Data Efficient Lithography Modeling with Residual Neural Networks and Transfer Learning

Yibo Lin[1], Yuki Watanabe[2], Taiki Kimura[2], Tetsuaki Matsunawa[2], Shigeki Nojima[2], Meng Li[1], David Z. Pan[1]

[1]ECE Department, University of Texas at Austin

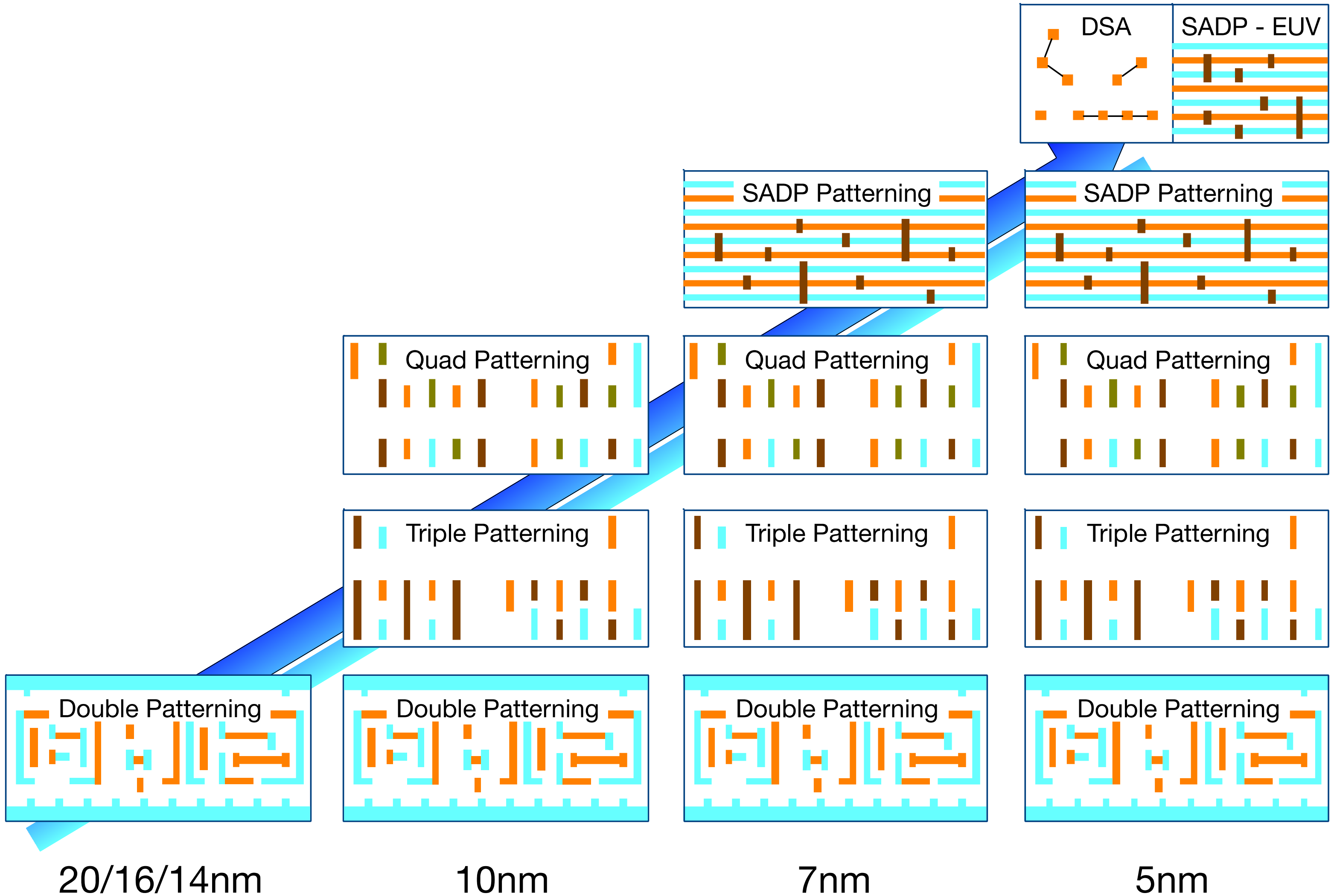
[2]Toshiba Memory Corporation

http://yibolin.com

Outline

- Introduction
- Problem Formulation
- Related Work
- Data Efficient Lithography Modeling
- Experiment Results
- Conclusion

