Using machine learning in optical CD metrology

Barak Bringoltz
Director of Technology and Machine Learning Group manager

Nova measuring instruments

barak-br@novami.com

25 YEARS

NOVA® PROCESS INSIGHT

Outline

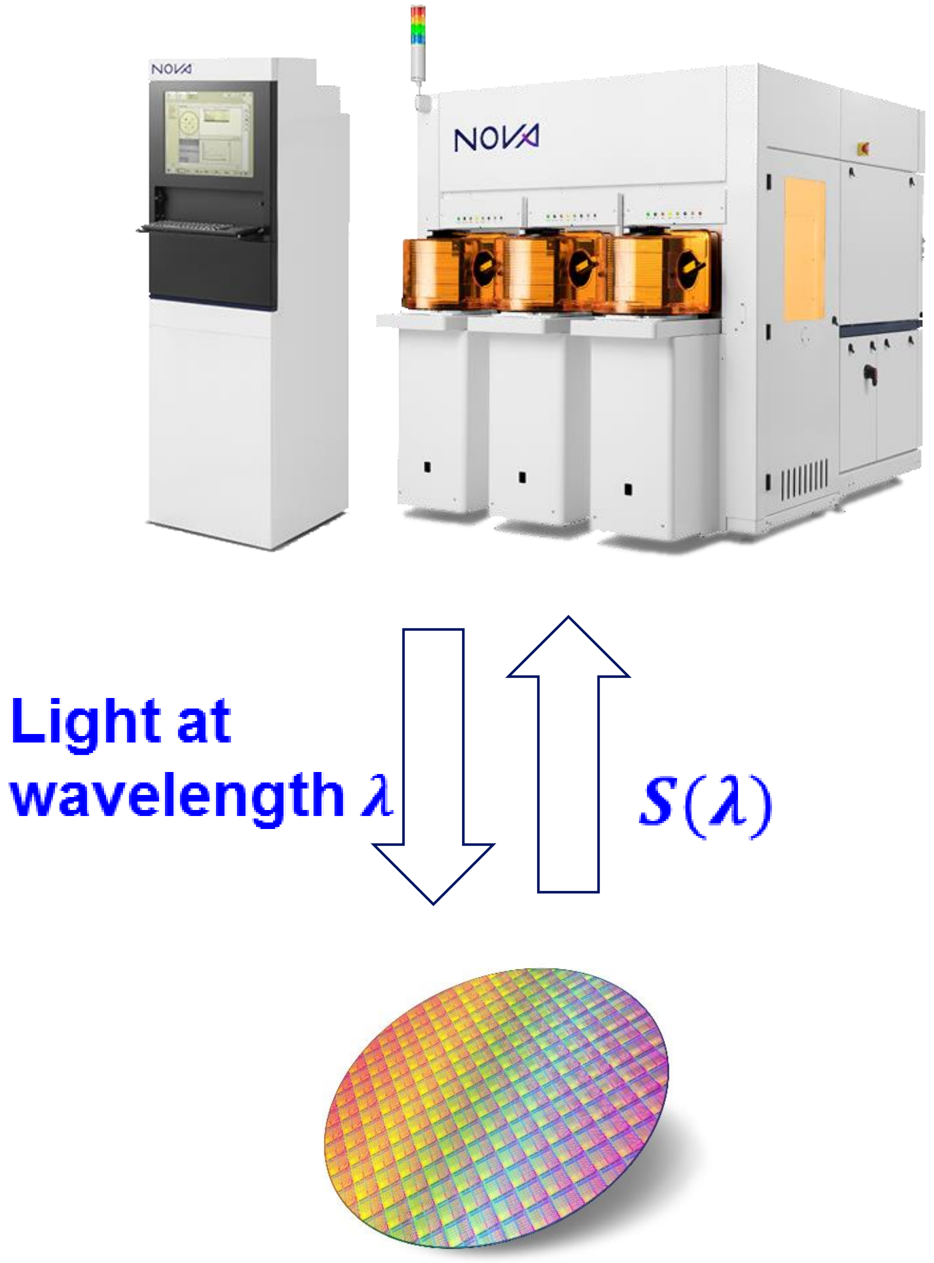
- Optical CD metrology and why machine learning in OCD.
- Nova's machine learning and big data solution.
- Performance:
  1. Basic accuracy performance.
  2. Budgeting accuracy and performance: spectral sensitivity, algorithm capacity, and data size and type.
- Summary

Why machine learning in OCD?

OCD:

Integrated metrology

- Less illumination and polarization modes.



Stand-alone or SA

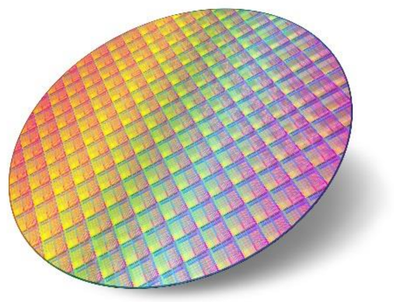
- Many illumination directions (azimuths and inclinations).
- Multiple polarization modes (full polarimetry).

Why machine learning in OCD?

OCD modeling:

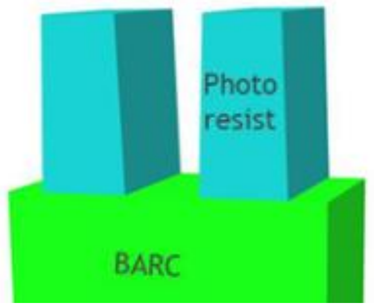
Light at wavelength λ

$S(\lambda)$



+

3D model:



Maxwell physical modeling:

$$\nabla \cdot \mathbf{E} = \frac{\rho_v}{\varepsilon} \quad \text{(Gauss' Law)}$$

$$\nabla \cdot \mathbf{H} = 0 \quad \text{(Gauss' Law for Magnetism)}$$

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} \quad \text{(Faraday's Law)}$$

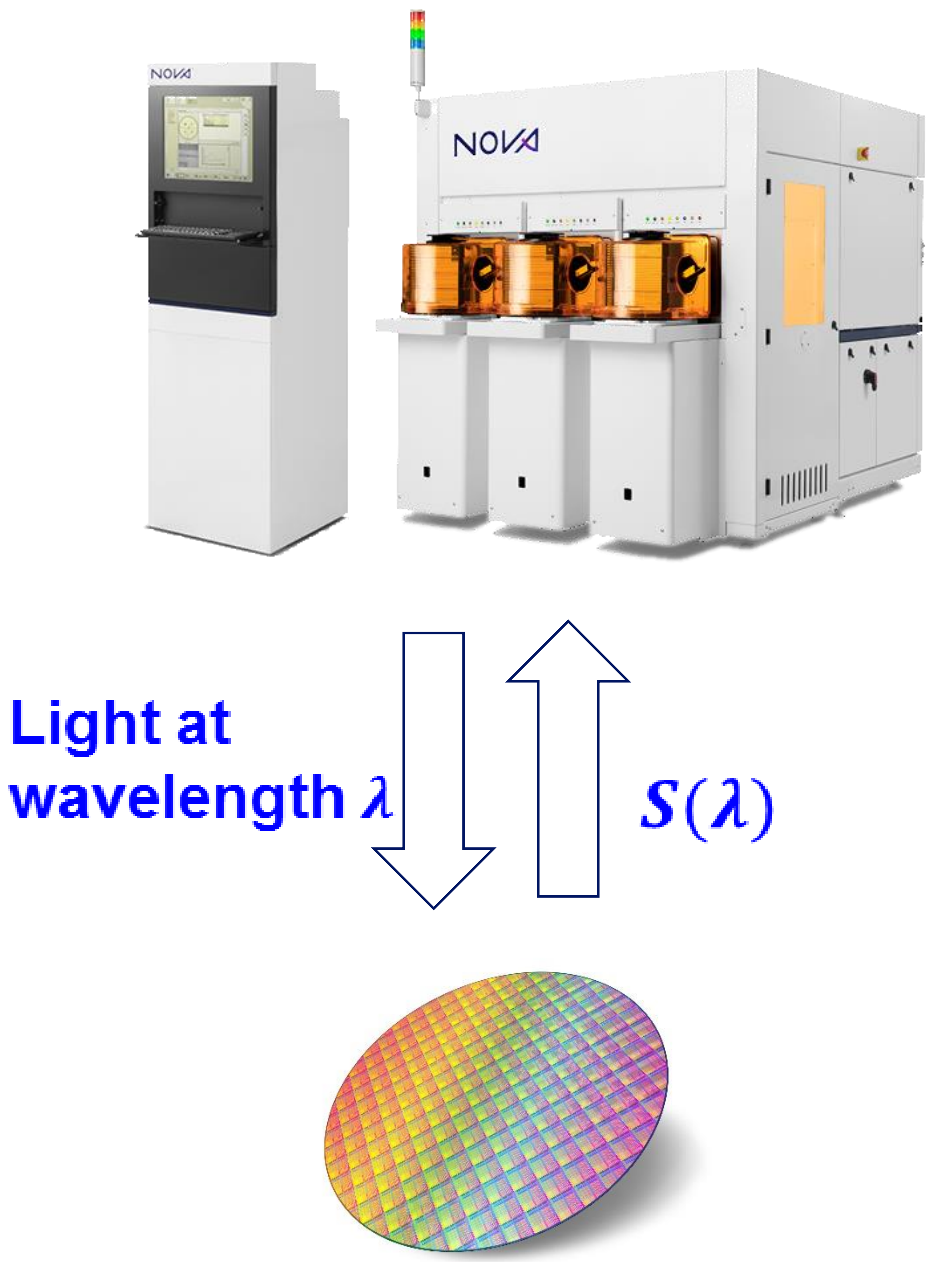
$$\nabla \times \mathbf{H} = \mathbf{J} + \varepsilon \frac{\partial \mathbf{E}}{\partial t} \quad \text{(Ampere's Law)}$$

HVM worthy model:

$S(\lambda) \rightarrow$ Model $\rightarrow CD$

Why machine learning in OCD?

OCD modeling with machine learning:

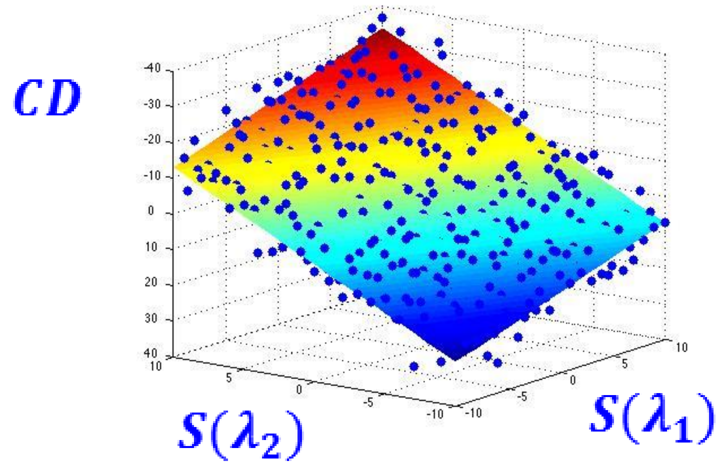


+

Data:

ID	$S(\lambda)$	CD
1	[0.54,0.43,...,0.32]	17.3A
2	[0.67,0.52,...,0.14]	16.3A
...	...	...
100	[0.33,0.15,...,0.28]	18.2A

Data science / modeling:



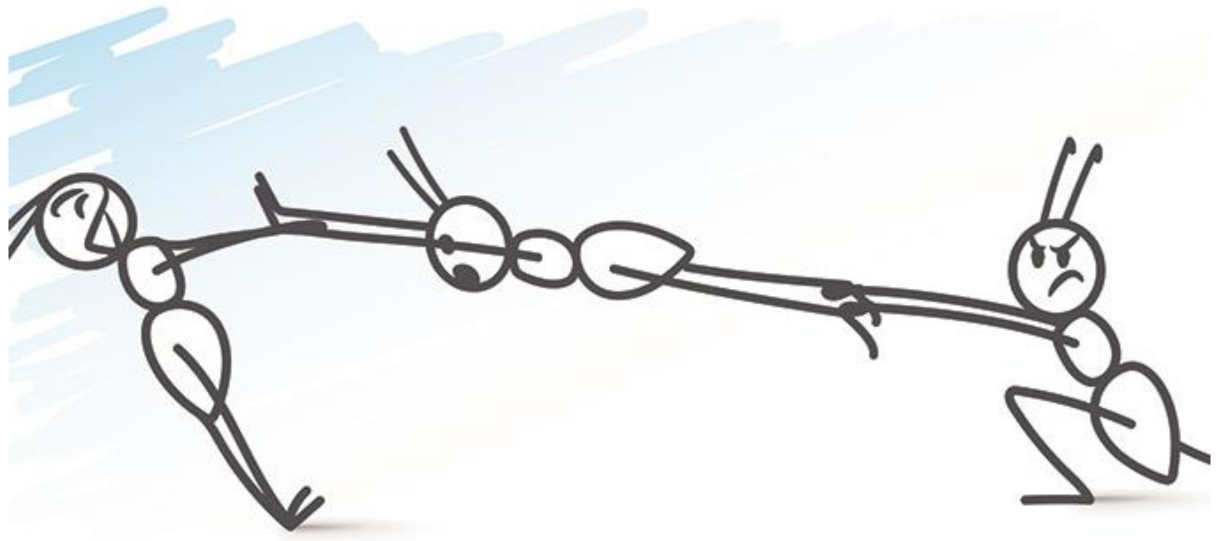
HVM worthy model:

$S(\lambda) \rightarrow$ Model $\rightarrow CD$

So … why machine learning in OCD?

Productive:

- FMP ~ sub-Angstrom
- TPT ~ measure every wafer and every die.



Accurate & robust:

- Spec ~ a few Angstroms
- Process splits and variations.

⇩

Integrated metrology tools

- *Measures every wafer.*
- *Less information in spectrum.*
- *Requires expert work.*

Automatic:

- Recipe creation time << operator shift.
- Reproducible, predictable.

