Timing-Driven Placement using Design Hierarchy Guided Constraint Generation [*]

Xiaojian Yang   Bo-Kyung Choi   Majid Sarrafzadeh

Computer Science Department, University of California, Los Angeles, CA 90095

xjyang,bkchoi,majid@cs.ucla.edu

Abstract

Design hierarchy plays an important role in timing-driven placement for large circuits. In this paper, we present a new methodology for delay budgeting based timing-driven placement. A novel slack assignment approach is described as well as its application on delay budgeting with design hierarchy information. The proposed timing-driven placement flow is implemented into a placement tool named Dragon (timing-driven mode), and evaluated using an industrial place and route flow. Compared to Cadence QPlace, timing-driven Dragon generates placement results with shorter clock cycle and better routability.

1. Introduction

Timing-driven placement is one of the most important steps to meet circuit performance in VLSI design. The problem has been studied for two decades, yet it remains challenging because of the dramatically increasing circuit size and the dominance of the interconnect delay in the deep sub-micron design.

Recent works on wirelength optimization suggested that hierarchical [1] or multi-level [2, 15] approach is indispensable to efficiently solve large-scale placement problem. Correspondingly, timing-driven placement problem using hierarchical or multi-level framework should draw research attention. Several recent works addressed this problem. Ou et al. [13] adopted a net-cut control method in min-cut placement. Halpin et al. [8] presented a linear programming based net length control for recursive bisection placement. Kahng et al. [9] extended the placement flow in [1] to incorporate direct minimization of the critical path. The above approaches shows the trend of the combination between traditional timing-driven approach and top-down placement flow. However, delay budgeting, as a previously effective approach, has not been applied into the top-down placement framework.

In this paper, we study the delay budgeting problem in hierarchical or multi-level placement flow. Our contributions in this paper are: First, we have designed a new model to describe the criticality of the interconnect in hierarchical multi-level placement. Second, we have extended the previously known zero-slack assignment algorithm to achieve flexibility based on the proposed criticality model, and finally we have implemented a timing-driven placement tool which aims at solving large-scale placement problems. It combines the delay budgeting approach and a new multi-level placement flow.

The rest of this paper is organized as follows. Section 2 gives preliminaries. Section 3 describes our novel model of net criticality for interconnection, In Section 4, we introduce a multi-level placement flow which considers the global interconnects of different design hierarchies and use this information to guide the slack assignment of the placement. We show our experimental results in Section 5, and conclude in Section 6.

2. Preliminaries

2.1 Timing constraint graph

We use the similar model as in [12, 14]. A *circuit* corresponds to a hypergraph $G(V, E)$, which consists of a set of cells $V = \{v_1, v_2, \ldots v_n\}$

and a set of nets $E = \{e_1, e_2, \ldots, e_m\}$. We form a *directed acyclic graph (DAG)* based on the hypergraph. Each node in the DAG corresponds to a non-sequential cell, an I/O pad, or an input/output pin of a sequenstial cell. We define the *source* nodes as the input pads or the output pins of the sequential cells, and the *sink* nodes as the output pads or the input pins of the sequential cells. There is a direct edge from node v_i to v_j if the output of cell v_i connects to the input of cell v_j. We assume that each cell has only one output, otherwise the node is duplicated to ensure the single output.

For each node v_i in the DAG, we define *arrival time* a_i and *required time* r_i as the latest input arrival time and earliest input required time of the cell which corresponds to node v_i, respectively.

We associate a *slack* s_i to node [1] v_i as the additional delay of v_i. Thus,

$$a_i + d_i + s_i \leq a_j \quad \forall \quad v_i \in V, \quad v_j \in FO_i \tag{1}$$

where FO_i is the set of all the fanouts of node v_i, and d_i is the internal delay of the cell corresponding to node v_i. Here we assume that the internal delay from different input pins to the output pin are the same. We will interchangeably use slack of a node and slack of a net in this paper.

Assuming that the arrival time of source nodes are zero, and the timing constraint of the circuit is T, we have

$$a_i = 0 \quad \forall \quad v_i \in V_{source} \tag{2}$$

$$a_i \leq T \quad \forall \quad v_i \in V_{sink} \tag{3}$$

where V_{source} and V_{sink} are the set of source nodes and sink nodes, respectively.

