IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE ELECTRONICS PACKAGING SOCIETY SECTION

Received June 29, 2020, accepted July 29, 2020, date of publication August 4, 2020, date of current version August 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3014156

Solder Joint Reliability Modeling by Sequential Artificial Neural Network for Glass Wafer Level Chip Scale Package

CADMUS C. A. YUAN AND CHANG-CHI LEE
Department of Mechanical and Computer-Aided Engineering, Feng Chia University, Taichung 40724, Taiwan
Corresponding author: Cadmus C. A. Yuan (cayuan@fcu.edu.tw)

This work was supported in part 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 Grant MOST108-2218-E-035-007.

ABSTRACT This article combines the sequential artificial neural network (NN) machine learning with finite element (FE) modeling to assess the solder joint thermal cycling performance. A glass wafer-level chip-scale package (G-WLCSP) is used for this study. This article investigates the network structure that can achieve prediction capability both inside and outside the design domain with the minimal required training dataset. First, a detailed FE model for G-WLCSP is developed to obtain the accumulated plastic strain per cycle for thermal-cycling loading. Three critical input parameters are defined to generate a dataset based on finite element analysis. Then, applying the supervised machine learning procedure, both the recurrent neural network (RNN) and the gate-network long short-term memory (LSTM) architecture are used to train the obtained dataset. The network complexity of the sequential NN model is carefully controlled to prevent numerical overfitting. Among the total 81 FE generated data pairs, only 27 data pairs have been applied to the sequential NN learning. These 27 data pairs are carefully selected to evenly distributed among the design domain. The average error norms after the learning are $1.213 \cdot 10^{-4}$ and $1.190 \cdot 10^{-4}$of RNN and LSTM, respectively. The prediction capability of the well-trained sequential NN model against the rest 54 data pairs has been tested and a similar scale has been obtained. Furthermore, the prediction capability is tested against the parameters outside the design domain. Approximately one order average error norm increased for both the well-trained RNN and LSTM model.

INDEX TERMS Wafer level package, chip-scale package, reliability, machine learning, recurrent neural networks, long short-term memory.

I. INTRODUCTION

Solder joint reliability is one of the most critical issues for most ball-grid array packaging types. To verify the reliability of the electronic packaging, a time-consuming accelerated thermal cycling test is required, which requires 3-4 months for a complete reliability test [1]. Because of the mechanical characteristics, 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 [2]. A decent electronic packaging FE model requires specialized modeling expertise and numerous experimental validation tests. A proper FE model with large material/geometrical nonlinearities under thermal cycling loading costs considerable time to accomplish a single result [3], [4], not to mention to achieve an optimal design of the packaging with large design parameters.

To maximize the utilization of these valuable FE results and accelerate the optimization of the design parameters, scholars have developed many methods. van Driel *et al.* [5] 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 and Chiang [6] and Liu *et al.* [7] applied the parametric method to study the solder fatigue failure mode under the thermal cycling loading. The model used in [7] had applied both material and geometrical nonlinearity with the precise description of the solder joint geometry. Yuan *et al.* [8], [9] applied the Taguchi matrix

The associate editor coordinating the review of this manuscript and approving it for publication was Koneru Ramakrishna.

to establish a response surface of the estimated solder joint lifetime based on a validated parametric finite element model. Che *et al.* [10] 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 and Zhang [11] applied RSM method to study the board level reliability of LED packaging. The main gap of the RSM method is that the accuracy highly depends on statistical theory, and the prediction capability will decrease when the design parameters move outside the fitting domain.

In recent years, the neural network method has been widely used in various research domains. It has been proven to be able to handle complex with high nonlinearity and multivariate relationships. The NN can convert the individual discrete samples into a network structure. Hence, 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. Chou *et al.* [15], [16] developed the basic framework of the combination of machine learning and finite element modeling, and has been applied the model the long-term reliability of wafer level packaging using rather large artificial neural network architecture with multiple key design features. Yuan *et al.* [17] applied the ANN training algorithm for the thermal response of the high power electronics. There are still two major scientific challenges for the NN application. One is the complexity of the network structure which can capture the response nature of the samples. Another is the size of the training dataset. Although a complicated NN structure with large number of neurons and hidden layers would be able to capture the nonlinearity nature, it is also required enormous training datasets to prevent the numerical overfitting. This vicious cycle limits the application of the NN due to the training sample size and long training time.

This article investigates the network structure that can achieve prediction capability both inside and outside the design domain with the minimal required training dataset. This article applies two sequential NN methods to model the solder joint risk of glass wafer level chip-scale packaging. An experimental validated FE method will be applied to generate the training datasets with respect to three major design parameters. The complexity of the NN network is carefully controlled to prevent overfitting, which allows using less training datasets. Only 27 from the 81 FE generated datasets have been applied for the training purpose, and the rest 54 ones are used for the validating of the trained NN model. Moreover, the prediction capability of the trained RNN and LSTM model will be tested by introducing the parameters that are outside the training domain.

