a-GIZO TFT neural modeling, circuit simulation and validation

Pydi Ganga Bahubalindruni a, Vítor Grade Tavares a,* Pedro Barquinha b, Cândido Duarte a, Nuno Cardoso a, Pedro Guedes de Oliveira a, Rodrigo Martins b, Elvira Fortunato b

a INESC TEC and Faculty of Engineering, University of Porto, Campus FEUP, Rua Dr. Roberto Frias, 378, 4200-465 Porto, Portugal
b CENIMAT/I3N, Departamento de Ciência dos Materiais, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa and CEMOP-UNINOVA, 2829-516 Caparica, Portugal

Article history:
Received 10 January 2014
Received in revised form 12 August 2014
Accepted 11 November 2014
Available online 29 December 2014

The review of this paper was arranged by Dr. Y. Kuk

Keywords:
a-GIZO TFT modeling
MLP
RBF
LS-SVM
Artificial neural networks
Verilog-A

Abstract
Development time and accuracy are measures that need to be taken into account when devising device models for a new technology. If complex circuits need to be designed immediately, then it is very important to reduce the time taken to realize the model. Solely based on data measurements, artificial neural networks (ANNs) modeling methodologies are capable of capturing small and large signal behavior of the transistor, with good accuracy, thus becoming excellent alternatives to more strenuous modeling approaches, such as physical and semi-empirical. This paper then addresses a static modeling methodology for amorphous Gallium–Indium–Zinc-Oxide – Thin Film Transistor (a-GIZO TFT), with different ANNs, namely: multilayer perceptron (MLP), radial basis functions (RBF) and least squares-support vector machine (LS-SVM). The modeling performance is validated by comparing the model outcome with measured data extracted from a real device. In case of a single transistor modeling and under the same training conditions, all the ANN approaches revealed a very good level of accuracy for large- and small-signal parameters ($g_{m}$ and $g_{d}$), both in linear and saturation regions. However, in comparison to RBF and LS-SVM, the MLP achieves a very acceptable degree of accuracy with lesser complexity. The impact on simulation time is strongly related with model complexity, revealing that MLP is the most suitable approach for circuit simulations among the three ANNs. Accordingly, MLP is then extended for multiple TFTs with different aspect ratios and the network implemented in Verilog-A to be used with electric simulators. Further, a simple circuit (inverter) is simulated from the developed model and then the simulation outcome is validated with the fabricated circuit response.

1. Introduction

Amorphous Gallium–Indium–Zinc Oxide – Thin Film Transistor (a-GIZO TFT) technology is becoming attractive in various electronic applications, namely with active matrix liquid crystal displays (AMLCD) [1] and Radio-frequency identification (RFID) tags [2]. They can be fabricated at low temperature [3–5], which results in low-cost transparent devices. Compared to organic [6] and a-Si:H [7] TFTs, a-GIZO TFT shows better electrical properties, such as higher mobility [3], and passive resistors can be realized, unlike in organic TFT technology [6]. All the advantages associated with a-GIZO TFTs are motivating factors to design circuits for various applications. The demand for ingenious electronic circuits with this technology propels the development of accurate device models for circuit simulations.

Transistor modeling can be broadly categorized into physical/semi-empirical [8–10], table-based [11] and empirical [12]. Physical modeling uses semiconductor physics to develop analytical equations that approximate the device behavior. Even though physical models are accurate and desirable for circuit design and process optimization, they are complex, with high development time and still too incipient for a-GIZO TFTs. This technology is not yet stable, with experiments being performed to attain solid and consistent electrical characteristics, by several means like employing different structures [13] or multiple channels [14] for the device, varying processing parameters or source/drain [15] or dielectric materials [16,17]. Therefore, whenever there is a modification of device structure and/or materials to obtain better performances, the complete device physics need to be studied to develop the physical model. At this stage, physical modeling for a-GIZO TFT is hard and is not suitable especially when circuit design is required on a faster pace. Nevertheless, few articles have already reported physical models for this device [8,9]. Semi-empirical modeling is essentially a physical model where some fitting (empirical) functions and parameters are added to
improve the model accuracy and generality. Consequently, it shares similar drawbacks with physical modeling. On the other hand, table based modeling is simple and fast, but high accuracy requires huge amount of data storage, and need interpolation algorithms to predict the output for the un-stored input. Empirical model approximates the device behaviour from the measured characteristics, irrespective of the device physics i.e. it is a black box modeling approach. Artificial neural networks (ANNs) can be considered as one type of empirical modeling.

