PERGAMON

Microelectronics Reliability 42 (2002) 653–667

MICROELECTRONICS RELIABILITY

www.elsevier.com/locate/microrel

Selected failure mechanisms of modern power modules

Mauro Ciappa

Physical Characterization Group, Integrated Systems Laboratory, Swiss Federal Institute of Technology (ETH), CH-8092 Zurich, Switzerland

Received 31 January 2002

Abstract

This paper reviews the main failure mechanisms occurring in modern power modules paying special attention to insulated gate bipolar transistor devices for high-power applications. This compendium provides the main failure modes, the physical or chemical processes that lead to the failure, and reports some major technological countermeasures, which are used for realizing the very stringent reliability requirements imposed in particular by the electrical traction applications. © 2002 Elsevier Science Ltd. All rights reserved.

1. Introduction

Thanks to the recent improvements made in handling large currents at high voltage and at high switching frequencies, insulated gate bipolar transistors (IGBT) have almost completely replaced bipolar power transistors (BPT) and they are challenging the position of gate turn-off thyristors (GTO) in their traditional fields of application. In the last five years, the need of increase the reliability of high-power IGBT multichip modules has been one of the most powerful drivers that forced engineers to design new products, especially intended for traction, for power transmission, and for power distribution applications. In order to cope this demand, various wide-band research projects, which involved both industry and academia, have been promoted on an international base. This is for example the case of the high-power semiconductors for railway traction applications (RAPSDRA) project and of the high-temperature IGBT- and MOSFET-modules for railway and automotive electronic applications (HIMRATE) project, both funded by the European Union. One among the most relevant goals of such research programs is the definition of realistic reliability requirements based on dedicated mission profiles. These studies confirmed that the reliability figures of merit required from power devices for railway traction are much more stringent than those needed by the components intended for the usual microelectronic applications. As an example, the useful lifetime of an IGBT module for railway traction is specified in at least 30 years, while the failure rate of every single IGBT module is not allowed to exceed 100 FIT. Under previous assumptions, the traditional reliability growth programs based on a posteriori failure rate assessment procedures are unuseful, as it is already well known from similar experiences with ULSI devices. In fact, even a simple characterization of the failure rate as function of the time with a reasonable level of confidence, would require many millions of cumulated component hours, what is almost impracticable in the case of complex devices and in particular for power modules (even at accelerated conditions). The strategy used for circumventing this limit bases on the well-known concept of built-in reliability. The main idea behind this approach is the continuous control of those process parameters, which may affect the reliability of the final product. This initial phase is followed by dedicated experiments and characterizations, which are intended to investigate the system response over the variation of a given parameter. Finally, the obtained information is returned into a feedback loop for finely tuning the process conditions. A very important step when realizing this strategy is the characterization and the classification of the failure mechanisms, which occur either during accelerated tests or in field applications. While wearout failure mechanisms can be attacked by adequate design

E-mail address: ciappa@mail.iis.ee.ethz.ch (M. Ciappa).

0026-2714/02/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0026-2714(02)00042-2

rules (and represent essentially a cost optimization problem), random failures are not necessarily related to a dominant failure mechanism. In fact, they express the random character of both the occurrence of physical processes and of the quality of manufacturing processes. Nevertheless, random failures play a very relevant role in determining the survival probability of a mature system.

In the following, we will shortly review the most frequent failure mechanisms observed to affect mainly IGBT modules. However, since the majority of the failure mechanisms listed here are package related and are driven by thermo-mechanical stresses, they can also be encountered in power modules using devices (e.g. power MOSFETs or diodes), which exhibit high densities of power dissipation.

2. Module architectures

Multichip modules for high-power IGBT devices are complex multilayered structures consisting of different materials, which have to provide a good mechanical stability, good electrical insulation properties, and good thermal conduction capabilities. The schematic cross-section through a module of type A (e.g. a standardized E2 package) is represented in Fig. 1a, and the related physical parameters are listed in Table 1. Starting from the bottom one can recognize the base plate, the direct copper bonded ceramic substrate, the silicon chip, the aluminum metalization (not shown), and the bond wire.

A type B module is schematically sketched in Fig. 1b. The most important structural difference with the type A module is the use of a strain buffer layer soldered on the top of the IGBT chip (layers a' and b' in Fig. 1b).

In both module types there are additional compliant layers, which are placed at the interface of materials with large differences in thermal expansion. This is the case of the solder layers.

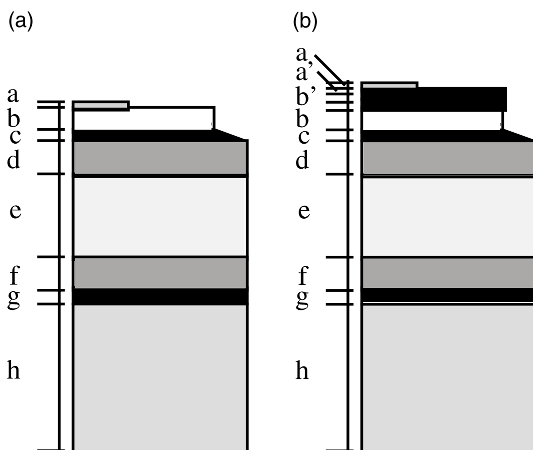

Fig. 1. Schematic representation of the multilayer in a multichip module of type A (a) and of type B (b).

Table 1
Thickness (t), CTE, typical length (L)

Material	t (μm)	CTE (ppm/°C)	L (mm)
Al	300	22	1
Si	250	3	12
Solder	100	Compliant	
Cu	280	Not relevant	
Al_2O_3 or AlN	1000	7 or 4	30–55
Cu	280	Not relevant	
Solder	180	Compliant	
Cu or AlSiC	4000	17 or 8	

Recently, a new generation of power modules has been developed, which does not make use of any base plate and, whose ceramic substrate (AlN) is directly mounted on the heat sink. The only solder layer in the module is the die attach. All other thermal and electrical contacts are established by mechanical pressure, such that the entire construction does not exhibit any bond wire. Even if in the past electrical conductive spring contacts have been proven to be highly rugged and to provide good long term reliability properties, at present, neither quantitative data about the lifetime, nor precise information about the failure mechanisms occurring in this technology are available. For this reason this new module architecture will not be considered in the following. This is also the case of the more traditional press-packaged devices.

3. Materials and thermomechanics

When considering thermal cycling of these multilayered structures and the consequent thermo-mechanical fatigue induced failure mechanisms, it is important to take into account all the factors, which play a role in determining thermo-mechanical stresses.

In first approximation, there are the mismatch in the coefficient of thermal expansion (CTE), the characteristic length of the layer, and the local temperature swing. The first two parameters for the relevant materials and layers in a multichip module are listed in Table 1. In Fig. 2a we represent the computed (one-dimensional approximation) relative temperature swing at different interfaces within two modules having the same geometry (as in Table 1), but with layers made of different materials. In particular, the first module includes a ceramic substrate of Al_2O_3 and a base plate of copper, while the ceramic substrate of the second module is AlN and the base plate AlSiC.

From Fig. 2b, it can be seen that, due to the good conductivity of AlN, the maximum temperature swing in the multilayer of type B is about 50% lower than in the

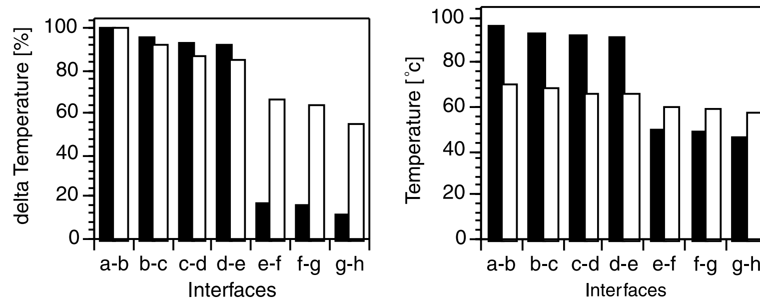

Fig. 2. (a) Temperature swing distribution at the interfaces of both stacks in Fig. 1 (black: type A, white: type B); a–b: Si, b–c: Si–solder, c–d: solder–Cu, d–e: Cu–ceramic, e–f: ceramic–Cu, f–g: Cu–solder, g–h: solder–base plate. (b) Temperature at the interfaces for a dissipated power of 100 W and a heat sink temperature of 40 °C (black: type A, white: type B).

