Proceedings of the World Congress on Engineering and Computer Science 2009 Vol II
WCECS 2009, October 20-22, 2009, San Francisco, USA

Bayesian Lifetime Modeling for Power Semiconductor Devices

Olivia Bluder *†   Alfred Waukmann†‡

Abstract— **Modeling and predicting lifetimes of smart power ICs has become more and more important during the last years. Higher demands on reliability and economy require prediction methods which save time and money. In this work two modeling approaches are discussed: (a) Bayesian linear models and (b) Bayesian linear models with mixed distributions. Both are based on the test parameters and include prior information. The information for the parameters of the prior distributions is taken from previous tests and the prior distribution itself is selected with help of global and local sensitivity analysis.**

Keywords: Bayesian inference, linear models, semiconductor lifetime

1 Introduction

In semiconductor industry the reliability of a device is essential, but gaining this information is not always straight forward. For the decision whether a device fulfills the requirements, lifetime tests are necessary. These tests are time and cost consuming, therefore reliable methods for predicting lifetime are needed.
The measured lifetimes of the devices under test (DUT) can not be modeled with known acceleration models like Arrhenius or Coffin Manson[1], therefore Bayesian Linear Models (LM) are used. The advantages of this method are: (a) available prior knowledge of previously measured data is integrated into the model, so not only current data have an impact on the model parameters. (b) higher flexibility and more information in the prediction.
In the first part the data and their characteristics will be described. Then follows the model definition, the prior selection and the global and local sensitivity analysis. In the last part of this work an advanced model is investigated and an outlook for further investigation topics will be presented.

2 Used Data

For this study datasets containing lifetimes of Smart Power ICs [2], tested with a temperature cycle stress test system [3] at KAI, have been used. Smart means that each device includes several protection functions against over-temperature, over-current, open load, etc. These devices are frequently used in automotive applications, e.g. to replace mechanical relays.
The power switches have been tested under different electrical stress conditions. The test system measures and records the state of every DUT. The Cycles to Failure (CTF) of each DUT are determined by the test parameters. The four most important parameters are:

- clamping voltage (V_{Cl}[V])
- peak current ($\hat{I}$[A])
- pulse length (t_p[μs])
- repetition time (frequency) (t_{rep}[ms])

The model is based on 8 tests, all with the same device type and the same package. In these tests the four above mentioned test parameters have been changed. According to previous investigations [4] it is known that the data follow a log-normal distribution, hence on the x-axis the logarithmic CTFs and on the y-axis the quantiles of the normal distribution are plotted to achieve linearity. The lifetimes of the DUTs for each test are shown in figure 1.

The gap between the first two tests (*T12*, *T15*) and the others is not attributable to significantly higher stress. This leads to the assumption that two failure mechanisms are dominating. To verify this, devices from both groups have been sent to the failure analysis (FA) with the results shown in figure 2.
Devices of the second group (right side) show a burn mark, but for the first group (left side) no obvious failure cause can be identified, because focused ion beam (FIB) analysis showed that cracks in the top metalization, which can be found at all devices from the first group, are only superficial. These results strengthen the theory of two failure mechanisms.

*Alpen-Adria-Universität Klagenfurt, Universitätsstr. 65-67, 9020 Klagenfurt, Austria, Tel: 0043-4242-34890-19, Email: olivia.bluder@k-ai.at
†KAI - Kompetenzzentrum für Automobil- und Industrieelektronik GmbH, Europastrasse 8, 9524 Villach, Austria
‡Carinthia University of Applied Sciences, Europastr. 4, 9524 Villach, Austria, Email: alfred.waukmann@k-ai.at
This work was jointly funded by the Federal Ministry of Economics and Labour of the Republic of Austria (contract 98.362/0112-C1/10/2005) and the Carinthian Economic Promotion Fund (KWF) (contract 18911 | 13628 |19100)