⇩

SA optical or high-resolution non-optical:

- *Much information, highly accurate.*
- *Typically measures few wafers per lot.*
- *Requires expert work.*

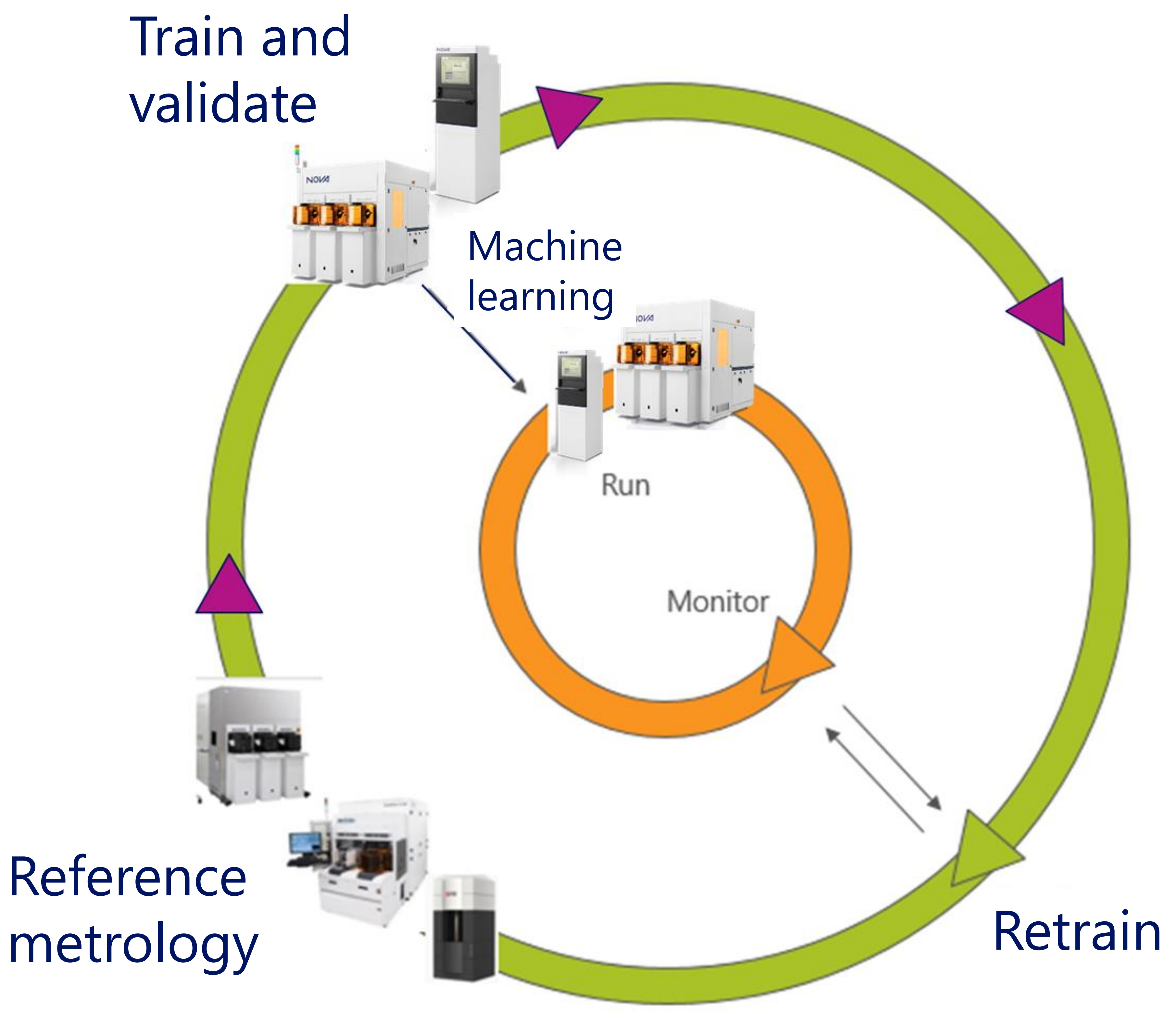
FMP = Fleet Measurement Precision
TPT = Throughput

A machine learning big data system solves this tension.

- Metrology solutions **built**, **tested**, **monitored** and **modified** **automatically**