Table 2
Differential elongation at the conditions of Fig. 2

Type A	
Si–bond wire	2 on 1000 μm
Si–Al_2O_3	4 on 12,000 μm
Al_2O_3–copper	28 on 55,000 μm
Type B	
Si–bond wire	0 on 1000 μm
Si–AlN	1 on 12,000 μm
AlN–AlSiC	7 on 30,000 μm

stack of type A. In converse, it can be seen that the most relevant temperature drop (about 80%) within the first multilayer occurs across the Al_2O_3 ceramic substrate. In the second multilayer, the largest temperature drop virtually occurs across the AlSiC base plate and the isothermal heat sink, such that all interfaces experience almost the full temperature swing.

From Table 2, it can be seen that the largest difference in CTE affects the aluminum (bond wires, metalization) and the silicon chip. In the first multilayer, the mismatch is worsened by the fact that both materials are in intimate contact. On the contrary, the strain buffer used in the second stack, which consists of an aluminum layer bonded onto a molybdenum plate (CTE 2.5 ppm/°C), dramatically reduces the thermo-mechanical stresses experienced by aluminum bond wires. On the second and third place in terms of CTE mismatch, one can mention the ceramic substrate and the base plate (especially the combination of Al_2O_3 and copper), and the silicon and the ceramic base plate (especially the combination of Silicon and Al_2O_3), respectively. Compliant solder layers separate the last two couples of materials.

Since the smaller is the lateral size of a layer, the smaller is the thermo-mechanical stress, several small-sized structures are preferred instead of single large plate (especially for brittle materials). Unfortunately, the size of almost all components is determined by physical constraints (e.g. size of the IGBT chip). The only degree of freedom concerns the ceramic substrate; such that in advanced packages it is partitioned in squares with a side length, which ranges from 30 to 55 mm. Small-sized ceramic substrates are normally used when combining AlN with copper base plates. This is also due to cost optimization reasons (Fig. 3).

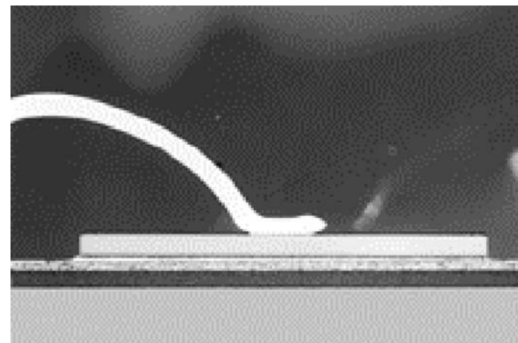

Fig. 3. Cross-section through the gate bond wire of an IGBT module of type A (cross-section, optical image 3×).

Table 2 summarizes the combined effect of the CTE mismatch, temperature swing, and size in the case of the stacks represented in Fig. 1a and b. The differential elongations, which have been computed according to the one dimensional approximation for a heat sink temperature of 40 °C and a dissipated power of 100 W, clearly identify the critical interfaces of both types of multichip module.

4. Bond wire fatigue

Multichip IGBT modules for high-power applications typically include up to 800 wedge bonds (Fig. 3). Since about half of them are bonded onto the active area of semiconductor devices (IGBT and freewheeling diodes), they are exposed to almost the full temperature swing imposed both by the power dissipation in the silicon and

by the ohmic self-heating of the wire itself. Emitter bond wires are usually 300–500 μm in diameter. The chemical composition of the wire can be different from manufacturer to manufacturer, however in all cases, the pure aluminum is hardened by adding some few thousand parts per million of alloying elements, such as silicon and magnesium, or nickel for corrosion control. The current capability of a bond wire decreases as well-known overproportionally with the length and just slightly depends on the substrate temperature. The maximum DC current capability of a bond wire is limited by melting due to ohmic self-heating. In a 1 cm long wire loop in air it is of 25 A for 300 μm (35 kA/cm^2) and of 60 A for 500 μm (30 kA/cm^2) aluminum wires. Under normal operating conditions the current within a single aluminum bond wire does not exceed 10 A, such that the maximum ohmic power dissipation is between 100 and 400 mW, depending on the wire diameter. During switching operation the current density distribution across the section of a bond wire is strongly inhomogeneous due to the skin effect. The wires are connected by ultrasonic wedge bonding either onto the aluminum metalization (with a thickness ranging from 3 to 5 μm), or onto the strain buffer.

In Fig. 4 is represented a cross-section through a wedge bond, before thermo-mechanical stress. The arrow indicates the transition from the non-bonded to the welded region, where the bond wire material cannot be distinguished from the aluminum metalization.

Failure of a wire bond occurs predominantly as a result of fatigue caused either by shear stresses generated between the bond pad and the wire, or by repeated flexure of the wire. The failure of a single or of multiple bond wires causes a change either in the contact resistance or in the internal distribution of the current, such that it can be traced by monitoring $V_{\mathrm{CE\,sat}}$ [1]. The observed failure mode can be different depending on the stress the devices are submitted. If the test is not interrupted after exceeding a pre-defined threshold, the end of life failure mode observed during power cycles is melting of the survivors bond wires. On the contrary, during high-voltage test or field operation, a frequently observed secondary failure mechanism is the triggering of parasitics.

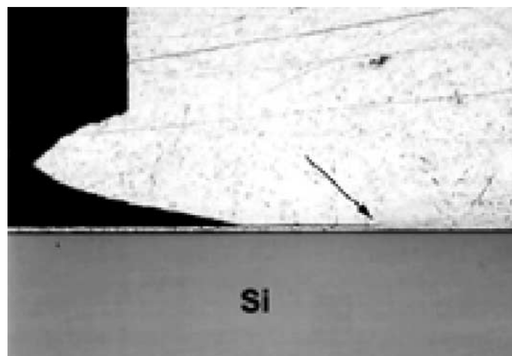

Fig. 4. Cross-section of a virgin wedge bond (tail side) on aluminum metalization, showing the transition to the interdiffused region (optical image, 120×).

4.1. Bond wire lift off

Bond wire lift off has been observed to affect both IGBT and freewheeling diodes. However, since power cycling experiments are usually performed with unipolar current sources, these last are often ignored. No bond wire lift off occurs at the wire terminations bonded onto copper lines. This is mainly due to the fact that copper lines do not experience large temperature swings. Additionally, the CTE mismatch between aluminum and copper is less severe than with silicon.

The fracture mechanics at bonded interfaces and the modeling of the crack propagation within the welded joint with time is a quite complex issue. There is experimental evidence that the crack leading to the failure is initiated at the tail of the bond wire (Fig. 4), and propagates within the wire material until the bond wire completely lifts off.

Polycrystalline metals exhibit yield strength. If the stresses exceed this value flow is very rapid. If they are below, the compliant behavior of the material depends on the amplitude of the applied stress and on the time. The kinetics of the flow process is controlled by effects at atomic scale, like the glide motion of dislocations and the diffusive flow of individual atoms [2]. Several attempts have been made to estimate the operating lifetime of bond wires by numerical simulation basing either on continuous mechanics models [3] or on quasi-atomistic models including grain boundaries [4]. Generally, the quantitative use of simulations is limited by the complexity of the three-dimensional structure of the bond, and due to the uncertainty in evaluating the initial stresses induced by the strong deformation of the wire through the bonding tool. This is the reason why the reliability of different types of solid-state welded contacts is still investigated experimentally.

