Microelectronics Journal 44 (2013) 152–162

Contents lists available at SciVerse ScienceDirect

Microelectronics Journal

journal homepage: www.elsevier.com/locate/mejo

Cost evaluation on reuse of generic network service dies in three-dimensional integrated circuits

Ji Wu [a,c], Gaofeng Wang [b,c,*]

[a] School of Computer Science, Wuhan University, Wuhan, Hubei 430072, China
[b] School of Computer Science and Education Software, Guangzhou University, Guangzhou, Guangdong 510006, China
[c] Institute of Microelectronics and Information Technology, School of Electronic Information, Wuhan University, Wuhan, Hubei 430072, China

ARTICLE INFO

Article history:
Received 13 May 2012
Received in revised form
26 August 2012
Accepted 28 August 2012
Available online 17 September 2012

Keywords:
Three-dimensional IC
Network-on-chip
Many-core processor design
Cost analysis
Reuse

ABSTRACT

Reuse of existing designs is one of the most effective means for cost reduction. Three-dimensional (3D) stacking technology makes possible reuse of dies in 3D stack. GNet – a 3D architecture for reusing generic network service dies – is herein proposed to construct a network-on-chip (NoC) by virtue of exploiting reuse of known good dies (KGDs). In GNet, generic network service dies (GNSDs) are KGDs that integrate several networks and can be directly used to bond with other dies in 3D stack. Flexible and configurable design of GNet makes it suitable to requirements in various application circumstances. A comprehensive cost model, which combines the design cost model, reuse model and fabrication model, is introduced to evaluate different architectures from the design phase to the fabrication phase. Experiments show that GNet is more cost efficient than the general 3D implementations.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Reuse of existing designs has always been one of the most effective means for cost reduction. The design cost data from the 2001 ITRS was analyzed [1], and it was found that some tool/methodology changes can significantly lower the design cost. As shown in Fig. 1, small block reuse (2500 to 24,999 gates) and large block reuse (25,000 to 100,000 gates) have the highest impact on design productivity.

Silicon technology is now at the stage where it is feasible to incorporate tens or even hundreds of millions of gates on a square centimeter of silicon. Granularity of reuse increases from thousands of gates in 1990s to millions of gates now. How to expand the reuse to a larger granularity is important to continue lowering design cost.

Network-on-chip (NoC) is chosen to be the popular architecture in many-core multiprocessor designs [2]. A large number of architectures on NoCs have been proposed. Since the concept on mesh-based packet switched NoC was proposed [2], the mesh-based topologies such as two-dimensional (2D) mesh, 2D torus and wormhole switching mechanisms have become the mainstream of NoCs [3–11]. Hence, it is possible to design a general-purpose network to meet the NoC demands based on the mesh-based topologies. This can push the granularity of reuse to tens of millions of gates or even more.

A 3D stacking architecture with interconnect service layer (ISL) was proposed in [12], which separates the network function from the system and provides interconnect service between the cache layer and the computing layer. Experiments were performed on a specific example of the ISL design, which illustrated die cost reduction and performance improvement brought by the 3D stacking architecture [12]. However, the 3D stacking architecture with ISL proposed in [12] has the following issues.

First, there is no comprehensive cost model for the 3D stacking architecture. Instead, the 3D cost model in [13], which is a fabrication model only, was used. Since large-scale reuse is the primary concern, non-recurring engineering (NRE) costs for design and manufacturing are important. Moreover, only one simple example without systemic analysis may not be convincing to illustrate the cost reduction brought by reuse of ISL.

Second, the method to verify the cost reduction lacks of the proper cost analysis on volume production. Moreover, the cost is calculated only by the die cost, while the saving die cost is calculated only by the reduction of through-silicon via (TSV) area overhead instead of a fabrication cost model.

Nevertheless, the major reduction of TSV area is due to the use of through silicon buses (TSBs). The TSBs are dynamic time-division multiple access (dTDMA) buses [14]. The TSBs can move data to both upper and lower layers, so the area occupied by TSBs

* Corresponding author. Tel.: +86 27 68778588 802; fax: +86 27 68778588 807.
E-mail addresses: jessie.wuji@gmail.com (J. Wu), gaofeng@whu.edu.cn (G. Wang).

