PHOTOMASK

BACUS—The international technical group of SPIE dedicated to the advancement of photomask technology.

BACUS N • E • W • S

APRIL 2022
VOLUME 38, ISSUE 4

SPIE Photomask Technology - Invited Paper

Uncertainty Quantification of Machine Learning Models: On Conformal Prediction

Inimfon I. Akpabio and **Serap A. Savari,** Texas A&M University, Mail Stop 3128 TAMU, College Station, TX 77843-3128, USA

ABSTRACT

Prediction intervals which describe the reliability of the predictive performance of machine learning models are important to guide decision making and to improve trust in deep learning and other forms of machine learning and artificial intelligence. Conformal prediction is a relatively recent, increasingly popular, rigorously proven and simple methodology to address this need for both classification and regression problems, and it does not use distributional assumptions like Gaussianity or the Bayesian framework; one new variant combines it with another technique to generate prediction intervals known as quantile regression. We will illustrate the strengths and limitations of different conformal prediction procedures for a regression problem involving line edge roughness (LER) estimation; LER affects semiconductor device performance and the yield of the manufacturing process. Low-dose images from the scanning electron microscope (SEM) are often used for roughness measurements because of relatively small acquisition times and resist shrinkage, but such images are corrupted by noise, blur, edge effects and other instrument errors. We consider prediction intervals for the deep convolutional neural network EDGENet, which was trained on a large dataset of simulated SEM images and directly estimates the edge positions from a SEM rough line image containing an unknown level of Poisson noise.

1. Introduction

The future of semiconductor device fabrication will increasingly depend on data processing, information extraction, and knowledge management.[1-3] Deep learning[4] and other forms of artificial intelligence are predicted to have an increasingly important role in semiconductor metrology and other aspects of manufacturing system performance.[2, 5] However, despite the growing interest in smart manufacturing there are barriers at this time to a complete realization of this vision. In December 2020 the National Science Foundation and the National Institute of Standards and Technology brought together many leaders throughout the manufacturing sector including the Senior Director of Industrial Innovation at Intel to discuss the acceleration of the implementation of artificial intelligence in manufacturing. The ensuing report describes real and perceived risks and a lack of transparency as two of the obstacles to the adoption of artificial intelligence.[6] Furthermore, Sections 5.8.4 and 5.8.6 of the 2020 International Roadmap for Devices and Systems (IRDS) report on Factory Integration discuss prediction engines and the need for them to incorporate indications of the quality of predictions.[2] One way to address both sets of concerns is by quantifying the uncertainty of the predictions obtained through machine learning models by means of prediction intervals which describe the reliability of predictive performance. Machine learning models can address two broad classes of problems, namely, regression problems or classification problems. In the case of a regression model estimating a single number a prediction interval specifies a range of values in which the output variable lies with high probability; while our focus in this paper is on regression, prediction intervals can also be defined for classification problems.[7]

There is a large literature on the design of prediction intervals. The most useful approaches for a particular application must satisfy certain criteria. First, they should be founded on assumptions which are well-suited to the application in order to offer valid coverage. Second, they should generate the shortest possible intervals to best guide decision making. Finally, they should have an acceptable computational complexity [2, p. 42].

Conformal prediction[7-13] is a comparatively recent and straightforward methodology with a mathematical foundation for constructing prediction intervals for machine learning models which has been attracting increasing attention. The underlying assumption of conformal prediction is a concept known as exchangeability, which roughly means that the past is representative of the future and the ordering of examples does not matter. This condition holds when all examples being considered are either sampled independently from the

TAKE A LOOK INSIDE:

INDUSTRY BRIEFS
—see page 10

CALENDAR
For a list of meetings
—see page 11

SPIE.

11855-11

EDITORIAL

Leveraging the renaissance

Patrick Naulleau, Center for X-ray Optics, Lawrence Berkeley National Laboratory

After a couple decades of the apparent commoditization of chip manufacturing, we appear to have entered a renaissance where the broader community is again recognizing the critical nature of advanced semiconductor manufacturing and the paramount value of technical leadership driven by fundamental science. This transition is evidenced by the various government initiatives around the globe including the CHIPS Act in the U.S. Although many of these initiatives were kicked off before the pandemic, the past two years have further driven home the critical importance of chip manufacturing by clearly bringing to light our society's ubiquitous dependence on semiconductors.

A key opportunity I see from this renewed public interest in advanced chip manufacturing is in building out the workforce of the future. Future innovation in this space depends on society allocating the brightest minds to this problem, and that starts with our students. Yet all of us serving on conference committees in this field are intimately aware of the challenges we face in finding graduate students dedicating their formative years to the field. We as a community should take this renaissance of awareness in semiconductor manufacturing as an opportunity to build out a robust academic network that will stand the test of time, educating and training our future workforce.

Hopefully, government funded research programs enabled by the new initiatives will have a significant impact on the academic network, but we cannot rely on the government alone for this role. Industry should also be instrumental in driving high-risk academic research, recognizing that the primary value of such engagements may not in fact be immediate deliverables, but rather in the fostering of a vibrant innovative ecosystem that will draw the best and the brightest to the field. Moreover, these responsibilities should not be confined to just those companies directly operating in the semiconductor manufacturing, materials, and tooling space, but they should also be shared by industry further up the value chain which is critically dependent on continued innovation in advanced chip manufacturing.

The ecosystem we strive to strengthen must include both students as well as young professors interested in building academic careers in this space. After all, at the root of addressing the lack of students is ensuring that we have a healthy supply of professors and academic programs to attract those students. Our industry should take this golden opportunity to re-invigorate the academic ecosystem in semiconductor manufacturing. In practice, what this means is being willing to fund open, exploratory, and long-term research. It also means outreach, engagement, and mentorship; academics strongly value collaboration and are drawn to challenging high-impact problems. Such problems for sure abound in this field, but we need to do a better job of making the academic community aware of that and leveraging their expertise.

This is clearly an exciting time to be in the field, so let's spread the word!

BACUS

N • E • W • S

BACUS News is published monthly by SPIE for *BACUS*, the international technical group of SPIE dedicated to the advancement of photomask technology.

Managing Editor/Graphics Linda DeLano
SPIE Sales Representative, Exhibitions, and Sponsorships Melissa Valum
BACUS Technical Group Manager Tim Lamkins

SPIE.

P.O. Box 10, Bellingham, WA 98227-0010 USA
Tel: +1 360 676 3290
Fax: +1 360 647 1445
SPIE.org
help@spie.org

same fixed but unknown probability distribution or satisfy some weaker conditions; see, e.g., [8, §3] for a detailed discussion of exchangeability. The circumstances to which conformal prediction can be applied are fairly general as it does not use distributional assumptions like Gaussianity or the Bayesian framework. We focus here on the *split* or *inductive* version of conformal prediction, which begins by fitting a regression model to training data and subsequently applies a devised "nonconformity score" on the strangeness of examples in a separate calibration set to specify reliable confidence levels in future test examples; the original transductive conformal prediction is a slower approach to finding prediction intervals. The effectiveness of conformal prediction is contingent on the underlying probability distribution and the nonconformity score, and there is often a trade-off between the width of prediction intervals and the computational cost of producing them. Another new development in this approach to uncertainty quantification shows that it is sometimes possible to combine the coverage guarantees of conformal prediction with an earlier and widely used methodology to constructing prediction intervals known as quantile regression[14-17] to obtain more efficient prediction regions.

