Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

DEVICE LEVEL MAVERICK SCREENING - DETECTION OF RISK DEVICES THROUGH INDEPENDENT COMPONENT ANALYSIS

Anja Zernig
Olivia Bluder

KAI - Kompetenzzentrum für Automobil- und Industrieelektronik GmbH
Europastraße 8
Villach, 9524, AUSTRIA

Jürgen Pilz

Alpen-Adria Universität Klagenfurt
Universitätsstraße 65-67
Klagenfurt, 9020, AUSTRIA

Andre Kästner

Infineon Technologies Austria AG
Siemensstraße 2
Villach, 9500, AUSTRIA

ABSTRACT

Reliable semiconductor devices are of paramount importance as they are used in safety relevant applications. To guarantee the functionality of the devices, various electrical measurements are analyzed and devices outside pre-defined specification limits are scrapped. Despite numerous verification tests, risk devices (Mavericks) remain undetected. To counteract this, remedial actions are given by statistical screening methods, such as Part Average Testing and Good Die in Bad Neighborhood. For new semiconductor technologies it is expected that, due to the continuous miniaturization of devices, the performance of the currently applied screening methods to detect Mavericks will lack accuracy. To meet this challenge, new screening approaches are required. Therefore, we propose to use a data transformation which analyzes information sources instead of raw data. First results confirm that Independent Component Analysis extracts meaningful measurement information in a compact representation to enhance the detection of Mavericks.

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 are very likely to occur as visualized by the bathtub curve in Figure 1.

The main reason for device failure during the early life stage results from the manufacturing process, where hundreds of single steps are performed and interacting with each other. 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 Burn-In (BI) testing, explained in more detail in Section 2.1, where the devices are tested under elevated stress conditions to simulate the early life. Due to undesirable side effects of the BI, like high costs and the need for extra equipment, a reduction of devices to be burned is preferable. More cost-efficient are statistical screening methods which are capable of detecting potential early life failures. A device which is suspicious compared to the majority represents a risk device, so-called Maverick.

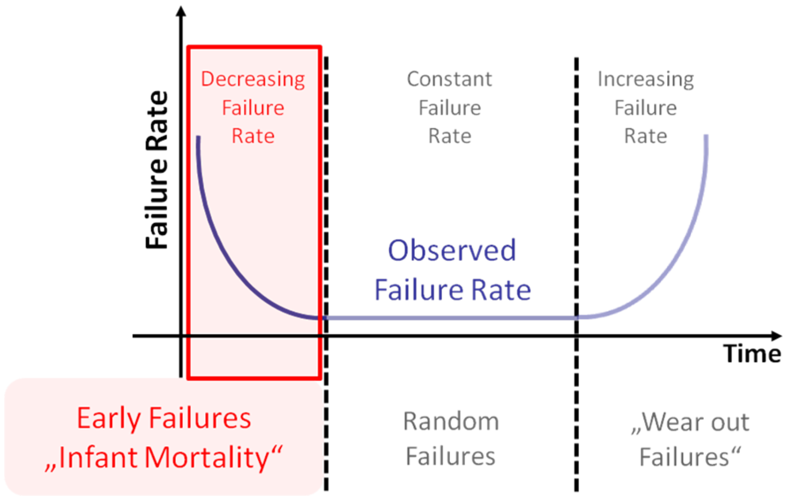

Figure 1: The bathtub curve describes the life stages of semiconductor devices. After surviving the early life stage, which is in focus of this paper, a constant failure rate dominates, where failures happen at random, before they enter the 'wear out' stage, where devices suffer from aging.

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, which are expected to contain new failure mechanisms, a distinction between reliable devices and Mavericks becomes increasingly challenging, as the threshold between them is less evident. Therewith, commonly known screening methods are insufficient, which opens the need for advanced methods to reliably detect Mavericks.