From past two decades, ANNs have become popular for semiconductor modeling. Litovski proposed ANN-MLP (multilayer perceptron) for MOSFET modeling [18] that has also been successfully applied to nanoscale MOSFETs [19,20]. ANN learns the function that needs to be estimated from the data provided to the network, during the training phase. They are universal approximators and can be built in very short-time, to a desired accuracy level, independently of the underlined device physics. Similarly, polynomials are also universal approximators, and can also be employed, however in ANNs the error is independent of the input dimensionality [21]. Despite the fact that ANN parameters lack explicit physical meaning, it meets most of the basic model requirements and benchmark tests suggested in literature [22–24], typically:

- The model should be simple, fast and easily implementable.
- The model and its higher order derivatives should be continuous in the complete region of operation to avoid convergence problems.
- Some physical basis adds value to the model.
- In order to evaluate the model, it is mandatory to test small signal parameters [22].

The main focus of this paper is then to develop an accurate, simple, continuous model with fast development time, which can effectively characterize all the physical properties of the device in the complete region of operation. Since ANNs have all the above mentioned properties and meet most of the model requirements and benchmarks, they have been adopted in this work. Different neural modeling methods namely, MLP, RBF and LS-SVM are tested to model the a-GIZO TFT.

In order to validate the model accuracy and generalization capability, small signal parameters (i.e. transconductance ($g_m$) and output conductance ($g_d$)) determined from the model are compared with the measured data. Based on the complexity (which directly impacts the simulation speed) and accuracy tradeoff, MLP modeling method is selected for circuits simulations, and is extended for multiple transistors, with different widths. Then, the MLP ANN is implemented in Verilog-A, the model output for testing data (that has not been seen by the network during training phase) is also compared against the measured characteristics to demonstrate the modeling ability as well as generalization capability. A simple circuit has also been simulated with the developed neural ANN to demonstrate the suitability of neural modeling methods namely, MLP, RBF and LS-SVM. The structure of the RBF and LS-SVM networks is similar, since the last also employs radial basis functions in the kernels. Actually, other kernels do exist for LS-SVM, but they did not show any improvement in relation to radial-basis; consequently, they were not considered in this work. MLP network topology with a single hidden layer is depicted in Fig. 1 and the RBF/LS-SVM in Fig. 2. The input layer consists of sensory units, which connect to the outside environment. The Hidden layer(s) maps the inputs data from input space to a hidden space. Finally, the output layer provides the response to input activation by a linear combination of the hidden-layer outputs.

2. ANN modeling methodologies

ANNs are potential candidates for function estimation problems. They comprise a set of interconnected neurons with data processing ability, which operate in parallel. Each neuron in the network has a set of inputs, synaptic weights and bias. A set of training samples are applied to the network so that it can learn the physical process, subjected to a cost-function minimization. MLPs and RBFs can be classified as empirical risk minimization networks, i.e., they work with training-data error during the learning phase, whereas LS-SVM employs structural risk minimization, since it operates with both training and generalization error during learning. In this sense, LS-SVM promises better generalization results. However, for all the considered cases, by proper training, weights and bias are determined to guarantee a specified performance goal. Structurally, these networks consist of a single input and output layers, one or more hidden layers, in case of an MLP, and a single hidden layer for RBF and LS-SVM.

2.1. Multilayer perceptron

In this work, MLP with a single hidden layer is employed, as this is a universal approximator when enough number of hidden neurons are used [25]. Back-propagation algorithm is used to train the network. Proper number of neurons in the hidden layer should be selected to ensure good training performance without overfitting. If the number of neurons are too high in the hidden layer, then, even though the network guarantees good training performance, its generalization capability will be inadequate because of overfitting. If the number of neurons in the hidden layer is too small, then the network has poor performance. So, the number of neurons should be selected properly, based on a trial and error process.

Mathematically, the MLP network hidden-layer outputs can be expressed by,

$$y_h = \text{Sig}(x \cdot \mathbf{w}_h + \mathbf{b}_h)$$

(1)

and the output of the MLP network given by,

$$y = y_h \cdot \mathbf{w}_o + \mathbf{b}_o$$

(2)

where $\text{Sig}()$ represents the tanh sigmoid function, $x$ represents the input vector, $\mathbf{w}_h$ and $\mathbf{b}_h$ denote the synaptic-weight vector, connecting inputs to the hidden neurons, and the hidden-neurons bias vector, respectively. The output layer weight vector and bias are represented by $\mathbf{w}_o$ and $\mathbf{b}_o$. The weights and biases are determined during the training phase.

