Productivity Boost and Optimum R_c Control in Wafer-Level Packaging enabled by HEXAGON

The megatrend of more miniaturized electronic devices highlights the importance of low and stable contact resistance (R_c). To classify the capability of HEXAGON as High-Throughput UBM/RDL Technology, we report the details of experimental work on throughput and R_c previously presented and published at ECTC 2023 by Evatec's ***Dr. Carl Drechsel***, ***Dr. Patrick Carazzetti***, ***Carl Wang***, ***Dr. Juergen Weichart*** and ***Ewald Strolz***, as well as ***Kay Viehweger*** from Fraunhofer IZM-ASSID (Moritzburg, Germany)[1].

INTRODUCTION

An impressive attribute of the digital transformation are the continuously growing amounts of data being processed. While the average monthly data volume per stationary broadband connection in Germany was only 47 GB in 2015, it had risen to 142 GB in 2019. In 2023 it was already 287 GB, which is about 10 GB per day. However, this does not include the mobile data volume, which averaged 7.2 GB per month in 2023 [2] and might increase tremendously in near future. Although these values differ considerably from country to country, there is nevertheless a clear global trend, which shows the requirement of data capacities that exceed our current average consumption by far: Operating an autonomous vehicle generates for example a data volume of around 5 GB per minute, which would amount to around 7.2 TB (= 7200 GB) per day. And if you want to estimate the amount of data being processed in a Smart Factory, you need to familiarize yourself with dimensions like Petabyte (1 PB = 10^6 GB) or Exabyte (1 EB = 10^9 GB).

The demand for higher data volumes goes hand in hand with a continuous trend for higher performance devices, increasing power efficiency and miniaturization. Recent progress is here explicitly based on wafer-level packaging (WLP) [3], where typically protective layers, electrical connections as well as the packaging itself are implemented before dicing the wafer into single chips. Hence, resulting packages are in similar dimension as the die itself, which is understood as wafer level chip-scale packaging (WLCSP). In advanced packaging this requires multi-layer and fine pitch packaging designs that push the dimensions of vertical interconnects and redistribution layer (RDL) technologies below 5 µm. But since smaller interconnects critically complicate the electrical performance requirements, the control of a low contact resistance (R_c) is becoming more and more important. Additionally, smaller scaling is in state-of-the-art packaging platforms coupled with higher organic loads that needs to be avoided, such as organic or oxide contamination of the metal interfaces.

Here we review hardware improvements that facilitate low R_c results (< 2.0 mOhm) on HEXAGON used in WLP for the creation of under bump metallization (UBM) and RDL. Keeping a low R_c, it is possible to reach a throughput of 80 wafers/hour (WPH). Key drivers for this optimization are an improved airlock design that effectively shrinks pumping and venting times, resulting in a cycle time of approx. 34 seconds per wafer. This allows other processes to be trimmed. Additionally, a new indexer rotation concept of the HEXAGON shortens the chamber-to-chamber transfer, which reduces the risk to re-contaminate the contact pads at the end of the ICP sputter etch step. The electrical performance relevant for R_c calculation is measured on single-contact Kelvin resistors with via diameters from 25 down to 10 micrometers. Experimental data show for all via diameters stable R_c results within a lot of 25 wafers at a high-throughput run of 80 WPH.


evatec
SEQ 300

I. EXPERIMENTAL METHODS

A. Overview on HEXAGON

To process the wafers, on which later the R_c is evaluated, a high-volume manufacturing HEXAGON platform, especially optimized for boosting the wafer handling speed, has been used. Figure 1 shows the configuration of the HEXAGON system.

