Identification of the Maximum Allowable Compression Load Per Solder Interconnect

Tae-Kyu Lee, Mohan Nagar, Gnyaneshwar Ramakrishna,
Adnan Mahmud, Cesar Escobar, David Chan
Cisco Systems
CA, USA
taeklee@cisco.com

ABSTRACT
Recent developments in high performance electronic devices require miniaturized but more complex configuration with higher power and higher functionality, consequently, an increasing number of interconnects are needed in a given size substrate and with the complexity, a package needs an increase in body size, which poses a constant challenge to the board assembly process due to the increase of the package warpage. The equally challenging process is the heatsink attachment, which need higher level of compression load to achieve a full contact at the already high coplanarity package, inducing a higher level of localized compression load per component and ultimately per solder joint. In this study, an attempt to identify the maximum quantitative allowable compression load per solder ball in a large component configuration with 600µm in solder ball diameter is investigated. Knowing the maximum compression load, which does not induce any potential degradation to the performance and stability of the solder joint is an important data point to assure a reliable heatsink attachment process. A series of solder ball compression tests are performed and the damage accumulation per solder ball loading conditions are evaluated. The correlation between compression load level, distribution of the compression load, and the damage accumulation are compared in a series of cross section analyses using optical polarized imaging and Electron–backscattered diffraction (EBSD) imaging. The analysis revealed the potential range of allowable compression load per solder ball and ultimately the maximum allowable compression load per component.

Key words: Compression load, solder ball, EBSD, microstructure

INTRODUCTION
The thermo-mechanical stability of interconnects in modern electronic components are the baseline of the reliability of the device and system. Along with the continuous and accelerated trend for miniaturization, more functionality, and higher performance with the demand for higher signal speed, these interconnects stabilities are constantly challenged in telecommunication and data center sector systems due to the increase into larger and more complex package configurations. [1-4] The increase rate of once Flip-chip BGA (FCBGA) components with single die configuration in a range of 45x45 to 55x55mm^2 body sizes are now commonly above 65x65mm^2 with 2.5D interposer technology and system in package (SIP) technology-driven configuration with even further increase in substrate size. [5] Various external stress factors exist, from electro-current driven degradation mechanisms in smaller interconnect size and contact area, more variable and vigorous end-use conditions, and higher level of external stresses induced by small form factors are a few examples. [5,6] The increase in body size and multiple silicon die configuration associated with 2.5D interposers or Fan-out package configuration not only introduce more interconnects but also increase the package coplanarity and dynamic warpages. Associated with high performance and complex functionality, these larger size components need reliable and stable heatsinks, which often induce higher compression load per component for full contact due to the warpage and higher coplanarity. Although compression load guidelines and application notes are established and used, the question of what level of maximum load is safe for localized solder joints in these large components needs an investigation to assure the long-term reliability of the component. The approach of observing the damage accumulation rate per given compression load can be segmented into several layers in a given component: The heatsink to thermal interface material (TIM) interface, solderbump and microbump interconnect interface, dielectric interfaces, die to die interfaces, and substrate to solder joint interface. To investigate the overall damage evaluation induced by a given compression load need an in-depth study per layer, which is not expected to be simple. But to provide a baseline for segmented interface investigation, the study presented here is solely on the solder joint level damage accumulation, which is at the interface between the solder bulk and package side interface. The interface between the solder bulk and the PCB pad interface not considered in this study. The approach taken in this study can be explained and listed as follows; (1) identification of allowable maximum compression load per solder joint by observing the contact region damage accumulation identification via stress intensity detection using EBSD analysis. Each solder joint is compressed with a given load and the contact area is cross sectioned to identify any damage development signatures. Multiple thermal cycles are applied after compression load to observe any sub-grain boundary development, which can be used as a signature of damage accumulation. (2) once the maximum compression load is identified, the level of compression load per a defined area with a grid of solder joints can be derived. But given the fact that in most cases the compression load is not uniformly applied, and also the

height of each solder ball varies, a way of calibrating the induced anisotropic compression needs consideration. (3) Once the titling/anisotropic compression is considered, the total maximum compression load can be realized.

