

# Improved model of the retardance in citric acid coated ferrofluids using stepwise regression

J F Lin<sup>1</sup> and X R Qiu<sup>2</sup>

<sup>1</sup> Department of Industrial Design, Far East University, No.49, Zhonghua Rd., Xinshi Dist., Tainan City 74448, Taiwan (R.O.C.)

<sup>2</sup> Graduate School of Mechanical Engineering, Far East University, No.49, Zhonghua Rd., Xinshi Dist., Tainan City 74448, Taiwan (R.O.C.)

E-mail: jacklin@cc.feu.edu.tw

**Abstract.** Citric acid (CA) coated  $\text{Fe}_3\text{O}_4$  ferrofluids (FFs) have been conducted for biomedical application. The magneto-optical retardance of CA coated FFs was measured by a Stokes polarimeter. Optimization and multiple regression of retardance in FFs were executed by Taguchi method and Microsoft Excel previously, and the  $F$  value of regression model was large enough. However, the model executed by Excel was not systematic. Instead we adopted the stepwise regression to model the retardance of CA coated FFs. From the results of stepwise regression by MATLAB, the developed model had highly predictable ability owing to  $F$  of  $2.55897 \times 10^7$  and correlation coefficient of one. The average absolute error of predicted retardances to measured retardances was just 0.0044%. Using the genetic algorithm (GA) in MATLAB, the optimized parametric combination was determined as [4.709 0.12 39.998 70.006] corresponding to the pH of suspension, molar ratio of CA to  $\text{Fe}_3\text{O}_4$ , CA volume, and coating temperature. The maximum retardance was found as  $31.712^\circ$ , close to that obtained by evolutionary solver in Excel and a relative error of -0.013%. Above all, the stepwise regression method was successfully used to model the retardance of CA coated FFs, and the maximum global retardance was determined by the use of GA.

## 1. Introduction

Citric acid (CA) coated  $\text{Fe}_3\text{O}_4$  ferrofluids (FFs) have been conducted for biomedical application such as magnetic resonance imaging (MRI) and tumor hyperthermia. It has been known that the surface of the magnetic nanoparticles (MNPs) can be stabilized in an aqueous dispersion by the adsorption of CA. This process, as described by Sahoo et al. [1], occurs by the CA being coordinated via the carboxylate functionalities. Srivastava et al. [2] synthesized CA coated  $\text{Fe}_3\text{O}_4$  MNPs of 6 nm particle size, which exhibits excellent magnetic properties. Effect of synthesis conditions on the properties of CA coated iron oxide nanoparticles was discussed [3].

In the present investigation,  $\text{Fe}_3\text{O}_4$  MNPs were prepared by an improved co-precipitation of  $\text{Fe}^{3+}/\text{Fe}^{2+}$  salts and then the MNPs were stabilized against agglomeration by surfactant encapsulating of CA [4]. Afterwards, the CA coated  $\text{Fe}_3\text{O}_4$  MNPs were used as the precursor of water-based FFs, and the retardance (magneto-optical property as the phase retardation of linearly birefringent medium, such as quartz and certain liquid crystals) of FFs were measured by a developed Stokes polarimeter [5]. It is known that Taguchi orthogonal design method may provide a powerful and efficient ways to find an optimal combination of factor levels that may achieve optimum. Hence, the Taguchi method with



range analysis was employed to find the parametric combination for CA coated FF with high retardance readily [4].

Optimization and multiple regression of retardance in FFs were executed by Taguchi method and Microsoft Excel, and the  $F$  value of regression model was large enough [6]. However, the modelling of retardance executed by Microsoft Excel was not in a systematic way. Further, using Taguchi-based measured retardances as the training data, an artificial neural network (ANN) model was developed for the prediction of retardance in CA coated FF. Based on the well-trained ANN model, the predicted retardance at excellent program from Taguchi method showed less error of 2.17% compared with a multiple regression (MR) analysis of statistical significance [7].

The aim of this study was to use the stepwise regression for modelling retardance of CA coated FFs. The stepwise regression by MATLAB was executed successfully for modelling retardance of CA coated FFs. The regression model had highly predictable ability and had a high correlation coefficient of one. Moreover, two optimization techniques, including evolutionary algorithm (in Microsoft Excel) and genetic algorithm solver (in MATLAB® Optimization Toolbox), were used to find the global maximum retardance value. Then the CA coated FFs could be more useful in practical applications.

