



Heriot-Watt University
Research Gateway

Machine Learning Based Edge Placement Error Analysis and Optimization

Citation for published version:

Ngo, AT, Dey, B, Halder, S, Gendt, SD & Wang, C 2023, 'Machine Learning Based Edge Placement Error Analysis and Optimization: A Systematic Review', *IEEE Transactions on Semiconductor Manufacturing*, vol. 36, no. 1, pp. 1-13. <https://doi.org/10.1109/TSM.2022.3217326>

Digital Object Identifier (DOI):

[10.1109/TSM.2022.3217326](https://doi.org/10.1109/TSM.2022.3217326)

Link:

[Link to publication record in Heriot-Watt Research Portal](#)

Document Version:

Peer reviewed version

Published In:

IEEE Transactions on Semiconductor Manufacturing

Publisher Rights Statement:

© 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.




General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

MACHINE LEARNING BASED EDGE PLACEMENT ERROR ANALYSIS AND OPTIMIZATION: A SYSTEMATIC REVIEW

Anh Tuan Ngo , Bappaditya Dey , *Member, IEEE*, Sandip Halder, Stefan De Gendt , *Member, IEEE*, and Changhai Wang, *Member, IEEE*

Abstract—As the semiconductor manufacturing process is moving towards the 3 nm node, there is a crucial need to reduce the edge placement error (EPE) to ensure proper functioning of the integrated circuit (IC) devices. EPE is the most important metric that quantify the fidelity of fabricated patterns in multi-patterning processes, and it is the combination of overlay errors and critical dimension (CD) errors. Recent advances in machine learning have enabled many new possibilities to improve the performance and efficiency of EPE optimization techniques. In this paper, we conducted a survey of recent research work that applied machine learning/ deep learning techniques for the purposes of enhancing virtual overlay metrology, reducing overlay error, and improving mask optimization methods for EPE reduction. Thorough discussions about the objectives, datasets, input features, models, key findings, and limitations are provided. In general, the results of the review work show a great potential of machine learning techniques in aiding the improvement of EPE in the field of semiconductor manufacturing.

Index Terms— Edge placement error, deep learning, machine learning, metrology, optical proximity correction, overlay, semiconductor, sub-resolution assist feature.

I. INTRODUCTION

MOORE'S Law states that transistor density will double every two years, which is still upheld until today, thanks to the constant innovations in the semiconductor industry. Advance technologies such as extreme ultraviolet (EUV) lithography and multiple patterning techniques [1], [2] are the driving forces to keep extending Moore's Law, as well as leading the whole industry moving towards 3 nm node and beyond. In these patterning schemes, especially in multi-patterning processes, placement error control between two process layers is strictly required to maintain high yield and

performance [3], [4], and the most important metrics to evaluate placement error is EPE. EPE can be defined as the relative displacement of the edges of features from their intended target location, and it is a directly representation of the pattern fidelity of fabricated structures. EPE is the combination of overlay errors and CD errors including OPC and stochastics [5]. The visualization of EPE and the distribution budget of its components are described in Fig. 1.

In [6], Jan Mulken et al proposed an analytical method to calculate the EPE, which is shown in (1). Here in this equation, $HalfRange_{OPC}$ is the half-range of the CD error due to optical proximity residuals. The second term σ_{PBA} stands for proximity bias average error, which is the feature-specific field average CD induced by scanner tool-to-tool variance. The third term, σ_{LWR} , refers to line width roughness, which is local errors that come from resist and photon stochastics in lithography process. The two terms inside the square root are the global errors of EPE. Overlay error ($\sigma_{overlay}$) is the placement shift error of the printed patterns, while σ_{CDU} is the CD uniformity that originates from etch and deposition steps.

$$EPE_{max} = \frac{HalfRange_{OPC}}{2} + \frac{3\sigma_{PBA}}{2} + \frac{6\sigma_{LWR}}{\sqrt{2}} + \sqrt{(3\sigma_{overlay})^2 + \left(\frac{3\sigma_{CDU}}{2}\right)^2} \quad (1)$$

EPE is the main challenge that inhibits the continuation of devices' size shrinkage in semiconductor industry. Without good value of the EPE, pattern-to-pattern electrical contacts can be poor, which would lead to fatal failures of IC devices such as short circuits or broken connections. Therefore, it is essential to develop effective EPE analysis and control solutions to ensure proper functionality of fabricated semiconductor devices. As the EPE consists of both overlay and CD errors, this goal can only be achieved by optimizing both overlay metrology and control methods, as well as mask design computational techniques in lithography process. Fig. 2 illustrates the EPE optimization flow.

Corresponding author: Bappaditya Dey.

Anh Tuan Ngo is with Heriot-Watt University, EH14 4AS Edinburgh, Scotland, UK. He is now completing an internship as a master thesis student at imec, Kapeldreef 75, 3001 Leuven, Belgium.

Bappaditya Dey (email: bappaditya.dey@imec.be) and Sandip Halder are with imec, Kapeldreef 75, 3001 Leuven, Belgium.

Stefan De Gendt is with the Department of Materials Engineering and Department of Chemistry, KU Leuven, 3001, Leuven, Belgium.

Changhai Wang is with Institute of Sensors, Signals and Systems, School of Engineering & Physical Sciences, Heriot-Watt University, EH14 4AS Edinburgh, Scotland, UK.

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>

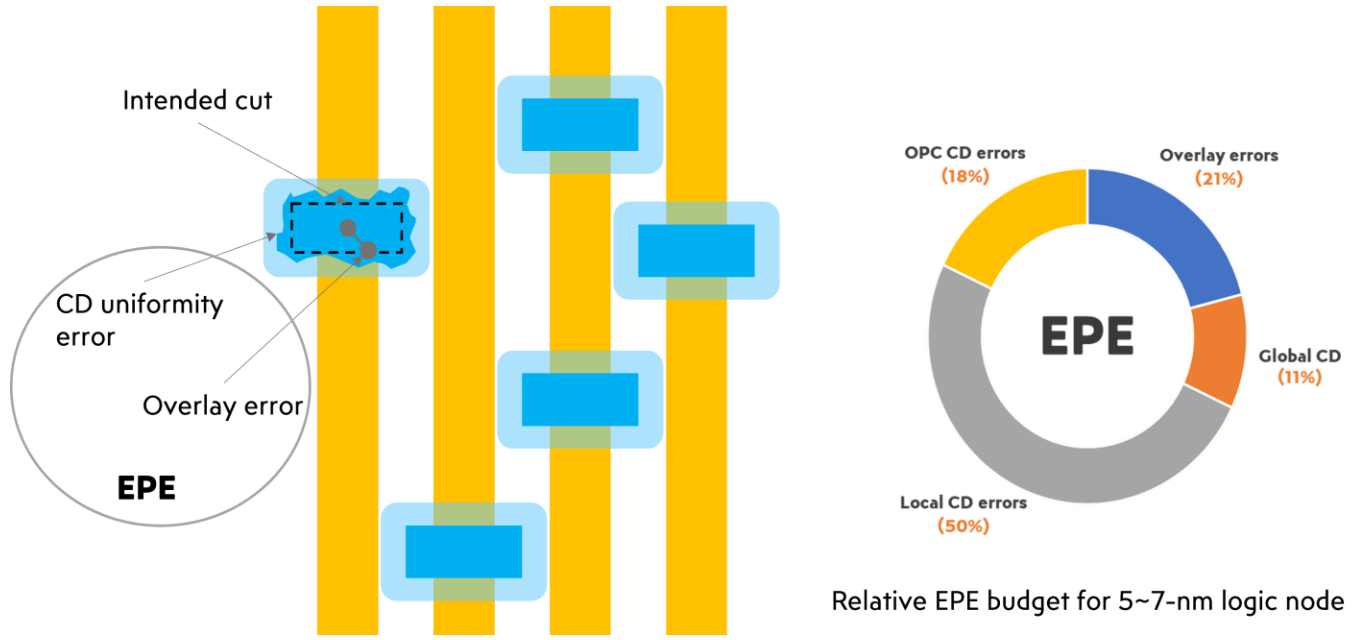


Fig. 1. EPE visualization and its distribution budget for 5~7-nm logic node [5].

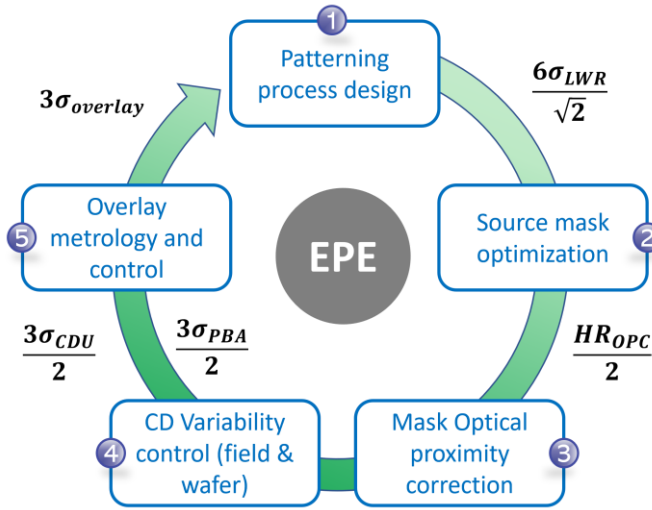


Fig. 2. EPE optimization flow [7].

