EPTC 2008
Singapore

STACKING TECHNIQUE OF KNOWN GOOD REBUILT WAFERS WITHOUT THRU-SILICON VIA COMMERCIAL APPLICATIONS

By Dr Christian Val, Dr Pascal Couderc and Dr Nadia Boulay

3D PLUS
641, Rue Hélène Boucher
78532 BUC – FRANCE
Tel. + 33 1 30 83 26 50 – Email : cval@3d-plus.com

A stacking technique of all types of bare dice has been developed. The various manufacturing steps as well as the main applications will be presented.

From 2001 to 2005, an important European programme, WALPACK, funded up to 20 M€ with St Microelectronics, CEA/LETI, Thales, and 3D Plus has allowed to establish the feasibility of a stacking technique totally wafer level process. A trademark registration :“Wirefree Die on Die (WDoD)” has been made for this process.

The prototyping and pre series has been made thanks to an agreement between 3D Plus and Philips Semiconductors, now NXP.
This wafer level process stacking technique has been named Wireless Die-on-Die (WDoD). It is based on the criteria which we used in the 1990 when we launched our mature and fully qualified 3-D technique:

- Use of any kind of standard dice, which means non modified and multi-sourcing
- Electrical tests of each level prior to stacking.
- Stacking of the die with different size in order to build a System in Package

The main steps are based on the following :

. Pick-and-Place of the dice on a sticky membrane to make a Known Good Rebuilt Wafer “KGRW”
Moulding to insulate the edges of the dice and rigidify the rebuilt wafer
. Redistribution layer (RDL)
. Use of double adhesive tape to stack the different KGRW
. Dicing of this stacking
. Collective electroless plating of the edges of the dicing streets
. Direct laser patterning of the edges.
. Electrical tests

This new approach of wafer stacking without thru-silicon via “TSV” therefore allows to stack any type of standard dice, whatever their sizes. It additionally allows to stack 10 levels per mm.(100µm per levels). The area is given by the larger die plus 100µm of polymer (epoxy resin) around it.

Applications which will be presented concern the smart cards, and particularly the SIM cards. As a matter of fact, memories capacities of the SIM cards will be significantly increased and we build modules comprising 4 FLASH memories plus the secure microcontroller within 500 µm, in order to place this module within the 800 µm thickness of the SIM card (Mega SIM).

Some SiP are under development for industrial products (Home Telecom, electricity switches). Other application for mobile phones are contemplated as well as highly density FLASH memories for nomad products.
Applications with MEMS in order to built “abandoned sensors” with energy, antenna and communication function are developed for avionics and industrial areas.
In conclusion, this Full Wafer Level Process approach does not need TSV and it stays versatile and low cost, since the process utilises some of the steps used in building of wafers and is totally parallel processing from the steps A to Z. A short comparison between the Package on Package “PoP” and the WDoD is given. It can be seen that the test and burn-in of one or two sub systems like in PoP can be easily made but in fine, the volume of the module Is much smaller.

INTRODUCTION

The fantastic development of the components interconnection allows to imagine « Systems in Package » of around a few mm3, which opens new perspectives for medical applications.
Since the use of hybrid modules, then Multichip Modules « MCM » which were identical to the previous ones but named differently, the coming of the interconnection in 3-D allowed to divide both the volume and weight of a module by a factor comprised between 25 and more than 100.
We can distinguish several technologies among the technologies used for the 3-D interconnection, (table figure 1)

