Hybrid modelling routine for metal-oxide TFTs based on particle swarm optimisation and artificial neural network

You Peng, Wanling Deng✉, Weijing Wu, Zhi Luo and Junkai Huang

An effective and robust hybrid algorithm consisting of particle swarm optimisation (PSO) and limited memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) method based on artificial neural network (ANN) is proposed for modelling flexible metal-oxide thin-film transistors (TFTs). The L-BFGS method as an optimiser is exploited to update the parameters of ANN and speed up the training process. A mutation strategy for PSO is derived to enhance the searching ability further. With the great global searching ability, PSO is implemented to find a hopeful initial position in solution space for the next ANN model. The simulation result shows a high accuracy not only in *I–V* curve fitting but also in small-signal parameter (g_m, g_d, etc.) predictions, which have not been exposed in the training process. The measured DC characteristics of In–Zn–O TFTs are used to verify the proposed ANN model, which has the benefits of rapid fitting from the L-BFGS algorithm and universal searching ability from PSO.

Introduction: Flexible metal-oxide thin-film transistors (TFTs) have unique properties of high mobility, controllable conductivity and good performance in low temperature so that they have been commonly used in LED electronic display screens and so on [1, 2]. As a large quantity of TFT materials have been investigated, traditional physics-based modelling approaches cannot meet the demands of fast and convenient modelling methods. Though physical models are accurate, they are not widely used in circuit design because it takes too much time to be developed. It is a huge challenge to build physics-based models when new technology keeps emerging under this circumstance [3]. Without going into the detailed device physics, artificial neural networks (ANNs) as data-orient modelling methods can be very good alternatives for modelling semiconductor devices.

Over the past decade, ANNs have been successfully used in modelling semiconductor devices [4]. The multi-layer perceptron (MLP) is one of multi-layer feedforward neural networks with one hidden layer, which is a universal approximator [5]. Because of its characteristics of simplicity and high efficiency, MLP can be considered as a potential method to build compact models for semiconductor devices. In this work, limited memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) method [6] is applied for MLP to update the weights and biases as an optimisation; this optimisation is numerically efficient and requires low memory. However, this algorithm is sensitive to the initial location because all the initial weights and biases are randomised, and therefore, this method may converge to a local best solution.

In this Letter, a particle swarm optimisation (PSO) is employed to initialise the weights and biases of MLP [7, 8]. The PSO has an excellent global searching ability so that it is suitable to search a hopeful initial position for the following L-BFGS method. In addition, we adopt a mutation strategy for PSO to enhance the global searching ability much further. The particle can be replaced by a new one to search the other area in solution space. Finally, for IZO-TFTs, the results of this hybrid modelling method are shown to demonstrate its flexibility and accuracy.

MLP model: Before feeding the data to MLP model, data preprocessing is necessary. The ANN algorithm finds trends by comparing the features in training data, but in this process, the feature with a bigger scale may dominate the other features completely. In this Letter, z-score normalisation is achieved to make the training data centring and scaling, i.e.

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

where x is the sample; μ and σ are the mean and standard deviations of the sample, respectively. After z-score normalisation, we can apply the logarithm of drain current (*Ids*) as the target of MLP model to avoid the negative numbers of *Ids*, so the inputs of the MLP model are standardised *W/L*, V_{gs} and V_{ds}.

The MLP feedforward neural network consists of an input layer, several hidden layers, and an output layer. The MLP model describes arbitrary complex function relationships as matrix calculus. The expression of an MLP has the following form:

$$a_k^l = g^l(a_k^{l-1} \cdot \omega_{jk}^l + b_k^l) \tag{2}$$

where a_k^l is the output of the kth synaptic at the lth layer; b_k^l is the kth bias at the lth layer; ω_{jk}^l is the weight from the kth neuron in the $(l-1)$th layer to the jth neuron in the lth layer; and g^l is the activation function of MLP that introduces non-linear properties to our network. Since the compact models of transistor devices should have smooth first- and higher-order derivatives in all the operation regions, the hyperbolic tangent function is a better choice for non-linear curve fittings.