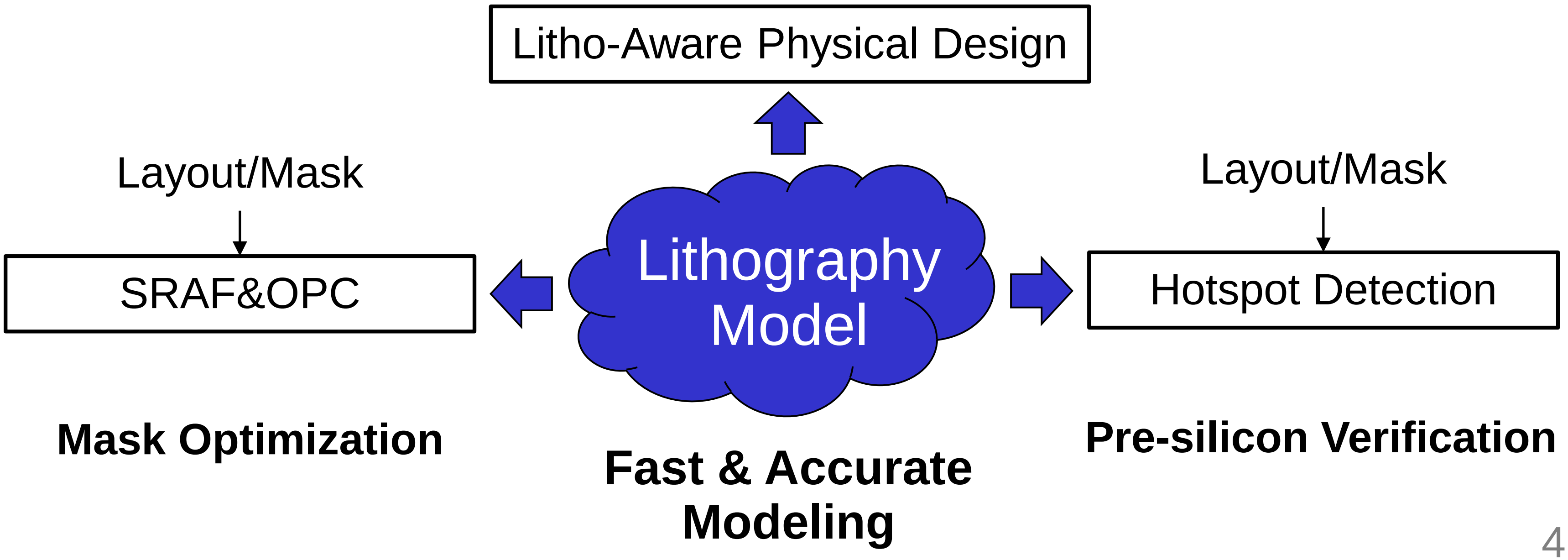
Advanced Lithography with Scaling

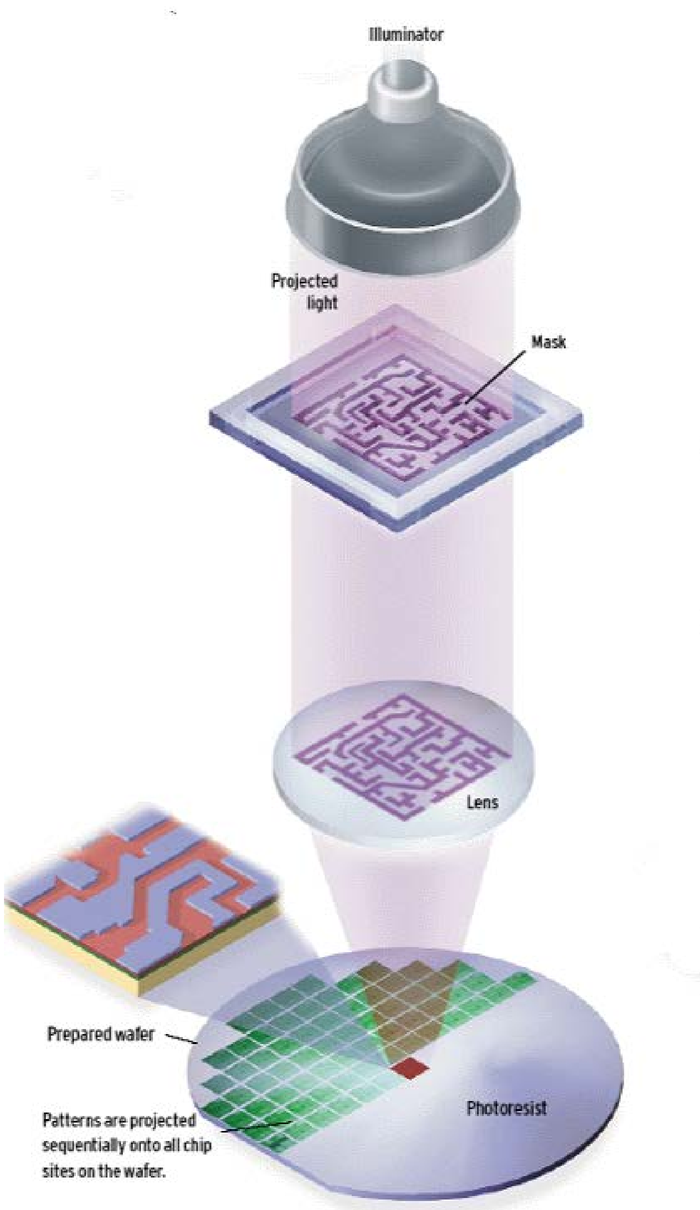


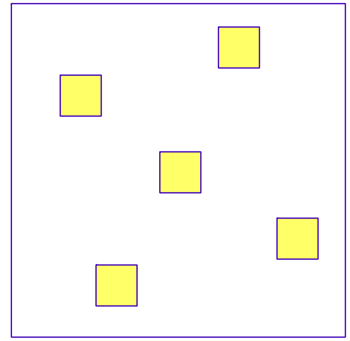
[Courtesy Mentor Graphics]

Design and Manufacturing Challenges

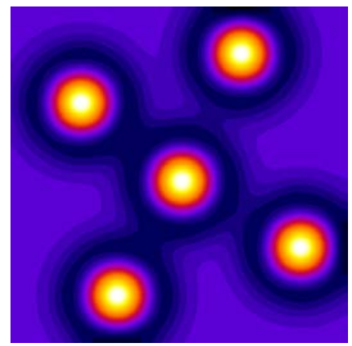
- Design Closure
  - Lithography aware physical design
  - Reduce turn-around time
- Manufacturing/yield Closure
  - Pre-silicon verification, e.g., hotspot detection
  - Fast and effective mask optimization, e.g., SRAF, OPC



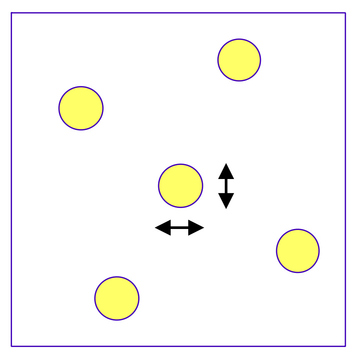




Contact Mask

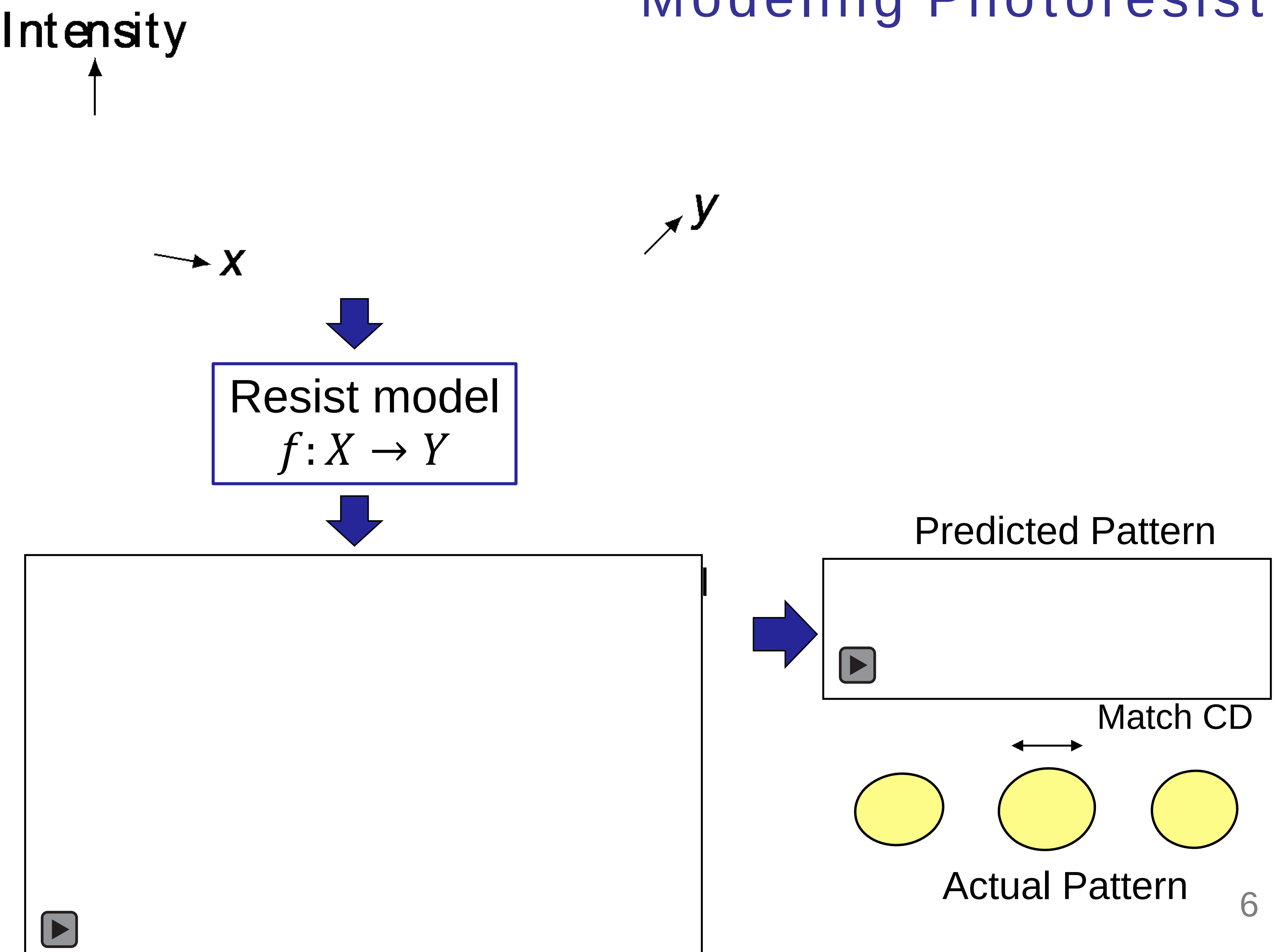


Aerial Image
(Light intensity map)



