ISSN: 2395-1680 (ONLINE)
DOI: 10.21917/ijme.2018.0094

FAILURE RATE ESTIMATION IN FIELD FOR A DEFECT, IN FUNCTION OF MANUFACTURING DEFECTIVITY DENSITY - CASE STUDY FOR A GATE OXIDE RUPTURE ON VALVE DRIVER IN AUTOMOTIVE SEMICONDUCTOR

Corinne Berges
Advanced Automotive Analog Quality, NXP Semiconductor, France

Abstract

Whatever industrial environment, failure rate prediction is important at different levels. Firstly, this prediction will be provided to customers when several failures with a same signature are observed in field: customers' request will fit with the number of future failures expected in field, from the number of already observed ones. A field modeling method is now typical for this case. But, before field step, failure rate prediction is performed during qualification for a new product: all the accelerated stress tests implemented in qualification aim field failure rate prediction: at this qualification step, we speak about reliability tests to guarantee product working during the full mission profile, or robustness tests to study product working limits until part breakage. In this paper, an innovative risk assessment method is presented for a gate oxide rupture defect, in automotive semiconductor industry. For the first time, a field risk assessment method uses results from a specific manufacturing test: features of this manufacturing test are not those of a reliability test nor a robustness one, and the most important input data for this assessment is defect density measured at manufacturing. A study case on a valve driver allows to precisely describe this innovative method, to identify its limits and to study possible implementation to other defects or products.

Keywords:

Automotive Semiconductor, Manufacturing Test, Industry, Failure Rate Prediction, HTOL

1. INTRODUCTION

In automotive semiconductor industry, as in any industrial environment type, failure rate prediction in customers' assembly line is looked for but this estimation is mainly important in field. When failure happens on assembly lines, defect is not latent; failure rate will no longer evolve past the assembly lines. At the contrary, when failure occurs in field, defect type is latent and failure rate may evolve in time. Latent failure prediction is all the more difficult for that. Typical latent failure rate prediction is performed from reliability test results, one of this reliability test that aims to reveal latent failures in die being the High Temperature Operating Life (HTOL) test. Unfortunately, HTOL test is not always efficient for a specific failure if test conditions are not the most appropriate ones to reveal the defect or for certain customers' application when parts are extremely solicited [1]. Another type of test exists when product breakage is looked for to study working limits. This sort of test, called robustness test, is possible during qualification steps, but not during manufacturing implemented on all the parts to be shipped to customers. HTOL test is not implemented either in manufacturing since HTOL test makes parts older. As an intermediate test, between reliability and robustness, that could be implementable in manufacturing, on all the parts to be shipped, it is possible to use a robustness test, with a very weak stress level. Breakage point is not reached but signs of early degradation are monitored: these ones could demonstrate their initial weakness [2]. Parts that do not show any degradation signs or only a low degradation level are shipped to customers, but these last ones may fail in field since they have some weakness: a specific methodology to predict failure rate on this type of parts is looked for. This is fitting with topic of this paper, and a real case study describes its implementation on an automotive semiconductor valve driver for which defect is a gate oxide rupture, screened in manufacturing step with a gate stress test. A first chapter will present the different failure rate prediction methods in field for a latent defect. Then, details on gate stress test will be given. Two failure rate prediction methods will be studied, from results for a similar product and from yield data. Finally, their limitations and possible application to other defects or products will be discussed.

2. LATENT FAILURE PREDICTION IN FIELD

Several methods to predict failure rate in field for a latent defect exist. The first one uses failures observed in field. A field modeling is performed from failure events whose unit is time spent or mileage covered in field by parts before they failed. At each failure event, failure probability is fitting with the ratio between the number of parts which already failed at this time or mileage, and the number of surviving parts that are the parts that did not fail yet [9] [10]. The curve obtained like that is modeled and a typical model is Weibull model. This modeling work is called a life distribution analysis. From the model performed on a specific time or mileage range, prediction at long term can be performed [3] [4]. A failure prediction for latent defects is always performed also in qualification step, before parts are manufactured on a large scale: reliability tests used for this prediction implement real application conditions, but with an acceleration in temperature and potentially in voltage. In these tests in which acceleration coefficient makes test duration equivalent to life duration, (for a car, 1000 test hours for 10 or 15 years), failures are no longer recorded in real time, which forbids any possibility to model life distribution, but counted at test end, which allow to estimate a maximum failure rate and quantity during whole life. Here, among latent failure prediction methods, innovation consists with using yield data on manufacturing tests and correlating them with defect density measures.