Fig. 5a shows a bond wire after lift off. Due to the spring action exerted by the aluminum wire loop, the wire looses the electrical continuity with the IGBT chip.

The close up into a footprint of a lifted bond wire in Fig. 5b, clearly indicates that the crack propagates within the wire material and not at the interface as it would be the case either of poor welding, or of delamination of the metalization layer. Furthermore, it can be seen, that welding just occurs at the periphery of the joint, while in the central region, the wire is not in contact with the metalization, as it can be deduced from the occurrence of reconstruction.

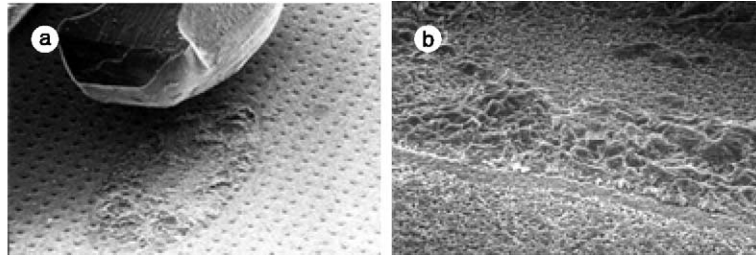

Fig. 5. (a) Bond wire lift off (SEM image, 40×). (b) Close view of the footprint of an aluminum bond wire after lift off (SEM image, 100×).

At present, two main technological countermeasures are common for facing the bond wire lift off failure mechanism. The first one makes use molybdenum–aluminum strain buffers [5], which are mounted on the top of the IGBT and of the diode chips, with the scope to eliminate thermo-mechanical fatigue by distributing the CTE mismatch of aluminum and silicon across a thick layer (Fig. 1b). The second solution is a symptomatic corrective action, which aims to avoid the physical separation of the wire from the bond pad, once the welding joint fails. This scope is achieved by gluing the bond wires with a coating layer. The coating consists of one or of multiple polymeric layers with graded hardness, which are painted onto the wires immediately after ultrasonic bonding. Fig. 6 reports the results of a very early experiment [6], where the efficiency of polymeric coatings in slowing down the consequences of the bond wire lift off is clearly shown. Alternative solutions like direct chip cooling for quenching large temperature swings at the chip surface have been demonstrated to be poorly effective.

The number of thermal cycles-to-failure N_f is usually modelled by using the simple bimetallic approach as an approximation for the thermo-mechanical stresses arising at the interface of a joint between aluminum and silicon when submitted to a temperature swing ΔT. This yields

$$\varepsilon_{tot} = L(\alpha_{Al} - \alpha_{Si})\Delta T \quad (1)$$

where α_{Al} and α_{Si} are the thermal expansion coefficient of aluminum and silicon, respectively, and L the typical length of the joint. Due to the large thermo-mechanical mismatch, the joint is operated in deep plastic regime. Thus, it can be reasonably assumed that the full strain is mainly given by the plastic strain

$$\varepsilon_{tot} = \varepsilon_{elastic} + \varepsilon_{plastic} \approx \varepsilon_{plastic} \quad (2)$$

Having the plastic strain, the mean number of cycles-to-failure N_f can be estimated by the Coffin–Manson law

$$N_f \propto \varepsilon_{plastic}^{-n} \quad (3)$$

where the exponent n is a positive number. After insertion of Eq. (1) into Eq. (3), we obtain

$$N_f = a(\Delta T)^{-n} \quad (4)$$

The coefficients a and n are sometimes estimated by numerical simulation, however they are more often provided by experimental measurements based either on thermal or on power cycling of real devices. This simple model has been shown to accurately predict the fatigue behavior of bond wires for thermal cycles, whose peak temperature does not exceed 120 °C. For higher temperature ranges, alternative models are necessary, which take into account the deviation from the power law. Furthermore, the computation of the lifetime of modules including a large amount of bond wires has also to consider the redundancy structure of the system [7].

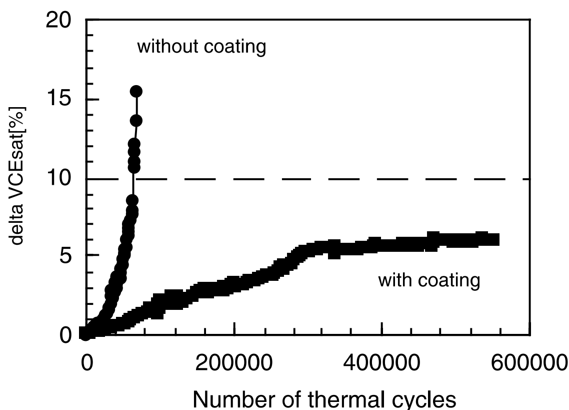

Fig. 6. Degradation of $V_{CE\,sat}$ in an IGBT module without and with a polymeric bond wire coating layer; $T_l = 65$ °C, $T_h = 125$ °C, $t_{on} = 0.8$ s, duty cycle 0.5.

4.2. Bond wire heel cracking

Bond wire heel cracking rarely occurs in advanced IGBT multichip modules. However, it can be observed mainly after long endurance tests and especially in cases where the ultrasonic bonding process was not optimized.

The failure mechanism is due again to a thermo-mechanical effect. In fact, when the wire is subjected to temperature cycles it expands and it contracts undergoing flexure fatigue. In the case of a typical bond wire length of 1 cm and of a temperature swing of 50 °C, the displacement at the top of the loop can be in the 10 μm range producing a change in the bending angle at the heel of about 0.05 °. An additional stress is introduced by the fast displacement of the bond wire (e.g. at the turn on) within the silicone gel, which can be considered as a very viscous fluid.

In those cases, where the temperature change within the bond wire is dominated by the ohmic self-heating, heel cracking can also be observed at the wire terminations welded on the copper lines of both IGBT chips and freewheeling diodes.

Figs. 7 a and 8a show two examples of heel cracking at a single and at a double bond, respectively. In the first case heel cracking and bond wire lift-off occur at the same time. However, while the adjacent bond has been completely removed by lift-off, the cracked wire still presents some electrical continuity with the chip. This is a clear indication of fact that even if the bonding parameters are not too close to the optimum, heel cracking is slower than the lift-off mechanism.

The couple of wires in Fig. 8a indicates that heel cracking preferably occurs at those locations where the aluminum wire has been previously damaged by the bonding tool. In fact, the bond wire at the left in Fig. 8a presents a thin crack at the same location where the crack fully developed in the wire at the right side. Additionally, it has to be mentioned that the temperature distribution in double bonds due ohmic self-heating and indirect heating through the chip is much more asymmetric than for single bonds.

The failure imaged in Fig. 7b could lead to a wrong identification of the failure mechanism. In fact, in this case, the wire rupture has not been caused by heel cracking, but by the shear stress arising due to the use of a too rigid bond wire coating layer (selectively removed in Fig. 8b).

The model of Schafft [8] enables to predict analytically the number N_f of thermal cycles to heel cracking due to bending stress. It also bases on the power law

$$N_f = A\varepsilon_f^n \tag{5}$$

where A and n are constants for a particular material and the wire strain ϵ_f is computed according to

$$\varepsilon_f = \frac{r}{\rho_0}\left(\frac{ar\cos\left((\cos\psi_0)(1-\Delta\alpha\Delta T)\right)}{\psi_0} - 1\right) \tag{6}$$

$\Delta\alpha$ is the mismatch in the CTE of aluminum and silicon, while ψ_0, ρ_0, and r are geometrical parameters defined in Fig. 8b. The values $A = 3.9 \times 10^{-10}$ and $n = -5.13$ are engineering estimates [9,10], which are usually encountered in microelectronic applications of aluminum bond wires with a diameter below 100 μm.

