*.

Note that the phase of the cos-function is in advance to that of the sin-function by 90°. In lock-in thermography, however, the surface temperature signals are at best in phase with the periodic power modulation, usually they are more or less delayed. This means that the 90°-signal $S^{90°}$ will be essentially negative. In order to have both the in-phase and the quadrature signal essentially positive, the $-$cos-function should be used instead of the $+$cos one in the quadrature channel.

$$K_j^{0°} = 2\sin\left(\frac{2\pi(j-1)}{n}\right) \quad K_j^{-90°} = -2\cos\left(\frac{2\pi(j-1)}{n}\right) \tag{2.10}$$

Then, the amplitude and the phase are:

$$A = \sqrt{(S^{0°})^2 + (S^{-90°})^2} \tag{2.11}$$

$$\Phi = \arctan\left(\frac{-S^{-90°}}{S^{0°}}\right)\left(-180° \text{ if } S^{0°} \text{ is negative}\right)$$

Throughout this book we will use this definition for the quadrature signal. It will be shown in Sect. 4.5.2 that in some cases, the $-45°$ phase component of the signal

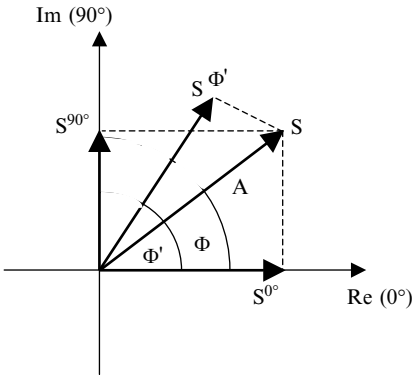

Fig. 2.4 Complex vector representation of the phase relations in the two-channel lock-in process and in retrieving an arbitrary phase component $S^{\Phi'}$

is needed. Applying the addition theorem to (2.9) using $\Phi' = -45°$ and regarding (2.6) leads to the following expression for the $-45°$ phase component:

$$S^{-45°} = \frac{1}{\sqrt{2}} \left(S^{0°} + S^{-90°} \right) \tag{2.12}$$

For a fixed sampling rate f_s the lock-in frequency is given by the number of samples per lock-in period n:

$$f_{\text{lock-in}} = \frac{f_s}{n} \tag{2.13}$$

Synchronization of lock-in and sampling frequency means that the number of samples per lock-in period n is an integer. According to the sampling theorem we need at least two samples per period, and we have to do this twice each period in order to get the in-phase and the quadrature signal. Thus, the minimum possible number of samples per period n is 4, if the two-phase lock-in correlation should be applied. For a fixed maximum sampling rate f_s, this sets an upper limit to the lock-in frequency:

$$f_{\text{lock-in}} \leq \frac{f_s}{4} \tag{2.14}$$

If only 4 signal values are measured within each lock-in period, we will speak of "4-point" correlation. In this case, according to (2.10) the values of the weighting factors K_j are the doubled values of both the sin-function and the $-$cos-function of 0°, 90°, 180°, and 270°, respectively. Hence, for the 4-point correlation the two correlation vectors are $K_j^{0°} = (0, 2, 0, -2)$ and $K_j^{-90°} = (-2, 0, 2, 0)$. For a larger number of samples per period the correlation vectors have to be calculated according to (2.10) and the lock-in frequency becomes (2.13).

The higher the number of samples per lock-in period, the more exactly does the sum (2.4) approximate the integral (2.2). Of course, for only 4 samples per period (4-point correlation, maximum possible lock-in frequency) the sine- and cosine-functions are only poorly approximated. Hence, for this case also eqs. (2.6–2.11), which are based on continuous signals, only approximately hold. Thus, in order to preserve the small-band lock-in correlation, the heat signal should actually be harmonic in this case. In serially measuring systems (see Sect. 2.3) this can be managed easily by introducing a band pass filter in the signal path before the ADC. In camera-based systems, however, it is not possible to introduce a filter. Here, the small band correlation may be guaranteed by ensuring that the heat introduction itself is harmonic. This is the reason why harmonic heat introduction is commonly used in NDT [2]. Note that in the early days of NDT in the mid-1990s the computation power was still limited. At that time, for reducing the numeric expense of the correlation procedure, for low lock-in frequencies the digitizing results were averaged over four equidistant periods within each lock-in period. These four averaged values were correlated according to the 4-point correlation procedure described above. In this so-called "4-bucket method" [29] the correlation function also contains higher harmonics, leading to the demand for harmonic heat introduction in NDT. However, for testing electronic devices, the harmonic introduction of heat is seldom

manageable, since in most cases a component has to be tested under well-defined biasing conditions. Hence, usually a certain supply voltage simply has to be switched on and off, leading to a square-wave function of the heat introduction. Possible errors caused by this non-harmonic heat introduction are estimated and discussed in Sect. 2.5. It has been tried by several authors to perform lock-in thermography by applying only a weak modulation to an essentially constant sample bias. In this case the lock-in thermography signal is proportional to the deviation of the locally dissipated power to the bias. However, it has been found that this measure only degrades the signal-to-noise ratio without providing significant advantages. Only in the case of IC failure analysis it may be necessary, for preserving a certain logical condition of the device, to modulate the bias between two values, see Sect. 6.1.

The procedure (2.4) using the weighting factors (2.10) is mathematically called a discrete Fourier transformation [31, 32]. It calculates the complex Fourier component at one single frequency $f_{\text{lock-in}} = f_s/n$ of the measured function $F_{i,j}$. Alternatively, for all measured data F_m also the so-called fast Fourier transform (FFT) procedure can be applied. In simplest form this procedure requires a total number of measurements of $M = 2^i$, i being an integer, but also variants working with arbitrary data format are available [31]. It calculates all Fourier components of this data vector, which belong to the frequencies $f_1 = f_s/M$; $f_2 = 2f_s/M$, etc. up to the Nyquist frequency $f_s/2$ (see Sect. 2.4). So FFT provides a complete harmonic analysis of a signal. The result of this analysis is a vector of complex numbers of length $M/2$ plus the average value. If the data values are real numbers, the imaginary part of the last element belonging to the frequency $f_s/2$ is always zero. This is due to the fact that, with only two samples per period, no two-phase correlation can be performed, see (2.14). Thus, the total number of meaningful values (including the mean value) is M, as can be expected from a mathematical transformation. Of course, this analysis is mathematically more demanding than the discrete Fourier transformation, but the special FFT algorithm is much less demanding than performing all discrete transformations separately. For $M = 2^i$ values per data vector the expense increases only proportional to $M * i$ and not to M^2, as for the discrete Fourier transformation. Hence, the larger the data files the more advantageous is the FFT method. If the measured signal contains only white noise, all Fourier components, which may be positive or negative, should appear with the same average magnitude. However, if it contains harmonic signal components, at the frequencies of these signals peaks appear which represent the magnitudes of the corresponding harmonic components. Just as for the discrete Fourier transformation result in (2.11), the two complex components can be combined to an amplitude, which is always positive, and a phase signal of the corresponding frequency. If the lock-in frequency does not exactly match one of the Fourier frequencies f_m, the amplitude values of both neighboring Fourier frequencies have to be evaluated. In this case, the procedure is equivalent to the asynchronous discrete Fourier transformation (asynchroneous correlation, see Sect. 2.4). Also the FFT procedure is very popular to extract harmonic signals from statistical noise. If it is used in a measurement setup, this procedure has the advantage to the discrete Fourier transform that it does not need any synchronization between the data sampling and the modulation of the signal. Hence, for a

series of measurements the harmonic analysis can be made by FFT after the measurement. The basic limitation of this procedure for lock-in thermography is that it cannot be evaluated on-line (parallel to the measurement, see Sect. 2.3) since all data must be present at once to be evaluated. Thus, in a lock-in thermography measurement all images would have to be stored, which is impractical for long lasting measurements. Another limitation is that the phase signal, which is very important in electronic device testing, is only meaningful if the excitation pulses are synchronized to the measurement events. This condition cancels the above-mentioned main advantage of no needed synchronization procedure. For these reasons, though FFT-based signal correlation is widely used in non-destructive testing, it is not further considered in this book concentrating on electronic device testing.

Another possibility which is not considered here is lock-in detection of higher harmonics. It will be demonstrated in Sect. 2.5 that, even if square-pulse excitation is applied, the basic information of lock-in thermography is lying in the basic harmonic. Higher harmonic detection with harmonic excitation is often used for detecting a non-linear response to an excitation action. In non-destructive testing, for example, second harmonic detection has been used for detecting plastic deformation by vibrothermography [7,33]. Second harmonic detection has been used also by Grauby and Forget [34] for distinguishing resistive from Peltier effects in microelectronic devices under harmonic load. However, we will show in Sect. 5.4 that these two components can also easily be distinguished with asymmetric square-pulse excitation and basic harmonic detection. Also Altes et al. [35] have used second harmonic detection under harmonic load in resistive probe-based scanning thermal microscopy for measuring temperature modulation, and third harmonic detection for the determination of thermal conductivity. Also here, if asymmetric square-pulsed excitation would have been used, basic harmonic detection would have been optimum for temperature modulation measurement, and IR-based lock-in thermography is anyway not feasible for performing thermal conductivity measurements. Therefore, higher harmonic detection will not be considered anymore in the following.

2.3 Lock-in Thermography

All considerations made up to now apply to any lock-in detection system. Lock-in thermography means that the heat in an extended sample is generated periodically and the lock-in correlation process described in the previous section is applied to the temperature signal of each pixel of an image of the surface of the sample under investigation. So lock-in thermography is a kind of active thermography since the sample is actively influenced during the observation [2]. Since the phase of the detected surface temperature modulation may depend on the lateral position, as a rule lock-in thermography uses the two-channel (in-phase and quadrature) lock-in correlation. Following to the previous considerations, the result of a lock-in thermography investigation may be displayed in different ways. The results of the in-phase correlation of all image positions yield the in-phase image (or 0° image), and that of the quadrature correlation yield the quadrature image (or −90° image).

Alternatively, these two images can be used to calculate the phase-independent amplitude image and the phase one if (2.11) is applied to all image positions. With these two images, the image of any phase component of the temperature modulation can be calculated, using (2.9) for all image positions, or systematic phase errors of the detection system can be corrected. Moreover, it is always useful to have a topography image in order to enable some orientation on the surface of the sample. In Chap. 5 (Measurement Strategies) we will discuss and demonstrate which of the different possible images is most appropriate to visualize different signal features.

In principle, performing lock-in thermography would require the parallel operation of one two-phase lock-in detection system for each pixel of an image. Of course, this can no longer be managed physically using analog electronics as shown in Fig. 2.3. This procedure, however, may be converted into a serial procedure. There are two essentially different approaches to perform lock-in thermography: 1. serially probing systems and 2. camera-based systems. In serially probing systems, a T-detector is successively scanned across the sample surface, and in each position a number of lock-in periods are measured and evaluated by a computer. In this procedure only one lock-in correlator is used, hence here either digital correlation or even classical analog correlation can be used. Figure 2.5 shows the general scheme of a serially measuring lock-in thermography system for investigating electronic components. The box called "signal conditioning" represents any kind of converting the temperature signal into a voltage including optional signal filtering. It is obvious that the time efficiency of serial systems is generally very poor, since always only one of many pixels can be probed at any time. For the classical digital lock-in correlation the maximum possible lock-in frequency is given by (2.14). Since the measurement time in each image position has to be at least one lock-in period (undersampling does not imply any advantage for serially probing systems), the minimum possible acquisition time for an image of $X \times Y$ pixel is $t_{\mathrm{acq}}^{\mathrm{min}} = XY/f_{\text{lock-in}}$. As the following chapters will show, from a technical point of view, in many cases the

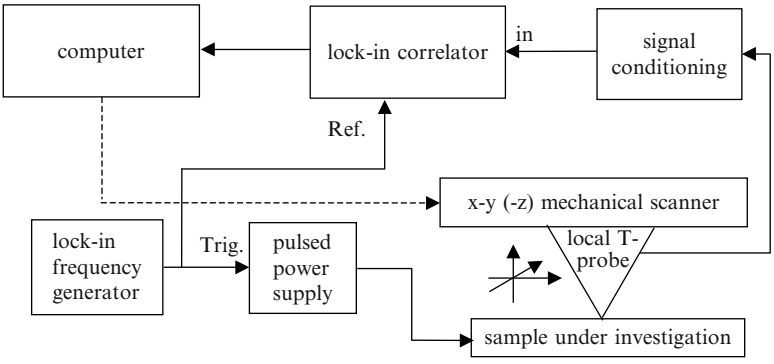

Fig. 2.5 Schematic of a serially probing lock-in thermography system for investigating electronic components. The temperature probing may be either contacting or non-contacting (see Sect. 3.1). Either the sample or the T-probe may be mechanically scanned. The frequency generator and the lock-in correlator may or may not be part of the computer system

highest possible lock-in frequency cannot be applied in practice because of sensitivity limitations, hence a lower lock-in frequency has to be used. For the same reason, it is mostly necessary to average over a number of lock-in periods. Thus, since at least 100×100 pixels are needed to form a meaningful image, serially measuring lock-in thermography systems often require measurement times of several hours to tens of hours (see Chap. 3). The major advantage of serially probing systems is that they are simpler to realize and therefore cheaper than camera-based systems. Moreover, in some cases (e.g. AFM probing) they provide a better spatial resolution than camera-based system do.

In camera-based systems, a camera is running with a certain frame rate f_{fr}, converting the whole 2-dimensional surface temperature distribution into an image containing the temperature information. This frame rate corresponds to the sampling rate f_{s} discussed in the previous section, except that the result of each camera measurement is a 2-dimensional array of T-data for the whole image at once. Thus, the lock-in correlation according to (2.4) has to be performed with all image data, which can be managed only digitally using a frame grabber and sufficiently high computation power.

As any lock-in measurement, lock-in thermography may be performed either with single phase correlation or with two-phase correlation. Single phase correlation means that the procedure (2.3) is performed only once with the phase of the correlation function selected to match the phase of the detected signal. As it is shown in Chap. 4, this is not optimum for lock-in thermography since, depending on the shape of the heat source and on the distance from the heat source, the phase position of the surface temperature modulation varies. If one is only interested to see point-like heat sources, it may be advantageous to display only the in-phase (0°) image, since this provides the best possible spatial resolution (see Sects. 4.3 and 5.1). However, performing only single phase correlation does not allow one to display the phase image or the 0°/−90° image (see Sects. 4.5.2 and 5.1), which are both inherently emissivity-corrected.

It should be mentioned that, if single phase correlation (e.g. 0°) is used, the image shows a 1.4 times better signal-to-noise ratio than the amplitude signal of a two-phase correlation system, since the latter signal contains noise components of both channels. However, also a two-phase system allows to display each single phase component separately, which is equivalent to single phase correlation. Therefore, there is no inherent advantage of single phase systems. In the following we will only consider only two-phase correlation as it was described in Sect. 2.2.