We will illustrate the strengths and limitations of different conformal prediction procedures for a regression problem involving line edge roughness (LER) estimation. The measurement of LER is necessary to understand and control semiconductor device performance and the yield of the manufacturing process (see, e.g., [18, pp. 82-92]). The critical-dimension scanning electron microscope (CD-SEM) is a standard tool for roughness measurements. It is ideal to work with low-dose SEM images because they reduce sample damage and acquisition time (see, e.g., Ref. 19). However, to determine edge geometry from such images requires techniques to account for their noise, blur, edge effects and other instrument errors, and there is a large literature on this subject with multiple approaches to the problem. Our goal in this paper is to investigate prediction intervals for a simple LER estimation strategy using a simulated dataset so that we know the ground truth. In References 20 and 21 our group proposed a deep convolutional neural network named EDGENet which inputs a 64 x 1024 SEM image containing one rough line which is corrupted by an unknown level of Poisson noise and directly outputs a matrix of dimension 2 x 1024 with the estimated left and right edge positions of the line; we will focus here on EDGENet and the simulated dataset used to study it.

In Section 2, we provide an overview of five methods to construct prediction intervals from simulated images. Our difficulties with a simplistic approach to applying quantile regression to the relatively large raw data in an image motivate a study of that methodology in an idealized situation where we reduce the input to a few key parameters about the image. In Section 3, we review the architecture of EDGENet and details about our dataset. In Section 4, we offer more information about the neural networks we design as tools in constructing prediction intervals as well as the results on coverage and efficiency; we will see that domain knowledge is important in exploiting the benefits of neural networks. In Section 5, we conclude the paper.

2. On Conformal Prediction and Conformalized Quantile Regression

Supervised machine learning is a branch of artificial intelligence which fits parametric functions to collections of input/output data pairs $z^i = (x^i; y^i)$. For the type of regression problems we consider, suppose x^i is the input vector of dimension d of instance z^i and y^i is output scalar of z^i. Machine learning algorithms try to estimate the relationship between the input and output based on an underlying model in the form of a parametric function, some loss criterion, and some approach to exactly or approximately solve the corresponding optimization problem. In many cases, once the model is fit the machine learning algorithm will offer a point prediction of the output for a new input example. The objective of predictive inference is to construct well-calibrated prediction bands with finite-sample coverage only assuming that the training instances $z^1; z^2; : : : ; z^n$ are exchangeable. In particular, if $\alpha \in (0, 1)$ is a prespecified miscoverage rate, then based on our training examples for a new test input X^{n+1} we wish to find a *marginal* prediction interval $C(X^{n+1})$ satisfying

$$P[Y^{n+1} \in C(X^{n+1})] \geq 1 - \alpha. \tag{1}$$

The term "marginal" means that the probability is over all random instances $Z^1; Z^2; : : : ; Z^{n+1}$, so this type of coverage occurs on average. Since there are no promises conditional on a specific observation $X^{n+1} = x^{n+1}$, the constructed confidence intervals may offer conditional over-coverage for certain observations and conditional under-coverage on others.[12,17] Nevertheless, the framework is popular because it only assumes the exchangeability of data and because it offers a diagnostic tool for and a comparison tool among regression models.[12,17]

Our focus here is on split or inductive conformal predictive strategies. In this scenario, we separate the input/output data pairs into three disjoint categories. The first category is the proper training set Z_T, which is used to train a machine learning model. Suppose the regression model fit to Z_T is g. The second category Z_C is a calibration set, which is processed to provide a list of nonconformity scores for a predetermined nonconformity measure which depends on g; we will mention two approaches for selecting that measure for conformal prediction. The list of nonconformity scores together with g determines prediction intervals for future examples. The third category Z_T offers test examples which are used to evaluate the predictive model.

The simplest nonconformity score η for an instance $z^i = (x^i; y^i)$ and a regression model g is the absolute value of the prediction error:

$$\eta(z^i) = |y^i - g(x^i)|. \tag{2}$$

Let k be the number of examples in the calibration set. Apply the nonconformity measure to each example in the calibration set and sort the resulting scores in nonincreasing order as $r_1; r_2; : : : ; r_k$. Given a miscoverage rate α let $m = \lfloor \alpha (k + 1) \rfloor$ be the index of the $(1 - \alpha)$-percentile nonconformity score. Then for a future input image $X^j = x^j$ the prediction interval

$$[g(x^j) - r_m,\ g(x^j) + r_m] \tag{3}$$

offers valid marginal coverage. The strength of this scheme is its simplicity. However, for the LER estimation problem we will consider with images corrupted by several possible levels of Poisson noise we would find it more informative if the prediction interval width varies according to the difficulty of edge detection for each input image. Problems like this motivate "normalized" nonconformity scores.

Normalized nonconformity scores generalize the previous approach by incorporating a model to adapt to features of individual inputs. To augment the earlier basic method, apply η to every example in the proper training set and fit a model γ using the input-output pairs $(x^i; \eta(z^i))$ for all $z^i \in Z_T$. Let β be an additional sensitivity parameter. Then the corresponding normalized nonconformity score η_N is[11]

$$\eta_N(z^i) = \frac{|y^i - g(x^i)|}{\gamma(x^i) + \beta}. \tag{4}$$

Apply the nonconformity measure to each example in the calibration set and sort the resulting scores in nonincreasing order as $\varrho 1; \varrho 2; : : : ; \varrho k$. Then for a future input image $X^j = x^j$ the prediction interval

$$[g(x^j) - \varrho_m(\gamma(x^j) + \beta),\ g(x^j) + \varrho_m(\gamma(x^j) + \beta)] \tag{5}$$

offers valid marginal coverage. The advantage of the additional flexibility is in the potential for more informative prediction intervals, and the price for that benefit is the need for more modeling and computation.

