

MIT Open Access Articles

Multi-strata subsurface laser die singulation to enable defect-free ultra-thin stacked memory dies

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Teh, W. H., D. Boning, and R. Welsch. "Multi-Strata Subsurface Laser Die Singulation to Enable Defect-Free Ultra-Thin Stacked Memory Dies." AIP Advances 5, no. 5 (May 2015): 057128.

As Published: <http://dx.doi.org/10.1063/1.4921205>

Publisher: AIP Publishing

Persistent URL: <http://hdl.handle.net/1721.1/120759>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of use: Creative Commons Attribution 3.0 unported license



Multi-strata subsurface laser die singulation to enable defect-free ultra-thin stacked memory dies

W. H. Teh,^{1,a} D. Boning,¹ and R. Welsch²

¹*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

²*Engineering Systems Division, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

(Received 11 February 2015; accepted 5 May 2015; published online 12 May 2015)

We report the extension of multi-strata subsurface infrared (1.342 μm) pulsed laser die singulation to the fabrication of defect-free ultra-thin stacked memory dies. We exploit the multi-strata interactions between generated thermal shockwaves and the preceding high dislocation density layers formed to initiate crack fractures that separate the individual dies from within the interior of the die. We show that optimized inter-strata distances between the high dislocation density layers together with effective laser energy dose can be used to compensate for the high backside reflectance (up to $\sim 82\%$) wafers. This work has successfully demonstrated defect-free eight die stacks of 25 μm thick mechanically functional and 46 μm thick electrically functional memory dies. © 2015 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported License. [<http://dx.doi.org/10.1063/1.4921205>]

Memory packaging is trending toward smaller form factor, increased heterogeneity, and higher performance. One method to address these challenges is to stack ultra-thin, high performance memory dies. However, ultra-thin dies generate higher sensitivity assembly interactions, which result in increased defect modes (e.g., chipping, die sidewall damage, microcracks) and decreased quality characteristics (e.g., die strength, kerf geometry, reliability). Advanced mechanical dicing or laser ablation dicing or combinations thereof (hybrid or sequential) based on dicing after grinding (DAG) or dicing before grinding (DBG) integration have been developed to help address the associated challenges.¹ Although they help, the fabrication of defect-free ultra-thin dies remains problematic because of surface perturbations (frontside and/or backside damage to the wafer depending on the integration approach) due to mechanical interactions or direct laser ablation with Si.

Stealth dicing (SD), a subsurface nanosecond pulsed, permeable laser die singulation technology,^{2–6} offers a potential solution. The principal one SD-layer method involves laser-induced “perforation” within the bulk Si, followed by fracture mechanics to physically “cleave out” the individual dies from within. SD avoids ablation defects from the use of conventional laser dicing operating at wavelengths that are highly absorbed by the materials to be diced. Those who have conducted experimental SD work have assessed processing quality based on photodiode characteristics,^{3,4,7} demonstrated the importance of focal plane depth,⁴ and reported SD-related defects,⁷ die strength,^{7,8} and stress distribution analysis.⁷ Most experiments have been performed to realize singulated 50 μm thick Si die using the SD After Backgrinding (SDAG) approach.^{3,4,7} Attempts have also been made to apply SD on 100 μm thick through-silicon via (TSV) wafers,⁹ subsurface machining of transparent materials,¹⁰ and to enable MEMS.¹¹ However, little work has been reported on enabling SD on high backside reflectance wafers where the amount of SD energy coupling into the wafer is extremely limited. In this letter, we report and demonstrate a multi-strata SD process on high backside reflectance (up to $\sim 82\%$) wafers for the fabrication of defect-free eight die stacks of 25 μm thick and 46 μm thick NAND memory dies. This is achieved by exploiting the

^a Author to whom correspondence should be addressed; also at: Leaders for Global Operations, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. Electronic mail: wenghong@mit.edu

