Thermal Failure Analysis by IR Lock-in Thermography

O. Breitenstein, J.P. Rakotoniaina  
Max Planck Institute of Microstructure Physics, Halle, Germany

F. Altmann, T. Riediger  
Fraunhofer Institute for Mechanics of Materials, Halle, Germany

O. Schreer  
Thermosensorik GmbH, Erlangen, Germany

Abstract

The basic limitations of thermal failure analysis by infrared (IR) thermography are that its spatial resolution is limited by the wavelength of the used IR radiation to about 5 µm and that the so-called IR-emissivity contrast usually obscures a weak thermal contrast in ICs. However, advantages of IR thermography against other popular thermal techniques like liquid crystal analysis and fluorescent microthermal imaging are its ease of operation without any surface preparation and the fact that it can also be applied as a backside analysis technique. Thermal IR microscopy has recently experienced a decisive technical improvement by the application of lock-in thermography, which is commercially available also for IC analysis now. This technique allows to display the phase image, which is completely free of the disturbing emissivity contrast. Due to the averaging nature of this lock-in technique local temperature modulations below 0.1 mK can be detected, leading to a detection limit for local heat sources well below 100 µW after 1/2 hr acquisition. The usefulness of lock-in thermography is given by the fact that it allows a relatively coarse (5 µm) but very sensitive localization of any leakage current or other local heat source in an IC with a very high success rate. By this technique some faults can be localized which are not visible e.g. in OBIRCH. In this contribution the technique of microscopic IR lock-in thermography is described, the basic principles of the interpretation of the results are reviewed, and some typical results illustrating the application of this technique are introduced. Ways to overcome the limited spatial resolution are discussed.

Introduction:

Thermal fault detection techniques

Thermal imaging techniques have been used for fault localization in ICs for at least two decades [1, 2]. Some types of local faults, like oxide pinholes, ESD damage, latch-ups, or shorts between neighbouring lines, are connected with local heat sources if a supply voltage is applied to the IC. If the positions of these heat sources can be localized precisely enough, a surface inspection by scanning electron microscopy (SEM) or a focused ion beam (FIB) analysis may reveal the physical origin of the faults. Until now the most popular thermal imaging techniques are liquid crystal microscopy (LCM [1]) and fluorescent microthermal imaging (FMI [2]). Less popular techniques are infrared thermography (IRT [2, 3]) and Schlieren imaging [4]. All these techniques have their inherent advantages and limitations. For performing LCM a light microscope with crossed polarizers is necessary. For the investigation the temperature of the sample has to be precisely stabilized or varied very slowly. Only at a certain well-defined sample temperature local heat sources appear dark in the image. However, also local inhomogeneities of the LC film thickness may appear dark, and especially near bonded wires it is hard to get a homogeneous film thickness, since these structures are attracting the LC material due to capillary forces [5]. So performing this technique and correctly interpreting the results needs a certain degree of experience of the operator. The temperature resolution of LCM is usually given as 0.1 K and the spatial resolution as 1 µm [1]. FMI, on the other hand, is much more straightforward to apply and to interpret, since within an extended temperature range the quenching of the UV-induced luminescence of the dye is predictable [2]. Since the quotient of the image with and without a bias applied to the sample is usually displayed (or its logarithm), also local inhomogeneities of the film thickness are effectively suppressed. By averaging the image data over many images and over a number of neighbouring pixels, a thermal resolution limit of 0.006 K was reported for a spatial resolution of 15 µm [6]. However, the nominal spatial resolution of FMI may be as low as 0.3 µm with a correspondingly increased thermal resolution limit [2]. One problem of FMI is the inevitable UV-induced bleaching of the dye, which limits the usable exposure time and may lead to certain artifacts in the image. Both for LCM and for FMI a foreign layer has to be applied to the surface, which is not always acceptable. Hence, these techniques cannot be used in in-circuit wafer testers. Another limitation of both techniques is that they cannot be used if the chips are mounted face-down (flip chip technology). Backside thermal imaging "through the substrate" is possible both with IR thermography and with Schlieren imaging.
Schlieren imaging relies in the angular deflection of parallel light near a local heat source, which is caused by the temperature dependence of the refractive index of the material. For this technique a thermal resolution limits of 0.01 K and 1 µm spatial resolution have been reported [4]. Schlieren imaging can only be applied with a special optical setup in reflection mode from the backside of an IC, or if the surface is coated by a special diffracting layer. Until now, it has only rarely been used for fault detection in ICs.

