PHOTOMASK

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

BACUS NEWS

MAY 2019
VOLUME 35, ISSUE 5

2nd Place Best Oral — PUV18

The self-driving photomask

J. P. Shiely, Synopsys, Inc., 2025 NE Cornelius Pass Rd., Hillsboro OR, USA

ABSTRACT

Machine learning (ML) has become increasingly powerful and several recent works have demonstrated the capability of neural networks to achieve performance gains for lithography applications. Much of the general literature on neural networks involves image classification. Application of neural networks to lithography requires increased scrutiny. How far can such a system be trusted, and how should we respond if the system fails? Neural networks can appear inscrutable and we lack understanding of why these systems generalize so well. On the other hand, the benefits neural networks appear to offer, in terms of reduced runtime or more accurate models, are compelling. This work will illustrate how two techniques, the Information Bottleneck (IB) and t-Distributed Stochastic Nearest Neighbors (t-SNE), that can improve our understanding of how neural networks work. We will use a multilayer perceptron for a simple resist model implemented with neural networks. We will then discuss how visualization methods can help assess the readiness of a neural network for a task, or help diagnose potential causes of failure.

1. Introduction

Machine Learning (ML) is used for diverse applications, including within the field of microlithography.[1-3] Benefits include faster calculations and improved representation of phenomena for which rigorous causal understanding is lacking. While machine learning has long been a foundational technology for Optical Proximity Correction (OPC) models,[4, 5] papers demonstrating powerful new methods have recently been published.[6-13]

Among many useful ML methods, we focus attention on Deep Learning (DL). DL is the training of neural networks with multiple hidden layers and very large numbers of free parameters. This neural network architecture is called a deep neural network (DNN). Beginning with the AlexNet paper[14] researchers overcame the plagues of prior generations of DNNs, including exploding gradients, vanishing gradients, and local minima. Back propagation with mini-batch gradient descent methods and some additional practical tricks (for example dropout, Adam optimization, batch normalization, etc.) surpassed prior limitations of the DNN training process.[15] The adoption of GPUs to accelerate the highly-parallelizable computations of DNNs also contributed to the success of DL.

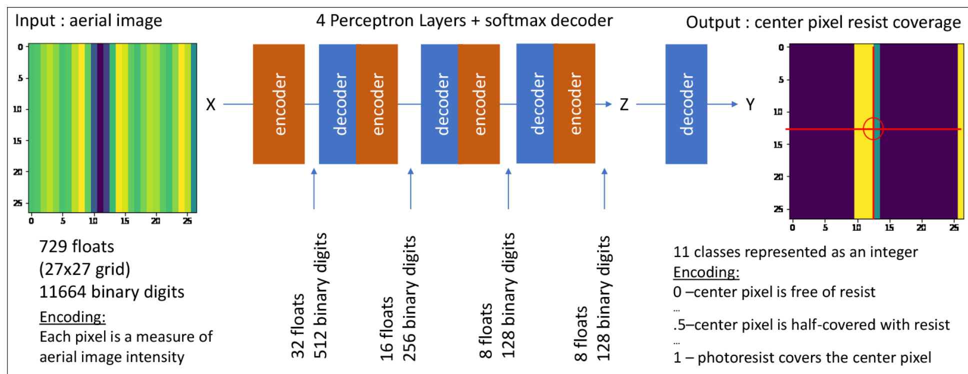

Figure 1. A simple experiment illustrates how the action of a neural network can be interpreted and visualized as a series of encodings of input samples, with sufficient statistics to accurately determine the output class with which sample is associated. The input aerial image X is a 27x27 pixel matrix of floats, with each element an aerial image intensity signal. The model is a 4-layer perceptron. The first layer encodes the input in a 32 float vector, compressing it. Succeeding neuron layers decode the previous layer, and generate new encodings. The final layer is a softmax decoder which receives the encoding from the final neuron layer and determines the output class. The output class is the percentage of resist coverage for the center pixel in the aerial image, in steps of 10% for easier visualization.

TAKE A LOOK INSIDE:

INDUSTRY BRIEFS
—see page 7

CALENDAR
For a list of meetings
—see page 8

SPIE.

10810-2

Editorial

Open Issues in Mask Technology as EUV Enters High Volume Manufacturing

Moshe Preil and James W Westphal, KLA Corporation

After years of optimistic projections and false starts, 2019 is finally the year that EUV will enter volume production. There are over 20 production scanners in the field. The combined cost of these tools, including associated equipment, installation and facilities is well into the billions of dollars, and this does not include the cost of new buildings currently under construction to support volume EUV production.

