Microelectronics Reliability 88–90 (2018) 85–90

Contents lists available at ScienceDirect

Microelectronics Reliability

journal homepage: www.elsevier.com/locate/microrel

Gaussian process regression approach for robust design and yield enhancement of self-assembled nanostructures

Pham Luu Trung Duong[a], Xuechu Xu[b], Qing Yang[b], Nagarajan Raghavan[a,*]

[a] Engineering Product Development Pillar, Singapore University of Technology and Design, 487372, Singapore
[b] Department of Optical Engineering, Zhejiang University, Hangzhou 310027, China

ARTICLE INFO

Keywords:
Gaussian process regression
Robust design
Design of experiment
Nanostructures
Logistic regression

ABSTRACT

Self-assembled nanostructures are increasingly used for nanoelectronic and optoelectronic applications due to their high surface area to volume ratio and their ability to break traditional lithography limits. However, they suffer due to poor yield and repeatability as the growth process is often not well studied or optimized. Gaussian process regression (GPR) is a machine learning technique that can be used for both regression and classification purpose. In the GPR framework, a probability measure is defined according to one prior belief about the response surface and the Bayesian rule is applied to combine the observations with prior beliefs to form a posterior distribution of the response surface, which is known as the "surrogate model". We propose here the use of GPR as an effective statistical tool to optimize the growth conditions of nanostructures so as to improve their yield, controllability and repeatability ensuring at the same time that the yield is not affected by process variations at the identified optimum process conditions. In effect, we are proposing a design for reliability and robust design strategy for optimization of self-assembled nanostructure growth. We present here a case study of cadmium selenide nanostructures making use of an extensive design of experiment result (available open source) to illustrate the proposed methodology. The prediction accuracy of GPR is compared with two other commonly used statistical models → binomial and multinomial logistic regression. The use of the GPR method resulted in much better accuracy of probabilistic prediction of the different nanostructures with fewer fitting parameters than the logistic regression method.

1. Introduction

Considering the limitations posed by lithography and the cost involved in the installation and use of extreme UV and other state-of-the-art lithography tools, there is a growing desire in the micro and nano community to leverage on *self-assembly* as an approach to achieve high resolution nano-features where the patterns are guided by nature and thermodynamics or at best using templates that are of larger spatial dimensions than the feature sizes. The key problem with self-assembled nanostructures is the variability in their growth and the intrinsic randomness in the process that is difficult to understand, model and optimize using a purely physics-based perspective. The lack of controllability and repeatability makes *self-assembly* an unattractive option for silicon CMOS technology.

However, there are niche applications in optoelectronics and sensing, where self-assembled structures are very desirable and cheap. Some specific applications that benefit from self-assembly include luminescence, lasing, biomedical imaging, sensing etc. Although there is some leeway in the extent to which growth control of nanostructures is required for these applications, it is still important to strive and ensure optimum process conditions are used so as to have a manufacturing-friendly recipe that has higher yield and improved sensitivity. To this end, it becomes necessary to use statistical and machine learning models to better quantify the relationship between the process parameters and the output metrics for nanostructure growth so that an optimal set of process conditions could be chosen that ensures high yield and robustness in the achieved outcome even in the presence of inherent randomness in the process parameters. One such approach is Gaussian process regression (GPR). This study makes an attempt to use GPR as a stochastic surrogate model for modeling the dependencies between the yield probability of obtaining certain types of nanostructures and their corresponding growth conditions. The accuracy of GPR will be compared with the binomial and multinomial Generalized Linear Model (GLM) logistic regression techniques, which were recently used by Dasgupta *et al.* [1] for the same purpose.

