

I/O Clustering in Design Cost and Performance Optimization for Flip-Chip Design

Hung-Ming Chen¹, I-Min Liu², Martin D.F. Wong³, Muzhou Shao⁴, and Li-Da Huang⁵

⁵Texas Instruments, Austin, TX 78714

⁴Synopsys Inc., Mountain View, CA 94043

²Cadence Design Systems Inc., San Jose, CA 95134

³ECE Department, University of Illinois at Urbana-Champaign, Urbana, IL 61801

¹Department of Electronic Engineering, National Chiao Tung University, Hsinchu, Taiwan

ABSTRACT

I/O placement has always been a concern in modern IC design. Due to flip-chip technology, I/O can be placed throughout the whole chip without long wires from the periphery of the chip. However, because of I/O placement constraints in design cost and performance, I/O buffer planning becomes a pressing problem. During the early stages of circuits and packaging co-design, I/O layout should be evaluated to optimize design cost and to avoid product failures.

In this paper, our objective is to better an existing/initial standard cell placement by I/O clustering, considering design cost reduction and signal integrity preservation. We formulate it as a minimum cost flow problem minimizing $\alpha W + \beta D$, where W is the I/O wirelength of the placement and D is the total voltage drop in the power network. The experimental results on some MCNC benchmarks show that our method achieves better timing performance and averagely over 30% design cost reduction when compared with the conventional design rule of thumb popularly used by circuit designers.

1. INTRODUCTION

With today's advanced integrated circuits (ICs) manufacturing technology in deep submicron (DSM) environment, we can integrate entire electronic systems on a single chip (SoC). Since more I/Os are needed in current designs, I/O placement has been a major concern in designing high-performance ICs. Flip-chip and multi-chip module (MCM) technologies now allow high-performance ICs and microprocessors to be built with many more I/O connections than in the past [6, 5], among which area-array bonded connection (Fig. 1) is considered a better choice [20, 15]. Since area-array style allows I/O buffers to be placed anywhere on the die, we need to be aware of I/O buffer placement constraints to better the design. Another con-

sideration in modern methodology is the cost for placing I/O buffer blocks in a design.

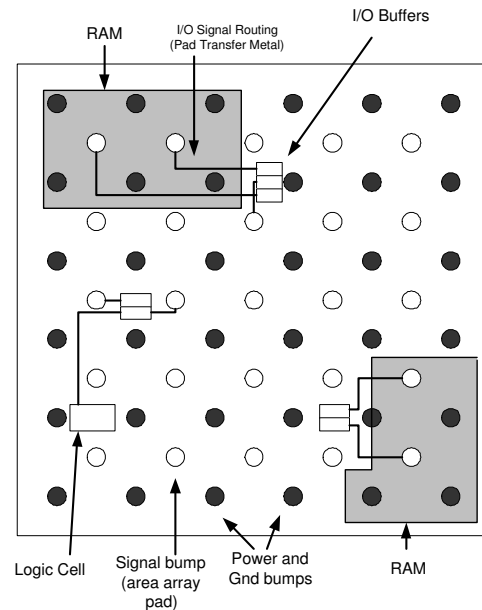


Fig. 1. Area-array footprint ASIC. The Vdd and Gnd bumps are uniformly distributed across the die with signal bumps in fixed interspersed locations. I/O buffers are associated with some specified signals bump and connected by pad transfer metal.

There were some approaches/methodologies for this problem. In [3, 22, 10, 24], similar methodologies for I/O cell placement and electrical checking using flip-chip technology have been presented. They also have graphic or interactive I/O placement tool to provide some constraints checking, trying to avoid hot-spot problem. Recently, [11] further developed a greedy algorithm to place I/O buffers in an ILP formulation of voltage drop constraint. In [13], they utilized area I/O flip-chip packaging for minimizing interconnect length, which is a major metric for cell and I/O placement optimization. How-

ever, those approaches failed to consider the building cost of I/O buffer blocks.

I/O buffers usually come with peripheral circuitry such as testing logic and electrostatic discharge(ESD). Thus there is a clearance region for standard cells outside of I/O buffers. With I/O buffers clustered in one slot, the clearance region is shared. In addition, power routing design is usually done with special care. With I/O buffers clustered, the design cost for power routing is reduced. If we just place I/O buffers in greedy ways [10, 21], more I/O buffer blocks will be generated, thus the design cost will be increased. Therefore, during the early stages of co-design of circuits and packaging[16, 18], the quality of I/O layout should be emphasized in design flow [7].

