ABSTRACT

One of the necessary requirements for the placement process is that it should be capable of generating routable solutions. This paper describes a simple but effective method leading to the reduction of the routing congestion and the final routed wirelength for large-scale mixed-size designs. In order to reduce routing congestion and improve routability, we propose blocking narrow regions on the chip. We also propose dummy-cell insertion inside regions characterized by reduced fixed-macro density. Our placer consists of three major components: (i) narrow channel reduction by performing neighbor-based fixed-macro inflation; (ii) dummy-cell insertion inside large regions with reduced fixed-macro density; and (iii) pre-placement inflation by detecting tangled logic structures in the netlist and minimizing the maximum pin density. We evaluated the quality of our placer using the newly released DAC 2012 routability-driven placement contest designs and compared our results to the top four teams that participated in the placement contest. The experimental results reveal that our placer improves the routability of the DAC 2012 placement contest designs and effectively reduces the routing congestion.

1. INTRODUCTION

One of the most crucial objectives of modern-day VLSI placement is the minimization of the final routed wirelength on the chip. Satisfying this objective has a detrimental effect on the performance of a placer and greatly affects factors such as congestion, delay, and timing. Despite being an important objective, the routed wirelength is usually only taken into account during the post-placement steps by employing techniques for local congestion minimization. However, techniques that are based on local congestion information have a relatively small impact on the quality of the final placement.

A number of approaches have been proposed for reducing routing congestion in placement. The approach in [14] discusses the use of integer linear programming to improve local routing congestion. In [15], the congestion estimation model is based on a stochastic algorithm that computes the horizontal and vertical track usage. The approach in APlace 2.0 [9, 10] describes a congestion estimation method [11] that is based on integrating congestion information into an objective function. The routability-driven placement approach [13] solves a sequence of unconstrained optimization problems, incorporating the wire density as a weighted penalty into an objective function. The authors in [17] examine the correlation between the Steiner tree and the routed wirelength in placement, and propose the ROOSTER algorithm in order to reduce the routed wirelength. The authors of RUDY [20] introduce a model that estimates the wire distribution on the chip by taking into consideration the wire density of each net. The approach proposed in CRISP [18] employs fast global routing to determine congested regions. The approach in [7] optimizes routability using pin density, routing overflow, and macro porosity consideration. The approach proposed in simPLR [12] applies look-ahead routing in a flat placement framework to reduce congestion. The authors of Ripple [5] use cell inflation and net-based movement for routability-driven placement.

The International Symposium on Physical Design (ISPD) 2011 contest [21] was organized to enhance research in VLSI and enable advances in routability-driven placement. The contest also introduced a benchmark suite of industrial ASIC designs that contain placement and routing blockages, varying metal width and spacing, non-rectangular fixed objects, and fixed 'Not in Image' objects. These objects have pins on metal layers on top of the ones used within cells for internal pins and routing. The Design Automation Conference (DAC) 2012 routability-driven placement contest [22] was a continuation of the ISPD 2011 contest [21] and introduced a benchmark suite of modern industrial ASICs, containing information for both placement and routing. The DAC 2012 contest [22] also introduced a routability metric that describes congestion more accurately than the ISPD 2011 contest metric, as well as two academic global routers for the evaluation of the results. In contrary to the global router [19], that was used to evaluate the routability of the ISPD 2011 contest [21] circuits, the routers NCTUgr and BFG-R of the DAC 2012 contest [22] have improved performance while being machine-independent.