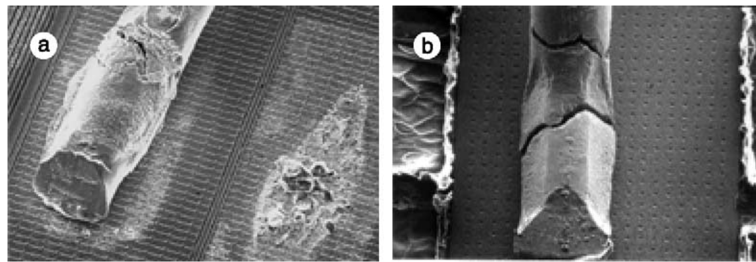

Fig. 7. (a) Bond wire heel cracking due to low-cycle fatigue stressing (SEM image, 25×). (b) Bond wire cracking due to improper bond wire coating (SEM image, 25×).

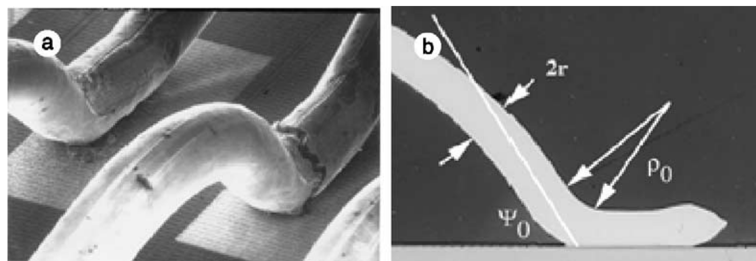

Fig. 8. (a) Heel cracking in a double wire bond. Crack initiation can also be observed in the double bond in the back (SEM image, 25×). (b) Parameter definition for the lifetime model.

5. Aluminum reconstruction

Although reconstruction of the aluminum metalization is an effect, which has been encountered since the early times of microelectronics [11,12], the occurrence of this degradation mechanism in IGBT multichip modules has been firstly reported in [13,14].

During thermal cycling of IGBT devices and of freewheeling diodes, periodical compressive and tensile stresses are introduced in the thin metalization film by the different CTEs of the aluminum and of the silicon chip. Due to the large thermo-mechanical mismatch between both materials and due to the stiffness of the silicon substrate, the stresses, which arise within the aluminum thin film during pulsed operation of the device can be far beyond the elastic limit. Under these circumstances, the stress relaxation can occur by diffusion creep, grain boundary sliding, or by plastic deformation through dislocation glide, depending on temperature and stress conditions. In the case of IGBT devices, the strain rate of the metalization is controlled by the rate of temperature change. Because the typical time constants for thermal transients in IGBT are in the range of the hundreds of milliseconds, if the devices are operated cyclically at maximum junction temperatures above 110 °C, the stress relaxation occurs mainly by plastic deformation at the grain boundaries. Depending on the texture of the metalization, this leads either to the extrusion of the aluminum grains or to cavitation effects at the grain boundaries. Fig. 9a and b shows how the metalization of an IGBT and of a freewheeling diode appears after reconstruction. In optical images reconstructed regions look dark, because of the light scattering due to the surface roughness. Reconstruction is more evident at the center of the chip, where the junction temperature reaches its maximum. It has been shown by infrared thermography [13] that surface reconstruction is negligible in those peripheral regions of the chip, where the maximum junction temperature does not exceed 110 °C. Fig. 9b shows that surface reconstruction sometimes occurs as a secondary mechanism in conjunction with bond wire lift-off. In fact, after release of the bond wires on the left side of the diode, the (pulsed) current has been carried by the bond wire on the right side only, by leading to an increase of the local temperature with consequent reconstruction of the metalization.

In Fig. 10a and b the emitter metalization of a virgin IGBT chip is compared with that of a similar device, which survived 3.2 million of cycles between 85 and 125 °C. After stress, it can be seen that non-columnar aluminum grains are extruded from the thin film surface, while voids are present at the boundaries of larger grains.

In field failures turning into a destructive burn out of the device, aluminum reconstruction may be less evident, due to remelting of the metalization as consequence of

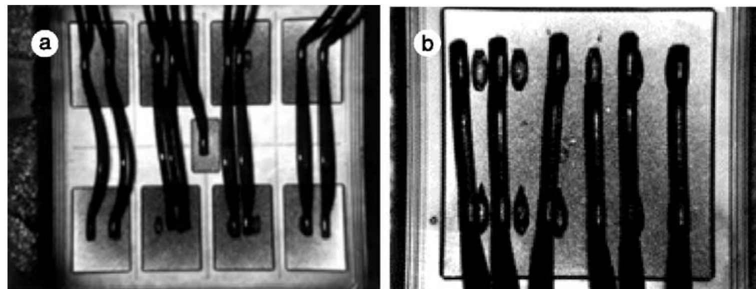

Fig. 9. (a) Reconstructed emitter and gate metalization of an IGBT (optical image, 4×). (b) Reconstructed metalization of a freewheeling diode (optical image, 5×).

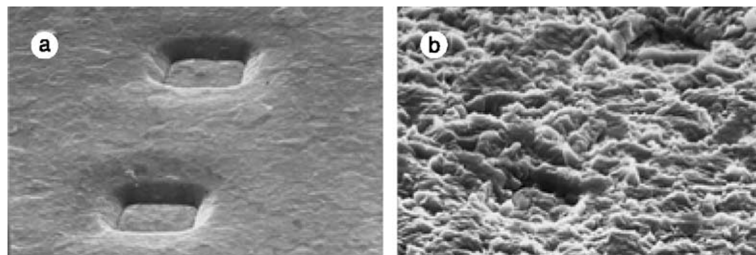

Fig. 10. (a) Emitter metalization of an IGBT chip before power cycling (SEM image, 1000×). (b) Reconstructed emitter metalization after 3.2 millions of power cycles between 85 and 125 °C (SEM image, 1000×).

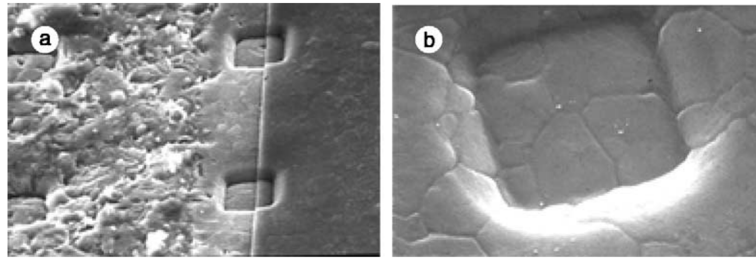

Fig. 11. (a) Reconstructed emitter metalization after removal of the polyimide passivation (SEM image, 800×). (b) Grain boundary depletion in a passivated emitter contact after power cycling (SEM image, 1500×).

the high-temperature levels that can be reached. In any case, aluminum reconstruction reduces the effective cross-section of the metalization and results into an increase of the sheet resistance of the aluminum layer with time. This effect contributes to the observed linear increase of V_{CE} as function of the number of cycles during power cycling tests. Aluminum reconstruction may become a reliability hazard in presence of pre-existing step coverage problems at the emitter contact vias. In this case, thermo-mechanical and electromigration effects can coalesce resulting into a complete depletion of the metalization from the wall of the via.

Fig. 11a shows the role of a compressive layer in suppressing reconstruction phenomena in aluminum layers. In fact, after selective removal of the polyimide passivation, it can be clearly seen that this overlayer has almost inhibited the extrusion of metal grains in the center of the image. Therefore, the use of compressive overlayers can be considered an effective countermeasure for controlling the increase of the sheet resistance of metalization layers submitted to large temperature swings. Fig. 11b shows a close view of the aluminum metalization at an emitter contact that has been coated with a compressive layer and then power cycled with a maximum temperature of 125 °C. As expected no reconstruction occurs. However, one can clearly see that the grain boundaries have been depleted as a consequence of cavitation effects. Voiding of the grain boundaries has been also observed in non-passivated metalization layers submitted to long power cycle testing with a maximum junction temperature below 100 °C.

