Microelectronics Reliability 78 (2017) 319–330

Contents lists available at ScienceDirect

Microelectronics Reliability

journal homepage: www.elsevier.com/locate/microrel

Assembly yield prediction of plastically encapsulated packages with a large number of manufacturing variables by advanced approximate integration method

Hsiu-Ping Wei [a], Bongtae Han [a,*], Byeng D. Youn [b], Hyuk Shin [c], Ilho Kim [c], Hojeong Moon [c]

[a] Mechanical Engineering Department, University of Maryland College Park, MD 20742, USA
[b] Mechanical and Aerospace Engineering, Seoul National University, Republic of Korea
[c] Package Development Team, Semiconductor R&D Center, Samsung Electronics, Republic of Korea

ARTICLE INFO

Article history:
Received 20 February 2017
Received in revised form 3 August 2017
Accepted 5 September 2017
Available online xxxx

Keywords:
Approximate integration scheme
Dimension reduction
Probability density function
Assembly yield
Warpage
Stacked die thin flat ball grid array

ABSTRACT

An advanced approximate integration scheme called eigenvector dimension reduction (EDR) method is implemented to predict the assembly yield of a plastically encapsulated package. A total of 12 manufacturing input variables are considered during the yield prediction, which is based on the JEDEC reflow flatness requirements. The method calculates the statistical moments of a system response (i.e., warpage) first through dimensional reduction and eigenvector sampling, and a probability density function (PDF) of random responses is constructed subsequently from the statistical moments by a probability estimation method. Only 25 modeling runs are needed to produce an accurate PDF for 12 input variables. The results prove that the EDR provides the numerical efficiency required for the tail-end probability prediction of manufacturing problems with a large number of input variables, while maintaining high accuracy.

1. Introduction

Epoxy molding compound (EMC) has been used extensively as a protection layer in various semiconductor packaging components. The mismatch of coefficient of thermal expansion (CTE) causes the warpage of components after molding, which is one of the most critical issues to board assembly yield. The warpage issue has become more critical as Package-on-Package (PoP) and fan-out wafer level package (FO-WLP) are widely adopted for portable devices.

The computer-aided engineering (CAE) tools, such as the finite element method (FEM), have been used extensively to predict the warpage. Typically, the CAE tools provide deterministic outputs, which establish quantitative relationships between the system response (i.e., warpage) and the input parameters such as geometries, material properties, process and/or environmental conditions, etc. The deterministic approaches have been proven effective for comparing competitive designs. In reality, the package warpage behavior shows statistical variations (or probability distributions) due to inherent manufacturing variabilities. The probabilistic aspect should be incorporated in prediction if the assembly yield is to be predicted.

The yield loss is in general a small probability event (i.e., tail-end probability) [1–3], especially for the large production volume. In many cases, even 0.1% yield loss would cause a significant profit loss. Based on the Six Sigma concept, the target is often to control the yield loss within 3 to 6 sigma, i.e., 6.67% to 3.4 ppm [4].

Fig. 1 shows a schematic illustration of the tail-end probability, where the statistical property of system performance (e.g., warpage) is represented by a probability density function (PDF). When a component has the performance exceeding or falling behind a certain specification, it cannot be processed further and is regarded as a failure. The probability of all possible failure, i.e., yield loss, is the area under the PDF where the performance does not satisfy the specification.

A technical approach critically required for the yield loss prediction is the uncertainty propagation analysis, which enables the intrinsically deterministic computational model to characterize the output distribution in the presence of input uncertainties. The most popular uncertainty propagation methods are “random sampling method” and “response surface method (RSM)”. When they are applied to complex manufacturing problems with a large number of input variables, however, they become impractical due to their own limitations.