As Fig. 2.6 shows, the 2-channel image correlation procedure consists of multiplying all incoming image information by the two sets of weighting factors K_{j} and adding the results to two frame storages, which had been zeroed before the measurement. Then after the measurement one image storage contains the in-phase (0°), and the other one the quadrature (−90°) image. Of course, also here the data are transferred and treated serially one after the other at a certain pixel rate of the frame grabber of the computer. The basic difference between serially probing lock-in thermography systems and those camera-based is that in serially probing systems the sample is scanned slowly only once during one measurement. In each position,

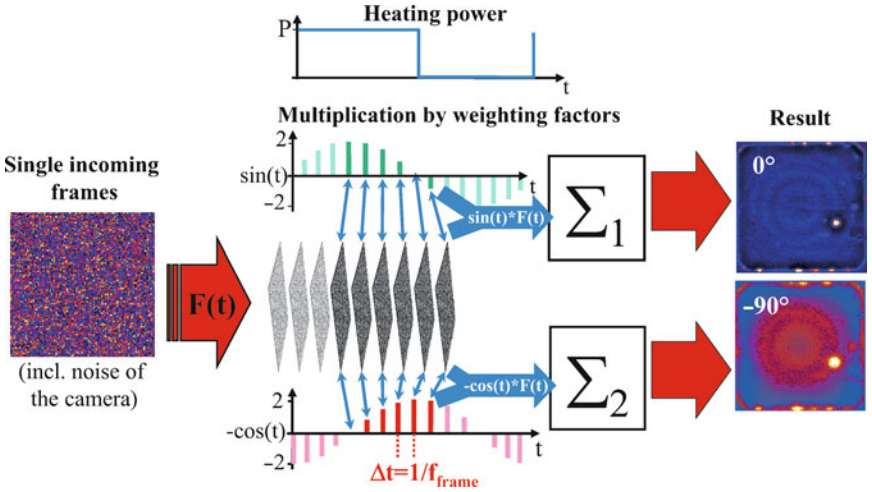

Fig. 2.6 Principle of the sin/−cos lock-in correlation procedure for camera-based systems, here shown for $n = 16$ frames per lock-in period. The images shown result from a real solar cell investigation

many T-measurements for one or several lock-in periods are performed. In camera-based systems, on the other hand, the image scan is performed "quickly" by the camera many times during one measurement, and each data point of each image belongs to one temperature value for a certain position at a certain time. Without using undersampling (see Sect. 2.4), the maximum possible lock-in frequency of camera-based systems is $f_{\mathrm{fr}}/4$ (see (2.14)). For lower lock-in frequencies, a larger number of frames has to be used in one lock-in period. Hence, in principle, a complete camera-based two-phase lock-in measurement could be performed using only 4 frames within one lock-in period requiring some 10 milliseconds, which is orders of magnitude faster than any serial system can do. Indeed, it had been shown that, if the detected signal is strong enough, a lock-in thermography image can be obtained within 10 ms [36]. As a rule, however, camera-based systems have to average over many lock-in periods just in order to improve their detection sensitivity. While serial systems may be both contacting and non-contacting, camera-based systems are, of course, always non-contacting. The noise properties of different serial and camera-based lock-in thermography systems will be compared in Sect. 3.1.

Here, a special remark should follow regarding the triggering mode of the sample under investigation. As mentioned above the simplest way of generating an amplitude modulated (a.c.) heat signal in electronic components is to pulse the supply voltage of the component. This measure reveals all heat sources present in this component. However, more complicated components like integrated circuits (ICs) may have a lot of other control inputs, which cause some complex electronic actions within these components. If these control inputs are used for sophisticated triggering, special modes of device testing of ICs become possible. In the simplest case, the supply voltage is permanently applied and the reference trigger is applied to a

certain control input of an IC. Then some current paths (e.g. gates) may become active if the trigger is in "high" position, and passive, if it is in "low" position. Some other gates may behave inversely. If the image in an appropriately selected phase position is displayed, the first class of gates will appear bright (positive contrast), and the second class will appear dark (negative contrast) with respect to the surrounding. Thus, the phase of the signal indicates whether a heat source is switched on or off by the trigger. The signal value of all regions not affected by this trigger will be zero, irrespectively whether continuous heat is generated there, or not. In this way, the action of different control signals on the operation of the circuit can be independently checked. One example of Sect. 6.1 demonstrates an application of this type of sophisticated triggering. Even complex control signals can be applied to one or several inputs, if the circuit under test is connected with an intelligent circuit tester. Whenever these controls are leading to internal circuit actions, which are phase-related to the detecting lock-in reference, these actions and only these can be visualized by lock-in thermography. This sophisticated triggering mode and phase detection greatly enhance the analytical possibilities of lock-in thermography in device testing of electronic components compared to the general imaging of all possible heat sources via simply pulsing the supply voltage.

2.4 Timing Strategies

Beginning with (2.4), only conventional synchronous lock-in correlation had been considered in Sect. 2.2. Hence it had been assumed that the lock-in trigger is derived from the frame trigger of the camera by using a frequency divider. Then each lock-in period contains a number of measurement events (frames, samples) of $n \geq 4$. This condition defines an upper limit of possible lock-in frequencies as $f_{\text{lock-in}} \leq f_{\text{frame}}/4$. It has been suspected that n must be an even number, but it is also allowed to be odd. If the frame rate of the camera is fixed, at least close to the limit $f_{\text{lock-in}} = f_{\text{frame}}/4$ the lock-in frequency can only be varied in certain steps ($f_{\text{frame}}/4$, $f_{\text{frame}}/5$, $f_{\text{frame}}/6$, etc.). Working exactly at a well-defined lock-in frequency is seldom necessary, but the condition $f_{\text{lock-in}} \leq f_{\text{frame}}/4$ often represents a serious experimental limitation. It will be shown in Chap. 4 that the spatial resolution of lock-in thermography improves with increasing lock-in frequency. Today special high-speed IR cameras are available, and for a given camera the frame rate can be further increased by working in sub-frame mode, hence by reducing the used image field. Nevertheless, especially for microscopic investigations, lock-in frequencies in the order of or even higher than the frame rate of the camera are often desirable. It will be shown below that this can be realized by using the so-called undersampling technique. For certain investigations, e.g. for thermomechanical stress investigations, where a sample is deformed by a free-running machine, it is generally not easy to synchronize the lock-in frequency to the lock-in correlation. Another possible case where it is hard to synchronize the lock-in periods to a free-running camera is sophisticated triggering in circuit tester-based IC failure analysis, which had been mentioned at the

end of Sect. 2.3. Here the circuit tester may easily deliver a trigger signal, which is synchronous to the heat dissipation due to certain local circuit activities, but it may be harder to trigger the tester externally from the lock-in thermography system. In this case FFT-based correlation may be used as described at the end of Sect. 2.2. However, as also described there, this kind of correlation has the disadvantage that it cannot be performed on-line. Hence, it is only useful for evaluating a limited number of lock-in periods. In the following, different strategies for organizing the lock-in timing in relation to the timing of the IR camera will be described, leading to different solutions for the lock-in correlation procedure also.

Again, the following considerations hold for any kind of digital on-line lock-in correlation. Let us first consider the simplest case of conventional synchronous lock-in correlation, where the lock-in trigger is derived from the trigger of the free-running digitizing events (camera frames in the case of lock-in thermography) by a frequency divider. Figure 2.7a shows this correlation for the typical case that the sampling rate f_s is considerably larger than the lock-in frequency $f_{\text{lock-in}}$. Here a sine-shaped signal with an amplitude of 2 is assumed to be digitized at certain equidistant measurement (sampling) events. In (a) the signal period is exactly 12 measurement periods long. The dark squares are symbolizing the measurement events, and the vertical dashed line marks the beginning of the second lock-in period. An alternative interpretation of this and the following figure is that the y-values of the dark squares represent the values of the weighting factors $K^{0°}$ of the correlation for the 0° (in-phase) correlation. Note that, according to (2.7) in Sect. 2.2, these factors are the doubled values of the sine function in the sampling moments. The values of $K^{-90°}$ belonging to the corresponding cosine function are not shown here. We always assume that the first measurement starts at $t = 0$, where the signal is at phase position $\Phi = 0°$. We see that in (a) after 12 samples there is the next (13th) sample again exactly at phase position $\Phi = 0°$ of the next period. Here, due to the synchronization of the lock-in trigger to the measurement events, the values of the weighting factors K_k are the same in each lock-in period, as it had been assumed for (2.4). This "synchronous correlation" is definitely the simplest way to perform lock-in thermography. Another advantage is that, if only a complete number of periods is evaluated, the sum of all weighting factors is always exactly zero. This ensures a perfect suppression of the d.c. part of the signal, which is called "topography image" in lock-in thermography.

Alternatively, the measurement events and the lock-in periods may also be not synchronized, which will be called in the following "asynchronous correlation". Note that our denomination "synchronous correlation" should not be confused with the terms "synchronous detection" or "synchronous demodulation", which are often generally used for any kind of lock-in detection and also imply asynchronous correlation. In Fig. 2.7b one lock-in period is somewhat larger than 12 measurement periods, hence there is no measurement event at the starting point of the following sine period. If $K(t)$ is the continuous correlation function in (2.2), which is phase-coupled to the signal to be measured, in (b) the discrete values of the weighting factors K_k in (2.3) have to belong to individual phase positions in each period. Therefore in this asynchronous correlation procedure (b), the weighting factors K_k

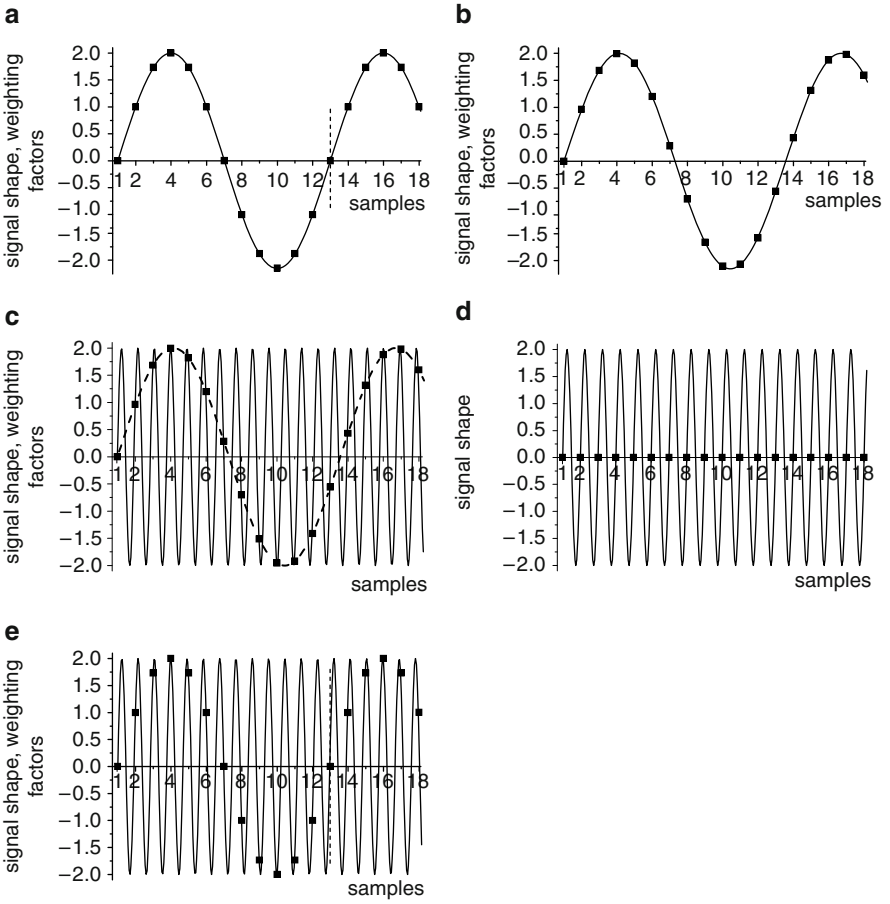

Fig. 2.7 (**a**) Timing diagrams of conventional synchronous correlation, (**b**) conventional asynchronous correlation, (**c**) asynchronous undersampling, (**d**) undersampling at a "forbidden frequency", (**e**) synchronous undersampling

have to change from period to period. This is practically realized by feeding a digital or an analog signal, which is synchronous to the signal generation process, to the computer which organizes the digitizing (sampling, imaging) events. The computer recognizes the frequency of this signal and also its phase, hence it recognizes and may predict the moments when each signal period starts. From these data and the knowledge of the moments of the internal digitizing events, the computer may calculate on-line in which phase position every sampling event is lying. Then the values K_k of the correlation function in (2.3) can either be taken from a look-up table or be calculated on-line for each period separately. In fact, this operation is analog to the operation of a conventional lock-in amplifier (see Fig. 2.3 in Sect. 2.2), where a reference signal of the measurement process is also fed to the lock-in amplifier. It is even possible to digitize an analog waveform, which follows

the expected signal shape, and use these data as the weighting factors for the correlation in (2.3). This also enables non-harmonic or even non-periodic signals to be correlated on-line. One problem of digital asynchronous correlation is that, as a rule, the sum of all weighting factors is not exactly zero. This diminishes the degree of d.c. suppression of the procedure, which is disturbing especially for low acquisition times implying only a limited number of lock-in periods. Therefore, in systems using asynchronous correlation, the d.c. part is usually subtracted from the image data before correlation [29].