Infrared thermography (IRT), which is based on the detection of thermally emitted radiation by an infrared (IR) camera, is the classical thermal imaging technique. Nevertheless, it also has only rarely been used for fault detection in ICs, since it shows two basic limitations. These are its limited spatial resolution and the so-called “emissivity contrast”. The spatial resolution of IRT is diffraction-limited to the order of 5 µm by the IR wavelength range detected (typically 3 ... 5 µm). The emissivity contrast comes from the fact that the intensity of the thermally emitted radiation equals that of a black body at that temperature, multiplied by the value of the so-called IR emissivity \( \varepsilon \) of the surface. The emissivity equals the absorbance of the surface (Kirchhoff’s law) and is generally low (\( \varepsilon < 0.1 \)) for highly reflecting metallized surfaces, hence these regions appear dark in the IR image. Silicon is quite transparent for IR radiation, but at high doping levels free carrier absorption may lead to a considerable IR emissivity of a bare Si surface. So even for an unpowered IC with a homogeneous surface temperature, the IR image is characterized by a strong contrast between metallized and non-metalлизed regions, which is no thermal contrast but “emissivity contrast”. If local heat sources are leading to a thermal contrast, this is superimposed to the emissivity contrast but may remain undetectable for weak power sources. Therefore, this technique has been used mostly for power devices until now [3]. In principle it is possible to generate an image of the local emissivity and to correct the IR image correspondingly [7], but this procedure needs thermostatting of the sample and its success may be limited.

A general problem of all microscopic thermal imaging techniques mentioned above is that they are steady-state techniques. Note that silicon shows a high thermal conductivity. So, if heat is produced locally, the lateral heat conduction instantly leads to a distribution of the increased temperature into the surrounding. This is the reason why microscopic thermal images always appear blurred. So, even if the optical resolution in an FMI experiment is below 1 micron, the effective spatial resolution of this and other steady-state techniques for investigating buried heat sources may be as poor as many microns (see e.g. Fig. 40 in [2]). This situation changes if time-dependent heat sources are considered, since heat conduction is a time-dependent process. If the power is switched on at \( t=0 \), at the beginning only the immediate surrounding of the source heats up. Only if the power persists dissipating, with increasing time after switching it on the heat diffuses more and more away from the heat source position, leading in thermal equilibrium to the blurred appearance of thermograms. Hence, if only the temperature response immediately after switching on the heat source would be observed, this image would be much less influenced by the lateral heat conduction and would appear less blurred. In principle, this is realized in lock-in thermography [8]. Moreover, due to its averaging nature the thermal sensitivity of lock-in thermography may be below 0.1 mK, which paves the way for the thermal investigation of many processes which have remained unidentified before, due to lack of sensitivity. Finally, by displaying the phase image, the emissivity contrast problem is overcome, hence even weak heat sources lying below metal layers become visible. Lock-in thermography is very easy to apply, and the results are straightforward to interpret. Thus, in spite of its limited spatial resolution of about 5 µm, lock-in thermography is a promising new tool for thermal failure analysis.

In the following the basic principles and the technique of lock-in thermography are described. Then the question of the real spatial resolution of this technique is discussed. A number of typical results are introduced, which demonstrate the wide applicability of this technique. Finally, different ways to overcome the limited spatial resolution are discussed and new technical developments of lock-in thermography are briefly mentioned.