Due to its random nature, the failure probability estimated from the random sampling method, e.g., Monte Carlo simulation (MCS), exhibits statistical variations [5]. The variations can be substantial when the tail-end probability is to be predicted. In order to ensure that the tail-end

* Corresponding author.
E-mail address: bthan@umd.edu (B. Han).

https://doi.org/10.1016/j.microrel.2017.09.006

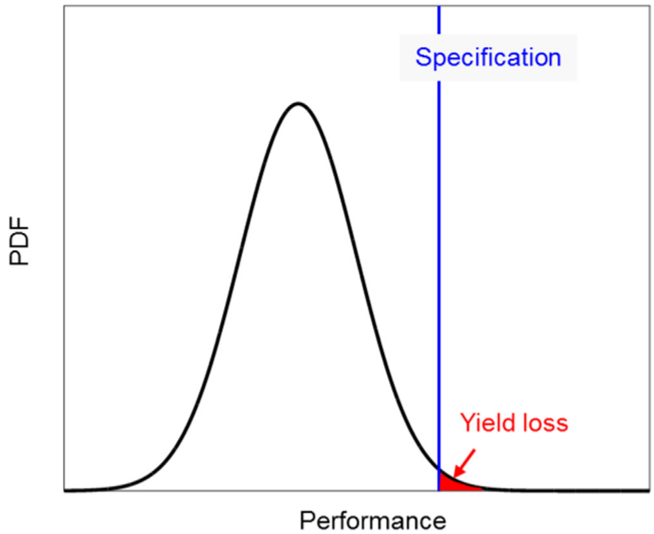

Fig. 1. Illustration of a yield loss (tail-end probability).

probability prediction falls within the specified accuracy tolerance, an extremely large number of model computations is required. This computational burden makes the random sampling impractical for the cases that require complex nonlinear computational models (e.g., visco-elastic analysis required for warpage prediction of plastically encapsulated components) [6].

The RSM has also been widely used in conjunction with the MCS [7,8] to reduce the computational burden. The RSM relies on Design of Experiments (DOE) to build computationally inexpensive mathematical response surface models, which can be used for the direct MCS. Two commonly used types of DOE are the Full Factorial Design (FFD) [9–11] and the Central Composite Design (CCD) [12–14]. Although the CCD can reduce the sample size of the FFD substantially, both types cannot avoid the challenge known as the curse of dimensionality (i.e., the computational costs increase exponentially as the number of random input variables increases). Due to this inherent limitation, the RSM has been applied to the designs with only a few input variables.

Another method for the uncertainty propagation analysis is “approximate integration scheme”. The scheme calculates the statistical moments of the output response by performing a multi-dimensional integration. Seo and Kwak proposed a numerical algorithm to perform the integration [15]. The algorithm also suffered from the curse of dimensionality as the FFD was used to select integration points. Rahman and Xu proposed the univariate dimension-reduction (UDR) method to cope with the curse of dimensionality [16]. With the method, a multi-dimensional integration is transformed into a series of one-dimensional integrations, and thus the computational cost increases only additively with the increased number of input variables. This additive increase makes the method attractive to the problems with a large number of input variables.

In a typical UDR implementation, however, a large number of numerical integration points are still required to ensure the accuracy of each one-dimensional integration result. For a large number of input variables, the method also can be computationally expensive. Youn et al. developed a method called “eigenvector dimension-reduction (EDR)” method [17] to relax the requirement of the UDR method. In the EDR method, the eigenvector sampling scheme was proposed to select a few sample points along the eigenvectors of the covariance matrix of the input variables, and the stepwise moving least square (SMLS) was implemented to interpolate and extrapolate the numerical integration points. As a result, the accuracy of statistical moment estimation by EDR remained virtually unaffected although the number of simulations was reduced substantially.

In this paper, the EDR method is implemented to predict the assembly yield of a plastically encapsulated package. A total of 12 manufacturing input variables are considered during the yield prediction, which is based on the JEDEC reflow flatness requirements. Section 2 provides a brief introduction of the EDR method. In Section 3, the details of an EDR implementation are described. The accuracy of the yield prediction is verified by the direct MCS in Section 4. Section 5 concludes the paper.

2. Eigenvector dimension reduction method

The eigenvector dimension-reduction (EDR) method estimates the complete probability density function (PDF) of a system response by (1) calculating the statistical moments and (2) constructing the corresponding PDF using the probability estimation methods.

