Bounding the Efforts on Congestion Optimization for Physical Synthesis

Davide Pandini†, Lawrence T. Pileggi‡, and Andrzej J. Strojwas‡
†STMicroelectronics, Central R&D, 20041 Agrate Brianza, Italy
‡Department of Electrical and Computer Engineering
Carnegie Mellon University, Pittsburgh, PA 15213 U.S.A.
email: davide.pandini@st.com, pileggi@ece.cmu.edu, ajs@ece.cmu.edu

1 Introduction

In DSM technologies interconnects play a crucial role in the overall performance of VLSI systems [1] and their impact must be carefully considered in order to satisfy the design constraints during all phases of the top-down design flow [2][3][4], which broadly speaking consists of logic synthesis and physical design. While the timing closure problem has been extensively studied in academia [5][6][7], and several commercial flows have been proposed by the EDA industry, of equal importance is the impact of wiring on defining the block or chip size, in particular for logic synthesis, which traditionally has attempted to minimize cell area in order to minimize the size of a block or IC, without considering the interconnect area. However, since the DSM wirelength trends do not scale with feature sizes, cell area minimization can no longer guarantee overall block or die area minimization. As a result, although the total active area of the synthesized netlist can be made smaller than the block area, the design may not be routable within the assigned chip or block size. In physical design, the required routing resources are modeled by congestion, and placement and routing can sometimes fix, or avoid, potential congestion problems [8][9]. However, attempting to solve such problems at the physical design stage is too late, since the optimization potentials of layout tools are limited by the structure of the synthesized netlist. Ideally, area minimization which includes the impact of wiring via congestion models would occur during logic synthesis, where structural modifications to the netlist are still possible. The methodology presented in [15] and [16] addressed the problem of congestion minimization during the synthesis technology dependent step, i.e., technology mapping, where it is possible to exploit placement information to capture the connectivity of the technology independent representation of the circuit, and then explore the trade-offs between area (and/or delay) and congestion minimization when a fixed amount of routing resources are available. However, in [15] it was demonstrated that excessive focus on congestion minimization during technology mapping would also yield very sub-optimal solutions in terms of both cell area (and/or delay). A priori estimation of the efforts on congestion minimization, along with the other traditional synthesis optimization objectives such as area and/or delay, is a very difficult problem, since not only does congestion depend on the structure of the gate level netlist, but it also depends on the amount of routing resources available (i.e., the die area and the number of metal layers), and there is a lack of accurate congestion models at the synthesis level. In this work, we derive bounds for the congestion minimization efforts during technology mapping, which effectively restrict the search space for an optimal weighting factor that controls the impact of congestion minimization with respect to the other synthesis optimization objectives.

This paper is organized as follows. In Section 2 previous work in layout-driven synthesis is overviewed, and in Section 3 the congestion-aware technology mapping approach is described. In Section 4 the bounds that restrict the congestion minimization efforts during the synthesis phase of the design flow are obtained, while in Section 5, results showing the effectiveness of such
bounds in reducing the search space for routable synthesized netlists are presented. Finally, our conclusions are summarized in Section 6.

2 Previous Work

Significant progress, both in academia and industry, has been made to provide logic synthesis with physical information. In [10], Pedram and Bhat integrated technology mapping with a companion placement to include wiring contribution in area and delay minimization. Other works describe attempts at modeling routability during logic extraction in the synthesis technology independent step, either by using topological information such as the fanout range [11], or a companion placement of the logic network nodes [12]. More recently, two approaches which address the congestion problem during logic synthesis were presented by Kutzschebauch and Stok [13][14], where congestion was taken into account by means of the physical coordinates obtained from placing the technology independent netlist. However, these works do not consider that different layout regions can have very different routing demands, and more importantly, they do not consider the sub-optimality in terms of area (and/or delay) that results from introducing congestion into the synthesis optimization objectives. In contrast, in [15] it was demonstrated that the impact of congestion minimization with respect to the other synthesis optimization objectives must be carefully controlled, while in [16] the global and local nature of congestion was analyzed, and a methodology that removes local congested areas by means of incremental remapping was presented.

