Solder joint reliability risk estimation by AI modeling

Cadmus Yuan and Chang-Chi Lee
Department of Mechanical and Computer-aided Engineering, Feng Chia University, Taichung, Taiwan
No. 100, Wenhwa Rd., Seatwen, Taichung, Taiwan 40724, R.O.C.
cayuan@fcu.edu.tw

Abstract

This paper studies AI modeling for the solder joint fatigue risk estimation under the thermal cycle loading of redistributed wafer level packaging. The artificial neural network (ANN), recurrent neural network (RNN) and vectorized-gate network long short-term memory (VN-LSTM) architectures have been trained by the same dataset to investigate their performance for this task. The learning accuracy criterion, the implementation of all neural network architecture, the learning results and result analysis would be covered.

Because the involvement of the time/temperature-dependent nonlinearity material characteristics, it is recommended that more than three hidden layers and a proper neural network architecture, which is capable of the sequential data processing, should be considered in order to guarantee the required accuracy and the satisfied convergence speed.

Keywords: Solder joint fatigue risk estimation, Time/temperature-dependent nonlinearity, ANN, RNN, LSTM, machine learning

1. Introduction

Solder joint reliability is one of the most critical issues for most ball-grid array packaging types. Because of the mechanical characteristics of the solder joint fatigue, the nonlinear finite element (FE) method has been often applied to predict the solder joint reliability when the structure is subject to the cyclic thermal loading.

However, a decent electronic packaging FE model requires specialized modeling expertise and numerous experimental validation tests. To maximize the utilization of these valuable FE results, scholars have developed many methods. van Driel et al. [1] had applied the response surface model (RSM) to optimize the product/process designs against the failure probability estimation. The RSM has been generated from the numerical approaches with the consideration of the nonlinearity of the geometrical and material properties. Liu et al. [2, 3] applied the parametric method to study the solder fatigue failure mode under the thermal cycling loading. The model used in [3] had applied both material and geometrical nonlinearity with the precise description of the solder joint geometry. Yuan et al. [4, 5] applied the Taguchi matrix to establish a response surface of the estimated solder joint lifetime based on a validated parametric finite element model. Che et al. [6] applied the parametric method to investigate the 3D IC packaging based on through-silicon interposer and silicon-less interconnection technology by finite element modeling with experimental validation. Zhang et al. [7] applied RSM method to study the board level reliability of LED packaging. Chou et al. [8, 9] model the long-term reliability of wafer level packaging using artificial neural network architecture with multiple key design features.

Besides the statistical approaches, this paper applies the three different artificial intelligence (AI) machine learning methods to explore the new way to utilize the FE results.

An artificial neural network (ANN) is based on a collection of connected units called artificial neurons or network nodes/cells, which loosely mimic the neurons in a biological brain. The connections, like the synapses in a biological brain, can transmit a signal to other neurons. An artificial neuron that receives a signal then processes it and can signal neurons connected to it. In ANN implementations, the signal at a connection is mostly defined by a real number, and the output of each neuron is computed by the non-linear activation function of the sum of its inputs. The connections from one node to others are called edges. Neurons and edges typically have a weight that adjusts as learning proceeds. The weight increases or decreases the strength of the signal at a connection. Practically, these artificial neurons are aggregated to form the artificial layers. Different layers may perform different transformations on their inputs. Signals that travel from the first input layer to the last output layer, possibly after traversing the layers multiple times.

In the 1940s, Hebb [10] created a learning hypothesis based on the mechanism of the neural plasticity that became known as unsupervised Hebbian learning. But the development of the neural network was blocked due to the computational power limitation. A key trigger for renewed the interest in neural network method was Werbos' backpropagation algorithm [11] that effectively solved the particular problem by making the training of multi-layer networks feasible and efficient.