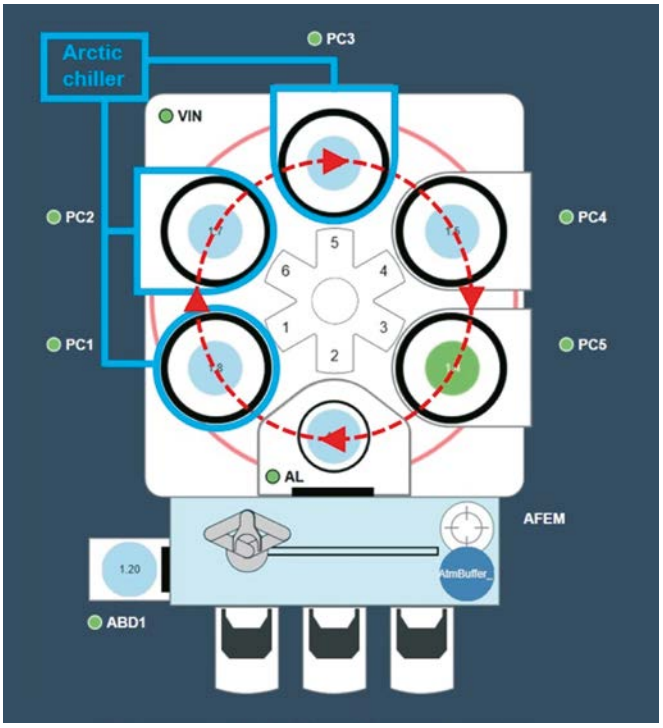

Figure 1: Configuration of HEXAGON platform, indicating pre-treatment (ABD, PC1, PC2, PC3) and PVD (PC4, PC5) modules.

The HEXAGON platform consists of two major components: An atmospheric front-end module (AFEM) and a vacuum indexer module (VIN). The AFEM is equipped with three loadport modules (LPMs) to introduce front opening unified pods (FOUPs) to the system, an aligner to correct eccentricity for every incoming wafer, a buffer as deposit place for pasting wafers and an atmospheric batch degasser (ABD). The VIN is equipped with an airlock (AL) to introduce wafers from AFEM and five process chambers (PCs). PC1 is an arctic cooling station to reduce the wafer temperature between ABD and the two inductively coupled plasma (ICP) sputter etch modules in PC2 and PC3. The ICP sputter etch procedure relies on physical bombardment with Ar+ ions in an ignited Ar plasma. Since ICP sputter etch processes are very critical to the thermal budget, the pedestal and the chamber shields are actively cooled to -30 °C by an external chiller unit. In PC4 and PC5 the PVD process of Ti as adhesion and Cu as seed layer for a final downstream process is performed.

B. Process Description

The procedure of the entire process, performed as BKM on single wafers, is described in Table I. Following the BKM, every wafer is first aligned and next transferred into the ABD, where a degassing phase takes place to avoid adsorbed water molecules on the wafer. There, a batch of 28 wafers can be processed simultaneously while a laminar N2 flow guarantees constant peak temperatures on the

Module	Physical process
LPM	FOUP is loaded to the AFEM & Wafer is taken
Aligner	Eccentricity check & notch alignment
ABD	20-30 min degassing at 120-130 °C
AL	Wafer is transferred from AFEM to VIN
PC1	Cooling to ≈ 50 % of ABD temperature
PC2	15 nm removal by ICP sputter etch
PC3	15 nm removal by ICP sputter etch
PC4	100 nm Ti PVD
PC5	300 nm Cu PVD
AL	Wafer is transferred from VIN to AFEM
LPM	Wafer is put back & FOUP is unloaded from AFEM

Table 1: Overview of BKM process flow

wafers, independently of the position and the total number of wafers loaded. The total degassing time and temperature is product specific and typically ranges 20-30 minutes and 120-130 °C. After degassing, wafers are transferred to the AL and thus get into the VIN. Next, they enter the first process chamber (PC1) and cool down to roughly half of the ABD peak temperature. This is important to ensure that the thermal budget of the ICP sputter etch process in PC2 and PC3 starts from a preferably low point. Since etching gives the highest amount to the thermal heat on the substrate, the total etch amount required is equally distributed to PC2 and PC3. In addition to pure Ar+ ion bombardment, the entire etching process also includes in-situ cooling and purging steps to reduce the peak temperature. After finishing the etching and the Ti/Cu PVD in PC4 and PC5, the wafers get back to AL and are therefrom re-transferred to the FOUP.

