Optimized feed-forward neural-network algorithm trained for cyclotron-cavity modeling

Masoumeh Mohamadian[1] Hossein Afarideh[1;1)] Mitra Ghergherehchi[2;2)]

[1] Energy Engineering and Physics Department, Amirkarbir University of Technology, Tehran, 15857-4413 Iran
[2] College of Information and Communication Engineering, Sungkyunkwan University, Suwon 440-746, Korea

Abstract: The cyclotron cavity presented in this paper is modeled by a feed-forward neural network trained by the authors' optimized back-propagation (BP) algorithm. The training samples were obtained from simulation results that are for a number of defined situations and parameters and were achieved parametrically using MWS CST software; furthermore, the conventional BP algorithm with different hidden-neuron numbers, structures, and other optimal parameters such as learning rate that are applied for our purpose was also used here. The present study shows that an optimized FFN can be used to estimate the cyclotron-model parameters with an acceptable error function. A neural network trained by an optimized algorithm therefore shows a proper approximation and an acceptable ability regarding the modeling of the proposed structure. The cyclotron-cavity parameter-modeling results demonstrate that an FNN that is trained by the optimized algorithm could be a suitable method for the estimation of the design parameters in this case.

Keywords: cyclotron cavity, CST, modeling, neural network

PACS: 29.20.dg, 07.05.Tp, 07.05.Mh **DOI:** 10.1088/1674-1137/41/1/017003

1 Introduction

1.1 The 10 MeV cyclotron

The IranCYC10 is an azimuthally varying field (AVF) 10 MeV cyclotron with straight sectors. It accelerates protons to different energies, and the applications are isotope production, nuclear reactions, and medical nuclear-spectroscopy studies. The 10 MeV cyclotron comprises two parallel accelerating electrodes called *dees*, with an oscillating electric field generated by the radio-frequency generator between them [1]; here, the charged particles are produced by an ion source located in the central region. The magnetic field of the main magnet of the cyclotron causes the particles to move in an approximately circular orbit.

The radius of the orbit is a function of the particle velocity; therefore, the radius increases with the energy, and the particles pursue a spiral path from the ion source to the edge of the magnet, where they are pulled out from the cyclotron by an electrostatic deflector. The extracted beam is guided by a beam-transportation system to the experimental stations. The neural-network system will help a designer to determine the values of a set of the parameters that are required to accelerate a particle to a certain frequency, and to obtain the desired scattering parameter of S11 for the reflected power. These parameters are related to the following components of the cyclotron:

- Tuner-disk diameter and gap: There are two coupling parts on both sides of the cavity, one of which is the tuner that is located at the end of the cavity and is connected to the dee. The tuner has two circular disks for the capacitance coupling.
- Coupler-disk diameter and gap (as described above).
- Stem dimensions

Figure 1 illustrates these parameters through a cross-sectional view of the cyclotron cavity. As one of the most widely used neural-network paradigms, the feed-forward neural network (FNN) is composed of a vast number of parallel simple processing units (neurons) that form a layered architecture. In a properly trained FNN, the complex relationships between the groups of related parameters are stored in the connected weights and the threshold values of each processing neuron. The FNN belongs to a large category of supervised neural networks that can be further categorized based on different training algorithms, e.g., conventional back-propagation (BP) algorithm, generalized regression algorithm, genetic algorithm (GA), etc. Among these algorithms, the

Received 29 February 2016, Revised 12 May 2016

1) E-mail: hafarideh@aut.ac.ir

2) E-mail: mitragh@skku.edu

BP algorithm is the most extensively studied, with major successes in varieties of process-modelling and real-time-control regimes [2, 3]; however, despite these successes, the BP algorithm is hampered by a number of inherent disadvantages when it is applied as a crude, gradient-descent optimization algorithm, including non-convergence, a slow convergent rate, and over-fitting [4].