The beauty of asynchronous correlation is that, with some restrictions (see below), the lock-in frequency may be arbitrary in relation to the sampling frequency f_s and may be even larger than f_s. For explaining the following terminology, let us make a short excursion to sampling theory. According to the Nyquist-Shannon sampling theorem, at least two samples per period are necessary for detecting a certain frequency. For example, a CD containing frequencies up to 22 kHz uses a sampling rate of 44 kHz. The maximum possible detected frequency being half the sampling frequency is called "Nyquist frequency". Since in lock-in thermography we have to detect two phase components of a signal shape independently, we need at least four samples per lock-in period, as it had been mentioned already several times. It had been demonstrated in Sect. 2.2 in the discussion of (2.14) that 4-point correlation actually consists of two independent 2-point correlations, each using half of the total sampling rate, since here always two of the four weighting factors are zero. Using more than two samples for sampling (or more that 4 for two-phase correlation; e.g. 12 in Fig. 2.7a) is called "oversampling". This measure is also often used in audio technique for improving the signal quality and accuracy. Also in lock-in thermography oversampling improves the accuracy of the harmonic correlation procedure, see Sect. 2.5. If less than two samples per lock-in period are taken (four samples for two-phase correlation), hence if the signal frequency is above the Nyquist frequency, different waveforms may become indistinguishable or "aliases" of each other, as the dashed curve in Fig. 2.7c shows. This curve represents the so-called sub-Nyquist frequency of a waveform having a frequency above the Nyquist frequency. Hence, in this case the frequency analysis of the digitizing procedure is not unique, and in the reconstruction of the signal certain "aliasing frequencies" appear, which have not been present in the original signal. For avoiding these distortions, frequencies higher than the Nyquist frequency have to be filtered out before digitizing by applying so-called anti-aliasing filters. However, if we already know the frequency of interest, we are allowed to sample also a signal with a frequency higher than the Nyquist frequency! Following Krapez [29] this procedure is called "undersampling". In Fig. 2.7c the signal frequency is slightly above the sampling frequency. As this figure shows, the magnitude of a signal with a frequency higher than the sampling frequency may be detected reliably. Here the measurement events occur also successively in 12 different phase positions of the signal period, but the measurements occur in subsequent periods of the signal. The correlation procedure is exactly the same as for asynchronous oversampling with about 12 samples per period, see Fig. 2.7b. However, some restrictions for choosing the signal (lock-in) frequency have to be considered. First, any multiples of the

Nyquist frequency are "forbidden frequencies" for undersampling. As an example, Fig. 2.7d shows sampling of a waveform at $f_{\text{lock-in}} = f_s = 2f_{\text{Nyq}}$. We see that this is a typical interference, hence the samples contain no information as to the amplitude of the periodic signal. For two-phase correlation this first interference occurs already at $f_{\text{lock-in}} = f_s/2 = f_{\text{Nyq}}/4$, since the sampling of each phase component occurs at $f_s/2$ (see Sect. 2.2). Another restriction may hold if the signal contains higher harmonics. Also these harmonics may generate spurious signals if they generate a common sub-Nyquist frequency as the basic harmonic. Frequencies where this may happen are called "aliasing collision frequencies" [37, 38]. In principle, also these frequencies should be avoided for undersampling. Note, however, that for square-pulse heat introduction the basic harmonic has the highest intensity anyway, and higher harmonics of the temperature modulation are naturally strongly damped (see Sect. 2.5). In fact, as long as only the location of local heat sources has to be imaged by lock-in thermography, it does not matter whether the basic harmonic or any higher harmonic is detected. So for electronic device testing aliasing collisions should not be harmful. For strictly harmonic heat introduction there are no aliasing collision frequencies at all, of course. Also therefore harmonic heat introduction is usually preferred in non-destructive testing (NDT), where undersampling is often used. Finally, it has to be mentioned that for undersampling a sufficiently low frame integration time of the IR camera has to be chosen. Note that for modern high-speed IR cameras the frame integration time may be close to the frame period, hence the camera is detecting photons for nearly the whole time between two images. This operation mode may be optimum for achieving a high signal-to-noise ratio, but in the case of undersampling the measured temperature modulation averages out during image integration over more than half a signal period. Therefore a frame integration time below half the period of the signal to be detected has to be chosen for reliably detecting the T signal in the case of undersampling.

Undersampling can also be performed synchronously as Fig. 2.7e shows. Synchronous undersampling has the advantage that, if only a complete number of correlation periods is evaluated, the sum over all weighting factors is exactly zero and the d.c. suppression of the correlation procedure is optimum. In addition, it can be realized relatively easy since the weighting factors are pre-defined and do not have to be calculated on-line. In Fig. 2.7e exactly after 12 sampling periods (at the 13th sample, which already belongs to the next period) the signal is at a phase position of 0° again, see the vertical dashed line. Also here the correlation occurs according to (2.4) and (2.10) with $n = 12$, hence the weighting factors are the same in all correlation periods. The condition that this happens is $f_{\text{lock-in}} = f_s(1 + 1/12)$, which can be realized phase-synchronously only by deriving both the lock-in frequency and the camera trigger from one and the same clock signal by using different frequency dividers. Hence, here the camera is not freely running at its own repetition frequency but is externally triggered by the computer, which also generates the lock-in trigger and thus organizes the synchronous undersampling procedure. If the camera cannot be triggered externally, the realization of synchronous undersampling is more complicated but can also be realized by using a special hardware using PLL (phase-locked loop) techniques [39].

External triggering of the camera opens a number of further timing options. Generally, synchronous undersampling with n samples per correlation period is realized by choosing:

$$f_{\text{lock-in}} = f_{\text{s}} \left(m + \frac{1}{n} \right). \tag{2.15}$$

Here m is an integer, which is sometimes called the order of undersampling [38], and n is the number of frames per correlation period, which has to be $n \geq 4$ for performing two-phase correlation. Undersampling of zeroth order ($m = 0$) is the conventional synchronous correlation procedure with $f_{\text{lock-in}} = f_{\text{s}}/n$. For undersampling of first order, subsequent samples are taken in subsequent periods of the signal, whereas for second order, samples are taken in every third period, and so on. Figure 2.8 shows the timing diagrams of synchronous undersampling of first and second order ($m = 1$ and 2) for $n = 4$ (4-point correlation). Again, the vertical lines mark the beginning of the next correlation period at the fifth sampling event. Synchronous timing with $f_{\text{lock-in}} = f_{\text{s}}(1-1/n)$ is also possible, but then the phase of the signal appears inverted if the evaluation according to Sect. 2.2 is applied (see below).

If the camera is triggered externally, small gaps in the desired frequency (e.g. between $f_{\text{s}}/4$ and $f_{\text{s}}/6$ for $m = 0$) can be closed by reducing the sampling rate, hence the frame rate of the camera, below the maximum possible one. It will be shown in Sect. 2.6 that the signal-to-noise ratio (SNR) increases with the square

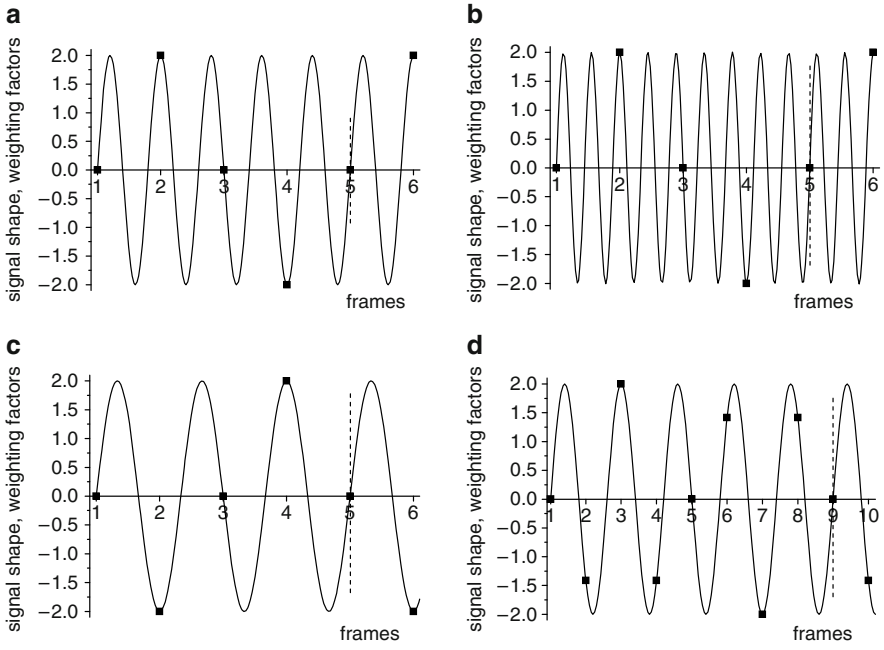

Fig. 2.8 Timing diagrams of 4-point undersampling (**a**) of first order, (**b**) of second order, and of synchronous multi-period correlation with (**c**) $u/n = 3/4$ and (**d**) $u/n = 5/8$

root of the frame rate. Hence, reducing the frame rate by 33% leads only to 18% degradation of the SNR, which usually can be tolerated. However, according to (2.15) there is a larger frequency gap by a factor of four between $f_s/4$ (the highest possible frequency for $m = 0$) and f_s (the lowest possible frequency for $m = 1$). Hence, if e.g. a frequency slightly above 1/4 of the maximum frame rate of the camera should be applied, according to (2.15) only undersampling of first order can be applied with the frame rate of the camera reduced by a factor of 4. This would degrade the SNR already by a factor of 2, which is not negligible anymore.

Fortunately, for filling this frequency gap a special variant of synchronous correlation can be applied, which may be called "synchronous multi-period undersampling". In conventional correlation four or more (generally n) samples are distributed over one signal period. For undersampling according to (2.15), one sample occurs in every one or more (generally m) signal periods. For synchronous multi-period undersampling, four or more (generally n) samples are distributed over two or more (here u) sample periods. Any lock-in frequency exactly fulfilling the condition $f_{\text{lock-in}} = (u/n)\, f_{\text{frame}}$ with u and n being integers with $n \geq 4$ and $u \geq 2$ can be used to fill the frequency gap between $f_{\text{frame}}/4$ and f_{frame}. The case $u = 1$ is the case of conventional correlation. Figure 2.8c shows the application of this technique for $u = 3$ and $n = 4$ ($f_{\text{lock-in}} = 0.75 f_s$). In this synchronous multi-period correlation procedure, after u periods of the lock-in frequency have passed, exactly n samplings have passed, and the phase relation between both is the same as at the beginning (next sampling at $\Phi = 0°$, see vertical dashed lines). The integers u and n should not be divisible by each other, otherwise the multi-period decays into several equivalent multi-periods, where $n < 4$ may appear. As for synchronous undersampling according to (2.15), also synchronous multi-period correlation has the advantage that the d.c. suppression is optimum if a complete number of correlation periods is evaluated, and that the weighting factors are pre-defined. Assuming again that at $t = 0$, where the first sampling occurs, the signal is at phase position $\Phi = 0$, the sampling moments for multi-period undersampling are in the following phase positions of the signal:

$$\Phi_j = 2\pi(j-1)\frac{u}{n}. \tag{2.16}$$

As in (2.4), j is the index of the samples in each period. Correspondingly, the following weighting factors have to be used in (2.4) for synchronous multi-period correlation instead of (2.10):

$$K_j^{0°} = 2\sin\left(\frac{2\pi u(j-1)}{n}\right) \qquad K_j^{-90°} = -2\cos\left(\frac{2\pi u(j-1)}{n}\right). \tag{2.17}$$

In fact, the case of $f_{\text{lock-in}} = 0.75 f_s$ ($u = 3, n = 4$), which is shown in Fig. 2.8c, corresponds to the above-mentioned case of $f_{\text{lock-in}} = f_s(1 - 1/4)$, which is equivalent to conventional 4-point correlation with inverted phase. Here the weighting factors according to (2.17) are $K^{0°} = (0, -2, 0, 2)$ and $K^{-90°} = (2, 0, -2, 0)$, which are the inverse of the double sin- and the cos-function. Figure 2.8d shows

another case with $u = 5$ and $n = 8$ ($f_{\text{lock-in}} = 0.625 f_s$). Here the values of the weighting factors for the 0° correlation do not follow anymore a simple sine function. For sufficiently filling the frequency gap between $1.1\,f_s$ (undersampling with $m = 1$ and $n = 10$) and $0.25\ f_s$ (the highest conventional lock-in frequency), synchronous multi-period undersampling with $u/n = 9/10$ (0.9), 3/4 (0.75), 5/8 (0.625), 4/10 (0.4), 3/8 (0.375), and 3/10 (0.3) can be used. Also the frequency gap between $f_{\text{lock-in}} = 1.25 f_s$ ($m = 1, n = 4$) and $f_{\text{lock-in}} = 2.25 f_s$ ($m = 2, n = 4$) can be filled by choosing $f_{\text{lock-in}} = 1{,}75 f_s$, where the same weighting factors as for $f_{\text{lock-in}} = 0.75 f_s$ are used, see above. For even higher frequencies, undersampling with $n = 4$ is generally most appropriate. By using these possibilities, together with conventional correlation and appropriate reduction of the externally controlled frame rate of the camera, any necessary lock-in frequency should be able to be realized by preserving the advantages of synchronous correlation (pre-defined weighting factors and optimum d.c. rejection) without compromising the signal-to-noise ratio. Remember that in any case of undersampling the frame integration time has to be smaller than half the signal period for avoiding signal averaging effects. This reduces the sensitivity for oversampling of high order and, together with the accuracy of the timing, finally limits the usefully attainable upper limit of the lock-in frequency.

2.5 Influence of Non-harmonic Heating

In the field of non-destructive testing (NDT) nearly all lock-in thermography publications have been based on the assumption that the introduction of heat into the investigated object has to be harmonic. Indeed, the whole theory of thermal waves, which the equations in Sect. 2.2 refer to, is based on the assumption of a harmonic introduction of heat. Therefore, enormous efforts have been made to ensure a most possible harmonic introduction of heat for lock-in thermography used in non-destructive testing. However, as mentioned in Chap. 1, in electronic device testing, as a rule, it is not useful to aspire a harmonic introduction of heat, because of the highly non-linear dependence of the properties of these devices on the supply voltage. Instead, a well-defined bias has to be switched on and off, leading to a square wave shape of the heat introduction. Here the question arises: Does this deviation from the harmonic introduction of heat possibly lead to erroneous results, if the evaluation formulas are based on the assumption of an harmonic heat-introduction? It had been mentioned in Sect. 2.4 that, in the case of oversampling, higher harmonics may lead to aliasing collisions [38], which are not really harmful at least for qualitative electronic device testing. The question remains, how large these errors are for conventional lock-in correlation, if really quantitative results should be obtained.

In this section we will analyze the influence of a non-harmonic (square wave) introduction of heat on the lock-in thermography results for different lock-in correlation conditions. First of all, it has to be emphasized that heat dissipation is an essentially linear process, as long as the thermal parameters of the material can be assumed to be temperature-independent. Hence, since the heat diffusion equation

is of first order (see Chap. 4), its solutions for different boundary conditions (heat sources) superimpose linearly. This also holds for the different spectral components of a square wave heat introduction, which for a basic frequency of $f = \omega/2\pi$ and a power of P_0 during the pulse can be described by a d.c. part and the sum of its Fourier components [32]:

$$P(t) = \frac{P_0}{2} + P_0\left(\frac{2}{\pi}\sin\omega t + \frac{2}{3\pi}\sin 3\omega t + \frac{2}{5\pi}\sin 5\omega t + \ldots\right) \tag{2.18}$$

The influence of the time-independent member $P_0/2$ will be treated in Sect. 4.1 and 4.2, discussing the initial heating phase of a lock-in thermography measurement. It will not be considered here. All harmonic members in (2.18) can be treated independently by the thermal wave formalism (see Chap. 4). Hence, for any heat source geometry the temperature field of the square wave heat-introduction is the superposition of different harmonic temperature fields for this geometry, belonging to the different harmonic frequencies f, $3f$, $5f$, etc. Since the square wave signal is symmetric in time, only odd harmonics have to be regarded.

