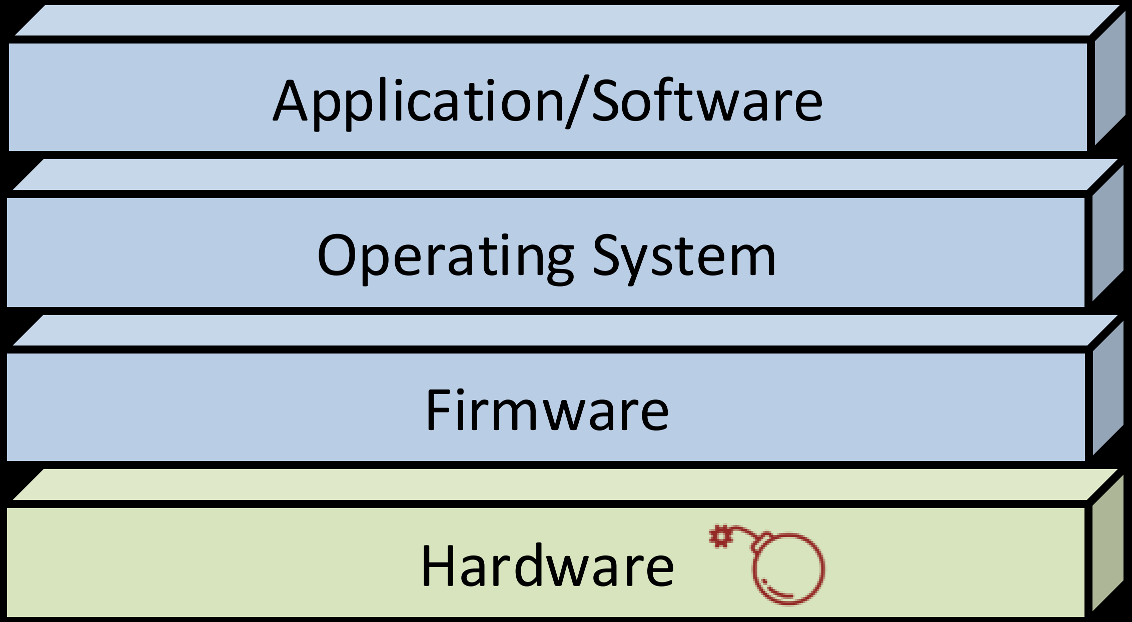
INSIGHT: Attacking Industry-Adopted Learning Resilient Logic Locking Techniques Using Explainable Graph Neural Network

Likhitha Mankali[1], Ozgur Sinanoglu[2], Satwik Patnaik[3]

[1]New York University, [2]New York University Abu Dhabi, [3]University of Delaware

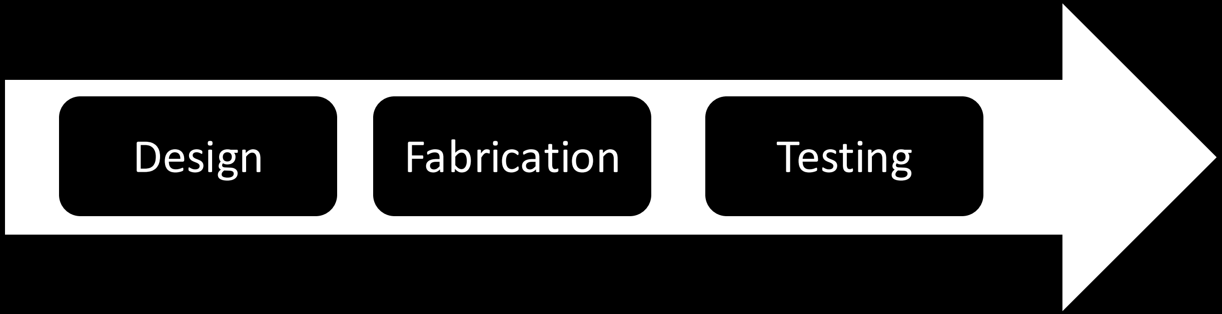
Semiconductors in Everyday Lives

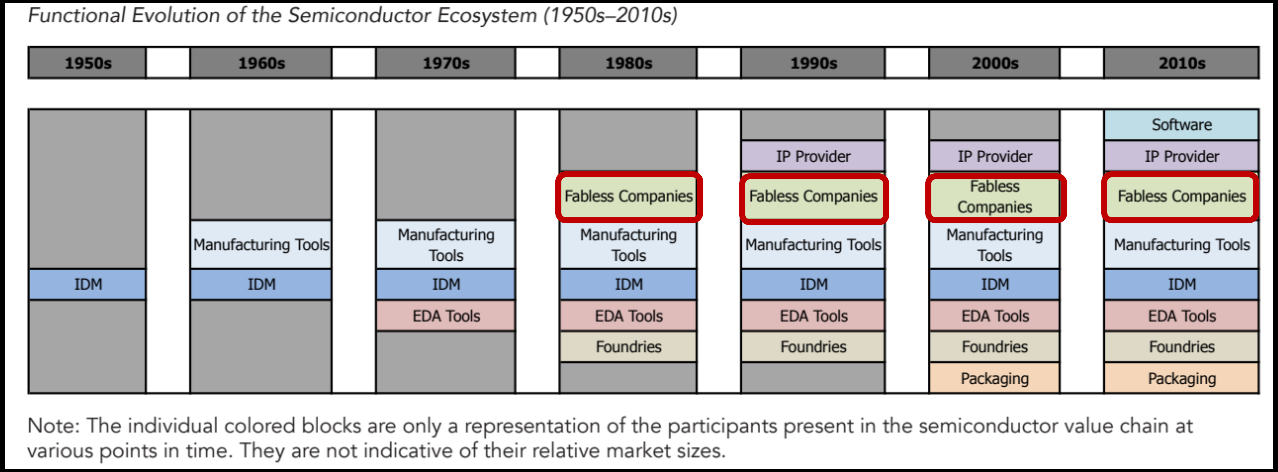
- Attacks can be launched via any abstraction layer
- Software bugs patched using updates
- What if hardware is compromised?