The loss function of this MLP is the mean square error between training data (I_{ds}) and prediction value ($\widehat{I_{ds}}$) of the MLP model, yielding

$$loss = \frac{1}{n}\sum_{i=1}^{n}(\log(I_{ds}) - \log(\widehat{I_{ds}}))^2. \tag{3}$$

Then, the L-BFGS algorithm is exploited as an optimiser to update the weights and biases of each layer iteratively to minimise the loss. Once done with training, MLP can be accurately generalised when it is presented with unseen data. Nevertheless, compact models require high accurate predictions in both *I–V* curves and small-signal parameters (i.e. g_m and g_d), which have a great influence on small-signal behaviours like a gain. With f as output and In_j as the jth input, the g_m and g_d can be obtained by the chain rule:

$$\frac{\partial f}{\partial In_j} = \omega_{kp}^3 \cdot [g_2' \cdot (\omega_{jk}^2 \cdot g_1' \cdot \omega_{ij}^1)] \tag{4}$$

where g_1 and g_2 are the first and second hidden layer activation functions, respectively; ω_{ij}^1, ω_{jk}^2 and ω_{kp}^3 are the node weights of the first hidden, second hidden and output layers, respectively; and the node's index is shown by the subscripts in (4).

Particle swarm optimisation: PSO is a meta-heuristic algorithm mimicking birds' cooperative foraging [9]. The PSO has the advantage of rapid convergence and the ability to search optimal solutions within global space. The same as MLP model, we define the mean-square-error function between model predictive value and experimental data as the fitness function of PSO. This fitness function influences the forward movement direction of each particle to minimise the errors. In addition, each particle updates its position in solution space by the following two positions; i.e. one is the own optimum p_{best} which is searched by itself, and the other is swarm's optimum g_{best} which is the global optimum in the whole population. In this Letter, the expressions for updating velocities and positions are, respectively, written by

$$V_i = wV_i + r_{i1} \cdot c_1(p_{\text{best}} - x_i) + r_{i2} \cdot c_2(g_{\text{best}} - x_i) \tag{5}$$

$$x_i = x_i + V_i \tag{6}$$

where V_i and x_i represent the velocity and position of the ith particle, respectively; w is an inertia factor that represents the degree of maintaining the former motion state. It helps particles to search a larger area in the former direction of updating. Moreover, r_{i1} and r_{i2} are two random numbers with values uniformly distributed in [0,1]; c_1 and c_2 are the acceleration factors which control the acceleration of particle.

In order to enhance the global search ability, a strategy of mutation is adopted to particle swarm. Every particle in each iteration is assigned to a random number φ_t (the random number of ith particles) with a uniform distribution between 0 and 1, and the threshold of φ_t is defined as 0.95 ($\varphi_t > 0.95$) which means every particle has 5% of mutation probability to become a new particle in solution space. After every 100 iterations, the Euclidean distance E_i between the ith particle and current g_{best} is calculated. Then we compared it to the mean of whole swarm Euclidean distance $\bar{E}$. The particle with E_i less than $\bar{E}$ has a 50% chance to mutate. On the one hand, this strategy makes half of the particles closer to g_{best} and mutate into new particles to search for other solutions. On the other hand, this method makes the rest particles continue to move towards g_{best}, so that they can converge to a local minimum.

Hybrid algorithm: As L-BFGS algorithm is sensitive to initial location, PSO is realised to search a hopeful initial position for L-BFGS method.

Since these two algorithms have the same cost function of mean square error, they are combined to get a higher accuracy. As a result, PSO is used to optimise the weights and biases of MLP model, and the best optimisation values are taken as the initial parameters. Fig. 1 plots the flowchart of this hybrid algorithm.

