Design Methodologies and Co-Design Options for Novel 3D Technologies

Tilman Horst, Robert Fischbach and Jens Lienig
Institute of Electromechanical and Electronic Design, Dresden, Germany
Email: [Tilman.Horst][Robert.Fischbach][Jens.Lienig]@tu-dresden.de

Abstract

Three-dimensional (3D) technologies enable faster, smaller and more energy efficient systems. Moreover, after more than two decades of research, the technologies are mature and ready for industrial application. However, only few developers make use of them and adoption by market is lagging behind. One of the main reasons is that 3D technologies are difficult to be integrated into existing two-dimensional (2D) design flows, with the latter often wasting the potential promised by 3D technologies. Therefore, 3D-compatible design flows, with an emphasis on co-design between various 3D hierarchy levels, are required. In this paper, we first give an overview of the most relevant and mature 3D technologies. Subsequently, we investigate related design methodologies in order to find so far neglected potentials for design optimization. The presented co-design options enable designers to better exploit the potential offered by novel 3D technologies.

1 Introduction

Novel three-dimensional (3D) technologies, such as through silicon vias (TSVs), microbumps (μ-bumps) or interposers, offer new ways for the integration of micro-electronic systems [1]. The exploitation of these technologies enables faster, smaller and more energy efficient systems. In addition to that, 3D technologies support heterogeneous systems, which integrate optical, microfluidic or other components diverging from traditional complementary metal-oxide-semiconductor (CMOS) technologies ("more-than-Moore"). Thereby, novel applications such as photonic integrated circuits (PICs), labs-on-a-chip or new generations of smart camera chips, which provide not just raw pixel values, but precise scenic information, are rendered possible. Moreover, recent industrial developments, such as memory cubes [2], interposer-based video cards [3] or interposer-based field-programmable gate arrays (FPGAs) [4], have proven that 3D technologies are mature and ready for industrial application. However, to fully exploit the potential revealed by 3D technologies, suitable design approaches are necessary [5]. These approaches have to enable concurrent design of all system components in order to assure system functionality and to meet globally optimized performance goals.

As depicted in Fig. 1, a wide range of electronic systems can be classified as heterogeneous or as 3D. This range can be further subdivided according to the level of integration, i.e. board level, package level and die level. Board-level integration comprises assembling techniques such as printed circuit boards (PCBs), multi-chip modules (MCMs) or package on package (PoPs). The common characteristic of these technologies is that they integrate already packaged components. The components can be active or passive, such as transistors, resistors, capacitors or entire integrated circuits (ICs).

On package level (Fig. 1, middle), 3D system in packages (SiPs) enable the integration of multiple bare dies within a single package. Using wire bond interconnect technology, this assembling technique allows for higher interconnect densities than aforementioned board-level integration. Other package-level techniques are interposer-based systems as well as TSV-based die stacks.

On die level (Fig. 1, right), monolithic 3D ICs consist of multiple active layers and therefore enable three-dimensionality inside the die. They employ nanoscale monolithic inter-tier vias (MIVs), rather than TSVs, to establish vertical interconnects. MIVs enable orders of magnitude higher interconnect densities compared to TSV-based systems [6]. However, as opposed to systems integrated on board level or package level, monolithic 3D ICs do not allow for heterogeneous integration, which restricts their functional diversity.

The focus of this paper is package-level integration, i.e., 3D-SiPs, TSV-based die stacks, as well as interposer-based systems. These systems consist of multiple unpackaged or bare dies that are integrated within one single package. This requires multiple design hierarchies to be optimized, subsequently labeled as “co-design”.

The goal of this survey paper is to point out the fundamental need for co-design with regard to 3D package-level integration and to present suitable design flows. In Section 2, we give an overview of the most important package-level 3D technologies. After that, in Section 3, we investigate appropriate 3D design methodologies in order to find unused potentials for design optimization. In Section 4, we summarize our findings.
2 3D Technologies

We introduce today’s most important 3D interconnect technologies on package level, such as wire bonding, microbumps and TSVs, in Subsection 2.1. Furthermore, we give an overview of the main package-level assembling techniques, such as 3D SiPs, interposer-based 2.5-dimensional (2.5D) systems and TSV-based 3D ICs, in Subsection 2.2.

