
17 Challenges and Future Directions of 3D Physical Design

Johann Knechtel, Jens Lienig, and Cliff C.N. Sze

CONTENTS

Abstract.....	357
17.1 Key Challenges for 3D Circuits and Approaches Revised.....	358
17.1.1 Technological Challenges.....	358
17.1.1.1 Diversity of Integration Approaches.....	358
17.1.1.2 Power Delivery and Thermal Management.....	360
17.1.1.3 Clock Delivery.....	362
17.1.1.4 Testing.....	363
17.1.2 Design Challenges.....	364
17.1.2.1 Layout Representations for 3D Circuits.....	364
17.1.2.2 Partitioning and Floorplanning.....	366
17.1.2.3 Placement.....	367
17.1.2.4 Clock-Distribution Networks.....	368
17.1.2.5 Routability Prediction.....	369
17.1.2.6 Routing.....	370
17.1.2.7 Multi-Physical Simulation and Verification.....	371
17.2 Prospective Directions for 3D Circuits and Design.....	373
17.2.1 Path-Finding: System-Level Design Exploration and Evaluation.....	374
17.2.1.1 Concepts and Approaches for Path-Finding.....	374
17.2.1.2 Flows and Tools for Path-Finding.....	374
17.2.2 Design Approaches and Standardization.....	375
17.2.2.1 Design Approaches for 3D Circuits.....	375
17.2.2.2 Standardization Trends and Examples.....	378
17.2.3 Demonstrators and Prototyping for 3D Circuits.....	378
17.2.3.1 Academic Efforts.....	378
17.2.3.2 Industrial Efforts.....	379
17.3 Summary and Conclusion.....	380
References.....	380

ABSTRACT

The concept of 3D integrated circuits (3D-ICs) provides new opportunities for meeting current and future design criteria, such as performance, functionality, delay, and power consumption. 3D-ICs are thus considered as a promising approach to spur both *More Moore* (i.e., further down-scaling of baseline CMOS device nodes) and *More-than-Moore* (i.e., diversification of functionality; heterogeneous system integration) [1,9] as shown in Figure 17.1. At the same time, 3D-ICs increase complexity for manufacturing and physical design notably.

Please cite as:

J. Knechtel, J. Lienig, C.C.N. Sze "Challenges and Future Directions of 3D Physical Design". In: Physical Design for 3D Integrated Circuits, A. Todri-Sanial, Ch. S. Tan (eds.) CRC Press, Boca Raton, FL, ISBN 978-1-498-71036-7, pp. 357-386, 2015. **357**

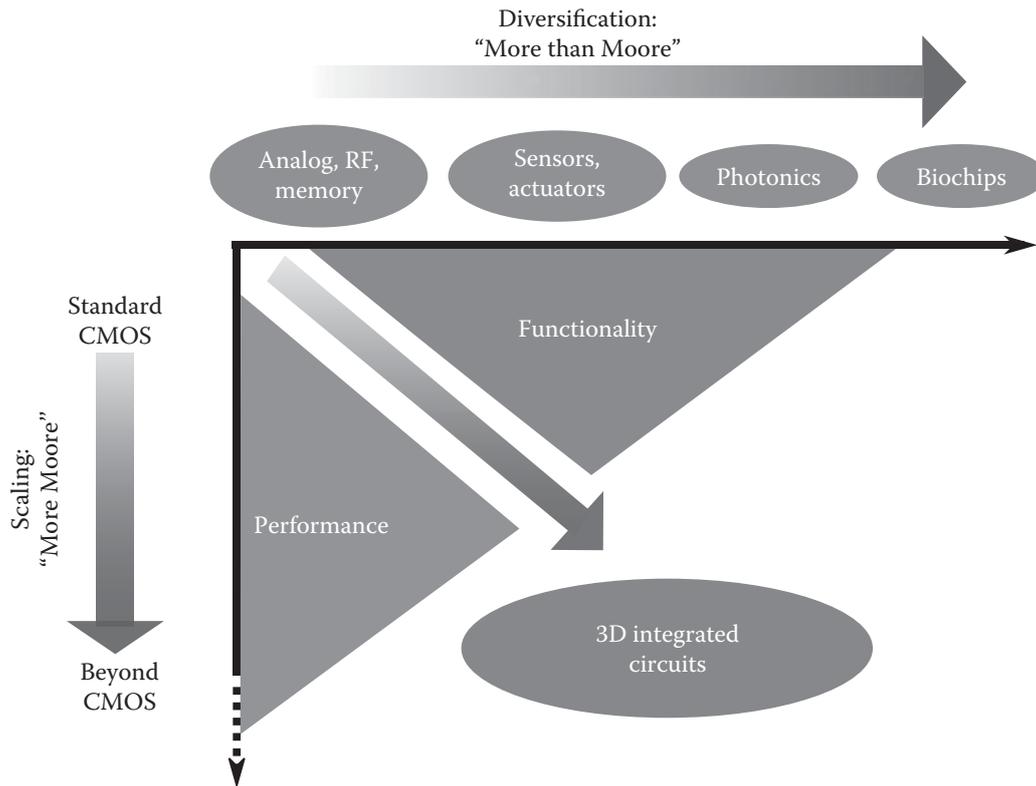


FIGURE 17.1 Besides the well-known trend for downscaling device nodes, slowly but surely reaching its limits for CMOS technology, the need for diversification has been acknowledged [9]. The concept of 3D integrated circuits is considered a promising option to combine both avenues.

In the previous chapters, various 3D design and application approaches have been discussed in detail. In this final chapter of the book, we aim to revise key technological and design challenges, and to point out prospective directions for further adoption of 3D-ICs.

17.1 KEY CHALLENGES FOR 3D CIRCUITS AND APPROACHES REVISED

The challenges for 3D circuits, as for any electrical device in general, can be classified into technological challenges and design challenges. In this section, we highlight both challenges and related approaches to provide an overview of the current state-of-the-art in 3D-IC design.

17.1.1 TECHNOLOGICAL CHALLENGES

Although much progress has been achieved in the recent years, technological challenges are still a hindering factor for mainstream adoption of 3D-ICs.

In the following, we first review different integration approaches for 3D devices, revealing the need for early and proper analysis of suitable technologies for 3D integration. Then, further key challenges, such as power delivery and thermal management, clock delivery and testing are reviewed.

17.1.1.1 Diversity of Integration Approaches

3D integrated devices are typically realized deploying one of the following approaches: package stacking, interposer-based packaging, through-silicon-via (TSV)-based 3D integration or monolithic 3D integration (Figure 17.2). Each of these approaches has its own scope of application, benefits and drawbacks, and requirements for design and manufacturing processes [108,128]. In the following, these approaches are briefly reviewed. Note that package stacking is not reviewed here since its concept is considered a well-known approach.

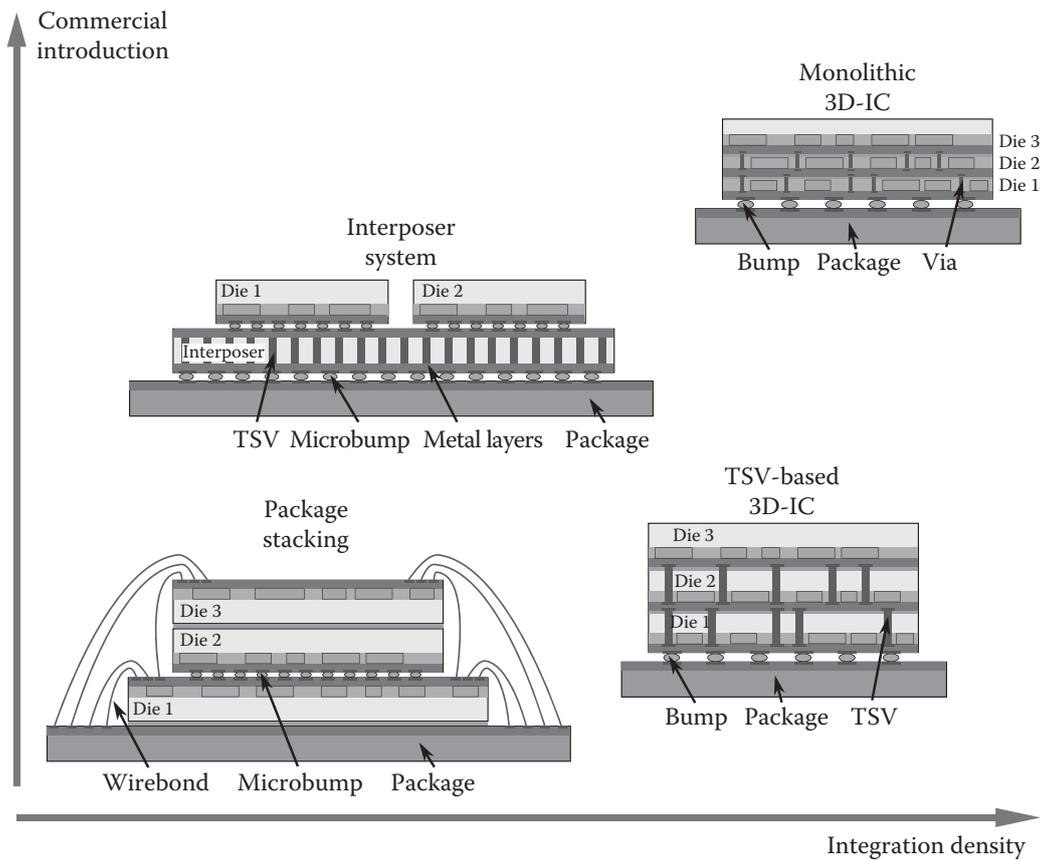


FIGURE 17.2 Evolution of 3D-integration technology. Originating with package stacking, 3D integration has evolved through interposer-based systems (also known as “2.5D integration”) toward TSV-based and monolithic 3D-ICs. Both integration density as well as design and manufacturing complexity have increased during this process. Furthermore, heterogeneous integration (mainly memory and logic) is the current key scope of application; homogeneous logic-on-logic integration is not yet foreseeable.

Interposer-based 3D packages are acknowledged as a cost-efficient driver toward 3D integration [67,98,140]. For such 3D packages, usually predesigned dies are stacked in lateral and/or vertical fashion on silicon carriers—the interposers—which comprise metal layers and TSVs for connectivity. Interposers are mainly realized as passive carriers, but can also include embedded components like decoupling capacitors or even glue logic [67]. Interposers support various integration scenarios and applications and are thus widely acknowledged in the industry.

The integration density of interposer-based 3D packages is the lowest among the approaches discussed here. Furthermore, the seemingly straightforward design of such packages is obstructed by the current lack of appropriate design tools [78]. For example, routing an interposer with its few metal layers, typically used to full capacity, falls short of expectations while using current tools. Further, planning and verification of signal integrity across the different domains in interposer-based devices is not sufficiently supported yet.

TSV-based 3D integration has evolved as a “prominent approach” for 3D integration; many researches and industry prototypes are based on TSV technology nowadays, for example, [3,15,29,37,56,102,124,126,141]. TSVs are metal plugs (mainly made of copper or tungsten) running through silicon dies, which are stacked and bonded, in order to interconnect them. Depending on the type/fabrication of TSVs, different design obstacles are occurring. Via-first and -middle TSVs occupy the active layer, thus resulting in placement obstacles; via-last TSVs and TSVs fabricated after bonding occupy the active layer as well as the metal layers, resulting in placement and routing obstacles [1,57]. There are manifold stacking configurations available, each having its advantages as well as disadvantages [1,128]. The classification mainly comprises wafer-to-wafer, die-to-wafer, and die-to-die stacking.

The concept of TSVs enables chip-level integration, while retaining the benefits of package-level integration [128]. Thus, TSVs are key enablers for 3D integration, as also proven by ample TSV-based prototypes, indicated earlier. Both heterogeneous and homogeneous 3D integration can be realized with TSV-based integration—an important feature to increase acceptance in the industry for such new technologies.

The integration density of TSV-based 3D devices is lower than that of monolithic 3D devices, but higher than that of interposer-based 3D packages. Due to their relatively large size and intrusive character, TSVs cannot be deployed excessively but have to be rather optimized in count and arrangement [63,64]. It is also notable that TSVs do not scale at the same rate as transistors, thus the TSV-to-cell mismatch will likely remain for future nodes and may even increase [101].

Monolithic 3D integration is recently gaining more attention [72,73,104], mainly thanks to advances in process technologies [13]. Active layers are built up sequentially rather than processed in separate and subsequently bonded dies. Due to very small vertical interconnects, monolithic integration enables fine-grain transistor-level integration; it provides the highest integration density among the three mentioned approaches. However, monolithic 3D-ICs also face further challenges, for example, the need for tools and knowledge for a low-temperature manufacturing process [13], or increased delays along with routing congestion [72,81].

Monolithic integration is nevertheless a promising approach, especially for high-density logic-integration [72]. Note that thermal management is even more challenging in such high-density logic-integration scenarios than it already is for “classical” 3D integration. A recent study by Samal et al. [113] has shown, however, that monolithic integration is superior in terms of heat dissipation compared to TSV-based integration.

Choosing the proper 3D-integration approach for a particular design is much more complex than handling the decisions typically required for classical 2D-ICs, for example, the selection of device nodes and packaging concept. As indicated earlier, each integration approach has its scope of application (mainly defined by the integration granularity) and benefits and drawbacks. Furthermore, given an abstract design description in early planning phases, one has to consider the following problems:

- In how many dies should the design be split up, and what technology/device node should be considered for each die?
- How is the system functionality of the design performing when it is split up into several components spread across separate dies? What is the appropriate partitioning strategy?
- How many interconnects are required between the components/subsystems, and what are the requirements for signal transfer? Which interconnect, bonding, and packaging technologies are applicable?
- How can the sub-systems and the overall device be tested?
- What package concept should be applied? Are constraints, such as thermal design, power, and signal integrity, met with the chosen package?
- In what order are the manufacturing steps to be conducted? Which manufacturing party is responsible for what deliverable?

Some of these earlier problems are interacting, and each respective decision does impact the overall performance, reliability and cost of the final 3D chip. It is apparent that addressing these complex problems requires experienced designers and well-defined project structures. Further, design tools that enable (1) a fast, yet accurate, exploration of the technological design space and (2) rapid evaluation of different configurations are crucial. Such tools have only recently become available; further details are discussed in Section 17.2.1.

17.1.1.2 Power Delivery and Thermal Management

The key advantage of 3D integration—high integration density thanks to vertical stacking of active layers or dies—also gives rise to significant challenges for power delivery and thermal management.

Assuming that a 3D-IC has d dies or active layers stacked, its potential power consumption is d times that of a classical 2D IC comprising one die. This is not entirely true, for example, because signal and clock interconnects have a large power consumption* but are much shorter in 3D-ICs than in 2D-ICs, which directly translates into power savings. Still, the *power density* and—along with it—the *heat-flow density* are notably increased in 3D-ICs. Power delivery is further impacted by the technological implications of stacking multiple dies: power/ground (P/G) TSVs contribute additional, notable resistance and inductance to the power-supply network [38]. Thermal management is similarly complicated by vertical stacking of dies: “thermal barriers” between dies are introduced, which are the layers required for die bonding. These bonding layers are, for example, made of BCB adhesive polymers, which have a thermal resistivity several hundred times that of silicon dies [105].