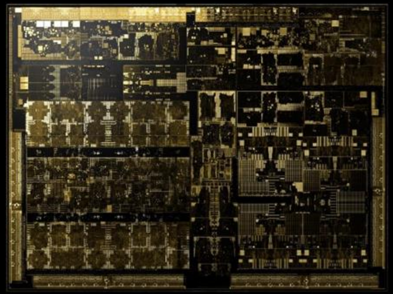
Protection of Hardware is essential

Globalized IC Supply Chain



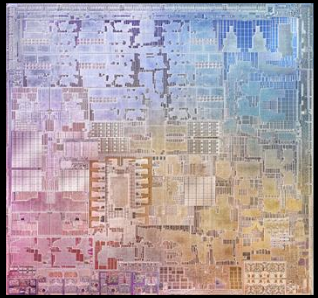


2018



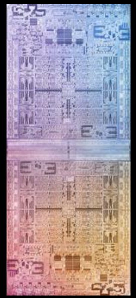
NVIDIA Xavier
(9B, 12nm)

2020

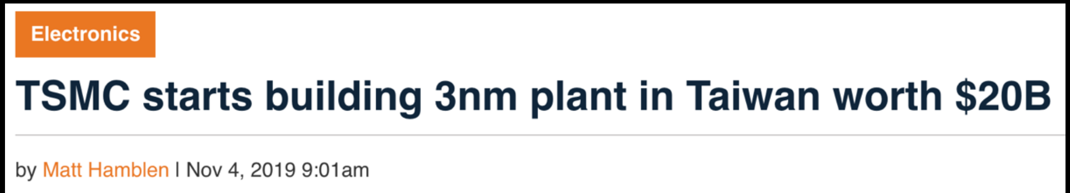


Apple M1
(16B, 5nm)

2022



Apple M1 Ultra
(114B, 5nm)

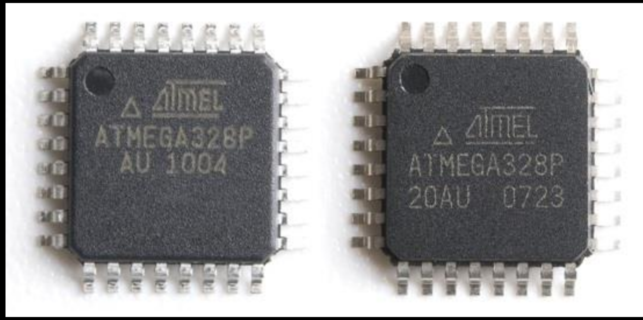


Electronics

TSMC starts building 3nm plant in Taiwan worth $20B

by Matt Hamblen | Nov 4, 2019 9:01am

Hardware Security Threats



IC counterfeiting

Hardware IP piracy

Hardware Trojans

Reverse-engineering

OPINION

The overlooked security risks of onshoring chip production

Here are four ways manufacturers can mitigate cybersecurity risks.

Published Dec. 19, 2023

Source: supplychaindive

Hardware IP Piracy

Are these threats real?

US DoJ indicates prominent IC design company sufferend a loss of around 8.75 billion dollars due to IP theft

DARPA

Automatic Implementation of Secure Silicon (AISS)

Dr. Lok Yan

A CROSS-LAYER FRAMEWORK FOR COST-EFFECTIVE INTELLECTUAL PROPERTY (IP) PROTECTION

EDA Forms The Basis For Designing Secure Systems

181 Shares 23 23 in 51

How to accelerate the design process at a lower cost and with less risk.

