

Received 17 December 2023; revised 29 February 2024; accepted 19 March 2024. Date of publication 21 March 2024; date of current version 19 April 2024.  
The review of this article was arranged by Editor K. Joshi

Digital Object Identifier 10.1109/JEDS.2024.3380572

# Machine Learning-Based Modeling of Hot Carrier Injection in 40 nm CMOS Transistors

XHESILA XHAFA<sup>1</sup>, ALI DOĞUŞ GÜNGÖRDÜ<sup>2</sup>, AND MUSTAFA BERKE YELTEN<sup>2</sup> (Senior Member, IEEE)

<sup>1</sup> LIRMM, CNRS, 34000 Montpellier, France

<sup>2</sup> Department of Electronics and Communications Engineering, Istanbul Technical University, 34467 Istanbul, Turkey

CORRESPONDING AUTHOR: M. B. YELTEN (e-mail: yeltenm@itu.edu.tr)

This work was supported by the Technological Research Council of Turkey through Project TÜBİTAK 1001 under Grant 118E253.

**ABSTRACT** This paper presents a machine-learning-based approach for the degradation modeling of hot carrier injection in metal-oxide-semiconductor field-effect transistors (MOSFETs). Stress measurement data have been employed at various stress conditions of both n- and p-MOSFETs with different channel geometries. Gaussian process regression algorithm is preferred to model the post-stress characteristics of the drain-source current, the threshold voltage, and the drain-source conductance. The model outcomes have been compared with the actual measurements, and the accuracy of the generated models has been demonstrated across the test data by providing the appropriate statistics metrics. Finally, case studies of degradation estimation have been considered involving the usage of machine-learning-based models on transistors with different channel geometries or subjected to distinct stress conditions. The outcomes of this analysis reveal that the established models yield high accuracy in such contexts.

**INDEX TERMS** Reliability, hot carrier injection, HCI, machine learning, ML, Gaussian process regressions, integrated circuits.

## I. INTRODUCTION

Integrated circuits have become an essential part of modern technology. Metal-oxide-semiconductor field-effect-transistors (MOSFETs) scale down in size at each process node so that the capabilities of integrated circuits grow incessantly. Nevertheless, as the scale of transistors comes down, their structure and performance become more vulnerable to degradation. Long-time exposure to large vertical and lateral electric fields causes the generation of traps near the semiconductor-oxide interface that reduce the charge carrier mobility and increase the threshold voltage,  $V_{TH}$ . This process is called the time-based degradation or aging of the device [1].

Different approaches have been proposed to combat the adverse effects of transistor aging at device, circuit, and system levels [2]. To find the right measure, developing a model for aging is essential. Physics-based models were proposed to describe the dynamics of prominent aging mechanisms, such as negative-bias-temperature instability (NBTI) and hot-carrier injection (HCI) [3]. Process variations

and quantum mechanical phenomena (such as gate-oxide tunneling) become intertwined in short-channel MOSFETs with reliability degradation [4]. Hence, the deterministic nature of device aging is replaced by stochastic processes [5], for which semi-empirical or black-box modeling approaches can be preferred.

Therefore, this paper uses machine learning methods to model HCI in n- and p-type MOSFETs belonging to a commercial 40 nm bulk CMOS process with a supply voltage of  $V_{DD} = 1.1$  V. Although the models will be developed based on radio-frequency (RF) transistors that are meant for high-frequency applications, the approach could be employed on any transistor type. Aging in this technology is known to cause reduced circuit performance, mainly in the voltage gain and frequency response [6]. Degradation in transistor quantities, such as the drain current and the threshold voltage, will be captured via Gaussian Process Regression. There have already been initial attempts where machine learning methods were instrumental in capturing NBTI [7], [8], [9], and HCI [8], [10] degradation impact.

The main contributions of the paper can be listed as (1) utilizing individual data points in transistor  $I$ - $V$  characteristics to elevate the accuracy of machine learning tools through larger data size, (2) estimating the degradation amount of transistors with different geometries and threshold voltages to better account for the increased impact of process variations in short-channel device technologies, (3) performing different degradation estimation exercises with the developed models to assess the capabilities of machine learning-based models (4) developing for the first time dedicated p-HCI models for p-MOSFETs that are important for analog and RF integrated circuits.