To account for these major challenges, many studies and approaches for power delivery and thermal management have been proposed. In the following, the most relevant are briefly explained.

Power-delivery networks for 3D-ICs should be designed considering the following:

- *Proper arrangement of TSVs:* Studies by Healy et al. [38,39] point out that a distributed topology for P/G-TSVs is superior to both single, large TSVs and groups of clustered TSVs. These and other studies, for example, [19,50], also favor irregular TSV placement, in particular such that regions drawing significant current can exhibit a higher TSV density.
- *Optimized power-delivery architectures:* To limit package impedance and external current supply, one can bring DC–DC converters closer to the logic circuitry, as demonstrated in [122] with a dedicated DC–DC die. Another possible architecture is the “multi-story power delivery” [46], where several power domains/supplies (e.g., one per die) reduce the respective load compared to a single, classical power supply. In general, design and optimization of P/G grids needs to account for the overall 3D power-delivery network, not only the part/die it is attached to [42,65].
- *Decoupling capacitor (decap) allocation:* To reduce power-supply noise, classical CMOS decaps and/or metal-insulator-metal (MIM) decaps can be deployed within each die [149] or even in dedicated decap dies [42]. Decap allocation has to be carefully analyzed to comprehend their impact on the complex power-supply noise distribution in 3D-ICs [60].

For thermal management of 3D-ICs, general approaches include

- *Low-power design:* Reducing overall power consumption and, thus, heat dissipation can be achieved by deployment of low-power circuitry.
- *Thermal-aware physical design:* Spreading high-power modules away from each other and arranging them in the dies closer to the heatsink, for example, during thermal-aware floorplanning [45,63,66] and placement [90], are simple but effective design measures.
- *Reducing thermal resistances:* Both internal paths (i.e., paths across and within dies) and external thermal paths (i.e., paths to the heatsink and the package) can be improved by new technologies and/or even simple design techniques. For example, TSVs can be grouped into TSV islands, which are then arranged and aligned such that they serve as effective “heat-pipes” [20]. A notable technology for reducing internal thermal resistance is the deployment of micro-fluidic channels [117].

Besides these separate approaches for power delivery and thermal management, some studies [19,65,71,76] investigated co-optimization of power delivery and thermal management and provided effective techniques. For example, their commonly proposed arrangement of P/G-TSV stacks (i.e., TSVs aligned across the whole 3D-IC) into high-power design regions is self-evident: regions with

* For modern 2D ICs, signal interconnects contribute nearly one-third of power consumption [119], and clock networks may consume even up to 50% of total power [150].

high current demand benefit from such P/G-TSV stacks in both terms of increased heat dissipation and reduced power-supply noise.

17.1.1.3 Clock Delivery

For 2D VLSI designs, the clock-network synthesis is one of the most important stages in the design flow because clock-distribution delivers the clock signals to clock sinks (such as latches, flip-flops, and registers) and synchronizes the calculation. Therefore, design automation for clock network is mostly used for block-level designs while designers manually craft the global clock distribution to perfection using SPICE-level simulation. It is not surprising that clock delivery is one major obstacle for general application of 3D-IC technology that, by its nature, targets designs with higher frequency (“More Moore”) at a higher cost than 2D designs.

One major challenge of 3D clock distribution is for the clock signals to arrive all clock sinks reliably in different dies. Typically, clock networks in different dies are connected by TSVs and it is shown that 3D clock networks with multiple TSVs yield shorter wirelength [62,99,143,144], which leads to lower power dissipation [147] and shorter clock latency. However, this argument has some potential caveats.

- *TSV reliability.* TSVs are less reliable than other 2D interconnect structures as they are subject to random open defects [86]. Therefore, different TSV-redundancy mechanisms have been proposed [41,55,86], using spare TSVs or TSV groups with reconfigurable routing. One straight-forward implementation of reliable 3D clock distribution is to have double TSVs along the whole clock network, which greatly increase the TSV overhead for clock routing. More research is needed for the industry to understand the tradeoff of TSV utilization between power/latency and fault tolerance.
- *TSV placement.* Other than the fault-tolerant TSV design for clocking, another challenge is TSV placement for clock distribution. Note that a lot of 3D physical-design algorithms extend from 2D algorithms and some assume that TSVs can be placed anywhere except in a set of pre-defined blockages. In reality, TSVs have to be specially designed in order to prevent known reliability problems [96,129] such as mechanical issues [52,96]. A recent work proposes decision-tree-based algorithms to select clock TSVs from a set of TSV arrays [145]. For high-performance 3D designs, it is expected to have specially designed and pre-placed clock TSVs according to a particular clocking style, which will be discussed in details in Section 17.1.2.4.
- *Floorplanning and hierarchical design.* A lot of clocking-automation research in the literature for 2D and 3D designs did not fully consider the fact that a good clock-distribution network has to support other physical-design stages in order to honor all the timing constraints. One example is that clock skew between clock sinks which are constrained by timing checks is much more critical than the global clock skew. This statement obviously also applies to 3D designs where designers usually put each functional partition/module in a single die, instead of separating it into more than one die because of the high connectivity inside the module. In this case, the interdie skew (among different functional blocks) usually is less important than the intradie skew (especially when the clock sinks are related by timing paths). Therefore, replacing the intramodule clock networks with TSVs connecting to different dies in order to reduce clock wirelength, power, and latency is probably not the right thing to do.

One potential application for 3D design is heterogeneous stacking, for instance, to stack dies from different-technology nodes. Microprocessor designers can design a new core and put it in a different die with new technology node while all other modules remain within the established technology node. For this case, 3D clock design (for both the new core design and TSV planning) has to be able to reuse the existing clocking structure. As a result, 3D clock delivery is usually highly dependent on the choice of the design flow and technology, and discussion on clock-synthesis algorithms without details of the design methodology is far from practical.

- *Testing.* As will be described in the next subsection, testing is vital on the road toward 3D adoption from an industrial perspective. Clocking is highly coupled with testability because prebond and midbond testing are very attractive DfT architectures. For example, a prebond-testable clock tree is presented in [74]. The idea of having a complete clock tree in each die basically limits 3D clock-network-synthesis research trying to replace long global clock wires on each die with TSVs. At the same time, TSVs connecting those clock networks on each die is a feasible clocking structure (Section 17.1.2.4).
- *Power/ground network synthesis.* It is well known that global clock wires have to be shielded to minimize coupling and noise, while one common choice to do so is to use power/ground wires as well as P/G TSVs. In some technologies, metal stacks and wirecodes are pre-defined and characterized. Hence, there is a need for co-optimization between P/G networks and global clock distributions.

In Section 17.1.2.4, the latest design challenges and state-of-the-art clock-distribution algorithms will be explained in detail.

17.1.1.4 Testing

Due to the inherently stacked arrangement of 3D circuits, testing is much more complicated than for 2D circuits and is still considered as a key obstacle for high-volume manufacturing of 3D circuits. Appropriate testing setups need to provide solutions for the following new problems [94]:

- Fault models and tests for wire-based and TSV-based interconnects with, for the latter, consideration of related intradie defects.
- Wafer probing on thinned dies, for dense arrangements of microbumps or TSVs and landing pads, considering stringent mechanical constraints.
- Design-for-Test (DfT) architectures, tailored for testing parts of the stack as well as for testing the whole stack.
- Optimization of the test flow for efficiency and limited cost-time overhead.

State-of-the-art studies which address the earlier problems are outlined next.

Fault models for both interposer-based interconnects [43] and TSV-based interconnects [26,44,85,89] have been proposed. The latter studies focus on specific types of interconnects and/or defects: Loi et al. [85] model and implement fault-tolerant 3D NoCs; Lung et al. [89] address fault-tolerant clock networks; Deutsch et al. [26] propose thermo-mechanical-stress-aware generation of test patterns; and Huang and Li [44] propose built-in self-repair scheme for TSVs.

Wafer probing, that is, early access of the dies' pads for testing purpose, is challenging in the context of 3D integration. More specifically, typical microbumps are too small, too densely arranged, and too fragile to be probed with conventional technologies [92]. Furthermore, the dies thinned for 3D stacking cannot be exposed to large probe weights, which range from 3 to 10 g per probe tip in conventional technologies [69]. New technologies, however, have been successfully developed and are becoming available. In [121], a lithographic-based MEMS, probe card was presented by Cascade Microtech, Inc. and IMEC. This technology is suitable for probing 40 μm - or smaller-pitched arrays, while inducing probe weights of only 1 g per tip. Another option is contact-less probing, as for example demonstrated by ST Microelectronics with capacitively coupled probing [115].

As for DfT architectures, testing facilities have provision for testing separate dies (i.e., *prebond testing*) and for testing the final stack (i.e., *postbond testing*). Further, testing the partial stack (i.e., *midbond testing*) is also relevant [92]. To enable such flexible facilities, modular setups are required. In practice, some wrapper circuitry is to be deployed on each die, which links with test facilities on other dies and, thus, across the whole 3D stack [75,94]. This is also acknowledged for the work-in-progress IEEE standard *PI838* [93]. The related wrapper enables controllability and observability at the die boundaries, which ensures interoperability between different dies (possibly even from

different manufacturers). Besides these wrapper components, existing testing facilities should be reused whenever possible. In *PI838* [93], for example, the well-known concepts of test-access ports and scan-chains are applied.

Prebond testability is also associated with another notion, that of integrating only *known-good-dies*. Full and proper testability of separate dies is crucial, and only after these tests succeed, can the 3D stack be safely constructed. In this context, design partitioning also largely affects DfT architectures and testability: the finer the partitioning granularity, the more signals of (partial) modules pass across dies (instead of remaining encapsulated within dies), and the more complex the DfT architectures will become, especially for prebond/known-good-die testing [75].

To optimize the test flow, Noia et al. [103] have studied effective and efficient scheduling of test patterns. Furthermore, they have shown that minor increase in test pins enables great reduction in test time. In another study by Chen et al. [21], it was shown that reducing time and cost for prebond testing is possible despite strictly limiting deployment of additional test pins. Agrawal et al. [8] have proposed a heuristic methodology for test-flow selection, which flexibly adapts for different scenarios of 3D integration.

17.1.2 DESIGN CHALLENGES

Besides technological challenges, design challenges are also still impeding the broad and successful adoption of 3D-ICs. Compared to classical 2D chips, a 3D-IC is a much more complex system; related design algorithms and simulation and verification have to account for complex (and sometimes conflicting) interactions of multiple design criteria and physical domains. Putting the initial (probably too optimistic) expectations for “straightforward 3D solutions” into perspective, researchers and industry experts are now concerned about more pragmatical approaches. This includes to carefully analyze the *scope and applicability* for 3D-ICs, considering the available manufacturing and design approaches and their cost-benefit trade-offs.

Many challenges for physical design of 3D-ICs stem from the simple fact that the solution space is notably increased by adding one dimension compared to 2D-ICs [30]. This naturally escalates complexity for all physical-design steps.

Next, we discuss aspects of design complexity, algorithms, simulation, and verification, following the simplified design flow as shown in Figure 17.3. We also look into state-of-the-art approaches for design challenges.

17.1.2.1 Layout Representations for 3D Circuits

Physical design automation of electronic devices is generally based on abstract models of the corresponding design problems. These models are computationally represented as data structures. The data structures, in combination with accordingly tailored operations, for example, direct access to adjacent blocks, are subsequently referred to as *layout representations*. For 2D floorplanning, it has been shown by Chan et al. [17] that deploying different layout representations induces only minor deviations of final design quality—this is true despite the fact that different mathematical

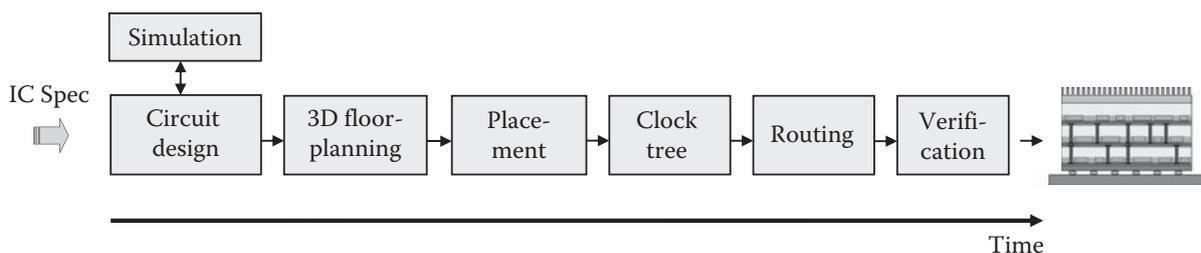


FIGURE 17.3 The major steps in the circuit design flow with a focus on physical design. Please note that this is a simplified view as in reality boundaries are blurred and iterations between these steps are common.

descriptions are at the heart of various representations. Chan et al. observed that a key bottleneck of floorplanning is to evaluate the actual layout (traditionally in terms of packing density and wire-length), and not to handle the abstract representations themselves.

For 3D design, this situation is even more intricate. 3D physical design has to consider a much more sophisticated set of design criteria [45,63,79]: wire-length, fixed outlines, thermal management, packing density, TSV management, power delivery, arrangement of massive interconnects within and across dies, 3D-stack-related noise coupling, etc. Given such complex and often interacting design criteria, it is apparent that 3D layout representations must be tailored carefully. Effective and efficient representations should provide the following features [30]:

- Inherent support for crucial constraints, for example, for spatial constraints given by vertical arrangements of modules interconnected across dies.
- An optimized solution space, that is, minimal redundancy while covering best solutions.
- A versatile set of operations applicable to different “granularities” of physical design. For example, such a set may include global operations, like swapping blocks across dies, and local operations, like shaping soft blocks.
- Fast transformation of abstract solutions into actual layouts and vice versa.
- Straightforward determination of correlations between abstract solutions and prospective design quality, for example, blocks adjacent in the abstract representation will also be adjacent in the actual layout. This is beneficial for speeding up layout evaluation, a key bottleneck for physical design.

In recent years, many effective 3D layout representations have been proposed, for example, layered transitive closure graph [68], T-tree [138], or Corblivar [63,66] (Figure 17.4). While some of these representations are derived from existing 2D representations, others are inherently developed for 3D integration.