In modern measurement schemes, overlay is measured by metrology tools which detect and scan specific targets or marks on scribe lines and/or within the field of the product wafers [4], [8]. Typically, in the semiconductor industry, a sampling approach for overlay metrology is employed, in which only a subset of wafers in the production line are measured by physical tools such as KLA optical-based overlay or ASML YieldStar diffraction-based overlay. This method suffers from a major drawback that a lot of wafers are not measured to detect abnormal overlay signatures. As a result, there is a great chance that underqualified wafers are not detected and can slip through to the next process steps. This can cause a serious problem in the later stages, since many dies will not yield, or wafers need to be scrapped [9]. One solution to tackle this problem is applying the virtual overlay metrology (VOM) method. There are tens of gigabytes of data generated every day from sensors in the lithography clusters

and other processing equipment in a typical semiconductor plant [10]. This amount of available data can be utilized by applying machine learning techniques to build a model which can predict the overlay property of each wafer, without actually measuring the overlay by physical metrology tools. Machine learning based VOM models can also find and link specific root causes of abnormal overlay excursions, enabling software tools or operators to take corrective and/or preventive actions in timely manner [9], [11]. Machine learning can also be used to aid overlay metrology tools in order to improve their measurement accuracy and applied in control schemes in the fabrication process to reduce the overlay error.

Regarding the problem of reducing EPE in lithography process, mask synthesis optimization is one of the most crucial tasks, because it can help optimize both the placement errors and the critical dimension of the desired patterns. The conventional methods such as the rule-based [12]–[14] and model-based techniques [15]–[17] for mask synthesis have been applied for many years. Rule-based approaches are capable of achieving high performance on simple designs, but they are not able to handle complex target patterns. Whereas in contrast, the model-based approaches can produce great masks with high accuracy of printed patterns, but they are very time consuming and computationally expensive. To ensure accurate performance with good runtime, different machine learning based mask optimization methods have been introduced [18]–[21]. One of these studies stated that a machine learning based model can approach a model-based method of a commercial software in terms of EPE performance but can produce the result up to 144 times faster [21]. This shows a great potential of machine learning in addressing mask optimization problems.

In this paper, a systematic review of applications based on machine learning techniques for the purposes of enhancing virtual overlay metrology, reducing overlay error, and

improving mask optimization methods for EPE reduction was conducted. We also thoroughly discussed the objectives, datasets, input features, models, key findings, and limitations of the work in the literature. Finally, future work in machine learning based techniques for EPE analysis and optimization is suggested.

II. METHODOLOGY

In this review, the Preferred Reporting Elements for Systematic Reviews and Meta-Analyses (PRISMA) guidelines was followed [22]. The reviewed papers were indexed in Google Scholar database, and the keywords for the queries were: 1. overlay, 2. mask optimization, 3. SRAF, 4. OPC, 5. analysis, 6. optimization, 7. metrology, 8. semiconductor, 9. machine learning, and 10. deep learning. In the search phrases, truncation was applied to the words ‘analysis*’ and ‘optimization*’ in order to seek for their variations or synonyms. Two search queries were used in this work, one was for overlay analysis and optimization, and the other was for EPE. They were constructed by using logical operators as follow: 1. for overlay analysis and optimization: (1) AND (5) AND (6) AND (7) AND (8) AND (9 OR 10), and 2. for mask optimization: (2) AND (3 OR 4) AND (8) AND (9 OR 10).

To filter the results returned from the search query, this review applied three inclusion and three exclusion criteria. The three inclusion criteria consist of: (i) articles published from 2010, (ii) peer-reviewed journal and conference publications, and (iii) articles that focused on machine learning based techniques for overlay – EPE analysis and optimization. The criteria used for exclusion included: (i) articles that did not propose new machine learning methods to analyze or optimize overlay – EPE, (ii) review or survey articles, (iii) articles that were not written in English.

III. RESULTS

Using the queries described before, the results returned 2110 and 599 articles for overlay analysis and mask optimization, respectively. In the initial screening after applying the inclusion and exclusion criteria on the title and abstracts of these articles, 2027 papers for overlay and 542 papers for EPE were filtered out. Finally, by applying the inclusion and exclusion criteria on the full text of those papers, 19 overlay papers and 17 mask synthesis papers remained. The selecting procedure is illustrated in Fig. 3.

For the articles in the topic of overlay analysis and optimization, they can be grouped in three categories, which are: (i) virtual overlay metrology, (ii) improvement of overlay metrology accuracy, and (iii) control scheme to improve overlay in manufacturing process. In terms of mask optimization tasks for EPE optimization, the publications are divided into two groups: (i) sub-resolution assist feature insertion, and (ii) optical proximity correction. The detail of which article belongs to which category is shown in Table I. Additionally, in Table II and Table III, the information about the reviewed articles was summarized, including the discussion about: (i) the objectives, (ii) the dataset and/or input features, (iii) the machine learning models, (iv) the results and key findings, and (v) the limitations.

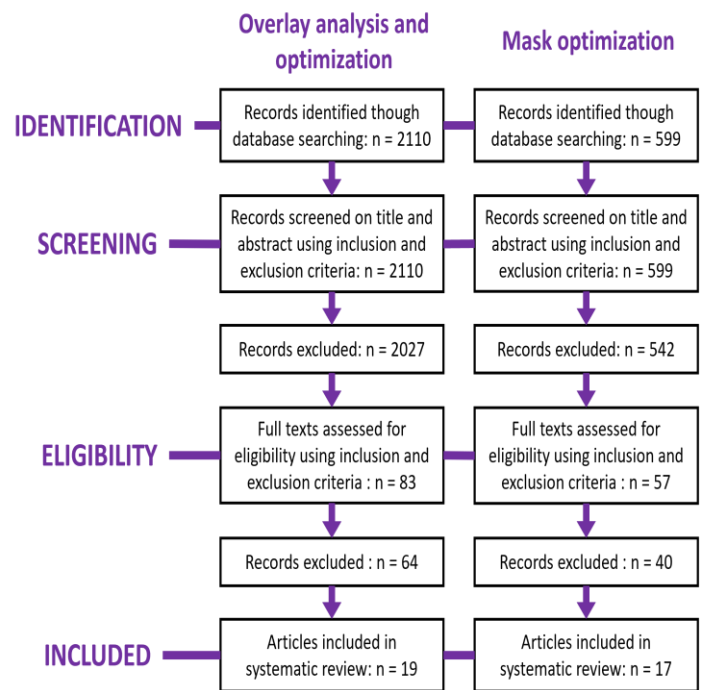


Fig. 3. Diagram of PRISMA selection procedure.

TABLE I
CATEGORIZATION OF THE REVIEWED ARTICLES

Paradigm	Categories	Articles
Overlay analysis and optimization	Virtual overlay metrology	[4], [9], [11], [23]–[30]
	Improvement of overlay metrology	[8], [31]–[33]
	Control scheme to improve overlay in manufacturing process	[34]–[37]
Mask optimization	Sub-resolution assist feature insertion	[21], [38]–[40]
	Optical proximity correction	[18]–[20], [41]–[50]

IV. DISCUSSION: MACHINE LEARNING BASED OVERLAY ANALYSIS AND OPTIMIZATION

Table II shows the summary results of 19 different studies about overlay analysis and optimization techniques using machine learning.

A. Virtual Overlay Metrology

VOM is one of the techniques that has attracted a lot of attention from researchers in terms of optimizing overlay in semiconductor process manufacturing. As it is almost impossible to measure every wafer of each lot in the lithography steps by physical measurement tools, VOM shows great potential to tackle this problem by leveraging different machine learning techniques with a huge amount of data from different sensors and the context of process steps to predict overlay signature of each wafer. Thus, it is possible to quickly detect wafers with significant overlay excursion and the steps through which those wafers have been processed.

TABLE II
SUMMARY OF MACHINE LEARNING BASED OVERLAY ANALYSIS AND OPTIMIZATION STUDIES

Studies	Objectives	Dataset/ Input Features	ML models	Results and Key Findings	Limitations
[4] (Cited by 22)	- Investigate the effect of SWA profiles on the overlay target. - Propose a model for predicting displacement offset by SWAs inputs.	The asymmetry in intensity between the +1 st and -1 st orders by using a grating with SWA α of 90°, 89°, 88°, 87°, and 86°. (The dataset is obtained at nominal process condition.)	ANN with Levenberg-Marquardt backpropagation learning	Improvement of the proposed ANN model compared to the conventional linear regression model with a wavelength of 650 nm: - MSE value: + ANN model: 0.0331 + Conventional model: 0.320 - Mean of the residual value: + ANN model: 0.135 + Conventional model: 0.887 (The results are obtained from modelling and simulation work.)	The authors only consider the slope sidewall grating for the top layer and vertical side wall for the bottom layer, while in reality, bottom layer grating can also have slope side wall.
[8] (Cited by 1)	Reducing the inaccuracy in misregistration measurements of the optical on product overlay at the after-develop inspection.	- Images of 1000 different sites from each wafer. - Various physical characteristics of the wafer (kernels, targets' accuracy). (The dataset is obtained at nominal process condition.)	- k-fold Cross-Validation for optimizing ML model's hyperparameters - Gradient Boosting	- Accuracy improvement compared to imaging-based overlay technique: + X: 15% + Y: 17% - Residual overlay error: + X m3s: < 1.6 nm + Y m3s: < 2 nm (The performance of ML models is verified by on-wafer measurements)	Small dataset.
[9] (Cited by 1)	Build a prediction model to provide overlay estimates for every wafer in the lot and to identify the main root causes for the significant variations in the implant-layer overlay.	- Average exposure fingerprints of the scanners. - Measured overlay of the first implant layer. - Context, scanner logging, etc. (The dataset is obtained at nominal process condition.)	- Multivariate linear regression - t-SNE	Predicted overlay correlation: + R-squared X: 0.67 + R-squared Y: 0.67 - t-SNE algorithm successfully identify wafers that have additional distortion contributions. (The performance of ML models is verified by on-wafer measurements)	Low R-squared correlation coefficient.
[11] (Cited by 10)	Virtual overlay metrology for fault detection.	- Wafer alignment metrology for all colors. - Wafer leveling metrology. - TWINSCAN Context. (The dataset is obtained at nominal process condition.)	Bayesian Interpolation for ANN	- Predicted overlay correlation R-squared: 0.7513 - Residual overlay error in 20 nm DRAM manufacturing process: + X m3s: 4.4 nm + Y m3s: 5.7 nm (The performance of ML model is verified by on-wafer measurements)	Low R-squared correlation coefficient.
[23] (Cited by 17)	- To build a correlation between context and TWINSCAN metrology to the measured overlay of YieldStar system. - Controller to improve the overlay.	- Pre & Post exposure sensor data. - Pre exposure context: + Chuck ID + Wafer sequence (The dataset is obtained at nominal process condition.)	Time series NARX	- Predicted overlay correlation R-squared: 0.8321 - Residual overlay error in 14 nm node process: + X m3s: 4.37 nm + Y m3s: 4.84 nm - The proposed control method can pick up on trained lot to lot wafer to wafer contributors to systematic and random overlay error. (The performance of ML model is verified by on-wafer measurements)	Information about the amount of data used for the training and validating of the model is not clearly specify.
[24] (Cited by 4)	Virtual overlay metrology for a series of implant layers, for the purpose of reducing the physical metrology process and optimizing overlay control.	- Average exposure fingerprints of the scanners. - Measured overlay of the first implant layer. - Context, scanner logging, etc. (The dataset is obtained at nominal process condition.)	Feed-forward ANN regression	- Predicted overlay correlation: + R-squared X: 0.65 + R-squared Y: 0.75 - Residual overlay error: + X m3s: 4.7 nm + Y m3s: 3.0 nm (The performance of ML model is verified by on-wafer measurements)	Short examining time period.
[25] (Cited by)	Propose a method using ML and OCD to monitor the	45 wafers were measured by SpectraProbe to get the Mueller Matrix information.	PCA regression algorithm	- R-squared correlation of SpectraProbe prediction data and HV-SEM reference data: + Tilt-X: 0.92	Small amount of training data.