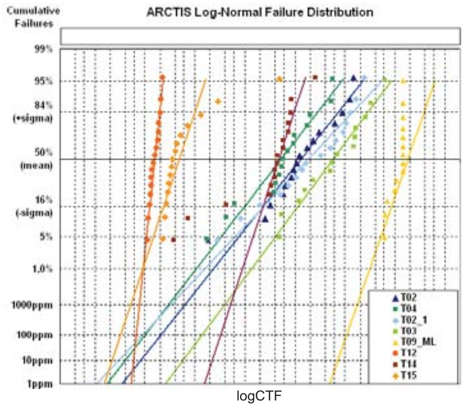

Figure 1: Results of tests

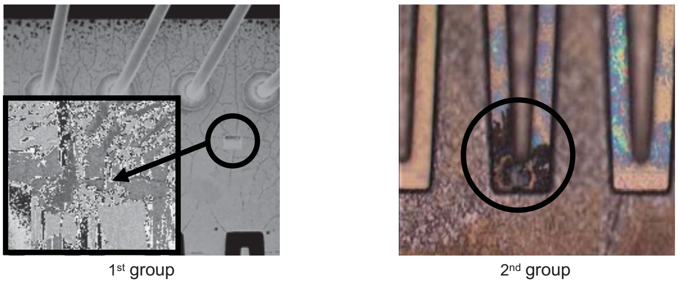

Figure 2: FA pictures

As a first modeling approach a linear model (LM) will be used, in section 7 the assumption of two failure mechanisms will be integrated into the model, which implies a mixture of two distributions.

3 Bayesian Linear Model Theory

Bayesian LMs are derived from the Bayesian law[5], which is:

$$p(\theta|\underline{y}) = p(\underline{y}|\theta) \cdot \frac{p(\theta)}{p(\underline{y})} \tag{1}$$

with $p(\theta|\underline{y})$ the posterior distribution (the distribution of the model parameters (θ) after measuring the CTF), $p(\underline{y}|\theta)$ the joint probability for the given data vector $\underline{y}$ and the model parameters, $p(\theta)$ the prior distribution of the model parameters (before measuring the CTF) and $p(\underline{y})$ the probability of the given data, which can be neglected under the assumption of proportionality, because it is constant. Furthermore the joint probability can be expressed by the likelihood function ($L(\cdot)$) of the data. The likelihood is the product of the probabilities of each value (y_i) from the given set of data $\underline{y}$ dependent on the model parameters (θ):

$$L(\theta|\underline{y}) = \prod_{i=1}^{n} p(y_i|\theta) \tag{2}$$

hence the Bayesian law converts to:

$$p(\theta|\underline{y}) \propto L(\theta|\underline{y}) \cdot p(\theta) \tag{3}$$

which states that the posterior distribution is proportional to the product of the likelihood of the data and the prior distribution of the model parameters. In the prior distribution the information of previous tests will be included.
LMs can be used for normal distributed data. Since the given data follow a log-normal distribution, a logarithmic transformation leads to normal distributed data. The transformed data ($\log_{10} CTF = \underline{y}$) are normal distributed with $\underline{\mu}$ the vector of means and $\Sigma = \sigma^2 I$ (variance times identity matrix) the covariance matrix:

$$\underline{y} \sim N(\underline{\mu}, \sigma^2 I) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp^{-\frac{1}{2\sigma^2}(\underline{y}-\underline{\mu})'(\underline{y}-\underline{\mu})} \tag{4}$$

Next the dependency on the four test parameters (the covariates) is integrated into the model by using a LM for the mean μ:

$$\underline{\mu} = \underline{\beta} X = \beta_0 + \beta_1 * V_{Cl} + \beta_2 * \hat{I} + \beta_3 * t_p + \beta_4 * t_{rep} + \epsilon \tag{5}$$

where X is the matrix of normalized covariates and with $\epsilon \sim N(\underline{0}, \sigma^2 I)$ the random errors.
Combining equations 4 and 5 leads to a likelihood function dependent on a total of six model parameters:

$$L(\underline{\beta}, \sigma^2|X, \underline{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp^{-\frac{1}{2\sigma^2}(\underline{y}-X\underline{\beta})'(\underline{y}-X\underline{\beta})} \tag{6}$$

The likelihood is only the first step needed to get a posterior distribution, an appropriate prior is also required.