Mask shop investment in EUV capable equipment, including writing, inspection, metrology, repair, review and cleaning tools as well as related infrastructure for storage, transportation, and pellicle support has also been substantial. However, in both mask shops and wafer fabs, key questions still remain unanswered even as High Volume Manufacturing (HVM) begins in the fab. These questions concern technical choices, financial considerations, and scheduling, and need to be answered to develop comprehensive, end-to-end strategies for mask inspection, use, and qualification strategies. For example, uncertainty over pellicle technology options and timing cascade into a series of questions related to reticle qualification flows throughout the lifetime of a mask.

Additional uncertainty comes from the lack of data on reticle contamination mechanisms during use in high power EUV exposure tools. Prior technologies dating back as far as i-line entered production with the assumption that pellicles would provide adequate protection for reticles over many thousands of exposures. Only later did the industry recognize the occurrence of progressive defects (haze or crystal growth, as well as outgassing from pellicle frames) and the impact on reticle quality.

EUV faces a similar learning curve. Even if pellicles are adopted, the frame is not sealed to the reticle surface, and the environment within the scanner is far from a perfect vacuum. Concerns over hydrocarbon deposition and reaction with intense EUV photons as well as with the out-of-band DUV present in the system, will require the development of careful monitoring and re-qualification plans. Reticle requalification cycles will be gated not just by the number of wafers exposed, but by the number of times a reticle is loaded and unloaded from the scanner and how long it sits in storage between cycles. Inspection strategies must take into account the difference between long exposure "trains" – reticles which expose thousands of wafers between changes – and short "trains", where a reticle may be used for only a few wafers before being placed back in storage. We anticipate that a combination of wafer-based and reticle-based inspection will be required to fully ensure reticle quality, especially if a pellicle solution is adopted which does not allow 193 nm based inspection.

The many tradeoffs and uncertainties need to be discussed in the context of a full, mask blank to wafer fab reticle qualification strategy for EUV volume manufacturing. One opportunity for such discussion was the recent edition of Photomask Japan, another one, the BACUS Photomask Symposium this coming September in Monterey. We hope to see you there!

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 Farlow

BACUS Technical Group Manager Marilyn Gorsuch

■ **2019 BACUS Steering Committee** ■

President
Peter D. Buck, *Mentor Graphics Corp.*

Vice-President
Emily E. Gallagher, *imec*

Secretary
Kent Nakagawa, *Toppan Photomasks, Inc.*

Newsletter Editor
Artur Balasinski, *Cypress Semiconductor Corp.*

2019 Annual Photomask Conference Chairs
Jed Rankin, *GLOBALFOUNDERIES Inc.*
Moshe Preil, *KLA-Tencor Corp.*

International Chair
Uwe F. W. Behringer, *UBC Microelectronics*

Education Chair
Frank E. Abboud, *Intel Corp.*

Members at Large
Michael D. Archuletta, *RAVE LLC*
Brian Cha, *Samsung Electronics Co., Ltd.*
Derren Dunn, *IBM Corp.*
Thomas B. Faure, *GLOBALFOUNDRIES Inc.*
Aki Fujimura, *DS2, Inc.*
Brian J. Grenon, *Grenon Consulting*
Jon Haines, *Micron Technology Inc.*
Naoya Hayashi, *Dai Nippon Printing Co., Ltd.*
Bryan S. Kasprowicz, *Photronics, Inc.*
Patrick M. Martin, *Applied Materials, Inc.*
Jan Hendrik Peters, *bmbg consult*
Stephen P. Renwick, *Nikon Research Corp. of America*
Douglas J. Resnick, *Canon Nanotechnologies, Inc.*
Thomas Scheruebl, *Carl Zeiss SMT GmbH*
Thomas Struck, *Infineon Technologies AG*
Bala Thumma, *Synopsys, Inc.*
Anthony Vacca, *Automated Visual Inspection*
Michael Watt, *Shin-Etsu MicroSi Inc.*
Larry Zurbrick, *Keysight Technologies, Inc.*

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

©2019

All rights reserved.

