

General

HPC Packaging Platform Including Compression Molding

M. Adler¹, K.-F. Becker¹, T. Braun¹, O. Hoelck¹, A. Gaebler¹, U. Maass¹, I. Ndip¹, K. Zoschke¹, M. Schiffer¹, M. Schneider-Ramelow²

¹ Fraunhofer Institute for Reliability and Microintegration IZM, ² Technical University Berlin

Keywords: High Performance Computing, Wafer Level Packaging, Silicon Interposer, Compression Molding, Epoxy Molding Compound, Warpage

<https://doi.org/10.4071/001c.143003>

Journal of Microelectronics and Electronic Packaging

Vol. 22, Issue 3, 2025

Efficient high-performance computing (HPC), especially for applications in the field of artificial intelligence, requires the use of specialized computing and memory chips, short transmission paths and high packing densities. A key technology for HPC is therefore the integration of processors and memory stacks in a single System-in-Package (SiP) using silicon-based interposers.

Manufacturing concepts have been developed for such a package, each of which utilizes different process flows to build an HPC module. Following a holistic approach, the process flow is evaluated from the processing of the interposer with high density organic redistribution layers, compression molding and balling. Evaluation methods were defined for the individual process steps and criteria established to enable the selection of suitable process steps and combinations for an optimized package. The process flow options were realized and analyzed for a reference HPC module. An important parameter for all these process steps is warpage, since an exceeding warpage can make processing impossible.

For compression molding, a key process step, the influence of different materials and process parameters on the warpage was therefore investigated.

The results of the process development are presented for the backend process flow with an emphasis on compression molding process step, an evaluation of the process variants is carried out according to the previously developed criteria and the most suitable variant of the process step for later production is selected on this basis.

I. INTRODUCTION

Artificial intelligence, big data and complex modeling require efficient high-performance computing modules. In addition to the development of ICs that can perform individual computing operations with high efficiency and speed, the interconnection of these chips to their environment is also one of the main challenges of high-performance computing (HPC).

For instance, short transmission paths between computing and memory chips and high signal integrity are required to take full advantage of a leading-edge chip architecture. Advanced packaging is therefore a central building block for the realization of HPC modules. One option for the advanced packaging of HPC modules is the integration of computing and memory chips in a single System-in-Package (SiP) (Matsuo et al. 2000; Lau 2021; Bhattacharya and Rao 2023). In addition to the integration of one computing and memory chip each, such a packaging platform can also enable the integration of a larger number of chips or chiplets and thus also makes a significant contribution to the emerging chiplet infrastructure.

According to this approach of heterogeneous integration in a SiP, one or more ICs that are optimized for specific computing operations such as stencil processing units (Datta et al. 2008; Krüger et al. 2011; Donofrio et al. 2009)

or GPUs can be assembled together with High Bandwidth Memories (HBM) on a silicon interposer (Chidambaram 2019; Wang et al. 2023). Interposers and chips are connected with microbumps with a pitch in the range of 50 μm (Lee et al. 2018). In addition to the use of silicon as interposer material, glass and organic interposers, possibly with integrated glass or silicon bridges, can also be considered (Li et al. 2016; Usman et al. 2017). Silicon interposers are currently wide-spread due to their ability to support high density wiring and existing process lines.

For the packaging of such HPC modules in a high-volume production, wafer level processes such as redistribution layer (RDL) processing and compression molding are key technologies. While wafer level packaging imposes major advantages such as good scalability, it also includes challenges such as warpage that increase with the size of wafers (Braun et al. 2024, 2023b; Braun and Hoelck 2022) which have to be addressed in a proper process evaluation.

II. PROCESS FLOWS FOR HPC MODULES

In order to investigate the manufacturing of an HPC module with the previously described setup of processing unit and HBM on silicon interposers and to analyze optimal process control, a simplified demonstrator setup is devel-

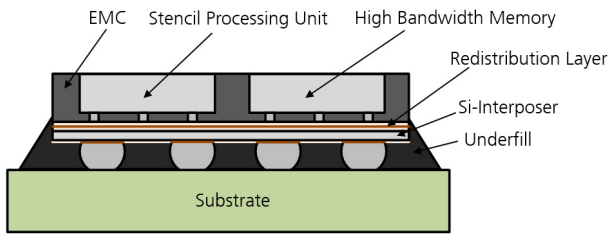


Figure 1. Schematic structure of the demonstrator

oped (Figure 1). A stencil processing unit and an HBM are integrated on a silicon interposer. In subsequent development steps, the platform can be extended to integrate a larger number of chips on an interposer using the chiplet approach.