The delay budgeting problem seeks to assign non-negative values to slacks. The assignment is *feasible* if the constraints of (1), (2) and (3) are satisfied.

The solution of delay budgeting problem can be converted into delay bounds for interconnects. Specifically, if we assume that a linear delay model[2] is adopted, the slack s_i of node v_i indicates that the maximum delay of the net driven by v_i is no larger than s_i. A placement is *timing feasible* if all nets satisfy the delay bounds converted from a feasible slack assignment (and thus the entire circuit satisfies the timing constraint). The conversion of delay bounds from slacks depends on the delay model, driving strength of the cell, and load characteristics.

2.2 Delay model

In this work, we use the delay model shown in Fig. 1.

$$d = I_i + r(c_e + c_j) + r_e c_j$$

where I_i is the intrinsic delay of gate i, r is the driver resistence of gate i, c_e and r_e are the capacitance and resistance of net e, and c_j is the input pin capacitance of gate j. If gate i drives multiple gates, the summation of c_j for all fanout gates replaces c_j.

We use wireload model to estimate the resistance of a net. Half perimeter of the bounding box is the estimated wirelength for a net. The

---

[1]It is also associated to the net which node v_i drives, under the assumption that each node has one output and the net delay is calculated using the bounding box of the net.

[2]The delay of a net is proportional to the half-perimeter of the bounding box of the net.
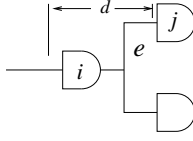
Fig. 1: Delay model used in this paper

intrinsic delays of the gates are not provided in our library file. Instead, we use the result from linear regression on the data in library look-up table.

3.   Slack Assignment

3.1   Previous work

Previous approaches on slack assignment include zero-slack assignment (ZSA) [12], weighted slack assignment [16], fast slack allocation [11], assignment with maximum flexibility [14], and maximum independent set algorithm (MISA) [3]. Efficient slack allocation for large circuits is still a hard problem, and it is the crucial step in delay budget based timing-driven placement.

In delay budgeting problem, the goal of slack assignment is to achieve a set of delay bounds for all nets. It is therefore desirable to assign non-zero delay to all nets. On the other side, evenly distributing slacks (ZSA) ignores net criticalities, thus over-constrains longer wires while giving freedom to shorter wires.

A number of approaches have been proposed to obtain uneven slack distribution. In [16] the slacks are assigned proportionally to the weights of the nets, which are derived from the circuit characteristics. The same method can be used in [12]. In [7, 14], $log()$ was adopted as an objective function to measure the quality of the slack assignment. In this work, we propose a new objective function which tries to better describe the slack requirement in placement problem.

3.2   A new objective function for slack assignment

In slack assignment, it is natural to expect that all nets are assigned to non-zero slacks, otherwise the delay budgets corresponding to this slack assignment will never be satisfied during the placement. Therefore, for a given net in the slack assignment problem, the maximum gain is obtained when we increase the slack from zero. However, if we keep increasing the slack for this net, the gain from the newly assigned slack decreases. If the slack of a net is beyond a certain threshold, there is no gain from allocating more slack on the net, i.e., the gain is zero. We seek for a good objective function to describe this feature. Previously used $log()$ is close but not an ideal candidate. We propose a new objective function:

$$f(x) = 1 - e^{-cx} \tag{4}$$

where c is a constant determined by the criticality of the net, and x is the slack assigned to this net.

Comparing this new objective function (4) with previously used $log(x)$, one can see that there is an upper bound for it but not for $log(x)$. This upper bound refrains from excessive slack allocation to a single net. Introducing the objective function helps modeling and solving slack assignment problem.
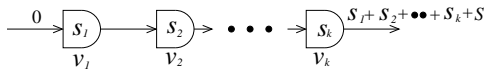
3.3   Slack assignment for paths

Fig. 2: Slack assignment problem in the single path

Consider a slack assignment problem for a single path (Fig. 2). There are k nodes on the path. Each node v_i is originally assigned a slack s_i.