The paper is organized in the following: Section 2 starts with an overview of the main screening methods used in semiconductor industry. Afterwards in Section 3, a novel approach is presented, using the Independent Component Analysis to filter informative knowledge from measurement data, introduced in Section 4. Further, the Negentropy is proposed to classify devices in good ones and Mavericks. Results are shown in Section 5. Finally, Section 6 summarizes the intermediate results and gives an outlook on future work. Current investigations show that this is a promising research direction for detecting Mavericks.

2 CURRENTLY APPLIED SCREENING METHODS IN SEMICONDUCTOR INDUSTRY

The detection of Mavericks using screening methods is a well-established procedure. Their main target is the identification of Mavericks already at an early stage of life. While usual test concepts are designed to detect functionality failures, the intention of screening methods is the detection of reliability issues. They are not a question of actual functionality, but of hidden evidence for failures during early life stage at customer level which is also a safety and warranty issue for some applications. Commonly, BI testing is performed to detect Mavericks but due to unwanted side effects, screening methods, based on cost-efficient statistical evaluations are preferable. This section gives detailed information about state of the art screening methods applied in semiconductor industry.

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 satisfactory, but with 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 harmless testing methods instead. Due to the fact that BI is well-known and broadly used among semiconductor manufacturers, a reduction of devices to be tested with BI is more realistic than replacing this method. This can be achieved if a reliable statistical classification of good devices and Mavericks can be performed beforehand, so that the BI is just needed for the Mavericks.

2.2 Part Average Testing

Part Average Testing (PAT) is a standard method based on the evaluation of measurement data regarding their distributions. Although a variety of different measurements can be used, electrical tests like diverse current and voltage measurements are preferable (Automotive Electronics Council 2011). The concept of PAT is the detection of 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 are calculated and set, based on the underlying distribution of the data. To be insensitive against outliers, robust parameter estimation is recommended. In general, the PAT limits are tighter than the lower (LSL) and upper (USL) specification limit, see Figure 2, to increase the detection rate.

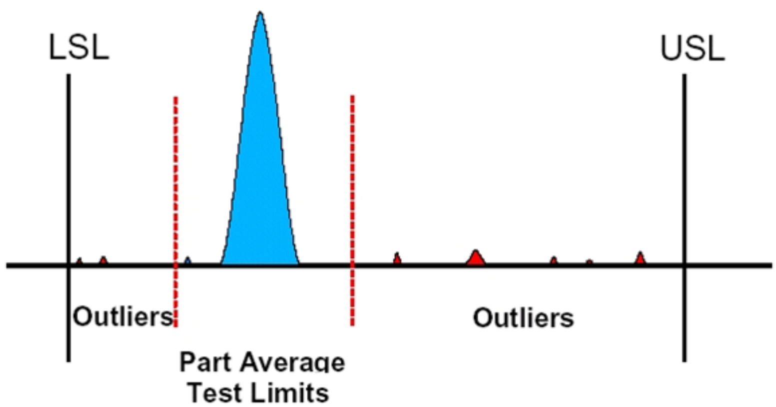

Figure 2: PAT limits, marked as dotted lines, are more severe than the lower (LSL) and upper (USL) specification limit. Outliers are detected and scrapped to guarantee the customer more reliable devices.

2.3 Good Die in Bad Neighborhood

Another commonly used method takes spatial dependencies of devices over the wafer into account, known as Good Die in Bad Neighborhood (GDBN). More accurately, a comparison of each device with its neighborhood indicates the devices potential risk. It can be observed that devices, which tend to cluster, often have the same risk behavior and therefore, supposedly good devices surrounded by or next to bad ones are more likely to fail. An example is given in Figure 3, where supposed good devices (in gray) are marked because they are in the neighborhood of bad devices (in black). The evaluation of good or bad can be done e.g. on the basis of the Unit Level Predictive Yield (ULPY) calculation (Riordan et al. 2005), taking a combination of yield per wafer and yield per lot into account. Again, devices outside specified limits are marked and scrapped in the next production step.