Resist Pattern

Modeling Photoresist



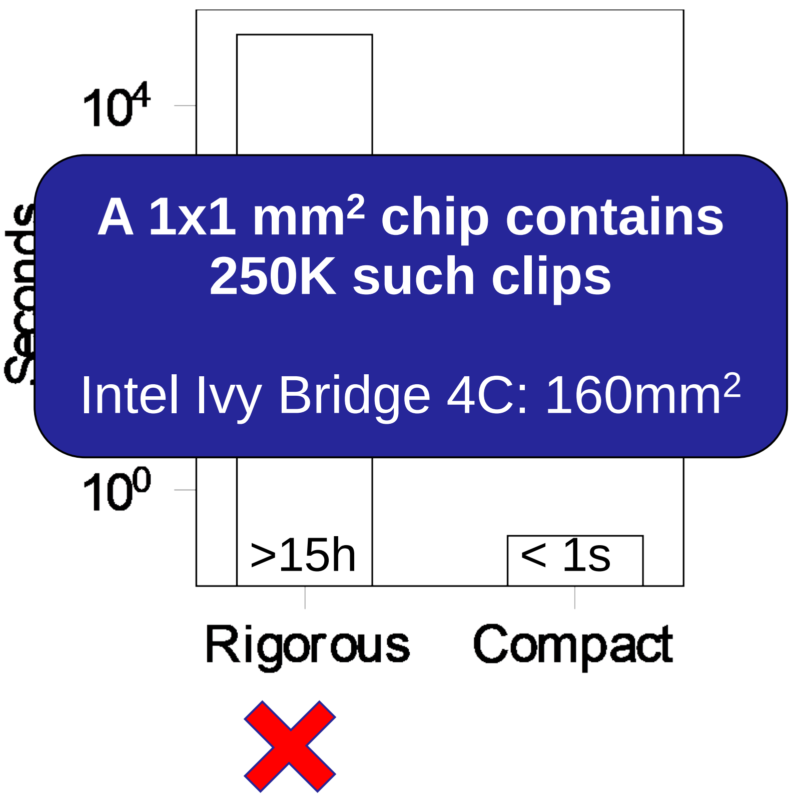
Challenges in Lithography Modeling

	Rigorous Simulation	Compact Model
Accuracy	High	Medium
Prediction Efficiency	Low	High
Data Demands	Medium	High

Rigorous simulation: physics-level simulation, e.g., Synopsys Sentaurus Lithography
Compact model: e.g., Mentor Graphics Calibre, machine learning models

Prediction Efficiency

For 1K 2x2 um^2 clips



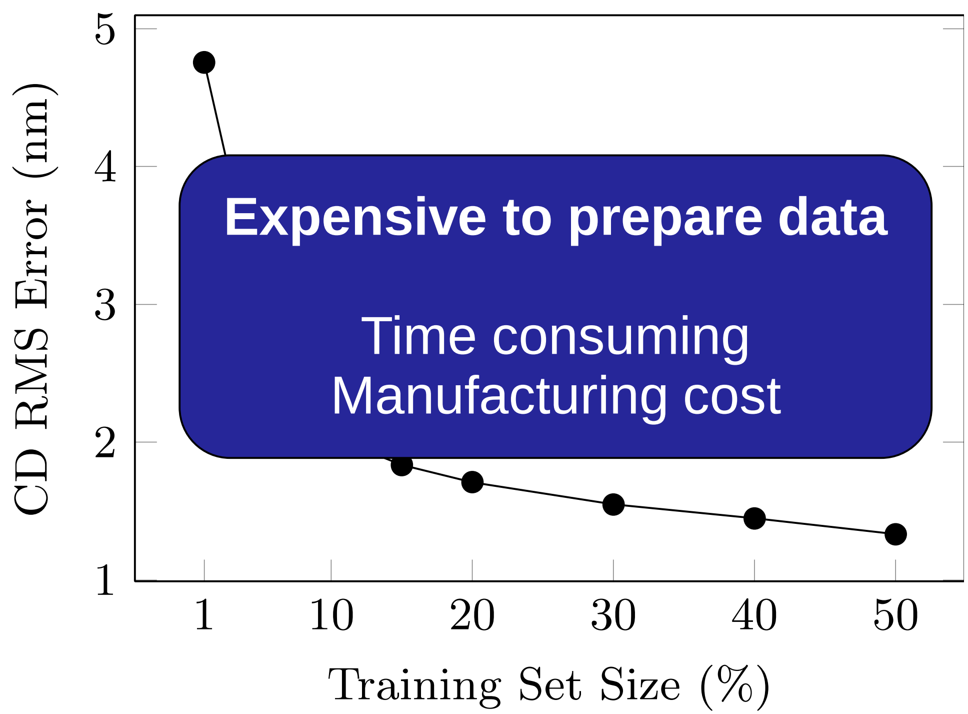
A 1x1 mm^2 chip contains 250K such clips

Intel Ivy Bridge 4C: 160mm^2

Data Demanding

Compact Model

High target accuracy → Require **big** training **data**



Expensive to prepare data

Time consuming
Manufacturing cost

Previous Study on ML-based Modeling

Convolutional neural networks

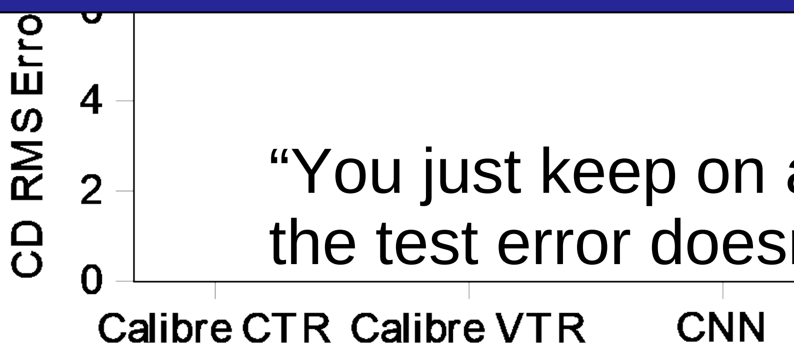
- [Watanabe+, SPIE'17]
- Task: threshold prediction
- 3 convolutional layers
- 2 fully connected layers