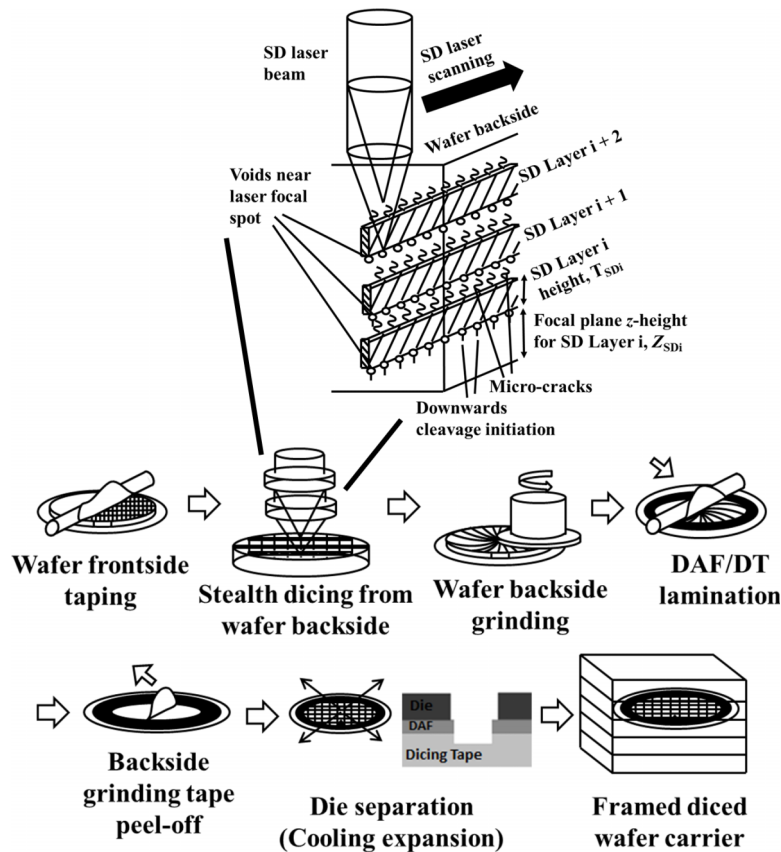


FIG. 1. Process schematic representing the partial stealth dicing before grinding (p-SDBG) integration flow. The magnified inset illustrates a representative multi-strata SD process.

multi-strata interactions between generated thermal shockwaves and the preceding multi-SD layers by optimizing the inter-strata distances, the effective SD laser dose, and the number of SD layers.

Fig. 1 shows a newly created partial-stealth dicing before grinding (p-SDBG) integration flow as a result of this work, spanning from frontside taping of the wafer to the final framed diced, ultra-thin 300 mm diameter wafer staged inside a specialty carrier. The three key processing modules involved in the p-SDBG integration flow are the SD module, the integrated wafer backgrinding module, and the die separation (DDS) module. Also, in Fig. 1, a representative multi-strata SD process is illustrated. The SD layer focal z-height, Z_{SDi} measures the vertical distance from the frontside of the wafer (facing down) to the z-focal plane of the SD laser, which is incident on the backside of the wafer. The SD layer height, T_{SDi} measures the vertical height of the SD-induced “dislocation belt” layer formed as a result of the SD laser scanning across the wafer.

For the SD laser source, a 90 kHz, 1.342 μm near-infrared wavelength pulsed laser is used. Multi-strata SD layers within the wafer are defined by translating the chuck table relative to the position of the laser with scanning speeds ranging from 50 to 900 mm/s and z-height focal points positioned from 25 to 200 μm , as measured from the wafer frontside surface. An integrated measurement laser operates at a near infrared wavelength of 830 ± 20 nm and is primarily used to detect the backside surface of the wafer in-situ during dicing in order to account for undesired wafer warpage effects on the z-height focal point positioning of SD scanning. A combination of precise z-direction spatial offset between the measurement laser’s focal spot and the SD laser’s focal spot, coupled with a calibrated displacement versus measured photodiode voltage curve (generated due to the reflection of the measurement laser beam from the wafer backside), allows not only wafer warpage compensation to the SD scanning line but also facilitates the “deep trace” capability of this integrated tool, i.e., the ability to form SD-layers deep within the wafer, beyond 500 μm , by