To set up the demonstrator module, the HBM and the processing unit, each with a thickness of 725 μm , are mounted on an interposer. The gap defined by the bump height with a distance of 35 μm between interposer and chips. It can either be filled with an underfill for thermo-mechanical stabilization or the underfilling is integrated into the encapsulation process by compression molding in the so-called molded underfill. The interposer wafer initially has a thickness of 725 μm . After thinning the wafer and exposing the through silicon vias (TSV), the final thickness is 100 μm . The thinning and complete processing of the interposer can take place either before or after compression molding. These two possible process variants are shown in Figure 2 and Figure 3. The 200 mm interposer wafer is fully processed before molding, including thinning and pad generation, in the process flow shown in Figure 2, and subsequently molded with a cap of 190 mm diameter, using the rim of the wafer for the sealing of the mold cavity. During this process, a glass carrier can be used to stabilize the 100 μm thin interposer. Handling during the balling process allows flexible wafer size, so the thin rim of the interposer wafer can be removed by sawing after carrier debonding to avoid uncontrolled fracturing.

For the process flow shown in Fig. 3, further process steps such as pad generation are necessary after molding. For these processes, a standard wafer size such as 200 mm is mandatory for handling and machine processes. Thus, the interposer wafer is cut down to 190 mm before molding to so that a 200 mm wafer can be molded. This allows handling of a stable 200 mm wafer without the necessity to remove a thin rim as in the first process flow.

The finished module has a size of 25x15 mm² and is assembled to an organic substrate after solder balls have been applied. The ball grid array layout is based on current standards; typical ball sizes are around 300 μm and typical pitches are 500 μm ("IPC 7095 - Design and Assembly Process Implementation for BGAs," n.d.). Here, too, underfilling and the mounting of a heat sink or lid with suitable thermal interface material on the back of the module for optimized heat dissipation follow.

The limit for further processing of the encapsulated wafers in the balling process at IZM is typically 2 mm total warpage for a mold wafer with $d = 200$ mm; the aim of

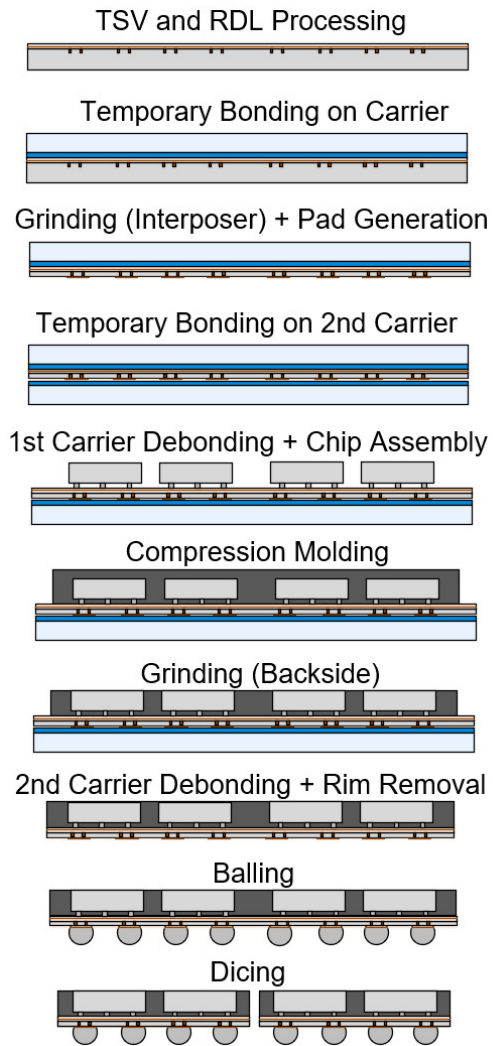


Figure 2. Process flow with interposer wafers thinned at the beginning

process development is to remain below this threshold value. A holistic assessment of the whole process chain is carried out to secure a high performance and reliability of the HPC packaging platform.