Arising from this classification, there are two ways to represent vertical dependencies. The first option is to deploy multiple instances of classical 2D representations, labeled “2.5D” in Figure 17.4. Here, additional mechanisms have to be implemented to consider vertical relations between modules placed among different dies, such as vertical alignment as well as overlapping and nonoverlapping constraints. These representations include a discrete z -direction, such as in the combined bucket and 2D array (CBA) approach in [24]. However, it is obvious that vertical dependencies should be

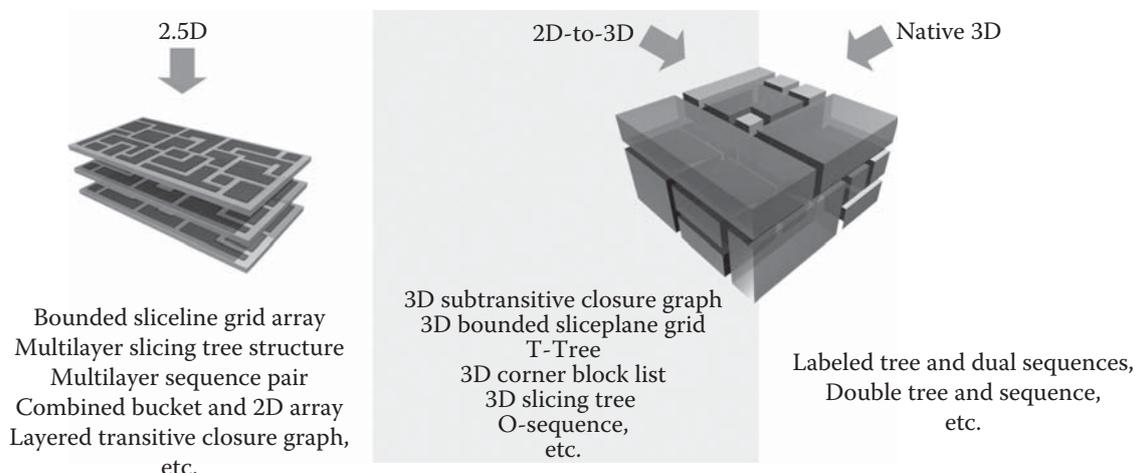


FIGURE 17.4 Categories and examples of layout representations tailored for 3D design [32]. Multiple instances of classical 2D representations are labeled “2.5D,” “2D-to-3D” characterizes former 2D representations adjusted to the third dimension and “Native 3D” representations are specifically designed for 3D design.

incorporated directly into the representation. Hence, more recent 3D representations consider multi-layer modules natively in all three dimensions (labeled “2D-to-3D” and “Native 3D” in Figure 17.4). An example for such genuine 3D representation is the 3D Slicing Tree described in [22].

Besides representations themselves, several studies focused on implications of 3D representations and respective design methodologies; some of these studies are outlined next. Wang et al. [131] have shown that consistent correlations between abstract and actual layouts (an important feature for 3D representations, as indicated earlier) are not easily achieved unless $P = NP$. This naturally increases complexity of 3D physical design. Fischbach et al. [30] have developed a methodology to evaluate and compare representations for particular needs. Their methodology is based on Monte-Carlo sampling and analysis of respective solution-space distributions. Quiring et al. [111] have proposed a meta-heuristic methodology, which aims to apply probabilistic optimization techniques (like the well-known simulated annealing) more effectively. The key idea is to track each (past) layout operations’ impact on relevant design criteria like thermal management. Then, (future) layout operations are deployed according to most prospective benefits for design quality.

17.1.2.2 Partitioning and Floorplanning

Partitioning divides the design into smaller blocks, each of which can be processed with some degree of independence and parallelism. A *divide-and-conquer strategy* can be implemented by laying out each block individually and reassembling the results as geometric partitions. Historically, this strategy was used for manual partitioning, but became infeasible for large netlists. In contrast, *netlist partitioning* can handle large netlists and redefine a physical hierarchy of an electronic system, ranging from boards to chips and from chips to blocks. Independent of the approach, the resulting partitions, subsequently also called *modules*, range from a small set of electrical components to fully functional ICs.

The initial partitioning step for 3D-ICs divides the circuit into several balanced partitions equal to the number of dies. The goal is, among others, to minimize the connections between dies. This translates into reducing the number of vertical interdie connections and decreasing the area overhead associated with TSVs, as discussed in earlier sections. After dividing the netlist or the circuit into multiple dies, 3D partitioning usually requires a subsequent *intradie partitioning*. Since both inter- and intra-die partitioning steps are extremely technology dependent, they are not further investigated here.

Floorplanning is closely related to partitioning. During floorplanning, the shapes and positions of the partitions/modules (such as digital and analog blocks) are determined. Thus, the floorplanning stage determines the external characteristics—fixed dimensions and external pin locations—of each module. These characteristics are necessary for subsequent placement (see Section 17.1.2.3) and routing steps (see Section 17.1.2.6), both of which determine the internal characteristics of the module.

Conventional floorplanning assumes a single 2D layer on which several modules must be arranged. 3D floorplanning includes new 3D-specific properties that must be represented in the underlying layout representations. For example, modules have vertical dependencies in addition to horizontal ones. As discussed in the previous subsection, layout representations should inherently consider these dependencies to facilitate efficient 3D floorplanning. For example, the 3D Slicing Tree described in [22] provides related features. As illustrated in Figure 17.5, different operations, such as module rotation and swapping, can be carried out efficiently to modify a given tree. A concatenation of these operations allows obtaining any possible slicing tree from any given slicing tree, thus enabling flexible 3D floorplanning. However, solutions from a 3D Slicing Tree are limited to slicing floorplans.

Besides considering vertical dependencies, 3D floorplanning should also account for reducing peak temperatures of 3D designs [63]. In addition to the increased power densities of stacked modules, peak temperatures are closely related to long wires on the chip due to interconnect power consumption [45].

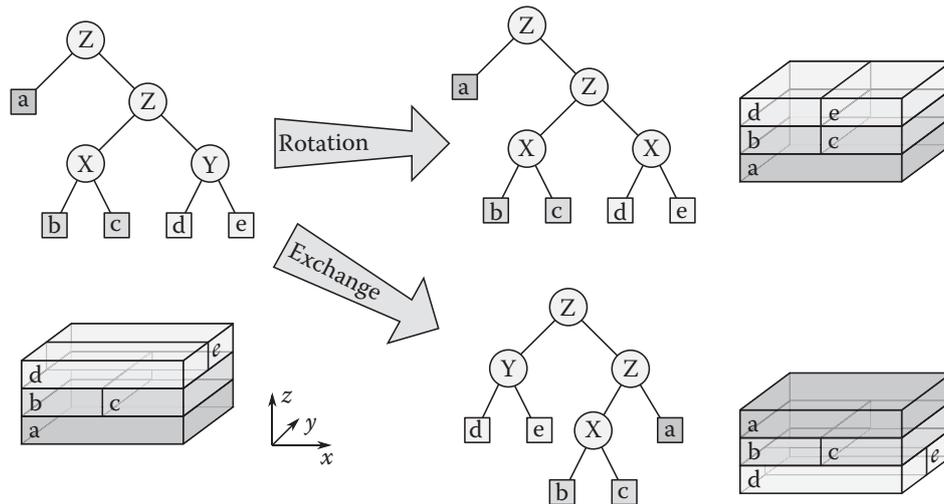


FIGURE 17.5 Illustration of 3D Slicing Tree operations to permute a given 3D floorplan [31]. A rotation alters an inner node (representing a cut through the normal plane) resulting in a physical rotation of modules contained in the subtrees of that node. An exchange swaps two subtrees resulting in a physical exchange of modules contained in these subtrees.

17.1.2.3 Placement

After floorplanning, the design is ready to be placed. That is, the next step in the design flow is to determine the location of each cell within its respective module (partition). The objective of placement is to determine the location and orientation of all cells, given solution constraints (e.g., no overlapping between cells) and optimization goals (e.g., minimizing total wirelength).

Depending on the applied partitioning approach for the 3D design, intra-die (2D) placement is limited to one die (layer), whereas interdie (3D) placement includes optimization between multiple dies. The latter approach requires new placement methodologies because of the “mismatch” between vertical and horizontal granularities: 3D layouts have limited flexibility in the third dimension due to both the relatively small number of dies and the scarce availability of TSVs. According to [36], this favors partitioning approaches (rather than force-directed techniques) at least during global placement. Accordingly, this work initially uses recursive bisectioning to perform global placement, with nets weighted according to the number of TSVs.

Quadratic placement approaches for inter-die placement require the “move force” to be modified such that cell overlap is eliminated in each die separately. More precisely, a move force should not be applied between two cells sharing the same x - and y -coordinate if they are located in different dies.

As mentioned previously, thermal constraints are crucial for reliable 3D designs. Hence, both intra- and interdie placement must spread cells such that a reasonable temperature distribution can be expected. However, due to the increased packing density in 3D-ICs, additional techniques are required to tackle the heat dissipation issue. For example, any vertical metal structure serves as “heat remover”—these structures play an important role in achieving a thermally solid design, and are in this context also called *thermal vias*.

Resulting from thermal constraints, 3D placement must not only place cells and consider regular TSVs but also take thermal vias into account. While intradie placement concerns only one die, the placement of a thermal via is affecting all dies; due to its aligned character, it creates a blockage throughout all dies. As such, thermal vias may represent a severe problem for cell placement and routing. Furthermore, cell placement and thermal-via placement are interacting because the position and size required for a thermal via depend on the thermal energy (i.e., power dissipation) of the cells nearby. Practical solutions, such as work presented by Goplen and Sapatnekar [35], have addressed these problems by designating specific areas within the circuit as potential thermal-via sites (Figure 17.6). Here, thermal conductivity of each region (site) can be considered a design

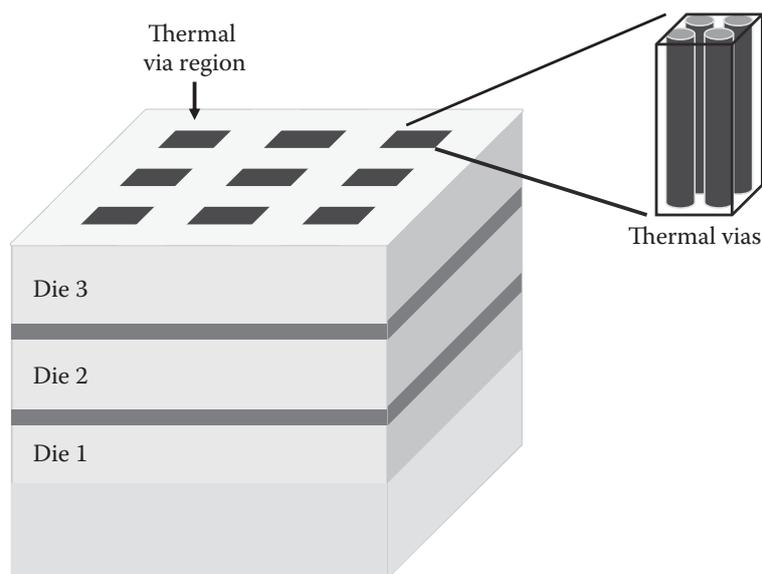


FIGURE 17.6 Regularly arranged thermal-via regions in a 3D-IC. Such regions unify the placement of a specific number of vias. The regions are sized according to the number of vias which are required for meeting the thermal requirements.

variable that is only subsequently translated into a precise number of thermal vias to be placed inside this region. Another advantage is the regularity of these sites (blockages), which can be addressed much easier than spread-out blockages, for example, during subsequent routing.

The aforementioned planning of thermal vias during placement applies also to regular TSVs, which—besides their electrical function—can provide heat-flow paths as well. An approach to grouping them into TSV islands (and thus to reduce their impact on placement, among others) has been presented in [64]. Further studies [12,23,90] consider detailed properties of TSVs in thermal- and wirelength-aware placement algorithms.

17.1.2.4 Clock-Distribution Networks

As mentioned in Section 17.1.1.3, 3D clocking is facing tremendous challenges such as TSV reliability, placement, as well as the dependence on design hierarchy, DfT, and power/ground synthesis. In fact, all the conventional objective and constraints on clock-network synthesis for 2D are getting more complicated, such as power, thermal, skew, slew, clock latency, jitter, glitch, and duty cycle. While Chapter 7 explains problems and algorithms for 3D clock networks, this subsection focuses on the future direction and challenges related to physical-design automation for 3D clock distribution.

Clock-skew minimization is one of the most important objectives for clock-network synthesis and is considered by almost all prior 3D clocking approaches, such as [61,62,89,146]. However, few works use SPICE simulation [109] and undergo Monte-Carlo simulations to obtain the clock-skew distribution [148]. Since 3D technologies are by its nature more fitted for high-performance designs with relatively higher manufacturing cost, it is expected that 3D clock networks have to be tailored for multiple-Gigahertz clock frequency and sub-20 ps of clock skew/jitter. Therefore, accurate timing simulation considering inter- and intra-die variability is a must for skew analysis to be realistic for 3D clock networks. A thorough study of clock variations for different number of TSVs is presented in [148], which is a very good starting point. Along this direction, a statistical clock-skew model considering inter- and intradie variability is presented in [133]. However, future work is urgently needed to understand how TSV usage affects the distribution of timing-impacting clock skew, which is the clock skew between each pair of clock sinks connected with real timing paths. In other words, variability analysis on clock-distribution network has to consider the full timing picture, which is tightly coupled with floorplanning and design hierarchy.

As mentioned previously, it is not likely that a logical partition/module is separated and placed in different dies. Therefore, there are many more timing paths between clock sinks on the same die, and it is more important to keep clock skew to be sub-20 ps for those clock sinks. In this case, we would like to maintain the common clock path of those clock sinks as much as possible and it is not preferable for their upstream clock paths to be on another die. In other words, it is desirable to keep the whole clock network of each module on the same die. An interesting research direction is to study the optimal TSV utilization and placement between the intact clock networks of partitions/modules.

With more variability and tighter skew/slew requirements of 3D designs, it is expected that H-tree and clock grid [109] are more preferable than the classical synthesis algorithms like Method of Mean and Median (MMM) and Deferred-Merge Embedding (DME) [146]. Pavlidis et al. [109] examine 3D clock structures having an H-tree on the middle plane. They found that the mix and match of H-tree, global rings, and local rings yield mixed clocking quality of results with no obvious winning architecture. This reinforces our discussion in Section 17.1.1.3 that each 3D clock-synthesis approach has to be driven and evaluated by a well-defined design methodology. One useful future work is to examine different 3D clock-grid structures where 2D global meshes are linked by TSVs. Since one of the most difficult tasks for 2D clock-grid design is the tuning of mesh wires and buffers driving the mesh, it is essential to achieve 3D clock-grid tuning in reasonable runtime. In summary, we are yet to see how different clock-synthesis algorithms perform in a real 3D hierarchical microprocessor-design flow with industrial timing analysis.

Fault tolerance and testing in 3D clock distribution is much more complex than that for 2D clocks because TSVs are subject to random open defects as mentioned in Section 17.1.1.3. There are different clock-synthesis algorithms considering TSV redundancy [88] and the introduction of fault-tolerant components [89]. While previous works derive fault-diagnosis test sequences to identify single and multiple defective TSVs [112], it is also important to use Monte-Carlo simulation for the timing of clock distribution.