## 2. Method

In this study, the methods including Taguchi orthogonal design method, stepwise regression method, evolutionary algorithm and genetic algorithm were briefly introduced.

### 2.1. Taguchi method

Taguchi method is straightforward and easy to apply to many engineering situations, making it a powerful yet simple tool. Subsequently, the optimal synthetic condition of CA coated  $\text{Fe}_3\text{O}_4$  FFs with high retardance was determined by the Taguchi orthogonal design method- $L_9(3^4)$ , i.e. four parameters with three levels, respectively, and nine tests of FFs with 1000 g/L, correspondingly [4]. Influence parameters were (A) pH value of suspension after coating (4.5, 5, 5.5), (B) molar ratio of CA to  $\text{Fe}_3\text{O}_4$  MNPs (0.03, 0.06, 0.12), (C) CA volume (10 ml, 20 ml, 40 ml), and (D) coating temperature (70 °C, 80 °C, 90 °C). Procedure of the chemical co-precipitation method for producing CA coated FFs could be found [4]. The retardance (magneto-optical property) of FF was measured by a Stokes polarimeter with a feasible algorithm [5]. Results of the retardance under 64.5 mT were as shown in table 1 [4]. Moreover, the influence sequence between these parameters on the retardance and the excellent program (parametric combination) of maximum retardance were decided by a simple range analysis.

**Table 1.** Orthogonal design and test results of retardance.

No	A	B	C	D	Retardance (deg.)
<b>A1</b>	1	1	1	1	7.8228
<b>A2</b>	1	2	2	2	11.9872
<b>A3</b>	1	3	3	3	29.3618
<b>A4</b>	2	1	2	3	4.0525
<b>A5</b>	2	2	3	1	23.6294
<b>A6</b>	2	3	1	2	18.1662
<b>A7</b>	3	1	3	2	-0.4628
<b>A8</b>	3	2	1	3	4.5877
<b>A9</b>	3	3	2	1	20.2021

### 2.2. Stepwise regression

While dealing with large number of independent variables, it is of significant importance to determine the best combination of these variables to predict the dependent variable [8]. Stepwise regression method is a routine statistic technique used for variable selection and is a combination of forward and backward procedures. The method starts with an initial model which includes a subset of all the candidate variables. And then, a so-called stepping procedure which iteratively altering the model

established at the previous step by adding or removing variables in accordance with the stepping criteria is used in order to add significant variables or remove insignificant variables.

The stepping procedure terminates when the model has been optimized, or when a specified maximum number of steps have been reached. The obtained model can be tested for validity through a variety of statistical methods. Usually, the evaluation of the model is made by using some criteria: correlation coefficient ( $R$ ), adjusted square of correlation coefficient (adjusted  $R^2$ ), standard error (SE), and  $F$  statistic [9]. The details on the use of stepwise model-building procedure could be found [8, 10].

In our previous study, optimization and multiple regression of retardance in FFs were executed by Taguchi method and Microsoft Excel, and the  $F$  value of regression model was large enough [6]. However, the modelling of retardance executed by Microsoft Excel was not in a systematic way, and the execution is time-consuming. Therefore, the stepwise regression method can be adopted to model the retardance of CA coated FFs efficiently.

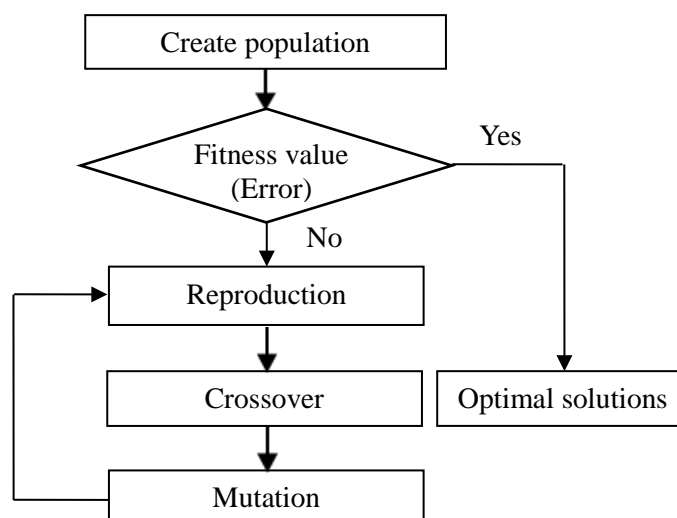
### 2.3. Evolutionary algorithm in Microsoft Excel

We can use the evolutionary solver [11] in Microsoft Excel to solve optimization problems. A genetic or evolutionary algorithm applies the principles of evolution found in nature to the problem of finding an optimal solution. In a “genetic algorithm,” the problem is encoded in a series of bit strings that are manipulated by the algorithm; in an “evolutionary algorithm,” the decision variables and problem functions are used directly.