It will emerge that possible errors introduced by a non-harmonic introduction of heat strongly depend on the parameters of the lock-in correlation. In any case, we will consider only standard sin/−cos lock-in correlations based on (2.4) and (2.10) in Sect. 2.2. If the lock-in frequency is small compared to the frame rate of the camera, the weighting factors (2.10) precisely approximate harmonic functions. Then the sum (2.4) is equivalent to the integral representation of the lock-in process. In this case, the sin/−cos correlation really works in the frequency-selective mode with respect to the basic harmonic, hence all higher harmonics are effectively suppressed, since they are leading to members of the kind

$$\int_0^{2\pi} \sin\omega t\,\sin(n\omega t+\varphi)\mathrm{d}t = \int_0^{2\pi} \cos\omega t\,\sin(n\omega t+\varphi)\mathrm{d}t = 0 (n = 3,5,7,\ldots) \tag{2.19}$$

which holds for any value of the phase shift φ of the T-response. Hence, as long as the number of frames per lock-in period is large and sin/− cos correlation is used, it is absolutely not necessary to manage a harmonic heat-introduction. This statement also holds for any kind of non-destructive testing.

The situation changes if the lock-in frequency reaches the order of the frame rate of the camera, or if the 4-bucket correlation is used. The possible errors caused by a non-harmonic heat-introduction also depend on the geometry of the heat source. The least critical case is a spatially extended heat source, leading to an essentially homogeneous heating of the whole sample. It will be shown in Sect. 2.7 (Calibration) that in this case, the thermal response drops with $1/f$. According to (2.18), the first higher harmonic of a square wave signal is the 3rd one, which itself has only $1/3$ of the intensity of the basic harmonic. Since even for the same power, its thermal response would be only $1/3$ of that of the basic harmonic, in sum the 3rd harmonic of a square wave signal then has a T-response of only 1/9 of the basic one.

Higher harmonics are quenched even more strongly. In fact, as it will be discussed in Sect. 2.7, the T-response then will be of triangular shape, which is close to the sinusoidal one.

The most critical case is, for instance, a point-like heat source at the surface of a 3-dimensional sample, because according to Chap. 4 its T-response in source position is independent of the frequency. Therefore, here the T-modulation in source position is really rectangular. The errors will turn out to additionally depend on the frame integration time of the camera (see Sect. 2.1) and on the phase of the frame integration, referring to the phase of the heating pulse. Hence, it is decisive whether the "electronic shutter" of the camera is open only for a small fraction of each frame period or during the whole frame period, and when the shutter opens, referring to the edge of the heating pulse. The origin of this error is sketched in Fig. 2.9, showing a rectangular T-modulation trace and the frame integration periods for a low frame integration time for a 4-point correlation, depicted for two different phase positions of the frame integration with respect to the pulse edge. It is obvious that the measured data are exactly the same in both cases, in spite of the different phase positions. Hence, the readings in the 0°- and in the quadrature channel are exactly the same. They are no longer linearly independent. This means that, in principle, they do not allow one to calculate amplitude and phase values independently. If the correlation factors are chosen according to (2.10), in both cases, the detected phase will be $-45°$. This is clearly wrong, since in the left image it is close to 0°, and in the right one it is close to $-90°$. Hence, in this special case the phase reading is independent on the real phase. For a longer frame integration time, if the integration periods will overlap with the pulse edges, the results will start to react on the real phase position, hence the results will become more realistic.

Now imagine that in the left part of Fig. 2.9 all the frame integration periods are starting even earlier, whereby periods 1 and 3 are crossing the rising and falling edges of the rectangular temperature track. Then the measured values will very sensitively depend on the exact starting position of the integration period with respect to the edges of the signal. In other words: If the edges of the signal are lying within a short integration period t_{int} of the camera, the sensitivity of the phase reading is

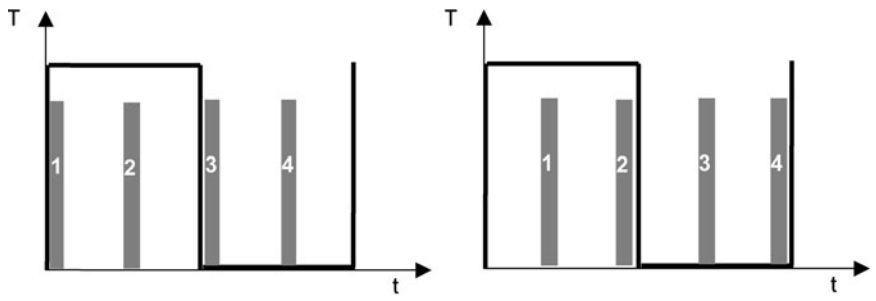

Fig. 2.9 Rectangular temperature trace, as it is measured e.g. in the position of a surface point heat source with a rectangular heat introduction, measured using the 4-point correlation at two different phases between frame integration and heating

increased by an order of $T/4t_{\text{int}}$, with $T = 1/f_{\text{lock-in}}$ being the lock-in period. This effect has been intentionally used by Ramspeck et al. [40] for performing dynamic carrier lifetime imaging of silicon wafers using an infrared-camera-based approach. This technique will be described in more detail in Sect. 3.5.3.

With increasing distance from the source position, the different spectral components are more strongly damped, the higher their frequency is (see Sect. 4.3), leading to a smoothing of the shape of the trace. This is demonstrated in Fig. 2.10, showing the simulation of the a.c. component of one period of the T-modulation at different distances from the considered point heat source. These data have been calculated up to the 99th harmonic. For a better comparison, all traces are independently scaled to the same amplitude of their basic harmonic. The distances r refer to the thermal diffusion length Λ, which in silicon is about 1 mm for a lock-in frequency of 30 Hz (see Sect. 4.1). The smoothing effect and also the expected phase delay with increasing distance are clearly shown. At a distance twice the thermal diffusion length the signal shape is already nearly sinusoidal.

In order to quantify possible errors in the amplitude and phase reading caused by a square-shaped heat introduction, the data of the temperature traces in Fig. 2.10 have been used to simulate different kinds of lock-in correlation. Amplitude and phase values for 4 and 8 frames per lock-in period have been calculated. In all cases, simulations followed for a short frame integration time of <10% of the frame period, similar to that shown in Fig. 2.9, and for a long frame integration time, filling 100% of the frame period. Note that the 4-point correlation with 100% frame integration time is equivalent to the 4-bucket correlation [29]. Figure 2.11 shows all data referred to the "exact" result, which holds for both a sinusoidal heat introduction and for a square-wave heat introduction and a large number of frames per lock-in period,

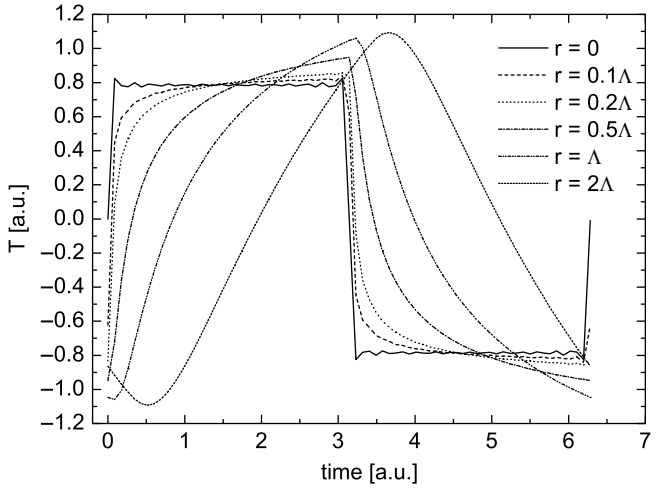

Fig. 2.10 Temperature traces in different distances r from a point source (referred to the thermal diffusion length Λ). All traces are separately scaled to the same amplitude of their basic harmonic

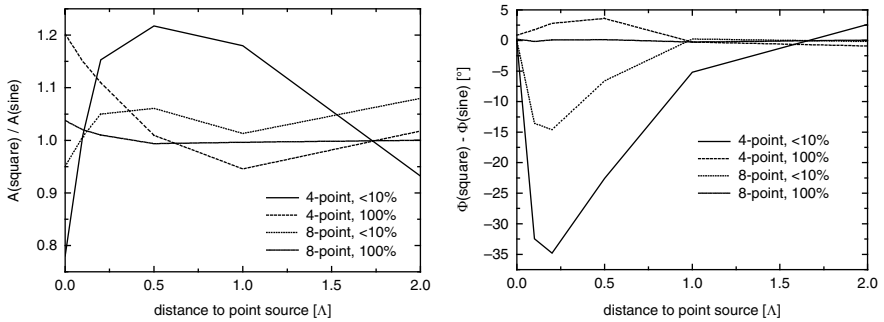

Fig. 2.11 Amplitude error (left) and phase error (right) of square-wave heat introduction referring to sinusoidal one, calculated for 4-point- and 8-point correlations with different percentages of the frame integration time relative to the frame period. The distance from the point source is given in units of the thermal diffusion length Λ

since in both cases the results were exactly the same. As expected, the 4-point correlation with $<10\%$ frame integration time produces the strongest errors ($\pm 20\%$ in amplitude, $-35 \ldots + 5°$ in phase), followed by the 4-point correlation with 100% frame integration time ($+20 \ldots - 7\%$ in amplitude, $-15 \ldots + 4°$ in phase), and the 8-point correlation with $<10\%$ frame integration time ($-5 \ldots + 6\%$ in amplitude, $-1 \ldots + 3.5°$ in phase). The results were best with the 8-point correlation with 100% frame integration time ($-1 \ldots + 3\%$ in amplitude, $\pm 1°$ in phase). With a larger number of frames per period, all these errors are approaching zero.

Note that the amplitude errors found here are always related to the amplitude values measured under the same conditions with a sinusoidal heat introduction. It will be shown in Sect. 4.3 that within the considered distance range the amplitude may vary by two orders of magnitude, depending on the spatial resolution of the system. Hence, the qualitative appearance of amplitude images hardly depends on the fact whether the heat introduction is sinusoidal or square-shaped. Only if results have to be evaluated quantitatively (see Sects. 4.5 and 5.3), or if phase positions have to be determined precisely, or if a signal of a well-defined phase has to be extracted from the measured data, for applying the square-wave heat introduction it is advisable to use >10 frames per lock-in period in order to keep the errors sufficiently small. The following conclusions may be drawn from these simulations:

1. If instead of a sinusoidal heating power a square-wave one is generated in the sample, this may cause quantitative errors in both the amplitude and phase reading.
2. For conventional synchroneous correlation (no undersampling) these errors are largest for the 4-point correlation and quickly reduce with increasing number of frames per lock-in period. Above 10 frames/period they become negligible.
3. Therefore, for low lock-in frequencies, the sin/$-$cos correlation is clearly superior to the 4-bucket correlation procedure, if square-wave heating is used.
4. The errors are largest in the vicinity of local heat sources, where the T-modulation still has the largest amount of higher harmonics.

5. The errors become lower, if the frame integration time covers the largest possible amount of the frame period. This measure is also advantageous for obtaining a good signal-to-noise ratio (see next section).
6. As long as the results are evaluated only qualitatively, the application of square pulses is equivalent to the sinusoidal heat introduction. Only if the results are to be evaluated quantitatively and a high accuracy is expected, or if a well-defined phase signal has to be displayed, it is advisable to work with $\geq$10 frames per lock-in period.
7. If undersampling is used, higher harmonics may be more disturbing, since they may lead to so-called aliasing collisions. These, however, are also only disturbing for quantitative evaluations.

2.6 Noise Analysis

The noise level of an IR-camera based temperature mapping system as well as of any other one is usually expressed by a "noise equivalent temperature difference" (NETD). The definition of NETD is given by the mean square of the deviation of different values measured of one pixel F_i (measured in digits) from its mean value $\bar{F}$ obtained in k successive measurements ($k \gg 1$):

$$\mathrm{NETD} = \frac{c}{k}\sqrt{\sum_{i=1}^{k}\left(F_i - \bar{F}\right)^2} = \frac{c}{k}\sqrt{\sum_{i=1}^{k}\left(F_i^N\right)^2} = c\sigma_{\mathrm{cam}} \tag{2.20}$$

(c = scaling factor of the camera, given e.g. in mK/digit; F_i^N = noise signal in digits of the i th measured value, related to its mean value $\bar{F}$; σ_{cam} = standard deviation of the camera noise, measured in digits). If the camera signal is given not in digits but in (m)K, as usual for modern thermocameras, c has to be set to 1 in the following. According to (2.20), NETD is given in units of mK. If all pixels of the detector array, in general, have the same noise properties, it does not matter whether (2.20) is calculated from successive measurements of one pixel or from evaluating n pixels of one image nominally showing a homogeneous constant temperature $\bar{T} = c\bar{F}$.