ensuring that the operating regime for the warpage compensation stays linear. Wafers used for the SD experiments were not backgrinded beforehand, and thus come in their original full thicknesses (775 μm) with dicing tape laminated on their frontside. The SD laser is incident on the wafer backside to avoid metallized frontside test element group (TEG) structures along the dicing streets that would block laser radiation. As a result, the SD laser is subjected to the challenges associated with high backside reflectance R from the wafers. Three types of 300 mm Si substrates are used for this work: patterned 2-D NAND memory wafers with measured absolute R of 13.4% (wafer A), 65% (wafer B), and 82.3% (wafer C) at the laser's operating wavelength. The p-SDBG approach is used to fabricate the ultra-thin memory dies for subsequent die stacking and wirebonding.

When the tightly-focused nanosecond SD laser pulse permeates through the Si wafer (without ablating the backside surface) and exceeds a peak power density (typically more than 100 MW/cm^2) during the condensing process, a highly nonlinear absorption effect occurs at the focal point due to the interactions between the Si medium and the laser field.^{2,4} A localized temperature field larger than 1000 K within the vicinity of the focal spot is established within nanoseconds. As a result, at the focal point vicinity, a void $\sim 1\text{-}3\ \mu\text{m}$ in size is formed due to the melting and vaporization of Si. Thereafter, a high dislocation density is generated due to the thermal shock wave produced upwards from the focal point because the absorption coefficient increases non-linearly with the increasing temperature.¹² As the SD laser scans in the horizontal direction, a dislocation "belt" layer known as the SD layer is formed. Fig. 2 plots the mean SD layer height T_{SD1} as a function of laser scanning speed, v (50 to 900 mm/s) for different R , with selected insets illustrating microstructural and dimensional transitions. It can be seen that as v increases from 50 mm/s to 900 mm/s, T_{SD1} decreases non-linearly across all wafer technologies. In addition, it is found that at significantly lower effective energy doses (achieved with higher v and higher R), the SD layer becomes less dense with dislocation damage. For example, at a laser average power of 2.0 W (PLE = 22.2 μJ), a sub-optimal "fishbone" SD layer microstructure arises at $v = 900\ \text{mm/s}$. When comparing across wafer technologies, C wafers exhibit a globally lower T_{SD1} than A/B wafers because of a higher R that limits the effective dose from entering the Si wafer from its backside. For the A/B wafers, despite scanning at 900 mm/s with a lower laser average power of 1.7 W (PLE = 18.8 μJ), a clear transition to the "fishbone" microstructure is not immediately obvious. As a result, for lower R wafers, the optimal v can be set at higher values, i.e., 700 mm/s for A/B instead of 500 mm/s

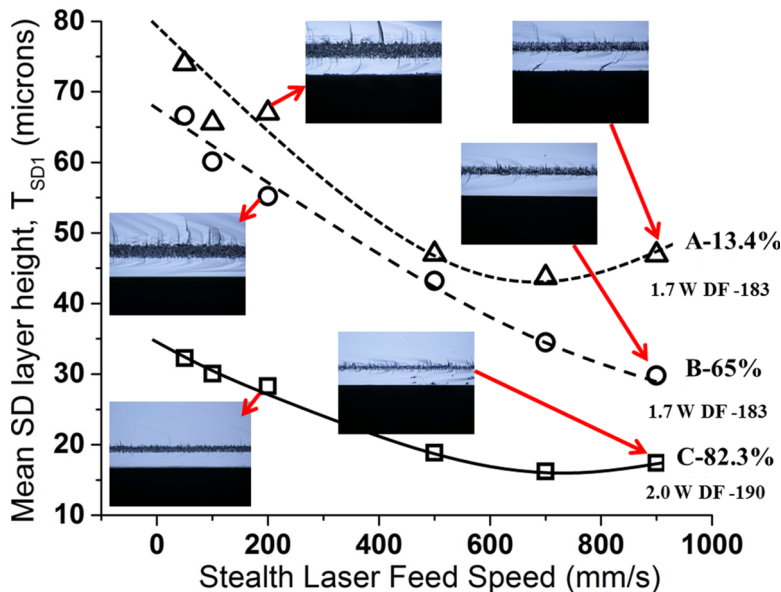


FIG. 2. Mean SD layer height T_{SD1} as a function of laser feed speed (50 to 900 mm/s) for different monitor wafer technologies (A, B, and C) with different backside reflectances (%). Insets: Selected microstructural and dimensional transitions observed by optical microscopy on full thickness die sidewalls.