The paper is organized as follows: Section II summarizes the fundamentals of HCI and machine learning. The modeling approach and results are described in Sections III and IV, respectively. Finally, conclusions are drawn in Section V.

## II. HCI AND MACHINE LEARNING

### A. HOT CARRIER INJECTION (HCI)

According to the classical HCI framework, as electrons are accelerated toward the drain in an n-MOSFET, they acquire sufficient kinetic energy to ionize lattice atoms in case of a collision [2]. More electrons are generated through this impact ionization, some of which attempt tunneling through the gate oxide. A number of these may get stuck in the oxide layer during tunneling and attract positive charges in the semiconductor to the surface that increase  $V_{TH}$ . Moreover, tunneling electrons break the bonds at the Si/SiO<sub>2</sub> interface, thereby creating traps that reduce the charge carrier mobility and reduce the drain-source current ( $I_{DS}$ ) as well as the device transconductance,  $g_m$ .

As the down-scaling in transistor channel geometries continues, new mechanisms causing HCI are discovered, which suggest that HCI is motivated by the defects generated via Si – H bonds breaking apart, similar to the situation in bias-temperature instability [11]. The difference is that in HCI, different vibrational modes of Si – H bonds are triggered based on the carrier energy level, and the resulting resonance leads to the rupture of bonds [12].

Recent studies in the literature aim to address the increased variability problem related to short-channel devices. In [13], variability of the charge carrier mobility and  $V_{TH}$  degradation due to HCI has been inquired, and the trap concentration dependence on the transistor channel geometry has been quantified. A semi-empirical model based on a physical relationship has been proposed; however, this model is not geared toward the use of circuit designers. In another study, a semi-empirical  $V_{TH}$  degradation model depending on multiple kinds of traps was presented [14]. Though this model is easier to employ in circuit design, it does not involve the channel geometry variation, suggesting that the impact of process variations was not properly accounted for.

As shown in literature [15], the largest degradation in short-channel process nodes occurs when the transistor experiences  $V_{STR} = V_{GS} = V_{DS}$ , where  $V_{GS}$  and  $V_{DS}$  are

the gate-source and drain-source voltages, respectively, and the bulk terminals are kept at the ground level. Finally, p-HCI observed in p-MOSFETs is less harmful than n-HCI since holes cannot attain sufficient energy for trap generation due to their heavier effective mass. Nevertheless, as the channel lengths scale down, the p-HCI impact also grows proportionally.

### B. MACHINE LEARNING (ML)

In the last two decades, increased computing power and the abundance of data have led to the unprecedented surge of machine learning (ML) in science and engineering [16], [17]. The main objective of this paper has been to explore how machine learning can be used to model time-based degradation phenomena and assert the accuracy of these models. For that purpose, a large amount of data must be acquired. Hence, a test chip consisting of differently-sized n-MOSFETs and p-MOSFETs has been designed [18]. The chip was then put under accelerated HCI stress conditions to observe performance degradation.

## III. HCI MODELING METHODOLOGY THROUGH ML

This section discusses the modeling of HCI phenomena in terms of transistor performance degradation. The approach adopted can be expressed as  $\mathbf{M} = \{(x_i, y_i)\}_{i=1}^N$ , where  $\mathbf{M}$  is the training set (matrix),  $x_i$  is the input data,  $y_i$  is the output (in this case, the post-stress drain/source current) and  $N$  is the number of samples. Performance degradation has been evaluated using three important transistor characteristics indicators,  $I_{DS}$ ,  $V_{TH}$ , and the drain-source conductance  $g_{on}$ . A separate model has been generated for each of these quantities using the data obtained during the HCI stress and measurement steps. When compared to the traditional physics-based analytical modeling [19], the ML approach puts less emphasis on micro-scale HCI damage (e.g., trap generation) but provides better coverage of macro-scale degradation on transistor performance characteristics subject to process variations.