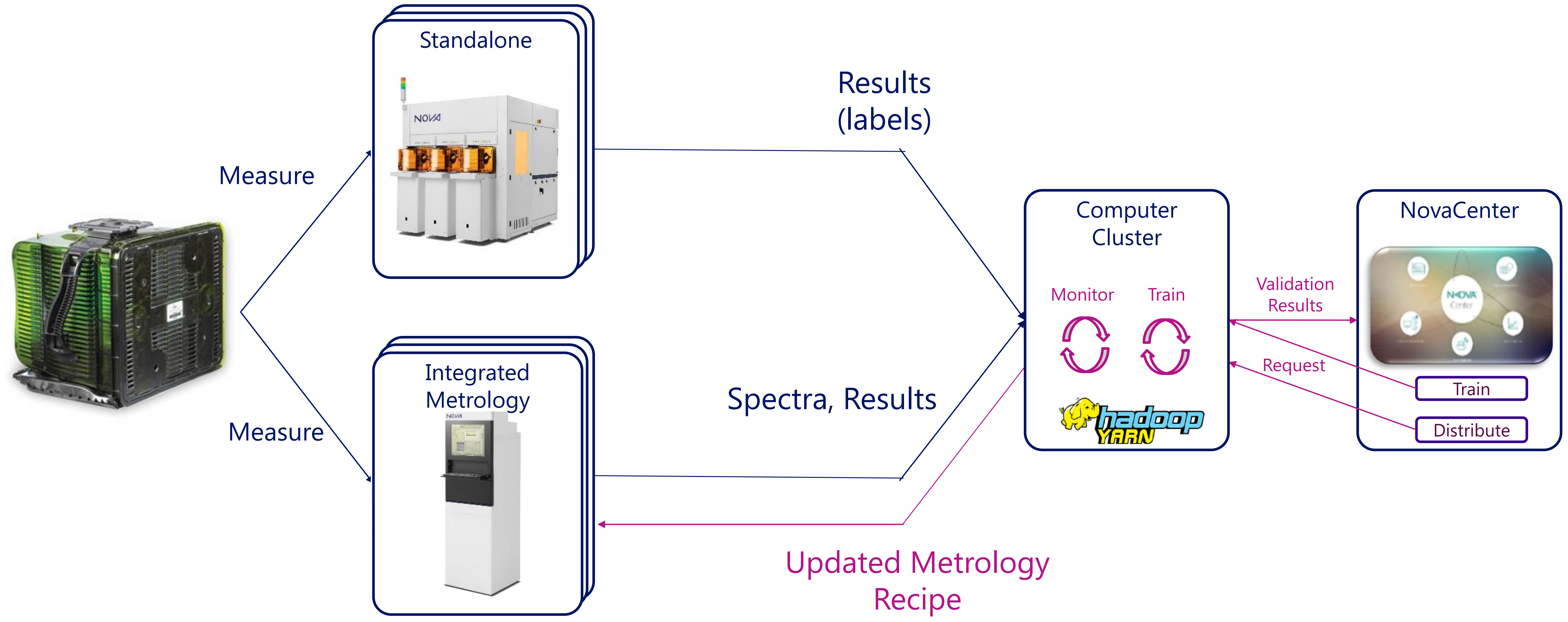
⇩

Accuracy, speed, productivity, and predictability

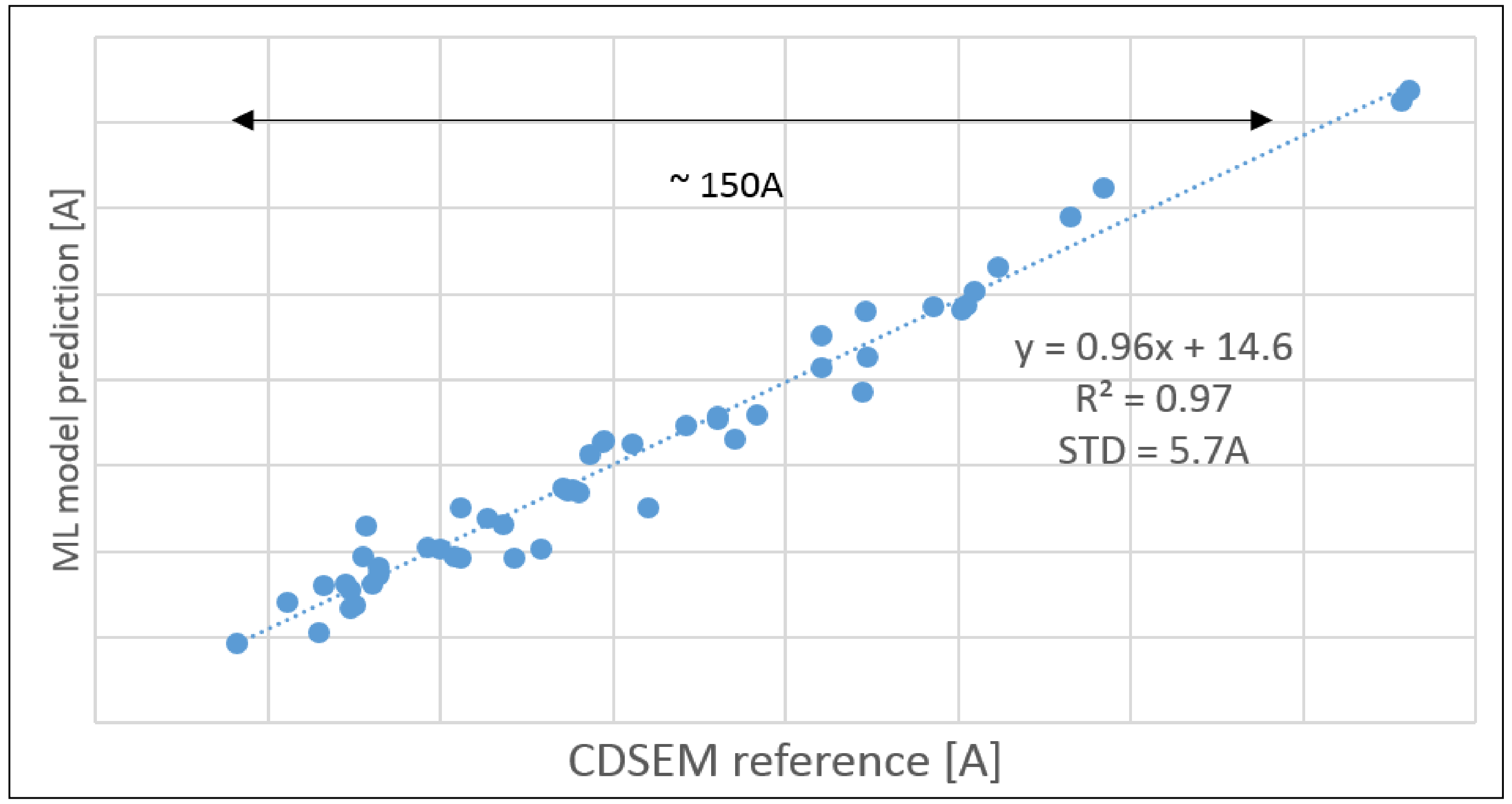


NOVA

The Standalone (SA) to Integrated (IM) Data Flow

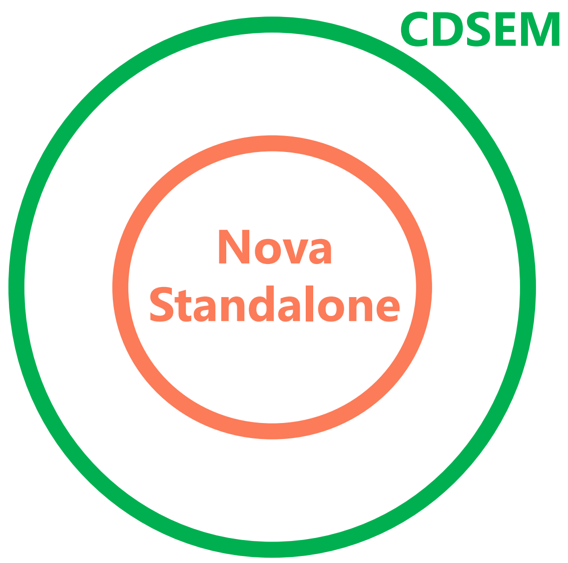


Accuracy performance of machine learning



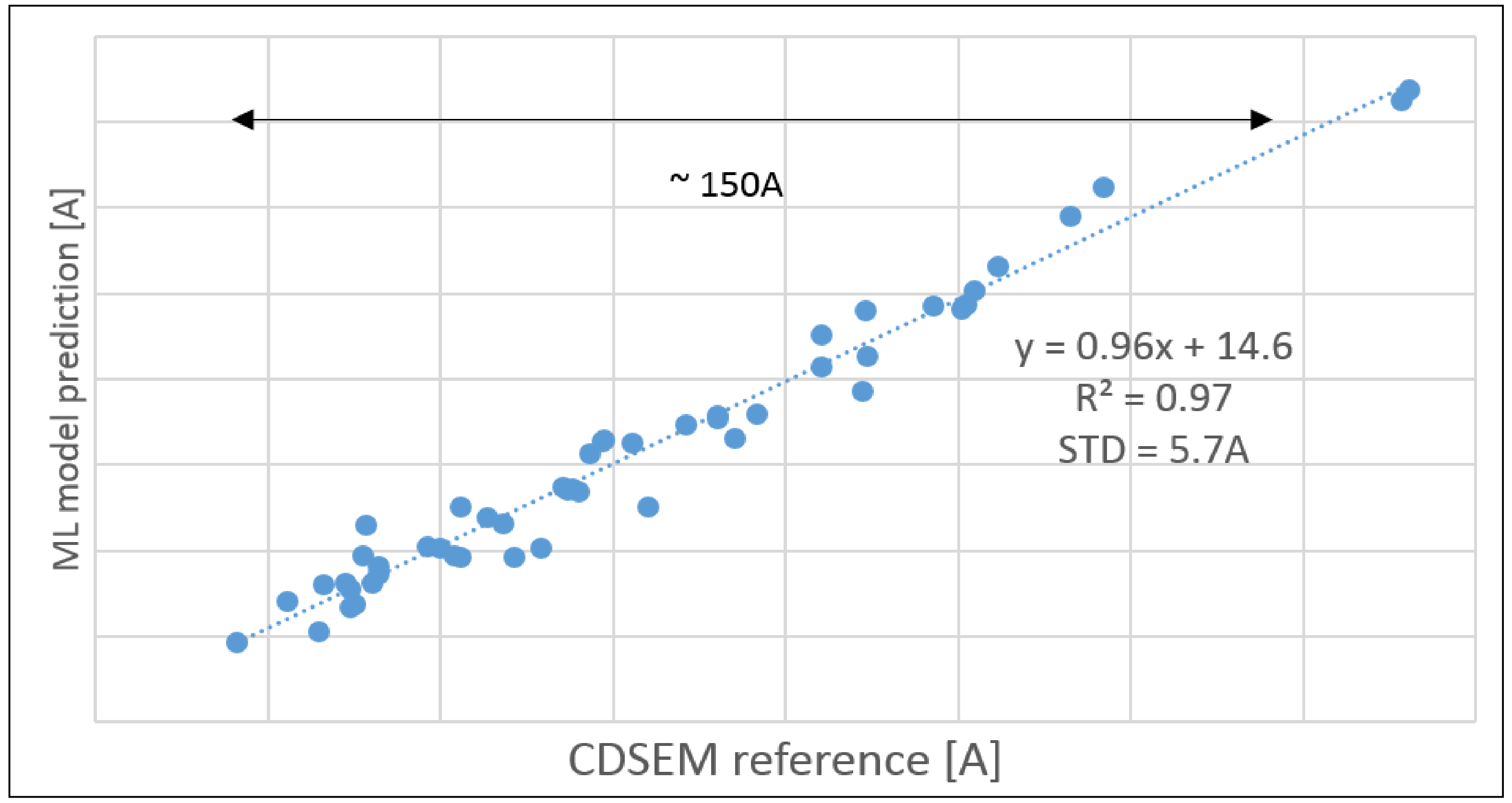
<u>Example A:</u>

Reference: CDSEM.
Inline tool: Nova Standalone.
Accuracy: $\mathbf{1\sigma \sim 5.7A}$
Train set: ~50dies DOE wafers.



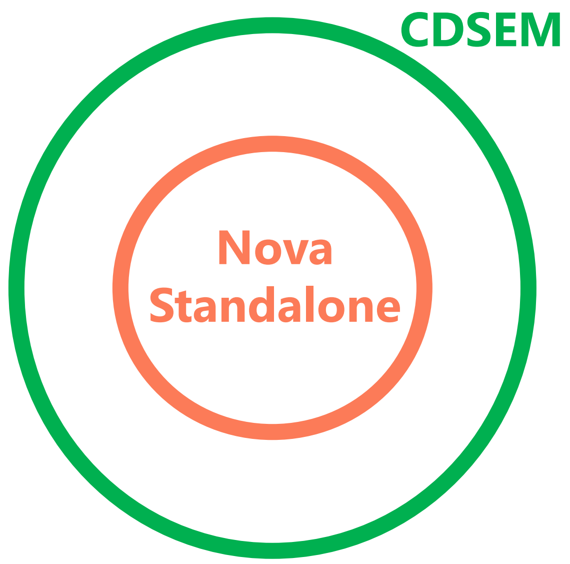
Accuracy performance of machine learning

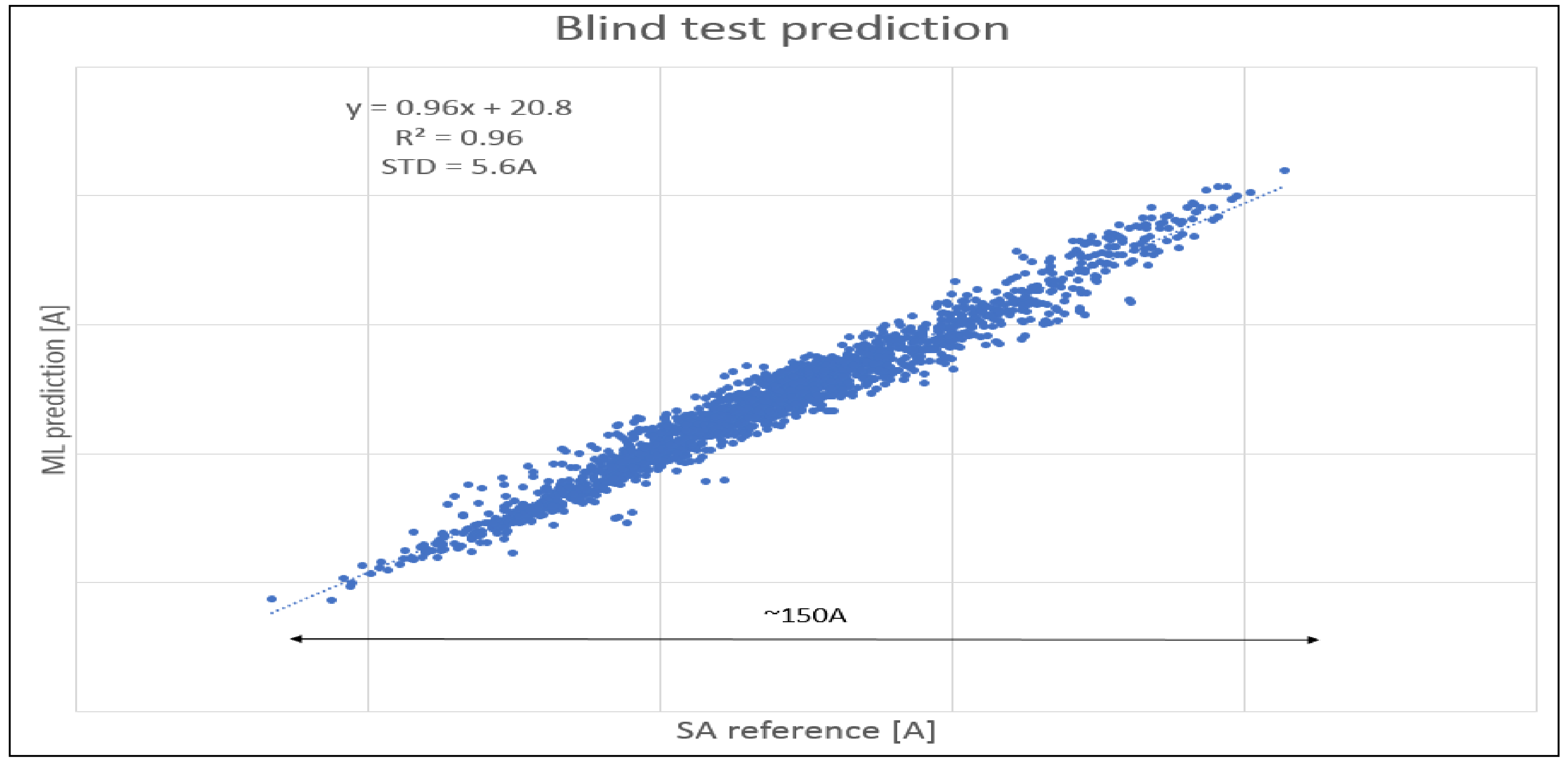
(also see SPIE 10585-32, 1014504, 97781W, JMM.15.4.044004)



<u>Example A:</u>

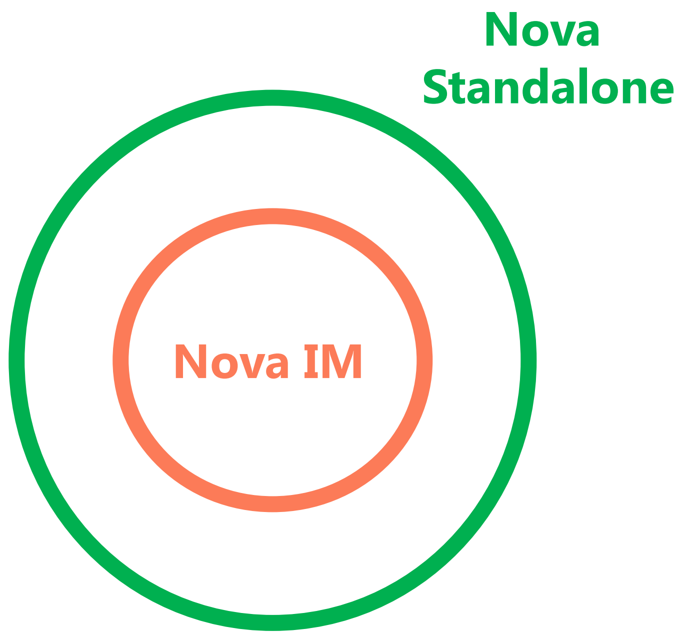
Reference: CDSEM.
Inline tool: Nova Standalone.
Accuracy: $1\sigma \sim 5.7A$
Train set: ~50dies DOE wafers.





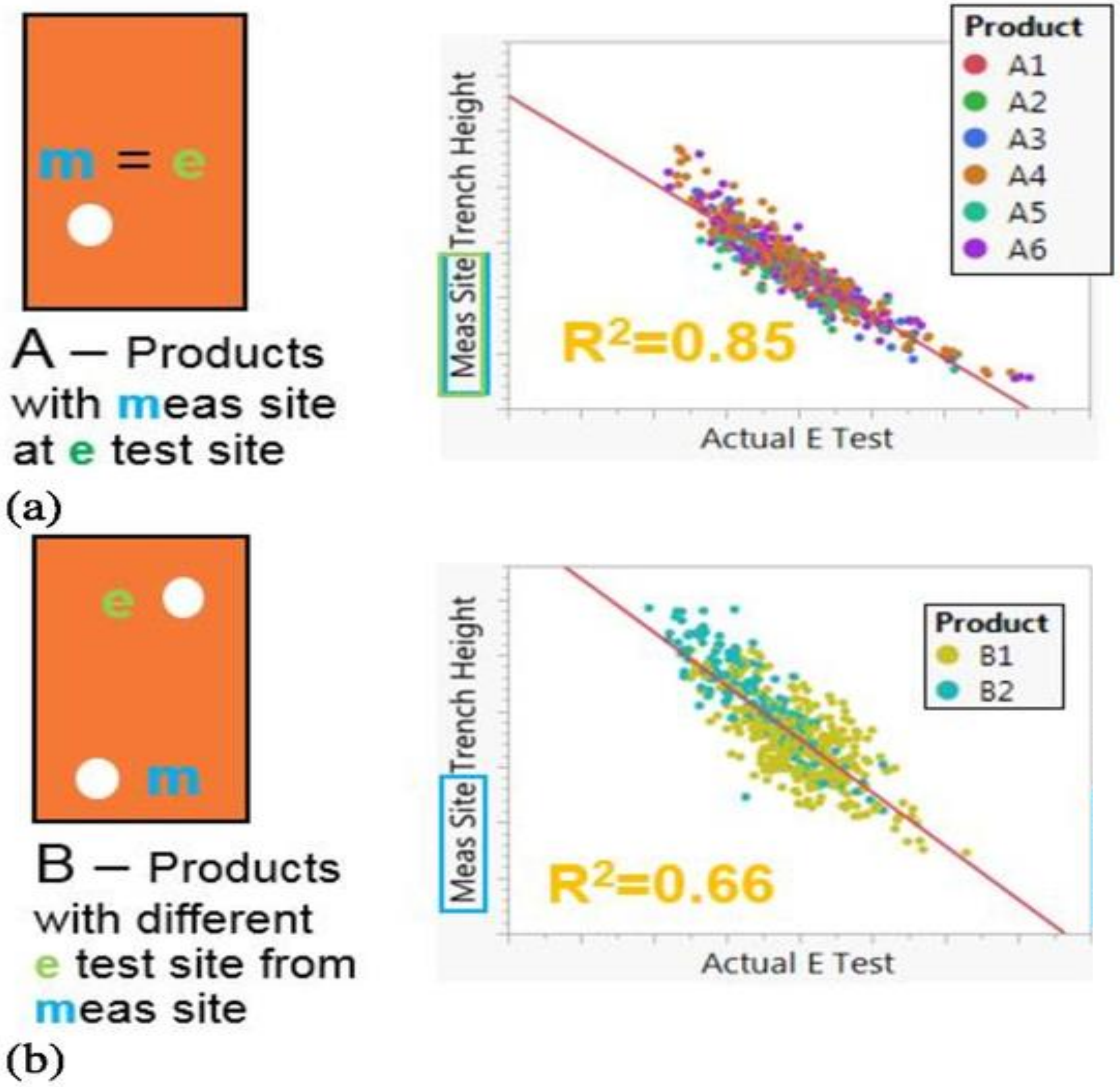
<u>Example B:</u>

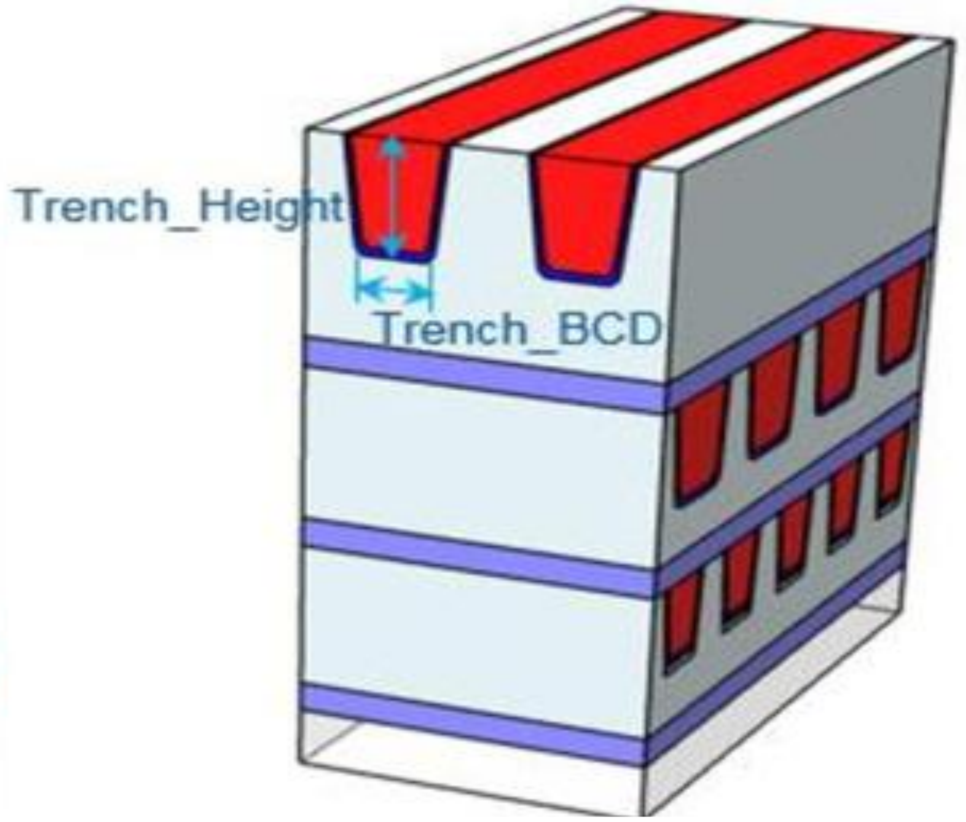
Reference: Nova SA physical model.
Inline tool: Nova IM.
Accuracy: $1\sigma \sim 5.6A$
Train set: 6000dies from POR sampling.