2)	channel hole tilting in-line.	<i>(The dataset is obtained at nominal process condition.)</i>		+ Tilt-Y: 0.94 - Measurement time was reduced by 40% when using the proposed OCD method compared to the HV-SEM method. <i>(The performance of ML model is verified by on-wafer measurements)</i>	
[26] <i>(Cited by 2)</i>	Propose a virtual overlay metrology which aims at predicting the overlay for a series of implant layers.	Overlay of another implant layer of the same wafer, exposure tool fingerprints, scanner logging, and process data. <i>(The dataset is obtained at nominal process condition.)</i>	- Lasso - Random Forests - Feedforward ANN	R-squared for both X and Y achieved approximately 0.8 when removed 10-par correctable errors from the overlay data. <i>(The performance of ML model is verified by on-wafer measurements)</i>	The prediction for some parameters such as Tx, Ty, and My are inaccurate (R square < 0.3).
[27] <i>(Cited by 0)</i>	Monitor in real time the process variation on wafer by the data from the Diffraction-based-overlay.	Film thickness, CD width, SWA under interest. <i>(The dataset is based on the study of process sensitivity)</i>	Feed-forward ANN regression	- Correlation of the predicted and simulated stack sensitivity: R-squared: 0.99 - Maximum prediction error: <5nm, which accounts for 3% of the nominal film thickness for that layer. <i>(The results are obtained from both simulation work and on-wafer measurements)</i>	- Small sample size (only five test wafers). - Training data is generated by simulation software.
[28] <i>(Cited by 4)</i>	Propose a model to generate dense computed overlay metrology.	- Customers APC and TWINSCAN context. - Leveling metrology deconvolved. - Alignment metrology - Residual vector per alignment metrology. <i>(The dataset is obtained at nominal process condition.)</i>	ANN with Bayesian regularization	- Successfully extracted four unique signatures from UVLS measurement data which serve as one of the inputs for the ML prediction model. - R-squared correlation of prediction data and the measured metrology data for overlay Y: 0.72 to 0.81. <i>(The performance of ML model is verified by on-wafer measurements)</i>	Small dataset, only ten wafers' metrology data was collected to use as training data.
[29] <i>(Cited by 4)</i>	Propose a model to enhance the alignment metrology by pairing alignment metrology with leveling metrology data in a wafer.	- Average sparse estimation across 4 colors using 32 marks on the wafer. - 4 out of the 11 leveling signatures. - Wafer quality per alignment metrology. <i>(The dataset is obtained at nominal process condition.)</i>	- PCA is used for dimensional reduction - Bayesian neural network with automated regularization	- The proposed model can capture > 1nm alignment metrology signature - Dense estimation performance of 3rd order HOWA model fit to the original 32 marks combine with leveling data outperformed the model which only fit to the original 32 marks. <i>(The performance of ML model is verified by on-wafer measurements)</i>	The authors only used 4 out of 11 leveling signatures as the input to the model, increasing the input signatures may also increase the performance of the model.
[30] <i>(Cited by 8)</i>	- Identify which processing steps that a group of wafers has in common. - Compute overlay signature.	- XY vector map of clamped wafers shape. - Wafer alignment metrology for all colors. - TWINSCAN context <i>(The dataset is obtained at nominal process condition.)</i>	- Local regression nonparametric equation with k-nearest neighbor search - Hierarchical clustering with dendrograms	- Wafer's overlay signature - Pointed out which prior processing steps the grouped wafers have in common. <i>(The performance of ML model is verified by on-wafer measurements)</i>	Lack of metric for grading the performance of the model which compute the overlay signature.
[31] <i>(Cited by 1)</i>	Develop a self-reference target method in overlay metrology.	Array of small individual targets carrying different dialed-in overlay bias in each target. <i>(The dataset is obtained at nominal process condition.)</i>	Custom ML regression	- m3s point to point delta to decap SEM for multiple wafers values are in acceptable range. - High correlation between IDM ASR and SEM per field. - Total Measurement Uncertainty: 0.2 nm - Move Acquire Measure: < 0.4 sec. <i>(The performance of ML model is verified by on-wafer measurements)</i>	Information about the amount of data used for the training and validating of the model is not clearly specify.
[32] <i>(Cited by 1)</i>	Improve the TIS to enhance the accuracy of overlay measurement.	- The ground truth from the calibrated overlay value. - 0° images - Contrast - Q-Merit - Kernel merits <i>(The dataset is obtained at nominal process condition.)</i>	Gradient Boosting	- TIS mean decrease: + DRAM: 40% + NAND: 81% - TIS 3σ decrease: + DRAM: 34% + NAND: 28% <i>(The performance of ML model is verified by on-wafer measurements)</i>	Training data was collected only from several sites on a wafer, thus limit the learning capacity of the model.
[33] <i>(Cited by 0)</i>	Propose a DBO model to reduce the measurement inaccuracy due to the sidewalls effect in the bottom gratings' targets.	18900 pupil images of 30 types of overlay targets using an optical scatterometry. These images are then applied PCA method to extract the features which are used as the input for the DBO model.	ANN with Bayesian regularization	- Using the dose level of 500 mJ/s·cm ² for each pupil image, the overlay MSE of the proposed model reduced 42.22% (from 2.25 nm ² to 1.3 nm ²) compared to the conventional DBO model. - Using the dose level of 1000 mJ/s·cm ² , the MSE was improved further to 0.4 nm ² . <i>(The results are obtained from modelling and</i>	The authors only consider the slope sidewall grating for the second layer and vertical side wall for the first layer, while in reality, first layer grating can also

		<i>(The dataset is obtained at nominal process condition.)</i>		<i>simulation work.)</i>	have slope side wall.
[34] <i>(Cited by 4)</i>	Monitoring the impact of non-lithography context on overlay.	More than 100 different context parameters per wafer (1200 lots with 4 wafers per lot). <i>(The dataset is obtained at nominal process condition.)</i>	Parametric clustering	- Successfully identify the correlation between non-litho context and overlay error. - Applying run-to-run control simulation can achieve overlay improvement (for a specific context value) as: + X-Overlay: 23.8% + Y-Overlay: 18.5% <i>(The results are obtained from both simulation work and on-wafer measurements)</i>	Not all available context parameters are examined. The authors only show an analysis example of one specific non-litho context for monitoring overlay signature.
[35] <i>(Cited by 0)</i>	Propose a method for higher alignment accuracy and a model to improve wafer edge overlay.	- For chucking improvement: parameters in iAS (Pitch, Roll, Bow Shape) - For edge overlay improvement: Wafers' overlay data which has intentional error. <i>(The dataset is obtained at nominal process condition.)</i>	- Bayesian optimization for chucking improvement - Circular Bessel function to improve edge overlay	- Chucking deformation was reduced about 1 nm (24%) by optimizing loading parameters with the proposed model. - Using 7th order Bessel circular showed well corrected overlay error for both center and edge of the examining wafers. <i>(The results are obtained from modelling and simulation work.)</i>	Information about the amount of data used for the training and validating of the model is not clearly specify.
[36] <i>(Cited by 14)</i>	Develop a control technique for overlay compensation with dynamic metrology delay.	10 variables including mask intra-field overlay errors involving rotation, translation, magnification, and wafer inter-field overlay errors such as expansion and rotation in x- and y- axes. <i>(The dataset is based on the study of process sensitivity)</i>	ϵ -SVR	- ϵ -SVR controller's improvement over conventional EWMA controller: + X-axis variation: 24% + Y-axis variation: 8% + X-axis overlay error reduction: 96% + Y-axis overlay error reduction: 97% - Under fast and deep delay, Lyapunov mapping function and Razumikhin condition can ensure the stability for the system. <i>(The results are obtained from modelling and simulation work.)</i>	When the variation in overlay variables is small, for some parameters, the ϵ -SVR performed worse than the conventional EWMA controller.
[37] <i>(Cited by 118)</i>	Propose a virtual overlay embedded run-to-run control system.	- 37 sensors parameters were collected from in two chucks in photolithography process, results in 1612 and 1563 data points for chuck 1 and chuck 2, respectively. - Summary of sensor parameters' statistics (148 inputs for 37 parameters) were used to train the VM model. <i>(The dataset is obtained at nominal process condition.)</i>	- Variable selection methods: stepwise linear regression, decision trees, GA, PCA, and KPCA - VM prediction models: linear regression, MLP, k-NN, SVR. - EWMA run-to-run controller	- The MASE of all regression models was less than 10% after training with data collected in 5 months, which satisfies the production engineers criteria. - Upon running with Monte Carlo simulation, the VM embedded EWMA decreased MSE and MASE on average 70% and 30%, respectively. <i>(The results are obtained from both simulation work and on-wafer measurements)</i>	The run-to-run control has not been verified by real equipment, which may contain more complicated factors compared to the simulation environment.