### Lock-in thermography technique

Lock-in thermography was invented in 1984 [9] and has been used extensively in non-destructive testing for “looking below the surface” of solid objects [10]. Meanwhile it represents also a standard technique for investigating shunting phenomena in solar cell research [11]. Lock-in thermography means that the power dissipated in the object under investigation is periodically amplitude-modulated, the resulting surface temperature modulation is imaged by a thermocamera running with a certain frame rate \( f_s \), and that the generated IR images are digitally processed according to the lock-in principle. Thus, the effect of lock-in thermography is the same as if each pixel of the IR image would be connected with a two-phase lock-in amplifier. Consequently, the two primary results of lock-in thermography are the image of the in-phase signal \( S^0(x,y) \) and that of the out-of-phase (or quadrature) signal \( S^{-90°}(x,y) \). In lock-in thermography often the -90° signal is used instead of the +90° one, since the latter is essentially negative [8]. From these two signals the image of the phase-independent amplitude \( A(x,y) \) and the phase image \( \Phi(x,y) \) of the surface temperature modulation can easily be derived:

\[
A(x,y) = \sqrt{S^0(x,y)^2 + S^{-90°}(x,y)^2} \tag{1}
\]

\[
\Phi(x,y) = \arctan\left(\frac{-S^{-90°}(x,y)}{S^0(x,y)}\right) \tag{2}
\]

Note that for \( \Phi \) the quadrant-correct arctan function has to be used, hence if \( S^0(x,y) \) is negative, 180° have to be subtracted from the pure arctan value. Both the 0°- and the -90°-image are proportional to the power of the pulsed heat source,
Therefore also in the amplitude image the contrast of a heat source is proportional to its dissipated power. The phase image, on the other hand, relies on the quotient of the 0°- and the -90°-image, hence it should be independent on the power of the heat source. In fact, the phase image is a measure of the time delay of the surface temperature modulation referred to the power modulation, which is indeed independent from the magnitude of the modulated power, as long as we have no superposition of the temperature fields of neighboured heat sources. Another property of the phase image is that it is also independent from the IR emissivity of the surface ε! Note that ε influences both S0° and S-90° in the same way, so according to (2) the quotient of both should be independent from ε. This property is widely used also for the application of lock-in thermography in non-destructive testing [10]. For the application of lock-in thermography in thermal failure analysis this property causes the effect that the phase image shows no emissivity contrast. Moreover, it implies a kind of “dynamic compression” so that local heat sources with different powers are displayed with a similar signal height. It will be shown below that these properties greatly simplify the interpretation of the results. Note, however, that in metallized regions having a low value of ε also the signal-to-noise ratios of both primary images decrease, hence metallized regions may appear more noisy in the phase image than non-metallized ones.

Fig. 1 shows the functional scheme of the lock-in thermography system TDL 384 M 'Lock-in’ of Thermosensorik GmbH (Erlangen, Germany [12, 13]), which is the only system on the market until now which is specialized to the microscopic investigation of electronic components. It is equipped with a Stirling-cooled cadmium-mercury-telluride (CMT) focal plane array (FPA) IR detector head, which is sensitive in the 3 ... 5 μm wavelength range and at 300 K object temperature it exhibits a noise level of about 20 mK at a frame rate of 130 Hz. The resolution of this camera is 384x288 pixel. Different IR lenses are available, the highest magnifying one leading to a pixel resolution of 5 μm, which is already close to the optical diffraction limit. The digital signal of the detector head is captured by a frame grabber board and cyclically written into a certain part of the memory (RAM) of the PC by direct memory access (DMA). The software picks up the data there and performs the lock-in correlation, the data display, and the system control. A hardware counter, which is controlled by the image trigger of the free-running IR detector head, derives the bias pulse trigger by dividing down the image trigger. The larger this division factor the lower the lock-in frequency is. Since at least 4 images are needed to perform the 2-channel lock-in correlation procedure, the highest possible lock-in frequency at full frame size is about 30 Hz. If a reduced frame size is selected, the frame rate of the camera increases and a higher lock-in frequency may be applied. As a special option so-called undersampling is possible, which allows to select at full frame size a lock-in frequency above 2 kHz, which will be discussed in the last chapter of this contribution. More details about the technique and application of lock-in thermography for investigating electronic components can be found in [8].

