Timing-Driven Analytical Placement According to Expected Cell Distribution Range

Jai-Ming Lin, You-Yu Chang, and Wei-Lun Huang

Speaker: Wei-Lun Huang
March 2024
Department of Electrical Engineering,
National Cheng Kung University, Tainan, Taiwan
Outline

- Introduction
- Our TDP Methodology
- Experimental Results
- Conclusion
Outline

- Introduction
- Our TDP Methodology
- Experimental Results
- Conclusion
Achieving timing closure is the largest concern to a modern design. Tape-out time will be delayed if the timing requirement is not met.

Most of placement algorithms target on minimization of wirelength and routing congestion without considering timing.

Timing closure can be achieved more quickly if a global placement algorithm can reduce WNS and TSN in addition to wirelength and routability. Moreover, chip cost can be reduced as well since it does not have to insert a lot of buffers to timing violations nets.
Multilevel Framework in Our Work

- Multilevel Framework is composed of three stages.
  - Bottom-up coarsening.
  - Initial placement.
  - Top-down refinement.

DEF / LEF / TF

Placement Prototyping
- Bottom-Up Coarsening
- Initial Global Distribution
- Top-down Refinement

Macro Placement

Initial Placement (quadratic programming)

\[ l_{\text{max}} = 3 \]

\[ l = 0 \]

\[ l = 1 \]

\[ l = 2 \]
**Motivation**

- **Lin et al.** [1] propose the design hierarchy-guided clustering which cluster objects layer by layer based on a design hierarchy tree.
- The score function to cluster two objects $o_i$ & $o_j$ is defined as:

$$S(o_i, o_j) = \rho(o_i, o_j) \times \Gamma(o_i, o_j) \times \left( \Phi(o_i, o_j) + \alpha \times \Psi(o_i, o_j) \right)$$

- $\rho(o_i, o_j)$: Area function.
- $\Gamma(o_i, o_j)$: Hierarchy function.
- $\Phi(o_i, o_j)$: Connectivity function.
- $\Psi(o_i, o_j)$: Indirect connectivity function.
- $\alpha$: The user-specified value.
- The value of score function is greatly determined by the combined area of $o_i$ & $o_j$.

The score function cannot get balance the combined area and pin-connectivity strength.

- Two objects with stronger connectivity cannot be clustered because they have a larger combined area.

Motivation (cont’d)

- Prohibition of large combined area.
- Balance of the connectivity and area term.
The classic multilevel placement placer [2] may generate longer wirelength since its formulation forces each bin to have an equal occupied area.

A smaller wirelength can be obtained if the spreading ranges of objects are limited.

Our Contributions

- Improve the classic analytical placement approach to obtain a smaller wirelength.
  - Propose a pin-connectivity-aware score function to consider the pin relationship between two objects to balance the connectivity and area during clustering.
  - Propose an approach to identify expected distribution ranges of objects before the analytical formulation is solved.

- Propose the net-based timing-driven placement (TDP for short) algorithm under the multilevel framework.
  - Re-define timing-critical nets and only adjust weights for TCNs to reduce its impact on wirelength.
  - Estimate the weight of a net according to its degree in addition to timing slack.
  - Propose a new historical net weight adjustment equation to get a stable weight each time a new placement is obtained.

- Experimental results demonstrate that our TDP can get better timing result than the previous timing-driven placers [3], [4].

Outline

.serialization

Introduction

Our Methodology

- Pin-connectivity-aware Score Function
- Expected Distribution Range in The Analytical Placement
- Timing-driven Global Placement Algorithm

Experimental Results

Conclusion
To balance pin-connectivity and object area, we propose pin-connectivity-aware score function \( \hat{S}(o_i, o_j) \) to improve the score function [1] as follows:

\[
\hat{S}(o_i, o_j) = G(\rho(o_i, o_j), \zeta(o_i, o_j)) \times \Gamma(o_i, o_j) \times (\Phi(o_i, o_j) + \alpha \times \Psi(o_i, o_j))
\]

- \( G(\rho, \zeta) \): To trade-off between area and pin connective according to the Gompertz curve function.
  - \( \rho(o_i, o_j) \): Area function.
- Pin-connectivity function \( \zeta(o_i, o_j) \): To estimate the true interconnection strength between \( o_i \) and \( o_j \) by considering the pin count.
The classic interconnection function $\Phi(o_i, o_j)$ simply counts the number of nets between two objects $o_i$ and $o_j$.