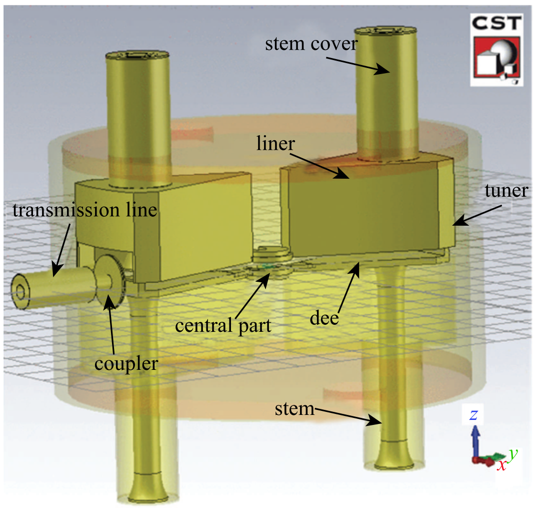

Fig. 1. (color online) Description of the different parts of the RF cavity of the cyclotron.

In this paper, a numerical FNN model that has been trained through an optimization of the neural-network parameters is proposed, and it is developed to simulate a cyclotron cavity with a high accuracy. Although a variety of training-algorithm types have been used in other fields to decrease the training time, they have not been used for the modelling of cyclotron-cavity designs. However, in Ref. [3] a neural network system was developed to determine the parameters of new operating modes for a cyclotron, to guide the operator in using this new operation mode, such as the currents to be applied to the coils of the magnetic lenses, correctors and the concentric coils. They did not perform any parameter estimation to design a specific part of the cyclotron, such as the RF cavity, as we do.

Also in this paper, the optimal BP algorithm is compared with the basic algorithm in terms of their approximation and generalization abilities, which are important regarding the application of the FNN for simulations. In fact, an attempt was made in this study to train a network to learn the relations between the cyclotron-cavity parameters with respect to the resonance frequency and the reflected power. For this purpose, the output from a CST simulation was considered for some of the parameter changes in limited situations, followed by the use of artificial neural-network training inputs with known outputs. An artificial neural network can therefore learn the relation between a parameter and the desired outputs, and a general structure that can be applied for any other situations and parameters was also performed. After this stage, the proposed designed network can obtain the resonance frequency and the reflected power of each structure for different parameters without time-consuming simulations.

The first part of this paper presents the background of the FNN structure and its working principles, the proposed optimal BP algorithm, and the conventional basic BP algorithm; the advantages of the optimal BP algorithm in relation to the conventional BP algorithm are also discussed, including approximation and generalization ability. In the second part, the performance of the proposed algorithm is extensively tested with the use of the simulated data of this study, followed by a comparison of the results regarding the convergence property and the average error of the test data that are for the attainment of the best FNN structure. In the last part, the modeled cavity is compared with the test results of the proposed algorithm, and a very sound fit is identified between them. As a result, the proposed optimal BP algorithm attains to our expected outcomes and we see the ability of the neural network with optimized parameters to estimate the cyclotron-cavity design parameters accurately and quickly.

2 Artificial neural network and training algorithm

2.1 Network architecture

Figure 2 schematically shows the typical two-hidden-layer FNN structure used in this paper, composed of one input layer with four input nodes, two hidden layers with optimized neurons, and one output layer with two neurons. For the introduction of the proposed model, it is proper to provide an overview of the perceptron type of neural network and its associated mathematical relations. In general, multi-layer perceptrons (MLPs) with two or more hidden layers are mostly used; for example, the following expression is for an MLP with two hidden layers:

$$\hat{y}(x)=\varphi^2\left(\sum_{i=1}^{N}w_i^2\varphi^1\left(\sum_{j=1}^{N}w_j^1x+b^1\right)+b^2\right), \qquad (1)$$

where w denotes the vector of the weights, x is the vector of the inputs, b is the bias, and φ is the activation function [5].

2.2 BP algorithm and its optimization

The principle of using the FNN for the modeling of the cyclotron cavity is the utilization of its function-approximation ability to describe the relation between the parameters in the cavity. To achieve this, the neural network is trained with the use of a number of samples (shown as input–output pairs) that are taken from a CST simulation such that the input–output relation of the network can properly fit the given samples. The conventional training method uses the BP algorithm or the gradient descent to minimize the mean square error between the given outputs in the samples and the neural-network outputs. For any set of input data and weights, a magnitude of error that is measured by an error function is derived [6]. The Delta Rule utilizes the error function of what is known as "gradient-descent learning," which consists of the amendment of the weights along the most direct path to minimize the error that is incurred. The change that is applied to a given weight is the negative derivative of the error that is proportional to that weight [7].