By targeting congestion during global placement we are able to minimize the routing congestion and the final routed wirelength. The remainder of the paper is structured as follows: Section 2 describes our placement flow, while Section 3 introduces the contest metric. Section 4 discusses our proposed technique of neighbor-based fixed-macro inflation to block narrow channels. Section 5 introduces the technique of dummy-cell insertion inside regions of reduced fixed-macro density. Section 6 describes the application of pre-placement inflation in our placer. Section 7 reports our experimental results. Section 8 concludes the paper.
2. PLACEMENT FLOW

We follow an analytical [7, 2, 16] placement approach. To estimate the routing congestion, we first decompose multi-pin nets into two-pin nets by FLUTE [3, 4]. Then, we divide the chip into global tiles and compute the congestion of each tile by taking into account the routing demand and the routing supply (separately for horizontal and vertical directions). As in [5], let \( \text{TileWidth} \) be the width (\( \text{TileHeight} \) the height) of the tile, and \( \text{WidthBB} \) be the width (\( \text{HeightBB} \) the height) of the bounding box of the two-pin net. Also, let \( \text{Ovlp} \) be the overlapping area between the tile and the bounding box of the net and \( \text{WireH} \) be the horizontal (\( \text{WireV} \) the vertical) wire area of the net. Finally, let \( \text{BlockageH} \) be the product of the ratio of horizontal (\( \text{BlockageV} \) the vertical) tracks which are occupied by each routing blockage and the overlapping area between the blockage and the tile. For the horizontal direction:

\[
\text{SupplyH} = (\text{TileWidth})(\text{TileHeight}) - \text{BlockageH},
\]

\[
\text{DemandH} = \frac{(\text{Ovlp})(\text{WireH})}{(\text{WidthBB})(\text{HeightBB})},
\]

\[
\text{CongestionH} = \frac{\text{SupplyH} - \text{DemandH}}{\text{SupplyH}}.
\]

For the vertical direction:

\[
\text{SupplyV} = (\text{TileWidth})(\text{TileHeight}) - \text{BlockageV},
\]

\[
\text{DemandV} = \frac{(\text{Ovlp})(\text{WireV})}{(\text{WidthBB})(\text{HeightBB})},
\]

\[
\text{CongestionV} = \frac{\text{SupplyV} - \text{DemandV}}{\text{SupplyV}}.
\]

We perform cell inflation to alleviate congested tiles. The inflation pattern is similar to the one described in [5]. In order to further reduce congestion, we propose blocking narrow channels on the chip by inflating fixed-macros that create such channels. We also propose inserting dummy-cells locally on the placement area, inside regions of reduced fixed-macro density. These two methods are applied during the last level of the placement framework. To our knowledge, there are no published papers that apply narrow channel reduction and dummy-cell local insertion to routability-driven placement. Furthermore, we perform pre-placement inflation at each level of the placement framework. In contrast to the in-placement inflation used in [1, 6], our pre-placement inflation takes place before placing the components at each level, and is not iteratively applied during that particular level.

Let \((x,y)\) be the vector of cell coordinates and \(\text{HPWL}(x,y)\) be the total half-perimeter wirelength on the placement area. We enable cell spreading by applying a uniform placement grid on top of the placement area. We define \(D_b(x,y)\) to be the cell density inside placement bin \(b\), and \(M_b\) to be the average cell density. The placement formulation is:

\[
\min \text{HPWL}(x,y)
\]

\[
s.t. \quad D_b(x,y) = M_b, \forall \text{placement bin } b.
\]

3. ROUTABILITY METRIC

The metric of DAC 2012 routability-driven placement contest [22] accounts for both routability and runtime. Let \(ACE\) be the average congestion of \(g\)-cell edges based on the histogram of \(g\)-edge congestion as described in [22, 29]. Then, \(ACE(x)\) computes the average congestion of the top \(x\)% congested \(g\)-cell edges. For the contest, the set of \(ACE\) values was computed using \(x \in 0.5, 1, 2, 5\). In order to simplify calculations, the contest routers NCTUgr and BFG-R report the set of \(ACE\) values. Based on the \(ACE\) metric, we calculate the peak-weighted congestion as:

\[
\text{PWC} = \frac{ACE(x)}{4}, \quad x \in 0.5, 1, 2, 5.
\]