3. GATE STRESS BETWEEN A RELIABILITY TEST AND A ROBUSTNESS ONE

Manufacturing test whose results are used here for this latent failure rate prediction, is a gate stress: it is implemented at probe step, directly on die, while die is not packaged yet. It screens gate

oxide rupture that is an early life defect: gate oxide has a weakness somewhere that may fail at any moment in field, most often the time in very early life in field. Older component becomes, lower failure risk is. So, failure rate does not evolve much past the first times spent or first mileages covered in field by the parts. Gate stress test consists in over-stressing transistor gates in voltage or current, and degradation signs are monitored. This test is not implemented on all the gates, in particular on the gates with the smallest surfaces, to avoid risk to break them. It is not implemented either on the largest ones that are less likely-to- fail, considered as more resistant [8]. In this paper, two levels of gate surface for gate stress test application is expressed by,

$$\text{Gate stress application for } sa\ (mm^2) < \text{gate oxide surface} < sb\ (mm^2)$$

This test is akin to a robustness one when breakage would not be expected, or to a burn-in or reliability one when part is stressed only for the very beginning of life. Stress level to be applied to parts is set according with the observed number of failed parts. Leakage level acceptable after gate stress test or leakage difference between before and after gate stress is also adjusted. But, number of field returns has also to be taken into account, which is customers' application-dependent. So, gate stress test adjustment is performed on stress level applied and according with customers' application. Gate stress test procedure is the following one: gate leakage current is measured before gate stress, then gate stress is performed, and finally, gate leakage is measured again after gate stress. Leakage current difference between before and after gate stress shows robustness difference of all the gates: gate stress differently degraded the parts. In manufacturing, a gate broken by gate stress shows an insufficient gate robustness and a defect somewhere on the gate oxide. A failure analysis is always performed on a gate oxide rupture. As an example, one failed part showed a sleep current too high. An emission microscopy (EMMI) technique revealed an issue on a pull-down component. A gate oxide rupture was found at edge of active area during a scanning electron microscope (SEM) analysis shown in Fig.1. Even if gates are not broken by gate stress, some degradation signs may be revealed in current leakage difference between before and after gate stress (delta leakage): electric measures have evolved after gate stress.

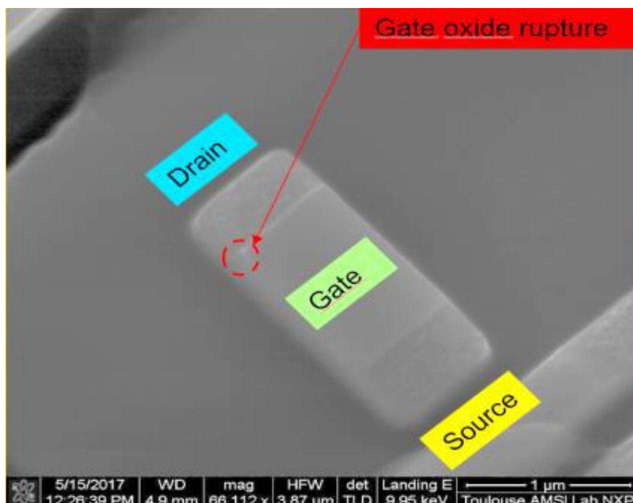

Fig.1. SEM picture of a Gate Oxide Rupture defect

Of course, failed parts are not shipped to customers. Parts for which electric characteristics did not evolve after gate stress, are no longer likely-to-fail in field and are launched into field: their gate are shown sufficiently robust. Parts for which there is a difference on gate leakage between before and after gate stress are shipped to customers in some extent that this difference stays below a specific threshold set by number of observed field returns. A too high threshold will generate field returns by gate oxide rupture. With a good threshold value, no field return by gate oxide rupture will be seen. Parts that evolved after gate stress are likely-to-fail, sensitive to this gate oxide rupture defect: it is on this part quantity that failure rate and quantity have to be estimated.

