Signal and Noise Modeling of Microwave Transistors Using Characteristic Support Vector-based Sparse Regression

Filiz GÜNEŞ, Mehmet Ali BELEN, Peyman MAHOUTİ, Salih DEMİREL

Electronic and Communication Engineering Dept., Yıldız Technical University, Istanbul, Turkey

gunes@yildiz.edu.tr, mbelen@yildiz.edu.tr, pmahouti@gmail.com, salihd@yildiz.edu.tr

Abstract. *In this work, an accurate and reliable S- and Noise (N) - parameter black-box models for a microwave transistor are constructed based on the sparse regression using the Support Vector Regression Machine (SVRM) as a nonlinear extrapolator trained by the data measured at the typical bias currents belonging to only a single bias voltage in the middle region of the device operation domain of (V_{DS}/V_{CE}, I_{DS}/I_C, f). SVRMs are novel learning machines combining the convex optimization theory with the generalization and therefore they guarantee the global minimum and the sparse solution which can be expressed as a continuous function of the input variables using a subset of the training data so called Support Vector (SV)s. Thus magnitude and phase of each S- or N- parameter are expressed analytically valid in the wide range of device operation domain in terms of the Characteristic SVs obtained from the substantially reduced measured data. The proposed method is implemented successfully to modeling of the two LNA transistors ATF-551M4 and VMMK 1225 with their large operation domains and the comparative error-metric analysis is given in details with the counterpart method Generalized Regression Neural Network GRNN. It can be concluded that the Characteristic Support Vector based-sparse regression is an accurate and reliable method for the black-box signal and noise modeling of microwave transistors that extrapolates a reduced amount of training data consisting of the S- and N-data measured at the typical bias currents belonging to only a middle bias voltage in the form of continuous functions into the wide operation range.*

Keywords

Scattering S-parameters, noise N-parameters, Support Vector Regression Machine (SVRM), Generalized Regression Neural Network (GRNN)

1. Introduction

Fast and accurate models of microwave devices and antennas are indispensable in contemporary microwave engineering. In today's RF and microwave technology, there is an ever-increasing demand for higher level of system integration that leads to massive computational tasks during simulation, optimization and statistical analyses, requiring efficient modeling methods so that the whole process can achieve the required reliability. However modeling still remains a major bottleneck for efficient RF-Microwave CAD. Among all component models, unreliable transistor models can easily lead to unsuccessful design because of their strong influence on the overall circuit performance. Thus, the efficiency of such models in terms of accuracy and speediness is critical to assure a reliable design.

Artificial Neural Network (ANN)s have emerged as the valuable tools to extend the repertoire of the statistical methods. Particularly the Back-Propagation Multi-Layer Perceptron (BPMLP)s have been employed for the non-linear interpolation based on "learning" from the measured or simulated data, in the fast, accurate and reliable modeling of both active and passive microwave devices [1–6]. Today's fast, accurate and reliable Signal S- and Noise N-black-box model of a microwave transistor can be achieved only by a single simple BPMLP with one hidden layer which is capable of the simultaneous generalization of 12 Scattering S- and Noise N- functions into the entire operation domain of the bias condition V_{DS}/V_{CE}, I_{DS}/I_C and the frequency f for all the configuration types [3]. However, these so-called Back Propagation Neural Networks BPNNs suffer from a number of disadvantages. The most important disadvantages can briefly be summarized as follows: (i) The variants of the back-propagation can be shown to converge to the local minima of the error surface; (ii) As most algorithms use (pseudo) random numbers, the result of training varies between runs even for identical networks and training datasets. Therefore BPNNs in modeling need statistical analysis of the results of a lot of runs.

In recent years, regression with Support Vector Machine (SVM)s and probability based neural networks named as Generalized Regression Neural Network (GRNN) remove these handicaps and facilitate to substantially reduce the expensive fine discrete training data with

their typical modeling applications to transistor [7–10], integrated microstrip and slot antennas in [11], [12], printed transmission lines [13–15] and vertical interconnects in microwave packaging structures [16].

The SVM based on the Structural Risk Minimization (SRM) principle is one of the most widely used learning algorithms, and achieves superior generalization performance for both classification and regression problems [16]. The SVM is systematic and properly motivated by the statistical learning theory [17]. Training of the SVM involves optimization of a convex objective function and globally minimizes to complete the learning process without suffering anymore from local minima [17], [18]. In addition, SVM can handle a large input, and can automatically identify a small subset consisting of informative points, namely Support Vector (SV)s [17]. Thus sparseness of solutions is obtained that the large amount of data to be fully characterized by the set of SVs, a subset of the training set. Especially the working principle of SVRM based on the small sample statistical learning theory is utilized in [13], [14] to build a knowledge-based SVRM model for the synthesis of printed transmission lines as fast as the coarse models and at the same time as accurate as the fine models. In those works, the SVs are determined using the coarse data generator which are the empirical synthesis formulae of the related transmission line and accuracy of the models is increased by supplying the corresponding fine values of the SVs from the fine data generator which are the full-wave EM simulators. Thus, amount of the expensive fine training data is reduced substantially as pointed out in [11–16].

In this paper, we worked out to determine the characteristic SVs for each S- and N- parameter using the sparse training data, that is subject to be generalized throughout a wide range of operation domain. Thus a fast, accurate and reliable black-box S- and N- parameter models are constructed based on the sparse regression using SVRM as a nonlinear extrapolator trained by the data measured at the typical bias currents of the single bias voltage in the middle region of the device operation domain of (V_{DS}/V_{CE}, I_{DS}/I_C, f). This modeling method is facilitated by the two features: The first one belongs to the black-box characterization parameters which depend upon the bias currents more than the bias voltages; the second is the superior generalization ability of the SVRM. On the other hand, the training time and memory of SVM are expensive and are strongly correlated to the number of training patterns ℓ, as $O(\ell^3)$ and $O(\ell^2)$, respectively [17–19]. Thus approximately at least 50% of the training data is reduced that corresponds to 1/8 and 1/4 of the training time and the memory complexity respectively of the SVRM as the nonlinear interpolator in the S- and N-modeling used in [3–8]. The proposed method is implemented on the modeling of a low-noise microwave transistor ATF-551M4 and VMMK-1225 as a case study.

The paper is organized as follows: Fundamentals of the SVRM are briefly given in the next section. Third section is devoted to the case study where SVRMs are used as nonlinear extrapolators in modeling of the microwave transistors together with error-metric as compared to the counterpart method (GRNN) performance. The paper ends with the conclusion.

