Device Level Maverick Screening
Application of Independent Component Analysis in Semiconductor Industry

Anja Zernig[1,3], Olivia Bluder [1], Andre Kästner [2] and Jürgen Pilz[3]
[1]*Kompetenzzentrum für Automobil- und Industrieelektronik GmbH, Europastrasse 8, 9524 Villach, Austria*
[2]*Infineon Technologies Austria AG, Siemensstrasse 2, 9500 Villach, Austria*
[3]*Alpen-Adria University Klagenfurt, Institute of Statistics, Universitätsstrasse 65-67, 9020 Klagenfurt, Austria*

1 INTRODUCTION

In semiconductor industry the demand on functional and reliable devices, commonly known as chips, grows as they are more and more frequently used in safety relevant applications such as airbags, aircraft control and high-speed trains. The most delicate period for devices is their early lifetime, where failures represent a high risk as visualized by the bathtub curve in Figure 1.

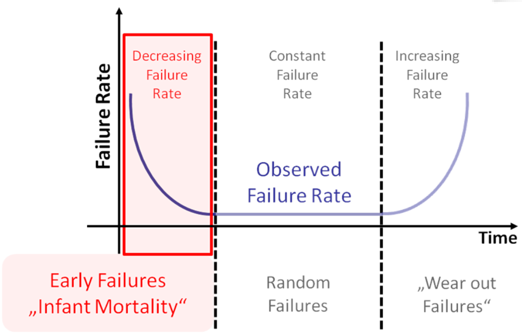

Figure 1: The bathtub curve describes the life stages of semiconductor devices.

The main reason for device failures during the early life stage comes from the manufacturing process. Most of these production failures can be detected with functionality tests, where mainly electrical connectivity is checked. Beside this, further tests regarding reliability of the devices, are performed. A generally accepted procedure for investigating reliability issues is the Burn-In (BI) test, where the devices are tested under accelerated stress conditions to simulate the early lifetime. Due to undesirable side effects of the BI, like high costs and the need for extra trained employees, a reduction of devices to be burned is preferable.

More cost-efficient methods are statistical screening methods which are capable of detecting potential early life failures. If a device is suspicious compared to the majority represents a risk device, so-called Maverick. Depending on the classification power of the screening method, detected Mavericks are immediately rejected or further investigated, e.g. with the BI. Due to the development towards sub-micron technologies, a distinction between reliable devices and Mavericks becomes increasingly challenging. Therewith, commonly known screening methods do not work as reliable as before. This opens the need for advanced methods to solidly detect Mavericks.

A promising approach which is investigated in the present PhD, is a combination of the Independent Component Analysis (ICA) followed by the Nearest Neighbor Residuals (NNR) method (Turakhia et al., 2005). The idea behind is to perform a data transformation to reveal masked information by applying the ICA and afterwards, taking spatial dependencies over the wafer into account with the NNR (cf. Figure 2). An ad hoc investigation shows that this is a promising research direction, but a thorough data analysis has to precede the ICA to guarantee reliable results.

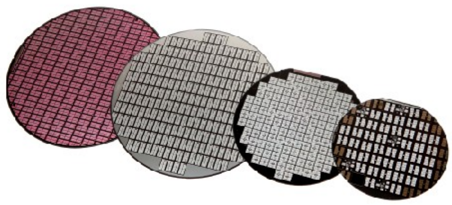

Figure 2: Wafers can be distinguished e.g. regarding the size and number of devices, which is considered with NNR.

2 STATE OF THE ART

Commonly, BI testing is performed to detect weak devices during their early lifetime. Currently, this is the only fully accepted method among semiconductor manufacturers to sort out early lifetime failures. Reliable screening methods, which can classify devices in good ones and Mavericks, consequently reduce the effort spent on BI. The detection of Mavericks using screening methods is a cost-efficient, fast and well-established procedure. Their main target is the identification of Mavericks already at an early stage, whereupon a lot of following tests and production steps can be skipped which results in a more efficient manufacturing sequence. While usual test concepts are designed to detect functionality failures, the intention of screening methods is the detection of reliability problems. They are not a question of actual functionality, but of hidden evidence for failures during early life stage at customer level. This is then a safety and warranty issue.