EXPERIMENTAL PROCEDURE

FCBGA components attached with 600μm diameter SAC305 solder balls on SOP surface finish pads with 508μm pad openings are used in this study. To identify the deformation rate per compression load, a series of solder balls are compressed individually with a 600μm diameter flat surface Cu rod compression pin, as shown in Figure 1. The compression pin was attached to a custom pin holder and force controlled with a multi-bond tester (Nordson Dage4000plus). The load was applied with a 30gf/sec displacement rate up to the target maximum load followed by 30 seconds holding. The load fluctuation during the holding period was ±0.1gf. After each compression load implementation, the top surface of each solder balls was observed with optical microscopy to identify the load induced deformation area as shown in the insert picture in Figure 1.

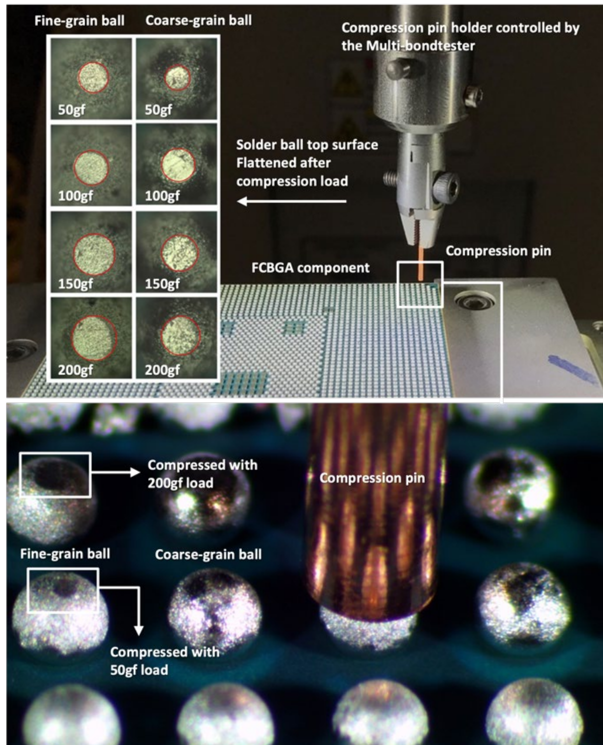

Figure 1. Compression load test set-up.

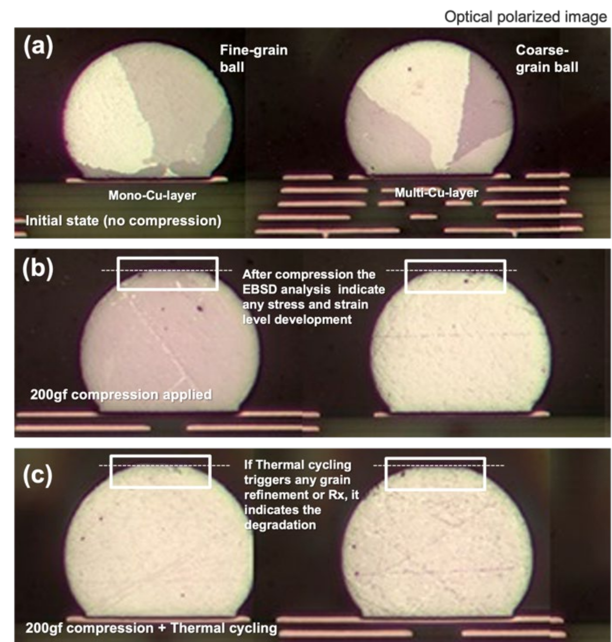

Figure 2. Cross-section optical images (a) no compression load applied fine-grain and coarse-grain solder balls, (b) after single compression with 200gf load, and (c) 200gf load applied then thermal cycled

The deformation impacted area are measured by the affected and flattened area diameter. As shown in Figure 2(a), coarse grain solder balls are on multi-layered Cu pads compared to the finer grain solder balls which are on top of mono-layered Cu structures. The difference between mono-layered and multi-layered Cu induced difference in thermal distribution and affects the heating and cooling rate during ball attachment process, which ultimately affects the grain size per solder balls. Each categorized solder balls are subject to compression load to a pre-selected load then went through the optical microscopy observation, then subjected to cross section for EBSD analysis as shown in Figure 2(b). Selected solder balls after compression load went through thermal cycling follow-up with EBSD analysis (Figure 2(c)). The compression load versus the deformation area correlation can be seen in Figure 3. Figure 3 presents the deformed and flattened top surface of selected solder balls after applying a certain level of compression load per fine grained solder balls and coarse-grain solder balls. This correlation is also used to identify the actual load experienced by each solder ball after applying a total compression load on an array of solder balls. As shown in the plot, the fine-grain solder balls have smaller diameter circles per given compression load compared to the coarse-grain solder balls. This is expected due to a finer grain structure have higher hardness values and resulted in less deformation and less flattening at the solder top region. The flattened surface diameters are collected to plot the correlation between the diameter and the applied load which is shown in Figure 3(b) and (c) per grain size condition. The load applied solder balls are subjected to cross section and EBSD analysis. Selected solder ball compression contact regions were analyzed with EBSD-GROD (grain reference