Global TSV planning and codesign are crucial for 3D physical design because clock-network synthesis is highly coupled with almost all other physical-design problems, for example, floorplanning, placement, routing, timing optimization. One example is co-optimization of clocking and power/ground networks, especially when the technology and design rules restrict that clock routing has to be shielded by P/G wires. Another good example is [118] where Shang et al. derive an electrical-thermal model for both signal and thermal TSVs, and use the model to generate thermal-reliable 3D clock trees. In fact, new algorithms are urgently needed to simultaneously plan signal, power/ground, clock and thermal TSVs during 3D physical design.

17.1.2.5 Routability Prediction

One of the last steps for physical design is signal routing, that is, defining the interconnects' geometry (Section 17.1.2.6). As a result of the routing stage, not only the interconnects are deployed but also electrical properties of the circuit are defined. In order to achieve good routing results, all previous design stages have to be optimized with regard to routability. Therefore, *evaluating routability* is an inherent part of most design stages. 3D circuits with complex interconnect topologies require new approaches for routability prediction.

Any routability-prediction method is valuable only if it allows computation in significantly less time compared to actual routing. This can only be achieved by using effective simplifications. These range from fast ("rough") estimations of routing paths to the time-consuming (but more accurate) global-routing procedure.

Global routing for 3D design, that is, global-routing algorithms that find routing paths in several interconnect layers while considering different types of vias, has been solved for some years. Various multilayer global routers are applicable to 3D circuits if vertical routing capacities (i.e., vias) within dies and between dies can be specified independently. This differentiation is necessary in order to respect the different properties of interdie vias (which are often implemented by TSVs) and conventional signal vias.

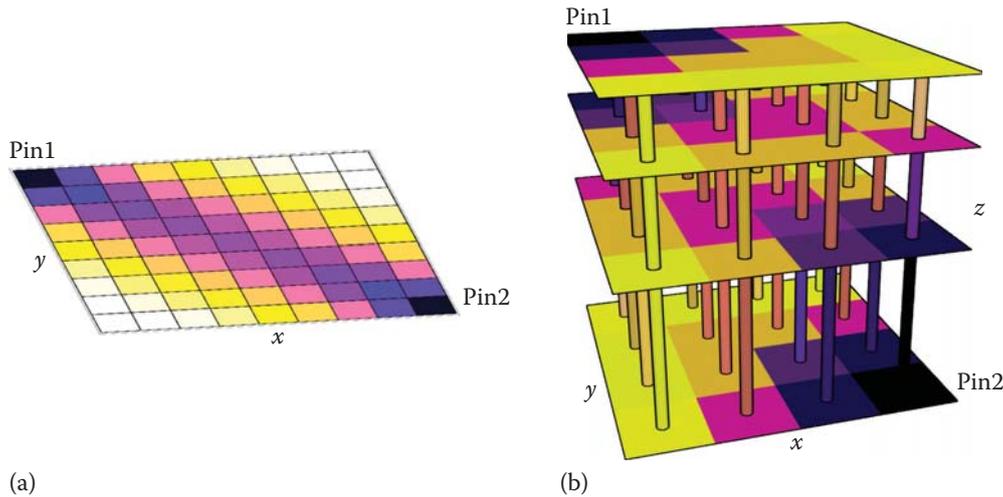


FIGURE 17.7 Routing-density distribution for a two-pin net in one layer (a) and extended 3D distribution in four dies (b). A darker color indicates a higher expected density. All routing paths are assumed to be of the same probability (i.e., no blockages exist).

However, fast 3D-routability prediction without performing global routing is still an open research topic. Such related fast methods are based on simple estimations (“informed guessing”) of routing paths. They require to extend routing-density distributions to 3D in order to adapt statistical estimations of routing demand to the requirements of 3D interconnect topologies.

Conventional 2D-routing-density distributions predict the routing demand, overflow, congestion, and thus, routability in a 2D plane (Figure 17.7a). This model must be extended by *vertical* routing capacities and routing-density distributions for each die to render it applicable for 3D-IC design.

A model capable of representing a 3D-routing-density distribution was presented in [31] and is depicted in Figure 17.7b. Depending on the level of abstraction, the layers of the density distribution either correspond to individual routing layers or to the combined layers of one die. Constraints such as blockages and varying densities of interdie vias are considered by means of a varying probability for routing paths [31]. Using this 3D-routing-density-distribution model, it is possible to predict the routing demand for 3D-ICs and to estimate the routing densities in each layer (die) as well as the expected (vertical) interdie-via density.

A more recent model [58] considers the TSVs’ impact on estimated routing topologies with particular focus on delay and power consumption.

17.1.2.6 Routing

A net is a set of two or more cell pins or terminals that has the same electrical potential in the final chip design. A circuit netlist includes all of the nets in a design. During the routing stage, all terminals of the nets must be properly connected while respecting constraints (e.g., design rules, routing resource capacities) and optimizing routing objectives (e.g., minimum total wirelength, maximum timing slack).

As already indicated, the main difference between regular (2D) and 3D routing is caused by the multi-die positions of net terminals that lead to net topologies that span more than one die (Figure 17.8). This requires expensive interdie vias (again, often implemented by TSVs) to be used in addition to regular signal vias, which connect metal layers within one die. Furthermore, 3D routing must take additional constraints into account, such as blockages introduced by thermal or interdie vias. These constraints require a much more sophisticated congestions management and blockage avoidance as it is applied for regular 2D routing. Additionally, the limited availability of interdie vias requires a careful allocation of this valuable resource among nets. The increased thermal impact on

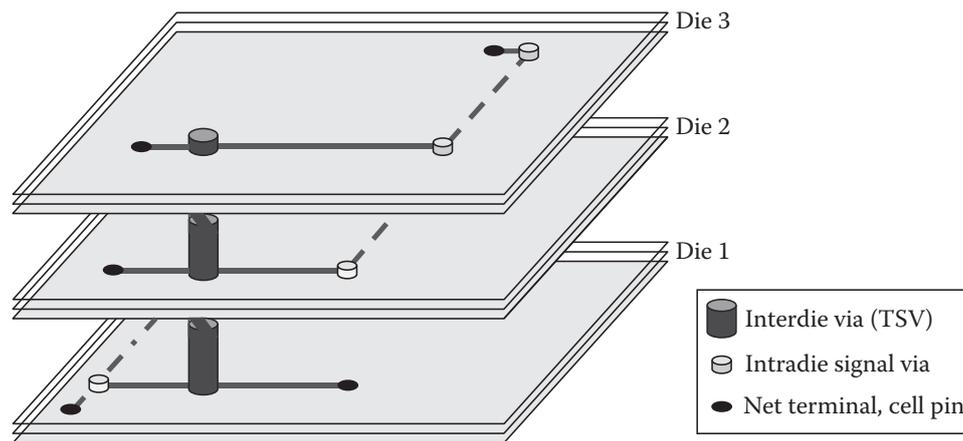


FIGURE 17.8 Example net route spanning multiple dies in a 3D design.

3D designs must also be considered during routing. For example, it is known that the delay of a wire increases with its temperature. Hence, critical nets must avoid hot regions of the chip.

Cong and Zhang presented a thermal-driven 3D router [25] using a multilevel-routing approach composed of recursive coarsening, initial routing, and recursive refinement. Its major feature is a thermal-driven via-planning algorithm. Based on this global view and capabilities for a multilevel scheme, the via-planning step effectively optimizes temperature distribution and wirelength using direct planning of the interdie vias instead of indirect planning through a routing-path search.

Another approach was presented by Zhang et al. [142]. It tackles the temperature-aware 3D-routing problem not only by using thermal vias but also by introducing the concept of *thermal wires*. Thermal wires are dummy objects with the function of spreading thermal energy in the lateral direction. Thermal vias perform the bulk of the conduction toward the heat sink, while thermal wires help distributing the heat paths among multiple thermal vias.

The well-known Steiner routing was also extended for 3D design. In [106], the authors propose a two-step flow: tree construction and tree refinement. The tree-construction step builds a delay-oriented Steiner tree under a given thermal profile. During tree refinement, TSVs are rearranged to further optimize the thermal distribution while preserving the routing topology and considering performance constraints.

17.1.2.7 Multi-Physical Simulation and Verification

Traditionally, physical design is separated from verification which aims to guarantee the intended functionality of a chip [54]. Simulation, on the other hand, is acknowledged as crucial part of physical design, for example, for thermal analysis of a chip. Verification is more detailed and complex than simulation, and typically leverages different simulation and analysis techniques itself. For example, electrical rule checking (ERC) verifies the correctness of power and ground interconnects, capacitive loads, signal transition times, etc. For proper handling of a 3D-IC's complex nature, simulation and verification techniques have to be deployed in a holistic manner in order to ensure design closure. The key reason for that requirement is given by the strong coupling of different physical domains in a 3D chip, and the resulting strong impact on overall design behavior and reliability [116].

The thermal, electrical, and mechanical domains are key subjects for multi-physical 3D-IC simulation and verification—with the domains' coupling being fortified by the high packing density in such chips [78,116] (Figure 17.9). Managing the *thermal domain* is much more challenging for 3D-IC design than for classical 2D design. With large thermal footprints, there is also an increasing impact on the *electrical domain*, that is, the behavior of active components. Since the leakage power of transistors is exponentially dependent on the temperature, a positive feedback mechanism arises, which, in worst-case scenarios, may lead to a thermal runaway and overheating of the 3D-IC.

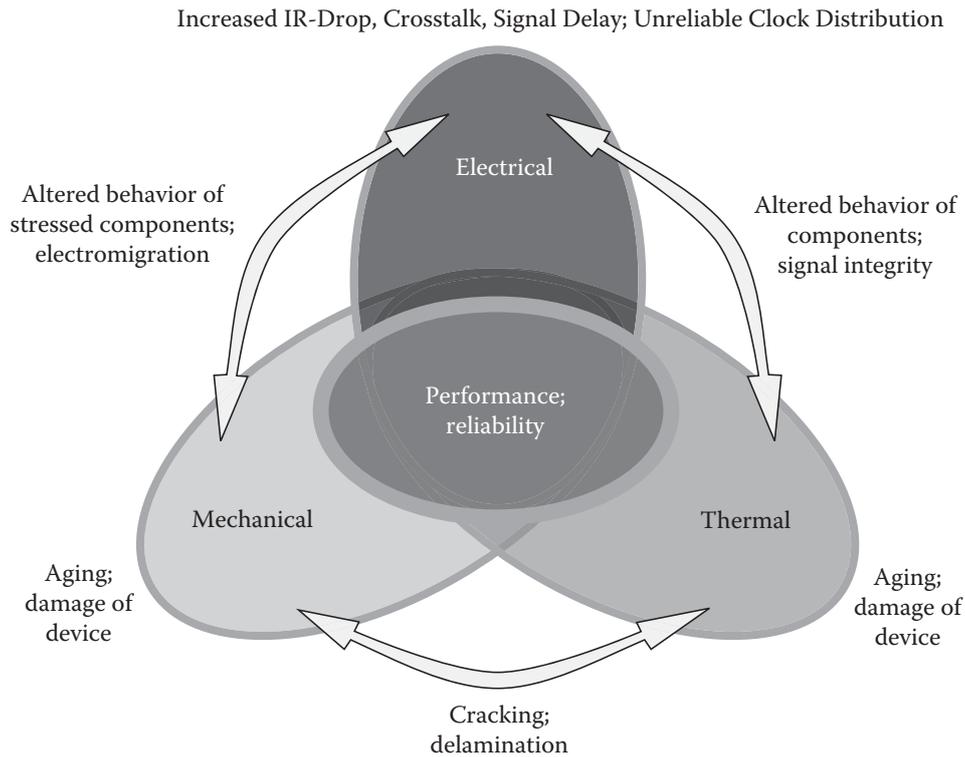


FIGURE 17.9 Coupling of the thermal, mechanical, and electrical domain in 3D-ICs and the related impact on device performance and reliability.

Besides, varying interconnect structures (e.g., metal layers “mixed” with TSVs) also impact the electrical domain: power and signal integrity, coupling, crosstalk and delays are all subject to varying interconnect structures’ properties, dominated by large discrepancies in geometry and size. The interfaces between (very large) TSVs and (very small) metal wires are especially prone to electromigration [77,107]. The *mechanical domain* is mainly influenced by the complex composition of 3D-ICs: material properties are varying strongly, for example, due to the “intrusion” of silicon chips by copper TSVs. The coefficient of thermal expansion for copper is approximately six times larger than for silicon, leading to notable thermo-mechanical stress in the surrounding of TSVs [53]. Such stress impacts both the performance and reliability of the chip; it even increases the likeliness for cracks or delamination [18].

These complex multi-physical interactions in 3D-ICs give rise to high demands on simulation and verification approaches. As indicated earlier, simulation and verification should be deployed into 3D physical design as early and holistic as possible. To do so, *hierarchical* modeling and simulation frameworks are a commonly accepted approach [116]. Such hierarchical frameworks include models and respective techniques for different levels of design abstraction or design phases [78]:

At the lowest level of abstraction, that is, for physical design and verification at transistor level, very detailed models are required. They must capture the composition of all devices with their specific geometries and material properties. Such models are also essential for evaluation and optimization of 3D interconnects technology, for example, for (individual) TSVs with regard to materials and geometries. The models are characterized by high accuracy, accompanied by large computational efforts for simulation. Typically, such models are implemented as fine-grain meshes of the chip’s structures, which are then deployed for finite element/finite difference analysis.

For “medium abstract” design phases, for example, place and route, models are more abstract; they are tailored to represent the system behavior. Therefore, their scope is the

(multi-physical) coupling between separate components and the resulting system behavior. For example, an arrangement of multiple TSVs is modeled such that its geometry, the signal crosstalk, the thermomechanical stress, etc., are captured, in order to evaluate the overall reliability and performance of the arrangement. It is often required to derive such behavioral models from low-level simulations where, for example, separate wires, TSVs, and/or active gates are considered. These simulations are independent from the actual design process and can be conducted in advance by experienced engineers, providing their findings into parametric models or design rules. The related models are typically implemented as equivalent networks; this principle can be applied to different physical domains and is sufficiently accurate yet computationally not as demanding as finite element analysis.

The highest level of design abstraction deals with functional or architectural design. Simulation at this abstract level is difficult; the 3D chip and its components can only be modeled as design blocks with abstract properties like design area, number of pins, power dissipation, timing constraints, etc. These properties are furthermore often given as estimations with inherent variations. However, for densely integrated 3D chips, some parameter variations (e.g., in power dissipation) may have a large impact on the final design (e.g., on chip reliability and needs for heat removal). Besides, architectural design requires to analyze many different and diversified 3D-chip compositions in order to determine the appropriate one. This leads inevitably to many analysis iterations, demanding fast computation and simulation. These two opposing requirements—sufficiently high accuracy, also reflecting parameter variations and versatile 3D-chip compositions, and fast computation—make simulation on this design level very challenging. Applied models vary, depending on the required accuracy, available time, and the complexity of the 3D design and chip. In general, models have to be scalable to address these challenges. They are typically implemented as equivalent networks, coarse-grain finite elements, or dedicated models. For the latter, many studies have been proposed in recent years, which also reflects the need for such custom-tailored models. For example, Kim et al. [58] proposed a TSV-aware wirelength distribution model, capable of predicting delays and power consumption.