Most commercial solver products are based on evolutionary algorithms. The evolutionary solver combines genetic algorithm methods such as mutation, crossover, and natural selection with classic methods drawn from linear programming, nonlinear optimization, and pattern search, to find better solutions in less time. Further, on smooth nonlinear problems where the GRG (Generalized Reduced Gradient) solver can find only a locally optimal solution, the evolutionary solver can often find a better solution or even a globally optimal (or close to globally optimal) solution, which is described by Frontline System [11]. Therefore, the optimized parametric combination and the maximum retardance of CA coated FFs could be obtained by the evolutionary algorithm in Microsoft Excel, which is capable of solving smooth and non-smooth global optimization problems.

### 2.4. Genetic algorithm

The genetic algorithm (GA) technique is based on the natural process of evolution to solve optimization and search problems. There are three main operators in GA which are reproduction, crossover and mutation [12]. GAs were pioneered by Holland in the 1960s and 70s [13] and have emerged as a highly effective technique for solving a wide variety of optimization problems [14].



**Figure 1.** Flowchart of genetic algorithm.

Figure 1 presents a simple flowchart representation of a typical GA solution procedure [14]. As shown, the process commences by creating an initial population of random solutions to the problem of interest. Each potential solution is then evaluated using a fitness function, i.e. an objective function which quantifies the optimality of each solution such that it can be ranked against that of all the other potential solutions. If the termination criteria are not satisfied, multiple candidate solutions are selected for modification via reproduction, crossover and mutation operations in order to create a new population pool. The optimality of each member of the population pool is then re-evaluated using the fitness function.

### 3. Results and discussions

Results of stepwise regression (first order and second order) on the Taguchi-based measured retardances of CA coated FFs are provided below. And the optimized parametric combination and the maximum retardance of CA coated FFs are obtained by genetic algorithm, which are compared with those results obtained by the evolutionary solver in Microsoft Excel.

#### 3.1. Results of stepwise regression

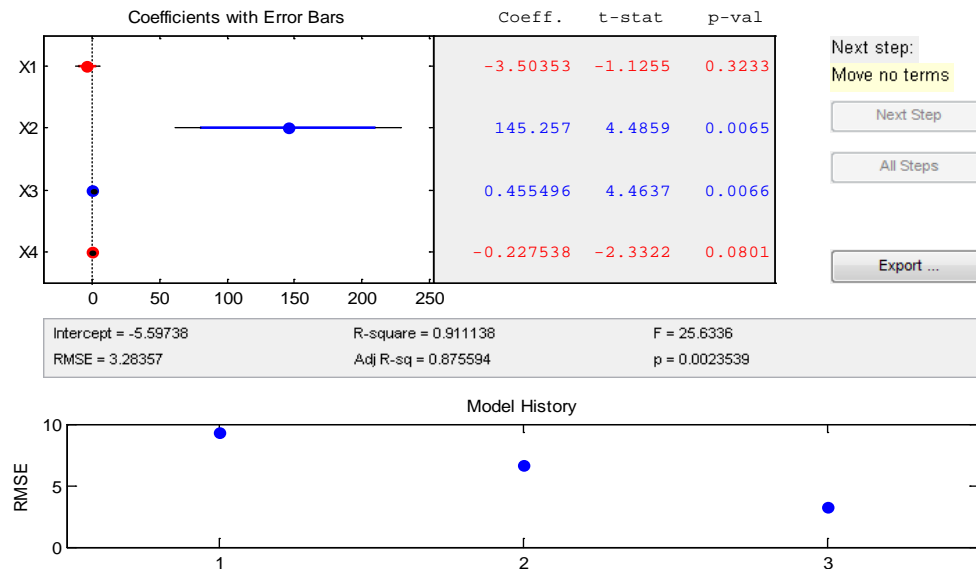
Based on the Taguchi-based measured retardances of CA coated FFs (as shown in table 1, the retardance value of the A7 sample is not considered due to the possible error), the results of first order stepwise regression executed by MATLAB are as shown in figure 2. We can see the meaningfulness degrees of input variables. This degree of meaningfulness is determined via being value of  $p$ -val ( $P$  value,  $P$  represents probability) below 0.05. Thus meaningfulness ranking of input variables is determined as variable 2 and variable 3, i.e. molar ratio of CA to  $\text{Fe}_3\text{O}_4$  MNPs and CA volume, respectively. The variable 2 (molar ratio of CA to  $\text{Fe}_3\text{O}_4$  MNPs) involved in this model represents the most significant influencing factor for the retardance of CA coated FFs. Based on the results obtained; equation of the first order stepwise regression model is given by equation (1):

$$\text{Retardance} = -5.59738 - 3.50353X_1 + 145.257X_2 + 0.455496X_3 - 0.227538X_4 \quad (1)$$