According to Sect. 2.3 ((2.4) and (2.10)), for a synchronous sin/−cos correlation, the 0°- and the −90° images are calculated from the measured values $F_{i;j}(x, y)$ as:

$$S^{0^\circ}(x, y) = \frac{2c}{nN}\sum_{i=1}^{N}\sum_{j=1}^{n}\sin\left(\frac{2\pi(j-1)}{n}\right)F_{i;j}(x, y) \tag{2.21}$$

$$S^{-90^\circ}(x, y) = \frac{-2c}{nN}\sum_{i=1}^{N}\sum_{j=1}^{n}\cos\left(\frac{2\pi(j-1)}{n}\right)F_{i;j}(x, y)$$

(n = number of frames per lock-in period; N = number of averaged lock-in periods). A typical quantity of images used for a lock-in measurement $F_{i;j}(x, y)$ is the sum of three components: 1. a steady-state ("topography") part $F^{\text{topo}}(x, y)$, which is assumed to be noise-free here and should be the same for all images, 2. the actually interesting local temperature modulation signal $\Delta F_j(x, y)$, which is phase-coupled to the periodic heat introduction. This is also assumed to be noise-free here and repeated in each lock-in period, and 3. a statistical noise contribution $F^N_{i;j}(x, y)$, which is related to the NETD of the camera via (2.20):

$$F_{i;j}(x, y) = F^{\text{topo}}(x, y) + \Delta F_j(x, y) + F^N_{i;j}(x, y) \tag{2.22}$$

Here, we have neglected the non-steady-state initial heating period at the beginning of each lock-in thermography measurement, which will be justified in detail in Sects. 4.1 and 4.2. Inserting (2.22) into (2.21) yields:

$$\begin{aligned} S^{0^\circ}(x, y) &= 0 + c\Delta F^{0^\circ}(x, y) + cN^{0^\circ}(x, y) \\ S^{-90^\circ}(x, y) &= 0 + c\Delta F^{-90^\circ}(x, y) + cN^{-90^\circ}(x, y) \end{aligned} \tag{2.23}$$

The first term containing the topography signal becomes zero due to the fact that, at least for the synchronous correlation, the sum of all weighting factors is exactly zero. This property is responsible for the perfect topography rejection of the synchronously working lock-in thermography technique. $\Delta F^{0^\circ}(x, y)$ and $\Delta F^{-90^\circ}(x, y)$ are the local amplitude values of the local temperature modulation signals in-phase and out of phase measured in digits, which we want to measure. $N^{0^\circ}(x, y)$ and $N^{-90^\circ}(x, y)$ are the noise values of the 0°- and −90°-signals (measured in digits) for this measurement:

$$\begin{aligned} N^{0^\circ}(x, y) &= \frac{2}{nN} \sum_{i=1}^{N} \sum_{j=1}^{n} \sin\left(\frac{2\pi(j-1)}{n}\right) F^N_{i;j}(x, y) \\ N^{-90^\circ}(x, y) &= \frac{-2}{nN} \sum_{i=1}^{N} \sum_{j=1}^{n} \cos\left(\frac{2\pi(j-1)}{n}\right) F^N_{i;j}(x, y) \end{aligned} \tag{2.24}$$

Since the values of $F^N_{i;j}(x, y)$ are statistically distributed, also $N^{0^\circ}(x, y)$ and $N^{-90^\circ}(x, y)$ are statistical, hence we cannot calculate these noise values. However, since we know the standard deviation of $F^N_{i;j}$ from (2.20), we may calculate the standard deviations of N^{0° and N^{-90°, with respect to a number of k different positions. We will call the standard deviations of N^{0° and N^{-90° the system standard deviations $\sigma^{0^\circ}_{\text{sys}}$ and $\sigma^{-90^\circ}_{\text{sys}}$. We know that the quantity of $F^N_{i;j}$ is normally distributed with the average value of 0. It has the standard deviation σ_{cam} (see 2.20). According to the laws of statistics [32] the variance (being the square of the standard deviation) of a sum of statistical quantities equals the sum of the variances of the elements of this sum:

$$\frac{1}{k}\sum_{l=1}^{k}\left(N_l^{0^\circ}\right)^2 = \left(\sigma_{\text{sys}}^{0^\circ}\right)^2 = \sum_{i=1}^{N}\sum_{j=1}^{n}\left(\frac{2}{nN}\sin\left(\frac{2\pi(j-1)}{n}\right)\sigma_{\text{cam}}\right)^2$$
$$= \frac{2\sigma_{\text{cam}}^2}{nN} \tag{2.25}$$
$$\frac{1}{k}\sum_{l=1}^{k}\left(N_l^{-90^\circ}\right)^2 = \left(\sigma_{\text{sys}}^{-90^\circ}\right)^2 = \sum_{i=1}^{N}\sum_{j=1}^{n}\left(\frac{-2}{nN}\cos\left(\frac{2\pi(j-1)}{n}\right)\sigma_{\text{cam}}\right)^2$$
$$= \frac{2\sigma_{\text{cam}}^2}{nN}$$

The latter identities hold because of the identity

$$\sum_{j=1}^{n}\sin\left(\frac{2\pi(j-1)}{n}\right)^2 = \sum_{j=1}^{n}\cos\left(\frac{2\pi(j-1)}{n}\right)^2 = \frac{n}{2} \tag{2.26}$$

which holds for each $n \geq 4$ because the average value of both the $\sin^2$ and the $\cos^2$ function is 1/2. Hence, according to (2.20) and (2.25), the system standard deviations of both the 0° and the -90° signals after lock-in correlation are:

$$\sigma_{\text{sys}}^{0^\circ;-90^\circ} = \sigma_{\text{sys}} = \frac{\sqrt{2}\,\text{NETD}}{c\sqrt{nN}} \tag{2.27}$$

For steady-state imaging techniques (e.g. fluorescent microthermal imaging [41]), the standard deviation between two successively measured frames is often used to give a measure of the noise level. For lock-in thermography, however, we have an even more convenient quantity in the phase-independent temperature modulation amplitude given by (2.11) in Sect. 2.3, which is usually measured in every lock-in thermography experiment. Thus, in order to measure the system noise level of a lock-in thermography system, only a steady-state scene has to be imaged (e.g. a black body at room temperature) in the lock-in mode. Then the image is dominated by the system noise. The phase-independent amplitude signal, averaged over a certain representative region implying $X \times Y$ pixel, is very similar to the definition of the standard deviation of the noise component of each lock-in signal component:

$$\langle A_{\text{noise}}\rangle = \frac{c}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y}\sqrt{\left(N^{0^\circ}(x,y)\right)^2 + \left(N^{-90^\circ}(x,y)\right)^2}$$
$$= \sqrt{2}\,c\,\sigma_{\text{sys}} = \frac{2}{\sqrt{nN}}\text{NETD} \tag{2.28}$$

The total number of frames involved in the correlation process $n \times N$ is given by the frame rate f_{frame} and the total acquisition time t_{acq} as

$$nN = f_{\text{frame}}\, t_{\text{acq}}, \tag{2.29}$$

leading to the final result:

$$\langle A_{\text{noise}} \rangle = \frac{2}{\sqrt{f_{\text{frame}}\, t_{\text{acq}}}} \text{NETD} \tag{2.30}$$

Thus, knowing NETD allows us to predict the averaged amplitude noise level after a certain acquisition time according to (2.30). For example, a thermocamera running at $f_{\text{frame}} = 60\,\text{Hz}$ with an NETD of 20 mK, after 30 s acquisition time in a lock-in system, should show an averaged amplitude noise level of $\langle A_{\text{noise}} \rangle = 0.93\,\text{mK}$. Note that the noise level depends on the frame rate of the camera and on the data acquisition time, but not on the lock-in frequency. This prediction, however, is based on the presence of white noise, which needs not be true.

In Chap. 3 a number of different lock-in thermography systems will be compared. This is a non-trivial task, since different measurements have been made using different data acquisition times and different spatial resolutions. For an averaging lock-in experiment, the signal-to-noise ratio improves with the square root of the lock-in integration time t_{int}. This integration time should not be confused with the frame integration time mentioned in Sect. 2.1, which is the time where the electronic shutter is opened to measure one image. In a camera based experiment t_{int} equals the total data acquisition time of the experiment t_{acq}, whereas $t_{\text{acq}} = XYt_{\text{int}}$ holds for an image having Y lines and X rows in a serially measuring experiment. The number of pixels per image XY also determines the noise level to be achieved within a certain data acquisition time. This becomes most obvious in serial experiments with $t_{\text{int}} = t_{\text{acq}}/XY$. But also in camera-based experiments, for a given detector chip size, the number of detected photons for each pixel is proportional to the pixel area. Hence, a larger pixel area is equivalent to a larger integration time. Thus, for comparing different lock-in thermography systems, for each system a universal "pixel-related system noise density" can be defined, which independently of the acquisition time and of the number of pixels describes the noise properties of a certain system:

$$N_{\text{sys}} = A_{\text{noise}} \sqrt{\frac{t_{\text{acq}}}{XY}} \tag{2.31}$$

Here, A_{noise} is the temperature noise amplitude averaged over a representative image region, and $X \times Y$ is the number of pixels of the resulting image. If the noise is measured in mK, this system noise density has the unit $\text{mK}\sqrt{\text{s}}$. This pixel-related system noise density N_{sys} will be used in Chap. 3 as an objective measure to compare the different lock-in thermography systems.

2.7 Calibration

If the temperature measurement system or the thermocamera used in a lock-in thermography system is correctly calibrated, and if there is no cross-talk between successive measurements, the results of the lock-in correlation according to (2.21) are also amplitude values correctly calibrated in units of K or mK. However, in order to check the function of the correlation procedure and to test lock-in thermography systems in general, a simple test device is required, which allows one to perform a functional testing and a quantitative calibration of lock-in thermography systems. Ideally, this test device should provide a homogeneous area, the surface temperature of which can be modulated sinusoidally with a well-defined amplitude. The simplest possible test device is a resistively heated metal stripe. It will be explained in detail in Sect. 4.1 that it is useful to connect this stripe to a heat sink via an insulation layer, providing a well-defined heat resistance between the metal stripe and the heat sink. This insulation layer also provides the necessary electrical insulation between metal stripe and heat sink. Owing to this insulation layer the heat generated by the electrical current in the metal stripe is only slowly fed away to the heat sink. After an initial heating phase the average temperature of the test device is stabilized. If the thermal time constant for this thermal relaxation process is below the inverse of the lock-in frequency, this test device can be regarded to be adiabatic with respect to the lock-in frequency, hence the surface temperature modulation is only governed by the heat capacity of the stripe and the amount of the introduced heat. It is independent of the heat resistance of the insulation layer (for details see Sect. 4.1). If this test device should be used for IR thermography, its surface has to be covered with a thin layer of high IR emissivity (e.g. graphite spray), since any metal is a poor IR emitter. The thickness of this black layer should be so small as to make its heat capacity negligible at the lock-in frequency. Hence its surface should immediately follow the surface temperature modulation of the test device itself. Figure 2.12 shows the cross section through such a resistively heated test device for IR lock-in thermography. The current in the metal stripe is flowing perpendicular to the image plane.

The heat may be introduced into such a resistively heated test device sinusoidally by exposing it to a sinusoidally alternating current. However, here the frequency of the temperature modulation would be double that of the heating current, since

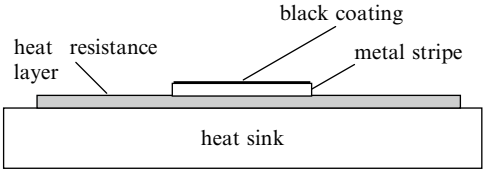

Fig. 2.12 Cross section through a resistively heated IR lock-in thermography test device. The heat resistance layer may be a sheet of paper, glued between the metal stripe and the heat sink

both the negative and the positive half-waves are generating heat. Therefore, it is more convenient to expose this test device to a pulsed d.c. current, hence to realize a square wave introduction of heat as it was discussed in Sects. 2.2 and 2.3. Then, after the initial heating phase, the device is linearly heated up within the first half-period, and in the second half-period, across the heat resistance it cools down linearly to the heat sink. Hence, for a square-shaped heat introduction, the surface temperature modulation in steady-state state has a triangle shape. Cooling begins owing to the steady-state temperature difference, which shows after the initial heating phase between the sample and the heat sink below the device. Therefore, as it will be described in Sect. 4.1 in more detail, the cooling rate is operative over the whole lock-in period and superimposes on the (higher) heating rate within the first half-period, resulting in a lower effective heating rate in the first half-period. Since in the steady-state the amount of heat generated in the first half-period has to be dissipated by the cooling rate operative over the whole period, the effective heating rate in the first half-period is exactly half the "adiabatic" heating rate (see Figs. 4.1 and 4.2 in Chap. 4). Hence, for a constantly dissipated electrical power P during the heating phase, in steady-state the expected real heating rate within the first half-period is exactly half the heating rate to be expected without the continuous cooling effect (m = mass of the metal stripe, c_p = specific heat):

$$\frac{\partial T}{\partial t} = \frac{1}{2}\frac{P}{mc_p} \tag{2.32}$$

Since this heating rate acts within one half of the lock-in period, the peak-to-peak temperature difference of the triangle-shaped temperature modulation of the sample is:

$$\Delta T_{\mathrm{triang}}^{\mathrm{pk-pk}} = \frac{P}{4mc_p f_{\mathrm{lock\text{-}in}}} \tag{2.33}$$

For a triangle-shaped signal having a pk-pk amplitude of 1, its basic harmonic has a pk-pk amplitude of $8/\pi^2 = 0.81$ [32], yielding an amplitude according to the definition in (2.5) of $4/\pi^2 = 0.405$. Thus, the basic harmonic of the temperature modulation amplitude of this calibration test device, if exposed to a pulsed current of I, is expected to be:

$$A_{\mathrm{test}} = 0.405\frac{I^2R}{4mc_p f_{\mathrm{lock\text{-}in}}} = 0.101\frac{I^2R}{mc_p f_{\mathrm{lock\text{-}in}}} \tag{2.34}$$

The correct dimensioning of both the heat resistance layer between the metal stripe and the heat sink, and the thickness of the black coating can easily be checked by measuring the dependence of the measured amplitude signal on the lock-in frequency: Only if both layers are correctly dimensioned the amplitude signal falls with $1/f_{\mathrm{lock\text{-}in}}$ as predicted by (2.34). If the heat resistance is too low, the surface temperature tends to stabilize even within each single heating pulse, leading to a sublinear decay of A_{test} with $f_{\mathrm{lock\text{-}in}}$, and finally for low $f_{\mathrm{lock\text{-}in}}$, to its independence of $f_{\mathrm{lock\text{-}in}}$. On the other hand, if the heat resistance is too large, the duration of the

initial heating phase increases, and the average stabilized temperature of the test device in operation would increase, too. This would not affect the result of the test measurement, since here it is assumed that the measurement time of the calibration measurement is generally large compared to the initial heating phase, or that this temperature drift is correctly compensated, as will be described in Sect. 4.2. If the thickness of the black coating is too large, the heat diffusion resistance from the surface of the test device to the surface of the coating will lead to an additional frequency-dependent quenching of the measured amplitude value with increasing frequency. Hence, instead of the $1/f_{\text{lock-in}}$ dependence expected from (2.34) an even stronger decay of A_{test} with $f_{\text{lock-in}}$ would follow at high frequencies.

2.8 Heat Dissipation and Transport Mechanisms in Solar Cells

Thermal investigations on solar cells are an important application field of lock-in thermography, see Sects. 6.2 and 6.3. The success of this technique in this field relies on the fact that all elementary processes occurring in solar cells are connected with the local heat generation or heat consumption (local cooling). Understanding these processes in detail is the key for understanding the different special lock-in thermography techniques, which have been developed in recent years for the detailed physical investigation of solar cells and modules. Therefore in this section, the elementary electronic processes, which may occur in a dark and illuminated solar cell, are described and related to their heat generation or consumption potential [42]. Reading this section requires basic knowledge in semiconductor physics [43] but is necessary for understanding especially Sects. 5.4, 6.2.2 and 6.3 in detail. Non-thermal processes in such devices, which nevertheless may also be investigated by lock-in thermography, will be described in Sect. 2.9.