orientation deviation) map and strain contour maps to identify any damage accumulation induced by the given compression load. Additional thermal cycles with -40 to 100°C thermal profile were applied to observe any sub-grain boundary development triggered by thermal release to identify the allowable maximum compression load per solder ball. Compared to single solder balls, to observe multiple solder balls in an array format, a 4x5 array of solder balls was cut out from the component and compression load was applied to the target BGA array with a 10x10mm contact area aluminum block. The correlation between each solder ball per overall compression load were identified by measuring each solder ball top deformation area diameter. For longer term compression load effect observation, the selected compression loads were applied with 37.7gf/sec and a holding time of 5minutes. The z-axis displacements were observed during the constant compression loading time until the saturation were reached.

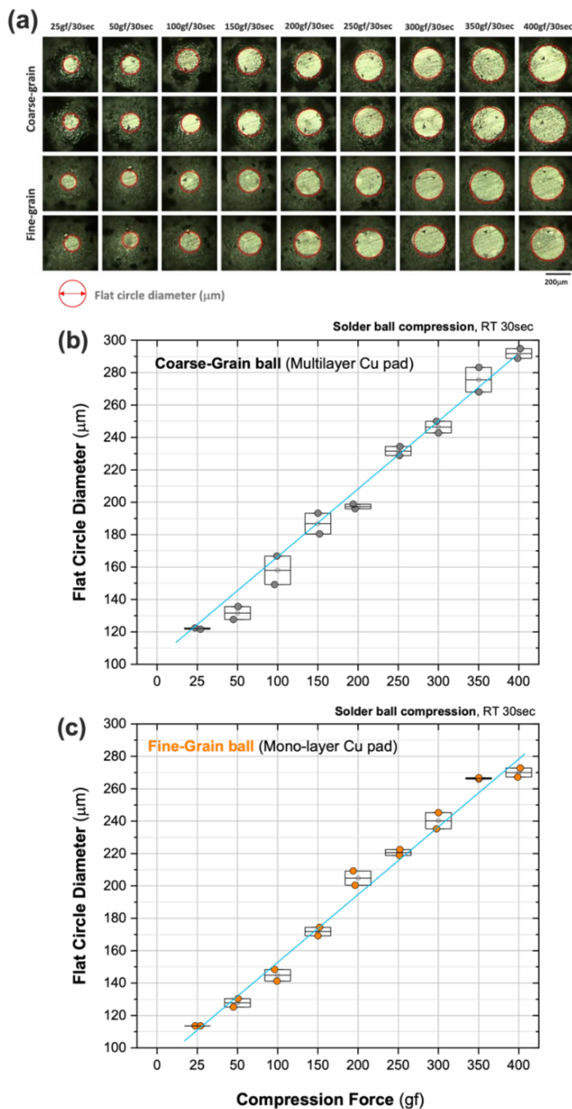

Figure 3. (a) Top view optical images per applied compression load. Compression load versus fattened circle diameter for (a) coarse-grain solder balls, and (c) fine-grain solder balls.

RESULTS AND DISCUSSION

To identify whether the given compression load is high enough to induce damage at the solder joints, an assessment was performed to identify any sub-grain structure development at the solder package interface region. As shown in Figure 4, during thermal cycling in Sn based solder alloys, an accumulation of defects initiates low angle sub-grain boundaries, which develops to high angle boundaries and eventually to crack initiation. [7] This phenomenon is closely associated with the initiation of sub-grain boundary developments. In other words, if the sub-grain boundary develops at the package side interface inside the solder joint bulk region, it indicates that the crack initiation is close to occur.