6. Brittle cracking and fatigue crack propagation

The brittle materials used in advanced IGBT multichip modules are the single crystal silicon, the thin insulating layers on it, and the ceramic substrate. One among the main assumptions in fracture mechanics of brittle materials is that the sharp stress concentration at pre-existing damages leads to the rupture under the influence of external mechanical stresses. Ultimate brittle fracture can occur suddenly without any plastic deformation, when an initial crack is present, whose length exceeds a critical size, which is a characteristic of every brittle material [15]. Failures due to brittle cracking are usually observed immediately after mounting or powering the device. However, even if the initial crack does not reach the critical length, it can develop by fatigue crack propagation under the influence of the applied stresses, until the threshold for brittle fracture is exceeded. This usually turns into early field failures, as in the case of the short circuits, which are due to the propagation of a crack through the polyoxide as a consequence of a pre-damage introduced during wire bonding.

Pre-existing defects can be originated for example by processing problems (e.g. during dicing), by assembly problems (e.g. hard wire bonding), or by soldering (e.g. voids in solder alloys). Fig. 12a shows a notch in the bottom side of an IGBT chip, which has been caused during diamond sawing of the silicon wafer.

There are different sources of stress, which can lead to brittle failures. One among these is the bending stress, which arises while mounting modules with a bowed base plate onto a flat heat sink. This failure cause is less frequent in advanced modules since the uncontrolled bimetallic warpage of the base plate is reduced by using partitioned ceramic substrates and by using bow-shaped base plates. Fig. 12b shows an unusual horizontal crack in the sub-surface region of an IGBT chip, which developed very likely as consequence of the peeling stress arising when mounting a module with an excessively convex base plate.

Fig. 13a shows the vertical cracks caused across an Al_2O_3 ceramic substrate by the horizontal tensile stress produced by the same failure cause as in previous case. In Fig. 13b a similar crack is represented but propagating from the border of a large void within the solder layer between the ceramic substrate and the base plate. Cracks across the ceramic substrate are particularly insidious, because they can transform with time into insulation failures, which can dramatically impair the partial discharge properties of a multichip module.

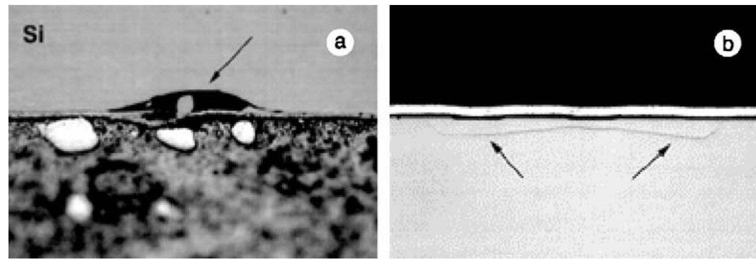

Fig. 12. (a) Notch in the silicon chip (micro-section, optical image, 250×). (b) Crack in the silicon chip due to bending stresses in the base plate (micro-section, optical image, 300×).

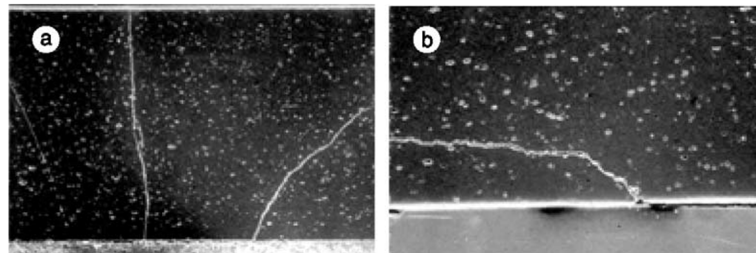

Fig. 13. (a) Vertical crack within an Al_2O_3 ceramic substrate, due bending stresses (micro-section, SEM image, 400×). (b) Crack within an Al_2O_3 ceramic substrate initiated from an inhomogeneity in the solder layer (micro-section, SEM image, 600×).

Because of the conservative design of the compliant layers, brittle cracking of the silicon chip and of the ceramic substrate due to thermo-mechanical mismatch only is unusual in advanced IGBT multichip modules. However, thermo-mechanics may concur with some pre-existing stresses in initiating and propagating the fracture. This is the case of extreme thermal shocks, where the thermal transient can be as fast, that it cannot be followed by the stress relaxation through the plastic deformation of the compliant layers.

Fig. 14 illustrates one among the most insidious cases of fatigue crack propagation leading to a latent short circuit between gate and emitter. When the liquid crystal microthermography has been performed before removing the bond wires, no hot spot could be detected, even at high levels of power dissipation. After careful removal of the emitter bond wires (but one) the failure mode was unaffected, but a hot spot has been found within the footprint of a bond wire (Fig. 14a). Selective etching of the metallization (Fig. 14b) revealed a damaged polyoxide with traces of interdiffusion between Aluminum and polysilicon. This kind of pre-damage is usually introduced during the bonding phase of the emitter bond wires, due to uncalibrated bonding tools. Pre-damaged modules often pass the final production tests, since the device does not exhibit any anomalous characteristic. Due to the thermal and thermo-mechanical stresses, which arise during operation, the microcrack in the polyoxide propagates and becomes decorated by the Aluminum metalization leading to a current leakage

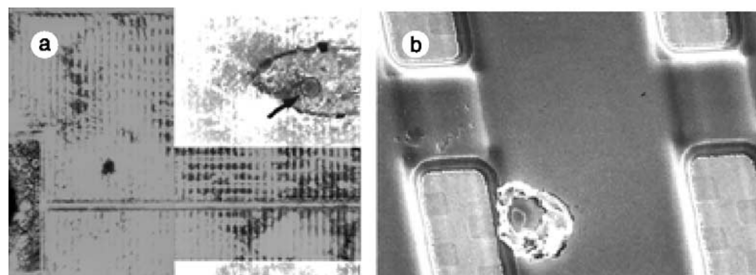

Fig. 14. (a) Localization of a leakage path between emitter and gate by liquid crystal microthermography (80×). (b) The hot spot is located below an emitter bond wire, and it is due to a mechanical damage of the polyoxide (SEM image 500×).

between the gate and the emitter, and thus to an early failure. In the case of gate drivers with a low output resistance, the conductive filament can be melted turning into an intermittent leakage current.

7. Corrosion of the interconnections

Corrosion of aluminum is a well-known failure mechanism since the early times of microelectronics. When pure aluminum (e.g. bond wires) is exposed to an oxygen containing atmosphere, a thin native Al_2O_3 surface layer is grown that passivates the metal. Aluminum is self-passivating also in pure water, where the native aluminum oxide is converted into a hardly soluble layer of aluminum hydroxide $Al(OH)_3$. When exposed to other solutions, aluminum hydroxide is amphotheric, i.e. it is dissolved both by strong acids (e.g. phosphoric, hydrofluoric and hydrochloric acid) and by strong bases (e.g. potassium hydroxide). This step is followed by the exposition of the bare aluminum surface to further chemical or electrochemical attacks. In converse, strong oxidizing agents (e.g. nitric acid) leave $Al(OH)_3$ unaffected. In presence of an electrolytic or of a galvanic cell aluminum is corroded according to the corresponding redox reaction [16]. The corrosion immunity of the aluminum as a function of the pH of the electrolyte and as function of the voltage applied in an electrolytic cell is described by the different regions of the related Pourbaix diagram [17]. In failure analysis both anodic and cathodic aluminum corrosion are found. Nevertheless, the cathodic and anodic corrosion mechanisms usually encountered in microelectronics are not expected to play a major role in IGBT devices. On the contrary, several galvanic corrosion mechanisms have been observed to affect in different ways the metallic components of a module. Among these there are bimetallic corrosion, thermo-galvanic corrosion, concentration cell (especially Oxygen cells), pitting corrosion, stress corrosion, and dealloying [18].

The identification of which driving force is promoting corrosion in IGBT multichip modules is a quite complex issue, since several causes may concur to the failure. In fact, in IGBT packages one is faced with multiple contamination sources, with different metals and alloys, with temperature gradients, as well with static and periodic mechanical stresses. Furthermore, the active devices and the bond wires are embedded in silicon gel, whose influence on the corrosion is not completely understood.

