3D Physical Design: Challenges and Solutions

Robert Fischbach, Jens Lienig, Tilo Meister
Dresden University of Technology
Institute of Electromechanical and Electronic Design, Dresden, Germany
www.ifte.de
Email: {fischbach, lienig, meister}@ifte.de

Abstract

Modern three-dimensional (3D) designs, in which the active devices are placed in multiple layers using 3D integration technologies, are helping to maintain the validity of Moore's law in today's nano era. However, progress in commercial 3D ICs has been slow due to multiple reasons. One of them is the lack of appropriate physical design (layout) tools that take the new constraints arising from the third dimension into account. In this paper, an overview of physical design's challenges in the new 3-dimensional context are presented. Specifically, we investigate the physical design steps of floorplanning, placement, routability prediction and routing. New 3D-tailored design methodologies are presented that are capable of addressing 3D-specific design challenges.

1 Introduction

Modern three-dimensional (3D) designs, in which the active devices are placed in multiple layers, are a promising option to keep Moore's law in effect in today's nanoscale era (Fig. 1). However, despite many advantages (such as heterogeneous integration), 3D designs are only slowly gaining practical importance. Obviously, 3D integration faces enormous challenges in both technology and physical design. Typical problems in 3D integration technologies are, among others, reliability, alignment accuracy, reuse of existing (2D) IP blocks, testability and thermal issues. While thermal integrity is a critical issue in all high-performance chips (system reliability is strongly dependent on the temperature), this problem is even more significant for 3D designs due to the higher power density in the stacked arrangement.

During physical design, all circuit components are instantiated with their geometric representations, resulting in a layout representation of the circuit. In other words, geometric images (shape, size, and metal layer) of all macros, cells, gates, transistors, etc., are assigned a location (*floorplanning*, *placement*) and have their interconnects laid out (*routing*). The result of physical design is a set of manufacturing specifications that must be subsequently verified and optimized (*verification/optimization*).

While the major step sequence in 3D physical design does not differ compared to 2D design (Fig. 2), each individual step has to take the special constraints of 3D integration into account. Hence, physical design of 3D circuits cannot be simply viewed as a stack of multiple 2D physical designs. Major issues are, for example:

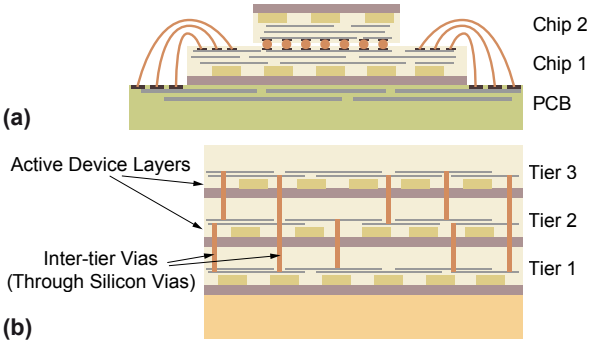

Figure 1: Schematics of 3D designs such as (a) 3D packages and (b) 3D integrated circuits. While 3D packages allow the vertical integration of various heterogeneous technologies, 3D integrated circuits enable denser circuits due to smaller distances between active devices in the third dimension.

- New net topologies spanning more than one tier are inherent to interconnect in 3D technologies.
- Inter-tier vias (Fig. 1) that interconnect tiers are not comparable to "regular" signal vias regarding their impact on circuit characteristics. Furthermore, the density of inter-tier vias is at least 100 times smaller than signal via density for all current and foreseeable 3D technologies [9].
- Interconnect and active components interact more tightly because of blockages constituted by thermal and inter-tier vias.
- A more complex heat management is necessary; thermal vias and other mechanisms are required to dissipate heat.
- The integration of existing (2D) IP blocks poses a severe challenge due to its limitation on the placement of inter-tier vias.