In summary, simulation and verification for 3D chips is challenging. The need to consider multi-physical coupling as well as the strong impact of technological configurations (like the number, size, and arrangement of separate dies) on each design phases are key issues. In general, hierarchical modeling and simulation frameworks are evolving as method of choice. Therein applied models have to be scalable. Further, the generation of parameterized models from low-level simulations is essential but not supported yet [116]. For verification, it seems practical to adapt available tools and leverage know-how from both classical 2D design and package design. For simulation, especially during high-level design phases, however, new approaches are required. In this context, Lim [79] reviewed key research needs for architectural floorplanning, to evaluate register-transfer-level (RTL)-based designs more accurately in terms of power, performance, and reliability. Lim argued that block-level modeling, TSV management, and chip/package coevaluation are crucial, and should be deployed as early and as effective as possible.

17.2 PROSPECTIVE DIRECTIONS FOR 3D CIRCUITS AND DESIGN

3D integration has been praised as a viable solution to keep up with the constantly increasing demands on electronic systems. The International Technology Roadmap for Semiconductors (ITRS) has prominently featured 3D-ICs for some years now: in the 2009 edition, for example, in the section on Interconnect and the section on Assembly and Packaging [1], or in the “More-than-Moore” whitepaper from 2010 [9].

Throughout these years, researchers and industry experts have been eager to cope with the many challenges arising from complex requirements for manufacturing and design of 3D-ICs, as

also discussed earlier. While challenges for mainstream commercialization still remain, they have mainly shifted from manufacturing to design infrastructure, as also confirmed by industry experts. For example, at the GSA 3D-IC Packaging Working Group meeting October 2014, Yazdani [136] argued that path-finding tools (Section 17.2.1) are much-needed to design and evaluate the 3D chip-package-board system.

In the following, we discuss prominent design and manufacturing approaches, which are considered to increase mainstream adaption of 3D-ICs.

17.2.1 PATH-FINDING: SYSTEM-LEVEL DESIGN EXPLORATION AND EVALUATION

Traditionally, high-level models of circuit components have been applied for evaluation of design options. As already discussed, this is much more complex for 3D-ICs than for classical 2D chips. With the ever increasing design complexity and the vast options for manufacturing and integration choices, system-level design of 3D chips cannot be conducted without considering physical-level details. Thus, it is necessary for system-level design tools to handle the complex interactions between performance, power, thermal management, process technology, floorplanning, system architecture, and even dynamic scheduling or workloads. Such extended system-level design exploration and evaluation is known as *path-finding*.

17.2.1.1 Concepts and Approaches for Path-Finding

Early concepts for path-finding in 3D design have already been proposed in 2009, for example, by Milojevic et al. [97]. Their main novelty was to link system-level design exploration with automated synthesis of RTL models and physical-design prototyping. This way, system engineers had been given the opportunity to evaluate their architectures on a much more detailed level, despite not necessarily being equipped with extensive know-how and time for actual physical design.

Research has resurged very recently, and several studies on practical path-finding methodologies have been presented. Martin et al. [95] proposed a methodology for early evaluation of electrical performance. In their study, they deployed building blocks (e.g., of large TSV arrays) using parameterized models. These blocks are then committed to fast electromagnetic solvers for analyzing signal crosstalk. With this methodology, the authors successfully evaluated interposer-based 3D devices and their TSV interconnects. A similar study was conducted by Yazdani and Park [137]; they showcased how to optimize system interconnects in 2.5D integration. More precisely, they conducted and evaluated the placement of buffer cells, arrangement of Cu pillar bumps and package BGA for Wide I/O memory integration on interposers. Thus, their tool enables planning of interconnect structures at early stages and for multiple chips integrated by state-of-the-art memory technology. Priyadarshi et al. [110] proposed transaction-level-based path-finding, complementing known RTL-based approaches. Their tool allows much faster design evaluation, since transaction-based modeling distinguishes computation and communication, thus hiding details not necessarily required for early design simulation. Additionally, they link thermal analysis to transactional modeling and simulation, thereby enabling efficient thermal-aware path-finding.

17.2.1.2 Flows and Tools for Path-Finding

A typical flow for path-finding tools covers three steps (Figure 17.10). Note that feedback loops between these steps are essential; capabilities for passing specifications top-down (e.g., physical constraints or technology details) as well as passing them bottom-up (e.g., simulation results) are needed for flexible and accurate path-finding.

1. *System-level design exploration*: A high-level description (e.g., given in SystemC) is generated. Already at this point, the technology and configuration for 3D integration have to be considered. For example, partitioning modules across separate dies can be modeled in these early phases, to help tackle the vast design space of 3D chips more efficiently.

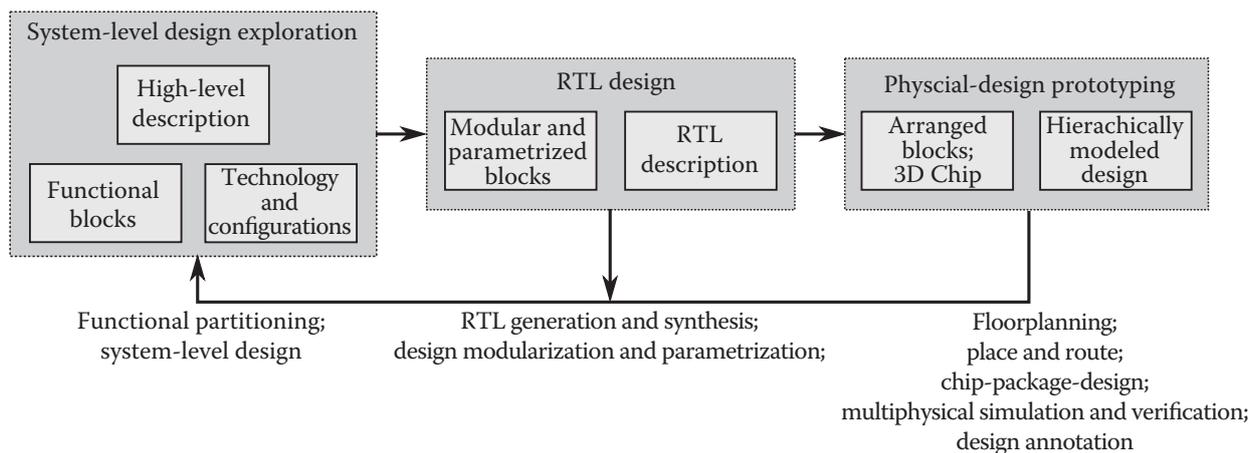


FIGURE 17.10 Path-finding flow, with each step’s components labeled in boxes and applied techniques labeled in the following. Note that details of technology (e.g., for a specific 3D integration approach) are to be considered and rendered more specifically for each step.

2. *RTL design*: From high-level descriptions, RTL models are derived. They are serving as “bridge” between system design and physical design. The models should be modularized, in order to represent the high-level design closely and to enable reuse for different arrangements of system components during path-finding. The models can furthermore be annotated during subsequent technology simulation, to provide technology feedback/guidance at this early phase. In short, a parametrized building-block model of the overall system is the objective of this step.
3. *Physical-design prototyping*: The RTL models are then fed to physical-design prototyping. In contrast to actual physical design, more abstract (and thus faster) techniques are applied to obtain estimates of the final design quality. For example, an important step of prototyping is floorplanning. Design blocks are usually annotated, for example, with power consumption and area, to enable more accurate estimates on, for example, thermal distribution and die sizes.

Besides the methodologies outlined in the previous subsection, commercial tools are becoming available. Note that such tools are usually modular and also rely on adapting legacy (2D) tools for simulation and verification.

17.2.2 DESIGN APPROACHES AND STANDARDIZATION

Due to the paradigm shift arising with 3D-ICs—the additional integration in the third dimension—physical-design automation cannot be considered as stand-alone process. In fact, all components of chip design (i.e., technology and manufacturing, system design, and physical-design automation) undergo a notable transition. This wide-ranging shift aggravates the need for reliable and effective design approaches and commonly established standards.

17.2.2.1 Design Approaches for 3D Circuits

Design approaches can be characterized by their *granularity*, that is, the applied partitioning scheme, defining which circuit parts are split and assigned to different dies [84]. On the opposite ends of the granularity scale, the approaches of transistor-level (finest-grain) integration versus core-level (coarsest-grain) integration can be found.

Only recently—mainly due to advances in monolithic manufacturing technologies—transistor-level integration becomes applicable [14,73,81,104]. Here, active layers are built up sequentially rather than processed in separate and subsequently bonded dies. This finest-grain integration style

is expected to provide large performance benefits due to shortest-path vertical coupling of transistors. Besides the high demands on very-small-scale vias and other related challenges, this style requires a full redesign, that is, completely prevents design reuse. It also faces further challenges, for example, the need for tools and knowhow for low-temperature manufacturing processes [13] or notably increased delays along with massive routing congestion [72,81].

For the other end of the integration scale, that is, for core-level integration, the efforts are comparable to traditional 2D chip design: only few intercore connects have to be realized by placing and wiring TSVs. Apart from that, the cores can be fully reused. In consequence, gained benefits are low: the properties of such a 3D-IC are still dominated by their stacked 2D chips.

Next, design approaches found in the middle of the granularity scale are reviewed: gate-level and block-level integration (Figure 17.11).

Gate-level integration means to partition cells across multiple dies and use TSVs whenever required for connecting cells across dies. This style promises significant wirelength reduction and great flexibility [84,100,102].

Its adverse effects include, for example, the massive number of necessary TSVs for random logic. Studies by Kim et al. [57] and Mak and Chu [91] reveal that partitioning gates between multiple dies can undermine wirelength reduction unless modules of certain minimal size are preserved and/or TSVs are downscaled. Another study [101] points out that layout effects can largely influence performance for highly regular blocks such as SRAM registers: a mismatch between TSV and cell dimensions introduces wirelength disparities while routing these regular structures to TSVs. Timing-aware placement of partitioned gates is required for design closure [70]; this timing issue is intensified by interdie variation mismatches [33]. Besides, partitioning a design block across multiple dies requires new prebond testing approaches [69,75]. After die stacking, a single failed die renders the whole 3D-IC unusable, thus easily undermining overall yield.

In summary, gate-level integration may be very promising in terms of design flexibility, performance, and wirelength reduction, but it faces many challenges and currently appears—like transistor-level integration—only applicable in a limited scope. Practical scenarios include devices with high demands on efficiency and low power, as demonstrated by, for example, 3D-ICs with complex modules like floating-point units and long-path multipliers [102,124,125].

Block-level integration promises to reduce TSV overhead by assigning only few global interconnects to them. This is possible since blocks typically subsume most of a design's connectivity and are linked by a small number of global interconnects [123].

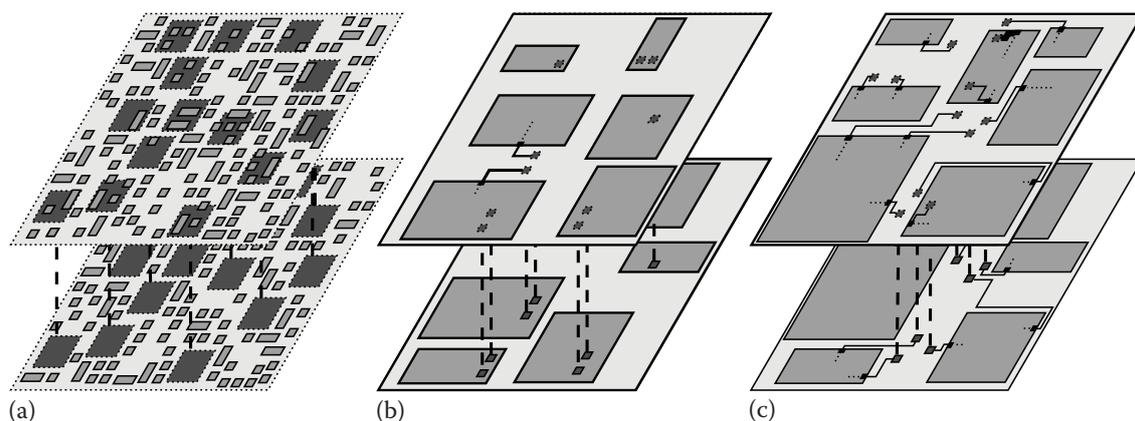


FIGURE 17.11 Design styles for 3D-ICs. TSVs are illustrated as dark-gray boxes and landing pads as dashed, dark-gray boxes. Gates or blocks are represented by light-gray boxes. Face-to-back stacking is considered; TSVs cannot obstruct blocks in lower dies but landing pads may overlap with blocks in upper dies, due to illustration perspective. (a) Gate-level integration, enlarged detail. (b and c) Block-level integration. (b) The *redesigned 2D* style uses predefined TSV sites within for 3D-design adapted blocks. (c) The *legacy 2D* style distributes TSVs between blocks, thus enabling reuse of available design blocks.

Sophisticated 3D systems combining many heterogeneous dies are anticipated in a whitepaper by Cadence [2]. Such devices require distinct manufacturing processes at different-technology nodes for fast and low-power random logic, several memory types, analog and RF circuits, on-chip sensors, microelectromechanical systems, and so on. Block-level integration is imperative for such heterogeneous 3D-ICs where modules cannot be partitioned among different-technology dies.

When assigning entire blocks to separate dies and connecting them with TSVs, we can distinguish two design styles.

- *Redesigned 2D (R2D) style*: 2D blocks designed for 3D integration; TSVs can potentially be embedded within the footprints (Figure 17.11b).
- *Legacy 2D (L2D) style*: 2D blocks not designed for 3D integration; TSVs are to be placed between blocks (Figure 17.11c).

Which style is appropriate also depends on the type of given *intellectual property (IP)* blocks. For hard blocks with predefined layout, applying L2D is mandatory. The fixed layout of such blocks cannot include large TSVs simply because the blocks' design was not accounting for TSVs. For soft blocks, that is, blocks given in behavioral description and synthesized during the design flow, the R2D style appears more appropriate but is still challenging in terms of TSV management, as elaborated next.

TSVs introduce design constraints and overheads, mainly due to their (to gates comparably large) dimensions and intrusive character when “injected” into silicon dies. Thus, inserting TSVs into densely packed design blocks is expected to complicate design closure since it (1) introduces placement and routing obstacles [57], (2) induces notable stress for nearby active gates [134], and (3) requires design tools to provide sophisticated TSV-related verification, for example, signal-integrity analysis considering coupling between TSVs [82,83,135].

Grouping TSVs into *TSV islands* is common practice and beneficial for several reasons [44,56,59,63,64,98,127,145]. For example, TSVs introduce stress in the surrounding silicon, which affects nearby transistors [10,47,134], but TSV islands do not need to include active gates. The layout of these islands can be optimized in advance [87,139]. Regular island structures help to limit stress below the yielding strength of copper [51], and limit stress generally to particular design regions [51,52,87]. Placing islands *between* blocks (i.e., applying the L2D style) may thus limit stress on blocks' active gates.