2.2. Radial basis function

Basically, RBF [26] performs curve fitting/approximation in a high dimensional space. During the training phase, RBF network finds a surface in a multi-dimensional space that provides the best...
fit to the training data. The hidden layer output can be described as
\( \Phi = G(||x - w h||) \).

RBF network output is given by (2); where Gaussian functions are the kernels chosen for this work. The input weight vector is again \( w h \) (centers of the Gaussian functions), \( x \) is the training data vector i.e. input to the network. The hidden-layer bias is given by \( b h = \frac{w h}{\sigma \sqrt{2\pi}} \), where \( \sigma \) is the spread or standard deviation of the basis function. This \( b h \) value was chosen such that the output of the hidden layer neuron is \( 0.5 \) when the distance between the input and the input weight is \( \leq \sigma \). The centers of the Gaussian functions (which directly gives the number of neurons in the hidden layer) are determined from the training samples. The output weight vector \( (w o) \) and output bias \( (b o) \) are calculated by solving linear constraints formed by the output layer [27].

In the Gaussian function, if \( \sigma \) is too small, then the basis function is highly localized, hence the network cannot guarantee good generalization. If it is too high, neurons respond in the same way to all the activation samples, consequently the network is not able to learn from the training samples. Hence, \( \sigma \) should be greater than the smallest difference between neighboring input training samples and less than the largest difference in the set, so that the basis function is neither localized nor too flat.

2.3. Least squares-support vector machine

Least Square-Support Vector Machine (LS-SVM) is a powerful technique for non-linear regression or functional approximation problems. LS-SVM for function estimation in primal and dual form are expressed in (4) and (5) respectively [21].

\[
\text{Min}_{w,b,e} J_p(w,e) = \frac{1}{2}w^T w + \gamma \sum_{i=1}^{N} e_i^2
\]

such that \( y_i = w^T \Phi(x_i) + b + e_i, \quad i = 1, \ldots, N \)

\[
L (w, b, e; x) = J_p (w, e) - \sum_{i=1}^{N} x_i [w^T \Phi(x_i) + b + e_i - y_i] \]

such that

\[
\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{N} x_i \Phi(x_i)
\]

\[
\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} x_i = 0
\]

\[
\frac{\partial L}{\partial e_i} = 0 \rightarrow \gamma e_i, \quad i = 1, \ldots, N
\]

\[
\frac{\partial L}{\partial y_i} = 0 \rightarrow w^T \Phi(x_i) + b + e_i - y_i = 0, \quad i = 1, \ldots, N
\]

The factor \( \gamma \) is a regularization parameter, which impacts the generalization capability of the network and \( e \) gives the accuracy of the model, \( x_i \) is Lagrange multiplier and \( \Phi \) is nonlinear transformation from input space to feature space. After solving (5), the resulting LS-SVM model for function estimation is expressed in (6)

\[
y(x) = \sum_{i=1}^{N} x_i K(x, x_i) + b
\]

where \( K(x, x_i) = \phi(x)^T \phi(x_i) \). In this work, the kernel \( (K) \) is a radial basis function.

3. Experimental setup and results

First, a single TFT static behavior is modeled with all the mentioned ANN methods. The inputs are gate-to-source voltage \( (V_{GS}) \), drain-to-source voltage \( (V_{DS}) \) and the output corresponds to the drain current \( (I_D) \). For the training data, \( V_{CS} \), \( V_{DS} \) are ranging from 0 to 10 V and 0 to 15 V respectively, in steps of 1 V. MATLAB 2011b is used to get the trained network for MLP and RBF, whereas, a Matlab toolbox [28] is used for LS-SVM.

3.1. Device structure

The transistors utilized in this work were fabricated at CENIMAT, FCT-UNL. The device structure and thickness of the materials are shown in Fig. 3. It corresponds to a bottom-gate staggered structure, where the gate, drain and source material is IZO, and the semiconductor layer is amorphous GIZO with 2:4:2 composition (Ga:In:Zn atomic ratio). The dielectric material is a multi-layered/multi-component structure (SiO₂ – Ta₂O₅ – SiO₂ – SiO₂). This structure is employed to minimize gate leakage currents, while allowing for a large dielectric constant, above 13. Source to gate or drain to gate overlap is 2–5 μm.