C. R_c Measurement

The R_c measurements are performed on four-terminal Kelvin resistors, built on 300 mm Si wafers. An overview of all steps necessary for the completion of a full R_c test vehicle is illustrated in Figure 2.

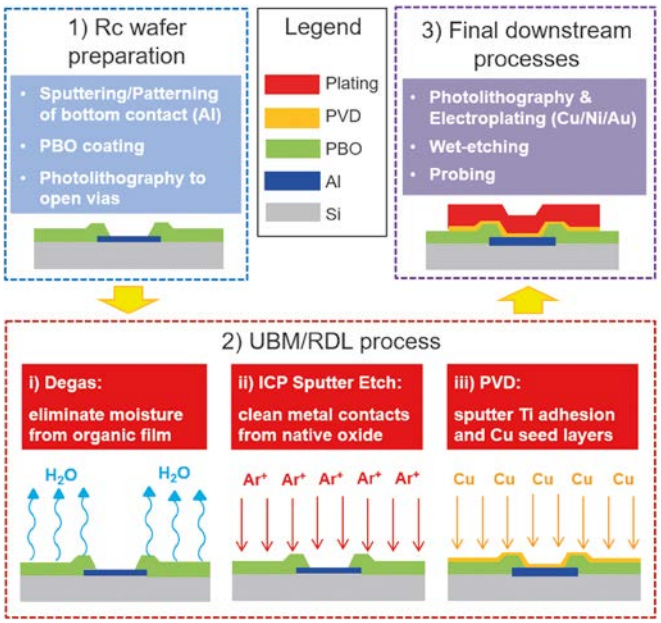

Figure 2: Fabrication flow indicating process steps of Rc test vehicles on 300 nm Si wafers.

Each R_c test vehicle consists of a patterned metallization layer (1 µm sputter-deposited Al) on a Si substrate, which acts as basic redistribution line (RDL0). On top of that a 9 µm thick passivation film of PBO is deposited. By means of photolithography, a contact opening via is created on top of the RDL0 (see Figure 2, step 1). The smaller the diameter of the opening pad d0, the more sensitive to the R_c is the Kelvin resistor. This is followed by the BKM flow (see Figure 2, step 2). Cleaning the pads and depositing Cu as RDL1 without oxidation is essential for an optimum operation of the device and is revealed by a low Rc. Next the final downstream process follows, which includes a thickening of RDL1 by 3 µm Cu electroplating and an addition of 2 µm Ni and 0.5 µm Au by selective plating of the probing pads to improve the reliability of the latter R_c measurement (see Figure 2, step 3).

An enlargement of a test vehicle used for this study and a sketched cross section of an opening via is shown in Figure 3. Each test vehicle has a total number of 121 cells, containing series of Kelvin structures differing in via diameter (d0), overlap size between RDL0 and RDL1 that is equal to 0.5(d1 – d0) and connector width c. The R_c measurement is performed at a voltage in the order of 1 mV with a vertical probe card. It is leveled by a fully automatic probe station (TEL Precio) and the high-precision measurement is conducted by an Advantest V93k test head with analog VI cards. A detailed description of the measurement procedure and calculation of the R_c can be found in [4].