2.1 Burn-In Testing

An established method to detect Mavericks is the Burn-In (BI), where devices are tested over several hours under increased, but still close to reality, test conditions, such as high temperature and high supply voltage. The reliability of this method regarding detection of early failures is acceptable, but contains the drawback of high costs, including testing time, special equipment requiring routine maintenance and extra trained employees. Moreover, the pre-damage caused during BI stress implies that the device is no longer 'virgin', which introduces an additional reason to search for other testing methods. Due to the fact that BI is currently the only fully accepted screening method among semiconductor manufacturers, the aim to reduce the number of devices tested with BI is more realistic than replacing this method. This can be achieved if the classification of good devices and Mavericks can be performed reliably.

2.2 Part Average Testing

Part Average Testing (PAT) is a standard method, based on the evaluation of data distributions. These distributions should be calculated in a robust way, meaning that the calculated distribution parameters are insensitive to outliers. Although a variety of different measurements can be used, electrical tests like diverse current and voltage measurements are preferable. The idea behind is to detect suspicious devices, which indicate some abnormality compared to the majority of the devices. These devices are then scrapped and not delivered to the customer. To decide whether a device is suspicious or not, upper and lower PAT limits, most often calculated on the basis of a normal distribution, have to be set. In generally, they are tighter than the lower (LSL) and upper (USL) specification limits, see 3. Further, these PAT limits can be divided into static and dynamic limits. The main difference between is that the static limits are calculated once, on the basis of a reasonable amount of data, and then are further applied for subsequent wafers. In contrast to this, the dynamic limits are updated for each wafer, which takes the variation between the wafers into account.

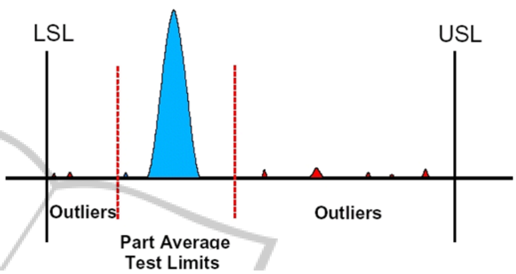

Figure 3: PAT limits, marked in red, are more severe than the lower (LSL) and upper (USL) specification limits.

A further powerful PAT variant is the multivariate PAT. Its focus is the detection of correlated failure mechanisms of two or more test measurements. Expert knowledge and experiences with different technologies are mandatory up to now, to take meaningful combinations of test measurements for the evaluation of the multivariate PAT.

2.3 Good Die in Bad Neighborhood

Another commonly used method, known as Good Die in Bad Neighborhood (GDBN), takes spatial dependencies of devices over the wafer into account. More accurately, a comparison of each device with its neighborhood indicates the devices potential risk. It is known that supposedly good devices surrounded by bad ones are more likely to fail than those surrounded by further good ones. An example is given in Figure 4, where a supposed good device is inked out (colored in gray) because it is surrounded by many bad devices (marked in black).

Further, it can be observed that devices with the same risk behavior tend to cluster and those at the edge of a wafer are more likely to fail due to the manufacturing process of a wafer. The evaluation of good or bad is done on the basis of the Unit Level Predictive Yield (ULPY) calculation (Riordan et al., 2005), taking a combination of yield per wafer (local yield) and yield per lot (stacked yield) into account:

$$\text{ULPY} = \sqrt{\text{local yield} \times \text{stacked yield}}. \quad (1)$$

Again, devices being outside specified limits are inked automatically and rejected in the next production step.

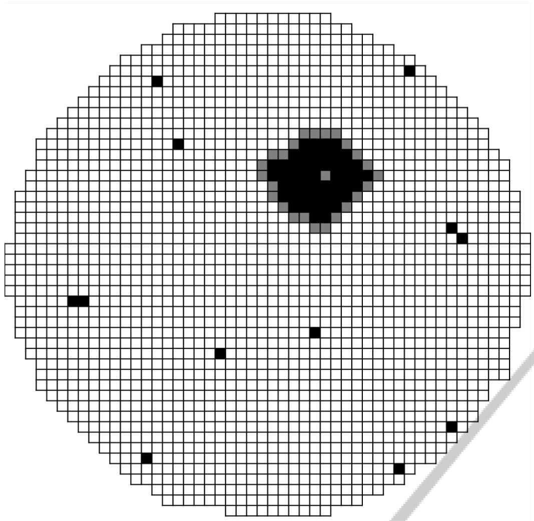

Figure 4: Good devices surrounded by bad devices (marked in black), are inked out (marked as gray devices).

