Building 3D-ICs: Tool Flow and Design Software Part 1

Robert Patti

11/14/2011 11:23 AM EST

Title-1
This is part 1 of a two part article. Part 2 can be found here

The industry's current enthusiasm for 3D-ICs is widespread and well warranted, but designing those 3D devices presents a challenge. Normal 2D tool flows, thoroughly honed and refined over many years, nonetheless fail to address some of the critical issues of 3D design. A new 3D design process is evolving gradually from that 2D heritage. When Tezzaron designed its first 3D circuits in 2003, the designers used standard 2D CAD tools and cobbled together a 3D DRC and LVS flow based on scripts. Today there are tools to handle a complete backend flow and strides are being made to enable true 3D design partitioning, synthesis, placement, and routing (see Figure 1).

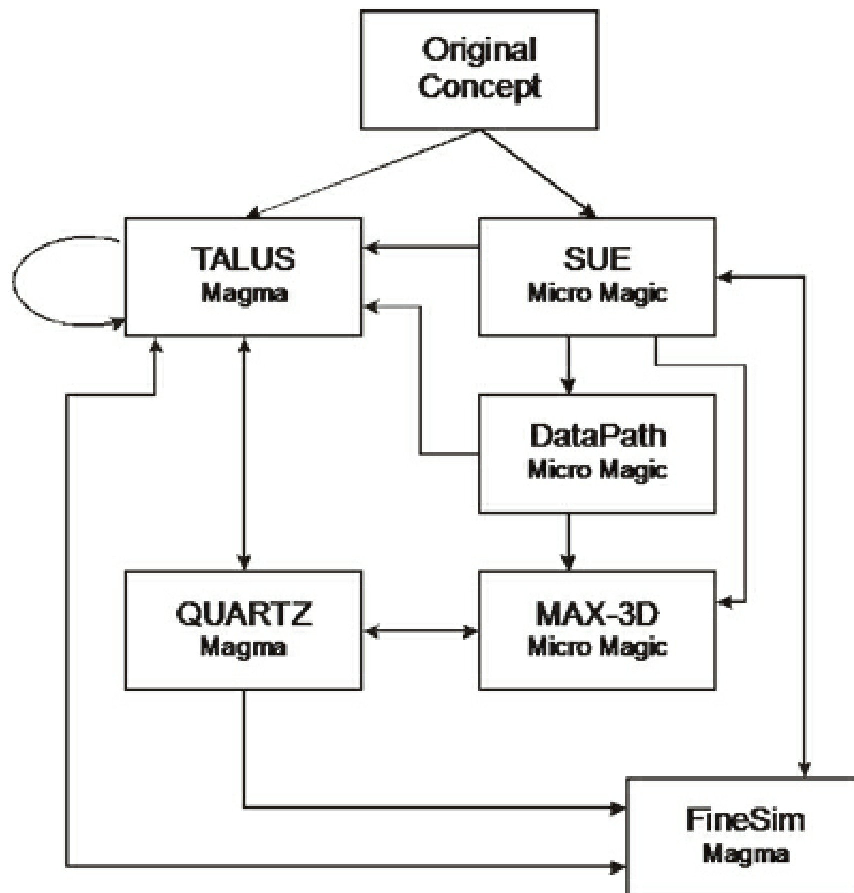

Figure 1 – Current 3D design flow

This article discusses the current state of 3D tools and software, describes a working flow, and identifies the areas where more progress is needed. We base the discussion on a specific next-generation demonstration device taken from a design that Tezzaron is prototyping with several partners. The demo design contains an

advanced ARM® processor stack, an "off the shelf" FPGA die, and a DRAM memory stack, all assembled onto an active silicon circuit board acting as an interposer, as shown in Figure 2 below.