3 Congestion-Aware Technology Mapping

The negative effects on routability of excessive efforts on area minimization during logic synthesis have been discussed in [15], where the technology mapping algorithm proposed by Keutzer [17] and implemented in [18], was extended to include congestion among the optimization objectives. An additional term representing the interconnect cost of a match was included in the technology mapping objective function by means of the congestion minimization factor \( K \) (for simplicity we consider here only area minimization):

\[
\text{COST}(m, v) = \text{AREA}(m, v) + K \cdot \text{WIRE}(m, v),
\]

where \( \text{AREA}(m, v) \) is the area cost of the match \( m \) at vertex \( v \) of the subject tree [17], and \( \text{WIRE}(m, v) \) is the corresponding wire cost [15]. The critical issue in this approach is the impact of congestion minimization into (1), controlled by the factor \( K \), which introduces a perturbation in the cost function, thus resulting in area (and/or delay) penalty. Uncontrolled congestion minimization may yield a very suboptimal solution in terms of cell area (and/or delay), and consequently an unroutable netlist within the assigned chip size. In contrast, if the impact of the interconnect cost in (1) is negligible, then the mapped netlist is structurally very similar to the minimum area solution (\( K = 0.0 \)), which may be structurally unroutable, as it was shown in [15]. However, the approach proposed in [13] did not address this critical problem. Instead, the results reported in [15] support the claim that excessive (or negligible) weight on congestion minimization may yield unroutable netlists within the fixed die size.

4 Estimating the Bounds for the Congestion Minimization Factor

In [15] it was demonstrated that it is not possible to estimate \textit{a priori} one global value for the congestion minimization factor \( K \), which yields a routable synthesized netlist. However, it was also demonstrated that there exists a range for the factor \( K \), by including the wire cost into the technology mapping cost function and controlling its impact with respect to the other synthesis optimization objectives, it is possible to synthesize structurally less congested netlists within fixed floorplan constraints. Consequently, for such range there exist the lower and upper bounds, i.e., \( K_{\text{min}} \) and \( K_{\text{max}} \). When \( K \leq K_{\text{min}} \), the structure of the mapped netlist is not significantly different from the minimum area netlist, since the wire cost has a negligible impact on the technology mapping cost function (1). On the other hand, when \( K \geq K_{\text{max}} \), the wire contribution dominates the other synthesis optimization objectives, thus introducing a large cell area penalty.

Given an unbond netlist of base gates, all the nets belong to the set \( N = \{n_1, ..., n_N\} \). Since in our technology mapping approach the network DAG is partitioned into subject trees [17], every multifanout net is split into a number of two-point source-sink nets. The two-point net set \( N = \{\tilde{n}_1, ..., \tilde{n}_p\} \) has cardinality \( p \) bounded by the relation:

\[
N \leq p = \sum_{i=1}^{N} |\text{fanouts}(n_i)| \leq N \cdot \text{NetMaxFanouts},
\]

where the function \( |\text{fanouts}(n_i)| \) yields the fanouts for net \( n_i \), and \( \text{NetMaxFanout} = \max_{1 \leq i \leq N}(|\text{fanouts}(n_i)|) \). The nets in \( N \) are sorted by their non increasing length according to the source-sink Manhattan distance, using the geometrical coordinates obtained from the initial placement of the technology independent network. The operator \( L_1 \) computes the Manhattan length for a two-point net, and the following relation holds:

\[
L_1(\tilde{n}_j) \geq L_1(\tilde{n}_{j+1}) \quad 1 \leq j \leq p-1,
\]

thus we can define \( \Delta Net_{\text{min}} = L_1(\tilde{n}_j) - L_1(\tilde{n}_{j+1}) > 0 \), such that:

\[
L_1(\tilde{n}_j) - L_1(\tilde{n}_{j+1}) \leq L_1(\tilde{n}_j) - L_1(\tilde{n}_{k+1})
\]

where \( 1 \leq k \leq p-1 \), and \( j \neq k \). The computational complexity of this step, given by the sorting algorithm, is \( O(p \log p) \), where \( p \) is expressed by (2). A similar non increasing order sorting procedure, with respect to the cell area, is performed on the match set \( M = \{m_1, ..., m_q\} \), whose cardinality is the number of minimum area instances of the library cells. Hence, the following relation holds:

\[
A(m_j) \geq A(m_{j+1}) \quad 1 \leq j \leq M-1,
\]

where the operator \( A(m_j) \) returns the area of the library cell corresponding to match \( m_j \). The minimum area difference between two matches is defined as: \( \Delta \text{Match}_{\text{min}} = A(m_j) - A(m_{j+1}) > 0 \), such that:

\[
A(m_j) - A(m_{j+1}) \leq A(m_j) - A(m_{k+1})
\]