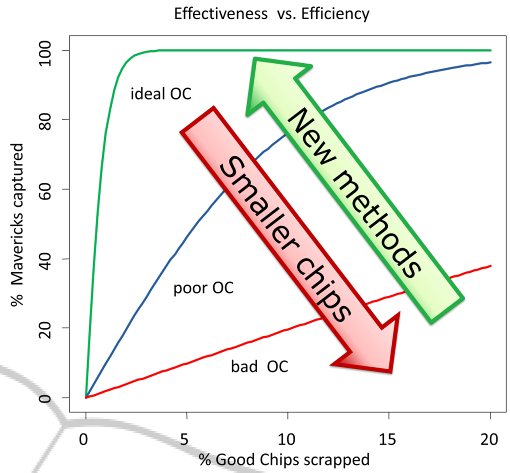

Figure 5: With ever smaller devices, a distinction between good devices and Mavericks becomes challenging. Therewith, the efficiency of currently applied screening methods decreases. New methods are expected to counteract. Consequently, the steeper the operating curve, the higher the classification power of the method.

3 OUTLINE OF OBJECTIVES

Due to the miniaturization of semiconductor devices, new challenges on screening methods arise. The classification in good devices and Mavericks does not work as accurate as before with screening methods described in Section 2. Other failure mechanisms appear or dominate the device failure. To counteract, advanced screening methods are needed with the aim to capture as many Mavericks as possible on the expense of only few good devices (misclassifications). To evaluate the classification power of a method, this ratio can be visualized as an operating curve (see Figure 5). The steeper the ascent of the curve, the more efficient the method works to separate good devices and Mavericks. A 100% Maverick detection without any misclassification forms the optimal case.

It is often not the development of a completely new method which leads to better results, but a combination of meaningful methods implemented in a reasonable order. It is important to analyze the data basis which will be used. Further, a transformation is often necessary to extract special features which lead to a better understanding of the underlying data. Section 5 presents the method of Independent Component Analysis (ICA), which is a promising approach to detect and separate latent information. Afterwards, post-processing steps can be applied. The Nearest Neighbor Residuals (NNR) method is proposed, which takes spatial dependencies over the wafer into account.

Generally, any combination of data analysis, data transformation and post-processing methods can lead to advanced screening quality. Therewith, the operating curve wants to be improved in a way that nearly all Mavericks can be identified, including hardly any misclassification.

4 RESEARCH PROBLEM

The intention of this PhD is to develop advanced screening methods, which can handle the new challenges on sub-micron devices. As a first approach the work flow schematically displayed in Figure 7 is examined.

First step is the evaluation and collection of meaningful data for this purpose. Although a variety of conventional test measurements are done during the production process, for many applications IDDQ measurements (Miller, 1999), i.e. measurements of the power supply current in the quiescent state, are more informative as e.g. functional voltage tests. Depending on the product and the technology, different numbers of measurements are taken. For the product under investigation, 577 IDDQ measurements per device are collected. When the power consuming elements are switched off, a perfect CMOS has IDDQ values in the range of some microamperes whereas higher values indicate a suspicious behavior of one or more transistors. Sub-micron technologies contain new failure mechanisms where it is expected that the currently used screening methods do not work as accurate as before. For instance, smaller devices have increased leakage current which makes it hard to find a threshold for separating good devices from Mavericks. Although, as a first attempt, IDDQ measurements seem to be a good choice, also other test measurements or a meaningful combination of them included in the calculation may be suitable.

High dimensional data often contain latent information, which becomes visible after a meaningful data transformation, preferably in a 2 or 3 dimensional space for complexity reduction as well as for visualizing purposes. Afterwards, a separation criterion to divide good devices and Mavericks has to be found. As a first attempt, the ICA is investigated, which is widely spread e.g. in the fields of speech recognition, image processing, text document analysis and biomedical applications. The applicability of ICA on device data is not that well investigated by now, but will be evaluated in this PhD.