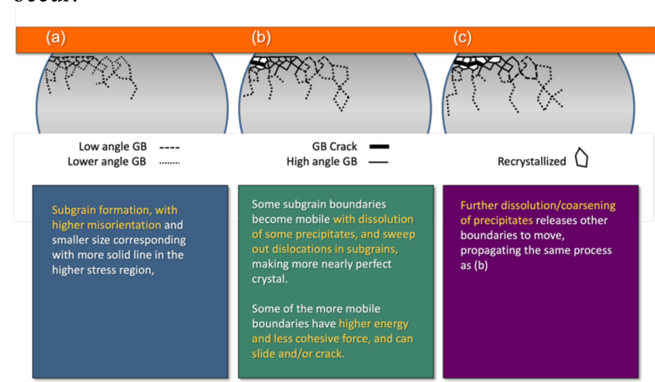

Figure 4. Damage accumulation and mechanism to crack initiation and propagation during thermal cycling. [7]

Since the test samples are not a solder joint but stand-alone solder balls, instead of observing the shoulder of the solder joint package side interface, the top of the stand-alone solder balls are the target region to be observed whether the compression load induced deformation is enough to cause the sub-grain boundary development to initiate. To observe whether a given compression load causes a sub-grain boundary, two steps of deformation process are considered, a simple mono-step compression to a pre-selected load level, then an additional thermal cycling stress step, which might trigger developments of sub-grain boundaries. Whether these steps induce sub-grain boundary development, the microstructure at the target regions were observed through EBSD analysis and assessed whether the compression load induced damage was enough to trigger a degradation at the solder ball. The EBSD grain reference orientation deviation (GROD) maps and the band contrast images revealed the sub-grain structure developments, and the strain contour map indicated any localized plastic deformation region. Figure 5 presents the compression load applied solder balls cross section SEM and EBSD images per compression load levels from 0gf (no compression) up to 250gf. The first column are the SEM images, second column are the EBSD inverse pole figure (IPF) images, and the third column are the Grain reference orientation deviation (GROD) map images, which

indicates the residual stress per loading condition. As shown in Figure 5(a) and (b) GROD map images for the no compression and 50gf applied solder balls, the top region of the solder balls did not reveal any specific increase of residual stress (indicated with white boxes). But once 100gf and 150gf load were applied as shown in Figure 5(c) and (d), a higher level of residual stress is shown at the top region and near the pad to solder ball interface as indicate by white arrows. Further, the level of residual stress shows an increase with 200gf and 250gf load conditions (Figure 5(e) and (f)). The top region reveals an increase of residual stress along with the neighbor region where the compression load was applied. For both 200 and 250gf load, not only the residual stress is increased but also the EBSD-IPF images show a localized deformation at the compression load applied top region per solder ball (indicated with white arrows in Figure 5(e) and (f)). The higher magnification EBSD images are presented in Figure 6 from the region indicated orange boxes in Figure 5(d)(e) and (f). Although the 150gf applied solder ball has an increase in localized residual stress shown in Figure 6(a), the spread is limited compared to the 200 and 250gf load applied solder balls in Figure 6(b) and (c). With 200gf loading condition, the contact area experience a higher residual stress region associated with sub-grain boundary development observed the EBSD GROD and strain contour maps. Based on this observation, the allowable compression load per solder ball can be derived as 150gf. In other words, with higher than 200gf load compression, the solder ball, and eventually the solder joint can initiate joint degradation. But with a simple compression load and the correlation to the residual stress is not sufficient to confirm the allowable load level per solder ball.

