Recursive bisection placement algorithm with the predicted wirelength

Hao Jie, Ma Hong, Peng Silong

(National ASIC Design Engineering Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: To obtain a better placement result, a partitioning-based placement algorithm with wirelength prediction called HI+PI is presented. A new method is proposed to estimate proximity of interconnects in a netlist, which is capable of predicting not only short interconnects but long interconnects accurately. The predicted wirelength is embedded into the partitioning tool of bisection-based global placement, which can guide our placement towards a solution with shorter interconnects. In addition, the timing objective can be handled within the algorithm by minimizing the critical path delay. Experimental results show that compared to Capo 0.5, mlPL 6, and NTU Place, HI+PI outperforms these places in terms of wirelength and run time. The improvements in terms of average wirelength over Capo 0.5, mlPL 6 and NTU Place are 13%, 3%, and 9% with only 19%, 91%, and 99% of their runtime, respectively. By integrating the predicted wirelength-driven clustering into Capo 0.5, the placer is able to reduce average wirelength by 3%. The timing-driven HI+PI can reduce the critical path delay by 23%.

Key words: hierarchy, interconnect placement, VLSI circuit wirelength prediction

Placement is an important step in the overall IC design process in deep submicron technologies. A lot of algorithms have been proposed in the past 30 years, including partitioning-based methods, analytical methods, iterative methods, etc. Partitioning-based placement algorithms determine cell locations by recursively dividing an initial region with successive bisections or quadrisections. Advances in partitioning research have provided a number of fast algorithms which produce extremely good results.

The placement problem can be driven by different objectives, such as timing, routability, thermal distribution, or a combination of them. The classical placement objective function is the total wirelength, which correlates well with global routing resource demand. Interconnection prediction is very important for early feasibility studies in design flow. During the past two decades, different wirelength prediction approaches have been proposed because of the need for early interconnect optimization in the design flow to achieve timing closure. Rent’s rule has been a basic tool to estimate the wirelength. In Refs. [9-11], wirelengths are estimated based on circuit characteristics. Most of the estimation techniques provide estimates of just the average wirelength. We develop an individual wirelength prediction based on circuit characteristics to estimate the wirelength of a layout design in advance before placement. In fact, the final wirelengths depend on placement algorithm’s wirelength predictions that are accurate for one placement algorithm may be inaccurate for another. In our placement tool, the predicted wirelength as a main optimization objective is embedded into placement flow. Experimental results show that the wirelengths of final placement can trend to predicted wirelengths. In this paper, we propose a novel partitioning-based placement algorithm for standard cells. The major contributions of this paper can be summarized as follows:

1) We find that the basic circuit characteristics (e.g. net degree and net area) and node level can be used to predict wirelengths in the final layout. These prelayout measures demonstrate good correlations with postlayout interconnect lengths. We propose to couple the wirelength predictions with our placement flows.

2) In conventional partitioning-based placements, partitioning tools are typically done with the min-cut objective. In order to reduce the lengths of intra-cluster interconnects, we introduce a new objective function that incorporates a predicted wirelength component for partitioning tools. Experiments show that we can obtain a better wirelength result with little loss in time.

1 Overall of Placement Tool

Fig 1 shows that our placement framework consists of four stages. We predict individual wirelengths for all nets based on circuit characteristics in stage 1. In stage 2, we partition both the chip region and the circuit netlists recursively by horizontal and vertical cut lines. In the partitioning phase, wirelength-driven clustering is performed to reduce the wirelengths of intra-cluster interconnects. After clustering, for better min-cut, wirelength-driven refinement is executed without loss of wirelength quality. The subcircuits after partitioning

![Fig 1] Framework of our placement tool
are assigned to rectangular bins. A bin-based simulated annealing is performed to improve the current placement. The final step simply spreads overlapped cells and makes local improvements to obtain the detailed placement.

2] Wirelength Prediction

In this section, we explain our wirelength prediction technique in detail. A circuit netlist can be modeled by a hypergraph \( H(V, E) \) with a node (cell) set \( V = \{ v_i \mid i = 1, 2, \ldots, n \} \) and a hyperedge (net) set \( E = \{ e_j \mid j = 1, 2, \ldots, m \} \). Each net \( v_i \) is a subset of \( V \) with cardinality \( |v_i| \geq 2 \). An edge \( (s, t) \in e_j \) is an output of source node \( s \) and an input of sink node \( t \), which is a subset of the hyperedge. The degree of the node \( v_i \), denoted by \( d(v_i) \), is the number of nets incident to it. The degree of net \( e_j \), denoted by \( d(e_j) \), is the number of nodes incident to it. Each node is associated with an area cost, \( area(v_i) \). Net area is the total area of nodes belonging to that net, namely, \( area(e_j) = \sum area(v_i) \).

