Integration and Packaging

December 2021

Contact Information:
Simone Mazza (simazza@ucsc.edu)
Ron Lipton (...)
Robert Patti (rpatti@nhanced-semi.com)

1 Introduction to technology

Simone Mazza, Ron Lipton

The recent development of silicon diode Low Gain Avalanche Detectors (LGADs) [**?**, **?**]. has enabled the design of granular (∼1x1 mm^2) fast-timing layers for the ATLAS and CMS tracking systems at the HL-LHC. These systems will allow the determination of the time-of-passage of minimum ionizing particles to a precision of better than 50 ps [**?**].

The essential design aspects of the LGAD can be described as: a region "p++" with a dopant concentration significantly greater than that of the bulk "p" region. This leads, after depletion, to an electric field large enough to provide amplification (by as much as a factor of 70) through multiplication of the signal. Because of this amplification, the "p" region can be made very thin (50 μm or less), leading to a fast signal and, in turn, precise timing. Several new technologies are being studied to overcome the LGAD granularity limitation as outlined in these references [**?**, **?**, **?**]. These technologies, when refined, will allow to produce finely segmented LGADs down to less than 50 μm scale maintaining the exceptional time resolution.

However once a very high granularity is achieved on the sensor it is necessary to develop a high density interconnect technology with the readout Electronics. The flip chip bump bonding procedure of sensor and ASIC has been proven to work down to the 50um scale but with issue of yield arising at lower scale. To overcome this limitation a 3D integration technology based on wafer to wafer bonding is proposed to connect the LGAD sensor to signal processing layers integrated within the readout wafer. The 3D integration technology is common in communication, computing and medical industries, however the option of 3D integration is currently not available in the international HEP community.

This 3D integration technology for high density interconnect would be of extreme interest for future HEP experiments. It will allow for finer pitch detectors that are crucial for particle detection very close to interaction points. This technology will allow a high density interconnect between sensor and electronics with a readout pitch that can be as low as 10um. The ultimate goal is to provide readout to LGAD sensors with the ability to reach 10um of position resolution and 10ps of time resolution.

Furthermore it would allow to build low mass trackers with thin layers of detectors/readout, after thinning is foreseeable to have a 50um sensor plus a 10um of electronics per layer. This

technology would make possible to build stacks of 3D interconnected sensors/readout that can be used for pattern recognition tracking [?] and for X-ray detection applications with thin sensors.

1.1 Review of companies available with respective capabilities

Robert Patty (Nhanched document summary), Rafi?

1.2 NHanched

NHanced Semiconductors has assembled 2.5/3D integrated circuits for more than a decade. We have shown active circuits stacked to a height of 8 layers, as seen in Figure 1 below. 3D assembly of test wafers has been demonstrated to a 3D stack height of 16 layers. NHanced has successfully assembled many heterogeneous 3D devices with elements such as memory, FPGAs, microprocessors, ROICs, and sensors.

1.2.1 Description of the 3D Integration Process

NHanced has produced dozens of different 3D devices and several Focal Plane Array (FPA) and Logic on Memory (LoM) parts for customers. We produced 3D IR sensors and several parts geared toward high energy physics and detection applications with research agencies, including Sandia National Laboratories. NHanced is also involved with the DARPA DAHI and CHIPS programs to develop mixed material 3D assembly techniques. This program envisions non-linear circuit performance gains, resulting in a very high Johnson figure of merit and very high circuit complexity. To date, NHanced has worked with more than 100 groups to produce 3D integrated circuits. NHanced 2.5D and 3D assembly technology covers several techniques including copper to copper diffusion bonding and covalent oxide bonding. NHanced can practice these and other assembly techniques in both wafer-to-wafer and die-to-wafer assembly. Correct operation and reliability of previously 3D assembled devices has been verified at temperatures down to 77°K. For Phase II of this effort, NHanced intends to employ wafer-to-wafer DBI assembly on the ROIC layer(s) and to also attach the IR detector. The DBI process has been demonstrated with maximum processing temperature as low as 125°C, which is in the range of most if not all 3/5 material IR detectors. For this project we anticipate doing the IR detector bonding at 180-200C

1.2.2 Vertical Interconnect Size and Pitch

NHanced's 3D technology stands out from other 3D integration technologies because of its very fine-grained small pitch vertical interconnect capability. Table 1 lists some of NHanced's interconnect sizes and pitches. The proposed flow plans to use SuperContact I TSV (through-silicon via) (1.2x 1.2um) for the ROIC portion of the 3D assembly as available from Jazz Semiconductor. NHanced has assembled devices with more than 10 million vertical interconnects per layer and has demonstrated 3D assemblies with up to 8 device layers. NHanced has developed and fabricated a number of different functional devices using this process. The fabricated devices have passed preliminary qualification, including –65 to +150°C temperature cycling. Most sensor development has been on 180nm and 130nm processes, with some work on 65nm and 55nm technologies. NHanced performs the 2.5/3D assembly process in its own facility in North Carolina.

