Different Methods for Theoretical Modeling and Simulation of Reactive Sputtering Process

K. György and M. Chindriș

Abstract—Depositing thin films of high quality and under precise control is done mainly by a process called sputtering. Reactive sputtering technique used for wear-resistant compound thin films deposition needs an optimized and reproducible process control. Modeling of the dynamic behavior of the reactive sputtering process is a key problem in many respects. A model is important for an active control of the process so that optimal deposition conditions can be maintained. This work is intended to present some theoretical aspects concerning the dynamic modeling and simulation of the process, namely analytically modeling, experimental modeling and the neural network modeling of reactive sputtering process.

Index Terms—identification, modeling, neural networks, nonlinear differential equations, parameter estimation, recursive estimation, simulation, sputtering, vapor deposition, white noise.

I. INTRODUCTION

In the present paper we will discuss some theoretical aspects concerning the dynamic modeling and simulation of reactive sputtering process. The material what will be sputtered, is in the form of a thin plate called target, is mounted in a chamber filled with argon gas at low pressure. The target is connected to negative high voltage, a glowing plasma forms just above the target. Positive argon ions produced in the plasma by collisions of electrons with argon atoms are accelerated toward the target. The fast moving ions collide with target and free atoms of target materials are expelled. The substrate is mounted in front of the target and some of the free atoms land on the substrate, building thin films. Reactive sputtering process is now used to deposit compound materials formed from the reaction between the sputtered target material and a reactive gas. The deposition process has to be stabilized in a specific window between the stable deposition modes at low and high partial pressure of reactive gas, called metallic mode and reactive mode. The problem of reactive magnetron sputtering is in the transition mode is the abrupt change from the metallic mode into the reactive mode accompanied by a marked increase of reactive gas partial pressure as well as a decrease of deposition rate. When measuring pressure–flow curves for increasing and decreasing reactive gas flow, the rate of change of the reactive gas supply strongly influences the width of the hysteresis region. This paper present three methods for modeling this nonlinear process. The first method is the analytically modeling, where is used some theoretically information about dynamic behaviors of process (for example mass flow balance). The analytically obtained dynamical model is a system with nonlinear differential equations when the input-output representation can contain hysteresis loop. This model is difficult to use in process control, so we have to determine some simpler model structures. So we can use simulating dates to identify one parameterized model or one neural network model. In both case (experimental model and neural network system) we will use one black-box model, for different input signal and construction structures.

II. ANALYTICALLY MODELING THE REACTIVE SPUTTERING PROCESS

High deposition rates are of essential importance in reactive magnetron sputtering. In reactive sputtering process the compound thin film is growing up not only on condensing surface but on the target surface itself. The process, named ‘target poisoning’ was theoretically studied by many researchers. A more realistic model was formulated by Berg and his co-workers and more recently by Ershov and Pekker. These models are based on reactive gas conservation in discharge process, describing the steady-state conditions on the target and condensing surfaces. In this model, they have made some simplifying considerations as follows: the ion current of the reactive gas is negligible due to the low concentration of reactive gas; the ion number of sputtered target material in the discharge is negligible small; the ion included secondary electron emission is uniform on both of pure and compound coated target surface; in dynamic equilibrium the flux of reactive gas atoms sticking to the target should be equal to the flux of reactive gas atoms, being removed from the target by sputtering. We have considered a sputtering model following the Berg-Ershov algorithm applied titanium target and N₂-Ar gas mixtures. The basic idea is to consider the three pumping or gettering mechanisms for reactive gas: the vacuum pump, the target and the condensed surface (chamber walls and substrate). The balance of yield the reactive gas is:

\[ q_{in} = q_{cons} + q_{p} = (q_{t} + q_{c}) + q_{p} \]  \hspace{1cm} (1)

where \( q_{in} \) is the mass flow of input reactive gas, \( q_{cons} \) the mass flow consumption in discharge process, \( q_{p} \) mass flow evacuated by pump system, \( q_{t} \) the flux absorbed on the target...
speed of vacuum system, $T$ temperature and $m_N$ is the mass of nitrogen atom, $p_N$ partial pressure of reactive gas. On the level of target and substrate (condensed surface), the flux of titanium and nitrogen atoms sputtered from the target surface and the flux of nitrogen atoms absorbed on the target and the flux of nitrogen gettering to the substrate for dynamic equilibrium in steady state condition is equal. The dynamical model we can start from variation in time of input flow rate:

$$\frac{dq_{in}}{dt} = \frac{dq_{c}}{dt} + \frac{dq_{p}}{dt}$$

If we use the relations from (2) to (4) and consider that $p_N$, $\theta_r$, $\theta_c$ are time variable quantities, then we can obtain one dynamical model, what in general can be written with the following differential equations:

$$\frac{dp_N}{dt} = \Delta q_{in} \cdot f(p_N(t), \theta_r(t), \theta_c(t), J(t));$$
$$\frac{dq_{in}}{dt} = \Delta q_{in};$$
$$\frac{dJ(t)}{dt} = 0;$$
$$R(t) = h(p_N(t), \theta_r(t), q_{in}(t), J(t))$$

where the input value is the variation of input flow $\Delta q_{in}$, the output value is the sputtering rate $R(t)$, the state values are partial pressure of reactive gas $p_N(t)$, the flow rate $q_{in}(t)$ and the ion current density $J(t)$, which is considered constant in time. Theoretically we can write this mathematical model with following general nonlinear form (input-state-output model):

$$\begin{bmatrix}
\frac{dx}{dt} \\
y(t)
\end{bmatrix} = \begin{bmatrix}
F(x(t), u(t), t) \\
G(x(t), u(t), t)
\end{bmatrix}$$

where $x(t)$ is the state vector, $u(t)$ is the input vector and $y(t)$ is the output value. Generally, the state vector is: $x(t) = [p_N(t), q_{in}(t), J(t)]^T$, but we can use one more complicated model where the state vector will be: $x(t) = [p_N(t), q_{in}(t), \theta_r(t), \theta_c(t), q_{in}(t), q_{in}(t), J(t)]^T$. In both case we can choose one output value for example: sputtering rate ($R$). Those mathematical models we can simulate in two cases. In the first case we can consider the input value $u(t) = \Delta q_{in}$ (practical model), in the second case the input value is $u(t) = \Delta p_N$ (theoretical model).

The most basic use of a system description is to simulate the system' response to various input scenarios. For simulation of the dynamical model of sputtering process was created one package program, which permits the simulation of steady state model and simulation of dynamical model. For simulation of dynamical models was used one numerical Runge-Kutta method (used to solve differential equations) and it is possible to change the direction of simulation and the model type. The used initial states value is from the data obtained at steady state model simulation. Theoretical results of simulation of the the theoretical and practical dynamical models described above, are presented in Fig.1 and Fig. 2.

![Fig.1. Sputtering rate ($R$) dependence versus mass flow rate ($q_{in}$) for constant ion current density ($J$), theoretical model ($u=dp_N$)](image)

Parameters used at the simulation are: $k=1.38 \times 10^{-23}$, $S=80.1/s$, $A_B=112.10^{-4}$ m$^2$, $A_A=10$, $At=1$, $\alpha_N=0.01$, $\alpha_M=0.1$, $T=300$ K. The presented simulation results and values will be used to parameter identification process and to neural network modeling (training neural network system).

### III. PARAMETER IDENTIFICATION WITH RECURSIVE METHODS

The analytically obtained dynamical model is a system with nonlinear differential equations when the input-output representation can contain hysteresis loop, so in this case we will use one black-box model, for different input signal and construction structure as: a) identification of input-output model for system when the input is partial pressure and the output is the sputtering rate; b) identification of input-output model for the system when the input signal is variation of input flow and the output is the sputtering rate. In both cases we can also estimate one input-state-output model when the vector states contain the partial pressure of reactive gas, the flow rate of the reactive gas and the ion current density.

The scope of identification methods is to determine the best model in the set guided by the data. The simplest input-output relationship is obtained by describing it as a linear difference
equation (ARX-AutoRegressive model whit eXogeneuos variable):
\[ A(q^{-1}) \cdot y(k) = B(q^{-1}) \cdot u(k-n_k) + e(k) \] (8)
where \( y(\cdot) \) is the output signal, \( u(\cdot) \) is the input signal, and \( e(\cdot) \) is the white-noise term. The \( A(q^{-1}) \) and \( B(q^{-1}) \) are polynomial in \( q^{-1} \) (delay operator)
\[ A(q^{-1}) = 1 + a_1 \cdot q^{-1} + \ldots + a_{ma} \cdot q^{-ma} \] (9)
\[ B(q^{-1}) = b_0 + b_1 \cdot q^{-1} + \ldots + b_{nb} \cdot q^{-nb} \] (10)

If we introduce the vectors:
\[ \theta = [a_1 \ a_2 \ldots a_{ma} \ b_0 \ b_1 \ldots b_{nb}]^T \] (11)
\[ \varphi(k) = [-y(k-1) \ldots -y(k-n_k) \ u(k-n_k) \ldots u(k-n_k-n_b)]^T \] (12)
then the autoregressive model can be rewritten as
\[ y(k) = \varphi^T(k) \cdot \theta + e(k) \] (13)

