Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio
©2023 ESREL2023 Organizers. *Published by* Research Publishing, Singapore.
doi: 10.3850/978-981-18-8071-1_P674-cd

Prediction of the Number of Defectives in a Production Batch of Semiconductor Devices

Ibrahim Ahmed

Energy Department, Politecnico di Milano, Milan, Italy. E-mail: ibrahim.ahmed@polimi.it

Piero Baraldi

Energy Department, Politecnico di Milano, Milan, Italy. E-mail: piero.baraldi@polimi.it

Enrico Zio

Energy Department, Politecnico di Milano, Milan, Italy. E-mail: enrico.zio@polimi.it
MINES Paris-PSL, CRC, Sophia Antipolis, France. E-mail: enrico.zio@mines-paristech.fr

Horst Lewitschnig

Infineon Technologies Austria AG, Villach, Austria. E-mail: horst.lewitschnig@infineon.com

The production of semiconductor devices requires to satisfy high reliability standards. For this reason, manufactured devices are subject to burn-in, i.e., extensive testing under accelerated stress conditions, which is costly and time-consuming. The present work develops a model for predicting the quality of the devices from data collected during the production process. The developed modelling approach is based on: *a*) a combination of Piecewise Aggregate Approximation (PAA) and Principal Component Analysis (PCA) for the extraction of features relevant for inferring the quality of the devices from signal measurements collected during the production; and *b*) a model based on Probabilistic Support Vector Regression (PSVR) for predicting the number of defective devices in the production batch. The model is validated on synthetic data, which emulate signal measurements collected during the production of semiconductor devices. The obtained results show that the proposed model is able of predicting the number of defective devices in a production batch with satisfactory accuracy.

Keywords: Semiconductor devices, Burn-in, Piecewise Aggregate Approximation, Principal Component Analysis, Probabilistic Support Vector Regression.

1. Introduction

In the semiconductor manufacturing industry, one of the main concerns of the manufacturers is to guarantee that the devices work when placed in service and do not fail at the beginning of their lives due to the presence of manufacturing defects (Kuo and Kuo, 1983). Burn-in (BI), which consists in testing the operation of some samples of the produced devices under accelerated stress conditions, such as high temperature and voltage, allows identifying the devices that are expected to fail during their early stages of life for removing them from the production and estimating the Early Life Failure Probability (ELFP) of the product population, i.e., the proportion of early failures, before the devices are delivered (Wood M. H., Muzik A, Huston H. H., 1993). Typically, the metric used is the upper one-sided confidence interval of the ELFP, computed by applying the Clopper-Pearson (CP) Estimator (Clopper and Pearson, 1934) to the number of BI-relevant failures. If the estimated value of ELFP is above a predefined target, given in terms of Parts Per Million (ppm), other devices of the production batches undergo BI until the desired quality level is guaranteed. BI can be very expensive and time-consuming to perform, particularly for new technologies, which require that all the produced devices undergo BI (100% BI) until the quality target is met. Considering that all semiconductor devices obtained from

tens of batches are BI-tested at the beginning of the production by a new process technology, BI cost and time can be very large (Hui and Lu, 1996). It is not surprising, then, that manufacturers are interested in reducing cost and time required for BI while maintaining the require quality target.

In this work, we exploit the possibility of using production data to predict the likelihood of early failures without performing BI. This is based on the observation that early-life failures are typically caused by defects that result from anomalies occurred during the manufacturing process (Baraldi et al., 2021). The data typically collected during the manufacturing process include: production and electric signals measured by sensors installed on the machines used for production, images of the devices taken during different production stages and results of electrical diagnostic tests. Specifically, the present work investigates the possibility of developing a model which, on the basis of data collected during production, estimates the number of defective semiconductor devices within a production batch.