In this paper, we study the problem of I/O clustering for flip-chip design and propose an algorithm to solve the problem with respect to design cost and performance optimization while preserving signal integrity. We formulate it as a min-cost maximum flow problem minimizing $\alpha W + \beta D$, where W is the I/O wirelength of the placement and D is the total voltage drop in the power network. This can be used in post placement optimization or interim/evaluation step in performance-driven placement methodology.

The rest of the paper is organized as follows. Section 2 describes the I/O placement considerations and problem formulation. The algorithm for I/O clustering in design cost and performance optimization is presented in Section 3. Experimental results are presented in Section 4 and Section 5 concludes the paper.

2. AREA ARRAY I/O BUFFER PLACEMENT IN DESIGN COST CONSTRAINED AND PERFORMANCE-DRIVEN PLACEMENT METHODOLOGY

In order to keep up the performance in technology advances, concurrent design of chip packaging and VLSI systems is applied to satisfy system specification and to optimize the design cost [7, 16, 18]. Flip-chip technology allows high-performance ICs and microprocessors to be built with many more power and I/O connections than in the past. In order to completely take advantage of this technology, we need to focus on the placement of highly power hungry buffers, namely I/O buffers.

The design will suffer mainly from hot-spot problem [11] and long interconnect length [13] if not carefully planning I/O buffers. From the footprint of ASIC in area-array design (Fig. 1 from [3]), I/O buffers are placed near signal bumps, one I/O buffer is connected to one signal bump. Those buffers also need to be placed near power bump to consume power, in order to avoid large IR drop¹ and long interconnections. Fur-

¹In this paper we only discuss the IR drop constraint for I/O placement, however other devices in VLSI design also have this constraint.

thermore, some areas can not be used for placing I/O buffers, such as RAMs. [11] lists some of the primary I/O placement constraints, mainly keeping voltage drop below the threshold in power sources when placing I/O buffers.

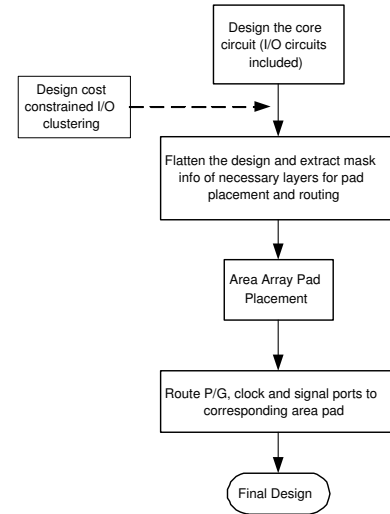


Fig. 2. Intrinsic area-array pad placement and routing flow from [21], and proposed clustering step.

On the other hand, generating minimal number of I/O buffer blocks is another major objective during cell placement for flip-chip design. If we can cluster I/O buffers, the clearance region for testing logic and ESD purposes can be shared, and power routing design cost can be reduced. Otherwise we will face more I/O buffer blocks by using greedy and intuitive approach, and the design cost is inevitably increased. Therefore, we need to find a way to handle the tradeoff between power distribution constraint violation, wirelength estimation, and design cost. Below we discuss I/O buffer placement for flip-chip design and problem formulation. Note that we can add this approach to an existing design flow in [21] to present a more complete methodology in design cost and performance optimization (Fig. 2).

A. I/O Buffer Placement for Flip-Chip Design

The analysis of the effect of I/O placement on the performance of power grids requires modeling the grids as well as the power sources and drains [17, 11]. For efficient analysis of power supply network, power grids are modeled as linear RC networks, power sources are modeled as simple constant voltage sources, and power drains are modeled as independent time-varying currents (Fig. 3).

The behavior of the system can be expressed in the modified nodal analysis (MNA) [19] formulation as the following ODE:

$$Gx + C\dot{x} = u(t) \quad (1)$$

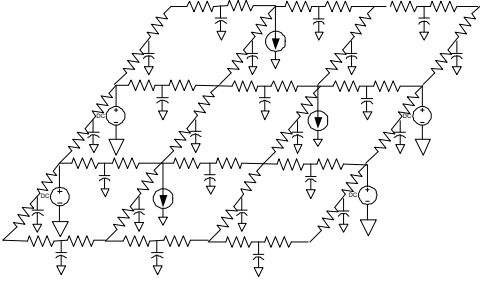


Fig. 3. Power supply network in area-array design for efficient analysis. Power grids are modeled as linear RC networks, power sources are modeled as simple constant voltage sources, and power drains are modeled as independent time-varying currents.