We propose a new pin connectivity function $\zeta(o_i, o_j)$ to measure the interconnection strength of $o_i$ and $o_j$ by considering their pin count numbers as follows:

$$\zeta(o_i, o_j) = \eta\left(\frac{U_{ij}}{\min(Y_i, Y_j)}\right)$$

- $U_{ij}$: The number of connections between $o_i$ and $o_j$.
- $Y_i (Y_j)$: The number of nets which are connected to $o_i (o_j)$.
- $\eta$ denotes the quadratic sigmoid function, where

$$\eta(x) = \begin{cases} 
1, & 1 \leq \theta \times x \\
1 - 2(\theta \times x - 1)^2, & 0.5 \leq \theta \times x \leq 1 \\
2\theta \times x^2, & 0 \leq \theta \times x \leq 0.5 \\
0, & \theta \times x < 0 
\end{cases}$$

- $\theta$ is a user-specified value ($\theta$ is set as 1.0625 in our experiment).
Pin-connectivity-aware Score Function (cont’d)

- Apply the Gompertz curve function \( G(\rho, \zeta) \) to balance the two values \( \rho(o_i, o_j) \) and \( \zeta(o_i, o_j) \) to avoid the result being overly determined by \( \rho(o_i, o_j) \) as follows:

\[
G(\rho, \zeta) = \beta + \delta \times e^{-\tau_1 e^{-\tau_2 \rho(o_i, o_j) \times \zeta(o_i, o_j)}}
\]

- \( \beta \): The Gompertz curve function bias, default as 0.1.
- \( \delta \): The magnification of the function, where \( \delta = \max\left(\frac{\text{Number of nets}}{\text{Number of objects}}, 1\right) \).
- \( \tau_1 \): The displacement along the x-axis, default as 40.
- \( \tau_2 \): The growth rate, default as 10.
Outline

Introduction

Our Methodology

- Pin-connectivity-aware Score Function
- Expected Distribution Range in The Analytical Placement
- Timing-driven Global Placement Algorithm

Experimental Results

Conclusion
Identify overflow zones $\Theta$’s after an initial placement is obtained.

- A bin $b_i$ is considered as overflowed if $D_{b_i} > M_{b_i}$.
- $\Theta$ is composed of overflow bins at contiguous locations.

Distribute objects from each overflow zone $\Theta$ to the corresponding diffusion region, denoted by $\hat{\Theta}$.

- A diffusion region is the largest area where the objects in $\Theta$ can be spread.
Procedure to Determine Expected Distribution Ranges of Objects

メディカル is set as Θ in the beginning.

メディカル diffusion region 1 is generated by gradually expanding the area of Θ ring by ring until $M_\Theta \geq D_\Theta$.
- $M_\Theta$ denotes the target placeable area in 1.
- $D_\Theta$ denotes the total cell potential in 1.

メディカル divide an overflow zone Θ into $\kappa$ sub-regions ($\kappa$ is a user-specified value and is set as 4 or 8 in our experiment) and spread objects to each direction according to the ratio of the area of original objects in the direction to the total area in Θ.
- The spreading speed in each direction should be different since objects distribution is unbalanced.
Objects are spread according to the following analytical placement formulation after all diffusion regions $\hat{\Theta}$’s are determined:

$$\min_{e_j \in E} \sum_{j} W_j(x, y) + \lambda \sum_{b} (D_b(x, y) - \hat{M}_b)^2$$

$\hat{M}_b$ is determined by the following equation:

$$\hat{M}_b = \begin{cases} M_b, & \text{if } b \in \hat{\Theta} \\ D_b, & \text{otherwise} \end{cases}$$
Outline

- Introduction
- Our Methodology
  - Pin-connectivity-aware Score Function
  - Expected Distribution Range in The Analytical Placement
  - Timing-driven Global Placement Algorithm
- Experimental Results
- Conclusion
Apply the wirelength-driven global distribution in the beginning.