The total (incremental) slack to be assigned along this path is S. For node v_i assigned with slack x_i, the gain of the slack is $1 - e^{-c_i x_i}$ where c_i is the coefficient of the objective function of node v_i, which will be described in Section 4.3. We want to solve the following problem:

$$\text{maximize} \quad f(\mathbf{x}) = \sum_{i=1}^{k} \left(1 - e^{-c_i(s_i + x_i)}\right)$$

$$\text{such that} \quad \sum_{i=1}^{k} x_i = S.$$

Let

$$g(x) = f(\mathbf{x})|_{x_k = S - x_1 - x_2 - \ldots - x_{k-1}}.$$

Then solving the problem is equal to solving the following system:

$$\begin{cases} \partial g(\mathbf{x})/\partial x_1 & = & 0 \\ \partial g(\mathbf{x})/\partial x_2 & = & 0 \\ \cdots \\ \partial g(\mathbf{x})/\partial x_{k-1} & = & 0 \\ x_1 + x_2 + \ldots + x_k & = & S \end{cases}$$

This system can be directly solved as:

$$x_k = \frac{S - \sum_{i=1}^{k-1} \left((\ln(c_i/c_k) - c_i s_i + c_k s_k)/c_i\right)}{1 + \sum_{i=1}^{k-1} (c_k/c_i)},$$

$$x_i = \frac{\ln(c_i/c_k) - c_i s_i + c_k s_k}{c_i} + \frac{c_k}{c_i} x_k \quad (1 \leq i \leq k-1).$$

The proposed slack assignment approach has the following features:

- It assigns slacks to the nodes based on their weights (which are reflected by c_i). Nodes with higher weights (thus smaller c_i in the objective function) will be assigned larger slacks.

- It can be extended to trees instead of paths. The similar linear system can be formed and solved, while it is not clear how to handle trees using proportional assignment method [16].

- It is fast and can be used in incremental assignment. Previously assigned slacks can be increased or decreased, constructing more reasonable slack distribution. Additional constraints can ensure positive incremental slacks. In our implementation we set s_i to zero for $i = 1, \ldots, k$, ensuring the slack assignment approach converges fast.

3.4   Modified ZSA slack assignment

For the entire circuit, we use modified ZSA to assign slacks. The algorithm starts from a timing analysis and computation of slacks for each node. Then a path with minimum positive slack on each node is identified. The proposed approach is then applied to optimally assign slacks on this path. The slack assigned to the node is added to its delay. An incremental timing analysis is followed and the new slacks for each node are updated. This completes one iteration and the next minimum slack path is identified.

4.   Multi-Level Placement with Predefined Hierarchy

In this section, we propose a new timing-driven placement flow based on delay budgeting method. Consider a top-down placement using recursive quadrasection approach (Fig. 3). We name the wires connecting different partitions by the first quadrasection *level-1* nets. Similarly, the wires connecting different partitions by the second quadrasection but not *level-1* nets are called *level-2* nets, and so on. It is obvious that,

after non-timing-driven placement, higher level nets (lower level number) have longer average length than lower level nets. It is therefore necessary to take the net length into consideration in delay budgeting process. Previous delay bound approaches have not considered the effect of global nets in the slack assignment.
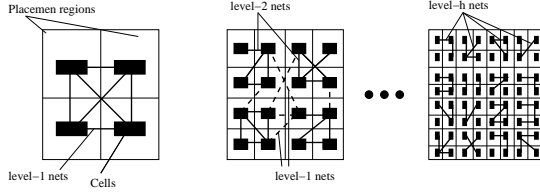

Fig. 3: Global nets of different level in top-down quadrasection placement

In delay budgeting for placement using net bound, it is desirable to assign larger slacks to potentially longer nets, as long as the slack assignment is feasible. If slacks are evenly distributed to nets with different lengths, the delay bounds will over-constraint the placement process and in some cases even cause the placement failure on finding a feasible solution.