Abbreviations:

ANN (Artificial Neural Network), DBO (Diffraction-Based Overlay), EWMA (Exponentially Weighted Moving Averages), GA (Genetic Algorithm), ϵ -SVR (Epsilon - Support Vector Regression), iAS (inline Alignment Station), k-NN (k-Nearest Neighbor), KPCA (Kernel Principal Component Analysis), m3s (mean plus 3 standard deviation), MASE (Mean Absolute Specification Error), ML (Machine Learning), MLP (Multi-Layer Perceptron), MSE (Mean Squared Error), NARX (Nonlinear Autoregressive Network with Exogenous Inputs), OCD (Optical Critical Dimension), PCA (Principal Component Analysis), SWA (Sidewall Angle), TIS (Tool Induced Shift), t-SNE (t-distributed Stochastic Neighbor Embedding)

Among all the papers that focus on VOM in our review, ANN based regression models were used most frequently [4], [11], [23], [24], [26]–[29]. ANNs are made up of a network of interconnected artificial neurons, which to some extent mimics the model of the biological brain. Although regression model based on ANN is complex and computationally expensive, it is a very powerful technique which can capture highly nonlinear relationships of the input features [51]. In those studies that used ANN, the performance in [23] was recorded as the best, with the correlation of predicted and measured overlay R-squared value of 0.8321. The authors of that work used an approach called time series NARX with the fitting function shown by (2), where $u(t)$ is the input signal (sensors data and input context) at time t , $y(t)$ is the corresponding

overlay prediction output at time t , n_u and n_y are the input and output sequence numbers, and $f()$ is the approximation function that can be learnt by an ANN.

$$y(t) = f(u(t - n_u), \dots, u(t - 1), u(t), y(t - n_y), \dots, y(t - 1)) \quad (2)$$

Aside from the ANN based approach, the traditional machine learning models such as linear regression, t-SNE [9], PCA regression [25], Lasso regression, and Random Forest [26] have also been employed to solve the VM overlay problems. The predicted overlay performance (R-squared value) of those models varies from 0.65 to 0.80.

B. Overlay Metrology's Accuracy Improvement

Applications of machine learning models in improving the accuracy of overlay metrology techniques have been successfully implemented in several studies. In [8], Gradient Boosting model was used to reduce the error in misregistration measurements of the optical overlay for after-develop inspection. Gradient Boosting is an ensemble machine learning model which takes the advantages of multiple weak Decision Tree predictors to obtain the strong one. Using this model, Verner et al has improved the accuracy of the conventional imaging-based overlay measurement by 15% and 17% on gratings X and Y, respectively. The Gradient Boosting model was also used by Boaz et al [32] to correct the TIS in optical overlay measurement, which is normally caused by the tool's interaction with target asymmetries, lens alignment, lens aberrations, and illumination alignment. The model developed in that work has reduced the mean of TIS for DRAM and NAND layers by 40% and 81%, respectively. Along with imaged-based overlay metrology, machine learning was also applied to improve the performance of the DBO measurement method. In [33], an ANN with Bayesian regularization model has been proposed to decrease the measurement inaccuracy caused by the sidewalls effect in the diffraction based overlay bottom gratings targets. Using the dose level of 500 mJ/s·cm² for each pupil image, the overlay MSE of the proposed model reduced 42.22% compared to the conventional DBO model.

C. Control Scheme to Improve Overlay in Manufacturing Process

To ensure high yield for the production line, the development of advanced overlay control scheme is very important. Recently, machine learning models have been integrated in various process control steps to improve the wafers' overlay accuracy.

Run-to-run control is one of the most popular methods used in semiconductor industry, and it has been applied to optimize various process steps such as chemical-mechanical planarization (CMP) [52], chemical - vapor deposition [53], plasma etching reactor [54], and lithography process [55]. In [34], Overcast et al has proposed several metrics with parametric clustering algorithms to identify which non-lithography processes (etch chambers, CMP processes, or high-temperature tools) significantly affect the overlay quality and stability. Once the notable non-lithography contexts have been detected, specific key numbers are calculated and fed to the run-to-run simulation software to check for the improvement of overlay performance. In [37], Kang et al introduced another run-to-run control scheme which embeds with a regression VOM model to reduce overlay error of the photolithography process. In this control scheme, the VM models were first developed by experimenting different ML regression models to find the optimal one for each measurement position on wafer. Then, the output of these models will be fed into the EWMA controller, which adjusts the controllable input parameters so that the prediction matches the measurement. The authors have verified this run-to-run method in a Monte Carlo simulation, and the results

showed a significant improvement compared to the system without the run-to-run controller.

As pointed out by Khakifirooz et al [36], the lack of real-time metrology data is a major limitation of the run-to-run controller. Therefore, the authors have developed a dynamic control scheme which is able to compensate the overlay error in real time. In that study, the ϵ -SVR optimization technique was used as the kernel for the online controller. ϵ -SVR is a strong regression technique which is proved to be able to find the global minimum in the optimization problem and avoid overfitting. The proposed model made use of 10 variables including intra-field and inter-field overlay errors, as well as considered stochastic time delay of the process, which has successfully outperformed the conventional EWMA controller.

D. Challenges and Future Work

As stated in [8], there are more than 1000 process steps to manufacture some leading-edge devices from a silicon wafer. In that process, a huge amount of data is generated by sensors from different tools. Thus, the biggest challenge in the field of machine learning based overlay analysis and optimization is to find which steps in the production line contribute the most to the overlay errors of a wafer. In other words, for the best performance of machine learning models, most relevant and significant features to train are crucial. Therefore, in the future, more advanced data analysis techniques need to be developed to identify which process steps have the most substantial impact on the overlay performance. From that, most meaningful features can be extracted from the data generated in those steps.

In addition, most of the reviewed studies above only considered the lithography-based processes in analyzing or optimizing overlay, while non-lithography-based processes also considerably contribute to the overlay performance of a wafer [34]. Hence, more work should be done to improve overlay outcome by combining data from both the lithography-based and non-lithography-based process steps.

Finally, many studies have trained their machine learning models with small datasets. This greatly limit the prediction accuracy of the models. For future work, machine learning models should be trained with larger datasets in order to improve their performance.

V. DISCUSSION: MACHINE LEARNING BASED MASK OPTIMIZATION TECHNIQUES

With the rapid growth of machine learning field and the abundance of available data, various advanced models have been introduced to optimize the mask synthesis computation task in the semiconductor manufacturing process. Table III shows the summary results of 17 different studies about mask optimization techniques using machine learning.

A. Sub-Resolution Assist Feature Insertion (SRAF)

Among mask optimization techniques, SRAF is one of the most effective approaches to enhance the through-process robustness of exposing masks in lithography process [56].