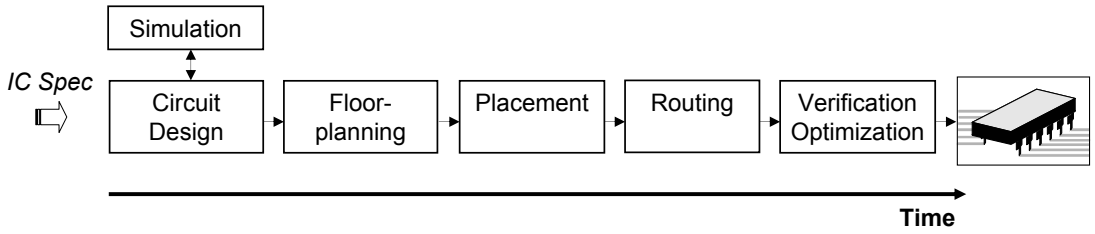

Figure 2: The major steps in the circuit design flow with a focus on physical design. It should be noted that modern VLSI design flows tend to blur the boundaries of these individual steps and have more integrated approaches. But for ease of explanation, these individual step notations are still used here.

In this paper, the major physical design steps floorplanning, placement and routing in the context of 3D designs will be discussed. Routability prediction is also investigated because it is an inherent part of these steps and it is greatly affected by the transition into a third dimension.

2 Floorplanning

During floorplanning, the shapes and positions of various modules (circuit partitions, 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 the subsequent placement (see Sec. 3) and routing (see Sec. 5) which both determine the internal characteristics of the module.

Conventional floorplanning assumes a single two-dimensional layer on which several modules must be arranged. A wide variety of different algorithmic approaches have been used in order to solve the floorplanning problem. 3D floorplanning includes new, 3D-specific characteristics that must be represented in the underlying data structures. For example, high output power modules need comprehensive consideration (i.e., thermal-driven floorplanning, [13]) and vertical dependencies arise in addition to horizontal ones.

There are two ways to represent vertical dependencies. The first possibility is the multiple usage of classical data structures, so-called 2.5D methods. Here, additional mechanisms have to be implemented to consider vertical relations between modules placed in different tiers, such as vertical alignment as well as overlapping and non-overlapping constraints. These representations include a discrete z-direction, such as in the combined bucket and 2D array (CBA) approach in [2].

It quickly became obvious that vertical dependencies must be incorporated directly into the data structure. This prevents invalid solutions without time-consuming evaluation and allows both that constraints are considered efficiently and solution space is minimized. More recent 3D data structures for floorplanning represent multilayer modules in all three dimensions. An example for such a 3D data structure is the 3D Slicing Tree described in [1]. As illustrated in Fig. 3, 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. However, solutions from a 3D Slicing Tree are limited to slicing floorplans.

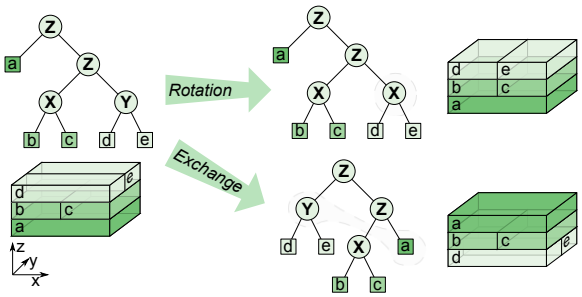

Figure 3: Illustration of 3D Slicing Tree operations to permutate a given 3D floorplan [5]. A rotation alters an inner node (representing a cut through the normal plane) resulting in a physical rotation of modules contained in the subtrees of that node. An exchange swaps two subtrees resulting in a physical exchange of modules contained in these subtrees.

3 Placement

After floorplanning to determine module outlines and pin locations, 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).

While 2D placement is limited to one (planar) layer, 3D placement requires optimizing the placement between multiple active layers. As mentioned, thermal constraints are crucial for 3D designs. Hence, 3D placement must ensure that thermal considerations are fulfilled. For example, the placer must spread cells such that a reasonable temperature distribution can be expected. However, due to the increased package density,

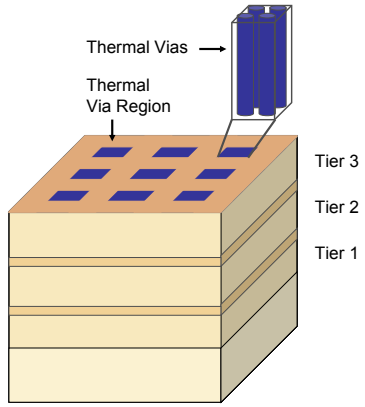

Figure 4: Regular thermal via regions in a 3D integrated circuit.

additional techniques are required to tackle the heat dissipation issue in 3D designs. Therefore, vertical metal structures that serve as "heat removers", so-called *thermal vias*, play an important role in achieving a thermal solution. Essentially, a thermal via acts as a pipe to transport the heat from a higher temperature region, such as a cluster of cells within the chip, to a heat sink, such as a thermal interface material with cooling fins above or below the chip.