To this aim, data-driven approaches for anomaly detection are considered. Their development requires to address the following challenges: *i*) dealing with multidimensional time-series containing the values of the signals measured during production or in electrical tests; and *ii*) estimating the model outcome uncertainty, which is caused by the intrinsic stochasticity of the number of defective devices in a production batch, the modelling error, and the signal measurement noise. To properly address these challenges, we develop a method that combines a feature extraction method based on Piecewise Aggregate Approximation (PAA) (Keogh et al., 2001) and Principal Component Analysis (PCA) with Probabilistic Support Vector Regression (PSVR) (Liu et al., 2013; Sollich, 1999a). The feature extraction method aims at extracting a reduced set of features correlated with the quality of the devices from the multidimensional time-series (Challenge *i*). The idea is to project the original data into a space of reduced dimensionality, where the number of time measurements associated to the production of a batch is reduced. This is obtained by applying PAA and, then, selecting the Principal Components (PCs) representing the largest variability of the data. With respect to Challenge *ii*), the choice of resorting to PSVR for the estimation of the number of defective devices in the production batch is motivated by its capability of properly representing the uncertainty associated to the model outcome.

The proposed method is validated on simulated data, which synthetically emulate real BI process data. The use of synthetic data is motivated by the need of verifying the performance of the method. A comparison is also made with the classical Support Vector Regression (SVR).

The rest of the paper is organized as follows. In Section 2, the problem of estimating the number of defective semiconductor devices within a production batch is formulated. In Section 3, the proposed method is presented and its application to a synthetic case study is given in Section 4. In Section 5, the results of the application of the developed method are presented and discussed. Finally, Section 6 presents the concluding remarks.

2. Problem Formulation

A set of S signals are measured during the semiconductor manufacturing process. They include physical quantities measured during the production process and electrical signals measured during diagnostic tests. These quantities are expected to contain information related to the quality of the production.

For simplicity of notation, we assume that the production of the generic batch l starts at time 0 and ends at time $T(l)$, and that all measurements are synchronously acquired every one time unit.

Since the production is performed on batch basis and the production cycle is based on several production stages (Fisher et al., 2012), a single batch is associated to the time series of the physical quantities measured during its production. The measurements taken at the generic time t (i.e., the present calendar time) are represented by the vector:

$$\mathbf{x}_t = (x_{t,1} \quad \cdots \quad x_{t,j} \quad \cdots \quad x_{t,S}) \qquad (1)$$

where the vector element $x_{t,j}$ indicates the measurements of the physical quantity j at time t. We assume the availability of a dataset $\mathcal{D}_{train} = [\boldsymbol{X}(l), n(l), y(l)]_{l=1 \cdots L_{train}}$ containing the

following data collected during the production of L_{train} batches:

- the matrix $\boldsymbol{X}(l) \in \mathbb{R}^{T(l)\times S}$ of the measurements collected during the production of batch l:

$$\boldsymbol{X}(l) = \begin{bmatrix} \mathbf{x}_1(l) \\ \mathbf{x}_2(l) \\ \cdots \\ \mathbf{x}_{T_l}(l) \end{bmatrix} = \begin{bmatrix} x_{11}(l) & x_{12}(l) & \cdots & x_{1S}(l) \\ x_{21}(l) & x_{22}(l) & \cdots & x_{2S}(l) \\ \vdots & \vdots & \ddots & \vdots \\ x_{T_l1}(l) & x_{T_l2}(l) & \cdots & x_{T_lS}(l) \end{bmatrix} \quad (2)$$

- the number, $n(l)$, of devices of the batch that have been randomly selected and have undertaken the BI test;
- the number, $0 \leq y(l) \leq n(l)$, of devices that have had BI-relevant failures.

Considering a test batch, which has not yet undertaken BI tests, and the signal measurements, $\boldsymbol{X}_{test}$, collected during its production, the objective of the present work is to predict the number of defective devices, $\hat{y}_{test}$, which are expected to have BI-relevant failures if n_{test} devices of the test batch are to undergo BI.