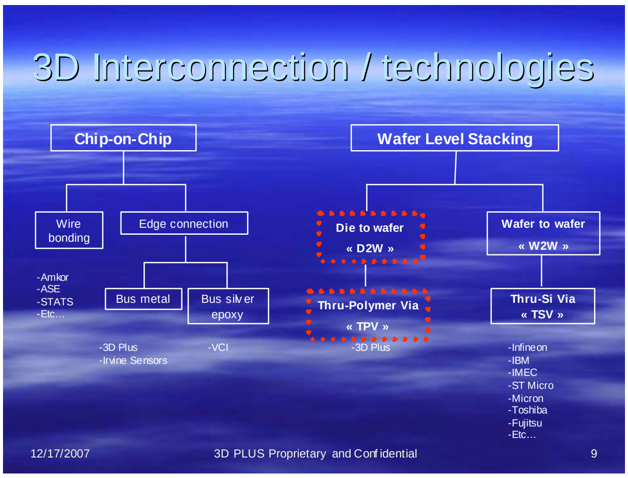

Figure 1

. The so-called Chip on Chip technology, with the chips interconnection carried out by wire bonding
. The Wafer Level Packaging technology with an interconnection carried out without wire bonding ; These lead to modules even denser than the above ones.

Among this last category, two families can be distinguished :

- The technology called Wafer to Wafer studied and developed by almost all the semiconductors manufacturers and the large laboratories worldwide, and which is based on the use of Thru Silicon Via « TSV » ; This technology presents the following disadvantages :
. Use of non standard wafer with the need for smallest possible Thru Silicon Vias. For instance, vias of 2µm diameter lead to a thickness of 20µm or less for the 8 or 10inches wafer
. Cost for a set of new masks at least K$ 500
. Copper’s TSV inside the silicon leads to a composite material which modifies its mechanical properties
. Almost impossible to make “System in Package” since die of different size

. Impossible to have 100% good wafer (very low global yield)

- The technology named Stacking of Known Good Rebuilt Wafer which is based on the Edge Connection and the Thru Polymer Via « TPV ». This technology avoids these disadvantages and adapts perfectly to the Systems In Package as well as to small and large volumes.
We have registered a trade mark : Wirefree Die on Die « WDoD » :
. Use of standard wafers without TSV
. Each layer testing prior to stacking, only “Known Good Rebuilt Wafer” are stacked
. Parallel processing from A to Z with the standard semi-conductor equipment
 - Very small form factor:
 - Thickness of each layer: 100µm (10 levels perm)
 - Size: 100 µm around the larger die

ULTRA LOW PROFILE 3-D TECHNOLOGY

This technology resulted in the registration of more than 23 patents worldwide. The flowchart is represented on

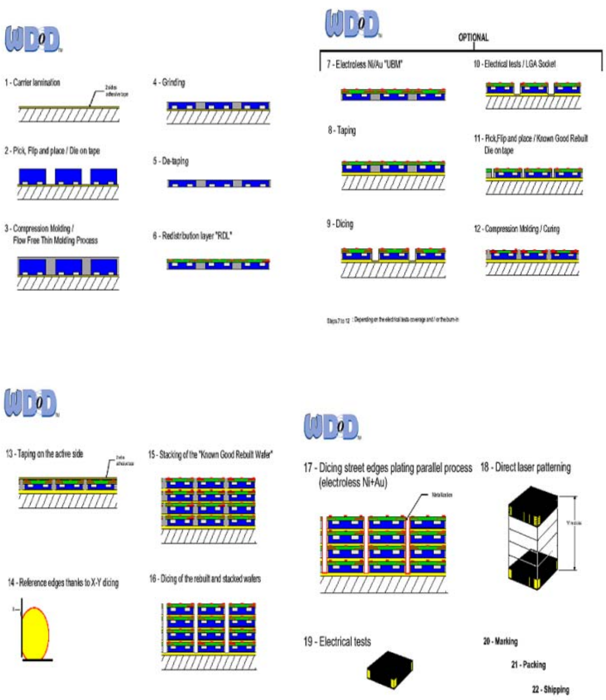

Figure 2

It can be observed that steps from 7 to 12 are optional, they are used only when the Burn-In is necessary, this is the case with some memories such as DDR or some analogic components or ASIC.
Figure 3 shows a « Known Good Rebuilt Wafer » of 6” diameter

« Rebuilt Wafer » built by compression molding

Figure 3

The status of this « WDoD » technology is the following :
- Proof of concept completed (2001 to 2004) with an European funding of 45M US$ with several companies: CEA/LETI, Thales, ST Microelectronics, Gemalto and 3D Plus
- Process development and optimization with NXP/Philips (4Q 2007)
- Functional prototypes/Qualifications (4Q 2008)
- Ramp up for the manufacturing (1Q 2009).