Since the last century, ANN techniques have been applied to the electronic packaging design. Subbarayan et al. [12] applied ANN to model the solder joint reliability and further applied to the reliability of the ball grid array packaging. Law et al. [13] had applied ANN in the thermal performance of QFN. Yang et al. [14] had applied ANN in conditional monitoring of power packaging. However, few of the researches paid attention to the ANN modeling technique. Yuan et al. [15] applied the neural network training algorithm for the thermal response of the high power electronics.

978-1-7281-6049-8/20/$31.00 ©2020 IEEE

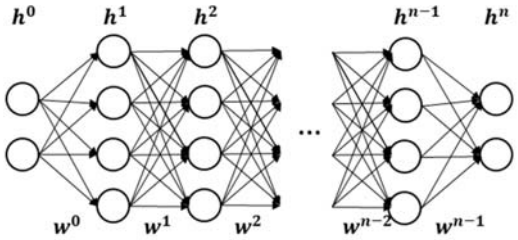

Figure 1: The general ANN structure with input ($\boldsymbol{h^0}$)/output layer($\boldsymbol{h^n}$), and the hidden layers ($\boldsymbol{h^1}$, . $\boldsymbol{h^2}$ … $\boldsymbol{h^{n-1}}$)

Basic RNN is a network of nodes, including the input, hidden, and output nodes. Similar to ANN, each node in RNN is connected with a fixed, directed connection to every other node. However, RNN will organize this network into successive layers, which is used to represent the characteristics of the sequential information. (Part of) the output of this layer will be taken to the next input for the layer. This makes each node has a time-varying real-valued activation.

This research applies the standard backpropagation through time (BPTT) for the RNN training. It is a general back-propagation for the feed-forward networks [11, 16]. The BPTT will be initialized by the very last time sequence and propagate its learning errors along with the invert sequential order.

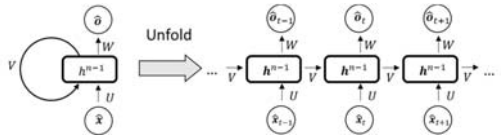

Figure 2: The RNN Architecture

Long short-term memory (LSTM) is a special artificial recurrent neural network (RNN) architecture with a feedback connection. A common LSTM unit is a cell, including an activation, input, output and a forget gate. The cell remembers values over arbitrary time intervals and the four gates regulate the flow of information into and out of the cell. LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series.

LSTM was proposed in 1997 by S. Hochreiter and J. Schmidhuber [17] by introducing the constant error carousel (CEC) units. The initial version of LSTM cells included input and output gates. In 1999, Gers et al. [18] introduced the forget gate (also called "keep gate") into LSTM architecture, enabling the LSTM to reset its own state. Greff et al. [19] in 2017 performed a large scale study for several variants of the LSTM architecture.

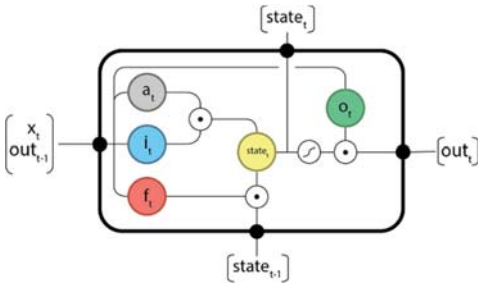

Figure 3: Typical unit cell of the *vanilla* LSTM [18]

As indicated in Figure 3, the conventional LSTM applied only matrices to convert the $\{x_t, out_{t-1}\}^T$ into the output of gates. Because of the nonlinear nature of the industrial application, this paper improves the matrix configuration to a complete neural network. The training technique would be described in this paper.

This paper will apply three neural network architectures, including the ANN, RNN, and VN-LSTM for the solder joint fatigue risk estimation of a wafer level packaging under thermal cycle loading. The dataset has been generated by a validated finite element model. This dataset consists of 81 orthogonal data points. Only 1/3 of the total dataset will be applied for machine learning and the rest for numerical stability testing and the accuracy validation.