III. INTERPOSER LAYOUT AND PROCESSING

To provide sufficient high wiring density for the interconnection of computing and memory devices on interposer, Fraunhofer IZM has further developed its multi-layer routing technologies (Zoschke et al. 2024). The multi-layer routing technology is originally adapted from well-known redistribution technology for fan in wafer level packaging featuring electroplated copper for the wiring and thin film polymer as intra- and interlayer dielectric. Standard redistribution technology typically has relaxed wiring density with 5 μm thick copper lines having minimum 12 μm lines/space and large vias in 10 μm thick polymer, realized by direct lithographic structuring using a photosensitive polymer.

However, for the interconnection of ICs with high IO-density, a higher wiring density is needed. To meet this

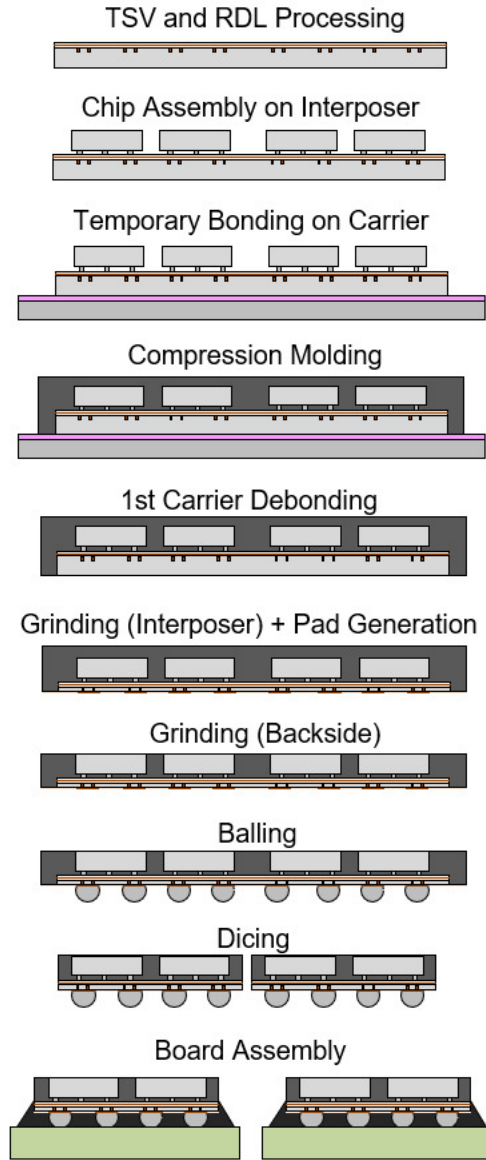


Figure 3. Process flow with initially unthinned interposer wafers

requirement, the multi-layer routing technology has been scaled towards higher density during the last couple of years.

Routing on an interposer device for HPC applications has to reach densities of 4 μm lines and space and beyond to meet the interconnection requirements of typical compute and memory chiplets. The required upscaling of routing density was enabled by reduction of layer thickness for copper and polymer. Thinner layers allow use of thinner photo resists and reduce final warpage, so that finer structures are possible for both, wiring and vias.

With a restriction of the maximum copper height to 2 μm per layer and maximum polymer thickness to 6 μm per layer, the layout rules shown in Figure 4 were defined. Based on that, a line pitch of 8 μm with symmetric width and spacing of 4 μm is possible over multiple layers. The polymer vias, which are generated by laser ablation using a pre-defined hard mask, were defined with a diameter of 8 μm . Secondary layout rules such as overlaps and via spaces

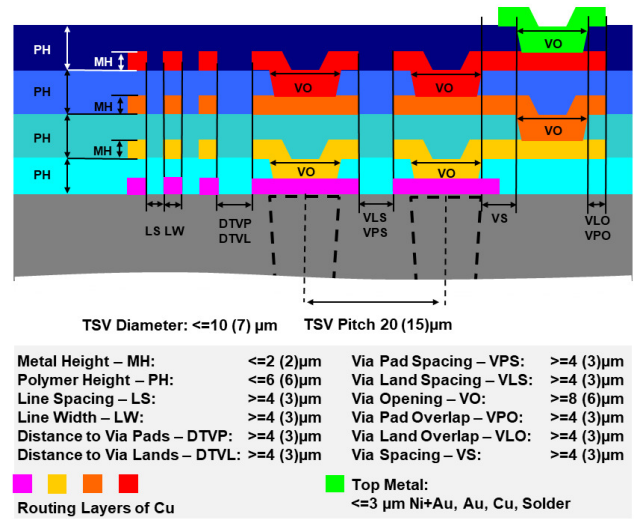


Figure 4. Design rules for fine pitch multi-layer wiring with 4 μm lines and space and 4 routing layers on TSV interposer (Zoschke et al. 2024)

can be also found in Figure 4. As it can be seen in the picture, the design rules support a possible TSV pitch of 20 μm for the interposer device. All numbers in brackets refer to the next technology release which can support a routing scenario of 6 μm pitch (3 μm lines and space) to support a 15 μm TSV pitch in the interposer device.