This article will be organized as follows: In ''Theory'' section, the sequential neural network architecture, including the RNN and LSTM will be addressed, as well as the definition of the average error norm. The design target of this research, a glass wafer level chip-scale packaging and its FE model will be described in the next two sections. The ''Supervised Machine Learning'' section, the selection of the training set, and RNN and LSTM training procedure and the validation, will be addressed. A brief conclusion will then summarize this article.

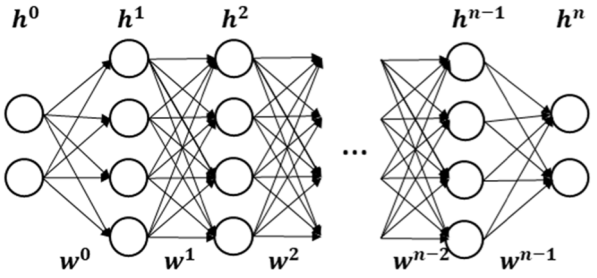

FIGURE 1. The general ANN structure with input (h^0), output layer(h^n), and the hidden layers (h^1, h^2 ... h^{n-1}).

II. THEORY

Artificial neural network (NN) is based on a collection of connected units called artificial neurons or network nodes/cells (Fig.1), 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. Learning is the adapting process of the network to better handle a particular task by considering the sample observations. Learning involves adjusting the weights and bias of the network to improve the accuracy of the result. Werbos' backpropagation algorithm [18] that effectively made the training of multi-layer networks feasible and efficient. Most NN models in the literature focus on the end effects, e.g. the lifetime under thermal loading, and the static junction temperature of the chip. This article focuses on modeling the solder joint risk by accumulating the intermediate states during the cyclic loading in order to gain more inside of the reliability problem. However, when the time/temperature-dependent material and/or geometrical nonlinearity and the thermal cyclic loading dominates the physical phenomena of the solder joint fatigue, it may not be reasonable to remain in the ANN method but seeking for sequential modeling techniques, in order to simultaneously consider the nonlinearity and time-dependency. Moreover, the complexity of the network structure, including the number of hidden layers and neurons at each layer, is controlled, to prevent the numerical overfitting and accelerate the network training.

As illustrated in Fig.1, the value of each i neuron at the l-th layer can be computed from $(l-1)$-th layer by the:

$$h_i = \varphi(\sum W_{ij} h_j + b) \tag{1}$$

where h_j are the neuron values from the $(l-1)$-th layer, the W_{ij} is the weightings and b is the layer bias and the φ is the activation function. By repeating (1), one can convert an input vector $\mathbf{h^0}$ to an output vector $\mathbf{h}^n$.

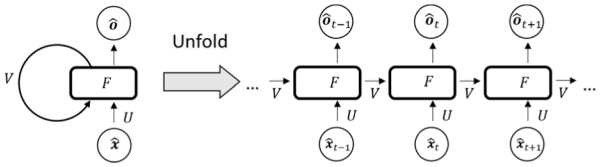

FIGURE 2. The RNN concept.

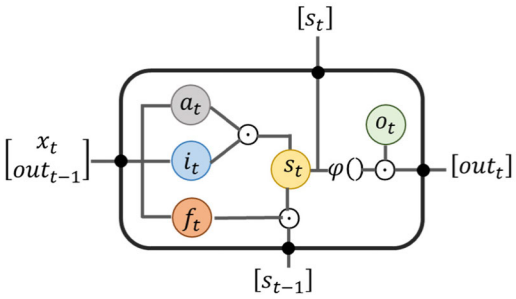

FIGURE 3. Typical unit cell of the LSTM.

The dataset is composed of multiple data pairs, which is a combination of the input and output values, $(\hat{t}^{input}, \hat{t}^{output})$. All data points in the dataset are assumed to be independent with each other.

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, as shown in Fig.2. 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 backpropagation through time (BPTT) for the RNN training [18], [19].

Long short-term memory (LSTM) is a special RNN architecture with a feedback connection. A standard LSTM unit is a cell, including an activation, input, output and forgets gate, as indicated in Fig.3. Via the state variable shown in Fig.3, 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 essential events in a time series.

This article defines each gate a_t, i_t, f_t and o_t are four independent neural networks, with the internal relationship of:

$$\begin{aligned} s_t &= a_t \odot i_t + f_t \cdot s_{t-1} \\ out_t &= \varphi_{out}(s_t) \odot o_t \end{aligned} \tag{2}$$

φ_{out} is the activation function for LSTM, and $\odot$ represents the Hadamard (elementwise) multiplication.

