Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 34 (2008) 588–595
Expert Systems with Applications
www.elsevier.com/locate/eswa

Combining radial basis function neural network and genetic algorithm to improve HDD driver IC chip scale package assembly yield

M.L. Huang, Y.H. Hung *

Department of Industrial Engineering and Management, National Chin-Yi Institute of Technology, 35, Lane 215, Chung-Shan Road, Taiping, Taaichung, 411, Taiwan, ROC

Abstract

In recent years, the future trend of micro HDD driver IC for large capacity micro HDD is to become lighter, thinner, shorter and smaller. Among all the options available for micro HDD driver IC's assembly, warpage is an important issue related to micro HDD driver IC manufacturability and reliability. The optimal packaging manufacturing process for driver IC for micro HDD is chip scale package (CSP). However, the production and assemble process for CSP is much more difficult. The aim of this study is to improve the lower warpage properties for 0.65 mm CSP assembly yield using a model based on a radial basis function network (RBFN), and the optimal HDD packaging process parameter design is achieved through a genetic algorithm (GA).

Keywords: Driver IC; Warpage; Assembly yield; RBFN; GA

1. Introduction

In recent years, cellular phones, PDA and other portable devices have grown rapidly in popularity, thanks to compact, thin, and light features that have become key requirements for design. Large storage capacity, as well as an attractive price, would be the major factors determining if hard disk drive (HDD) will become increasingly more popular in growing consumer electronics markets (Coughlin et al., 2004). Furthermore, this trend today continues to accelerate. Generally, the driver IC remains a key component of micro HDD in providing the necessary speed and control capacity that allow a large of amount data to be read and coded by HDD at multi-megabyte/s rate.

Inevitably, the future trend of micro HDD driver IC for large capacity micro HDD is to become lighter, thinner, shorter and smaller. For satisfying the lower profile design the vertical dimension "TH" (*Z*-axis) control of driver IC is crucial, owing to the limited space in micro HDD (shows in Fig. 1). Chip scale package (CSP) technology is being developed to achieve miniaturized packaging systems with improved space limitations. The advantages to using a CSP over direct chip attach (DCA) include: easier handling, more protection for the chip, simpler board assembly and reduced total package costs (Baliga, 1998). The CSP manufacturing processes of micro HDD driver IC are described as follows: Wafer back-grinding ⇒ Wafer saw ⇒ Die attach ⇒ Wire bonding ⇒ Molding ⇒ Marking ⇒ Post mold cure ⇒ Solder paste printing ⇒ Re-flow ⇒ Singulation ⇒ Packing.

Among all the options available for micro HDD driver IC's assembly, warpage is an important issue related to micro HDD driver IC manufacturability and reliability. The warpage has been studied widely, from peripheral packages (e.g., SOP and QFP) to area-array packages (e.g., BGA and CSP), and from linear elastic studies to viscoelastic studies (Egan, Kelly, O'Donovan, & Kennedy, 2003; Li, 2003; Miyake, Yoshida, Baik, & Park, 2001; Ume, Martin, & Gatro, 1997; Xueren & Tong, 2004; Xie

* Corresponding author. Tel.: +886 4 23924505; fax: +886 4 23934620.
E-mail address: hys502@ncit.edu.tw (Y.H. Hung).

0957-4174/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2006.09.030

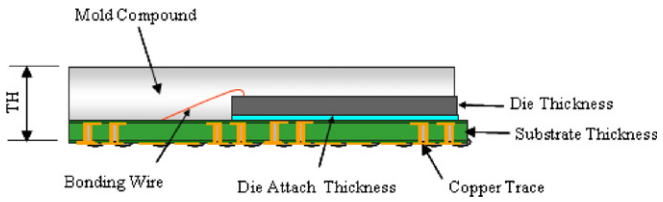

Fig. 1. Cross-sectional view of CSP package for micro HDD driver.

& Yi, 2002). In fact, the micro HDD driver IC package is more critical than is the case for the normal IC package concerning the reliability issues, as the assembly processes are much more numerous and the packages much thinner (Hung, Huang, & Chang, 2006). Upon the whole, for an ultra-thin CSP assembly, lowering warpage is a very important issue (Mertol, 2000), especially since warpage will result in a serious problem for CSP assembly when the overall package height is less than 1.0 mm. In recent years, quite a number of researchers have been conducted based on the finite element method (FEM), computational fluid dynamics (CFD) and heat transfer simulations for electronic packages and in predicting the warpage situation (Driel et al., 2003; Song, Zhang, Wang, & Diao, 2000). However, the execution time for these types of analyses can be in the order of hours or days for the warpage evaluation. Another paper presented the influence of the properties of materials and structure on the warpage of CSP packages. Mertol studied the low stress and low package warpage for the robust design of overmolded CSP on a flex-tape carrier with 280 solder balls, for the second level interconnect, with the overall package height of less than 1.2 mm by the Taguchi method (Mertol, 2000). The advantage of using the Taguchi method is in the reduction of both production cost and time. It concerns minimizing the effect of uncertainty or variation in design parameters (Phadke, 1989). However, most often, the Taguchi method has been applied to analyzing only linear systems under the assumption of the addition of individual factor effects; the goal is to find a design point that is robust in relation to variation in control or noise variables (Tong, Su, & Wang, 1997). Complicating matters is the fact that uncertainty is present in complex engineering systems such as micro HDD driver IC's CSP assembly. It has been a difficult task to accurately predict the micro HDD warpage situation. This often prevents the application of formal optimization techniques that can require many such evaluations.