The aim is to find a reliable combination of data analysis techniques, followed by a meaningful data transformation (e.g. ICA) and a final classification method to detect Mavericks.

5 METHODOLOGY

5.1 Independent Component Analysis

High dimensional data often mask informative features which may help to explain an underlying process. Independent Component Analysis (ICA) performs a transformation of observable data $\mathbf{x}$, i.e. test measurements, into a new representation of sources $\mathbf{s}$. This can be obtained by applying a transformation matrix $\mathbf{A}$ (mainly called mixing matrix) to reveal latent structures. In other words, it is expected that the measured data are mixtures of sources that want to be recovered. Mathematically this can be written as follows:

$$\mathbf{x} = \mathbf{A}\mathbf{s}. \tag{2}$$

With a simple inversion of the mixing matrix $\mathbf{A}$, the sources can be calculated:

$$\mathbf{A}^{-1}\mathbf{x} = \mathbf{W}\mathbf{x} = \mathbf{s}. \tag{3}$$

Due to the fact that both, the sources and the mixing matrix, are unknown, conventionally solving the equation is not possible. This means that $\mathbf{A}$ or $\mathbf{W}$ have to be estimated, leading to approximations for the sources as well:

$$\hat{\mathbf{s}} = \hat{\mathbf{W}}\mathbf{x}. \tag{4}$$

The idea behind ICA is to separate measurement data into statistically independent sources. Statistically independent data contain the most information because they do not include any redundancy. To achieve independence, either the non-gaussianity can be maximized or the mutual information can be minimized, as will be outlined in the next section.

5.1.1 Pre-processing for ICA

To perform a reliable ICA, main emphasis lies on data preparation of the test measurements $\mathbf{x}$. Various preprocessing methods are available. Commonly used techniques (Naik and Kumar, 2011) are centering followed by whitening. Centering means a subtraction of the mean from the data. The ensuing whitening data, $\mathbf{x}_w$, is a linear transformation of the measurements whereby $\mathbf{x}_w$ is uncorrelated with a unit variance:

$$\mathrm{Var}\{\mathbf{x}_w\} = \mathrm{E}\{\mathbf{x}_w\mathbf{x}_w^T\} = I. \tag{5}$$

The advantage of whitened measurements is the reduction of the computational complexity of ICA; the number of variables to be estimated decreases from n^2 for matrix $\mathbf{A}$ to $\binom{n}{2}=\frac{n(n-1)}{2}$ for a resulting orthogonal matrix $\mathbf{A}_w$. One method to obtain whitened data $\mathbf{x}_w$ is the Singular Value Decomposition with

$$\mathbf{x}_w = \mathbf{V}\mathbf{D}^{-\frac{1}{2}}\mathbf{V}^T\mathbf{x}, \tag{6}$$

where $\mathbf{V}$ contains the eigenvectors of the covariance matrix $\mathrm{E}\{\mathbf{x}\mathbf{x}^T\}$ and $\mathbf{D}$ is the diagonal matrix of eigenvalues. This modifies the mixing matrix $\mathbf{A}$ to an orthogonal mixing matrix $\mathbf{A}_w$ as shown in the following equation:

$$\mathbf{x}_w = \mathbf{V}\mathbf{D}^{-\frac{1}{2}}\mathbf{V}^T\mathbf{A}\mathbf{s} = \mathbf{A}_w\mathbf{s} \tag{7}$$

with,

$$\begin{aligned} \mathrm{E}\{\mathbf{x}_w\mathbf{x}_w^T\} &= \mathrm{E}\{\mathbf{A}_w\mathbf{s}\mathbf{s}^T\mathbf{A}_w^T\} \\ &= \mathbf{A}_w\mathrm{E}\underbrace{\{\mathbf{s}\mathbf{s}^T\}}_{=I}\mathbf{A}_w^T \\ &= \mathbf{A}_w\mathbf{A}_w^T \\ &= I. \end{aligned} \tag{8}$$

Here, $\mathrm{E}\{\mathbf{s}\mathbf{s}^T\} = I$ can be assumed without loss of generality because ICA is insensitive to the variance. The new representation in Equation 7, containing now an orthogonal matrix $\mathbf{A}_w$, has the previously mentioned advantage of a decrease in complexity. Instead of the Singular Value Decomposition, also a PCA can be performed to get uncorrelated data with unit variance. Geometrically spoken, just a rotation of the matrix has to be found to get the desired independent data. Therefore, numerical optimization algorithms, like the gradient descent, can be used and optimized using a quantitative measure of non-gaussianity, like the kurtosis and the Neg-entropy.