where x is a vector of node voltages and source currents, G is the conductance matrix, C includes the capacitance terms, and $u(t)$ includes the contributions from the sources and the drains. Applying backward euler (BE) numerical integration, we can express the resultant linear equations as:

$$Ax(t+h) = u(t+h) + x(t)C/h \quad (2)$$

where $A = G + C/h$. The system matrix A can be shown to be symmetric, and further reformulated to be nonsingular M -matrix [2]. Since the DC solution is a prudent, conservative, and practical approach to the problem, the system equation becomes $Ax = u$, where $A = G$. What we care is the voltage drop in power grids, so we reformulate the equation as $A\delta = b$, where $\delta = V - x$ is the vector of voltage drops and b is the vector of current sources. In other words, b can be seen as the following:

$$b_i = \sum_{k=1}^n d_{ik} I_k, \forall i \quad (3)$$

where I_k is the current associated with buffer io_k and $d_{ik} = 1$ if io_k consumes the power from node p_i , $d_{ik} = 0$ otherwise, n is the number of I/O buffers. Therefore, the relationship between the voltage drop at node p_j and all the entries of the vector b is:

$$\delta_j = \sum_{i=1}^m a_{ji}^{-1} b_i, \forall j \quad (4)$$

where a_{ji}^{-1} is the element on the row j and column i of the inverse of system matrix A^{-1} , m is the number of nodes. The problem can be formulated as placing a given set of I/O buffers while suppressing the voltage drop to be under the user-specified voltage drop thresholds, denoted by δ_{max} .

B. Problem Formulation

Problem 2.1 ICDCPO (I/O Clustering in Design Cost and Performance Optimization): Given an existing/initial stan-

ard cell placement, a set of I/O buffers (which has corresponding set of signal bumps) $IO = \{io_1, \dots, io_n\}$ and the current I_i associated with I/O buffer io_i , a set of power bumps $P = \{p_1, p_2, \dots, p_m\}$, a user-specified voltage drop threshold vector δ_{max} , the system matrix A for power network, a certain building cost for I/O buffer blocks, and a set of nets $N = N_1 \cup N_2 \cup \dots \cup N_k$, find a solution to simultaneously suppress the design cost, the I/O wirelength for the placement, and voltage drop threshold violation for power network.

We divide the whole die into bins based on power bumps. Each bin has a certain amount of area for accommodating I/O buffers, obtained from the dead space or other pre-planned free space in existing placement and the building cost of buffer block. For some bins which are occupied fully or partially by memory blocks, the area of corresponding bins will be zero or much less than a certain amount. Thus we use P to represent the set of power bump bins as well. We define $H = \{h_1, \dots, h_n\}$ to be the set of regions that the buffer io_i can possibly draw current from (shown in Fig. 4), similar to [12]. Each region contains a set of power bumps that the corresponding I/O buffer can use.

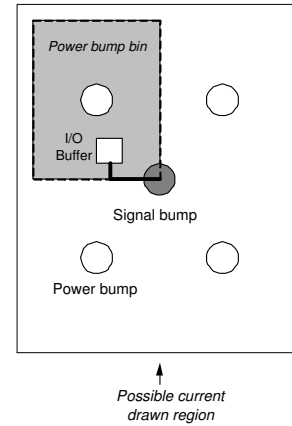


Fig. 4. The relationship between signal bump, power bump, power bump bin, I/O buffer possible positions, and possible current drawn region.

In next section, we introduce a cost function to minimize the I/O wirelength and total voltage drop in power network, and present an algorithm to solve the proposed problem. Note that the I/O wirelength we mention in this paper is wirelength estimation of connecting I/O buffer, signal bump, and I/O port of corresponding logic cells. It is not total wirelength estimation for the whole placement.

3. THE I/O CLUSTERING ALGORITHM IN DESIGN COST AND PERFORMANCE OPTIMIZATION

We first construct a network with embedded cost function and run a min-cost flow algorithm [1] to obtain the solution.

The network graph $G = (V, E)$ is constructed as follows, also see Fig. 5 for illustration.