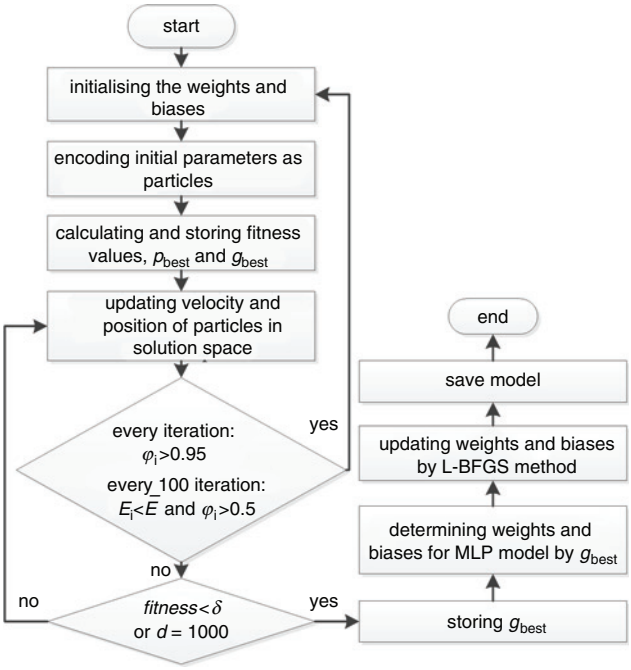

Fig. 1 *Flowchart of the hybrid algorithm*

Model verification and discussions: IZO-TFTs as a kind of metal-oxide TFTs were fabricated to obtain experimental data, as shown in Fig. 2. The devices were made with various channel widths of 5, 10, 15, 20, 30 and 40 µm, and with channel length fixed to 5 µm. The fabrication of the devices can be found in [10].

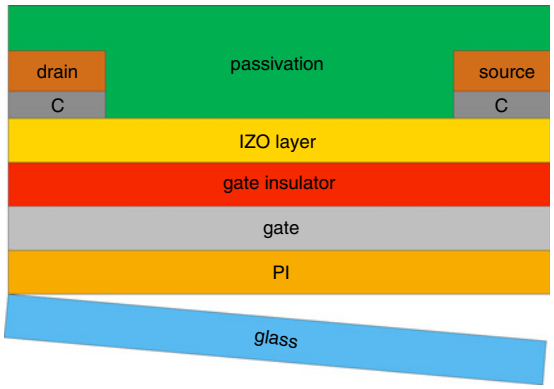

Fig. 2 *Schematic diagram of a flexible IZO-TFT*

We determine the number of nodes by developing ANNs with different amounts of synaptic and choosing the model with the best performance. The best number of neurons in the structure with a single hidden layer is 15, since this model achieves the highest accuracy of 95.8%. Because of the limitation of the single hidden layer structure, adding more hidden nodes or changing the activation function cannot improve the accuracy. To solve this problem, we build MLP with two hidden layers. The reason is the first hidden layer mainly works in the subthreshold region and the other affects the above-threshold region [11]. The structure of this MLP can be denoted as 3-15-20-1, the accuracy of which can be over 97.5% as shown in Fig. 3.

This MLP model with the size of 3-15-20-1 can be regarded as an approximator for semiconductor devices. However, in some cases, the modelling of g_m and g_d cannot fit well with experimental data even though I–V fitting has great accuracy. This problem results from an encounter with local optimum. The L-BFGS optimiser depends too much on the initial weights and biases. Although the model gets an acceptable accuracy, a better position in solution space represents that a higher accuracy can be reached. In this Letter, the PSO is implemented to initialise the weights and biases for MLP and achieve its full potential of non-linear fitting ability.