The effective spatial resolution

If there are physically determined blurring mechanisms like the lateral heat conduction for lock-in thermography, even in absence of optical limitations the pixel distance does not equal the effective spatial resolution of the investigation. Here we define the effective spatial resolution as the minimum distance up to which a point heat source can be localized or up to which a sharp edge of an extended heat source can be determined. In absence of Peltier effects, electrical heat dissipation always leads to an increase of the temperature of the object under test. So, if periodic bias pulses are applied to a sample, at the beginning of the experiment its mean temperature will slowly increase until equilibrium between electric heating and heat dissipation to the environment has established. In this “quasi equilibrium state” periodic heating superimposes to permanent cooling to the surrounding. For small ICs the time constant to reach this quasi equilibrium state is in the order of seconds, hence for a typical lock-in measurement lasting many minutes the sample is in this state for most of the time. Then it is possible to describe the pulsed heat sources as harmonically oscillating heat sources implying subsequent heating and cooling periods (thermal wave approach). The non-steady-state heat diffusion equation in a 3-dimensional isotropic and homogeneous solid is:

$$c_p \rho \frac{\partial T}{\partial t} = \lambda \frac{\partial^2 T}{\partial r^2}$$

(3)

Its solution for a harmonically oscillating point heat source with the power $P(t) = P_0 \sin(\omega t)$ is:

$$T(r, t) = \frac{A}{r} \exp\left(-\frac{r}{\Lambda}\right) \exp\left(i \omega (t - \frac{r}{\Lambda})\right)$$

(4)

with $A = \frac{P_0}{2 \pi \lambda}$ and $\Lambda = \sqrt{\frac{2 \lambda}{\rho c_p \omega}}$

(5)
Here \( r \) is the radial distance to the point source, \( \omega \) is \( 2\pi f_{\text{lock-in}} \), \( \lambda \) is the heat conductivity of the material, \( c_p \) is its specific heat, \( \rho \) is its density, \( A \) is an amplitude factor, and \( \Lambda \) is the so-called thermal diffusion length. According to (5) \( \Lambda \) depends on the properties of the material and reduces with \( 1/\sqrt{f_{\text{lock-in}}} \). For silicon \( \Lambda \approx 1 \text{ mm} \) for \( f_{\text{lock-in}} = 30 \text{ Hz} \). The decisive point is that \( \Lambda \) does not represent a limit for the spatial resolution if the point heat source is close to the surface. According to (4) the T-modulation amplitude diverges in source position and then drops with \( 1/r \). Of course, in reality there is no divergence of the thermal signal for two reasons: Any heat source has a finite extension and the signal is always averaged across one detector pixel. But as long as the extension of the heat source is small against the pixel resolution, the model of a point source is appropriate to describe the situation. So, if a point heat source would be at the surface of the sample, it would appear "sharp" in the lock-in thermogram, even if \( \Lambda \) is large against the pixel distance. Of course, there will be a "halo" around this peak, the extension of which will depend in \( \Lambda \). Fig. 2 shows a simulation (frequency 30 Hz; \( \Lambda \approx 1 \text{ mm} \), spatial resolution: 5 \( \mu \text{m} \)) of the amplitude signal of a point heat source in the center (a) and of an extended homogeneous area of power dissipation to the right with its edge in the center of the trace (b). Both heat sources are assumed to lie at the surface of a silicon sample, which is thick against the image dimensions. We see that the effective spatial resolution of lock-in thermography strongly depends on the geometry of a heat source. While the signal maximum of a point-like heat source is really sharp with a half maximum width of 7 \( \mu \text{m} \), the edge of the extended heat source is only hardly recognizable. The physical reason for this effect is that at the edge of a sufficiently large laterally extended heat source plane thermal waves are running horizontally, whose amplitude do not diverge anymore in source position but decays exponentially with the thermal diffusion length \( \Lambda \). Hence, for this source geometry \( \Lambda \) really represents a limit for the spatial resolution of lock-in thermography.

