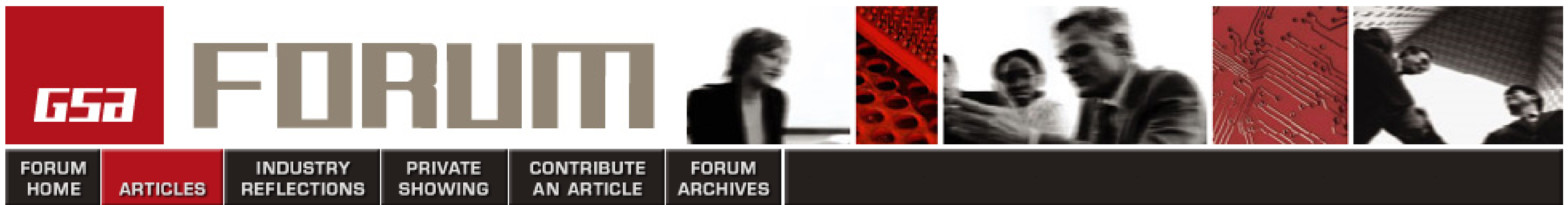


ARTICLES

Monolithic 3D IC: an Emerging Technology to Keep Us on Moore's Law

Zvi Or-Bach, President & CEO, Monolithic 3D, Inc.

In the 1960s, James Early of Bell Labs proposed three-dimensional (3D) structures as a natural evolutionary step for integrated circuits (ICs). Since then many attempts have been made to develop such a technology. Some have even declared the coming decade is the "3D Era" for both transistor shape and layer stacking. However, through silicon via (TSV) connected layer stacking is proving to be expensive, mostly due to the large TSV size and keep-out-zone driven pitch. It is also relatively limited in vertical connection density (about 104/mm2).

The "Holy Grail" of 3D-IC has been monolithic 3D layering. This technology allows a second transistor layer to be constructed directly over the base wafer using ultra-thin (<100nm) silicon, enabling a very rich vertical connectivity. Although it has been difficult to overcome process temperature limitations imposed by the underlying interconnect materials, it now seems that monolithic 3D solutions, in both memory and logic, are becoming practical.

This technology has the potential of becoming the technology that keeps us on track with Moore's Law, the doubling of the number of on chip transistors that can effectively be produced in volume. With two-dimensional scaling getting harder and costlier and with next generation lithography - Extreme UltraViolet (EUV) - tools continually delayed, monolithic 3D-IC seems to be a very compelling option.

First Adopters

NAND memories drive high volume semiconductor manufacturing today. Samsung announced in August 2013 Mass Production of the industry's First 3D Vertical NAND Flash. As other memory vendors adopt this path, the International Technology Roadmap for Semiconductors (ITRS) for NAND Flash technology is suggesting that increasing the number of 3D layers per node is the path to future scaling.

A very compelling advantage of scaling in 3D for memory products, as presented in an IEEE 3D-IC conference tutorial1, is the ability to process multiple layers simultaneously, supporting higher memory capacity at about a constant cost. This is clearly represented in the acronym BiCS used by monolithic 3D memory innovator Toshiba, standing for Bit Cost Scalable memory.

Further evidence of the viability comes from equipment vendors such Applied Materials and Lam Research gearing up to support this emerging market. Tools under development, such as Etch and Deposition equipment, allow an increasing number of layers to be processed simultaneously with each new technology node.

Next Adopters

It appears that DRAM vendors are looking for a monolithic 3D path as well. For example, SK Hynix recently licensed BeSang's monolithic 3D IC technology. Most vendors utilize polysilicon transistors for their 3D NAND transistor stacks, but the leakage of such transistors would be too high for a DRAM implementation. BeSang developed a monolithic 3D technology that utilizes mono-crystalline thin layers, constructing vertical transistors as a second layer on top of a prefabricated device wafer that includes copper interconnect.

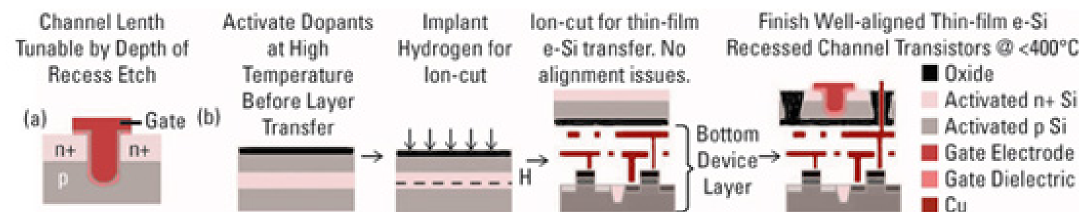
The challenge for monolithic 3D technology is the 400°C process temperature limitation imposed by the use of copper (or aluminum) for interconnection. This is the main reason TSV-based 3D IC technology, which allows each wafer to be independently processed, emerged. With this approach, a wafer is thinned, then placed in a 3D configuration, and connected to the substrate with TSV using a low temperature (<400°C) process.