3. Method

The proposed method for the prediction of the number of defective devices is depicted in Fig. 1. It consists in a feature extraction method and a Probabilistic Support Vector Regression (PSVR) algorithm for the estimation of the number of defective semiconductor devices.

3.1. Feature extraction

The training data, $\boldsymbol{X}(l)$ with $l = 1, \cdots, L_{train}$, and test data, $\boldsymbol{X}_{test}$, are preprocessed to extract vectors of reduced dimensionality, $\mathbf{z}(l)$ and $\mathbf{z}_{test}$, respectively, containing features related with the quality of the production batch. To this aim, two sequential steps of feature extraction are performed: *1)* reduction by PAA of the number, $T(l)$, of time measurements associated to the production of the lth batch and *2)* reduction by PCA of the number, S, of the measured signals.

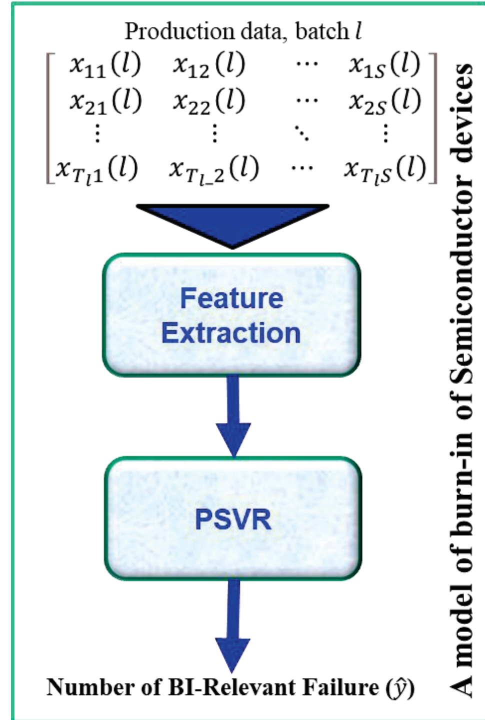

Fig. 1. Proposed method for the prediction of the number of defective semiconductor devices.

3.2. Probabilistic Support Vector Regression

One of the most effective formulation of Support Vector Machine (SVM) for regression is the ε-insensitive Support Vector Regression (ε-SVR) (Vapnik, 1995). When a limited number of training patterns is available, the principle of Structural Risk Minimization (SRM) is implemented within the ε-SVR implementation (Gao et al., 2002) to balance the trade-off between variance and bias. A probabilistic framework for ε-SVR (ε-PSVR) based on the Bayesian Evidence Framework is adopted in this paper. It is based on the interpretation of the SVR outcome as the maximum a posteriori (MAP) solution of the inference problem when Gaussian priors and an appropriate likelihood function are used (Gao et al., 2002). This probabilistic interpretation of the SVR enables the use of Bayesian methods to set the three main hyperparameters of the model, which are regularization parameter, C, sparsity control parameter, ε and kernel parameter, γ (Sollich, 1999b), and allows obtaining a prediction interval as outcome (Lin and Weng, 2004).

4. Case Study

The developed model has been validated considering a synthetic case study. The simulation for generating the data of the synthetic case study is based on the following assumptions:

i) the values of the signals measured during the production of a generic batch, $\boldsymbol{X}(l)$, depend on the quality of the production batch, represented by the ground truth ELFP, $p(l)$;
ii) the time, $T(l)$, needed to produce a batch is proportional to the number of BI-tested devices $n(l)$:
$$T(l) = \xi \cdot n(l). \tag{3}$$
We set the processing time required to produce a device, ξ, equal to $8 \times 10^{-3}\ s/device$;
iii) the batch-specific quality target level, p_{target}, for BI at 90% confidence level (CL) is 23 ppm.