2. Theory

As indicated by Figure 1, ANN with n layers can be represented as the composition of $n-1$ functions $f_i: E_i \times H_i \to E_{i+1}$, where E_i, H_i and E_{i+1} are inner product spaces for all $i \in \{n-1\}$. The $\boldsymbol{h} \in E_i$ and $\boldsymbol{\theta} = \{\boldsymbol{w^{n-1}}, \boldsymbol{b^{n-1}}\} \in H_i$ are input and parameters. The output of such NN as be defined as $F: E_i \times (H_1 \times \ldots \times H_{n-1}) \to E_n$ [20] by:

$$F(\boldsymbol{h^0}; \boldsymbol{w^{n-1}}, \boldsymbol{b^{n-1}}) = f_{n-1} \cdot f_{n-2} \cdots f_1(h^0) \quad (1)$$

Eq. (1) can convert an input vector $\boldsymbol{h^0}$ to an output vector $\boldsymbol{h^n}$. The number of neurons of each layer is $l_0, l_1, \cdots, l_n$, accordingly. Each neuron can be written as a scalar form as h_I^m, which indicates the I-th neuron at the m-th layer, where $0 \le m \le n$ and $0 \le I \le l_m$.

3. Establish the dataset

A glass wafer level chip scaled packaging (G-WLCSP) model has been used to generate an orthogonal data set. This model is designed to estimate the G-WLCSP solder joint risk during the thermal cycle loading [4, 5]. As shown in Figure 4, the IC has been redistributed onto a glass wafer and then the trace forming and solder bumping process. Figure 4 (a) shows the wafer level picture and (a') is the

detail view. The individual packaging can be obtained after the dicing from the distributed wafer, as shown in Figure 4 (b). From the schematic cross-section view in Figure 4 (c), the chip has been attached by UV adhesive to the glass wafer, and the epoxy filler has been applied between chips to protect the sidewall of the chip from cracking.

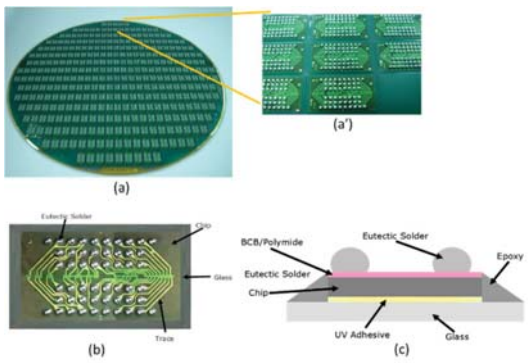

Figure 4: Glass distributed wafer level packaging (a) wafer view with the detail (a'), (b) single device (c) schematic cross-section [4, 5].

All materials applied in the finite element modeling are assumed as linear besides the solder joint and the PI. The eutectic solder joint and PI are considered as the temperature-dependent, elastic-plastic materials [3]. Only one half of the full-scaled two-dimensional finite element approximation model is used owing the symmetrical condition, and the finite element result is used by the commercial finite element code ANSYS® (version 15). Figure 5 shows the finite element model with the mesh density of the most critical solder joint.

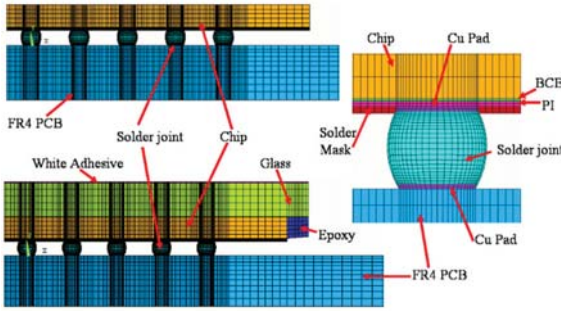

Figure 5: Finite element model for conventional WLCSP and proposed glass WLCSP

4. Machine learning

4.1 AI learning target

Figure 6 shows the statistic result of the total dataset obtained previously. It indicates that the average plastic strain incremental becomes stable after the third cycle. Among the 81 data points, the average strain increment of the last three cycles is 3.01%, with the standard derivation of 1.17%. The empirical Coffin-Mason equation has been applied to convert the strain increment to the reliability cycles, as

$$N_f(\Delta\varepsilon_p) = 0.4405 \cdot (\Delta\varepsilon_p)^{-1.96} \tag{2}$$