In this paper we consider a neural network-based model on input-output pairs $(x^i; \eta(z^i))$ which is inspired by [9, Equation (16)]. EDGENet is effective at predicting LER for the dataset of interest, but it performs better with lower noise levels. Therefore, it is reasonable to study the relationship between xi and $-\ln |y^i - g(x^i)|$: For our problem x^i is an image

of dimension 64 x 1024 and less complex models are typically faster, so our approach to considering the relationship has two parts. We use an autoencoder[22] to reduce the dimensionality of the input image x^i to a feature vector $f(x^i)$ with 64 elements from which it is possible to recover an approximation to x^i. We next fit a second neural network model ϕ between $f(x^i)$ and $-\ln|y^i - g(x^i)|$. Our normalized nonconformity score η_N is then

$$\eta_N(z^i) \;=\; \frac{|y^i - g(x^i)|}{e^{-\phi(f(x^i))}} \tag{6}$$

and the corresponding prediction interval for a future input $X^j = x^j$ is

$$[g(x^j) - \varrho_m \cdot e^{-\phi(f(x^j))},\; g(x^j) + \varrho_m \cdot e^{-\phi(f(x^j))}]. \tag{7}$$

We will provide more details about the neural network architectures from which we obtain f and ϕ in Section 4. Observe that if it were possible to do this prediction perfectly, then all nonconformity scores would be one, and y^j would be an endpoint of a predicted interval for input x^j. Therefore, this approach should be an indicator of the difficulty of making the initial prediction $g(x^j)$ from input x^j.

A normalized conformal prediction interval for input x^j is forced to be centered at $g(x^j)$, and this may artificially limit the efficiency of that interval. Could it be more effective to directly estimate upper and lower bounds of prediction intervals while maintaining the marginal coverage guarantees of conformal prediction? For some machine learning problems this strategy works.[15-17] Here the upper and lower bounds are based on quantile regression,[14] which is a widely used methodology for nonparametric probabilistic forecasting initially introduced to predict conditional quantiles; quantile regression subsequently became a tool to study the relationships among response and predictor variables that were not captured by other regression techniques.[23] While popular, quantile regression does not generally promise validity.[15] Our experience is that this approach to devising prediction intervals is not as immediately amenable to successful implementation as conformal prediction, but it is never-the-less promising. We will begin by discussing conformalized quantile regression (CQR)[15] and a variation of it called CQR-r.[17]

In these methods, the proper training set is used to train an upper and a lower conditional quantile function. To elaborate, the conditional cumulative distribution function of random variable Y given random vector X is $P[Y \le y \mid X = x]$. Inverting this function leads to conditional quantile functions. For $\varepsilon \in [0, 1]$,

$$q_\epsilon(x) = \inf\{y :\; P[Y \le y \mid X = x] \ge \epsilon\}. \tag{8}$$

Then for a miscoverage rate α and parameter $\delta \in (0, \alpha)$ we have

$$P[q_\delta(x) \le Y \le q_{1-\alpha+\delta}(x) \mid X = x] \ge 1 - \alpha. \tag{9}$$

We will consider $\delta = 0.5\alpha$. For any $\varepsilon \in (0, 1)$, the standard[14,15] loss function used to estimate $q_\varepsilon(x)$ from training data is the pinball loss function:

$$\rho_\epsilon(y,\; \hat{y}) = \begin{cases} \epsilon(y - \hat{y}) & \text{if } y - \hat{y} \ge 0 \\ (1-\epsilon)(\hat{y} - y) & \text{if } y - \hat{y} < 0 \end{cases} \tag{10}$$

In general, a quantile regression method is based on some parametric linear or nonlinear family of functions $\varphi(\cdot)$ that can be realized; if N is the number of proper training images, x^i is training image i, and y^i is the ground truth LER corresponding to x^i, then we ideally wish to minimize over all φ

$$\frac{1}{N}\sum_{i=1}^{N} \rho_\epsilon(y^i, \varphi(x^i)). \tag{11}$$

We focus here on neural network models. One typically trains neural networks by applying a form of gradient descent to perform the optimization. Observe that $\rho_\varepsilon(y; \hat{y})$ is not differentiable along the line $y = \hat{y}$, and this potentially hinders the training process.[24] Therefore, we also consider the following smooth approximation of the pinball loss[25] which incorporates a positive smoothing parameter α:

$$s_{\epsilon,a}(y,\; \hat{y}) = \epsilon(y - \hat{y}) + a \ln\left(1 + \exp\left(\frac{\hat{y} - y}{a}\right)\right) \tag{12}$$

We estimate $q_\varepsilon(\cdot)$ by a function which offers an exact or approximate minimum to (11) using the original pinball loss or a smooth version. We let $\hat{q}_\varepsilon(\cdot)$ denote the estimated conditional quantile function.

In the next section we will discuss EDGENet. We created EDGENet with the belief that an automated, accurate and direct estimate of edge positions is an effective strategy to obtain a good point prediction of LER from an input image. Our first approach to investigating quantile regression is based on the assumption that an accurate estimate of edge positions should also be useful for good predictions of LER quantiles conditional on an input image. One objective of conformalized quantile regression is to easily modify deep neural networks to produce valid prediction intervals.[16] It would be computationally expensive to train a deep convolutional neural network from scratch to estimate an LER quantile, so we opt for a transfer learning[26] strategy. In particular we modify EDGENet to obtain our initial estimates of quantiles from the proper training set. We will describe our approach in Section 4.

Since quantile regression does not generally guarantee coverage, the initial lower and upper bounds, $\hat{q}_{0.5\alpha}(x)$ and $\hat{q}_{1-0.5\alpha}(x)$ need adjustment. CQR and CQR-r use different modi cations. CQR employs the following nonconformity score η_Q:

$$\eta_Q(z^i) \;=\; \max\{\hat{q}_{0.5\alpha}(x^i) - y^i,\; y^i - \hat{q}_{1-0.5\alpha}(x^i)\}. \tag{13}$$

Apply this nonconformity measure to each example in the calibration set and sort the resulting scores in nonincreasing order as $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k$. Then for a future input image $X^j = x^j$ the prediction interval

$$[\hat{q}_{0.5\alpha}(x^j) - \varepsilon_m,\; \hat{q}_{1-0.5\alpha}(x^j) + \varepsilon_m] \tag{14}$$

offers valid marginal coverage.

Ref. 16 proposes a variation of CQR which requires an estimate of the regression median in addition to the upper and lower quantiles; since this is more complex we do not consider this approach in this paper. CQR- r[17] is a variant of the scheme in Ref. 16 that does not need an estimate of the regression median. Here the nonconformity score η_{Q-r} is

$$\eta_{Q-r}(z^i) \;=\; \max\left\{\frac{\hat{q}_{0.5\alpha}(x^i) - y^i}{\hat{q}_{1-0.5\alpha}(x^i) - \hat{q}_{0.5\alpha}(x^i)},\; \frac{y^i - \hat{q}_{1-0.5\alpha}(x^i)}{\hat{q}_{1-0.5\alpha}(x^i) - \hat{q}_{0.5\alpha}(x^i)}\right\}. \tag{15}$$

Apply this nonconformity measure to each instance in Z_C and sort the resulting scores in nonincreasing order as $\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \ldots, \tilde{\varepsilon}_k$. Then for a future input image $X^j = x^j$ the prediction interval

$$[(1+\tilde{\varepsilon}_m)\hat{q}_{0.5\alpha}(x^j) - \tilde{\varepsilon}_m\hat{q}_{1-0.5\alpha}(x^j),\; (1+\tilde{\varepsilon}_m)\hat{q}_{1-0.5\alpha}(x^j) - \tilde{\varepsilon}_m\hat{q}_{0.5\alpha}(x^j)]. \tag{16}$$

provides valid marginal coverage. The conformalized quantile regression schemes are the most challenging ones to implement and learning quantiles from raw image data appears to require more computational resources than we used with our transfer learning scheme. However, we believe that quantile regression has the potential to be a useful tool when combined with domain knowledge. Therefore, we examine an idealized scenario where instead of using raw images we describe them by a few key features. In particular, instead of training on simulated images we train on the parameters we describe in Section 4 that we use to generate those images together with the LER estimate produced by EDGENet.