The most obvious local heat source is Joule heat, which is generated if any current flows through a material with a finite specific resistance. In the volume the generated local Joule-type power density due to a local current density $J(x, y, z)$ is:

$$p_{\text{J}}^{\text{v}}(x, y, z) = \rho J^2(x, y, z). \tag{2.35}$$

Here ρ is the specific resistance (resistivity) of the material. If the current flows in a thin layer with a specific sheet resistance (sheet resistivity) of ρ_{s}, which happens for example if the current flows in a solar cell horizontally in the thin emitter layer or in a contact grid line on top, the area-related generated Joule-type power density is:

$$p_{\text{J}}^{\text{s}}(x, y) = \rho_{\text{s}}\, J_{\text{s}}^2(x, y). \tag{2.36}$$

Here $J_{\text{s}}(x, y)$ is the sheet current density in units of A/m. Typical resistivities of the bulk material of silicon solar cells are in the range of 1 Ωcm and a typical value of the generated current density under non-concentrated solar illumination (having a

power density of 100 mW/cm^2) is about 30 mA/cm^2. Regarding a typical thickness of a cell of 0.2 mm, this corresponds to an area-related dissipated power density due to the current flowing vertically in the bulk material of about 0.018 mW/cm^2. This is a really negligible value compared to the illumination power density. Thus, the contribution of the volume current in the bulk of solar cells to the generated heat can be neglected for standard solar cells. This does not hold anymore for the horizontal current in the emitter or in the grid lines. Their contribution has to be regarded in any electro-thermal simulation of solar cells according to (2.36) [42]. Note that ohmic shunts, which may be responsible for very dominant local heat sources (see Sect. 6.2.1), also have to be described by (2.35) or, in the case of extended edge shunts, by (2.36).

Solar cells are constructed to feed away typical currents of 30 mA/cm^2 with negligible Joule-type power losses. If the cell does not have any series resistance problems, even the losses according to (2.36) are weak compared to other heat generation mechanisms. Then the dominant heat exchange occurs as a result of the light irradiation and of the vertical current flow across the horizontal p-n junction, which can be illustrated in an energy band diagram. Figure 2.13 shows a simplified band diagram of a p-n junction under forward bias condition, which is exposed to monochromatic illumination with a photon energy of $h\nu$ (h = Planck constant, ν = frequency of light). All important elementary heat generation or consumption mechanisms are indicated in this figure by arrows. Since these arrows symbolize the movement of electrons with the y-axis being their energy, arrows pointing downwards are processes of energy loss, which is generation of heat, and arrows pointing up are processes of energy gain, which is heat consumption, also called Peltier cooling. For holes in the valence band the energy axis is inverted, hence for them an

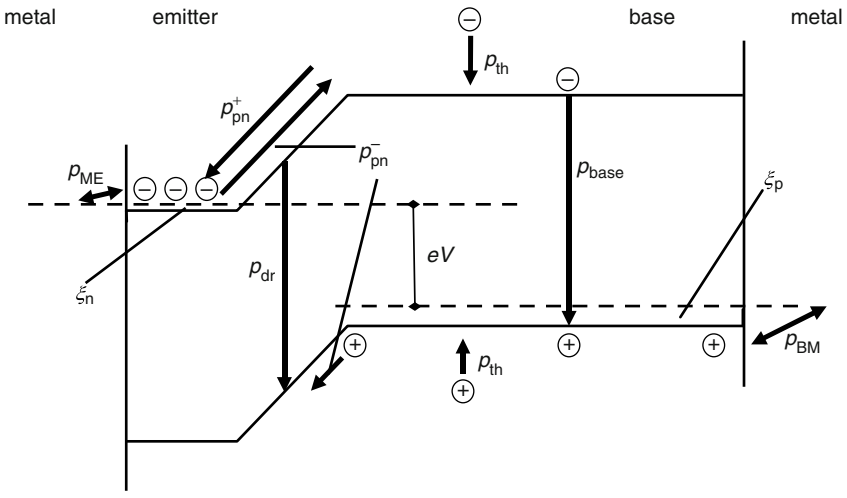

Fig. 2.13 Simplified band diagram of an illuminated p-n junction under forward bias (solar cell in operation) with heat dissipation mechanisms indicated (reproduced with permission of American Institute of Physics)

upward arrow means heat generation and a downward arrow means heat consumption. The dashed lines indicate the majority carrier Fermi levels in the n^+- and the p-side (emitter and base region), respectively, the difference of both being the bias energy eV (e = elementary charge). In the emitter the Fermi level is lying within the conduction band, hence the emitter is degenerately doped. The energy differences between the n- and p-side Fermi levels and the conduction and the valence band edge in these regions are the two Fermi energies ξ_n and ξ_p, respectively. Since the emitter is degenerately doped, ξ_n is negative. The two contacts are symbolized here as vertical lines at the left and right, separating the semiconductor emitter and base from the contact metals. In the following, it will not be considered at which depth in the device (which is the x-axis in Fig. 2.13) an amount of heat is generated or consumed. Hence, the actual solar cell is assumed to be thermally thin, see Sect. 4.1, so all heat generation or consumption mechanisms are assumed to appear at the same lateral position. Exceptions from this assumption will be discussed at the end of this section and especially in Sect. 5.4. Of course, if a thin film solar cell on glass is considered, the whole device may be thermally thick. Nevertheless, also then the active region has a negligible thermal thickness.

If charge is transported in semiconductors by carriers, the electrons in the conduction band and holes in the valence band contributing to this transport have a certain mean energy of ε_e and ε_h above the corresponding band edge. These values depend on temperature T, on the form of the density of states in the two bands, on the doping concentration, and on the nature of the scattering mechanism relevant in the transport process. For a box-shaped density of states, non-degeneracy, and an energy-independent scattering process, the average kinetic energy would be 3/2 kT, which is 1/2 kT for each degree of freedom. A further, purely thermoelectric effect contributes significantly to the kinetic energy term. It is due to the carrier-phonon interaction and is called "phonon drag". The magnitudes of these energies will be discussed in more detail in Sect. 5.4. In the case of non-degeneracy (base region) these are mean kinetic energies, but in the case of degeneracy (emitter region) they also have a potential energy contribution. Here we only will use the quantities ε_e^n, ε_h^n, ε_e^p, and ε_h^p as additional energies for thermalized free electrons and holes in the emitter and in the base, respectively. These kinetic energies were not yet considered in [42] but will be taken into account in the following.

It was assumed in Fig. 2.13 that only in the p-type base excess electrons as minority carriers are generated by irradiation with light with an energy of $h\nu$ being larger than the gap energy (E_g = 1.1 eV for silicon). This excitation event leaves behind a "hot" free hole and a "hot" free electron. By generating phonons, within less than 1 picosecond after the absorption event these carriers loose their excess kinetic energy and thermalize to close to the edge of the bands, where they only have the mean kinetic energy ε_h^p and ε_e^p. Hence, the thermal energy dissipated by this absorption process to the lattice is ($h\nu - E_g - \varepsilon_h^p - \varepsilon_e^p$). This thermalization energy is indicated in Fig. 2.13 by two arrows, both labelled p_{th} (note that the energy counting for holes is inverse, hence both arrows indicate heat generation). If an electron were excited exactly from the edge of the valence band deeper into the conduction band, this process would leave behind a "cold" hole of zero kinetic energy. Thereby

the number of holes increases by 1 but the kinetic energy of the whole hole ensemble does not increase. Thereby the hole temperature, which is a measure of their mean kinetic energy, slightly decreases. This is equivalent to cooling the hole ensemble by an amount of $\varepsilon_{\mathrm{h}}^{\mathrm{p}}$, which instantly couples to the lattice. In the same way generating a "cold" electron cools the electron ensemble by $\varepsilon_{\mathrm{e}}^{\mathrm{p}}$. Even if $\mathrm{h}\nu = E_g$ holds, the two generated "cold" carriers consume a heat energy of $\varepsilon_{\mathrm{h}}^{\mathrm{p}} + \varepsilon_{\mathrm{e}}^{\mathrm{p}}$. Thus, if we assume that all absorbed photons lead to a photocurrent density J_{ph} (i.e. the internal quantum efficiency equals 1), the thermal power per unit area, which is generated by thermalization of absorbed photons in the bulk, is:

$$p_{\mathrm{th}} = \frac{J_{\mathrm{ph}}}{e}\left(h\nu - E_{\mathrm{g}} - \varepsilon_{\mathrm{h}}^{\mathrm{p}} - \varepsilon_{\mathrm{e}}^{\mathrm{p}}\right). \tag{2.37}$$

The assumption that all absorbed photons contribute to the photocurrent needs not to be true. In reality, a certain fraction of the photo-generated minority carriers recombines before reaching the p-n junction. Since this fraction is independent of the bias, it only leads to another constant heat contribution, which is exactly $h\nu$ for each absorbed photon not leading to a photocurrent. This contribution will be neglected in the following, but it can easily be regarded at the end of the calculation. We also will neglect the influence of photons absorbed in the emitter, since this emitter is usually very thin (about 300 nm) and shows a very low minority carrier lifetime. Moreover, we will neglect hole injection into the emitter, since this process is improbable due to the very asymmetric doping concentrations.

If the electrons photo-generated in the bulk reach the p-n junction, they inevitably "fall down" the barrier, i.e. they are attracted by the electric field in the junction and move to the emitter region. Note that the photocurrent is a reverse current for the p-n junction. By acceleration in the electric field, the electrons gain kinetic energy, which they suddenly loose by emitting phonons. This process, which will be discussed in more detail in Sect. 5.4, is another kind of thermalization and actually a Peltier heating process, which occurs at the p-n junction. The amount of heat per unit area generated by this process is:

$$p_{\mathrm{pn}}^{+} = \frac{J_{\mathrm{ph}}}{e}\left(E_{\mathrm{g}} - \xi_{\mathrm{n}} - \xi_{\mathrm{p}} - eV + \varepsilon_{\mathrm{e}}^{\mathrm{p}} - \varepsilon_{\mathrm{e}}^{\mathrm{n}}\right). \tag{2.38}$$

Here $(E_g - \xi_n - \xi_p - eV)$ is the value of the energy barrier between emitter and base at a forward bias of V, the other terms in the bracket are a possible difference of the mean kinetic energy of the electrons before and after crossing the p-n junction. Note that the Peltier effect is a pure heat transport mechanism. Hence, a certain Peltier heating is always balanced by a certain Peltier cooling at another position. It will be discussed in Sect. 5.4 where the cooling belonging to the heating in (2.38) comes from. Note also that the two heat sources p_{th} and p_{pn}^{+} are dissipated at the same lateral location where the light is absorbed. This holds exactly for p_{th} and at least approximately for p_{pn}^{+} since, even for a large diffusion length, the electrons may diffuse laterally in the bulk only over a distance in the order of the cell thickness before they flow through the p-n junction. Therefore, these

two heat dissipation mechanisms, together with recombination of photo-generated minority carriers, have been called "local power dissipation" mechanisms [15]. For all other mechanisms involving majority carriers the carriers are free to move laterally between the moment of their generation and that of heat dissipation, which will be discussed later on.

Since this p-n junction is under forward bias, also the electron current opposite to J_{ph} flows from the emitter across the barrier into the base, which is called the "diffusion current" J_{diff}. This current is symbolized in Fig. 2.13 as an upward arrow named p_{pn}^{-}, hence here electrons gain energy. This current would also flow at this bias without any illumination, hence this is a so-called "dark current". According to the detailed balance principle, these opposite current flows (J_{ph} and J_{diff}) can be described independently from each other. Another contribution to the dark forward current is the so-called "depletion region current" J_{dr}, which is also often called "recombination current". It is due to the recombination of electrons and holes in the depletion region between emitter and base region. Since this recombination occurs predominantly in the middle of the p-n junction, both electrons and holes have to gain an energy of approximately half the barrier height. Hence for the diffusion current the electrons have to gain an amount of energy of the barrier height per charge, minus the difference of the mean kinetic energies in both regions. Also for the depletion region current the electrons and holes together have to gain an amount of energy of the barrier height per charge. We will assume that up to the recombination event their kinetic energy is preserved. Since this energy gain consumes thermal energy, this process is a Peltier cooling effect (for details see Sect. 5.4). The amount of Peltier cooling at the p-n junction per unit area is:

$$p_{\mathrm{pn}}^{-} = -\frac{J_{\mathrm{diff}}}{e}\left(E_{\mathrm{g}} - \xi_{\mathrm{n}} - \xi_{\mathrm{p}} - eV + \varepsilon_{\mathrm{e}}^{\mathrm{p}} - \varepsilon_{\mathrm{e}}^{\mathrm{n}}\right) - \frac{J_{\mathrm{dr}}}{e}\left(E_{\mathrm{g}} - \xi_{\mathrm{n}} - \xi_{\mathrm{p}} - eV\right). \quad (2.39)$$

If thermalized electron-hole pairs recombine non-radiatively in the depletion region, they generate a heating power per unit area of $(E_g + \varepsilon_{\mathrm{e}}^{\mathrm{n}} + \varepsilon_{\mathrm{h}}^{\mathrm{p}})$ per electron:

$$p_{\mathrm{dr}} = \frac{J_{\mathrm{dr}}}{e}\left(E_{\mathrm{g}} + \varepsilon_{\mathrm{e}}^{\mathrm{n}} + \varepsilon_{\mathrm{h}}^{\mathrm{p}}\right). \quad (2.40)$$

In the same way, if due to the diffusion current electrons reach the base region, they inevitably recombine there or at the base contact, leading to a generated heating power per unit area of:

$$p_{\mathrm{base}} = \frac{J_{\mathrm{diff}}}{e}\left(E_{\mathrm{g}} + \varepsilon_{\mathrm{e}}^{\mathrm{p}} + \varepsilon_{\mathrm{h}}^{\mathrm{p}}\right). \quad (2.41)$$

Even if an electron-hole pair recombines directly from the band edges, it generates only an energy amount of E_g. However, this process consumes carriers having zero kinetic energy. Then, in the same way as discussed above for the absorption of light with $h\nu = E_g$, the electron–hole ensemble cools down by an amount of $(\varepsilon_{\mathrm{e}}^{\mathrm{p}} + \varepsilon_{\mathrm{h}}^{\mathrm{p}})$. Here we consider only non-radiative recombination where the whole

recombination energy is converted to heat. The case of radiative recombination will be discussed in Sect. 5.4.