The main applications concern memories and SiP.
The most important production so far is carried out as per the Package on Package (PoP) technology.
A comparison is presented hereafter between the PoP, the Wafer to Wafer (W2W) and the WDoD.

Comparison of PoP / W2W and WDoD Technologies

The PoP usually contains two subsystems (Figure 4):

. The top PoP usually integrates stacked memory devices in a fine-pitch ball grid array (FBGA); we call it : subsystem 1,
. The bottom PoP typically contains a logic device of some sort; we call it: subsystem 2 .

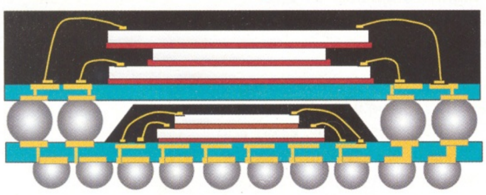

Figure 4 : Assembled Package-on-Package

Among the PoP main advantages presented hereafter, test and burn-in are considered essential.

Wirefree Die on Die (WDoD)

The 3D Plus wafer level package (WDoD) combines the main advantages of PoP technology with those of a smaller full wafer level package.

As can be seen on Figure 5, the two subsystems are the same :

. Top package with 2 memory devices such as Flash and DDR.

- . Bottom package with digital logic lus passive devices, such as PICS from NXP. The WDoD technology enables the same electrical test and burn-in strategy as the PoP however, the WDoD package is:

. Smaller *(die size of the largest device plus 100µm around 2 or 4 sides, 100µm between each level)*
. Wirefree
. Solder ball free for internal subsystems connections *(the handset makers must surface- mount the 2 subsystems on the PC board simultaneously, the same way as PoP)*

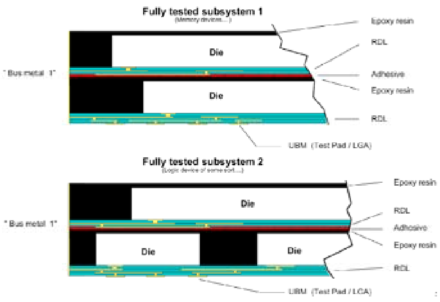

Figure 5: PoP and WDoD package

Figure 6 shows the 2 subsystems:

. Subsystem 1 contains two or more memories such as Flash, SDRAM, DDR, SRAM etc. The thickness of each level is 100µm, therefore if the subsystem contains 3 devices, the total thickness, including the external layer with the UBM pads, will be 350µm. The UBM pads will be used to make the electrical test and burn-in contact via a LGA socket.
. Subsystem 2 contains single or multiple devices to integrate high density digital logic with, or without, passive components (R,L,C). Here again, the external UBM pads enable testing and burn-in via a LGA socket.

The process to build subsystems 1 and 2 is illustrated in Figure 2.

The first 6 steps are similar to those utilized for new wirefree manufacturing processes announced by Freescale, NEC, Infineon, etc.

Step 4 can be avoided if thin wafers are available (thickness between 50 and 70µm). In this case, the epoxy resin stays over the back of the die and no grinding is needed. If thin wafers are not available, step 4 is required. Once the die have been thinned down to 80µm, a CMP process is used to release the stress inside the silicon.

Once the redistribution layer has been applied, there are two process options.

Option 1 presents the process for the test and burn-in of one device. When the dicing step has

been completed, we have a micro-package which can be placed in a standard LGA socket.
After test, we will have a Known Good Rebuilt Die (KGRD). This (KGRD) is now treated like a die and the same process is used:
. Pick, flip and place the (KGRD) on to a lamination film (step 11); note that the accuracy of this pick and place is not critical because there is no RDL process requirement.
. Once the active side has been mounted on the lamination film, the devices are molded and steps 13 up to 18 can be carried out.

Option 2 enables the use of 2 or more stacked rebuilt wafers in order to form a Subsystem Stacked Rebuilt Wafer (SSRW).The manufacturing flow for the SSRW is exactly the same as the standard WDoD process from steps 13 to18.