AUGUST 3RD, 2020 - BY: **ADAM CRON**

source: semiengineering

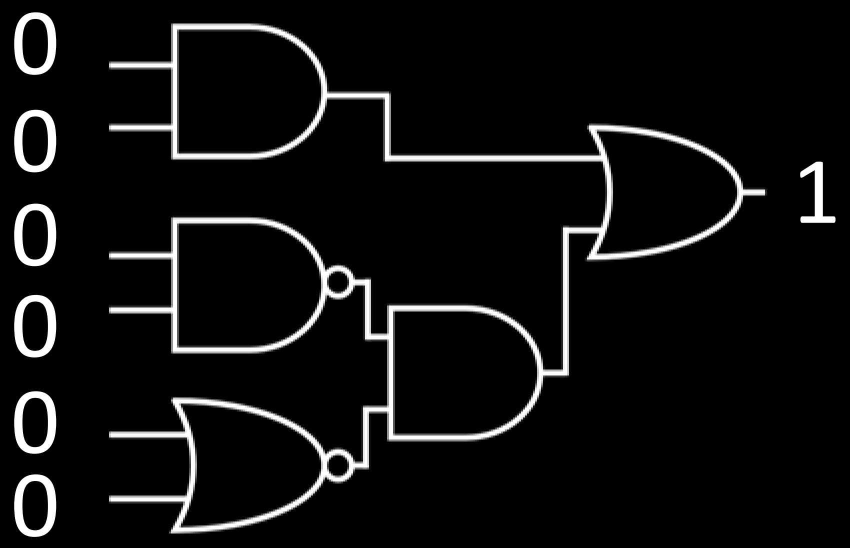
SYNOPSYS®

Mentor Graphics®

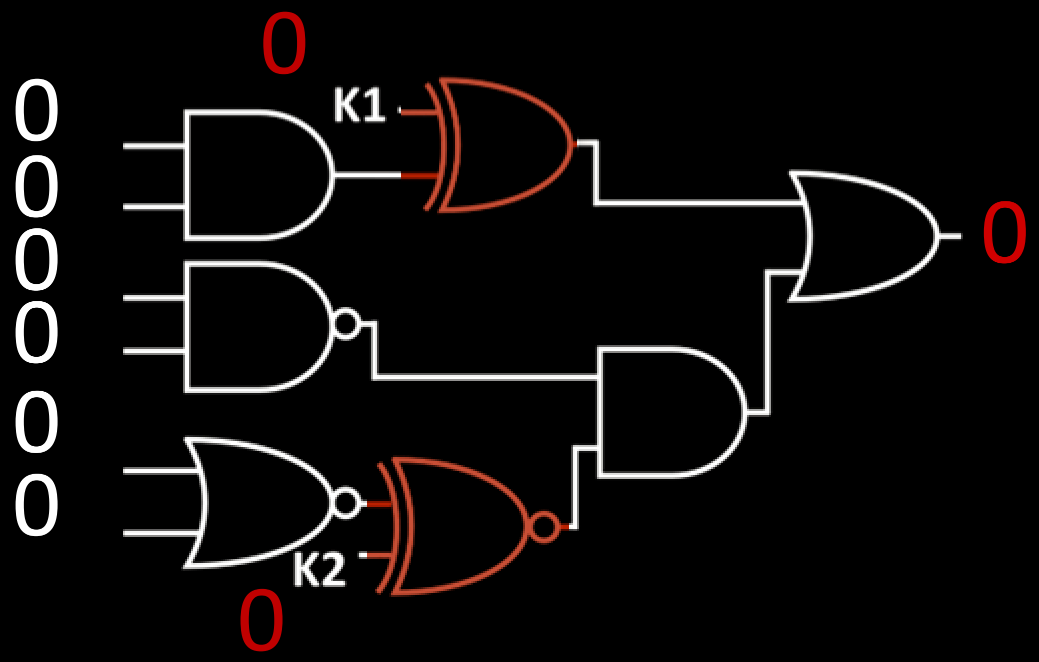
SAHARA

STRUCTURED ARRAY HARDWARE for AUTOMATICALLY REALIZED APPLICATIONS

Logic Locking



Original Design



Logic-locked Design

Tamper-proof Memory

Key inputs

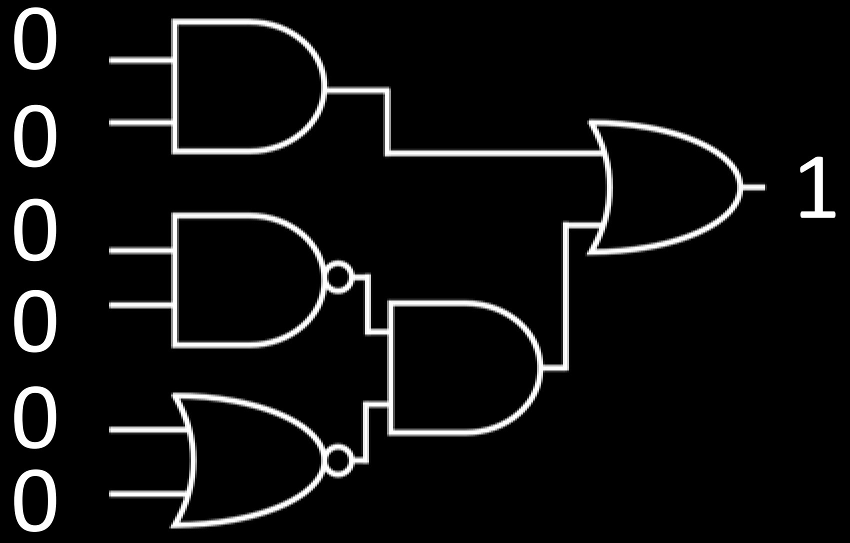
Original Design

Logic Locking

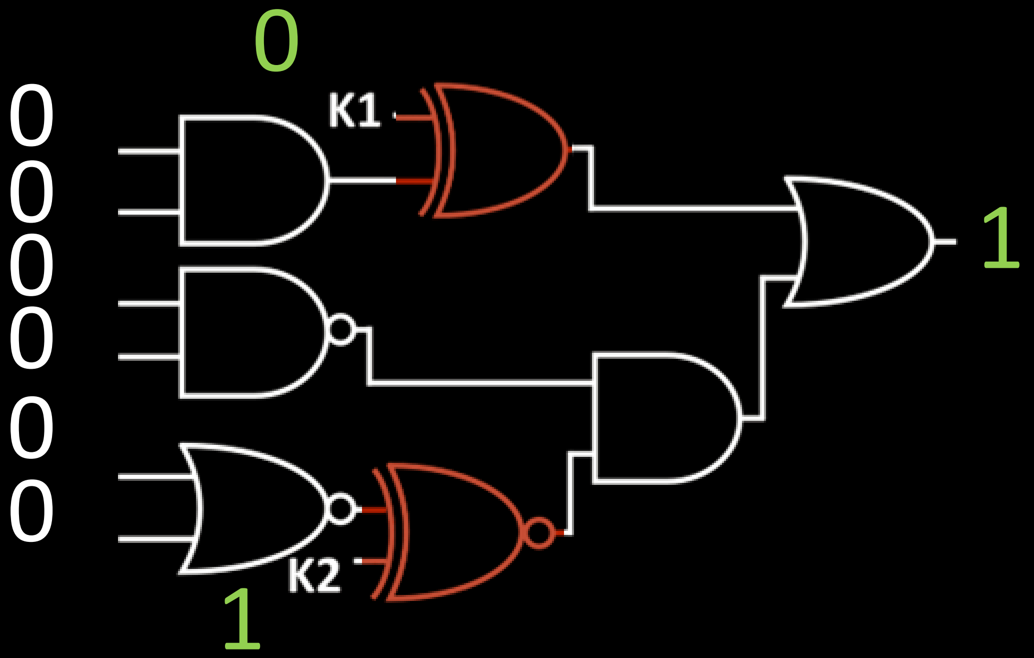
Locked Design

Logic locking hides the functionality of a design by inserting key-gates

Logic Locking



Original Design



Logic-locked Design

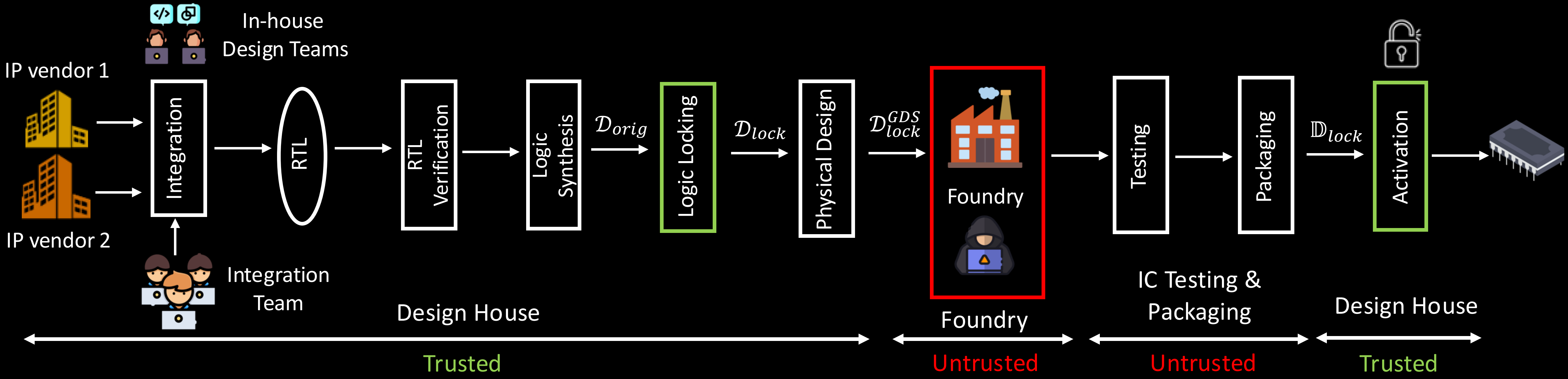
Tamper-proof Memory

Key inputs

Original Design → Logic Locking → Locked Design