Using (1), the routing congestion is:

\[
\text{RC} = \max(100, \text{PWC}).
\]

Let \(PF\) be the penalty factor that scales the \(\text{HPWL}\) to account for routing congestion. In the DAC 2012 placement contest \(PF=0.03\) so that for every 1% excess routing congestion, there is a 3% wirelength penalty. Also, let \(\text{RuntimeFactor}\) be the runtime factor that penalizes placers based on their runtime. Using (2), the contest metric is defined as the scaled wirelength of the placement solution as follows:

\[
\text{ContestMetric} = \frac{\text{HPWL}(1 + PF(\text{RC} - 100))(1 + \text{RuntimeFactor})}{4}.
\]

4. NARROW CHANNEL REDUCTION BY APPLYING NEIGHBOR-BASED FIXED-MACRO INFLATION

A common factor of congestion in modern industrial designs is the existence of fixed macros on the placement area. These macros are responsible for creating narrow channels as illustrated in Figure 1. The existence of narrow channels on the placement area contributes to the routing congestion on the chip. This is due to the fact that cells placed in a narrow regions have difficulty in escaping during placement, thus leading to an increase of the routing congestion in their neighborhood.

We propose blocking narrow channels on the chip in order to reduce routing congestion. Our method is implemented by inflating fixed macros based on their distance to neighboring fixed macros. Our neighbor-based fixed-macro inflation for blocking narrow channels is further described as follows:

For each one of the fixed macros, we define a set of right-side neighbors that consists of all fixed macros located on the right side of this particular macro, and whose distance from the macro is smaller than a right-threshold value.

Similarly, we define a set of left-side neighbors of the macro, with distance smaller than a left-threshold value, a set of top-side neighbors of the macro, with distance smaller
than a top-threshold value, and a set of bottom-side neighbors of the macro, with distance smaller than a bottom-threshold value. The four threshold values, that determine the closest neighbors on each side of the macro, may not necessarily be equal. We also define four rates of inflation for the macro, each rate associated to the degree of inflation we apply to the macro towards the respective side. For example, to completely block a narrow channel on the right side of the macro, we set the right-side rate of inflation of the macro equal to 100%.

5. Dummy-Cell Insertion Inside Regions of Reduced Fixed-Macro Density

The existence of narrow channels on the placement area is not the only factor to take into account in order to improve routability. Another factor is the existence of large empty regions on the design. As illustrated in Figure 2, when a design has the majority of its macros located at the periphery, then the empty part of the placement area may become congested during the placement of cells. To avoid this type of congestion, we identify large empty regions on the chip and insert dummy cells inside them. To identify such regions, we apply a coarse grid on the placement area and determine bins with reduced fixed-macro density. Then, we insert dummy cells inside the bin with the smallest fixed-macro density value. More specifically, after applying a coarse grid on top of the placement area, we calculate the overlap of each fixed macro with the bins of the grid. Then, we identify the bin that has the smallest amount of overlap with fixed macros. Finally, we insert dummy cells in this particular bin.

6. Pre-Placement Inflation

To identify regions characterized by routing congestion, we usually employ techniques that are based on pin density, net density, and/or fast routing. Then, we inflate cells located inside congested regions. In order to reduce congestion, we perform pre-placement inflation at each level of the placement framework.

6.1 GTL-Based Inflation

We adopt the Group of Tangled Logic (GTL) metric that was introduced in [8] to detect tangled logic structures in a netlist. Let $T(C)$ be the net cut of cell cluster $C$, $|C|$ be the number of cells in $C$, $A_C$ be the average pin count of cells in $C$, and $p$ be the Rent exponent. The GTL score of $C$ is defined as follows:

$$GTL(C) = \frac{T(C)}{A_C|C|^p 2^\frac{N}{c}}.$$

The GTL score curve of $C$ is illustrated in Figure 3 and corresponds to the growth of $C$. During the growth of a cluster from a seed cell, the score of the cluster is modified by iteratively adding highly connected neighbors. The curve contains a distinct trough if a tangled logic structure appears during the growth. The position of the trough indicates when the cluster growth has reached the most tangled logic structure. The GTL-based technique for cluster growth imposes a significant runtime overhead to the algorithm. In order to find tangled logic structures, the algorithm selects a large number of seed cells and carries out the growth process multiple times. As reported in [8], the runtime for a circuit with 800K vertices may take up to 141 minutes, even if eight parallel threads are used. This particular circuit size only corresponds to the median among the benchmarks of the ISPD 2011 contest [21].