4. FAILURE RATE ESTIMATION OF GATE OXIDE RUPTURE IN FIELD, AFTER GATE STRESS IN MANUFACTURING: CASE STUDY

Two methods to estimate failure rate for field gate oxide rupture after gate stress in manufacturing, have been designed.

- A first method uses test data from another product: on this other product, what gate oxide ruptures on stressed gates with high leakage differences, were observed in field? But, failure rate depends on product technology and design, and on customers' application. Then, failure rate extrapolation from a product to another one is risky.
- A second method does not have this significant drawback and uses on one hand leakage difference measures for gate stress test in manufacturing, on the other hand a possible field failure rate estimation: method basis is a proportionality relation between leakage difference level that is accepted on parts to be shipped to customers, and field failure rate. It is obvious that yield value is linked to this leakage difference level, and the fact that yield is a function of defect density is also used: so, in this method, defect density and yield will be spoken about.

A case study deals with a valve driver semi-conductor device for which gates with oxide areas between sa mm^2 and sb mm^2 are stressed in manufacturing steps. Risk taken not testing the gate oxide areas superior to sb mm^2 is estimated: this problem is equivalent to study failure rate for gate oxides superior to sb mm^2. Gate oxide rupture is an early life failure, so failure rate does not evolve much past the first mileages in field: so, in a first study, an instantaneous estimation is sufficient.

5. FAILURE RATE ESTIMATION OF GATE OXIDE RUPTURE IN FIELD, AFTER GATE STRESS IN MANUFACTURING, FROM OTHER PRODUCT DATA

Implemented methodology is the following one: in another product than the one for which this estimation is wanted, gates with oxide areas superior to sb mm^2 and not-stressed in manufacturing, so that field gate oxide rupture returns are counted: an instantaneous risk linked to these gates, not taking account the gate oxide rupture latent nature, is estimated by the ratio between the number of returns and the total number of parts in field. But, already told, the most important limitation for this

method is that the product on which estimation is computed may be very different, in term of design, wafer manufacturing technology and customer's application than the studied product, while field gate oxide rupture failure rate is fully dependent on these three items: technology, design and application. Failure rate estimation method computing this only one ratio between return quantity and shipped part volume would be valid only for a non-latent failure, and means that failure rate will not evolve in time. At that step, it is interesting to check error made by this assumption. So, a field modeling is performed on the field gate oxide ruptures observed for the product in Fig.2. The model that provides the best fitting on the data is a Threshold Weibull one. The very low Beta parameter inferior to 0.3 confirms that gate oxide rupture is an early life defect in Table.1. However, failure rate and quantity do evolve in time: usage of an instantaneous failure rate estimation is not the best option but stays valid at estimation instant or at very short term with a low error.

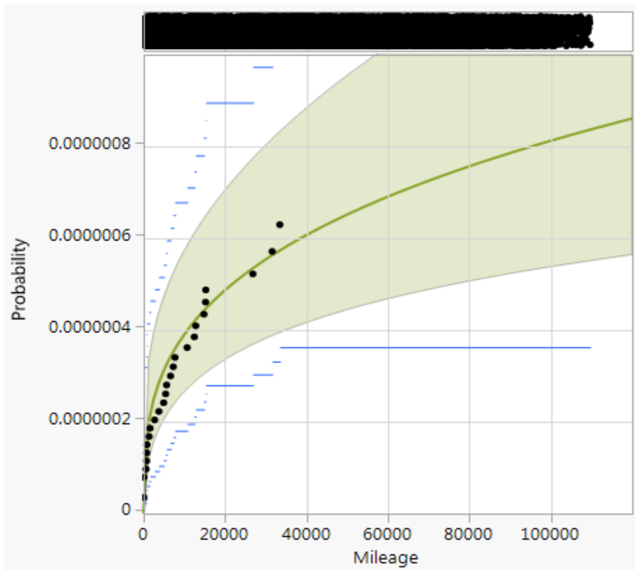

Fig.2. Field modeling for gate oxide ruptures on the automotive valve driver

Table.1. Field modeling parameters for a Weibull model

Parametric Estimate – Threshold Weibull				
Parameter	**Estimate**	**Std Error**	**Lower 95%**	**Upper 95%**
Location	55.6357381	9.03362133	37.9301656	73.3413105
Scale	3.14649344	0.62220095	1.92700198	4.3659849
Threshold	-4.441e-16	0	-	-
Weibull α	1.4531e+24	1.3127e+25	2.9707e+16	7.1077e+31
Weibull β	0.31781411	0.06284591	0.22904339	0.51894083