3 A NOVEL SCREENING APPROACH

Common screening methods like those in Section 2 are expected to work less efficient on future technologies with smaller structures, which provides the need for advanced screening methods. As it is discussed

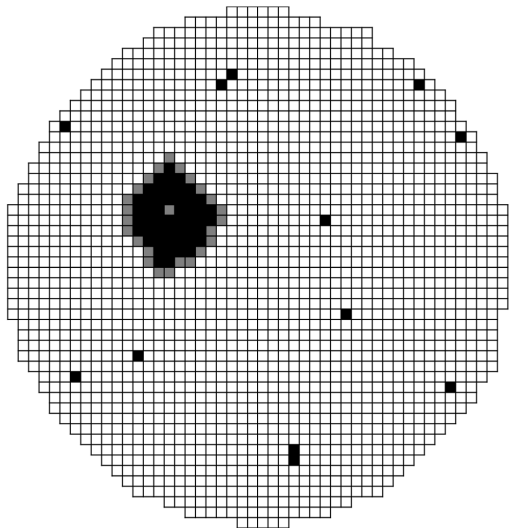

Figure 3: The method of GDBN takes spatial dependencies into account. Good devices in the neighborhood of bad devices (marked in black) are preventatively scrapped (marked as gray devices).

in (Turakhia et al. 2005), the Independent Component Analysis is proposed to perform a valuable data transformation, separating informative and non-informative content of measurement data. After a successful separation, the Negentropy is applied to investigate the output of the transformation.

3.1 Data Transformation with Independent Component Analysis

High dimensional data often mask informative features which may help to explain or discover an underlying process. Assuming that the measured data are mixtures of sources that want to be recovered, a data transformation to reveal this information is needed. Independent Component Analysis (ICA) (Hyvärinen et al. 2001) performs such a transformation of observed data $\mathbf{x}$ into a new representation of sources $\mathbf{s}$. The observed data $\mathbf{x}$ are various measurements of each device over the wafer. The transformation can be obtained by applying a transformation matrix $\mathbf{A}$, called mixing matrix, to reveal latent structures. The model of ICA can mathematically be written as

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{pmatrix} = \mathbf{A} \begin{pmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \vdots \\ \mathbf{s}_n \end{pmatrix}, \tag{1}$$

$$\text{with } \mathbf{x}_i = (x_{i1}, \ldots, x_{im}),\ \mathbf{s}_i = (s_{i1}, \ldots, s_{im}),\ 1 \le i \le n \text{ and } \mathbf{A}^{(n \times n)},$$

where $x_{ik}, 1 \le k \le m$, is one measurement per device. With a simple inversion of the mixing matrix $\mathbf{A}$, the respective source value s_{ik} can be calculated, but due to the fact that both, the sources and the mixing matrix, are unknown, conventional solving of (1) is not possible. Thus, $\mathbf{A}^{-1}$ has to be estimated, leading to approximations for the sources as well. The best approximation in terms of ICA is achieved when the sources $\mathbf{s}_i$ are as independent from each other as possible.

For a successful ICA it is common to start with data-centering (Naik and Kumar 2011), which means the subtraction of the mean from the data. Ensuing, data-whitening is performed to achieve a decorrelation of the data by transforming the measurements with known covariance matrix in an uncorrelated representation of the measurements, $\mathbf{x}_w$, with the identity as new covariance matrix:

$$\mathrm{Var}\{\mathbf{x}_w\} = \mathrm{E}\{\mathbf{x}_w\mathbf{x}_w^T\} = I, \text{ with } \mathbf{x}_w = \mathbf{V}\mathbf{D}^{-\frac{1}{2}}\mathbf{V}^T\mathbf{x}, \tag{2}$$

where the whitened data $\mathbf{x}_w$ are obtained by the Singular Value Decomposition. $\mathbf{V}$ contains the eigenvectors of the covariance matrix $\mathrm{E}\{\mathbf{x}\mathbf{x}^T\}$ and $\mathbf{D}$ is the diagonal matrix of eigenvalues as denoted in (2). This modifies the mixing matrix $\mathbf{A}$ to an orthogonal mixing matrix $\mathbf{A}_w$, with the advantage of a reduction in computational complexity. This means that 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 $\mathbf{A}_w$. Henceforth, just a rotation of the matrix has to be found to get the desired independent sources. Numerical optimization algorithms, like the gradient descent, can be used and optimized using a quantitative measure of non-Gaussianity, like the kurtosis or the Negentropy (Comon 1994).