To solve the problem of increasing runtimes we incorporate the cluster growth method to the multilevel placement engine. A bottom-up clustering of cells is initially performed and all cells are treated as objects. The best pair of objects is identified and clustered into a new object and the process continues iteratively until the netlist becomes sufficiently coarse. We integrate GTL scoring into the clustering process with a small runtime overhead and each object is associated to a GTL score curve. When a pair of objects is...
Although the problem formulation in (4) is nonlinear, it can be proved that Algorithm 1 solves this problem optimally by iteratively assigning whitespace to cells of increased pin density, until there is no whitespace left.

**Algorithm 1 Pin Density-Based Inflation.**

**Input:** Pin number \( \{p_i\} \) and original area \( \{A_i\} \) of cells, available amount of whitespace \( W \) for cell inflation.

**Output:** Inflated areas \( \{A'_i\} \) of cells.

**Step 1:** Sort cells in descending order by making use of their pin density s.t. \( p_i/A_i \geq p_{i+1}/A_{i+1}, \forall i \).

**Step 2:** Find the largest index \( k \) s.t. \( I_k \leq W < I_{k+1} \), where \( I_k = A_k(\sum_{i=1}^{k} p_i)/p_k - (\sum_{i=1}^{k} A_i) \).

**Step 3:** Compute the maximum pin density after inflation \( d_{\text{max}} = (\sum_{i=1}^{k} p_i)/(W + \sum_{i=1}^{k} A_i) \).

**Step 4:** Compute the inflated area \( A'_i = \begin{cases} p_i/d_{\text{max}} & (i \leq k) \\ A_i & (k < i) \end{cases} \).

**Theorem 1.** The solution given by Algorithm 1 is the optimal solution of the nonlinear programming problem (4).

**Proof.** The pin density \( d_{\text{max}} \) is the maximum pin density after inflation, because

\[
W < I_{k+1} \Leftrightarrow W < A_{k+1} \sum_{i=1}^{k+1} p_i/p_{k+1} - \sum_{i=1}^{k} A_i
\]

\[
\Leftrightarrow W + \sum_{i=1}^{k} A_i < A_{k+1} \sum_{i=1}^{k+1} p_i/p_{k+1} \Leftrightarrow \frac{W + \sum_{i=1}^{k} A_i}{\sum_{i=1}^{k+1} p_i} < \frac{A_{k+1}}{p_{k+1}}
\]

\[
\Leftrightarrow \frac{1}{d_{\text{max}}} < \frac{A_{k+1}}{p_{k+1}} \Leftrightarrow \frac{p_{k+1}}{A_{k+1}} < d_{\text{max}}.
\]

Thus,

\[
\begin{cases}
p_i/A'_i = d_{\text{max}} & (i \leq k), \\
p_i/A'_i < p_i/A'_k = d_{\text{max}} & (k < i).
\end{cases}
\]