Neural networks (NNs) have been used in a large number of applications and have proven to be effective in performing complex functions in various fields. NNs can be structured to perform classification (Raimundo & Narayanaswamy, 2001), to approximate equations (Joo et al., 2001), and to predict values (Freeman & Skapura, 1991; Winquist, Hornsten, Sundgren, & Lundstrom, 1993). The widely used algorithms of NN in approximate model are the back propagation neural networks (BPNNs) and radial basis function networks (RBFNs). In predictive modeling, the goal is to map a set of input patterns onto a set of output patterns. BPNNs have accomplished this task by learning from a series of data sets related to the system and then applying what was learned to approximate or predict the corresponding output. BPNNs, in conjunction with a statistical experimental design, have been widely used to construct semiconductor manufacturing prediction models (Ho, Xie, Tang, Xu, & Goh, 2001; Liau & Chen, 2005; Lo & Tsao, 2002). Adaptation or learning is a major focus of BPNN research that provides a degree of robustness to the NN model. The most successful BPNN learning model till now is the least mean square (LMS) algorithm, due to its high prediction accuracy having been excellently proven. However, despite the practical success, BPNNs suffered from increasing convergence. Additionally, building a BPNNs model is complicated by the presence of many training factors. Training factors typically involved in building a BPNNs model may include: the hidden neuron, training tolerance, initial weight distribution and function gradient. The most difficulty often arises from the nature of randomness in the initial weight distribution (Kim, Kim, & Park, 2005).

The other RBFN is a neural network, approached by viewing the design as a curve-fitting (approximation) problem in a high dimensional space. Learning is equivalent to finding a multidimensional function that provides a best fit to the training data, with the criterion for "best fit" being measured in some statistical sense. The RBFN is typically composed of three factors: the number of pattern units, the width of a radial basis function and the initial weight distribution between the pattern and output layers. Although the RBFN may require more neurons than standard feed-forward BPNN, they can often design in a fraction of the time that it takes to train standard feed-forward networks. One of the advantages of RBFN is the fact that linear weights associated with the output layer can be treated separately from the hidden layer neurons. This layer fulfils both the functions of information compression and pattern recognition. Consequently, the whole network's training time is enormously reduced. The aim of this study is to optimize the lower warpage properties for 0.65 mm CSP assembly using a model based on a RBFN-GA.

2. Experiment design

Our earlier study (Hung et al., 2006), concerned the lower warpage (less than 100 μm) issues on the vertical dimension of driver IC in the specification of 0.65 mm, based on the Taguchi method. The nine control factors included in a Taguchi orthogonal array $L_{27}(3^{13})$ experiments are: die thickness, die size, die attach thickness, mold thickness, mold compound, substrate thickness, cure temperature, cure time and package size for the driver IC for micro HDD. In particular, to satisfy the lead-free materials requirement, based on the different CTE α2 quality characteristics, different materials of mold compound were selected to observe the warpage change of the driver IC (shows as Table 1). Three mold compound (CTE α2) experimental levels are displayed in Table 2. The signal-to-noise

Table 1
Control factors and their levels

Control factors	Variables	Levels		
		1	2	3
Die thickness (mm)	A	75 μm	100 μm	125 μm
Die size (mm)	B	6 × 6 mm	9 × 9 mm	12 × 12 mm
Die attach thickness (mm)	C	20 μm	30 μm	40 μm
Mold thickness (mm)	D	300 μm	350 μm	400 μm
Mold compound	E	M1/25	M2/30	M3/35
Substrate thickness (mm)	F	110 μm	130 μm	150 μm
Cure temperature (°C)	G	170	175	180
Cure time (h)	H	2h	4h	6h
Package size (mm)	I	10 × 10 mm	12 × 12 mm	14 × 14 mm

Table 2
Mold compound material properties

	M1	M2	M3
CTE α1 (ppm/C)	8	8	8
CTE α2 (ppm/C)	25	30	35
Tg (C)	140	140	140
Filler content (wt%)	89	89	89
Young's modulus			
25.5 25.5	25.5	25.5	
Specific gravity	2.01	2.01	2.01
Curing temperature (C)	175 ± 5	175 ± 5	175 ± 5
Curing time (h)	2 ~ 6	2 ~ 6	2 ~ 6

(S/N) ratio (η) is an index of robustness in experimental processing, and the definition of S/N ratio for the smaller-the-better (STB) response by Phadke (1989) is as follows:

$$\eta = \mathrm{SN}_{\mathrm{STB}} = -10\log_{10}(\mathrm{MSD}) \quad (1)$$

where $\mathrm{MSD} = \left(\sum_1^n y_i^2\right)/n$, y_i is the ith observation and n is the number of observations in each combination.