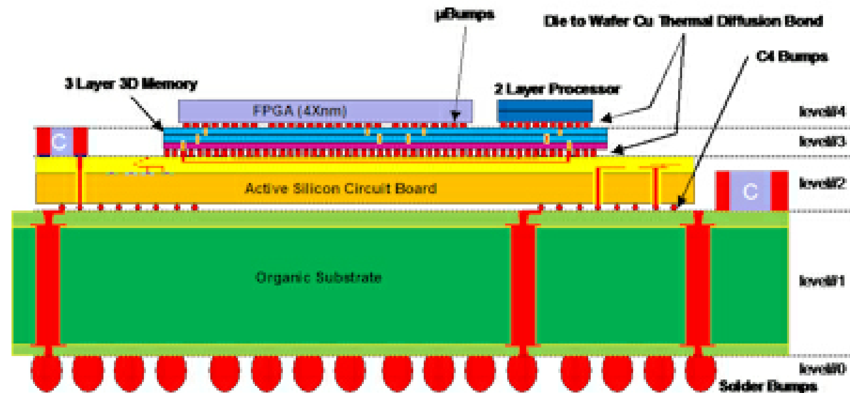

Figure 2 – The demo device

The Parts

The CPU is an implementation of an ARM® Cortex™-A processor, part of a new class of highly scalable ultra-efficient processors. An out-of-order execution unit with superscalar pipelining is tightly coupled to a low-latency level-2 cache. We take advantage of the multicore capability to support 1 to 4 layers of processing for a scalable design implementation. Each layer of the 3D processor stack contains two CPU cores, their respective L1 caches, and a shared 2MByte L2 cache. The processor layers communicate vertically via a 1024 bit ultra wide low latency bus interface block. The interface block isolates powered-down cores and copes with bus masters of differing frequencies. The gate count per core is about 2 million gates, excluding the cache memories.

The FPGA is a standard 2D device. This could just as well have been an ASIC or other logic element. It provides the glue in our demo system.

The DRAM, a three layer 8Gbit memory, is representative of a next generation Tezzaron memory. It acts as the main memory in this system.

The final element in our demonstration system is an active interposer or silicon circuit board (SiCB). Beyond signal redistribution, our demo system takes advantage of the SiCB to provide power regulation, decoupling, and signal termination.

The Processor

The 3D processor is assembled using a wafer to wafer process to produce a frontside to frontside bonded wafer pair. Wafer to wafer bonding allows the smallest, densest vertical interconnect structures and opens many avenues for design optimization.