### A. STRESS CONDITIONS AND DATA NORMALIZATION

To develop a model that describes and predicts the behavior of 40 nm CMOS RF transistors, stress tests at seven different conditions have been carried out on five differently-sized n- and p-MOSFETs. The stress conditions and the chosen device geometries can be seen in Table 1. Here, stress voltages are aimed to cause appreciable degradation while respecting the transistor breakdown voltages. Channel geometries comply with their typically preferred values in analog and RF integrated circuits. Finally, stress times are determined to represent different levels of HCI damage on devices.

The temperature has been kept at  $T = 105$  °C during the stress periods. The reason for this choice stems from the fact that Si – H bond-breaking activity increases at elevated temperatures due to increased charge carrier energies [20].

**TABLE 1. HCI stress conditions and test device sizes.**

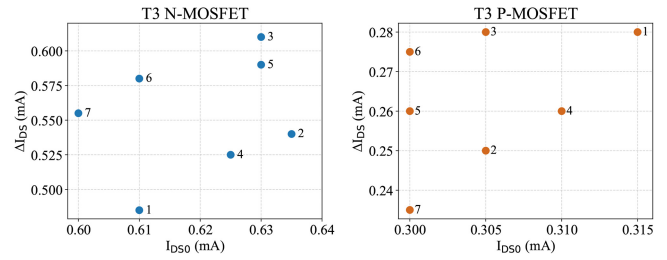
	$V_{STR}$	$ V_{BS} $	$t_{STR}$ (s)	Device Sizes
1.	1.8 V	0 V	3600	$(W/L)_{T_1} = 9.6 \mu\text{m}/60 \text{ nm}$
2.	1.9 V	0 V	3600	$(W/L)_{T_2} = 19.2 \mu\text{m}/60 \text{ nm}$
3.	2 V	0 V	3600	$(W/L)_{T_3} = 76.8 \mu\text{m}/60 \text{ nm}$
4.	1.9 V	0.3 V	3600	$(W/L)_{T_4} = 19.2 \mu\text{m}/120 \text{ nm}$
5.	1.9 V	0 V	1800	$(W/L)_{T_5} = 76.8 \mu\text{m}/120 \text{ nm}$
6.	1.9 V	0 V	7200	
7.	1.9 V	0.3 V	1800	

So, 105 °C was chosen in accordance with this relationship. Later, the post-stress measurements were performed at room temperature,  $T = 25$  °C, for two reasons. First of all, it is known that the bias temperature instability (BTI) effect is also triggered at high temperatures. So, in order to distinguish between these two competing mechanisms,  $I_{DS}$  measurements were recorded at a lower temperature so that BTI degradation reduces due to recovery; hence, the only remaining degradation is due to HCI. The other reason is that the typical operation temperature for these devices is room temperature, so the natural choice is 25 °C.

The variables used in the models are (1) the stress voltage  $V_{STR}$ , (2) the bulk-source voltage  $V_{BS}$ , (3) the stress time  $t_{STR}$ , (4) the width and (5) length of the transistors, (6)  $V_{GS}$ , and (7)  $V_{DS}$  during post-stress measurements. Moreover, for each transistor, the  $I_{DS} - V_{GS}$  and  $I_{DS} - V_{DS}$  characteristics before stress (time-zero) are measured so that the initial  $I_{DS}$  for the device under test,  $I_{DS0}$ , can be included as the variable (8), which represents the process characteristics of the transistor. The output of the constructed model is the post-stress  $I_{DS}$  measured at  $T = 25$  °C, building the ninth column in the input data matrix  $\mathbf{M}$ .