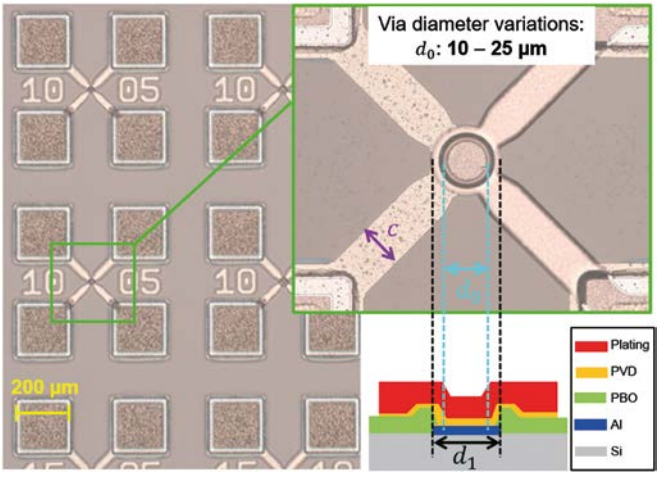

Figure 3. Enlargement of Rc test vehicle with four-terminal Kelvin resistors

As standard qualification procedure, for each experiment condition, a lot of 25 wafers is processed with the described BKM. Before the lot is started, a single Al pasting wafer is etched for 90 s to ensure similar conditions in PC2 and PC3. PC4 and PC5 are warmed up with PVD processes on 4 dummy wafers. In the lot, the R_c test vehicles are located in slots 1, 13 and 25, while all other slots are filled with PBO dummy wafers.

II. HARDWARE FEATURES OF HEXAGON

The optimized, high throughput of HEXAGON is the consequence of a very short cycle times. The cycle time itself can be divided into handling time and process time.

The handling time denotes the duration a PVD system needs to transfer all wafers to the next process chamber between completion of the last process and start of the next process. In an indexer system, it consists of a synchronized downward pedestal movement, a 60° clockwise rotation of the indexer carousel and an upward pedestal movement.

The process time is product specific and depends mostly on the required etch amount and PVD thicknesses. However, in standard applications of UBM/RDL deposition, the etching sequence is usually the process time bottleneck. Within the ICP sputter etch sequence, there is also time for pumping and purging before and after pure etching in order to prevent contamination of the etched pads and to keep the thermal excursion as low as possible.

A. Optimized Indexer Unit

The reduced handling time of HEXAGON has its origin in the design of the central rotating device. While in previous indexer systems the support paddles on the indexer carousel were connected to a motor by gears (see Figure 4a), the new design places the paddles directly on the rotary table driven by a central servo motor (see Figure 4b). Only this causes a time gain of approx. 2 s. Additionally, the potentially higher attrition of the gear wheel components is counteracted. The total handling time was shortened to approx. 9 s.

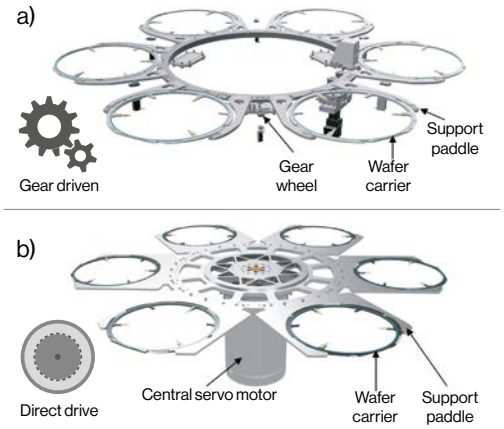

Figure 4. Comparison between previous and new design of central rotating device; a) previous gear driven indexer unit; b) new direct drive approach.

B. Optimized Airlock Cycle

HEXAGON also uses an optimized AL design. Compared to an earlier indexer tool design, the AL volume V0 has been reduced to 25%. Instead of one large turbo pump, two smaller ones, each with its own pump valve, are installed (see Figure 5).

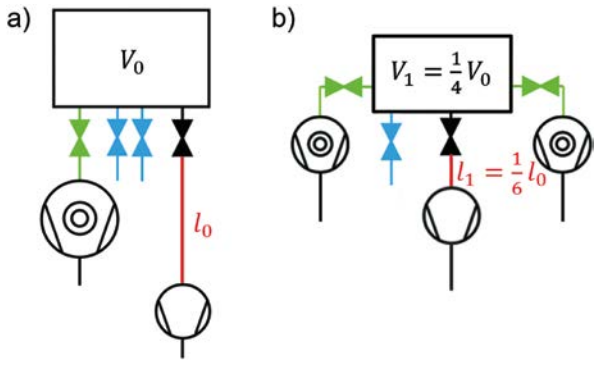

Figure 5. Scheme of AL in a) previous design and b) new design. Different colors indicate pumping valves (green), venting valves (blue) and fore-vacuum valves (black). The pumping line to fore-vacuum pump is marked in red.