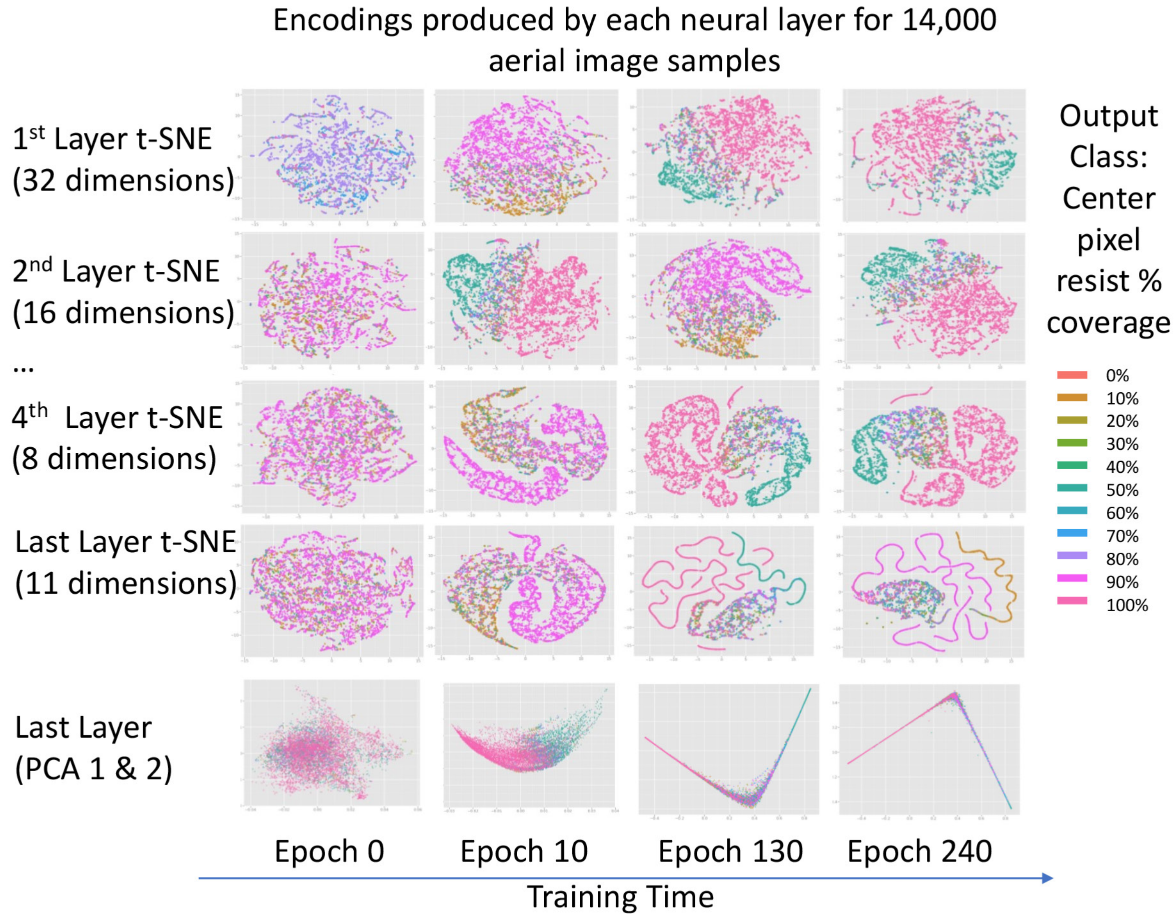

Figure 2. From left to right, we observe a scatter plot of the t-SNE projection as it develops during the course of training. From top to bottom we observe the progression from the first layer and downward to the final neuron layer. At epoch zero for all layers we initially see very little organization in the encodings. Samples with different classes are scrambled with each other. As training proceeds, the encodings from each layer evolve very differently from each other. The first layer, near the input, has compressed the image from 729 dimensions to 32. The compression scheme needs to preserve as much information about the input as is necessary to enable the subsequent layers to refine the classification of a sample. The second and third layers compress to 16 and then 8 dimensions, respectively, and begin to develop more effective separation between the classes. The encoding provided by the final layer is able to represent many of the points on nearly one-dimensional manifolds, but there remains confusion about how to distinguish the classes of many of the samples with pixel coverage between 30 and 70%. The encoding from the final 8-dimensional layer has become so simple that it can be rendered well with the first two principle components, as seen on the bottom four plots. By epoch 240, the encoding has organized most of the samples along a simple bent line with class confusion concentrated at the apex.

Application of DNNs to real world problems requires assessment of their limitations as well as their capabilities. A person who is excited about the idea of self-driving cars may also be cautious about riding one. The benefits of offered by a DNN must be compared to the costs encountered when it fails. In semiconductor manufacturing a failure could be as costly as a bad photo-mask, bad silicon, or a delay in time-to-market. It is easy to foresee risks in the range of millions of dollars.

The chief risk is that a DNN performs well during training, responds poorly when encountering a situation beyond its prior experience. Such a model is said to "generalize poorly." The most surprising fact about DNNs is how well they can generalize. They generalize so well that they exceed our understanding.[16] at the same time they can be easily fooled[17] Without a clear theory for why they generalize so well, it becomes difficult to determine the boundaries within which they can be trusted.

Despite these uncertainties, the potential benefits of neural networks should not be overlooked. Our intention is to use DNNs where beneficial. However we believe that this requires a methodology for assessing when they can be trusted.

This paper we will briefly recount the recent history of DL and DNNs, and two kinds of methods to help humans interpret them : information bottleneck (IB), and t-SNE embeddings. We will conduct an experiment, applying IB and t-SNE to training a neural network model of a photo-resist process. Visualizations will show what the network is doing while it learns. Through these visualizations we hope to make a contribution that will help the reader understand how DNNs work, and in particular in the context of problems of lithographic interest.