3.2 Negentropy as Decision Criterion for Maverick Detection

In context of ICA, the Negentropy is used as a measure of non-Gaussianity, to optimize the transformation of measurement data $\mathbf{x}_i$ towards independent sources $\mathbf{s}_i$, see (1). Here, too, the Negentropy is used as an indicator for deviations from a Gaussian distribution, but applied to the final sources, gained by ICA, to decide, whether a device is an outlier, i.e. a Maverick, or not. Detailed steps can be found in the application part of this paper in Section 4.

Standard outlier detection methods, like PAT (see Section 2.2), need a predefined threshold (e.g. 6σ limits) to define a decision boundary between good devices and Mavericks. In this paper a new strategy, avoiding hard limits, is developed, using the Negentropy $J(\mathbf{y})$ defined by (Hyvärinen et al. 2001):

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \qquad \text{with} \begin{cases} J(\mathbf{y}) = 0 & \text{if } \mathbf{y} \text{ is Gaussian} \\ J(\mathbf{y}) > 0 & \text{if } \mathbf{y} \text{ is non-Gaussian.} \end{cases} \tag{3}$$

In words, the Negentropy represents the deviation of the data under investigation $\mathbf{y}$ from a Gaussian data set $\mathbf{y}_{gauss}$ of the same variance as the data. Beside different approximations for the Entropy H, a general definition via the probability density function f with support set $\mathbb{X}$, can be given as

$$H(\mathbf{y}) = -\int_{\mathbb{X}} f(\mathbf{y}) \log f(\mathbf{y}) d\mathbf{y}. \tag{4}$$

Since the Gaussian distribution has the highest possible value of entropy for a fixed variance, substracting the entropy of an arbitrary data set always results in a non-negative value for the Negentropy. This property is used in Algorithm 1, where the data under investigation are each single source $\mathbf{s}_i$. Therewith, a new strategy for Maverick detection is provided.

4 APPLICATION OF ICA AND NEGENTROPY ON SEMICONDUCTOR DATA

In this section the previously described approach is applied to a semiconductor data set, consisting of one wafer containing 443 devices. For each device, 577 IDDQ measurements are available. From the mathematical point of view, as these measurements are highly correlated and therefore contain redundant information, the dimensionality is reduced. Also in practice, fewer but representative measurements would be preferable as an investigation of all measurements during the process chain is not feasible. In addition, the quality of the ICA transformation is improving when ICA is applied on a subset of the data instead of using all of them, as this is dependent on the ratio between measurements and devices. Hence, the ICA is applied to a selected subset of measurements and the resulting sources are investigated by using the Negentropy as a detector of Mavericks, described in Algorithm 1.

ALGORITHM 1: THIS PSEUDOCODE DESCRIBES WHETHER A DEVICE IS CLASSIFIED AS MAVERICK OR NOT BY USING THE NEGENTROPY AS DECISION CRITERION.

1. Generate a Gaussian distributed random variable $\mathbf{y}_{gauss} \sim N(\mu, \sigma^2_{\mathbf{s}_i})$
2. Calculate Negentropy J as defined in (3) and (4) with $\mathbf{s}_i$ as the data under investigation
3. **While** $\mathbf{s}_i$ is not empty
    - **find** $\arg\min_{\mathbf{s}_{ik}} J(\mathbf{s}_i \setminus \mathbf{s}_{ik})$ for $k = 1, \ldots, m$
    - **save** $\mathbf{s}_{ik}$ as outlier and delete it from $\mathbf{s}_i$