The Taguchi experiment results combined (e.g., production processes parameters): the die thickness is thick, the die size is 6 mm × 6 mm, the die attach thickness is 30 μm, the mold thickness is 350 μm, the mold compound is level 'M1', the substrate thickness is 130 μm, the cure temperature is 170 °C, the cure time is 6 h, and the package size is 10 mm × 10 mm, respectively. However, according to Taguchi experiment parameters, the average value of the production quality level of warpage is $\bar{x} = 85$ μm, and the standard deviation is $s = 10.12$. Although Taguchi experiment could improve production quality effectively, the current production level is close to the upper limit of quality standard, thus, unable to reach the optimal design for production parameters.

Since large variability of CTE characteristic (shown as Fig. 2) exists among different mold thickness and mold compound type or substrate thickness and mold thickness for the different CTE α2 quality characteristics, the selection of a proper combination of the above four materials is essential for controlling the warpage level. However, the 0.65 mm driver IC's CSP packaging design problem that accurately handles uncertainty for these highly complex systems, further adds to the computational burden because each evaluation of the objective function for these systems can require sampling multiple points in the design space. Therefore, a well-trained RBFN was constructed to simulate and predict the output response for various control factor-level setting. The RBFN was trained using 27 data points generated by a three level Taguchi method. Each point represented a complete experiment combination of parameters.

3. Driver IC RBFN modeling

3.1. RBF network

RBFNs are composed of simple elements operating in parallel. These elements are inspired by biological nervous

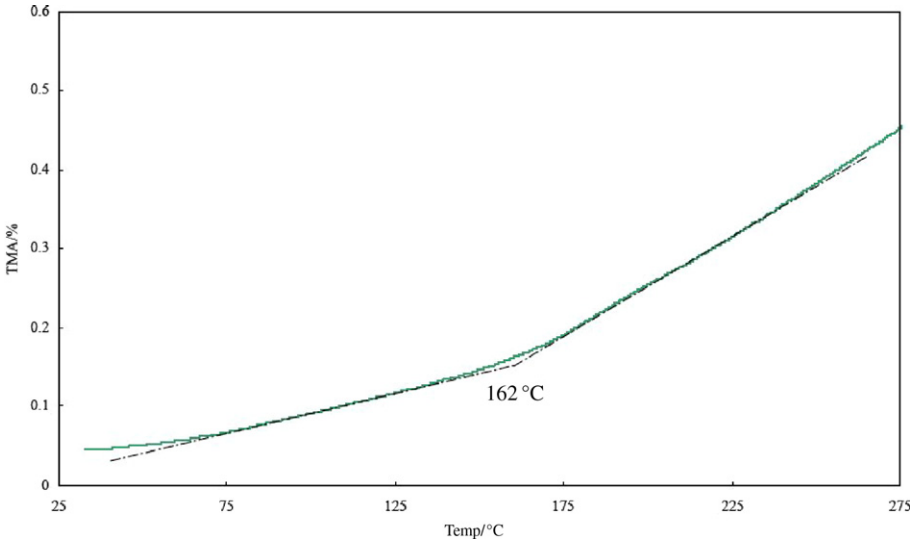

Fig. 2. The T_g curve between CTE ⟨alpha⟩ 1 and ⟨alpha⟩2.

systems. As in nature, the network function is determined largely by the connections between elements (neurons). Commonly RBFNs are adjusted, or trained, so that a particular input leads to a specific target output. Moody and Darken (1989) define a learning rule as a procedure for modifying the values of the connections (weights and biases) between neurons for a RBFN. The design of a RBFN in its most basic form consists of three separate layers: input layer, hidden layer and output layer. The hidden layer contains a number of RBF neurons, and each of them represents a single radial basis function. The output layer provides the response of the network to the activation patterns applied to the input layer. The architecture of the RBFN is shown in Fig. 3.