In this case the least-squares criterion is
\[ V(\theta) = \sum_{k=1}^{N} (y(k) - \varphi^T(k) \cdot \theta)^2 \] (14)
where \( N \) is the number of samples. This criterion is a quadratic function in \( \theta \), therefore it can be minimized analytically, which gives the least squares estimate:
\[ \hat{\theta} = \left[ \sum_{k=1}^{N} \varphi(k) \cdot \varphi^T(k) \right]^{-1} \cdot \left[ \sum_{k=1}^{N} \varphi(k) \cdot y(k) \right] \] (15)

This method can be use for time invariant linear systems and before identification we have to collect large data sets. In many cases it is necessary to have a model of the system available on-line while the system is in operation. The model should then be based on observations up to the current time. The on-line computation of the model must also be done in such a way that the processing of the measurements from one sample can be completed during one sampling interval. Otherwise the whole building cannot keep up with the information flow. Identification techniques that comply with this requirement will be called recursive identification methods, since the measured input-output data are processed sequentially as they become available. We consider the (8) autoregressive model and the weighted least square criterion:
\[ V(\theta(t)) = \sum_{k=1}^{t} \beta(t,k) \cdot (y(k) - \varphi^T(k) \cdot \theta(k))^2 \] (16)
where the \( \beta(t,k) \) is the weighting sequence.
\[ \beta(t,k) = \prod_{j=k+1}^{t} \lambda(j) \] (17)

The estimate of the recursive least-square method will be described by the following algorithm as:
\[ \hat{\epsilon}(k+1) = y(k+1) - \varphi^T(k+1) \cdot \hat{\theta}(k) \]
\[ K(k+1) = \frac{F(k) \cdot \varphi(k+1)}{\lambda(k) + \varphi^T(k+1) \cdot F(k) \cdot \varphi(k+1)} \] (18)
\[ F(k+1) = \frac{1}{\lambda(k)} \cdot (F(k) - K(k+1) \cdot \varphi^T(k+1) \cdot F(k)) \]
\[ \hat{\theta}(k+1) = \hat{\theta}(k) + K(k+1) \cdot \epsilon(k+1) \]

where: \( \epsilon \) is the measurement vector, \( \theta \) is the estimated parameters vector (ARX model) and \( \lambda(t) \) is the weighting factor (0<\( \lambda < 1 \)) and \( F(k) \) is the covariance matrix. To use the recursive algorithms, initial values for their start-up are required (\( F(0) \) and \( \theta(0) \)). The initial conditions can be interpreted so that \( \theta(0) \) is the mean and the \( F(0) \) is the covariance matrix of the prior distribution. This means that the \( \theta(0) \) is what we guess the parameter vector to be before we see the data, and \( F(0) \) reflects our confidence in this guess. An important reason for using adaptive methods and recursive identification in practice is that the properties of the system may be time varying. This handled in a natural way in the weighted criterion by assigning less weight to older measurements that are no longer representative for the system. This relationship (8) can be use for dynamic systems described with input-output model, but a linear system can always represent in state space form. Consider now one linear system with three states variable \( x_1, x_2, x_3 \) and two input variable \( u_1, u_2 \), this system can be characterized by the following linear state space model:
\[ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{bmatrix}_{k+1} = A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{bmatrix}_k + B \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}_k + e_k \] (19)

For autoregressive model we can transform this model as:
\[ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{k+4} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_k + \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}_k + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}_k \] (20)

In this case the measured vector and the parameter matrix are:
\[ \varphi(k) = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ u_1 \\ u_2 \end{bmatrix}_k \]
\[ \theta = \begin{bmatrix} A^T \\ B^T \end{bmatrix} \] (21)
and the parameter estimation can be realized with no recursive (15) and the recursive (18) least-squares methods.

**Simulation results**

In the first step of the investigations the performances of the parameter identifications methods for reactive sputtering process simulations was evaluated, using the Matlab environment. In the first we study the theoretical and practical model and estimate in the first one input-output model (ARX) with recursive method, where the model is dimensioned with \( n_p=3, m_a=2 \) and \( n_b=1 \) parameters. The results are presented in Fig. 3 and Fig.4, where we can observe the white noise perturbed simulated values and the estimated outputs.
The results of recursive identification were obtained with constant forgetting factor and the initial conditions it was adjustment at approximate values, which we obtained with no recursive method. We can represent the model parameter variations in both cases (Fig.5 and Fig.6).

We can observe that in the second case, when the input is \( dq_{in} \), the results of identification are considerable poor, the estimation error is great, it is provided from the abrupt change of the signal output. If we use just a fraction of simulated data (data behind and after abrupt changes) we can improve these results.