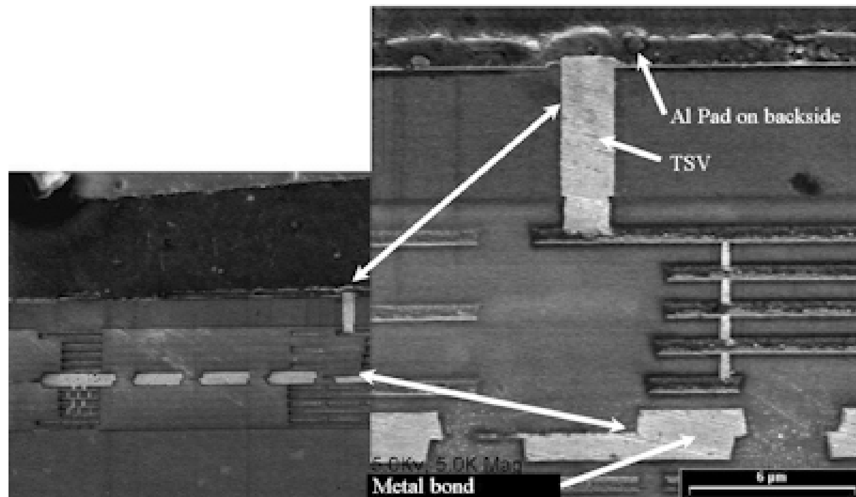

Figure 3 – Details of a two layer logic device

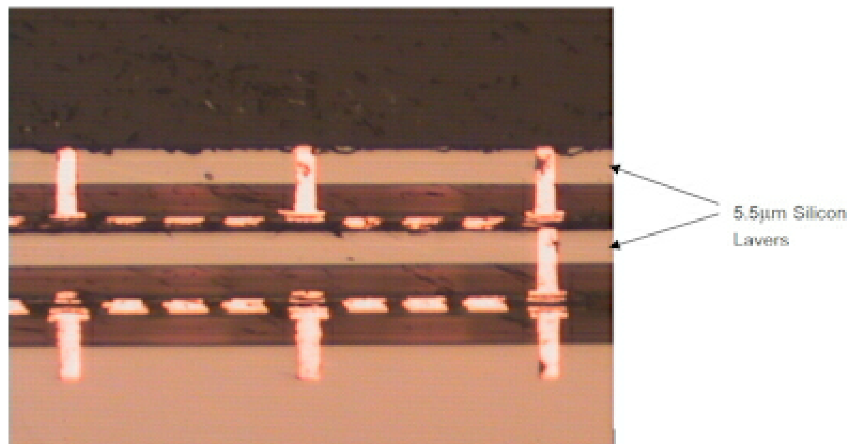

Figure 4 – Three layers of silicon, top two layers thinned to 5.5µm

On the down side, wafer to wafer bonding does not provide the yield benefit that die to die techniques can offer with a Known Good Die (KGD) approach. Overall, a wafer to wafer 3D design has about the same yield as a 2D equivalent when looking at the total area of all layers.

We scale the performance of the processing element in this demo device by adding a pair of CPU cores in each layer of the 3D stack. This reduces our partitioning effort and allows us to treat the design effort more like a 2D design with extensions. Our early 3D designs were done this way because of the primitive state of 3D tools eight years ago. In this case, the 3D enhancement of the 2D design can be relatively brute force; TSVs and vertical interconnect planning can be inserted manually. The synthesis and placement are done in 2D in this case. However, to provide insight into a more sophisticated and powerful flow, let us explore some alternatives.

Partitioning and Placement

The CPU begins as a Verilog module. It can be partitioned in a few different ways. For instance, the CPU logic could be put on one level and the cache on a separate level. In this case the manufacturing process for the cache could be done in a different node or technology than that of the CPU core itself. Process separation is one of the most beneficial aspects of 3D integration.

Alternatively, we could use recent tool advances to create a 3D optimized placement of the CPU internals themselves. Using Magma's Talus tool, with its unified data model and highly flexible TCL interface, we can

easily optimize the Verilog into partitioned regions. While the tool itself is not presently 3D aware, we can use it to create identical regions in a flat 2D model and define interconnect elements that mirror between regions. The mirroring in fact forces the 3D interconnect to align vertically. Another approach is to use the Talus tool to do full block physical synthesis and early partitioning and then import the .DEF files to Micro Magic's 3D-FloorPlanner (see Figure 5 and Figure 6 below).

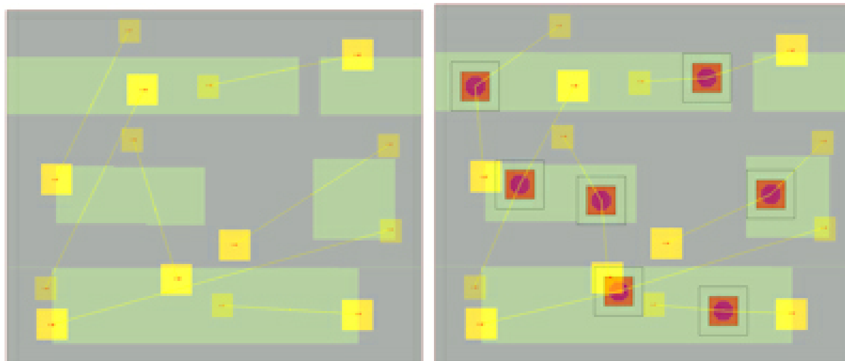

Figure 5 – TSV insertion example using Micro Magic 3D MAX auto placement

Left: Top down view of two layers to be vertically connected
Right: TSVs inserted into areas without obstructing transistors, placed to minimize wire length