These TFTs are fabricated at room temperature by RF magneton sputtering, being the patterns of the composing layers defined by lift-off and dry-etching processes. An SU-8 passivation layer on top of the structure was deposited by spin-coating. Final devices were annealed in air, for one hour, at 150 °C. Electrical characteristics (sub-threshold swing, turn-on voltage, on-off current ratio, mobility and leakage current) and stress behavior of these TFTs, with respect to different processing parameters, are reported in [29–31]. The fabricated transparent chip, which contains the TFTs, is shown in Fig. 4. Measurements \( (I_D) \) as a function of \( V_{GS} \) and \( V_{DS} \) have been taken using a semiconductor parameter analyzer Keithley 4200-SCS, and probe station Cascade Microtech M150 under darkroom conditions.

3.2. ANN modeling results

From a large signal perspective, the modeled drain current may display a good agreement, on average, with the measured data, but the correspondent small signal parameters \( (g_m \text{ and } g_d) \) may show otherwise. It is fundamental to test \( g_m \) and \( g_d \), since they have a direct impact on the small-signal behavior of the device (gain). Furthermore, if overfitting occurs, the variations in the characteristic predicted by the model will be emphasized in small signal parameters due to their derivative nature, and thus helps to infer about the generalization capability of the network. For this reason, the small signal parameters are presented for all the above mentioned
modeling methods, with the TFT ($W = 160 \mu m$) operating in the linear and saturation regions, as shown in Fig. 5. The figure of merit used for relative comparison among the different ANN approaches is defined by the MARE (mean absolute relative error) as in (7), where $I_{\text{meas}}(i)$ and $I_{\text{model}}(i)$ refer to the measured and modeled drain current, respectively. The $g_m$ and $g_d$ are determined from numerical differentiation and the resulting MARE for each modeling approach is presented in Table 1, together with the corresponding ANN complexity.

Table 1

<table>
<thead>
<tr>
<th>Method</th>
<th>$g_m$</th>
<th>$g_d$</th>
<th>$g_m$</th>
<th>$g_d$</th>
<th>$g_m$</th>
<th>$g_d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP (15)</td>
<td>1.4</td>
<td>3.4</td>
<td>1.6</td>
<td>4.8</td>
<td>1.2</td>
<td>2.3</td>
</tr>
<tr>
<td>RBF (60)</td>
<td>0.7</td>
<td>5.5</td>
<td>1.45</td>
<td>6.2</td>
<td>0.4</td>
<td>2.5</td>
</tr>
<tr>
<td>LS-SVM (176)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5. Small signal parameters (a) $g_m$: when $V_{DS} = 1.5$ V (b) $g_m$: when $V_{DS} = 10$ V (c) $g_d$: when $V_{GS} = 8$ V (d) $g_d$: when $V_{GS} = 1.5$ V.

Fig. 4. (a) Fabricated transparent chip (b) micrograph of a single a-GIZO TFT.
From these results it can be observed that all ANNs tested seem to be potential candidates for modeling the non-linear behavior of the semiconductor. It can be noticed that the LS-SVM presents the best performance, however at a cost of a higher complexity. RBF network shows a relatively less performance compared to both MLP and LS-SVM, with lower complexity than LS-SVM, but higher than MLP. Finally, MLP expresses better performance than the RBF, but slightly less than LS-SVM. Yet, it accomplishes the results with minimal complexity. Therefore, this method represents the best option for circuit simulations, since its complexity is minimal with a comparable accuracy to LS-SVM. This way, a better simulation speed is expected, as the number of computations involved is smaller.

3.3. Circuit level simulation and validation

Once the optimum weights and bias are determined, MLP network (using multiple TFTs, with widths = 10, 60, 80, 100, 120, 160, 180, 220, 260 and 300 μm) is implemented in Verilog-A to enable the creation of a generic cell. A brief description of the verilog-A code is presented in annex.
The I/V characteristic of a TFT, never seen by the network during the training phase, with a width of 200 μm, is generated from the Verilog-A ANN model using Cadence Spectre simulator. The simulation results are further compared with the measured data, as shown in Fig. 6b (the I/V measurement setup is shown in Fig. 6a). The voltages $V_{GS}$ and $V_{DS}$ are varied from 2 to 10 V and 0.5 to 14.5 V, in steps of 1 V, respectively. From these results it can be understood that the network is capable of predicting the I/V relation for any aspect ratio that lies within the training range.