The statistical moments are the characteristics of a distribution. The 1st moment, μ, is the mean, which represents the central tendency of the distribution, and the 2nd moment is the standard deviation, σ, which represents the spread of the distribution. The 3rd and 4th moments are skewness, β_1, and kurtosis, β_2, which indicate the symmetry and the peakedness of the distribution, respectively. The m^{th}-order statistical moment of a system response is defined as

$$E\{[\mathrm{Y}(\mathrm{X}_1,\ldots,\mathrm{X}_N)]^m\} \equiv \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\{\mathrm{Y}(\mathrm{X}_1,\ldots,\mathrm{X}_N)\}^m f_{\mathrm{X}_1,\ldots,\mathrm{X}_N}(\mathrm{X}_1,\ldots,\mathrm{X}_N)d\mathrm{X}_1\cdots d\mathrm{X}_N \tag{1}$$

where $E(\cdot)$ is the expectation operator; $\mathrm{Y}(\mathrm{X}_1,\ldots,\mathrm{X}_N)$is the system response with N random input variables, $\mathrm{X}_1,\ldots,\mathrm{X}_N$ (i.e., N dimensions); and $f_{\mathrm{X}_1,\ldots,\mathrm{X}_N}(\mathrm{X}_1,\ldots,\mathrm{X}_N)$ is the joint probability density function. In this paper, the capital letters are used to denote the input variables.

To tackle the mathematical challenge associated with the multidimensional integration in Eq. (1), Rahman and Xu proposed the additive decomposition [16] to transform the multidimensional response function $\mathrm{Y}(\mathrm{X}_1,\ldots,\mathrm{X}_N)$ into a series of one-dimensional functions. The approximated system response function, then, can be expressed as [16]:

$$Y(X_1,\ldots,X_N) \approx Y_a(X_1,\ldots,X_N) = \sum_{j=1}^{N} Y\left(\mu_1,\ldots,\mu_{j-1},X_j,\mu_{j+1},\ldots,\mu_N\right) - (N-1)\cdot Y(\mu_1,\ldots,\mu_N) \tag{2}$$

where Y_a is the approximated system response function obtained by the additive decomposition, μ_j is the mean value of an input variable, X_j, $Y(\mu_1,\ldots,\mu_{j-1},X_j,\mu_{j+1},\ldots,\mu_N)$ is the system response of the input variable, X_j, while the other input variables are kept as their respective mean values, and $Y(\mu_1,\ldots,\mu_N)$ is the system response with all input variables are fixed as their mean values.

Substituting Eq. (2) into Eq. (1) yields:

$$\begin{aligned} &E\{[\mathrm{Y}(\mathrm{X}_1,\ldots,\mathrm{X}_N)]^m\} \cong E\{[\mathrm{Y}_a(\mathrm{X}_1,\ldots,\mathrm{X}_N)]^m\} \\ &= E\left\{\left[\sum_{j=1}^{N} Y\left(\mu_1,\ldots,\mu_{j-1},X_j,\mu_{j+1},\ldots,\mu_N\right) - (N-1)\cdot Y(\mu_1,\ldots,\mu_N)\right]^m\right\} \end{aligned} \tag{3}$$

Using the binomial formula, the right-hand side of Eq. (3) can be rewritten as [16]:

$$\begin{aligned} &E\left\{\left[\sum_{j=1}^{N} Y\left(\mu_1,\ldots,\mu_{j-1},X_j,\mu_{j+1},\ldots,\mu_N\right) - (N-1)\cdot Y(\mu_1,\ldots,\mu_N)\right]^m\right\} \\ &= E\left\{\sum_{i=0}^{m}\frac{m!}{i!(m-i)!}\left[\sum_{j=1}^{N} Y\left(\mu_1,\ldots,\mu_{j-1},X_j,\mu_{j+1},\ldots,\mu_N\right)\right]^i\left[-(N-1)\cdot Y(\mu_1,\ldots,\mu_N)\right]^{m-i}\right\} \\ &= \sum_{i=0}^{m}\frac{m!}{i!(m-i)!}E\left\{\left[\sum_{j=1}^{N} Y\left(\mu_1,\ldots,\mu_{j-1},X_j,\mu_{j+1},\ldots,\mu_N\right)\right]^i\right\}\left[-(N-1)\cdot Y(\mu_1,\ldots,\mu_N)\right]^{m-i} \end{aligned} \tag{4}$$