On the other hand, it indicates that the difference between the experimental and simulation result has $2 \cdot \Delta = 437$ cycles, which means the cycle prediction accuracy is about $\Delta = 218.5$. Based on Eq.(2), define a problem of:

$$f = \underbrace{\arg}_{\Delta\epsilon_{accu}\in\mathbb{R}} \{min[max(N_f(\Delta\epsilon_{avg} \pm \sigma_\epsilon \pm \Delta\epsilon_{accu})] - \Delta\} \tag{3}$$

By the regression computation of Eq. (3), $\Delta\epsilon_{accu}\sim 0.18\%$ is obtained, and it is assigned as the fitting accuracy requirement for machine learning.

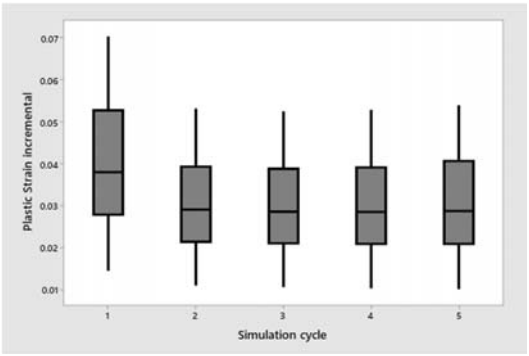

Figure 6: The boxplot of the plastic strain incremental of the 81 data points.

4.2 The neural network design and learning results

This applies ANN, RNN, and VN-LSTM to learn the risk estimation of the solder joint and the accuracy will be then investigated by the complete dataset obtained from Section 3.

27 and 9 data points of the total dataset has been chosen as the training and testing set, respectively. The training set has been selected carefully to ensure the $\hat{\boldsymbol{t}}_k^{input}(k = 1,2,\ldots,27)$ are nearly even distributed within the total dataset. The training and testing set will be fixed through the studies of ANN, RNN, and VN-LSTM.

The machine learning has been done by the in-house software, which has been developed based on .net framework 4.6.1. The learning will be terminated before the overfitting. The accuracy will be computed by applying the trained model to the total dataset with 81 datapoints.

The ANN architecture will be applied first, in order to study the impact of the sequence of the parameters. Three major design parameters, will be applied. The "Parameter alias" of these three parameters will be mentioned in the

following section. The output of the ANN is defined as the average plastic strain incremental of the last three cycles.

The structure of (3,4,4,1) is assigned to the ANN architecture, which means there are three parameters at the input layer, two hidden layers consisted of four neurons and one parameter at the output layer.

We choose the rectified linear unit (ReLU) as the activation function with the formation of $\varphi(x) = \max(0, x)$, in order to represent the nonlinearity of the data and fast convergent rate.

Table 1 shows the ANN modeling results. Due to the time/temperature-dependent nonlinearity nature of the dataset, it is expectable that none of them satisfied the accuracy requirement. Nevertheless, the parameter sequence of "1-2-3" and "2-1-3" perform better than others, which will be carried out to the next phase.

Table 1. ANN modeling results

Parameter sequence	Accuracy (%)
1-2-3	0.338
2-1-3	0.395
2-3-1	0.435
3-2-1	0.425
3-1-2	0.421

In the RNN learning, two more input parameters and are introduced. They are the current temperature and the number of cycles. The output of the RNN is the plastic strain increment, which is also the recurrent parameter, as listed in Table 2.