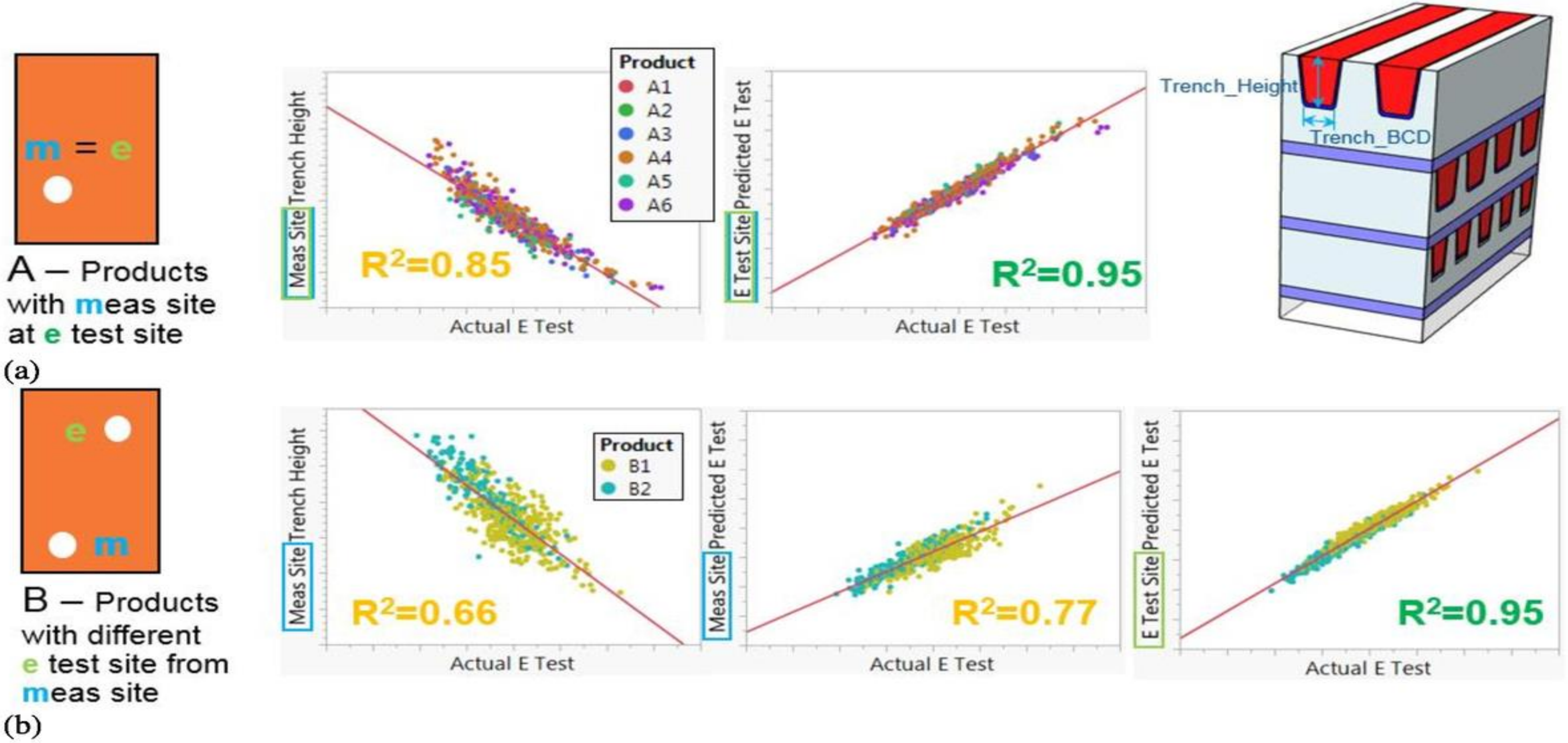
NOVA

Accuracy performance of machine learning-from SPIE Advanced Lithography P. Timoney *et al.* 10585-32 (see also 1014504, 97781W, JMM.15.4.044004)

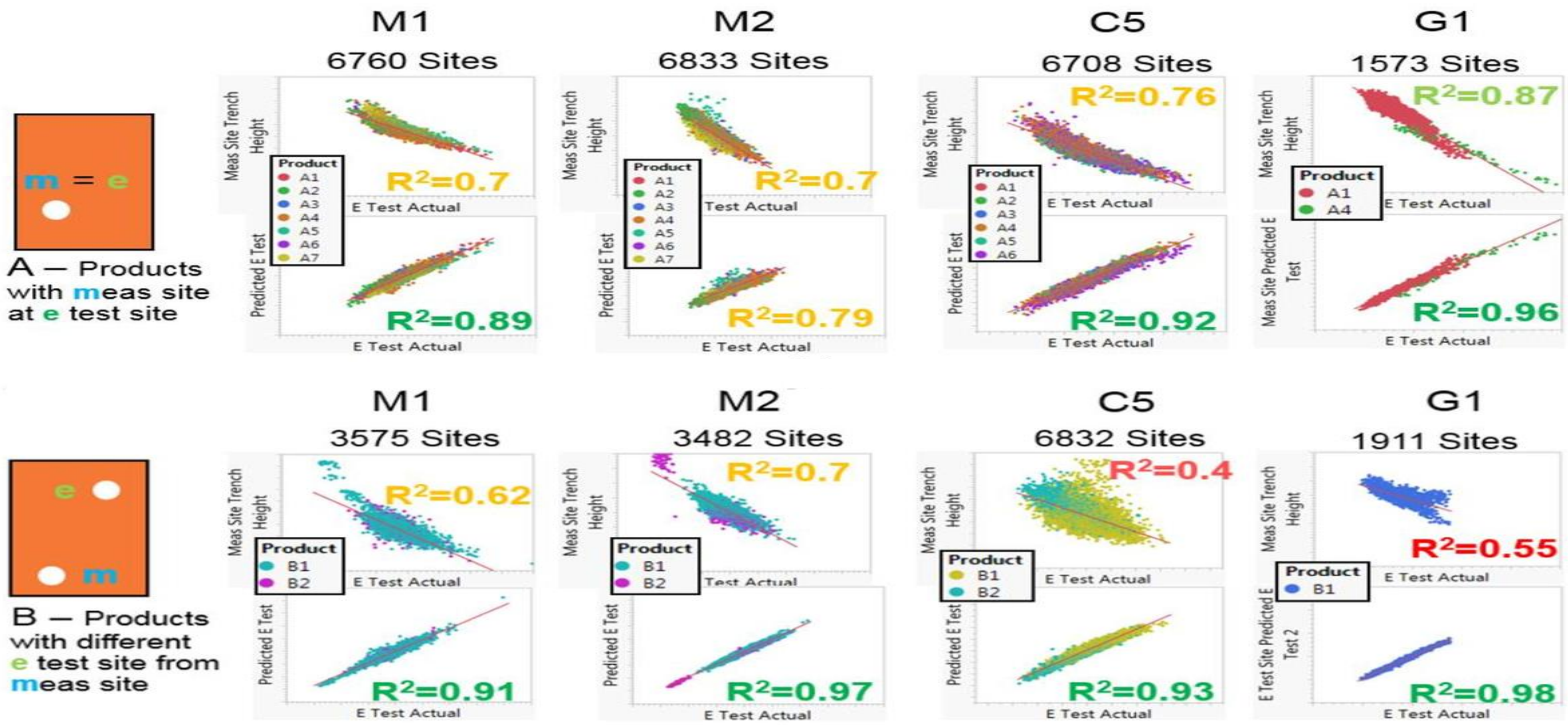




Accuracy performance of machine learning-from SPIE Advanced Lithography P. Timoney *et al.* 10585-32 (see also 1014504, 97781W, JMM.15.4.044004)



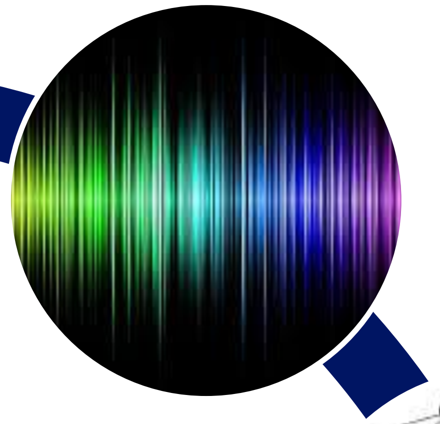
Accuracy performance of machine learning-from SPIE Advanced Lithography P. Timoney *et al.* 10585-32 (see also 1014504, 97781W, JMM.15.4.044004)



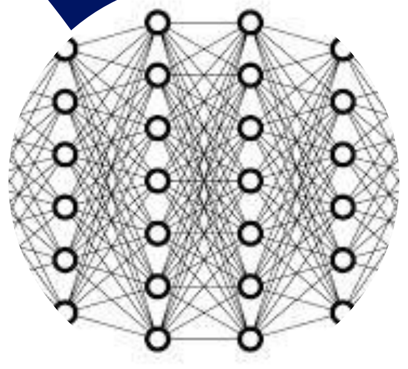
NOVA

Error budgeting accuracy and robustness of machine learning

Robustness and Accuracy



Spectral sensitivity



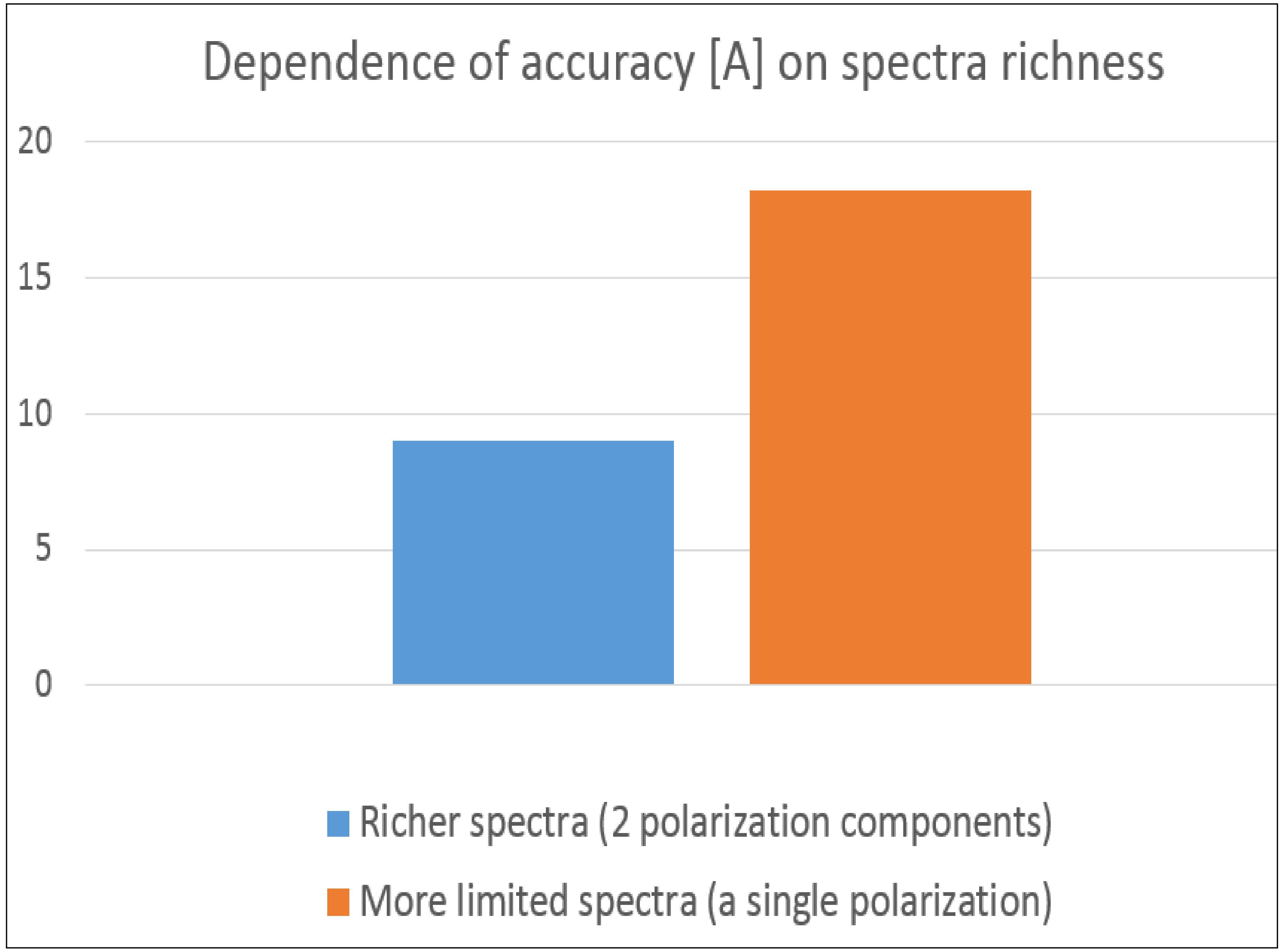
Algorithm capacity (simple models vs. DNN)

