Attacking Split Manufacturing from a Deep Learning Perspective

Haocheng Li[1], Satwik Patnaik[2], Abhrajit Sengupta[2], Haoyu Yang[1], Johann Knechtel[3], Bei Yu[1], Evangeline F. Y. Young[1], Ozgur Sinanoglu[3]

[1]The Chinese University of Hong Kong
[2]New York University
[3]New York University Abu Dhabi

Split Manufacturing

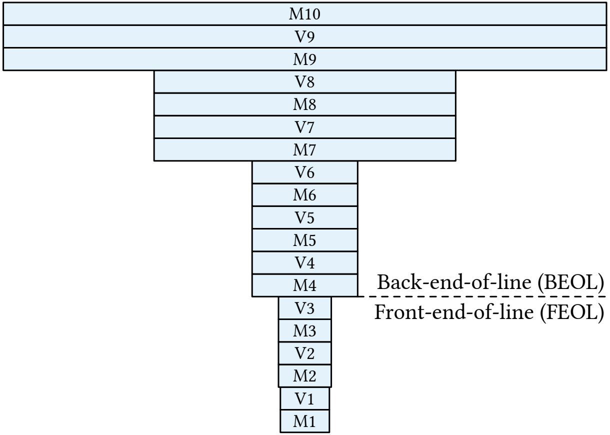


Figure 1: Wire width in Nangate 45 nm open cell library.

- Hardware is vulnerable with un-trusted foundries [a,b].
- Split manufacturing safeguards chip designs [c,d].

---

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

Threat Model

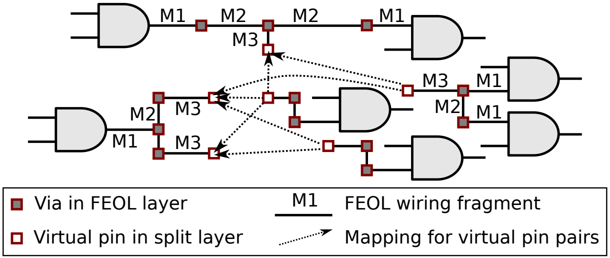


Figure 2: Two source fragments and three sink fragments.

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

Objective: correct connection rate [a]

$$CCR = \frac{\sum_{i=1}^{m} c_i x_i}{\sum_{i=1}^{m} c_i}, \qquad (1)$$

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]

Contributions

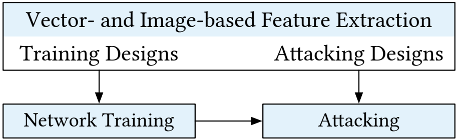


Figure 3: Attack flow.

- 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.

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

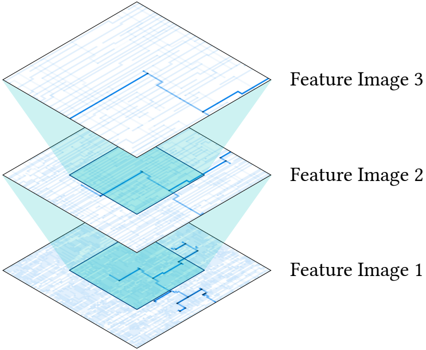


Figure 4: Layout Image Scaling.

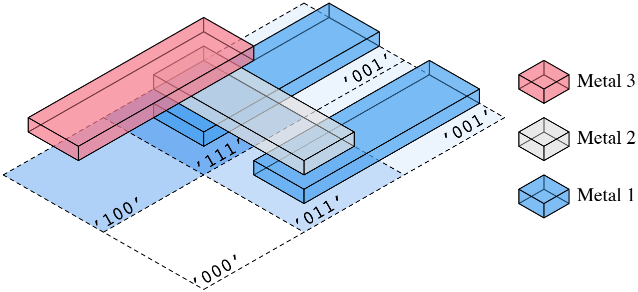


Figure 5: Layout Image Representation.

Sample Selection

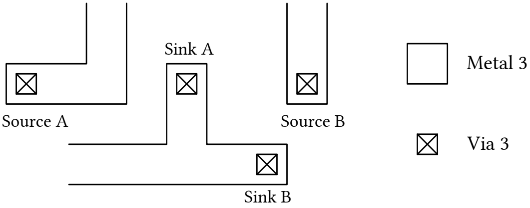


Figure 6: All VPPs are considered as candidates except VPP (Source A, Sink B).

Table 1: VPP Preferences

Sink	Source	Sink Prefers Source	Source Prefers Sink	Direction Criterion
A	A	✓	✗	✓
A	B	✓	✓	✓
B	A	✗	✗	✗
B	B	✓	✓	✓

Model Architecture

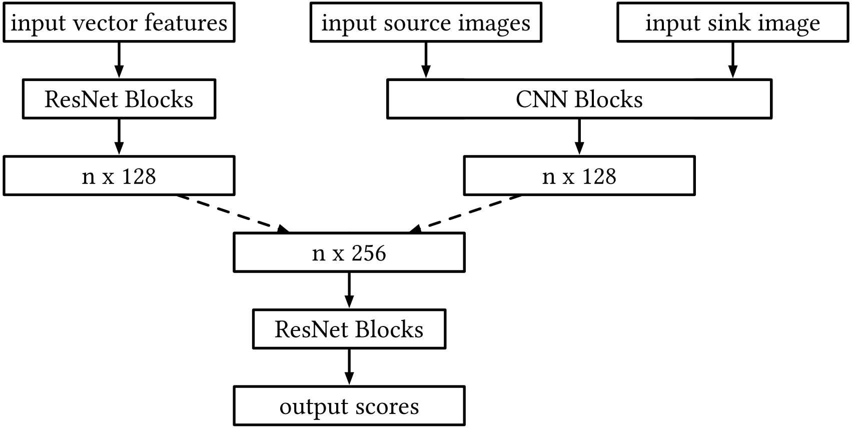


Figure 7: Neural Network Structure.