Individual wirelength is dependent on three factors in this paper: 1) The net degree and the net area, which are the bases of the wirelength; 2) The shape distribution of a node level; 3) The range of a node level. Wirelength prediction is performed in three phases in which we calculate three different weights for all the nets according to the above three factors. The final predicted wirelength is obtained by combining these weights.

2.1] Basic wirelength

Obviously, the net with a larger net degree and net area tends to have a larger fan-out range. It is intuitively that larger fan-out nets usually correspond to longer net lengths. We use net degree and net area of net as its basic wirelength. The area factor of net \( e \) is computed by

\[
A(e) = \sum \frac{area(v_i)}{d(e)}
\]

The basic wirelength of net \( e \) is the combination of the area factor and the number of nodes belonging to the net and is given as

\[
L_{\text{base}}(e) = 1 + \frac{A(e)}{A_{\text{max}}} \cdot \frac{d(e)}{d_{\text{max}}}
\]

where \( A_{\text{max}} \) is the largest node area amongst all nodes; \( d_{\text{max}} \) is the largest net degree over all nets.

2.2] Shape distribution of node level

Define level \( (v) \), the level of node \( v \), as the maximum topological depth over all directed paths beginning at a primary input (primary input) and terminating at node \( v \). \( \text{level}(v) \) is the number of nodes at level \( (v) \). Fig 2 shows the shape of the mj73 circuit in the MCNC (Microelectronics Center of North Carolina) benchmark. In Fig 2, the number of nodes at level 1 is maximal and the placement tool needs more sites for the nodes at level 1. So the wirelength of nets with nodes at level 1 may dilate during the placement phase. The nets connecting the nodes with larger \( \text{level}(v) \) will have larger fan-out ranges in the final layout. The factor \( DNL(e) \) used to measure the dilatability due to the shape distribution of node levels for net \( e \) can be calculated as

\[
DNL(e) = \sum \frac{\text{level}(v_i)}{d(e)}
\]

For example, referring to Fig 3, \( \text{num. level}(u) = 1 \), \( \text{num. level}(x_i) = \text{num. level}(x_j) = 5 \), \( \text{num. level}(x_k) = 2 \).

Since \( DNL(u, x_i, x_j) \) is larger than others, it indicates that net \( (u, x_i, x_j) \) has a stronger dilatability than others.

2.3] Range of node level

If node levels are within a limited range, the wirelength of the output net will likely to be less than that of a net with widely distributed node levels. To capture the ranges of the node levels of net \( e \), we define the factor \( RNL(e) \) as

\[
RNL(e) = \sum \frac{\text{level}(v_i) - \text{level}(v_j)}{d(e)}
\]

If the nodes connecting the net have the same node level, the RNL factor of the net is 0. In Fig 3, from our definitions, \( RNL(u, x_i) = 1 \), \( RNL(u, x_j) = 2 \). That is, the metric, net \( (u, x_i) \) has a longer wirelength than net \( (u, x_j) \). Note that sometimes the RNL and DNL are approximately “orthogonal” when the RNL factor of a net is 0, the DNL factor of the net may be very large.

The predicted result of the apex2 circuit is shown in Fig 4. In these figures, on the \( x \)-axis are net id and on the \( y \)-axis are net lengths. The solid line represents predicted w
lengths, and the points represent the actual wirelengths (in percentages) in the final placement using our HJ-P1. The shape distribution of node level DNL, and the range of node level RNL, in estimation of circuits are presented in Figs 4 (b) and (c). As shown in Fig. 4, wirelengths of long interconnects increase rapidly with an increase in RNL. DNL is more efficient for short interconnects.

![Wirelength Distribution](image)

**Fig. 4** Prediction results of apex2. (a) L\textsubscript{svp} vs actual wirelength; (b) DNL vs actual wirelength; (c) RNL vs actual wirelength; (d) Connectivity vs actual wirelength; (e) Edge separability vs actual wirelength

2.4 Individual wirelength

- The wirelength of net e, L\textsubscript{svp}, is its basic wirelength L\textsubscript{base} + a \times L\textsubscript{base} \times DNL + b \times L\textsubscript{base} \times RNL,

\begin{align}
L\textsubscript{svp} (e) &= a \times L\textsubscript{base} (e) \times DNL (e) + b \times L\textsubscript{base} (e) \times RNL (e) \\
L\textsubscript{svp} (e) &= L\textsubscript{base} (e) \\
\end{align}

where a, b are parameters that control the trade-off between DNL and RNL. L\textsubscript{base} is the largest L\textsubscript{base} (e) amongst all nets.

