Least Squares Support Vector Machines for Modelling Electronic Devices

Karsmakers P.[1,2], Leroux P.[1,3], De Brabanter J. [2,4], Suykens J.A.K.[2]

March 13, 2008

[1]"K.H.Kempen (Associatie K.U.Leuven)", Departement IBW, B-2440 GEEL
Peter.Karsmakers@khk.be, Paul.Leroux@khk.be
[2]"K.U.Leuven", ESAT-SCD/SISTA, B-3001 Heverlee
Johan.Suykens@esat.kuleuven.be
[3]"SCK-CEN", the Belgian Nuclear Research Centre, B-2400 Mol
[4]"KaHo Sint Lieven (Associatie K.U.Leuven)", Departement II, B-9000 Gent
Jos.Debrabanter@kahosl.be

Abstract

The modeling of electronic components is of paramount importance in the area of circuit simulation. In this paper we would like to use data-driven black-box modeling for this purpose, where we assume no prior information whatsoever of the electronic device. Black-box models could be used for both modeling poorly known devices as for model simplification of existing models. It is seen that using the Least-Squares Support Vector Machines (LS-SVMs) as a methodology we can construct a Non-linear AutoRegressive with eXogenous inputs (NARX) model for an electronic device. To asses this methodology for its use in modeling electronic components we describe two applications in the experimental section. The first experiment handles the modeling of an exclusive OR logic gate and the second explains modeling a laser driver exposed to radiation. Both experiments show that accurate and robust models can be constructed with only a relatively small amount of data.
Keywords: *NARX, LS-SVM, identification, black-box modeling.*

1 Introduction

The modeling of electronic components is of paramount importance in the area of circuit simulation. As circuit devices become more and more complex, their performance more time critical and in cases where poorly known physical effects

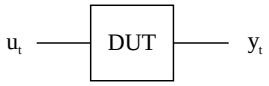

Figure 1: Black-box model

come in to play, the use of conventional idealized and expensive circuit simulation can become ineffective. Especially in fast digital circuits, where parasitic effects are activated, the use of non-ideal models is necessary but due to size restrictions the complexity is limited. Instead of using the transistor level description of gates to formulate a model one could try to find one model for a certain logic gate directly. In this paper we would like to use data-driven black-box modeling for this purpose, where we assume no prior information whatsoever of the electronic device. As shown in Fig. 1 we establish an input signal (or possible more) at our device under test (DUT) and measure the corresponding outputs. Starting from these measurement we construct the black-box model which can be used for future simulations. This type of modeling is the most general approach for poorly known devices whereas the model complexity allows the balancing between accuracy (and generalization to unseen data) and computational efficiency. In this scenario, the Non-linear AutoRegressive with eXogenous inputs (NARX) model structure could provide a useful modeling approach. Such models are an extension of the commonly used ARX models to the non-linear case. For most electronic devices we have to build models which can include non-linear effects. To handle those non-linearities we illustrate the use of Least-Squares Support Vector Machines (LS-SVMs). LS-SVMs are a reformulations to the standard Support Vector Machines (SVMs) [2], which where introduced within the context of statistical learning theory and structural risk minimization, which leads to solving linear KKT systems. Besides regression and classification also problems of unsupervised learning, recurrent networks can be solved within this framework [1]. The use of LS-SVM for modeling dynamical systems has already been studied in applications such as control [1], or load forecasting [8].

This paper is organized as follows. In section 2 we will start by explaining the LS-SVM identification process to build a NARX model. To asses this methodology for its use in modeling electronic components we describe two applications in the experimental section. The first experiment handles the modeling of an exclusive OR logic gate and the second explains modeling a laser driver exposed to radiation.