According to the CP estimator and based on the above assumptions, it is possible to show that the minimum number of devices to be sampled and BI-tested to meet the batch-specific quality target if BI-relevant failures are not observed $(y(l) = 0)$ is 100150. Then, considering a generic *l*th production batch, the simulation to generate values of the quantities $\boldsymbol{X}(l)$, $n(l)$, $y(l)$ is based on the following procedure:

a. Sample the ground truth ELFP, $p(l)$, from a uniform probability distribution, $f_p(p)$, in the range [0, 23 ppm] for batches without BI relevant failures, or in the range [23 ppm, 40 ppm] for batches with BI relevant failures.
b. Sample the number $n(l)$ of BI-tested devices from a uniform discrete probability distribution, $f_N(n)$, in the range $[100150, 150000]$. The range is chosen so as to ensure that any production batch made by a random sample, $n(l)$, taken from $f_N(n)$ with $y(l) = 0$ BI-relevant failures produces an estimate of $p(l)$ smaller than the required quality target level.
c. The number of BI-relevant failures, $y(l)$, is binomially distributed:
$$y(l) \sim \boldsymbol{Bin}[n(l), p(l)]. \tag{4}$$
d. Generate the matrix $\boldsymbol{X}(l)$ of the signal measurements. We assume that $S = 8$ signals are measured for a time period of duration $T(l)$ obtained from Eq. (3). The matrixes of measurements, $\boldsymbol{X}(l)$, of size $[T(l), S]$, are obtained by simulating at each time step *t* the vector $\mathbf{x}_t(l)$ of the signal measurements using:
$$\mathbf{x}_t(l)^T = \mathbf{A}\boldsymbol{\beta}(t) + \boldsymbol{\vartheta} \tag{5}$$
where $\boldsymbol{A}$ is a matrix of size $[S, 3]$ representing the system process; $\boldsymbol{\beta}(t)$ is a column vector of length 3 representing independent input parameters correlated to the quality of the batch, such as raw materials or machine types, and controlled and uncontrolled inputs (e.g., tool/machine parameters, operator variability); and $\boldsymbol{\vartheta}$ is a column vector of length *S* representing the measurement noise:
$$\mathbf{A} = \begin{bmatrix} 0.9532 & 0.9321 & 0.0000 \\ 0.0000 & 0.2341 & 0.8643 \\ 0.9523 & 0.0000 & 0.9346 \\ 0.0164 & 0.0298 & 0.9453 \\ 0.7655 & 0.0000 & 0.0129 \\ 0.5473 & 0.4323 & 0.8120 \\ 0.0000 & 0.9842 & 0.8239 \\ 0.0000 & 0.0476 & 0.8788 \end{bmatrix}. \tag{6}$$
The elements of the matrix $\mathbf{A}$ are reported in Eq. (6), whereas the elements of the vector of the independent and identically distributed measurement noises are sampled from a Gaussian distribution with mean equal to zero and standard deviation equal to 0.1, i.e., $\boldsymbol{\vartheta} \sim \mathcal{N}(\mathbf{0}, 0.1\mathbf{I})$. The independent input parameters, $\boldsymbol{\beta}(t) = [\beta_1 \ \ \beta_2 \ \ \beta_3]^T$, are time-dependent quantities influenced by the quality of the production batch:
$$\boldsymbol{\beta}(t) = \boldsymbol{\beta}_{baseline}(t) + \vec{\boldsymbol{\zeta}}_{\tilde{s}}\delta p(l) \tag{7}$$
where the elements of the baseline, $\boldsymbol{\beta}_{baseline}(t)$, are:
$$\begin{gathered} \beta_1(t) = 2\cos(0.0799t)\sin(0.0723t) \\ \beta_2(t) = \sin(0.2991t) + 2\cos(0.1013t) \\ \beta_3(t) = U(0, 2) \end{gathered} \tag{8}$$
and the function $\vec{\boldsymbol{\zeta}}_{\tilde{s}}\delta p(l)$ is built by randomly sampling from a uniform discrete distribution in which one of the three process parameters, $\tilde{s} = 1, 2, 3$, is influenced by the batch quality ($\vec{\boldsymbol{\zeta}}_{\tilde{s}}$ is the $\tilde{s}$th column of the identity matrix) and, then, by adding to the sampled process parameter $\beta_{\tilde{s}}(t)$ a term proportional to $p(l)$ through a multiplication factor $\delta = 60000$.