2. Support Vector Regression Machines

SVRM builds a nonlinear function between a given input and its corresponding output in the training data. This continuous nonlinear relation can be used to predict outputs for the given inputs not included in the training data. The nonlinear function is learned by a linear learning machine in a kernel induced feature space. As in the classification case, the learning algorithm minimizes a convex cost function and its solution is sparse. In order to explain the mathematical framework of the SVRMs, let us consider a training dataset:

$$D = (\vec{x}_i, y_i), i = 1, 2, \dots \ell , \tag{1}$$

where $\vec{x}_i \in R^n$, $y_i \in R$, SVRM tries to find the mapping function $f(\vec{x})$ between the input variable vector $\vec{x}$ and the desired output variable y, which are operation parameters of the device (V_{DS}/V_{CE}, I_{DS}/I_C, f), and magnitude and phase of each S- or N- parameter, respectively in our case. In the SVRM model, the regression function $f(\vec{x})$ is expressed as follows [14–18]:

$$f(\vec{x}) = \vec{w}^T . \vec{\phi}(\vec{x}) + b \tag{2.1}$$

where

$$\vec{x}^T = (x_1, x_2, \dots x_n) \Rightarrow \vec{\phi}^T(\vec{x}) = (\phi_1(\vec{x}), \phi_2(\vec{x}), \dots \phi_N(\vec{x})). \tag{2.2}$$

This step is equivalent to mapping the x-input space into a new space $F = \{\vec{\phi}(\vec{x}) / \vec{x} \in X^n\}$. Thus, $f(\vec{x})$ is a nonlinear function in the n dimensioned x-input space and linear function in the N dimensioned F-feature space. Dimension of the feature space N may be smaller than or equal to the dimension of the input space n. However, strategy to follow in the selection of the F-feature space is to convey the essential information of the original data into its representation in the new space. Thus, a SVRM can be constructed in two steps: First, a fixed nonlinear mapping vector $\vec{\phi}(\vec{x})$ transforms the data into a feature space F, and then the linear machine built in this feature space is used to perform regression on the data. In this manner, we will refer to the quantities $\vec{w}$ and b in (1) as the weight vector and bias:

$$\vec{w}^T = (w_1, w_2, \dots w_N) . \tag{3}$$

The regression function in (2.1) is built by determination of $\vec{w}$ and b by the strategy "seeking to optimize the generalization bounds". This is relied on defining a loss function that ignored errors that were within a certain distance of the true value. This type of function is referred to

as ε-insensitive loss function. The use of ε-insensitive loss function has the advantage of ensuring the existence of a global minimum and the optimization of a reliable generalization bound [17], [18]. Linear ε-insensitive loss function is given as follows [17], [18]:

$$L_\varepsilon(\vec{x}, y, f) = \begin{cases} 0 & |f(\vec{x}_i) - y_i| < \varepsilon \\ |y_i - f(\vec{x}_i, y_i)| - \varepsilon & \text{otherwise} \end{cases} \quad (4)$$

where ε is predefined. Equation (4) defines ε tube so that if the predicted value is within the tube the loss is zero, while if the predicted point is outside the tube, the loss is the magnitude of the difference between the predicted value and target with the radius ε of the tube. In order to minimize the sum of the linear ε-insensitive losses the function can be given as [17], [18]:

$$\frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{\ell} L_\varepsilon(\vec{x}_i, y_i, f) \quad (5)$$

to control the size of $\|\vec{w}\|$ for a fixed training set. C is a parameter to measure the trade-off between the complexity and losses. Since the given objective function (5) is a convex, it has no local minima and it guarantees the global minimum which is one of the advantages of the SVRM on the other regression methods, especially neural networks. Substituting (4) into (5) and introducing two slack variables ξ_i and $\hat{\xi}_i$, equation (5) is transformed into the following soft margin primal optimization problem [17], [18]:

$$\min_{\vec{w}, b, \xi_i, \hat{\xi}_i} \{\frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{\ell}(\xi_i + \hat{\xi}_i)\} \quad (6a)$$

subject to

$$\begin{aligned} (\langle \vec{w}, \vec{\phi}(\vec{x}_i) \rangle + b) - y_i &\leq \varepsilon + \xi_i \quad \xi_i, \hat{\xi}_i \geq 0, \\ y_i - (\langle \vec{w}, \vec{\phi}(\vec{x}_i) \rangle - b) &\leq \varepsilon + \hat{\xi}_i \quad i = 1, 2, ..., \ell \end{aligned} \quad (6b)$$

The optimization problem in (6a), (6b) can be solved more easily in its dual formulation. In this stage, a Lagrangian function L is constructed combining the soft margin primal objective function in (6a) with the constraints in (6b) and introducing Lagrangian multipliers $\alpha_i, \hat{\alpha}_i, \gamma_i, \hat{\gamma}_i$ as follows:

$$\begin{aligned} L = &\frac{1}{2}(\vec{w}^T \cdot \vec{w}) + C(\sum_{i=1}^{\ell}\xi_i + \sum_{i=1}^{\ell}\hat{\xi}_i) \\ &- \sum_{i=1}^{\ell}\hat{\alpha}_i[y_i - (\vec{w}^T \cdot \vec{\phi}(\vec{x}_i) - b + \varepsilon + \hat{\xi}_i] \\ &- \sum_{i=1}^{\ell}\alpha_i[(\vec{w}^T \cdot \vec{\phi}(\vec{x}_i) - b - y_i + \varepsilon + \xi_i] \\ &- \sum_{i=1}^{\ell}(\hat{\gamma}_i\hat{\xi}_i + \gamma_i\xi_i) \end{aligned} \quad (7)$$

The function L has a saddle point with respect to the primal $\vec{w}, b, \xi_i, \hat{\xi}_i$ and dual positive variables $\alpha_i, \hat{\alpha}_i, \gamma_i, \hat{\gamma}_i$

at the optimal solution. By substituting the saddle point conditions with respect to the primal variables into the Lagrangian function L, we have the following equivalent dual space objective function to be maximized [17], [18]:

$$\begin{aligned} W(\alpha_i, \hat{\alpha}_i) = &\max_{\vec{w}, b, \xi_i, \hat{\xi}_i} L(\vec{w}, b, \xi_i, \hat{\xi}_i, \alpha_i, \hat{\alpha}_i, \gamma_i, \hat{\gamma}_i) \\ &\sum_{i=1}^{\ell} y_i(\hat{\alpha}_i - \alpha_i) - \varepsilon \sum_{i=1}^{\ell}(\hat{\alpha}_i + \alpha_i) \\ &-\frac{1}{2}\sum_{i,j=1}^{\ell}(\hat{\alpha}_i - \alpha_i)(\hat{\alpha}_j - \alpha_j)\langle \vec{\phi}(\vec{x}_i), \vec{\phi}(\vec{x}_j)) \rangle \end{aligned} \quad (8a)$$

subject to

$$0 \leq \alpha_i, \ \hat{\alpha}_i \leq C, \ \sum_{i=1}^{\ell}(\hat{\alpha}_i - \alpha_i) = 0, \ i = 1, 2, ..., \ell. \quad (8b)$$

In (8a, 8b) the dual variables $\gamma_i, \hat{\gamma}_i$ are expressed in terms of $\alpha_i, \hat{\alpha}_i$ and C through the saddle point conditions.

The corresponding Karush-Kuhn-Tucker (KKT) complementarity conditions are [17], [18]:

$$\alpha_i(\langle \vec{w}, \vec{\phi}(\vec{x}_i) \rangle + b - y_i - \varepsilon - \xi_i) = 0 , \quad (9a)$$

$$\hat{\alpha}_i(y_i - \langle \vec{w}, \vec{\phi}(\vec{x}_i) \rangle - b - \varepsilon - \hat{\xi}_i) = 0 , \quad (9b)$$

$$\xi_i \cdot \hat{\xi}_i = \alpha_i \cdot \hat{\alpha}_i = 0, \ (\alpha_i - C) \cdot \xi_i = (\hat{\alpha}_i - C) \cdot \hat{\xi}_i = 0 . \quad (9c)$$

