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

SURVEY OF RECENT ADVANCED STATISTICAL MODELS FOR EARLY LIFE FAILURE PROBABILITY ASSESSMENT IN SEMICONDUCTOR MANUFACTURING

Daniel Kurz

Department of Statistics
Alpen-Adria-Universitaet Klagenfurt
Universitaetsstraße 65-67
9020 Klagenfurt, AUSTRIA

Horst Lewitschnig

Infineon Technologies Austria AG
Siemensstraße 2
9500 Villach, AUSTRIA

Jürgen Pilz

Department of Statistics
Alpen-Adria-Universitaet Klagenfurt
Universitaetsstraße 65-67
9020 Klagenfurt, AUSTRIA

ABSTRACT

In semiconductor manufacturing, early life failures have to be screened out before delivery. This is achieved by means of burn-in. With the aim to prove a target reliability level and release burn-in testing of the whole population, a burn-in study is performed, in which a large number of items is investigated for early life failures. However, from a statistical point of view, there is substantial potential for improvement with respect to the modeling of early life failure probabilities by considering further available information in addition to the performed burn-in studies. In this paper, we provide ideas on how advanced statistics can be applied to efficiently reduce the efforts of burn-in studies. These ideas involve scaling the failure probability with respect to the sizes of the different products, as well as taking advantage of synergies between different chip technologies within the estimation of the chips' failure probability level.

1 INTRODUCTION

With the usage of semiconductor products in many safety-critical applications like personal and public transportation systems (cars, planes, trains, etc.) or environmentally friendly energy generators (offshore wind parks, photovoltaics, etc.), there is an increased focus on the reliability of semiconductor devices. Deviations caused by weak semiconductors produce huge direct and total economic costs.

The bathtub curve as depicted in Figure 1 is the most widely accepted model for describing the development of the chips' hazard rate $\lambda(t)$ over time (Wilkins 2002). Based on that, the lifetime of electronic devices can be divided into three different periods of life:

- the early life phase, which starts with an increased $\lambda(t)$ and ends up with a reduced hazard rate (that is, the majority of latent defects is assumed to fail within the early life phase),
- the useful life, in which $\lambda(t)$ is constantly low, and
- the wear-out phase, in which $\lambda(t)$ is increasing.

With the aim to reduce the hazard rate of the produced devices already before the delivery, the chips' early life has to be eliminated. This is done by putting the final chips under accelerated temperature and voltage

978-1-4799-7486-3/14/$31.00 ©2014 IEEE

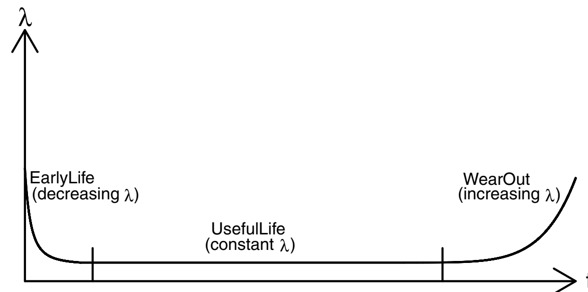

Figure 1: Bathtub curve describing the behavior of the hazard rate of semiconductor devices over time.

stress conditions for a certain period of time. We refer to this as burn-in (BI), see (Gerstle and Lee 2005), (Kuo et al. 1998), (Jensen and Petersen 1982). Note that the actual time of the BI process for a chip depends on the application. Typical values are:

- up to 10 hours for consumer products,
- up to 48 hours for automotive devices and
- up to 96 hours for aviation modules.

However, the whole population of a product has to be put under BI stress (100% BI), which involves high efforts in terms of costs, time and engineering resources. For that reason, it is essential to optimize the handling of BI (Riordan et al. 2005), (Kuo and Kuo 1983).

From a statistical point of view, BI time can be reduced by investigating a sample of the devices for an evidence of a reduced failure rate within a certain time point of early life. In general, we distinguish between two approaches for reducing the BI duration (Barlow and Proschan 1975).