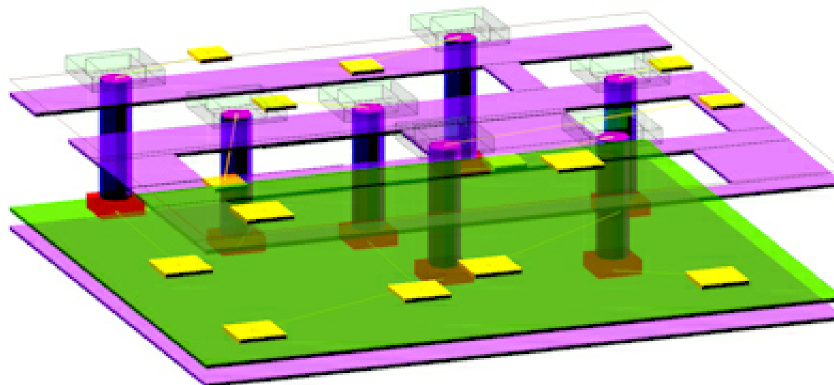

Figure 6 – Orthogonal view of 2 layer circuit after TSV insertion

This tool auto inserts TSVs into the floor plan, optimizes the TSV placement, and further optimizes the cell floor plan. One of the benefits of the Micro Magic environment is that its MAX-3D tool allows true 3D physical editing. The design can be tweaked as necessary to improve the partitioning and pre-route critical nets. After the Micro Magic adjustments, the database can once again be transferred to the Magma Talus environment for routing. The advantages of partitioning the entire design in 3D is increased performance and lower power. In theory, 3D can reduce the length of wires by a factor of N where N is the number of layers used. This is overly optimistic, but as long as the 3D interconnect structures (face to face or TSV) are reasonably fine grained, the wire shortening is significant. The longest nets tend to be clocks, data busses, and global controls – all of which have the most repeaters and consume the most power. Applying 3D wiring to these nets can easily improve the system speed and power by 30 to 50%.

Someday, the ultimate 3D tool will take into account both separate processes and true 3D placement optimization. Today, however, this tool is somewhere over the horizon.

Routing

We have now successfully partitioned our processor and done the initial placement. The vertical interconnects in the placed design are modeled as standard cells or blocks and these are fixed to prevent the router from breaking the inherent 3D wiring that the placement created. Additional cells and blocks can either float or be fixed, as desired. We finish the routing with the Magma Talus router. The Talus router now sees just a simple series of fixed regions that have "wormhole" connections – that is, the 3D interconnect points provide connectivity between the fixed regions with no obvious wire length. The 3D interconnect is modeled as a lumped fixed element in the wiring network. As mentioned previously, the vertical interconnect locations are locked down, but the Talus tool is free to make final adjustments to the clustering and balancing of the cells and blocks as necessary and complete the routing process. A key to making this flow work is the extensive flexibility that Talus offers. We have glossed over the obvious strengths of the Talus tool at handing a multi-gigahertz, high performance core processor through physical synthesis, placement, and routing. At 28nm these tasks are no small issue. The need for a clock guard ring and ultra tight signal matching is paramount. Needless to say, if these basic requirements of state of the art tools were not already present in the Talus tool, the 3D extensions and use would be substantially diminished.

Verification

The design verification involves design rule checking (DRC) and comparing the layout versus schematic (LVS). Magma's Quartz, which already supports 3D DRC and LVS decks, can check the 2D and 3D design requirements. The key to 3D verification tools is that they must support the simultaneous use of multiple databases and technology files. Quartz, like Talus, has a built in multi-database capability that fulfills this basic requirement. To simplify and allow reuse of the existing 2D rule decks provided by the foundries, the design is treated as a set of 2D designs, each of which is separately checked and verified. The whole is then checked as a 3D assembly of the 2D designs. The 3D rule sets should be provided by the 3D assembly house – in this case, Tezzaron

DRC

As mentioned previously, 3D DRC is performed by first checking each of the 2D layers with the DRC deck supplied by its respective wafer foundry. The use of unadulterated foundry decks is important. If we were to significantly customize a deck for use with 3D, the continuing maintenance and revisions from the foundries would make updating that deck a never-ending chore. Error results are thus generated separately for each of the silicon layers in the 3D stack.