From the results of figure 2, the  $R^2$  value is 0.911138. This indicates that the current equation is able to describe 91.1138 percent of the variation in the response variable; this doesn't guarantee the appropriateness of the first order model. In addition,  $t$  statistic ( $t$  Stat) in table 2 is equal to the ratio of respective standard error to the coefficient value and tests the significance of individual regression coefficients for a confidence interval as 95%. Under the normality assumption, for  $n$  number of observations,  $t$  statistic follows student's  $t$  distribution with  $(n-2)$  degrees of freedom (DF), and critical value of  $t$  can also be found out by using TINV function in Microsoft Excel which comes out to be 2.36. As can be seen that  $X_2$  coefficient ( $4.4859 > t$  critical value of 2.36) has the strongest effect on the regression model followed by the  $X_3$  coefficient.  $P$  value in table 2 is the probability that for a particular confidence level as 95%, a random variable having a student's  $t$  distribution is greater than absolute value of observed  $t$  statistic and hence significant.

For testing whether the simple stepwise regression model gives any useful information about the response variable or not, as shown in figure 2.  $F$  statistic is the ratio of mean squares due to regression model and error, and is used to test the significance of regression model for a confidence level. The  $F$ -test is executed by comparing the observed  $F$  statistic with the appropriate critical value obtained from the standard  $F$  distribution table. Critical value of  $F$  for the current regression model can be found out from standard  $F$  distribution table by  $F(4, 4, 0.05)$  or by using FINV function in Microsoft Excel, which comes out to be 6.388. As can be seen that the observed  $F$  value of 25.6336  $>$  6.388, it implies the regression model is significant. Figure 2 shows the corresponding probability value  $P(F)$ . It can be deduced that the regression model is statistically significant as  $P$  value (0.0023539, chance of 0.23539% that such model  $F$ -value could occur due to noise) is less than 0.05 (i.e. 95% confidence level). Compared to the results obtained by multiple linear regression in Microsoft Excel [6], the determination coefficient  $R^2$  and the observed  $F$  value of the first order stepwise regression model are larger than those of 0.888 and 7.954, respectively [6]. It can be deduced that the first order stepwise

regression model is statistically significant and the first order stepwise regression is more reliable than the multiple linear regression.



**Figure 2.** Results of stepwise regression (first order).

**Table 2.** Results of stepwise regression (first order).

	Coefficient	<i>t</i> Stat	<i>P</i> value
$X_1$ (pH)	-3.50353	-1.1255	0.3233
$X_2$ (Molar Ratio)	145.257	4.4859	0.0065
$X_3$ (CA Volume)	0.455496	4.4637	0.0066
$X_4$ (Temperature)	-0.227538	-2.3322	0.0801

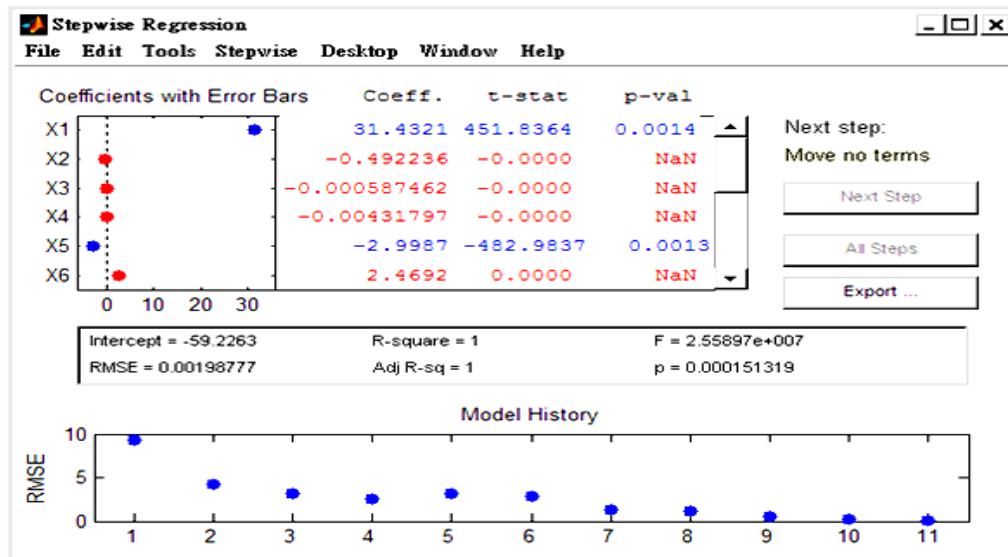
Subsequently, we aim to increase the precision of regression model and precede the second order stepwise regression. The results of stepwise regression by MATLAB are shown in figure 3, the model has highly predictable ability owing to  $F$  of  $2.55897e+7$  and has a high correlation coefficient  $R$  of one, which is suggested that this regression model is statistically significant and gives useful information about the response variable. The average absolute error of predicted retardances is less just as 0.0044%. The equation for the second order stepwise regression model is given by equation (2):

