The Road towards Fully Hybrid CMOS Imager Sensors.

Joeri De Vos, Anne Jourdain, Wenqi Zhang, Koen De Munck, Piet De Moor, Antonio La Manna
Imec, Kapeldreef 75, 3001 Heverlee, Belgium
E-mail: Joeri.DeVos@imec.be

Abstract

Monolithic imagers contain the photosensitive elements as well as the read-out IC (ROIC) on the same substrate. Backside thinning on carrier enables efficient collection of photo-generated carriers through back illumination, resulting in almost 100% fill factor. This contrary to front side illumination where light loss is introduced by reflection on metal interconnects. Together with an optimized backside ARC coating, high quantum efficiency (QE) can be achieved. Hybrid imagers consist of a detector array that is produced separately and hybridized on a ROIC. A fully-hybrid backside illuminated imager has more flexibility because the detector array and the ROIC can be separately optimized to the needs of the application leading towards further improvement on QE and inter pixel cross talk. Fully processed thinned diode arrays were flip-chipped onto the ROIC by means of an Indium bump per pixel. The choice of the bump type is very critical for yielding imager assemblies, or more in general, 3D assemblies. The Indium bump process has however limited fab compatibility to evolve towards a production mature hybrid imager process. Therefore an alternative electroplated CuSn micro bump process is described. We report an average daisy chain yield above 90% for die-to-die assemblies with CuSn bumps. Measurements were performed on a dedicated 1M bump area array test design with very long daisy chains of bumps on a 20µm pitch. Processing aspects like choice of plating seed layer, the influence of cleaning agents and seed layer etchants on the micro bump performance are being discussed. Finally, the impact on the daisy chain yield after thermal cycling is shown.

Keywords: monolithic imager, hybrid imager, backside illumination, high performance 20µm pitch CuSn micro bump interconnection

I. Introduction

CMOS image sensor arrays have developed to the point where they are offering an advantageous alternative to CCDs for many applications. Their high suitability as a detector for a broad wavelength range promises a very significant reduction in weight, volume and power consumption [1]. Its applications are very wide going from high volume, low cost cameras for mobile phones to niche markets like space for earth observation. An overview is given in figure 1. For high volume, low cost products like mobile phone cameras, front side (FS) illuminated monolithic imagers are used, while in future, more advanced applications will evolve towards fully hybrid backside (BS) illuminated imagers.

![Figure 1. Imager Market and its applications](image-url)
II. CMOS imagers overview

For monolithic imagers the read-out IC (ROIC) and the photosensitive diodes are implemented on the same substrate. Illumination from the front side of the device through the backend has the disadvantage that the incident light is partly reflected on the metals in the backend and on the backend dielectrics (limited fill-factor (FF)). In the end, less than 50% of the incident light is converted and collected as electrons in the ROIC; in other words, quantum efficiency (QE) is less than 50%. For applications where every photon counts, like nighttime earth observation, this is too much.

The solution is to illuminate the imager from the backside after thinning it down to a Si thickness in the range of 30 μm. Optimal thickness is a trade-off between cross talk requirements and QE in the long wavelength range (1μm). This solution offers maximal light coupling to the substrate as FF is very high (no reflecting metals) and optimized ARC can be used matched to the refractive index of the substrate. As a result BS illuminated imagers have a wider spectral response than FS illuminated imagers. On the other hand, however, BS illuminated imagers require more complex post processing because the detector array has to be thinned down.

Monolithic imager wafers are typically permanently bonded on a carrier wafer before thinning. Dedicated wafer surface preparation, followed by a low temperature molecular oxide-oxide bonding and post-bond annealing temperatures below 250°C, result in very high bonding energies [2, 3]. This technique avoids adhesive or thermocompressive bonding techniques, but at the other hand it needs extra processing steps to open wire bond pads from the BS (imager side) of the wafer.

For hybrid imagers, by splitting up the processing of the detector array and the ROIC, each part can be separately optimized to the needs of the application. As such special substrates or epi layers can be chosen, pixel-separating trenches can be added for crosstalk reduction, etc. This additional degree of freedom has the advantage to produce photodiodes, which are not available as standard for monolithic imagers fabricated in commercial foundries. The hybrid approach on the other hand is more costly but the ROIC does not have to withstand all the post CMOS processing like for monolithic imagers.