4 Prior Selection

Selecting the prior distribution is essential for Bayesian inference, because the influence on the posterior distribution of the parameters can be significant. If there is no scientific reason for selecting one dedicated prior distribution, basically there is no restriction for it, even an improper prior (does not fulfill the conditions for a distribution function) can be used. Nevertheless, selection must be made carefully because a bad prior can lead to a bias in the model.
Prior selection can be supported by global sensitivity analysis, a method for comparing resulting posterior distributions of possible prior distributions. In this work a set of uninformed and informed priors is used. Uninformed means that no knowledge about the parameters is given, informed means knowledge, e.g. mean and standard deviation, from given data or from experts is available.
Possible distributions for the β_is are the following:

- diffuse normal: $\beta_i \sim N(0, 10^6)$
- informed uniform: $\beta_i \sim U(m_i - 3 * s_i, m_i + 3 * s_i)$
- informed normal: $\beta_i \sim N(m_i, s_i^2)$
- non centralized student t with 1 df: $\beta_i \sim nct(m_i, 1)$
- gamma or negative gamma distributions: $\beta_i \sim \pm Gam(a_i, b_i)$

The prior information for the means $\underline{m} = (6.71, -13.85, -23.88, -16.41, 6.28)$ and the standard deviations $\underline{s} = (0.13, 1.04, 1.65, 1.15, 0.93)$ of the model parameters are extracted from the given data.
When normal priors for the β_is and an inverse gamma (IG) prior for σ^2 are used, than the resulting posterior distributions for the parameters can be calculated analytically ($\beta_i|\underline{y} \sim t$ and $\sigma^2|\underline{y} \sim IG$), but in all other cases the posterior distribution needs to be simulated numerically. This has been done with the slice sampling algorithm in MATLAB[1], with a sample size of 10000 and a burn in period of 1000.
The densities of the resulting posterior distributions for the five β parameters show similar characteristics, therefore only β_0 is visualized in figure 3. The global sensitivity analysis shows a division into two groups. Densities with less variation descend from highly informed priors (gamma and normal) and the flatter ones from diffuse and little informed priors (diffuse normal, uniform and non-centralized t). Furthermore, the posteriors differ only in shape not in location, this means that all assumed prior distributions are acceptable, but the choice for the degree of information integrated needs to be made.

[1] MATLAB R2008, MathWorks

For calculating the parameters of the prior distributions reliable data have been used and it is intended that the prior contains as much information as possible without manipulating the results, hence an informed normal distribution will be used as prior.
σ^2 requires a global sensitivity analysis too. As possible priors informed lifetime distributions (inverse gamma, log-normal and Weibull) are considered, because they are restricted to $\mathbb{R}^+$. For completeness also a diffuse normal and a uniform prior are used. Although the shapes and the degrees of information of the used priors differ significantly, the influence on the posterior distribution of σ^2 is negligible. Therefore the inverse gamma distribution will be used, because it is the conjugate prior for normal distributed data. In Bayesian modeling conjugate means that the posterior distribution is from the same family as the prior distribution[5].

After selecting a proper prior local sensitivity analysis needs to be performed, this means evaluating if the posterior distribution is sensitive to reasonable changes in the parameters of the prior distribution. Sensitive posteriors can lead to big variations in the model and poor prediction quality will be the result. Reasonable changes according to Gill [5] are shifts in prior mean of plus/minus one prior standard deviation respectively multiplying/dividing the prior standard deviation by two. If the resulting posterior of the transformed prior shows significant differences in location or shape, a less informed prior for this parameters should be chosen.
Figure 4 shows the resulting posterior densities of two parameters after local sensitivity analysis. Only β_0 and σ^2 are visualized, since they show the biggest variations. The results shown in figure 4 indicate that shifting the mean has only slight influence on the posterior distribution. Using a transformed standard deviation has no influence on σ^2, but reducing the prior standard deviation of β_0 by 50% leads to a reduce of 25% in the posterior distribution, hence this posterior distribution is sensitive to transformations. This problem can be solved by using a less informed prior (e.g. use two times the variance). With this modificated prior the effect for the mean reduces and the sensitivity to transformations in the standard deviation vanishes.