As with option 1, step 10 enables subsystem 1 to be tested and burned-in in a micro-package format, courtesy of the UBM pads which have a pitch used in a standard socket.
Subsystem 2 is manufactured exactly the same way as subsystem 1 utilizing the full wafer level WDoD process.

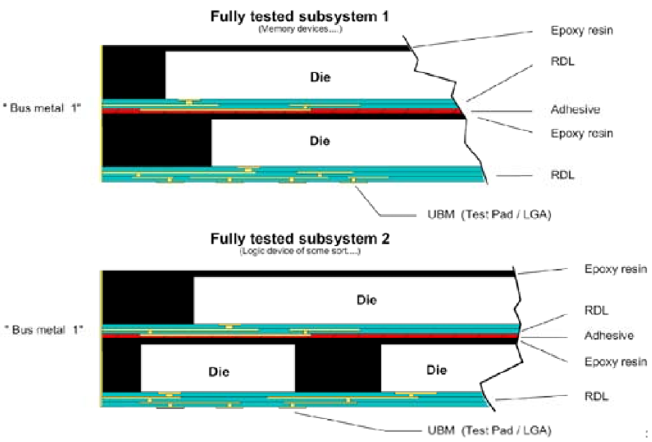

Figure 6: Subsystem 1 and subsystem 2 after testing and burned-in and before stacking

Figure 7 illustrates the two stacked subsystems. Again, the standard WDoD process is used,
specifically steps 13 to 18. The two stacked rebuilt wafers subsystems are stacked (step 15) and diced (step 16). This second dicing (cutting line N° 2) destroys "Bus metal 1" and again exposes each RDL copper trace. The layout of the external RDL used to test subsystem 2 remains compatible with the test and the surface mounting of this PoP.

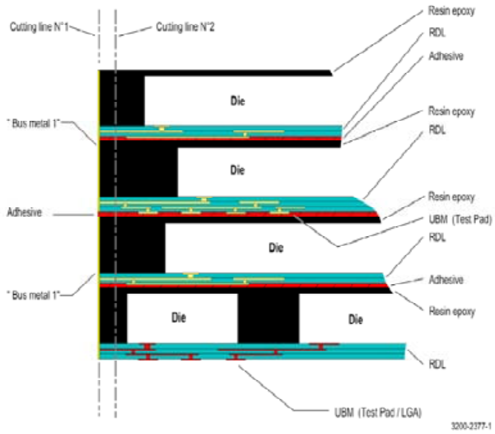

Figure 7 Subsystems stacked rebuilt wafer

Figure 8 shows the finished PoP, with the second "Bus metal". The package can now be tested before shipping. The total thickness for this LGA example is 450µm. If a BGA format is required , the diameter of the solder balls has to be added. Compared to PoP, *the epoxy resin around the die is well-balanced thereby minimizing warpage.*

The surface mounting process is comparable to that of the micro package or LGA package. It can be seen that only one package, instead of two, needs to be surface mounted. *The warpage of the bottom package of the PoP is a critical factor in determining the yield for successful package stacking. "If the warpage is too large, open solder joints may appear between bottom package and motherboard, as well as between bottom package and top package, making the stacking fail," Wei Lin from Amkor said.*