3. On Edgenet and the Simulated Dataset

We initially designed the deep convolutional neural network EDGENet to directly estimate left and right line edge positions from a 64 x 1024 noisy SEM image with pixel size 0:5 x 2 nm and containing one rough line and an unknown level of Poisson noise as illustrated in Figure 1; the original EDGENet outputs a matrix of dimensions 2 x 1024 specifying the predicted line edge positions, which are reported with pixel-level precision and not with subpixel-level precision. EDGENet uses seventeen convo-

lutional layers[28] with filter dimension 3 x 3 x input depth. The initial four convolutional layers each use 64 filters, convolutional layers five through eight each use 128 filters, convolutional layers nine through twelve each use 256 filters and convolutional layers thirteen through sixteen each use 512 filters. Each convolutional layer of EDGENet except for the last layer is followed by a batch normalization layer[29] and a dropout layer[30] with dropout probability of 0.2 for regularization. The final convolutional layer uses a single filter to output the 2 x 1024 matrix of estimated edge positions. Figure 2 depicts the sizes of the output volumes or tensors corresponding to each convolutional layer. We use the mean absolute error (MAE) loss criteria to train EDGENet; i.e., if N is the number of training images, x^i is training image i, $\varphi(\cdot)$ is a parametric nonlinear function that can be realized by EDGENet and ζ^i is the matrix of ground truth edge positions corresponding to x^i, then we ideally wish to minimize over all φ

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} \|y^i - \varphi(\zeta^i)\|_1. \tag{17}$$

In order to facilitate the training process for our quantile regressors, we modified EDGENet by adding an extra layer to compute the LER values associated with the left and right edges.

The training of deep convolutional neural networks requires large datasets, and our dataset consists of simulated SEM images. The first step in the generation of the dataset is to apply the Thorsos method[31,32] with normally distributed random variables to simulate rough line edges or linescans. Each linescan follows a Palasantzas spectral model,[33] which is described by three parameters: σ is the line edge roughness (LER), i.e., the standard deviation of edge positions, a represents the roughness (or Hurst) exponent and ξ denotes the correlation length:

$$PSD(f) = \frac{\sqrt{\pi}\Gamma(\alpha+0.5)}{\Gamma(\alpha)} \cdot \frac{2\sigma^2\xi}{(1+(2\pi f\xi)^2)^{\alpha+0.5}}. \tag{18}$$

Every simulated edge is 2.048 microns or, equivalently, 1024 pixels long. Our edges can take on eight possible LER values (σ = 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8 nm), nine possible Hurst/roughness exponent values (α = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) and 35 possible correlation length values (ξ = 6, 7, ..., 40 nm). This results in 2520 possible combinations of parameters (σ, α, ξ), and for each one we generated eight edges.

By applying the SEM simulator ARTIMAGEN[34,35] with the parameter settings suggested by examples within the ARTIMAGEN library to a line of width 10 nm or 15 nm having two of the previously constructed rough edges we generate an image of dimension 64 x 1024 pixels with pixel width 0.5 nm and pixel height 2 nm which incorporates random backgrounds, a fixed edge effect, fine structure and Gaussian blur; we created 10080 such images. The locations of the lines within the images vary. The ARTIMAGEN library does not provide fractional edge positions, so the edge positions are rounded. Because rounding operations are not differentiable they were removed in the implementation of the new final layer of EDGENet as well as in the actual LER computations.

These 10080 images form our original image set. The ARTIMAGEN library enables us to generate ten noisy images for each original image having Poisson noise with electron density per pixel in the range {2, 3, 4, 5, 10, 20, 30, 50, 100, 200}. The resulting 100800 images form our noisy image dataset. From the noisy and original images we constructed a supervised learning dataset of pairs of matrices (x^i; y^i) for the training of the initial version of EDGENet, where the input x^i is a noisy image and the output array y^i has dimension 2 x 1024 with entries consisting of the edge positions in the corresponding original image. In the enlarged EDGENet that we consider here the output y^i is now either the LER of the left edge of the original image or the LER of the right edge of the original image.

4. Auxiliary Neural Networks, Experiments, and Results

4.1 An Autoencoding Approach to Normalized Conformal Prediction

Autoencoders are neural networks that are typically trained to reconstruct their input[22,36] and are used to learn salient latent representations without the need for training labels; they are a powerful and popular unsupervised learning tool. We designed our autoencoder network to compress the dimensions of the noisy image x^i from 64 x 1024 to 1 x 64 in order to realize a robust approximation $f(x^i)$ that encapsulates the "difficulty" of each noisy image for use in a second neural network that predicts – ln $|y^i - g(x^i)|$. Our architecture consists of twelve convolutional layers with filters of size 3 x 3 x input depth. Following the typical approach to constructing autoencoders, there is some symmetry in the design of pairs of layers. The first and eleventh layers each have 64 filters, the second and tenth layers each have 128 filters, the third and ninth layers each have 256 filters, the fourth and eighth layers each have 512 filters, the remaining intermediate layers each contain a single filter and aid in compression, and the final layer also contains a single filter. Each layer except for the three compression layers and the twelfth layer is followed by batch normalization and dropout layers with a dropout rate of 20% to curb over fitting. The fifth and sixth layers are each followed by 2 x 2 max-pool layers to help provide the desired dimension reduction. The output of the seventh layer is a tensor of dimension 1 x 1 x 64 which is reshaped to a row vector with 64 elements. All subsequent convolutional layers are preceded by up-sampling layers of sizes 4 x 4, 2 x 4, 2 x 2 and 1 x 2, respectively, to reproduce at the final layer a tensor of dimension 1 x 64 x 1024 which is reshaped to an image of dimensions 64 x 1024 from the compressed representation.

We use the mean square error (MSE) loss criteria to train the autoencoder; i.e., if N is the number of training images, x^i is training image i, and $\psi(\cdot)$ is a parametric nonlinear function that can be realized by the autoencoder, then we ideally wish to minimize over all ψ

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N} \|x^i - \psi(x^i)\|_2^2. \tag{19}$$

The second neural network architecture is a relatively shallow fully-connected neural network with four layers. All layer utilize rectified linear units (ReLU) nonlinearities as their activation functions. The first layer contains 128 units and accepts inputs of size 1 x 64. The second and third layers each have 64 units and apply L2 regularization. The final layer has two units which predict – ln $|y^i - g(x^i)|$ for both the left and right edges. We trained the network with mean square error loss using the Adam[37] optimizer with a learning rate of 0.001 for 24 epochs.