The NN model prediction capability will be estimated by the average error norm(e), by:

$$e = \frac{\sqrt{\sum (o_j - n_j)^2}}{n} \tag{3}$$

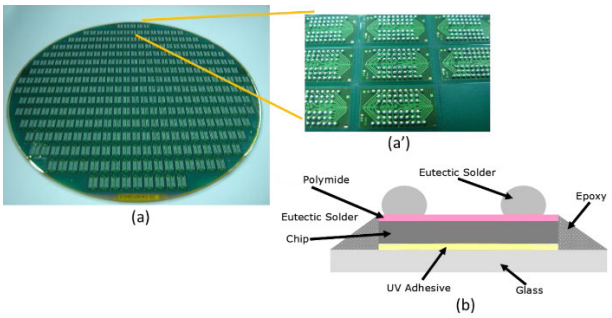

FIGURE 4. Glass distributed wafer level packaging (a) wafer view with the detail (a'), (b) schematic cross-section [8], [9].

where o_j and n_j are the expected values and the model outputs, respectively. n is the size of the dataset.

III. GLASS WAFER LEVEL CHIP SCALE PACKAGING (G-WLCSP)

A glass wafer level chip scale package (G-WLCSP) model has been used as the design target of this research. As shown in Fig.4, the IC has been redistributed onto a glass wafer and then the trace forming and solder bumping process. Fig.4 (a) shows the wafer level picture and (a') is the detail view. From the schematic cross-section view in Fig.4(b), 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.

The CTE of the glass and chip are similar (TABLE 1), the failure risk between chip and glass interface is low. But the combination of chip and glass introduces large stiffness. Due to the thermal mismatch of the glass wafer, chip and printed circuit board (PCB), the solder joints are subject to high deformation during the thermal loading. Therefore, a polyimide (PI) layer is applied under the solder joint the absorb the deformation energy during the thermal loading. Therefore, from the perspective of the solder joint reliability risk, the chip, glass and PI thickness are the most crucial design parameters [8], [9].

On the other hand, from the manufacturability consideration, a thin glass wafer will cause the excess warpage and handling thin chip remains extra cost. Moreover, a thick polyimide (PI) will induce yield-loss in the spin-coating process.

It is required to provide an optimal solution of the glass, chip and PI thickness with both reliability and currently available manufacturability, but also provide a model that can predict the solder joint reliability when the manufacturing capability improves in the future. Hence, this research selects the sequential NN method to build the solder joint regression model.

IV. FINITE ELEMENT MODELING

All materials applied in the finite element modeling are assumed as linear besides the solder joint and the polyimide (PI). The eutectic solder joint and PI are considered as the temperature-dependent, elastic-plastic materials [7] as shown in Fig.5. The rest of the material properties are considered as linear and listed in TABLE 1. Only one half of the full-scale

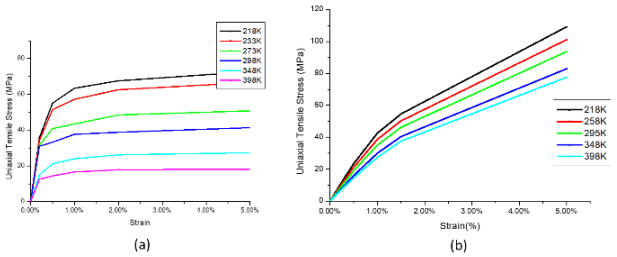

FIGURE 5. Nonlinear mechanical response of (a) solder joint (63Sn/37Pb) and (b) PI within different temperature.

TABLE 1. Material properties.

	Young's modulus(MPA)	CTE (ppm)	Poisson's ratio
Solder joint (63Sn/37Pb)	Temperature dependent and nonlinear (Fig.5)	23.9	0.35
Cu	76,000	17	0.35
Solder mask	3,400	30	0.35
FR4 PCB	18,200	16	0.19
PI(Stress Buffer Layer)	Temperature dependent and nonlinear (Fig.5)	150	0.4
BCB	3,000	50	0.34
Si	112,400	2.62	0.28
White adhesive	0.7	300	0.45
Glass	63,000	3.25	0.28
Epoxy	80	250	0.34

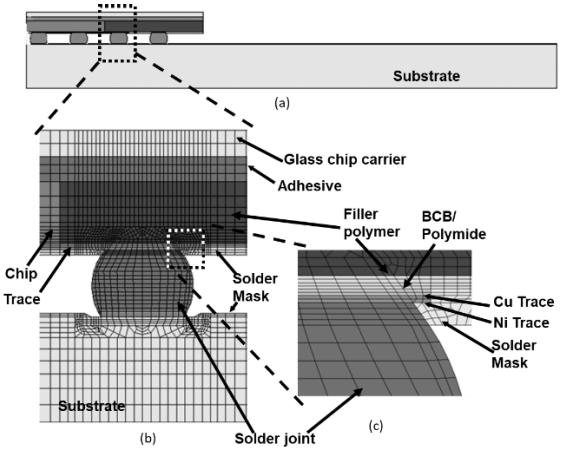

FIGURE 6. Finite element model for proposed G-WLCSP.

finite element approximation model is used owing to the symmetrical condition, and the finite element result is used by the commercial finite element code ANSYS®(version 15). Fig.6 shows the finite element model with the mesh density of the most critical solder joint.