Further benefits of both R2D and L2D styles are described next.

- Design-for-Test (DfT) structures are key components of existing IP blocks and can be used to realize prebond and postbond testing [69].
- Block-level integration can efficiently reduce critical paths, thus simultaneously limiting signal delay, increasing performance and reducing power consumption [11,59,70,84].
- With block-level integration, critical paths are mostly located within 2D blocks—they do not traverse multiple active layers, which limits the impact of process variations on performance [34].
- For yield-optimized matching of “slow dies” and “fast dies,” based on accurate delay models with process variations considered [28], block-level integration is mandatory. That is because this matching approach assumes that dies can be delay-tested before stacking, which is only possible when all dies encapsulate self-contained modules.
- Modern chip design mostly relies on predesigned and optimized IP blocks. Existing IP blocks and physical-design automation tools do not account for 3D integration. Even when such tools appear, it will take IP vendors much time and money to upgrade their extensive portfolios for 3D integration. Thus, redesigning existing IP blocks to be spread out on multiple dies (as proposed in gate-level integration) is not practical; in contrast, reusing them as legacy blocks (as proposed in block-level integration) is convenient.

These wide-ranging considerations suggest that block-level integration is a more practical approach for general 3D-IC design; for dedicated applications, other styles may also be considered.

17.2.2.2 Standardization Trends and Examples

Despite the great interest and recent achievements in design and manufacturing of 3D-ICs, this technology is still not yet available in high-volume applications. Besides the aforementioned need for design, verification and test tools, another pressing concern is the lack of standard definitions. However, as for any successful technology in the chip industry, standards will be required for increasing acceptance and establishing supply chains and “ecosystems” [2].

Efforts for standardization initially focus on I/O and interfaces, while later on heterogeneous and/or interposer-based assembling and supply chains need to be addressed. For example for the latter, the JEDEC Multiple Chip Packages Committee is “currently developing mixed technology pad sequence and device package standards to enable SRAM, DRAM, and Flash memory to be combined into a single package that may also contain processor(s) and other devices” [49]. Standards already available and widely acknowledged in the industry cover memory integration and testing, briefly reviewed next.

For memory integration, the JESD229 standard [48], more commonly known as *Wide I/O*, is a prominent example. It defines memory integration with one up to four memory dies stacked on top of a controller die. The standard is considered mature; two versions are available, the first being published December 2011 and the second (*WideIO2*) August 2014. Devices fulfilling the standard provide high-bandwidth memory interfaces, namely four (up to eight for *WideIO2*) 128-bit-wide memory channels. The standard covers details on functionality, AC and DC characteristics, packages, and micropillar signal assignments. Several studies proposed designs based on *Wide I/O* and/or tools for related verification [40,53,98,121,137].

With JESD235 [49], better known as *High Bandwidth Memory (HBM)*, an alternative standard for 3D memory integration is available and currently adopted by industry, for example, by SK Hynix [120].

Yet another memory standard, the *Hybrid Memory Cube (HMC)*, has recently gained more attention. The related consortium was founded in October 2011 by co-developers Altera, Micron, Open-Silicon, Samsung, and Xilinx. The first specification was released in May 2013 [3], and the second version in December 2014 [7]. This update most notably increases data rates for single channels from 1920 up to 3840 MB/s.

For testing, standards are currently under development. As indicated in Section 17.1.1.4, IEEE standard P1838 [93] evolves as prominent example. The proposed wrapper architecture enables controllability and observability at the die boundaries, which ensures interoperability between different dies, possibly even from different manufacturers. Besides these wrapper components, existing test facilities are proposed for reuse whenever possible: IEEE 1149.x for test access, IEEE 1500 for die test, and IEEE P1687 for internal debugging. Some studies addressed the implementation and review of such standardized DfT structures [94,98].

17.2.3 DEMONSTRATORS AND PROTOTYPING FOR 3D CIRCUITS

3D integration has been eagerly discussed and investigated for many years now. There have been many efforts for demonstrators and prototyping of TSV-based 3D-ICs and interposer-based 2.5D systems, both from academic and industrial groups. The following gives a brief overview on demonstrators and outlines prototyping platforms.

17.2.3.1 Academic Efforts

The project *3D-MAPS: A Many-Core 3D Processor with Stacked Memory* [37,56,80] is a prominent example for large-scale demonstrators driven by academia (at Georgia Institute of Technology). In 2010, the first version was developed: a 64-core memory-stacked-on-processor system running at

277 MHz with 64 GB/s memory bandwidth. That bandwidth was confirmed in measurements of taped-out chips; up to 63.8 GB/s was achieved while overall power consumption was approx. only 4 W. The processor was fabricated in 130 nm by GLOBALFOUNDRIES and the TSV technology (1.2 μm size, 2.5 μm pitch, and 6 μm height) was provided by Tezzaron Semiconductors. Note that TSVs had been deployed there only for power delivery and external interconnects. In 2012, a second version was proposed with 128 cores, embedded in two logic dies, and stacked with three DRAM dies.

Another sophisticated academic 3D demonstrator was developed in 2012 at the University of Michigan: *Centip3De*, a large-scale 3D chip with a cluster-based near-threshold computing architecture [29]. The chip comprises a logic die (130 nm) with 64 ARM Cortex-M3 cores, and an SRAM die; both are interconnected via face-to-face bonding. Again, TSVs are deployed from Tezzaron's technology and are used for connecting the chip to the package. A notable result of this demonstrator is a $>3\times$ improvement in energy efficiency (measured in DMIPS/W) over traditional chips and operation. Furthermore, the chip can fully operate under a fixed thermal design power of only 250 mW.

Besides these large "flagship" demonstrators, further academic studies have proposed tools, designs, and some also measurements from taped-out 3D chips, for example, [124,141].

A prototyping platform projected for 3D chips was proposed in *FlexTiles: Self Adaptive Heterogeneous many core based on Flexible Tiles* [6,16]. Here, the (so far only conceptual) 3D chip comprises a FPGA die and a many-core die. The FPGA die provides dynamic reconfigurability at runtime, that is, dynamic adaption of system functionality by reloading and/or reconfiguring IP blocks. The many-core die shall comprise general processing cores and dedicated DSPs. The intended scope of the prototyping platform is to evaluate adaptive and heterogeneous designs with state-of-the-art technology. For example, an (so far on two FPGAs instead of the 3D chip) evaluated use case is a "smart" camera, which dynamically adapts for low-power scenarios. Objectives of the project are the definition and development of a heterogeneous many-core with self-adaptation capabilities, its virtualization layer and its tool chain ensuring programming efficiency and low power consumption.

17.2.3.2 Industrial Efforts

Already in 2010, Xilinx has presented the *Virtex-7 FPGA* family [27]. Here, the FPGA is split into four dies (manufactured in 28 nm), which are assembled side-by-side onto a passive silicon interposer (manufactured in 65 nm including TSVs). With this 2.5D approach, cost are limited, while yield and performance are increased compared to previous high-end FPGA systems. For example, the bandwidth-per-watt ratio is over $100\times$ that of standard FPGA interconnects.

Intel presented an energy-efficient, high-performance 80-core system with stacked SRAM in 2011 [15]. The demonstrator provides one tera-FLOPS while consuming less than 100W. At that time, TSVs were not necessarily very reliable, so they had been sparsely (pitch 190 μm) deployed for power delivery and external routing to the package. The SRAM die and the 80-core logic die had been interconnected via face-to-face metal bonding.

IBM demonstrated a 3D version of a processor with up to three dies of eDRAM in 2012 [132]. With deployment of known-good dies, it uses 50 μm pitch of μC4 bumps to join the front side of the processor to the TSV connections on the back side of the thinned memory chips. The demonstrator is based on 45 nm technology and runs at 2 GHz. It achieves a data bandwidth of approx. 56 GB/s.

For 3D memory integration, the industry has passed the prototyping stage, and is recently approaching high-volume manufacturing. For example, since end of 2014, SK Hynix offers HBM modules [120]. With 128 GB/s, these modules provide approx. $4.5\times$ the bandwidth of state-of-the-art GDDR5 modules. Alok Gupta, PE at Nvidia, presented at the 3D ASIP 2014 conference [5] a GPU-on-interposer system comprising four HBM modules and achieving a bandwidth of 1 TB/s. Another example are the efforts of Samsung: since 2013, the company is mass-producing so-called vertical-NAND (V-NAND) memory [114]. Here, as the name suggests, the transistors are vertically arranged, that is, the gate and insulator are circularly wrapped around the channel. These transistors are then repeatedly processed onto many stacked layers. With this dedicated design, achieving

integration density in the vertical dimension instead of traditionally in the plane, wider bit lines and the deployment of “old” but much more reliable process nodes (e.g., 30 nm) are feasible. These two measures effectively reduce cell-to-cell interferences and small-scale patterning issues, which are major concerns for modern memory technology.

Besides the mentioned Virtex-7 FPGAs from Xilinx, further interposer-based prototypes haven been presented by GLOBALFOUNDRIES. In cooperation with Open-Silicon, a prototype containing two ARM Cortex A-9 chips (28 nm) stacked onto a 65 nm-and-TSV-embedded silicon interposer was presented in 2013 [4]. This demonstrator was mainly intended as proof-of-concept and also to establish EDA flows for design, verification, and test of interposer systems. GLOBALFOUNDRIES’ Packaging Director Alapati stated the transition for interposer-based systems as well as for 3D-ICs to high-volume manufacturing in 2015 [130]. Further interposer-based prototypes comprising 20 and 14 nm chips have been presented as well.

17.3 SUMMARY AND CONCLUSION

In comparison to classical 2D chips, 3D-ICs and even interposer-based 2.5D systems are much more complex. As discussed in this chapter (and throughout the book), both design and manufacturing engineers have to cope with quite different and challenging requirements and objectives.

In recent years, much research and development effort, from academia and industry, has been undertaken. Slowly but surely 3D integration is making the transition from a “hyped new technology” toward a viable option for keeping up with constantly increasing demands on performance, functionality, power consumption, and cost of electronic systems. Very recently, few companies (e.g., Samsung and SK Hynix) have introduced 3D-integrated memory products on the market, while other companies (e.g., GLOBALFOUNDRIES) have established tool-chains and design flows to enable high-volume manufacturing of interposer-based systems and even 3D-ICs in very near future.

This implies that technology and manufacturing concerns have been mainly addressed, at least for TSV- and interposer-based systems. It is common consensus that no “show-stoppers” are blocking the adoption of such 3D integration. However, for large-scale heterogeneous integration and especially for logic-on-logic integration, key concerns remain. For example, thermal management, power and clock delivery, testing along with yield and cost are still obstructing such sophisticated 3D systems.

In the recent years, design challenges have largely shifted toward high-level design issues. Besides the fact that 3D-EDA tools are only slowly reaching the market, high-level design features are yet insufficiently supported. Such sought-after features include: multi-physical simulation and verification of the chip-package-board system, including different types of active layers and interconnects; path-finding for efficient design exploration and evaluation; standards-based design of modules and interfaces, for example, NoCs or test structures.

Overall, 3D integration is nevertheless on an “accelerating trajectory,” and the next few years will bring this integration approach with its versatile options more and more into mainstream chip development.

REFERENCES

1. International technology roadmap for semiconductor. <http://www.itrs.net/Links/2009ITRS/Home2009.htm>, 2009. Accessed on December 2014.
2. 3D ICs with TSVs—Design challenges and requirements. http://www.cadence.com/rl/Resources/white_papers/3DIC_wp.pdf, 2010. Accessed on December 2014.
3. Hybrid Memory Cube Specification 1.0. http://hybridmemorycube.org/files/SiteDownloads/HMC_Specification%201_0.pdf, January 2013. Accessed on December 2014.
4. Open-Silicon and GLOBALFOUNDRIES Demonstrate Custom 28 nm SoC Using 2.5D Technology. <http://www.open-silicon.com>, Nov 2013. Accessed on December 2014.
5. 3D ASIP 2014: All Aboard the 3DIC Train!. <http://www.3dincites.com/2014/12/3d-asip-2014-addresses-3d-benefits-challenges-solutions/>, 2014. Accessed on December 2014.