The RBFN has been defined, and the equation of the conventional RBFN form given, by Moody and Darken (1989):

$$f(x) = \sum_{i=1}^{n} w_i \phi_i(x) + b, \quad x = [x_1, x_2, \ldots, x_p]^T, \quad w = [w_1, w_2, \ldots, w_n]^T \tag{2}$$

where ϕ is a nonlinear activation function, x is the input vector and x_j is the jth input pattern, w_i is the weights of the ith activation function and b is the basis of the output layer neuron. The Gaussian function is chosen as the activation function by a center (μ) and a width (σ^2). Therefore, the nonlinear transformation function of the ith class of RBF is given as follows:

$$\phi_i(x) = e^{-\frac{\|x-\mu_i\|^2}{2\sigma_i^2}} \tag{3}$$

where e is the exponential function, μ_i is the ith RBF center determined by the Clustering algorithm, and σ_i is the widths of the ith RBF input patterns. Hence, the output function of the RBFN is also the linear combination of Gaussian functions. The output layer transfer function is linear, and given as follows:

$$f_j(x) = \sum_{i=1}^{n} w_{ji}\phi_i(x) + b_j = \sum_{i=1}^{n} w_{ji} e^{-\frac{\|x-\mu_i\|^2}{2\sigma_i^2}} + b_j \tag{4}$$

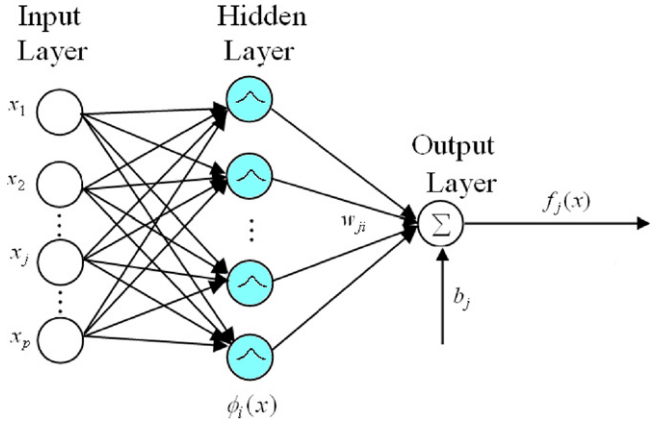

Fig. 3. The architecture of the radial basis function network.

where $f_j(x)$ is the output of the jth output layer neuron, w_{ji} is the weight of connection between ith hidden layer neuron and the ith output layer neuron, ϕ_i is the output of the ith hidden layer neuron, and b_j is the basis of the jth output layer neuron.

In the learning phase, there are two separate stages: unsupervised learning and supervised learning. The difference between supervised and unsupervised training is that, external prototypes are used as target outputs for specific input patterns, and the network is given a learning algorithm to follow and calculate new connection weights that bring the output closer to the target output. In the first stage, unsupervised learning, it is a process where a set of weights is defined that produces a desired response as a reaction to certain input patterns (Moody & Darken, 1989). The weights and biases are modified only in response to network inputs. There are no target outputs available. Most of these algorithms perform clustering operations. They categorize the input patterns into a finite number of classes. A very simple and effective clustering method is the K-means clustering method that works by starting out from a set of K initial center points and uses an iterative algorithm that minimizes the sums of distances from each object to its cluster centroid, over all clusters. This algorithm moves objects between clusters until the sum cannot be decreased any further. We denote the jth input pattern by x_j which has an associated target value t_j. The algorithm proceeds as follows:

(a) Randomly assign each pattern vector x_j to one of k subsets.
(b) Compute the mean vector of each subset.
(c) Reassign each point to subset with the closest mean vector.
(d) Continue until no further reassignments, and loop back to (b).

The final (equilibrium) locations of the centers μ_k are then used as center points for the hidden nodes in the RBFN. As the second training stage, the LMS algorithm is employed to perform the supervised training, to ensure the accuracy of the network. It is much more efficient than either the conjugate gradient algorithm or the variable learning rate algorithm. Therefore it is frequently used in implementations for the RBFN's training stage. Finally, the HDD RBFN model performance was measured by the typical root mean-squared error (RMSE) expressed as

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{m}(t_j - f_j(x))^2}{m}} \tag{5}$$

where m denotes the number of testing patterns.

3.2. HDD RBFN molding

The orthogonal array used in the study is $L_{27}(3^9)$ with four repetitions for each combination. There are a total

of 108 observations in our study. Based on the results of the Taguchi experiment, the data set contained nine input variables (control factors) and one target variable (η). Whatever, there are two kinds of input variable in the data set: seven continuous variables (factor A, C, D, E, F, G and H) and two nominal variables (factor B and I). Generally, the Driver IC RBFN training can be made more efficient if certain pre-processing steps are performed on the network input variables and targets. Before training, it is often useful to scale the input variables and targets so that they always fall within a specified range. In this study, the data set was pre-processed with a transformation encoding, one binary coding scheme was applied to the nominal variables and the other continuous variables were rescaled in range [0,1·]. For example, the "die size" (factor B) and the "package size" (factor I) variables were encoded as: a binary code "level 1" $\rightarrow (0,0,0)$, "level 2" $\rightarrow (0,+1,0)$ and "level 3" $\rightarrow (+1,+1,+1)$ in Table 1. Regarding the output, the target (η) was rescaled in range $[0,1]$. These features are listed in Table 3. Next, the training data were randomly selected from the data set, and set as the output of Driver IC RBFN to construct the nonlinear network models. In this study, 80% input patterns (21 observations) are used for training and the remaining 20% observations (six observations) are used for testing. The Driver IC RBFN simulation was performed using MATLAB® software.