2 Building a model for an electronic device

To model an electronic device we use a non-linear autoregressive with exogenous inputs (NARX) model structure [4]. In this model structure, the device output characteristics (voltage, current) at a given time step are explained by the evolution of the output at previous time instances as well as by the effect of the exogenous input variables. More formally this results in

$$y_t = f(x_t) + e_t, \tag{1}$$

where f is an unknown smooth function, y_t denotes a voltage or current at time step t, e_t is assumed to be white noise and $x_t \in \mathbb{R}^n$ is the regression vector

$$x_t = [y_{t-1}, ..., y_{t-q}, u_t, u_{t-1}, ..., u_{t-r}], \tag{2}$$

where u_t specifies the input variable of the considered device and the positive integers q, r denote the number of lags of respectively the output and input. The prediction at time t is given by $f(x_t) = y_t - e_t$. In the context of linear system identification, the unknown function f in (1) is identified by parameterizing f as a function of a parameter vector $w \in \mathbb{R}^n$, and then using a data sample $\mathcal{D} = \{x_t, y_t\}_{t=1}^N$ to find an estimate $\hat{w}$. The estimate is usually computed as a solution of an optimization problem [5]. When f is parameterized as a linear function of the form

$$f(x_t) = w^T x_t + b, \tag{3}$$

then the model in (1) is an ARX model, and $\hat{w}, \hat{b}$ are commonly obtained as a solution of a least-squares optimization problem. Linear models assume a linear effect from the input variable x_t to the output y_t. For most electronic devices such a linear relation is not sufficient and we therefor have to build models which can include non-linear effects in function f. To handle those non-linearities we illustrate the use of Least-Squares Support Vector Machines (LS-SVMs) [1] to infer the parameters of the NARX model (1).

2.1 Least Squares Support Vector Machine for non-linear modeling

Least-Squares support vector machines belong to the class of kernel methods that use positive-definite kernel functions to build a non-linear representation of the original inputs in a high-dimensional feature space. Let us consider a simple linear model of the following form

$$f(x_t) = w^T \varphi(x_t) + b, \tag{4}$$

where $x_t \in \mathbb{R}^n$ is the regression vector, $w \in \mathbb{R}^{n_\varphi}$ is an unknown parameter vector, $b \in \mathbb{R}$ is a bias term and $\varphi(\cdot) : \mathbb{R}^n \to \mathbb{R}^{n_\varphi}$ some mapping to a high and possibly infinite dimensional feature space. Using this linear in the parameters formulation

we can use a whole range of well studied stable linear solvers to infer an estimate of w and still find a non-linear relation between x and y because of the non-linear feature map $\varphi(\cdot)$. For example suppose that the feature map $\varphi(\cdot)$ is known in advance, we can first compute our features $z_i = \varphi(x_i)$ and then use linear regression techniques to solve $y_i = w^T z_i + e_i$ using the data set $\{z_i, y_i\}_{i=1}^N$ to find $\hat{w}$. However, in the context of LS-SVMs, the feature map $\varphi(\cdot)$ does not have to be known explicitly but rather is implicitly defined by using a kernel function. This fact is a key element of Support Vector Machines. A kernel is a function from $X \times X \to \mathbb{R}$, where usually $X \subseteq \mathbb{R}^n$. A commonly used kernel function is the Gaussian Radial Basis Function (RBF) kernel given by

$$K(x_i, x_j) = \exp\left(-\frac{||x_i - x_j||_2^2}{\sigma^2}\right), \tag{5}$$

where σ is a hyper parameter. Mercer's theorem shows that for every positive-definite kernel function K, there exists a feature map $\varphi : \mathbb{R}^n \to \mathbb{R}^{n_\varphi}$ from the input space to the feature space such that

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) = \sum_{k=1}^{n_\varphi} \varphi_k(x_i)\varphi_k(x_j). \tag{6}$$

LS-SVM regression estimation involves primal and dual model formulations. Given the training data set $\mathcal{D} = \{x_t, y_t\}_{t=1}^N$ the goal is to estimate the model (1), where f is parameterized as in (4), we can formulate the following optimization scheme to infer our parameters

$$\begin{aligned} \min_{w,b,e} L_P(w, e) &= \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^N e_i^2 \\ &\text{such that} \quad y_i = w^T \varphi(x_i) + b + e_i, \quad i = 1, ...N. \end{aligned} \tag{7}$$

where the regularization constant $\gamma > 0$ is included to control the bias-variance trade-off. The above statement is in fact the same formulation as is used in case of ridge regression [3] in the feature space defined by $\varphi(\cdot)$. However one should be aware of the fact that in some cases w becomes infinite dimensional, such is the case for the Gaussian RBF kernel (5), and therefore the above problem formulation cannot be used to solve the problem. We therefor perform the computations in an other space, called the dual space of Langrangian multipliers after applying Mercer's theorem. Consider the Lagrangian of (7) given by