Model Architecture

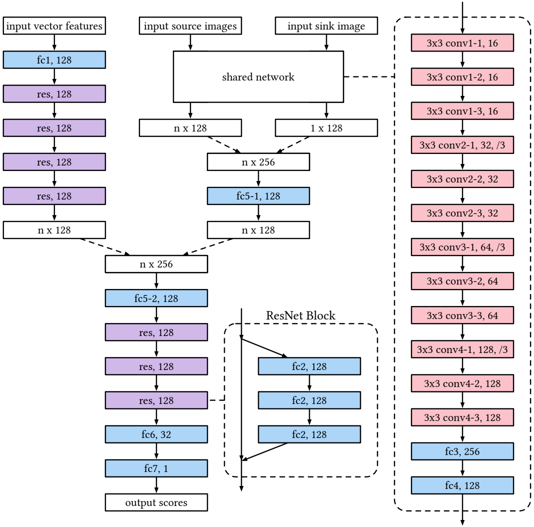


Figure 8: Neural Network Architecture.

Softmax Regression Loss

The loss of the two-class classification is

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

whose partial derivative is

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

The partial derivative in the last FC layer is

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

We propose the following *softmax regression loss*

$$l_c = -\log\frac{e^{s_t}}{\sum_{j=1}^n e^{s_j}}, \tag{5}$$

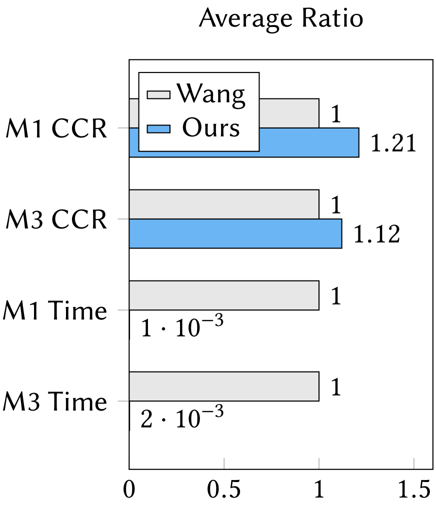
whose partial derivative is

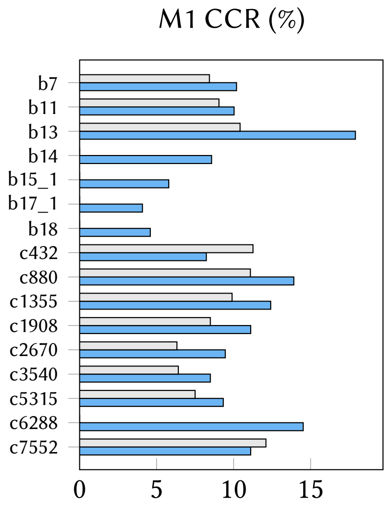
$$\frac{\partial l_c}{\partial s_j} = \begin{cases} \dfrac{e^{s_j}}{\sum_{j=1}^n e^{s_j}} - 1 & \text{if } j=t, \\ \dfrac{e^{s_j}}{\sum_{j=1}^n e^{s_j}} & \text{otherwise.} \end{cases} \tag{6}$$

The partial derivative in the last FC layer is

$$\frac{\partial l_c}{\partial w_i} = \frac{\sum_{j=1}^n e^{s_j}x_{i,j}}{\sum_{j=1}^n e^{s_j}} - x_{i,t}. \tag{7}$$

Experimental Results





Experimental Results

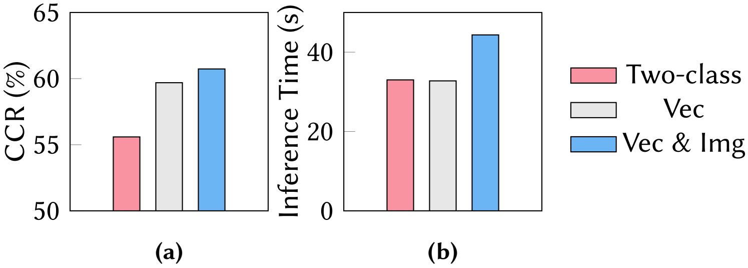


Figure 9: Comparison between different settings of techniques used.

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

- Bi, Yu, Jiann Yuan, and Yier Jin (2015). "Beyond the interconnections: Split manufacturing in RF designs". In: *Electronics* 4.3, pp. 541–564.
- Durvaux, François and François-Xavier Standaert (2016). "From improved leakage detection to the detection of points of interests in leakage traces". In: *Annual International Conference on the Theory and Applications of Cryptographic Techniques.* Springer, pp. 240–262.
- McCants, C (2011). "Trusted integrated chips (TIC)". In: *Intelligence Advanced Research Projects Activity (IARPA), Tech. Rep.*
- Shamsi, Kaveh, Travis Meade, Meng Li, David Z. Pan, and Yier Jin (2019). "On the approximation resiliency of logic locking and IC camouflaging schemes". In: *IEEE Transactions on Information Forensics and Security* 14.2, pp. 347–359.
- Wang, Yujie, Pu Chen, Jiang Hu, Guofeng Li, and Jeyavijayan Rajendran (2018). "The cat and mouse in split manufacturing". In: *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 26.5, pp. 805–817.