We must find ℓ Lagrangian multipliers $\alpha_i, \hat{\alpha}_i$. For this purpose, firstly the nonzero Lagrangian multipliers are determined using the KKT conditions and substituted into (8a) and (8b), then the obtained equation is maximized with respect to the Lagrangian multipliers $\alpha_i, \hat{\alpha}_i$ in the feature space F. From (9a) and (9b) of the KKT conditions, it follows that for only the samples satisfying $|f(\vec{x}_i) - y_i| \geq \varepsilon$, the Lagrangian multipliers may be nonzero, and for the samples of $|f(\vec{x}_i) - y_i| < \varepsilon$, the Lagrangian multipliers vanish. Since products of α_is and $\hat{\alpha}_i$s are zero in (9c) that means that at least one of these terms is zero. The sampling data $(\vec{x}_i, y_i)$ that comes with the non-vanishing Lagrangian multipliers are called Support Vector SVs, thus there can be at most ℓ SVs. These SVs may be named as "Characteristic SV"s in our case, since they facilitate to express a transistor characteristic parameter which is either signal or noise parameter, analytically in a wide range of the operation domain. Then the mapping function $f(\vec{x})$ between the input variable space and the desired output variable can be expressed in terms of the SVs within the kernel $K(\vec{x}, \vec{x}_i)$ space as follows:

$$f(\vec{x}) = \sum_{i=1}^{n_{SV}} \alpha_i y_i K(\vec{x}, \vec{x}_i) + b \quad (10)$$

$$K(\vec{x}, \vec{x}_i) = \langle \vec{\phi}(\vec{x}), \vec{\phi}(\vec{x}_i) \rangle$$

where Radial Basis Function (RBF) is chosen as the kernel $K(\vec{x}, \vec{x}_i)$ function in the case study and n_{SV} is the number of the characteristic SVs ($n_{SV} \leq \ell$) and b is determined

using KKT conditions (10a) and (10b). Thus we have for $f(\vec{x})$ an analytical function mapping the n-dimensioned x-input space and one-dimensional y_i-output space, in terms of n_{SV} characteristic SVs within the RBF kernel domain as follows:

$$f(\vec{x}) = \sum_{i=1}^{n_{SV}} \alpha_i \, y_i \exp\{-\|\vec{x} - \vec{x}_i\|^2 / 2\sigma^2\} + b \qquad (11)$$

where n_{SV} is number of the characteristic SVs obtained by applying the ε-tube SVRM selection process into the training data set $D = (\vec{x}_i, y_i)$, $i = 1, 2, \ldots \ell$ where $\vec{x}_i \in R^n$, $y_i \in R$. In the next section a small-signal transistor will be characterized as a black-box for use in small-signal amplification. In Sec. 4, as a case study, $f(\vec{x})$ will be obtained for magnitude and phase of each Scattering (S-) or Noise (N-) parameter for typical LNA transistors ATF551M4 and VMMK1225 using their manufacturer's data belonging to the typical currents of only a single bias voltage in the middle region of the device operation domain of (V_{DS}, I_{DS}, f). Furthermore error-metric analysis of both of the two transistors will also be given as compared to the counterpart method GRNN which has also been worked out for the transistor modeling by our research group in [8–10].

3. A Small-Signal Microwave Transistor

3.1 Black-Box Characterization Parameters

The black-box representation of a small-signal transistor with its terminations and port reflections is shown in Fig. 1. Here signal and noise performance of a small-signal transistor are given by scattering $\vec{S}$ (V_{DS}, I_{DS}, f), noise $\vec{N}$ (V_{DS}, I_{DS}, f) functions in the device operation domain of bias condition (V_{DS}, I_{DS}) and frequency f. Thus the measured $\vec{S}$, $\vec{N}$ data at the discrete frequencies throughout the operational band at a bias condition (V_{DS}, I_{DS}) can be arranged in a table-form function as follows:

$$\begin{bmatrix} f_1 & \vec{S}^{(1)} & \vec{N}^{(1)} \\ f_2 & \vec{S}^{(2)} & \vec{N}^{(2)} \\ . & & \\ f_i & \vec{S}^{(i)} & \vec{N}^{(i)} \\ . & & \\ f_l & \vec{S}^{(l)} & \vec{N}^{(l)} \end{bmatrix} \qquad (12.1)$$

where $\vec{S}^{(i)}$, $\vec{N}^{(i)}$ are the $\vec{S}$, $\vec{N}$ parameter data at sampling frequency f_i which can be expressed in the following form:

$$\vec{S}^{(i)} = \left[|S_{11}|^{(i)}, \angle S_{11}^{(i)}, |S_{12}|^{(i)}, \angle S_{12}^{(i)}, |S_{21}|^{(i)}, \angle S_{21}^{(i)}, |S_{22}|^{(i)}, \angle S_{22}^{(i)} \right]^t,$$
$$\vec{N}^{(i)} = \left[F_{\min}^{(i)}, |\Gamma_{opt}|^{(i)}, \angle \Gamma_{Opt}^{(i)}, R_n^{(i)} / 50 \right]^t, \quad i = 1, 2, \ldots \ell \qquad (12.2)$$

where i is each data sample, φ is phase value for the selected S/N parameters, $|\cdot|$ is magnitude value for the selected S/N parameter.

In this work, the $\vec{S}$, $\vec{N}$ data defined by (12.1), (12.2) belonging to the typical currents of only a single bias voltage in the middle region of the device operation domain of (V_{DS}, I_{DS}, f) is found to be sufficient for training SVRM. Then, the characterization vectors $\vec{S}$ (V_{DS}, I_{DS}, f), $\vec{N}$ (V_{DS}, I_{DS}, f), at any desired frequency f of any bias condition (V_{DS}, I_{DS}) can be obtained from the network output by inputting that (V_{DS}, I_{DS}) and the frequency f. Thus the transistor under consideration is characterized continuously all throughout the device operation domain using greatly reduced amount of sampling data.

3.2 Performance Measure Functions

Once the characterization $\vec{S}$, $\vec{N}$ functions are determined, the signal and noise performance of that transistor are known for any input and output (Γ_S, Γ_L) termination couple at any operation condition (V_{DS}, I_{DS}, f). There are four functions measuring the performance of the small-signal transistor (Fig. 1) [2]:

(1) Transducer power gain G_T of an active device is defined as the ratio of the power delivered to the load P_L to the available maximum power P_{AVS} from the source that is the function of the (Γ_S, Γ_L) termination couple and $\vec{S}$ parameters as follows:

$$G_T = \frac{P_L}{P_{AVS}} = G_T(\Gamma_S, \Gamma_L, \vec{S}) = \frac{(1 - |\Gamma_L|^2)}{|1 - S_{22}\Gamma_L|^2} |S_{21}|^2 \frac{(1 - |\Gamma_S|^2)}{|1 - \Gamma_{in}\Gamma_S|^2}. \qquad (13)$$

(2), (3) M_{in} and M_{out} are mismatching functions that measure net powers entering into the input and load, respectively. Their dependences to the device signal $\vec{S}$ and the termination (Γ_S, Γ_L) couple can be given as follows:

$$M_{in} = \frac{P_{in}}{P_{AVS}} = M_{in}(\Gamma_S, \Gamma_L, \vec{S}) = \frac{|1 - \Gamma_{in}\Gamma_S|^2}{(1 - |\Gamma_S|^2)(1 - |\Gamma_{in}|^2)}, \qquad (14a)$$

$$M_{out} = \frac{P_L}{P_{AVout}} = M_{out}(\Gamma_S, \Gamma_L, \vec{S}) = \frac{|1 - \Gamma_{out}\Gamma_L|^2}{(1 - |\Gamma_L|^2)(1 - |\Gamma_{out}|^2)} \qquad (14b)$$

where the input Γ_{in} and output Γ_{out} reflection coefficient are given in terms of the device signal $\vec{S}$ and the termination (Γ_S, Γ_L) couple:

$$\Gamma_{in} = \Gamma_{in}(\Gamma_L, \vec{S}) = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{11}\Gamma_L}, \qquad (15a)$$

$$\Gamma_{out} = \Gamma_{out}(\Gamma_S, \vec{S}) = S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - S_{22}\Gamma_S}. \qquad (15b)$$