for C, and thereby improving the SD throughput time. These optimal speeds for a set of given conditions are extracted not only qualitatively from microstructural observations, but also from the non-linear dependency plotted in Fig. 2 where T_{SD1} starts to plateau beyond a certain scan speed. The plateauing of T_{SD1} can be explained by the fact that as irradiation pulses separate further and further from one another as scan speed increases, it reaches a point where no overlap of individual irradiation pulses begins to occur. When this happens, T_{SD1} remains similar because the effective dose becomes a constant thereafter. One can expect the “fishbone” structure to emerge when the vertical microcracks stabilize in size while T_{SD1} starts to decrease and plateau off.

Fig. 3 plots T_{SD1} as a function of laser average power (1.0 W to 2.2 W, i.e., PLE from 11.1 μJ to 24.4 μJ) for wafers with different R . For the C wafer ($R = 82\%$), two passes of SD processing were necessary in order to facilitate manual separation using the scribe and break technique for cross-sectional inspection. The results in Fig. 3 show qualitative and quantitative evidence that as LPE increases, T_{SD1} increases non-linearly. At lower effective energy doses (achieved with lower LPE and higher R), the SD layer becomes less dense with dislocation damage and vertical microcracks becoming more prominent. For example, at a laser average power of 1.0 W (PLE = 11.1 μJ), the suboptimal “fishbone” SD layer arises at $v = 500$ mm/s. Similar to the results in Fig. 2, when comparing across wafers, C wafers have a generally lower T_{SD1} than A/B wafers technologies because of its higher R . For the A/B wafers, despite a low laser average power of 1.0 W (PLE = 11.1 μJ) with a higher scan speed at 700 mm/s, there is no obvious transition to the “fishbone” microstructure. Therefore, for lower R values, the optimal PLE can be set at a lower value, i.e., 1.7 W for A/B wafers instead of 2.0 W for C, and thereby improving laser lifetime (cost of ownership) for SD processing. In addition to qualitative observations, the optimal PLE for a given SD condition can also be validated from the non-linear plot shown in Fig. 3. From Fig. 3, it can be seen that T_{SD1} starts to increase as PLE increases but begins to plateau beyond a certain point, thus resembling a sigmoidal curve (this is more apparent for A/B wafers given the PLE sweeping range). The decreasing sensitivity of T_{SD1} to PLE as the latter increases reinforces the need to fully comprehend the minimal “safety” T_{SD1} (or T_{SDi} if using multi-passes) to initiate crack fracture