Fig. 15a and b shows aluminum bond wires with no strain buffer that have been corroded at different grades during power cycle tests, which lasted over one million of cycles. This kind of corrosion has been encountered during power cycles performed at low voltage (typically 8 V), as well during lifetime tests at high voltage. The corroded areas were mainly located at those sites of the bond where the wire suffered the most severe deformation, at the heel of the bond wire, and at the top of the wire loop.

These corrosion events have been observed to occur in conjunction with the local formation of gaseous inclusions within the silicone gel (Fig. 16b) that can be sometimes noticed during high-temperature operation of the devices. After package opening by wet chemistry, no corrosion by-products are left at the attacked locations.

This combination of symptoms leads to the conclusion that the observed bond wire corrosion is strongly correlated with the mechanical stresses, which arise either due the thermo-mechanical cycling, or due to residual deformation stresses in the bond. The absence of reaction by-products and the corrugated surface of the corroded bond wires, may indicate that the corrosion occurs at the grain boundaries of the aluminum. Once completely separated the grains get loose and are removed during package opening. In summary, these indications are compatible with the stress corrosion failure mechanism. Intergranular corrosion is observed also to occur in the adjacent aluminum metalization, but with a less destructive effect than in the much thicker bond wires. This is probably due either to the lower mechanical stress, or to the beneficial effect of the additional alloying elements (silicon and copper). The nature and the source of the contamination are not completely understood. The corrosion could be due to the presence

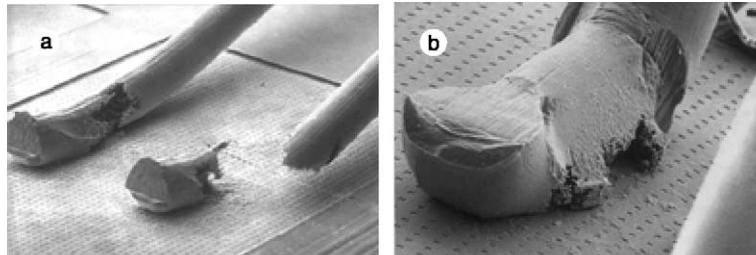

Fig. 15. (a) Rupture of emitter bond wires due to stress corrosion (SEM image, 30×). (b) Detail of a corroded emitter bond wire (SEM image, 80×).

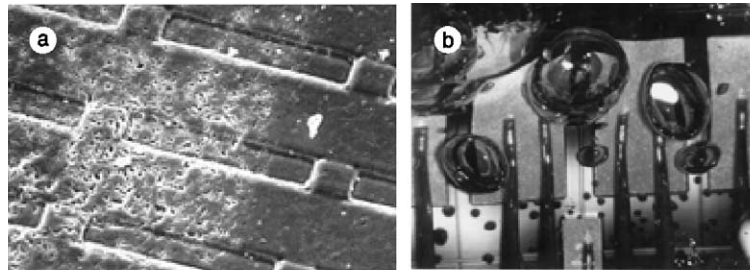

Fig. 16. (a) Corroded emitter bond pad close to an emitter bond wire (SEM image, 160×). (b) Formation of gaseous inclusions into the silicone gel during power cycling (optical image, 8×).

of chlorides originated either from process residuals, or by thermal segregation of the silicone gel.

Recently, the incidence of this failure mechanism has been mitigated by more effective cleaning processes after assembly, by the control of the water content in the silicone gel, and by the use of corrosion-hardened bond wires.

8. Solder fatigue and solder voids

A main failure mechanism of IGBT multichip modules is associated with the thermo-mechanical fatigue of the solder alloy layers. The most critical interface is represented by the solder between the ceramic substrate and the base plate, especially in the case of copper base plates [19]. In fact, at this location one finds the worst mismatch in the CTEs, the maximum temperature swing combined with the largest lateral dimensions (see Table 1 and Fig. 3). Nevertheless, fatigue phenomena occurring in the solder between the silicon chip and ceramic substrate cannot be neglected. This is also the case of process-induced voids, which can both interact with the thermal flow and with the crack initiation within the solder layer.

8.1. Voids

Both gross voids and extended fatigue-induced cracks can have detrimental effects on dissipating devices. In fact, they can significantly increase the peak junction temperature of an IGBT or of a diode and therefore accelerate the evolution of several failure mechanisms including bond wire lift off and solder fatigue. Furthermore, since the heat flow within an IGBT module is almost one-dimensional, when a relatively large void is present in a solder layer, the heat must flow around it by creating a large local temperature gradient such that the heat dissipation performances of the assembly are degraded. On the contrary, if the large void is broken up into many smaller voids, the perturbation to the heat flow is less evident and has a much smaller impact on the

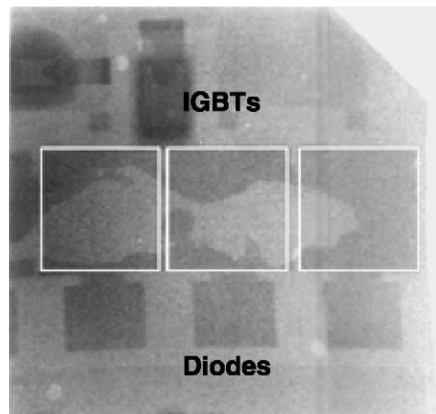

Fig. 17. X-ray microscopy image of an IGBT module, which shows a large void immediately below three IGBT chips (0.8×). The void is located in the die attach layer.

overall thermal resistance of the multilayer. Critical sizes and the most critical sites of contiguous voids in power devices have been investigated experimentally and by numerical simulation in [20]. Fig 17 shows an X-ray image of a module, which failed due to inhomogeneous current sharing caused by a local increase of the thermal resistance as consequence of large contiguous solder voids in the die attach layer.

Since IGBT are vertical devices the die attach has to provide at the same time an efficient thermal and electrical conduction path. Therefore, the most insidious voids within the die attach are those, which hinder the thermal flux to the heat sink without inducing any noticeable reduction of the current distribution within the semiconductor. They are for instance edge cracks or shallow voids and delaminations at the interface with the ceramic substrate.

In advanced assembly processes, special care is taken to avoid the formation of gaseous inclusions within solder layers, by using e.g. vacuum ovens and clean processes. During the packaging phase the control of the temperatures profiles during soldering and during the successive annealing steps is essential for avoiding an

excessive growth of brittle intermetallic layers. Nevertheless, the quality of solder joints between large size plates is still considered a critical issue.

8.2. Solder fatigue

The most frequent solders used in advanced IGBT multichip modules are based on tin–silver, indium, or tin–lead alloys. They have excellent electrical properties and as soft solders they exhibit good flow characteristics. For sake of simplicity, solders are often modeled as a single homogeneous phase. However, when copper is soldered for example with a standard lead–tin alloy, the bond is mainly provided through the formation of a Cu_5Sn_6 intermetallic phase located close to the copper plate [21]. Two additional distinct phases, one tin rich and one lead rich, are formed in the central part of the solder layer upon solidification. During power cycling, these phases coarsen rapidly due to the high homologous temperature at which the alloy is operated. Since the copper phase is much more brittle than the tin–lead phases, thermo-mechanic fatigue cracks often propagate within the copper rich intermetallic. Due to the larger CTE mismatch and to the higher temperature, fatigue cracks are found preferably in the vicinity of the intermetallic layer immediately below the ceramic substrate. Metallographic preparations have shown [22] that cracks initiate as expected at the border of the solder joint, where the shear stress reaches its maximum. Additionally, crack formation is highly promoted by the presence of sharp angles at the edges of the ceramic substrate. This problem requires a dedicated engineering of the solder fillets. Thermal cycle tests with ΔT up to 100 °C have shown that the number of cycles to the failure of the solder between ceramic substrate and base plate just weakly depends on the temperature swing. Because of the very severe conditions imposed by this kind of accelerated test, the results can hardly be extrapolated to real operating conditions. In fact, with a junction temperature swing of 100 °C and the typical material constants listed in Table 1, the expected plastic strain can be estimated in about 50 μm. Since this value has the same order of magnitude of the thickness of the solder layer, it can be expected that the failure mechanism leading to the degradation of the solder during the accelerated test is not representative for the lower temperatures encountered in field applications. In fact, during field operation, the most critical system (Al_2O_3 on copper) rarely experiences temperature swings over 30 °C. Finally, in recent experiments, severe thermal cycles have shown some adhesion failures of AlN ceramic substrates due to the peeling of the copper metalization [23]. However, this unusual failure mechanism is not expected to impact the field reliability of IGBT devices.