4.1 Achieving Hierarchies

The authors in [17] proposed a new methodology of multi-level placement flow. The main method is to partition the circuit into small size clusters, and then execute multi-level placement based on the known design hierarchy. We use the similar flow in this work. At the beginning of placement, a recursive bisection step is applied on the circuit. After every two bisections (i.e. one quadrasection), the global nets at this level are identified and labeled. The bisection step stops when the average number of cells for each cluster is less than a threshold. The remaining unlabeled nets are labeled as the highest level number. The net level labels will guide the delay budgeting process.

4.2 Estimating Net Lengths

Estimating the length of external nets in hierarchical placement has been studied previously. We use the same method as [6, 5]. In a hierarchical placement flow, the average length of the *level-k* nets can be estimated as the following:

$$\overline{l_k} = \frac{4(\frac{4\lambda}{3} - \frac{1}{3\lambda}) + 4\lambda}{6}$$

where $\lambda = 2^{H-k}$ and H is the total number of hierarchical levels.

The estimated wirelength of the nets are converted into net criticality, and then the weight in delay budgeting algorithm.

4.3 Delay Budgeting considering Design Hierarchy

We use the following function to convert the estimated net length to the node weight[3] for slack assignment:

$$c_i = \frac{\max_{v_j \in V}(\sqrt{l_j \cdot f_j})}{\sqrt{l_i \cdot f_i}} \quad (6)$$

where l_i is the estimated wirelength and f_i is the fanout of the net which node v_i drives. In general, a lower hierarchical level number or a larger fanout indicates a longer net. Therefore a lower coefficient c_i is given for allocating a larger slack in the later delay budgeting algorithm.

4.4 Multi-level Placement using net bounds

In this modified multi-level placement flow, we use simulated annealing approach to improve the placement. At each level, the cells are clustered based on their hierarchical labels and these clusters are placed into

---

[3]The node which drives the net will be assigned the weight

rectangular placement regions. The clusters can be entirely exchanged with other clusters during the annealing. The cost function is:

$$WL + \lambda_1 \sum_{i=1}^{|E|} max(l_i - b_i, 0)$$

where l_i is the current length and b_i is the bound of net i. WL is the total bounding box wirelength. λ_1 is a parameter for adjustment between wirelength optimization and delay optimization. Both items in the cost function can be incrementally updated when two clusters are exchanged.

The entire placement flow is shown in Fig. 4. We use hMetis [10] as the partitioning tool. The original circuit is recursively bipartitioned till the pre-calculated level. Global nets of each level are identified and are assigned to an estimated wirelength. The local nets within partitioned clusters are regarded as the *level-H* nets. The coefficient c in (4) is determined by the estimated wirelength and the net fanout using (6). Then the modified ZSA algorithm is called to assign slacks according to the new objective function. Such slacks are then converted into delay bounds in terms of net length.

Next, delay bounds are used to evaluate the placement quality in the multi-level placement flow. At each level, the optimization step tries to reduce the combined cost of total bounding box wirelength and total delay violation. The placement meets the timing constraints if the total delay violation is zero. In general a lower delay violation value corresponds to a better placement in terms of delay.
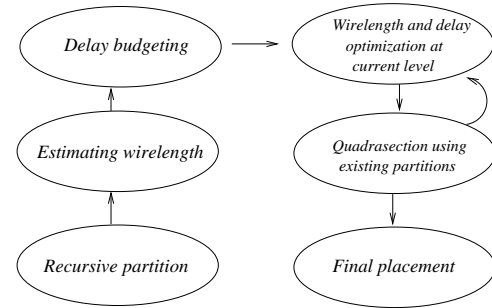

Fig. 4: Multi-level placement flow with a prior hierarchy information

5. Experimental Results

We have implemented (a) the proposed new slack assignment approach, (b) a simulated annealing engine with cost function including delay bound violation, and (c) the multi-level placement flow with pre-defined hierarchy information and built a timing-driven Dragon. To evaluate our approach, we test our placement tool using an industrial place and route flow. The results after global and final routing are reported, and they are compared to the industrial place and route outputs.

Specifically, we compare our placement tool with Cadence QPlace (Silicon Ensemble 5.3). Both Dragon and QPlace read the same LEF/DEF files and the outputs are fed into Cadence WRoute. After global and final routing, we use Pearl timing analysis tool to show the timing result. The final routed wirelength and the minimum slack are reported.