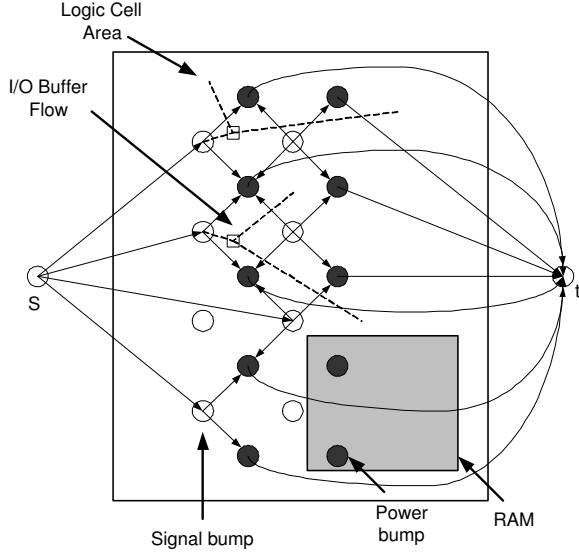


Fig. 5. Network construction for ICDCPO. Some signal bump (corresponding I/O buffer) vertex io_i only connects to power bump bin vertices which are inside the possible current drawn region for io_i . Note that for each power bump bin vertex there is a specified capacity indicating the number of I/O it can accommodate. The dashed lines in the figure represent the connection between I/O buffers, signal bumps, and other logic cells. We use the corresponding wirelength (I/O wirelength) as a part of cost function.

1. $V = \{s, t\} \cup IO \cup P$, where s is the source vertex, t is the sink vertex. IO and P are defined in the problem.
2. $E = \{(s, io_i) | io_i \in IO\} \cup \{(io_i, p_j) | io_i \in IO, p_j \in P \cap h_i\} \cup \{(p_j, t) | p_j \in P\}$, where h_i is the corresponding possible current drawn region for io_i .
3. Edge capacity: $U(s, io_i) = 1, U(io_i, p_j) = 1, U(p_j, t) =$ upper bound of the number of I/O buffers that bin p_j can accommodate, computed from the dead space or other pre-planned free space in the placement.
4. Cost function: $C(io_i, p_j) = \left\lceil \alpha W_{ij} + \beta I_i \sum_{k=1}^m a_{kj}^{-1} \right\rceil$, where W_{ij} is the I/O wirelength estimation for I/O buffer io_i placing at bin p_j (along with the computation with other internal logic modules or cells²), a_{kj}^{-1} is the element on the row k and column j of the inverse of system matrix A^{-1} . For other edge $e \in E, C(e) = 0$.

Any flow in the network can be mapped into an I/O clustering solution for a subset of given I/O buffers. If a flow f exists and $|f| = n$, we can assign all I/O buffers to buffer blocks in

²This is only an estimation from the center of the bin to other modules or cells, not including the effect of further change in total wirelength after I/O buffer placement.

given power bump bins. And since the cost of the flow is the cost for the solution of I/O buffer placement, minimum cost flow guarantees a solution with minimum total cost $\alpha W + \beta D$, where W is the I/O wirelength and D is the total voltage drop in power network. The total capacities of edges going from source vertex s is n , so the maximum flow $|f_{max}| = n$. We have the following theorems to show the effectiveness and integrality property [1] of min-cost maximum flow, and present the proposed algorithm for solving ICDCPO.

Theorem 3.1 A min-cost flow f in G corresponds to an I/O clustering solution to ICDCPO problem with minimum total cost: $\alpha W + \beta D$. A min-cost maximum flow assigns all I/O buffers in IO with minimum total cost.

Theorem 3.2 If all edge capacities and supplies/demands of nodes are integers, the min-cost flow problem always has an integer min-cost flow.

Algorithm for solving ICDCPO

- 1 Construct the network graph G .
- 2 Assign capacities U and cost C .
- 3 Apply min-cost maximum flow algorithm on G .
- 4 Derive the corresponding I/O clustering solution.

Finding a min-cost maximum flow in a network is a classical problem for which several polynomial-time optimal algorithms are available [9, 1]. We use capacity scaling algorithm to solve the network in $O((m \lg U)(m + n \lg n))$ time [1], where $n = |V|$, $m = |E|$, and U is the upper bound of the edge capacity.

Here we have presented an approach to clustering I/O buffers. Since we estimate utilizable space for I/O buffer blocks in power bump bins, we need to move part of the existing cells around to accommodate the blocks. We can either use overlap removal in [8], applying bisection technique, or use mixed mode placement like [23], treating I/O buffer blocks as small macros.

4. EXPERIMENTAL RESULTS

We have implemented our algorithm and run on 650MHz Pentium III machine. The existing cell placements based on some MCNC benchmarks (in Table I) are obtained from the placer *FENG SHUI* [14], with aspect ratio 1.0.