4. Results and discussion

Too great a number of neurons cause over-training to influence the accuracy of the Driver IC RBFN model until the number of neurons has been optimized. In order to optimize the number of neurons for Driver IC RBFN, the SSE for the Driver IC RBFN models' predictions of the training set, and the RMSE for the Driver IC RBFN models' predictions of the testing set, were assessed for ascertaining the different number of neurons. An over-training was caused in testing set when neurons number more than 20 (shown as Fig. 5). Therefore, Fig. 4 and Fig. 5 show that 20 neurons should be extracted for the Driver IC RBFN model. Besides, the width increased incrementally from 1.0 to 10.0 with an increment of 0.5. The resulting RMSE variations are shown in Fig. 6. As shown in Fig. 7, the SSE for the prediction of the Driver IC RBFN is the least and the testing RMSE is 0.0034 when the width is 7.5 and the number of neurons is 20.

4.1. Optimal HDD drive IC packaging process parameters

Through the use of the Driver IC RBFN model, the prediction model between HDD drive IC packaging process parameters and warpage was generated. Moreover, the optimal HDD drive IC packaging process parameter

Table 3
A summary of the pre-processed data set based on the Taguchi experiment's

No.	Factors													Response	
	A	B			C	D	E	F	G	H	I			S/N (dB)	Normalized
1	0.6	0	0	0	0.50	0.75	0.71	0.73	0.94	0.33	0	0	0	−39.6031	0.218606
2	0.6	0	1	0	0.50	0.75	0.86	0.87	0.97	0.67	0	0	0	−40.0154	0.185057
3	0.6	1	1	1	0.50	0.75	1.00	1.00	1.00	1.00	0	0	0	−42.2508	0.003165
4	0.6	0	1	0	0.75	0.88	0.71	0.73	0.94	0.67	0	1	0	−38.9822	0.269128
5	0.6	1	1	1	0.75	0.88	0.86	0.87	0.97	1.00	0	1	0	−38.8970	0.27606
6	0.6	0	0	0	0.75	0.88	1.00	1.00	1.00	0.33	0	1	0	−39.6946	0.211161
7	0.6	1	1	1	1.00	1.00	0.71	0.73	0.94	1.00	1	1	1	−40.4239	0.151818
8	0.6	0	0	0	1.00	1.00	0.86	0.87	0.97	0.33	1	1	1	−40.9959	0.105275
9	0.6	0	1	0	1.00	1.00	1.00	1.00	1.00	0.67	1	1	1	−42.2897	0
10	0.8	0	1	0	0.75	1.00	0.71	0.87	1.00	0.33	0	0	0	−40.3799	0.155398
11	0.8	1	1	1	0.75	1.00	0.86	1.00	0.94	0.67	0	0	0	−41.2926	0.081133
12	0.8	0	0	0	0.75	1.00	1.00	0.73	0.97	1.00	0	0	0	−40.1646	0.172917
13	0.8	1	1	1	1.00	0.75	0.71	0.87	1.00	0.67	0	1	0	−40.1477	0.174292
14	0.8	0	0	0	1.00	0.75	0.86	1.00	0.94	1.00	0	1	0	−40.6657	0.132143
15	0.8	0	1	0	1.00	0.75	1.00	0.73	0.97	0.33	0	1	0	−41.1828	0.090067
16	0.8	0	0	0	0.50	0.88	0.71	0.87	1.00	1.00	1	1	1	−38.3892	0.31738
17	0.8	0	1	0	0.50	0.88	0.86	1.00	0.94	0.33	1	1	1	−40.6728	0.131565
18	0.8	1	1	1	0.50	0.88	1.00	0.73	0.97	0.67	1	1	1	−41.1682	0.091255
19	1	1	1	1	1.00	0.88	0.71	1.00	0.97	0.33	0	0	0	−38.5958	0.300569
20	1	0	0	0	1.00	0.88	0.86	0.73	1.00	0.67	0	0	0	−38.0205	0.34738
21	1	0	1	0	1.00	0.88	1.00	0.87	0.94	1.00	0	0	0	−39.1645	0.254294
22	1	0	0	0	0.50	1.00	0.71	1.00	0.97	0.67	0	1	0	−40.1116	0.17723
23	1	0	1	0	0.50	1.00	0.86	0.73	1.00	1.00	0	1	0	−40.1037	0.177873
24	1	1	1	1	0.50	1.00	1.00	0.87	0.94	0.33	0	1	0	−40.9754	0.106943
25	1	0	1	0	0.75	0.75	0.71	1.00	0.97	1.00	1	1	1	−40.6653	0.132176
26	1	1	1	1	0.75	0.75	0.86	0.73	1.00	0.33	1	1	1	−40.0880	0.17915
27	1	0	0	0	0.75	0.75	1.00	0.87	0.94	0.67	1	1	1	−39.3314	0.240714