A G-WLCSP structure, fabricated by the currently available technologies, of the die size of $5.77 \times 10.38 \times 0.3$ mm^3 with a glass thickness of 0.5 mm has been attached to a test board with 1.2 mm thickness. The 63Sn/37Pb solder ball with

TABLE 2. Parametric model settings.

Parameter name	Upper bound	Lower bound
Die thickness	250 um	500 um
Glass thickness	300 um	800 um
PI thickness	15 um	40 um

0.45 mm in diameter has been applied onto the die-side pad with its opening of 0.37 mm, and the solder joint stand-off height has been reduced to 0.35 mm after the reflow. This G-WLCSP is subjected to a thermal cycle loading between -40° and 125° with a ramp rate of 11°/min and a dwelling time of 15 min. The initial stress-free reference temperature equals 25°. 21 samples have been put into the thermal cycling oven. According to the Weibull (63.2% failure cycle) distribution, an average of 1,007 cycles has been reported. The accumulated plastic strain of each thermal loading cycle is taken as the solder fatigue risk indicator, following the Coffin-Manson equation. The finite element fatigue life prediction gives 1,444 cycles, with the average plastic strain per cycle of 0.01608. A good agreement has been achieved.

Eighty-one parametric finite element models with the control of the mesh density of the farthest solder joint corner are then executed to build the total dataset. Compared to the validated finite element model, no simplification has made and the temperature-dependent material nonlinearity remains. These simulations consist of 5 sequential thermal cycling loadings with all the abovementioned conditions. Key parameters with the levels and noise factors are listed in TABLE 2. Each simulation comprises a complete five thermal cyclic loading.

There only 27 data points will be applied for the sequential NN training, and the remaining 54 data points will be used to validate the accuracy of the RNN and LSTM model.

V. SUPERVISED MACHINE LEARNING

This article applies RNN, and LSTM architectures to learn the task of the risk estimation of the solder joint using a small portion of the total dataset.

The training set has been selected carefully to ensure these data pairs are nearly even distributed within the parameter domain of the total dataset. Based on the design parameter ranges listed in TABLE 2, each design parameter has been split into three groups, and each group consists of three levels. The combination of three groups of three design parameters introduced 27 data pairs, listed in TABLE 3.

The training dataset will be fixed through the rest studies of RNN, and LSTM. Afterward, the accuracy of the trained model will be computed not only against these 27 training data pairs but the rest 54 data pairs.

There are five input parameters for sequential NN learning for RNN and LSTM. Besides the geometrical design parameters that listed in TABLE 2 and TABLE 3, there are two computation parameters used, including the temperature change and the plastic strain per cycle ($\Delta\epsilon_{pl}$). The concept of the simplest hypothesis that consistent with the data, following Ockham's razor [21], has been applied to this research.

TABLE 3. The 27 selected data pairs for the network training.

Parameter name / Data pair	Die thickness (um)	Glass Thickness (um)	PI Thickness (um)	Plastic strain per cycle($\Delta\epsilon_{pl}$)
1	235	270	35	0.0107
2	250	300	45	0.0105
3	265	330	40	0.0128
4	235	470	20	0.0209
5	250	500	30	0.0212
6	265	530	25	0.0241
7	235	770	10	0.0385
8	250	800	20	0.0391
9	265	830	15	0.0431
10	360	270	20	0.0180
11	375	300	30	0.0180
12	390	330	25	0.0208
13	360	470	10	0.0281
14	375	500	20	0.0282
15	390	530	15	0.0314
16	360	770	35	0.0397
17	375	800	45	0.0399
18	390	830	40	0.0438
19	485	270	10	0.0261
20	500	300	20	0.0260
21	515	330	15	0.0286
22	485	470	35	0.0289
23	500	500	45	0.0291
24	515	530	40	0.0323
25	485	770	20	0.0491
26	500	800	30	0.0493
27	515	830	25	0.0527

Two different RNN structures of (5,5,5,1) and (5,5,5,5,1), as illustrated in Fig.7, where the geometrical design parameters, i.e., die thickness, glass thickness and the PI thickness, as well as the loading condition, the ΔT, and the $\Delta\epsilon_{pl}$ per cycle will be used as the inputs. The structure of (5,5,5,1) represents in Fig.7 (a), where two hidden layers with equally five neurons and 1 output are listed. Equation (1) will be applied to compute from input to output. Notably, the output ($\Delta\epsilon_{pl}$) from the current time step will be treated as the input for the next time step, as the nature of RNN that illustrated in Fig.2. The same concept applies to the structure (5,5,5,5,1).