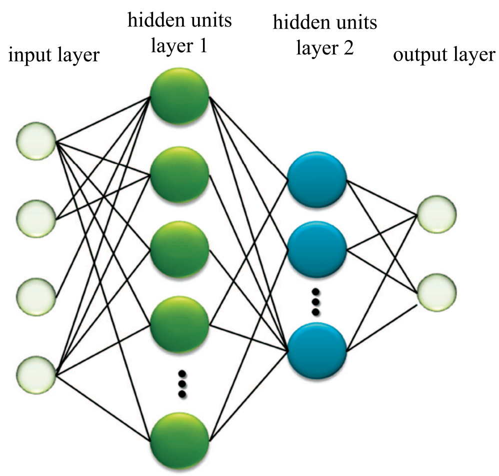

Fig. 2. (color online) Network structure of a two-hidden-layer feed-forward neural network (FNN).

The error function is often given as the sum of the squares of the differences between all of the targeted desired outputs and the response of the neural-network-output layer. The error function is usually taken as the following equation:

$$\min \quad Err(..) = \frac{1}{M}\sum_{(X,y)\in T}(\hat{y}(X)-y)^2, \tag{2}$$

where $\hat{y}$ is the desired output in each sample, y is the corresponding FNN output, and M is the number of the input–output in a designated set of training data according to $T=\{(X_i,Y_i)\}_{i=1}^{M}$, where each X_i the input matrix and Y_i is the desired output response of the ith input. It becomes evident that the input–output relation in an FNN is fully determined by all of the weights and biases; for example, in this modeling, the desired output of the authors is the resonance frequency of the cavity and its scattering parameters.

The intrinsic idea of back-propagation is the incorporation of a non-linear multi-layer perceptron system that is capable of modifying the weights with the error function of the Delta Rule [7]. The BP algorithm uses a simple iteration formula to sequentially update W, starting from a random vector W_0, until a stable solution is obtained. The iteration formula is as follows:

$$W_i(t+1) = W_i(t) - \eta\frac{\mathrm{d}Err}{\mathrm{d}W_i}, \tag{3}$$

where η is the learning rate (LR) and $\frac{\mathrm{d}Err}{\mathrm{d}W_i}$ is the gradient vector of the error function with respect to the weighting vector W. The error function, as well as the gradient vector, can be explicitly expressed in terms of the weighting vector W. Each weight multiplies to the previous layer-output delta and is employed as an input of the activation function of the next layer; also, it obtains the gradient of the weight and then a ratio of the gradient of the weight is subtracted. This ratio η affects the speed and quality of the learning; that is, this ratio is the "learning rate." The greater the η value, the faster the neuron trains; the lower the η value, the higher the accuracy of the training. The sign of the gradient of a weight indicates where the error is increasing [8]. All of the components in the gradient vector can be analytically expressed as follows [9]:

$$\begin{aligned}\frac{\delta Err(..)}{\delta b_k^1} &= \frac{1}{M}\sum_{(X,y)\in T}(\hat{y}(X)-y)\\ &\quad\cdot f^{2'}(W_{:,k}\times a_k^1+b^2)\cdot W_{:,k}\cdot f^{1'}(W_{:,k}\times a^1+b_k^1)\\ \frac{\delta Err(..)}{\delta W_{k,l}} &= \frac{1}{M}\sum_{(X,y)\in T}(\hat{y}(X)-y)\\ &\quad\cdot f^{1'}(W_{k,l}\times a^1+b_k^2)\cdot a_1,\end{aligned} \tag{4}$$

where $\boldsymbol{y}$ is the output vector, $\boldsymbol{X}$ is the input matrix, $\boldsymbol{W}$ is the weight matrix of the hidden and output layer, b^i is the bias of the layers, f^i is the activation function of each layer, a^1 is the output hidden layer, and k is the number of hidden-layer neurons [9].