2.1 Interconnect Technologies

Interconnect technologies are the key enabler for 3D integration. However, they add constraints to the package and die design such as minimum interconnect pitches, prohibited wire intersections, increased parasitics, silicon overhead and so forth. Moreover, the higher the interconnect density enabled by a specific technology, the more complex is the design of a system using this technology. In this section, we introduce wire bonding, μ-bump and TSV interconnect technologies (Fig. 2) and discuss their impact on the design.

2.1.1 Wire Bonding

Wire bonding has a long history in packaging (e.g. micromechanical systems (MEMS), memory die stacking) and is considered to be very mature [7]. Interconnectivity is established by stretching a wire between contact pads of unpackaged dies and substrate. Therefore, wires are typically several mm in length, which results in high inductance and significant crosstalk in case of closely spaced wires [7]. Typically, one to four lines of wire bonds can be arranged at the periphery of the die. The inner area of the die cannot be accessed, which restricts the maximum number of interconnections. Wire diameters range from 15 μm to 30 μm depending on the current that is supposed to flow through the wire. Pitches range from 50 μm to 70 μm [8].

2.1.2 Microbumps

In principle, a microbump (μ-bump) is similar to a controlled collapse chip connection (C4) bump used in flip chip packaging technology, although on a smaller scale. μ-bumps are used to establish die-to-substrate or die-to-die connections. Accordingly, μ-bumps can be either used to create packages containing multiple dies arranged on the substrate in a two-dimensional (2D) fashion or to realize 3D ICs consisting of two dies stacked to each other in a face-to-face (F2F) manner.

If multiple dies are attached to a common substrate using μ-bumps, the substrate acts as a redistribution layer (RDL), which connects the dies to each other. The RDL thereby provides flexibility for the position, orientation and input/output (I/O) configuration of individual dies. If two dies are stacked onto each other F2F, no RDL is available. This constrains the physical position of interconnected I/Os of both dies and results in a complex chicken-and-egg problem as will be discussed in Section 3.

As opposed to wire bond technology, μ-bumps enable connectivity across the whole die area and not only peripheral interconnections. Therefore, large interconnect counts can be realized. Currently, μ-bump pitches around 50 μm are common in industrial applications [9]. This is equal to 400 interconnects per mm². μ-bumps are an active field of research to enable ever smaller diameters and pitches. In 2016, 7.6 μm μ-bumps were reported [10]. However, when going beyond 10 μm, μ-bump formation poses major challenges and becomes inefficient [11]. Alternatives like wafer-to-wafer direct bonding or hybrid bonding techniques must be considered. These technologies enable contact pitches of around 1 μm or even less [11], [12].

2.1.3 Through-Silicon Vias

A through-silicon via (TSV) penetrates and goes all the way through the silicon substrate of a die. Thereby, dies can be contacted on either side and connectivity between face-to-back (F2B) aligned dies is enabled. As opposed to μ-bumps, which enable vertical connectivity between a maximum of two F2F aligned dies, TSVs allow for vertical inter-die connectivity across any number of layers.

The first step in TSV manufacturing is hole formation, followed by TSV metallization. Deepening the hole can be done by plasma etching or laser drilling [13]. In general, the TSV diameter is fairly scalable and depends on the deepness of the hole. That is, thin substrates and shallow TSVs enable smaller diameters, and in return deep TSVs for thick substrates require wider diameters [12]. Diameters typically range from 10 μm in case of large TSVs and thick wafers (70 μm or more) down to 2 μm in case of extremely thinned wafers. A common aspect ratio for TSV diameter and substrate thickness is 10:1. For a typical TSV diameter of 5 μm, a 10 μm TSV pitch can be assumed [12], [14].

Due to a different coefficient of thermal expansion (CTE) of Cu and Si, TSVs are likely to induce thermo-mechanical stress, which requires keep-out zones (KOZs) in order to ensure correct behavior of surrounding gates. This typically causes large silicon overhead to go along with TSVs [14].

TSVs additionally require bonding technologies like μ-bumps or alternatives in order to eventually connect a TSV with the contact pad of another die.

2.2 Assembling Technologies

The 3D assembling technique has far-reaching consequences for the design approach that has to be used for the design of the system. The most important criterion is the existence of a redistribution layer (RDL) such as a package substrate or an interposer. If a RDL exists, the individual dies can be designed independently and only the RDL needs to be custom-designed. Nonetheless, custom-design of individual dies is optional and could help to improve the system performance. With regard to design reusabil-
ity. RDLs are an important prerequisite because they make sure the different components can be connected properly.