This section may be summarized as follows:
1. If a small (point-like) heat source is lying at or close to the surface, it can be sharply localized thermally up to an accuracy of one image pixel, nearly independent from the thermal diffusion length \( \Lambda \).
2. However, the extension of the halo around the signal maximum, which is due to lateral heat conduction, depends on the value of \( \Lambda \).
3. The more extended a heat source is, the more blurred it appears in the lock-in thermogram, and the more the effective image resolution will depend on \( \Lambda \). Small heat sources generally appear sharper than extended ones. So, for accurately imaging extended heat sources choosing a higher lock-in frequency (leading to a lower value of \( \Lambda \) at the cost of a lower signal height) is advantageous.
4. If heat sources are lying not at the surface but at a certain depth below the surface, even for point-like heat sources the depth below the surface governs the effective spatial resolution.

### Examples

In the first example an IC is examined, which failed with raised power consumption. By lock-in thermography a temperature difference image (amplitude image) of the whole chip surface was taken. The amplitude of the pulsed supply...
voltage was 5 V and the current flow had a value of 90 µA. The measurement was taken over a time period of 8 min with an lock-in frequency of 24 Hz. By zooming in the region with the thermal emission spot the defect position located precisely enough for the following physical failure analysis (Fig. 3). By TEM analysis starting with a FIB-prepared cross section on 1.8 µm thickness at the thermal spot position a gate oxide (GOX) breakdown was found (Fig. 4). The GOX breakdown has a diameter of about 100 nm and has melted regions in the poly Si above and the Si substrate underneath.

The next example is a IC, which was affected by electrically damages that occurred by its processing due to electrostatic discharges (ESD) in a direct way to the chip surface. The specific problem of this measurement was a very weak thermal emission of the flaw, which was over-radiated by the local halo of an area with high power consumption. With the standard lock-in frequencies of the TDL 384 M system of up to 30 Hz (in consequence of the frame rate of the IR camera) this halo could not be reduced appropriately. However with the new undersampling option of the TDL system it was possible to set the lock-in frequency to much higher values. The acquisition of one single image is here specially timed and expanded over several periods. Due to the high measuring frequency of 176 Hz and the weak failure signal a acquisition time of 2 h was necessary (Fig. 5, 6).

Fig. 4: TEM images of the found gate oxide (GOX) breakdown (Top – overview image, Bottom – detail image)

Fig. 5: Top – amplitude image with 15 Hz (the halo of the main thermal radiation source masks the failure spot), Middle – amplitude image with 176 Hz (a little second spot is now visible), Bottom – topography image with superposed amplitude spots

Fig. 6: SEM image of the cross sectional area (prepared by FIB) which shows the conductor layers; the ESD stroke has cracked the above passivation layer and the discharge channel has filled with melted metal of the two conductor layers and connected them electrically (Image by EMPA Duebendorf)
The following example shows a substrate of a multichip module (MCM) (Fig. 7, 8). The MCM consists of two ICs which are connected electrically together by a substrate. The substrate contains three redistribution Cu-layers in the package. The MCM failed after 480 h HTOL (High Temperature Operating Life) test. By electrical analysis a low resistive short in the distribution system between two connection pins was found. Due to the optical transparency of the substrate down to the infrared light range, after mechanical dissection it was possible to detect thermal emissions on any layer of it. So the failure found on an inner Cu-layer as a thin path of Cu between two adjacent conduction lines. The measurement was taken with a lock-in frequency of 20 Hz and a supply voltage of 1.5 V which caused a current of 0.3 mA.