where \( 1 \leq k \leq M-1 \), and \( j \neq k \). Typically, the number of cells in a standard cell library is much less than the number of nets in modern VLSI circuits, thus the computational complexity of this step is dominated by the net sorting procedure. In order to estimate the lower bound \( K_{\text{min}} \), we examine the following case: minimum area cost difference between two match candidates, and maximal wire-length difference between these two candidates. Consider the following expressions:

\[
\text{AREA}(m_1, v) + K \cdot \text{WIRE}(m_1, v)
\]

\[
\text{AREA}(m_2, v) + K \cdot \text{WIRE}(m_2, v)
\]

corresponding to different matches \( m_1 \) and \( m_2 \) at the same vertex \( v \), which include both the cell area and the wire cost, and suppose the tree covering algorithm [15] must choose one of the two solu-
tions represented in (3) and (4). The congestion minimization factor $K$ can be expressed by:

$$K = \frac{\text{AREA}(m_2, v) - \text{AREA}(m_1, v)}{\text{WIRE}(m_1, v) - \text{WIRE}(m_2, v)}$$  \hspace{1cm} (5)$$

and in order to estimate $K_{\text{min}}$ we focus on the smallest numerator and the largest denominator as possible. The wirelength contributions in (5) are obtained by the following expression:

$$\text{WIRE}(m, v) = \sum_{v_i \in \text{fanins}(m, v)} \left( L_1(v_i, v) + \sum_{v_k} L(v_i, v_k) \right)$$

where the operator $L_1(v_i, u)$ computes the Manhattan distance between the geometrical coordinates of vertices $v$ and $u$, the function $\text{match}(v)$ yields the match at vertex $v$, and the function $\text{fanins}(m, v)$ returns all the fanin vertices of match $m$ at vertex $v$. After the net sorting step, the following relation holds: $L_1(\tilde{n}_p) \leq L_1(v, u) \leq L_1(\tilde{n}_1)$, $\forall v, u$, and consequently the following relations are satisfied:

$$\text{WIRE}(m_1, v) \leq \sum_{v_i} \left( L_1(\tilde{n}_1) + \sum_{v_k} L(\tilde{n}_k) \right) \leq \text{MFI} \cdot (1 + \text{MFI}) \cdot L_1(\tilde{n}_1)$$ \hspace{1cm} (6)$$

$$\text{WIRE}(m_2, v) \geq \sum_{v_i} \left( L_1(\tilde{n}_p) + \sum_{v_k} L(\tilde{n}_p) \right) \geq s_{\text{MFI}} \cdot (1 + s_{\text{MFI}}) \cdot L_1(\tilde{n}_p)$$ \hspace{1cm} (7)$$

where:

$$\text{MFI} = \max_{1 \leq i \leq M} \{|\text{fanins}(m, v)|\}$$

$$\text{mfi} = \min_{1 \leq i \leq M} \{|\text{fanins}(m, v)|\}$$

respectively. The minimum area difference between two matches is given by $\Delta \text{Match}_{\text{min}}$, and the following relation holds true:

$$0 < \Delta \text{Match}_{\text{min}} \leq \text{AREA}(m_2, v) - \text{AREA}(m_1, v).$$  \hspace{1cm} (8)$$

Hence, the lower bound $K_{\text{min}}$ can be obtained by using relations (6), (7), and (8):

$$K_{\text{min}} = \frac{\Delta \text{Match}_{\text{min}}}{\text{MFI} \cdot (1 + \text{MFI}) \cdot L_1(\tilde{n}_1) - \text{mfi} \cdot (1 + \text{mfi}) \cdot L_1(\tilde{n}_p)}.$$  \hspace{1cm} (9)$$

In contrast, the estimation of the upper bound $K_{\text{max}}$ is more complicated. The maximum area cost difference between two mappings is given by the difference between the area of the unbound netlist and the minimum area solution. The minimum wire cost difference between two different matching solutions is bounded by $\Delta \text{Net}_{\text{min}}$, and $K_{\text{max}}$ can be expressed as:

$$K_{\text{max}} = \frac{\text{UnboundNetlistArea} - \text{MinimumAreaMapping}}{\Delta \text{Net}_{\text{min}}}.$$  \hspace{1cm} (10)$$