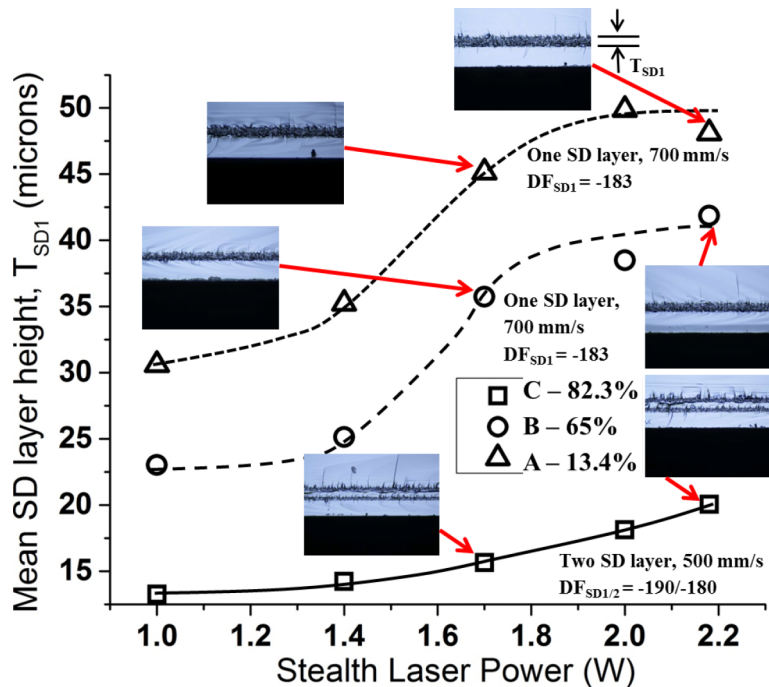


FIG. 3. Mean SD layer height T_{SD1} as a function of pulse laser energy (11.1 μJ to 24.4 μJ) for different wafers (A, B, and C) with different backside reflectances (%). Inset: Selected microstructural and dimensional transitions observed by optical microscopy on full thickness die sidewalls.

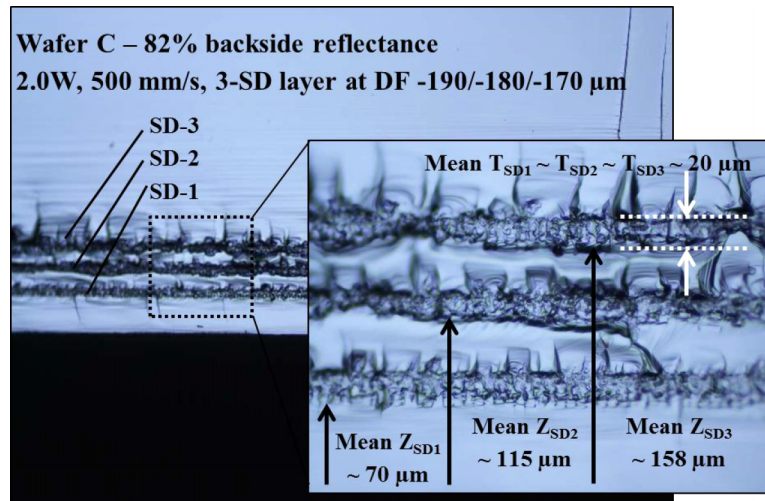


FIG. 4. Cross-sectional optical micrographs of the developed three-strata stealth dicing process before backgrinding. The inset image shows a magnified optical image demonstrating well controlled definition of the three SD layers, SD1-SD3, with no undesired defects.

without unnecessarily high PLE. It is clear from Figs. 2 and 3 that T_{SD1} can be well-controlled by using different PLEs in combination with different scanning speeds, with optimal conditions usually close to the vicinity of the rising edges shown in Fig. 3.

Fig. 4 shows cross-sectional optical micrographs of the developed three-strata SD process for C wafers (highest R so that the process can also encompass A/B wafers, i.e., those with lower R values) before backgrinding. The inset of Fig. 4 shows a magnified optical image of the boxed region demonstrating well controlled definition of the three SD layers, SD1-SD3, with no undesired defects (e.g., frontside ablation, interference, cleavage defects). Additionally, a total of 22 runs with two SD-processed C-type wafers per run over a period of ~ 2 weeks were performed to characterize the run-to-run (RtR) and within wafer (WIW) variation of the developed three-strata SD process. It was found that all three SD layer heights have a well-controlled grand mean of $\sim 19 - 20 \mu\text{m}$ with a RtR mean variability (one-sigma) of $\sim 1.3 - 1.4 \mu\text{m}$. As for the WIW variation, all three SD layer heights have a grand mean of $\sim 1.4 - 2.3 \mu\text{m}$ with a variability (one-sigma) of $\sim 0.6 - 0.4 \mu\text{m}$. At the same time, it was found that the SD layer focal plane z -height for SD1, SD2, and SD3 layers have respective well-controlled grand means of $69 \mu\text{m}$, $115 \mu\text{m}$, and $158 \mu\text{m}$. Its RtR mean variability (one-sigma) ranges between $3.2 - 4.0 \mu\text{m}$. As for the WIW variation, all three SD layer heights have a grand mean of $\sim 1.9 - 3.4 \mu\text{m}$ with a variability (one-sigma) of $\sim 0.6 - 1.3 \mu\text{m}$. These values demonstrate the potential for SD technology and p-SDBG to enable controlled fabrication of thinned, singulated die measuring $25 \mu\text{m}$ and below in thickness, because the size and the positioning of the SD “damaged” layers within Si has a very low variation; much lower than that of backgrinding.