Resulting from this thermal constraint, 3D placement must not only place cells and inter-tier vias but also take the positions of thermal vias into account. While cell placement usually concerns only one tier, the placement position of a thermal via is affecting all layers because it creates a blockage area throughout all layers (tiers). As such, it may represent a severe problem for cell placement and routing. Furthermore, cell placement and thermal via placement are interacting because the position and size of a thermal via depend on the thermal energy of the cells in its neighborhood. This requires a full-chip thermal analysis at the granularity of the individual thermal via which would lead to an unreasonable thermal simulation matrix.

However, practical solutions presented in literature over the last couple of years, such as [7], have addressed these problems by designating specific areas ("regions") within the circuit as potential thermal via sites (Fig. 4). Subsequently, the placer can be limited to determine the density with which these regions are filled with thermal vias. One advantage of this approach is the large granularity with which the thermal analysis method could work. The thermal conductivity of each region (element) can be treated as a design variable that is only subsequently translated into a precise number of thermal vias placed inside this region. Another advantage is the regularity of these blockage areas which can be considered much easier by the subsequent routing procedure.

Besides accounting for thermal issues, 3D placement also requires new placement methodologies. One of the first 3D placement approaches is presented in [4]. Here, a netlist hypergraph is embedded into the layout area. A recursive bipartitioning procedure is used to assign nodes of the hypergraph to the partitions with partition's capacities taken into account. The partitioning in vertical direction (z direction) correspondents to tier assignment. Thermal constraints are not considered in this approach.

A 3D-specific analytical placer, based on a force-directed method, is presented in [6]. It takes thermal objectives directly into account by using Finite Element Analysis (FEA) which discretizes the design space into regions with thermal boundary conditions. Attractive forces are created between connected cells which are proportional to the squared Euclidean distances between cells. Multiplicative factors for these distances are chosen to be higher in the vertical (z) direction in order to discourage inter-tier vias. While repulsive forces are commonly used to prevent cell overlaps, this 3D version of an analytical placer avoids hot spots by including thermal criteria to define these forces. Specifically, the temperature gradient is used to determine magnitudes and directions of the repulsive forces.

It is obvious that 3D layouts have limited flexibility in the third dimension due to both the relatively small number of feasible tiers and the scarce availability of inter-tier connections (vias). According to [8], this favors partitioning approaches (rather than force-directed solutions) at least during global placement. Accordingly, this work initially uses recursive bisectioning to perform global placement. This step is followed by coarse legalization in which the methods of FastPlace [11] are generalized. Finally, a non-overlapping layout is generated using detailed legalization. Thermal effects are incorporated through thermal resistance reduction nets. They are attractive forces that aim for keeping high power nets to remain close to heat sinks. Subsequently, this placement provides a good trade-off between conventional objectives such as the overall wirelength and 3D-specific goals such as the number of inter-tier vias and temperature.

4 Routability Prediction

One of the last steps in a physical design flow is routing, i.e., defining the interconnect geometry (see Sec. 5). As a result of the routing stage, not only the geometry of interconnects is determined but also electrical properties of the circuit are defined. In order to achieve good routing results, all previous design steps have to be optimized with regard to routability. Therefore, evaluating routability is an inherent part of most physical design steps. Current development towards 3D designs with complex interconnect topologies requires new approaches for routability prediction.

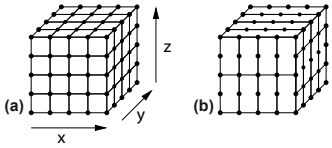

Figure 5: (a) General graph modeling interconnect resources and (b) a graph with a reduced number of edges by assuming preferred directions for routing [5].