TABLE III
SUMMARY OF MACHINE LEARNING BASED MASK OPTIMIZATION STUDIES

Studies	Objectives	Dataset/ Input Features	ML models	Results and Key Findings	Limitations
[18] (Cited by 1)	Propose a ML based model for etch bias prediction, and using the output of that model to introduce an ILT method for mask optimization to minimize EPE	- For etch bias model: micro-loading, macro-loading, visible area, blocked area, ILS, stress terms, other ADI model signal properties. - For ILT method: AEI target and the compact AEI model proposed above. (The dataset is obtained at nominal process condition.)	- For etch bias prediction model: ANN regression - For ILT method: N/A	- Etch bias prediction model performance: + RMS Error: 0.86 nm (50% improvement over conventional model) + Range: 7.9 nm (40% improvement over conventional model) - ILT etch correction method can output smooth contour which helps to improve EPE, and it converge quite fast in only 3 iterations. + Average EPE: 0.03 nm + Worse case EPE: < 0.2 nm (The results are obtained from modelling and simulation work.)	For the ML-based ILT method, the performance in term of EPE was quite similar to the conventional method without significant improvement.
[19] (Cited by 3)	Propose a ML based OPC model	- PFT signals - Initial EPE (The dataset is obtained at nominal process condition.)	RF regression	- The proposed ML-OPC model using RFR improve the RMS error by 54% compared to the state-of-the-art OPC model using MLP. - Segment grouping improved the RMS error by 15%. - Using the proposed ML-OPC model as an initial guess in combination with the MB-OPC model, the iteration time reduced by 56%. (The results are obtained from modelling and simulation work.)	Initial guess for EPE has a great impact on the performance of the model. If a bad guess is chosen, the performance will be degraded.
[20] (Cited by 14)	Propose a fast OPC approach which is capable of generating manufacture-friendly mask	Vector of 57 sampling points for each mask feature. (The dataset is obtained at nominal process condition.)	NKR	- The proposed algorithm reduced the average [EPE] by 62% compared to the initial target pattern. - The computational time of the proposed method is two-fold faster than the PBOPC software - The mask manufacturability is 47% and 27% improved for Metal and Poly layers in comparison with the PBOPC output. (The results are obtained from modelling and simulation work.)	- The choice of observation points density is empirical and greatly affect the performance of the algorithm. - The performance of the proposed method is still noticeably worse than the commercial PBOPC software.
[21] (Cited by 29)	Propose a SRAF generation technique	Multi-channel images with three different channels: Target patterns, horizontal SRAFs, and vertical SRAFs. (The dataset is obtained at both nominal condition and process sensitivity study.)	CGANs	- CGAN could reduce the EPE and PV band of the mask without SRAF by 86.23% and 13.06%, respectively. - The runtime of CGAN model is ~14.6x faster than the LS_SVM model proposed in [38] and ~144x faster than the model-based approach by Calibre software. (The results are obtained from modelling and simulation work.)	The authors used outputs from the model-based technique of Calibre software as training data, which means that the performance of the proposed model cannot exceed the performance of that software, in terms of EPE as the criteria.
[38] (Cited by 20)	Propose a SRAF generation technique	SRAF labels and feature vectors. (The dataset is obtained at both nominal condition and process sensitivity study.)	- LGR - SVC	- For one training set, the proposed LGR and SVC models reduced the PV band by 13.72% and 14.08%, respectively, compared to the mask without SRAF. The EPEs are also reduced by 86.82% and 86.41% by LGR and SVC. - Regarding the runtime, proposed method produces the result 10 times faster than the model-based approach from Calibra software. (The results are obtained from modelling and simulation work.)	The authors used outputs from the model-based technique of Calibre software as training data, which means that the performance of the proposed model cannot exceed the performance of that software, in terms of EPE as the criteria.
[39] (Cited by 21)	Propose a SRAF insertion technique	- CCAS feature in the features extraction stage. - Low-dimension features of grids obtained from last stage. (The dataset is obtained at both nominal condition and	- SODL algorithm for features extraction stage - ILP	- Compared to the ISPD'16 model in [40], EPE and PV band of the new proposed method improved 3.5% and 11.8%, respectively. - In term of SRAF's total area on mask, the proposed method's SRAF area is only 79.1% of the SRAF area generated by ISPD'16 model.	The run time of this proposed method is 1.572 times higher than the run time of the comparing method in [40].

		<i>process sensitivity study.</i>)		<i>(The results are obtained from modelling and simulation work.)</i>	
[40] (Cited by 34)	Propose a machine learning framework for the SRAF generation	- Feature vector represents the optical conditions of the grid point with respect to the target patterns. - SRAF label for each grid. (The dataset is obtained at both nominal condition and process sensitivity study.)	- Decision Tree - Logistic Regression	The performance of the complete mask generated by the proposed framework is comparable with the commercial model-based framework (even better in EPE criterion), but the speed is 10 times faster. (The results are obtained from modelling and simulation work.)	The PV band performance of the proposed method is degraded compared to the model-based method, which should be improved in the future.
[41] (Cited by 3)	Extract the contour of precise EPE metrology on ADI wafer regardless the e-beam scanning direction to the pattern edge	- SEM images taken by 0°/45°/90° scan of 4 blocks which contain 25 repeating pattern each on IMEC Suez wafer. - Feature vectors based Local contrast and Local Gray Level Ratio. (The dataset is obtained at nominal process condition.)	ERTs	Differences of the contour extracted by proposed method compared to the reference for: - Critical Dimension: 0.40% - Die-to-Database EPE: 11.04% (The results are obtained from modelling and simulation work.)	In this paper, the training edges was synthesized from different directions for different patterns, which is not as accurate as scanning the same pattern from all directions.
[42] (Cited by 58)	Propose a HBM based OPC model	- Training dataset: All edge types + 5000 features vector extracted from layout A using CCAS. - Testing dataset: 5000 features vector extracted from layout B using CCAS. (The dataset is obtained at nominal process condition.)	HBM	- HBM resolved the overfitting issue occurred in LGR and SVC. - For the testing layout, HBM model outperformed LGR and SVC by 5% and 112% in terms of RMSEs for the extracted samples - HBM's performance is comparable with the output from the model-based method used in Calibre software with 10 iterations. (The results are obtained from modelling and simulation work.)	Small number of train and test layout: only one layout was used for training and one layout was used for testing.
[43] (Cited by 91)	Propose a GAN based OPC model	A training layout library of 4000 instance was generated based on 32nm M1 layout topologies specification. (The dataset is obtained at both nominal condition and process sensitivity study.)	GAN	- Compared to the ILT model in [45], L2 error of the proposed GAN model decreased by 9.2%, while the contour area variations decrease around 1%. - Mask optimization time of GAN based method is more than 2x faster compared to ILT model in [45]. (The results are obtained from modelling and simulation work.)	- Hard to achieve good convergence for training. - The performance is not consistent among all the test cases.
[44] (Cited by 9)	Propose a mask synthesis framework which consists of 3D mask model, imaging model, resist model, and mask synthesis model	Target images and contours. (The dataset is obtained at both nominal condition and process sensitivity study.)	Reinforcement Learning	- Produced mask that has good pattern fidelity. - Good prediction accuracy of 3D mask and imaging models. - ML mask synthesis framework showed good symmetry and consistency properties. (The results are obtained from modelling and simulation work.)	The authors tried different loss functions to optimize different individual quality metrics which are pattern fidelity and image contrast. Loss function which optimizes for more than one metrics can be the way to improve the performance of this framework in the future.
[45] (Cited by 52)	Mask optimization approach	- Run time - PV band - Image difference - EPE (The dataset is obtained at both nominal condition and process sensitivity study.)	Gradient Descent	The proposed method can generate mask which outperformed model from the first-place winner of the ICCAD 2013 contest by 7% and 11%, respectively. (The evaluation is based on number of EPE violation and PV band). (The results are obtained from modelling and simulation work.)	Sub-resolution assist features in this model are generated using rule-based method, which need to be improved in the future.
[46] (Cited by 16)	Propose a fast pixel-based optical proximity correction model	Fragments from the desired pattern: - Convex corners - Concave corners - Edges (The dataset is obtained at nominal process condition.)	NKR	- Using the proposed method, the average EPE of a 90-nm metal layer decreased 81%, and the pattern error decreased 92%, compared to the initial mask. - The runtime of the proposed method is 48% and 61% less than the Calibre pxOPC software for the 90- and 45-nm metal layers, respectively. (The results are obtained from modelling and simulation work.)	Only nominal condition was considered in this work, process variation should be examined as the future work.
[47] (Cited)	Evaluate the performance of ILT	Defocus and dose process variation.	ANN with Adam	- Under the nominal exposure condition, model train by Adam can achieve EPE of 2.3 nm, 30%	Information about the amount of data used for

by 0)	by applying Adam algorithm	(The dataset is obtained at both nominal condition and process sensitivity study.)	optimizer	smaller than model train by SGD. - Adam has a larger common process window than SGD method. (The results are obtained from modelling and simulation work.)	the training and validating of the model is not clearly specify.
[48] (Cited by 3)	Improve OPC model performance by integrating new metrology method and deep learning prediction model	Critical dimension and edge placement gauges from several different patterns. (The dataset is obtained at nominal process condition.)	CNN	- Using proposed MXP metrology with conventional FEM+ model showed ~20% accuracy improvement - Using MXP metrology with proposed Neuron model gain ~30% accuracy improvement. This method also improved wafer critical dimension prediction 4 to 8 nm on 2D patterns. (The performance of ML model is verified by on-wafer measurements)	The way MXP extract the contour is not so clear.
[49] (Cited by 0)	Propose a new OPC model based RNN network	- PFT signals - Segment type and direction - Location difference (The dataset is obtained at nominal process condition.)	RNN with Attention Mechanism	The EPE measure with M1 layer in 28 nm node of the proposed RNN-OPC is 36% smaller than the EPE of the ML-OPC model in [57]. (The results are obtained from modelling and simulation work.)	The runtime of this RNN-OPC is much longer than the compared MLP-OPC method.
[50] (Cited by 2)	Propose ML based models for etch bias prediction	ADI and AEI SEM Images, layouts, coordinates. (The dataset is obtained at nominal process condition.)	Comparing 3 methods: - SNN - ENN - RF	- Compared to the conventional VEB method, all three proposed ML methods (SNN, ENN, and RF) showed better performance. - Predicted Etch bias RMSE of SNN, ENN, and RF models improved 8.34%, 15.01%, 12.69% respectively over the VEB method. - Training and predicting time of the proposed methods are also faster than the VEB method. (The results are obtained from both simulation work and on-wafer measurements)	The edge fragments used to measure the etch bias did not cover the rounded corner, which leads to the information loss.

Abbreviations:

ADI (After Develop Inspection), AEI (After Etch Inspection), CCAS (Concentric Circle Area Sampling), CGANs (Conditional Generative Adversarial Networks), CNN (Convolutional Neural Network), ENN (ensemble Fully-connected Neural Networks), ERTs (Extremely Randomized Trees), HBM (Hierarchical Bayes Model), ILP (Integer Linear Programming), ILS (Image Log Slope), ILT (Inverse Lithography Technology), LGR (Logistic Regression), MXP (Metrology of Extreme Performance), NKR (Nonparametric Kernel Regression), RF (Random Forest), RNN (Recurrent Neural Network), SNN (Single Fully-connected Neural Network), SODL (Supervised Online Dictionary Learning), SVC (Support Vector Classification)

SRAFs are small features placed around the target patterns (Fig. 4) in order to create a region with high features density, which can improve the depth-of-focus of those desired patterns [58]. As a result, the resolution of the printed patterns when using mask assisted SRAFs is improved considerably compared to the masks without SRAFs.