The first approach aims at fitting a probability distribution (e.g. a Weibull distribution $Wb(a,b)$ (Weibull 1951) with scale parameter $a > 0$ and shape parameter $b < 1$) to (censored) failure times recorded from the BI analysis. In this way, BI duration can be lowered given an evidence of a reduced failure rate (Ooi et al. 2007), (Reliability 2007).

The idea of the second approach is to demonstrate a target failure probability level p_{target} for the produced devices in a BI study. As long as p_{target} is not met, the rest of the produced devices undergo 100% BI screening. Once p_{target} is reached, 100% BI screening is released. To retrieve information on the devices' early life failure probability level p, we investigate a large sample of the stressed devices for BI relevant failures (e.g. metalization residues, particles in oxide, crystal defects, etc.). Based on the number of observed failures in the BI study, the product's failure probability level p can be estimated. If this estimation is below the predefined target failure probability level, full BI testing is released and a BI monitoring procedure is initiated. The interested reader finds further information on how to assess the lifetime distribution of early life failures given the outcomes of a BI study in (Kurz et al. 2014b).

Since all produced items have to be put under BI stress until the target failure probability level has been verified, it is of particular importance to reach p_{target} before the start of the product's ramp-up phase. Thus, an efficient handling of BI studies is required (e.g. to avoid a restart of a BI study in case of occurred failures). This is directly related to the need of an efficient modeling of the devices' failure probability p. In this paper, we provide an overview on novel estimation methods for the chip failure probability p, which include further available information for improving the efficiency of BI studies. These methods involve

- a model for BI failures, which are tackled by countermeasures implemented in the chip production process (Kurz et al. 2014c),

- an estimation concept for the failure probability p, which takes advantage of synergies (e.g. comparable chip layers) among the different chip technologies (Kurz et al. 2014e),
- an approach for handling BI studies on multiple reference products with different chip sizes (Kurz et al. 2014d), and
- a novel area scaling model, which is capable of scaling differently reliable chip subsets separately from each other (Kurz et al. 2014a).

All of these models then contribute to a reduction of the efforts of BI testing (e.g. less burnt devices, reduced BI equipment, less engineering resources, etc.), and lead to a faster closure of the BI study.

2 EXACT ESTIMATION OF EARLY LIFE FAILURE PROBABILITIES

In a BI study, a large number of n items is randomly selected from a product's population and inspected for early failures. From a statistical point of view, this random experiment can be described as a Bernoulli trial assuming a sufficiently large number of items in the underlying population. Therefore, the random number of failures X within the drawn sample can be modelled to follow a binomial distribution, $X \sim Bi(n, p)$. The aim of a BI study is then to demonstrate (with high certainty) that the true p is below a target failure probability level p_{target}. This requires to assess an upper bound for p at a certain confidence level (CL) given k failures out of n inspections in the BI study.

In general, there are two groups of methods for assessing an upper bound $\widehat{p}$ at the $(1-\alpha)\cdot 100\%$ CL: the group of exact methods ensuring a coverage probability $P(p < \widehat{p}) \geq 1-\alpha$ and the class of approximative methods generally providing a lower $\widehat{p}$, while having a coverage probability smaller than $1-\alpha$ for selected values of p (Brown et al. 2001), (Brown et al. 2002). With regard to the assessment of the reliability of semiconductor products, an underestimation of p is seen as much more critically than an overestimation. For that reason, we prefer the application of an exact method for assessing an upper bound for p.

The Clopper-Pearson (CP) estimation model is the most widely accepted method for assessing an exact estimator of p at the $(1-\alpha)\cdot 100\%$ CL (Clopper and Pearson 1934), (Thulin 2013). We find the CP estimator of p under the condition that $P(X \leq k|n, \widehat{p}) = \alpha$. With regard to the computation of $\widehat{p}$, we exploit that $P(X \leq k|n, \widehat{p}) = P(p > \widehat{p}|k, n)$ with $(p|k, n) \sim Be(k+1, n-k)$ as derived in (Kurz et al. 2014b). Hence, the CP estimator of p is computed as the $(1-\alpha)\cdot 100\%$-quantile of a $Be(k+1, n-k)$-distribution. Supposing k failures in the BI study, we can then determine the required number of inspections for reaching a predefined target failure probability level p_{target} by solving $P(X \leq k|n, p_{target}) = \alpha$ with respect to n.