Regardless of the distribution of independent random variables, based on the central limit theorem, their sum converges to a Gaussian distribution for a sufficiently large sample size. Conversely, making the measurements as non-Gaussian as possible will return these independent sources. This implies the assumption that the sources are independent and further, that at most one of the measurements is Gaussian. The usage of the kurtosis as evaluation criterion for non-gaussianity is quite popular because it is computationally easy to implement. A kurtosis value of zero implies a perfect Gaussian distribution of the underlying data, whereas a nonzero value indicates a deviation from the Gaussian distribution. Unfortunately, the kurtosis is very sensitive to outliers. A more robust criterion is so-called Neg-entropy which is defined as a measure of gaussianity, reflecting the deviation of the data from a Gaussian distribution. The disadvantage of this method is that the probability density function of the data has to be known in advance. The uncertainty about the underlying probability density function can be compensated using approximations instead. Another procedure is the Infomax-principle (Bell and Sejnowski, 1995), which stands for information maximization and will be realised by minimizing mutual information.

Further pre-processing steps, which often result in dimension reduction techniques, can be performed. Depending on the data and the application, some of them use PCA, Projection Pursuit (Friedman, 1987), filtering, stochastic search variable selection like Bayesian networks and wavelet transformation.

5.1.2 Post-processing after ICA

After the ICA has been performed, the resulting sources have to be evaluated. This can be again a form of filtering, e.g. the separation in informative sources and noise. As a first attempt, the NNR method is used, which takes spatial dependencies of devices over the wafer into account, see Equation 9. From each device value, $\mathrm{v}(x_i, y_j)$, the median of the surrounding devices is subtracted, whereas m and n are dependent on the neighborhood:

$$\mathrm{NNR}(x_i, y_j) = \mathrm{v}(x_i, y_j) - \mathrm{med}(\mathrm{v}(x_{i+m}, y_{j+n})). \quad (9)$$

The size of the neighborhood (8, 24 or more) for calculating the NNR is a further topic which will be investigated in this PhD. Previous investigations have shown that the nearest 24 surrounding devices ($m, n = \{-2, -1, 1, 2\}$) are a good choice. However, first evaluations of the NNR on the sources have shown that the number of surrounding devices taken into calculation needs to be determined for this project.

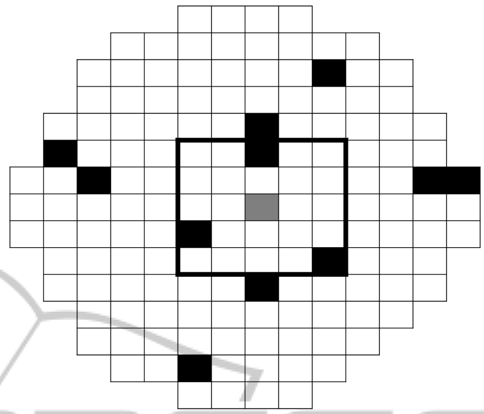

Figure 6: To calculate the NNR value for each device (here marked in gray), just the pass devices are taken into account because for the electrical fail devices (marked in black) no value is available. The black square shows the involved devices for a 24-based neighborhood.

As visualized in Figure 6, only pass devices are considered for the NNR calculation and therefore it happens that not all surrounding neighbors are available. In this case it might be useful to consider either just the given values or other significant ones instead. As a first attempt, the gap can be filled up with devices on the main x-y-directions, starting from device (x_i, y_i). Those from the diagonal directions are only considered if further devices are needed. Additionally, depending on e.g. the distance, weights can be incorporated.

5.2 Further Considerations

As outlined in Section 5.1.1, various pre-processing methods are listed, where the most meaningful one has to be determined to provide a useful starting position for the ICA method itself. Generally, the number of sources is unknown, implying that there can be more sources than measurements (underdetermined) or vice versa (overdetermined system of equations) (Naik and Kumar, 2011). For the first case, the calculation of a pseudo-inverse is necessary. An application can be found e.g. in bio-signal processing, where the number of electrodes are limited compared to the active muscles involved. For an overdetermined system of equations dimension-reducing pre-processing steps can be performed, see Section 5.1.1. A reduction of the measurements to the number of expected sources is preferable. For ICA itself, different MATLAB® packages, like the *FastICA* (Hyvärinen et al., 2001), are available.