5 Model Definition

After selecting the prior the full Bayesian LM can be defined as:

$$\begin{aligned} \underline{y} &\sim N(\underline{\mu}, \sigma^2 I) \\ \underline{\mu} &= X\underline{\beta} + \epsilon \\ \beta_i &\sim N(m_i, s_i) \\ \sigma^2 &\sim IG(a, b) \end{aligned} \tag{7}$$

Equation 3 shows that the joint posterior distribution of data and model parameters is proportional to the likeli-

ISBN:978-988-18210-2-7

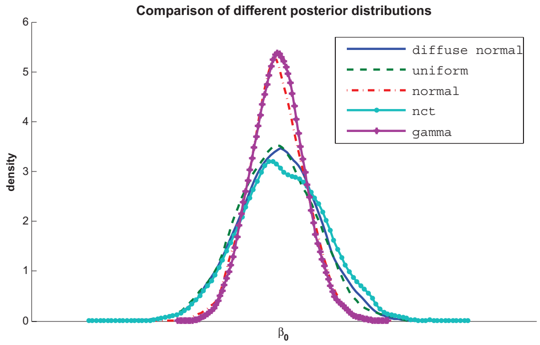

Figure 3: Simulated posterior distributions for β_0

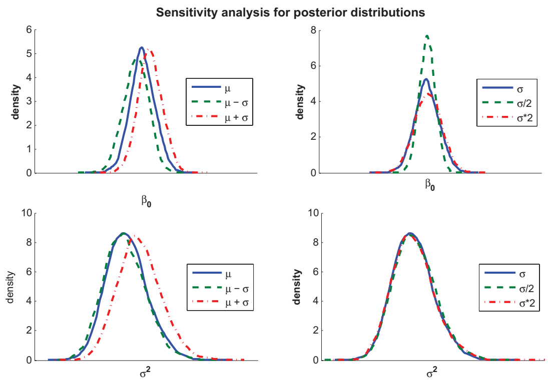

Figure 4: Results of local sensitivity analysis

hood times the prior distribution of the parameters. Integrating the specific joint posterior distribution of the model defined in equation 7 with respect to $\underline{\beta}$ and σ^2, respectively, leads to student t distributions for the β_is and to an inverse gamma distribution for σ^2. The summary statistics for the model parameters are given in table 1.

Table 1: Summary statistics of posterior distributions

parameter	mean	st.dev	Quantiles 5%	95%
β_0 (intercept)	6.71	0.08	6.58	6.84
β_1	-13.82	0.91	-15.30	-12.32
β_2	-24.03	1.14	-26.02	-22.21
β_3	-16.38	0.73	-17.61	-15.22
β_4	6.38	0.54	5.54	7.29
σ^2	0.71	0.05	0.64	0.79

The standard deviations of the model parameters vary between 1-8% of the mean and the percentages of the simulation errors are in the same range, this means that simulation results are reliable, although they might be improved since $< 5\%$ simulation error is desired.

6 Posterior Predictive Distribution

The main intention for finding an appropriate model for CTFs of semiconductor devices is to predict reliable lifetimes. In case of Bayesian LMs predictions are no point estimates but posterior predictive distributions of the data. The output is the distribution of new data after observing and including the information of the old data. For a given set of new data (y_{new}), the posterior predictive distribution is[5]:

$$\begin{aligned} p(y_{new}) &= \int_{\Theta} p(y_{new}, \theta | Y) d\theta \\ &= \int_{\Theta} \frac{p(y_{new}, \theta | Y)}{p(\theta|Y)} p(\theta|Y) d\theta \\ &= \int_{\Theta} p(y_{new}|\theta, Y) p(\theta|Y) d\theta \end{aligned} \tag{8}$$

which is the integral over the product of the joint probability of data and model parameters $(p(y_{new}|\theta, Y))$ and the posterior distribution of the model parameters $(p(\theta|Y))$. For the Bayesian LM defined in equation 7 with a new set X_{new} of covariates the posterior predictive distribution of the resulting $\log_{10} CTF$ $(= \underline{\tilde{y}})$ is:

$$p(\underline{\tilde{y}}) = \int_{\Theta} p(\underline{\tilde{y}}|\underline{\beta}, \sigma^2, X_{new}, X, \underline{y})\ p(\underline{\beta}, \sigma^2|X, \underline{y})\ d\theta \tag{9}$$