As a next step, an inverter is simulated with the developed Verilog-A model for a power supply of 10 V and input voltage sweep of 0 to 10 V. The schematic and the fabricated circuit are shown in Fig. 7(a) and (b) respectively. The fabricated inverter was also characterized under the same test set-up as the simulation environment. The circuit response, both from simulation and measurements are shown in Fig. 8, with DC and transient analysis. These results demonstrate that the ANN model is capable of predicting the behavior of devices, never seen during training, was assessed by comparing the I/V characteristics generated through simulation with the actual measurements. Results demonstrated that the model is able to anticipate the behavior of TFTs with any dimensions that lay within the training range. Later, an inverter was designed and simulated using the developed Verilog-A model, subsequently validated with the measured circuit response.

4. Conclusions

Results for modeled drain current and small signal parameters have shown good agreement with the measured data, implying that the ANNs (MLP, RBF, LS-SVM) are potential candidates for TFT modeling. Among the three ANN techniques, MLP is the most suitable approach for circuit simulations, since it resulted in minimal complexity with good accuracy. The MLP model was then implemented in Verilog-A for circuit simulations with Cadence Spectre. The model ability for predicting the behavior of devices, never seen during training, was assessed by comparing the I/V characteristics generated through simulation with the actual measurements. Results demonstrated that the model is able to anticipate the behavior of TFTs with any dimensions that lay within the training range. Later, an inverter was designed and simulated using the developed Verilog-A model, subsequently validated with the measured circuit response.

Acknowledgments

This work is funded by the ERDF through the Programme COMPETE and by the Portuguese Government through FCT – Foundation for Science and Technology, project Ref. CMU-PT/SIA/0005/2009, FCOMP-01-0124-FEDER-013070, PEST-C/CTM/LA0025/2011 strategic project, and also by the European Research Council, through the Advanced Grant INVISIBLE (ERC-2008-AdG 228144). The work of P. Bahubalindruni and C. Duarte is also partially supported by the FCT under Grants BD/62678/2009 and BD/28163/2006, respectively.

Appendix A. Verilog-A code

```
#include “constants.h”
#include “disciplines.h”
module gizo-tft(dg,s); //Module declaration
input g, d, s; //Input/Output terminals declaration
electrical d, g, s;
parameter integer NI, NO; //No. of inputs (vgs, vds) and ops
(Id-Drain current) //Define parameters
parameter integer NNHL; //No. of neurons in hidden layer

real hlayer-w[0:NI*(NI-1)], hlayer-b[0:NNHL-1], hlayer-y[0:NNHL-1], hlayer-v[0:NNHL-1]; //hidden layer
real olayer-w[0:NO*(NO-1)], olayer-b[0:NO-1], olayer-v[0:NO-1]; //output layer
real vgs,vds, width, inputs[0:1];
integer i, j, ii, jj;

// Define pre and post processing parameters
real train-input-range[0:(NI-1)], train-output-range, id;

//Preprocessing inputs
for (i = 0; i < (NI); i++) //Calculating local field and non
-linear activation of hidden neurons
begin
hlayer-v[i] = 0;
hlayer-v[i] = hlayer-v[i] + width*hlayer-w[NI*i] + vgs*hlayer-w[NI*i] + id;
hlayer-v[i] = hlayer-v[i] + hlayer-b[i];
hlayer-y[i] = tanh(hlayer-v[i]);
end
for (ii = 0; ii < (NO); ii++) //Calculating local field and output of op neurons
begin
olayer-v[ii] = 0;
for (jj = 0; jj < (NNHL); jj++) //Calculating local field and non
-linear activation of hidden neurons
begin
olayer-v[ii] = olayer-v[ii] + hlayer-y[jj] *olayer-w[jj];
end
olayer-v[ii] = olayer-v[ii] + olayer-b[ii];
id = olayer-v[ii];
end
//Post processing to bring the drain current to normal scale
l(d,s) <<(id);
l(g,s) <<0;
l(g,d) <<0;
end
endmodule
```

References


Choi S, Han M. Effect of channel widths on negative shift of threshold voltage, including stress-induced hump phenomenon in InGaZnO thin-film transistors under high-gate and drain bias stress. Appl Phys Lett 2012;100. 043503.