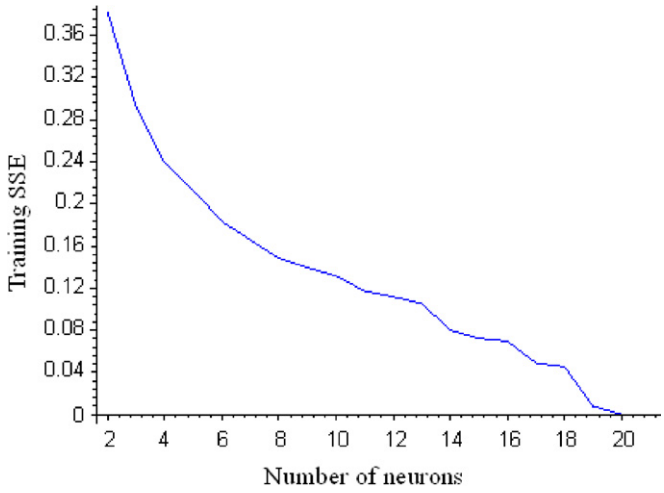

Fig. 4. Training errors for selected number neurons of RBFN mode.

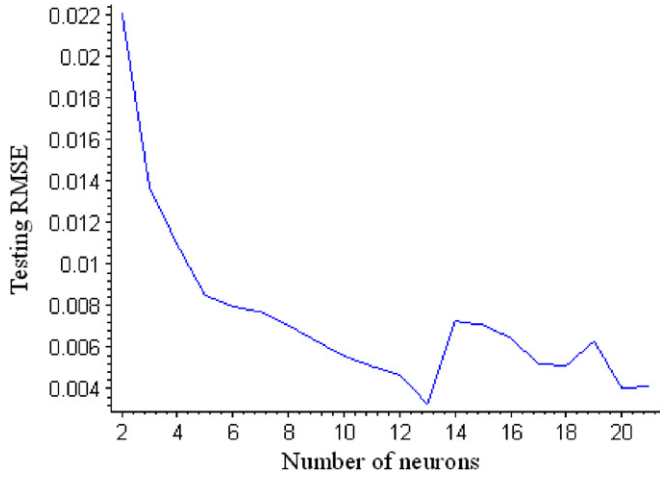

Fig. 5. Testing errors for selected number neurons of RBFN model.

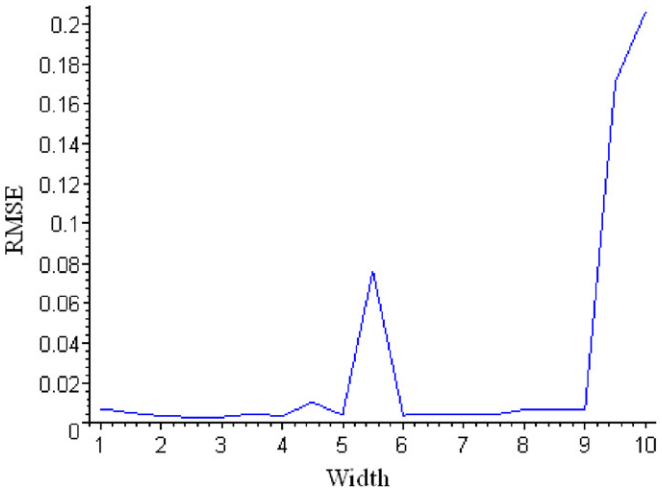

Fig. 6. RMSE errors for selected width.

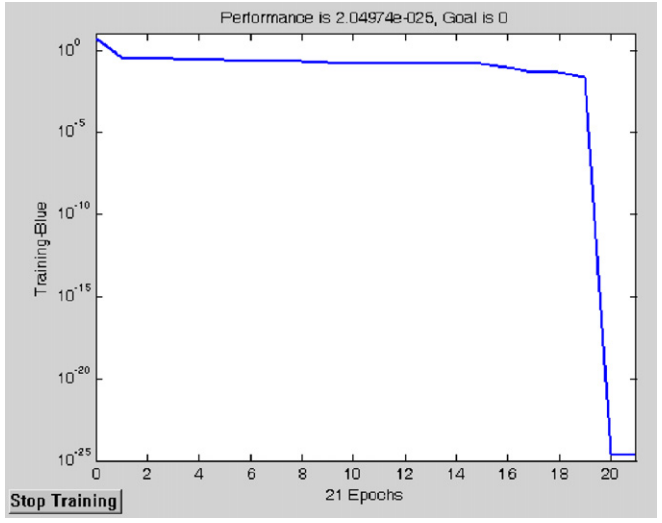

Fig. 7. Training SSE for RBFN model.

designs were obtained by genetic algorithm (GA) implementation in this section. GA is widely used to search for optimized parameters satisfying given constraints (Goldberg, 1989). The advantage of the GA approach lies in the ease with which it can handle arbitrary kinds of constraints and objectives. When solving multi-objective problems, GA provides many satisfactory solutions in terms of the objectives, and then allows the decision maker to select the best alternative. Therefore GA is most useful for problems involving multimodal design spaces.