Machine learning has been done by the in-house developed software based on.net framework 4.6.1. The overfitting will be strictly prevented by the network structure design and by continuously comparing the loss function between the training and testing sets. The accuracy will be computed by applying the trained model to the total dataset with 81 data points. 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.

With the learning rate of 0.2, Fig.8 shows the typical error norm (L2Norm) during the machine learning procedure. The Y-axis and X-axis represent the error L2norm and the iteration epochs. The black and red lines of Fig.8 represent the error of the training and testing sets, respectively. The testing set is organized by randomly choosing from the 54 remaining data points. These testing datasets only used to monitor the overfitting probability and not involved in the

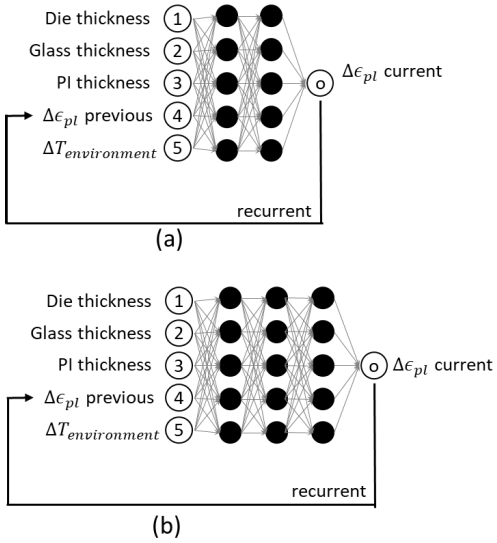

FIGURE 7. The (a) (5,5,5,1) and (b) (5,5,5,5,1) RNN structure.

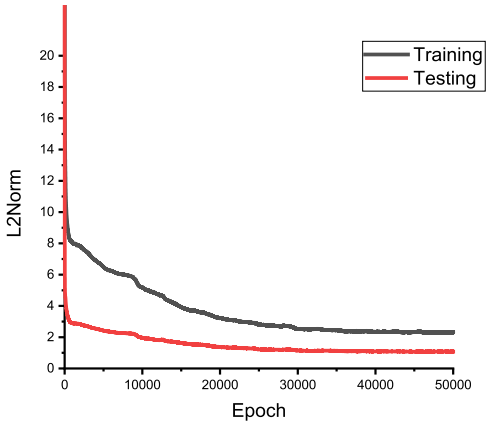

FIGURE 8. Typical RNN L2 Norm change of structure (5,5,5,5,1).

training algorithm. Due to the initial guessing of the network weightings, the error will be high in the beginning. With the help of BPTT algorithm, the error will be reduced gradually.

LSTM consists of four independent networks as its gates. Fig.9 visualizes a (5,5,5,1) gate network structure of LSTM, where the four gates, including a_t, i_t, f_t and o_t are represented by four independent gate-network, with the size of (5,3,1), (5,5,5,5,1), (5,5,1) and (5,5,1), respectively. The black square represents one single LSTM cell. When the input parameters, enter the LSTM cell, these will be treated as the inputs of those independent gate networks. In this research, two kinds of i_t gate network structures have been applied, as listed in TABLE 4. By (2), one can compute the output parameter of LSTM cell, i.e., the $\Delta\epsilon_{pl}$. Similar to RNN, the current $\Delta\epsilon_{pl}$.output will be taken as the input for the next computation step. Note that hyper tangent equation is chosen for the φ_{out} in (2).

A typical error norm (L2norm) decreasing curve of the training and testing sets during the learning procedure, as shown in Fig.10. The Y-axis and X-axis represent the error

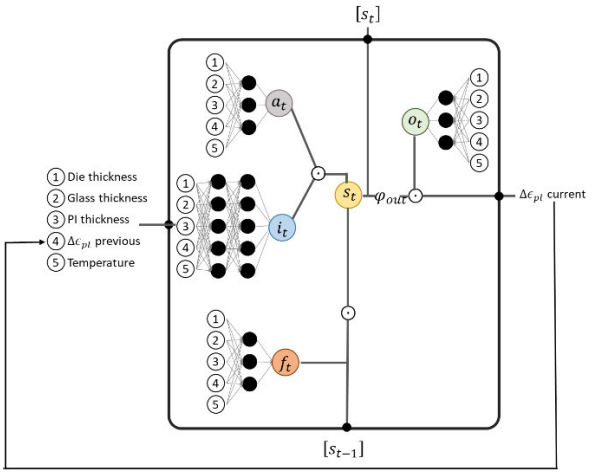

FIGURE 9. Gate-Network LSTM structure for (5,5,5,1).