1.2.3 DBI Processing Backgrounder

A key step in the DBI process is the direct bonding of two surfaces (dies or wafers) using Van der Waals forces. The direct bond process, shown here using two silicon wafers, is a two-stage kinetic

reaction. At room temperature, the wafers are aligned and brought into contact. The wafers have a natural dipole layer of moisture (water) adsorbed onto the surface that initially repels the opposite surface. The wafers “float” above each other until a gentle force, usually a mechanical pin, presses lightly at the wafer edge as shown in Figure 2. This overcomes the electrostatic repulsion so that the water molecules bond through Van der Waals forces. Subsequent heating to high temperatures is necessary to achieve high bond strength through the formation of covalent Si-O-Si bonds (Equation 1 and Figure 3). This reaction requires a high thermal budget that is not suitable for 3D IC fabrication. However, modifying the surface chemistry allows the formation of chemical bonds at much lower temperatures. Such surface modification technology has been reported and patented. One version simply requires a plasma treatment followed by an aqueous ammonium hydroxide rinse. A second method first prepares a metal structure(s) interconnecting to a surface (via) or backside (TSV) metal, then deposits oxide, followed by CMP to expose the metal/SiO2 coplanar surface. This method is a single mask level process if the seed layer used for electroplating is blanket-etched following electroplating. As with the Damascene metal fill method, these unit process steps are similar to those used in volume foundry interconnect stack fabrication manufacturing. This method forms the basis of the DBI process. Figure 4 is a schematic of the process flow using copper as the DBI metal. In the case of copper, special CMP techniques must be used to eliminate the dishing that normally occurs in the soft copper metal in the presence of the hard oxide during CMP. Subsequent heating of the bonded structures in a standard clean room oven forms a monolithic, low resistance metal-metal interface. Typically, a 300°C anneal is used for formation of a low resistance Cu-Cu interface. Since the activated and terminated oxide layers are bonded together with high strength, the M-M interface is subject to sufficient internal pressure so that when the copper expands at elevated temperatures, a reliable metallic bond results, even when the anneal temperature is lower than a standard anneal for Cu. NHanced has produced commercial devices using a maximum processing temperature of 150°C.

2 Advantages in respect to current available packaging

Robert Patty (Nhanched document summary), Simone Mazza, Ron Lipton In traditional 2D electronic circuits, each die is packaged separately. The packaged chips are laid out on a circuit board and connected to one another with thousands of small copper interconnects as seen in Figure 5.

2.1 Footprint

Devices that share a package take up less space than if they were packaged separately. A multi-layer 3D chip may be no larger than a single traditional 2D chip, thus offering appreciable size reduction.

2.2 Speed

Devices assembled in 3D are much closer together than chips on a circuit board. The vertical 3D connections in the device stack are only a fraction of the length of circuit board wires. Shorter distances allow electronic signals to travel more quickly from one component to another, so a 3D assembly can demonstrate higher performance than an equivalent circuit board. Past work has shown 3-5x improvement in latency or speed.

2.3 Performance

3D integration drastically lowers the parasitics of the interconnect wiring. NHanced's TSVs typically have only 2-3fF of capacitance and less than 3 ohms of resistance. In sensor applications, using DBI rather than micro-bumps has been shown to reduce noise and improve signal gain by 30% or more. This was proven in a true "apples to apples" comparison in a DoE program with FermiLab. Identical ROICs and detectors were assembled in stacks, the only difference being the detector to analog ROIC connection technology. The DBI attached sensor produced superior results with no downside effects. The only concern in employing DBI bonding is the processing temperature, which today is no less than 125°C.

2.4 Power

The shorter connections in 3D assemblies also save significant power. Today, most of the power expended in systems is consumed by driving signals in wire, whether on-chip or across circuit boards. Additionally, when signals are external to a packaged device, the devices must have protection on the signal drivers and receivers to avoid ESD (electrostatic discharge) damage. ESD protection adds capacitance, and thus consumes more energy, when signals are driven in the interconnect wiring. 3D integration eliminates the need for ESD protection on the wiring between chips. Past work has shown 90% or more power reduction in the energy consumed by the interconnect.

2.5 Heterogeneous Integration

The devices in a 3D assembly are manufactured separately, so they can vary widely. Each device may come from a different supplier, be built in a different process, incorporate different materials, etc. The components may differ in scale, voltage, dimensions, and any number of requirements. Best of class technology options can be applied to each functional layer separately.

2.6 Robustness

3D assembled layers will interface with one another through DBI connections, which have very few mechanical weaknesses. In fact, 3D devices display robustness similar to a counterpart monolithic 2D part before packaging. However, 2D devices need to interface through wire bonding or solder bumped connections, package to circuit board soldering, and the circuit board itself, all of which have mechanical weaknesses. Because 3D technology greatly reduces the number of packages involved in a system, it better suited to withstand harsh environments than traditional circuit board collections of packaged 2D counterparts.

2.7 Security

An additional benefit of 3D is improved security. 3D devices integrated with DBI are inseparable after assembly. Attempting to separate the layers of a device would destroy them. Further, the outermost layers are assembled facing inward, thus obscuring their active surfaces and making reverse engineering much more difficult.