6 STAGE OF THE RESEARCH

A first and promising approach to detect Mavericks is depicted in Figure 7.

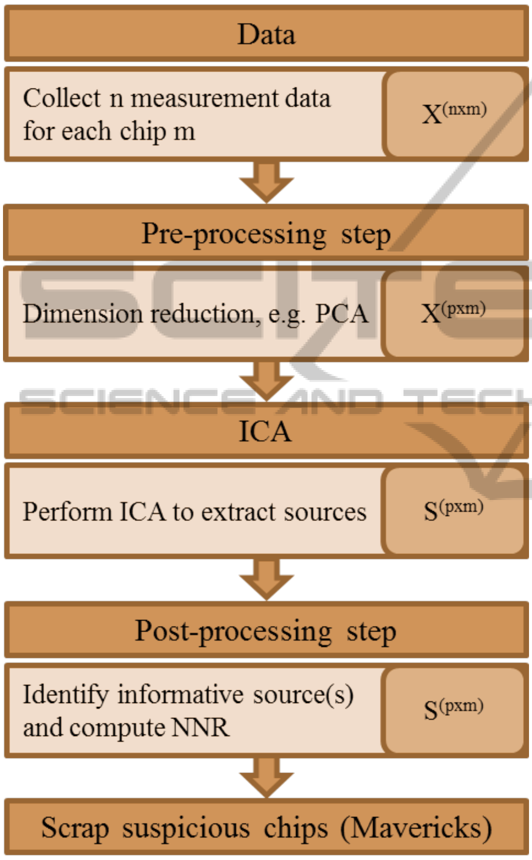

Figure 7: The flowchart provides an overview of the proposed way of proceeding.

The proposed way of proceeding starts with the collection of meaningful data, followed by their analysis and preparation as a pre-processing step for ICA. After ICA is performed, the most informative source or even sources have to be identified. It is expected that an additional NNR calculation reveals further latent information to finally detect suspicious devices.

6.1 Data and Pre-processing Step

As explained in Section 4, the data under investigation are 577 IDDQ measurements per device. Figure 8 shows the first five of them. The correlation matrix C shows strong similarity between the measurements,

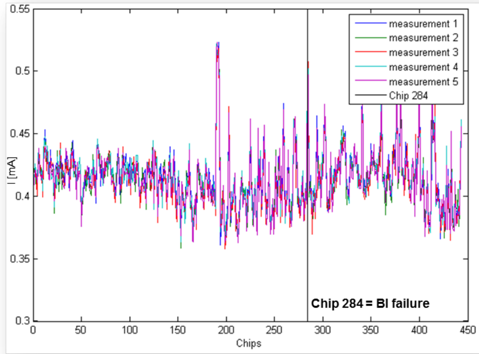

Figure 8: Each IDDQ measurement can be visualized as a signal over all devices. This figure shows the first 5 (superimposed) IDDQ measurements taken from overall 577 available, used as a dimension-reduced basis for the following ICA.

with $C_{ij} > 0.97$. Also the PCA has shown, that already one component explains more than 90 % of the variability in the measurement data. A dimension reduction of the measurement data is recommended as a pre-processing step. For this, different techniques will be investigated. With knowledge about the number of hidden sources in the measurements, the size of dimension reduction would be determined but this insight is generally not given. Nevertheless, due to the high correlations between the measurements, a huge amount of measurements is at least mathematically negligible, i.e. contains no additional information. This implies the assumption of an overdetermined system, where more measurements are available than sources expected. As outlined in Section5.2, a pre-processing step in terms of a dimension reduction is proposed.

6.2 ICA on 5 IDDQ Measurements

As a first attempt, the ICA calculation has been performed on the basis of the first five IDDQ measurements (see Figure 8). The ICA is performed using the implemented MATLAB function *FastICA*. The (5×5) mixing matrix $\mathbf{A}$ and de-mixing matrix $\mathbf{W}$ are calculated and applied to the measurements (see Equation 4). The resulting 5 sources are visualized in Figure 9.