$$\begin{aligned} L_D(w, b, e_i; \alpha) = &\frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^N e_i^2 \\ &- \sum_{i=1}^N \alpha_i (w^T \varphi(x_i) + b + e_i - y_i), \end{aligned} \tag{8}$$

where $\alpha_i \in \mathbb{R}$ are Lagrangian multipliers. The first order conditions for optimality are:

$$\begin{cases} \frac{\partial L_D}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{N} \alpha_t \varphi(x_i), \\ \frac{\partial L_D}{\partial b} = 0 \rightarrow 0 = \sum_{t=1}^{N} \alpha_t, \\ \frac{\partial L_D}{\partial e_t} = 0 \rightarrow \alpha_t = \gamma e_t, \ \ t = 1, ..., N. \\ \frac{\partial L_D}{\partial \alpha_t} = 0 \rightarrow y_t = w^T \varphi(x_t) + b + e_t, \ \ t = 1, ..., N. \end{cases} \tag{9}$$

Combining the first and the last condition yields

$$y_t = \sum_{i=1}^{N} \alpha_i \varphi(x_i)^T \varphi(x_t) + b + e_t. \tag{10}$$

Now, using Mercer's theorem to replace the dot product $\varphi(x_i)^T \varphi(x_t)$ yields

$$y_t = \sum_{i=1}^{N} \alpha_i K(x_t, x_i) + b + e_t, \tag{11}$$

for a certain positive-definite kernel function $K(\cdot, \cdot)$. Replacing the third expression from (9) into (11) gives

$$\begin{aligned} y_t &= \sum_{i=1}^{N} \alpha_i K(x_t, x_i) + b + \frac{\alpha_t}{\gamma}, \ \ t = 1, ..., N, \\ 0 &= \sum_{t=1}^{N} \alpha_t, \end{aligned} \tag{12}$$

with unknowns $\alpha_t, \ \ t = 1, ..., N$, and b. If we define the kernel matrix $\Omega_{ij} = K(x_i, x_j)$ and write the previous two equations in matrix form gives the linear system

$$\begin{bmatrix} \Omega + \frac{1}{\gamma} I & 1_N \\ 1_N^T & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix}, \tag{13}$$

where $\alpha = [\alpha_1, ..., \alpha_N]^T$ are the dual variables, $y = [y_1, ..., y_N]^T$, I is the identity matrix and 1_N is a vector containing N ones. Solving for α and b yields the estimate for f expressed in dual form

$$f(x_t) = \sum_{i=1}^{N} \alpha_i K(x_i, x_t) + b. \tag{14}$$

2.2 Parameter selection

Besides our estimation of the dual variables α one also has to consider the tuning of the kernel function K, a regularization constant γ, which are called hyper parameters and the determination of the number of lags which are necessary to correctly

model the underlying device. This tuning of course is critical in the process of obtaining a good model.

When solving system (13) a certain choice for a kernel function K and a chosen regularization constant γ is made. The LS-SVM training process additionally consists of doing an optimal selection of the kernel parameters (σ in case of an RBF-kernel) and γ. These parameters are chosen in such a way that the model shows a good generalization ability. Formally, we try to maximize the performance of the model on unseen data, not used for the training stage (out of sample data). Hence, we train our model with some initial hyper parameter values on the training set and sequentially compute the Mean Squared Error (MSE) of the residuals, $\hat{e}_t = y_t - \hat{y}_t$, of the model on the unseen validation data. By iteratively adapting the hyper parameters we try to minimize this cost using some optimization scheme. In our simulations we use the Nelder-Mead simplex [6] optimization method for this purpose. The MSE can be computed on a separate validation set or can be calculated using cross validation as is done frequently in the machine- and statistical learning community [7]. Cross validation works by dividing the training set into m parts. The model is estimated using $m-1$ parts, and the predictive performance is assessed on the remaining part by using the MSE. The process is repeated m times and the performance measurements are averaged.