$$\text{Retardance} = -59.2263 + 31.4321X_1 - 2.9987X_1^2 + 0.00801827X_3^2 - 0.000804091X_4^2 - 0.0458668X_1X_4 + 1.8636X_2X_4 \quad (2)$$

As shown in figure 3 and table 3, we can see the meaningfulness degrees of input variables. This degree of meaningfulness is determined via being value of  $p$ -val below 0.05. Thus meaningfulness ranking of input variables is determined as  $13(X_2X_4)$ ,  $7(X_3^2)$ ,  $5(X_1^2)$ ,  $1(X_1)$ ,  $11(X_1X_4)$ , and  $8(X_4^2)$ . As shown in figure 3, the RMSE (Root Mean Square Error) value is 0.00198777 and is much lower than that of 3.28357 in the first order stepwise regression as shown in figure 2. Further, the  $R^2$  value is determined as one and is higher than 0.911138 in the proposed first order stepwise regression (in figure 2) and 0.9999998 in the multiple nonlinear regression model [6]. It is indicated that the ability of prediction in the second order stepwise regression model is excellent.

The significance  $t$ -test of each variable is listed in table 3. It is seen that the significant influence of each term on the retardance of CA coated FFs is given by:  $X_1 > X_3 > X_4$ . This indicates that the effect of coating temperature on the retardance is of the least significance. Although the effect of the coating temperature on the retardance is not significant, the cross correlations of coating temperature and other two parameters as pH value of suspension and molar ration of CA to  $\text{Fe}_3\text{O}_4$  MNPs ( $X_2X_4$  and  $X_1X_4$ )

are significant. The relative influence of cross correlations on the retardance of CA coated FFs (from examining the magnitude of the  $P$  value or from  $t$  Stat checking the significance level of the cross items) is given by:  $X_2X_4 > X_1X_4$ . Although the item of  $X_2$  seems not significant in second order stepwise regression, the  $X_2$  plays an important role in the influence of retardance judging from the  $X_2X_4$ . In addition, factors at higher order (i.e.  $X_3^2$  and  $X_4^2$ ) in equation (2) denote positive and negative quadratic relationships, respectively. Therefore, the effect of each factor on the experiment indices must be taken into account during the experiment.



**Figure 3.** Results of stepwise regression (second order).

**Table 3.** Results of stepwise regression (second order).

	Coefficient	t Stat	P value
$X_1$	31.4321	451.8364	0.0014
$X_1^2(X_5)$	-2.9987	-482.9837	0.0013
$X_3^2(X_7)$	0.00801827	4644.0641	0.0001
$X_4^2(X_8)$	-0.000804091	-86.9664	0.0073
$X_1X_4(X_{11})$	-0.0458668	-156.2040	0.0041
$X_2X_4(X_{13})$	1.8636	6389.4874	0.0001

### 3.2. Results of optimization using genetic algorithm

Based on equation (2), the evolutionary solver in Microsoft Excel is used to solve optimization problem firstly. The optimized parametric combination and the maximum retardance of CA coated FFs are obtained as [4.7037 0.12 39.99991 70.0021], corresponding to the pH of suspension, molar ratio of CA to  $\text{Fe}_3\text{O}_4$ , CA volume, coating temperature, and  $31.716^\circ$ , respectively. Also, the GA in MATLAB is used to solve the optimization problem. As shown in figure 4, the optimized parametric combination and maximum retardance of CA coated FFs are determined as [4.709 0.12 39.998 70.006] and  $31.712^\circ$ , close to those obtained by Microsoft Excel and a relative error of -0.013%. Above all, we successfully model the retardance in CA coated FFs by stepwise regression and the maximum global retardance is determined by the GA.