Table 2. The parameter setting of the RNN and VN-LSTM

Parameter name	Parameter Alias	Parameter function
Die thickness	1	Input
Glass thickness	2	Input
PI thickness	3	Input
Plastic strain increment from the previous state	4	Recurrent / output
Temperature	5	Input
Number of cycles	6	Input

The purpose of the RNN is to study the impact of the neural network structure against accuracy. ReLU with the learning speed of 0.2 has been chosen as the activation function. RNN learning has been carried out by in-house code and the help of GAIN. Figure 7 shows the typical learning of RNN. The learning results are depicted in Table 3. It shows the same unsatisfied result if the RNN structure is similar to the ANN, and also indicated that (6,5,5,1) does not provide sufficient degree-of-freedom to represent the nonlinearity. On the other hand, the structure (6,5,5,5,1) consisted of three hidden layers, could provide enough accuracy. Note that, in order to compensate for the uncertainty of the weighting initialization, the results are the average of multiple learnings.

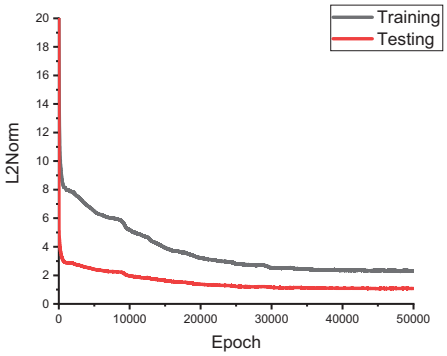

Figure 7: Typical RNN L2 Norm change

Table 3. The learning result of RNN

NN structure	Parameter sequence	Accuracy (%)
(6,5,5,1)	"1-2-3-4-5-6"	0.2974
(6,5,5,1)	"2-1-3-4-5-6"	0.3319
(6,5,5,5,1)	"1-2-3-4-5-6"	0.0282
(6,5,5,5,1)	"2-1-3-4-5-6"	0.0411

To verify the learning freedom of VN-LSTM which indicated in Section 2.4, a combination of gate network structure, activation function, and initial learning rate, which differs from the *vanilla* LSTM, has been chosen. the ReLU with a high initial learning rate is introduced. Note that Hyper tangent is chosen for the φ_{out}.

5. Conclusion

The AI architectures and the machine learning processes have been implemented in order to establish a reliable regression model for the solder joint fatigue risk estimation of G-WLCSP. An orthogonal dataset has been collected based on a parametric study of the experimentally validated finite element model.

The ANN, RNN, and the vectorized-gate network LSTM (VN-LSTM) methods have been applied with the consideration of the time-dependent nonlinearity of the G-WLCSP structure. Note that, compared to the *vanilla* LSTM, VN-LSTM can learn highly nonlinear data because of the gate network design. A learning accuracy requirement of 0.18% has been established based on the experimental result and database though the Coffin-Manson empirical equation.

Acknowledgments

This research is partially supported by the "Developing Smart Automation Mechatronics System and Internet of Things for Shoe Manufacturing" project of Feng Chia University, sponsored by the Ministry of Science and

Technology, under the grand no. MOST108-2218-E-035-007.

References

[1] W. D. van Driel, G. Q. Zhang, J. H. J. Janssen, and L. J. Ernst, "Response Surface Modeling for Nonlinear Packaging Stresses," *Journal of Electronic Packaging,* vol. 125, no. 4, pp. 490-497, 2003.

[2] C.-M. Liu and K.-N. Chiang, "Solder shape design and thermal stress/strain analysis of flip chip packaging using hybrid method," presented at the International Symposium on Electronic Materials and Packaging (EMAP2000) Hong Kong, China, , 2000.

[3] C.-M. Liu, C.-C. Lee, and K.-N. Chiang, "Enhancing the Reliability of Wafer Level Packaging by Using Solder Joints Layout Design," *IEEE Transactions on Components and Packaging Technologies,* vol. 29, no. 4, pp. 877-885, 2006.