6. FAILURE RATE ESTIMATION OF GATE OXIDE RUPTURE IN FIELD, AFTER GATE STRESS IN MANUFACTURING FROM MEASURED YIELD

This second method is based on the assumption of a relation between yield loss due to gate stress (breakage and before-after leakage difference), and failure quantity observed in field: higher accepted yield loss is, smaller field failure quantity will be. Less parts that show degradation signs, less likely-to-fail parts in field will be: assumption taken is consistent. Furthermore, relation between overall yield loss and manufacturing defect density is taken into account: indeed, a manufacturing issue or weakness, or a particle are often failure root cause.

In the case study:

- Si is total surface of the gates for which a gate stress test has already been implemented (each gate has an area between sa mm² and sb mm²);
- Sf will be total surface for the ones for which gate stress is decided: Sf fits with sum of Si and area of the gates with an area superior to sb mm².
- Question is still the same one: failure quantity estimation due to the fact that gates with an area superior to sb mm² are not gate stressed is looked for.
- Yield loss for the surface Si is measured and known: the one for Sf has to be estimated. This fits with the first step of the methodology.

6.1 DEFECT DENSITY MEASUREMENT

Defect density is measured for each type of gate oxide: positively-doped gates are distinguished from the negatively-doped ones, and separately measured. Type of components is taken into account: for example, gates in logic components that are never stressed because of their too small gate area, are not measured in defect density. Different measures are performed also on capacitor gates. Anyway, taking into account each different gate defect density for each gate type makes far more complex yield loss estimation, because of mathematical complexity of defect density models. Then, a mean defect density value will be obtained from relation between yield and total gate oxide surface. And it will be always possible to check this mean value with the different measures to guarantee value consistency.

6.2 DEFECT DENSITY MODEL, A FUNCTION OF YIELD AND GATE AREA

A Murphy model expresses function of mean defect density D_0 with total yield Y and critical area A_C [5]. Yield fits with the overall yield, not only due to gate stress leakage difference. Here, critical area is gate oxide area:

$$D_0 = \frac{\alpha\left(\sqrt[\alpha]{1/Y} - 1\right)}{A_C} \quad (1)$$

Parameter α translates variability or uniformity of defect density:

- When defect density is uniform, this parameter takes a value superior to 10: thus, Murphy model is equivalent to an Exponential model;
- When variability is extreme, that is to say that clustering phenomena are observed in defects, Parameter α is equal to 1, and Murphy model is called and equivalent to a Seeds Exponential model.

Actually, with current values for the product $D_0 \times A_C$ in case study, all the different models are very similar which shown in Fig.3. Then the most typical of these models, that is maybe

Murphy model, is used for this estimation. In this model, parameter α is set to traduce a medium variability: α is adjusted equal to 5:

$$D_0 = \frac{5\left[\left(\frac{100}{Yield}\right)^{0.2} - 1\right]}{Die\,Size\left(cm^2\right)} \quad (2)$$

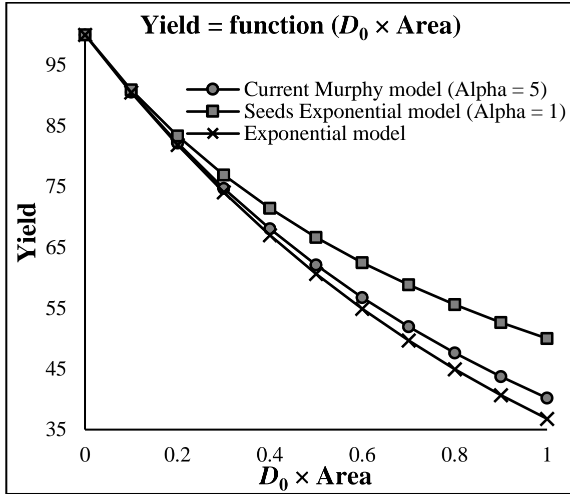

Fig.3. Yield models are similar for the current $D_0 \times A_C$

So, with this relation and with the assumption taken about defect density variability, it is now possible to link observed yield with a medium defect density D_0, and to compare the found value D_0 with each defect density measured per gate oxide type, to check consistency of this method.