The average absolute energy of the electrons (and holes!) in the contact metals, which contribute to current transport, is close to the position of the Fermi level in these metals. If electrons or holes enter the semiconductor from a contact metal, their energy increases by a potential energy contribution, which is the corresponding Fermi energy ξ_n or ξ_p, respectively, and by a kinetic energy contribution, which for holes entering the base is ε_h^p and for electrons entering the emitter is ε_e^n. If holes or electrons leave the base or the emitter, the same amount of Peltier heat is released. So also at these barriers Peltier cooling or Peltier heating may occur, depending on the current direction. Only if the current flows homogeneously in a cell and the contacts cover the whole area, at each position the net current density between metal and emitter J_{ME} and between base and metal J_{BM} both equal the difference between photocurrent and dark current ($J_{diff} + J_{dr} - J_{ph}$). However, whenever lateral currents are flowing in the device, the local current densities at the contacts may be different. Therefore, the general expressions for Peltier heating or cooling for a current density at the emitter contact of J_{ME} and at the base contact of J_{BM} (both counted positively for a dark forward current and negatively for a reverse current) are [44]:

$$p_{ME} = \frac{-J_{ME}}{e}(\xi_n + \varepsilon_e^n), \tag{2.42}$$

$$p_{BM} = \frac{-J_{BM}}{e}(\xi_p + \varepsilon_h^p). \tag{2.43}$$

This whole set of equations (2.37)–(2.43) describes the heat dissipation and heat transport of the most important mechanisms in a p-base silicon solar cell. In an n-base cell the complementary processes occur. Let us apply these equations to the special case of a current flowing in the dark under forward bias. We will assume here a homogeneous device without any lateral current flow. Joule heating will be neglected. Then only the two dark current contributions J_{diff} and J_{dr} are flowing. The bias dependence of these two currents is described by the following equations [45]:

$$J_{diff} = J_{01}\left(\exp\left(\frac{eV}{kT}\right) - 1\right), \tag{2.44}$$

$$J_{dr} = J_{02}\left(\exp\left(\frac{eV}{2kT}\right) - 1\right).$$

The magnitudes J_{01} and J_{02} are called the "saturation current densities" of the diffusion and the depletion region (recombination) current, respectively. Note that, due the different slopes of these two current contributions, J_{diff} dominates for high forward bias and J_{dr} dominates for medium and low forward bias. J_{diff} is a local measure of the minority carrier lifetime in the bulk. Generally, J_{diff} is an areal current. Regions of low lifetime show a locally increased value of J_{diff} [45]. Local sites of strongly increased J_{dr} are the origin of so-called "non-linear shunts" in solar cells, see Sect. 6.2.1.

For describing the dissipated power density in a homogeneous thermally thin solar cell in the dark, the sum of (2.39)–(2.43) has to be considered with $J_{\mathrm{ME}} = J_{\mathrm{BM}} = J_{\mathrm{diff}} + J_{\mathrm{dr}}$, yielding a total dissipated power per unit area of:

$$p_{\mathrm{dark}} = p_{\mathrm{pn}}^{-} + p_{\mathrm{dr}} + p_{\mathrm{base}} + p_{\mathrm{ME}} + p_{\mathrm{BM}} = V\,(J_{\mathrm{diff}} + J_{\mathrm{dr}})\,. \tag{2.45}$$

As can be expected already from the energy conservation law, the total dissipated power density just equals the product of current density and voltage. All terms containing Fermi energies and kinetic energies, which are related to the Peltier effects, cancel. If the cell contains any kind of inhomogeneity, this holds not anymore locally in any position but only for the whole cell, see below.

If a homogeneous cell is illuminated and a photocurrent J_{ph} flows, also (2.37) and (2.38) have to be added to (2.45), and J_{ph} has to be regarded in (2.42) and (2.43) as a negative current contribution. The result is:

$$p_{\mathrm{ill}} = p_{\mathrm{dark}} + \frac{J_{\mathrm{ph}}}{e}\,(\mathrm{h}\nu - eV)\,. \tag{2.46}$$

Figure 2.14 shows the result of (2.46) using (2.44)–(2.45) for assumed typical values of $J_{01} = 1.5 \times 10^{-12}\ \mathrm{A/cm^2}$, $J_{02} = 2 \times 10^{-6}\ \mathrm{A/cm^2}$, $J_{\mathrm{ph}} = 30\ \mathrm{mA/cm^2}$, and $\mathrm{h}\nu = 1.4$ eV ($\lambda = 880$ nm) [42]. It has been assumed here that all absorbed photons are leading to a photocurrent. We see that both under short circuit condition ($V = 0$, $J = -J_{\mathrm{ph}}$) and under open circuit condition ($V = 0.6\,\mathrm{V}$, $J = 0$) the complete amount of irradiated power ($h\nu$ per irradiated photon) is dissipated into heat. In between, there is a minimum of the dissipated power at the so-called maximum power point at about $V = 0.52\,\mathrm{V}$, which is due to the fact that here the maximum amount of electrical power is generated by the solar cell and fed out to the electric load.

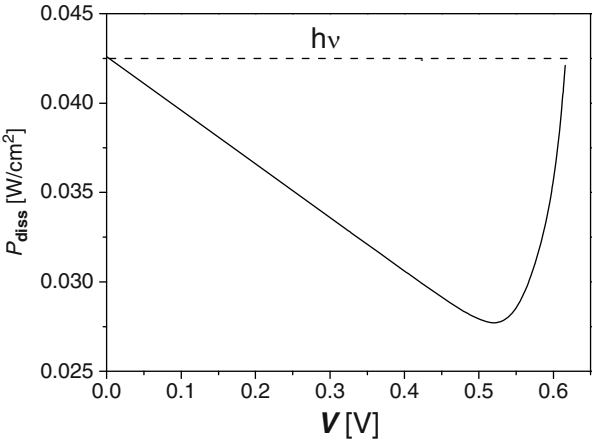

Fig. 2.14 Dissipated thermal power density under monochromatic illumination as a function of the cell bias V (with permission of American Institute of Physics)

Equations (2.35)–(2.43) describing the different elementary heat dissipation and transport mechanisms hold generally, but (2.45) and (2.46) only hold for strictly vertical current flow. If any lateral current flows, e.g. due to local contacts or due to any kind of inhomogeneity in the cell, the elementary mechanisms can be observed separately. For example, the observation of Peltier effects will be presented in Sects. 5.4 and 6.3. The fact that different heat sources may be acting at different sites is the base of lock-in thermography on solar cells. This will be demonstrated in Sects. 6.2 and 6.3.

2.9 Carrier Density Imaging

Until now, only thermal measurements have been considered for lock-in thermography measurements. It had been mentioned in Sect. 2.1 that, according to Kirchhoff's law, the probability of a surface to emit radiation (i.e. its emissivity ε) equals the absorption probability α at this wavelength. Since mid-range radiation in the wavelength range 3–5 µm typically used for lock-in thermography corresponds to a photon energy of 0.25–0.41 eV, which is much smaller than the energy gap of silicon of 1.1 eV, silicon and also most other semiconductors are nominally transparent to infrared light. Hence, in principle, thermography on these materials should be impossible. Fortunately, in doped semiconductors there is another source of IR absorption, which is free carrier absorption. Whenever electromagnetic radiation penetrates a semiconductor having a free carrier concentration of n or p, respectively, the electric and magnetic fields accelerate the free carriers in this body, which is connected with a certain heat generation by absorption of electromagnetic energy. For n-type material having a free carrier concentration of n, the corresponding absorption coefficient can be described by:

$$\alpha_{\mathrm{fc}} = K_n \lambda^2 n. \tag{2.47}$$

Here K_{n} is a material constant, discussed in more detail below, and λ is the wavelength of the radiation. An equivalent equation holds for p-type material. For n-Si, $K_n \approx 10^{-18}\,\mathrm{cm}^2/\mu\mathrm{m}^2$, and for p-Si $K_{\mathrm{p}} \approx (2-2.7) \times 10^{18}\,\mathrm{cm}^2/\mu\mathrm{m}^2$ holds [46]. Hence, free carrier absorption/emission becomes stronger with increasing wavelength and is directly proportional to the free carrier concentration. If both electrons and holes are present in a semiconductor (e.g. under optical excitation), their absorption coefficients simply add up. Only by this free carrier absorption of optically excited or highly doped semiconductor regions, which corresponds to an equivalent free carrier IR-emission, thermal IR-investigations on bare semiconductor devices become possible.

On the other hand, the measurement of free carrier absorption or emission in the IR-range also enables the measurement of the free carrier concentration of semiconductor materials via (2.47). If this measurement is performed by using a thermal camera, images of free carrier concentrations can be obtained. In contrast to actual

temperature measurements, this type of measurements may be called "non-thermal" IR-camera based measurements. Note that "thermal" lock-in thermography is performed on complete electronic devices, whereas "non-thermal" measurements are typically performed on wafers. This section will introduce the physical basics of such measurements; their practical realisation will be described in Sect. 3.5, and typical applications will be introduced in Sect. 6.4.

If we generate excess electrons and holes of density Δn, Δp, respectively, in a silicon wafer, by irradiation with light of an energy $h\nu$ greater than the band gap, they thermalise within picoseconds to a thermal equilibrium with the crystal lattice occupying electronic states above/below their respective band edge. We assume in the following excitation levels which lead to excess carrier densities much smaller than typical doping concentrations of wafers used for solar cells ($10^{16}\,\mathrm{cm}^{-3}$). Then the quasi-Fermi-levels are distant from the band edges (non-degenerate case) and the excess free carriers are distributed over the electronic states given by their respective density of states following a Boltzmann distribution with the temperature T_w of the wafer. If the light irradiation is performed periodically, the weak additional absorption and emission of this excess free carrier distribution can be detected by a thermal camera using the lock-in principle as outlined in Sects. 2.2–2.6. This may be done in a distinct absorption or emission mode, if we select the experimental conditions such that either the absorption or the emission of free carriers dominates. Such a measurement may then be converted to a distribution of the excess carrier lifetime.

As already mentioned, this kind of "non-thermal" investigation is performed on bare wafers, in contrast to the "thermal" lock-in thermography performed on complete electronic devices. Note, however, that in these "non-thermal" investigations also the wafer temperature is modulated (any light absorption also leads to a temperature increase), which leads to a modulation of the thermal IR-radiation even if the free carrier concentration would remain constant. Hence, in reality the IR-camera measures both the "thermal" (temperature-induced) and the "non-thermal" (free carrier-induced) modulation of the device radiation in parallel. It will be described in Sect. 3.5 how these two contributions can be separated from each other.

Historically, single spot infrared absorption of free excess carriers had been used to determine their density and according to

$$\tau = \frac{\Delta n}{G} \tag{2.48}$$

the minority carrier lifetime, where G denotes the generation rate in $\mathrm{cm}^{-3}\,\mathrm{s}^{-1}$. This has been done by methods measuring the decay of carrier concentrations via IR-absorption [47–49] or modulated excitation (Modulated Free Carrier Absorption, MFCA) [50]. This technique was further developed to a setup allowing 2D lifetime mappings [51]. The IR-absorption by free carriers excited in a small spot was measured in this case using a 1,550 nm laser beam transmitting the sample within the generation area. Depending on lateral resolution requirements and wafer size, MFCA maps often took several hours of measurement time per sample. With the

introduction of highly sensitive, fast CCD-cameras, operating in the infrared, the idea to overcome the necessity of scanning the sample was first suggested by Bail et al. [12], developed in parallel by Riepe et al. by implementing lock-in techniques and published in [13].

In this section, we introduce the basic concepts and the two principal measurement modes, absorption and emission of IR-radiation, by free carriers. Calibration and noise considerations will be outlined together with the description of specific measurement setups in Sect. 3.5.

An excess free electron gas in a wafer with temperature T_w interacts with a thermal radiation field by absorption and by emission of photons. If we want to detect excess carriers with a thermal camera, we can in principle make use of both interactions. First, we have to keep in mind that within the detection ranges where suitably sensitive cameras are available, namely the 3–5 (mid-wave) and the 8–10 µm (long-wave) wavelength bands (see Sect. 2.1), there are no pronounced absorption bands of the silicon crystal lattice itself. As mentioned above, a wafer of usual thickness of a few hundred micrometers and low to medium doping concentration is basically transparent to radiation coming from the background. Thus, a big part of the radiation leaving the wafer will be background radiation. If the doping concentration of a wafer is only at about $10^{16}\,\mathrm{cm}^{-3}$, as for typical solar cell material, the free carrier absorption according to (2.47) only weakly influences the background radiation penetrating the wafer. The thermal radiation, which the camera images if focused on a wafer surface, is thus dependent on three factors: (1) the radiation coming from the background, which is influenced by the absorption in the wafer, (2) the free carrier emission in the wafer, and (3) the relation of the temperature of the wafer T_w to that of the background, T_b, which strongly influences the ratio between (1) and (2).

Let us first consider absorption and emission of radiation by free carriers in silicon in some more detail. The radiation transmitted through the sample is proportional to the radiation of the background and the absorption spectrum of the free carriers. In more detail than (2.47), the absorption coefficient of free carriers with density N is given by [52]:

$$\alpha_{fc}(\lambda) = \frac{e^3 N}{4\pi^2 \varepsilon_0 c^3 n\, m^{*2} \mu} \lambda^2, \tag{2.49}$$

where λ is the wavelength, e the elementary charge, ε_0 the real part of the dielectric constant, c the speed of light, n the refractive index, m^* the effective mass, and μ is the mobility. If by periodic irradiation of light the free carrier density N is modulated by an amount ΔN being small compared to N, the steady-state free carrier absorption due to the net doping concentration is modulated by a small amount. If this modulation is detected by lock-in techniques, the steady-state contribution of $\alpha_{fc}(\lambda)$ cancels and only the modulated part due to ΔN is measured. Replacing $\frac{\Delta N}{m^{*2}} = \frac{\Delta n}{m_e^{*2}} + \frac{\Delta p}{m_h^{*2}}$, with $\Delta n = \Delta p$ and m_e^* and m_h^* being the effective masses of electrons and holes, yields the dependence of the free carrier absorption on the modulated excess minority carrier density Δn. Schroder et al. [46] found experimentally a fair agreement of (2.49) and the absorption coefficients for silicon

at room temperature. From (2.49) we note that, in principle, a camera detecting at longer wavelength is of advantage. Whether this advantage can be utilized for CDI/ILM depends on the actual state of the development of IR-cameras for the different wavelength ranges concerning sensitivity and NETD (for an overview of the status in 2009 see Sect. 3.1). Since for thermal lock-in thermography measurements on electronic devices usually mid-wave cameras are preferred owing to their better spatial resolution capability (see Sect. 3.4), these cameras are also often used for non-thermal (carrier density) measurements.

