Enhanced Wafer Matching Heuristics for 3-D ICs

Vasilis F. Pavlidis, Hu Xu, and Giovanni De Micheli
Integrated Systems Laboratory, EPFL, Lausanne, Switzerland
{vasileios.pavlidis, hu.xu, giovanni.demicheli}@epfl.ch

Pre-bond test has been identified as a vital step for the wafer level integration of 3-D ICs [1], [2]. The data obtained during this step can guide the subsequent manufacturing stages to improve the functional or parametric yield of the 3-D stack. The existing methods, however, do not relate directly the performance of the resulting circuits with sales revenues. More importantly, methods that consider the distribution of speed of the assembled 3-D stacks neglect the partition of the critical path delay across the layers of the stack. In other words, a physical layer that does not include any critical path does not primarily determine the performance of the system. Consequently, for a method that aims at maximizing the profit that can be made from a 3-D system, this layer should be treated differently.

A heuristic is proposed to maximize the profit obtained by *wafer-to-wafer* (W2W) matching, where the speed of the resulting 3-D circuits is considered in the matching process. To this end, the delay distribution of the 3-D circuits is mapped to a price profile. The objective of the heuristic is to maximize profit, although this may incur a loss in functional yield. This loss can be counterbalanced by a shift in the distribution of 3-D stacks toward the faster (*i.e.,* more profitable) bins, thereby increasing the total revenues. Another important parameter used to guide the matching procedure is related to the partition of the design across the layers. To consider the partition of the critical path(s) among the layers of the stack a vector a_{coef}, which contains the individual contributions of each layer in the total path delay is employed. For n layers, $a_{coef} = [a_1 \ldots a_n]$, where $a_i \leq 1$ and $\sum_{i=1}^{n} a_i = 1$.

Prior to exploring any wafer matching techniques, specific assumptions on the available information related to the wafers are needed. In the simulations performed, a Gaussian delay distribution with mean μ and standard deviation σ is assumed. For modeling purposes, the number of dies in each wafer is determined through analytic expressions [4]. The parameters used in these expressions are D_0 = 0.5 defects/cm^2, and the defect clustering parameter is a = 0.5. A clearance radius of 3 mm, for the 300 mm^2 wafers considered herein, is also assumed. The size of the wafer repositories is set to the typical size of 25 wafers. The critical path partition a_{coef}, the die area A_{die}, and the number of layers are the variables used to explore the efficiency of the partition-driven heuristic. In this abstract results for two design partitions of a four layer 3-D circuit are considered, where a_{coef1} = [0.5 0.5 0.0 0.0] and a_{coef2} = [0.1 0.0 0.8 0.1].

Four heuristics are compared. The case of agnostic wafer matching is not used as the reference case, since blind matching produces the lowest yield or profit. The compared approaches include: *i)* the heuristic in [1] used to improve the functional yield of a 3-D stack (denoted as "FY"), *ii)* the technique in [3] for W2W integration, employing the notion of left precedence (denoted as "LP"), *iii) die-to-wafer* (D2W) integration considering the delay distribution of the dies similar to [2] and where this approach is extended to manage more than two layers (denoted as "PY") and is adapted to support wafer level integration, and *iv)* the new partition-driven heuristic (notated as "PD") where maximizing profit based on the speed of the dies is the objective.

As shown in Fig. 1, although the increase in profit is limited for all heuristics, the improvement offered by PD can be more than 1.5× as compared to PY, which is the second best approach. For a larger circuit area in Fig. 1b, the PD achieves an average increase in profit of 7.34% and up to 9.58% for linear, quadratic, and exponential price profiles, respectively.

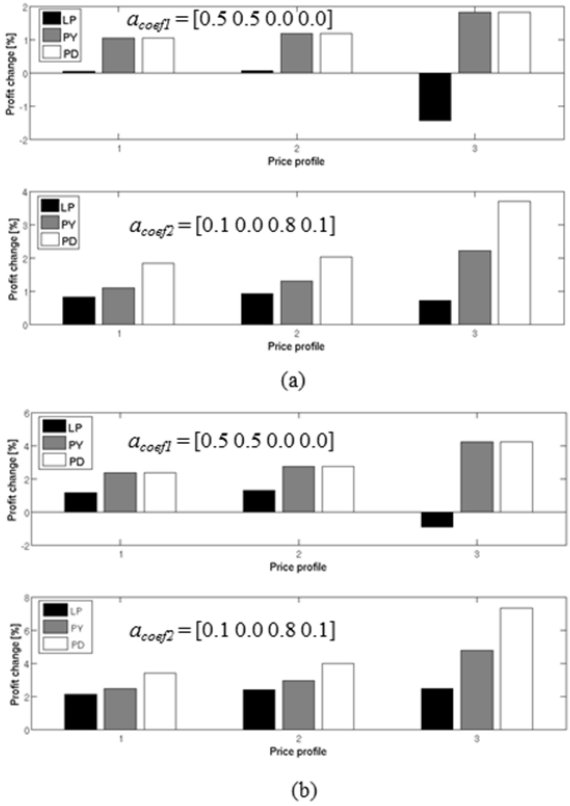

Fig. 1 Comparison of the proposed heuristic (PD) with existing approaches (FY, LP, PY), where the importance of considering the design partition is depicted. The FY is used as the reference method. The 3-D circuit consists of four layers and the circuit area is (a) 50 mm^2 and (b) 100 mm^2, respectively.

This work is funded in part by the Swiss National Science Foundation (No. 260021_126517), European Research Council Grant (No. 246810 NANOSYS), and Intel Braunschweig Labs, Germany.

REFERENCES

[1] J. Verbree, E. J. Marinissen, P. Roussel, and D. Velenis, "On the Cost Effectiveness of Matching Repositories of Pre-Tested Wafers for Wafer-to-Wafer 3D Chip Stacking," *Proceedings of the IEEE European Test Symposium*, pp. 36-41, May 2010.

[2] C. Ferri, S. Reda, and R. I. Bahar, "Strategies for Improving the Parametric Yield and Profits of 3D ICs," *Proceedings of IEEE International Conference on Computer Aided Design*, pp. 220-226, November 2007.

[3] S. Reda, G. Smith, and L. Smith, "Maximizing the Functional Yield of Wafer-to-Wafer 3-D Integration," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* Vol. 17, No. 9, pp. 1357-1362, September 2009.

[4] D. K. de Vries, "Investigation of Gross Die Per Wafer Formulas," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 18, No. 1, pp. 136-139, February 2005.

978-1-4673-0697-3/12/$31.00 ©2012 IEEE