To calculate the mixing matrix, symmetric orthonormalization is recommended, which calculates the sources in parallel. Consider, that for reproducibility the continual ambiguities of scaled and permuted sources remain.

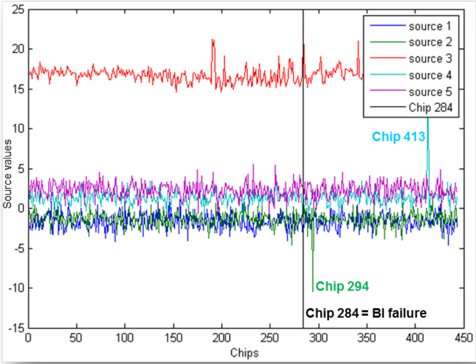

Figure 9: With a quadratic de-mixing matrix **W** the dimension of the resulting sources remains the same. Source 3 (red) is conspicuous compared to the remaining sources and is therefore assumed to contain the most information. Source 2 and 4 show one peak each and thus will be investigated in more detail.

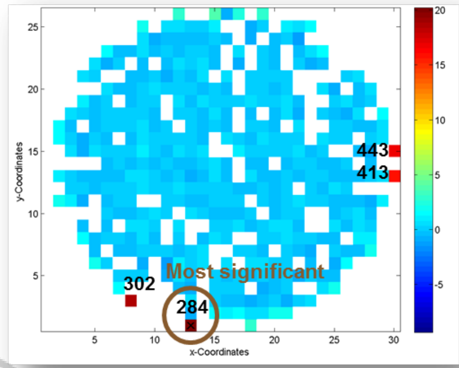

Figure 10: The wafer map represents the resulting NNR calculation on source 3 (see Figure 9) with 8 neighbors (NNR8). Device 284 is the most significant one.

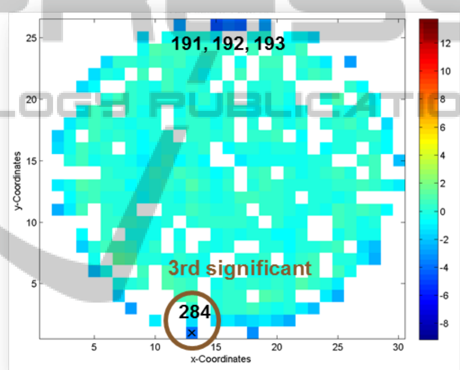

Figure 11: The wafer map represents the resulting NNR calculation on source 3 (see Figure 9) with 24 neighbors. In contrast to NNR8, device 284 is the third significant one.

6.3 Post-processing Step

While each of the measurements seems to contain the same information (optically (see Figure 8) and indicated by high correlations), the ICA transformed sources show clearly separated signals (see Figure 9). In contrast to other applications of ICA, e.g. investigations of EEG signals, where the pathway of a standard signal is known, there is no reference signal for IDDQ measurements which can be used on a comparative basis. Without the identification of one specific significant source, the NNR is calculated for each of the five sources based on 8 as well as on 24 neighbors. Source 3 (see Figure 9, colored in red), which is conspicuous compared to the remaining sources, even provides 7 Mavericks from totally 8 detected ones, see Figure 10 and Figure 11.

Therewith, source 3 seems to include the most information. Source 4 reveals device 413, whereas this device has been found with source 3 as well. Source 2 is responsible for the suspicious device 294. NNR8 has to be considered carefully because for devices on the edge of the wafer, just the available neighborhood is used, without any gap filling adaptation (c.f. Section 5.1.2). This means that for some NNR calculations, only few surrounding devices are available, especially if just an 8-based neighborhood is chosen. Nevertheless, both figures show clearly suspicious devices. Device number 284, additionally marked with a black cross, is extremely suspicious in even both NNR calculations. Altogether, device 191, 192, 193, 284, 302 and 443 has been detected in source 3, device 413 in source 3 as well as in source 4 and device 294 is just suspicious in source 2 (see Figure 9). To finally judge these 8 suspicious devices the results from the BI study are provided in the next section.

