Hot Nitrogen Deballing of Ball Grid Arrays

Stoyan Stoyanov[1], Alexander Dabek[2], and Chris Bailey[1]

[1] CMRG, University of Greenwich, London SE10 9LS, United Kingdom
[2] Cassidian-EADS, Woerthstr. 85, 89077 Ulm, Germany

Email: s.stoyanov@gre.ac.uk

Abstract: Ball Grid Array (BGA) packages are increasingly adopted in high reliability electronics equipment. The main reliability concern is that lead-free (Pb-free) packaged BGAs bring the risks of failures due to tin whiskers growth phenomena associated with tin or tin-rich alloys. Replacing Pb-free solder balls of BGA components with tin-lead solder alloy materials is the most effective risk mitigation strategy. Post-manufacturing processes that can be used to remove (deballing) and then deposit back (reballing) BGA solder balls have been recently developed and increasingly put in practice. This paper reports on the assessment of the thermal responses of BGAs subjected to hot nitrogen (N_2) deballing and details the respective conclusions about the risk of thermally induced damage.

1. INTRODUCTION

The supply of electrical and electronic parts manufactured with tin-lead (SnPb) solder has substantially reduced in recent years as result of legislations such as the European Union (EU) Directive on Restriction of Hazardous Substances (RoHS) and the Waste Electrical & Electronic Equipment (WEEE) Directive [1,2]. As a result, lead-free (Pb-free) solder interconnects and component finishes are now becoming universal. Despite existing exemptions from regulations that ban the use of lead and other hazardous materials in electronic parts, manufacturers of electronic equipment used in high reliability and long service life applications have little choice but to adopt increasingly lead-free (Pb-free) packaged commercial-of-the-shelf components in their products.

For the defence and high reliability communities the introduction of Pb-free parts, materials and processes is problematic. The substitution of SnPb solder interconnections and plating on electrical and electronic part termination finishes with Pb-free tin alloys or pure tin has reintroduced the potential for electrical short-circuit malfunctions due to the tin whisker growth phenomenon, something which has largely been forgotten about for over a century [3].

The risk of whisker induced short circuits is a major reliability concern. There are two basic mitigations that can be applied in design to minimise the effects of tin whiskers. Firstly, conformal coating should be applied to circuit panels. It will not necessarily stop whisker growth, but it can prevent short-circuits in the sense that nodes adjacent to the whisker may be protected. The second mitigation strategy is to avoid pure tin finishes and use only Sn-Pb as surface passivation or interconnect material. This requires that RoSH compliant parts with Pb-free solder interconnects or pure Sn finish on component terminations in the case of leaded components, as supplied from component manufacturers, to be stripped and replaced with SnPb.

Ball Grid Array (BGA) components are commonly used in many applications due to the various advantages of this packaging technology (e.g. reduced size, excellent electrical and thermal performance, large I/O, etc) [4]. In the case of BGA components, processes that use localised thermal load to remove Pb-free solder balls and then deposit tin-lead solder balls on a ball-by-ball basis have been developed recently and commercialised [5]. Such processes are ideally suited for repair and rework but also for replacing Pb-free with SnPb solder balls for the purpose of tin whisker risk mitigation.

There is very little knowledge about the thermo-mechanical impact on BGAs from such processes although they are generally regarded as being most safe due to localised thermal effects. The motivation of this work is to assess quantitatively the thermal behaviour of BGA packages subjected to deballing by means of hot nitrogen heating and partial vacuum removal of the solder material from the BGA pads.

2. HOT NITROGEN DEBALLING PROCESS

The vacuum de-balling principle is shown in Figure 1. In the case of the studied process, the nozzle is maintained at a temperature of 250°C, and the hot nitrogen (at 250°C) is injected through injectors integrated outside the nozzle so that they can produce a focused nitrogen flow jet capable of melting an individual solder ball in the package ball array without the nozzle coming into contact with the solder ball. The nozzle aperture is chosen to be smaller than the solder ball. The hot nozzle is lowered from above the solder ball and the focused nitrogen flow gradually heats and melts the solder material. For a ball with diameter 500 um this time is about 4 seconds.
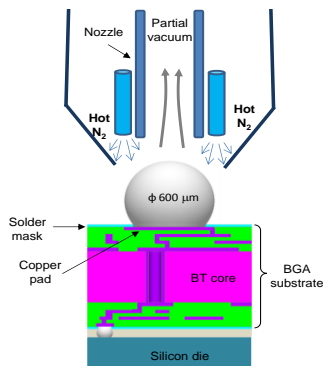

Fig. 1. Hot N_2 vacuum deballing.