Apply the timing-driven global placement when the netlist of the current level contains enough number of fine objects.

$$
\min \sum_{e_j \in E} \omega_j W_j(x, y) \\
+ \lambda \sum_b (D_b(x, y) - M_b)^2
$$

$$
\omega_j \text{ denotes the weight of } W_j(x, y) \text{ for } e_j.
$$
A net $e_j$ is considered as a timing-critical net (TCN) if one segment $s_i$ of $e_j$ belongs to a timing violation path $p_u$ and its delay $\pi_i$ meets the following condition:

- $\zeta$ is a user-specified parameter ($\zeta$ is set as 0.05 in our experiment).

$$\pi_i > \zeta \times \text{path delay of } p_u$$

The weight $\omega_j$ of a TCN $e_j$ is proportional to the slack of the path $p_u$ ($p_u \in P_j$) which has the smallest value as follows, where $P_j$ denotes a set of paths which contain the segment $s_i$ of $e_j$.

$$\omega_j \propto \max_{p_u \in P_j} \left( \frac{\sigma_u}{\sigma_{min}}, \Omega \right)$$

- $\sigma_u$ denotes the timing slack of $p_u$ and $\sigma_{min}$ represents the minimum slack among all timing violation paths.
- $\Omega$ is a user-specified parameter ($\Omega$ is set as 0.2 in our experiment).
The effect of weight adjustment will be reduced as the degree of a net increases.

The function $L(d_j)$ adjust the weight of $e_j$ according to its degree by a logistic function as follows:

$$L(d_j) = \frac{1}{1 + e^{\left(\frac{d_j - \mu_2}{\mu_1 - \mu_2}\right)}}$$

- $d_j$: the degree of $e_j$.
- $\mu_1, \mu_2$: user-specified values ($\mu_1$ and $\mu_2$ are set as 3, 5 in our experiment, respectively).
- The value of $L(d_j)$ is in the range of $[0,1]$ and the value reduces as the value $d_j$ increases.

The weight $\omega_j$ of a net $e_j$ is adjusted by the following equation:

$$\omega_j = \begin{cases} 1 + L(d_j) \max_{u \in E_j} \left( \frac{\sigma_u}{\sigma_{\min}}, \Omega \right), & \text{if } e_j \text{ is a TCN} \\ 1, & \text{otherwise} \end{cases}$$
Outline

- Introduction
- Our Methodology
- Experimental Results
- Conclusion
Experimental Results

Environments:

<table>
<thead>
<tr>
<th>Linux Workstation</th>
<th>Programming Language</th>
<th>CPU</th>
<th>Memory</th>
<th>System</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C++</td>
<td>Intel® Xeon® Silver 4110 2.1GHz</td>
<td>400GB</td>
<td>Cent OS 6.9</td>
</tr>
</tbody>
</table>

This experiment is divided into two parts.

- Effect of improved analytical placement approach with respect to classic analytical placement approach.
- Comparison with other timing-driven placers.
Effect of Improved Analytical Placement Approach

- Benchmarks: Industrial designs.
  - Placement legalization and signal routing are completed by IC Compiler 2.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th># Movable Macros</th>
<th># Preplaced Macros</th>
<th># IO Pads</th>
<th># Standard Cells</th>
<th># Nets</th>
<th># Pins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cir. 1</td>
<td>29</td>
<td>2</td>
<td>684</td>
<td>111339</td>
<td>116928</td>
<td>477911</td>
</tr>
<tr>
<td>Cir. 2</td>
<td>81</td>
<td>5</td>
<td>1219</td>
<td>340130</td>
<td>365605</td>
<td>1328667</td>
</tr>
<tr>
<td>Cir. 3</td>
<td>48</td>
<td>101</td>
<td>19323</td>
<td>497704</td>
<td>509005</td>
<td>2182323</td>
</tr>
<tr>
<td>Cir. 4</td>
<td>146</td>
<td>36</td>
<td>190</td>
<td>1425892</td>
<td>1471214</td>
<td>6037846</td>
</tr>
<tr>
<td>Cir. 5</td>
<td>855</td>
<td>69</td>
<td>263</td>
<td>5555610</td>
<td>5797026</td>
<td>22340831</td>
</tr>
</tbody>
</table>
Demonstrate the effectiveness of two techniques associated with wirelength.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Classic Analytical Placement</th>
<th>with Pin-connectivity-aware Score Function</th>
<th>with Expected Distribution Range of Objects</th>
<th>Our Analytical Placement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WL</td>
<td>RT</td>
<td>WL</td>
<td>RT</td>
</tr>
<tr>
<td>Cir. 1</td>
<td>0.622</td>
<td>205.73</td>
<td>0.634</td>
<td>149.91</td>
</tr>
<tr>
<td>Cir. 2</td>
<td>1.453</td>
<td>1158.76</td>
<td>1.417</td>
<td>1085.73</td>
</tr>
<tr>
<td>Cir. 3</td>
<td>1.799</td>
<td>333.17</td>
<td>1.764</td>
<td>257.85</td>
</tr>
<tr>
<td>Cir. 4</td>
<td>8.877</td>
<td>3051.78</td>
<td>8.713</td>
<td>2985.94</td>
</tr>
<tr>
<td>Cir. 5</td>
<td>42.447</td>
<td>24569.8</td>
<td>40.432</td>
<td>20320.6</td>
</tr>
<tr>
<td>Nor.</td>
<td>1.16</td>
<td>1.12</td>
<td>1.13</td>
<td>0.95</td>
</tr>
</tbody>
</table>