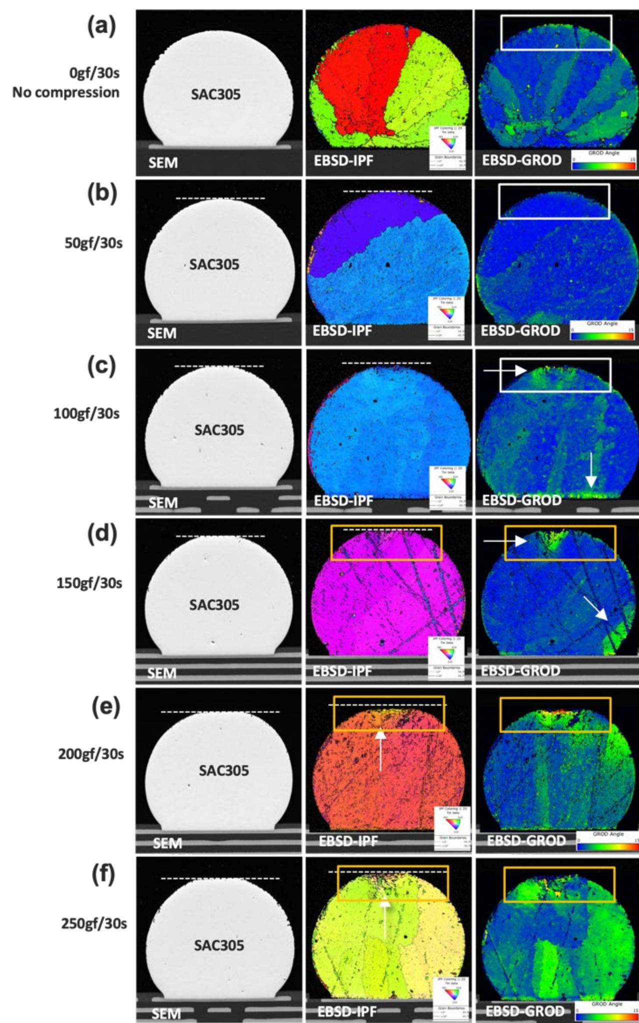

Figure 5. Compression load applied solder balls cross section SEM and EBSD images per compression load levels from 0gf (no compression) to 250gf/30s.

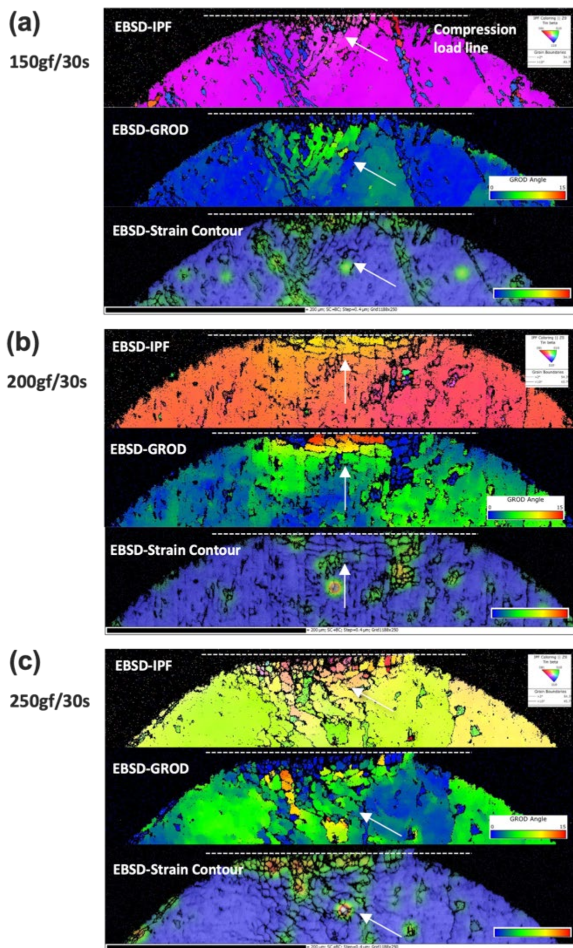

Figure 6. Higher magnification compression load applied solder balls cross section SEM and EBSD images per 150gf, 200gf, and 250gf compression load from orange boxes indicated in Figure 5(d)(e) and (f).