TABLE 4. The gate network structure, activation function and the initial learning rate in LSTM.

Gate	Network structure	Activation function
Activate (a_t)	(5,3,1)	Sigmoid
Input (i_t)	(5,5,5,1) and (5,5,5,5,1)	ReLU
Forget (f_t)	(5,3,1)	Hyper tangent
Output (o_t)	(5,3,1)	Sigmoid

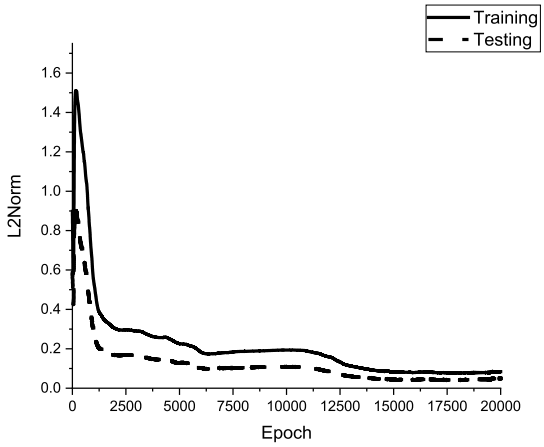

FIGURE 10. Typical LSTM L2 norm versus epoch.

L2norm and the iteration epochs. The solid and dashed lines of Fig.10 represent the error of the training and testing sets, respectively. The error norm is high at the beginning of the learning process, and the error norm decreases as the iteration continues. However, comparing to Fig.8, and a rugged convergence history can be found in LSTM. This phenomenon is because the L2 Norm of LSTM is a combination of the performance of all independent gates.

Using (3), the average error norm of the RNN and LSTM model has been computed against the training and validating sets, listed in TABLE 5. Comparing the training and the validating results, no significant difference can be discovered. When the training dataset is evenly distributed over the

TABLE 5. The performance of RNN and LSTM model.

	Data points	RNN Model (5,5,5,1)	RNN Model (5,5,5,5,1)	LSTM Model (5,5,5,1)	LSTM Model (5,5,5,5,1)
Training	27	$1.706\cdot 10^{-3}$	$1.432\cdot 10^{-4}$	$1.110\cdot 10^{-3}$	$1.357\cdot 10^{-4}$
Validating	54	$1.218\cdot 10^{-3}$	$1.213\cdot 10^{-4}$	$7.830\cdot 10^{-4}$	$1.190\cdot 10^{-4}$

TABLE 6. The top 10 worst cases of the performance of RNN and LSTM model within the design range.

Die thickness (um)	Glass thickness (um)	PI thickness (um)	Plastic strain per cycle($\Delta\epsilon_{pl}$) FEM	RNN	LSTM	Errors RNN	LSTM
375	770	40	0.0425	0.0390	0.0388	-0.0035	-0.0037
235	300	40	0.0106	0.0117	0.0122	0.0011	0.0016
265	270	45	0.0103	0.0112	0.0118	0.0009	0.0015
235	330	45	0.0110	0.0122	0.0124	0.0012	0.0014
250	270	40	0.0105	0.0112	0.0119	0.0007	0.0014
265	300	35	0.0124	0.0134	0.0136	0.0010	0.0011
375	830	35	0.0443	0.0433	0.0432	-0.0010	-0.0011
250	330	35	0.0129	0.0140	0.0140	0.0011	0.0010
515	270	20	0.0258	0.0245	0.0248	-0.0013	-0.0010
485	530	45	0.0297	0.0306	0.0307	0.0009	0.0009

TABLE 7. The performance of RNN and LSTM model outside the design range.

Die thickness (um)	Glass thickness (um)	PI thickness (um)	Plastic strain per cycle($\Delta\epsilon_{pl}$) FEM	RNN	LSTM	Errors RNN	LSTM
660	1120	45	0.0597	0.0721	0.0723	0.0124	0.0126
590	980	10	0.0616	0.0674	0.0680	0.0059	0.0064
590	980	45	0.0539	0.0594	0.0603	0.0055	0.0064
515	835	10	0.0522	0.0558	0.0559	0.0036	0.0037
515	835	45	0.0441	0.0487	0.0475	0.0045	0.0034
660	1120	10	0.0787	0.0811	0.0810	0.0023	0.0022
590	120	10	0.0257	0.0227	0.0235	-0.0030	-0.0022
590	120	45	0.0206	0.0177	0.0180	-0.0023	-0.0020
90	120	10	0.0092	0.0152	0.0073	0.0060	-0.0019
90	35	10	0.0090	0.0066	0.0071	-0.0025	-0.0019

total space, the well-trained model will be able to describe the rest of the domain. Moreover, the (5,5,5,1) contains too little information and doesn't perform well, in both RNN and LSTM models. However, an increase of 1 more hidden layer will contribute approximately one order accuracy. The top 10 worst cases of RNN and LSTM are shown in TABLE 6, where the "Errors" are the direct error, in order to meet the requirement of Eq. (3). A similar trend of the RNN and LSTM can be identified by comparing the difference between the FEM results and RNN/LSTM predictions.