    end
4. Run step 1 to 3 for multiple times.
5. **for** $l = 1 : m$
    - **if** outlier $\mathbf{s}_{ik}$ at position l of all runs is equal **then**
        - outlier $\mathbf{s}_{ik}$ is a Maverick
    - **else**
        - break
    - **end**

    end

4.1 Data Preparation and Subsequent ICA

Although a variety of conventional 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 assumed to be more informative than e.g. functional voltage tests (Automotive Electronics Council 2011). The wafer of investigations consists of 443 devices. For each device 577 IDDQ measurements $(\mathbf{x})$ are collected as shown in the following matrix:

$$
\begin{array}{c} \\ \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{577} \end{array}
\begin{array}{c}
\begin{array}{cccc} \text{Device 1} & \text{Device 2} & \ldots & \text{Device 443} \end{array} \\
\begin{pmatrix}
\text{IDDQ}_{1,1} & \text{IDDQ}_{1,2} & \ldots & \text{IDDQ}_{1,443} \\
\text{IDDQ}_{2,1} & \text{IDDQ}_{2,2} & \ldots & \text{IDDQ}_{2,443} \\
\vdots & \vdots & \ddots & \vdots \\
\text{IDDQ}_{577,1} & \text{IDDQ}_{577,1} & \ldots & \text{IDDQ}_{577,443}
\end{pmatrix}
\end{array}
$$

Due to high correlations of the 577 IDDQ measurements, a dimension reduction is recommended as a pre-processing step for the subsequent ICA. Therefore, a simulation study was performed which supports the selection of a subset of 5 measurements $\mathbf{x}_1, \ldots, \mathbf{x}_5$, visualized in Figure 4. On the reduced data set of 5 IDDQ measurements, the ICA is performed using the implemented MATLAB® function *FastICA* (Hyvärinen et al. 2001). The mixing matrix $\mathbf{A}^{(5\times5)}$ and de-mixing matrix $\mathbf{A}^{-1}$ are calculated and applied to each of the 5 measurements $\mathbf{x}_i$ (cf. Section 3.1). The resulting 5 sources, $\mathbf{s}_1$ to $\mathbf{s}_5$, are visualized in Figure 5.

While each of the measurements seems to contain the same information, indicated by high correlations, see Figure 4, the ICA transformed measurements show clearly separated sources $\mathbf{s}_i$, as visualized in Figure 5. In contrast to other applications of ICA, e.g. investigations of EEG signals, where the pathway of the desired sources is known, there is no reference signal for transformed IDDQ measurements which can be used on a comparative basis. Without this knowledge yet, each resulting source has to be investigated regarding its information content. Results are given in Section 5.

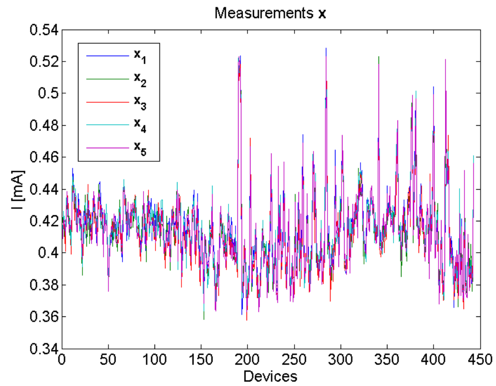

Figure 4: Each IDDQ measurement $\mathbf{x}_i$ can be visualized as a signal over all devices. This figure shows 5 (superimposed) IDDQ measurements taken from overall 577 available, used as a dimension-reduced basis for subsequent ICA.

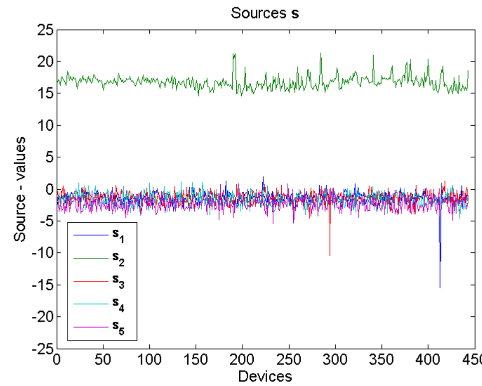

Figure 5: With a quadratic de-mixing matrix $\mathbf{A}^{-1}$ single sources can be revealed from initially superimposed measurements, see Figure 4. Compared to the remaining sources, Source 2 (green) has a conspicuous offset.