Under the partial vacuum condition inside the nozzle tube, the liquid solder is taken from the pad as the nozzle moves back vertically and away from the package. It takes approximately another 8 seconds to transport the solder to the reservoir, and for the nozzle to start de-balling the next joint. It is a fairly slow process compared with the laser-assisted processes for de-balling BGA components. For example, in the case of laser assisted de-balling the typical ball removal time is significantly shorter, in the range of 1 sec and below. With hot nitrogen vacuum de-balling, the time required to de-ball a single solder ball from the BGA package is approximately 12 seconds implying 2

hours 40 minutes to complete a typical 800 ball package. The main advantage over laser-assisted deballing is that the heating with hot nitrogen de-balling is more gradual and equipments are capable to collect large amounts of molten solder efficiently and without need for frequent empting of the equipment's reservoirs, thus making this process suitable for large arrays and BGAs with large solder balls.

3. MODEL DEVELOPMENT

A typical lidded BGA package construction is investigated. The selected XilinxVirtex-4 flip-chip BGA has body size 27.0 x 27.0 mm and 688 ball connections at 1.00 mm pitch. The nominal substrate thickness is 0.75 mm.

Finite element analysis is used to predict the thermal response of this BGA structure at both global and local level. All simulations are realised using finite element analysis software ANSYS [6].

3.1. Finite Element Models

Two types of finite element models are developed to enable realising a sub-modelling approach to the problem of analysing the thermal behaviour of the BGA package to hot nitrogen deballing. The global model developed to represent the entire package (see Figure 2, left) captures all main constructional details of the component but omits details associated with the explicit inclusion in the model of the flip-chip first-level solder joints.
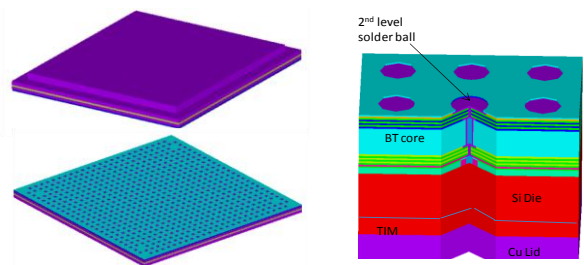

Fig. 2. Global (left) and local (right) FEA thermal models of the de-balling processes. Cut-off view revealing inside details of the copper macro via and the first-level solder joint beneath.

The local model is required to enable precise thermal evaluation in the vicinity of the removed ball. Given the process characteristics, localised thermal effects are expected to dominate the package responses. The local model is illustrated in Figure 2

(right). The developed local model represents a package section within the die perimeter. The local model captures a sub-domain of 3x3 mm section that encloses 9 copper pads/solder balls on the BGA substrate (an array of 3 by 3 solder balls). The pad in the centre of the array is the one subjected to process simulation.

The local model assumes the presence of shortest signal path. The copper path is in the form of a straight line connecting the processed second-level solder ball and the BGA first-level solder joints beneath, and includes micro vias through the dielectric layers and a macro via in the BT core. This scenario addresses the question about the thermal impact of the de-balling thermal loads on the first-level interconnects and die assuming an extreme case for the internal copper structure of the BGA substrate.

Thermal material properties used in the model are listed in Table 1. The copper content in metal layers of the BGA substrate is 65% evaluated as volumetric fraction.

