Attacking Split Manufacturing from a Deep Learning Perspective

Haocheng Li\textsuperscript{1}, Satwik Patnaik\textsuperscript{2}, Abhrajit Sengupta\textsuperscript{2}, Haoyu Yang\textsuperscript{1}, Johann Knechtl\textsuperscript{3}, Bei Yu\textsuperscript{1}, Evangeline F. Y. Young\textsuperscript{1}, Ozgur Sinanoglu\textsuperscript{3}

\textsuperscript{1}The Chinese University of Hong Kong
\textsuperscript{2}New York University
\textsuperscript{3}New York University Abu Dhabi
Split Manufacturing

Hardware is vulnerable with un-trusted foundries $^{ab}$.

Split manufacturing safeguards chip designs $^{cd}$.

$^a$[Durvaux and Standaert 2016]
$^b$[Shamsi et al. 2019]
$^c$[McCants 2011]
$^d$[Bi, Yuan, and Jin 2015]

Figure 1: Wire width in Nangate 45 nm open cell library.
Threat Model

Available: FEOL design, cell library, database of layouts generated in a similar manner.

Objective: correct connection rate \( ^a \)

\[
CCR = \frac{\sum_{i=1}^{m} c_i x_i}{\sum_{i=1}^{m} c_i} ,
\]

\( m \) is the number of sink fragments, \( c_1, c_2, \ldots, c_m \) are the numbers of sinks in every fragment, \( x_i = 1 \) when a positive virtual pin pair (VPP) is selected for the \( i \)-th sink fragment, \( x_i = 0 \) when a negative VPP is selected for the \( i \)-th sink fragment.

\(^a\) [Wang et al. 2018]
Design and train a deep neural network to predict the missing BEOL connections.

The neural network makes use of both vector-based and image-based features.

Propose softmax regression loss to select best connection among variable-size candidates.

Figure 3: Attack flow.
Vector-based Features

- Distances for VPPs along both directions.
- Numbers of sinks connected within the fragments.
- Maximum capacitance of the driver and pin capacitance of the sinks.
- Wirelength and via contribution in each FEOL metal layer.
- Driver delay according to the underlying timing paths.
Image-based Features

Figure 4: Layout Image Scaling.

Figure 5: Layout Image Representation.
Sample Selection

Figure 6: All VPPs are considered as candidates except VPP (Source A, Sink B).

Table 1: VPP Preferences

<table>
<thead>
<tr>
<th>Sink</th>
<th>Source</th>
<th>Sink Prefers Source</th>
<th>Source Prefers Sink</th>
<th>Direction Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B</td>
<td>A</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>B</td>
<td>B</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Model Architecture

[Diagram showing a neural network structure with input vector features, input source images, and input sink image.]

Figure 7: Neural Network Structure.
Figure 8: Neural Network Architecture.
Softmax Regression Loss

The loss of the two-class classification is

$$l_r = -\frac{1}{n} \left( \log \frac{e_{s_i}^+}{e_{s_i}^- + e_{s_i}^+} + \sum_{j\neq t} \log \frac{e_{s_j}^-}{e_{s_j}^- + e_{s_j}^+} \right), \quad (2)$$

whose partial derivative is

$$\frac{\partial l_r}{\partial s_j^+} = - \frac{\partial l_r}{\partial s_j^-} = \begin{cases} - \frac{e_{s_j}^-}{n \left( e_{s_j}^- + e_{s_j}^+ \right)} & \text{if } j = t, \\ \frac{e_{s_j}^+}{n \left( e_{s_j}^- + e_{s_j}^+ \right)} & \text{otherwise}. \end{cases} \quad (3)$$

The partial derivative in the last FC layer is

$$\frac{\partial l_r}{\partial w_i^+} = - \frac{\partial l_r}{\partial w_i^-} = \frac{1}{n} \left( \sum_{j=1}^{n} \frac{e_{s_j}^+ x_{i,j}}{e_{s_j}^- + e_{s_j}^+} - x_{i,t} \right). \quad (4)$$

We propose the following *softmax regression loss*

$$l_c = -\log \frac{e_{s_t}}{\sum_{j=1}^{n} e_{s_j}}, \quad (5)$$

whose partial derivative is

$$\frac{\partial l_c}{\partial s_j} = \begin{cases} \frac{e_{s_j}}{\sum_{j=1}^{n} e_{s_j}} - 1 & \text{if } j = t, \\ \frac{e_{s_j}}{\sum_{j=1}^{n} e_{s_j}} & \text{otherwise}. \end{cases} \quad (6)$$

The partial derivative in the last FC layer is

$$\frac{\partial l_c}{\partial w_i} = \frac{\sum_{j=1}^{n} e_{s_j} x_{i,j}}{\sum_{j=1}^{n} e_{s_j}} - x_{i,t}. \quad (7)$$
Experimental Results

Average Ratio

M1 CCR
- Wang
- Ours

M3 CCR

M1 Time

M3 Time

1.21
1.12
1

1 \cdot 10^{-3}

2 \cdot 10^{-3}

Wang

Ours

M1 CCR (%)

b7
b11
b13
b14
b15_1
b17_1
b18
c432
c880
c1355
c1908
c2670
c3540
c5315
c6288
c7552

0 5 10 15
Experimental Results

Figure 9: Comparison between different settings of techniques used.
Conclusion

- Demonstrate vector-based and image-based features.
- Process these heterogeneous features simultaneously in a neural network.
- Propose a softmax regression loss.
Thanks!

Questions?
References I