The statistics of the benchmarks are summarized in Table 1. The circuits are acquired from a website provided by [4]. We use a $0.18\mu m$ standard-cell library[4] as the LEF file. We determine the clock cycles for the circuits by the following way. For each circuit we first use QPlace (non-timing-driven mode) to place it. After routing and timing analysis, the minimum cycle period is recorded as the cycle of the design. Therefore for each circuit, the minimum slack by QPlace non-timing-driven

---

[4]Downloaded from Artisan Components Inc. website under academic research program.

mode is zero (see Table 2). This clock cycle is then used as the constraint for QPlace timing-driven mode.

circuits	cells	nets	rows	routing layers	clock cycle
matrix	3,083	3,200	56	4	3.89
VP2	8,714	8,789	100	4	4.57
32-MAC	8,902	9,115	101	5	3.85
64-MAC	25,616	26,017	134	5	7.67

Table 1: Tested circuit statistics, including number of cells, number of nets, number of rows, number of routing layers and clock cycle (ns).

Table 2 shows the results from four different placement approaches: QPlace non-timing-driven, QPlace timing-driven, Dragon with ZSA delay budgeting, and Dragon with the new slack assignment. Dragon with hierarchy based slack assignment consistantly achieves the best among four approaches, including QPlace in timing-driven mode[5].

Placement		Results	Circuit			
			matrix	VP2	32-MAC	64-MAC
QPlace	NTD	WL	126618	496905	511763	2163585
		slack	0.00	0.00	0.00	0.00
	TD	WL	124980	517596	520468	unroutable
		slack	-0.25	-0.19	0.18	unroutable
TD Dragon	ZSA	WL	118823	430470	470711	2618864
		slack	0.00	0.12	0.11	-1.32
	New	WL	118833	424379	482252	2655655
		slack	0.09	0.28	0.46	0.97

Table 2: Routed wirelength (micron) and minimum slack (ns) of four placement runs: QPlace non-timing-driven(NTD), QPlace timing-driven(TD), Dragon with ZSA delay budgeting, and Dragon with the new slack assignment. "unroutable" shows the routing failure. New slack assignment is the method proposed in Section 3.

Fig. 5 shows the comparison between QPlace's results and ours[6]. QPlace's results are from timing-driven mode. Dragon uses design hierarchy guided new objective function in slack assignment. Our results are better than QPlace's except for the wirelength of 64-MAC.

6. Conclusion

In this paper, we have presented a new methodology of timing-driven placement. A novel slack assignment approach is described as well as its application on delay budgeting with design hierarchy information. The proposed timing-driven placement flow is evaluated within an industrial place and route flow. The advantage of the new methodology has been confirmed by experimental results.

There are a number of unclear issues in delay budgeting based timing-driven placement approach. One open problem is determination of the timing constraint during the placement. Neither very tight nor very loose constraint is appropriate for delay optimization. It is useful to find the proper value without accurate delay model and routing information. Another issue is the exact cost function for delay violation. Questions such as whether we should penalize larger violations remain unanswered.

7. Acknowledgments

---

[5]For two benchmarks QPlace timing-driven mode creates worse delays than non-timing-driven mode. The reason is possibly that the tight constraint mis-guides the optimization during placement.

[6]In terms of running time, Dragon is about 10 - 15 times slower than QPlace, including the whole step of slack assignment and timing-driven placement.
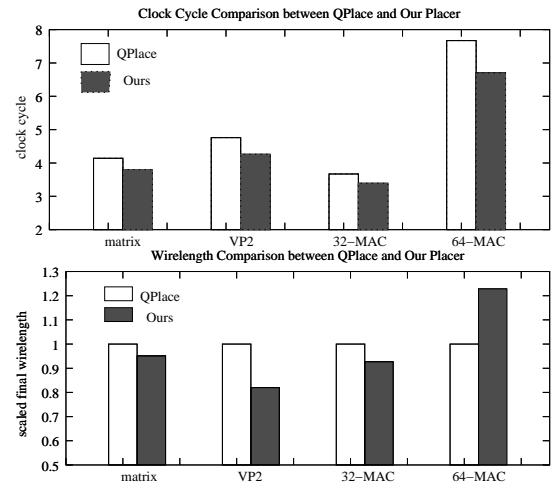

Fig. 5: Clock cycle period and routed wirelength comparison between QPlace and Dragon. Wirelengths are scaled with QPlace's result as unity. The QPlace's non-timing-driven result is used to replace timing-driven result for 64-MAC.