(4) The noise figure F of an active device is defined as the ratio of signal-to-noise ratios available at input and output; the noise vector $\vec{N}$ describes the dependence of the
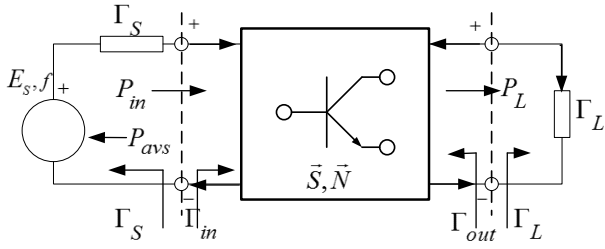
Fig. 1. Black-box representation of a small-signal transistor and its port impedances.

transistor noise figure F on the input termination (source) reflection coefficient Γ_S. These are linked through the following relationship:

$$F(\Gamma_S, \vec{N}) = F_{min} + 4\frac{R_N}{Z_0}\frac{|\Gamma_S - \Gamma_{opt}|^2}{(1 - |\Gamma_S|^2)|1 + \Gamma_{opt}|^2}. \qquad (16)$$

4. Case Study

In this work, the proposed method is implemented to modeling of the signal and noise parameters of the two transistors ATF551M4 and VMMK 1225 and the library LIBSVM for Support Vector Machines is employed [20], [21]. ATF551M4 previously has been modeled using SVRM in the interpolation process in [7]. On the other hand, VMMK1225 is properly selected since it is a typical LNA transistor with the given manufacturer's data in a large operation bandwidth of 2–45 GHz at the bias currents from the lower range up to the upper range of I_{DS}= 5, 10, 15, 20 mA of the bias voltages of V_{DS}= 2, 3, 4 V [22].

Firstly let us consider ATF551M4. The 24 S- data for ATF551M4 transistor are supplied ranging in the 0.1–18 GHz at each of the bias (V_{DS}, I_{DS}) condition where the bias current is given as I_{DS}= 10, 15 and 20 mA of each bias voltage of V_{DS}= 2, 2.7, 3 V by the manufacturer's data sheets. Nevertheless, 15 N-data are supplied ranging in the 0.5–10 GHz at each of the same bias (V_{DS}, I_{DS}) conditions. In Tab. 1, the training and test data sets are given for the SVRM modeling with respect to the bias condition. The data belong to 2.7 V DC bias condition are given to SVRM model for training purpose only, while 2 and 3 V DC bias data are used after training process for test/validation purpose. Thus by this means, the total measured data provided by manufacturer datasheets is separated in two different datasets for training and test process. In Tab. 2 number of characteristic SVs is given for each SVRM model of the magnitude and phase of each S- and N- parameter and the spread parameter σ of the radial kernel function in (11) is taken to be equal to 0.1 for the lowest Mean Absolute Error MAEs using (17) and (18) taken place in Tab. 3, 4. As seen from the results given in Tab. 3, 4, the overall performance of SVRM model is higher than its counterpart method GRNN, particularly in S- parameter domain, where the variations of parameters are more complex than N- parameters for same DC bias conditions over the operation frequency bandwidth.

V_{DS}/I_{DS}	10 mA	15 mA	20 mA
2 V	Test	Test	Test
2.7 V	Training	Training	Training
3 V	Test	Test	Test

Tab. 1. Training and test data set for ATF551M4.

Scattering Parameters	Number of SVs (Total Training Data:72)	
	Mag.	Phase
S_{11}	31	72
S_{21}	65	72
S_{12}	72	72
S_{22}	64	72
Noise Parameters	Number of SVs (Total Training Data:45)	
	Mag.	Phase
F_{min}	45	-
Γ_{opt}	42	45
$R_N/50$	45	-

Tab. 2. Number of the characteristic SVs in modeling of ATF551M4.

Parameter	Error Metric			
	Mean Absolute Error MAE		Relative Mean Error RME	
	SVRM	GRNN	SVRM	GRNN
S_{11}	0.009	0.018	0.013	0.007
S_{21}	0.279	0.433	0.088	0.015
S_{12}	0.007	0.005	0.132	0.052
S_{22}	0.056	0.314	0.228	9.441

Tab. 3. Error metric of scattering parameters for SVRM extrapolation of ATF551M4.

Parameter	Error Metric			
	Mean Absolute Error MAE		Relative Mean Error RME	
	SVRM	GRNN	SVRM	GRNN
F_{min}	0.023	0.016	0.050	0.036
$|\Gamma_{opt}|$	0.006	0.007	0.016	0.018
φ_{opt}	0.726	0.971	0.019	0.020
Γ_{opt}	0.031	0.067	0.018	0.029
$R_N/50$	0.002	0.001	0.028	0.010

Tab. 4. Error metric of noise parameters for SVRM extrapolation of ATF551M4.

Comparative Error Metrics in Tab. 3 and 4 use the definitions given by (17) and (18) in calculation of Mean Absolute Error MAE and Relative Mean Error RME for both the SVRM and GRNN methods:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|T_i - P_i|, \qquad (17)$$

$$RME = \frac{1}{N}\sum_{i=1}^{N}\frac{|T_i - P_i|}{|T_i|} \qquad (18)$$

where P_i and T_i are the ith predicted and target phasor value, respectively and N is the sampling number for the validation of the SVRM or GRNN model.

In Figs. 2, 3, the predicted scattering parameters obtained from the Characteristic SV–based sparse regression

are compared on the Smith chart and polar plane with the target values for bias conditions (2 V,10 mA) and (3 V, 20 mA), respectively. Furthermore in Figs. 4 and 5, predicted results of the typical parameters S_{11} phase and S_{21} magnitude are compared with all the test samples belonging to the bias voltages V_{DS} of 2 V and 3 V, respectively.
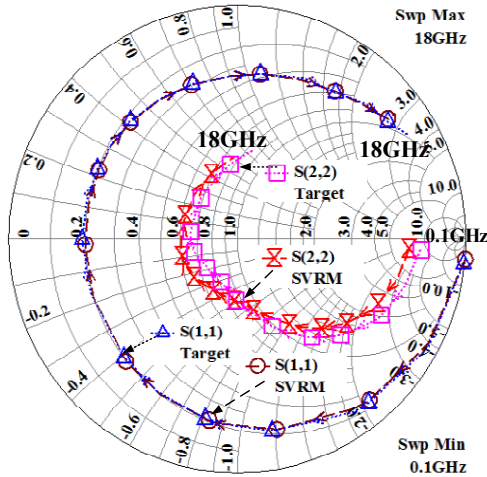

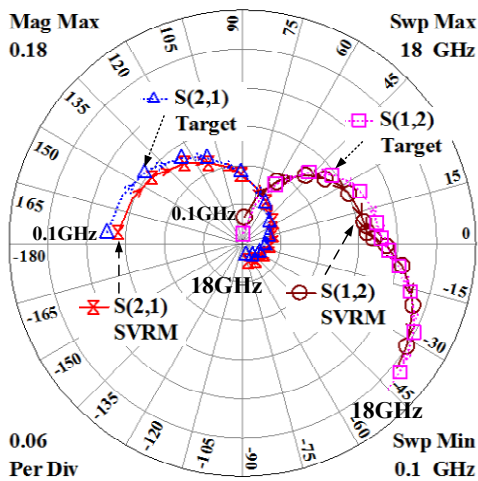**Fig. 2.** S_{22} and S_{11} predicted values vs. measurements at the bias condition (2 V, 10 mA) for ATF551M4.