In figure 2, a schematic overview is given of the described CMOS imagers.
III. Fully hybrid CMOS imager

The process integration of a hybrid backside illuminated 1024x1024 pixel imager is presented in detail in [4, 5]. The sensor array, thinned to 30µm, is flip chipped on top of a ROIC with 15µm diameter Indium micro bumps, where the pixel pitch is 22.5µm. The detector IC was produced at imec and the ROIC at a commercial foundry, both in a 0.35µm technology. In figure 3 the hybrid processing scheme is schematically overviewed. In contrast the monolithic imager, the hybrid imager is thinned on a temporary carrier and further flipped on a second temporary carrier before bump definition.

![Figure 3. Hybrid backside illuminated CMOS imagers: general overview of the processing modules.]

The hybrid approach is more expensive but also more advanced. Figure 4 shows a glimpse of the influence of the different parameters on the QE of the imager. The substrate of the described CMOS imagers [4, 5] has an epi layer with a dopant concentration that is changing over depth. The effect of this graded epi on the basic parameters of a photodiode is calculated in analytical form based upon the Shockley approximation. The described model, contains multiple parameters related to the condition of the substrate (diffusion length (bulk recombination centers), surface recombination, graded epi, backside surface implantation). In addition the effect of a multilayer ARC was included.

![Figure 4. Simulation of the effect of several important parameters on the spectral response.]

Without applying an ARC layer on the backside of the imager, theoretically the maximum achievable QE is 61% (Ideal plain Si response curve in figure 4). To reach a high broadband QE, an optimized ARC was developed. The ZnS / MgF$_2$ stack, show a less than 3% light loss in the wavelength range of interest (figure 5).

![Figure 5. Reflectivity simulations and measurements on the optimized ZnS / MgF$_2$ stack. Global reflective light loss is below 3% over the wide optical range (400nm-850 nm).]

Applying the developed ARC on the imager’s backside together with the surface

![Figure 6. Excellent broadband QE was measured. The results are in agreement with simulations [6].]
treatments (post-grinding clean, implantation and laser annealing), a QE above 80% was measured between wavelengths 400nm-850nm (figure 6).

Figure 7. Tilted SEM cross section view of a detector pixel after W contact plug formation. The filled deep trenches for cross-talk reduction are clearly visible and introduce a lateral built-in electrical field. The vertical electrical field is formed by the graded epitaxial layer in the Si bulk.

Figure 8. The response of both trenched (right) and non-trenched (left) imagers, acquired by single pixel illumination. For the single pixel illumination of the trenched device (right), all the incoming light is collected by the central pixel, while for the non-trenched (left), a significant portion of signal is collected by neighbouring pixels, concluding that introducing pixel separating trenches dramatically reduces, or better, avoids crosstalk [8].

Next to the different degrees of freedom BS illumination offers, the cross talk between neighbouring pixels can be improved by introducing pixel-separating trenches [7]. The deep trenches filled with poly-Si provide a lateral drift field (figure 7) which counteracts minority carrier diffusion between pixels and thus blocks inter-pixel diffusion, thereby substantially reducing electrical crosstalk. The last can be defined as the ratio of the output signal of neighbouring pixels to the output signal of the central pixel when illuminating only the central pixel [8]. For the non-trenched imager, a large part of the signal is captured by the neighbouring pixels (figure 8).

IV. Alternative CuSn micro bumping

In previous section, excellent results were shown using Indium bump interconnections. Indium is however soft by its nature and smears out easily which can lead towards shorted bumps. The bumps are defined by ultrasonic agitation [5,6], increasing the probability of lifting bumps resulting in open connections. In a detector array each unconnected bump will show up in the image as a black dot. It is clear that the Indium bump process has limited fab compatibility to evolve towards a production mature hybrid imager process.