Assume \( d_{\text{max}} \) is not the optimal pin density. Then, there is a solution \( A''_i \) s.t. \( p_i/A''_i < d_{\text{max}}, \forall i \). However, in such case, the second constraint of (4) is violated, since:

\[
\sum_{i=1}^{n} (A''_i - A_i) = \sum_{i=1}^{n} A''_i + \sum_{i=k+1}^{n} A''_i - \sum_{i=1}^{n} A_i
\]

\[> \sum_{i=1}^{k} \frac{p_i}{d_{\text{max}}} + \sum_{i=k+1}^{n} A''_i - \sum_{i=1}^{n} A_i
\]

\[= (W + \sum_{i=1}^{k} A_i) + \sum_{i=k+1}^{n} A''_i - \sum_{i=1}^{n} A_i
\]

\[= W + \sum_{i=k+1}^{n} (A''_i - A_i) \geq W.
\]

By contradiction, the solution given by Algorithm 1 is an optimal solution of (4). \( \square \)
Intuitively, all inflated cells will have the same pin density in an optimal solution; otherwise, if an inflated cell does not achieve maximum pin density, it will be inflated less. The effect of pre-placement inflation is discussed in Section 7. Pre-placement inflation improves the total number of overflows compared to a naive inflation with the same amount of whitespace allocated.

7. EXPERIMENTAL RESULTS

The placers were evaluated on the DAC 2012 routability-driven contest benchmarks, using the contest router NTUgr [22]. Table 1 reports the scaled wirelength of the contest top four teams Ripple, NTUplace4, mPL12, and simPLR. For each placer, the scaled wirelength is calculated as $HPWL(1 + PF(RC – 100))$, where $PF = 0.03$. The $HPWL$ is obtained from the output of the global router. Table 2 reports the placement runtimes of the contest top four teams Ripple, NTUplace4, mpl12, and simPLR. For each placer, the runtime is in seconds. For the contest placers, the results were obtained on the contest organizers’ 64-bit Intel Xeon CPU X7560 at 2.27GHz (placer binaries were not made available to us). Our placer was evaluated on a machine with the Intel Core i7 Processor at 2.67GHz with 8GB main memory and the average runtime of our placer was estimated by applying a scaling factor.

Our placer minimizes the routing congestion and improves routability. In terms of scaled wirelength, we obtain improved results compared to the winner of the DAC 2012 routability-driven placement contest [22]. In terms of placement runtime, our placer is required smaller placement runtimes than the analytical placers of the contest. However, no direct comparison can be made due to the fact that the contest binaries were not available to use, therefore we used a different machine for our placer.

8. CONCLUSION

Our proposed placer incorporates narrow channel reduction, dummy-cell insertion inside regions of reduced fixed-macro density, and pre-placement inflation in order to reduce routing congestion and improve the routability of large-scale mixed-size designs. The quality of our placement tool on industrial circuits is evaluated using the global routers of the DAC 2012 placement contest [22] and our results compare favorably to the top four teams that participated in the contest.

9. ACKNOWLEDGMENT

This research was supported in part by the NSF Computing Research Association (Prime Award Number 1136996).

10. REFERENCES


[16] G.-J. Nam and J. Cong. Modern Circuit Placement:
Table 1: Experimental Results on the contest placers Ripple, NTUplace4, mPL12, and simPLR. The results for the contest placers were obtained on the contest organizers’ 64-bit Intel Xeon CPU X7560 at 2.27GHz. The results of our placer were obtained on an Intel Core i7 Processor at 2.67GHz with 8GB main memory.