In case of stacked dies without a RDL, custom-design of all dies is essential and design reuse is not an option. In order to ensure the functionality of die stacks, elaborated design methodologies are required.

In this section, we present the most important integration schemes for heterogeneous 3D systems (Fig. 3) and outline the most important characteristics with regard to an appropriate design approach for each assembling technique.

### 2.2.1 Traditional SiP and 3D SiP

Packaging technologies allow for heterogeneous integration of different components, such as memory dies, processors or MEMS devices. Traditional 2D SiPs make use of a RDL, which carries and connects laterally aligned system components (passives and bare dies). 3D SiPs, on the other hand, allow for vertical stacking of multiple dies (up to five [7]) without using an intermediate RDL. The prevailing interconnect technology for traditional as well as 3D SiPs is wire bonding. Direct stacking requires closely coordinated design of all dies, whereas the RDL in traditional 2D SiPs provides more degrees of freedom for the design of individual dies.¹

3D packaging emerged from traditional SiP and relies on sophisticated supply chains and packaging infrastructure [12]. SiPs enable design flexibility, short time-to-market, low risk and low cost [15]. Overall, SiPs allow for the integration of multiple devices in a small package, but due to wire bonding interconnect technology, the integration density is limited and the system performance is furthermore reduced by wire parasitics [16].

### 2.2.2 Interposer-Based 2.5D systems

In interposer-based systems, the interposer serves as an integration platform for the system components, i.e. bare dies. That is, the interposer acts as a carrier for the components and RDL, but does not necessarily feature functionality in itself. The dies attached to the interposer may stem from different manufacturing processes (i.e. analog, digital, memory), labeling the interposer a heterogeneous integration technology. Moreover, interposers support design reuse and improve compound yield by enabling to pick known-good dies [17]. An interposer can be either organic or anorganic. However, anorganic interposers (e.g. Si) can be fabricated with standard back end of line (BEOL) wafer processes to enable fine-pitch interconnects, while organic interposer have more limited interconnect capabilities [18]. To attach and connect dies to Si interposers, μ-bumps with pitches around 50 μm are employed [19]. However, interposer technologies remain expensive. Research is done in order to replace Si substrates by glass substrates [18], [20], aiming to reduce costs by a factor of ten [20].

### 2.2.3 Stacked 3D ICs

Stacked 3D ICs consist of multiple dies, which are assembled by vertical stacking. Stacked 3D ICs allow for heterogeneous integration of dies from different technologies (e.g. analog, digital, memory). Dies are directly stacked onto each other without an intermediate RDL. Accordingly, the contact pads must match each other exactly, leaving little or no freedom for the I/O configuration of individual dies. Therefore, design reuse of dies across projects is restricted. In order to enable vertical interconnectivity, TSVs in combination with μ-bumps or other bonding technologies are used. A special case of stacked 3D ICs are F2F bonded dies. Due to the F2F configuration, TSVs are not required, but the stack is restricted to a maximum of two dies. Stacked TSV-based 3D ICs already in production (i.e. memory cubes) use interconnect pitches of 55 μm [21], [22]. Academic studies succeeded to scale the interconnect pitch below 5 μm by using bumpless bonding techniques [23]. 3D ICs enable footprint reduction of about 40 % [24]. However, the high capacitance of TSVs reduce the overall performance improvements [6].

### 3 Co-Design Options of 3D Designs

Designing 3D integrated systems consisting of multiple dies is a classic “chicken-and-egg problem”: Package-level design requires at least the die footprints as input. At the same time, footprint generation and reasonable I/O configuration require a package layout as entry point.

Coping with this chicken-and-egg situation requires a high level of incorporation between packaging and IC design tools, subsequently labelled as “die/package co-design”. However, historically package and IC design are two isolated processes, leaving little room for this co-design.²

Current 3D design methodologies are referred to as “throw-over-the-wall” because ICs are designed and

---

¹Intersecting wires at package level should be avoided.

²Initially, packages contained only one bare die, and mainly straightforward pin fan-out was pursued at package level. Therefore, as the tools developed, there was no need for die/package co-design.
handed over without exactly knowing the conditions of usage at package level. This involves several problems. For example, tough or even unresolvable pin assignment problems may arise. Moreover, while each die might be optimized to the last detail using conventional 2D flows, the overall system is unlikely to reach the same degree of optimization and therefore renders the previous effort obsolete.