WL: wirelength \((\times 10^7 \mu m)\).
RT: runtime (sec).
## Comparison with Other Timing-driven Placers

- **Benchmarks:** ICCAD 2015 contest benchmark suite.
- **STA engine:** an open-source timer, OpenTimer.
- **Placement legalization is performed using Innovus.**
- **The timing slack was estimated by the evaluation script provided by the ICCAD 2015 contest.**

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>#Cells</th>
<th>#Nets</th>
<th>#Pins</th>
<th>#Rows</th>
</tr>
</thead>
<tbody>
<tr>
<td>superblue1</td>
<td>1209716</td>
<td>1215710</td>
<td>3767494</td>
<td>1829</td>
</tr>
<tr>
<td>superblue3</td>
<td>1213253</td>
<td>1224979</td>
<td>3905321</td>
<td>1840</td>
</tr>
<tr>
<td>superblue4</td>
<td>795645</td>
<td>802513</td>
<td>2497940</td>
<td>1840</td>
</tr>
<tr>
<td>superblue5</td>
<td>1086888</td>
<td>1100825</td>
<td>3246878</td>
<td>2528</td>
</tr>
<tr>
<td>superblue7</td>
<td>1931639</td>
<td>1933945</td>
<td>6372094</td>
<td>3163</td>
</tr>
<tr>
<td>superblue10</td>
<td>1876103</td>
<td>1898119</td>
<td>5560506</td>
<td>3437</td>
</tr>
<tr>
<td>superblue16</td>
<td>981559</td>
<td>999902</td>
<td>3013268</td>
<td>1788</td>
</tr>
<tr>
<td>superblue18</td>
<td>768068</td>
<td>771542</td>
<td>2559143</td>
<td>1788</td>
</tr>
</tbody>
</table>
We compare our TDP with state-of-the-art timing-driven placers including DREAMPlace4.0 [4] and Differentiable TDP [3].

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WNS</td>
<td>TNS</td>
<td>WNS</td>
</tr>
<tr>
<td>superblue1</td>
<td>-14.10</td>
<td>-85.03</td>
<td>-10.77</td>
</tr>
<tr>
<td>superblue4</td>
<td>-12.78</td>
<td>-144.38</td>
<td>-8.49</td>
</tr>
<tr>
<td>superblue5</td>
<td>-26.76</td>
<td>-95.78</td>
<td>-25.21</td>
</tr>
<tr>
<td>superblue7</td>
<td>-15.21</td>
<td>-63.86</td>
<td>-15.22</td>
</tr>
<tr>
<td>superblue10</td>
<td>-31.88</td>
<td>-768.75</td>
<td>-21.97</td>
</tr>
<tr>
<td>superblue16</td>
<td>-12.11</td>
<td>-124.18</td>
<td>-10.85</td>
</tr>
<tr>
<td>superblue18</td>
<td>-11.87</td>
<td>-47.25</td>
<td>-7.99</td>
</tr>
<tr>
<td>Nor.</td>
<td>1.32</td>
<td>1.63</td>
<td>1.01</td>
</tr>
</tbody>
</table>

- **WNS**: worst negative slack ($\times 10^3$ ps).
- **TNS**: total negative slack ($\times 10^5$ ps).


Outline

- Introduction
- Our Methodology
- Experimental Results
- Conclusion
Conclusion

✧ We have proposed an improved analytical placement algorithm.
  ✧ Pin-connectivity-aware score function.
  ✧ Expected distribution range of objects.

✧ We have proposed a TDP algorithm based on an improved analytical placement algorithm.
  ✧ Determine timing net weights only for TCNs.
  ✧ Estimate net weights based on the degree of a net in addition to consideration of timing slacks.
  ✧ Propose a new equation to update net weights based on their historical values.

✧ The experimental results have demonstrated that our approach can obtain better results than the previous approaches.
Thank You For Your Attention