This independent (parallel) processing has its own advantages, however the use of thick layers (>50 μm) greatly limits vertical connectivity as it requires development of new processing flows, and is still too expensive for broad market adoption. On the other hand, monolithic 3D IC provides 10,000x the vertical connectivity and would bring many additional benefits as recently presented in the IEEE 3D IC conference.

BeSang's innovative solution is to perform the high temperature process on a donor wafer constructing N-P-N layers. Those layers are transferred on top of the pre-processed base wafer, detaching the donor wafer leaving generic N-P-N layers on top, and completing the processing of vertical transistors.

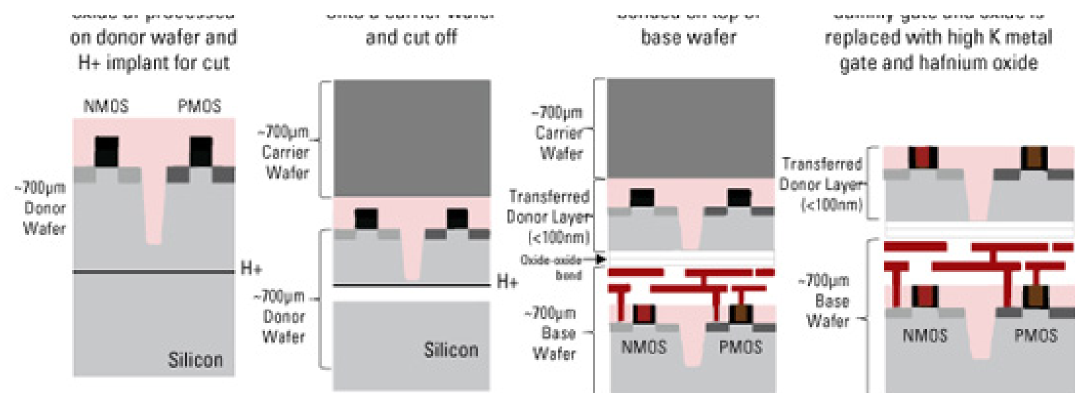
An alternate innovative flow for monolithic 3D-IC was presented at the IEEE 3D-IC Conferences of 2012 and 2013. These flows utilize the industry standard Ion-Cut layer transfer technology also known as Smart-Cut®. Ion-Cut has been demonstrated below 400ºC[2] and is a volume production qualified process due to its two decades of use for SOI wafer manufacturing. It is estimated that with the reuse of substrates, ion-cut would cost less than $60 per layer.

Figure 1: (a) A Recessed Channel Transistor (b) Process Flow for monolithic 3D Logic. Bottom device layer with Cu/low k does not see more than 400ºC



This alternative flow allows construction of conventional and well understood high performance horizontal transistors. The proven transistor structure, which has been used for many years in DRAM devices, the Recessed Channel Array Transistor (RCAT) is a good example. Figure 1 describes the RCAT process[3], which constructs the RCAT transistors commonly used in DRAM manufacturing since the 90nm node. The RCAT is quite competitive with standard planar transistors[4] and looks like the inverse of a FinFET.

Figure 2: Process Flow for Gate Replacement Process



As can be seen in Figure 2, high temperature dopant activation steps are performed before transferring a bilayer n+/p silicon layer atop Cu/low k using ion-cut. The transferred layers are un-patterned, so that no misalignment issues occur during wafer bonding. After bonding, sub-400ºC etch and deposition steps are used to define the transistor. This is enabled by the unique structure of the device. These transistor definition steps directly utilize the alignment marks of the bottom Cu / low k base substrate with conventional nanometer precision, as transferred silicon films are thin (usually sub-100nm) and transparent. Sub-50nm diameter through-layer connections with no dielectric liner needed can be produced due to the thin transistor layer, conventional STI (Shallow Trench Isolation), and excellent alignment.

Recently the industry moved to Hi-K metal gates (HKMG), and later fully adopted the gate last (gate replacement) approach to avoid exposing the Hi-K material, such as hafnium oxide, to high temperature. This could be used for monolithic 3D as illustrated in Figure 2. The first step is to process the dummy gate stack transistors at the conventional high temperature protocol on a donor wafer. The next step is to use ion-cut and a carrier wafer to cause a face-up transfer on top of a base wafer. Finally, perform the gate replacement by removing the H+ damaged oxide and replacing with a HKMG stack using low temperature deposition and etch processes.

It is interesting to point out that just as scaling challenges are mounting, some of the trends associated with dimensional scaling make monolithic 3D easier. For example, the amount of silicon associated with a transistor structure was measured in cubic microns in the early days of the IC industry and has now scaled down to tens of cubic nanometers. The new generation of advanced transistors have thicknesses in nanometers, as seen in the 7nm thin body of Fully Depleted Silicon On Insulator (FD-SOI) transistor recently released to production by ST Micro.