At the same time, Fig. 5 shows top view optical micrographs of the frontside surface of SD-singulated dies with well defined, high quality SD kerfs (identified by the arrows) initiated along the dicing streets regardless of the presence of complex TEG structures. There are no signs of kerf geometric defects such as kerf width, kerf loss, kerf perpendicularity, and kerf straightness issues when using the developed three-strata SD process. Kerf width measures about $2 \mu\text{m}$ wide on average with near zero kerf loss observed as expected. Post-SD, the static loading from backgrinding will “finish the job” of full kerf separation of individual dies, which was originally initiated by frontside directing crack fractures initiated from “within” due to the multi-strata interactions between generated thermal shockwaves and the preceding SD-layers formed as the laser scans horizontally. Finally, Fig. 6 shows SEM micrographs of SDBG-integrated defect-free memory dies progressively stacked with single-sided bonding pads using (a) two four-die blocks and (b) one eight-die block. The respective insets show magnified SEM images to illustrate the integrity of the

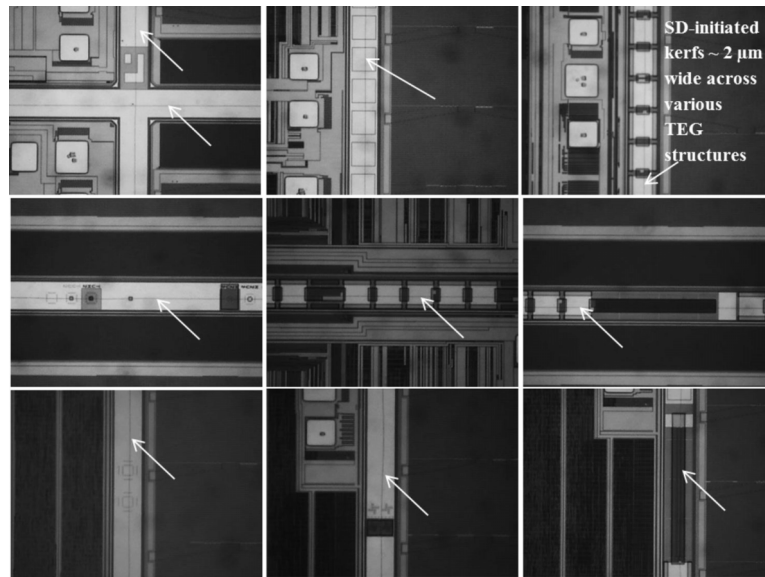


FIG. 5. Top view optical micrographs of frontside surface of singulated dies on a full 300 mm memory wafer with well-defined SD kerfs (identified by the arrows) initiated along the dicing streets, even across complex TEG structures. There are no signs of kerf geometric defects such as kerf width, kerf loss, kerf perpendicularity, and kerf straightness issues.