Furthermore, the pump line l0 between the fore-vacuum valve and the roughing pump was shortened by mounting the roughing pumps directly on the main frame of the tool. All in all, this enabled the pumping time to be roughly halved. The number of venting valves at the AL has been reduced from 2 to 1, since the reduction of the AL volume does not require more venting capacity. Even so, the venting time at AL is also reduced by about half.

By the optimization of AL, a cycle time of approx. 34 s can be reached, which corresponds to a peak throughput of approx. 105 WPH when the system runs in dry-cycle mode. A pressure-time analysis of the AL, including pumping, handling time, venting and wafer exchange is illustrated in Figure 6.

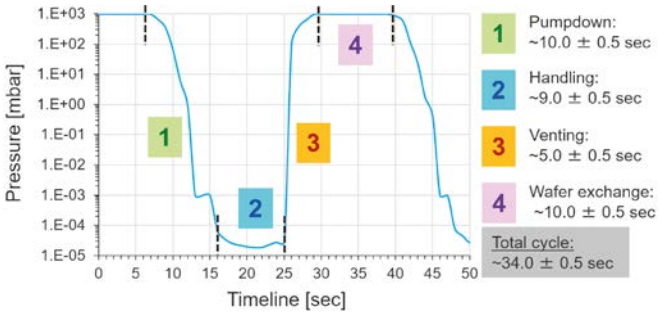

Figure 6. Pressure vs. time analysis of a single AL cycle time.

III. THERMAL MODEL AND R_c ANALYSIS

A. Thermal Model

The temperature flow over the whole BKM stack can be simulated with a thermal model [5]. Applying it to throughputs of 44.2, 54.5, 69.2 and 80.3 WPH (see Figure 7), allows to determine the peak temperature after ICP sputter etch process and over the full stack process. From this, a correlation between peak temperatures and the respective throughput can be determined (see Figure 8).

B. R_c Analysis

For all throughputs, on which the thermal model is applied, the R_c is measured on a lot of 25 wafers. It indicates on which values and how constant the R_c remains in the face of higher chamber temperatures during continuous operation. Figure 9 shows for different throughputs the averaged result for via diameters of 10, 15, 20 and 25 µm on wafers in slots W#1, W#13 and W#25.

All R_c results for investigated throughputs are in the order of 1-2 mΩ. Only the run at 44.2 WPH shows higher values at a via diameter of 10 µm for wafers #13 and #25. Regarding to the examined via diameters, a larger diameter generally leads to lower R_c values. Within a FOUP, the R_c values remain extremely stable at 80.3 WPH, while a slight increase can be observed at 69.2 WPH and a significant increase at 44.2 WPH. For 54.5 WPH the R_c values even decrease very slightly. The statistical error ranges are very small for the measurements with via diameters of 20 and 25 µm, for 15 µm they are a bit larger. For 10 µm, on the other hand, the error ranges are more than twice as large. This arises from the systematically more inaccurate measurement conditions for smaller via diameters.

Overall, the R_c results for high-throughput runs on HEXAGON reach a very low level (< 2 mΩ). Although the peak temperature is highest at 80.3 WPH (see Figure 9), no negative impact on the R_c behavior can be detected, which is attributed to faster process and handling times, reducing the potential time for a recontamination of the etched pads.