Additionally, the technique that can help the network out of local minima and improve the convergence property is the use of a momentum term. The speed of the convergence of a network can be improved by increasing the η value; but, increasing η will usually lead to an increase of the network instability, whereby the weight values oscillate erratically as they converge on a solution. Instead of changing η, most standard back-propagation algorithms use a momentum term to increase the speed of the convergence while avoiding instability. The momentum term is added to Eq.(3) so that the change in weight occurs in the previous weight. By adding the fractions of previous weight changes, the weight changes can be maintained on a faster and more-even path [10]. With the momentum m, Eq.(5) shows what the weight update at a given time t becomes, as follows:

$$W_i(t+1) = W_i(t) - \eta \frac{\mathrm{d}Err}{\mathrm{d}W_i} + m \times \Delta W_i. \qquad (5)$$

According to the above discussion, an optimal BP algorithm was developed using proper network parameters such as optimized learning rate, appropriate momentum value, suitable learning function and so on. These were implemented via minimizing the error function for such inputs to achieve desired performance.

2.3 Dataset

The input data from this network are some of the cavity parameters. Table 1 illustrates some of them, including stem, dee, and the coupling-disk parameters.

Table 1. Parameters and adjustable ranges in the cyclotron-cavity modeling.

parameter	adjust range
stem diameter	155 to 215
tuner diameter	30 to 130
tuner gap	0.6 to 3
coupler diameter	30 to 130
coupler gap	0.6 to 3
dee & liner distance	5 to 20

These are the main parts of the cyclotron structure. Figure 3 shows the cavity structure beside the coupling and the tuning capacitive disks on the left, and the electric-field distribution on the right.

The desired network output is the resonance frequency and the reflected power, or scattering parameters such as S11.

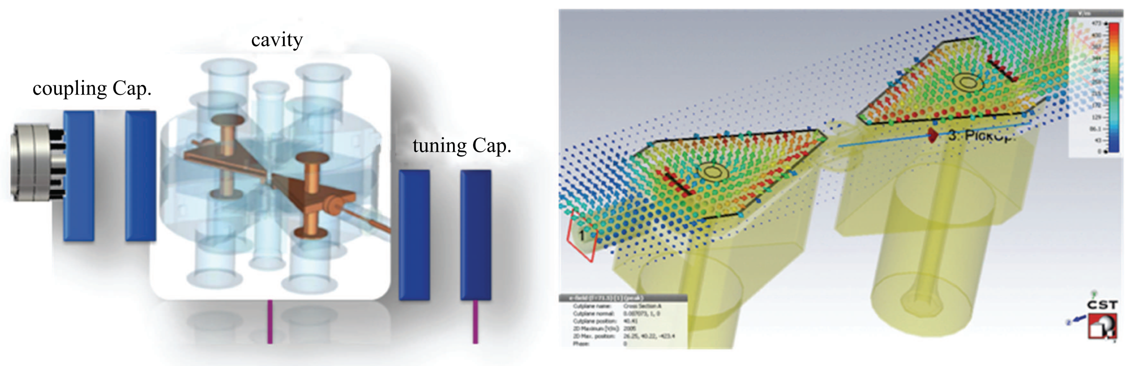

Fig. 3. (color online) Structure of the RF cavity of the cyclotron with the capacitive coupling and tuning disks on the left, and the electric-field distribution on the right.

2.4 Dataset preparation

The input is a vector of the different possible values of the cavity parameters called X_t, which is defined by $X_t \equiv$ cavity parameters. For the first step of this study, 100 samples are considered. The target value for this vector is defined by $y_t \equiv$ resonant frequency and the reflected power at this frequency. The target is the neural-network output, and the aim is to reach it. The set of training data is defined by $S_{\text{training}} = \{(X_t, y_t)\}_{t=1}^{4\times 101}$ that are the input/output pairs of 101 samples that the target value provided from a MWS CST Studio software simulation.