Fig. 3. S_{12} and S_{21} predicted values vs. measurements at the bias condition (2 V, 10 mA) for ATF551M4.
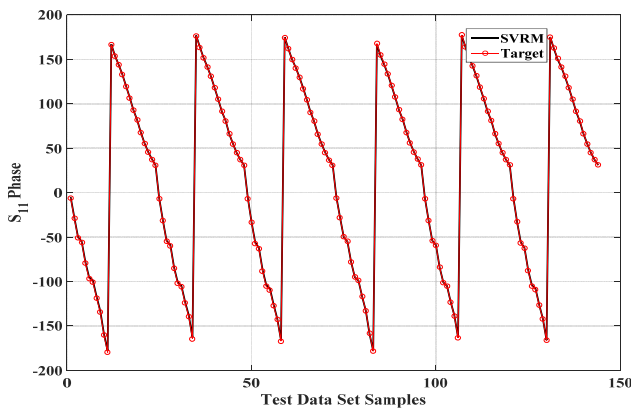
Fig. 4. S_{11} phase throughout the entire test data domain of ATF551M4.
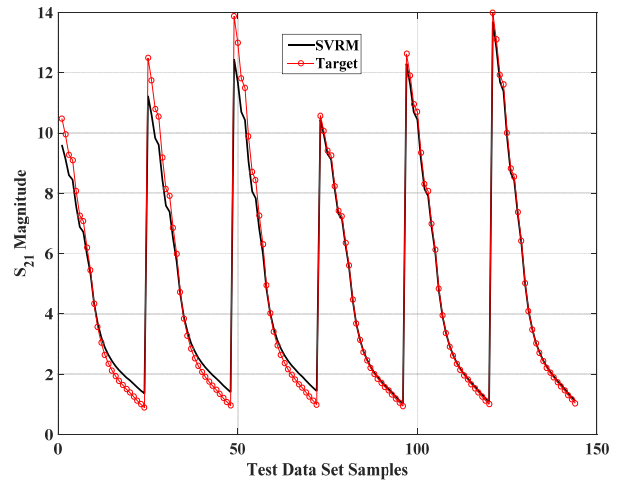

Fig. 5. S_{21} magnitude throughout the entire test data domain of ATF551M4.

In Fig. 6, the predicted Noise parameters of ATF551M4 using the Characteristic SV–based sparse regression are compared with the target values for bias condition (2 V, 10 mA).

Similar to modeling of ATF551M4, the Characteristic Support Vector-based Sparse Regression is also implemented to the modeling of another LNA transistor VMMK1225 within a large operation bandwidth of 2–45 GHz at the bias currents from the lower ranges up to the upper range of I_{DS} = 5, 10, 15, 20 mA of the bias voltages of V_{DS} = 2, 3, 4 V [22]. 692 data belonging to the typical bias currents of 5, 10, 15, 20 mA at the middle bias voltage 3 V are used for the S-parameter training and the model is tested the rest 1384 data as given in Tab. 5. The N-parameters of the VMMK1225 are given within the range of 2–17 GHz frequency in the manufacturer's datasheet. Similarly the N-parameter model is built by training with the central 64 data and tested with the rest 128 data as in Tab. 5. Similarly in case of modeling of ATF551M4, the data belong to 3 V DC bias condition are given to SVRM model for training purpose only, while 2 and 4 V DC bias voltage values are used after training process for test/validation purpose. Thus by this mean, the
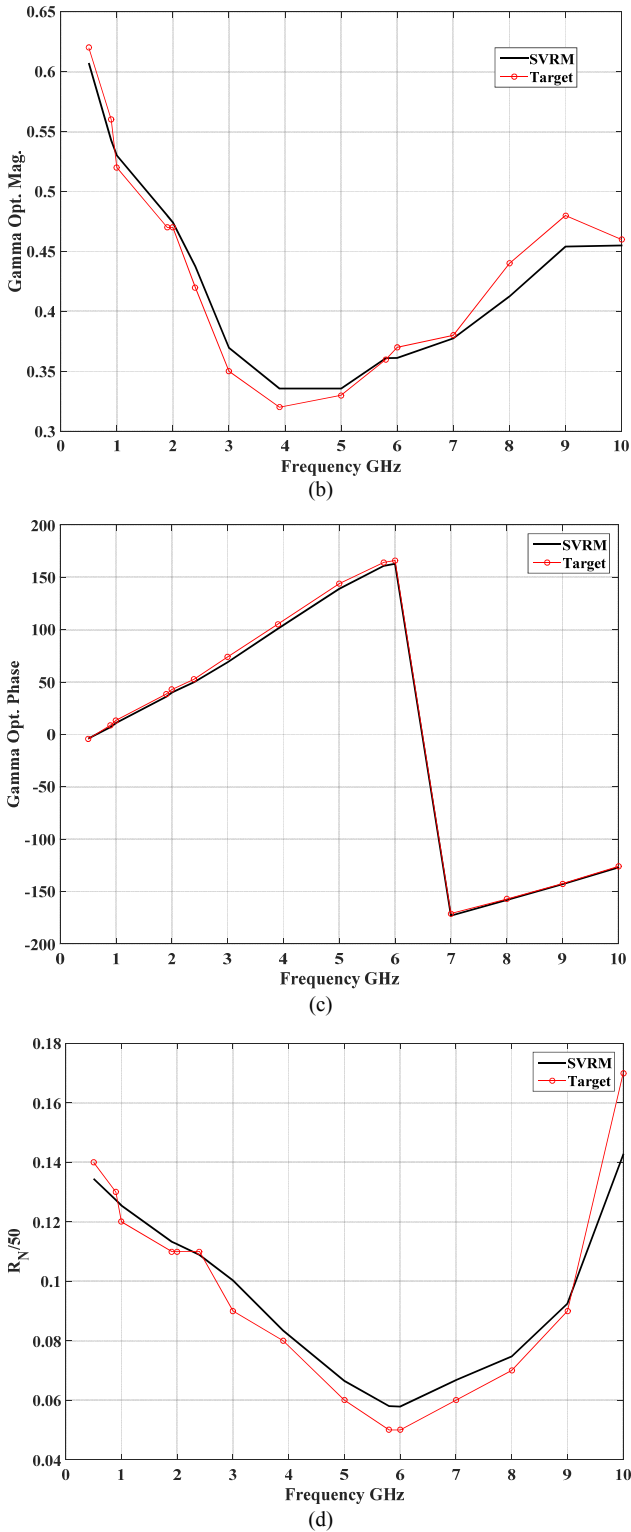

(a)
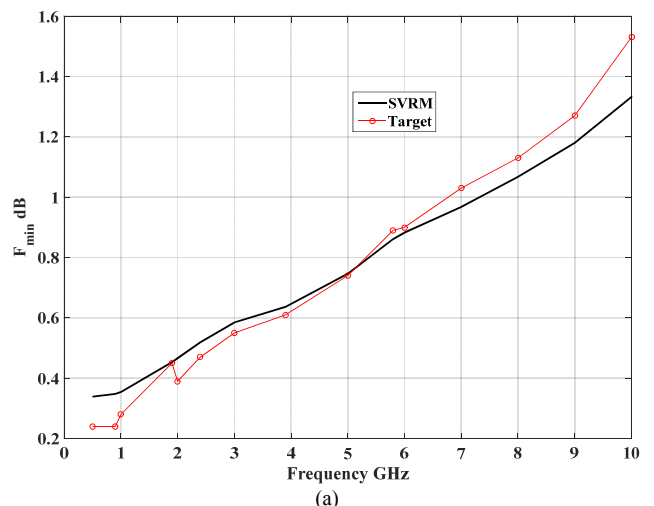
Fig. 6. Extrapolated prediction of SVRM vs. measurements at the bias condition (2 V, 10 mA) of ATF551M4: (a) F_{min}, (b) $|\Gamma_{opt}|$, (c) Γ_{opt} phase, (d) $R_N/50$.