In this section, we suggest three different methodologies to cope with the described chicken-and-egg situation. The first two approaches, namely chip-driven package design and die/package co-design, divide the overall 3D design problem into 2D design problems that are solvable using conventional flows. Key is the proper definition of the physical interface between the 2D dies and the package as well as the inter-die interfaces. The third approach, labelled 3D IC design flow, is considered in case of extraordinary high interconnect counts and accordingly complex interfaces between system components.

### 3.1 Chip-Driven Package Design

In the chip-driven package design methodology, depicted in Fig. 4, the package design and die design are two sequential processes. Firstly, dies are designed using a conventional RTL-to-GDS 2D design flow. Then, the package designers use the data previously generated by the IC designers to carry out the package design. The individual dies can be designed immediately before package design or can be “recycled” from previous projects. That is, design reuse is supported by this design methodology.

The actual package design process starts with the generation of a package-level schematic. In the package-level schematic, each die is represented by a block and by pins. The inter-die connectivity is expressed by lines between corresponding pins of different dies. After schematic generation, the physical implementation of the package can start by arranging the dies on the package substrate. For this task, the previously generated die footprints are imported from the 2D design flow.

Once all dies are allocated, the redistribution layer (RDL) routing is performed. Between placement and routing, optimization loops are possible. Examples for optimization goals are wire length and minimum redistribution layer count in order to reduce package cost.

The last step of the flow is the electrical simulation and verification (see Fig. 4). This includes package-level design rule check (DRC) and layout versus schematic (LVS). Moreover, parasitics of package interconnects can be extracted for electrical simulations. Subsequent optimization cycles assure that the system is meeting all electrical constraints. For example, if the simulation reveals crosstalk to be critical, this might be reduced by introducing grounded shielding between affected wires.

The physical interface of dies used at package level must comply with the package-level design rules. For example, bond pads must have the right pitches, so that they can be accessed by interconnect technologies like wire bonding properly. Package designers need to communicate these rules to the IC designers.

Because package and IC design are performed sequentially, this approach provides very limited co-design opportunities. Moreover, an important prerequisite for this approach is the existence of a RDL like a package substrate or an interposer. The RDL is required to reorganize and match misaligned signal paths. Accordingly, the approach is not appropriate for stacked package configurations which do not possess a RDL.

A possible implementation of this design methodology, which also goes more into the details regarding how to extract information from different sources of design data, can be found in [25].

### 3.2 Die/Package Co-Design

Using the chip-driven package design methodology depicted in Subsection 3.1, the dies are handed over for assembly with minimal interaction between die and package design teams. This “one-way” approach is likely to result in performance losses or overdesign. Another issue of the chip-driven style is that a RDL is crucial to assure connectivity between independently designed dies. That is, TSV-based or F2F bonded die stacks are impossible to be designed employing the approach from Subsection 3.1. However, these issues can be mitigated by performing a die/package co-design [26].

As depicted in Fig. 5, in contrary to the chip-driven approach, the package and die design flows are running in parallel providing manifold options for co-design. For the 2D die design, conventional flows can be used. Though, slight modifications are required to enable interaction with package design flows.

The package design starts, like in the chip-driven methodology, with a package schematic. However, for the following die placement die footprints are not available yet. Therefore, estimated die dimensions are provided by the 2D flow and dummy dies are created in order to perform an early package-level die placement. The next step, after dummy-die placement, is package-level pin assignment. The package-level pin assignment requires the pinouts of...
the dies, which depend on the intra-die floorplans. Because at this point, neither the pin assignment, nor the pinouts, nor the floorplans are determined, cross-hierarchical silicon/package co-design is greatly encouraged. Apart from appropriate methods, this co-design process also requires cross-hierarchical data structures representing the particulars of all design elements (e.g., dies and package) efficiently [27].

Once the co-design problem is solved iteratively, the package design flow continues with the RDL routing. As the previous pin assignment has been co-optimized, superior routing results can be expected. Once the die and package layouts are finished, package-level verification and simulation routines can be initiated. More information about possible silicon/package co-design options can be found in [28], [29] and [30].

### 3.3 3D IC Design for High Interconnect Counts

The methodologies investigated in Section 3.1 and 3.2 assume that the overall design problem is decomposed into multiple functional entities or dies. Subsequently, each entity and all the interfaces in between are specified and implemented almost independently. The system decomposition is done at early stages of the design process and in rather high levels of the design hierarchy.