All pattern definition steps are done by mask aligner lithography and thin photo resists. Lithography based on mask aligner is a feature of this technology, because it is especially useful for realization of large devices which exceed typical stepper reticle sizes, so that no reticle stitching is required. In extreme cases full wafers can be re-routed and finally used as one device (Zoschke et al. 2017). Details of the process, especially related to the direct laser structuring approach of the polymer vias are described in (Zoschke et al. 2019).

IV. ENCAPSULATION BY COMPRESSION MOLDING

A. COMPRESSION MOLDING AND WARPAGE

Compression molding is a well-established process for large-area encapsulation in the area of wafer level and panel level packaging in volume production (Braun et al. 2023a). A variety of mold system suppliers are offering equipment for these processes, for this study a TOWA compression molding system in cavity down configuration has been used. For encapsulation at wafer level, an epoxy molding compound (EMC), an epoxy-based polymer filled with SiO_2 particles, is dosed into a cavity in wafer form (lower half of the tool). The components to be encapsulated are then placed on a carrier in the open mold (upper mold half), the mold is closed, the cavity is evacuated and the EMC is pressed into the cavity by moving the mold onto the component carrier (see Figure 5). The EMC hardens under pressure and temperature during the In-Mold Cure (IMC) and

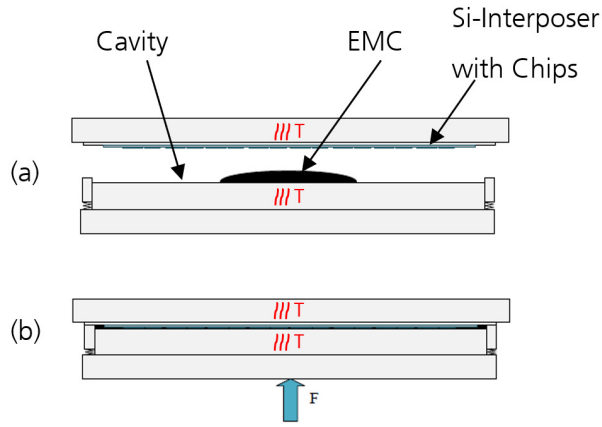


Figure 5. Principle of compression molding: (a) placing interposer and EMC in cavity; (b) closing cavity for In-Mold Cure (IMC)

Table 1. Material properties of the selected EMCs

Parameters	EMC A	EMC B	EMC C
Micrograph			
EMC type	Granular	Liquid	Liquid
Filler content [wt%]	82	82	83,5
Maximum filler size [μm]	20	10	5
CTE1 [ppm/K]	9	15	13
T_g [$^{\circ}\text{C}$]	190	138	150
G' [GPa]	19	16	21
In-Mold Cure IMC	10 min @ 130 $^{\circ}\text{C}$	10 min @ 110 $^{\circ}\text{C}$	10 min @ 120 $^{\circ}\text{C}$
Post-Mold Cure PMC	60 min @ 150 $^{\circ}\text{C}$	60 min @ 150 $^{\circ}\text{C}$	60 min @ 150 $^{\circ}\text{C}$

achieves sufficient mechanical strength to be removed from the cavity. The final cross-linking takes place in a subsequent post-mold cure (PMC) in an oven outside the cavity.

The epoxy molding compound must have a sufficiently high flowability for processing in compression molding in order to fill the entire cavity and all gaps between the components without voiding. In particular, in order to achieve the narrow gaps between chips and interposer as a molded underfill, small filler sizes and low viscosity are required at the same time. The materials shown in Table 1 were selected for the process discussed here. Material C in particular is characterized by a small filler size, but at significantly higher costs than materials A and B.