For a case study, we will be focusing on cadmium selenide (CdSe)

* Corresponding author.
E-mail address: nagarajan@sutd.edu.sg (N. Raghavan).

https://doi.org/10.1016/j.microrel.2018.07.062
Received 30 May 2018; Accepted 5 July 2018

based nanostructures here. There has been an extensive design of experiment (DOE) done recently on CdSe-based nanostructures [1]. It is known that CdSe can exhibit different types of one dimensional morphology such as nanowires, nanobelts and nanosaws. The yield of these different morphology types strongly depends on the growth condition. This paper makes an attempt to use Gaussian process regression as a stochastic surrogate model for modeling the dependencies between the yield probability of nanostructures and their growth conditions. In this work, we will use GPR to model the link between the probabilities of obtaining the different specific types of CdSe nanostructures with three process parameters → temperature (*T*), pressure (*P*) and distance of the source from the substrate (*D*).

Gaussian process regression models, also known as Kriging, can approximate detailed mechanistic models and are capable of simplifying highly nonlinear and computationally expensive problems. They have been used for various machine learning and control tasks, such as dimensionality reduction [2], nonlinear classification, system identification, and regression [3]. Rather than defining a parametric function for an input-output relationship as in the standard regression approach, the GPR puts a prior to infinitely many functions and then combines this prior belief with observation via the Bayes' rule to obtain posterior distribution of functions, which ends up in a surrogate model. As a result, many complex model responses can be fitted by this non-parametric approach. In surrogate modeling, it is normal to expect that the input points that are close to each other should lead to a similar response value. These similarities are encoded in GPR using the *covariance* functions, which are also known as *kernel* functions. The introduction of the kernel requires different hyper-parameters to be optimized.

This paper is organized as follows. In Section 2, the details of the DOE for CdSe nanostructures are introduced. In Section 3, we present a brief overview of the Gaussian process regression technique. Section 4 presents the results of the GPR for modeling the probabilities of obtaining specific types of CdSe nanostructures for different growth (process) conditions. We also compare the accuracy of GPR with the binomial and multinomial Generalized Linear Model (GLM) logistic regression. Since the growth conditions cannot be controlled very precisely, the process parameters are considered as random variables and sampling is carried out using the GPR surrogate model through Monte Carlo simulations of the process parameter space to obtain the average nanostructure yield probabilities and their variances. The average yield probabilities are then optimized to find the growth conditions that would allow maximal growth of a particular nanostructure of interest with acceptable variance of the yield probabilities. Section 5 concludes the findings of the study in terms of the validity and effectiveness of GPR as a nanoscale yield modeling and optimization tool.

2. Design of Experiment for CdSe growth

As mentioned earlier, the growth of CdSe almost always gives a stochastic distribution of three different nanoscale structures → nanobelts, nanosaws and nanowires. Depending on the actual growth conditions (process parameters → *P*, *T* and *D*), the preferred type of nanostructure can be very different. An extensive DOE was carried out by Dasgupta *et al.* in Ref. [1] for CdSe. They carried out a 5 × 9 full factorial experiment with five levels of source temperature (630 °C, 700 °C, 750 °C, 800 °C, 850 °C) and nine levels of pressure (4, 100, 200, 300, 400, 500, 600, 700 and 800 mbar). For a specific combination of source temperature and pressure, 4–6 substrates were placed downstream of the source to grow the nanostructures. Three experimental runs were conducted at different locations along the furnace tube with each of the 45 combinations of temperature and pressure. The total number of substrates processed with CdSe growth for 135 (=45 × 3) runs was 415. However, for the source temperature of 850 °C, no morphology was observed in several experimental trials. Therefore, 67 experimental runs at this temperature were discarded. Considering the remaining 348 trial runs as an experimental test case, the design matrix is a 348 × 3 matrix with normalized process parameters for *T*, *P* and *D*. For each deposition process on the substrate, the SEM was used to count 180 individual nanostructures, which comprised of a mix of nanowires, nanobelts and nanosaws. The fraction of these structures obtained was considered as the output quantity of interest in this DOE.