Data size and type

Reference error

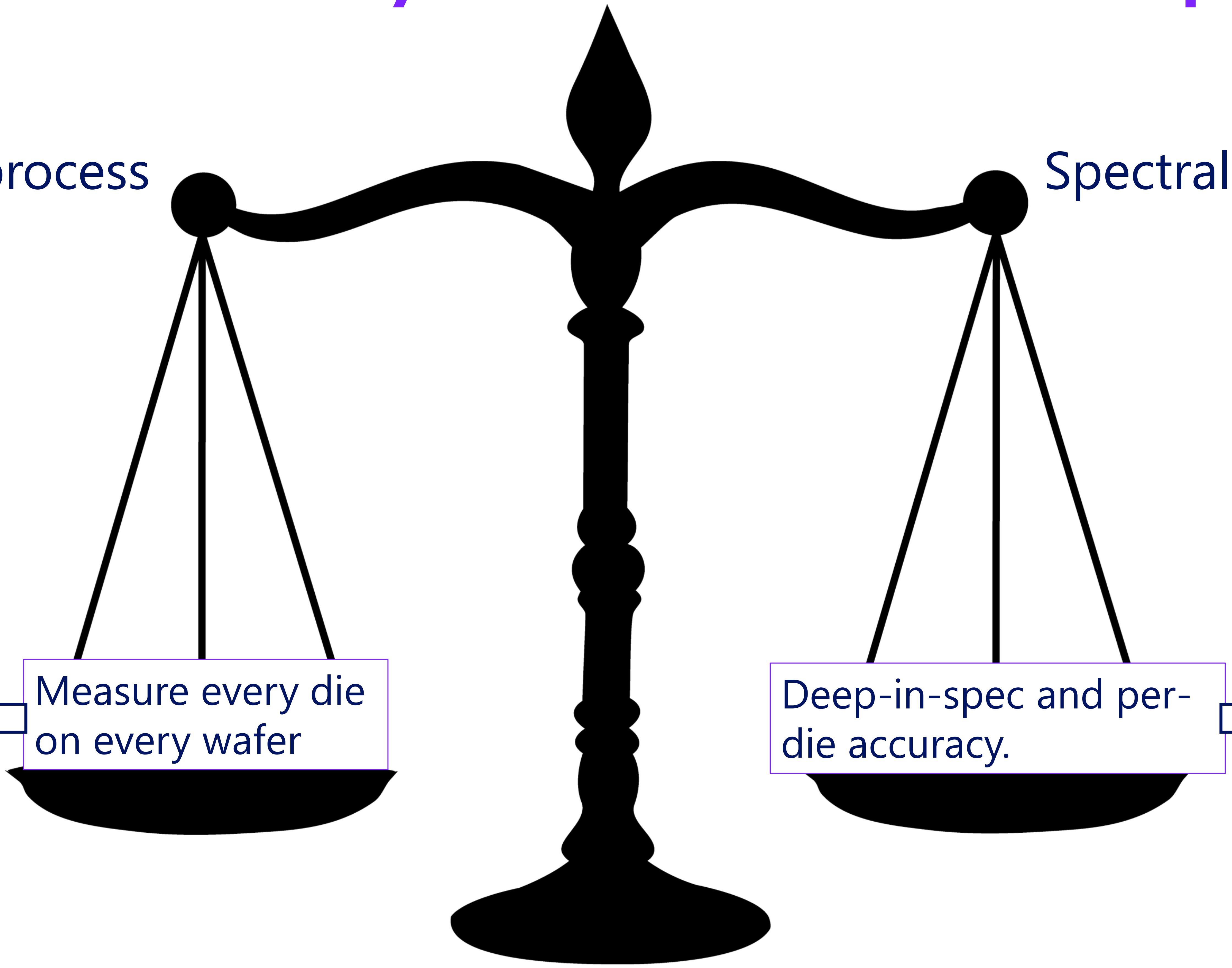
Today

Accuracy: spectral sensitivity

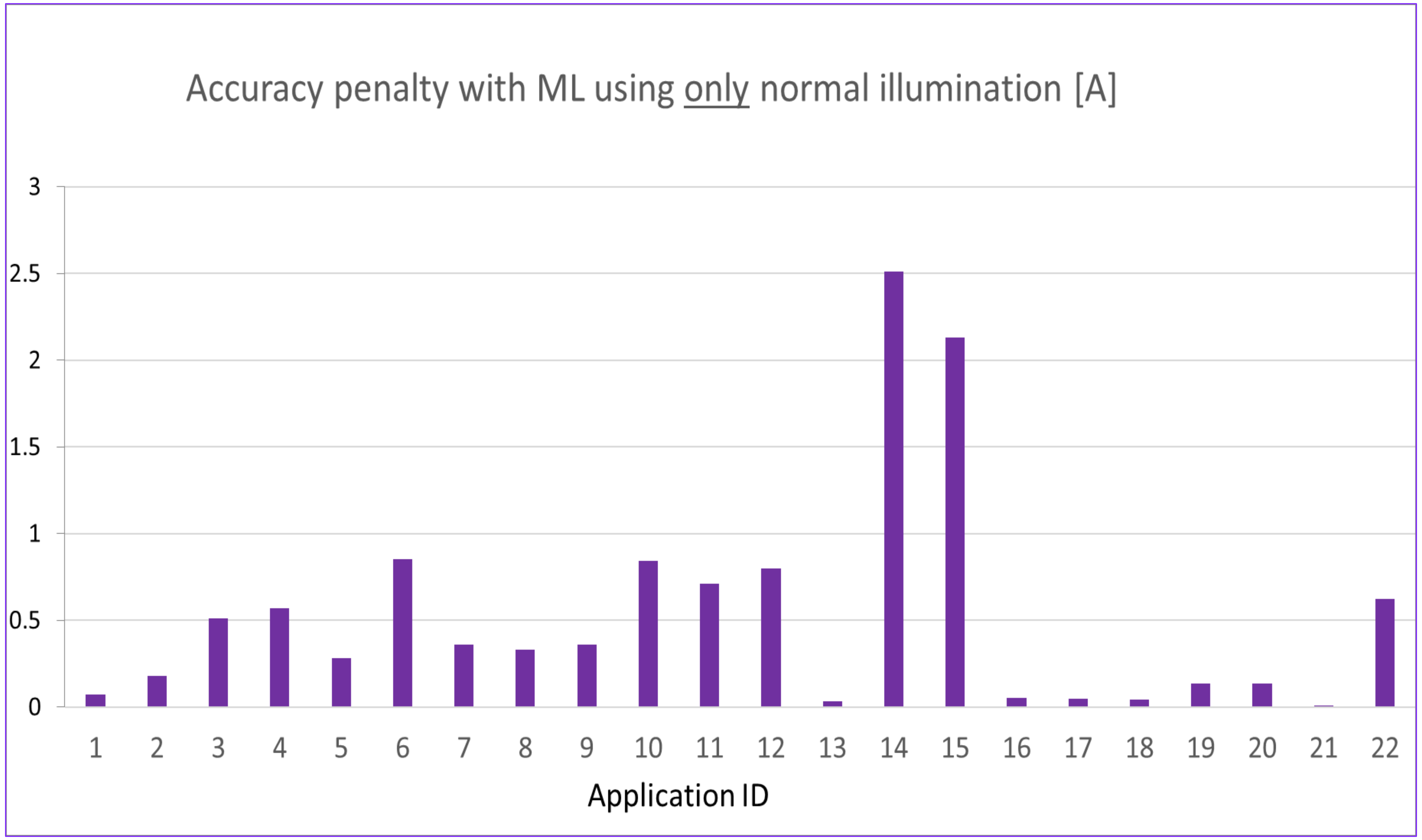


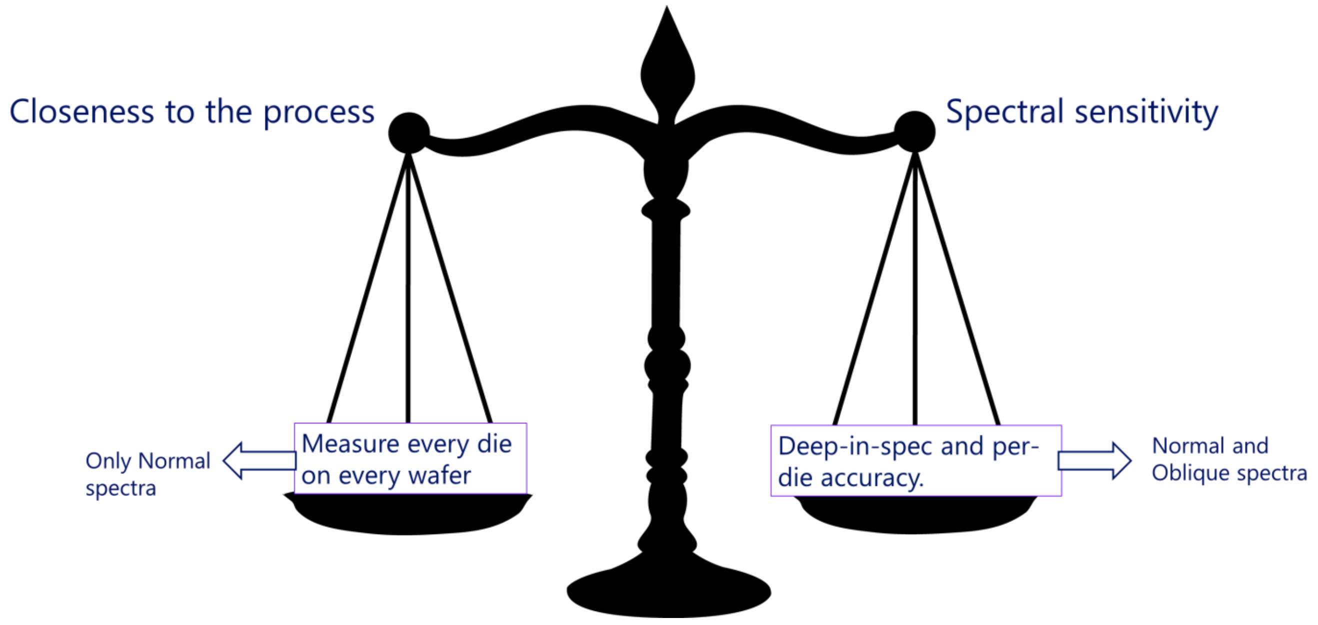
- Transfer of a physical model solution on Nova SA to a Nova IM.
- *Clearly ML is not `black magic':*
  - More spectral information improves accuracy by 2x: especially Normal channel vs Normal channel & Oblique.
  - Have other examples where we see how spectral information reflects the *underlying physics.*

Machine learning helps balance spectral sensitivity vs. closeness to the process



Balancing spectral sensitivity and closeness to the process

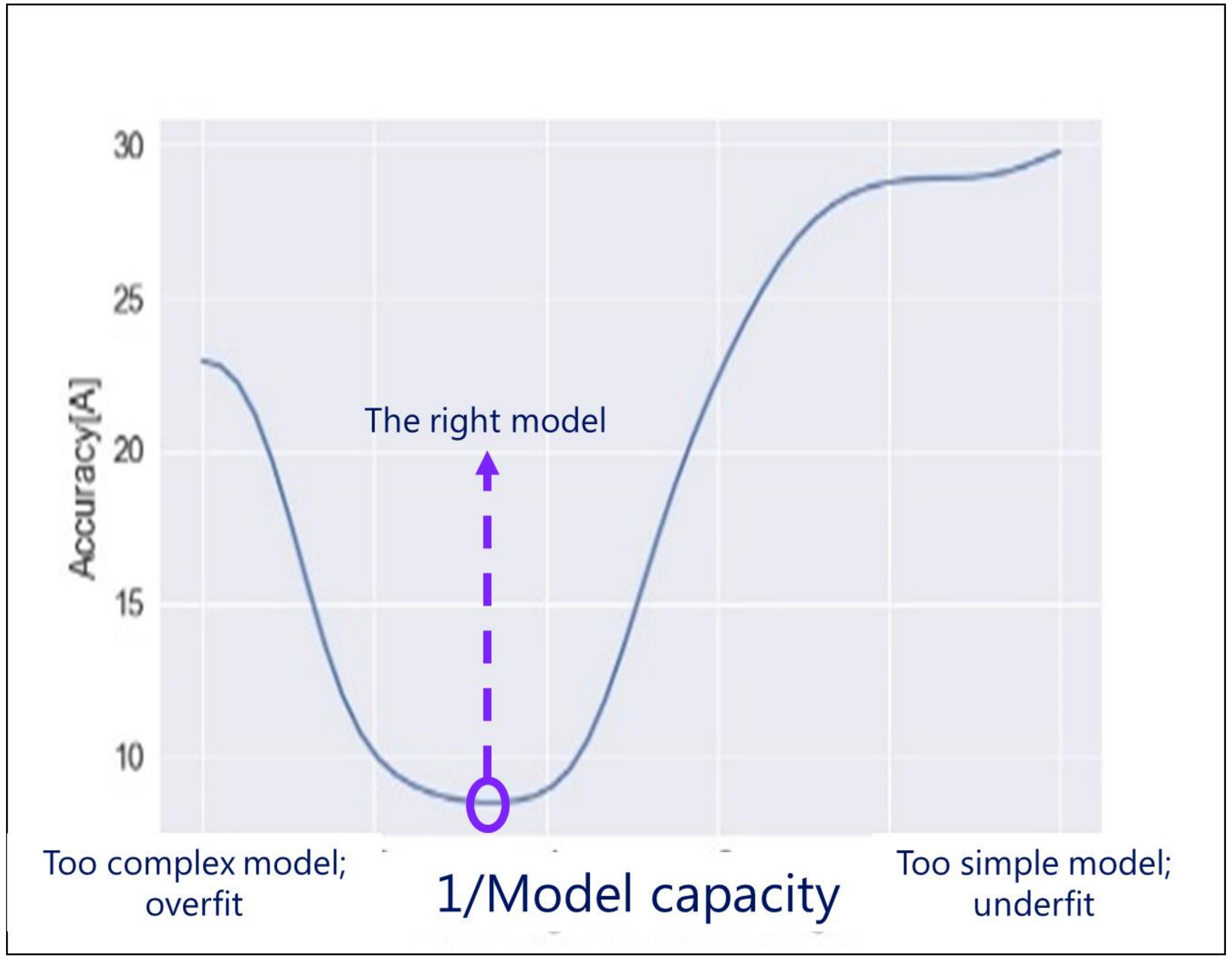




- Using machine learning we can balance spectral sensitivity and closeness-to-the-process.
- Customers are able to balance the two as per their specific needs.