Most proposed wirelength predictions have poor results for long interconnects. Figs 4 (d) and (e) are the predicted results using connectivity and edge separability, which are not sensitive for long connections. Fig. 4(a) shows that our approach has good predictions for long interconnects.

3 Recursive Bisection-Based Global Placement

- Recursive bisection-based placement algorithms seek to decompose a given placement instance into smaller instances by subdividing the placement region, assigning cells to subregions, reformulating constraints and cutting the netlist (see Fig. 1). The top-down placement process can be viewed as a sequence of passes where each pass examines all blocks and, if required, divides them into two smaller blocks using min-cut partitioning. Such netlist decomposition is typically done with the min-cut objective. A novel clustering-refinement multilevel partitioning algorithm by incorporating a min-wirelength objective is proposed in this paper. The global placement algorithm is as follows:

- GLOBAL PLACEMENT (H (V, E), Layout) / /Q: the queue of placement bins
- While bin size is big enough
  - do Q -> a bin;
  - Choose a (horizontal or vertical) cut-line for the bin;
  - Partitioning attempts to split each bin roughly in half;
  - Build partitioning hypergraph from netlist and cells contained in the bin;
  - Partition the bin into smaller bins using WL. PAR-
Recursive bisection placement algorithm with the predicted wirelength

\[ \text{slack}(s, t) = \text{REQ}(t) - \text{ARR}(s) - d(s, t) \quad (10) \]

where \( d(s, t) = r(s, t) (c(s, t) / 2 + c(T_s)) \), \( c(s, t) = (c_w(s, t) + c_i) l(s, t) \), \( r(s, t) = r_i l(s, t) / w(s, t) \), \( l(s, t) = L_{pp}(c) / d(e) \). \( w(s, t) \), \( c_w \), \( c_i \), and \( r_i \) are wire width, area capacitance, fringing capacitance and resistance for unit width wire, respectively; \( T_s \) is the subtree rooted at \( v \); \( c(T_s) \) is the capacitance of a dc-connected subtree in \( T \), rooted at \( T_s \). Finally, according to Eq. (10), we can obtain the net weight with timing objective:

\[ \text{weight}(c) = 1 - \sum_{s \in c} \text{slack}(s, t) \quad (11) \]

where \( s_{\text{max}} \) is the largest slack sum over all nets. For optimizing timing objectives, as the net weight \( \text{weight}(c) \) instead of \( L_{pp} \) is embedded into the timing clustering phase. The net with small slack will be protected by the timing-driven clustering, which can effectively minimize critical path delay during placement.

3.3 Wirelength-driven refinement

In the refinement phase, the IM algorithm\(^{[12]} \) is used to reduce the cuitzize. We change the original gain of IM by introducing a predicted wirelength objective to reduce the degeneration of the final wirelength. The wirelength-aware gain of IM swapping is given as

\[ g_w = \lambda L_{pp} + (1 - \lambda) g_n \quad (12) \]

where \( \lambda \) is a parameter that controls the trade-off between wirelength and the original gain of IM, which is less than 1 and larger than 0.

3.4 B-in-based swapping

After clustering for minimizing the wirelengths of intra-partition interconnects, bin-based simulated annealing is conducted to find a good location for each partition to be placed in, thus, minimizing the total wirelength between bins. There are three types of moves in bin-based simulated annealing: horizontal switch, vertical switch, and diagonal switch. These moves switch two adjacent bins.

4. Detail Placement

In the detailed placement step, overlaps between cells have to be resolved to obtain a legal placement. When placing cells to remove overlaps, we have to consider two conflicting factors: the total wirelength and the legality of result. We use a greedy heuristic to obtain a legal placement, while the local swapping tries to reduce the wirelength of the placement.

After global placement, we divide each row in the placement region into placeable segments based on the overlap of the blockages with the row. We now use a greedy heuristic to bring every row in the placement region to within its capacity by moving the standard cells. Once the rows have been brought under capacity, we move the cells among the placeable segments to satisfy their respective capacities. The cells are then assigned to legal positions within each segment.

After all overlaps are removed, greedy local improvement
is performed. We try to switch adjacent standard cells to see if total wirelength can be improved. This step will cure the wirelength loss caused by the blind cell spreading step of legalization. For each row, all the bins from left to right are traversed and an attempt is made to place in the bin onto the site array. If there are blocked sites, we ignore it and go to the right side of the chunk of blocked sites.