In special cases this distribution can be calculated, but since simulation data of the involved distributions are available, sampling from them is also a solution.
A first check concerning model quality can be made by predicting the CTF for the given data and looking at the goodness of the fit (GoF) with a Bayesian χ^2 test. This test is similar to Pearsons χ^2 test[6], but with the difference that it is performed various times (in this work

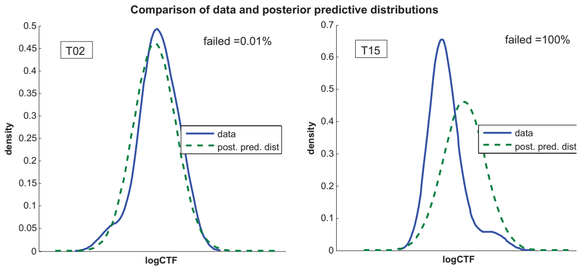

Figure 5: Comparison of predicted and real density of two datasets

10000) and the percentage of failed tests is the indicator for the model quality The comparison of the predicted and the real density of two representative datasets is shown in figure 5. The used model fits the data of test *T02* almost perfectly, because only 0.01% of the GoF tests fail. In contrast to this result 100% of the GoF tests fail for test $T15$. In total, three posterior predictive distributions fit the data well (failed GoF $< 3\%$), two show weaknesses (failed GoF $\approx 35\%$) and three show poor fitting quality (failed GoF $> 87\%$).
One reason for bad quality in the five cases was already mentioned in section 2. The data show a division into two groups with different failure mechanisms and additionally a transition zone, this means that some datasets belong to both groups, e.g. $T14$. The mathematical reasons for bad model quality for these tests are:

- $T12$ and $T09$ have smaller σ^2 than other tests
- $T05$, $T14$ and $T15$ do not behave "well", they show a mixture of two distributions

This implies that the model fits well for datasets of the second group in figure 1. Calculating the posterior predictive distribution of the next two performed tests ($T13$ and $T16$), which are mainly part of the second group, verifies the assumption. The percentage of failed GoF tests is 0% and 5%, respectively. This means the predicted values are reasonable.
The investigations show that the proposed Bayesian LM can be used for lifetime tests which are part of the second group, but model improvement is needed since weaknesses for the first group of tests are observed. One solution is to adapt the model more to the mixed behavior of the data, this will be addressed in the next section.

7 Bayesian Linear Models with Mixed Distribution

In section 2 the theory about two dominating failure mechanisms was introduced. This assumption leads to the idea of using Bayesian LM with a mixture of distributions for this data. The used model consists of a combination of two normal distributions with a mixing proportion (π), this is:

$$\underline{y} \sim \pi * N(\underline{\mu_1}, \sigma_1) + (1-\pi) * N(\underline{\mu_2}, \sigma_2) \qquad (10)$$

where μ_1 and μ_2 are modeled with LMs:

$$\begin{aligned} \underline{\mu_1} &= \underline{\beta} * X \\ \underline{\mu_2} &= \underline{\gamma} * X \end{aligned} \qquad (11)$$

Using this model more than doubles the number of model parameters, i.e. from 6 to 13. The simulation with diffuse priors showed high variations in the model parameters, therefore less informed priors than for the Bayesian LM will be considered, these are uniform distributions on a pessimistically chosen interval. The information for the intervals is extracted from the performed tests, the observations in figure 1 and the previously used Bayesian LM. This leads to the following priors:

$$\begin{aligned} \pi &\sim U(0,1) \\ \beta_0 &\sim U(2,6) \\ \gamma_0 &\sim U(6,10) \\ \beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3 &\sim U(-100,0) \\ \beta_4, \gamma_4 &\sim U(0,100) \\ \sigma_1, \sigma_2 &\sim U(0,100) \end{aligned} \qquad (12)$$

The prior distributions of β_0 and γ_0 contain the information that 10^6 seams to be the border between the two groups, π is a weighting parameter, hence it has to be chosen from the interval [0,1] and the priors of the other parameters only restrict the sign.

The simulation results (see table 2) show that the model parameters of the first distribution (β_i's and σ_1) tend to vary more than the others. A possible explanation is the lack of data, because out of the 123 data points only 35 are part of the first group.
The intention for using a mixture of distributions was to increase the quality of the model, but the higher variation in the parameters is already an indicator that the model will not fulfill the expectations.