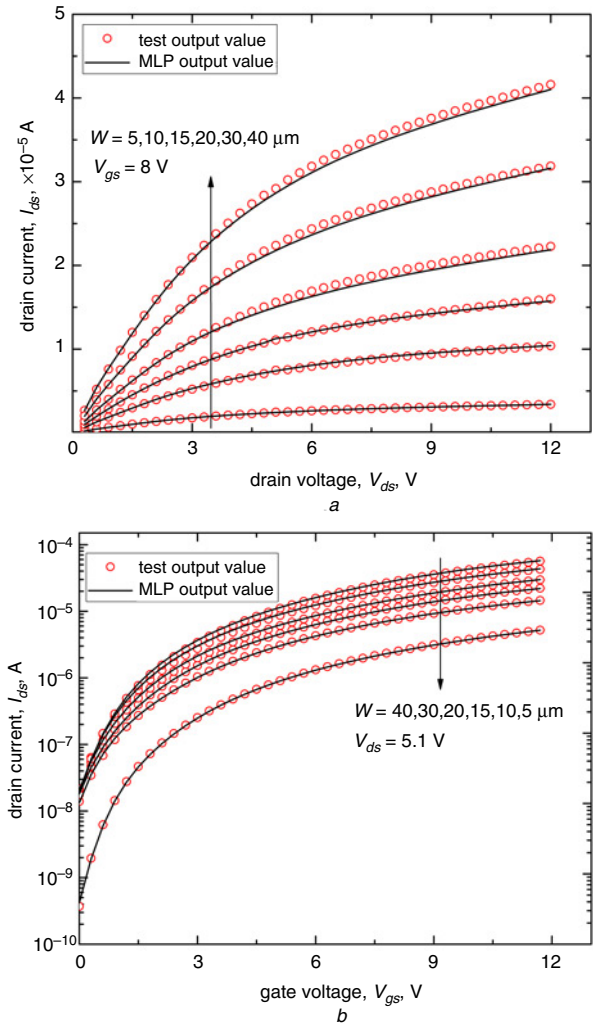

Fig. 3 *Simulation values using 3-15-20-1 MLP model with various channel widths of 5, 10, 15, 20, 30, 40*

a Output characteristics
b Transfer characteristics

The initial swarm contains m particles which are randomly generated. Every particle contains a set of weights and biases for MLP whose dimension is 401. Higher dimension implies larger swarm. However, efficiency and accuracy should be balanced. Too small swarm (m < 50) may miss the best position in space, and too large swarm (m > 100) will waste computing resources and time significantly, which cannot improve the accuracy of this model. Here, we set m as 80 in this work. For the sake of encouraging particles to search a larger area, we set the value of the acceleration factor as 2 and w as 0.8. Besides, we limit the maximum number of iterations d to 1000, and set δ as δ = 0.1, which represents 4×10^{-11} A^2 of the mean square error between PSO output and experimental data. If *fitness* < δ or d = 1000, we terminate the iteration and determine the weights and biases for MLP by g_{best}.

In order to verify the validity of the hybrid model, the same dataset is fed to train the MLP model and PSO-MLP model separately. Then, we compared the simulation results of IZO-TFT I–V characteristics and small-signal parameters with experimental data. The modelling method with PSO gets higher accuracy. As shown in Fig. 4. The excellent agreement between predictive values and measured ones proves the accuracy of our hybrid model. Besides it, the small-signal curve fitting results are good in a wide range of operation regions. The simulations show that this hybrid algorithm combines the advantages of both global searching abilities from PSO and local convergence ability from the L-BFGS algorithm.

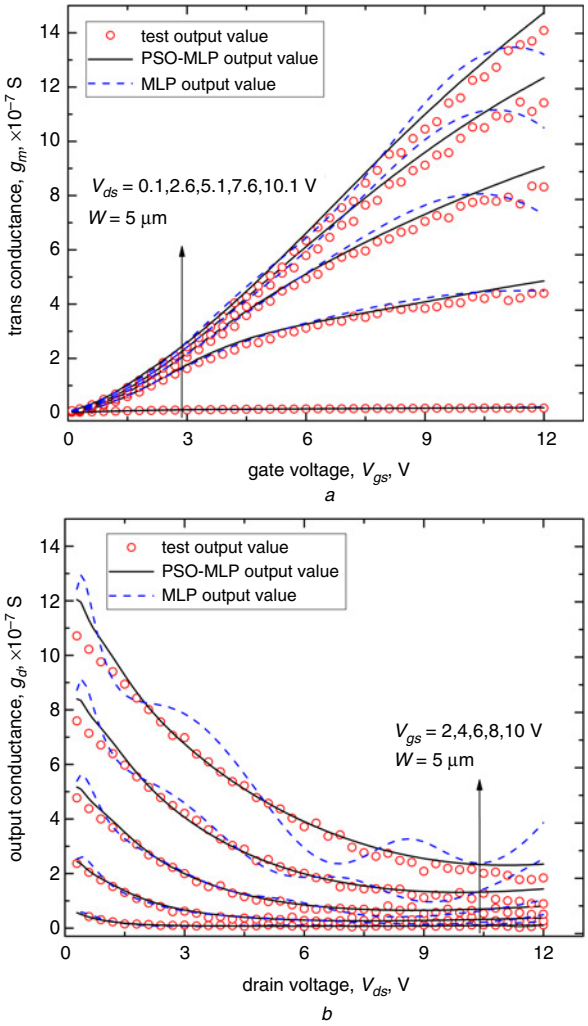

Fig. 4 *Comparisons of experimental data versus simulation using MLP model and PSO-MLP method*

a Transconductance (g_m)
b Output conductance (g_d)

Conclusion: A PSO-MLP modelling method for metal-oxide TFTs is implemented. A strategy of PSO mutation is proposed and proved for model efficiency and accuracy enhancement. This modelling method involving particle swarm optimisation and the L-BFGS method has the advantages of these two algorithms. With its accuracy and efficiency, this method is expected to play a positive role in modelling and IC designs for transistors.