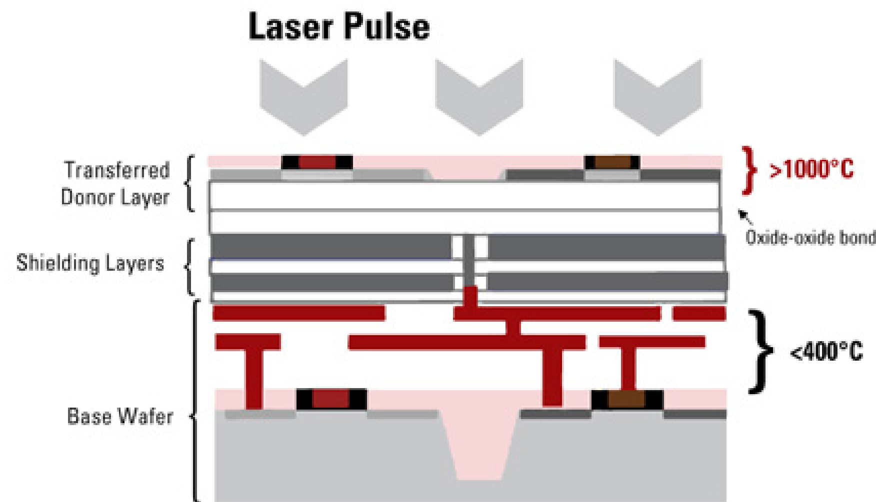
Dimensional scaling has also brought down the amount of time required for transistor activation/annealing, allowing sharper transistor junction definition. In the early days of the industry, annealing and activation took tens of minutes of high temperature furnace based processes. These were replaced later by Rapid Thermal Processing (RTP), which is now being replaced with laser annealing and activation processes. The amount of heat associated with transistor formation has reduced dramatically with scaling, as less volume of silicon gets heated for far less time. Unlike furnace heating or RTP annealing, laser annealing allows the heat to come from the top and directs to only on a small part of the wafer at any given time.

These trends help make it practical to protect the first strata interconnect from the high temperature processing required for the second strata transistor formation. As the high temperature is projected on a small amount of silicon for a very short time and for a small part of the wafer, the total amount of thermal energy required for activation/annealing is now very small.

One of the three most newsworthy topics and papers included in the 2013 IEDM Tip Sheet for the "Advances in CMOS Technology & Future Scaling Possibilities" track was a monolithic 3D chip fabricated using a laser (reported by Solid State magazine "Monolithic 3D

chip fabricated without TSVs"). "To build the device layers, the researchers deposited amorphous silicon and crystallized it with laser pulses. The researchers then used a novel low-temperature chemical mechanical planarization (CMP) technique to thin and planarize the silicon, enabling the fabrication of ultrathin, ultraflat devices. The monolithic 3D architecture demonstrated high performance – 3ps logic circuits, 1T 500ns nonvolatile memories and 6T SRAMs with low noise and small footprints, making it potentially suitable for compact, energy-efficient mobile products."

Figure 3: Smart-cut®



Furthermore, in early October an alternative was presented in the IEEE 3D-IC and IEEE S3S conferences based on simulation work. Smart-cut® was suggested for the formation of the second strata (and not amorphous silicon crystallization) with innovative shielding layers to protect the first strata interconnect during high temperature pulsed laser processing, as illustrated above.

Currently there are at least three different laser annealing systems offered on the market. The shielding layers could be adjusted according to the preferred choice of the laser annealing system. The simulations show that if an excimer laser is used, even without these shielding layers, the first strata interconnect layers are not adversely impacted by the laser annealing process. Worth noting are the good results of IMEC when utilizing Excico laser annealing for 3D memory enhancement, as announced in late October 2013 - Laser thermal anneal to boost performance of 3D memory devices. In conclusion, while dimensional scaling is becoming more difficult, it allows Monolithic 3D to become easier.. The industry should be able to keep scaling one way or the other, potentially even both, while continuing to enjoy the benefits.

About the Author

Zvi Or-Bach is the founder President and CEO of MonolithIC 3D™ Inc., Or- Bach has a history of innovative development including the breakthrough of monolithic 3D IC and fast-turn ASICs for over 30 years. Prior to MonolithIC 3D, Or-Bach founded eASIC in 1999 and served as the company's CEO for six years. Earlier, Or-Bach founded Chip Express in 1989 (recently acquired by Gigoptix) and served as the company's President and CEO for almost 10 years, bringing the company to $40M revenue, He holds over 100 issued patents, primarily in the field of 3D integrated circuits and semi-custom chip architectures. He is the Chairman of the Board for Zeno Semiconductors, Bioaxial and VisuMenu.

Note: smart-cut® is a registered Trademark of Soitec.

Back to Articles Home

Forum Home | Articles | Industry Reflections | Private Showing | Forum Archives | GSA Home