Figure 6 shows the comparison in prediction quality and confirms the assumptions. Quality increase for one test (right side) means at the same time decrease for a test which was well explained by the Bayesian LM. Hence using a mixture of distributions does not lead to satisfying results.

ISBN:978-988-18210-2-7

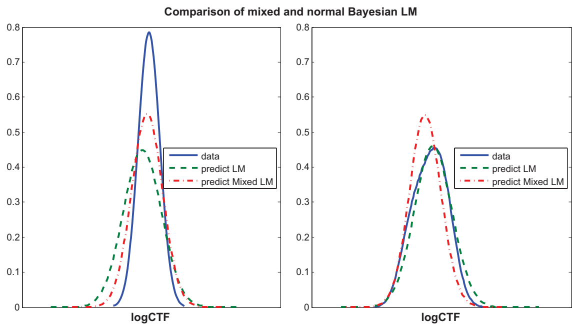

Figure 6: Comparison of predictions of the two LMs investigated

Table 2: Summary statistics of posterior distributions for model with mixed distribution

parameter	mean	st.dev	Quantiles 5%	95%
π	0.16	0.06	0.06	0.27
β_0	4.73	0.61	3.89	5.92
β_1	-3.45	2.37	-7.45	-0.26
β_2	-7.98	2.07	-11.01	-4.31
β_3	-5.61	4.12	-14.68	-1.32
β_4	2.85	1.53	0.29	5.11
σ_1	0.19	0.18	0.09	0.35
γ_0	6.85	0.11	6.58	7.03
γ_1	-13.55	1.39	-16.28	-11.30
γ_2	-24.58	1.98	-27.82	-21.32
γ_3	-14.40	1.49	-16.82	-11.96
γ_4	5.77	0.82	4.31	7.09
σ_2	0.60	0.06	0.49	0.71

8 Conclusions and Future Work

This work showed that modeling the lifetimes of power semiconductor devices with Bayesian LMs is possible, but restricted to a specific range, where tested devices have the same failure mechanisms. Expanding the prediction range showed the poor extrapolation quality of the model for tests with probably other failure mechanisms. As a step of improvement a Bayesian LM with mixed distributions was used, but no significant increase in quality could have been observed, hence further improvements and/or other model assumptions are needed.
Among all possible new or advanced approaches four have been chosen to be the most promising for the given data, these are:

- include accurate temperature measurements
- using non-linear models and incorporate known physical relationships and models for parameters
- add more prior information into the Bayesian LM with mixed distributions, e.g. model the mixing proportion (π) dependent on a parameter
- consider censored data (adapt the likelihood function)

References

[1] Escobar, L. A., Meeker, W. Q., "A Review of Accelerated Test Models, *Statistical Science*, V21, N4, pp. 552-577, 11/06

[2] Glavanovics, M., Estl, H., Bachofner, A., "Reliable Smart Power Systems ICs for Automotive and Industrial Application - The Infineon Smart Multichannel Switch Family", *43. International Conference Power Electronics, Intelligent Motion, Power Quality (PCIM 2001)*, Nürnberg, Germany, 6/01

[3] Glavanovics, M., Köck, H., Eder, H., Košel, V., Smorodin, T., "A new cycle test system emulating inductive switching waveforms", *12th European Conference on Power Electronics and Applications (EPE 2007)*, Aalborg, Denmark, pp.1-9, 9/07

[4] Bluder, O., "Statistical Analysis of Smart Power Switch Life Test Results", *Diploma thesis*, Alpen-Adria-University of Klagenfurt, Austria, March 2008

[5] Gill, J., *Bayesian methods*, Chapman & Hall/CRC, Boca Raton FL, 2008.

[6] Hamada, M. S., Wilson, A. G., Reese, C. S., Martz, H. F., *Bayesian Reliability*, Springer Science + Business Media, New York, 2007.

[7] Voiculescu, S., Guerin, F., Charki, A., "Bayesian Estimation in Accelerated Life Testing Application on Exponential-Arrhenius Model", *Reliability and Maintainability Symposium (RAMS 2006)*, Newport Beach CA, USA, p. 325-330, 1/06