The GA is a method for solving both constrained and unconstrained optimization problems, and one that is based on natural selection, the process that drives biological evolution. The GA repeatedly modifies a population of individual solutions. At each step, the GA selects individuals at random from the current population to be parents and uses them to produce the children for the next generation. In GA optimization, the factors involved are: the size of the initial population, the crossover probability, the mutation probability, and a fitness function. A fitness function is used to evaluate individuals, since reproductive success varies with fitness. The optimal HDD packaging process parameter design, achieved through the procedures of executing GA is simplified as follows.

Step 1: All of the HDD drive IC packaging process parameters will vary within a known range and encoded as a binary string (chromosome). We use 10-bit encoding which results in a 130-bit chromosome for each parameter.
Step 2: Set the GA's operating conditions: the generation size was set to 1500, the size of the initial population was set to 100, the crossover and mutation probabilities were set to 0.8 and 0.01, respectively.
Step 3: To create a random initial population.
Step 4: Scoring each member of the current population by computing the individual's fitness value. The fitness function can be evaluated as RMSE calculated in the checking sample or otherwise according to the specificity

Table 4
Comparison of the Taguchi result, RBFN-GA optimized, and adjustment result

Method	Factors' level													Prediction S/N (dB)
	A		B		C	D	E	F	G	H	I			
Orthogonal array 20th run	125	0	0	0	40	350	30	110	180	4	0	0	0	−38.0205
Taguchi's result	125	0	0	0	30	350	25	130	170	6	0	0	0	−36.956
RBFN-GA's result	114.88	−0.072	0.189	−0.099	35.61	327.71	27.10	123.76	176.3	3.999	−0.0408	0.2691	0.0847	−35.512
Adjustment-GA result	115	0	0	0	35	330	27.0	120	176	4	0	0	0	−35.882

of the task. The GA algorithm meeting a given fitness function is expressed as

$$Fitness = 10000 \times \frac{-1}{\mathrm{SN}_{\mathrm{prediction}}} \tag{6}$$

Step 5: Selection of members: the members of the new population are selected based on their fitness. Elite children are the individuals in the current generation with lower fitness values. Roulette wheel selection is employed in this algorithm. These individuals automatically survive to the next generation.
Step 6: Production of children from their parents. Dependent on the crossover rate, crossover of the bits from each chosen chromosome occurs at a random position, where there is an interchange between the two parts. Proceed through the chosen chromosome bits and flip them in dependence to the mutation rate.
Step 7: Replace the current population with the children, to form the next generation.
Step 8: Steps 4, 5, 6 and 7 are repeated until a stopping criteria is met.

The comparisons on parameter settings among the Taguchi method, and RBFN-GA are listed in Table 5. Although the parameter of RBFN-GA could elevate the production standard to −35.512 dB, it is still limited to the specifications (such as package size) and parameters (such as die size, die attach thickness, and cure temperature) of the molds setting in practice. We use the parameters from RBFN-GA to adjust parameters of adjustment-GA in Table 4, and predict the output result of −35.882 dB by using the RBFN model. The results are sent to a semiconductor manufacturer for three tests, each drawing 30 samples. The production standard C_{pk} of the tests is shown in Table 5. The final parameter setting from adjustment-GA result is 115 μm, 6 mm × 6 mm, 35 μm, 330 μm, 27.0 (ppm/C), 120 μm, and 176 (°C), 4 h, and 10 mm × 10 mm for die thickness, die size, die attach thickness, mold thickness, mold compound, substrate thickness, cure temperature, cure time, and package size, respectively. According to the experimental results, the production parameters derived from RBFN-GA have excellent control over production variation, and the overall production competency index C_{pk} reaches over 1.80. This proves that this study could effectively improve the production competency and the competitive advantages of semiconductor manufacturers.

Table 5
Adjustment-GA experiments' result

Batch		Max	Min	$\bar{x}$	s	C_{pk}
1	(mil)	2.57	1.24	2.00	0.3617	1.8451
	(μm)	65	31	51	9.19	1.79
2	(mil)	2.48	1.40	2.02	0.3113	2.1229
	(μm)	63	36	51	7.91	2.06
3	(mil)	2.28	1.18	1.77	0.2992	2.4857
	(μm)	58	30	45	7.6	2.42

5. Conclusion

The warpage is an important issue related to micro HDD Driver IC manufacturability and reliability, especially, when the size of Driver IC for micro HDD becomes smaller and thinner. This study combines RBFN with Taguchi method to structure a well-trained prediction model and further to search for the optimal HDD drive IC packaging process parameter design through GA. According to the experimental results, the production parameters derived from RBFN-GA have excellent control over production variation, and the overall production competency index C_{pk} reaches over 1.80. This proves that this study could effectively improve the production competency and the competitive advantages of semiconductor manufacturers.