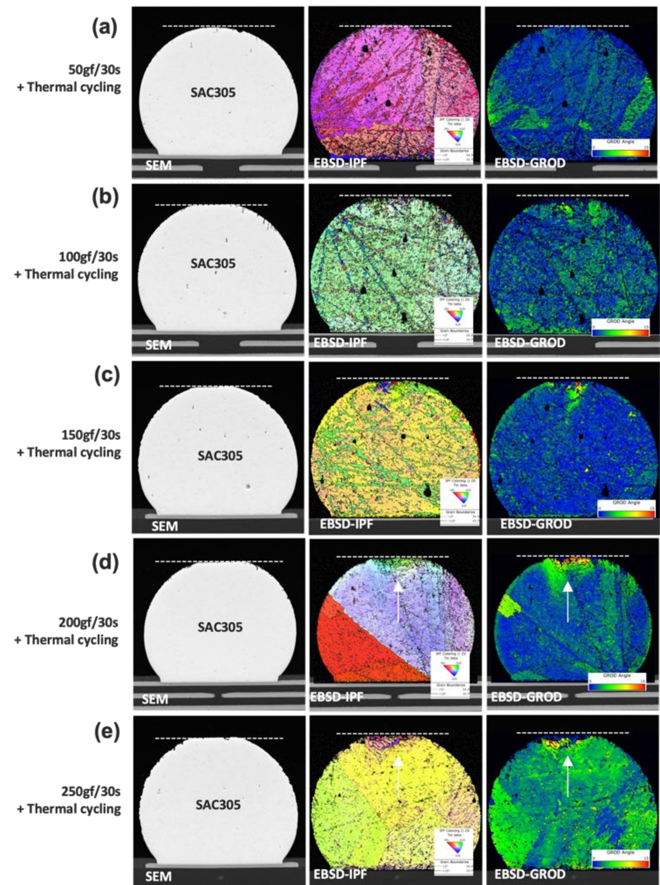

Figure 7. Compression load applied and thermal cycling added solder balls cross section SEM and EBSD images per compression load levels from 50gf to 250gf.

The load applied region needs to expose to thermal cycling to see whether the given load induces a sub-grain boundary development. For identification, the once compressed solder balls are subjected to thermal cycling and the cross sections per loading condition are observed as shown in Figure 7 and 8. Figure 7 shows the SEM and EBSD images after selected loading condition and added thermal cycling. 20 cycles of -40 to 100°C thermal profile cycles are added after compression loading. As shown in Figure 7(a)and (b), the 50gf and 100gf compressed and thermal cycled solder balls did not reveal any sub-grain development and increased residual stress regions. But with compression load with 150gf, a localized region begins to reveal a higher residual stress region and a sub-grain development as presented in Figure 7(c) and Figure 8(b). The level of damage accumulations are more developed with 200 and 250gf loading conditions as shown in Figure 7(d)(e) and Figure 8(c)(d) at the white arrow indicated regions. Based on these series of observations the allowable compression load can be 100gf per solder ball since the 150gf induce no significant damage at the single compression load but reveal a possible degradation microstructure once the thermal cycling occurs.

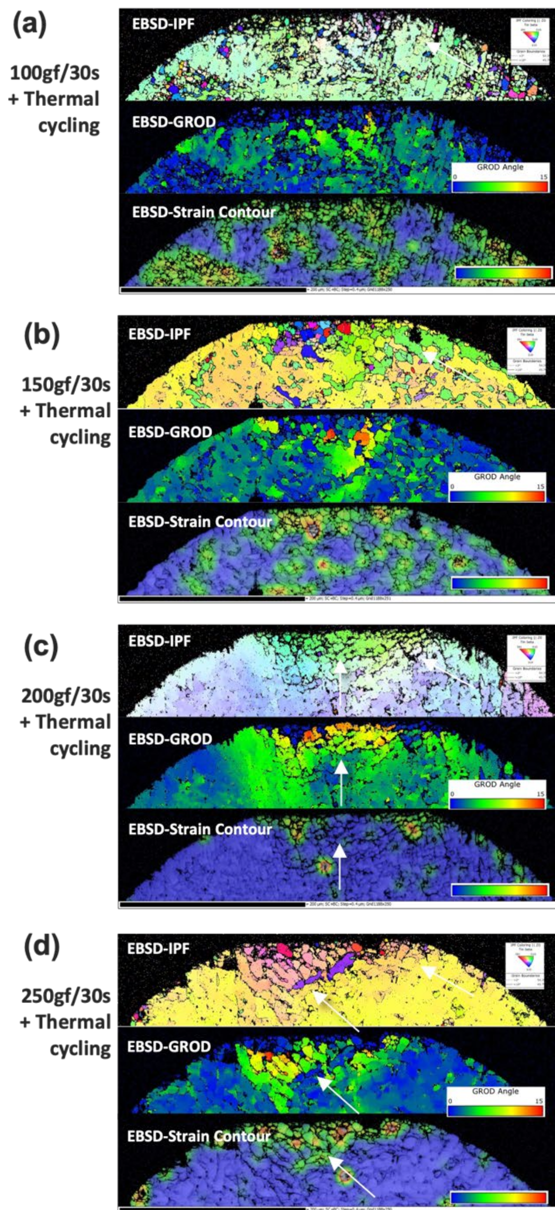

Figure 8. Higher magnification compression load applied then thermal cycling added solder balls cross section SEM and EBSD images per 100gf, 150gf, 200gf, and 250gf compression load from orange boxes indicated in Figure 7(b)(c)(d) and (e).