We have adopted the following abstract model of I/O regimes from [4] for our experiments:

- I/O buffers must be placed exactly at pad locations, and any I/O buffer can be placed at any pad location.

TABLE I
NUMBER OF CELLS, NETS, AND I/O TERMINALS IN SOME MCNC
STANDARD CELL PLACEMENT BENCHMARKS.

Benchmark	Cells	Nets	I/Os
struct	1952	1920	64
biomed	6514	7052	97
industry1	3085	2594	814
industry2	12637	13419	495

- No two I/O buffers can occupy the same location.
- For a design with I/O buffers and a rectangular core layout region, we fix pad locations with an array of locations spaced uniformly within the core layout region.

The number of power bumps and signal bumps are scaled from IBM SA-27E area-array copper technology [3].

Our approach has been compared with a conventional design rule of thumb popularly used by circuit designers [21], which is to greedily minimize wirelength and IR-drop when placing I/O buffers. To be more specific, the area-array pads are placed at fixed sites on the top layer and each of the I/O ports is routed to the closest pad. In this way, all I/O buffers can have the least signal integrity constraint violations and the I/O wirelength is minimized.

Table II shows the experimental results on MCNC benchmarks summarized in Table I. The voltage drop threshold violation (VDTV) percentage shown in the table is obtained by the number of nodes whose voltage drop exceeds the threshold normalized by total number of nodes. From the design cost comparison, we obtain much less number of I/O buffer blocks (average 31.5% reduction) with slight increase percentage of VDTV in power nodes. The gain in design cost reduction is due to the insertion of I/O buffer block building cost in solving the problem. The reason of slight increase in VDTV is that we use worst case IR-drop estimation, assuming all buffers draw maximum current at the same time, but the situation virtually never arise in reality. In practice, designers will waive such small violations.

The table also shows the I/O wirelength comparison results, and the wirelength estimation has been described in Section 3. We obtain better I/O timing performance by smaller I/O wirelength. The tradeoff coefficients α and β are used based on the importance of the two objectives. Here we adjust the coefficients so that these two terms are about equal weights. As can be seen in the table, more I/O wirelength reduction has been reported for more I/O buffers (industry1), indicating the better performance of the proposed approach in large number of I/Os.

5. CONCLUSION

In this paper, we have presented an I/O clustering step in design cost and performance optimization for high-end flip-chip design. We formulate the problem as a min-cost maximum flow problem and the experimental results are encouraging. With slight increase percentage of voltage drop threshold violation, we can automate the I/O buffer block generation, obtaining better timing performance and much less design cost.

REFERENCES

- [1] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin. “*Network Flows - Theory, Algorithms, and Applications*”. Prentice-Hall, 1993.
- [2] A. Berman and R.J. Plemmons. “*Nonnegative Matrices in the Mathematical Sciences*”. Academic Press, NY, 1996.
- [3] P.H. Buffet, J. Natonio, R.A. Proctor, Y.H. Sun, and G. Yasar. “Methodology for I/O cell Placement and Checking in ASIC Designs Using Area-Array Power Grid”. In *IEEE Custom Integrated Circuits Conference*, pages 125–128, 2000.
- [4] A.E. Caldwell, A.B. Kahng, S. Mantik, and I.L. Markov. “Implications of Area-Array I/O for Row-Based Placement Methodology”. In *IEEE Symposium on IC/Package Design Integration*, pages 93–98, 1998.
- [5] L. Cao and J.P. Krusius. “A New Power Distribution Strategy for Area Array Bonded ICs and Packages of Future Deep Sub-Micron ULSI”. In *IEEE Electronic Components and Technology Conference*, pages 1138–1145, 1997.
- [6] A. Chandrakasan, W.J. Bowhill, and F. Fox, editors. “*Design of High-Performance Microprocessor Circuits*”. IEEE Press, 2001.
- [7] K.-Y. Chao and D.F. Wong. “Signal Integrity Optimization on the Pad Assignment for High-Speed VLSI Design”. In *Proceedings IEEE International Conference on Computer-Aided Design*, pages 720–725, 1995.
- [8] W. Choi and K. Bazargan. “Incremental Placement for Timing Optimization”. In *Proceedings IEEE International Conference on Computer-Aided Design*, 2003.
- [9] T.H. Cormen, C.E. Leiserson, and R.L. Rivest. “*Introduction to Algorithms*”. The MIT Press, 1990.