2. Related Work

After a long period of skepticism concerning the practicality of training DNNs, confidence has soared since the 2012 AlexNet paper[14] (which currently has more than 27,000 citations). Since then, papers have demonstrated impressive performance of classification of images, translation of literature, voice recognition, and other tasks.

Paradoxically, at the same time that our confidence in the predictions of neural networks is growing, we lack a firm understanding of why some models generalize and others do not. Machine learning is essentially an inductive process. Classically a model is probably better

if it explains the available data with fewer free parameters. According Vapnik-Chervonenkis statistical learning theory[18] the error bounds on a model's prediction are proportional to the log of the \hypothesis space" of the model. A model with a larger hypothesis space has more capacity to learn. But, adding new parameters very rapidly explodes the possible number of parameter combinations and therefore the size of the hypothesis space. This should rapidly expand the error bounds of the model, leading to models that appear accurate but are badly over-t. The original AlexNet included more than 62 million free parameters,[14] trained on a set of just over a million images. Its generalization performance exceeded expectations. Several papers followed AlexNet with significantly more parameters and showed strong generalization.

DNNs have enough model capacity to memorize all of the data encountered during training. A recent paper[16] demonstrated that popular neural networks architectures feature in recent publications can be trained to "t" random data. Such a network cannot be said to have learned anything. Yet for unknown reasons, if the data is not random and includes a signal rather than pure noise, trained neural networks perform surprisingly well.[16, 19] No single theory has emerged to convince all experts why this is so, although it is an active topic of research.

Although it is not universally accepted, and may not apply to all neural networks, the Information Bottleneck theory (IB)[20] can be useful for quantifying the goals and actions of the hidden layers of a DNN. We also find that IB can be combined with the t-SNE method, to allow us to visualize what a neural network is doing during training and inference. The analysis of the layers according to IB, visualized using t-SNE, will be the content of this paper.

2.1 Information Bottleneck

We will briefly review IB theory and then illustrate its application to the task of modeling the lithographic formation of a structure in photo-resist.

To understand IB we start with a simple description of the supervised learning problem. Suppose we have a group of aerial images of sampled from specific locations on a wafer, and a collection of resist contours extracted from SEM images or simulations. We consider our aerial image data a collection of samples x_n of a random variable X, drawn from a real generator of aerial images according to a probability distribution $P(X)$. Likewise the samples of the photoresist contours resulting from each input sample x_n are samples y_n of a random variable Y with probability distribution $P(Y)$.

A resist model is a function $f(x_n)$ that returns a value $\hat{y}_n$. The model f is parameterized by values $\theta \subset \Theta$. The θ parameters are learned during a training procedure, utilizing the collection of training data pairs (x_n; p_n). The x values are passed into the model, and the resulting predictions $\hat{y}$ are compared to the true answer y. The different between the correct answer and the prediction is evaluated according to some loss function L. The training process finds the model parameter values θ, resulting in a model $f(x;\theta)$ that can reliably predict $\hat{y}$ close to y and minimize the loss function L. In a good model, predicted values of $\hat{y}_n$ will closely approximate the observed values of y_n. If we withhold some of the samples during the training process, we can evaluate how well the network generalizes to new data not encountered during training.

According to IB, the network generalizes because it learns to discriminate between more or less valuable data in the input, based on how determinative it is in relation to classifying the output. The IB method suggests that a neural network generalizes well when it "compresses" the information content of the input X as much as it can without losing the information it needs to predict Y. The reason that compression is important is that it reduces the number of dimensions needed for representation. If the dimensionality required for representation of all of the anticipated input samples X can be compressed, then the problem of mapping from X to Y is greatly simplified and the size of the data set required to learn this mapping could be reduced. If X is compressed too much, we lose the information we need to classify the sample's class membership in Y.

The network learns what it can forget, or even what it should ignore, about the input, subject however to the requirement that it can still accurately predict the output class. Forgetting some of the input information is what helps enable generalization : the earlier network layers are forcing the later layers to do the classification job with less. One consequence is that the dimensionality of the problem is reduced.

The success of the network will be determined by the final layer. We call the inputs to the final layer a new random variable Z. We want the dimensionality of Z to be small, so that there are relatively fewer parameters to be optimized. On the other hand, if the input is not adequately represented then the final layer will lack the statistical power to accurately decide the classes. Mathematically this is equivalent to the condition of minimizing $H(Z)$ subject to the constraint that we also maximize $P(Y|Z)$. By this account, during the training procedure we are learning the parameters that define the encoding of X in terms of the encoding Z, or $P(Z|X)$, as well as the free parameters of the decoder $P(Y|Z)$.[21]

More generally, IB suggests that a DNN possessing multiple hidden layers can be considered a sequence of encoder/decoder pairs. Each layer is decoding the results of the previous layer and then reorganizing them into a new encoding that it thinks will be useful for the task of the next layer. The layers only directly communicate in the forward direction when predictions are made. However during training layers communicate backward to the preceding layer to inform it how effectively it is doing its job, so that it can adjust and perform better in the future.