6.3 SURFACE SUMMING

But, above all, previous formula linking defect density, yield and gate oxide area, is used to estimate yield for another gate oxide area. So, in the step above, yield observed for the initial gate oxide area Si (for gates with a surface between sa and sb mm²) provided a medium value D0 for defect density. Now, for D_0, yield is estimated for a new surface Sf, including the gate oxide areas superior to sb, with the previous ones. Finally, a new yield will be calculated for this surface Sf. But, this is possible since chosen defect density model (similar Murphy, Exponential or Seeds models) is nearly linear for the current values of $D_0 \times A_C$, so that summing of the surfaces can be performed, without obligation to solve an equation that would be complex as shown in Fig.4 and Fig.5.

6.4 RELATION BETWEEN GATE STRESS LEAKAGE DIFFERENCE AND FIELD FAILURE QUANTITY

From above, at this step, it was possible to estimate a value for the overall yield for the surface total Si of the gates with oxide areas between sa and sb, another value for this same overall yield for the gate oxide areas also superior to sb. Distinction is also introduced between overall yield or overall yield loss, and yield loss only due to gate stress leakage difference [6].

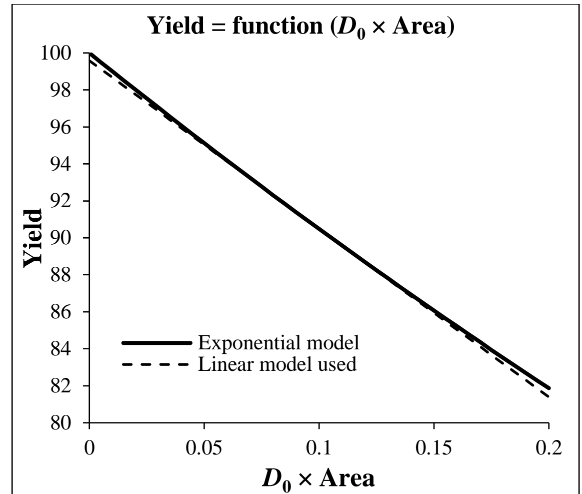

Fig.4. Linearity of the yield model for the current $D_0 \times A_C$

Linear model built on 2 points:
Point 1 area = 4.212 mm^2
Point 2 area = 5.203 mm^2
($D_0 = 2$)

Fig.5. Linearity parameters of the yield model for the current $D_0 \times A_C$

Now, for the valve driver component, a proportionality coefficient is calculated, on one hand from observed yield loss due to gate stress leakage difference, on another hand from field failure rate quantity. For another product, the most similar possible than valve driver, decrease of failure rate obtained by increasing surface of gates subjected to gate stress test, is measured: this provides a quantification of gate stress benefit on failure rate, for gate sizes superior to sb mm². Then, this quantification can be applied to valve driver: field failure quantity avoided by that fits with risk not to implement a gate stress test on the gate oxide surfaces superior to sb, which was looked for. Study may be performed for long term: same field failure rate estimation can be needed when defect density is improved, for a mature technology. It will be enough to find and apply a proportionality between failure quantity estimated for a defect density value, and new failure quantity with a defect density fitting with maturity for this technology.