3 BI STUDIES AND COUNTERMEASURES

In general, a sample with zero failures out of n devices is required in order to reach the target failure probability level and therefore, to release 100% BI testing. The classical procedure in case of failures in the BI study is to implement countermeasures (CM's) (e.g. optical inspections, process and design measures, increase of the test coverage at electrical testing and post-processing, etc.) in the production process and to repeat the BI study.

However, the implemented CM's reduce the likelihood for early life failures. Therefore, our new approach is to consider the effectivenesses of the CM's (which are determined together with the quality department to avoid an overestimation) within the assessment of the early failure probability level p, see (Kurz et al. 2014c), (Lewitschnig and Lenzi 2014). In this way, the required number of additional inspections n_{add} for reaching the target failure probability level p_{target} can be essentially reduced and therefore, a restart of the BI study is not necessary any more. This further means to earlier release full BI testing of new products and thus, to reduce the overall effort of BI.

Example. Suppose a target failure probability level of $p_{target} = 23$ ppm at 90% CL, which at least requires $n \approx 100$k inspections (with zero failures). Let us further assume to have $k = 1$ failure out of $n = 100$k tested devices, which would actually demand the BI study to be restarted. More efficiently, a CM is introduced

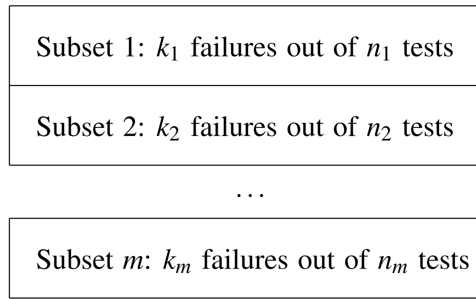

Figure 2: Illustration of data situation after integrating synergies among different chip technologies.

in the production process, which we further assess to be 80% effective in detecting the observed failure mode already before the BI. Based on the model in (Kurz et al. 2014c), we can then reestimate the failure probability level and determine the required number of additional inspections for reaching p_{target}. In this example, we obtain $\widehat{p} = 27.39$ ppm at 90% CL taking account of the effectiveness of the CM and therefore, just have to burn additional $n_{add} \approx 19.1$k devices (with zero failures) to reach p_{target}, instead of burning 100k pieces again. This leads to less BI efforts.

4 BI STUDIES WITH SYNERGIES BETWEEN DIFFERENT CHIP TECHNOLOGIES

BI studies are performed individually for each chip technology. However, if we partition the devices into several chip subsets (e.g. logic and DMOS, or package and die, etc.), one often observes synergies among the different technologies (that is, the technologies are comparable with respect to selected parts of the chips). Thus, the idea is to exploit these synergies by merging the BI information of the comparable chip subsets, which finally means to obtain a certain number of failures and a certain number of inspections for each partition of the product, see Figure 2. We then apply the model as presented in (Kurz et al. 2014e) to compute an estimate of p at the $(1-\alpha)\cdot 100\%$ CL given the data in Figure 2, and finally, to determine the required number of additional inspections for reaching the target failure probability level (in case of occurred failures). Moreover, an extended version of the model in (Kurz et al. 2014e) enables to combine synergies with CM's implemented in the chip production process.