4.2 Transfer Learning for Conformalized Quantile Regression

Our quantile regression network builds upon the pretrained EDGENet convolutional neural network. The original EDGENet architecture outputs an array of line edge positions. By simply adding a custom layer to compute standard deviations EDGENet is transformed into a network from which we can directly compute the left and right edge LER values. This is useful because it enables us to train the network end-to-end with input/output pairs of the form (x^i, LER values), which is needed to perform quantile regression. We also utilize a form of transfer learning[26] in which we freeze the weights on the initial layers of EDGENet so that the quantile regression network can fully exploit prior information learned through the edge prediction task. In particular, we freeze the weights from the first fifteen layers and retrain the remainder of the network using the pinball loss function. Given a desired miscoverage rate α we train two separate networks for the quantile functions $q_{0.5\alpha}$ and $q_{1-0.5\alpha}$. One of the known challenges in quantile regression is the issue of quantile crossing, where the upper quantile network output could be smaller than the lower quantile network output (see, e.g., Ref. 24). The smooth pinball loss function alleviates this problem.[24] Another challenge is the quantile regression networks tend to undercover and their performance depends on network hyperparameters;[38] this also motivates the use of separate networks to train upper and lower quantiles. We use pinball loss for training the lower quantiles and smooth pinball loss for training the upper quantiles.

4.3 Quantile Regression Based on the Defining Parameters

Quantile regression[14] is an important tool for nonparametric probabilistic forecasting and can be helpful in describing the relationships among

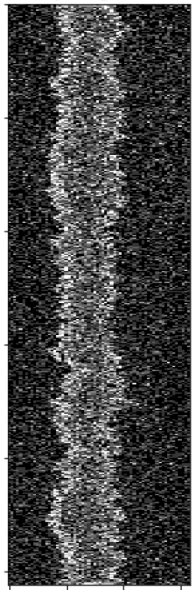

Figure 1. A noisy SEM image of dimension 64 x 1024. The image has one line with two edges. The aspect ratio of the image has been scaled to improve viewing. Reprinted with permission from Ref. 27.

the pinball loss function in both cases as well as a batch size of sixteen in training each network. We train each network for about 90 epochs.

4.4 Experiments and Results

We next describe and compare the experimental results of five different approaches to constructing prediction intervals: conformal prediction, normalized conformal prediction, quantile regression, conformalized quantile regression (CQR), and a second conformalized quantile regression scheme (CQR-r). For each of these schemes we perform four groups of experiments; namely, we study the outcomes of each scheme on the left and right edges at miscoverage rates of 0.1 and 0.05. One of the messages we would like to deliver is the importance of modeling. In addition to these five schemes we will discuss preliminary results for six other approaches for the left edge data at a miscoverage rate of 0.1. Five of these approaches are "ensemble" schemes[13] which use the previous models in a simple way to improve upon quantile regression. The last scheme is intended to demonstrate the benefits of incorporating domain knowledge.

We partition the simulated dataset of 100800 input/output pairs with all values normalized to the range (0,1) into a proper training set, a calibration set and a test set. The calibration and test sets together correspond to the 11520 input images with correlation length ξ in the set {10, 20,

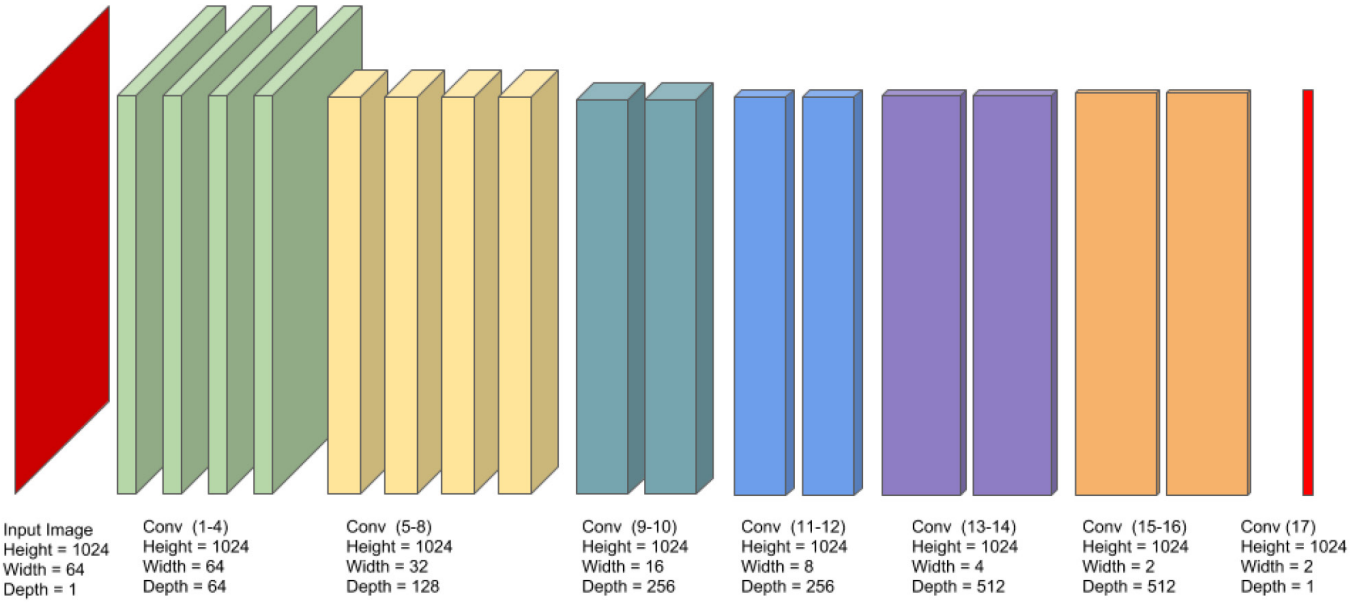

Figure 2. The 17 convolutional layers of EDGENet, which inputs a noisy SEM image of dimension 64 x 1024 and outputs a matrix of dimension 2 x 1024. Reprinted with permission from Ref. 21.

response and predictor variables.[23] Therefore, domain knowledge can potentially assist in improving quantile regression. To test this idea, we considered an idealized situation for producing a quantile regression prediction interval for the left edge with 90% coverage. To avoid crossing quantiles and low coverage we utilize separate networks to train for upper and lower quantiles. Instead of using the 64 1024 raw image data, our input to our quantile regression networks is a vector with seven elements. Six of these are the parameters used to create the simulated image { the three parameters of the Palasantzas spectral model, the noise, the line width and the position. The last element is the left edge LER value predicted by EDGENet. Our quantile regression fully-connected neural networks each have six layers and apply rectified linear units (ReLU) nonlinearities as their activation functions. For both networks the first two layers each contain 32 units, the third layer has sixteen units, the fourth layer has eight units, the fifth layer has four units, and the final layer has a single unit that outputs the LER quantile prediction. We use

30, 40} nm. The proper training set consisted of the remaining 89280 input/output pairs, which matches the proper training set of EDGENet. The output depends on the network we train; for the autoencoder the network output is a noisy image. The calibration and test sets are the same size and were carefully constructed to as closely as possible approximate exchangeability; the 1152 original images with correlation length ξ in the set {10, 20, 30, 40} nm were randomly split into two subsets each of size 576. All 5760 noisy images corresponding to one subset of original images are the inputs for the calibration set, and the remaining 5760 noisy images are inputs for the test set.