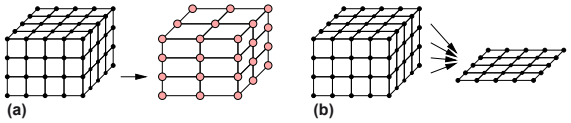

Figure 6: Simplifications of the routing graph. The coarsening of the routing grid is shown at the left while combining layers is illustrated at the right.

In order to investigate routability prediction in the 3D context, we must first address the issue of *modeling routing resources* for 3D ICs.

Routing resources can be modeled by a rectilinear graph with weighted edges as shown in Fig. 5a. Nodes correspond to regions in the design and weighted edges correspond to routing resources in those regions. Each of the nodes is assigned a location (x, y, z). Here, x and y are continuous coordinates within a layer, while z is discrete and specifies the interconnect layer. In practice, x and y are often artificially restricted to discrete values, and preferred directions for interconnects are assumed. These two measures simplify the design process by limiting the number of nodes as well as reducing the number of edges of the graph (Fig. 5b).

This general model of routing resources is valid for 3D designs as well as for previous 2D technologies. This observation would imply that the complexity of routing prediction does not increase for 3D ICs. However, during 3D routing prediction, the number of interconnect layers that have to be taken into account is larger, design objectives differ compared to previous technologies, and additional 3D-specific constraints must be considered.

As discussed, routability prediction is characterized by a trade-off between accuracy and runtime. Balancing this trade-off is even more important for 3D routability prediction as solutions spaces are generally larger. Depending on the required accuracy, the routing graph has to be simplified by an appropriate degree. For 2D systems, the routing graph is simplified by combining routing layers (Fig. 6b) and/or by coarsening the routing grid (Fig. 6a). The former simplification discards any information about vias. The latter – which reduces the x- and y-resolution by an arbitrary factor – is the basic concept of global routing.

While merging all routing layers has proven to be effective for routability prediction of 2D ICs, this simplification is not applicable to 3D ICs due to the omission of inter-tier vias. These vias must be included during 3D routability prediction because of their impact on routing (blockages, electrical properties, etc.). Therefore, routing layers of different tiers have to be treated separately. This requires a new intermediate level of simplification when combining routing layers (Fig. 7). In a first step (from (a) to (b)), only layers of the same tier are combined. Only if a further simplification is necessary, layers of different tiers are merged (from (b) to (c)). In this case, however, characteristics of thermal and inter-tier vias are omitted and the routability prediction is only of limited use.

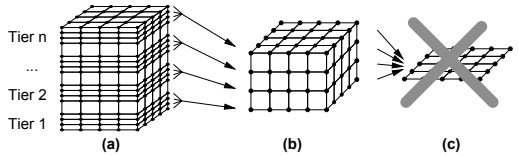

Figure 7: Simplifications of the routing graph for 3D ICs. In addition to the simplification shown in Fig. 6b, an intermediate level of simplification (b) is necessary. Simplification level (c) is only of limited use for 3D ICs due to its omission of thermal and inter-tier vias.

After reviewing the problem of modeling routing resources for 3D designs, the process of *routability prediction* is discussed next. Here, the adaptation of 2D methods, such as global routing and other, often faster methods, is required.

The problem of adapting global routing to the third dimension, i.e., global routing algorithms that find routing paths in several interconnect layers while considering vias, has been solved for some years. Various multilayer global routers are applicable to 3D circuits if vertical/via routing capacities within tiers and between tiers can be specified independently. This differentiation is necessary in order to respect the different properties of inter-tier vias and conventional signal vias.

However, the challenges of 3D routability prediction without performing global routing have not been solved yet. These fast methods of routability prediction are based on simple estimations ("guessing") of routing paths. To address this challenge, we introduce an effective model next. It extends routing density distributions to three dimensions in order to adapt statistical estimations of routing demand to the requirements of 3D interconnect topologies.