4.2 Investigation of ICA Sources based on Negentropy

The Negentropy as a measure of non-Gaussianity is used to detect the Mavericks provided by each source $\mathbf{s}_i$, visualized in Figure 5 or plotted as probability plot for a Gaussian distribution in Figure 6. Therefore, the pseudocode in Algorithm 1 was applied to each source: Iteratively, one device after each other is rejected, to lead the distribution of the source towards Gaussianity. Theoretically, the rejection of devices is stopped, when hardly any further improvement is achieved. In other words, devices are classified as outliers, as long as the pathway of Negentropy values converges to zero. Here, step 3 from the algorithm is repeated for 10 times (see Figure 7). The pathway of the Negentropy values differs from run to run. This results from step 1, where for each run a new Gaussian distributed random variable is generated and needed for calculating the Negentropy. To make a decision whether an outlier is a Maverick or not, from the 10 runs just the devices in common and in the same selected order are classified as the final Mavericks (cf. step 5 in Algorithm 1). For instance, in Source 2, 12 devices fulfill this criterion and are therewith classified as Mavericks. Results for all sources are given in Section 5.

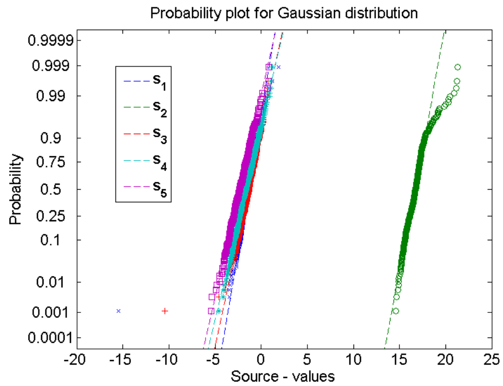

Figure 6: The probability plot for each source illustrates devices, which deviate from a Gaussian distribution. Obviously there is one outlier in Source 1 and 3 and multiple outliers in Source 2.

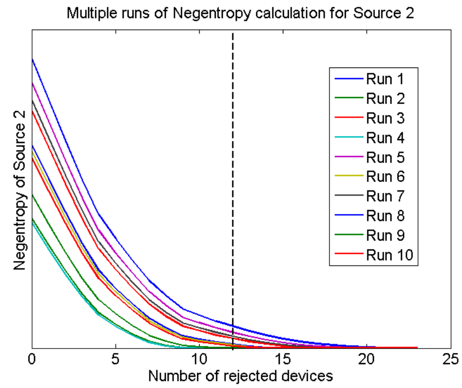

Figure 7: Following the pseudocode in Algorithm 1, 10 runs of Negentropy calculation for Source 2 are performed and 12 devices are classified as Mavericks.

5 RESULTS

In the previous section, Mavericks have been detected by applying the ICA on a subset of semiconductor measurement data to reveal informative sources. Afterwards, each source has been investigated regarding Mavericks, using the Negentropy from Section 3.2 as decision criterion, following the pseudocode in Algorithm 1. However, to finally judge if the devices are correctly classified as Mavericks, they have to be compared to the results from the corresponding BI study, which provides the true information about the behavior of all devices during their early life stage (cf. Section 2.1). The results are provided in Table 1.

Table 1: Classification power of sources: For each source the Mavericks are figured out by applying the Negentropy as decision criterion. Subsequently, the results of the BI Study are used to evaluate the classification power of the sources. Source 2, which differs from the remaining sources, provides remarkable results. Source 1 and 3 reveal two further Mavericks, but cannot ensure a BI failure.