Logic locking hides the functionality of a design by inserting key-gates

Threat Model



Attacker's Resources

- Locked design
  - Obtained by reverse-engineering the chip

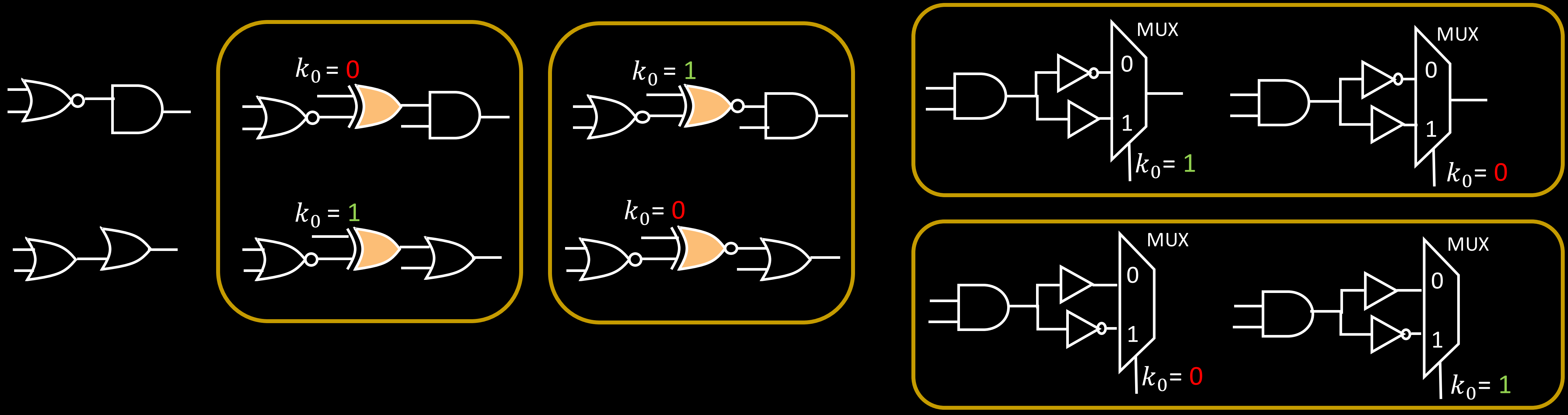
Attacker's Capabilities

- Analyze reverse-engineered locked design

Attacker's Objective

- To recover the secret key → Hardware IP piracy

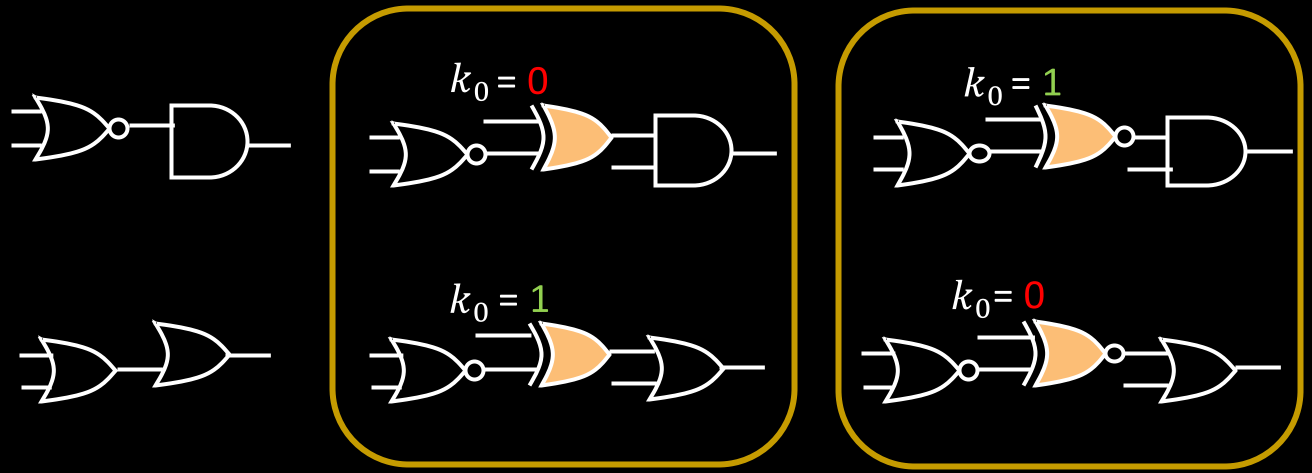
Learning Resilient Logic Locking



Removes correlation between key-gate type and key-value

N. Kavand, et al., ICCAD'22; L. Alrahis, et al.; TIFS'21; N. Limaye, et al., TCAD'21; N. Limaye, et al., TCAD'22; A. B. Chowdhury, et al., DAC'23; F. Wang, et al., ISPD'23