In the following, the 2D density approach presented in [10] (depicted in Fig. 8a) is used as an example to illustrate our method. The 2D-routing-density distribution originally enables the prediction of routing demand, overflow, congestion and thus, routability of a 2D circuit. However, for reasons mentioned earlier, this approach is not sufficient to predict routability of 3D circuits. It must be extended by vertical routing capacities and routing

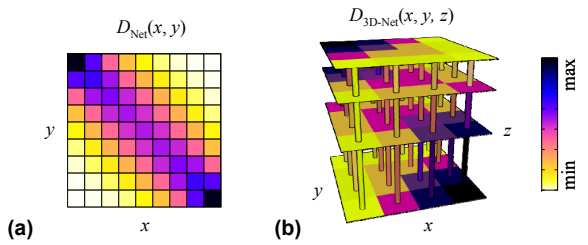

Figure 8: (a) Routing density distribution for a two pin net in one layer and (b) extended three dimensional routing density distribution for a two pin net in four tiers. A darker color indicates a higher expected density. All routing paths are assumed to be of the same probability.

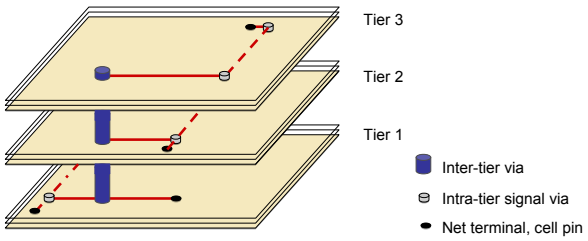

Figure 9: Example route for a net in a three-tier 3D design.

density distributions for each tier to render it viable for 3D ICs.

Our model to represent a 3D routing density distribution is shown in Fig. 8b [5]. Depending on the level of simplification, the layers of the wiring-density-distribution either correspond to individual routing layers or to the combined layers of one tier. Constraints such as blockages and varying densities of inter-tier vias are considered by this model by means of a varying probability of routing paths. Using our 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 (tier) as well as the expected (vertical) inter-tier via density.

5 Routing

A net is a set of two or more cell pins that have the same electrical potential in the final chip design. The circuit netlist includes all of the nets in the design. During the routing stage, all terminals of the nets in the circuit netlist 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 mentioned, the main difference between regular (2D) and 3D routing is caused by the multi-tier position of net terminals that lead to net topologies which span more than one tier (Fig. 9). This requires expensive inter-tier vias to be used in addition to regular signal vias which connect metal layers within one tier. Furthermore, 3D routing must take additional constraints into account, such as blockages introduced by thermal and inter-tier vias. These constraints require a much more sophisticated congestions management and blockage avoidance as is common for regular 2D routing. Additionally, the limited availability of inter-tier vias requires a careful allocation of this valuable resource among nets. The increased impact of temperature on 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 the hottest regions of the chip.

Next, two routing approaches are discussed that consider these 3D-specific constraints. Due to their promising results, they point to a new class of routers targeted for 3D designs.

Cong and Zhang present in [3] a thermal-driven 3D router using a multi-level routing approach composed of a recursive coarsening, an initial routing, and a recursive refinement process. Its major feature is a thermal-driven via planning algorithm. Based on this global view and capabilities of a multi-level planning scheme, the via planning step effectively optimizes temperature distribution and wirelength using direct planning of the inter-tier vias instead of indirect planning through a routing path search. This approach allows effectively controlling the chip temperature. It should be noted that any inter-tier via is also considered as a thermal via. The approach adds dummy inter-tier vias when the signal inter-tier via number is not sufficient to bring down the chip temperature to a required level.

An initial routing solution is built using a 3D minimum spanning tree (MST) for each multi-pin net. Obstacles, such as thermal via regions, are avoided by using a simple maze routing algorithm. Then, the number of dummy inter-tier vias to be inserted is estimated using binary search. The upper bound of dummy inter-tier via numbers that can be inserted into each device layer is estimated by the amount of whitespace between the blocks. During each refinement stage, the inter-tier vias are refined first to minimize wirelength and temperature. This via-refinement process includes two steps, inter-tier via number distribution and signal inter-tier via assignment. These steps try to optimize temperature and wirelength, respectively. After inter-tier via refinement, wires are also adjusted according to the updated via positions.

Experimental results suggest that the presented methodology is effective in using dummy inter-tier vias to keep the chip temperature below a required level without major increases in wirelength and runtime.

Another approach is presented in [12]. 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 objects with the function of spreading thermal energy in the lateral direction.