To assess the quality of the model (using different numbers of lags) we check the prediction performance of the model using the MSE of the residuals on a different independent validation set. We start by predicting the first out-of-sample value, next to predict the second out-of-sample value we use the first prediction, and continue by iterative simulation.

3 Experiments

3.1 NARX EXOR model

In this section we address the modeling of an EXOR module built in Metal Oxide Semiconductor (MOS) technology using the technique described in the previous sections. We generate output signals for the identification process via SPICE simulations based on the HSPICE BSIM level 13 model, which is a detailed MOS model including parasitics. In our first experiment we used the Exclusive OR logic component, build out of 4 NAND gates see Fig. 2. To simplify the experimental setting we consider a Single Input Single Output (SISO). We accomplish this kind of configuration by keeping one EXOR input at a constant voltage. The output of the module is loaded with a capacitor of 0.01pF. Hence, we have voltages at the gates input and output ports as the input (u_t) and the output (y_t) signals, respectively. The simulated outputs are used either directly, which could be useful for model simplification, or corrupted by noise, to reproduce identification from

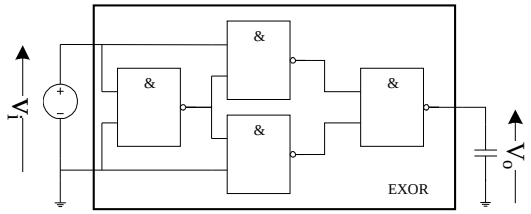

Figure 2: The first experiment tries to model an Exclusive OR logic component, built out of 4 NAND gates. In order to simplify the experimental setting we only consider one varying input while the other is set to the reference voltage. At the output we used a load capacitor with a capacitance of 0.01pF.

measured data.

To start our experiment we have to define an identification sequence which perturbs the device in such a way that a proper identification is possible. Although we will not go into detail, we notice that the selection of a suitable identification signal is a critical element in the identification of a non-linear system. Because we assume digital signals at the input of the gate we opted for a Pseudo Random Binary Sequence (PRBS) at $1.33e9$ bps. We kept the number of samples low on purpose to show the ability of the identification method to generalize well with only a limited number of samples, although more data probably would yield a better accuracy. In numbers the identification sequence has 200 samples for the training process and 100 additional instances for the model selection process, the level variations are wide enough to highlight the non-linear nature of the electronic component. Among all the identified models, the one with the smallest MSE on a separate test set is the one with the most faithful static and dynamic behavior and can be considered the final model. In Fig. 3 we have plotted the MSE on the test sequence for the different models as is seen in the figure we obtained $r = 7, q = 9$ which are respectively the number of lags of the input and output. The response of the model on a third separate test sequence is shown in Fig. 4. On the same PC it took 0.43s to compute a sequence of 2000 samples in HSPICE 2002.2.2 while using the final NARX model the output was given after 0.035s in MATLAB. In order to check the robustness of the identification methodology to corrupted output sequences we perform a series of experiments where we superimpose white noise with different variance on the simulated output signal generated by HSPICE and use the resulting signal for identification. In this way, we observe that the identification process up to certain Signal to Noise Ratio (SNR) is rather insensitive to the added noise. In Fig. 6 the MSE on a validation sequence is plotted in function of the SNR of the HSPICE output signal.

The identification process is hardly affected by the noise as long as the SNR has values larger than 20dB. If we apply an SNR with a lower value then the behavior will differ more from the original value. This is seen in Fig. 7 and Fig. 5, where

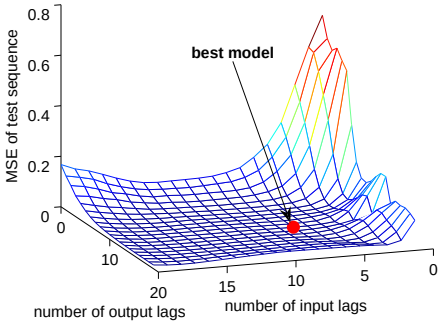

Figure 3: In this figure the MSE of the different NARX models, with different number of lags for input (r) and output (q), are shown. The one with the lowest MSE is considered the most faithful model.

respectively the noisy output identification sequence and the response of the model identified from such a noisy sequence is shown. In this example, the output identification sequence has an SNR of 12.3 dB, and the response of the identified model to a test sequence still tracks the reference output fairly well. The extension to a Multiple Input Multiple Output (MIMO) model is straightforward. For example, our EXOR model can be extended by adding the second input and the output current to the regression vector x_t.