Example. Let us again suppose a target failure probability level of $p_{target} = 23$ ppm at 90% CL and a BI study with $k = 1$ failure out of $n = 100$k tested devices. Classically, the BI study has to be restarted. However, we observe that the failure is located in a chip subset, which has been inspected for 500k times with no failures in the course of BI testing of some related technology. This eventually means to have $k_1 = 1$ failure out of $n_1 = 600$k inspections for the first subset and $k_2 = 0$ failures out of $n_2 = 100$k tests for the second part of the product. Applying the estimation concept of (Kurz et al. 2014e), we then find that $\widehat{p} = 26.71$ ppm at 90% CL and that we just have to burn an additional number of $n_{add} \approx 18.5$k devices (with zero failures) for finalizing the BI study. If the failure is additionally tackled by a CM with 80% effectiveness, we can again reestimate the failure probability level based on (Kurz et al. 2014e) and find that we just have to extend the running BI study by $n_{add} \approx 3.6$k items in order to reach p_{target}. Briefly, a restart of the BI study is not necessary any more, implying less efforts associated with BI.

5 AREA SCALING OF EARLY LIFE FAILURE PROBABILITIES

5.1 Classical area scaling

Products from the same chip technology typically just differ with respect to their chip sizes. To avoid BI studies for all of these products, the idea is to assess the technology's failure probability level on a reference product with chip size A [mm^2] and scale the estimated failure probability level to a follower product with chip size A' [mm^2]. Modeling a chip as a serial system of equally reliable chip areas, this

means to compute

$$\widehat{p}' = 1 - (1 - \widehat{p})^{A'/A}, \tag{1}$$

where $\widehat{p}$ and $\widehat{p}'$ are the estimates of the failure probability of the reference and the follower product. Thus, if p_{target} is met for the reference product, full BI testing can be skipped for follower products with size $A' \leq A$. However, if $A' > A$, the additional number of tests n_{add} in the BI study of the reference product required for reaching p_{target} for the follower product is given by

$$P\Big(X \leq k | n + n_{add}, 1 - (1 - p_{target})^{A/A'}\Big) = \alpha. \tag{2}$$

Example. Suppose to have a reference product with chip size $A = 12.64$ mm^2, for which the target failure probability of $p_{target} = 23$ ppm at 90% CL is reached with $k = 0$ failures out of $n \approx 100$k devices. Referring to Eq. (1), the failure probability of a follower product with size $A' = 15.42$ mm^2 is then estimated as $\widehat{p}' = 28.06$ ppm at 90% CL and therefore, $n_{add} \approx 22.5$k additional inspections are required in the BI study of the reference product for reaching p_{target} for the follower product.

5.2 Area scaling with multiple reference products

In the classical situation, there is a single reference product for which a BI study is performed. However, there are also cases, in which we have to deal with multiple reference products with different chip sizes. For each of the reference products, a BI study is performed. The idea is then to use the information of all the BI studies when scaling the failure probability to a follower product (as introduced in Section 5.1). This requires a model for adding failures on the differently sized reference products. Such a model is presented in (Kurz et al. 2014d). The main idea of this model is to treat multiple reference products as being built up of a different number of equally sized parts with equal failure probability, which is then estimated according to the BI information of all of the reference products. In this way, exact estimates of the failure probability of the follower products can be obtained. Moreover, the model in (Kurz et al. 2014d) enables to determine the required number of additional inspections in each of the BI studies for reaching the target failure probability level for the follower product.

Example. Suppose that there are two reference products with chip sizes $A_1 = 5$ mm^2 and $A_2 = 7.5$ mm^2. Let us further assume that the performed BI studies show $k_1 = 0$ failures out of $n_1 = 100$k and $k_2 = 1$ failure out of $n_2 = 100$k, respectively. We then treat the two reference products to be built up of two and three parts of size 2.5 mm^2 with equal failure probability $p_{2.5mm^2}$, respectively. According to (Kurz et al. 2014d), we find that $\widehat{p}_{2.5mm^2} = 7.80$ ppm at 90% CL. Using Eq. (1), the failure probability p' of a follower product with chip size A' [mm^2] is then estimated by using $\widehat{p}' = 1 - (1 - \widehat{p}_{2.5mm^2})^{A'/2.5}$. Therefore, for

$$A' \leq 2.5 \cdot \log\big(1 - p_{target}\big) / \log\big(1 - \widehat{p}_{2.5mm^2}\big), \tag{3}$$

we obtain $\widehat{p}' < p_{target}$ and therefore, we can skip full BI testing of the follower product. Taking e.g. $A' = 10$ mm^2 and $p_{target} = 23$ ppm at 90% CL, we again make use of the model in (Kurz et al. 2014d) to find that additional $n_{add} \approx 59$k items in the BI study of the larger reference product are needed to reach p_{target} for the follower product.