Table 1. Thermo-mechanical properties.

| BGA Material	Density (kg/m³)	Thermal Conductivity (W/m.K)	Specific Heat (J/kg.K)
Silicon	2330	146 at 20°C 99 at 125°C	712
Copper	8900	380	385
Solder mask	1700	0.33	920 at 20°C 1380 at 125°C
Solder (Sn/Pb)	8510	53	185
Underfill	1300	0.2	1655
BT Core	2000	0.5	900
Dielectric	2000	0.337	900

3.2. Modelling Assumptions

The total duration for removing one solder ball is 12 sec. During the first 4 sec the second-level solder ball is heated and the remaining 8 sec are used to suck the solder through the nozzle under partial vacuum. No contact between the liquid solder material and the package pad during the latter interval of 8 sec occurs.

The analysis assumes the BGA is pre-heated to 65°C. During processing the package is kept flipped upside down and resting on a temperature chuck maintained at temperature 65°C; hence fixed

temperature (at 65°C) is used as boundary condition (BC) at this surface. Ambient air temperature is 22°C and natural convection to ambient air is specified as BC at the relevant external boundary of the package. Temperature boundary conditions for the local thermal model at the cut-off-boundary come from the global model.

The processed solder ball is not explicit in the models. Instead suitable boundary conditions at the package pad and for ambient air are used in the form of prescribed (imposed) time-temperature relationships. The definition of these time-temperature curves is on the basis of data for the hot nitrogen temperature (250°C) and the timings/speed of nozzle movement towards and away from the ball. The imposed temperature boundary condition for the pad is illustrated with the graph in Figure 3. This BC for the pad covers the first 4 sec (from total duration 12 sec) of the process. The ambient temperature data above the de-balled pad is given with the graph in Figure 4. This is defined over time 4 to 12 sec when no solder is in contact with the pad.
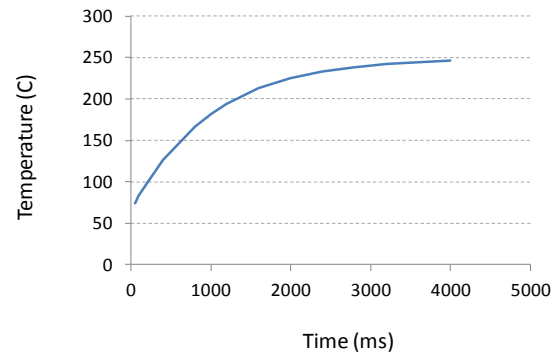
Fig. 3. Time-temperature boundary condition for the de-balled pad in the global/local models over the first 4 sec of the de-balling process.
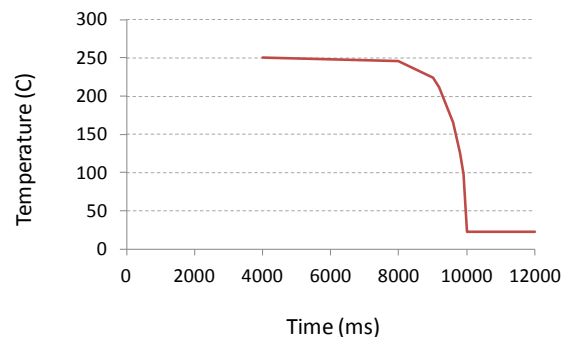
Fig. 4. Time-temperature profile of the ambient air in the close proximity of the de-balled pad over the processing time from 4 sec to 12 sec.

Because nitrogen injection is continuous but the nozzle moves away from the pad after 4 sec, the time interval 4 sec to 12 sec requires changing the deballed pad BC. The new BC for the pad is the one of convection and radiation to ambient air. The temperature of the "ambient" air (in fact injected N_2), defined with the relationship in Figure 4, is due to the continuous nitrogen injection. The ambient temperature change away from the processed pad is modelled as having a spatial distribution where temperature of the ambient becomes 22°C two pads away from the de-balled pad.