In addition to the electrical insulation of the individual contacts (e.g. whisker formation, short circuits due to conductive contamination), the EMC also serves to mechanically reinforce the assembled components under operating conditions, both during further processing and in the subsequent application. Therefore, sufficient material stiffness is required for stabilization and good thermo-mechanical adaptation of the EMC to the other materials present in the package. However, excessive stiffness can be critical, as warped wafers can no longer be processed if the stiffness

Table 2. Full-factorial experimental design

Factor	-	+
EMC	EMC A	EMC B
Overmold Thickness	60 μm	200 μm
Process Flow	Unthinned Interposer	Thinned Interposer

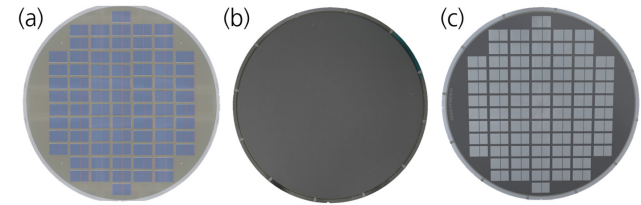


Figure 6. Demonstrator setup: (a) mounting the chips on interposer; (b) molding; (c) backgrinding to reveal chip backsides

is too high, for example if vacuum chucking and therefore processing of the warped wafer is not possible.

In order to keep warpage generally low, the CTE of the EMC should deviate as little as possible from the CTE of the silicon interposer (2.6 ppm/K). A difference in the CTE creates internal stresses when cooling from the processing temperature, resulting in warpage of the wafer at handling temperature. Since polymers generally have a significantly higher CTE than silicon, a low CTE is an important target value when selecting the EMC. Since the exact warpage and its limits depend on a variety of parameters such as stiffness, CTE and component structure, the warpage after compression molding and the processability of the warped wafers are analyzed in the experiment.

B. EXPERIMENTAL DESIGN OF MOLDING STUDY

To compare the process variants presented above, the process is run through as a demonstrator production. In order to also consider warpage as an important quality criterion, a test plan is developed in which the molding compound and the overmold thickness are examined as parameters in addition to the two variants of the process flow. The choice of material and the overmold thickness can potentially have a decisive influence on warpage, as they determine the degree of contraction of the upper layer of the structure and thus the internal stresses introduced.

The parameters of the full factorial test plan are shown in Table 2. The more cost-effective material variants EMC A and B are examined first.

For this demonstrator run to investigate the mechanical behavior, dummy chips made of silicon are mounted with a die attach material on the silicon wafers, which serve as interposers (Figure 6). This structure deviates from the final product, where the interposer is covered by a multi-layer RDL and chips are connected to the RDL with bumps, the bulk

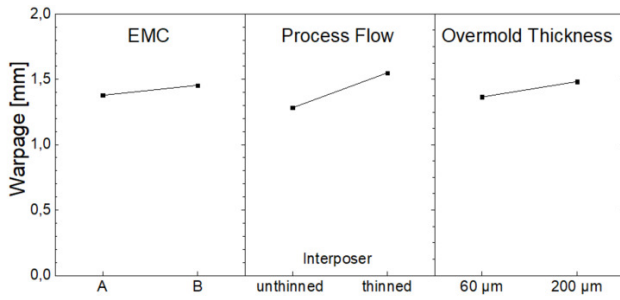


Figure 7. Main effect plot showing the largest effect for process flow

Compression molding is carried out using a Y-120 wafer molding machine (TOWA, Kyoto, Japan). The material manufacturer's recommendations according to Table 1 are used as the processing temperature and time. Directly after molding and after each process step, the warpage is measured using a precision 3-coordinate measuring system (OMS 600, Mahr GmbH, Göttingen, Germany).

C. EVALUATION

After molding, a warpage of the wafers in the range of 1.1 mm to 1.7 mm results for all process variants. The most pronounced effect on warpage can be observed for the factor process flow (Figure 7). A significant difference is reasonable due to the different mechanical structure. However, since additional steps for processing the interposer follow for the unthinned interposers, even a lower warpage can also be critical here.

The other parameters examined, EMC and overmold thickness, have a less significant influence on warpage. The slight increase in warpage associated with thicker overmold thickness is generally plausible, as more EMC also leads to greater stress due to the greater contraction of the upper layer of the package structure when cooling from the processing temperature. Although the average warpage values for both types of EMC are similar in the initial state, some differences become apparent especially after the following process steps.