5.3 Area scaling with differently reliable chip subsets

Classically, area scaling is done assuming each chip subset (e.g. logic, DMOS, analog, etc.) to have the same failure probability per mm^2. However, the different chip subsets show individual production and testing conditions (e.g. different test coverage). Thus, the assumptions of the classical area scaling might not be reflected by the number of failures on each of the chip subsets. In this case, the chip subsets have

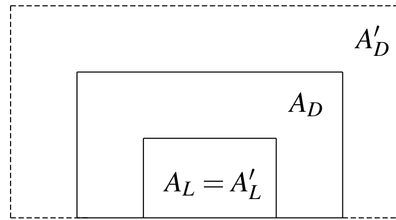

Figure 3: Illustration of separate area scaling of logic and DMOS.

to be treated separately within the failure probability scaling. This means to generalize the classical area scaling as defined by Eq. (1) using (Kurz et al. 2014a)

$$p' = 1 - \prod_{i=1}^{m} (1 - p_i)^{A'_i/A_i}, \tag{4}$$

where A_i [mm^2] and A'_i [mm^2] are the sizes of the i-th chip subset on the reference and the follower product, and p_i denotes the failure probability of the i-th chip subset, $i = 1, \ldots, m$. In (Kurz et al. 2014a), we then show how to assess estimators of the subset failure probabilities p_i given the number of failures in each of the chip subsets. By scaling the subsets separately from each other, we finally achieve a more efficient estimation of the failure probability of the follower product, especially if the reference and the follower product are only different with respect to the sizes of some of the subsets. In this way, the required number of inspections needed for reaching the target failure probability for the follower product can be essentially reduced in comparison to the classical area scaling.

Example. Suppose that we partition a reference product into logic and DMOS areas with sizes $A_L = 2.5$ mm^2 and $A_D = 5$ mm^2. Let us further assume to have a low-ohmic follower product with a larger DMOS of size $A'_D = 10$ mm^2 and equal logic, i.e. $A'_L = A_L = 2.5$ mm^2. The situation is illustrated in Figure 3. In the BI study of the reference product, we observe $k = 1$ failure out of $n = 100$k tested devices. In this case, the classical area scaling as defined by Eq. (1) assesses the failure probability of the follower product as $\widehat{p}' = 64.82$ ppm at 90% CL assuming an equal failure probability level per mm^2 for logic and DMOS. However, by having a more differentiated look on the observed failure, we find that the failure is in the logic, which provides strong evidence that the logic has a higher ppm-level per mm^2 in comparison to the DMOS. This further means that the classical area scaling overestimates the failure probability of the DMOS and in the same way, the failure probability level of the follower product. With the model in (Kurz et al. 2014a), we overcome this drawback by inferring estimates of the failure probabilities of logic and DMOS according to the given failure constellation (i.e. one failure in the logic and no failures in the DMOS) and scaling the DMOS separately from the logic. In this way, we obtain $\widehat{p}' = 51.96$ ppm at 90% CL, which is significantly lower than the original estimate. This further means that a reduced number of additional inspections is needed to reach the target failure probability for the follower product.

6 CONCLUSIONS AND OUTLOOK

6.1 Conclusions

In this paper, an overview on novel statistical models for assessing the early life failure probability of semiconductor devices has been provided. These models take account of additionally available information, including effectiveness values of countermeasures implemented in the chip production process, synergies between different chip technologies and the chip sizes of the tested products, in order to efficiently estimate the chips' failure probability level. In particular, these methods ensure a more efficient handling of BI studies and therefore lead to reduced efforts associated with BI.