0026-2692/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.mejo.2012.08.007

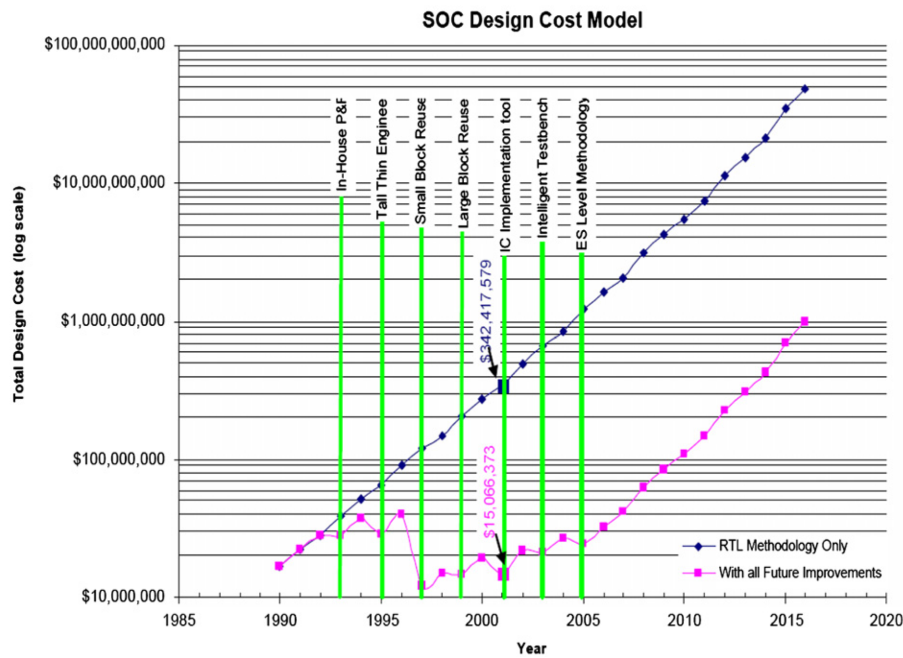

Fig. 1. Impact of design technology on system implementation cost [1].

in ISL can be halved. However, the additional area consumed by dTDMA buses such as the arbitrator and additional ports of the routers connecting the buses was not accounted for. Furthermore, dTDMA buses are designed for network-in-memory (NiM), which include interconnections only between cores and caches [14], whereas the NoC includes both communications of cores and accesses of caches that are more congested.

Third, an additional coarse-grained mesh was used for 3D designs with ISL in [12], while only fine-grained routers were used in 2D cases and 3D cases without ISL. The performance gain could come from the additional design instead of ISL itself. If the coarse-grained mesh is used, the performance gain can also be observed for 2D cases and 3D cases without ISL.

3D stacking is one of the most promising enabling technologies for die reuse. In 3D stacking, separate KGDs are bonded together by vertical interconnections. The KGDs can be manufactured at different foundries. There are technologies facilitating the reuse of dies. A concept of generic memory layer that would be adaptable to the needs of many 3D designs was introduced [15]. Re-distribution layers (RDLs) have been adopted to handle the interface issue of connections, which used micro bumps and RDLs to achieve the connection between adjacent dies [16].

In this work, the reuse granularity is expanded from component to die by virtue of a novel architecture GNet, which uses generic network service die (GNSD) to construct an NoC. The GNSDs are designed, manufactured and tested in prior, and thus known good dies (KGDs) that can be stacked with other dies into 3D IC directly. The GNet can be reused in different designs. A flexible, scalable and configurable network for the architecture is designed. To the authors' knowledge, this is the first architecture to reuse dies in 3D IC.

Cost is often the dominant factor for decision making on architecture selection and identifying the optimal design among various design options. Dong [17] proposed a fabrication cost estimation method for 3D ICs, and demonstrated the optimum partitioning strategy for future 3D IC designs through case studies of many-core microprocessor designs. A conclusion was drawn that when core number of a many-core microprocessor is large (more than 16 in [17]), the cost breakdowns of 2D designs are higher than those of 3D designs. Therefore, the comparison is herein performed only between GNet and the general 3D implementations since the 2D designs are expected to be phased out by 3D designs in future.

A comprehensive cost model, integrating design cost model, reuse model and fabrication model, is herein introduced to evaluate different architectures from the design phase to the fabrication phase. By using this cost model to evaluate the costs of various GNet designs, a conclusion is drawn that the lowest cost 3D IC would obey the following rule: the areas of individual dies should be around an optimal area. Through experiments, the approximate optimal area is identified and four GNSDs are obtained for different processes. Through comparative study on GNet and the general 3D implementation, it is found that GNet is more cost efficient than the general 3D implementation.

The rest of this paper is organized as follows. GNet – the 3D architecture for reusing generic network service die – is introduced in Section 2. Section 3 presents the cost model for evaluating architecture performance. Section 4 discusses the design optimization. The implementation details and experimental evaluations are presented in Section 5, followed by a conclusion in Section 6.

2. Architecture

A conception of the generic network service die (GNSD), which is a pre-designed and pre-verified specific die for providing general-purpose network service for a large variety of interconnection applications, is herein introduced. The architecture using a GNSD to interconnect several functional dies to construct an NoC is herein called GNet. The GNet can integrate some popular kinds of NoCs and cater for different designs. The GNSDs are KGDs that can provide generic network service in a 3D IC stack. They are designed, manufactured and tested in prior, and can be readily stacked with other dies into 3D IC directly. Moreover, they can be sold to those who need NoCs in their designs. The users of GNSDs do not need to design the network by themselves. All what they

Download English Version:

https://daneshyari.com/en/article/10364884

Download Persian Version:

https://daneshyari.com/article/10364884

Daneshyari.com