Since the (5,5,5,5,1) performs well in both RNN and LSTM architecture, this structure will be applied to study when the design parameters located outside the predefined dataset. These design parameters of the pre-defined dataset have been listed in TABLE 2. 10 extra design parameter combinations which are located outside the pre-defined parameter domain, are selected to test the predictability of the sequential neural network models in TABLE 7, where the "Errors" are direct errors. As listed in TABLE 7, different combination parameters will induce different levels of errors in both RNN and LSTM, and there is no clear evidence to tell whether RNN and LSTM are better than the other. Following (3), the average error norms of RNN and LSTM are $1.533\cdot 10^{-3}$ and $1.453\cdot 10^{-3}$. Comparing to the average error norm within the pre-defined dataset (TABLE 5 and TABLE 6), the one outside the design parameter will increase the average error norm in approximately one order.

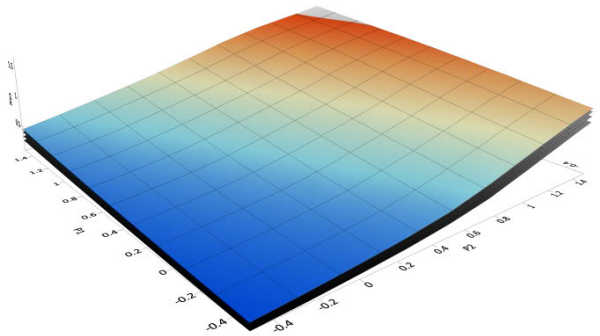

FIGURE 11. The response surface of parameter 1 and 2 with respect to the strain increment of the LSTM model.

Fig.11 shows the LSTM plot of both P1 and P2 axis of the range of [−0.5,1] with fixed P3 (The PI thickness) values, and the Z-axis is the accumulated plastic strain per cycle($\Delta\epsilon_{pl}$). The nonlinear nature of has been depicted along the P1 and P2 axis. It shows that the ReLU activation function with proper multiple hidden layers can model the nonlinearity by turning off some of the neurons within specific layers. Moreover, the surface is smooth even outside the domain of the total 81 data pairs and is continuous and piecewise differentiable at all points. It could contribute the expand the design domain and find the design optimal.

VI. CONCLUSION

This article investigates the network structure that can achieve prediction capability both inside and outside the design domain with the minimal required training dataset. Three key reliability design parameters, including the die, glass and PI thicknesses, for the G-WLCSP has been modeled by the sequential NN with finite element modeling to assess the solder joint thermal cycling performance. First, a detailed finite element model for G-WLCSP is developed to obtain the accumulated plastic strain per cycle for thermal-cycling analysis. Then critical input parameters are defined to generate a data set based on finite element analysis. Then, applying the supervised machine learning procedure, both the recurrent neural network (RNN) and the modified long short-term memory (LSTM) architecture are used to train the obtained dataset. Moreover, this article studies the minimal required NN structure and training data pairs to achieve acceptable accuracy.

The RNN and LSTM architectures have been applied. Besides the three key parameters, two more learning parameters, including the temperature cycling range and plastic strain per cycle, are introduced. From the 81 data pairs obtained by the validated FE model, only 27 has been selected for the NN training purpose. The training of NN has been accomplished by the backpropagation through time method.

The training results show that the sequential neural network with 1 extra hidden will improve approximately 1 order of average error norm. The average error norms of RNN and LSTM with 3 hidden layers are $1.432{\cdot}10^{-4}$ and $1.357{\cdot}10^{-4}$, respectively. Because these 27 data pairs have been carefully chosen that uniformly distributed along with the design parameter domain, the average error norms of the rest 54 data pairs for RNN and LSTM are $1.213{\cdot}10^{-4}$ and $1.190{\cdot}10^{-4}$, respectively. Extra design parameter combinations which are located outside the pre-defined parameter domain, are selected to test the predictability of the well-trained sequential NN. The result indicated that the well-trained model will exhibit approximately 1 order average error norm increasing for the parameter outside the training parameter domain.