As many as 30464 combinations of the first eight columns, along with the corresponding current measurement in the ninth column, are designated by leveraging the alternatives depicted in Table 1.  $V_{GS}$  and  $V_{DS}$  are varied between 0 and  $V_{DD}$ , whereas  $I_{DS0}$  can assume any value depending on the impact of process variations on the particular transistor. Furthermore,  $V_{STR} = V_{GS} = V_{DS}$  has been adopted for the worst-case degradation conditions. Two input matrices,  $\mathbf{M}_N$  and  $\mathbf{M}_P$ , of the size  $30464 \times 9$  are constructed that will be used toward modeling the  $I_{DS} - V_{GS}$  characteristics of n-MOSFETs and p-MOSFETS, respectively so that n-HCI and p-HCI degradation could be properly captured. Each row of these matrices is a sample for which data is acquired from the test transistors that are chosen from several different chips belonging to separate wafers with distinct process variation characteristics. Extensive graphical representation of these samples in the form of pre- and post-stress  $I_{DS} - V_{GS}$  and  $I_{DS} - V_{DS}$  characteristics could be visualized in [18].

Fig. 1 depicts the observed HCI degradation amounts of  $I_{DS}$  for  $T_3$  (both the n-MOSFET and the p-MOSFET) with  $(W/L)_3 = 76.8 \mu\text{m}/60 \text{ nm}$  in  $I_{DS}$  corresponding to the seven stress conditions, as shown in Table 1. Time-zero currents ( $I_{DS0}$ ) of both transistors vary by about 5%, which



**FIGURE 1.  $\Delta I_{DS}$  vs.  $I_{DS0}$  for  $T_3$  (both the n-MOSFET and the p-MOSFET) under the seven HCI stress conditions, as presented in Table I.**

quantifies the extent of the process variations recorded for those transistors in this data set.

Normalization is important when preparing the data for training since the input variables are often in different numerical scales and ranges. For example, the transistor channel length ( $L$ ) is at the order  $10^{-9}$  m while the stress time is in the range of  $10^3$  seconds. This imbalance affects learning performance negatively. Normalization creates a new input data set, which still retains the statistics of the initial data but removes the numerical complications that can adversely impact the accuracy of the model. Hence, the standard score normalization method has been applied to all available data.

The relationship between the input (voltage) and output (current) is exponential. Taking the logarithm of the output data yields a better modeling performance since the learning algorithm does not have to learn the extra nonlinearity arising from the exponential nature, thereby making the model less complex and more accurate. Also, it avoids the problem of getting negative predicted output values, which would not make sense in the case of the drain current output. Moreover, to avoid undefined results for zero currents in logarithmic expressions, all current values have been shifted up by a constant of 0.01 before training and then shifted down by the same amount to correctly extract the predicted current values.

## B. GAUSSIAN PROCESS REGRESSIONS

The approach for machine-learning-based modeling of HCI degradation should be chosen according to the distribution characteristics of transistor properties. Traditionally, process parameters of MOSFETs follow the Gaussian distribution. Hence, the Gaussian process regression (GPR), a Bayesian regression approach, is a natural and suitable choice since it involves several random variables sharing a joint Gaussian distribution [21]. Each of these random variables will have a different weight based on their impact on the model outcome that will be shaped by hyperparameters. These hyperparameters are used in different kernel functions (covariance functions) and optimized to yield the lowest possible training and test errors based on the outcomes of the employed training and test samples via a modeling tool [22]. In this work, MATLAB<sup>®</sup> regression learner toolbox has

been adopted to determine the optimal values of these hyper-parameters where a plethora of learning algorithms can be simultaneously trained. The training has been performed for different kernel functions, such as rational quadratic, squared exponential, matern 5/2, and exponential. Results have been compared, and the best-performing kernel functions are found as the rational quadratic and squared exponential, of which the latter choice has been adopted in this work.