The choice of the bump type is indeed very critical for high yielding imager stacking, or more in general, 3D stacking. Therefore an alternative electroplated CuSn micro bump process is described making use of a dedicated 20 µm pitch test design its evaluation. The influence of processing materials like seed layer etchants and cleaning agents on the electrical performance of the daisy chains is discussed. Further Ti/Cu versus TiW/Cu seed layers for electroplating are compared. The bump process needs to be compatible with the top metallization (Aluminium) of the detector and ROIC.

A. Design and process

The test design mainly consists of very long daisy chains of 1766 bumps. On 44 positions of the die, 10 interwoven daisy chains (figure 9 right) were measured. The electroplated CuSn microbumps are 10µm in diameter on a 20µm pitch. To avoid edge effect on the outer structures of the die, dummy bumps were foreseen until about 100µm from the edge of the die. The 2.0cm x 1.9cm top die is assembled on the landing die of 2.1cm x 2.1cm with 925000 bumps in total.

Both on top and landing die the bumps are connected by Al metal wires (figure 9 left). On the landing die the daisy chains are connected to the Al measurement pads. The top die is populated with Cu bumps, while the CuSn bumps are located on the landing die. In Figure 10 the bump process is explained. On an Al finished CMOS like die, the seed layer is deposited and followed by lithography. The bumps are consequently electroplated and the photoresist is stripped. Cu seed is etched followed by adhesion layer etch and an optional wafer level Cu clean.
The presence of exposed Al is an important parameter for choosing the appropriate seed layer for electroplating. This is applicable for our relatively simple test design but also for more advanced back end processes that use an Al passivation metal on top of a Cu damascene process. The seed etch process is compared for Ti/Cu and TiW/Cu seed layers. As shown in figure 11, when using Ti/Cu as seed, severe galvanic effects were observed when removing the Ti layer on the landing die. During Ti etch, the etch rate dropped when the Al measurement pads are exposed. In order to remove the Ti layer between bumps to prevent shorts, the etch time needed to be increased, resulting in a non-uniform attack of the Al measurement pads. With a TiW/Cu seed no damage on the Al measurement pads was observed.

Next to the galvanic effect, the bump undercut after seed layer etch was inspected. Cross-section TEM pictures of the bumps with both seed layers are shown in figure 12 and figure 13. The TiW/Cu based bump has no undercut, while on the Ti/Cu bump at each side more than 1µm undercut was observed. In combination with bump diameters of 10µm or lower for future applications, it is clear that TiW/Cu is preferred as seed layer.

For Cu seed layer removal, two different seed layer etchants were compared. They both successfully remove the Cu layer but interfere differently with the bumps (figure 14). Looking at the Cu bumps, applying etchant A resulted in a very smooth Cu surface, while etchant B ended in a rough Cu surface. The latter, in contrary to etchant A, does not attack Sn. From these physical inspections only the incompatibility of the Cu seed etchant A with the presence of Sn can be concluded. In the next section the electrical performance of the daisy chains is investigated comparing both etchants on Cu bumps, for a fixed seed etchant B on CuSn bumps.
Figure 14. Influence on bump performance of two types of Cu seed etchants. Etchant A is incompatible with Sn bumps.

B. Electrical evaluation

In this section electrical results are described based on two types of measurements:

- 2 point probe (2PP) measurements on 440 daisy chains composed of 1766 bumps each with daisy chain yield extraction. Daisy chains with a chain resistance above 1kΩ are considered open. If the inter chain resistance is less than 1E5 times of its corresponding chain resistance, the interwoven chain pair is considered shorted.

- 4 point probe (4PP) measurements on 44 daisy chains with 1766 bumps to investigate daisy chain resistance and extract bump resistance.

As explained in the previous section, a TiW/Cu seed layer for electroplating was used. Flip chip assembly is performed on SET FC150 at 5MPa. Neither flux nor underfill was used during assembly. In figure 15 an idea of the assembly accuracy is shown where less than 2µm misalignment can be achieved.

Figure 15. Cross-section view of an epoxy molded and lapped sample. Misalignments of less than 2µm can be achieved with SET FC150.

The data underneath covers a set of experiments. To generate some statistics, per experiment 3 stacks were assembled. Besides the Cu seed etchant comparison, we looked into the TiW etch process window. The process of reference recipe (POR) versus 25% over etch time (OE) are compared. Next, the influence of a wafer level Cu clean is investigated. This clean removes copper oxide, passivates the surface, and prevents further oxidation.