The four steps procedure described above has been applied to generate a training dataset of $L_{train} = 1400$ training batches, $\mathcal{D}_{train} =$

$[\boldsymbol{X}(l), n(l), y(l)]_{l=1,\ldots,1400}$, containing 700 batches without BI relevant failures and 700 batches with BI relevant failures, and a test dataset made by $L_{test} = 600$ test batches, $\mathcal{D}_{test} = [\boldsymbol{X}(l), n(l), y(l)]_{l=1,\ldots,600}$, containing 300 batches without BI failures and 300 batches with BI failures. The training dataset (which has been divided into 70% and 30% for training and validation sets, respectively) has been used to develop the proposed model and perform the feature extraction phases; the test dataset has been used to assess the method performance.

5. Results and Discussions

The feature extraction step of Section 3.1 has been performed by setting the time series reduction length to 300 and the number of PCs to 4, which allows explaining 99.81% of the data variability. This way, the original time-series $\boldsymbol{X}(l)$ made by $T(l)$ values of 8 signals is projected into a vector $\mathbf{z}(l)$ of length 4. Then, the extracted features are used to train the PSVR model. Since the proposed method provides in output a real number, the predicted number of defective devices, $\hat{y}$, is obtained by approximation to the nearest integer.

The performances of the trained model have been computed on the test set. Fig. 2 shows the predicted versus groundtruth number of defective devices. Notice that one source of the prediction error, i.e., the variations between the predicted and groundtruth number of defective devices, is the stochasticity of the process due to the fact that the BI relevant failure follows a Binomial distribution. The overall rate of correct prediction $\left(\hat{y}(l) = y(l)\right)$ is 45% and the RMSE is close to 1.86. Fig. 3 shows the distribution of the prediction error $(y - \hat{y})$ in the test set for the batches without BI relevant failures (blue) and those with BI relevant failures (red). The rate of correct predictions of the number of defective devices of batches without BI failures is approximately four times larger than that of the batches with BI failures, and errors larger than 2 units occur only for batches with BI failures. A further analysis of the errors made on batches with BI failures has shown that they tend to occur on batches characterized by large values of $y(l)$. This is due to the fact that the PSVR model considers the similarity among the test batch and the selected support vectors extracted from the training batches, and, since the number of training data $[\boldsymbol{X}(l), y(l)]$ with large $y(l)$ is small (the fraction of batches having $y(l)$ more than 7 is 3.75%), the estimation is less accurate. Other possible sources of errors in the prediction of the number of defective devices are:

- the output of the model, i.e., the number of defective devices in a batch, is an integer random variable, whereas PSVR models provide in output continuous random variables;
- PSVR assumes that the uncertainty of the output can be modeled using a Gaussian variable, whereas one of the main sources of uncertainty of the model output is the number of defective devices in a batch which follows a binomial random variable.

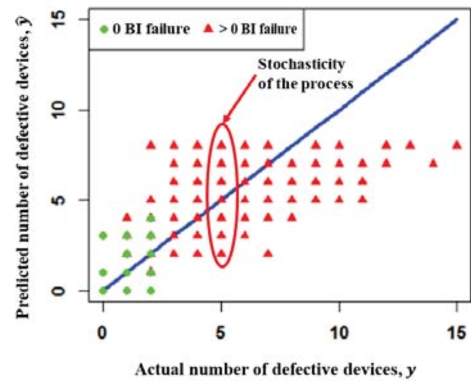

Fig. 2. Predicted versus groundtruth number of defective devices of the test batches.

The performance of the developed method has been compared to that of classical Support Vector Regression (SVR). Table 1 reports the values of root mean square error (RMSE) obtained by the two methods. It can be noticed that the proposed method outperforms the SVR method.