Thermal vias perform the bulk of the conduction to the heat sink, while thermal wires help distributing the heat paths among multiple thermal vias.
Similar to the approach described above, an initial routing solution is generated using a 3D minimum spanning tree (MST) for each net. Routing congestion for subsequent two-pin decomposition is estimated over each lateral routing edge utilizing a probabilistic congestion prediction. This estimation is based on the assumption that a two-pin net with pins in different tiers has an equal probability of using any inter-tier via position within its bounding box.
Afterwards, a recursive bipartitioning scheme is applied to define the exact positions of the inter-tier vias. With a known signal inter-tier via distribution, a temperature analysis can be performed to obtain an initial temperature map. Moreover, once the inter-tier vias are fixed, the task is reduced to a 2D routing problem in each tier that can be solved by maze routing. Its cost function contains an additional temperature term, so that the resulting route will have an incentive to choose a low temperature path. This strategy not only limits the temperature impact on the delay of signal nets, but also reduces congestion in hot regions. This is beneficial since more thermal vias may be inserted later into these regions to reduce the chip temperature.
Finally, linear programming is used to insert thermal vias and thermal wires based on a thermal analysis. This process is performed iteratively; each time after the insertion, a rip-up-and-reroute procedure resolves the lateral routing congestion and overflow. This process stops when there are no temperature and congestion violations or no further improvement is possible.
Experimental results in [12] show that this temperature-aware 3D routing can efficiently resolve the contentions between placing thermal vias/wires and routing, generating a solution satisfying both congestion and temperature requirements.

6 Summary

This paper summarizes state-of-the-art layout approaches in floorplanning, placement, routability prediction and routing for 3D designs. While the complexity arising from the added third dimension is an enormous challenge, it is shown that new 3D-tailored design methodologies are increasingly capable of addressing this task.

7 References

[1] L. Cheng, L. Deng, and M. D. F. Wong. Floorplanning for 3-D VLSI design. In *ASPDAC*, pages 405–411, 2005.

[2] J. Cong, J. Wei, and Y. Zhang. A thermal-driven floorplanning algorithm for 3D ICs. In *ICCAD*, pages 306–313, 2004.

[3] J. Cong and Y. Zhang. Thermal-driven multilevel routing for 3D ICs. In *ASPDAC*, pages 121–126, 2005.

[4] S. Das, A. Chandrakasan, and R. Reif. Design tools for 3-D integrated circuits. In *ASPDAC*, pages 53–56, 2003.

[5] R. Fischbach, J. Lienig, and T. Meister. From 3D circuit technologies and data structures to interconnect prediction. In *SLIP*, pages 77–84, 2009.

[6] B. Goplen and S. Sapatnekar. Efficient thermal placement of standard cells in 3D ICs using a force directed approach. In *ICCAD*, pages 86–89, 2003.

[7] B. Goplen and S. Sapatnekar. Thermal via placement in 3D ICs. In *ISPD*, pages 167–174, 2005.

[8] B. Goplen and S. Sapatnekar. Placement of 3D ICs with thermal and interlayer via considerations. In *DAC*, pages 626–631, 2007.

[9] ITRS. International technology roadmap for semiconductors. Technical report, ESIA, JEITA, KSIA, TSIA and SIA, 2009.

[10] J. Lou, S. Krishnamoorthy, and H. S. Sheng. Estimating routing congestion using probabilistic analysis. In *ISPD*, pages 112–117, 2001.

[11] N. Viswanathan and C. C.-N. Chu. FastPlace: Efficient analytical placement using cell shifting, iterative local refinement,and a hybrid net model. *IEEE T. Comput. Aid. D.*, 24(5):722–733, 2005.

[12] T. Zhang, Y. Zhan, and S. S. Sapatnekar. Temperature-aware routing in 3D ICs. In *ASPDAC*, pages 309–314, 2006.

[13] P. Zhou, Y. Ma, Z. Li, R. Dick, L. Shang, H. Zhou, X. Hong, and Q. Zhou. 3D-STAF: scalable temperature and leakage aware floorplanning for three-dimensional integrated circuits. In *ICCAD*, pages 590–597, 2007.