3.2 Modeling a laser driver exposed to radiation

In this example we consider a real world data set where a VCSEL (Vertical Cavity Surface Emitting Laser) driver IC, implemented in a 0.7μm CMOS technology, was tested before and after exposure to a gamma radiation dose of 3.7kGy.

The effect of ionizing radiation on CMOS integrated circuits has been thoroughly described in literature [9]. On one hand, radiation damage is incurred in the ubiquitous silicon oxide (SiO2) material. On the other hand, chemical bond rupture at the Si-SiO2 interface creates amphoterous interface states. For a PMOS transistor these two effects make the negative threshold voltage decrease further under radiation. For a NMOS the positive gate oxide charges cause an initial decrease of the threshold voltage, which is followed by a subsequent increase, caused by the charged interface states.

The field oxide suffers from the same degradation mechanism as the gate oxide. The deposited positive charge may be sufficient to create inversion in the underlying p-substrate, possibly forming a parasitic conductive leakage channel between two n-doped regions [10].

For the design of the radiation tolerant VCSEL driver, the radiation effects de-

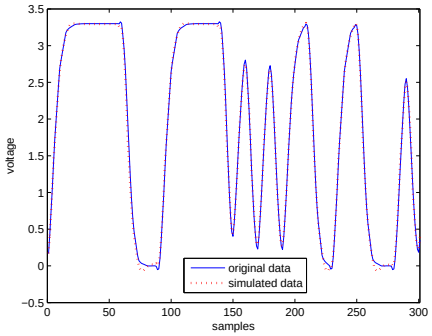

Figure 4: The response of the model on an independent test sequence is shown and compared to the response generated by the HSPICE model.

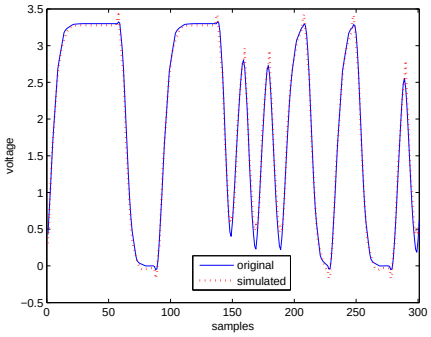

Figure 5: The same independent test set as is shown in Fig. 4 is used to compare the response of the NARX model (trained on corrupted measurement data) to the response generated by the HSPICE model.

scribed above are handled at both schematic level and layout level. The radiation induced shifts in the MOS transistor threshold voltage, which would cause a drift in the VCSEL current, are tackled with a current correcting feedback mechanism. The parasitic leakage is handled at the layout level [11].

The high level circuit schematic of the laser driver is shown in Fig. 8. The driver core is a differential amplifier which is compatible with both differential and single-ended TTL (Transistor-Transistor Logic) and low-voltage TTL inputs. The driver converts the input bits into a pulsed current pattern for the VCSEL diode. The value of this current is determined by the biasing of the driver. A control loop continuously monitors the forward current through the VCSEL and adjusts the bias of the driver in order to keep this current constant. This control is implemented via an exact replica of the driver since the pulsed current of the driver itself would complicate a direct measurement. Capacitor C_{fb} creates a dominant Miller pole to stabilize the feedback loop. It also determines the speed of the control loop, but since the only function of the loop is to correct for very slow radiation or temperature induced shifts in the device characteristics, this time constant is not critical. For more details on the design and characterization of the driver, the reader is referred to [12].

The aim of the developed model in this example is to estimate the eye diagram of the output signal without requiring the entire response to the applied PRBS signal. The model also deduces the main quality factors of the eye, like the vertical eye opening and the horizontal jitter value.