8. References

[1] A. E. Caldwell, A. B. Kahng, and I. L. Markov. "Can Recursive Bisection Alone Produce Routable Placements?". In *Design Automation Conference*, pages 477–482. IEEE/ACM, June 2000.

[2] T. F. Chan, J. Cong, T. Kong, and J. R. Shinnerl. "Multilevel Optimization for Large-Scale Circuit Placement". In *International Conference on Computer-Aided Design*, pages 171–176. IEEE, 2000.

[3] C. Chen, X. Yang, and M. Sarrafzadeh. "Potential Slack: An Effective Metric of Combinational Circuit Performance". In *International Conference on Computer-Aided Design*, pages 198–201. IEEE, 2000.

[4] Y. C. Chou and Y. L. Lin. "A Performance-Driven Standard-Cell Placer Based on a Modified Force-Directed Algorithm". In *International Symposium on Physical Design*, pages 24–29. ACM, April 2001.

[5] P. Christie and D. Stroobandt. "The Interpretation and Application of Rent's Rule". *IEEE Transactions on VLSI Systems*, 8(6):639–648, 2000.

[6] W. E. Donath. "Placement and Average Interconnection Lengths of Computer Logic". *IEEE Transactions on Circuits and Systems*, 26(4):272–277, April 1979.

[7] T. Gao, P. M. Vaidya, and C. L. Liu. "A New Performance Driven Placement Algorithm". In *International Conference on Computer-Aided Design*, pages 332–335. IEEE/ACM, 1991.

[8] B. Halpin, C. Y. Chen, and N. Sehgal. "Timing Driven Placement using Physical Net Constraints". In *Design Automation Conference*, pages 780–783. IEEE/ACM, 2001.

[9] A. B. Kahng, S. Mantik, and I. L. Markov. "Min-Max Placement For Large-Scale Timing Optimization". In *International Symposium on Physical Design*, pages 143–148. ACM, April 2002.

[10] G. Karypis, R. Aggarwal, V. Kumar, and S. Shekhar. "Multilevel Hypergraph Partitioning: Application in VLSI Domain". In *Design Automation Conference*, pages 526–529. IEEE/ACM, 1997.

[11] W. K. Luk. "A Fast Physical Constraint Generator for Timing Driven Layout". In *Design Automation Conference*, pages 626–631. IEEE/ACM, 1991.

[12] R. Nair, C. L. Berman, P. S. Hauge, and E. J. Yoffa. "Generation of Performance Constraints for Layout". *IEEE Transactions on Computer Aided Design*, 8(no.8):860–874, Aug 1989.

[13] S. L. Ou and M. Pedram. "Timing-driven Placement Based on Partitioning with Dynamic Cut-net Control". In *Design Automation Conference*, pages 472–476. IEEE/ACM, June 2000.

[14] M. Sarrafzadeh, D. A. Knol, and G. E. Tellez. "A Delay Budgeting Algorithm Ensuring Maximum Flexibility in Placement". *IEEE Transactions on Computer Aided Design*, 16(11):1332–1341, 1997.

[15] M. Wang, X. Yang, and M. Sarrafzadeh. "Dragon2000: Fast Standard-cell Placement for Large Circuits". In *International Conference on Computer-Aided Design*, pages 260–263. IEEE, 2000.

[16] H. Youssef, R. Lin, and E. Shragowitz. "Bounds on Net Delays for VLSI Circuits". *IEEE Transactions on Circuits and Systems*, 39(11):815–824, November 1992.

[17] C-C. Zhang, J. Cong, Z. Pan, and X. Yuan. "Physical Hierarchy Generation with Routing Congestion Control". In *International Symposium on Physical Design*, pages 36–41. ACM, April 2002.