Once the maximum allowable compression load per solder ball as 100gf is identified, a maximum load per component with a BGA array can be derived. For example, if the component is configured with 10x10 solder array BGA configuration, the maximum allowable compression load will be 100ea x100gf = 10,000 gf = 10kgf and for 50x50 solder array component maximum load will be 2500ea x 100gf = 250000 gf = 250kgf. But this simple calculation based on the single solder ball allowable maximum compression load can be misleading. Since the applied total compression load in most cases are not-uniform, thus the realistic allowable compression load is expected to be lower. Figure 9 is a 4x5 array BGA component compressed with 3kgf load for 30 seconds. In an ideal condition, all of the 20 solder balls are expected to experience the average compression load of 150gf, which is close to the allowable compression load per SAC305 solder balls. But observed in Figure 9, the flatten circle diameter varies from 0 to 270µm in diameter instead of the ideal condition with a 170µm diameter flatten surface diameter.

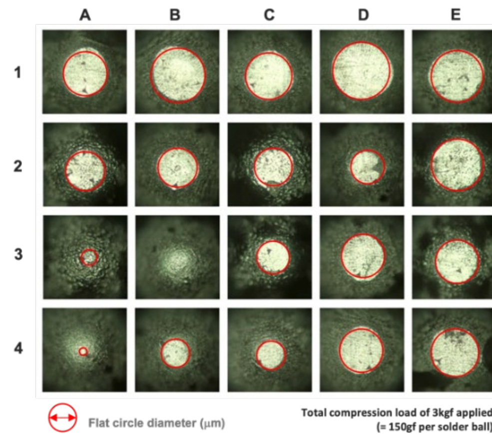

Figure 9. Top view optical images of 4x5 array BGA component compressed with 3kgf load for 30 seconds.

Figure 10 presents the compression load induced flatten surface diameter correlation for both ideal and measured compression load condition. Given the fact that the large diameter flatten circles are around 250-270µm, the maximum compression load which was applied to those solder balls are already over 300-350gf using the conversion plot in Figure 3(b). This level of compression load is enough to trigger and initiate a solder joint degradation and it is expected that this level of load will negatively impact the solder joint reliability. Thus, the allowable compression load needs to be lower than the 150gf per solder ball setting, which is lowering the 3kgf total compression load.

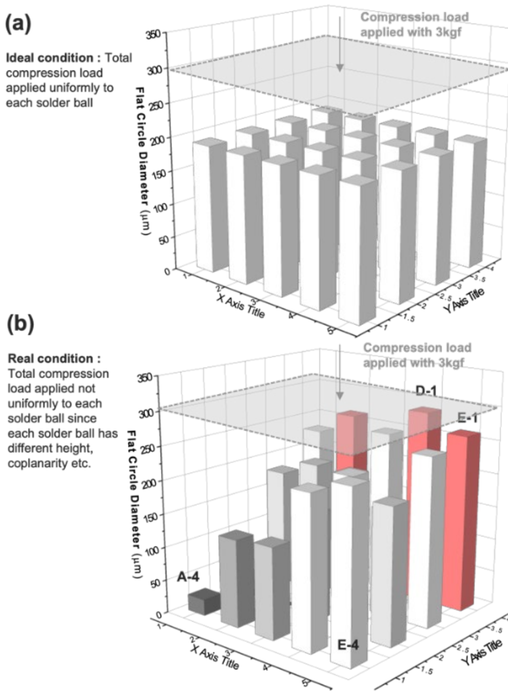

Figure 10. Compression load induced flatten surface diameter correlation per (a) ideal condition and (b) real and measured compression load condition.