2.2 t-SNE

The t-Distributed Stochastic Neighbor Embedding (t-SNE) method[22] converts high-dimensional data sets into two-or three-dimensional data sets that can be visualized as scatter-plots. We may consider structure in a data set to be the tendency of certain samples to cohere in high-dimensional space with nearby similar samples. The t-SNE method reduces the dimensionality of the data set, while preserving the high-dimensional structure. It is similar to the Principal Component Analysis (PCA) dimensionality reduction approach, but is not restricted to linear distance mappings between samples, and tends to be more effective at keeping dis-similar points far apart rather than grouping similar points close together. Intuitively, what t-SNE is doing is similar to representing a globe as a two-dimensional printed map, except that the globe may be of much higher than three dimensions, and the mapping from high- to low-dimension may be non-linear.

For our application, t-SNE provides us a way to visualize the encodings that are the outputs of each layer in the deep neural network. Each neuron in the encoding is a dimension in the encoding-space. Any given input sample will be located at some point in the high-dimensional encoding space, and t-SNE provides us a way to see where that sample ends up on the encoding-map.

3. Approach

Our approach will be to combine the encoder-decoder and information-theoretic ideas from IB, and the low-dimensionality representation of high-dimensional data provided by t-SNE. This will enable is to present, within the limits of a two-dimensional manuscript, the evolution of the neural network during the training process.

We propose a simple resist model. The input to the model is the aerial image. Due to the band-limiting characteristics of the lens aperture, we can represent this signal without information loss on a finite grid with sufficient sampling of 20nm. We also assume that the influence range of the aerial image upon the resist format is finite. For computational convenience we choose an aerial image surrounding each resist pixel that is 27x27 pixels. The aerial image sampling pitch is 20nm so the total influence range is 540nm on a side. The output of the model is the fraction of a small region in the center of the image that is "inside" the photoresist after all of the resist processing steps. For example a value of 0 for a pixel indicates that a box surrounding the pixel is completely free of resist after development, and a value of 1 is completely covered. Intermediate values are predicted with a relatively granular classification drawn from the set f0.1 0.2 0.3 0.4 0.5

0.6 0.7 0.8 0.9g. In principle we can make the classification arbitrarily ne, but to simplify color visualization we retain only 11 output classes.

The mathematical form of the model is a multilayer perceptron (MLP). To classify a single pixel of the photoresist, the MLP receives input from the local 27x27 pixel surrounding aerial image. An MLP consists of an input layer, one or more hidden layers, and an output layer. Each layer consists of a set of neurons, each connected to all of the outputs of the preceding layer. Neurons activate based on the collective weighted contribution of their inputs, and according to a non-linear response function. We use the sigmoid activation, although it should be mentioned that most state-of-the-art neural networks use rectified linear units (ReLUs). Also, as a side note, state of the art neural network performance is often achieved using a different model architecture based on the Convolutional Neural Network. For the purpose of this paper, a simple sigmoid-activated MLP retains the important functional characteristics that illustrate how neural networks function. The techniques can be easily extended to other activations and layered architectures. The model is illustrated in Fig.1.

Following the path suggested by the "Information Bottleneck" (IB) theory, we now examine each layer of the network as a decoder and encoder. The output of each layer is a new encoding of the inputs received from the preceding layer. The mechanism for the encoding is simple linear algebra in the form of a matrix of weights that captures the strength of interaction between each input activation and each neuron in the current layer. The weights are learned during a training process which follows the back-propagation method. Mispredictions of the output class during training motivate adjustments in the weights of the network, in a proportion determined by the application of recursive application of the chain rule through each layer in the network. We limit our analysis to look only at the neuron activations of each of the hidden layers.

If the IB theory is correct then each layer in the neural network is filtering out some information about x unnecessary for the task of predicting y. Thus at the input we should have a larger information content to represent the greater range of possible aerial image configurations, and the information content should diminish, layer by layer. As the network learns to throw out information that is not necessary for its task, the task itself is simplified. We will test in the next section whether this is the case for our resist neural network.

4. Results

We used a training set of 14,000 aerial image sample points, and their associated resist coverage classifications as extracted from simulated resist contours, to train an MLP resist model. The input aerial image was represented as a 27x27 matrix of oats, sampled at 20nm. The output Y was a classification representing the fractional coverage of the vicinity of the pixel, in steps of 0.1, from a completely uncovered 0 to a completely covered 1. The MLP included 4 hidden layers and a softmax output layer for the final classier. Neuron activations were sigmoids. The model was trained by stochastic gradient descent using the Adam optimizer in TensorFlow.