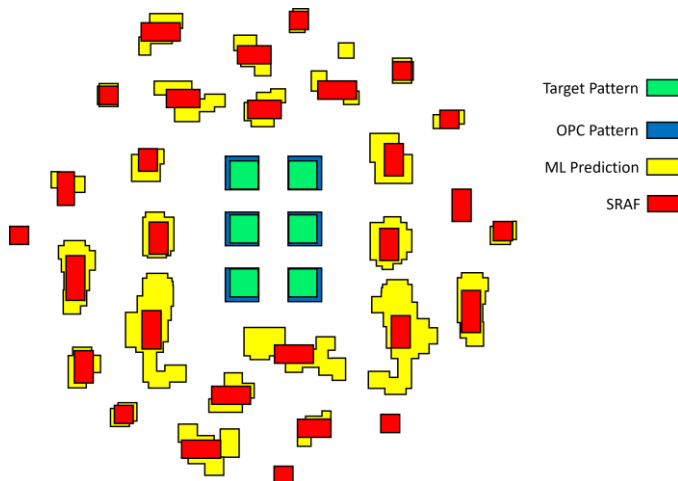


Fig. 4. Optimized lithographic masks with SRAFs and OPC patterns [40].

Most of the reviewed studies in this survey used traditional ML models such as Logistic Regression (LGR) [38], [40], Support Vector Classification (SVC) [38], Supervised Online Dictionary Learning (SODL) [39], or Decision Tree [40] for the insertion of SRAFs. In those studies, each layout clip of training and testing patterns is divided into small pixels, and these pixels' feature vectors are sampled by using the Concentric Circle Area Sampling (CCAS) scheme. Then, all the pixels are classified using trained ML models to predict whether a pixel should be inserted as SRAF or not. The results in [38] indicate that the SVC-based model performed approximately 5% better than LGR-based model for EPE, while in [40], X. Xu et al stated that LGR has better performance compared to Decision Tree. In [39], Geng et al showed that EPE and PV band of the SODL-based model improved 3.5% and 11.8%, respectively, compared to the LGR-based model in [40]. In terms of the runtime, machine learning based models in those studies have achieved up to 10x speed increase compared to the model-based technique used in commercial software tools.

Besides traditional ML algorithms, modern Deep Learning algorithm has also been used for SRAFs insertion. In [21], Alawieh et al introduced a SRAFs insertion framework which used Conditional Generative Adversarial Networks (CGANs) to generate SRAF features. Using CGAN, the model can be trained to translate images of original layout domain to the

domain which has layout with SRAFs. The experimental results have shown that the CGAN-based model could achieve similar performance of the SVC-based model proposed in [38], but with around 14.6x speed increase in the runtime.

B. Optical Proximity Correction

Light diffraction and interference are the main contributors to the distortion errors like rounded corners or shortened line-ends in photolithography process, especially in advanced technology nodes, where the critical dimensions are in the range of only several nanometers. To compensate for those errors and minimize the EPE, OPC is widely applied. Fig. 5 demonstrates the role of OPC in photolithography process.

In previous studies, several traditional ML models such as Nonparametric Kernel Regression [20], [46], Random Forest Regression [19], and Gradient Descent [45] have been introduced for OPC optimization. These models first fragment the edges of target patterns, then extract the features from the fragmented layout. Because the diffraction effects are extremely complicated toward the sub-resolution domain, the traditional ML models have immense complexity and often suffer from the overfitting issue [39], [59]. To overcome the overfitting problem, a hierarchical Bayes model (HBM) with CCAS sampling scheme has been proposed by Matsunawa et al [42]. In that work, a generalized linear mixed model was trained while considering four different edge types, which are normal, convex, concave and line-end edge. The results from HBM model can be comparable to a 10-iteration conventional model-based method.

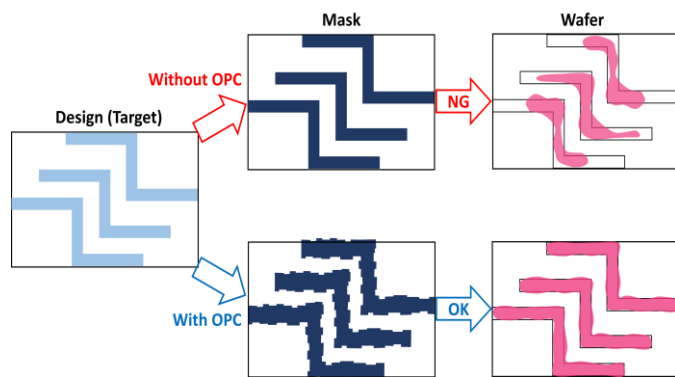


Fig. 5. The role of OPC in photolithography process [42].

Recently, supervised deep learning techniques, including ANN [18], [47], CNN [48], RNN [49], ENN [50], and GAN [43] have been implemented with promising results. However, the limitation of supervised learning models is that they need a training dataset generated from other mask generation tools for training. Thus, their performance can hardly surpass the available tools' output. One solution for this was proposed by Peng Liu in [44], where Reinforcement Learning with ANN was used for mask synthesis. This approach directly optimizes the mask based on the user's desired quality metrics with the rewarding-punishing mechanism. The proof-of-concept evaluation provided by the author has shown that Reinforcement Learning for mask optimization is a potential direction that can be explored in the future.

C. Challenges and Future Work

Despite a great number of successful applications of ML techniques on mask synthesis and optimization, there are still some difficult challenges. For example, in SRAF insertion, each pixel in a layout clip is classified one by one. Since one clip normally has the width and height in micrometer scale, while the precision of SRAF features needs to be in the nanometer region, the number of pixels required can be enormous. This eventually becomes very computationally expensive, especially with complex ML models. One approach to deal with this issue in SRAF generation is to adapt end-to-end methods with an advance deep learning-based architecture such as GAN. GAN models can directly map the images of target patterns to that of the mask patterns, hence effectively reducing the computational power without compromising much accuracy.

The challenge of OPC optimization lies in the high complexity of diffraction and process effects, especially when the fabricated patterns are increasingly denser, while the sizes become smaller and smaller. Even the state-of-the-art models based OPC techniques are yet to be optimal to capture the physical phenomena. As most of the previous ML-based OPC models rely on the output from model-based OPC for training, this becomes a performance bottleneck for the further improvement of OPC techniques. Model-based OPC is heavily relied on lithography simulators which solve Maxwell's equations by using numerical methods such as finite differences and finite elements. Although these methods have proven to be generally accurate, they are very computationally expensive. Besides, their obtained solutions are discrete or have limited differentiability [60]. Furthermore, choosing optimal boundary conditions and fine-tuning the geometry for these discretizing approaches heavily rely on past experience of the design templates, and due to the constraints on computational power and time, only a small number of parameters are adjusted in order to find the desired responses [61]. Physics-informed ML [62] approaches can be a prospective direction to transcend the aforementioned limitations. Physics-informed ML is a type of ML methods which embed the physical laws governing the given dataset with a ML model. This ML method overcomes the low data availability problem and is very effective in solving ill-posed and inverse problems compared to conventional mesh-based solvers. Fig. 6 shows an illustration of a physics-informed ML algorithm. Here in this figure, the left part represents the conventional ML network, whose output is the surrogate solution of the physical model, whereas the right part contains the mathematical function of the physics prior. The loss function is the sum of supervised loss of the conventional ML network and the unsupervised loss of the physical model. The network is trained until the loss is smaller than a pre-determined/desired threshold ϵ_0 .

Physics-informed ML has been applied in various fields, such as geophysics [63], [64], molecular simulations [65], material sciences [66], and quantum chemistry [67]. In the domain of electromagnetism, Ref [68], and Ref [69] have successfully built wave-propagating simulators by leveraging

physics-informed ML models, which integrate Maxwell's equations with neural networks. Applying these simulators in developing OPC and inverse lithography models is expected to be able to improve the performance of mask generation tasks in the future.

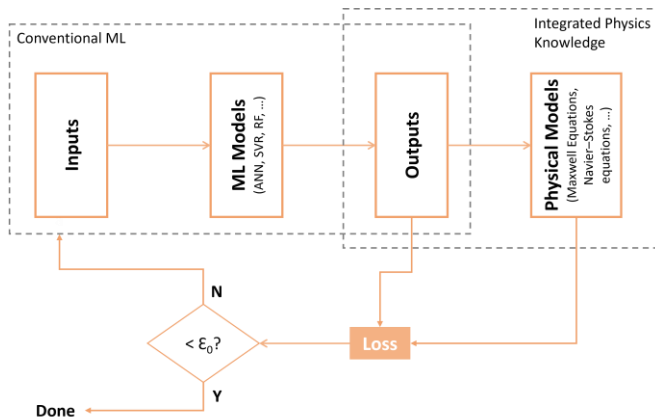


Fig. 6. Illustration of a physics-informed ML algorithm [62].

VI. CONCLUSION

In this paper, the recent advances of machine learning applications in EPE analysis and optimization problems in semiconductor manufacturing have been reviewed. ML models have proved to be beneficial in a number of tasks, including virtual overlay metrology, overlay metrology's accuracy improvement, overlay control scheme, SRAFs insertion, and OPC. Currently, there are a great number of ongoing studies with the effort of pushing further the effectiveness of Machine Learning models to reduce EPE and enable higher yield for the manufacturing process. This area of research will significantly contribute to the roadmap toward 3nm node of the semiconductor industry.

Nevertheless, applications of ML techniques in EPE analysis and optimization are still in the very early stage of development. There are a variety of challenges that are yet to be overcome, which offer a considerable scope for future research work.