Submitted: *11 December 2019* E-first: *13 February 2020*
doi: 10.1049/el.2019.4001
One or more of the Figures in this Letter are available in colour online.

You Peng, Wanling Deng, Zhi Luo and Junkai Huang (*Department of Electronic Engineering, Jinan University, Guangzhou 510630, People's Republic of China*)

✉ E-mail: dwanl@126.com

Weijing Wu (*State Key Laboratory of Luminescent Materials and Devices, South China University of Technology, Guangzhou 510640, People's Republic of China*)

References

1 Zhang, L., Huang, C., Li, G., *et al.*: 'A low-power high-stability flexible scan driver integrated by IZO TFTs', *IEEE Trans. Electron Devices*, 2016, **63**, (4), pp. 1779–1782, doi: 10.1109/TED.2016.2529656
2 Jeong, J.K.: 'The status and perspectives of metal oxide thin-film transistors for active matrix flexible displays', *Semicond. Sci. Technol.*, 2011, **26**, (3): 034008, doi: 10.1088/0268-1242/26/3/034008
3 Bahubalindrun, P., Tavares, V., Barquinha, P., *et al.*: 'Ingazno TFT behavioral model for IC design', *Analog Integer. Circuits Signal Process.*, 2016, **87**, (1), pp. 73–80, doi: 10.1007/s10470-016-0706-4
4 Hayati, M., Rezaei, A., and Seifi, M.: 'CNT-MOSFET modeling based on artificial neural network: application to simulation of nanoscale circuits', *Solid-State Electron.*, 2010, **54**, (1), pp. 52–57, doi: 10.1016/j.sse.2009.09.027
5 Hornik, K., Stinchcombe, M., and White, H.: 'Multilayer feedforward networks are universal approximators', *Neural Netw.*, 1989, **2**, (5), pp. 359–366, doi: 10.1016/0893-6080(89)90020-8
6 Liu, D., and Nocedal, J.: 'On the limited memory BFGS method for large scale optimization', *Math. Program.*, 1989, **45**, (1-3), pp. 503–528, doi: 10.1007/BF01589116
7 Ismail, A., Jeng, D.-S., and Zhang, L.: 'An optimised product-unit neural network with a novel PSO-BP hybrid training algorithm: applications to load-deformation analysis of axially loaded piles', *Eng. Appl. Artif. Intell.*, 2013, **26**, (10), pp. 2305–2314, doi: 10.1016/j.engappai.2013.04.007
8 Hou, C., Yu, X., Cao, Y., *et al.*: 'Prediction of synchronous closing time of permanent magnetic actuat-or for vacuum circuit breaker based on PSO-BP', *IEEE Trans. Dielectr. Electr. Insul.*, 2017, **24**, (6), pp. 3321–3326, doi: 10.1109/TDEI.2017.006475
9 Eberhart, R., and Kennedy, J.: 'A new optimizer using particle swarm theory'. Proc. of the Sixth Int. Symp. on Micro Machine and Human Science, Nagoya, Japan, October 1995, pp. 39–43, doi: 10.1109/MHS.1995.494215
10 Chen, Z., Xu, W., Wu, J., *et al.*: 'A new high-gain operational amplifier using transconductance-enhancement topology integrated with metal oxide TFTs', *IEEE J. Electron Devices Soc.*, 2018, **7**, (1), pp. 111–117, doi: 10.1109/JEDS.2018.2883585
11 Zhang, L., and Chan, M.: 'Artificial neural network design for compact modeling of generic transistors', *J. Comput. Electron.*, 2017, **16**, (3), pp. 825–832, doi: 10.1007/s10825-017-0984-9