Artificial neural networks

- [Shim+, SPIE'17]
- Task: resist height prediction
- 5 hidden layers

Neural networks are getting **deeper** for higher accuracy

AlexNet-8, VGG-19, ResNet-101, ResNet-1202



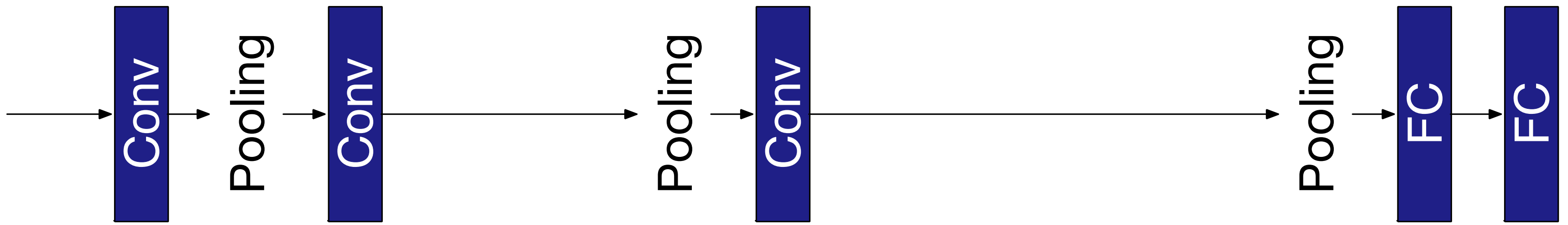
"You just keep on adding layers, until the test error doesn't improve anymore."

– Yoshua Bengio

Pitfalls in Deeper Neural Networks

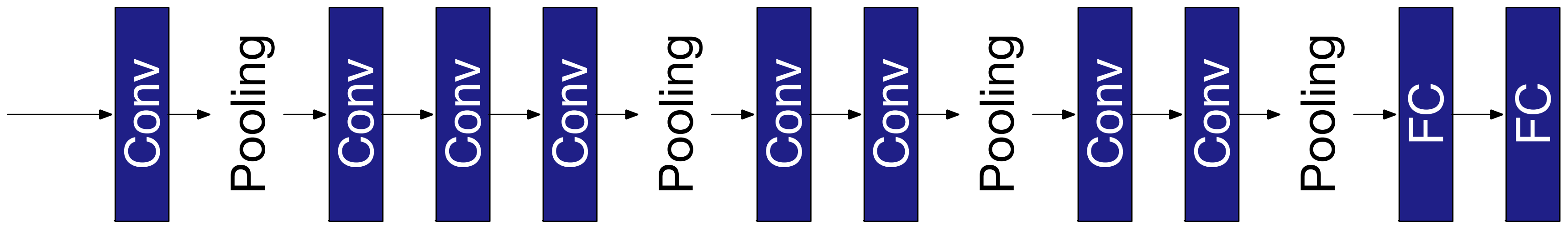
- Larger model capacity

CNN-5



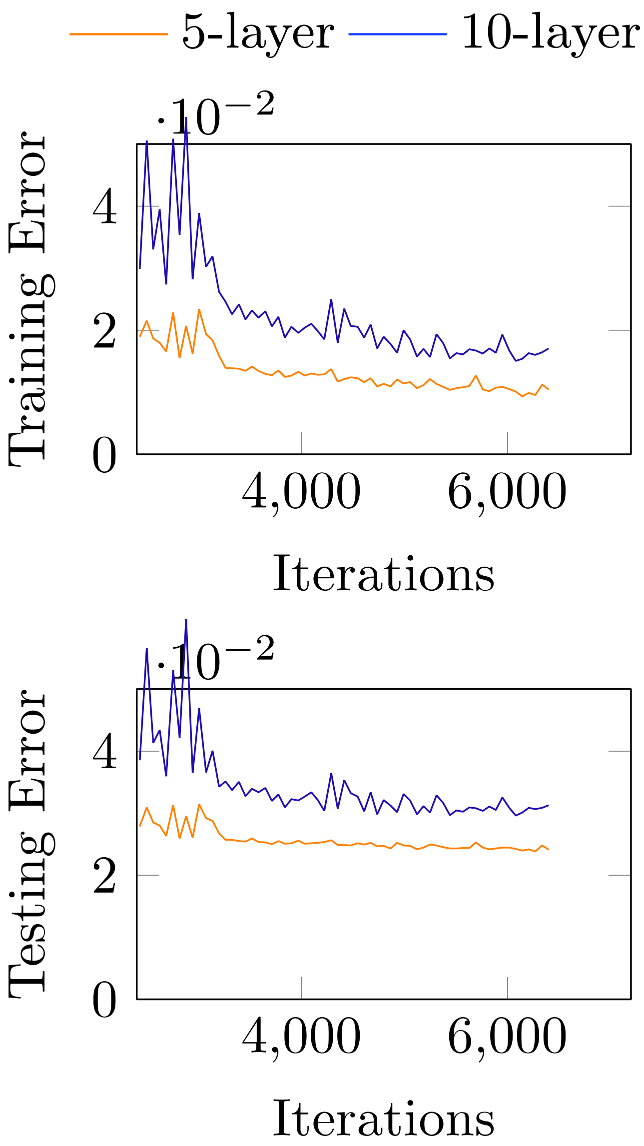
[Watanabe+, SPIE'17]

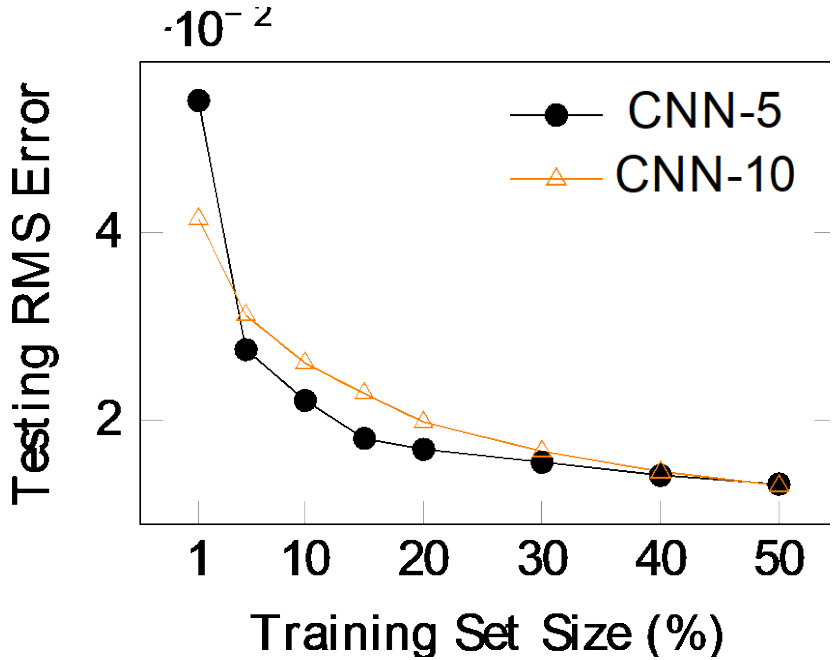
CNN-10



Pitfalls in Deeper Neural Networks

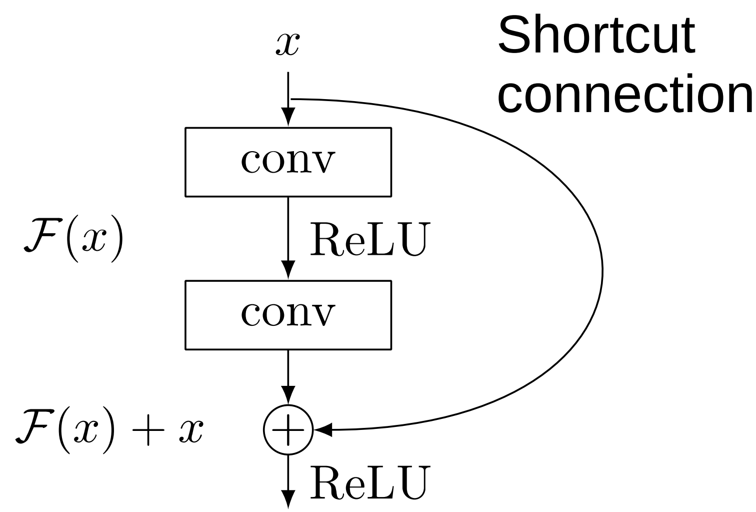
- Larger model capacity
- Gradient vanishing