All models were trained using an Intel Xeon E5-2680 v4 processor running at 2.4GHz and a Tesla K80 GPU. Training hyper parameters such as the batch size, optimizers and the learning rate require optimization and fine tuning. Except for our quantile regression networks in the case where we apply domain knowledge we use a batch size of eight input-output pairs to account for memory constraints. In all cases we employ

Table 1. Coverage and interval length statistics for the LER of the left edge when the miscoverage rate α equals 0.1.

Method	Coverage (%)	Average interval length (nm)	Interval length (nm)		Ratio of interval length to original LER	
			Minimum	Maximum	Minimum	Maximum
Conformal Prediction (CP)	90.22	0.135	0.135	0.135	0.065	0.487
Normalized CP	90.15	0.153	0.006	1.588	0.008	2.716
Quantile Regression	85.94	0.197	0.033	0.496	0.056	0.521
CQR	89.55	0.215	0.052	0.515	0.071	0.570
CQR-r	89.51	0.225	0.038	0.566	0.065	0.595
Defining Parameters	89.77	0.103	0.032	0.239	0.028	0.501

Table 2. Coverage and interval length statistics for the LER of the right edge when the miscoverage rate α equals 0.1.

Method	Coverage (%)	Average interval length (nm)	Interval length (nm)		Ratio of interval length to original LER	
			Minimum	Maximum	Minimum	Maximum
Conformal Prediction (CP)	89.22	0.186	0.186	0.186	0.079	0.684
Normalized CP	88.96	0.241	0.017	1.109	0.012	2.141
Quantile Regression	83.26	0.224	0.047	0.573	0.055	0.731
CQR	88.61	0.251	0.074	0.600	0.079	0.823
CQR-r	88.26	0.265	0.056	0.677	0.065	0.864

the Adam[37] optimizer for stochastic gradient descent with a learning rate of 0.001 during training.

Tables 1 and 2 respectively summarize the performance of the main five approaches for the left and right edges when the miscoverage rate is 0.1. Tables 3 and 4 likewise respectively describe the results of the main five approaches for the left and right edges when the miscoverage rate is 0.05. Notice that the prediction intervals for the right edge tend to be wider than those for the left edge. As Table 2 illustrates, the coverage for the right edge is uniformly below 90%; we feel that EDGENet has more difficulties processing right edge data than left edge data. Furthermore, the authors of Ref. 17 report the undercoverage of conformalized quantile regression schemes for some real datasets, and they recommend using 10-30% of all training examples for the calibration set while we used about 6.06% of all training examples for the calibration set. The simple conformal prediction scheme achieves good coverage but cannot offer information about the difficulty of the regression task for a specific instance. Our implementation of normalized conformal prediction has reasonable coverage. Its average prediction interval length is higher than that of simple conformal prediction, but it can produce significantly narrower intervals for less difficult instances. Our analysis reveals that normalized conformal prediction intervals are smaller than their simple conformal prediction counterparts 60% of the time for left edges and 45% of the time for right edges further confirming the observation that EDGENet has more difficulty with the right edges. Our initial approach to quantile regression tends to grossly undercover and has a relatively large average interval length. There are two main factors contributing to this. First, quantile regression neural networks are known to have a tendency to undercover and their performance is sensitive to network hyperparameters.[38] Second, our transfer learning scheme uses the pre-trained weights of EDGENet and attempts to learn both left and right edge statistics with relatively few available parameters from raw image data. Nevertheless, despite these challenges CQR and CQR-r offer significant coverage improvements and show the potential of conformalization techniques in producing coverage guarantees.

While these five schemes offer the beginning of a study of prediction intervals, there are endless ways to improve on modeling. In Table 5 we consider five simple ensemble strategies. Each one of the first three produces a new quantile regression algorithm which is founded in part on our initial quantile regression scheme and which significantly outperforms it. Observe from Tables 1-4 that normalized conformal prediction produces both very narrow and very wide intervals while our conformalized quantile regression algorithms avoid these extremes. Nevertheless, all three schemes offer good coverage. Therefore, it is desirable to take advantage of the best of both worlds, and this leads us to examine the following three schemes for the left edge data and 90% coverage. The Ensemble 1 scheme uses the shorter predicted interval between normalized conformal prediction and CQR. The Ensemble 2 scheme uses the shorter predicted interval between normalized conformal prediction and CQR-r. The Ensemble 3 scheme uses the shortest predicted interval among all three algorithms. As Table 5 demonstrates, each scheme has a better coverage than our initial quantile regression scheme while offering a better expected length than simple conformal prediction. It is also possible to consider ensemble strategies involving conformal prediction. Observe that if two models offer miscoverage rates α_1 and α_2, then the miscoverage rate of their ensemble will be between max $\{\alpha_1, \alpha_2\}$ and $\alpha_1 + \alpha_2$. For Ensemble 4 we use the shorter predicted interval between the

normalized conformal prediction scheme generated for $\alpha_1 = 0.1$ and the conformal prediction scheme generated for $\alpha_2 = 0.05$. For Ensemble 5 we use the shorter predicted interval between the normalized conformal prediction scheme generated for $\alpha_1 = 0.05$ and the conformal prediction scheme generated for $\alpha_2 = 0.1$. Table 5 shows that Ensembles 4 and 5 improve both upon the coverage and the average interval length attained with Ensembles 1-3.

If there are benefits to combining machine learning models, then there are potentially greater benefits to combining machine learning models with domain knowledge. Here we will look at an instance where our domain knowledge is perfect and again focus on the left edge data with 90% coverage. In this case we replace the raw image data with a vector with seven elements. Six of these are the parameters used to create the simulated image { the three parameters of the Palasantzas spectral model, the noise, the line width and the position. The last element is the left edge LER value predicted by EDGENet. These vectors are our inputs to two quantile regression networks, which are described in the previous subsection. We report our result in Table 1, and it is clearly superior to all of the previous approaches that we have discussed. Furthermore, we made no attempt to optimize the model and we did not conformalize this quantile regression algorithm.