Fig.2 shows the encodings produced by the activations of the first layer. Each of one of the 14,000 aerial images is represented by a single not with a position on the scatter plot, and the target output value is encoded by color. The placement of the samples is randomly scattered in iteration 0 of the training process (called "Epoch 0" in the DL literature). As training proceeds, inputs that should be classified similarly are clustered close to other similar inputs. The first layer compresses the original 27x27 image into 32 neuron activation values, compressing the data more than 95%. Clusters are still extended as clouds in the first layer in order to preserve information that can distinguish similar inputs. Further compression is achieved in the second layer which reduces the representational space for the input to 16 activations. The third layer, not shown, compresses the input again, down to 8 activations. By layer 4 the data is now represented in 8-dimensional space and similar samples, with regard to output class, are pushed close to each other while pushed apart from samples of other classes. We can see that although this layer started off scattered randomly in Epoch 0 it has started to develop rich structure, and is organizing the samples into subsets of the entire t-SNE space and within well-defined clusters or manifolds.

According to IB theory, we expect the final layer prior to classification to retain the mutual information $I(X,Y)$ sufficient to classify the output. The t-SNE representation shows this network partially succeeds and partially fails. Some of the classes are well-defined, and show up as ribbons in t-SNE space. These ribbons are effectively some kind of line traceable in 8-d encoding space and then smashed down into 2D by t-SNE. A confusion region is also clear, where samples with different classes remain indistinguishable from each other. We would not expect to have high confidence about the networks classifications of samples with encodings that project into the confusion region.

In the PCA representation of the encodings in the final layer, this lower confidence region can be localized near the apex of a bent line along which every sample can be represented. This scatter plot illustrates the simplicity of representation arrived at by the neural network during the training process, as it ranked each 27x27 input image sample point with a position along the nearly one-dimensional line.

5. Discussion

Visualizing the encodings of each layer and interpreting them in light of information theory helps us understand what the network is doing when it accurately predicts a result for a sample, and also helps us diagnose causes of Mispredictions.

For each sample, we can trace back through the network to examine its encoding in each layer. If a sample is predicted correctly, in early layers the encoding on the t-SNE map should be well-separated from samples that have different classes, and it should find other similarly-classified samples in its vicinity. Conversely, for a Mispredictions, the representation of the sample in the early layers may not be able to distinguish it from other dissimilar samples, if necessary information has been forgotten in the encoding, or the information may be missing in the raw input to make an accurate classification. On the other hand, if the sample is encoded at some layer of the neural network in such a way that it appears isolated from other samples on the t-SNE map, this may indicate that the model has not been trained on similar data and the encoding may be less trustworthy. More training data may be required, or the training data must be evaluated to ensure that it is more representative of the data that will really be encountered during testing or production. Alternately, the sample may be encoded in a t-SNE region that is well-represented by other samples, but those other samples are uniformly of a different class. This sample should be examined to ensure that the expected class is correct, for it is possible that it has been mislabeled. If the sample label is confirmed, we can examine the encoding for the sample in each layer and see if the sample seems to be encoded in such a way that it is not clustered near other samples of the same class.

6. Conclusion

Neural networks are extremely competent at fitting sampled data during the training process, and surprisingly effective at generalization to correctly classify new samples. They have a large enough representational capacity to effectively "memorize" the training data set, this does not appear to be what they actually do. Currently we lack a definitive explanation of how neural networks generalize, but we have applied some promising theories to microlithography problems and gained insight. We applied the Information Bottleneck theory to interpreting the training process for a resist model. Our demonstration illustrated that the network may generalize because successive layers learn which aspects of the input they can "forget" while still accurately predicting the output. Forgetting unnecessary information enables the network to reduce the dimensionality of the problem, preventing rote memorization of the inputs. Layers are coaxed, during training, into efficiently decoding their inputs and encoding them for the next

layer. We also applied the t-SNE method to visualize the encodings of each layer as a scatter plot, layer-by-layer and epoch-by-epoch. This provided some insight into why the networks work well for some samples, and how the network could break down for other samples. The resulting plots can be interpreted as a story, where over time each layer passes on to the next as implied representation of its inputs. The simplification is subject to the constraint that the final layer can still perform an accurate classification.

7. Acknowledgments

Robert Lugg assisted me in generating the aerial image and resist dataset used for illustrating the neural network training process.

8. References

[1] Cummings, K. D., Frye, R. C., and Rietman, E. A., “Using a neural network to proximity correct patterns written with a cambridge electron ebeam microfabricator 10.5 lithography system,” *Applied Phys. Lett. 57,* 1431 (1990).

[2] Frye, R. C., Cummings, K. D., and Rietman, E. A., “Proximity effect corrections in electron beam lithography using a neural network,” *IEEE Int. Conf. on Systems, Man, and Cybernetics* (1990).