6. Flextiles: Self adaptive heterogeneous manycore based on flexible tiles. <http://flextiles.eu>, 2014. Accessed on December 2014.
7. Hybrid Memory Cube Specification 2.0. http://hybridmemorycube.org/files/SiteDownloads/20141119_HMCC_Spec2.0Release.pdf, February 2014. Accessed on December 2014.
8. M. Agrawal, K. Chakrabarty, and B. Eklow. Test-cost optimization and test-flow selection for 3D-stacked ICs. Technical report, Electrical and Computer Engineering, Duke University, Durham, NC, 2012.
9. W. Arden, M. Brillout, P. Coge, M. Graef, B. Huizing et al. More-than-Moore white paper. Technical report, ITRS, 2010.
10. K. Athikulwongse, A. Chakraborty, J.-S. Yang, D. Z. Pan, and S. K. Lim. Stress-driven 3D-IC placement with TSV keep-out zone and regularity study. In *Proc. Int. Conf. Comput.-Aided Des.*, San Jose, CA, pp. 669–674, 2010.
11. K. Athikulwongse, D. H. Kim, M. Jung, and S. K. Lim. Block-level designs of die-to-wafer bonded 3D ICs and their design quality tradeoff's. In *Proc Asia South Pacific Des. Autom. Conf.*, Yokohama, Japan, pp. 687–692, 2013.
12. K. Athikulwongse, M. Pathak, and S. K. Lim. Exploiting die-to-die thermal coupling in 3D IC placement. In *Proc. Des. Autom. Conf.*, San Francisco, CA, pp. 741–746, 2012.
13. P. Batude, M. Vinet, B. Previtali, C. Tabone, C. Xu et al. Advances, challenges and opportunities in 3D CMOS sequential integration. In *Proc. Int. Elec. Devices Meeting*, Washington, DC, pp. 7.3.1–7.3.4, 2011.
14. S. Bobba, A. Chakraborty, O. Thomas, P. Batude, T. Ernst et al. Celoncel: Effective design technique for 3-D monolithic integration targeting high performance integrated circuits. In *Proc. Asia South Pacific Des. Autom. Conf.*, Yokohama, Japan, pp. 336–343, 2011.
15. S. Borkar. 3D integration for energy efficient system design. In *Proc. Des. Autom. Conf.*, San Diego, CA, pp. 214–219, 2011.
16. R. Brillu, S. Pillement, A. Abdellah, F. Lemonnier, and P. Millet. Flex-Tiles: A globally homogeneous but locally heterogeneous manycore architecture. In *Proc. Workshop on Rapid Sim. and Perform. Evaluation*, Vienna, Austria, pp. 3:1–3:8, 2014.
17. H. H. Chan, S. N. Adya, and I. L. Markov. Are floorplan representations important in digital design? In *Proc. Int. Symp. Phys. Des.*, San Francisco, CA, pp. 129–136, 2005.
18. Y. S. Chan, H. Y. Li, and X. Zhang. Thermo-mechanical design rules for the fabrication of TSV interposers. *Trans. Compon. Packag. Manuf. Technol.*, 3(4):633–640, 2013.
19. H.-T. Chen, H.-L. Lin, Z.-C. Wang, and T. T. Hwang. A new architecture for power network in 3D IC. In *Proc. Des. Autom. Test Europe*, Grenoble, France, pp. 1–6, 2011.
20. Y. Chen, E. Kursun, D. Motschman, C. Johnson, and Y. Xie. Through silicon via aware design planning for thermally efficient 3-D integrated circuits. *Trans. Comput.-Aided Des. Integr. Circ. Syst.*, 32(9):1335–1346, 2013.
21. Y.-X. Chen, Y.-J. Huang, and J.-F. Li. Test cost optimization technique for the pre-bond test of 3D ICs. In *VLSI Test Symposium*, Maui, HI, pp. 102–107, 2012.
22. L. Cheng, L. Deng, and M. D. F. Wong. Floorplanning for 3-D VLSI design. In *Proc. Asia South Pacific Des. Autom. Conf.*, Shanghai, China, pp. 405–411, 2005.
23. J. Cong, G. Luo, and Y. Shi. Thermal-aware cell and through-silicon-via co-placement for 3D ICs. In *Proc. Des. Autom. Conf.*, pp. 670–675, 2011.
24. J. Cong, J. Wei, and Y. Zhang. A thermal-driven floorplanning algorithm for 3D ICs. In *Proc. Int. Conf. Comput.-Aided Des.*, pp. 306–313, 2004.
25. J. Cong and Y. Zhang. Thermal-driven multilevel routing for 3-D ICs. In *Proc. Asia South Pacific Des. Autom. Conf.*, Shanghai, China, pp. 121–126, 2005.
26. S. Deutsch, K. Chakrabarty, S. Panth, and S. K. Lim. TSV stress-aware ATPG for 3D stacked ICs. In *Proc Asian Test Symp.*, pp. 31–36, 2012.
27. P. Dorsey. Xilinx stacked silicon interconnect technology delivers break-through FPGA capacity, bandwidth, and power efficiency. Technical report, Xilinx, Inc., 2010.
28. C. Ferri, S. Reda, and R. I. Bahar. Strategies for improving the parametric yield and profits of 3D ICs. In *Proc. Int. Conf. Comput.-Aided Des.*, pp. 220–226, 2007.
29. D. Fick, R. G. Dreslinski, B. Giridhar, G. Kim, S. Seo et al. Centip3De: A cluster-based NTC architecture with 64 ARM Cortex-M3 cores in 3D stacked 130 nm CMOS. *J. Solid-State Circ.*, 48(1):104–117, 2013.
30. R. Fischbach, J. Lienig, and J. Knechtel. Investigating modern layout representations for improved 3D design automation. In *Proc. Great Lakes Symp. VLSI*, pp. 337–342, 2011.

31. R. Fischbach, J. Lienig, and T. Meister. From 3D circuit technologies and data structures to interconnect prediction. In *Proc. Int. Workshop Syst.-Level Interconn. Pred.*, pp. 77–84, 2009.
32. R. Fischbach, J. Lienig, and M. Thiele. Solution space investigation and comparison of modern data structures for heterogeneous 3D designs. In *Proc. 3D Syst. Integr. Conf.*, pp. 1–8, 2010.
33. S. Garg and D. Marculescu. 3D-GCP: An analytical model for the impact of process variations on the critical path delay distribution of 3D ICs. In *Proc. Int. Symp. Quality Elec. Des.*, pp. 147–155, 2009.
34. S. Garg and D. Marculescu. Mitigating the impact of process variation on the performance of 3-D integrated circuits. *Trans. VLSI Syst.*, 21(10):1903–1914, 2013.
35. B. Goplen and S. Sapatnekar. Thermal via placement in 3D ICs. In *Proc. Int. Symp. Phys. Des.*, pp. 167–174, 2005.
36. B. Goplen and S. Sapatnekar. Placement of 3D ICs with thermal and interlayer via considerations. In *Proc. Des. Autom. Conf.*, pp. 626–631, 2007.
37. M. B. Healy, K. Athikulwongse, R. Goel, M. M. Hossain, D. H. Kim et al. Design and analysis of 3D-MAPS: A many-core 3D processor with stacked memory. In *Proc. Cust. Integr. Circ. Conf.*, pp. 1–4, 2010.
38. M. B. Healy and S. K. Lim. Power delivery system architecture for many-tier 3D systems. In *Proc. Elec. Compon. Technol. Conf.*, pp. 1682–1688, 2010.
39. M. B. Healy and S. K. Lim. Power-supply-network design in 3D integrated systems. In *Proc. Int. Symp. Quality Elec. Des.*, pp. 223–228, 2011.
40. A. Heinig, R. Fischbach, and M. Dittrich. Thermal analysis and optimization of 2.5D and 3D integrated systems with Wide I/O memory. In *Proc. Therm. Thermomech. Phenom. Electr. Syst. Conf.*, pp. 86–91, 2014.
41. A.-C. Hsieh and T. T. Hwang. TSV redundancy: Architecture and design issues in 3-D IC. *Trans. VLSI Syst.*, 20(4):711–722, 2012.
42. G. Huang, M. Bakir, A. Naeemi, H. Chen, and J. D. Meindl. Power delivery for 3D chip stacks: Physical modeling and design implication. In *Proc. Electri. Perf Elec. Packag. Sys.*, pp. 205–208, 2007.
43. L.-R. Huang, S.-Y. Huang, K.-H. Tsai, and W.-T. Cheng. Parametric fault testing and performance characterization of post-bond interposer wires in 2.5-D ICs. *Trans. Comput.-Aided Des. Integr. Circ. Syst.*, 33(3):476–488, 2014.
44. Y.-J. Huang and J.-F. Li. Built-in self-repair scheme for the TSVs in 3-D ICs. *Trans. Comput.-Aided Des. Integr. Circ. Syst.*, 31(10):1600–1613, 2012.
45. W.-L. Hung, G. M. Link, Y. Xie, N. Vijaykrishnan, and M. J. Irwin. Interconnect and thermal-aware floorplanning for 3D microprocessors. In *Proc. Int. Symp. Quality Elec. Des.*, pp. 98–104, 2006.
46. P. Jain, T.-H. Kim, J. Keane, and C. H. Kim. A multi-story power delivery technique for 3D integrated circuits. In *Proc. Int. Symp. Low Power Elec. Design*, pp. 57–62, 2008.
47. H. Jao, Y. Y. Lin, W. Liao, B. Wu, B. Huang et al. The impact of through silicon via proximity on CMOS device. In *Proc. Microsys. Packag. Assemb. Circ. Technol. Conf.*, pp. 43–45, 2012.
48. JEDEC Solid State Technology Association. JEDEC Standard: JESD229 Wide I/O. <http://www.jedec.org/standards-documents/results/jesd229>, December 2011. Accessed on December 2014.
49. JEDEC Solid State Technology Association. JEDEC: 3D-ICs. <http://www.jedec.org/category/technology-focus-area/3d-ics-0>, December 2014. Accessed on December 2014.
50. M. Jung and S. K. Lim. A study of IR-drop noise issues in 3D ICs with through-silicon-vias. In *Proc. 3D Sys. Integr. Conf.*, pp. 1–7, 2010.
51. M. Jung, J. Mitra, D. Z. Pan, and S. K. Lim. TSV stress-aware full-chip mechanical reliability analysis and optimization for 3D IC. In *Proc. Des. Autom. Conf.*, 2011.
52. M. Jung, J. Mitra, D. Z. Pan, and S. K. Lim. TSV stress-aware full-chip mechanical reliability analysis and optimization for 3-D IC. *Trans. Comput.-Aided Des. Integr. Circ. Syst.*, 31(8):1194–1207, 2012.
53. M. Jung, D. Z. Pan, and S. K. Lim. Chip/package mechanical stress impact on 3-D IC reliability and mobility variations. *Trans. Comput.-Aided Des. Integr. Circ. Syst.*, 32(11):1694–1707, 2013.
54. A. B. Kahng, J. Lienig, I. L. Markov, and J. Hu. *VLSI Physical Design: From Graph Partitioning to Timing Closure*. Springer, 2011.
55. U. Kang, H.-J. Chung, S. Heo, S.-H. Ahn, H. Lee et al. 8Gb 3D DDR3 DRAM using through-silicon-via technology. In *Proc. Int. Solid-State Circ. Conf.*, pp. 130–131, 131a, 2009.
56. D. H. Kim, K. Athikulwongse, M. B. Healy, M. M. Hossain, M. Jung et al. 3D-MAPS: 3D massively parallel processor with stacked memory. In *Proc. Int. Solid-State Circ. Conf.*, pp. 188–190, 2012.
57. D. H. Kim, S. Mukhopadhyay, and S. K. Lim. Through-silicon-via aware interconnect prediction and optimization for 3D stacked ICs. In *Proc. Int. Workshop Sys.-Level Interconn. Pred.*, pp. 85–92, 2009.

58. D. H. Kim, S. Mukhopadhyay, and S. K. Lim. TSV-aware interconnect distribution models for prediction of delay and power consumption of 3-D stacked ICs. *Trans. Comput.-Aided Des. Integr. Circ. Syst.*, 33(9):1384–1395, 2014.
59. D. H. Kim, R. O. Topaloglu, and S. K. Lim. Block-level 3D IC design with through-silicon-via planning. In *Proc. Asia South Pacific Des. Autom. Conf.*, Sydney, Australia, pp. 335–340, 2012.
60. K. Kim, J. S. Pak, H. Lee, and J. Kim. Effects of on-chip decoupling capacitors and silicon substrate on power distribution networks in TSV-based 3D-ICs. In *Proc. Elec. Compon. Technol. Conf.*, pp. 690–697, 2012.
61. T.-Y. Kim and T. Kim. Clock tree synthesis for TSV-based 3D IC Designs. *Trans. Des. Autom. Elec. Syst.*, 16(4):48:1-48:21, 2011.
62. T.-Y. Kim and T. Kim. Clock tree embedding for 3D ICS. In *Design Automation Conference (ASP-DAC), 2010 15th Asia and South Pacific*, Taiwan, pp. 486–491, January 2010.
63. J. Knechtel. Interconnect Planning for Physical Design of 3D Integrated, Circuits, volume 445 of Fortschritt-Berichte VDI. VDI-Verlag, Düsseldorf, Germany, 2014.
64. J. Knechtel, I. L. Markov, and J. Lienig. Assembling 2-D blocks into 3-D chips. *Trans. Comput.-Aided Des. Integr. Circ. Syst.*, 31(2):228–241, 2012.
65. J. Knechtel, I. L. Markov, J. Lienig, and M. Thiele. Multiobjective optimization of deadspace, a critical resource for 3D-IC integration. In *Proc. Int. Conf. Comput.-Aided Des.*, pp. 705–712, 2012.
66. J. Knechtel, E. F. Y. Young, and J. Lienig. Structural planning of 3D-IC interconnects by block alignment. In *Proc. Asia South Pacific Des. Autom. Conf.*, Singapore, pp. 53–60, 2014.
67. J. H. Lau. TSV interposer: The most cost-effective integrator for 3D IC integration. In *SEMATECH Symposium*, Taiwan, 2011.
68. J. H. Law, E. F. Y. Young, and R. L. S. Ching. Block alignment in 3D floorplan using layered TCG. In *Proc. Great Lakes Symp. VLSI*, pp. 376–380, 2006.
69. H.-H. S. Lee and K. Chakrabarty. Test challenges for 3D integrated circuits. *Des. Test Comput.*, 26(5):26–35, 2009.
70. Y.-J. Lee and S. K. Lim. Timing analysis and optimization for 3D stacked multi-core microprocessors. In *Proc. 3D Sys. Integr. Conf.*, pp. 1–7, 2010.
71. Y.-J. Lee and S. K. Lim. Co-optimization and analysis of signal, power, and thermal interconnects in 3-D ICs. *Trans. Comput.-Aided Des. Integr. Circ. Syst.*, 30(11):1635–1648, 2011.
72. Y.-J. Lee and S. K. Lim. Ultrahigh density logic designs using monolithic 3-D integration. *Trans. Comput.-Aided Des. Integr. Circ. Syst.*, 32(12):1892–1905, 2013.
73. Y.-J. Lee, P. M., and S. K. Lim. Ultra high density logic designs using transistor-level monolithic 3D integration. In *Proc. Int. Conf. Comput.-Aided Des.*, pp. 539–546, 2012.
74. D. L. Lewis and H.-H. S. Lee. A scan-island based design enabling prebond testability in die-stacked microprocessors. In *Int. Test Conf.*, pp. 1–8, 2007.
75. D. L. Lewis and H.-H. S. Lee. Test strategies for 3D die stacked integrated circuits. In *Proc. Des. Autom. Test Europe 3D Workshop*, 2009.
76. Z. Li, Y. Ma, Q. Zhou, Y. Cai, Y. Xie et al. Thermal-aware P/G TSV planning for IR drop reduction in 3D ICs. *Integration*, 46(1):1–9, 2013.
77. J. Lienig. Electromigration and its impact on physical design in future technologies. In *Proc. Int. Symp. Phys. Des.*, pp. 33–40, 2013.
78. J. Lienig and M. Dietrich, editors. *Entwurf integrierter 3D-Systeme der Elektronik*. Springer, Berlin and Heidelberg, Germany, 2012.
79. S. K. Lim. Research needs for TSV-based 3D IC architectural floorplanning. *J. Inf. Commun. Converg. Eng.*, 12(1):46–52, 2014.
80. S. K. Lim, H.-H. Lee, and G. Loh. The 3D-MAPS processors. <http://www.gtcad.gatech.edu/3d-maps/index.html>, 2010. Accessed on December 2014
81. C. Liu and S. K. Lim. A design tradeoff study with monolithic 3D integration. In *Proc. Int. Symp. Quality Elec. Des.*, 2012.
82. C. Liu, T. Song, J. Cho, J. Kim, J. Kim et al. Full-chip TSV-to-TSV coupling analysis and optimization in 3D IC. In *Proc. Des. Autom. Conf.*, 2011.
83. C. Liu, T. Song, and S. K. Lim. Signal integrity analysis and optimization for 3D ICs. In *Proc. Int. Symp. Quality Elec. Des.*, pp. 42–49, 2011.
84. G. H. Loh, Y. Xie, and B. Black. Processor design in 3D die-stacking technologies. *Micro*, 27:31–48, 2007.
85. I. Loi, F. Angiolini, S. Fujita, S. Mitra, and L. Benini. Characterization and implementation of fault-tolerant vertical links for 3-D networks-on-chip. *Trans. Comput.-Aided Des. Integr. Circ. Syst.*, 30(1):124–134, 2011.