V_{DS}/I_{DS}	5 mA	10 mA	15 mA	20 mA
2 V	Test	Test	Test	Test
3 V	Training	Training	Training	Training
4 V	Test	Test	Test	Test

Tab. 5. Training and test data set for VMMK-1225.

Scattering Parameters	Number of SVs (Total Training Data: 692)	
	Mag.	Phase
S_{11}	685	692
S_{21}	181	692
S_{12}	676	666
S_{22}	31	692
Noise Parameters	Number of SVs (Total Training Data:64)	
	Mag.	Phase
F_{min}	23	-
Γ_{opt}	38	34
$R_N/50$	64	-

Tab. 6. Number of the characteristic SV in modeling of VMMK1225.

Parameter	Error Metric			
	Mean Absolute Error MAE		Relative Mean Error RME	
	SVRM	GRNN	SVRM	GRNN
S_{11}	0.024	0.036	0.03	0.047
S_{21}	0.124	0.495	0.066	0.26
S_{12}	0.014	0.014	0.22	0.22
S_{22}	0.101	0.106	0.22	0.23

Tab. 7. Error metric of scattering parameters for SVRM extrapolation of VMMK1225.

Parameter	Error Metric			
	Mean Absolute Error MAE		Relative Mean Error RME	
	SVRM	GRNN	SVRM	GRNN
F_{min}	0.048	0.104	0.093	0.153
$|\Gamma_{opt}|$	0.039	0.030	0.095	0.076
φ_{opt}	0.914	2.385	0.012	0.029
Γ_{opt}	0.262	0.011	0.095	0.096
$R_N/50$	0.010	0.104	0.085	0.153

Tab. 8. Error metric of noise parameters for SVRM extrapolation of VMMK1225.

total measured data provided by manufacturers datasheets is separated in two different datasets for training and test process. The reduced number of characteristic support vector corresponding to each scattering and noise parameter is given in Tab. 6. As seen from the results given in Tab. 7, 8, similarly to the case of modeling of ATF551M4,

again for VMMK1225 the overall performance of SVRM modeling method is higher than its counterpart method, GRNN. Thus, one can infer that S- and N- parameter modeling based on Characteristic Support Vector-based Sparse Regression, SVRM is a fast and reliable modeling method requiring substantially reduced measurements and human effort.

Tables 7 and 8 give the error metrics of S- and N-parameters compared with the values resulted from the

counterpart method GRNN respectively. In Figs. 7, 8, the predicted scattering parameters obtained from the characteristic SV–based sparse regression are compared on the Smith chart and polar plane with the target values for bias

condition (4 V, 20 mA). As is seen from Fig. 7, 8, the SVRM based regression model has a very accurate response over the wide frequency bandwidth. Likewise in Fig. 7. 8, the response of SVRM model for all test data
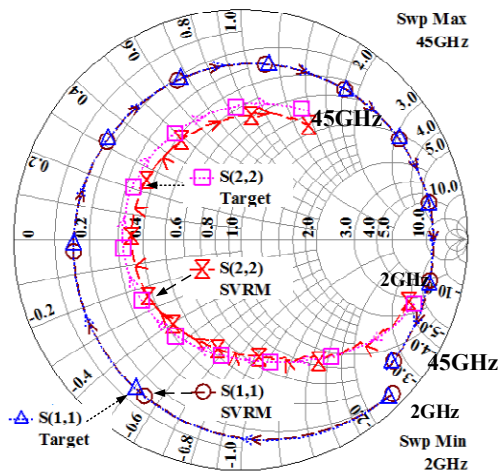

Fig. 7. S_{11} and S_{22} predicted values vs. measurement at the bias condition (4 V, 20 mA) of VMMK1225.
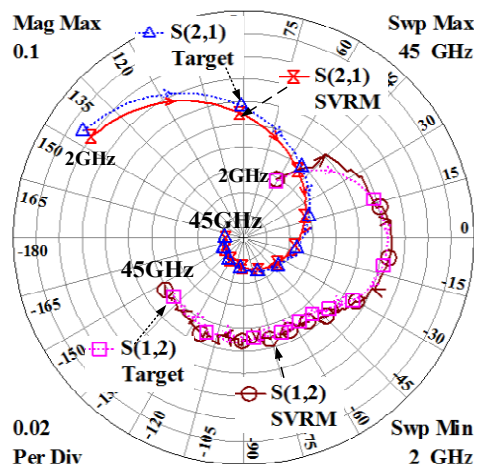
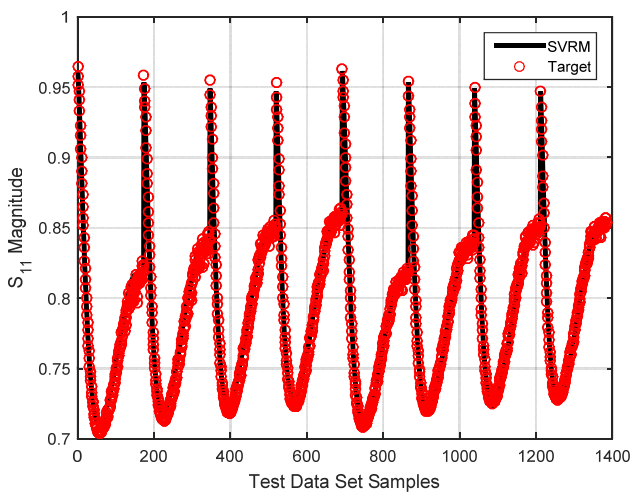**Fig. 8.** S_{12} and S_{21} predicted values vs. measurements at the bias condition (4 V, 20 mA) of VMMK1225.