[3] Zach, F. X., “Neural-network-based approach to resist modeling and opc,” **Proc. SPIE 5377** (2004).

[4] Striniman, J. P. and Rieger, M. L., “Fast proximity correction with zone sampling,” **Proc. SPIE 2197** (1994).

[5] Cobb, N., “Fast optical and process proximity correction algorithms for integrated circuit manufacturing (dissertation),” (1998).

[6] Mardiris, V., Karafyllidis, I., Soudris, D., and Thanakilakis, A., “Neural networks for the simulation of photoresist exposure process in integrated circuit fabrication,” *Modeling and Simulation in Materials Science and Engineering 5* (1997).

[7] Nakayamada, N., Nishimura, R., Miura, S., Nomura, H., and Kamikubo, T., “Electron beam lithographic modeling assisted by artificial intelligence technology,” **Proc. SPIE 10454** (2017).

[8] Nakayamada, N., Kamikubo, T., Anze, H., and Ogasawara, M., “Modeling of resist surface charging effect on ebm-8000 and its comparison with ebm-6000,” **Proc. SPIE 8701** (2013).

[9] Shim, S. and Shin, Y., “Etch proximity correction through machine learning driven etch model,” **Proc. SPIE. 9782** (2016).

[10] Shim, S., Choi, S., and Shin, Y., “Machine leaning-based resist 3d model,” **10147** (2017).

[11] Watanabe, Y., Kimura, T., Matsunama, T., and Nojima, S., “Accurate lithography simulation model based on convolutional neural networks,” **10147** (2017).

[12] Lan, S., Liu, J., Wang, Y., Zhao, K., and Li, J., “Deep learning assisted fast mask optimization,” **10587** (2018).

[13] Agudelo, V., Fühner, T., Erdmann, A., and Evanschitzky, P., “Application of artificial neural networks to compact mask models in optical lithography simulation,” **Proc. SPIE. 8683** (2013).

[14] A. Krizhevsky, I. S. and Hinton, G., “Imagenet classication with deep convolutional neural networks,” *Advances in Neural Information Processing 25* (2012).

[15] Goodfellow, I., Bengio, Y., and Courville, A., [*Deep Learning*], The MIT Press, Cambridge, Mass. (2016).

[16] Zhang, C., Bengio, S., Hardt, M., Recht, B., and Yinyals, O., “Understanding deep learning requires rethinking generalization,” *International Conf. on Learning Rep. (ICLR)* (2017).

[17] Nguyen, A., Yosinski, J., and Clune, J., “Deep neural networks are easily fooled : high confidence predictions for unrecognizable images,” *Computer Vision and Pattern Recognition 15,* 368-377 (2015).

[18] Vapnik, V., [*Statistical Learning Theory*], Wiley Interscience, Hoboken, N.J. (1998).

[19] Neyshabur, B., Bhojanapalli, S., McAllester, D., and Srebro, N., “Exploring generalization in deep learning,” *Conf. on Neural Info. Proc. Systems (NIPS)* (2017).

[20] Tishby, N., Pereira, F., and Bialek, W., “The information bottleneck method,” *Allerton Conference on Communication, Control and Computing 37,* 368-377 (1999).

[21] Schwartz-Ziv, R. and Tishby, N., “Opening the black box of deep neural networks via information,” in [*Why and when deep learning works: looking inside deep learning*], Skilling, J., ed., 53-71, Intel Collaborative Research Institute for Computational Intelligence, Dordrecht (1989).

[22] van der Maaten, L. and Hinton, G., “Visualizing data using t-sne,” *Journal of Machine Learning Research 9,* 2579-2605 (2008).

N • E • W • S

Sponsorship Opportunities

Sign up now for the best sponsorship opportunities

Photomask Technology + EUV Lithography 2019

Contact: Melissa Farlow,
Tel: +1 360 685 5596; **melissaf@spie.org**

Advanced Lithography 2020

Contact: Teresa Roles-Meier,
Tel: +1 360 685 5445; **teresar@spie.org**

Advertise in the BACUS News!

The BACUS Newsletter is the premier publication serving the photomask industry. For information on how to advertise, contact:

Melissa Farlow,
Tel: +1 360 685 5596
melissaf@spie.org

BACUS Corporate Members

Acuphase Inc.
American Coating Technologies LLC
AMETEK Precitech, Inc.
Berliner Glas KGaA Herbert Kubatz GmbH & Co.
FUJIFILM Electronic Materials U.S.A., Inc.
Gudeng Precision Industrial Co., Ltd.
Halocarbon Products
HamaTech APE GmbH & Co. KG
Hitachi High Technologies America, Inc.
JEOL USA Inc.
Mentor Graphics Corp.
Molecular Imprints, Inc.
Panavision Federal Systems, LLC
Profilocolore Srl
Raytheon ELCAN Optical Technologies
XYALIS

Industry Briefs

TSMC Makes Progress on 5nm with Complete Infrastructure Design and Risk Production

By **Eric Hamilton,** TECHSPOT

TSMC announced that it has completed infrastructure design on its 5nm process node that will leverage the company's second generation of extreme ultraviolet (EUV), as well as deep ultraviolet (DUV) lithography. TSMC's 5nm chips will be aimed at SoC designs, 5G mobile applications, AI, and high performance computing.