6.4 Investigation of Detected Mavericks

To verify the actual behavior of the previously detected 8 devices during their early lifetime, a BI-study was performed. Results from Backend and BI testing reveal that 5 devices (192, 193, 294, 302, 443) failed in the standard Backend test flow before BI and 2 devices (191 and 284) failed during BI after 2 hours and 12 hours, respectively. Altogether, 7 out of the 8 detected Mavericks failed indeed. Just device number 413 survived the 96h BI test, although it is suspicious in even two sources. BI survivors are not necessarily devices no longer containing any risk. A lifetime investigation of these devices may explain their longtime behavior. Nevertheless, 7 out of 8 is a promising intermediate result. Further, a more precise investigation of the BI failure mechanisms is planned to examine a possible connection between the theoretically detected Mavericks and their failure type. Possibly different sources might indicate different failure mechanisms but this is an open question yet.

Since source 3 itself detects most of the Mavericks, an automated detection instead of a visual assessment to identify a meaningful source **s**, is desired. Calculating the L_2-norm for each source, Equation 10, reflects the applicability at least for source 3, see Table 1.

$$||\mathbf{s}_i||_2 = \left(\sum_{j=1}^{577} (\mathbf{s}_{ij})^2 \right)^{1/2} \quad \text{for } i = 1, \ldots, 5. \qquad (10)$$

Table 1: Calculation of the L_2-norm for the 5 sources.

Source	L_2-norm
1	39.9
2	35.6
3	353.5
4	35.7
5	52.9

Source 3 takes the highest norm value. Unfortunately, rearranging the remaining sources regarding their descending order of the norm values does not fit to the observed outcome. Beside source 3, source 2 and 4 detected Mavericks and therewith were expected to get higher values than source 1 and 5. An investigation of the sources regarding their four moments, namely mean, variance, skewness and kurtosis, identifies source 3 as the most non-Gaussian source whereas the remaining 4 sources are close to noise. Further evaluations have to be done to quantify this information.

7 EXPECTED OUTCOME

The aim of this PhD is to develop an advanced screening method to detect Mavericks on sub-micron technologies with higher integration density, where currently existing screening techniques do not work as accurate as before on larger structures. To judge the classification power of the new method, its operating curve will be evaluated and compared to those of the already existing screening methods. As an interim result on the currently investigated data, Table 2 compares the classification power between the proposed method and the commonly used dynamic PAT.

First investigations (see Section 6) have shown that a meaningful combination of pre- and post-processing methods improves the performance of ICA, while ICA itself has free selectable optimization criteria as well, depending on the application. With a new improved approach to detect Mavericks, the effort spend on BI can be reduced and a relevant amount of time and money can be saved. Additionally, lifetime investigations of Mavericks which survive the BI may even give information about an appropriate BI time.

Table 2: Comparison of the classification power between the proposed method (see flowchart in Figure 7) and the dynamical PAT (DPAT) on the 8 suspicious devices (191, 192, 193, 284, 294, 302, 413, 443).

classification	proposed method	DPAT
correct	7	1
incorrect	1	7

ACKNOWLEDGEMENTS

Special thanks goes to Johannes Kaspar from Infineon Technologies Austria AG for his valuable discussions and for being at hand with his expert knowledge anytime.

This work is funded by the Federal Ministry for Transport, Innovation and Technology (BMVIT) funding scheme Talente of the Austrian Research Promotion Agency (FFG, Project No. 839342) and by Infineon Technologies Austria AG.

REFERENCES

Bell, A. J. and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159.

Friedman, J. H. (1987). Exploratory projection pursuit. *Journal of the American Statistical Association*, 82:249–266.

Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. John Wiley & Sons.

Miller, A. C. (1999). Iddq testing in deep submicron integrated circuits. In *ITC International Test Conference*.

Naik, G. R. and Kumar, D. K. (2011). An overview of independent component analysis and its applications. *Informatica*, 35:63–81.

Riordan, W. C., Miller, R., and St. Pierre, E. R. (2005). Reliability improvement and burn in optimization through the use of die level predictive modeling. In *Annual IRPS*. 43rd IEEE Annual IRPS.

Turakhia, R. P., Benware, B., Madge, R., Shannon, T. T., and Daasch, W. R. (2005). Defect screening using independent component analysis on iddq. In *VTS'05*. 23rd IEEE VLSI Test Symposium.