In the further investigation of the first process flow, in which not yet fully processed interposers are overmolded, an increase in warpage can be seen for almost all of the wafers examined after the interposer has been thinned and the backside subsequently ground (compare with Figure 3). This can be explained by the decrease in the stiffness of the overall structure.

Grinding of the backside to reveal the die backsides could have the effect that stress is reduced since the layer of plain EMC with comparably high CTE is removed. However, the fact that backgrinding causes an increase in warpage can be explained by a further decrease in stiffness, while internal stresses generated by the surrounding ring of EMC around the interposer remain. Interestingly, this behavior is more pronounced for EMC A, even though its CTE is lower compared to EMC B. Other factors such as high mechanically relevant shrinkage and low stress additives (LSA) may have an impact on this effect.

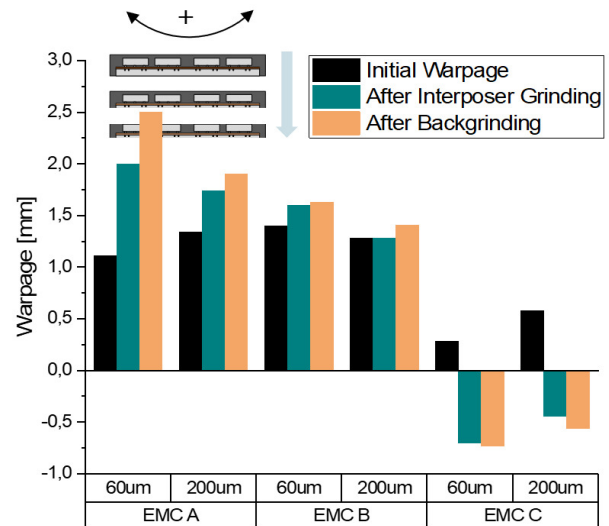


Figure 8. Warpage with initially unthinned interposer wafers after each process step

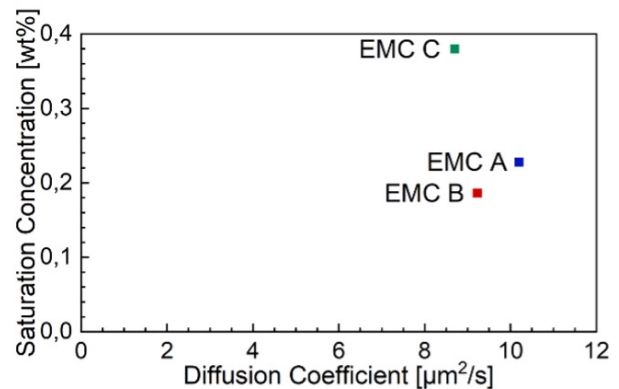


Figure 9. Saturation concentration and diffusion coefficient as parameters of moisture absorption of EMC A-C

Further processability proves to be problematic with this process variant; for example, the suction of the wafer to oxidize the bottom side of the interposer is not possible with the system used.

EMC C is tested as a possible alternative material. Figure 8 shows that a significant reduction in warpage can be achieved. After thinning the interposer, the measured warpage is negative, i.e. the wafer is warped in the direction of the interposer. This can possibly be explained by the expansion caused by the absorption of moisture. A higher maximum moisture absorption is supported by material analysis, that shows a significantly higher saturation concentration of water in EMC C compared to the other materials, even though the diffusion coefficient, indicating the absorption kinetics, is lower. Figure 9 displays these values, gained from dynamic vapor sorption analysis with a Discovery SA (TA Instruments, New Castle, DE), see (Braun et al. 2015) for metrology.

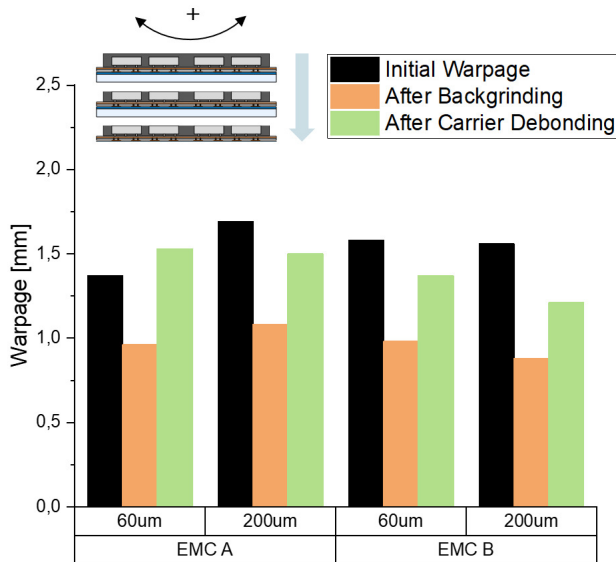


Figure 10. Warpage with initially thinned interposer wafers after each process step