IV. NEURAL NETWORK MODELING

Neural networks are very sophisticated modeling techniques capable of modeling extremely complex functions. In particular, neural networks are nonlinear. For many years linear modeling has been the commonly used technique in most modeling domains since linear models have well-known optimization strategies. Where the linear approximation was not valid the models suffered accordingly. Neural networks also keep in check the curse of dimensionality problem that bedevils attempts to model nonlinear functions with large numbers of variables. The analytically obtained dynamical model is a system with nonlinear differential equations, so in this case we will use one black-box model, for different input signal and construction structure as: a). identification of neural network model for system when the input are partial pressure \( p_N \) and ion current density \( J \), and the output is the sputtering rate \( R \); b) identification of input-output model for the system when the input signal are variation of input flow \( q_{in} \) and ion current density \( J \) and the output is the sputtering rate \( R \) or partial pressure of reactive gas \( p_N \).

The first stage of model identification is to train a neural network to represent the forward dynamics of the plant. The prediction error between the plant output \( Y_p \) and the neural network output \( Y_n \) is used as the neural network training signal (Fig.7).

Generally we can say that an assumed relationship among observed signals a model of the system. So the model has to be constructed from observed data. The neural network plant model uses previous inputs and previous plant outputs to predict future values of the plant. For this process was used one feed-forward network, with one hidden layers whit \( N_1 \) number of neuron. The generalized scheme for this model is represented by the Fig. 8, where \( x_1 \ldots x_{N_0} \) are the input signals, \( \omega_{1,0} \) is the weight vector between hidden layers and input layers; \( \omega_{2,1} \) is weight vector between output layer and hidden layers and \( a \) is the output signal of neural network.

Once the number of layers, and number of units in each layer, has been selected, the network's weights and thresholds must be set so as to minimize the prediction error made by the network. This is the role of the training algorithms. This have to automatically adjust the weights and thresholds in order to minimize this error. This process is equivalent to fitting the model represented by the network to the training data available. The error of a particular configuration of the network can be determined by running all the training cases through the network, comparing the actual output generated with the desired or target outputs. The most common error functions are the “sum-squared error”. The best-known example of a neural network training algorithm is back propagation.
Simulation results

In the first step of the investigations the performances of the neuronal network modeling for reactive sputtering process simulations was evaluated, using the Matlab environment. We can start the study, using in the first the theoretic model: neuronal network training when the input set is the simulated value for partial pressure for different ion current density. The training parameters are: number of epochs is 100, the goal is 0. The hidden layer has ten “logsig” neurons, the output layer has one “linear” neuron and the training algorithm is the Levenberg-Marquardt back propagation method. The training performance and the training errors are presented by Fig. 9 and Fig. 10.

This result was obtained for one set inputs when the ion current density is constant, if we want to learn for a few set of inputs then the time learning is considerable great, but de performance is also acceptable. The simulated outputs and the targets for five sets of inputs is presented by Fig. 11.

In the other case we can study neuronal network modeling for mathematical model with qin and J input signals. In this case we have to introduce one supplement input what is represent the direction of simulation. If this input is “1” then signal input qin increasing, if this input is set at “-1” then the input flow is decreasing. In the first we try to train the neuronal network for three set of date, when the inputs is increased. The neuronal network is the same in this case the training errors (Fig.12) and the simulated and the targets output (Fig. 13) are represented by following figure.

We can observe at this graphics the abrupt change from the metallic mode to the reactive mode and the hystersys loop. In conclusion for this result, we can say that this model we can represent with following block diagram represented in Fig.14.
We can observe that this model hasn’t a strong dynamic character, because the output signal depends of the current values of input signals. The variation in time of input signal is present just with direction of simulation, this means that we have to know the early or lately values in time. To eliminate those insufficiencies we can define one neural network model with additional inputs, for example \((q_{in}(k), q_{in}(k-1), \ldots)\). The dynamic character of model will be more emphasis, if this additional input vector has a large dimension. The results of this training neural network model are same, the difference is that this model structure is more reliable.

V. CONCLUSIONS

The present investigations show that the parametric identification methods at nonlinear reactive sputtering process can be used just with some restriction. Success of identification process depends on the type model set (selection between input-output and physically parameterized state space models) and the size of the model (the order of a state-space model or the degrees of the polynomials in a model). To choose a model structure for reactive sputtering process a priori information can be the dynamical model obtained with analytic method. But this model contained nonlinearities, so in this case we will use one black-box model for different input signal. In this paper for measured dates it was used simulated dates with additional white noise signal from numerical simulation of dynamic model. The present investigations show that the neuronal network modeling can be used with success for nonlinear reactive sputtering process. The learning performances depends of the type model set (selection between input-output signals) and the neuronal network structure (for example the number of neuron in the hidden layer). The proposed neuronal network model is constructed with a feed forward multilayered neural-network system for one complex nonlinear system, with know input and output signals.

VI. REFERENCES