4. THERMAL MODELLING RESULTS

4.1. Thermal Behaviour at Global Package Level

The global thermal model is used to provide knowledge how far spatially the temperature front from the heated ball under removal spreads away from the processed pad location and to define the thermal boundary conditions at the cut-off-boundary of the local thermal model. The global simulation is set to predict the thermal behaviour of the package under the sequential removal of three consecutive solder balls, hence total process time that the model simulated is 36 seconds. This enables to investigate to what extend, if any, temperature fronts propagate away from the processed pad, over what time periods, and therefore to gain knowledge which balls and pads might be thermally affected.
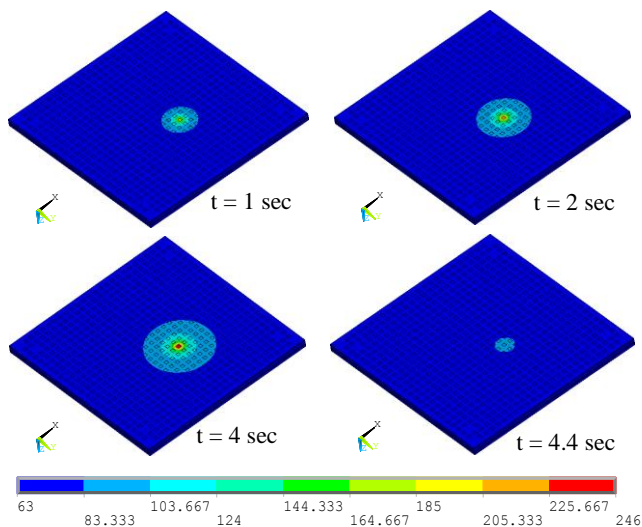
Fig. 5. Temperature (°C) contours at global package level and at discrete time steps during hot N_2 ball removal. Results refer to the removal of one solder ball.

Figure 5 illustrates the temperature contour results at package level during de-balling of one out of the three simulated ball removals, and therefore results in the figure are within the time interval 0 to 12 seconds only.

Modelling results show that at the end of each 12 second time window for a ball removal, the temperature at package level drops back to its initial state given with the temperature of the hot plate on which the package sits. This means that at the beginning of each solder ball removal the localised region is not at elevated temperature from the previous ball de-balling, hence in this case the individual solder ball removals are independent events from thermal interaction point of view. However, the model shows that the temperature front propagates to some extent during de-balling. As a result of this the package structures in near proximity undergo thermal fluctuations. These gradually fade away as the hot nozzle moves further away from any fixed pad under observation.

As evident from Figure 5, once a ball is taken from the pad approximately 4 sec after it has been heated, and when there is no more a full contact between solder and the pad, the temperature at that same pad drops sharply. Half second after the ball is detached, and despite having ambient air in the local region still around 240°C from the hot nitrogen, the temperature at that local region of the bare pad drops to around 85°C and below (see contour plot at time 4.4 sec). The reason for this quick temperature drop is the heat loss through conduction from the pad to the package (maintained at 65C). The heat loss through conduction dominates completely over the convection heat transfer. For the remaining time that makes up the full 12 seconds of the time window per solder ball, the entire localised region gets to almost a uniform temperature equal to the hot plate temperature on which the package sits. From the contour plot corresponding to time 4 sec (the highest temperature time point) it is evident that the contour line for the 83°C temperature front stretches no further than 4 to 5 pads away from the de-balled pad.

Figure 6 shows the temperature change at a fixed spatial point defined at the centre pad in the local model. In this plot, the temperature history is over 36 sec thus corresponding to the sequential removal of three-in-row solder balls, each taking 12 sec to process. The plot provides the temperature history at

the location of the last (among the three considered in this case) processed copper pads. It is evident that as the hot N_2 melts and then nozzle removes the solder material of each ball, in sequential manner pad by pad and with repetition rate 12 sec, the temperature at the "fixed-for-observation" pad is influenced by the processing taking place at the pad next to the adjacent and at the adjacent pad itself. On the graph, this behaviour is observed over the interval 0 to 12 sec for the next to the adjacent pad and over the period 12 to 24 sec for the adjacent ball removal. The peak temperatures predicted at the pad under observations are 108.5°C and 142°C respectively for the two balls removed prior to reaching the pad under observation. The graph section over the time duration 24 to 36 sec corresponds to the de-balling of the ball at the pad where this temperature data is considered. The temperature peak at the pad subjected to solder de-balling peaks at near 247°C at time 28 sec (4 sec into the re-balling of that specific ball). The detailed thermal behaviour at this pad is analysed with the local thermal model of de-balling.
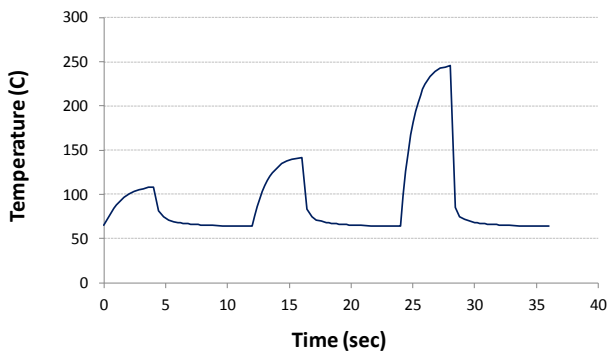