For the process variant in which already processed interposer wafers are overmolded (cf. Figure 2), there is a clear reduction in warpage after the chip backsides have been exposed by grinding (Figure 10). This decrease is expected, as the tensile stress from the “pulling” EMC layer is relieved. Following debonding of the encapsulated wafer from the glass substrate—onto which the thinned interposer was originally mounted for processing—an increase in warpage occurs again, as anticipated, due to the reduced structural stiffness.

At no point in the process were issues observed with wafer handling via vacuum suction. This indicates that the process flow is advantageous compared to the previously described variant, despite the higher initial warpage after molding.

The process variant using a pre-thinned and completed interposer meets the <2mm warpage requirement for balling for both molding compounds used, and is further considered lower risk, as sensitive steps such as the final processing of the interposer’s bottom side are no longer required on an already warped, molded wafer.

This process flow was successfully validated in experiments using both materials A and B, as well as with both tested overmold thicknesses (60µm and 200µm). To further reduce the warpage without considering the material costs, the use of EMC C is also possible here, but not necessary.

Figure 11 shows an example of a module after molding, backgrinding, application of solder balls and dicing.

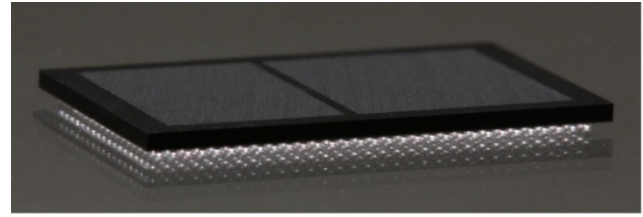


Figure 11. Demonstrator module after balling and separation ready for PCB assembly

V. CONCLUSION AND OUTLOOK

HPC modules require a suitable packaging platform to enable communication between multiple chips and to protect the components. Such a platform ranges from interposers with high-density electrical redistribution layers and solutions for encapsulation to balling and mounting on a substrate. The work presented here deals with suitable processes for the production of HPC modules with a focus on encapsulation. Encapsulation by compression molding is suitable for increasing the robustness of the interposer wafers during the process and for meeting the reliability requirements for assembly and operation in the application. However, the encapsulation process also introduces a warping that influences handling in the subsequent process steps or makes it impossible if the warpage is too large. With sufficient knowledge about the factors influencing the warpage, about the change in warpage during and after the respective process steps and about the limits of tolerable warpage, optimized parameters for the encapsulation process can be found. In the investigations carried out for this purpose with non-functional components, the process variant in which compression molding is carried out on a thinned interposer wafer was successfully demonstrated up to balling and separation, and both EMCs used were qualified as suitable encapsulation materials.

The next step is to investigate the molded underfill of the gap between the chip and interposer. In parallel, work on the simulation, design and construction of functional interposers with high-density redistribution layers is ongoing and the effect of warpage and stress on package level after the assembly on board has to be evaluated.

Submitted: January 06, 2025 EDT. Accepted: July 07, 2025 EDT.

Published: September 15, 2025 EDT.