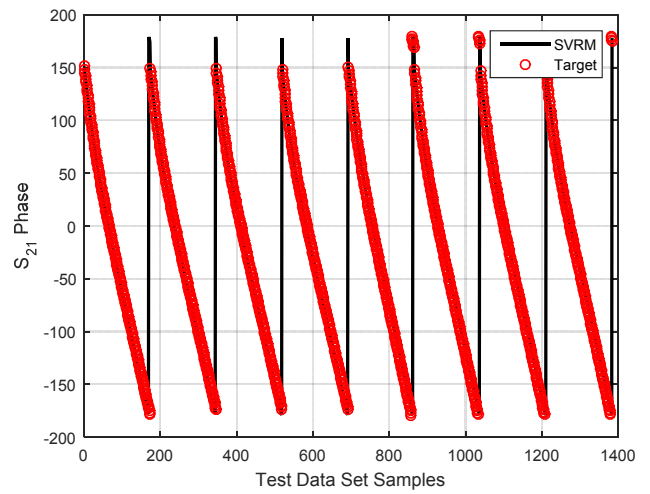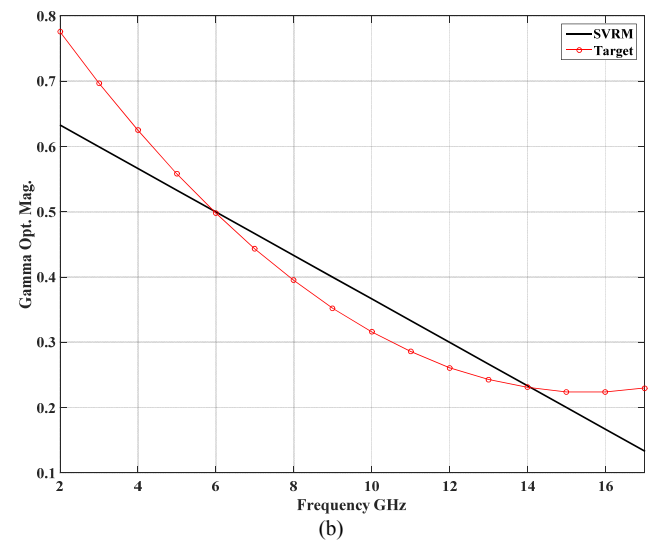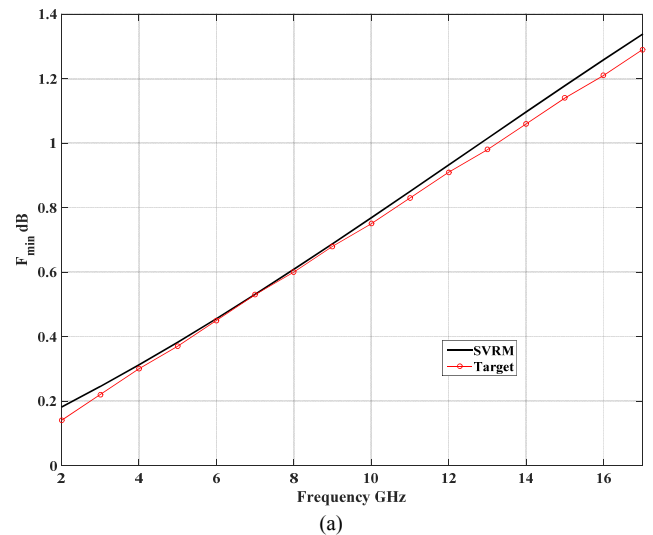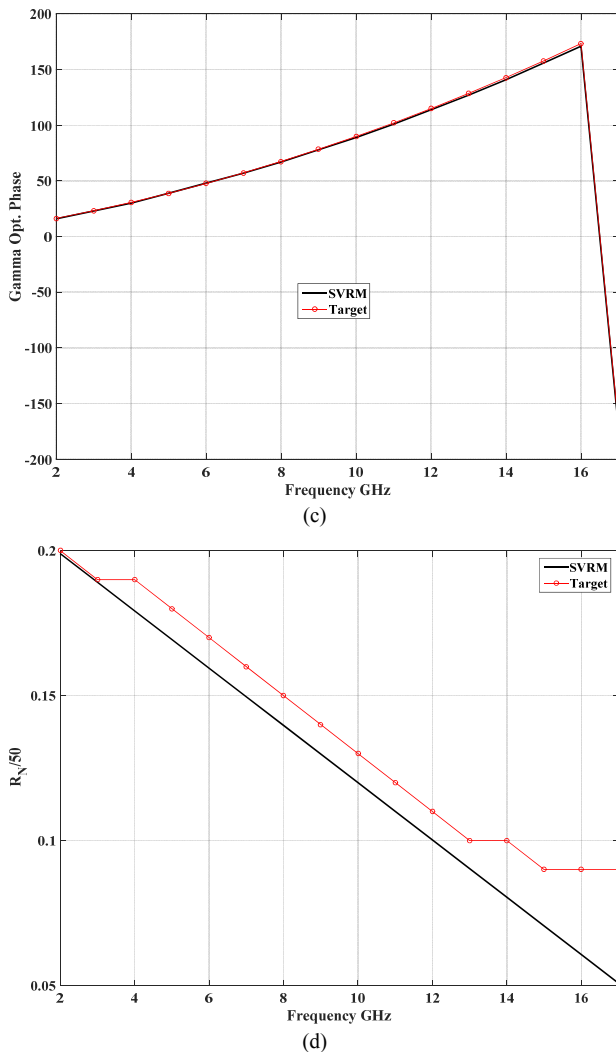**Fig. 9.** S_{11} magnitude throughout the entire test data domain of VMMK1225.

Fig. 10. S_{21} phase throughout the entire test data domain of VMMK1225.

(a)

(b)

(c)

(d)

Fig. 11. Extrapolated prediction of SVRM vs. measurements at the bias condition (4 V, 20 mA) of VMMK1225: (a) F_{min}, (b) $|\Gamma_{opt}|$, (c) Γ_{opt} phase, (d) $R_N/50$.

samples given in Tab. 14, 15 fit snugly into targeted values. Furthermore in Figs. 9 and 10, predicted results of the typical parameters S_{11} magnitude and S_{21} phase are compared with all the test samples belonging to the bias voltages V_{DS} of 2 and 4 V, respectively. In Fig. 11, the predicted noise parameters of VMMK1225 using the characteristic SV–based sparse regression are compared with the target values for bias condition (4 V, 20 mA). Once again, the obtained results suggest that the sparse modeling of microwave transistors with SVRM method is an effective and efficient method for transistors S- and N- parameters.

5. Conclusion

SVRMs are novel linear learning machines in the kernel-induced feature space that combine the convex optimization theory with the generalization, thus they guarantee the global minimum in the optimization procedure. Furthermore the solution has sparseness, in other words the solution can be expressed as a continuous function in terms of an informative subset of the training data so called

Characteristic Support Vectors. This sparseness of solution facilitates a fine and fast sparse regression for suitable regression problems such as modeling of microwave transistors; at the same time it also provides a significant tool to characterize the large amount of data. These two features provide SVRMs distinctive superiorities to the other classification and regression methods, such as commonly used BPNN methods.

In this work first time in the literature, SVRMs are employed successfully as nonlinear extrapolators in the black-box modeling of Scattering and Noise parameters of a microwave transistor as counterparts to the Generalized Regression Neural Network GRNNs [8–10]. Thus this work can be considered mainly in the following significant contribution in the transistor modeling: Magnitude and phase of each characterization S- or N- parameter can be expressed as a continuous function in the throughout device operation domain of (V_{DS}, I_{DS}, f) using only a subset of the reduced training data so-called Characteristic SVs. In the modeling process, the data measured at the typical currents of only a single bias voltage in the middle region of the device operation domain of (V_{DS}/V_{CE}, I_{DS}/I_C, f) is found to be sufficient for training SVRM. Furthermore a detailed error-metric analysis is made between the two nonlinear extrapolators SVRM and GRNN. From this comparative error-metric analysis, it can be observed that SVRM is superior to the counterpart GRNN as the extrapolation performance, furthermore the SVRM extrapolation results in an analytical expression while GRNN predicts on the probability bases. It can be concluded that in this work a revolutionary approach is put forward in the transistor modeling since Characteristic Support Vector based-sparse regression results in an accurate analytical expression using a great amount reduction of training data and human effort.