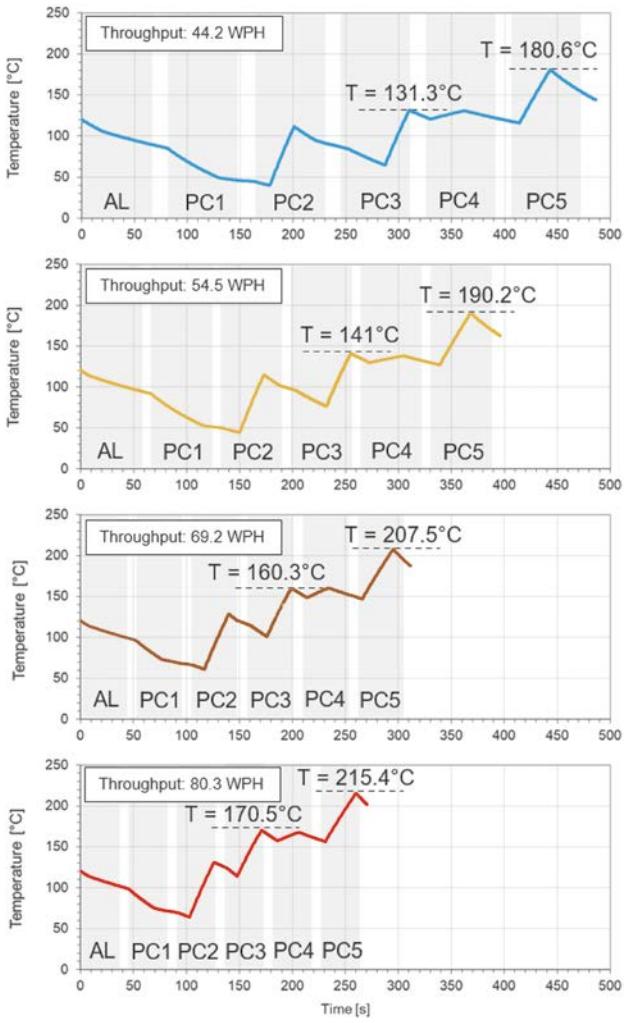

Figure 7. Thermal model for different throughputs, indicating the maximum temperature after ICP sputter etch and for the entire process.

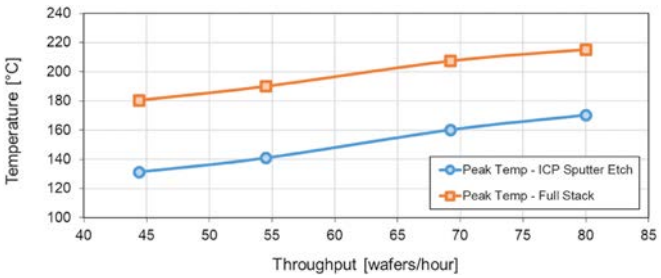

Figure 8. Peak temperatures after ICP sputter etch process and within total stack for different throughputs.

