

Congestion Mitigation During Placement

Kanad Chakraborty and Natesan Venkateswaran

kanad/natesan@us.ibm.com

IBM Server Development Division, Hopewell Junction, NY 12533

Abstract- High post-placement congestion in complex ASICs and microprocessors may pose severe constraints on the wiring resources, thereby causing wireability, timing and noise problems. *Linear* wirelength-based mincut partitioning algorithms have some built-in advantages for reducing congestion. We present a mathematical model of congestion and experimentally investigate various congestion mitigation techniques used in conjunction with linear wirelength-based placement. The experimental results validate our congestion model. Our placement tool, CPlace[©], is a clustering-based mincut partitioner that optimizes a linear wirelength objective.

I INTRODUCTION

This paper addresses the importance of various techniques for minimizing congestion during placement of ASICs, built using deep submicron technology. As is well known [1, 4], the minimization of congestion during placement is of prime importance in trying to (a) satisfy timing constraints, (b) meet noise-related design requirements, and (c) achieve wireability after placement.

II CONGESTION

A Modeling of congestion

Congestion can be modeled using the concepts of *wiring demand* and *wiring supply*.

The wiring demand in a region of a chip or a macrocell is a function of (a) the number of pins in that region; (b) blockage due to circuits¹ belonging to that region; (c) connection between pins in that region; (d) connection between pins outside the region to those inside; and (e) connections that pass through the region without touching any pin in that region.

We define the *wiring demand* $D_{w_i}(S)$ of a wiring layer w_i in a region S of a chip or a macrocell as the minimum number of tracks of w_i needed to bring signals into and out of all the circuits present in S .

Suppose the circuits present in region S are labeled 1 through C_S . Since one wiring track on wiring layer w_i is needed for every pin belonging to w_i in each of the circuits, D_{w_i} is obtained by summing the wiring demand for pins with the following types of wiring demands: (a) the utilization of the wiring resources due to circuits, assuming the absence of any nets, (b) the utilization, over all nets, of wire segments in S of a net j with all or some pins in S , and (c) the utilization, over all nets, of wire segments in S of a net j that runs through S but has no pin inside S .

¹In this paper, we use the term 'circuit' or 'cell' interchangeably to mean either a leaf cell, macrocell, or a random-logic block.

Expressed mathematically:

$$D_{w_i}(S) = \sum_{j=1}^{C_S} U_{w_i}(S, j) + \sum_{j=1}^{C_S} \sum_{k=1}^{P_j} F(j, k) L_{w_i}(S, j, k) + \sum_{j=1}^N H(S, j) M_{w_i}(S, j) \quad (1)$$

In the above equation, $U_{w_i}(S, j)$ denotes the number of tracks of layer w_i utilized by the circuit j in S ; P_j denotes the number of external pins of circuit j ; $F(j, k) = 1$ iff pin k of circuit j belongs to wiring layer w_i , otherwise $F(j, k) = 0$; $L_{w_i}(S, j, k)$ denotes the number of tracks of layer w_i utilized by wire segments connecting to pin k of circuit j in region S ; N represents the total number of nets; $H(S, j) = 1$ iff net j satisfies condition (c) above, and $H(S, j) = 0$ otherwise; $M_{w_i}(S, j)$ denotes the number of tracks of w_i utilized by wire segments of net j in region S .

We define the *wiring supply* $R_{w_i}(S)$ of a wiring layer w_i in a region S of a chip or a macrocell as the total number of wiring tracks of w_i present in S . This total wiring resource in a region is a constant, unaffected by movement of circuits into or out of the region.

Expressed mathematically:

$$R_{w_i}(S) = T_{w_i}(S) \quad (2)$$

In the above equation, $T_{w_i}(S)$ denotes the total number of tracks of wiring layer w_i in region S . Congestion of a layer w_i in S is given by the fraction $D_{w_i}(S)/R_{w_i}(S)$, and can be expressed as a percentage.

B Congestion mitigation techniques used in placement

Congestion-driven placement and congestion-aware wiring have been investigated by various researchers [1, 2, 3]. In this work, a clustering-based mincut partitioning tool, CPlace, that tries to minimize the *linear* wirelength is used. It is observed that a linear wirelength objective captures congestion more effectively [4] than a quadratic wirelength objective. The main problem with a quadratic wirelength objective is that it tries to minimize the lengths of long wires *at the expense* of short ones. This is illustrated by a simple example: suppose two wires are of lengths l_1 and l_2 , where $l_1 < l_2$. Then $(l_2 - x)^2 + l_1^2 < (l_1 - x)^2 + l_2^2$, where x is the amount by which either of the wires is shrunk. Therefore, to a quadratic algorithm, there is a better payoff in shortening a long wire by an amount x as compared to shortening a short wire by the same amount. This type of behavior would cause a strongly connected cluster to be spread out, thereby causing congestion and wiring difficulties for the router. Three congestion mitigation techniques that can be used with CPlace are as follows:

Testcase	Technique used	Wirelength (mm)	Avg. Edge Congestion		Max. Edge Congestion	
			Horizontal	Vertical	Horizontal	Vertical
Test1	Global	9106.369 (1.00)	1309 (1.00)	1307 (1.00)	1816 (1.00)	2120 (1.00)
	Global+SF	9294.246 (1.02)	1340 (1.02)	1331 (1.02)	1840 (1.01)	2140 (1.01)
	Global+Detailed	8588.449 (0.94)	1229 (0.94)	1237 (0.95)	1667 (0.92)	2033 (0.96)
Test2	Global	20563.739 (1.00)	1535 (1.00)	1860 (1.00)	2115 (1.00)	2872 (1.00)
	Global+SF	21040.783 (1.02)	1574 (1.02)	1901 (1.02)	2093 (0.99)	2761 (0.96)
	Global+Detailed	19218.701 (0.93)	1446 (0.94)	1722 (0.94)	2030 (0.96)	2665 (0.93)

Note: The relative performance w.r.t. **Global** are shown in parentheses.

TABLE I
Results illustrating the comparison among three congestion mitigation techniques

1. **Global placement using linear mincut partitioning (Global):** In this approach, the linear wirelength objective is used together with a clustering-based mincut partitioning algorithm. This generates a legal placement (i.e. without overlaps).
2. **Global placement as in (1) above, together with spread factor (Global+SF):** In this technique, a constraint is imposed on the number of circuits that can be mapped to each region during global placement, so as to allow adequate wiring supply to each region.
3. **Global placement as in (1) above, followed by detailed placement (Global+Detailed):** Detailed placement after the global phase is the iterative improvement phase that repeats circuit movements such as module swaps and module insertions to reduce the wirelengths, thereby improving wireability.

III EXPERIMENTS

The three techniques have been benchmarked using several testcases. Each testcase denotes either a full chip (ASIC) or a macrocell within an ASIC, containing a mix of large and small leafcells, macrocells, and random-logic blocks. All these testcases represent actual chip designs in various stages of progress within IBM. Most of our testcases have high percentage utilization due to circuits and are thus good candidates for studying congestion mitigation techniques. Results for two sample testcases are shown in Table I. Test1 consists of 31359 circuits and 32402 nets, with 77% utilization. Test2 consists of 110969 cells and 109539 nets, with 65% utilization.

The goal of a congestion mitigation technique is to ensure that the wiring supply meets the wiring demand across all regions in the layout. Per Equation (2), wiring supply is a constant. With regard to the three components of the wiring demand (denoted as (a), (b) and (c) in Section 2.1), mapping fewer circuits per region (**Global + SF**) could cause an overall increase in the wiring demand by decreasing (a) and increasing (b) and (c). In **Global+Detailed**, in order to decrease the wirelength, the detailed placement minimally perturbs the global placement and usually moves circuits around within local regions. Therefore, it could only cause a small increase in (a), but a significant reduction of (b) and (c). This has the effect of

causing an overall reduction in wiring demand, thereby lowering the congestion.

IV CONCLUSION

Our mathematical modeling and experimental results lead us to conclude that global followed by detailed placement consistently produces very promising results in mitigating congestion. We are presently investigating ways to achieve accurate and realistic technology-dependent congestion modeling; and use congestion mitigation techniques to meet design constraints (timing, wireability, noise etc.) at various stages of chip design.²

REFERENCES

- [1] P. N. Parakh, R.B. Brown, and K.A. Sakallah, "Congestion-driven quadratic placement," *Proceedings of the Design Automation Conference 1998 (DAC '98)*, pp. 275-278.
- [2] S. Mayrhofer, et al., "Congestion-driven placement using a new multi-partitioning heuristic," *Proceedings of the IEEE International Conference on Computer-Aided Design 1990 (ICCAD '90)*, pp. 332-335.
- [3] C.L.E. Cheng, "RISA: accurate and efficient placement routability modeling," *Proceedings of the IEEE International Conference on Computer-Aided Design 1994 (ICCAD '94)*, pp. 690-695.
- [4] C.J. Alpert, et al., "Faster minimization of linear wirelength for global placement," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 17, no. 1, January 1998, pp. 3-13.

²We are grateful to Wilm Donath and Prabhakar Kudva of IBM T.J. Watson Research Center, Yorktown Heights, for reviewing our paper.