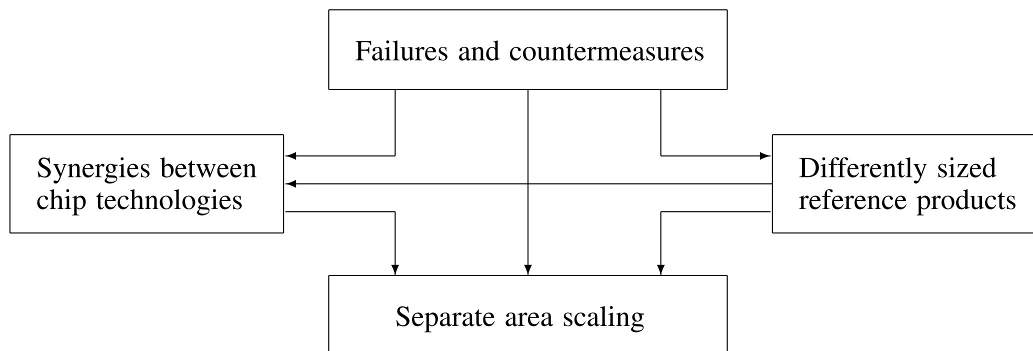

Figure 4: Illustration of possible model interactions.

6.2 Outlook

Originally, the presented models have been developed separately from each other. Nevertheless, in future, we plan to further investigate possible combinations of the proposed methods. For instance, the modeling of synergies between different chip technologies, which do not match in the size of the synergetic components, would require a combination of the model for chip synergies and the model for multiple differently sized reference products. Figure 4 summarizes the possible model interactions.

Apart from that, another aim in the future is to investigate the influence of the production process onto the occurrence of early life failures. Data from various process steps are recorded and can therefore be used to identify those levers in the production process, which significantly contribute to early life failures. In this way, failures can be detected already before the BI, which eventually reduces the risk of failure occurrences in a BI study.

ACKNOWLEDGMENTS

The work has been performed in the project EPT300, co-funded by grants from Austria, Germany, Italy, The Netherlands and the ENIAC Joint Undertaking. This project is co-funded within the programme ”Forschung, Innovation und Technologie für Informationstechnologie” by the Austrian Ministry for Transport, Innovation and Technology.

REFERENCES

Barlow, R., and F. Proschan. 1975. *Statistical Theory of Reliability and Life Testing*. New York: Holt, Renerhart & Winston.

Brown, L. D., T. T. Cai, and A. DasGupta. 2001. “Interval Estimation for a Binomial Proportion”. *Statistical Science* 16 (2): 101–133.

Brown, L. D., T. T. Cai, and A. DasGupta. 2002. “Confidence intervals for a binomial proportion and asymptotic expansions”. *The Annals of Statistics* 30 (1): 160–201.

Clopper, C. J., and E. S. Pearson. 1934. “The use of confidence or fiducial limits illustrated in the case of the binomial”. *Biometrika* 26 (4): 404–413.

Gerstle, D., and P. Lee. 2005. “Impact of Burn-In on Power Supply Reliability”. *Power Electronics Technology*: 20–25.

Jensen, F., and N. E. Petersen. 1982. *Burn-In*. New York: Wiley.

Kuo, W., W. T. K. Chien, and T. Kim. 1998. *Reliability, Yield, and Stress Burn-in*. Norwell, MA: Kluwer Academic Publishers.

Kuo, W., and Y. Kuo. 1983. “Facing the headaches of early failures: A state-of-the-art review of burn-in decisions”. In *Proceedings of the IEEE*, Volume 71, 1257–1266.

Kurz, D., H. Lewitschnig, and J. Pilz. 2014a. "An Advanced Area Scaling Approach for Semiconductor Burn-in". *To appear in: Microelectronics Reliability*.

Kurz, D., H. Lewitschnig, and J. Pilz. 2014b. "Advanced Bayesian Estimation of Weibull Early Life Failure Distributions". *Quality and Reliability Engineering International* 30 (3): 363–373. DOI: 10.1002/qre.1577.