However, this upper bound is quite loose and not of practical interest, since from all our experimental results, a few percent of cell area penalty with respect to minimum area may yield an unroutable netlist within the assigned tight floorplan constraints. In order to obtain an upper bound of practical interest, we should estimate a $K_{\text{max}}$ value, which effectively restricts the search space, and at the same time does not exclude potentially routable netlists. This is a difficult task, since at this stage we only have the initial placement of the technology independent netlist, and have not performed any matching and covering as yet. The function $\text{AREA}(m, v)$ not only depends on the fanins of the match at vertex $v$, but it also depends on the logic depth of vertex $v$, i.e., its distance from the leaf vertices. In order to obtain a practical estimation of the upper bound we propose the following expression:

$$K_{\text{max}} = \frac{A(m_1) \cdot (1 + \text{MFI}) - A(m_2) \cdot (1 + \text{mfi})}{\Delta \text{Net}_{\text{min}}}.$$  \hspace{1cm} (10)$$

which in practice has performed well on all our experiments. It is worth pointing out that $K_{\text{min}}$ and $K_{\text{max}}$ have two different behaviors. When the minimum area solution yields an unroutable netlist within the fixed size, by increasing $K$ we include congestion among the synthesis optimization objectives. Therefore, small changes with respect to the structure of the minimum area netlist may yield a routable solution, while keeping the cell area penalty limited. However, since the cell area penalty imposes the effectiveness of this approach, the lower bound $K_{\text{min}}$ must be as small as possible. Therefore, by estimating $K_{\text{min}}$ with (9), we obtain a lower bound where for $K \leq K_{\text{min}}$ the netlist structure is not significantly different from the minimum area solution, and if the minimum area netlist is unroutable, then all the netlists obtained with $0 < K \leq K_{\text{min}}$ are also unroutable. On the other hand, as $K$ increases, the structure of the mapped netlist changes more significantly, but the associated cell area penalty decrement may soon yield unroutable netlists, such that when $K \geq K_{\text{max}}$, the netlist is unroutable within the fixed floorplan constraints. The estimation of both $K_{\text{min}}$ and $K_{\text{max}}$ can be performed efficiently, thus effectively restricting the search space for a routable mapped netlist using the cost function (1) within the range:

$$K_{\text{min}} < K < K_{\text{max}}.$$  \hspace{1cm} (11)$$

Not every $K$ within range (11) guarantees a routable solution, as it was shown in [15]. However, when the minimum area netlist is unroutable, the netlists obtained with $0 < K \leq K_{\text{min}}$ are unroutable as well, thus reducing the search space. Routable solutions may exist for $K \geq K_{\text{max}}$, but since a few cell area penalty percents may exceed the available routing resources under tight floorplan constraints, the upper bound expressed by (10) works well in practice.

## 5 Experimental Results

The bounds for the congestion minimization factor $K$ expressed by (9) and (10) have been applied to the routability results obtained on the IWLS93 benchmarks SPLA and PDC, with the congestion-aware technology mapping algorithm presented in [15]. For completeness, the routability ranges are reported in Table 1. In our experiments we have considered all the single-output combinational library cells up to four inputs of a commercial cell library in 0.18\textmu m technology, thus $\text{mfi} = 1$, and $\text{MFI} = 4$ respectively. The library cell sizes used for the bound estimation are reported in Table 2, and yield: $\Delta \text{Match}_{\text{min}} = A(\text{ND2HS}) - A(\text{IVHS}) = 4.096$.

### Table 1. Routability range obtained in [15]

<table>
<thead>
<tr>
<th>IWLS93 benchmark</th>
<th>$K_{\text{min}}$</th>
<th>$K_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPLA</td>
<td>0.00025</td>
<td>0.005</td>
</tr>
<tr>
<td>PDC</td>
<td>0.0010</td>
<td>0.01</td>
</tr>
</tbody>
</table>

### Table 2. Library cell sizes

<table>
<thead>
<tr>
<th>Library Cell</th>
<th>IVHS</th>
<th>ND2HS</th>
<th>NR4HS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area (\textmu m$^2$)</td>
<td>8.192</td>
<td>12.288</td>
<td>32.768</td>
</tr>
</tbody>
</table>
The net length values* from the initial placement of the technology independent netlists are reported in Table 3, and the results for

<table>
<thead>
<tr>
<th>IWLS93 benchmark</th>
<th>Net Maximum Length</th>
<th>Net Minimum Length</th>
<th>$\Delta Net_{min}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPLA</td>
<td>92352</td>
<td>128</td>
<td>64</td>
</tr>
<tr>
<td>PDC</td>
<td>98048</td>
<td>128</td>
<td>66</td>
</tr>
</tbody>
</table>

Table 3. Net length for bound estimation

$K_{min}$ and $K_{max}$ obtained from (9) and (10) are shown in Table 4. These bounds effectively restrict the search space for a $K$ value which yields a routable solution. Moreover, it is important to note that according to the routability results reported in [15], no routable solution exists outside the range expressed by relation (11).