The number of cycles to the failure of large solder joints due to thermo-mechanical fatigue can be simply modeled by a Coffin–Manson-like power law of the form

$$N_f = 0.5\left(\frac{L\Delta\alpha\Delta T}{\gamma x}\right)^{1/c} \tag{7}$$

In Eq. (7), L represents the typical lateral size of the solder joint, $\Delta\alpha$ the CTE mismatch between the upper and the lower plate, ΔT the temperature swing, c is the fatigue exponent. x and γ are the thickness and the ductility factor of the solder, respectively. The values $\gamma = 1.1$, and $c = -0.49$ are conservative engineering estimates usually encountered for the In-70% Pb-30%, Sn-40% Pb-60%, and Sn-10% Pb-90% solder alloys [9,10]. From Eq. (7) one can also derive some simple design rules for minimizing the fatigue of solder joints. In fact, it can be easily seen that the lifetime is improved by reducing the size of the solder joint, by matching the CTE of the materials, by reducing the edge voids, and by increasing the thickness of the solder (compatibly with the requirements imposed by the thermal resistance).

9. Burnout failures

Device burnout is a failure mode, which is very frequently observed either as the final act of wear out, or as consequence of a failure cause occurring randomly. Burnout is often associated with a short circuit condition, where a large current flows through the device (or through a portion of it), while it is supporting the full line voltage. Sustaining a short circuit over a long time interval inevitably leads to thermal runaway and finally to a fast destruction of the device. In fact, since IGBTs do not require any $\mathrm{d}i/\mathrm{d}t$ snubbering, the device itself limits the current increase rate. Therefore, after the failure the current may increase at a rate up to 10 kA/μs, leading to a current maximum in the 100 kA range and to a decay within 100 μs [24]. In this case, the main part of the stored capacitive energy is released in few hundreds of nanoseconds reaching a peak power up to 100 MW. The capacitive energy is dissipated by the ohmic components of the circuit, i.e. mainly by the bond wires and by the silicon chip. As consequence of the adiabatic heating process, the bond wires evaporate, by producing a preferential conductive path for arching through the module. The resulting shock wave rapidly propagates through the silicon gel by leading to the catastrophic destruction of the device. Advanced IGBT multichip module [25] have been expressly designed for minimizing the consequences of such an explosion in order to match the tight requirements in terms of safety imposed by traction applications [26].

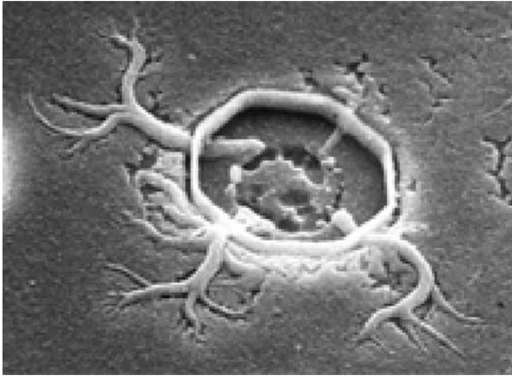

Fig. 18. Polysilicon filamentation in an IGBT as a consequence of a short circuit between gate and emitter, due to pre-damaged insulation during wire bonding (SEM image, 2000×).

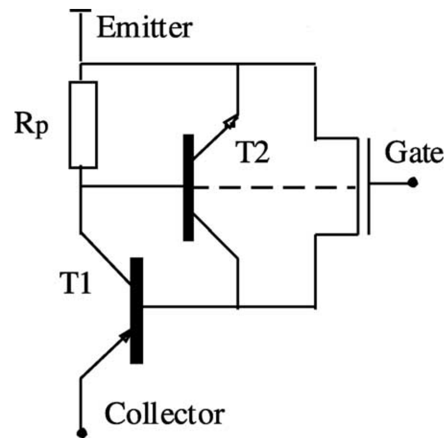

Fig. 19. Simplified equivalent circuit of an IGBT.

They are many system, environmental and wear out related causes, which may turn into a short circuit condition. Among these there are operation of the device outside the safe operating area, gate unit malfunction, inhomogeneous current sharing [27], overheating due to the degradation of the thermal resistance, dielectric breakdown, and cosmic ray irradiation.

Fig. 18 shows a characteristic polysilicon filamentation, which occurred during high-voltage testing of an IGBT module, because of the short circuit between gate and emitter. The root cause of this failure is a pre-damage introduced in the polyoxide by the bonding tool. The relatively small damage produced is due both to the very localized fracture in the dielectric, as well by the short circuit detection and consequent current limitation in the tester.

9.1. Latch up

The latch up is a failure mechanisms inherent to IGBT devices. This phenomenon is of special relevance, because most of the root causes mentioned above activate this mechanism, such that it plays an important role in determining the availability of a power system. Nevertheless, it has to be noted that the latch up is mainly a problem related to the ability of a certain device design to survive stresses out-of-specification. Thus, strictly speaking, it is a robustness issue rather than a reliability concern. The latch up mechanism (static and dynamic) manifest itself through a sudden collapse of the collector to emitter voltage, and once this failure mechanism is activated the device cannot be longer controlled through the gate. The failure mode associated with latch up is always a generalized low-ohmic short circuit of collector, emitter, and base.

Fig. 19 represents the simplified equivalent circuit of an IGBT, which takes into account just quasi-static effects. Under normal forward operating conditions the voltage drop caused by the collector current over the p-emitter diffusion (represented by the resistor R_p) is almost negligible. Therefore, the parasitic BPT T2 is in the non-conducting state and the IGBT device is controlled by the electron flow injected into the base of T1 through the MOS transistor. On the contrary, if the collector current reaches the critical value for which the voltage across R_p exceeds 0.6 V, T2 enters in conduction and provides the base current to T1. Since the additional base injection turns into an increase of the collector current, this effect is regenerative and leads to the thermal destruction of the device, which is not controlled by the gate voltage anymore. This simple quasi-static model illustrates how latch up may arise in n-channel IGBTs, while forcing the collector current to increase. This situation can occur in an IGBT module, if the number of operating cells within a module is reduced with time, due to a degradation mechanism, as for example bond wire lift-off. Nevertheless, more complex physical and numerical models are required for taking into account all dynamic effects, which concur in triggering this failure mechanism [28,29].

Fig. 20a represents a melted pit on an IGBT device, which resulted from the latch up event during a lifetime test at high voltage. During failure analysis the module showed clear evidence of distributed bond wire lift off. The melted path usually crosses the IGBT chip down to the die attach and a silicon-solder alloy is formed. Craters have been often observed in immediate vicinity of melted emitter bond wires, indicating that those bond wires were still attached shortly before the latch up event. This fact also suggests that the local current density is increased also in consideration of the sheet resistance degradation due to the reconstruction of the metalization. Fig. 20b, shows the effect of a latch up event, which occurred in an IGBT device during a long-term frequency test. The latch up interested just some

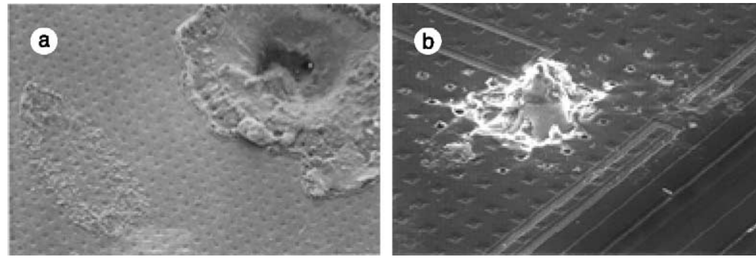

Fig. 20. (a) Melted pit in an IGBT due to a latch up event, which occurred in conjunction with bond wire lift off (SEM image, 50×). (b) Same effects than in the previous image but localized to some few cells (SEM image, 90×).

few cells and the melted area has been kept under control by external limitation of the collector current.