3. Gaussian process regression (GPR) model

Gaussian process regression (GPR), just as any other meta-modeling approach, aims at approximating the response of a model/experiment given a finite set of observations. Within this context, let us consider a system, whose behavior we want to study, which can be represented by a response surface that maps *M*-dimensional inputs to the one dimensional output space. A GPR model is a generalization of the multivariate Gaussian random variable to an infinite dimension. It is described by the equation:

$$\mathcal{M}(\mathbf{x}) = h(\mathbf{x})^T\boldsymbol{\beta} + f(\mathbf{x}) \tag{1}$$

where $\boldsymbol{x}$ is the multi-dimensional input variable, M($\mathbf{x}$) is the model output,$f(\mathbf{x})$ is a zero mean Gaussian random process with a covariance function $R(\cdot;\boldsymbol{\theta})$, $h(\mathbf{x})$ is a vector of basis functions and $\boldsymbol{\beta}$ is a vector of coefficients. The term $h(\mathbf{x})^T\boldsymbol{\beta}$ is also called a *trend*. The hyper parameters, $\boldsymbol{\theta}$, associated with the covariance function, $R(\cdot;\boldsymbol{\theta})$, are unknowns that need to be estimated from the available observation data. The covariance function encodes the assumption about the response surface that is being approximated. Different covariance functions are available in the literature [4–6]. Given the trend and covariance function with unknown hyper parameters, it is possible to obtain an arbitrary number of realizations of the prior Gaussian process. The GPR framework uses the Bayes' rule to condition the prior on the available observation to make a prediction at a new point as:

$$y^{(i)} \sim N\left(h(\mathbf{x}^{(i)})^T\boldsymbol{\beta} + f(\mathbf{x}^{(i)}), \sigma^2\right) \tag{2}$$

where σ is the measurement noise and *N*() is the Gaussian distribution.

In this work, the popular squared exponential covariance function ($\boldsymbol{R}(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta})$) with automatic relevance determination mechanism (ARD) is utilized (Eq. (3)).

$$R(\mathbf{x}, \mathbf{x}'; \theta) = \sigma_f \exp\left(-\frac{1}{2}\sum_{l=1}^{M}(x_l - x_l')^2/\theta_l\right) \tag{3}$$

The vector $\boldsymbol{\theta} = (\sigma_f, \theta_1, \ldots, \theta_M)$ is a vector of hyper-parameters. Note that the ARD uses different length scales for different inputs to characterize the inputs that may have different impact on the output. It is normal to expect that similar inputs should lead to similar responses in a surrogate model. Hence, the covariance in Eq. (3) is used to characterize this similarity through the length scale, θ_i. The length scales characterize how far one needs to move along a particular input space to make the responses uncorrelated. This mechanism determines the relevance of an input which is inversely proportional to the length scale.

Therefore, the GPR model is completely defined by the coefficients vector and the hyper parameters. The σ_f in the hyper-parameter vector is used to characterize the degree of variation of the responses. Assuming that there are *P* observation data from the experiments $\{\mathbf{x}^{(i)}, y^{(i)}\}_{i=1}^{i=P}$, the optimal hyper-parameters, $\widehat{\theta}$, will be learnt from the observation data by maximizing the log-likelihood expression (*LKL*) given by:

$$LKL(\boldsymbol{\beta}, \theta) = \frac{1}{2}(\mathbf{y} - H\boldsymbol{\beta})^T[K(\mathbf{x}^{(i)}, \mathbf{x}'^{(i)}; \theta) + \sigma I_p](\mathbf{y} - H\boldsymbol{\beta}) - \frac{P}{2}\log(2\pi) - \frac{1}{2}\log|K(\mathbf{x}^{(i)}, \mathbf{x}'^{(i)}; \theta) + \sigma I_p| \tag{4}$$

where I_p is an identity matrix (size P), *H* is a matrix of size $P \times r$, *r* is the number of basis functions and *K* is the covariance matrix of the