Accuracy: model capacity

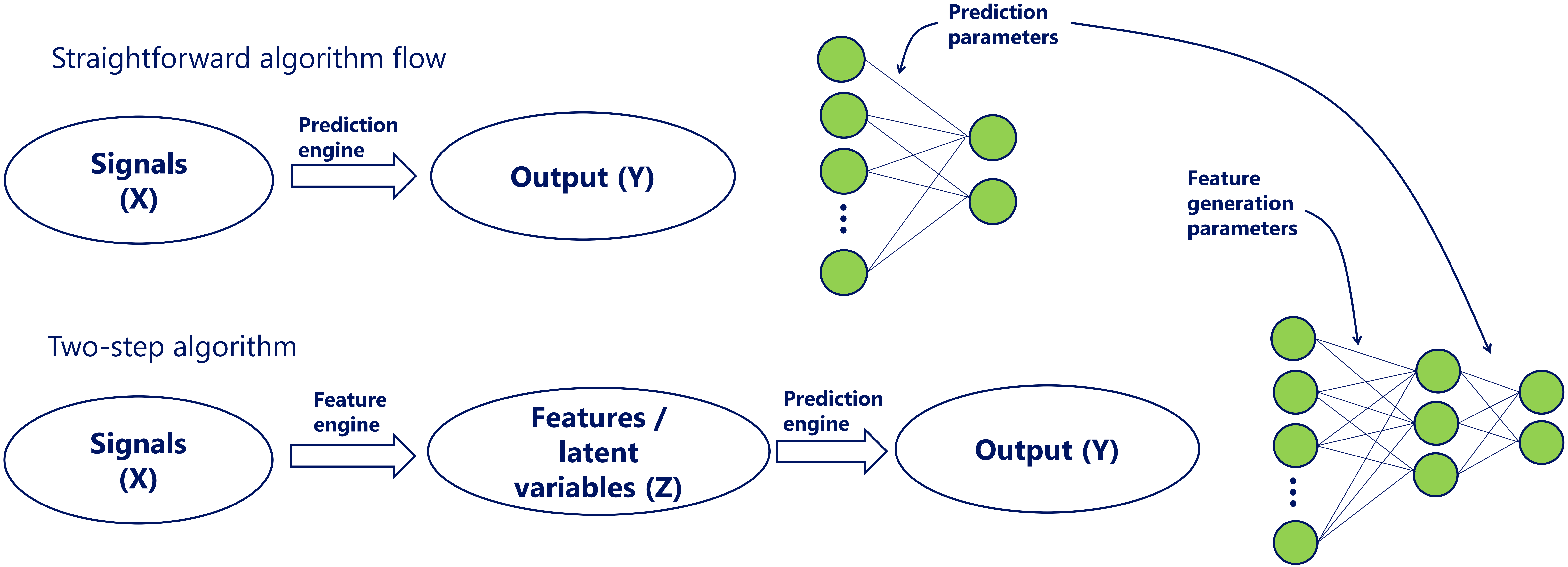
<u>Example A: changing model capacity with a single hyper-parameter</u>



- Transfer of a physical model solution on Nova SA to a Nova IM.
- Here, model capacity was modified by regularization on a fixed data size.
- Simplicity of model setup makes it easy to *<u>automate.</u>*

Accuracy: model capacity

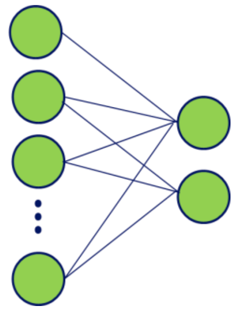
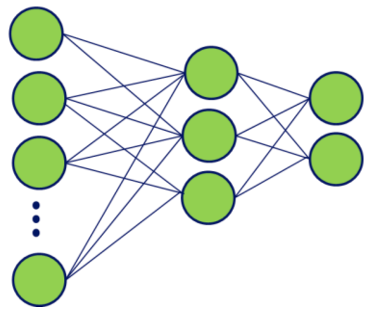
Example B: changing model capacity by changing the number of hyper-parameters



Accuracy: model capacity

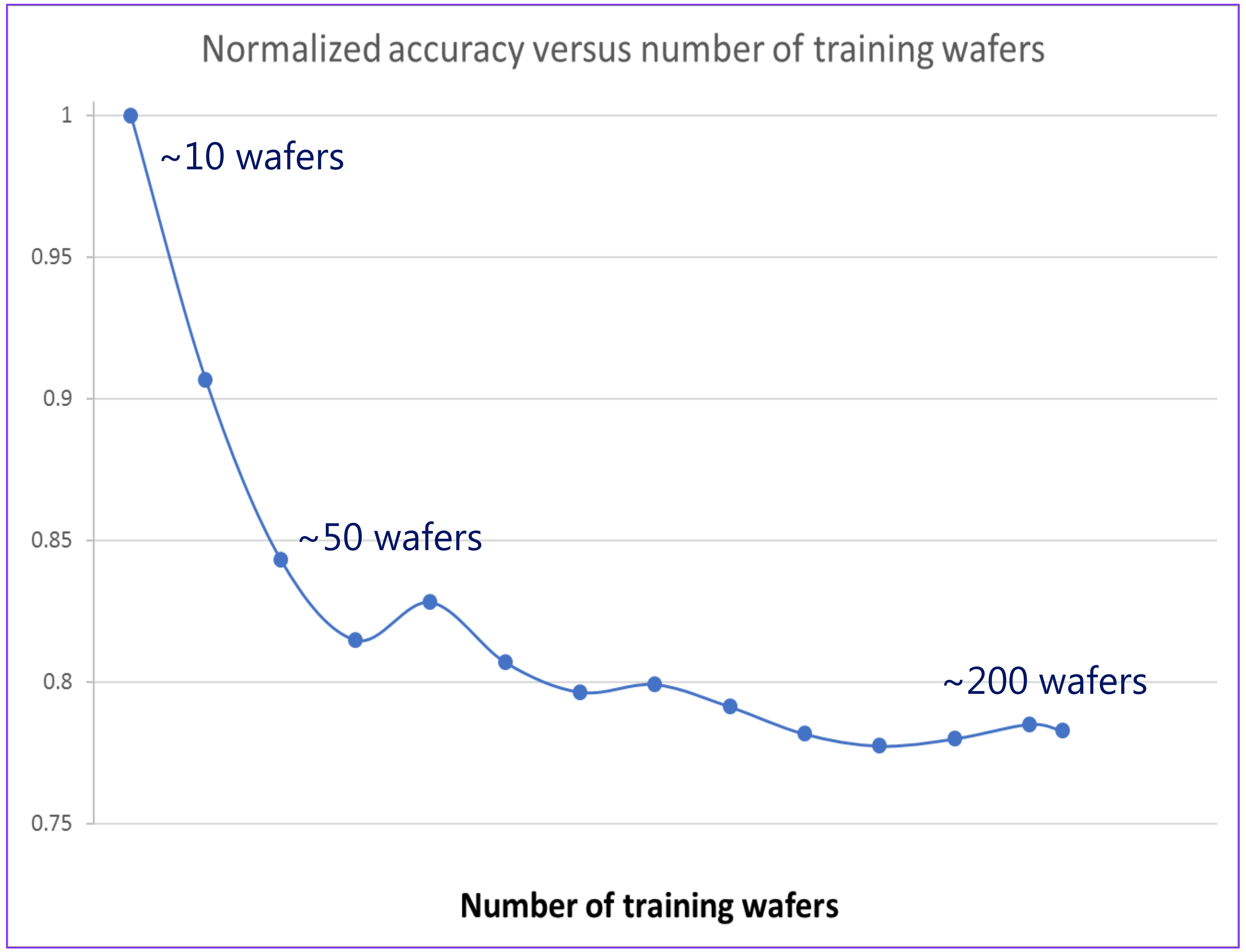
Example B: changing model capacity by changing
the number of hyper-parameters

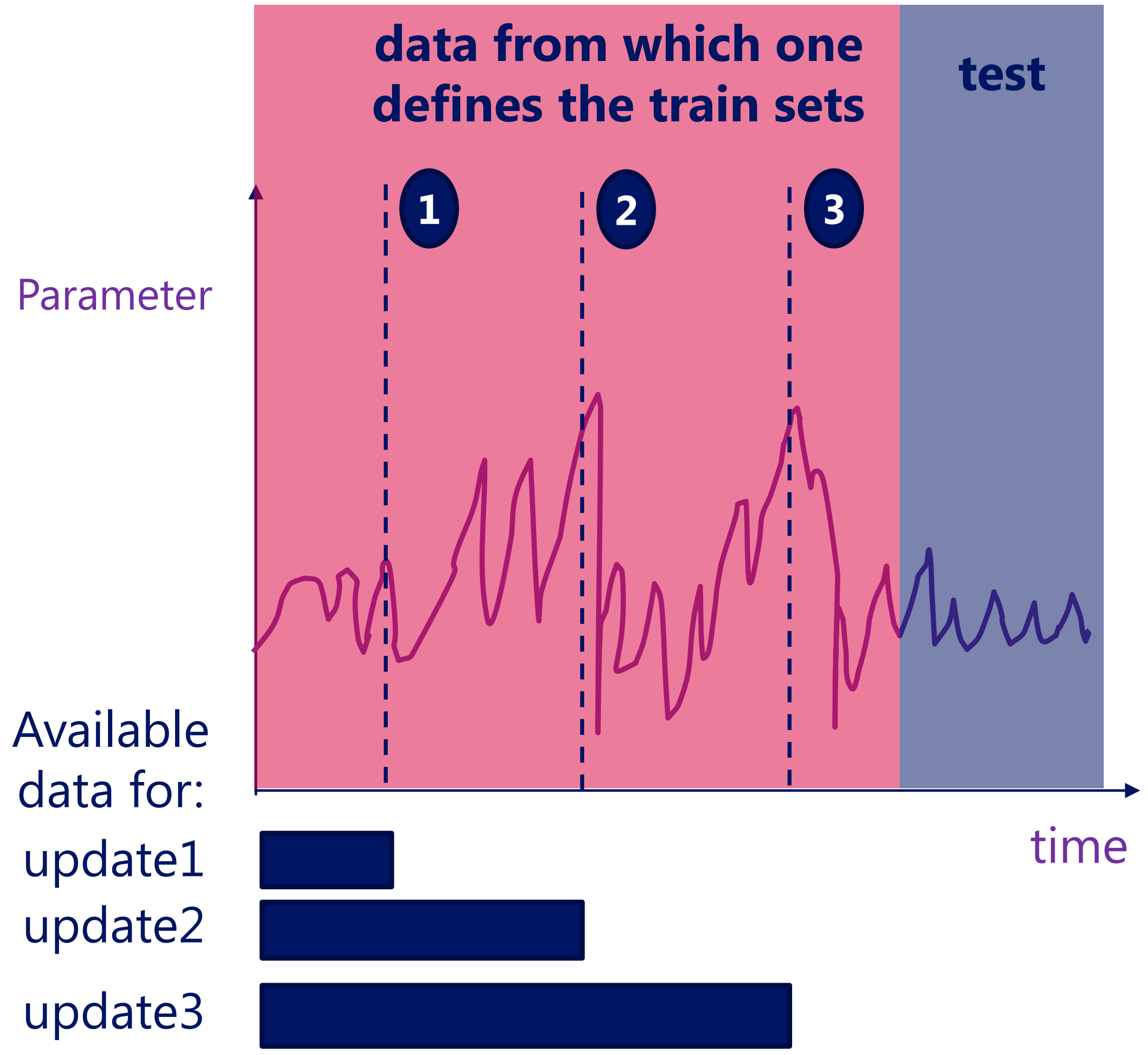
<u>CMP thickness case, from Standalone to Nova IM (train and test sets each comprised of production data of ~75 wafers)</u>

Attribute		
STD in Angstrom	23	16
R2	0.96	0.98
Slope	0.92	0.97

Accuracy: data size

Example below: transfer of a physical model solution on SA to Nova IM.