<table>
<thead>
<tr>
<th>Circuit</th>
<th>Contest Ripple</th>
<th>Contest NTUplace4</th>
<th>Contest mPL12</th>
<th>Contest simPLR</th>
<th>Our Placer</th>
</tr>
</thead>
<tbody>
<tr>
<td>sb19</td>
<td>1.70</td>
<td>1.53</td>
<td>2.46</td>
<td>1.66</td>
<td>1.51</td>
</tr>
<tr>
<td>sb14</td>
<td>2.31</td>
<td>2.26</td>
<td>2.67</td>
<td>2.48</td>
<td>2.45</td>
</tr>
<tr>
<td>sb16</td>
<td>2.74</td>
<td>2.80</td>
<td>3.01</td>
<td>3.47</td>
<td>2.74</td>
</tr>
<tr>
<td>sb9</td>
<td>2.97</td>
<td>2.55</td>
<td>3.22</td>
<td>2.75</td>
<td>2.50</td>
</tr>
<tr>
<td>sb3</td>
<td>4.27</td>
<td>3.62</td>
<td>4.66</td>
<td>3.90</td>
<td>3.60</td>
</tr>
<tr>
<td>sb11</td>
<td>3.58</td>
<td>3.42</td>
<td>4.52</td>
<td>3.98</td>
<td>3.40</td>
</tr>
<tr>
<td>sb6</td>
<td>3.56</td>
<td>3.42</td>
<td>3.95</td>
<td>3.53</td>
<td>3.40</td>
</tr>
<tr>
<td>sb2</td>
<td>7.39</td>
<td>6.24</td>
<td>1.33</td>
<td>8.24</td>
<td>6.14</td>
</tr>
<tr>
<td>sb12</td>
<td>3.42</td>
<td>3.12</td>
<td>5.40</td>
<td>3.63</td>
<td>3.04</td>
</tr>
<tr>
<td>sb7</td>
<td>4.45</td>
<td>3.99</td>
<td>5.24</td>
<td>1.73</td>
<td>3.95</td>
</tr>
<tr>
<td>avg.</td>
<td>1.09</td>
<td>1.00</td>
<td>1.41</td>
<td>1.46</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Table 2: Experimental Results on the contest placers Ripple, NTUplace4, mPL12, and simPLR. The results for the contest placers were obtained on the contest organizers’ 64-bit Intel Xeon CPU X7560 at 2.27GHz. The results of our placer were obtained on an Intel Core i7 Processor at 2.67GHz with 8GB main memory and the average runtime of our placer was estimated by applying a scaling factor.

<table>
<thead>
<tr>
<th>Circuit</th>
<th>Contest Ripple</th>
<th>Contest NTUplace4</th>
<th>Contest mPL12</th>
<th>Contest simPLR</th>
<th>Our Placer</th>
</tr>
</thead>
<tbody>
<tr>
<td>sb19</td>
<td>2309</td>
<td>8450</td>
<td>11087</td>
<td>981</td>
<td>9911</td>
</tr>
<tr>
<td>sb14</td>
<td>2806</td>
<td>9341</td>
<td>10006</td>
<td>1247</td>
<td>7539</td>
</tr>
<tr>
<td>sb16</td>
<td>2757</td>
<td>8573</td>
<td>13670</td>
<td>1128</td>
<td>9435</td>
</tr>
<tr>
<td>sb9</td>
<td>4307</td>
<td>13129</td>
<td>14910</td>
<td>1821</td>
<td>12736</td>
</tr>
<tr>
<td>sb3</td>
<td>5432</td>
<td>14144</td>
<td>19294</td>
<td>2283</td>
<td>12924</td>
</tr>
<tr>
<td>sb11</td>
<td>3745</td>
<td>15263</td>
<td>18284</td>
<td>2342</td>
<td>14723</td>
</tr>
<tr>
<td>sb6</td>
<td>4944</td>
<td>11179</td>
<td>20508</td>
<td>2484</td>
<td>17121</td>
</tr>
<tr>
<td>sb2</td>
<td>6686</td>
<td>17466</td>
<td>23900</td>
<td>3125</td>
<td>18741</td>
</tr>
<tr>
<td>sb12</td>
<td>7635</td>
<td>34831</td>
<td>26107</td>
<td>3459</td>
<td>19245</td>
</tr>
<tr>
<td>sb7</td>
<td>11285</td>
<td>29593</td>
<td>22233</td>
<td>3025</td>
<td>17243</td>
</tr>
<tr>
<td>avg.</td>
<td>2.37</td>
<td>7.24</td>
<td>8.67</td>
<td>1.00</td>
<td>8.07</td>
</tr>
</tbody>
</table>