TABLE II
EXPERIMENTAL RESULTS OF OUR APPROACH ON MCNC BENCHMARKS SUMMARIZED IN TABLE I, COMPARED WITH A GREEDY APPROACH. WITH SLIGHT INCREASE PERCENTAGE IN VOLTAGE DROP THRESHOLD VIOLATION, DESIGN COST (DC) REDUCTION CAN BE OBTAINED WITH LESS I/O WIRELENGTH.

Benchmark	# power bump	# signal bump	Greedy			I/O Clustering			Comparison	
			Time (sec)	VDTV (%)	# i/o buf block	Time (sec)	VDTV (%)	# i/o buf block	dc imp (%)	I/O WL (x)
struct	306	272	<1	1.96	60	<1	1.96	52	15.4	0.62
biomed	342	306	<1	1.1	86	1	1.1	78	10.3	0.99
industry1	1116	1050	1	0.1	669	32	0.98	455	47.0	0.002
industry2	676	625	<1	0.1	405	11	1.1	264	53.4	0.98
Average				0.82			1.28		31.5	

- [10] R. Farbarik, X. Liu, M. Rossman, P. Parakh, T. Basso, and R. Brown. "CAD Tools for Area-Distributed I/O Pad Packaging". In *IEEE Multi-Chip Module Conference*, pages 125–129, 1997.
- [11] J.N. Kozhaya, S.R. Nassif, and F.N. Najm. "I/O Buffer Placement Methodology for ASICs". In *IEEE International Conference on Electronics, Circuits and Systems*, pages 245–248, 2001.
- [12] I-Min Liu, H.-M. Chen, T.-L. Chou, A. Aziz, and D.F. Wong. "Integrated Power Supply Planning and Floorplanning". In *Proceedings IEEE Asia and South Pacific Design Automation Conference*, pages 589–594, 2001.
- [13] R.J. Lomax, R.B. Brown, M. Nanua, and T.D. Strong. "Area I/O Flip-Chip Packaging to Minimize Interconnect Length". In *IEEE Multi-Chip Module Conference*, pages 2–7, 1997.
- [14] P.H. Madden. "Reporting of Standard Cell Placement Results". *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 21(2):240–247, February 2002.
- [15] V. Maheshwari, J. Darnauer, J. Ramirez, and W.W.-M. Dai. "Design of FPGAs with Area I/O for Field Programmable MCM". In *Proceedings ACM Symposium on Field Programmable Gate Arrays*, pages 17–23, 1995.
- [16] Joel Mcgrath. "Chip/Package Co-Design: The bridge between chips and systems". In *Advanced Packaging*, June 2001.
- [17] S.R. Nassif and J.N. Kozhaya. "Fast Power Grid Simulation". In *Proceedings ACM/IEEE Design Automation Conference*, pages 156–161, 2000.
- [18] J.C. Parker, R.J. Sergi, D. Hawk, and M. Diberardino. "IC-Package Co-Design Supports Flip-Chips". *EE Times*, November 2003. <http://www.eedesign.com/story/OEG20031113S0055>.
- [19] L.T. Pillage, R.A. Rohrer, and C. Visweswariah. "*Electronic and System Simulation Methods*". McGraw-Hill, 1995.
- [20] P.A. Sandborn, M.S. Abadir, and C.F. Murphy. "The Tradeoff Between Peripheral and Area Array Bonding of Components in Multichip Modules". *IEEE Transactions on Components, Packaging, and Manufacturing Technology - Part A*, 17(2):249–256, June 1994.
- [21] C. Tan, D. Bouldin, and P. Dehkordi. "Design Implementation of Intrinsic Area Array ICs". In *Proceedings 17th Conference on Advanced Research in VLSI*, pages 82–93, 1997.
- [22] G. Yasar, C. Chiu, R.A. Proctor, and J.P. Libous. "I/O Cell Placement and Electrical Checking Methodology for ASICs with Peripheral I/Os". In *IEEE International Symposium on Quality Electronic Design*, pages 71–75, 2001.
- [23] H. Yu, X. Hong, and Y. Cai. "MMP: A Novel Placement Algorithm for Combined Macro Block and Standard Cell Layout Design". In *Proceedings IEEE Asia and South Pacific Design Automation Conference*, pages 271–276, 2000.
- [24] P.S. Zuchowski, J.H. Panner, D.W. Stout, J.M. Adams, F. Chan, P.E. Dunn, A.D. Huber, and J.J. Oler. "I/O Impedance Matching Algorithm for High-Performance ASICs". In *IEEE International ASIC Conference and Exhibit*, pages 270–273, 1997.