According to early numbers on an Arm Cortex-A72 core, TSMC's 5nm process will deliver 1.8 times the density and a 15 percent gain in clock speeds compared to 7nm, and that's based on process refinements alone. TSMC also notes that its second generation of EUV will both simplify the manufacturing process and present excellent yield learning, allowing the process to mature more quickly.

https://www.techspot.com/news/79541-tsmc-makes-progress-5nm-complete-infrastructure-design-risk.html

SK Hynix Plans to Spend $107 Billion Building Four Memory Chip Plants

By **Heekyong Yang and Ju-min Park,** Reuters

South Korea's SK Hynix Inc. said it would spend $107 billion building four factories, as the memory chip maker seeks to maintain its competitiveness in the face of Chinese efforts to become a leading chipmaking nation.

The chip fabrication plants will be built on a 4.5 million square meter site south of Seoul beginning 2022, complementing two existing domestic factories that will receive a separate 55 trillion won ($49 billion) investment over the next decade.

The plans for the factories, producing DRAM and next-generation chips, come as chipmakers prepare for a surge in demand to power new technology such as fifth-generation (5G) communication networks and artificial intelligence, even as a slowdown in smartphone sales kills off a two-year chip boom.

https://www.reuters.com/article/us-sk-hynix-investment/sk-hynix-plans-to-spend-107-billion-building-four-memory-chip-plants-idUSKCN1QA073

Semiconductor Equipment Revenues to Drop 17% in 2019 on 29% Capex Spend Cuts

By **Robert Castellano,** The Information Network

The semiconductor equipment market grew 37.3% in 2017 on the heels of capex spend by memory companies in order to increase bit capacity and move to more sophisticated products with smaller nanometer dimensions. Unfortunately, these companies overspent resulting in excessive oversupply of memory chips. As memory prices started dropping, these companies put a halt to capex spend, and global equipment revenues increased only 13.9% in 2018. As excess inventory continues to increase in 2019, capex spend by these companies is projected to drop 29%, which will result in a significant reduction in equipment revenues in 2019.

https://seekingalpha.com/article/4251198-semiconductor-equipment-revenues-drop-17-percent-2019-29-percent-capex-spend-cuts

Bruker Announces Acquisition of RAVE LLC

PRNewswire

Bruker Corporation announced that it has acquired the semiconductor mask repair and cleaning business of RAVE LLC, a leading provider of nanomachining and laser photomask repair equipment. For calendar year 2018, the acquired business was profitable and had revenues of approximately $25 million. Financial details of the transaction were not disclosed, and the business has now become part of Bruker's semiconductor metrology division.

The acquired business, which will continue to be operated in Delray Beach, Florida, has built a strong reputation around its comprehensive portfolio of nanoprobe and laser-based photomask repair products, as well as CO2 cryo-cleaning technology for photomask and wafer applications. The acquisition adds to Bruker's leadership position in automated atomic force microscopy (AFM) for semiconductor photomask and wafer metrology. The combined offering uniquely positions Bruker's Semiconductor division to offer unmatched solutions for nanomachining mask repair and cleaning, as well as metrology for advanced nodes for EUV and multi-patterning lithography.

https://www.prnewswire.com/news-releases/bruker-announces-acquisition-of-rave-llc-300823469.html

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

2019

The 35th European Mask and Lithography Conference, EMLC 2019
17-19 June 2019
Hilton Hotel Dresden
Dresden, Germany

SPIE Photomask Technology + EUV Lithography
15-19 September 2019
Monterey Conference Center and
Monterey Marriott
Monterey, California, USA

2020

SPIE Advanced Lithography
23-27 February 2020
San Jose Marriott and
San Jose Convention Center
San Jose, California, USA

SPIE is the international society for optics and photonics, an educational not-for-profit organization founded in 1955 to advance light-based science, engineering, and technology. The Society serves nearly 264,000 constituents from 166 countries, offering conferences and their published proceedings, continuing education, books, journals, and the SPIE Digital Library in support of interdisciplinary information exchange, professional networking, and patent precedent. SPIE provided more than $4 million in support of education and outreach programs in 2018. **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

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

You are invited to submit events of interest for this calendar. Please send to lindad@spie.org; alternatively, email or fax to SPIE.