	# of detected devices with measure of Negentropy	correctly classified (according to BI)
Source 1	1	0
Source 2	12	12
Source 3	1	1
Source 4	0	0
Source 5	0	0

Summarized in Table 1, Source 2 detects the most Mavericks and, in addition, all of them are correctly classified. For the remaining sources hardly any Mavericks are detected as expected from the probability plot in Figure 6. They are already close to a Gaussian distribution and therefore may be negligible as the aim is to concentrate important information in one source, which then finally needs to be screened for Mavericks.

6 SUMMARY AND FUTURE WORK

State of the art screening methods are expected to be less reliable regarding Maverick detection for submicron technologies. This opens the need of advanced screening methods to guarantee the customer a reliable product. In this paper the method of ICA is proposed to filter information about risk devices from measurement data which then provides a basis for subsequent Maverick detection. After transforming the measurement data in sources by applying the ICA, the measure of Negentropy is used as a new concept of outlier detection. First results confirm that a closer look in this direction is valuable.

For future work, investigations on Negentropy are planned, because the classification results differ dependent on the applied approximation for the Entropy. In addition, analysis regarding an optimal number of runs, used in Algorithm 1, will be done. Further room for improvement provides ICA itself as this method is an optimization problem. Additionally, verification tests on different data sets will be performed. With a reliable approach to detect Mavericks, the effort spend on BI could be reduced significantly, since just the Mavericks have to be burned, and therewith a relevant amount of time and money can be saved.

ACKNOWLEDGMENTS

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

Automotive Electronics Council 2011. "Guidelines for Part Average Testing". AEC-Q001 Rev-D.

Comon, P. 1994. "Independent component analysis, a new concept ?". *Signal Processing* 36 (3): 287–314.

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

Miller, A. C. 1999. "IDDQ testing in deep submicron integrated circuits". In *ITC International Test Conference*. 28 (2): 724–729.

Naik, G. R., and D. K. Kumar. 2011. "An Overview of Independent Component Analysis and Its Applications". *Informatica* 35: 63–81.

Riordan, W. C., R. Miller, and E. R. S. Pierre. 2005. "Reliability improvement and burn in optimization through the use of die level predictive modeling". In *Reliability Physics Symposium*, 43rd Annual.: 435–445.

Turakhia, R., B. Benware, R. Madge, T. Shannon, and R. Daasch. 2005. "Defect Screening Using Independent Component Analysis on IDDQ". In *Proceedings of the 23rd IEEE Symposium on VLSI Test*, VTS '05: 427–432.

AUTHOR BIOGRAPHIES

ANJA ZERNIG is a Ph.D. student at the Alpen-Adria Universität of Klagenfurt (Department of Statistics), Austria and works at KAI - Kompetenzzentrum für Automobil- und Industrieelektronik GmbH, Austria. Her research field is the reliability of screening methods used to detect risk devices. Her email address is anja.zernig@k-ai.at.

OLIVIA BLUDER is a Junior Researcher at KAI - Kompetenzzentrum für Automobil- und Industrieelektronik GmbH, Austria. She received her M.S and Ph.D. in statistical analysis and Bayesian modeling of Smart Power semiconductor lifetime at the Alpen-Adria Universität of Klagenfurt (Department of Statistics), Austria. Her email address is olivia.bluder@k-ail.at.

JÜRGEN PILZ is full Professor and Head of the Department of Statistics at the Alpen-Adria Universität of Klagenfurt, Austria. He holds M.S. and Ph.D. degrees in Mathematical Statistics at the Technical University Bergakademie Freiberg, Germany. His research interests are Bayesian modeling, computer science, spatial statistics, design of experiments and geoinformatics. His email address is Juergen.Pilz@aau.at.

ANDRE KÄSTNER is a Staff Engineer in the field of reliability at Infineon Technologies Austria AG, Austria. He received his M.S in physics at the University of Hamburg and Ph.D. in electrical engineering at the Technical University Braunschweig, Germany. His email address is andre.kaestner@infineon.com.