In figure 16 splits on the CuSn bump side are shown in a yield chart. The process condition of the top die with Cu bump was fixed (Cu seed etchant A, POR TiW etch, Cu clean). All conditions show high daisy chain yield (87%-99%). Neither increasing the TiW over etch time nor adding a Cu clean or the combination of both show a significant change in yield. Also the chain resistance (240-260Ω) measured by 4PP in figure 17 is stable under the mentioned conditions.
Figure 17. Daisy chain resistance: splits on landing die process conditions for a fixed top die process condition. 44 daisy chains of 1766 bumps were measured by means of 4 point probing. Each condition consists of three assemblies.

Figure 18. Daisy chain yield: splits on top die process conditions for a fixed landing die process condition. 440 daisy chains of 1766 bumps were measured by means of 2 points probing. Each condition consists of 3 assemblies. Wafer level Cu clean on Cu bumps is crucial for yielding.

In Figure 18, splits on the Cu bump side are shown in a yield chart. The process condition of the landing die with CuSn bump was fixed (Cu seed etchant B, POR TiW etch, no Cu clean). Obviously, the wafer level Cu clean is crucial as removing it from the process on the Cu bumps resulted in not yielding daisy chains. Copper oxidation prevented intermetallic formation during assembly. With Cu clean, a daisy chain yield above 90% is obtained. The clean was applied at wafer level before die singulation. Even a time interval of 4 months between wafer level clean and assembly did not result in a change of performance.

Neither increasing the TiW over etch time nor changing Cu seed etchant show a significant change in daisy chain yield. Looking at the chain resistance in figure 19 an increase in resistance is not negligible when over etching TiW. The Cu clean is not able to remove all copper oxide introduced by a longer TiW etch resulting in a higher bump resistance. With Cu seed etchant B on Cu bumps, all five assemblies show a stable resistance. The corresponding yield also passed 90%.

Figure 19. Daisy chain resistance: splits on top die process conditions for a fixed landing die process condition. 44 daisy chains of 1766 bumps were measured by means of 4 points probing. Each condition consists of 3-5 assemblies.

From the average daisy chain resistance of 260Ω, single bump resistance can be calculated. Taking routing lines into account, a bump resistance of 58mΩ (10µm diameter) was extracted.

We can conclude that although it introduces surface roughness, no influence on electrical performance was observed, so etchant B can be fixed as a unified process for etching Cu seed layers for both types of bumps, Cu and CuSn. For TiW the
POR recipe is kept and a wafer level Cu clean is mandatory for the Cu bumps.

The daisy chain yield of all experiments is varying between 87.5% and 99.5%. The daisy chain structures are very sensitive to failures as they were designed for high bump yield. It should be however noted that this daisy chain yield is not a direct measure for the bump yield. The daisy chain yield will always be smaller than the actual bump yield. This bump yield can be derived by calculation from the daisy chain yield. An average daisy chain yield above 90% for the CuSn assemblies implies a defect density that is below 60 for 1 million bumps, which is an excellent yield result. No pre-assembly inspections were performed to exclude any failures.

Finally, five stacks with the new reference CuSn bumping process were cycled between -65°C and 150°C, with 10min hold time. Neither change in yield (figure 20) nor chain resistance was observed after cycling. Measurements were done up to 3000 thermal cycles.

Considering these results, we can conclude that for most applications CuSn micro bumps are a more reliable and more production like option than Indium bumps. However for advanced applications like space, more aspects need to be investigated (i.e. whisker formation, cryogenic temperature, mechanical reliability under vibrations, etc ...) but this lies beyond the topic of this paper.

**V. Conclusion**

In this paper, an overview is given on the different types kind of CMOS imagers. For high-end applications, high performance fully hybrid imagers are the preferred option. Next to the influence of the different parameters that can influence the detector's performance, special attention is given to the need of a high yielding, high density bumping method. Considering the yield and stability of the process, together with its -65°C/150°C thermal cycling behaviour, the electroplated CuSn micro bump process is a mature candidate for connecting each detector pixel to its corresponding read-out circuit.

**References**