[4] C.-A. Yuan, C. N. Han, and K.-N. Chiang, "Design and analysis of novel glass WLCSP structure," in *Proc. 5th International Conference on Thermal and Mechanical Simulation and Experiments in Microelectronics and Microsystems (EuroSimE)*, Brussels, Belgium, 2004, pp. 279-285.

[5] C.-A. Yuan, C. N. Han, M.-C. Yew, C.-Y. Chou, and K.-N. Chiang, "Design, analysis, and development of novel three-dimensional stacking WLCSP," *IEEE Transactions on Advanced Packaging,* vol. 28, no. 3, pp. 387-396, 2005.

[6] F. X. Che, X. Zhang, and J.-K. Lin, "Reliability study of 3D IC packaging based on through-silicon interposer (TSI) and silicon-less interconnection technology (SLIT) using finite element analysis," *Microelectronics Reliability,* vol. 61, pp. 64-70, 2016.

[7] J. Zhang and G. Zhang, "Geometric effects of LGA solder joint on board level reliability in 2-pad ceramic LED assembly," presented at the 13th China International Forum on Solid State Lighting (SSLChina), Beijing, China, 15-17 Nov. 2016, 2016.

[8] P. H. Chou, H. Y. Hsiao, and K. N. Chiang, "Failure Life Prediction of Wafer Level Packaging using DoS with AI Technology," presented at the 2019 IEEE 69th Electronic Components and Technology Conference (ECTC), 2019.

[9] P. H. Chou, K. N. Chiang, and S. Y. Liang, "Reliability Assessment of Wafer Level Package using Artificial Neural Network Regression Model," *Journal of Mechanics,* vol. 35, no. 6, pp. 829-837, 2019.

[10] D. O. Hebb, "The organization of behavior," ed: New York: Wiley, 1949.

[11] P. Werbos, "Beyond Regression:" New Tools for Prediction and Analysis in the Behavioral Sciences," PhD, Harvard University, 1974.

[12] G. Subbarayan, Y. Li, and R. L. Mahajan, "Reliability Simulations for Solder Joints Using Stochastic Finite Element and Artificial Neural Network Models," *Journal of Electronic Packaging,* vol. 118, no. 3, 1996.

[13] R. Law, R. Cheang, Y. Tan, and I. A. Azid, "Thermal Performance Prediction of QFN Packages using Artificial Neural Network (ANN)," presented at the 2006 Thirty-First IEEE/CPMT International Electronics Manufacturing Technology Symposium, Petaling Jaya, 2006.

[14] S. Yang, D. Xiang, A. Bryant, P. Mawby, L. Ran, and P. Tavner, "Condition Monitoring for Device Reliability in Power Electronic Converters: A Review," *IEEE Transactions on Power Electronics,* vol. 25, no. 11, pp. 2734-2752, 2010.

[15] C. Yuan, Y.-J. Hong, C.-C. Lee, K.-N. Chiang, and J.-H. Huang, "Application of Artificial and recurrent neural network on the steady-state and transient finite element modeling," presented at the 20th International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE), Hannover, Germany, 24-27 March 2019, 2019.

[16] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Internal Representations by Error Propagation," Institute for Cognitive Science, University of California, San Diego (CA)1985.

[17] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation,* vol. 9, no. 8, pp. 1735-1780, 1997.

[18] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: continual prediction with LSTM," presented at the 9th International Conference on Artificial Neural Networks: ICANN '99, 1999.

[19] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A Search Space Odyssey," *IEEE Trans Neural Netw Learn Syst,* vol. 28, no. 10, pp. 2222-2232, Oct 2017.

[20] A. L. Caterini, "A Novel Mathematical Framework for the Analysis of Neural Networks," Master of Mathematics, Department of Applied Mathematics, University of Waterloo Waterloo, Canada, 2017.