7. EVALUATION AND COMPARISONS

The first method described in this paper uses only data from another product: it is the main feature of the method, but also a major drawback, since estimation dependency on technology, product design and customer's application, should forbid data usage from a product to another one. The second method, based on test yield measured and field gate oxide ruptures observed, does not have this disadvantage. However, it presents some limitations. Indeed, for a great part of the approach, no data from another product is used, except for estimation of failure quantity decrease by increase of total stressed gate surface. Mathematical complexity of such a model as Murphy one, may be considered as another real limitation: as a reminder from above chapters, surface summing is only possible because model is nearly linear for

current small values of D0× area, which widely simplifies computation. It is the same case for defect density: a medium value D0 is used, but anyway, there would not be no real possibility to use each elementary defect density measure for each type of gate oxide: how would be the formula integrating elementary defect density? Furthermore, in term of sample size, it would become not significant to work per type of oxide, number of fitting field failures becoming smaller: from a statistical point, computation or estimation significance would benefit a greater number of failures, contrary to high quality level required by automotive industry. Gate oxide rupture latency is not directly taken into account, but a parallel field modeling may give some indications about failure rate evolvement in time: it is quite reasonable to think that failure rate for gate oxide area superior to sb will evolve in time as an area between sa and sb. Lastly, results of the two methods do not present confidence level: but adding a risk level and a confidence intervals is easy and recommended [7].

8. CONCLUSION

Whatever industry, typical field risk estimation methodology is based on accelerated stress tests or on field customer return records. The innovative risk assessment method described in this paper uses manufacturing test results and a relation estimation between defect density, manufacturing yield and field failure quantity. Difficult point stays bridge estimation between what happens in manufacturing (yield loss) and field (failure rate). To quantify this relation, when no data is available on the studied component, it is always possible to observe it on another device, but this comparison decreases method efficiency, above all if studied defect occurrence for studied part depends on manufacturing technology, product design and customer's application in field, which is the typical case. However, choosing a risk level and applying it to generate confidence intervals allow to bypass this drawback and reinforce analysis consistence. This statistical aspect makes method still more applicable to any product.

REFERENCES

[1] C. Berges, A. Feybesse and W.A.R. Othman, "Reliability and Risk Assessment from Accelerated Test Result and Field Modeling: Case Study for Automotive Analog Parts and Sensors", *Proceedings of 23rd International Symposium on the Physical and Failure Analysis of Integrated Circuits*, pp. 323-327, 2016.

[2] N.A. Dumin, K. Liu and S.H. Yang, "Gate Oxide Reliability of Drain-Side Stresses Compared to Gate Stresses", *Proceedings of 40th Annual Reliability Physics Symposium*, pp. 221-224, 2002.

[3] C. Berges, Y. Chandon and P. Soufflet, "Reliability Analysis from Field Data and Prediction Models for Customer risk Assessments: Case Studies and Strategy", *Proceedings of IEEE 21st International Symposium on Physical and Failure Analysis of Integrated Circuits*, pp. 455-459, 2014.

[4] C. Berges, Y. Chandon and R. Gubian, "Innovative methodology for failure rate estimation from quality incidents, for ISO26262 standard requirements", *Proceedings of IEEE 19th International Symposium on Physical and Failure Analysis of Integrated Circuits*, pp. 233-237, 2012.

[5] G.S. May and C.J. Spanos, "*Fundamentals of Semiconductor Manufacturing and Process Control*", Wiley-IEEE Press, 2006.

[6] T.S. Barnett, M. Grady, K.G. Purdy and A.D. Singh, "Combining Negative Binomial and Weibull Distributions for Yield and Reliability Prediction", *IEEE Design and Test of Computers*, Vol. 23, No. 2, pp. 110-116, 2006.

[7] A. Agresti and B.A. Coull, "Approximate Is Better than "Exact" for Interval Estimation of Binomial Proportions", *The American Statistician*, Vol. 52, No. 2, pp. 119-126, 1998.

[8] M.A. Khalil and C.L. Wey, "High-Voltage Stress Test Paradigms of Analog CMOS ICs for Gate-Oxide Reliability Enhancement", *Proceedings of IEEE VLSI Test Symposium*, pp. 175-179, 2001.

[9] C. Berges and J. Goxe, "Benefits of Field Failure Distribution Modeling to the Failure Analysis", *Microelectronics Reliability*, Vol. 53, No. 9-11, pp. 1194-1198, 2013.

[10] M.M. Alam and K. Suzuki, "Lifetime Estimation using only Failure Information from Warranty Database", *IEEE Transactions on Reliability*, Vol. 58, No. 4, pp. 573-582, 2009.