REFERENCES

- [1] D. Z. Pan, L. Liebmann, B. Yu, X. Xu, and Y. Lin, "Pushing multiple patterning in sub-10nm: Are we ready?," in *2015 52nd ACM/EDAC/IEEE Design Automation Conference (DAC)*, 2015, pp. 1–6. doi: 10.1145/2744769.2747940.
- [2] M. A. van de Kerkhof, J. P. H. Benschop, and V. Y. Banine, "Lithography for now and the future," *Solid. State. Electron.*, vol. 155, pp. 20–26, 2019, doi: https://doi.org/10.1016/j.sse.2019.03.006.
- [3] Y. Jiao and D. Djurdjanovic, "Stochastic Control of Multilayer Overlay in Lithography Processes," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 3, pp. 404–417, 2011, doi: 10.1109/TSM.2011.2142329.
- [4] H.-F. Kuo and A. Faricha, "Artificial neural network for diffraction based overlay measurement," *IEEE Access*, vol. 4, pp. 7479–7486, 2016, doi: 10.1109/ACCESS.2016.2618350.
- [5] K. Bhattacharyya, "Tough road ahead for device overlay and edge placement error," in *Metrology, Inspection, and Process Control for Microlithography XXXIII*, 2019, vol. 10959, pp. 1–8. doi: 10.1117/12.2514820.
- [6] J. Mulken et al., "Patterning control strategies for minimum edge placement error in logic devices," in *Proc.SPIE*, Mar. 2017, vol. 10145. doi: 10.1117/12.2260155.
- [7] J. Mulken et al., "Holistic approach for overlay and edge placement error to meet the 5nm technology node requirements," in *Metrology, Inspection, and Process Control for Microlithography XXXII*, 2018, vol. 10585, pp. 375–388. doi: 10.1117/12.2297283.
- [8] A. Verner et al., "Optical overlay measurement accuracy improvement with machine learning," in *Metrology, Inspection, and Process Control for Microlithography XXXIV*, 2020, vol. 11325, p. 113251Z. doi: https://doi.org/10.1117/12.2551850.
- [9] L. van Dijk, K. M. Adal, M. Chastan, A. Lam, and R. van Haren, "Excursion detection and root-cause analysis using virtual overlay metrology," in *Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV*, 2021, vol. 11611, p. 1161132. doi: https://doi.org/10.1117/12.2581561.
- [10] Y. Li, "Selecting training points for one-class support vector machines," *Pattern Recognit. Lett.*, vol. 32, no. 11, pp. 1517–1522, 2011, doi: https://doi.org/10.1016/j.patrec.2011.04.013.
- [11] H.-G. Lee et al., "Virtual overlay metrology for fault detection supported with integrated metrology and machine learning," in *Metrology, Inspection, and Process Control for Microlithography XXIX*, 2015, vol. 9424, p. 94241T. doi: https://doi.org/10.1117/12.2085475.
- [12] J.-S. Park et al., "An efficient rule-based OPC approach using a DRC tool for 0.18 $\mu\text{m}/\text{m}$ ASIC," in *Proceedings IEEE 2000 First International Symposium on Quality Electronic Design (Cat. No. PR00525)*, 2000, pp. 81–85. doi: 10.1109/ISQED.2000.838858.
- [13] J. Wang, A. Wei, P. Verma, and W. Wilkinson, "Optimization of rule-based OPC fragmentation to improve wafer image rippling," in *31st European Mask and Lithography Conference*, 2015, vol. 9661, pp. 79–94. doi: 10.1117/12.2194755.
- [14] Y. Li, S.-M. Yu, and Y.-L. Li, "Intelligent optical proximity correction using genetic algorithm with model- and rule-based approaches," *Comput. Mater. Sci.*, vol. 45, no. 1, pp. 65–76, 2009, doi: https://doi.org/10.1016/j.commatsci.2008.04.031.
- [15] M. Harb and H. Abdelghany, "Hybrid OPC technique using model based and rule-based flows," in *Proc.SPIE*, Apr. 2013, vol. 8683. doi: 10.1117/12.2011497.
- [16] M. Lam, C. Clifford, A. Raghunathan, G. Fenger, and K. Adam, "Enabling full field physics based OPC via dynamic model generation," in *Proc.SPIE*, Mar. 2017, vol. 10143. doi: 10.1117/12.2261222.
- [17] P. Fanton et al., "Resist 3D model based OPC for 28nm metal process window enlargement," in *Proc.SPIE*, Mar. 2016, vol. 9778. doi: 10.1117/12.2218916.
- [18] K. Hooker, L. Zavyalova, S. Huang, and L.-J. Chen, "Using machine learning etch models in OPC and ILT correction," in *Design-Process-Technology Co-optimization XV*, 2021, vol. 11614, p. 116140B. doi: https://doi.org/10.1117/12.2587225.
- [19] J. Cho, G. Cho, and Y. Shin, "Optimization of machine learning guided optical proximity correction," in *2018 IEEE 61st International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2018, pp. 921–924. doi: 10.1109/MWSCAS.2018.8623985.
- [20] X. Ma, S. Jiang, J. Wang, B. Wu, Z. Song, and Y. Li, "A fast and manufacture-friendly optical proximity correction based on machine learning," *Microelectron. Eng.*, vol. 168, pp. 15–26, 2017, doi: https://doi.org/10.1016/j.mee.2016.10.006.
- [21] M. B. Alawieh, Y. Lin, Z. Zhang, M. Li, Q. Huang, and D. Z. Pan, "GAN-SRAF: Sub-resolution assist feature generation using conditional generative adversarial networks," in *Proceedings of the 56th Annual Design Automation Conference 2019*, 2019, pp. 1–6.
- [22] S. Morton, A. Berg, L. Levit, J. Eden, and others, *Finding what works in health care: standards for systematic reviews*. Washington (DC): National Academies Press, 2011.
- [23] E. Schmitt-Weaver et al., "Overlay improvements using a real time machine learning algorithm," in *Metrology, Inspection, and Process Control for Microlithography XXVIII*, 2014, vol. 9050, pp. 496–502. doi: https://doi.org/10.1117/12.2046914.
- [24] L. van Dijk, F. Hasibi, M. Larrañaga, A. Pastol, A. Lam, and R. van Haren, "Smart implant-layer overlay metrology to enable fab cycle time reduction," in *Metrology, Inspection, and Process Control for Microlithography XXXIII*, 2019, vol. 10959, pp. 298–310. doi: https://doi.org/10.1117/12.2515185.
- [25] M. Meng, L. Tu, J. Mi, H. Zhou, and X. Zou, "Machine learning and hybrid metrology using HV-SEM and optical methods to