P. Chakraborty, et al., TIFS'21; A. Alaql, et al., TVLSI'21; L. Alrahis, et al., TCAS-II'22; D. Sisejkovic et al., JETC'21; L. Alrahis, et al., DATE'22;

Learning Resilient Logic Locking



TRLL | TRLL+ | ALMOST
RGLock | TroMUX | UNSAIL
SimLL

Removes correlation between key-gate type and key-value

✗ - Failed to attack

SAIL	SnapShot	SCOPE	OMLA	MuxLink
✗	✗	✗	✗	✗

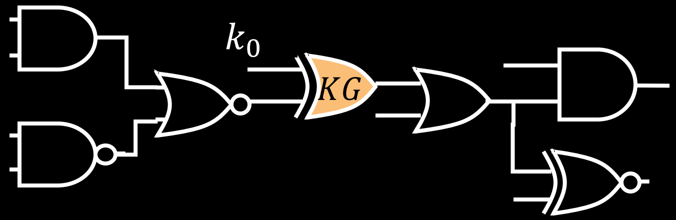
N. Kavand, et al., ICCAD'22; L. Alrahis, et al.; TIFS'21; N. Limaye, et al., TCAD'21; N. Limaye, et al., TCAD'22; A. B. Chowdhury, et al., DAC'23; F. Wang, et al., ISPD'23
P. Chakraborty, et al., TIFS'21; A. Alaql, et al., TVLSI'21; L. Alrahis, et al., TCAS-II'22; D. Sisejkovic et al., JETC'21; L. Alrahis, et al., DATE'22;