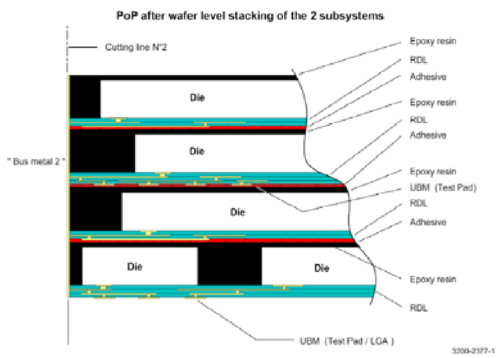

Figure 8: WDoD/PoP after wafer level stacking of the two subsystems

Comparisons of PoP, Wafer to Wafer, and WDoD

In the table shown in Figure 9 the three major 3-D technologies are compared. It can be seen that:
. Wafer to Wafer(W2W) is not compatible with the System in Package,
. PoP and WDoD are on the same position for "***components*** sourcing" and "test and burn-in"
. Cost is now lower with PoP, as the volumes scale for WDoD, the efficiencies of wafer level process (vs. non-parallel processing) will enable it to be equally or more cost effective .
. W2W offers the smallest size, with WDoD being only slightly larger with (Die size plus 100µm). The PoP is significantly larger due to the need to fan out the ball pattern on around the mold cap of the lower package in order to enable the package stack.
. The total height of WDoD and W2W is, and will remain, significantly lower than PoP due to the solder ball interconnect methodology utilized for PoP.

		WAFER LEVEL PACKAGE	
	PoP	Wafer to Wafer "W2W"	Wirefree Die on Die"WDoD"
Stacking of different size of the die	++	Very Difficult	+++
More than 1 Die/Level	+	Very Difficult	+++
Sourcing flexibility	+++	Almost impossible	+++
Test and / or burned-in before stacking	+++	Impossible	+++
Package size	+	+++	++
Package height	+	+++	++
Cost	+++	+	++

Figure 9 : Comparison of PoP, W2W and WDoD
(1 cross : bad – 3 crosses : good)

APPLICATIONS

For the small volumes 3D Plus carries out the design and the manufacturing.
For large volumes of the commercial type (mobile phones, etc) a Joint Development Agreement was signed in 2007 with NXP/Philips.

Abandoned Sensors/autonomous sensors

. For a medical application and for very small volumes we have integrated 5 die on a 0.6mm thickness and a total volume of 3 mm3.
. With an European program "e-Cubes", 3D Plus is making prototypes of abandoned sensor modules (7x7x7 mm) with 30 dice and sensors to measure the mechanical aging of the structure of aircrafts.
. Extremely dense micro-systems for medical applications are contemplated thanks to a future European project (7e PCRD) – see figure 10

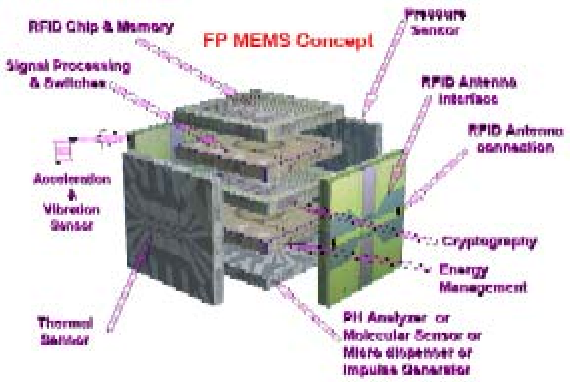

Figure 10

SiP with memories

Memories and memory-based SiP naturally represent a considerable market.
On Figure11 we present a stacking of 8 SRAM memories on a 1mm thickness.

Figure 11

The FLASH memories represent an important part of the applications for the nomad products.

We can distinguish:

. The FLASH media with a stacking of FLASH memories + the micro controller; discussions are in progress with the most important manufacturers of Nomad products.

. The smart cards: for applications such as Mega SIM, it will be necessary to stack FLASH memories. This System in Package will be used inside the SIM card, i.e. within the 800 µm thickness of this card. The 5 levels (4 FLASH memories + secure controller + oscillator + passive components) must absolutely have a thickness lower than 550 µm in order to be embedded in the SIM card.

CONCLUSION

Miniaturisation for commercial applications demands very high interconnection densities and low costs.

Wiser for former experiences, multi-chip modules, wafer scale integration, 3-D modules for Space, Defense and professional (3D Plus) applications, we learned that the yield constituted an important part of the production costs.

The WDoD process allows to stack Known Good Rebuilt Wafer (KGRW) only.

The main large-volume applications, which are, contemplated concern the SiP for smart cards, the SiP for mobile phones and stacked FLASH memories with micro controller for nomad applications

This extremely important densification of 10 to 20 levels per mm with only 100 µm around the largest die allows to launch extremely ambitious applications with memories and Systems in Package.