Cases have been reported [30], where catastrophic burnout can be caused by second breakdown. However, since second breakdown is basically related to avalanche carrier multiplication at high electron current regimes, it mainly affects p-channel IGBT devices.

9.2. Cosmic rays

Catastrophic burnout of IGBT devices can also be initiated through local self-sustaining filamentary discharges produced in the silicon by recoil nuclei, which result either from neutron scattering, or from the decay of neutron-activated isotopes within the semiconductor. At normal operating conditions, high-energy neutrons are usually associated to terrestrial cosmic radiations [31]. A universal curve has been derived from the Zeller model [32], which predicts the failure rate of bipolar devices (thyristors, GTO, diodes) as function of an electric field parameter. IGBT devices show an increased sensitivity to cosmic ray in respect to thyristors, GTO, and diodes. In fact, the measured failure rate exceeds by at least an order of magnitude the value predicted by the universal curve.

10. Conclusions

Electric traction has provided a particularly demanding application domain where severe thermal and power cycle constraints are engendered by the manyfold application profiles. The increased demand in terms of device performances and reliability has to be faced concurring with the continuous trends towards cost reduction. In the past, the selection of power modules was primarily based upon current and voltage ratings, only. Today's high-power applications have to take in the same consideration the effects of the power dissipation in particular on the long term reliability of the system. The investigation of the failure mechanisms, which occurred in early development stage of new power modules enabled the designers to implement efficient corrective actions for realizing the specified mission profiles.

Together with the optimization of the silicon design to minimizes the losses and to increase the maximum operating temperature, the thermal management of the modules is one among the future challenges in power applications. This is leading to the introduction of new packaging materials and to the use of integrated cooling technologies.

Acknowledgements

I would like to thank Paolo Malberti for his continuous technical advise and Giovanni Nicoletti and Paolo Scacco for preparing some among the samples presented in this paper.

References

[1] Cova P, Ciappa M, Franceschini G, Malberti P, Fantini F. Thermal characterization of IGBT power modules. Microelectron Reliab 1997;37:1731–4.
[2] Frost H, Ashby M. The plasticity and creep of metals and ceramics. New York: Pergamon Press; 1982.
[3] Hager C. Lifetime estimation of Al wire bonds based on computational plasticity. ETH Thesis Nr. 13763.
[4] Ramminger S, Türkes P, Watchutka G. Crack mechanism in wire bonding joints. Microelectron Reliab 1998;38:1301–5.
[5] Hamidi A, Beck N, Thomas K, Herr E. Reliability and lifetime evaluation of different wire bonding technologies for high power IGBT modules. Microelectron Reliab 1999;39:1153–8.
[6] Scacco P, Ciappa M. Reliability Laboratory of the Swiss Federal Institute of Technology (ETH), Zurich (ETH), unpublished results.
[7] Ciappa M, Fichtner W. Lifetime prediction of IGBT modules for traction applications. IEEE Int Reliab Symp 2000;38:210–6.
[8] Schafft H. Testing and fabrication of wire bonds electrical connections—A comprehensive survey. National Bureau of Standards, Tech. Note, vol. 726, 1972. p. 106–9.

[9] Ciappa M. Package reliability in microelectronics: a review. Proceedings of International Workshop on Electronics and Detector Cooling, WELDEC, vol. 1, 1994. p. 133–49.

[10] Reliability analysis/assessment of advanced technologies, Technical Report RADC-TR-90-72.

[11] Santoro C. Thermal cycling and surface reconstruction in aluminum thin films. J Electrochem Soc 1969;116:361–4.

[12] Philofsky E, Ravi K, Hall E, Black J. Surface reconstruction of aluminum metalization—a new potential failure mechanism. IEEE Int Reliab Phys Symp 1971;9:120–8.

[13] Ciappa M, Malberti P. Plastic-strain of aluminum interconnections during pulsed operation of IGBT multichip modules. Qual Reliab Eng Int 1996;12:297–303.

[14] Malberti P, Ciappa M, Cattomio R. A power-cycling-induced failure mechanism of IGBT multichip modules. Int Symp Testing Failure Anal 1995;21:163–8.

[15] Lall P, Pecht M, Hakim E. Influence of temperature on microelectronics and system reliability. Boca Raton, FL: CRC Press; 1997.

[16] Ciappa M. Proceedings of the International Course on Failure Mechanisms and Failure Analysis of Semiconductor Devices, Swiss Federal Institute of Technology (ETH), 2–4 April 1996.

[17] Winston R, editor. Uhlig's corrosion handbook. New York: Wiley; 2000.

[18] Ciappa M. Some reliability aspects of IGBT modules for traction applications. In: Hartung-Gorre, editor. Series in microelectronics, vol. 111, 2001.

[19] Herr E, Frey T, Schlegel R, Stuck A, Zehringer R. Substrate-to-base solder joint delamination in high power IGBT modules. Microelectron Reliab 1997;37:1719–22.

[20] Ciampolini L, Ciappa M, Malberti P, Regli P, Fichtner W. Modelling thermal effects of large contiguous voids in solder joints. Microelectron J 1999;30:1115–23.

[21] Olsen D, Wright R, Berg H. Effects of intermetallics on the reliability of tin coated Cu, Ag, and Ni parts. IEEE Int Reliab Phys Symp 1975;13:80–6.

[22] Rodriguez M, Shammas N, Plumpton N, Newcombe D, Crees D. Static and dynamic finite element modeling of thermal fatigue effects in IGBT modules. Microelectron Reliab 2000;40:455–63.

[23] Mitic G, Beinert R, Klofac P, Schultz H, Lefranc G. Reliability of AlN substrates and their solder joints in IGBT power modules. Microelectron Reliab 1999;39:1159–64.

[24] Zeller H. RAPSDRA project BE 95-2105, 30/36m Report.

[25] Stockmeier T, Schlapbach U. 1200A, 3300 V IGBT power module exhibiting very low internal stray inductance. Proceedings of the PCIM Conference, PCIM, 1998. p. 331–7.

[26] Gekenidis S, Ramezani E, Zeller H. Explosion tests on IGBT high voltage modules. International Symposium on Power Semiconductor Devices and ICs, ISPSD, vol. 11, 1999. p. 129–32.

[27] Palmer P, Joyce J, Stark B. Measurement of chip currents in IGBT modules. European Conference on Power Electronics and Applications, EPE, vol. 7, 1997. p. 406–11.

[28] Baliga B. Modern power devices. New York: Wiley; 1987.

[29] Bauer F, Stockmeier T, Lendenmann H, Dettmer H, Fichtner W. MOS controlled power switches above 2000 V: MCT versus IGBT. Proceedings of the 26th International Power Conversion Conference, PCIM, 1993. p. 312–26.

[30] Iwamuro N, Okamoto A, Tagami S, Motoyama H. Numerical analysis of short circuit safe operating area for p-channel and n-channel IGBTs. IEEE Trans Electron Dev 1991;38:303–8.

[31] Findeisen C, Herr E, Schenkel M, Schlegel R, Zeller H. Extrapolation of cosmic ray induced failures from test to field conditions for IGBT modules. Microelectron Reliab 1998;38:1335–9.

[32] Zeller H. Cosmic ray induced failures in high power semiconductor devices. Microelectron Reliab 1997;37:1711–8.