Preliminary Problem Modeling

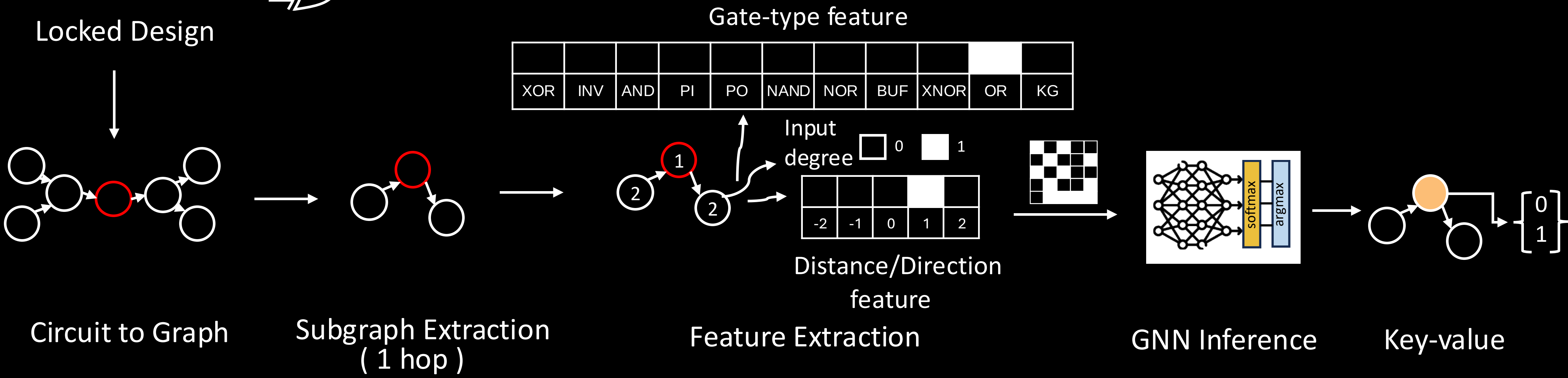
GNN-based Key-Prediction

- Maps the problem of key-prediction to GNN-based node classification

Preliminary problem modeling obtains KPA of **~50%** (random guess)

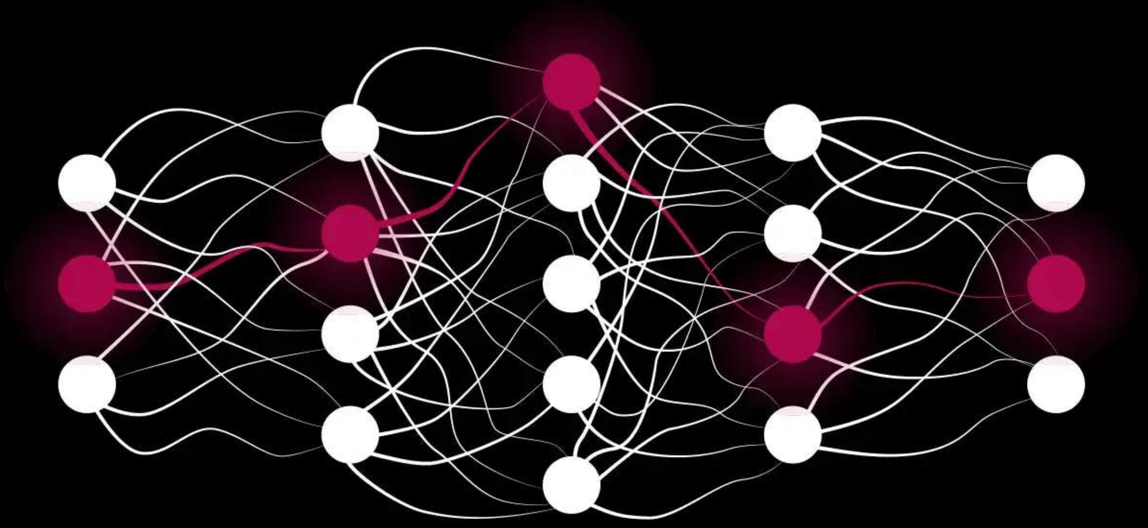


Locked Design



We employ explainable ML to find the reasons behind failure of the attack?

Why Explainable ML?



Black-box

Complex computations

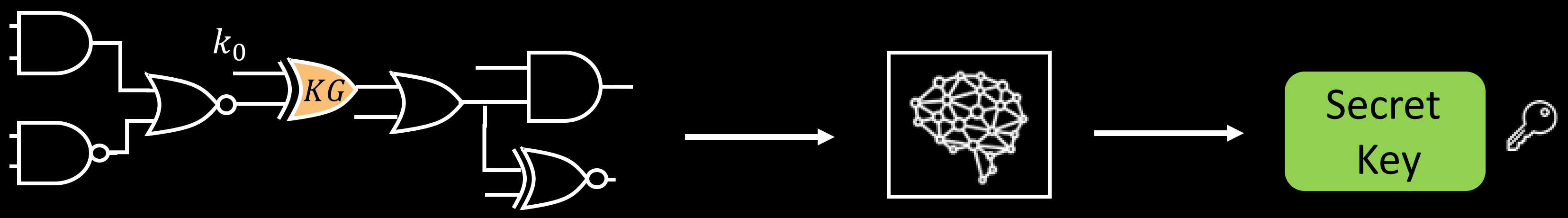
Outputs only accuracy

Explainable ML provides reasons behind the prediction

Important Features

Important Nodes

INSIGHT



Locked Design

Explainable GNN -based Framework

Challenge 1: To select a suitable explainer

Solution 1: Perform ablation study on explainers

Observation

GNNExplainer

- is more suitable for our work (provides better explanations)
- is computationally efficient (**600x** better than SubgraphX)