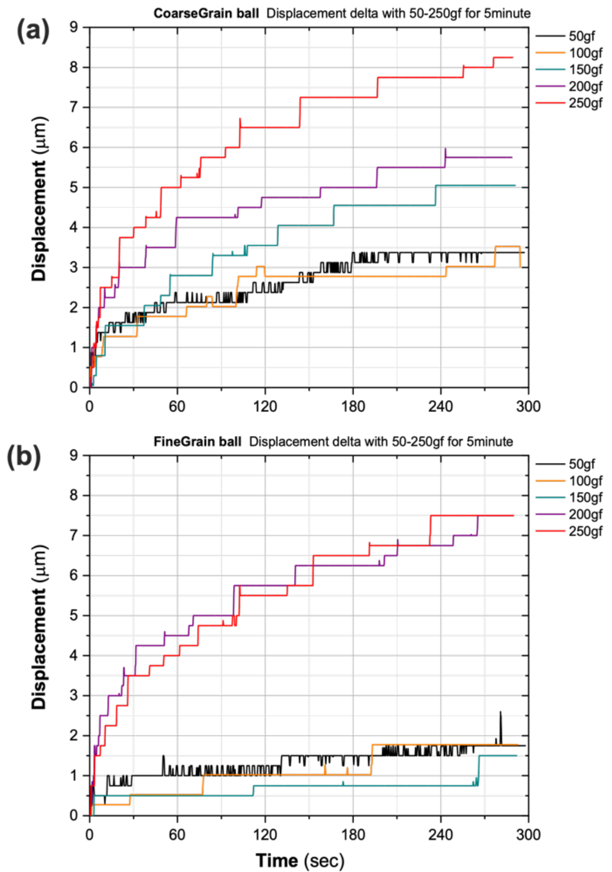

Figure 11. Z-axis displacement changes during applied constant compression load per (a) coarse-grain and (b) fine-grain solder balls.

For example, a 60x60mm^2 FCBGA component with 3340 solder balls can endure 334kgf of compression load in ideal condition, but in reality the coplanarity of the component after assembly, the solder joint height difference, Cu pad configuration induced solder ball grain size, and the compression load distribution during compression induce a different amount of actual load per solder joint at specific locations. Due to these factors, the 334kgf needs to be drastically reduced to a lower allowable compression load limit. The approach to find the maximum compression load introduced and discussed in this study is still work in progress and need to consider several additional factors. But the approach which is investigated in this study to derive the quantitative load limit per solder ball provides the right direction to get the accurate estimation for an allowable maximum compression load per a given component. Among additional factors the long-term compression load impact needs to be considered. As shown in Figure 11, the applied compression load duration was increased to 300 seconds to reach the z-axis displacement saturation per coarse and fine-grain solder balls. The displacement for the fine-grain solder balls shows a lower level of displacement (Figure 11(b)) compared to the coarse -grain solder balls (Figure 11(a)). An overall assessment for even longer-term creep performance per solder joint needs a closer investigation to confirm the allowable maximum compression load.

CONCLUSION

An approach to identify the maximum allowable compression load per large component were proposed and performed. Knowing the maximum compression load, which does not induce any potential degradation to the performance and stability of the solder joint, is an important parameter to assure a reliable heatsink attachment process. A series of solder ball compression tests are performed and the damage accumulation per solder ball loading conditions were evaluated. The correlation between compression load level, distribution of the compression load, and the damage accumulation are compared in a series of cross section analyses using optical polarized imaging and Electron–backscattered diffraction (EBSD) imaging. The maximum compression load per single SAC305 solder ball was identified as 150gf with 30 second duration and 100gf considering the sub-grain microstructure development after applying thermal cycling. But the derived allowable maximum compression load per component cannot be simply derived with multiplying the allowable maximum load per single solder ball by the solder ball numbers. To get to the realistic condition embedded compression load condition, the distribution of the actual experienced load per solder balls need to be considered. The study presented here has additional and various consideration factors, which need to

implement for an accurate and final quantitative maximum load per component, but the first step and identification of the allowable maximum compression load per solder ball provides a direction towards a valid assessment methodology and baseline.

REFERENCES

1. J. Glazer, Int. Mater. Rev. 40, 65 (1995)
2. D.R. Frear and P.T. Vianco, Metall. Trans. A 25A, 1509 (1994)
3. V. Fiori, K. Ewuame, S.Lallois-Garreignot, H.Jaouen, and C.Tavernier, Microelectronics Reliability, 54, 764 (2014)
4. H.G. Song, J.W. Morris Jr., and F. Hua, JOM 56, 30 (2002)
5. J. Savic, M. Nagar, W. Xie, M. Ahmad, D. Senk, A. Bansal, N. Islam, P. Oh, R. Pendse, H. Choi, and S.Lee, Proceeding of IEEE 62nd Electronic Components and Technology conference, San Diego, CA, 450-456 (2012)
6. H.K. Kim and K.N. Tu, Phys. Rev. B 53, 16027 (1996).
7. Tae-Kyu Lee, Thomas Bieler, Choong-un Kim and Hongtao Ma, Fundamentals of Lead-Free Solder Interconnect Technology, Chapter 5, Springer, 131 (2015)