5. Conclusion

The digital transformation of semiconductor manufacturing is underway. To maximize the rewards of this opportunity the semiconductor industry must have better tools to apply deep learning and other forms of artificial intelligence in decision making. Distribution-free prediction intervals with coverage guarantees are one approach to address this need, and this topic is an active area of research in the data sciences. The most successful techniques are likely to leverage advances in multiple domains.

6. Acknowledgments

The authors used the Texas A&M University High Performance Research Computing Facility to conduct part of the research. They also thank Dr. Narendra Chaudhary for sharing some of his experiences and Mr. Prasad Hugar for assisting with some of the experiments.

7. References

[1] 2020 IRDS. International Roadmap for Devices and Systems. 2020 Edition. *Metrology*. (2020).

[2] 2020 IRDS. International Roadmap for Devices and Systems. 2020 Edition. *Factory Integration*. (2020).

[3] 2020 IRDS. International Roadmap for Devices and Systems. 2020 Edition. Yield Enhancement. (2020).

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, 521(7553), 436-444 (2015).

[5] B. D. Bunday, N. G. Orji, and J. A. Allgair, "High volume manufacturing metrology needs at and beyond the 5 nm node," **Proc. of SPIE**, 11611, 116110F (2021).

[6] Symposium: Strategy for Resilient Manufacturing Ecosystems Through Artificial Intelligence. Report from the First Symposium Workshop: Aligning Artificial Intelligence and U.S. Advanced Manufacturing Competitiveness. December 2 and 4, 2020. Facilitated by UCLA. Supported by the National Science Foundation and the National Institute of Standards and Technology. March 2021.

[7] V. Vovk, A. Gammerman, and G. Shafer, *Algorithmic Learning in a Random World*, Springer Science+Business Media, New York (2005).

[8] G. Shafer and V. Vovk, "A tutorial on conformal prediction," *J. Mach. Learn. Res. 9*, 371-421 (2008).

[9] H. Papadopoulos and H. Haralambous, "Reliable prediction intervals with regression neural networks," *Neural Networks 24*, 842-851 (2011).

[10] V. Balasubramanian, S.-S. Ho, and V. Vovk, *Conformal Prediction for Reliable Machine Learning*, Morgan Kaufmann, Waltham, MA (2014).

[11] H. Linusson, "An introduction to conformal prediction," *The 6th Symposium on Conformal and Probabilistic Prediction with Applications (COPA 2017)*, Tutorial 1 (2017).

[12] J. Lei, M. G'Sell, A. Rinaldo, et al., "Distribution-free predictive inference for regression," *J. Am. Stat. Assoc. 113*(523), 1094-1111 (2018).

[13] I. Cortés-Ciriano and A. Bender, "Deep confidence: a computationally efficient framework for calculating reliable prediction errors for deep neural networks," *J. Chem. Inf. Model. 59*(3), 1269-1281 (2018).

[14] R. Koenker and G. Bassett Jr., "Regression quantiles," *Econometrica*, 46(1), 33-50 (1978).

[15] Y. Romano, E. Patterson, and E. J. Candès, "Conformalized quantile regression," *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, 3543-3553 (2019).

[16] D. Kivaranovic, K. D. Johnson, and H. Leeb, "Adaptive, distribution-free prediction intervals for deep networks," *Proc. of the Twenty Third Int. Conf. on Artificial Intelligence and Statistics, PMLR 108*, 4346-4356 (2020).

[17] M. Sesia and E. J. Candès, "A comparison of some conformal quantile regression methods," *Stat* 9(1), e261 (2020).

[18] H. J. Levinson, *Principles of Lithography*, Third Edition, SPIE Press, Bellingham, Washington (2010).

[19] S. Babin, "Automated, model based contour and CD extraction in CD-SEM metrology," one-page abstract for Lithography Workshop 2018, Ketchum, Idaho (2018).

[20] N. Chaudhary, S. A. Savari and S. S. Yeddulapalli, "Automated rough line-edge estimation from SEM images using deep convolutional neural networks," **Proc. of SPIE**, 10810, 108101 (2018).

[21] N. Chaudhary, S. A. Savari and S. S. Yeddulapalli, "Line roughness estimation and Poisson denoising in scanning electron microscope images using deep learning," *J. Micro/Nanolith. MEMS MOEMS,* 18(2), 024001 (2019).

[22] G. E. Hinton and R. S. Zemel, "Autoencoders, minimum description length and Helmholtz free energy," *NIPS'93: Proc. of the 6th Int. Conf. on Neural Inf. Proc. Sys.,* 3-10 (1993).

[23] B. S. Cade and B. R. Noon, "A gentle introduction to quantile regression for ecologists," *Front. Ecol. Environ.* 1(8), 412-420 (2003).

[24] K. Hatalis, A. J. Lamadrid, K. Scheinberg, and S. Kishore, "Smooth pinball neural network for probabilistic forecasting of wind power," arXiv.org preprint arXiv:1710.01720v1[stat.ML] (2017).

[25] S. Zheng, "Gradient descent algorithms for quantile regression with smooth approximation," *Int. J. Machine Learning and Cybernetics* 2(3), 191 (2011).

[26] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proc. of the IEEE*, 109(1), 43-76 (2021).

[27] N. Chaudhary and S. A. Savari, "Towards a visualization of deep neural networks for rough line images," **Proc. of SPIE**, 11177, 111770S (2019).

[28] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in M. A. Arbib, ed., *The Handbook of Brain Theory and Neural Networks*, pp. 255-258, MIT Press, Cambridge (1998).

[29] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in F. Bach and D. Blei, eds., *Proceedings of Machine Learning Research,* Vol. 37, pp. 448-456, Microtome Publishing, Brookline (2015).

[30] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. of Machine Learning Research*, 15(1), 1929-1958 (2014).

[31] E. I. Thorsos, "The validity of the Kirchoff approximation for rough surface scattering using a Gaussian roughness spectrum," *J. Acoust. Soc. Am.*, 83(1), 78-92 (1988).

Table 3. Coverage and interval length statistics for the LER of the left edge when the miscoverage rate α equals 0.05.

Method	Coverage (%)	Average interval length (nm)	Interval length (nm)		Ratio of interval length to original LER	
			Minimum	Maximum	Minimum	Maximum
Conformal Prediction (CP)	95.45	0.199	0.199	0.199	0.095	0.715
Normalized CP	95.33	0.214	0.009	2.220	0.010	3.798
Quantile Regression	88.75	0.292	0.091	0.776	0.110	1.122
CQR	95.40	0.333	0.132	0.818	0.139	1.264
CQR-r	95.50	0.350	0.108	0.932	0.131	1.346

Table 4. Coverage and interval length statistics for the LER of the right edge when the miscoverage rate α equals 0.05.