It is noted that the Fitness function as @FitFun75 is obtained by changing the sign of equation (2) and is used to find the minimum value. The results obtained by nine times of executions of GA in MATLAB are listed in table 4, and figure 4 is corresponding to the results of the second execution of GA.





**Figure 4.** A graph output using the GA in MATLAB® optimization.

**Table 4.** Optimized parametric combination and maximum retardance.

No	pH	Molar Ratio	CA Volume	Coating Temp.	Retardance (deg)
1	4.693	0.12	39.999	70.005	31.714
2	4.709	0.12	39.998	70.006	31.712
3	4.706	0.12	39.994	70	31.684
4	4.722	0.12	39.999	70.024	31.710
5	4.673	0.12	39.997	70.213	31.680
6	4.697	0.12	39.998	70.001	31.714
7	4.693	0.119	39.998	72.979	31.273
8	4.71	0.12	39.998	70.108	31.703
9	4.72	0.12	39.999	70.487	31.664

In overall, Taguchi method is a discontinuous optimization approach, which can only obtain the local discontinuous optimal solution of pre-selected parameter level value, and cannot find out the global optimal solution [4]. The integration of Taguchi method with various approaches including numerical simulation, principal component analysis (PCA), artificial neural network (ANN), response surface methodology (RSM), and genetic algorithm (GA) could be used to enhance the efficiency of the optimization process. In addition, it is noted that GA based on a neural network model has been applied successfully to optimize complicated bioprocesses [4].

In future work, the predicted retardance value obtained by the trained ANN model [7] could be adopted as the fitness value, further, the GA operation such as select, crossover, and mutation are executed to find the global retardance of CA coated FFs. Moreover, the multiple-objective optimization including maximum retardance and low dichroism of CA coated FFs [4] or in other double-layer coated FFs is supposed to be performed in future. Therefore, the practical biomedical

applications such as hyperthermia (magnetic inductive heating of cancer tumour) and magnetic resonance imaging (MRI) could be executed.

#### 4. Conclusions

The stepwise regression method was successfully used to model the retardance of CA coated FFs. According to the results of second order stepwise regression, the  $F$  value was obtained as high as  $2.55897 \times 10^7$  and had a correlation coefficient of one; the developed regression model had highly predictable ability. Therefore, the retardance value, corresponding to each parametric combination, built by four parameters with three levels in Taguchi method, could be precisely determined by regression equation. Also, using the genetic algorithm in MATLAB, the optimized parametric combination was determined as [4.709 0.12 39.998 70.006]. The maximum retardance was found as  $31.712^\circ$ , close to that obtained by the evolutionary solver in Microsoft Excel and a relative error of -0.013%.

#### Acknowledgements

The financial support provided by the Ministry of Science and Technology in Taiwan under MOST 104-2221-E-269-006 and 105-2221-E-269-001 is gratefully acknowledged.

#### References

- [1] Sahoo Y et al 2005 *J. Phy. Chem. B* **109** 3879
- [2] Srivastava S et al 2011 *J. Colloid Interface Sci.* **359** 104
- [3] Li L et al 2013 *Microel. Eng.* **110** 329
- [4] Lin J F et al 2014 *J. Magn. Magn. Mater.* **372** 147
- [5] Lin J F and Lee M Z 2012 *Opt. Commun.* **285** 1669
- [6] Lin J F et al 2015 *Proc. SPIE* **9302** 930229
- [7] Lin J F and Sheu J J 2016 *J. Magn. Magn. Mater.* **404** 201
- [8] Cevik A et al 2010 *Adv. Eng. Softw.* **41** 611
- [9] Jiao L and Li H 2010 *Chemometr. Intell. Lab. Syst.* **103** 90
- [10] Hsieh K L and Lu Y S 2008 *Expert Syst. Appl.* **34** 717
- [11] Frontline System 2016 <http://www.solver.com/genetic-evolutionary-introduction>
- [12] Yusup N et al 2012 *Expert Syst. Appl.* **39** 9909
- [13] Holland J H 1975 *Adaptation in Natural and Artificial Systems* (Cambridge, MA: MIT Press)
- [14] Lin W L et al 2009 *IEEE J. L. T.* **27** 4136