During the measurement the driver is perturbed by a 127-bit PRBS signal at a bitrate of roughly 155Mbps. This device has one input (V_{in}) and output (I_{out}). To

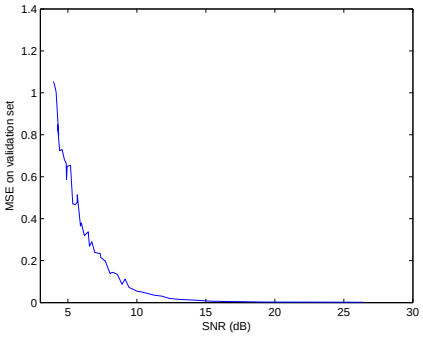

Figure 6: The MSE on a validation sequence is plotted in function of the SNR of the HSPICE output signal.

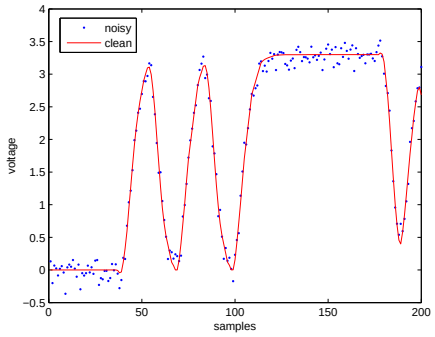

Figure 7: This figures shows an example of an output sequence with additive white noise. The SNR is 12.3 dB

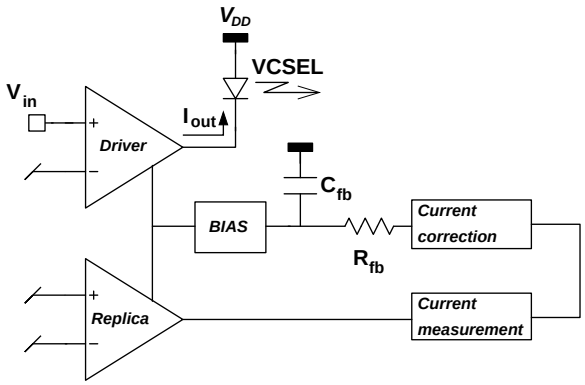

Figure 8: High-level circuit schematic of the VCSEL driver

train the model we used the k first bits out of the PRBS. The hyper parameters were tuned using 10-fold cross validation and the assessment of the models is made on 10 validation bits (not used during the training stage). The model characteristic values extracted from the simulated data are represented in Table 1. In the table it is seen that we already obtain a good accuracy when we are using 40 bits. In this case we obtained $r = 3, q = 27$ which are the number of lags of the input and output after the model selection step. The final eye diagram is constructed using the remaining 77 bits. In Fig. 9 it is seen that the model gives an accurate prediction of the eye diagram.

By limiting the number of required bits in the measurement a significant time saving may be realized. This becomes even more important for larger PRBS sequences (up to $2^{31} - 1$ bits) which are often required by the system specifications of optical transmission links. Moreover, a $2^{31} - 1$ PRBS is not available in every data generator. In this case, it could be interesting to investigate whether the model could

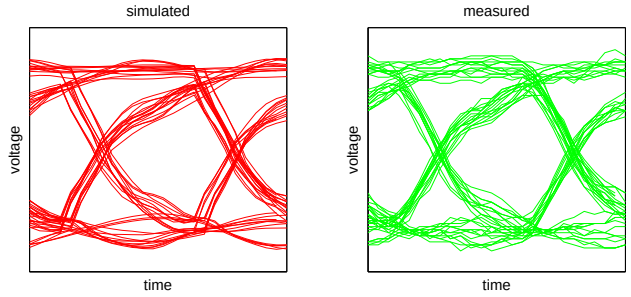

Figure 9: This figure shows the eye diagram of the laser driver which processes digital data at a bitrate of 154.95e6 bps. The output of the NARX model (left) on an unseen random bit sequence is compared to the measured output data (right) of the laser driver.

Table 1: This table shows in the first column the mean squared error on a independent test set, in the second column the standard deviation (std) of the jitter J_s and in the last column the vertical opening in the middle of the eye denoted by V_s for data which is simulated from a model constructed from different training set sizes. As a reference the standard deviation of the jitter and vertical eye opening for the measured data is $std(J_m)$ =1.94e-10 and $V_m = 1.56$.