REFERENCES

[1] *Temperature Cycling (JESD22-A104)*, JEDEC Solid State Technol. Assoc., Arlington, VA, USA, 2009.
[2] X.-J. Fan, H. B. Wang, and T. B. Lim, "Investigation of the underfill delamination and cracking in flip-chip modules under temperature cyclic loading," *IEEE Trans. Compon. Packag. Technol.*, vol. 24, no. 1, pp. 84–91, Mar. 2001.
[3] X. Fan, M. Pei, and P. Bhatti, "Effect of finite element modeling techniques on solder joint fatigue life prediction of flip-chip BGA packages," presented at the Electron. Compon. Technol. Conf. (ECTC), 2006.
[4] X. J. Fan, B. Varia, and Q. Han, "Design and optimization of thermo-mechanical reliability in wafer level packaging," *Microelectron. Rel.*, vol. 50, no. 4, pp. 536–546, Apr. 2010.
[5] W. D. van Driel, G. Q. Zhang, J. H. J. Janssen, and L. J. Ernst, "Response surface modeling for nonlinear packaging stresses," *J. Electron. Packag.*, vol. 125, no. 4, pp. 490–497, Dec. 2003.
[6] 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 Int. Symp. Electron. Mater. Packag. (EMAP) Hong Kong, 2000.
[7] C.-M. Liu, C.-C. Lee, and K.-N. Chiang, "Enhancing the reliability of wafer level packaging by using solder joints layout design," *IEEE Trans. Compon. Packag. Technol.*, vol. 29, no. 4, pp. 877–885, Dec. 2006.
[8] C.-A. Yuan, C. N. Han, and K.-N. Chiang, "Design and analysis of novel glass WLCSP structure," in *Proc. 5th Int. Conf. Thermal Mech. Simulation Exp. Microelectron. Microsyst. (EuroSimE)*, Brussels, Belgium, 2004, pp. 279–285.
[9] C.-A. Yuan, C. Nan Han, M.-C. Yew, C.-Y. Chou, and K.-N. Chiang, "Design, analysis, and development of novel three-dimensional stacking WLCSP," *IEEE Trans. Adv. Packag.*, vol. 28, no. 3, pp. 387–396, Aug. 2005.
[10] 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," *Microelectron. Rel.*, vol. 61, pp. 64–70, Jun. 2016.
[11] 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 Int. Forum Solid State Lighting (SSLChina), Beijing, China, Nov. 2016.
[12] G. Subbarayan, Y. Li, and R. L. Mahajan, "Reliability simulations for solder joints using stochastic finite element and artificial neural network models," *J. Electron. Packag.*, vol. 118, no. 3, pp. 148–156, Sep. 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 31st IEEE/CPMT Int. Electron. Manuf. Technol. Symp., Petaling Jaya, Malaysia, 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 Trans. Power Electron.*, vol. 25, no. 11, pp. 2734–2752, Nov. 2010.
[15] 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 IEEE 69th Electron. Compon. Technol. Conf. (ECTC), 2019.
[16] P. H. Chou, K. N. Chiang, and S. Y. Liang, "Reliability assessment of wafer level package using artificial neural network regression model," *J. Mech.*, vol. 35, no. 6, pp. 829–837, Dec. 2019.
[17] 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 Int. Conf. Thermal, Mech. Multi-Phys. Simulation Exp. Microelectron. Microsyst. (EuroSimE), Hannover, Germany, Mar. 2019.

[18] P. Werbos, "Beyond regression: New tools for prediction and analysis in the behavioral sciences," Ph.D. dissertation, Committee Appl. Math., Harvard Univ., Cambridge, MA, USA, 1974. [Online]. Available: https://zenodo.org/record/1262035#.Xypo1TozaUk
[19] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," Inst. Cognitive Sci., Univ. California, San Diego, CA, USA, 1985.
[20] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," presented at the 9th Int. Conf. Artif. Neural Netw., 1999.
[21] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Harlow, U.K.: Pearson, 2016.

CADMUS C. A. YUAN received the Ph.D. degree from the Department of Power Mechanical Engineering, National Tsing Hua University, Hsinchu, Taiwan, in 2005, with the major of solid mechanics and numerical methods.

From 2007 to 2010, he was a Senior Engineer with NXP Semiconductors, The Netherlands. From 2010 to 2012, he was the Senior Scientist at TNO, The Netherlands. From 2012 to 2015, he was the Program Manager of the State Key Laboratory of Solid-State Lighting, Beijing, China. He is currently an Assistant Professor with the Department of Mechanical and Computer-Aided Engineering, Feng Chia University, Taichung, Taiwan. His research interests include the development of machine learning algorithms for industrial and engineering applications, in the domain of the design/manufacturing optimization and reliability modeling.

CHANG-CHI LEE received the B.S. degree in mechanical computer-aided engineering from Feng Chia University, Taichung, Taiwan, in 2019, where he is currently pursuing the M.S. degree in mechanical engineering. His research interest includes the application of the sequential neural network method for the industrial application.