- monitor channel hole tilting in-line for 3D NAND wafer production," in *Metrology, Inspection, and Process Control for Microlithography XXXIV*, 2020, vol. 11325, p. 1132500. doi: <https://doi.org/10.1117/12.2551622>.
- [26] F. Hasibi, L. van Dijk, M. Larrañaga, A. Pastol, A. Lam, and R. van Haren, "Towards fab cycle time reduction by machine learning-based overlay metrology," in *34th European Mask and Lithography Conference*, 2018, vol. 10775, p. 107750X. doi: <https://doi.org/10.1117/12.2500239>.
- [27] H. Chen *et al.*, "Real time process monitoring using diffraction-based overlay measurements from YieldStar," in *2020 International Workshop on Advanced Patterning Solutions (IWAPS)*, 2020, pp. 1–4. doi: 10.1109/IWAPS51164.2020.9286803.
- [28] C. Hwang *et al.*, "Smart overlay metrology pairing adaptive deep learning with the physics-based models used by a lithographic apparatus," in *Optical Microlithography XXXI*, 2018, vol. 10587, p. 105870B. doi: <https://doi.org/10.1117/12.2297513>.
- [29] E. Schmitt-Weaver and K. Bhattacharyya, "Pairing wafer leveling metrology from a lithographic apparatus with deep learning to enable cost effective dense wafer alignment metrology," in *Optical Microlithography XXXII*, 2019, vol. 10961, pp. 35–40. doi: <https://doi.org/10.1117/12.2514455>.
- [30] E. Schmitt-Weaver *et al.*, "Computational overlay metrology with adaptive data analytics," in *Metrology, Inspection, and Process Control for Microlithography XXXI*, 2017, vol. 10145, p. 101450V. doi: <https://doi.org/10.1117/12.2258039>.
- [31] Y. Feng *et al.*, "Fast in-device overlay metrology on multi-tier 3DNAND devices without DECAP and its applications in process characterization and control," in *Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV*, 2021, vol. 11611, pp. 144–149. doi: <https://doi.org/10.1117/12.2583416>.
- [32] B. Ophir *et al.*, "Machine learning for Tool Induced Shift (TIS) reduction: an HVM case study," in *Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV*, 2021, vol. 11611, p. 116110E. doi: <https://doi.org/10.1117/12.2590247>.
- [33] C.-H. Su, Z.-H. Lin, Y.-S. Lin, and H.-F. Kuo, "Enhancement of Diffraction-Based Overlay Model for Overlay Target With Asymmetric Sidewall," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 3, pp. 373–382, 2020, doi: 10.1109/TSM.2020.3004040.
- [34] M. Overcast *et al.*, "Understanding overlay signatures using machine learning on non-lithography context information," in *Metrology, Inspection, and Process Control for Microlithography XXXII*, 2018, vol. 10585, p. 105851U. doi: <https://doi.org/10.1117/12.2303487>.
- [35] T. Kikuchi, A. Sugimoto, S. Eto, A. Okutomi, and N. Morita, "Novel overlay correction using inline alignment station (iAS) for scanner," in *Optical Microlithography XXXIII*, 2020, vol. 11327, pp. 232–236. doi: <https://doi.org/10.1117/12.2551973>.
- [36] M. Khakifirooz, C.-F. Chien, and Y.-J. Chen, "Dynamic support vector regression control system for overlay error compensation with stochastic metrology delay," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 1, pp. 502–512, 2019, doi: 10.1109/TASE.2019.2935179.
- [37] P. Kang, D. Kim, H. Lee, S. Doh, and S. Cho, "Virtual metrology for run-to-run control in semiconductor manufacturing," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2508–2522, 2011, doi: <https://doi.org/10.1016/j.eswa.2010.08.040>.
- [38] X. Xu *et al.*, "Subresolution assist feature generation with supervised data learning," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 37, no. 6, pp. 1225–1236, 2017, doi: 10.1109/TCAD.2017.2748029.
- [39] H. Geng, W. Zhong, H. Yang, Y. Ma, J. Mitra, and B. Yu, "SRAF insertion via supervised dictionary learning," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 39, no. 10, pp. 2849–2859, 2019, doi: 10.1109/TCAD.2019.2943568.
- [40] X. Xu, T. Matsunawa, S. Nojima, C. Kodama, T. Kotani, and D. Z. Pan, "A machine learning based framework for sub-resolution assist feature generation," in *Proceedings of the 2016 on International Symposium on Physical Design*, 2016, pp. 161–168. doi: <https://doi.org/10.1145/2872334.2872357>.
- [41] Y. Okamoto *et al.*, "Improvement of EPE measurement accuracy on ADI wafer, the method of using machine learning trained with CAD," in *Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV*, 2021, vol. 11611, p. 116111W. doi: <https://doi.org/10.1117/12.2584709>.
- [42] T. Matsunawa, B. Yu, and D. Z. Pan, "Optical proximity correction with hierarchical Bayes model," in *Proc.SPIE*, Mar. 2015, vol. 9426. doi: 10.1117/12.2085787.
- [43] H. Yang, S. Li, Z. Deng, Y. Ma, B. Yu, and E. F. Y. Young, "GAN-OPC: Mask optimization with lithography-guided generative adversarial nets," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 39, no. 10, pp. 2822–2834, 2019, doi: 10.1109/TCAD.2019.2939329.
- [44] P. Liu, "Mask synthesis using machine learning software and hardware platforms," in *Optical Microlithography XXXIII*, 2020, vol. 11327, p. 1132707. doi: <https://doi.org/10.1117/12.2551816>.
- [45] J.-R. Gao, X. Xu, B. Yu, and D. Z. Pan, "MOSAIC: Mask optimizing solution with process window aware inverse correction," in *2014 51st ACM/EDAC/IEEE Design Automation Conference (DAC)*, 2014, pp. 1–6. doi: 10.1109/DAC.2014.6881379.
- [46] X. Ma, B. Wu, Z. Song, S. Jiang, and Y. Li, "Fast pixel-based optical proximity correction based on nonparametric kernel regression," *J. Micro/Nanolithography, MEMS, MOEMS*, vol. 13, no. 4, p. 43007, 2014, doi: <https://doi.org/10.1117/1.JMM.13.4.043007>.
- [47] D. Yu, Y. Liu, and C. Hawkinson, "The application of a new stochastic search algorithm 'Adam' in inverse lithography technology (ILT) in critical recording head fabrication process," in *Optical Microlithography XXXIV*, 2021, vol. 11613, p. 116130N. doi: <https://doi.org/10.1117/12.2583508>.
- [48] W. Yuan *et al.*, "Metrology and deep learning integrated solution to drive OPC model accuracy improvement," in *Optical Microlithography XXXII*, 2019, vol. 10961, p. 109610N. doi: <https://doi.org/10.1117/12.2516236>.
- [49] Y. Kwon and Y. Shin, "Optical Proximity Correction Using Bidirectional Recurrent Neural Network With Attention Mechanism," *IEEE Trans. Semicond. Manuf.*, vol. 34, no. 2, pp. 168–176, 2021, doi: 10.1109/TSM.2021.3072668.
- [50] Y. Meng, Y.-C. Kim, S. Guo, Z. Shu, Y. Zhang, and Q. Liu, "Machine Learning Models for Edge Placement Error Based Etch Bias," *IEEE Trans. Semicond. Manuf.*, vol. 34, no. 1, pp. 42–48, 2020, doi: 10.1109/TSM.2020.3042803.
- [51] A. Y.-L. Chong, "Predicting m-commerce adoption determinants: A neural network approach," *Expert Syst. Appl.*, vol. 40, no. 2, pp. 523–530, 2013, doi: <https://doi.org/10.1016/j.eswa.2012.07.068>.
- [52] J. Yi, Y. Sheng, and C. S. Xu, "Neural network based uniformity profile control of linear chemical-mechanical planarization," *IEEE Trans. Semicond. Manuf.*, vol. 16, no. 4, pp. 609–620, 2003, doi: 10.1109/TSM.2003.818987.
- [53] X. A. Wang and R. L. Mahajan, "Artificial neural network model-based run-to-run process controller," *IEEE Trans. Components, Packag. Manuf. Technol. Part C*, vol. 19, no. 1, pp. 19–26, 1996, doi: 10.1109/3476.484201.
- [54] E. S. Hamby, P. T. Kabamba, and P. P. Khargonekar, "A probabilistic approach to run-to-run control," *IEEE Trans. Semicond. Manuf.*, vol. 11, no. 4, pp. 654–669, 1998, doi: 10.1109/66.728563.
- [55] C. E. Chemali, J. Freudenberg, M. Hankinson, and J. J. Bendik, "Run-to-run critical dimension and sidewall angle lithography control using the PROLITH simulator," *IEEE Trans. Semicond. Manuf.*, vol. 17, no. 3, pp. 388–401, 2004, doi: 10.1109/TSM.2004.831530.
- [56] L. D. Barnes, B. D. Painter, and L. S. M. III, "Model-based placement and optimization of subresolution assist features," in *Proc.SPIE*, Mar. 2006, vol. 6154. doi: 10.1117/12.656691.
- [57] S. Choi, S. Shim, and Y. Shin, "Machine learning (ML)-guided OPC using basis functions of polar Fourier transform," in *Proc.SPIE*, Mar. 2016, vol. 9780. doi: 10.1117/12.2219073.
- [58] J. Fung Chen, T. Laidig, K. E. Wampler, and R. Caldwell, "Optical proximity correction for intermediate-pitch features using sub-resolution scattering bars," *J. Vac. Sci. Technol. B Microelectron. Nanom. Struct. Process. Meas. Phenom.*, vol. 15, no. 6, pp. 2426–2433, Nov. 1997, doi: 10.1116/1.589660.
- [59] Y. Lin, X. Xu, J. Ou, and D. Z. Pan, "Machine learning for mask/wafer hotspot detection and mask synthesis," in *Photomask Technology 2017*, 2017, vol. 10451, p. 104510A. doi: <https://doi.org/10.1117/12.2282943>.
- [60] M. Chiaramonte, M. Kiener, and others, "Solving differential equations using neural networks," 2013. [Online]. Available: [© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See <https://www.ieee.org/publications/rights/index.html> for more information.](https://cs229.stanford.edu/proj2013/ChiaramonteKiener-
</div>
<div data-bbox=)

- SolvingDifferentialEquationsUsingNeuralNetworks.pdf
- [61] W. Ma, Z. Liu, Z. A. Kudyshev, A. Boltasseva, W. Cai, and Y. Liu, "Deep learning for the design of photonic structures," *Nat. Photonics*, vol. 15, no. 2, pp. 77–90, 2021, doi: 10.1038/s41566-020-0685-y.
 - [62] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nat. Rev. Phys.*, vol. 3, no. 6, pp. 422–440, 2021, doi: 10.1038/s42254-021-00314-5.
 - [63] W. Zhu, K. Xu, E. Darve, and G. C. Beroza, "A general approach to seismic inversion with automatic differentiation," *Comput. Geosci.*, vol. 151, p. 104751, 2021, doi: <https://doi.org/10.1016/j.cageo.2021.104751>.
 - [64] D. Li, K. Xu, J. M. Harris, and E. Darve, "Coupled Time-Lapse Full-Waveform Inversion for Subsurface Flow Problems Using Intrusive Automatic Differentiation," *Water Resour. Res.*, vol. 56, no. 8, p. e2019WR027032, 2020, doi: <https://doi.org/10.1029/2019WR027032>.
 - [65] W. Jia *et al.*, "Pushing the Limit of Molecular Dynamics with Ab Initio Accuracy to 100 Million Atoms with Machine Learning," 2020.
 - [66] K. Shukla, P. C. Di Leoni, J. Blackshire, D. Sparkman, and G. E. Karniadakis, "Physics-Informed Neural Network for Ultrasound Nondestructive Quantification of Surface Breaking Cracks," *J. Nondestruct. Eval.*, vol. 39, no. 3, p. 61, 2020, doi: 10.1007/s10921-020-00705-1.
 - [67] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, "Ab initio solution of the many-electron Schrödinger equation with deep neural networks," *Phys. Rev. Res.*, vol. 2, no. 3, p. 33429, Sep. 2020, doi: 10.1103/PhysRevResearch.2.033429.
 - [68] P. Zhang, Y. Hu, Y. Jin, S. Deng, X. Wu, and J. Chen, "A Maxwell's Equations Based Deep Learning Method for Time Domain Electromagnetic Simulations," in *2020 IEEE Texas Symposium on Wireless and Microwave Circuits and Systems (WMCS)*, 2020, pp. 1–4, doi: 10.1109/WMCS49442.2020.9172407.
 - [69] J. Lim and D. Psaltis, "MaxwellNet: Physics-driven deep neural network training based on Maxwell's equations," *APL Photonics*, vol. 7, no. 1, 2022, doi: 10.1063/5.0071616.