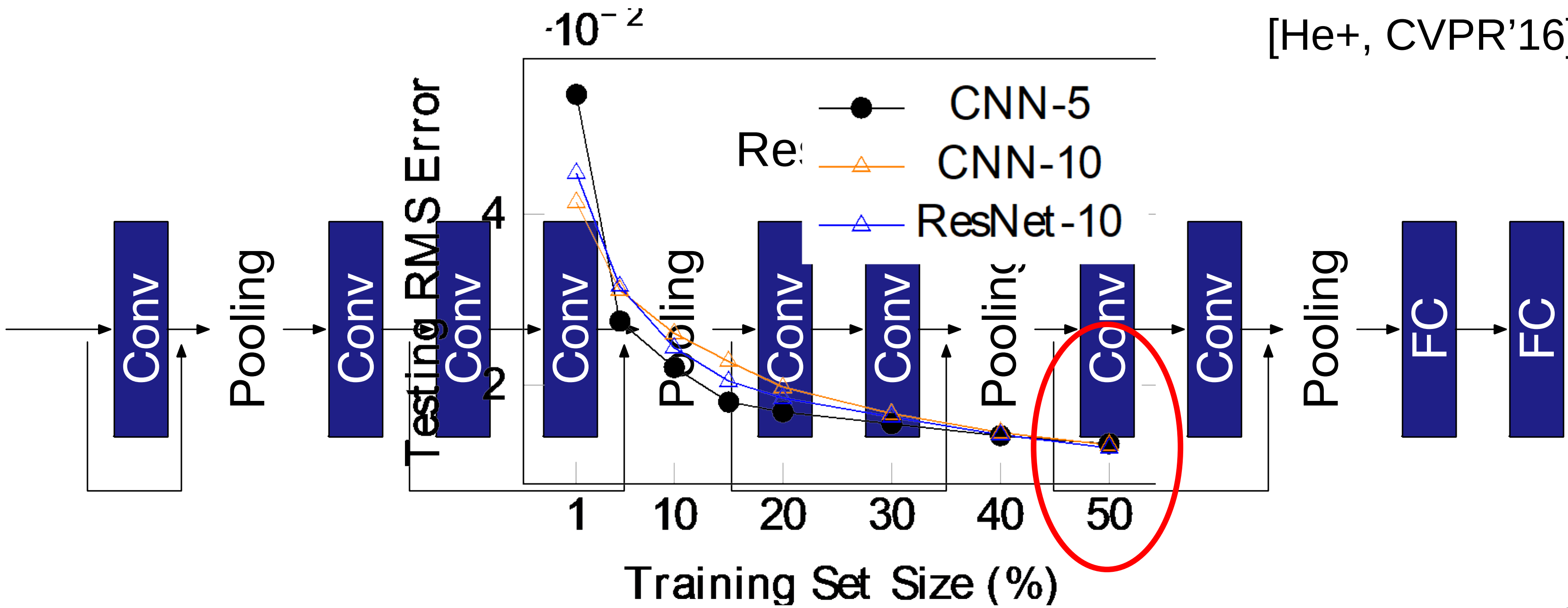


Pitfalls in Deeper Neural Networks

- Larger model capacity
- Gradient vanishing
  - ResNet
- Overfitting
- Require **MORE** data



[He+, CVPR'16]

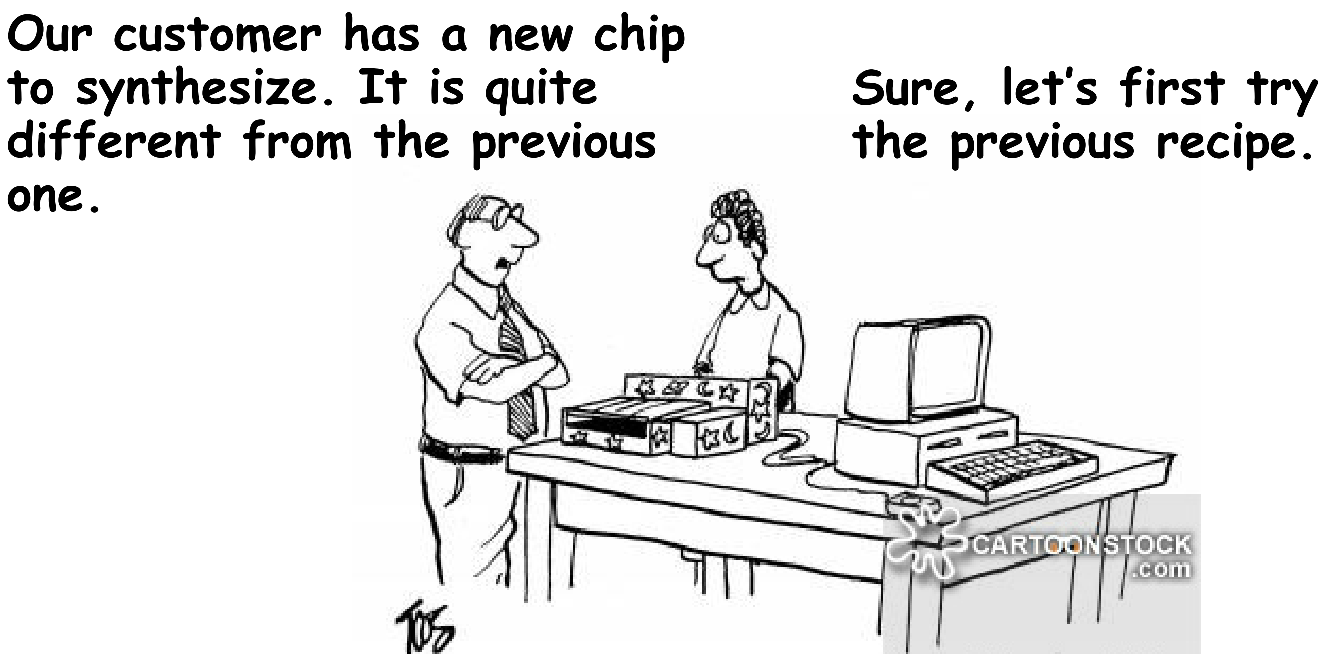


Do We Have Big Data?