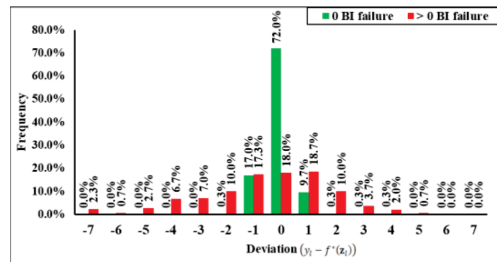

Fig. 3. Distribution of the error in the test batches without BI relevant failures (green) and with BI relevant failures (red).

Table 1. RMSE of the methods.

Model	RMSE
Support Vector Regression (SVR)	1.9391
Proposed method (PSVR)	**1.8586**

6. Conclusions

In this paper, we have proposed a method for the prediction of the number of defective semiconductor devices in a production batch. The proposed method starts with a step of feature extraction and selection, which is performed using PAA and PCA. In the second step, the extracted features are fed to a model based on PSVR for the prediction of the number of defective devices.

The method has been applied to synthetic data mimicking the statistics of a real burn-in process of semiconductor devices. The obtained results shows that the proposed method is capable of predicting the number of defective devices and outperforms another state-of-the-art method in terms of RMSE.

Acknowledgement

This work has been performed within the iRel40 European co-funded innovation project, granted by the ECSEL Joint Undertaking (JU) under grant agreement No 876659. The funding of the project comes from the Horizon 2020 research programme and participating countries. National funding is provided by Germany, including the Free States of Saxony and Thuringia, Austria, Belgium, Finland, France, Italy, the Netherlands, Slovakia, Spain, Sweden, and Turkey.

Disclaimer

The document reflects only the author's view and the JU is not responsible for any use that may be made of the information it contains.

References

Baraldi, P., Medici, S., Ahmed, I., Zio, E., Lewitschnig, H., 2021. A Method Based on Gaussian Process Regression for Modelling Burn-in of Semiconductor Devices, in: Proceedings Ofthe 31st European Safety and Reliability Conference. Angers, France, pp. 2619–2626.

Clopper, C.J., Pearson, E.S., 1934. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. Biometrika 26, 404–413.

Fisher, G., Seacrist, M.R., Standley, R.W., 2012. Silicon crystal growth and wafer technologies. Proc. IEEE 100, 1454–1474. https://doi.org/10.1109/JPROC.2012.2189786

Gao, J.B., Gunn, S.R., Harris, C.J., Brown, M., 2002. A probabilistic framework for SVM regression and error bar estimation. Mach. Learn. 46, 71–89.

Hui, Y. V., Lu, W.L., 1996. Cost optimization of accelerated burn-in. Int. J. Qual. Reliab. Manag. 13, 69–78.

Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S., 2001. Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases. Knowl. Inf. Syst. 3, 263–286.

Kuo, W., Kuo, Y., 1983. Facing the Headaches of Early Failures: A State-of-the-Art Review of Burn-In Decisions, in: Proceedings of the IEEE. pp. 1257–1266.

Lin, C., Weng, R.C., 2004. Simple Probabilistic Predictions for Support Vector Regression. National Taiwan University, Taipei.

Liu, J., Seraoui, R., Vitelli, V., Zio, E., 2013. Nuclear power plant components condition monitoring by probabilistic support vector machine. Ann. Nucl. Energy 56, 23–33.

Sollich, P., 1999a. Probabilistic methods for Support Vector Machines, in: Advances in Neural Information Processing Systems 12 (NIPS 1999).

Sollich, P., 1999b. Probabilistic interpretations and Bayesian methods for Support Vector Machines, in: IEE Conference Publication, Artificial Neural Networks. pp. 91–96. https://doi.org/10.1049/cp:19991090

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer, New York.

Wood M. H., Muzik A, Huston H. H., R.H., 1993. Burn-In, in: Microelectronics Manufacturing Diagnostics Handbook. Springer Science+Business Media, LLC.