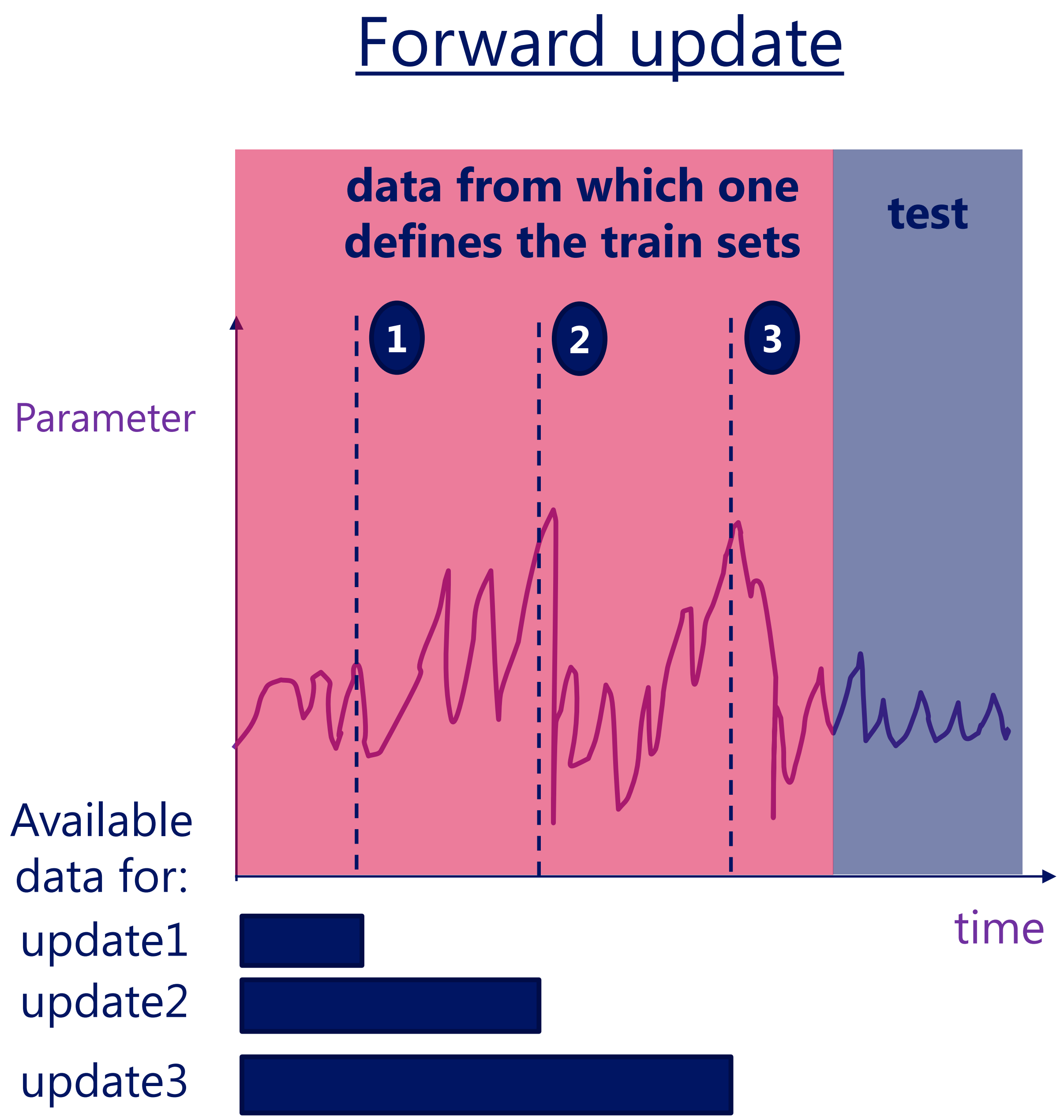


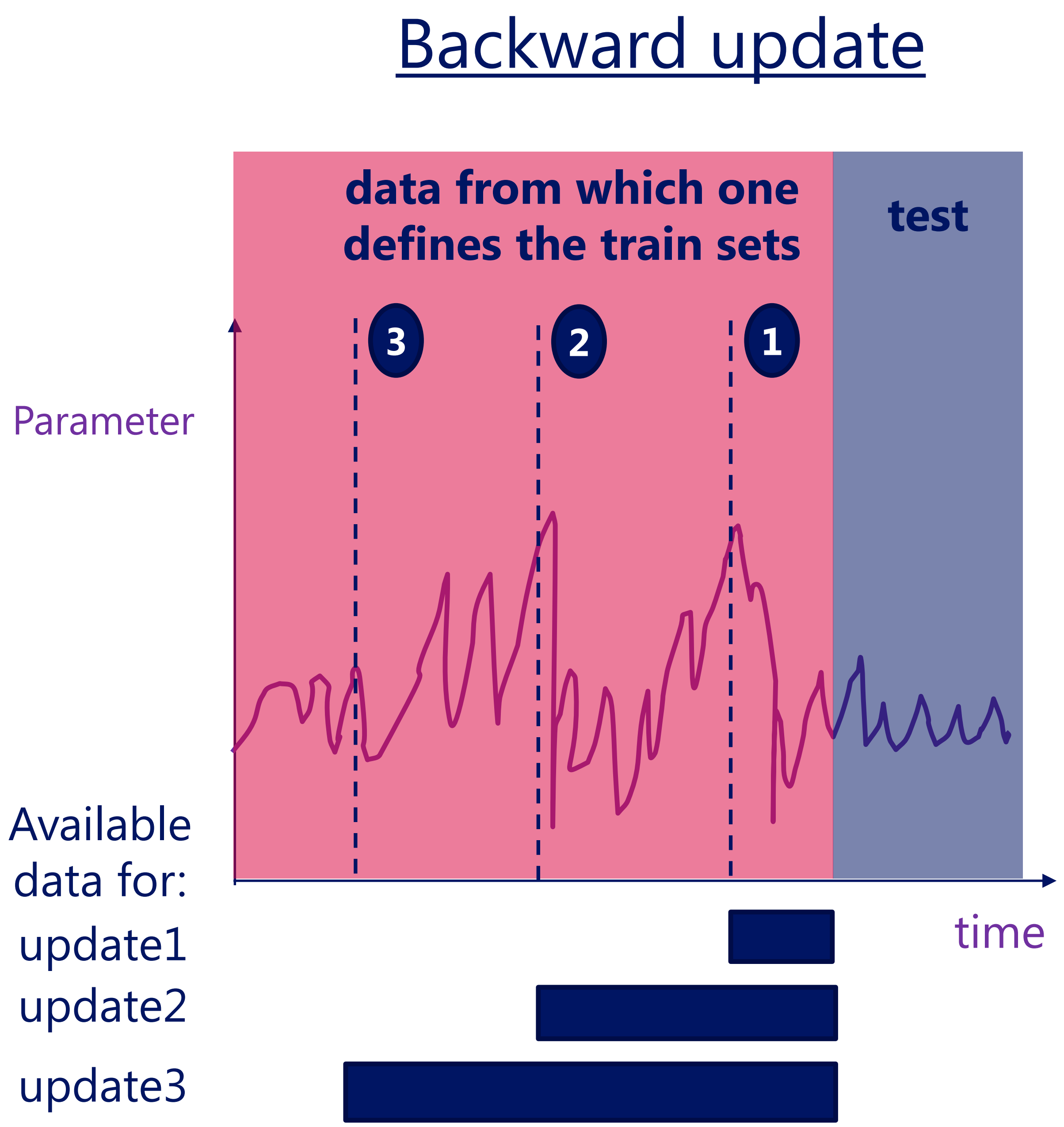


- An extreme case: can gain 15% accuracy by increasing train set from 10 to 50 wafers.
- Model retrain, enabled by a big data system, is important here.

Accuracy: data size and type

We find it is important to test how the blind test accuracy differs between two different update methods

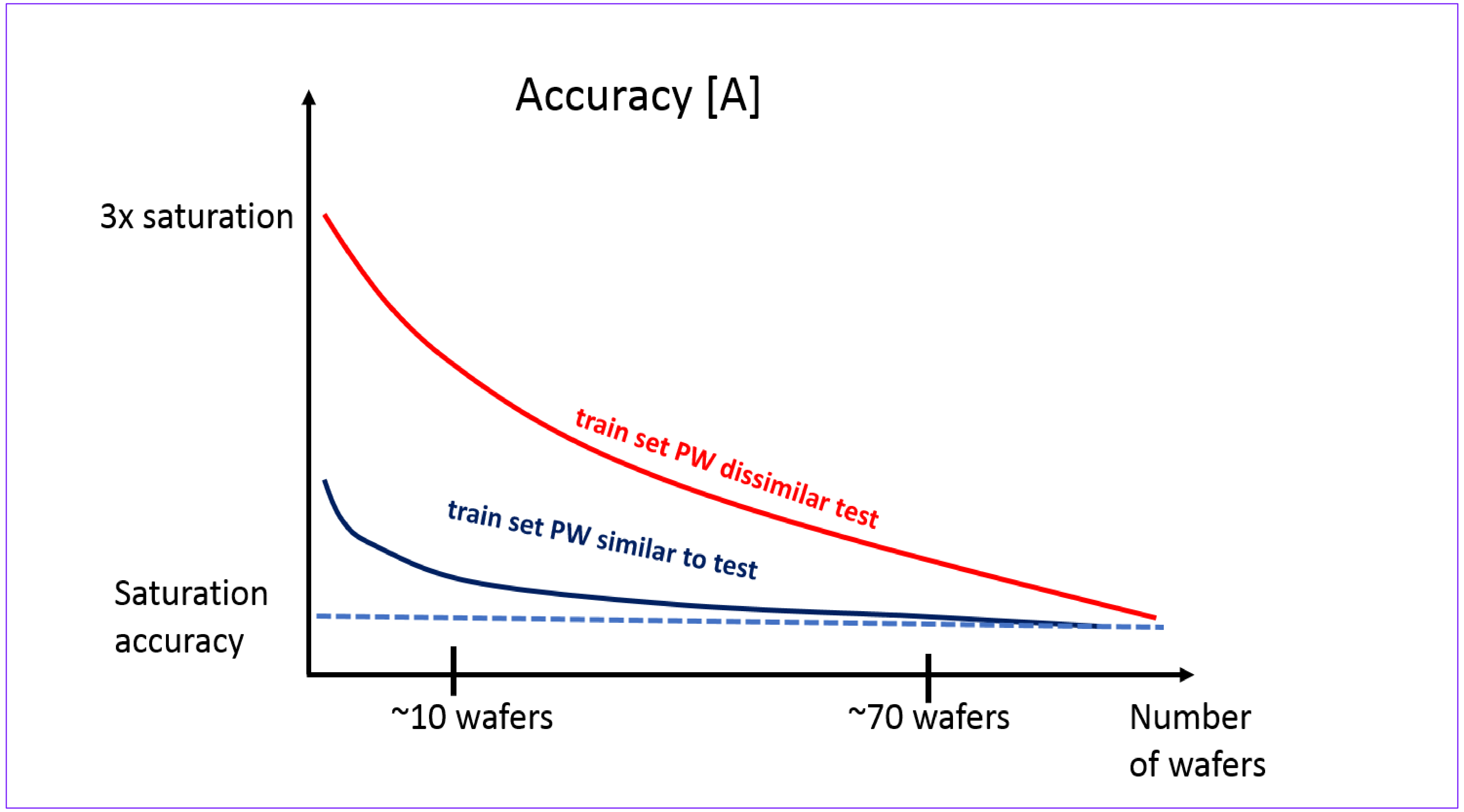




→ Any sizeable difference between forward and backward indicates process instability and the need for a dynamic update and control system.

Accuracy: data size and type

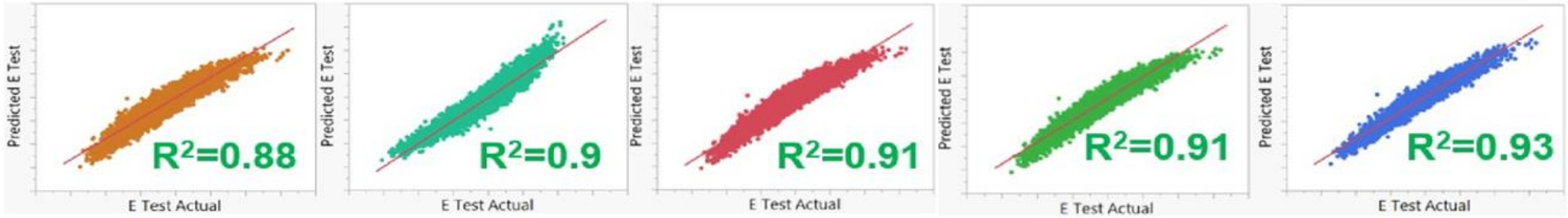
Example below: transfer of a physical model solution on SA to Nova IM.

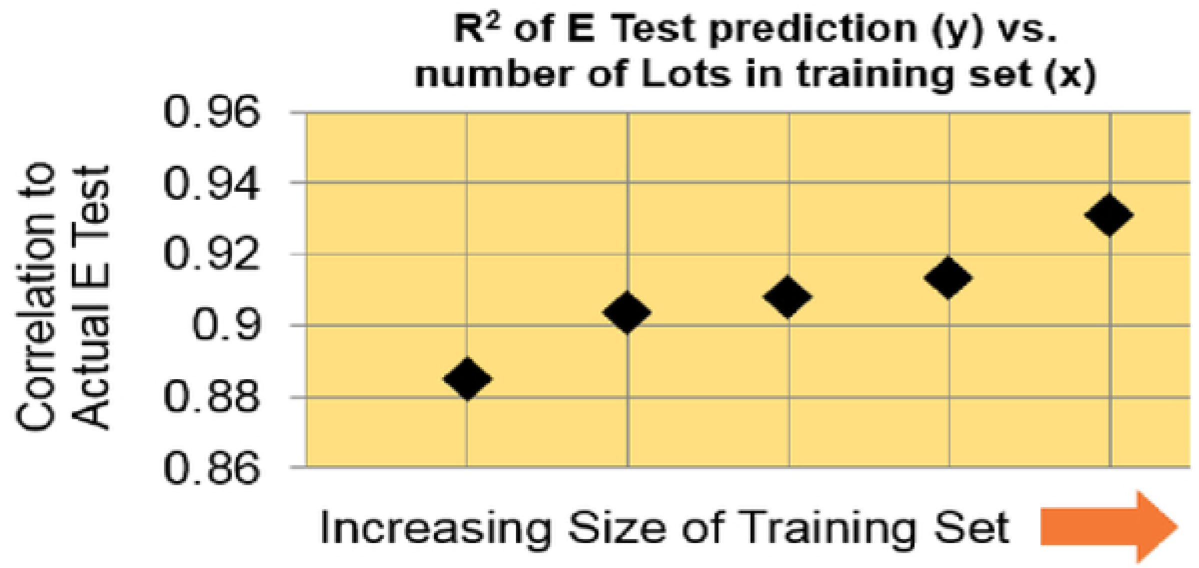


- Process Window (PW) is important. Plot is a sketch of what happens when the PW drifted with time.
- Clearly if PW(train) ~ PW(test), accuracy is better.
- Model retrain, enabled by a big data system, is important here.