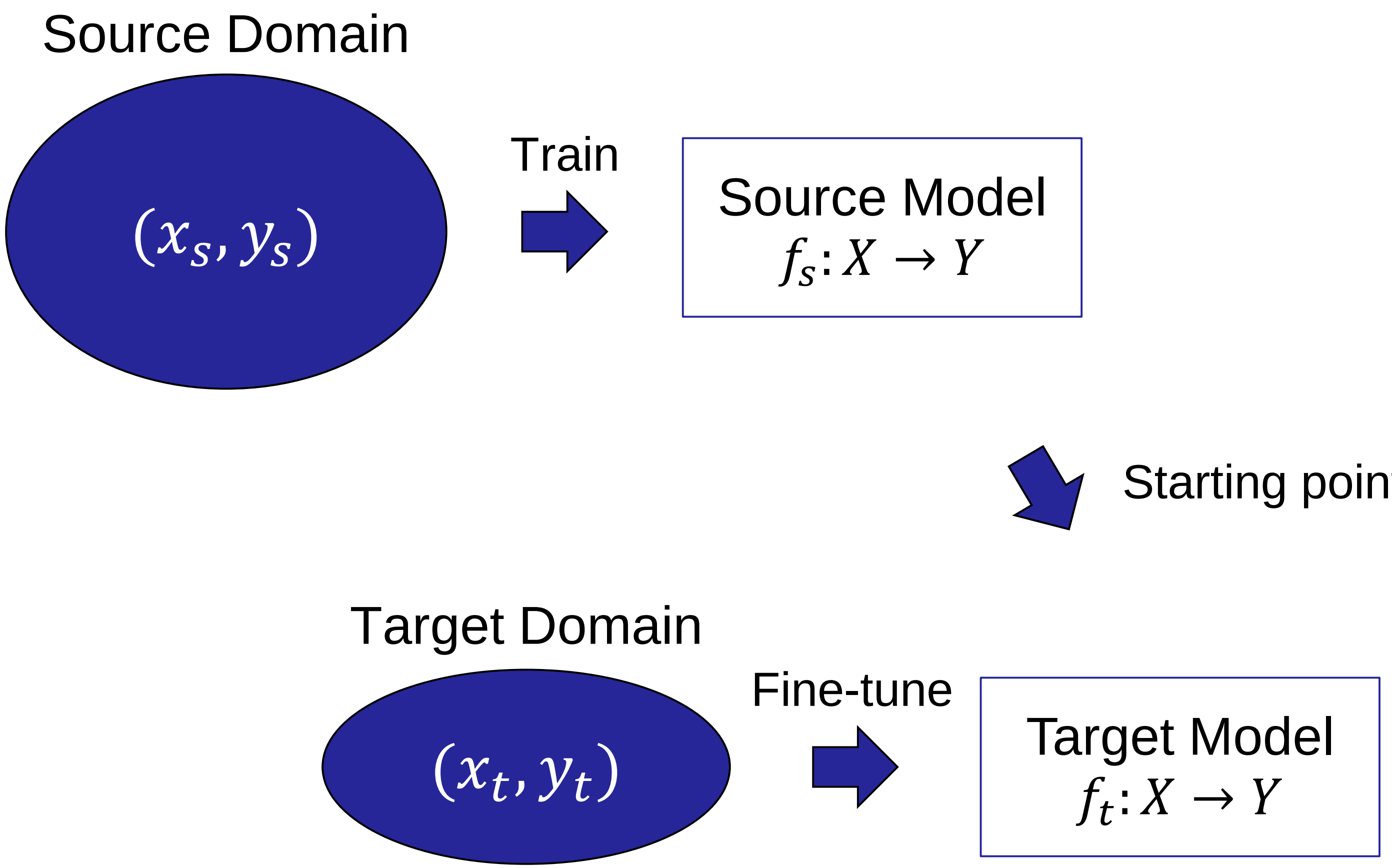
- Yes, but **old data**
- Different design rules
- Different manufacturing configurations

Can We Use Old Data?