3 CST simulation

In this work, some of the parameters of the cavity are used for the cyclotron-cavity modeling, including the dimensions of the stem, dees, and couplings. The computer model of our 10 MeV cyclotron was developed using the general-purpose simulation software CST STUDIO SUITE [11, 12]. The ranges of the parameters that were used are listed in Table 1. Magnetic-field modeling and beam dynamics were used to determine the orbital frequency of the ions equal to 17.75 MHz. As the RF cavities will be operated in the 4$^{\text{th}}$ harmonic mode, the resonance frequency must be 71 MHz; here, it is arranged so that normal conducting RF cavities [4, 13, 14] can be used for ion-beam acceleration in the IranCYC10 cyclotron. The geometric model of the cavity that is housed inside the valley of the magnetic system of the IranCYC10 cyclotron was developed in CST. A number of models with differences regarding the above-mentioned parameters were studied for this paper, so that the results could be used to train the artificial neural network to learn the relations between these parameters in terms of the obtained results.

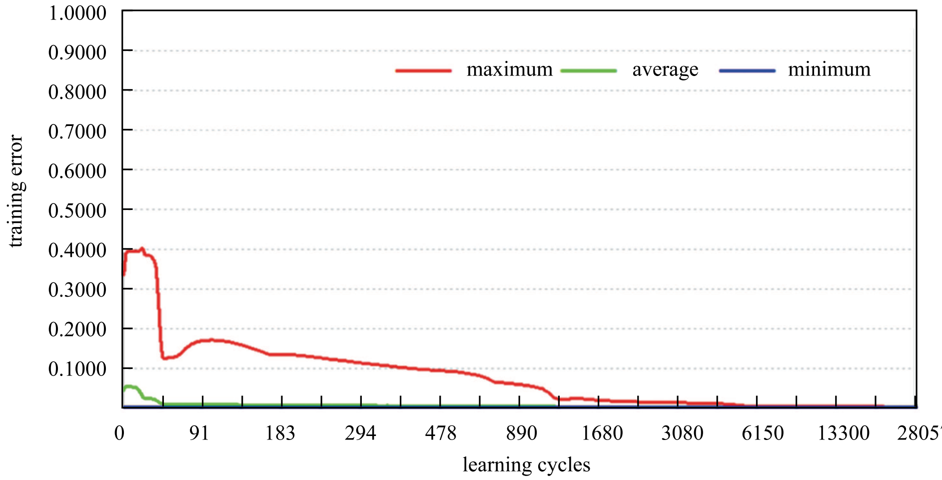

Fig. 4. (color online) Training error versus learning cycles (maximum, minimum, and average error).

4 Results and discussion

The best method for determining a neural-network structure, including the number of layers and each neuronal layer and the importance of the connections, comprises a survey of the hidden-layer growth and an optimization of the neuron quantity and their connections. An attempt was therefore made to change network parameters such as the hidden layers and the number of neurons in each step, via a consideration of the error function and the convergence of the network, to achieve the optimal design.

As illustrated in Fig. 4, the training error decreased through an increasing of the iteration numbers, and the training stopped with an average training error of 1×10^{-4}, while the maximum error is 1.079×10^{-3}. The network information is expressed in Table 2; these are for the network structure with four neurons in the input layer, two neurons in the output layer, 12 neurons for the 1st hidden layer, and 10 neurons for the 2nd hidden layer.

Table 2. Maximum, minimum, and average error of trained neural network.

parameter	value
learning rate	0.6
momentum	0.8
number of training data	101
min.training error	2.7×10^{-7}
max.training error	1.079×10^{-3}
average training error	$9.97e\times10^{-5}$

One of the network parameters that is usually helpful for the annealing of the learning rate over time is the above-mentioned learning rate, which is a training parameter that controls the sizes of the weight and the bias changes in the learning of the training algorithm. Also, similar to the annealing schedules for the learning rates, an optimization can sometimes benefit from those momentum schedules where the momentum increases in the later stages of the learning. As described in subsection 2.2, the momentum parameter is used to prevent the system from converging to a local minimum or the saddle point. The adaptive learning rate, for which the final learning rate and the momentum value are 0.6 and 0.8, respectively, is used in this structure.

The linear normalized regression between the network outputs and the corresponding targets is shown in Fig. 5. The red and green dots indicate the predicted resonance frequency (70.623 MHz to 73.57 MHz) and the reflected power (−53.78 dB to −2.606 dB) at this frequency, respectively. Since the values are normalized, this curve just shows a sound conformity between the predicted and true outputs.