Fig. 6. Temperature at the centre of the pad over period of 36 sec capturing the de-balling of three sequential solder joints. The pad with the temperature data is the one processed from time 24 to 36 sec.

The main conclusion is that the removal of a ball causes thermal cycles at the adjacent pad that have a peak temperature 142°C reached from an initial 65°C over 4 sec. The second main conclusion is that under the process conditions assumed and modelled in this simulation, at the end of any interval of 12 sec the entire localised region around the removed ball comes back to thermal equilibrium. The model predictions in the graph of Figure 6 are useful particularly in the sense they provide a quantification of the magnitude of these other thermal cycles.

4.2. Thermal Behaviour at Local Package Level

To fully understand the thermal behaviour at local package level, temperature predictions obtained with the local thermal model of de-balling are used. Figure 7 provides results on the localised thermal phenomena by illustrating the temperature contours at the local package domain at 1 sec and 4 sec. In the time interval from 4.4 sec to 12 sec the temperature in the entire local domain converges to isothermal state at temperature 65°C.

Model results confirm that the thermal heat dissipates mainly through the top most metal layers of the BGA substrate. Thermal effects are much localised. The temperature front from the pad does not propagate substantially down towards the die. At the high temperature time point, 4 sec into the de-balling, first-level solder joints stay unaffected by the thermal load.
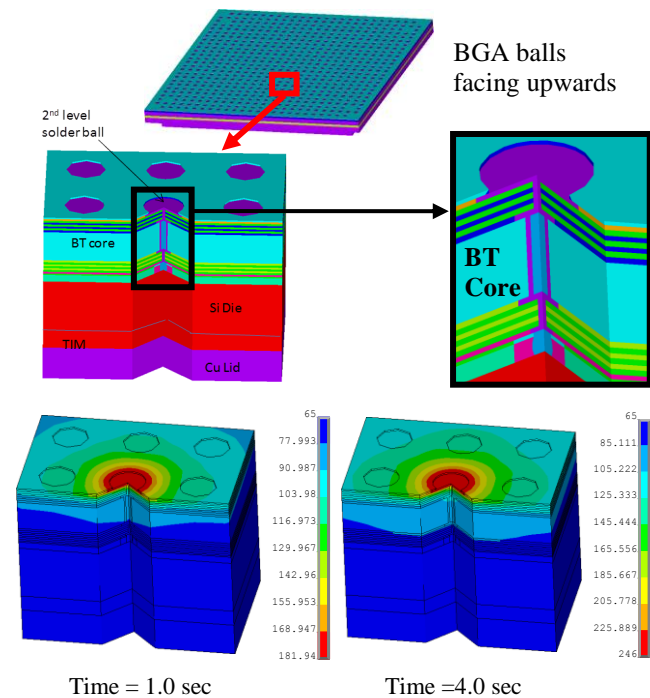
Fig. 7. Temperature (°C) results at local level at discrete time points during de-balling.