- It depends…
- How different the old data is from the new data?
- Worth trying

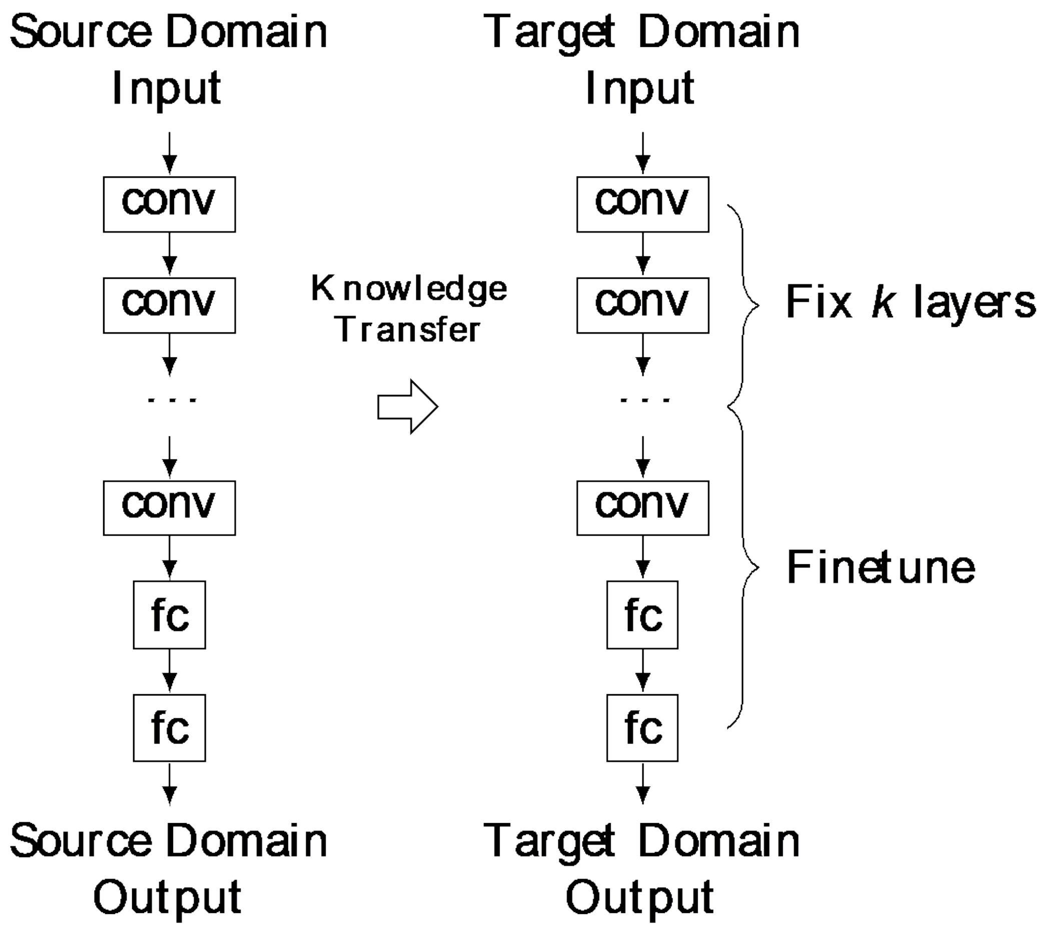


Transfer Learning from Source to Target



Transfer Learning from Source to Target

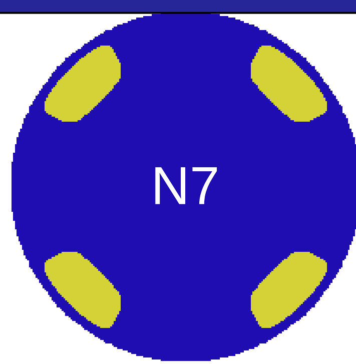
TF_k Scheme



Technology Transition from N10 to N7

Contact Layer Design Rules [Liebmann, SPIE'15]		
	N10	N7
Patterning	LELE	LELELE

	N10	$N7_a$	$N7_b$
Design Rule	A	B	B
Optical Source	A	B	B
Resist Material	A	A	B



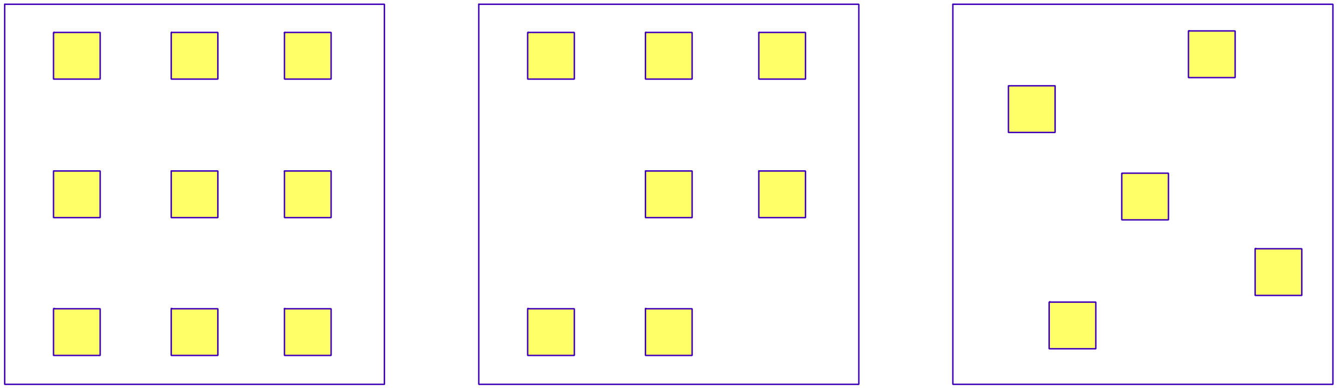
Resist A

Resist B

Different dissolution slopes

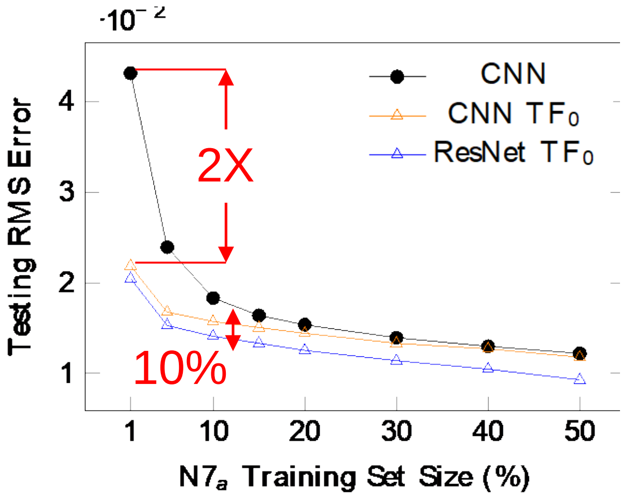
Technology Transition from N10 to N7

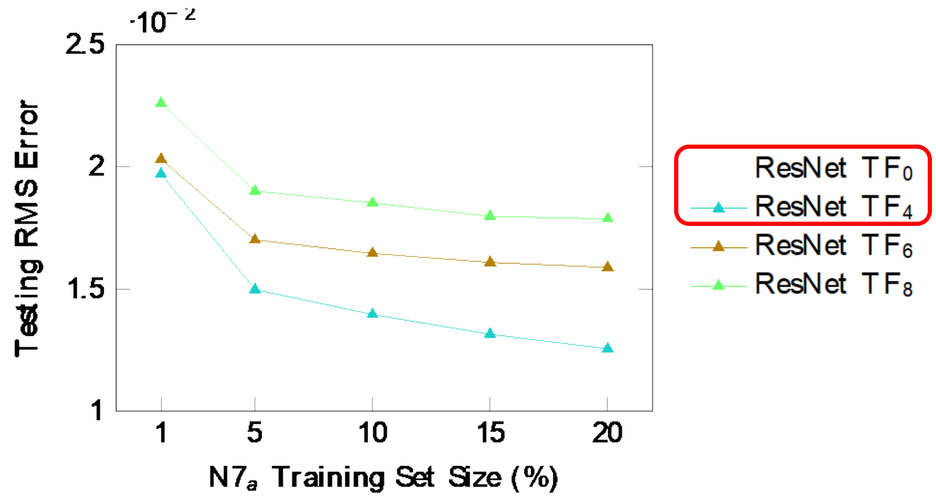
- Python 2.7
- Tensorflow 1.2.1
- GeForce GTX 1080
- SRAF, OPC, Aerial image: Mentor Graphics Calibre
- Rigorous simulation: Synopsys Sentaurus Lithography
- Average 10 trials of different random seeds
- ~30K clips for each N10 and N7 dataset