References

Baliga, J. (1998). Making room for more performance with chip scale packages. *Semiconductor International*, 85–92.
Coughlin, T. M. Hanlon, P., & Waid, D. (2004). Looking for storage in all the small places, Entertainment and digital media storage report. http://www.tomcoughlin.com/techpapers.htm.
Driel, W. D. V. et al. (2003). Prediction and verification of process induced warpage of electronic packages. *Microelectronics Reliability, 43*, 765–774.
Egan, E., Kelly, G., O'Donovan, T., & Kennedy, P. (2003). Response surface methodology for enhancing theoretical models: Application to warpage predication of CSP BGAs. In: *Proceeding of the EuroSimE conference* (pp. 221–228). France.
Freeman, J., & Skapura, D. (1991). *Neural networks, algorithms, applications, and programming techniques*. Addison-Wesley, pp. 89–111.
Goldberg, D. E. (1989). *Genetic algorithms in search optimization and machine learning*. Reading MA: Addison-Wesley.

Ho, S. L., Xie, M., Tang, L. C., Xu, K., & Goh, T. N. (2001). Neural network modeling with confidence bounds: A case study on the solder paste deposition process. *IEEE Transaction on Electronics Packaging manufacturing, 24*(4), 323–332.

Hung, Y. H., Huang, M. L., & Chang, C. H. (2006). Optimizing the controller IC for micro HDD process based on Taguchi methods. *Microelectronics Reliability, 46*, 1183–1188.

Joo, B. S., Choi, N. J., Lee, Y. S., Lim, J. W., Kang, B. H., & Lee, D. D. (2001). Pattern recognition of gas sensor array using characteristics of impedance. *Sensors and Actuators B Chemical, 77*(1/2), 209–214.

Kim, B., Kim, D. W., & Park, G. T. (2005). Prediction of plasma-induced DC bias using polynomial neural network. *Vacuum, 79*, 111–118.

Li, Y. (2003). Accurate predictions of flip chip BGA warpage. In: *Proceeding of the 53rd ECTC conference* (pp. 549–553).

Liau, L. C. K., & Chen, B. S. C. (2005). Process optimization of gold stud bump manufacturing using artificial neural networks. *Expert Systems with Applications, 29*, 264–271.

Lo, Y. L., & Tsao, C. C. (2002). Integrated Taguchi method and neural network analysis of physical profiling in the wirebonding process. *IEEE Transaction on Components and Packaging Technology, 25*(2), 270–277.

Mertol, A. (2000). Application of the Taguchi method to chip scale package (CSP) design. *IEEE Transaction on Advance Package, 23*(2), 266–276.

Miyake, K., Yoshida, T., Baik, H. G., & Park, S. W. (2001). Viscoelastic warpage analysis of surface mount package. *ASME Journal of Electronic Packaging, 123*, 101–104.

Moody, J., & Darken, C. (1989). Fast learning in networks of locally-tuned processing units. *Neural Computation, 1*, 281–294.

Phadke, M. S. (1989). *Quality engineering using robust design*. Englewood Cliffs, NJ: Prentice-Hall.

Raimundo, I. M., & Narayanaswamy, R. (2001). Simultaneous determination of relative humidity and ammonia in air employing an optical fibre sensor and artificial neural network. *Sensors and Actuators B – Chemical, 74*(1-3), 60–68.

Song, Y. H., Zhang, K. F., Wang, Z. R., & Diao, F. X. (2000). 3-D FEM analysis of the temperature field and the thermal stress for plastics thermal forming. *Journal of Materials Processing Technology, 97*, 35–43.

Tong, L. I., Su, C. T., & Wang, C. H. (1997). The optimization of multi-response problems in Taguchi method. *International Journal Quality Reliability Management, 14*, 367–380.

Ume, C., Martin, T., & Gatro, J. (1997). Finite element analysis of PWB warpage due to the solder masking process. *IEEE Trans-CPMT-A, 20*(2), 295–306.

Winquist, F., Hornsten, E. G., Sundgren, H., & Lundstrom, I. (1993). Performance of an electronic nose for quality estimation of ground meat. *Measurement Science and Technology, 4*(12), 1493–1500.

Xie, D., & Yi, S. (2002). Reliability studies and design improvement of mirror image CSP assembly. *Microelectronics Reliability, 42*, 1931–1937.

Xueren Z., & Tong Y.T. (2004). Advanced warpage prediction methodology for matrix stacked die BGA during assembly processes. *Electronic components and technology conference* (pp. 593–600).