The temperature front from the deballed BGA pad does not propagate substantially down towards the die, and most critically towards the first-level solder balls in the package. The temperature at first-level solder joints located beneath the processed BGA pad increases from the initial temperature condition (65°C) to only 72°C over the 4 seconds period of

solder ball heating. Results provided with Figure 8 show that even if direct conductive path in the form of macro-via is presented in the BT core of the package substrate beneath the processed pad, the temperature range at the via remains low (in the range 90°C to 106.6°C) during the deballing process. The heat from the melted balls dissipates dominantly through the nearest metal planes of BGA substrate. Heat transfer in vertical direction and towards the centre of the package is extremely limited. Therefore, a major conclusion from this study is that the process of solder ball removal using hot nitrogen does not impose any risks for the integrity of the first-level BGA interconnects and for die related damage.
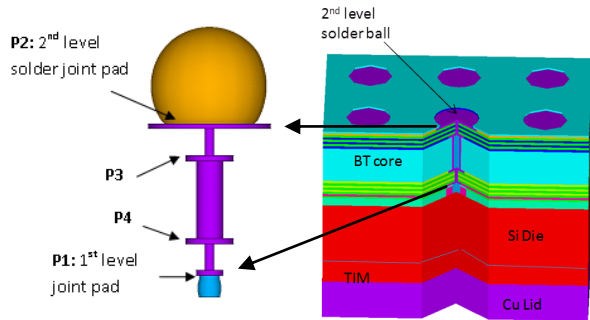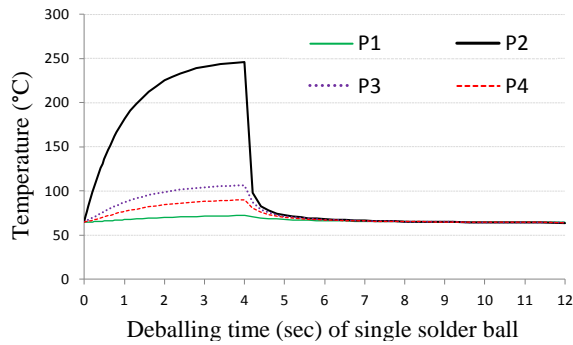
Fig. 8. Temperature graph over de-balling time (12 sec) at four different locations defined along the shortest copper path from second-level ball to first-level solder joint.

The only risk indicated with the thermal results, if any, is for the integrity of the solder mask in the vicinity of the deballed pad which is heated to 250°C.

Preliminary temperature measurements using thermo-couple (TC) placed in the BGA substrate indicated that peak temperature is about 81°C. Measurements agree well with the model predictions but while model assumes only localised heating in the processed pad region, the TC data suggests hot nitrogen may in fact spread over region of more than few BGA pads.

5. CONCLUSIONS

Thermal modelling of hot nitrogen deballing of BGAs confirmed that the process causes only localised thermal effects, thus making this process a very safe choice. A principal conclusion is that first-level solder interconnects and IC dies in flip-chip BGA components are not affected by the heat transfer from vacuum de-balling.

The temperature front from the processed BGA substrate pad propagates laterally only as far as few adjacent pads. Therefore, a negligible impact on inducing thermo-mechanical stresses outside the local region of the processed pad should be expected. In the extreme case of direct heat path between a second-level ball and first-level solder interconnects the heat conducted towards package die is still minimal. In the case of the investigated package temperature at first-level interconnects increases marginally. The main disadvantage of the de-balling is the slow rate of processing but given that the process brings no risk of damage this may be accepted with many applications.

REFERENCES

[1] European Union, (February 13, 2003). "Directive 2002/96/EC of the European Parliament and of the Council of 27 January 2003 on Waste of Electrical and Electronic Equipment," Official Journal of the European Union, pp. L37/24-L37/38

[2] European Union, (February 13, 2003), "Directive 2002/95/EC/ of the European Parliament and of the Council of 27 January 2003 on the Restriction of the Use of Certain Hazardous Substances in Electrical and Electronic Equipment," Official Journal of the European Union, pp. L37/19-L37/23

[3] D. Pinsky, M. Osterman, and S. Ganesan, "Tin Whiskering Risk Factors", IEEE Transactions on Components and Packaging Technologies, Vol. 27, No. 2, June 2004, pp. 427-431

[4] J. Lau, "Ball Grid Array Technology", Pub. McGraw-Hill Professional, 1994

[5] http://www.pactech.com

[6] ANSYS 12.0, www.ansys.com