Naturally, the interconnect density at top levels of the design hierarchy, i.e., system or block level, is lower than at bottom levels, i.e., gate level [12]. Therefore, the interconnect count between the entities or dies is low. However, technologies like F2F direct wafer bonding enable contact pitches of 1 μm, or, accordingly, one million interconnects per mm² [11]. Completely exploiting this potential requires the systems to be decomposed at low levels of the design hierarchy, i.e., gate level rather than system or block level. This requirement renders the methodologies from Section 3.1 and 3.2 impractical for systems using superior interconnect technologies and accordingly high numbers of interconnects.

Taking these considerations into account, we suggest a unified full 3D IC design flow as depicted in Fig. 6 for systems with high interconnect counts between the dies. Unlike the two previous approaches, this flow is not composed of two separate package and die flows. As the full 3D IC design flow treats the 3D design problem more holistically, there is no need for die/package co-design.

Similar to existing 2D design flows, this 3D flow starts with a register-transfer level (RTL) description of the entire system. Initially, all hard macros such as memory blocks are placed in appropriate layers of the 3D stack; these areas are subsequently blocked. Hard macro blocks are already implemented as 2D layouts and therefore cannot be split into multiple layers retroactively.

The initial macro placement is followed by gate-level 3D place and route in accordance with the 3D timing analysis. Inter-die connections between stacked dies behave differently compared to conventional intra-die connections and therefore need to be considered during 3D timing analysis. In the first instance, similar to 2D verification, DRC and LVS can be performed on each die separately during simulation and verification stage. However, signal and power integrity analysis must be performed on the whole system accounting for perturbations due to inter-die connections.

Overall, this proposed flow involves several steps, such as 3D place and route, proper clock tree synthesis, power distribution network generation as well as adequate electrical and thermal analysis, which already exist for 2D ICs, but need to be re-engineered for 3D systems. Apart from empowering the tools to operate in a 3D space, they also have to incorporate a number of new 3D specific require-
ments, such as thermal-aware design, steady power delivery from the bottom tier to the top tier, and prevention of couplings due to closer vertical proximity between devices [31]. Because appropriate 3D tools are not available yet, state-of-the-art 3D IC design flows similar to the proposed one utilize conventional 2D place and route tools. Here, a 2D placement is generated and subsequently split into multiple tiers as a post-process [32] [33]. Hence, this approach allows to benefit from the well-developed features of existing tools. However, this workaround involves many drawbacks and is restricted to a maximum of two dies stacked onto each other.

4 Summary and Conclusions
This paper introduced the most important 3D manufacturing technologies, followed by an investigation of suitable design methodologies for the 3D integration of bare dies. The chip-driven methodology, investigated in Subsection 3.1, proposes a sequential procedure starting with the dies, followed by the package-level design. The advantage of this approach is that conventional design tools can be used one after another requiring a minimum of tool modifications. Moreover, chip designs can be reused across projects in two or more different packages. However, the “one-way” character of this design approach does not allow for global optimization routines and squanders away the potential of novel 3D technologies. The die/package co-design methodology, proposed in Subsection 3.2, provides co-design opportunities by parallelizing chip and package design. More precisely, the pin assignment stage of the package-level design incorporates the floorplanning stage of a conventional 2D IC design flow. Thereby, package-level optimization goals can be considered during die design and, likewise, die-level issues can be incorporated at package level.

The aforementioned design methodologies need the overall system to be decomposed into several entities or dies at early design stages. This only allows for system partitioning at high levels of the design hierarchy, such as system or block level. However, in order to fully exploit high-density interconnect technologies, system decomposition at gate level is required. Therefore, in Subsection 3.3, a full 3D IC design flow is proposed for designing systems with extraordinary high numbers of inter-die connections. This flow requires new tools, customized for the requirements of 3D design, which are not available at present.

By comparing the design flows in this paper with the ones that are currently used in industrial applications, potentials for design improvements become visible and a better design optimization is enabled. Eventually, this allows to harness the full potential of novel and mature 3D technologies.

Acknowledgement
This work has received funding from the ECSEL Joint Undertaking under grant agreement No 737465 (named MICROPRINCE). This Joint Undertaking receives support from the European Union’s Horizon 2020 research and innovation programme and Germany, Belgium, Ireland.

5 References