3 Foreseen applications for 3D integration

Ron Lipton

3.1 3DIC SIPM

The Silicon Photomultiplier (SIPM) has become a staple phtotodetector for high energy physics, replacing the photomultiplier in many applications. Although inherently a digital device, the current generation of SIPMs rely on analog summing of the micropixels and resistive quenching of the avalanche. A 3DIC version of the SIPM can incorporate much more sophisticated processing including active quenching for each pixel, digital timing and windowing, intermicropixel communication and digital pattern recognition and readout. Such a device can be tailored to the application and can be more selective and much more powerful that the usual analog SIPM. Prototypes are under development by the Pratte group at Sherbrooke University [2].

3.2 Edgeless Tile Arrays

Silicon based sensors can be wafer scale, such as the 8” sensors for the CMS High Gran- ularity Calorimeter but electronic integrated circuits are generally limited by the size of the photomask reticule, typically 2×3 cm or less. This mismatch forces us to engineer complex multi-chip modules where the geometry and functionality of the pixel array is limited by the availability and routing of the external interconnects. The combination of 3D electronics, which provides dense interconnects and vertical through-silicon readout, with active edge sensors that limit dead regions normally associated with silicon detectors, allow us to build fully active tiles. Tiles would be assembled into an array only after they have been fully tested. The only external interconnect are the backside readout bump bonds that are bonded to the detector motherboard. Tiles can populate complex shapes with near-optimal tiling and low dead area. All fabrication processes are wafer-scale, which lowers the processing costs. An active edge sensor array with dummy readout has been fabricated to demonstrate the concept.

3.3 Small Pixel Induced Current Detectors

3D integration allows for small pixels with minimal capacitance associated with the inter- connect and electronic processing in complex, multi-tier readout integrated circuits. The resulting signal/noise can exceed that provided by LGADs, which typically have much higher load capacitance. The induced current pulse is prompt, with very fast rise time, and, combined with low capacitance, has the potential to provide excellent time resolu- tion [5]. A fast transimpedance amplifier coupled to the small pixel should provide both excellent time and spatial resolution. In addition to the time resolution provided by the central pixel the detector can also utilize the transient currents determined by weighting field coupling to nearby pixels. Shapes of these output signals depend on the geometry of charge motion within the silicon with respect to the electrode location. The signals shapes can be used to provide track angle information and remove off-angle background tracks. The design of such a device can be flexible, a thin detector to optimize radiation hardness, or a thicker detector to provide information on track angle or charge deposition pattern.

3.4 Double Sided and Small Pixel LGADs

The basic concept of the double sided LGAD consists of a double-sided silicon detector with a gain layer on the electron-collecting side and an array of small (3D-integrated) electrodes on the hole-collecting (anode) side [3]. We assume that the detector is thick compared to the anode pitch. The electron-collecting cathode will observe a fast rise-time signal due to the avalanche in the nearby gain layer while the anodes can provide information on the depth and location of the charge deposition. This complementary approach allows us to measure timing with coarse segmentation

in the cathodes and therefore lower the total power and complexity for the timing layer. The anode signal shapes reflect an initial peak due to the primary ionization and a secondary peak a few nanoseconds later due to holes generated in the gain layer. We can use the anode signals to either measure total collected charge for position resolution or the current pulse shape to measure the depth of the charge deposition at the pixel position.

3.5 3DIC for high performance Pattern Recognition

High performance pattern recognition capability will become more important in the fu- ture. Traditionally, pattern recognition capability has been implemented in either FPGAs or conventional ASIC. Adding a "third" dimension opens up the possibility for new archi- tectures that could significantly enhance pattern recognition capability. The 3DIC based architectures allow massive three dimensional network for data communication with much shorter traces and very low parasitic capacitance, with flexible algorithm cells distributed throughout the network. With this kind of data communication network, pattern recog- nition become much easier and one could even mimic the detector structure for pattern recognition, such as track finding or particle flow over multiple detector layers using both position information and time of arrival information. The basic algorithm cells could be as simple as Content Address Memory cell for simple matching, or could be as sophisticated as NN cells to form a high performance NN network. One simple example [4] is using 3DIC as a way to implement associative memory structures for fast track finding applications.

3.6 Possible use in oncoming experiments (EIC, X-rays ...)

Simone Mazza

4 Preliminary results (FNAL past results, possible UCSC near future results)

Ron Lipton, Simone Mazza

5 Path for future development

6 Conclusions

3D electronics and sensor integration provide a variety of technologies that can meet the needs of future particle physics experiments. Combining these capabilities with silicon tech- nologies developed for HEP, such as low gain avalanche diodes and active edge sensors, will allow us to design sophisticated detector systems that can meet the increasing challenges of next generation experiments. Crucial to this development is collaboration with partners in laboratories, universities and industry that can provide cost-effective implementations of 3D technologies. The Snowmass study provides an opportunity to explore the potential of these technologies for next generation experiments.

References