GNNExplainer

SubgraphX

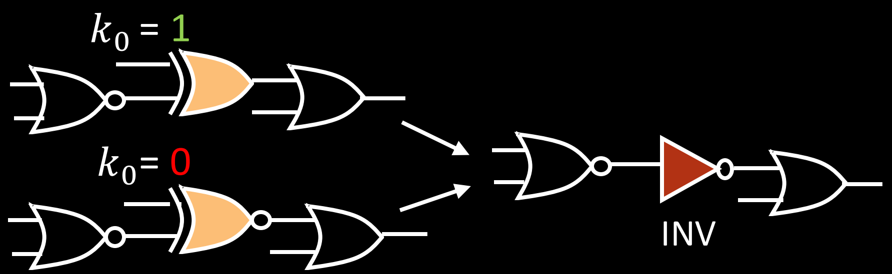
ZORRO

PGExplainer

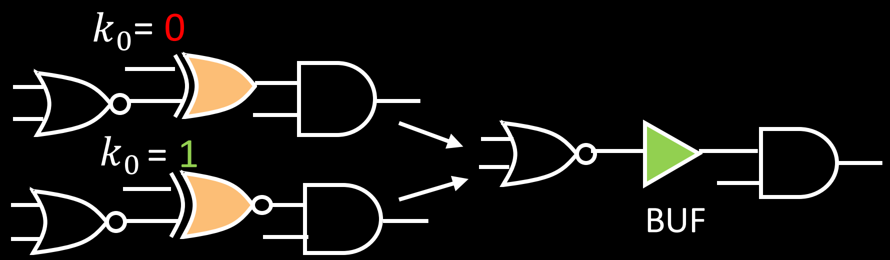
D. Luo, et al., NIPS'20; H. Yuan, et al., ICML'21; Z. Ying, et al., NeurIPS'19; T. Funke, et al., TKDE'23

INSIGHT

Challenge 2: To identify reasons behind the failure of attack through explanations



Solution 2: We map key-prediction problem to INV/BUF prediction problem



Design	b14_C	b15_C	b17_C	b20_C	b21_C	b22_C
Solution 2	99.78	99.68	99.06	99.53	99.53	99.53

KPA improves by **1.86x**

INSIGHT

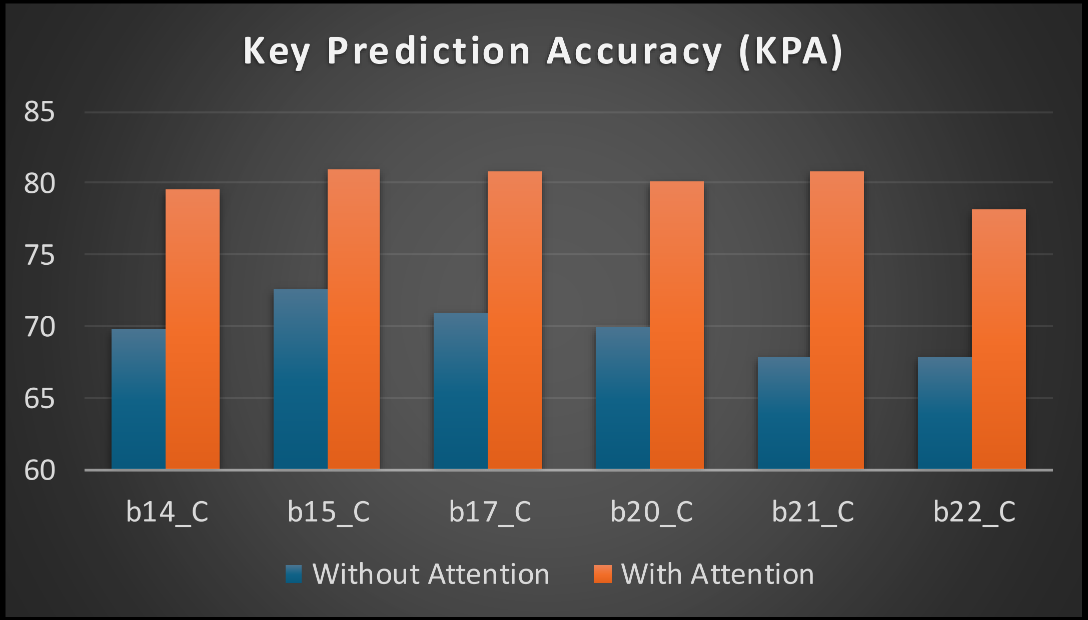
Semiconductor industry re-synthesizes designs upon logic locking