86. I. Loi, S. Mitra, T. H. Lee, S. Fujita, and L. Benini. A low-overhead fault tolerance scheme for TSV-based 3D network on chip links. In *Proc. Int. Conf. Comput.-Aided Des.*, pp. 598–602, 2008.
87. K. H. Lu, X. Zhang, S.-K. Ryu, J. Im, R. Huang et al. Thermo-mechanical reliability of 3-D ICs containing through silicon vias. In *Proc. Elec. Compon. Technol. Conf.*, pp. 630–634, 2009.
88. C.-L. Lung, J.-H. Chien, Y. Shi, and S.-C. Chang. TSV fault-tolerant mechanisms with application to 3D clock networks. In *Int. SoC Des. Conf.*, pp. 127–130, 2011.
89. C.-L. Lung, Y.-S. Su, H.-H. Huang, Y. Shi, and S.-C. Chang. Through-silicon via fault-tolerant clock networks for 3-D ICS. *Trans. Comput.-Aided Des. Integr. Circ. Syst.*, 32(7):1100–1109, 2013.
90. G. Luo, Y. Shi, and J. Cong. An analytical placement framework for 3-D ICs and its extension on thermal awareness. *Trans. Comput.-Aided Des. Integr. Circ. Syst.*, 32(4):510–523, 2013.
91. W.-K. Mak and C. Chu. Rethinking the wirelength benefit of 3-D integration. *Trans. VLSI Syst.*, 20(12):2346–2351, 2012.
92. E. J. Marinissen. Challenges and emerging solutions in testing TSV-based 2 1/2D- and 3D-stacked ICs. In *Proc. Des. Autom. Test Europe*, pp. 1277–1282, 2012.
93. E. J. Marinissen. Status update of IEEE Std P1838. In *Proc. Int. Workshop Testing 3D Stack. Integr. Circ.*, 2014.
94. E. J. Marinissen, C.-C. Chi, J. Verbree, and M. Konijnenburg. 3D DfT architecture for pre-bond and post-bond testing. In *Proc. 3D Sys. Integr. Conf.*, pp. 1–8, 2010.
95. B. Martin, K. Han, and M. Swaminathan. A path finding based SI design methodology for 3D integration. In *Proc. Elec. Compon. Technol. Conf.*, 2014.
96. A. Mercha, G. Van der Plas, V. Moroz, I. De Wolf, P. Asimakopou-los et al. Comprehensive analysis of the impact of single and arrays of through silicon vias induced stress on high-k/metal gate CMOS performance. In *Proc. Int. Elec. Devices Meeting*, pp. 2.2.1–2.2.4, 2010.
97. D. Milojevic, T. E. Carlson, K. Croes, R. Radojicic, D. F. Ragett et al. Automated pathfinding tool chain for 3D-stacked integrated circuits: Practical case study. In *Proc. 3D Sys. Integr. Conf.*, pp. 1–6, 2009.
98. D. Milojevic, P. Marchal, E. J. Marinissen, G. Van der Plas, D. Verkest et al. Design issues in heterogeneous 3D/2.5D integration. In *Proc. Asia South Pacific Des. Autom. Conf.*, Yokohama, Japan, pp. 403–410, 2013.
99. J. Minz, X. Zhao, and S. K. Lim. Buffered clock tree synthesis for 3D ICs under thermal variations. In *Proc. Asia South Pacific Des. Autom. Conf.*, Seoul, Korea, pp. 504–509, 2008.
100. R. K. Nain and M. Chrzanowska-Jeske. Fast placement-aware 3-D floor-planning using vertical constraints on sequence pairs. *Trans. VLSI Syst.*, 19(9):1667–1680, 2011.
101. V. S. Nandakumar and M. Marek-Sadowska. Layout effects in fine-grain 3-D integrated regular micro-processor blocks. In *Proc. Des. Autom. Conf.*, pp. 639–644, 2011.
102. G. Neela and J. Draper. Logic-on-logic partitioning techniques for 3-dimensional integrated circuits. In *Proc. Int. Symp. Circ. Syst.*, pp. 789–792, 2013.
103. B. Noia, K. Chakrabarty, S. K. Goel, E. J. Marinissen, and J. Verbree. Test-architecture optimization and test scheduling for TSV-based 3-D stacked ICs. *Trans. Comput.-Aided Des. Integr. Circ. Syst.*, 30(11):1705–1718, 2011.
104. S. Panth, K. Samadi, Y. Du, and S. K. Lim. High-density integration of functional modules using monolithic 3D-IC technology. In *Proc. Asia South Pacific Des. Autom. Conf.*, Yokohama, Japan, pp. 681–686, 2013.
105. J.-H. Park, A. Shakouri, and S.-M. Kang. Fast thermal analysis of vertically integrated circuits (3-D ICs) using power blurring method. In *Proc. ASME InterPACK*, pp. 701–707, 2009.
106. M. Pathak and S. K. Lim. Performance and thermal-aware steiner routing for 3-D stacked ICs. *Trans. Comput.-Aided Des. Integr. Circ. Syst.*, 28(9):1373–1386, 2009.
107. M. Pathak, J. Pak, D. Z. Pan, and S. K. Lim. Electromigration modeling and full-chip reliability analysis for BEOL interconnect in TSV-based 3D ICs. In *Proc. Int. Conf. Comput.-Aided Des.*, pp. 555–562, 2011.
108. V. F. Pavlidis and E. G. Friedman. *Three-Dimensional Integrated Circuit Design*. Morgan Kaufmann Publishers Inc., Burlington, MA, 2008.
109. V. F. Pavlidis, I. Savidis, and E. G. Friedman. Clock distribution networks in 3-D integrated systems. *Trans. VLSI Syst.*, 19(12):2256–2266, 2011.
110. S. Priyadarshi, W. R. Davis, M. B. Steer, and P. D. Franzon. Thermal pathfinding for 3-D ICs. *Trans. Compon, Packag, Manuf. Technol.*, 4(7):1159–1168, 2014.
111. A. Quiring, M. Olbrich, and E. Barke. Improving 3D-floorplanning using smart selection operations in meta-heuristic optimization. In *Proc. 3D Syst. Integr. Conf.*, pp. 1–6, 2013.
112. J. Rajski and J. Tyszer. Fault diagnosis of TSV-based interconnects in 3-D stacked designs. In *Int. Test Conf.*, pp. 1–9, 2013.

113. S. K. Samal, S. Panth, K. Samadi, M. Saedi, Y. Du et al. Fast and accurate thermal modeling and optimization for monolithic 3D ICs. In *Proc. Des. Autom. Conf.*, 2014.
114. Samsung. 3D vertical-NAND memory. <http://www.samsung.com/global/business/semiconductor/html/product/flash-solution/vnand/overview.html>, December 2014. Accessed on December 2014.
115. M. Scandiuozzo, S. Cani, L. Perugini, S. Spolzino, R. Canegallo et al. Input/output pad for direct contact and contactless testing. In *Proc. Europ. Test Symp.*, pp. 135–140, 2011.
116. P. Schneider, A. Heinig, R. Fischbach, J. Lienig, S. Reitz et al. Integration of multi physics modeling of 3D stacks into modern 3D data structures. In *Proc. 3D Syst. Integr. Conf.*, pp. 1–6, 2010.
117. D. Sekar, C. King, B. Dang, T. Spencer, H. Thacker et al. A 3D-IC technology with integrated micro-channel cooling. In *Proc. Int. Interconn. Technol. Conf.*, pp. 13–15, 2008.
118. Y. Shang, C. Zhang, H. Yu, C. S. Tan, X. Zhao, and S. K. Lim. Thermal-reliable 3D clock-tree synthesis considering nonlinear electrical-thermal-coupled TSV model. In *Design Automation Conference (ASP-DAC), 2013 18th Asia and South Pacific*, Yokohama, Japan, pp. 693–698, January 2013.
119. R. S. Shelar and M. Patyra. Impact of local interconnects on timing and power in a high performance microprocessor. *Trans. Comput.-Aided Des. Integr. Circ. Syst.*, 32(10):1623–1627, 2013.
120. SK Hynix Inc. SK Hynix HBM Graphics Memory. http://www.skhynix.com/inc/pdfDownload.jsp?path=/datasheet/Databook/Databook_Q4'2014_Graphics.pdf, December 2014. Accessed on December 2014.
121. K. Smith, P. Hanaway, M. Jolley, R. Gleason, E. Strid et al. Evaluation of TSV and micro-bump probing for wide I/O testing. In *Proc. Int. Test Conf.*, pp. 1–10, 2011.
122. J. Sun, J.-Q. Lu, D. Giuliano, T. P. Chow, and R. J. Gutmann. 3D power delivery for microprocessors and high-performance ASICs. In *Proc. Appl. Power Electr. Conf.*, pp. 127–133, 2007.
123. D. Sylvester and K. Keutzer. A global wiring paradigm for deep submicron design. *Trans. Comput.-Aided Des. Integr. Circ. Syst.*, 19(2):242–252, 2000.
124. T. Thorolfsson, S. Lipa, and P. D. Franzon. A 10.35 mw/gflop stacked SAR DSP unit using fine-grain partitioned 3D integration. In *Proc. Cust. Integr. Circ. Conf.*, pp. 1–4, 2012.
125. T. Thorolfsson, G. Luo, J. Cong, and P. D. Franzon. Logic-on-logic 3D integration and placement. In *Proc. 3D Syst. Integr. Conf.*, pp. 1–4, 2010.
126. R. Topaloglu. Applications driving 3-D integration and corresponding manufacturing challenges. In *Proc. Des. Autom. Conf.*, pp. 214–219, 2011.
127. M.-C. Tsai, T.-C. Wang, and T. T. Hwang. Through-silicon via planning in 3-D floorplanning. *Trans. VLSI Syst.*, 19(8):1448–1457, 2011.
128. R. R. Tummala. *System on Package: Miniaturization of the Entire System*. McGraw-Hill Professional, 2008.
129. G. Van der Plas, P. Limaye, I. Loi, A. Mercha, H. Oprins et al. Design issues and considerations for low-cost 3-D TSV IC technology. *J. Solid-State Circ.*, 46(1):293–307, 2011.
130. F. von Trapp. GLOBALFOUNDRIES has its 3D Ducks in a Row. <http://www.3dincites.com/2014/11/globalfoundries-3d-ducks-row/>, November 2014. Accessed on December 2014.
131. R. Wang, E. F. Y. Young, and C.-K. Cheng. Complexity of 3-D floor-plans by analysis of graph cuboidal dual hardness. *Trans. Des. Autom. Elec. Syst.*, 15(4):33:1–33:22, 2010.
132. M. Wordeman, J. Silberman, G. Maier, and M. Scheuermann. A 3D system prototype of an eDRAM cache stacked over processor-like logic using through-silicon vias. In *Proc. Int. Solid-State Circ. Conf.*, pp. 186–187, 2012.
133. H. Xu, V. F. Pavlidis, and G. De Micheli. Effect of process variations in 3D global clock distribution networks. *J. Emerg. Tech. in Comp. Sys.*, 8(3):20:1–20:25, 2012.
134. J.-S. Yang, K. Athikulwongse, Y.-J. Lee, S. K. Lim, and D. Z. Pan. TSV stress aware timing analysis with applications to 3D-IC layout optimization. In *Proc. Des. Autom. Conf.*, pp. 803–806, 2010.
135. W. Yao, S. Pan, B. Achkir, J. Fan, and L. He. Modeling and application of multi-port TSV networks in 3-D IC. *Trans. Comput.-Aided Des. Integr. Circ. Syst.*, 32(4):487–496, 2013.
136. F. Yazdani. Readiness of 2.5D/3D IC package design environment. <http://www.gsaglobal.org/wp-content/uploads/2012/04/3D-IC-Readiness-BroadPak.pdf>, 2014. Accessed on December 2014.
137. F. Yazdani and J. Park. Pathfinding methodology for optimal design and integration of 2.5D/3D interconnects. In *Proc. Elec. Compon. Technol. Conf.*, pp. 1667–1672, 2014.
138. P.-H. Yuh, C.-L. Yang, and Y.-W. Chang. Temporal floorplanning using the T-tree formulation. In *Proc. Int. Conf. Comput.-Aided Des.*, pp. 300–305, 2004.
139. C. Zhang and L. Li. Characterization and design of through-silicon via arrays in three-dimensional ICs based on thermomechanical modeling. *Trans. Electron Devices*, 58(2):279–287, 2011.
140. C. Zhang and G. Sun. Fabrication cost analysis for 2D, 2.5D, and 3D IC designs. In *Proc. 3D Syst. Integr. Conf.*, pp. 1–4, 2012.

141. T. Zhang, K. Wang, Y. Feng, Y. Chen, Q. Li et al. A 3D SoC design for H .264 application with on-chip DRAM stacking. In *Proc. 3D Syst. Integr. Conf.*, pp. 1–6, 2010.
142. T. Zhang, Y. Zhan, and S. S. Sapatnekar. Temperature-aware routing in 3D ICs. In *Proc. Asia South Pacific Des. Autom. Conf.*, Yokohama, Japan, pp. 1–6, 2006.
143. X. Zhao, D. L. Lewis, H.-H.S. Lee, and S. K. Lim. Pre-bond testable low-power clock tree design for 3D stacked ICs. In *Proc. Int. Conf. Comput.-Aided Des.*, pp. 184–190, 2009.
144. X. Zhao and S. K. Lim. Power and slew-aware clock network design for through-silicon-via (TSV) based 3D ICs. In *Proc. Asia South Pacific Des. Autom. Conf.*, Taiwan, pp. 175–180, 2010.
145. X. Zhao and S. K. Lim. TSV array utilization in low-power 3D clock network design. In *Proc. Int. Symp. Low Power Elec. Design*, pp. 21–26, 2012.
146. X. Zhao, J. Minz, and S. K. Lim. Low-power and reliable clock network design for through-silicon via (TSV) based 3D ICs. *Trans. Compon. Packag. Manuf. Technol.*, 1(2):247–259, 2011.
147. X. Zhao, J. R. Tolbert, C. Liu, S. Mukhopadhyay, and S. K. Lim. Variation-aware clock network design methodology for ultra-low voltage (ULV) circuits. In *Proc. Int. Symp. Low Power Elec. Design*, pp. 9–14, 2011.
148. X. Zhao, S. Mukhopadhyay, and S. K. Lim. Variation-tolerant and low-power clock network design for 3D ICS. In *Electronic Components and Technology Conference (ECTC), 2011 IEEE 61st*, pp. 2007–2014, May 2011.
149. P. Zhou, K. Sridharan, and S. S. Sapatnekar. Congestion-aware power grid optimization for 3D circuits using MIM and CMOS decoupling capacitors. In *Proc. Asia South Pacific Des. Autom. Conf.*, pp. 179–184, Yokohama, Japan, 2009.
150. Q. K. Zhu. *High-Speed Clock Network Design*. Kluwer Academic Publishers, Boston, CA, 2003.