Testing digital ICs is also possible by supplying them with a permanently applied supply voltage. In this case only the voltage on the control inputs is triggered by the lock-in clock. Concerning to this, permanently existing heat sources in the IC, which are not affected by the trigger signal, do not appear in the lock-in thermogram. Only heat sources affected by the trigger signal are detected here. So in the phase image can be distinguished, if a heat source is emitting in on-state or in off-state of the trigger (positive or negative response). Fig. 9a, b show an example where some complementary acting heat sources are present. The investigated IC contained some power transistors in the centre and some control logic around to drive them. In the phase image the bright regions correspond to positive responding heat sources and the dark regions correspond to the inversely reacting heat sources.

Fig. 7: Top – amplitude image, focussed in the depth of the substrate; Bottom – IR topography image with superposed amplitude spot

Fig. 8: Light microscopically images, taken after milling the substrate back to the appropriate layer

Fig. 9a: Topographical image of a device under functional test (Phase image in Fig. 9b)
Summary and outlook

Lock-in IR thermography, which already is established as a standard technique in solar cell research, has the potential to enter the field of IC failure analysis. In comparison with the most popular thermal failure analysis techniques liquid crystal microscopy (LCM) and fluorescence microthermal imaging (FMI) it shows the following advantages:

1. It requires no foreign layer at the surface, hence it also may be used on a wafer scale.
2. The measurement procedure and the interpretation of the results are very straightforward: One needs no thermostating and no shading of the setup, there is no degradation, and heat sources appear simply bright in the images.
3. It also may be used as a backside analysis technique.
4. Due to the dynamic nature of the measurement lateral heat conduction is considerably suppressed. Hence, the effective spatial resolution is improved compared to steady-state thermal imaging techniques.
5. The thermal sensitivity may be below 100 µK (depending on the measure time and the F# number of the IR lens), which is 1-2 orders of magnitude below that of previous techniques. Thus, it can be applied to investigate also weak heat sources, which have been unaccessible by thermal methods previously.

One of the general problems of IR microscopy, the IR emissivity contrast artifact, is overcome in lock-in thermography by displaying the phase image, which shows no emissivity contrast. Thanks to the "undersampling option" recently developed for the Thermosensorik TDL 384 M 'Lock-in' thermography system the lock-in frequency can be chosen above 2 kHz [14], which also for extended heat sources brings the effective spatial resolution more close to the best possible pixel resolution of 5 µm. However, lock-in thermography still shows the following limitations:

1. The optical spatial resolution is diffraction-limited by the IR wavelength used to the order of 5 µm.
2. Since metallized layers show a weak IR emissivity, the signal-to-noise ratio and thus the thermal sensitivity is considerably degraded in metallized regions.

Limitation (1) is the most serious one. On the one hand, according to our experiences already the localization of a fault to an accuracy of 5 µm is very helpful for failure analysis. In many cases knowing the layout already gives an indication where exactly the fault could be. On the other hand, there are other better resolving techniques available like FMI. This technique can also be used in the form of lock-in FMI [13], which again improves its effective spatial resolution for extended heat sources. Then lock-in IR thermography can be used for a coarse localization of heat sources, and a sub-micron localization may be performed on demand afterwards e.g. by lock-in FMI.

Limitation (2) may be overcome by covering the surface with a sufficiently thin and structureless black layer. We have found that colloidal bismuth evaporated a a residual air pressure of 10⁻² mbar is appropriate for this purpose [14]. Of course, this measure removes advantage (1).

As a whole, lock-in IR thermography shows a number of advantages against previous thermal imaging techniques, which makes it a promising alternative to them. With this article we hope to support the introduction of this technique into the field of thermal failure analysis.

Acknowledgements

The authors like to thank Achim Lindner from Micronas GmbH Freiburg / Germany and Jürgen Schulz from Melexis GmbH Erfurt for their support for this work. Also we like to thank Peter Jacob from EMPA Duebendorf and EM Microelectronic Marin SA / Switzerland for providing samples and FIB images.

References

4. www.optomet.com
5. J. Schulz, MELEXIS GmbH (Erfurt), personal communication.
7. www.quantumfocus.com
12. www.thermosensorik.com