Challenge 3: To tackle logic re-synthesized designs

Observation: Explainer analysis indicates different importance scores for the gates around key-gate

Solution 3: Added attention layer to the GNN

Incorporating attention increases KPA by **10%** for re-synthesized designs



INSIGHT

Challenge 4: To tackle insufficient training data

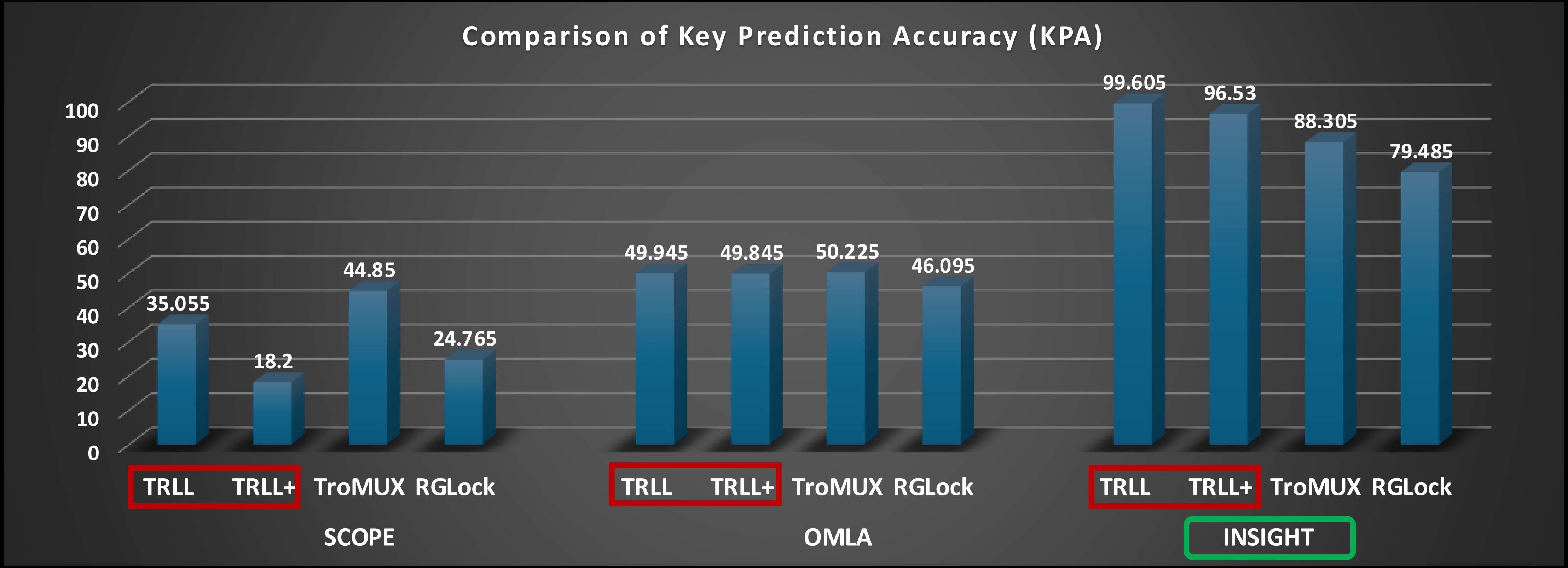
Solution 4: We incorporate two approaches
- Data augmentation
- Semi-supervised learning

Design	b14_C	b15_C	b17_C	b20_C	b21_C
No Data Augmentation	69.05	72.13	66.67	66.38	67.72
Solution 4	76.56	82.79	71.87	72.56	74.83
Improvement (x)	1.11 x	1.15 x	1.08 x	1.09 x	1.10 x

Data augmentation increases KPA by **1.10x**

Semi-supervised learning increases KPA by **1.29x**

Results



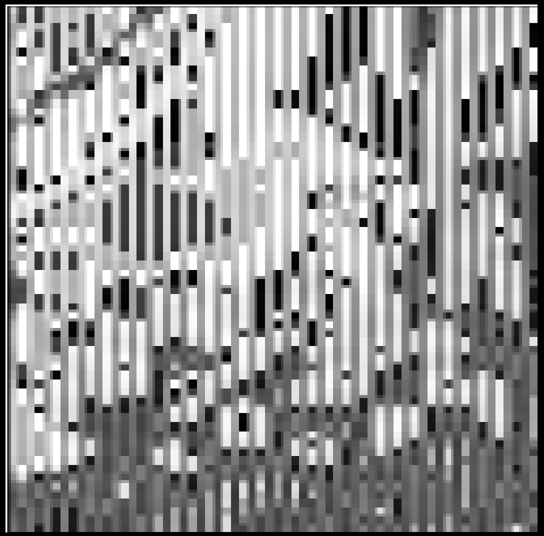
INSIGHT achieves KPA of **2.96x** and **1.86x** than SCOPE and OMLA

Results (Real-World Application)

Gaussian Blurring Example

Input

Golden Output



OMLA’s Output

INSIGHT’s Output

INSIGHT recovers secret key in practical designs

Thank You!

Likhitha Mankali

lm4344@nyu.edu

NYU TANDON SCHOOL OF ENGINEERING

CONTACT