In thermal equilibrium, free carriers also emit infrared radiation. According to the law of detailed balance they have to emit the same amount of power as they absorb at any energy interval to stay at the same temperature. Therefore the absorption coefficient has to equate the emission coefficient (Kirchhoff's law)

$$\alpha_{\mathrm{fc}}(\lambda) = \varepsilon_{\mathrm{fc}}(\lambda). \tag{2.50}$$

If the background with temperature T_{b} can be assumed to emit black body radiation, the number of photons absorbed by free carriers S_{abs} is proportional to the integral of Planck's law (see (2.1)), written for the photon flux density, over the considered range of wavelength which is defined by the spectral range of the camera:

$$S_{\mathrm{abs}} \propto \int_{\lambda_{\mathrm{min}}}^{\lambda_{\mathrm{max}}} \frac{\alpha_{\mathrm{fc}}(\lambda, \Delta n)}{\lambda^4 \left(e^{hc/k\lambda T_{\mathrm{b}}} - 1\right)} \mathrm{d}\lambda. \tag{2.51}$$

The emission by the free carriers depends on their temperature, which is equal to the temperature of the wafer T_{w}, and the emission coefficient of the free carriers is:

$$S_{\mathrm{emi}} \propto \int_{\lambda_{\mathrm{min}}}^{\lambda_{\mathrm{max}}} \frac{\varepsilon_{\mathrm{fc}}(\lambda, \Delta n)}{\lambda^4 \left(e^{hc/k\lambda T_{\mathrm{w}}} - 1\right)} \mathrm{d}\lambda. \tag{2.52}$$

The signal which can be detected by a thermal camera is in general composed of both contributions:

$$\begin{aligned} S_{\mathrm{CDI}} = S_{\mathrm{emi}} - S_{\mathrm{abs}} = d(A_{\mathrm{w}}) \cdot \int_{\lambda_{\mathrm{min}}}^{\lambda_{\mathrm{max}}} \frac{\varepsilon_{\mathrm{fc}}(\lambda, \Delta n)}{\lambda^4 \left(e^{hc/k\lambda T_{\mathrm{w}}} - 1\right)} \mathrm{d}\lambda \\ - d(A_{\mathrm{b}}) \cdot \int_{\lambda_{\mathrm{min}}}^{\lambda_{\mathrm{max}}} \frac{\alpha_{\mathrm{fc}}(\lambda, \Delta n)}{\lambda^4 \left(e^{hc/k\lambda T_{\mathrm{b}}} - 1\right)} \mathrm{d}\lambda, \end{aligned} \tag{2.53}$$

where A_{w} is the surface area on the wafer which corresponds to one camera-pixel and A_{b} is the corresponding surface area on the background. Here d is a positive factor which accounts for the surface dependence of the emission and absorption signal and contains pre-factors of Planck's formula. In a practical setup $A_{\mathrm{w}} \approx A_{\mathrm{b}}$.

With the linear dependence of α_{fc} on Δn and the quadratic dependence on λ from (2.49) and with (2.50), this results in an approximate expression for the camera signal:

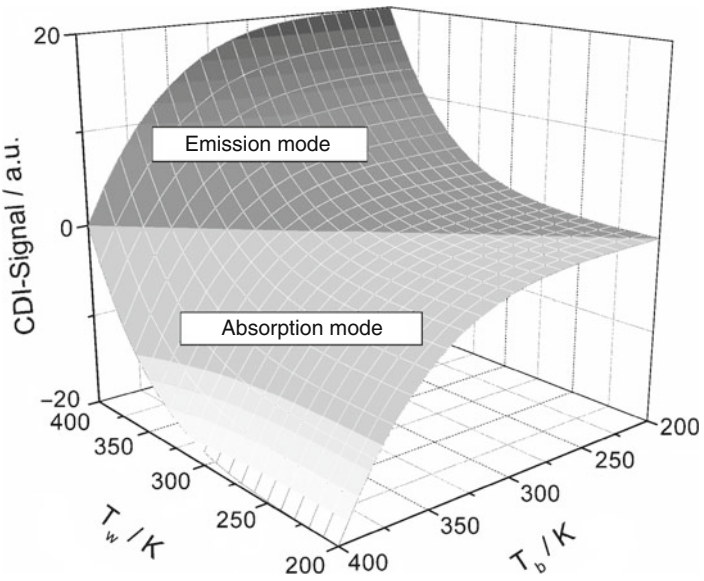

Fig. 2.15 Theoretical dependence of the camera signal on background temperature T_b and wafer temperature T_w according to (2.54)

$$S_\mathrm{CDI} \propto \Delta n \cdot \int_{\lambda_\mathrm{min}}^{\lambda_\mathrm{max}} \left(\frac{1}{\lambda^2 (e^{hc/k\lambda T_\mathrm{w}} - 1)} - \frac{1}{\lambda^2 (e^{hc/k\lambda T_\mathrm{b}} - 1)} \right) \mathrm{d}\lambda . \tag{2.54}$$

In principle, if T_w equals T_b, the absolute value of the absorption signal and the emission signal are equal and a vanishing net signal results. If the background is at a higher temperature than the wafer, an absorption signal can be detected (absorption mode). A cooled background causes a signal which is dominated by the emission of the free carriers, which corresponds to the emission mode. The theoretical dependence of the detected signal on background temperature T_b and wafer temperature T_w is plotted in Fig. 2.15, for an experimental confirmation see Sect. 3.5.

From Fig. 2.15 it is apparent that, in the emission mode, an increase of the wafer temperature T_w results in a strong signal increase, while for lower background temperature T_b the signal increase is asymptotically limited. The signal in the absorption mode, on the other hand, rises strongly with background temperature, while the effect of lowering the wafer temperature is limited.

For the detection of carrier densities the lock-in principle is realised by a chopped optical generation. The thermal radiation coming from the wafer is measured in two different states, first with irradiation of the sample by a light source with $h\nu > E_\mathrm{g}$, generating excess free carriers, and second without any irradiation. If the two images obtained are subtracted from each other, the difference is proportional to the signal related to free carriers, since all other parameters determining the absolute

IR-radiation coming from the sample remain unchanged when switching the generation illumination on/off. This procedure is preferably done with a fast enough sequence of irradiated and unirradiated periods to suppress possible temperature drifts and other changes in the IR background radiation.

Historically, the first realisation of a carrier imaging technique has been to have a black body with emissivity close to 1 and a temperature $T_b > T_w$ behind the wafer [12]. The dominating effect is then that the excess carriers of temperature T_w absorb radiation emitted from the background, the resulting image shows areas with high excess carrier density dark; in areas with low excess carrier density more of the bright background radiation is visible. Figure 2.16, left, shows such an example for a multicrystalline silicon wafer. The sample has been damage etched and SiN-passivated. The displayed area is $100 \times 100\,\mathrm{mm}^2$. The measurement time was 30 s. Compared to the strong background radiation the emission by the colder excess carriers themselves can be neglected.

This "absorption mode" was developed in parallel by two groups, first published in [12] without using the lock-in principle and named "Infrared Lifetime Imaging" (ILM). In a subsequent publication a setup implementing lock-in technique was reported [13, 53] and this variant called "Carrier Density Imaging" (CDI). At first, both terms were used in parallel in literature. Since, however, ILM had included the option to apply the lock-in principle as well and was subsequently developed to this state, later a combined abbreviation was introduced, to clarify, that both terms refer to the same technique. In the following, we will mostly give examples measured with the original "CDI" setup and thus call the technique CDI/ILM.

To use the infrared emission of free carriers instead of their infrared absorption had already been proposed by White et al. [54]. They introduced a scanning method to detect the infrared emission by optically generated free carriers in heated samples. Later Bail et al. [12] used an infrared camera to detect the emission of free carriers.

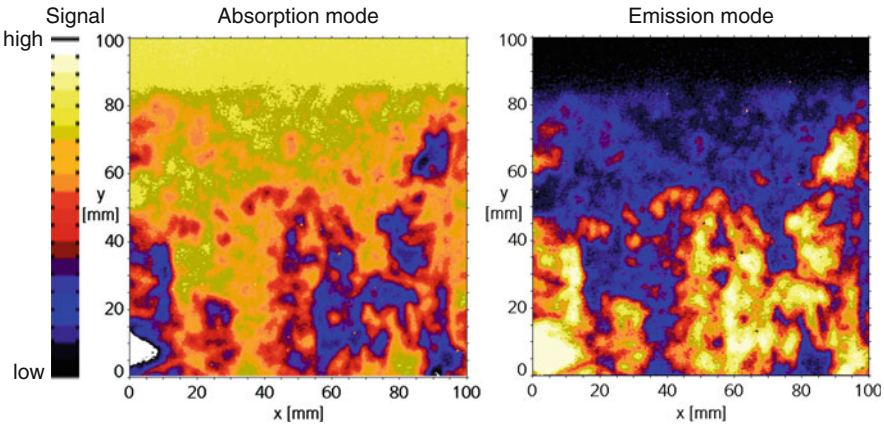

Fig. 2.16 Measured camera image for a surface-passivated multicrystalline wafer in absorption (*left*) and emission mode (*right*)

A high-lifetime FZ wafer was heated with a hot air fan to 350°C to increase the emission of the free carriers and to enable the detection of IR-radiation. Experimental setups which allowed measurements at moderately elevated temperature were introduced in [55] and [56]. In Sect. 3.5 we explain experimental realisations of the different measurement modes in some more detail. For the example of Fig. 2.16, the signal measured in the emission mode is compared to the one measured in the absorption mode (Fig. 2.16, right). As expected, intensities are reversed. Using appropriate calibration (see Sect. 3.5) the excess carrier lifetimes calculated from emission images are in good agreement with the ones calculated from the absorption images. This has been demonstrated e.g. in [55]: The mean lifetime over the total area calculated for two example images was 61 µs for the emission and 65 µs for the absorption measurement. The difference between the absolute lifetime values derived from both techniques was thus only about 6% for this example.

The advantage of emission CDI/ILM is the possibility to increase signal strength by using higher wafer temperatures. As increasing background temperature for absorption CDI/ILM fails because of saturation of the CCD-chip of the camera, the wafer temperature can be raised to moderate temperatures without saturation problems in emission mode since it shows a low emissivity. This allows to retrieve carrier lifetime images with higher signal-to-noise ratio or in a shorter time. If fast measurements are necessary, e.g. for in-line application, the emission mode is an interesting option. Figure 2.17 (courtesy of [57]) shows a comparison of emission CDI measurements realised at different wafer temperatures (40 °C, 59 °C and 100 °C) with a measurement time of only 1 s. The 100 × 100 mm^2 wafer had been etched to remove saw-damage and is SiN_x-passivated. The mean lifetime averaged over the whole wafer was about 18 µs. Noise dominates the measurement at low temperature

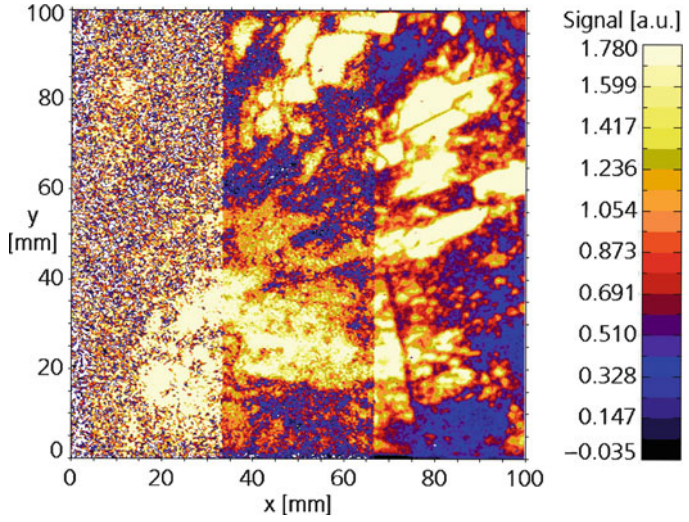

Fig. 2.17 Emission image of a multicrystalline sample with SiN_x-passivation at different wafer temperatures. From left to right: 40 °C, 59 °C, 100 °C. Measurement time was 1 s only

Table 2.1 Quantitative evaluation of the resolution improvement at enhanced temperatures. Over the same homogeneous lifetime area the standard deviation of the signal strength was calculated

Wafer Temperature T_w (°C)	Standard Deviation $\sigma/\sigma_{26°C}$
26	1
40	0.298
59	0.188
100	0.097

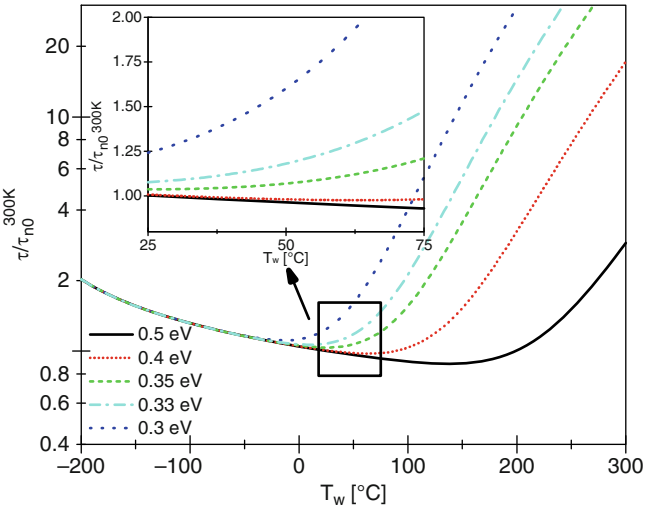

Fig. 2.18 Temperature dependence of minority carrier lifetime limited by Shockley-Read-Hall-recombination for different defect levels. Capture cross sections for electrons and holes are assumed to be equal. τ_{n0}^{300K} is the low injection lifetime for electrons at 300 K. The insert magnifies the most relevant part (*bold square*)

(40 °C), whereas a high signal to noise ratio can be achieved at wafer temperatures of 59 °C and 100 °C.

For a quantitative evaluation of the increase of resolution achieved by an increase of the sample temperature, the variation of the signal from an area of good lifetime was observed for different temperatures (Table 2.1, courtesy of [57]). Compared to the measurement at room temperature the standard deviation at 100 °C reduces to approximately 1/10, already at 60 °C a reduction of more than 1/5 is achieved.

For measurements in the emission mode we have to take into account that the temperature the wafer is heated to may at the same time influence the excess carrier lifetime, which we wish to determine. In Fig. 2.18 (courtesy of [57]) we exemplify the changes in carrier lifetime over the relevant temperature range for recombination limiting defects with defect energy level at various distance ΔE from the conduction or valance band edge calculated according to the Shockley-Read-Hall-recombination mechanism. For shallow defect levels the lifetime varies

significantly, whereas the temperature dependence for deep defect levels is weak in the temperature range under consideration (20–80 °C). Detrimental defects are mostly deep defects, so especially the lifetimes in low performance areas, which are of special interest for material qualification, may be determined at temperatures up to 80 °C and compared with some caution with lifetimes at ambient temperature. Please note in addition, that the temperature of solar cells in operating conditions are mostly above ambient temperature. The measurement of lifetime at these enhanced temperatures allows therefore a reasonable assessment of the expected solar cell performance.