REFERENCES

- Bhattacharya, S., and V. S. Rao. 2023. "Multi-Chiplet Heterogeneous Integration Packaging for Semiconductor System Scaling." *IEEE VLSI Technology and Circuits*. <https://doi.org/10.23919/VLSITechnologyandCir57934.2023.10185396>.
- Braun, T., K.-F. Becker, O. Hölck, H. Walter, and M. van Dijk. 2015. "Influence of Humidity on Reliability of Plastic Packages." In *European Microelectronics Packaging Conference (EMPC)*.
- Braun, T. et al. 2023a. "Trends on Large Area Advanced Packaging." In *40th International Conference of Photopolymer Science and Technology*.
- Braun, T., and O. Hoelck. 2022. "Manufacturing for Reliability of Panel-Level Fan-out Packages." In *Reliability of Organic Compounds in Microelectronics and Optoelectronics*, edited by W. D. van Driel and M. Yazdan Mehr, 517–31. Zurich: Springer.
- Braun, T., O. Hölck, M. Adler, M. Obst, S. Voges, and K.-F. Becker. 2024. "How to Manipulate Warpage in Fan-out Wafer and Panel Level Packaging." *IEEE 74th Electronic Components and Technology Conference (ECTC)*.
- Braun, T., O. Hölck, M. Voitel, M. Obst, S. Voges, and K.-F. Becker. 2023b. "A Closer Look to Fan-out Panel Level Packaging." *7th IEEE Electron Devices Technology & Manufacturing Conference (EDTM)*.
- Chidambaram, V. 2019. "Active Silicon Interposer for Heterogeneous Integration: System Scaling and Cost Effectiveness." In *IEEE Electron Devices Technology and Manufacturing Conference*.
- Datta, K., M. Murphy, V. Volkov, S. Williams, J. Carter, and L. Oliker. 2008. "Stencil Computation Optimization and Auto-Tuning on State-of-the-Art Multicore Architectures." *Proceedings of the 2008 ACM/IEEE Conference on Supercomputing*.
- Donofrio, D., L.R. Olike, J. Shalf, M.F. Wehner, C. Rowen, and J. Krueger. 2009. "Energy-Efficient Computing for Extreme-Scale Science." *IEEE Computer* 42:62–71. <https://doi.org/10.1109/MC.2009.353>.
- "IPC 7095 - Design and Assembly Process Implementation for BGAs." n.d. <http://www.ipc.org>.
- Krüger, J., D. Donofrio, J. Shalf, M. Mohiyuddin, S. Williams, and L. Oliker. 2011. "Hardware/Software Co-Design for Energy-Efficient Seismic Modeling." *SC '11: Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis*.
- Lau, J. 2021. *Semiconductor Advanced Packaging*. Singapore: Springer Nature.
- Lee, J., C.Y. Lee, C. Kim, and S. Kalchuri. 2018. "Micro Bump System for 2nd Generation Silicon Interposer with GPU and High Bandwidth Memory (HBM) Concurrent Integration." *IEEE 68th Electronic Components and Technology Conference (ECTC)*.
- Li, L., P. Chia, P. Ton, M. Nagar, S. Patil, and J. Xue. 2016. "3D SiP with Organic Interposer for ASIC and Memory Integration." *IEEE 66th Electronic Components and Technology Conference (ECTC)*.
- Matsuo, M., N. Hayasaka, K. Okumura, E. Hosomi, and C. Takubo. 2000. "Silicon Interposer Technology for High-Density Package." *Proceedings of IEEE 50th Electronic Components and Technology Conference (ECTC)*.
- Usman, A., E. Shah, N.B. Satishprasad, J. Chen, S. Bohlemann, and S.H. Shami. 2017. "Interposer Technologies for High-Performance Applications." *IEEE Transactions on Components, Packaging and Manufacturing Technology* 7:819–28.
- Wang, H., J. Ma, Y. Yang, M. Gong, and Q. Wang. 2023. "A Review of System-in-Package Technologies: Application and Reliability of Advanced Packaging." *Micromachines* 14 (6): 1149. <https://doi.org/10.3390/mi14061149>.
- Zoschke, K., M. Güttler, L. Böttcher, A. Grübr, D. Husmann, and J. Schemmel. 2017. "Full Wafer Redistribution and Wafer Embedding as Key Technologies for a Multi-Scale Neuromorphic Hardware Cluster." In *19th Electronics Packaging Technology Conference*. <https://doi.org/10.1109/EPTC.2017.8277579>.
- Zoschke, K., P. Mackowiak, H.D. Ngo, C. Tschoban, C. Fritsche, and K. Kröhnert. 2019. "High-Density Flexible Substrate Technology with Thin Chip Embedding and Partial Carrier Release Option for IoT and Sensor Applications." In *IEEE 69th Electronic Components & Technology Conference (ECTC)*.
- Zoschke, K., H. Oppermann, M. Schiffer, I. Ndip, K.F. Becker, and M. Adler. 2024. "Key Technologies and Design Aspects for Wafer Level Packaging of High Performance Computing Modules." In *IEEE 74th Electronic Components and Technology Conference (ECTC)*. <https://doi.org/10.1109/ECTC51529.2024.00340>.