	MSE	$std(J_s)$ [s]	V_s [V]
10 bits	0.6699	7.11e-10	0.46
20 bits	0.0877	2.5e-10	1.36
30 bits	0.0533	2.3e-10	1.41
40 bits	0.0383	2.22e-10	1.45
50 bits	0.0313	2.15e-10	1.46
60 bits	0.0293	2.11e-10	1.48
70 bits	0.0261	2.07e-10	1.49

be used to estimate the performance of the circuit in response to a $2^{31} - 1$ PRBS, based on its response to a significantly lower PRBS.

4 Conclusion

In this paper we investigated the use of Least-Squares Support Vector Machines for the use of modeling electronic devices. In our experiments we have shown that this identification method can produce accurate and robust models. Which allows for both modeling poorly known devices as for model simplification of existing models. For the laser data set we have shown that by limiting the number of required bits in the measurement a significant time saving may be realized.

5 Acknowledgments

Research supported by GOA AMBioRICS, CoE EF/05/006; (Flemish Government): (FWO): PhD/postdoc grants, projects, G.0407.02, G.0197.02, G.0141.03, G.0491.03, G.0120.03, G.0452.04, G.0499.04, G.0211.05, G.0226.06, G.0321.06, G.0553.06, G.0302.07. (ICCoS, ANMMM, MLDM); (IWT): PhD Grants,GBOU (McKnow), Eureka-Flite2 - Belgian Federal Science Policy Office: IUAP P5/22,PODO-II,- EU: FP5-Quprodis; ERNSI; - Contract Research/agreements: ISMC/IPCOS, Data4s, TML, Elia, LMS, Mastercard. JS is a professor.

References

[1] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor and J. Vandewalle, *Least Squares Support Vector Machines*, World Scientific, Singapore, 2002.

[2] V. Vapnik, *Statistical Learning Theory*, Wiley and Sons, 1998.

[3] G.H. Golub, C.F. Van Loan, *Matrix Computations*, Baltimore MD: Johns Hopkins University Press, 3th edition, 1996.

[4] J. Sjöberg, Q. Zang, L. Ljung, A. Benveniste, B. Deylon, P. Glorennec, H. Hjalmarsson, A. Juditsky, "Non-linear black-box modelling in system identification: A unified overview", "*Automatica*", vol 31 (12), pp. 1691-1724, 1995.

[5] L. Ljung, *System Identification: Theory for the User*, Englewood Cliffs, NJ:prentice hall, 1987.

[6] J. A. Nelder, R. Mead, "A simplex method for function minimization", *Computer Journal*, vol. 7, 308-313, 1965.

[7] G. Golub, M. Heath and G. Wahba, "Generalized cross-validation: A method for choosing a good ridge regression parameter", *Techonometrics*, vol. 21, pp. 215-223, 1979.

[8] M. Espinoza , J.A.K. Suykens, R. Belmans, B. De Moor, "Electric Load Forecasting - Using kernel based modeling for nonlinear system identification", *IEEE Control Systems Magazine, Special Issue on Applications of System Identification*, vol. 27(5), pp. 43-57, 2007.

[9] T. R. Oldham and F. B. McLean, "Total Ionizing Dose Effects in MOS Oxides and Devices", *IEEE Transactions on Nuclear Science*, vol. 50, no. 3, pp. 483-499, June 2003.

[10] F. Faccio and G. Cervelli, "Radiation-Induced Edge Effects in Deep Submicron CMOS Transistors", *IEEE Transactions on Nuclear Science*, vol.52, no. 6, pp. 2413-2420, December 2005.

[11] W. J. Snoeys, T. A. P. Gutierrez, G. Anelli, "A new NMOS layout structure for radiation tolerance", *IEEE Transactions on Nuclear Science*, vol. 49, no. 4, pp. 1829-1833, August 2002.

[12] P. Leroux, S. Lens, M. Van Uffelen, W. De Cock, M. Steyaert, F. Berghmans, "Design and Assessment of a Circuit and Layout Level Radiation Hardened CMOS VCSEL Driver", *IEEE Transactions on Nuclear Science*, vol. 54, pp. 1055-1060, August 2007.