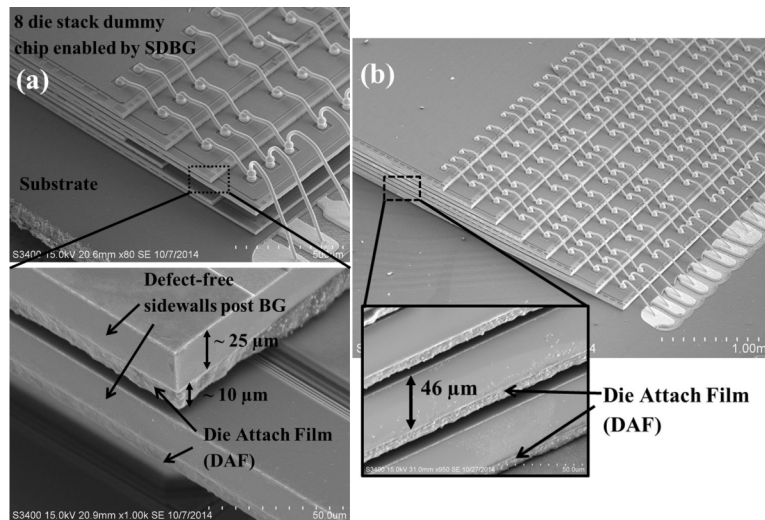


FIG. 6. Angled side-view SEM micrographs showing defect-free memory dies progressively stacked with single-sided bonding pads using (a) two four-die blocks and (b) one eight-die block. The respective insets show magnified SEM images to illustrate the integrity of the sidewalls/edges and the flush profile across the 25 μm and 46 μm thick die to the 10 μm thick DAF.

defect-free sidewalls/edges and the flush profile across the 25 and 46 μm thick die to the 10 μm thick DAF, both of which are characteristics enabled by an optimal SD process and SDBG integration flow.

The authors thank SanDisk Semiconductor Shanghai and DISCO for discussions and support. One of the authors (W.H.T.) thanks the Noyce Foundation and MIT LGO for his Robert N. Noyce full scholarship support.

¹ W.-S. Lei, A. Kumar, and R. Yalamanchili, *J. Vac. Sci. Technol.*, B **30**, 040801 (2012).

² E. Ohmura, F. Fukuyo, K. Fukumitsu, and H. Morita, *J. Achievements in Materials and Manufacturing Engineering* **17**, 381 (2006).

- ³ M. Kumagai, N. Uchiyama, E. Ohmura, R. Sugiura, K. Atsumi, and K. Fukumitsu, [IEEE Trans. Semicond. Manuf.](#) **20**, 259 (2007).
- ⁴ E. Ohmura, M. Kumagai, M. Nakano, K. Kuno, K. Fukumitsu, and H. Morita, [J. Advanced Mechanical Design, Systems, and Manufacturing](#) **2**, 540 (2008).
- ⁵ T. Monodane, E. Ohmura, F. Fukuyo, K. Fukumitsu, H. Morita, and Y. Hirata, [J. Laser Micro/Nanoengineering](#) **1**, 231 (2006).
- ⁶ K. Fukumitsu, M. Kumagai, E. Ohmura, H. Morita, K. Atsumi, and N. Uchiyama, in *Proc. of 4th Int. Congress on Laser Adv. Mat. Processing (LAMP 2006)*.
- ⁷ K. Fukuyo, K. Fukumitsu, and N. Uchiyama, in *Proc. of 6th Int. Symp. on Laser Precision Micofabrication (LPM 2005)*.
- ⁸ C. Miyazaki, H. Shimamoto, T. Uematsu, and Y. Abe, *IEEE Proc. 3DIC* **1**, 2009.
- ⁹ W.-T. Chen, M.-C. Lee, C.-T. Lin, M.-H. Yang, and J.-Y. Lai, in *Proc. of 7th Int. Microsystems, Packaging, Assembly and Circuits Technology Conference (IMPACT)* (2012), p. 271.
- ¹⁰ Y. Liao, Y. Shen, L. Qiao, D. Chen, Y. Cheng, K. Sugioka, and K. Midorikawa, [Opt. Lett.](#) **38**, 187 (2013).
- ¹¹ M. Birkholz, K. E. Ehwald, M. Kaynak, T. Semperowitsch, B. Holz, and S. Nordhoff, *J. of Optoelectronics and Adv. Mat.* **12**, 479 (2010).
- ¹² H. A. Weakliem and D. Redfield, [J. Appl. Phys.](#) **50**, 1491 (1979).