<table>
<thead>
<tr>
<th>IWLS93 benchmark</th>
<th>$K_{min}$</th>
<th>$K_{max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPLA</td>
<td>0.00000222</td>
<td>2.304</td>
</tr>
<tr>
<td>PDC</td>
<td>0.0000209</td>
<td>2.234</td>
</tr>
</tbody>
</table>

Table 4. Bounds obtained from (9) and (10)

In order to further limit the search space for less congested netlists, we would ideally generate a tighter lower bound, i.e., increase the value of $K_{min}$, and a tighter upper bound, i.e., decrease the value of $K_{max}$. By examining the net length distributions of the technology independent representations of several circuits, we observed that most nets are short ones, and the few long nets are localized in the distribution tail. As a consequence, a tighter estimation of the congestion minimization factor $K$ can be obtained in practice by neglecting both the shortest and the longest nets. Our heuristic does not consider all the net lengths with a number of instances less than 5% of the peak distribution, that corresponds to the net length with the maximum number of instances. We obtain a restricted net length range, as shown in Table 5, where most net instances are within this range. The net length values reported in Table 3 and in all the other Tables in Section 5 are expressed in DBU, i.e., Database Units, and are consistent with all the experimental results reported in [15].

<table>
<thead>
<tr>
<th>IWLS93 benchmark</th>
<th>$\text{NetLen}_{min}$</th>
<th>$\text{NetLen}_{max}$</th>
<th>Net instances within the restricted length range</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPLA</td>
<td>128</td>
<td>2624</td>
<td>81.1%</td>
</tr>
<tr>
<td>PDC</td>
<td>128</td>
<td>2688</td>
<td>79.3%</td>
</tr>
</tbody>
</table>

Table 5. Net length restricted range

\[
K_{min}^{R} = \frac{\Delta \text{Match}_{min}}{MFI \cdot (1 + MFI) \cdot \text{NetLen}_{max} - mfi \cdot (1 + mfi) \cdot \text{NetLen}_{min}} \quad (12)
\]

\[
K_{max}^{R} = \frac{A(m_1) \cdot (1 + MFI) - A(m_2) \cdot (1 + mfi)}{\text{NetLen}_{max} - \text{NetLen}_{min}} \quad (13)
\]

where with respect to (11) the following relation holds:

\[K_{min}^{R} < K_{min} < K_{max}^{R} < K_{max}\]

The restricted bounds of the congestion minimization factor $K$ obtained with (12) and (13) using the values of Table 5, are reported in Table 6. Our heuristic has significantly reduced the search space for a less congested netlist, and at the same time, according to the results presented in [15] and summarized in Table 1, the routability range has been fully captured.

The net length values reported in Table 3 and in all the other Tables in Section 5 are expressed in DBU, i.e., Database Units, and are consistent with all the experimental results reported in [15].

\[
\begin{array}{|c|c|c|}
\hline
\text{IWLS93 benchmark} & K_{min}^{R} & K_{max}^{R} \\
\hline
\text{SPLA} & 0.000078 & 0.059 \\
\text{PDC} & 0.000076 & 0.058 \\
\hline
\end{array}
\]

Table 6. Restricted $K$ bounds

6 Conclusions

The wiring congestion is a very important design factor which directly impacts the time-to-market of complex System-on-Chip designs, and must be considered both globally in logic synthesis, and locally in physical design. However, traditional synthesis focused on area minimization may produce structurally unroutable circuits, and hence yield larger logic block areas due to wiring congestion. While global congestion can be effectively addressed during technology mapping, the suboptimality obtained by including congestion into the synthesis optimization objectives must be carefully evaluated. In this paper we have derived practical bounds for controlling the congestion minimization impact during logic synthesis, with respect to the other optimization objectives like area and/or delay. Our results demonstrate that such bounds effectively restrict the search space for synthesizing structurally more routable circuits within fixed floorplan constraints. In conclusion, our approach yields a practical methodology, which incorporates congestion directly within the synthesis multi-objective optimization.

References