5. Experiment

Our placement algorithm is implemented in C programming language and compiled with gcc 4.0 on a Linux PC with an Intel Pentium IV 1.5 GHz CPU and 768 MB memory. These algorithms are tested using a set of benchmarks published in ISPD 2002, ISPD 2005, BSCAS89 and PEKO. In the first experiment, we compare HJ-P1 with three well-known academic placement tools, namely Capo10.5, mPL6 and N'TU/Place. Moreover, the wirelength-driven clustering algorithm is an individual soft package which can embed other open source tools. For the second experiment, we integrate the wirelength clustering with Capo10.5 due to its code availability. Our wirelength prediction tool estimates only HPWL in all the experiments. For the last experiment, we compare the timing-driven HJ-P1 with the original HJ-P1.

5.1 Comparison with other placers

For all experiments, we give an average of 10 runs and use the ISPD2002 benchmark with 20% total white space and random pad locations. The statistical information of benchmarks is listed in Tab. 1. Tab. 1 shows that HJ-P1 obtains a HPWL improvement of 13% versus Capo10.5 and is 5.24 times faster than Capo10.5. The runtime of Capo10.5 for the ibm16 circuit is very long. HJ-P1 (WL) reduces runtime by 10% and 9% over mPL6 and N'TU/Place with 3% and 1% shorter wirelengths, respectively.

<table>
<thead>
<tr>
<th>Tab 1</th>
<th>The resulting HPWL and runtime for different placers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circuit</td>
<td>#cell</td>
</tr>
<tr>
<td>ibm01</td>
<td>12752</td>
</tr>
<tr>
<td>ibm02</td>
<td>19601</td>
</tr>
<tr>
<td>ibm07</td>
<td>45926</td>
</tr>
<tr>
<td>ibm08</td>
<td>51309</td>
</tr>
<tr>
<td>ibm09</td>
<td>53395</td>
</tr>
<tr>
<td>ibm10</td>
<td>69429</td>
</tr>
<tr>
<td>ibm11</td>
<td>70558</td>
</tr>
<tr>
<td>ibm12</td>
<td>71076</td>
</tr>
<tr>
<td>ibm16</td>
<td>18348</td>
</tr>
<tr>
<td>ibm18</td>
<td>210613</td>
</tr>
<tr>
<td>Ratio</td>
<td>1</td>
</tr>
</tbody>
</table>

5.2 Integration with Capo

We integrate our wirelength-driven clustering with Capo10.5, which is based on min-cut placement techniques. In addition, the wirelength-driven clustering using Connectivity is embedded into Capo10.5 as a comparison with our method. Tab. 2 shows the results without and with wirelength-driven clustering based on the PEKO benchmark with 10% total white space, uniform cell sizes, and random pad locations, and the ISPD 2005 benchmarks. From Tab. 2, the original Capo10.5 averages 3% more wirelength than Capo10.5 with wirelength-driven clustering, and runs faster. The experimental results also reveal the fact that our wirelength prediction is superior to Connectivity for placement quality.

<table>
<thead>
<tr>
<th>Tab 2</th>
<th>HPWL and runtime comparison of original Capo and Capo with wirelength-driven clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circuit</td>
<td>Capo10.5 (WL)</td>
</tr>
<tr>
<td></td>
<td>HPWL / (10^6 μm)</td>
</tr>
<tr>
<td>peko01</td>
<td>1.43</td>
</tr>
<tr>
<td>peko05</td>
<td>4.11</td>
</tr>
<tr>
<td>peko10</td>
<td>14.11</td>
</tr>
<tr>
<td>peko15</td>
<td>45.92</td>
</tr>
<tr>
<td>peko18</td>
<td>38.73</td>
</tr>
<tr>
<td>adapt0c1</td>
<td>87.43</td>
</tr>
<tr>
<td>adapt0c2</td>
<td>96.89</td>
</tr>
<tr>
<td>bigble1</td>
<td>101.14</td>
</tr>
<tr>
<td>Ratio</td>
<td>1</td>
</tr>
</tbody>
</table>

5.3 Timing-driven HJ-P1

The results without and with timing-driven clustering are shown in Tab 3. After placement (using the BSCAS89 benchmark), we perform global and detailed routing, RC extraction and timing analysis using commercial tools. The electrical parameters have been chosen to resemble a typical 90 nm process. Tab. 3 shows that HJ-P1(T) with timing-driven clustering reduces the critical path delay by about 23%.
null