Accuracy: data size-from SPIE Advanced Lithography P. Timoney *et al.* 10585-32

- E Test results from one of the GlobalFoundries products



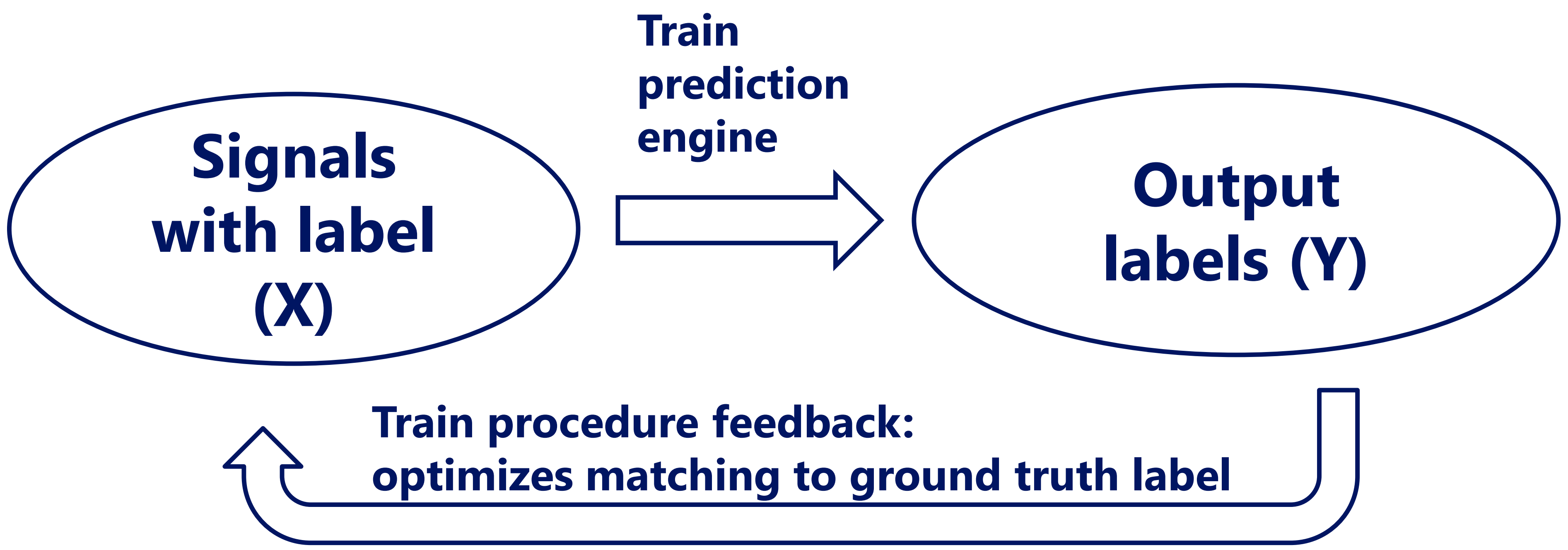


Change with increasing training data size is moderate.

Algorithm performance: repeatability

- Modify model setup to balance accuracy with repeatability.

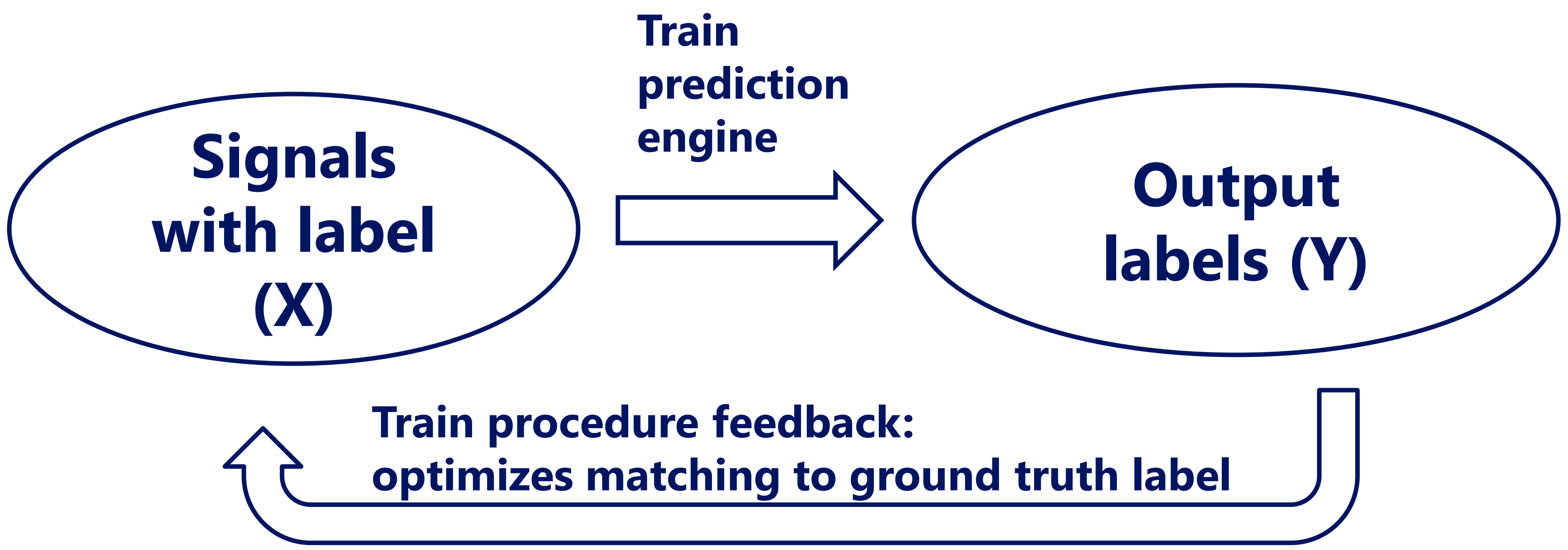
<u>Typically in machine learning</u>



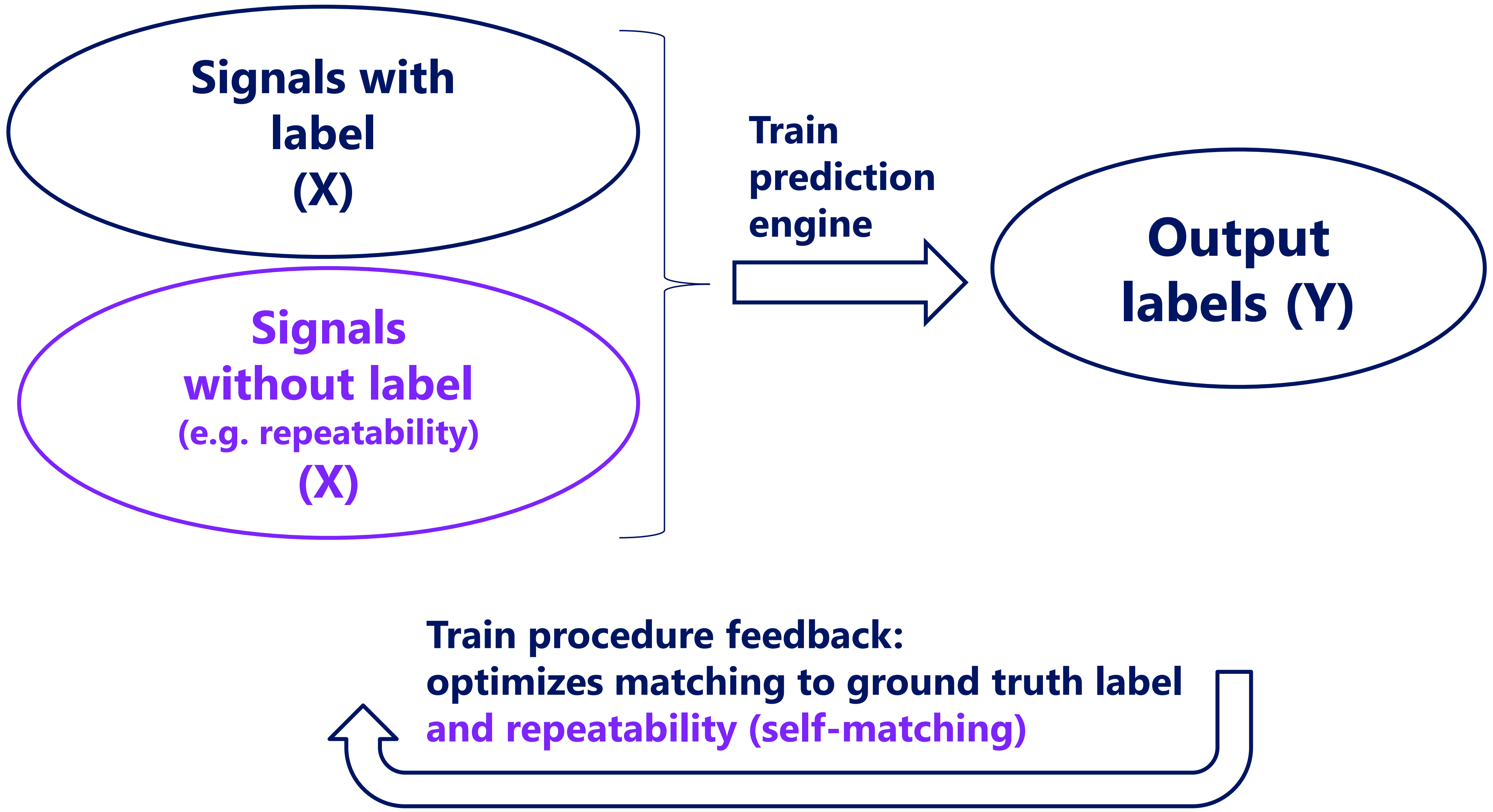
Algorithm performance: repeatability

- Modify model setup to balance accuracy with repeatability.

Typically in machine learning



Machine learning in OCD metrology



Repeatability optimization

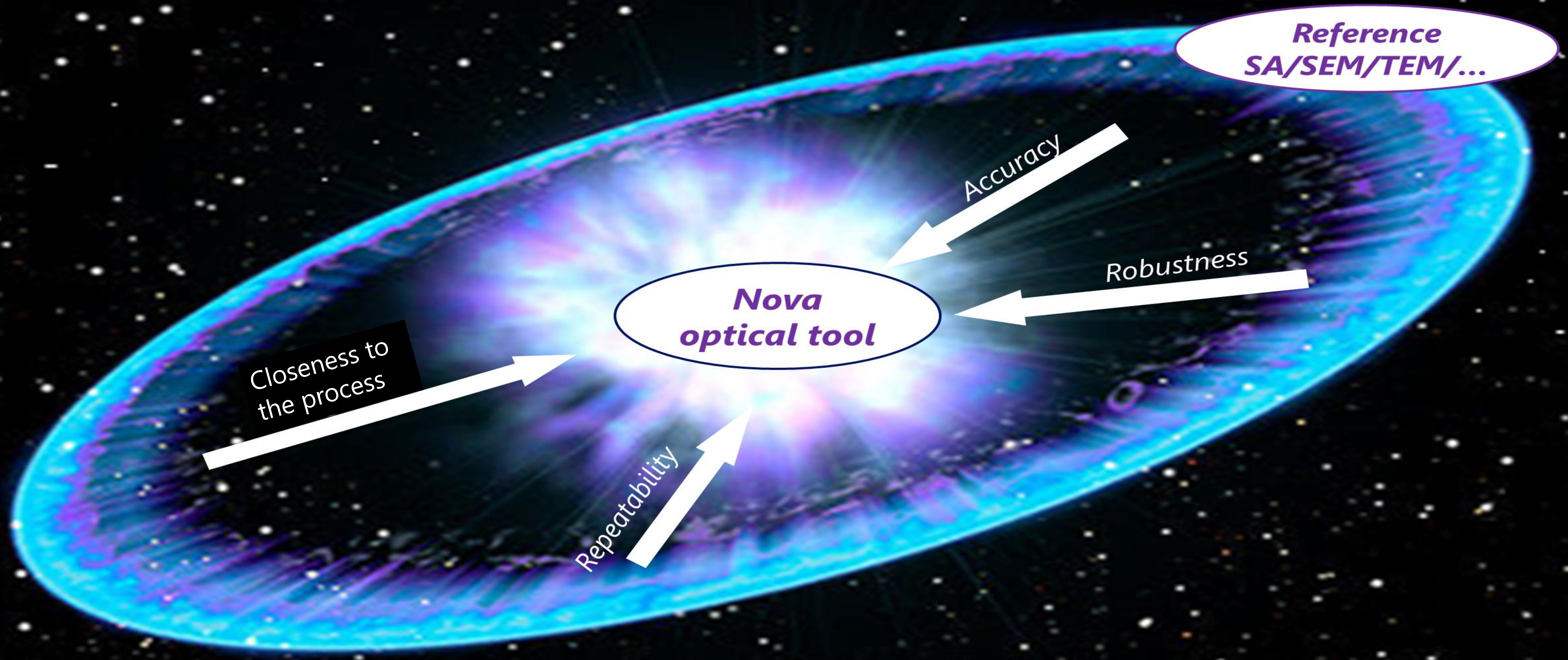
- Modify model setup to balance accuracy with repeatability.

Layer	Accuracy change [relative in percent]	Repeatability improvement [relative in percent]
Logic	-21	+136
	-17	+133
	0	+40
	-18	+123
	-17	+155
Memory	-13	+123
	-7	+11

Accuracy penalties are all less than 10% of customer spec

Improvement are significant, can be 50%-100% of customer spec

In summary: machine learning is valuable for OCD



But ….

There are still unsolved problems, mainly:

- Interpretability: the `black box' issue.
- Reference cost.

⇨

Different approaches:

- Combining physical modeling and machine learning.
- Another alternative: see talk by Noam Tal at APC2018.

Thanks to all my co-authors at Nova: the machine learning and big data group at Nova

Eitan, Ilya, YongHa, Noam, Oded, Shay, Ariel, Eylon, and Tal

And to the GLOBALFOUNDRIES and Nova authors of

SPIE Advanced Lithography – Metrology, Inspection, and Process Control for Microlithography – P. Timoney *et al*. 10585-32

P. Timoney, T. Kagalwala, E. Reis, H. Lazkani, J. Hurley, H. Liu, B. Kang, P. Isbester, N. Yellai, M. Shifrin, Y. Etzioni