Kurz, D., H. Lewitschnig, and J. Pilz. 2014c. "Decision-Theoretical Model for Failures Tackled by Countermeasures". *IEEE Transactions on Reliability* 63 (2): 583–592. DOI: 10.1109/TR.2014.2315952.

Kurz, D., H. Lewitschnig, and J. Pilz. 2014d. "Failure Probability Estimation for Differently Sized Products in Semiconductor Manufacturing". *To appear in: Applied Stochastic Models in Business and Industry*.

Kurz, D., H. Lewitschnig, and J. Pilz. 2014e. "Modeling of chip synergies for failure probability estimation in semiconductor manufacturing". *To appear in: Journal of Applied Statistics*.

Lewitschnig, H., and D. Lenzi. 2014. "GenBinomApps: Clopper-Pearson Confidence Interval and Generalized Binomial Distribution". *R-package*.http://cran.r-project.org/web/packages/GenBinomApps/index.html.

Ooi, M. P. L., Z. A. Kassim, and S. N. Demidenko. 2007. "Shortening Burn-In Test: Application of HVST and Weibull Statistical Analysis". *IEEE Transactions on Instrumentation and Measurement* 56 (3): 990–999.

Reliability Edge. 2007. "Guidelines for Burn-in Justification and Burn-in Time Determination". *ReliaSoft Publishing* 7 (2).

Riordan, W. C., R. B. Miller, and E. R. S. Pierre. 2005. "Reliability Improvement and Burn In Optimization through the Use of Die Level Predictive Modeling". *International Reliability Physics Symposium*: 435–445.

Thulin, M. 2013. "The cost of using exact confidence intervals for a binomial proportion". *ARXIV*. arXiv: 1303.1288v1 [math.ST].

Weibull, W. 1951. "A Statistical Distribution Function of Wide Applicability". *Journal of Applied Mechanics*: 293–297.

Wilkins, D. J. 2002. "The Bathtub Curve and Product Failure Behavior. Part One - The Bathtub Curve, Infant Mortality and Burn-in". *HotWire: The eMagazine for the Reliability Professional* 21. http://weibull.com/hotwire/issue21/hottopics21.htm.

AUTHOR BIOGRAPHIES

DANIEL KURZ received a bachelor degree in technical mathematics and data analysis at the Alpen-Adria University (AAU) of Klagenfurt/Austria in 2010 and a master degree in technical mathematics with main focus on statistics at the AAU Klagenfurt/Austria in 2012. He has been involved in research topics from the semiconductor industry since 2010 and is currently working toward the PhD degree in statistics at the AAU Klagenfurt/Austria. His research focus is on reliability models, Bayesian statistics and statistical decision theory. His email address is daniel.kurz@aau.at.

HORST LEWITSCHNIG received a master degree in technical physics at the Technical University of Vienna in 1998 and a PhD in statistics at the Alpen-Adria University in Klagenfurt/Carinthia in 2009. He works in electronic industry since 1998, mainly in quality for design. His research focus is on statistical models that are needed in the electronic industry and covers reliability, data analysis and stochastic modelling. His email address is horst.lewitschnig@infineon.com.

JÜRGEN PILZ is Professor of Applied Statistics at the Alpen-Adria-University (AAU) of Klagenfurt, Austria. He has authored more than 100 publications in international journals and conference proceedings in the areas of Bayesian statistics, spatial statistics, environmental and industrial statistics, statistical quality control and design of experiments. He is the author of six books which appeared in internationally renowned publishing houses such as Wiley, Springer and Chapman and Hall, and authored several book chapters.

He obtained his M.Sc., PhD and D.Sc. degrees from Freiberg Technical University in Germany in 1974, 1978 and 1988, respectively, in the areas of mathematics and statistics. Since 2007 he serves as head of the department of Applied Statistics at AAU Klagenfurt, Austria. He is elected member of the International Statistical Institute (ISI) and the Institute of Mathematical Statistics (IMS). He is on editorial boards of several international statistics journals. His email address is juergen.pilz@aau.at.