Method	Coverage (%)	Average interval length (nm)	Interval length (nm)		Ratio of interval length to original LER	
			Minimum	Maximum	Minimum	Maximum
Conformal Prediction (CP)	96.89	0.233	0.233	0.233	0.112	0.839
Normalized CP	94.41	0.313	0.022	1.441	0.016	2.781
Quantile Regression	92.60	0.358	0.103	0.912	0.159	1.122
CQR	95.45	0.385	0.131	0.940	0.182	1.216
CQR-r	95.26	0.340	0.115	1.017	0.177	1.251

Table 5. Coverage and interval length statistics for five ensemble schemes for the LER of the left edge. For the first three ensemble strategies the miscoverage rate of all models is approximately 0.1. The Ensemble 1 scheme applies the minimum length interval from Normalized Conformal Prediction and CQR. The Ensemble 2 scheme applies the minimum length interval from Normalized Conformal Prediction and CQR-r. The Ensemble 3 scheme applies the minimum length interval from all three approaches. The Ensemble 4 scheme applies the minimum length interval from Normalized Conformal Prediction from Table 1 and from Conformal Prediction from Table 3. The Ensemble 5 scheme applies the minimum length interval from Conformal Prediction from Table 1 and Normalized Conformal Prediction from Table 3.

Method	Coverage (%)	Average interval length (nm)	Interval length (nm)	
			Minimum	Maximum
Ensemble 1	87.66	0.133	0.006	0.431
Ensemble 2	86.82	0.134	0.006	0.471
Ensemble 3	86.89	0.133	0.006	0.431
Ensemble 4	88.92	0.131	0.006	0.199
Ensemble 5	89.22	0.127	0.009	0.135

[32] C. A. Mack, "Generating random rough edges, surfaces, and volumes," *Applied Optics*, 52(7), 1472-1480 (2013).

[33] G. Palasantzas, "Roughness spectrum and surface width of self-affine fractal surfaces via the K-correlation model," *Phys. Rev. B,* 48(19), 472-478 (1993).

[34] P. Cizmar, A. E. Vladár, B. Ming, and M. T. Postek, "Simulated SEM images for resolution measurement," *Scanning*, 30(5), 381-391 (2008).

[35] P. Cizmar, A. E. Vladár, and M. T. Postek, "Optimization of accurate SEM imaging by use of artificial images," **Proc. of SPIE**, 7378, 737815 (2009).

[36] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," arXiv.org preprint arXiv:2003.05991v2[cs.LG] (2021).

[37] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," *Proc. 3rd Inter. Conf. on Learning Representations (ICLR 2015)*, May 9 Conference Poster Session Board 11 (2015).

[38] N. Tagasovska and D. Lopez-Paz, "Single-model uncertainties for deep learning," arXiv.org preprint arXiv:1811.00908v3[stat. ML] (2019).

Industry Briefs

New Fab construction

Intel has indicated in the past that a new fab site in Europe could attract investment of $95 billion over a decade as a number of fabs are built there.

https://www.electronicsweekly.com/news/business/intel-chooses-magdeburg-european-fab-site-2022-02/

Taiwan's UMC to spend $5 billion on new chip plant in Singapore

https://www.thestar.com.my/tech/tech-news/2022/02/24/taiwan039s-umc-to-spend-5-billion-on-new-chip-plant-in-singapore

Industry reports cover trends in Foundry Growth, Semi Industry Capex, Global Semiconductor Sales.

https://www.semiconductor-digest.com/2022-to-mark-the-third-year-in-a-row-of-20-growth-for-the-foundry-market/

https://www.semiconductor-digest.com/semi-industry-capex-forecast-to-jump-24-and-reach-over-190-billion-this-year/

https://www.semiconductor-digest.com/global-semiconductor-sales-increase-26-8-year-to-year-in-january/

Reviving U.S. semiconductor production, at the apex of the electronics industry, is unlikely to succeed without also rebuilding the more basic domestic ecosystem of companies in chip assembly and test, according to industry experts.

https://www.eetimes.com/reshoring-chip-industry-risks-failure-with-just-more-fabs/#

2022 Chip Forecast: Mixed Signals. Jim Feldhan, president of Semico Research, sat down with Semiconductor Engineering to talk about the outlook for the semiconductor market.

https://semiengineering.com/2022-chip-forecast-mixed-signals/

Unsolved Issues In Next-Gen Photomasks. Semiconductor Engineering sat down to discuss optical and EUV photomasks issues, as well as the challenges facing the mask business.

https://semiengineering.com/unsolved-issues-in-next-gen-photomasks/

Join the premier professional organization for mask makers and mask users!

About the BACUS Group

Founded in 1980 by a group of chrome blank users wanting a single voice to interact with suppliers, BACUS has grown to become the largest and most widely known forum for the exchange of technical information of interest to photomask and reticle makers. BACUS joined SPIE in January of 1991 to expand the exchange of information with mask makers around the world.

The group sponsors an informative monthly meeting and newsletter, BACUS News. The BACUS annual Photomask Technology Symposium covers photomask technology, photomask processes, lithography, materials and resists, phase shift masks, inspection and repair, metrology, and quality and manufacturing management.

Individual Membership Benefits include:

- Subscription to BACUS News (monthly)
- Eligibility to hold office on BACUS Steering Committee

spie.org/bacushome

Corporate Membership Benefits include:

- 3-10 Voting Members in the SPIE General Membership, depending on tier level
- Subscription to BACUS News (monthly)
- One online SPIE Journal Subscription
- Listed as a Corporate Member in the BACUS Monthly Newsletter

spie.org/bacushome

CALENDAR

2022

SPIE Advanced Lithography + Patterning
24-28 April 2022
San Jose, California, USA
www.spie.org/al

Photomask Japan
26-28 April 2022
Digital Forum
www.photomask-japan.org

EMLC 2022
20-23 June
Leuven, Belgium
https://www.emlc-conference.com/

SPIE Photomask Technology + Extreme Ultraviolet Lithography
25-29 September 2022
Monterey, California, USA
www.spie.org/puv

SPIE, the international society for optics and photonics, brings engineers, scientists, students, and business professionals together to advance light-based science and technology. The Society, founded in 1955, connects and engages with our global constituency through industry-leading conferences and exhibitions; publications of conference proceedings, books, and journals in the SPIE Digital Library; and career-building opportunities. Over the past five years, SPIE has contributed more than $22 million to the international optics community through our advocacy and support, including scholarships, educational resources, travel grants, endowed gifts, and public-policy development. www.spie.org.

SPIE.
International Headquarters
P.O. Box 10, Bellingham, WA 98227-0010 USA
Tel: +1 360 676 3290
Fax: +1 360 647 1445
help@spie.org • spie.org

Shipping Address
1000 20th St., Bellingham, WA 98225-6705 USA

SPIE.EUROPE
2 Alexandra Gate, Ffordd Pengam, Cardiff,
CF24 2SA, UK
Tel: +44 29 2089 4747
Fax: +44 29 2089 4750
info@spieeurope.org • spieeurope.org

You are invited to submit events of interest for this calendar. Please send to lindad@spie.org.