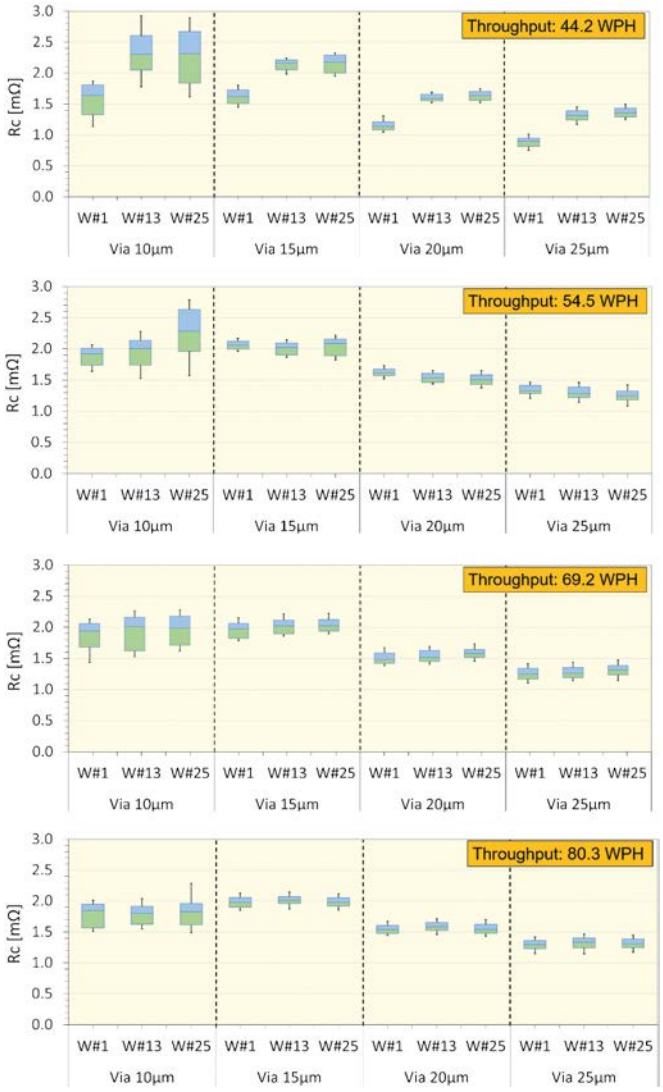

Figure 9. Rc results for different throughputs and via diameters. W#1, W#13 & W#25 indicating the slot position of the test wafer within a FOUP.

CONCLUSION

Due to continuous miniaturization in high-end and advanced packaging, the control of R_c has uttermost importance. In this work we discussed the achievement of low R_c results on a HEXAGON system used for UBM/RDL in WLP. At the same time, the throughput is increased to 80 WPH based on hardware improvements.

The hardware optimization can be divided into a new design of the central rotating device and the AL. The first leads to a reduction of the handling time, the second leads to a reduction of the AL cycle time, i.e. the duration the AL needs to pump down, handle, vent and exchange a wafer. Based on experimental data, the R_c was measured for throughputs between 44.2 and 80.3 WPH on single-contact Kelvin resistors with via diameters of 25 down to 10 µm, located in the 1st, 13th and 25th slot of a FOUP. Furthermore, the temperature profile of a test wafer was simulated for all throughputs, from which the peak temperature after etching process and for the full stack process is determined.

It could be proven that an increase in throughput is associated with an increase of the peak temperatures, but at the same time no increase in R_c is observable. In particular, the via diameters of 10 and 15 µm show a lower R_c at high throughput. The comparison of the R_c results for the 1st, 13th and 25th wafer exhibits no increase for wafers within one full lot. In summary, we provide evidence that an increase in throughput on a HEXAGON platform goes hand in hand with an optimum R_c control. This performance is achieved even though the full stack peak temperature exceeds 215 °C.

REFERENCES

1. C. Drechsel, P. Carazzetti, C. Wang, K. Viehweger, J. Weichart and E. Strolz, "Optimum R_c Control and Productivity Boost in Wafer-Level Packaging Enabled by High- Throughput UBM/RDL Technology", 2023 IEEE 73rd Electronic Components and Technology Conference (ECTC), Orlando, FL, USA, 2023, pp. 126-131.
2. German federal network agency: Telecommunications Annual Report 2023.
3. B. Keser and S. Kroehnert, "Embedded and Fan-Out Wafer and Panel Level Packaging Technologies for Advanced Application Spaces", Wiley, Hoboken, NJ, USA, 2022.
4. P. Carazzetti, J. Weichart, A. Erhart, M. Elghazzali and E. Strolz, "Evaluation of ICP Sputter Etch with Reducing Atmosphere for the Improvement of Rc in UBM/RDL Applications", 2020 IEEE 70th Electronic Components and Technology Conference (ECTC), Orlando, FL, USA, 2020, pp. 1711-1717.
5. P. Carazzetti et al., "Impact of Process Control on UBM/RDL Contact Resistance for Next-Generation Fan-Out Devices", 2017 IEEE 67th Electronic Components and Technology Conference (ECTC), Orlando, FL, USA, 2017, pp. 909-916.