Acknowledgments

References

[1] ZHANG, Q. J., GUPTA, K. C. *Neural Networks for RF and Microwave Design*. Artech House Publishers, 2000. ISBN-10: 1580531008

[2] GÜNEŞ, F., GÜRGEN, F., TORPI, H. Signal – noise neural network model for active microwave devices. *IEE Proceedings Circuits, Devices and Systems*, 1996, vol. 143, no. 1, p. 1–8. DOI: 10.1049/ip-cds:19960150

[3] GÜNEŞ, F., TORPI, H., GÜRGEN, F. Multidimensional signal-noise neural network model. *IEE Proceedings Circuits, Devices and Systems*, 1998, vol. 145, no. 2, p. 111–117. DOI: 10.1049/ip-cds:19981712

[4] GIANNINI, F. LEUZZI, G., ORENGO, G., et al. Small-signal and large-signal modelling of active devices using CAD-optimized neural networks. *International Journal on RF and Microwave Computer-Aided Engineering*, 2002, vol. 12, no. 1, p. 71–78. DOI: 10.1002/MMCE.10007

[5] MARINKOVIC, Z. D., MARKOVIC, V. V. Temperature-dependent models of low-noise microwave transistors based on neural networks. *International Journal on RF and Microwave Computer-Aided Engineering,* 2005, vol. 15, no. 6, p. 567–577. DOI: 10.1002/mmce.20102

[6] MARINKOVIC, Z., PRONIC-RANCIC, O., MARKOVIC, V. Small-signal and noise modeling of class of HEMTs using knowledge-based artificial neural networks. *International Journal on RF and Microwave Computer-Aided Engineering,* 2013, vol. 23, no. 1, p. 34–39. DOI: 10.1002/MMCE.20631

[7] GÜNEŞ, F., TOKAN, N. T., GÜRGEN, F. Signal - noise support vector model of a microwave transistor. *International Journal on RF and Microwave Computer-Aided Engineering,* 2007, vol. 17, no. 4, p. 404–415. DOI: 10.1002/mmce.20239

[8] GÜNEŞ, F., ÖZKAYA, U., ULUSLU, A. Generalized neural-network based efficient noise modeling of microwave transistors. In *International Symposium on Innovations in Intelligent Systems and Applications* (*INISTA 2011*). Istanbul (Turkey), 2011, p. 212–216. DOI: 10.1109/INISTA.2011.5946067

[9] MAHOUTİ, P., GÜNEŞ, F., DEMİREL, S., et al. Efficient scattering parameter modeling of a microwave transistor using generalized regression neural network. In *20th International Conference on Microwaves, Radar, and Wireless Communication (MIKON)*. Gdansk (Poland), 2014, p. 1–4. DOI: 10.1109/Mikon.2014.6899968

[10] GÜNEŞ, F., MAHOUTİ, P., DEMİREL, S., et al. A cost–effective GRNN–based modeling of microwave transistors with a reduced number of measurements. *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, 2015. Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/jnm.2089

[11] ANGIULLI, G., CACCIOLA, M., VERSACI, M. Microwave devices and antennas modeling by Support Vector Regression Machines. *IEEE Transactions on Magnetics*, 2007, vol. 43, no. 4, p. 1589–1592. DOI: 10.1109/TMAG.2007.892480

[12] CHAUHAN, N., MITTAL, A., KARTIKEYAN, M. V. Support vector driven genetic algorithm for the design of circular polarized microstrip antenna. *International Journal on Infrared and Millimeter Waves*, 2008, vol. 29, no. 6, p. 558–569. DOI: 10.1007/s10762-008-9354-9

[13] GÜNEŞ, F., TOKAN, N. T., GÜRGEN, F. A consensual modeling of the expert systems applied to microwave devices. *International Journal on RF and Microwave Computer-Aided Engineering*, 2010, vol. 20, no. 4, p. 430–440. DOI: 10.1002/mmce.20448

[14] TOKAN, N. T., GÜNEŞ, F. Knowledge-based support vector synthesis of the microstrip lines. *Progress in Electromagnetics Research*, 2009, vol. 92, p. 65–77. DOI: 10.2528/PIER09022704

[15] GÜNEŞ, F., TOKAN, N. T., GÜRGEN, F. A Knowledge-based support vector synthesis of the transmission lines for use in microwave integrated circuits. *Expert Systems with Applications*, 2010, vol. 37, no. 4, p. 3302–3309. DOI: 10.1016/J.ESWA.2009.10.021

[16] XIA, L., MENG, J. C., XU, R. M., et al. Modeling of 3-D vertical interconnect using support vector machine regression. *IEEE Microwave and Wireless Components Letters*, 2006, vol. 16, no. 12, p. 639–641. DOI: 10.1109/LMWC.2006.885585

[17] VAPNIK, V. N. An overview of statistical learning theory. *IEEE Transactions of Neural Networks*, 1999, vol. 10, no. 5, p. 988–999. DOI: 10.1109/72.788640

[18] CRISTIANINI, N., SHAWE-TAYLOR, J. *An Introduction to Support Vector Machines (and Other Kernel-Based Learning Methods).* Cambridge University Press, Cambridge, 2000. ISBN: 0521780195

[19] CAMPS-VALLS, G., GOMEZ-CHOVA, L., CALPE-MARAVILLA, J., et al. Robust support vector method for hyperspectral data classification and knowledge discovery. *IEEE Transactions on Geoscience and Remote Sensing,* 2004, vol. 42, no. 7, p. 1530–1542. DOI: 10.1109/TGRS.2004.827262

[20] KIM, D., CHO, S. Pattern selection for support vector regression based response modeling. *Expert Systems with Applications*, 2012, vol. 39, no. 10, p. 8975–8985. DOI: 10.1016/J.ESWA.2012.02.026

[21] CHANG, C. C., LIN, C. J. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2011, vol. 2, p. 27:1-27:27. Software available at http://Www.Csie.Ntu.Edu.Tw/~Cjlin/Libsvm.

[22] VMMK1225 0.5 to 26 GHz recommended as a low noise E-PHEMT in a wafer scale package. *Datasheet.* [Online] Cited 2016-04-13. Available at: http://www.avagotech.com/docs/AV02-1082EN.

About the Authors ...

Filiz GÜNEŞ received her M.Sc. degree in Electronics and Communication Engineering from the Istanbul Technical University. She attained her Ph.D. degree in Communication Engineering from the Bradford University in 1979. She is currently a full professor in Yıldız Technical University. Her current research interests are in the areas of multivariable network theory, device modeling, computer aided microwave circuit design, monolithic microwave integrated circuits, and antenna arrays.

Mehmet Ali BELEN received his M.Sc. degree in Electronics and Communication Engineering from the Süleyman Demirel University in 2011. He has been currently in Ph.D. program of Yıldız Technical University. His current research interests are in the areas of multivariable network theory, device modeling, computer aided microwave circuit design, monolithic microwave integrated circuits, and antenna arrays.

Peyman MAHOUTİ received his Ph.D. degree in Electronics and Communication Engineering from the Yıldız Technical University in 2016. The main research areas are optimization of microwave circuits, broadband matching circuits, device modeling, computer-aided circuit design, and microwave amplifiers.

Salih DEMİREL has received M.Sc. and Ph.D. degrees in Electronics and Communication Engineering from Yıldız Technical University, Istanbul, Turkey in 2006 and 2009, respectively. He has been currently working as an associate professor in the same department. His current research interests are among of microwave circuits especially optimization of microwave circuits, broadband matching circuits, device modeling, computer-aided circuit design, microwave amplifiers.