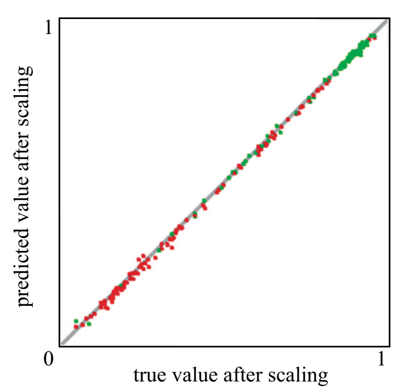

Fig. 5. (color online) Linear normalized regression between the network outputs and the corresponding targets. Red and green dots indicate the predicted resonance frequency (70.623 MHz to 73.57 MHz) and the reflected power (−53.78 dB to −2.606 dB) at this frequency, respectively.

Also, a general goal of statistical modelling is the identification of the relative importance of the illustrative variables with respect to their relation to one or more response variables [15]. This is an indication of the parameters that can help with the reaching of a desired output that is quicker and more accurate.

The aim of the relative-importance analysis is the division of the explained variance among multiple predictors to attain a better understanding of the role that is played by each predictor. The relative importance of the predictor or the input variables is the contribution of each of the variables for the prediction of the dependent variable [16, 17].

The relative importance of the input data is considered and expressed in Fig. 6. According to this analysis, the tuner-gap distance followed by the tuner diameter are the most important regarding the network training. The relative sensitivity to the cavity parameters, which is also illustrated by the red color in Fig. 6, is also considered. The general aim of the sensitivity analysis is similar to that of an evaluation of the relative importance of the illustrative variables, with the exception of a few important distinctions. For both analyses, the relationships between the illustrative and response variables, as demonstrated by the model, are of interest, and it is expected that the neural network provides the explanation for this relation. Using Garson's algorithm, an idea of the magnitude and the sign of the relationship between the variables that are relative to each other is obtained; conversely, the sensitivity analysis allows us to obtain information about the relation between the variables instead of a definite explanation [18]. The effect of the tuner-gap distance regarding the results could therefore be greater than that of the other parameter, whereby it could help with an appropriate tuning of the cavity-resonance frequency.

The sensitivity coefficients demonstrate the change of the systemic outputs that are due to the variations of the parameters that affect the system. A large sensitivity to a parameter proposes that the systemic performance can severely change with a slight variation of the parameter, whereas a small sensitivity proposes a small performance change. A few methods can be used for the consideration of the sensitivity of the ANN model [19].

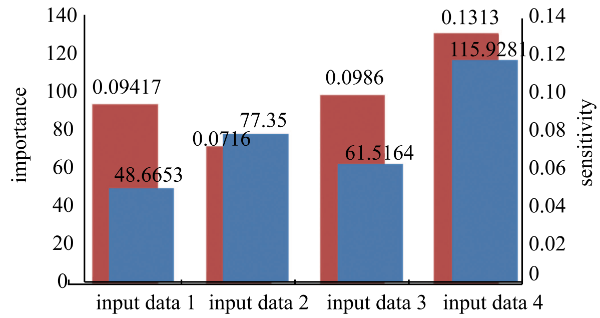

Fig. 6. (color online) Relative importance and sensitivity of input data vs. coupler, tuner diameters, and gap distances.

The simulated results for the four example structures have been reported in terms of a comparison with the proposed-neural-network outputs, and Table 3 shows these examination results. As can be seen, the trained neural network was used for an estimation of the desired outputs of this study with respect to the parameters of the different systems and with acceptable errors. The considered outputs of this research are parameters that are very different from each other, but the proposed neural network could estimate both of them with almost the same error function.

The modeling of such structures that comprise many effective parameters, where a nonlinear relationship exists between these parameters and the desired outputs, is not suitable via conventional methods, and the results are not satisfactory. But a machine-learning methods, like that which is used in this study, can be used to very effectively estimate non-linear relations, and this can also be seen in the results of the present research study.