After the 2D DRC processing is complete, 3D DRC is run. The TCL interface to Quartz allows simple scripts, supplied by the 3D assembly house, to check all of the 2D layers and then process the 3D rules as well. 3D DRC must look at the top and bottom of each of the 2D layers and the interconnect between the layers, such as bonding or backside metals. If necessary, 3D DRC also performs checks for layer specific pieces of the 3D interconnect – for example, TSVs might not be part of the 2D rules from the foundry. An important nuance is that there are multiple sets of silicon layers that may or may not have identical GDS layers. Micro Magic's MAX-3D tool can specify unique GDS numbering for the 3D layer stackup as part of its 3D tech file. In this way metal1 on the first layer of silicon can be unique from metal1 on the second and third layers, etc.

The 3D GDS file written from MAX-3D contains all of the layers that are required to check the design to assure that, mechanically, the design can be correctly fabricated. The 3D tech file and the 3D DRC files should be written and supported by the 3D assembly house.

LVS

Talus provides a post place and route Verilog netlist. This netlist can be used for simulation and physical verifications such as LVS. Quartz LVS operates in 3D much as Quartz DRC does. A script interface checks each of the 2D designs against a set of 2D netlists and then an overall 3D check is done, comparing a top level netlist against the extracted 3D interconnect. Again, the original LVS decks as supplied by the foundries can be used for the 2D checks, eliminating the need for a cumbersome and error prone rewrite of the extraction and comparison rules. The 3D LVS is very simple in that there are no devices – just blocks with pins and wiring between those blocks. This greatly simplifies the 3D assembly house's task of creating 3D LVS decks. The 3D LVS can be generalized to easily accommodate 2D layers with differing technology.

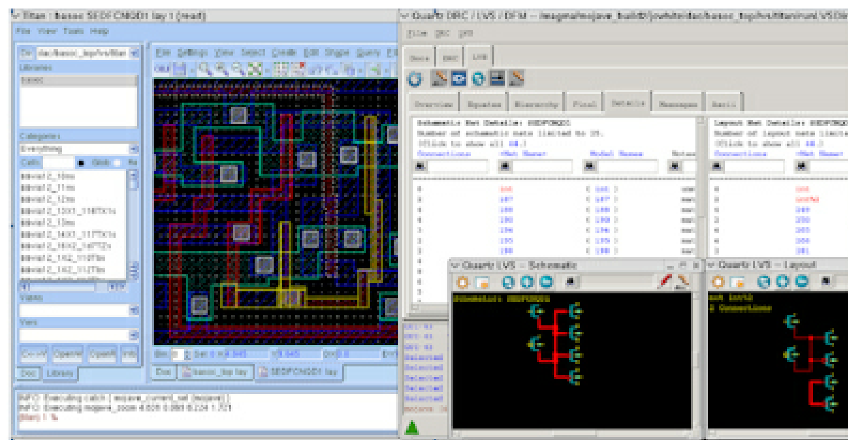

Figure 7 – Quartz LVS screen shot

Part 2 of this article can be found here

Robert Patti, CTO and VP of Design Engineering; Director, Tezzaron Semiconductor (Singapore) Pte Ltd Previous experience: Founder and President of the predecessor company, ASIC Designs, Inc.; Member of Technical Staff for Tellabs, Inc. Member of IEEE. Former Vice-Chairman of JEDEC's DDRIII / Future Memories Task Group. Holder of 16 US patents, numerous foreign patents and many more pending patent applications in deep sub-micron semiconductor chip technologies; BS (EE/CS) and BS (Physics), Rose-Hulman Institute of Technology.

If you found this article to be of interest, visit **EDA Designline** where – in addition to my blogs on all sorts of "stuff" – you will find the latest and greatest design, technology, product, and news articles with regard to all aspects of Electronic Design Automation (EDA).

Also, you can obtain a highlights update delivered directly to your inbox by signing up for the EDA Designline weekly newsletter – just **Click Here** to request this newsletter using the Manage Newsletters tab (if you aren't already a member you'll be asked to register, but it's free and painless so don't let that stop you [grin]).