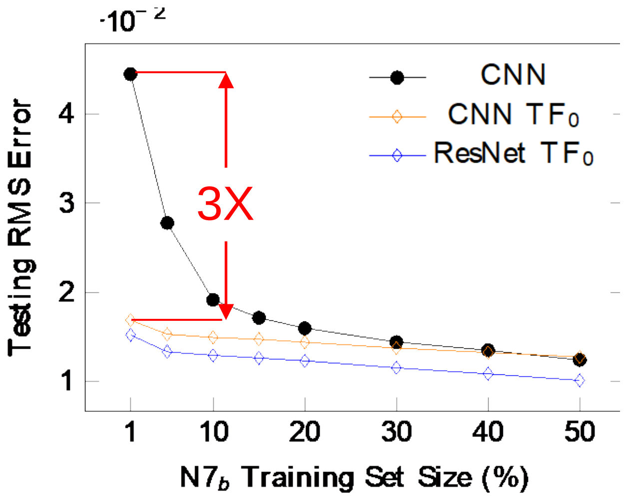
Explore Knowledge Transfer

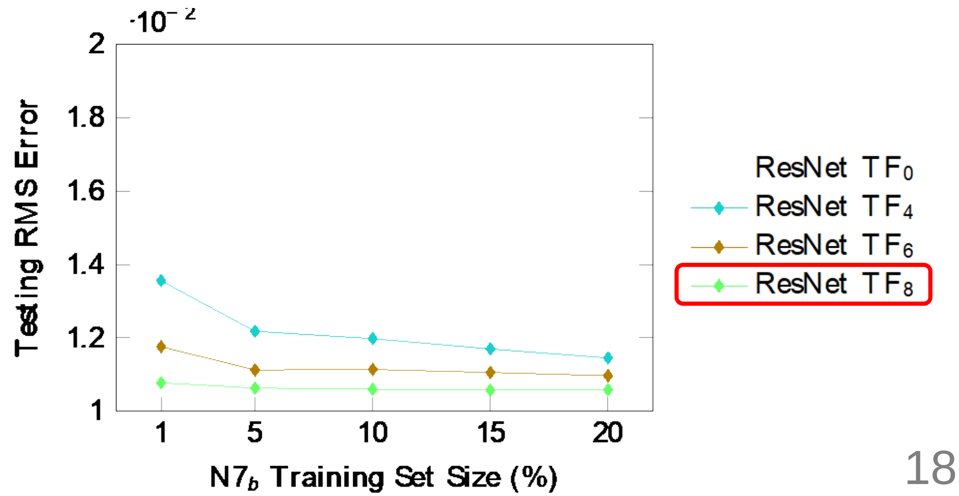
From N10 to N7





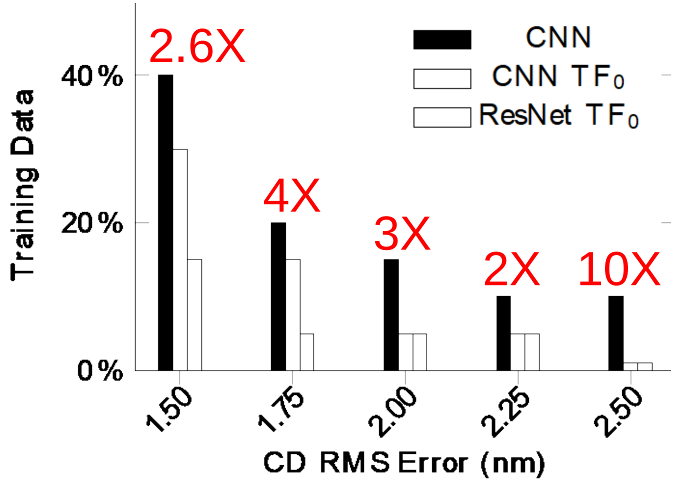
From $N7_a$ to $N7_b$





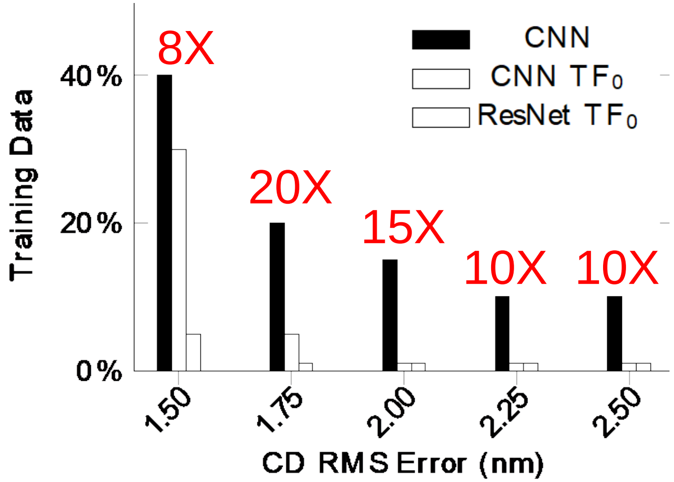
Data Reduction from Knowledge Transfer

From N10 to $N7_b$



2~10X reduction on training data

From $N7_a$ to $N7_b$



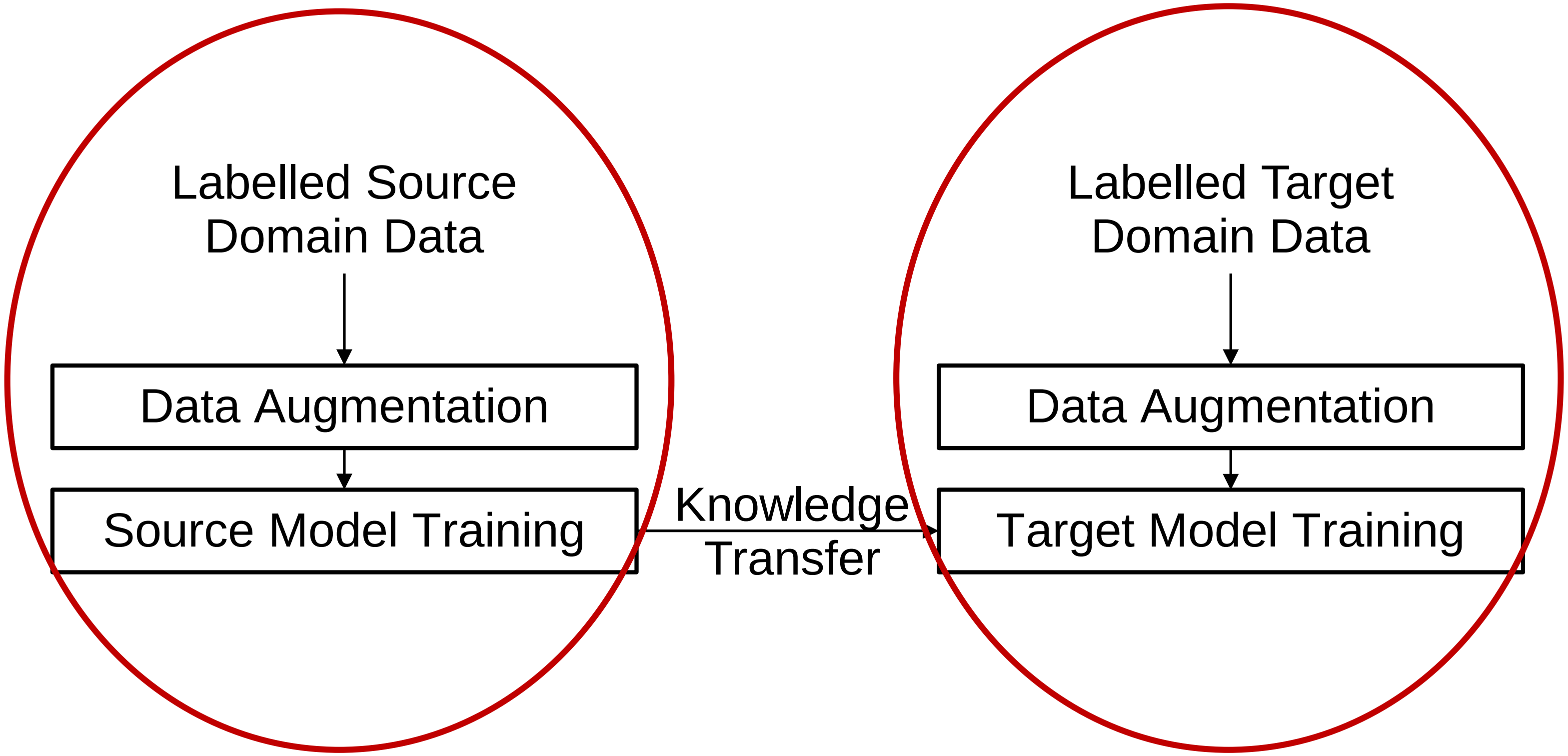
8~20X reduction on training data

Explore Knowledge Transfer

	N10→$N7_b$	**$N7_a$→$N7_b$**
Dataset Similarity	Medium	High
Knowledge Transfer	Medium	High
Data Reduction	2~10X	8~20X

- Improve data efficiency
- Less cost for data preparation
- Less turn-around time
- Prototyped by Toshiba Memory Corp.

Conclusion



- Transfer learning & ResNet
- Improve data efficiency
- 2~10X reduction of training data
- Reduce turn-around time
- Increase modeling accuracy

Future Directions

	Transfer Learning	Active Learning	Semi-supervised Learning
Labelled Old Data	✔		
Labelled New Data	✔		✔
Label Querying		✔	

ML is more than classification/regression
Better understanding about data

ML for
SRAF&OPC
P & R…

Thanks