Table 3. Instances for the testing of the trained proposed neural network in comparison with the CST-simulation results.

pattern No.	neural network inputs				CST simulation output		neural network outputs	
	tuner Dia.	tuner gap	coupler Dia.	coupler gap	Freq.	S11	Freq.	S11
1	55	1.7	115	1.9	71.07	−53.78	71.0381	−52.0083
2	55	1.7	120	1.9	70.96	−45.9	70.9594	−45.9856
3	50	1.6	50	1.9	72.429	−9.3436	72.3768	−9.4174
4	50	2	95	1.95	72.508	−12.365	72.4876	−12.2516

5 Conclusion

Many factors affect the design of an RF cavity, thereby changing the resonant frequency and the scattering parameters. A selection of the appropriate values for each parameter, especially those that are variable in practice after the manufacturing process, such as the tuner, is therefore very important. The formulation of an algorithm for the estimation of these values will result in a high efficiency that will facilitate a faster achievement of the desired results.

The results here show that a numerical modelling of the cyclotron cavity for which the optimized FNN algorithm is used can produce proper results with acceptable errors. The overall average and the maximum error are within 9.97×10^{-5} and 1.079×10^{-3}, respectively, and these are sufficiently accurate for the provision of reliable simulation results that provide an understanding of the relation of the parameters of the cavity to its resonant frequency. The simulated results of some of the example structures are also consistent with the CST-simulation results; therefore, the FNN that is trained by the optimized algorithm can be used to develop a parametric cyclotron-cavity design. While this research does not completely cover the effects of all of the cavity parameters, it shows the capability of an FNN estimation regarding the cavity design. This structure can be used for most AVF cyclotron cavities.

References

1 M. S. Livingston, J. P. Blewett, *Particle Accelerators* (New York: McGraw-Hill, 1962)

2 J. Xia, Rusli, A. S. Kumta, IEEE T Plasma SCI, **38**(2): 142–148 (2010)

3 M. Abd El- Kawy, M. Shaker Ismail, M. Abdel-Bary, M. M. Ouda, Arab J. Nucl. Science and App., **45**(3): (2012)

4 R. Fletcher, *Practical Methods of Optimization* (UK: Chichester Wiley, 1987)

5 L. Fausett, *Fundamentals of Neural Networks* (New York: Prentice Hall 1994)

6 J. Y. F. Yam, T. W. S. Chow, IEEE T Neural Networks, **8**: 806–811, (1997)

7 J. L. McClelland, D. E. Rumelhart, *Explorations in Parallel Distributed Processing - a Handbook of Models, Programs, and Exercises* (MIT Press, Cambridge, 1988), p. 126–130

8 P. J. Werbos, *The Roots of Backpropagation From Ordered Derivatives to Neural Networks and Political Forecasting* (New York, NY: John Wiley & Sons, 1994)

9 M. Mohamadian, H. Afarideh, F. Babapour, IAENG Int. J of Com. Science, **42**(3): 265–274 (2015)

10 S. I. Gallant, *Neural Network Learning and Expert Systems* (MIT Press, Cambridge, 1993)

11 Y. Jongen et al, Proc of Cyclotrons, China (2010)

12 CST Studio Suite 2014 (CST Microwave Studio)

13 M. Mohamadian, M. Salehi, H. Afarideh, M. Ghergherechi, J. Chai, Proc. 12th International Computational Accelerator Phys. Conf., China, (2015)

14 A. W. Chao, K. H. Mess, M. Tigner, F. Zimmermann, *Handbook of Accelerator Physics and Engineering* (2nd Edition, World Scientific Publishing Co., 2013)

15 S. L. Ozesmi, U. Ozesmi, Ecological Modelling, **116**: 15–31 (1999)

16 O. M. Ibrahim, J Applied Sciences Research, **9**(11): 5692–5700 (2013)

17 S. Tonidandel, J. M. LeBreton, J. B. Psychol, J Business and Psychology, **26**(1): 1–9 (2011)

18 G. D. Garson, Artificial Intelligence Expert, **6**: 46–51 (1991)

19 R. Gunawan, Y. Cao, L. Petzold, Biophys J., **88**(4): 2530–2540 (2005)