#### IV. MODELING OF THE DEGRADED CURRENT DUE TO HCI

The proposed modeling approach starts by acquiring the  $I_{DS} - V_{GS}$  characteristic, from which the degradation in the drain saturation current,  $I_{Dsat}$ , can be extracted. Then,  $V_{TH}$  of the transistors is modeled and extracted based on predictions. Lastly,  $g_{on}$  will be captured as an indicator of mobility degradation. The reason for these multiple models stems from the need for higher accuracy. As the input design space of a machine-learning-based model grows, the modeling tools strive for a balanced error across the whole range of input parameters. Hence, the overall modeling error rises. However, for device quantities, such as  $V_{TH}$  and  $g_{on}$ , the model inputs can be chosen from only the region mandatory for the model. For example, the  $I_{Dsat}$  model covers the full range of gate-source voltages extending from 0 to  $V_{DD} = 1.1$  V, which is much broader than the necessary interval to extract  $V_{TH}$ . Hence, it is possible to yield a more accurate model for  $V_{TH}$  by narrowing down the gate-source voltage range and keeping  $V_{DS}$  low. Similar arguments hold for  $g_{on}$ , where  $I_{DS} - V_{DS}$  characteristics are captured by modeling only the triode region, corresponding to small  $V_{DS}$  values since that section is sufficient to find out the conductance of device.

##### A. GENERIC $I_{DS} - V_{GS}$ MODELING

Before employing GPR in the MATLAB<sup>®</sup> regression learner toolbox, all available data has been normalized, and 20% of it has been randomly separated as test data in the pre-processing stage. The 5-fold cross-validation has been adopted as an error measure: All training data is incorporated into the training process and separated into five batches, where four batches are utilized during training. The remaining batch is left for validation. This process is repeated until all batches are used as training and testing data sets. Finally, the logarithm of the output current has been taken, such that the resulting modified output current has improved linearity. In post-training, the exponential of the anticipated outputs has been calculated to find the true RMSE and obtain the predicted drain-source current results. The resulting accuracy of the generated n-HCI model is  $RMSE_{train} = 1.14 \times 10^{-5}$ , which would translate into an error of 11.4  $\mu A$  for the predicted output current. After model generation, the test data has been fed into the model, and the resulting  $RMSE_{test} = 1.832 \times 10^{-5}$  is recorded, which maps to an error of 18.32  $\mu A$  in the predicted output current.

A common measure of performance degradation due to the HCI phenomenon can be acquired by observing  $I_{Dsat}$ . In

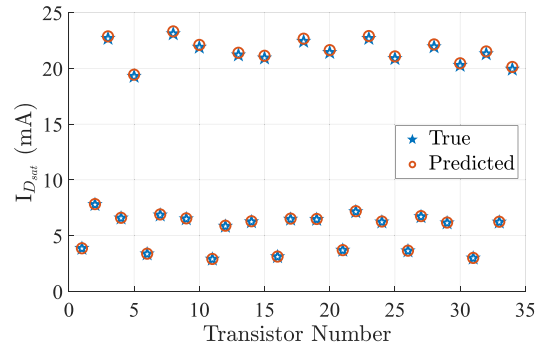


FIGURE 2. True and predicted  $I_{Dsat}$  after degradation of all tested n-MOSFETs.

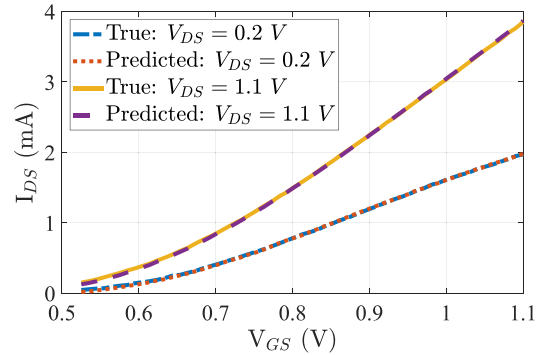


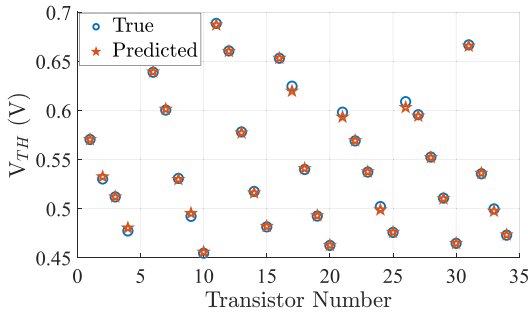
FIGURE 3. True and predicted  $I_{DS} - V_{GS}$  curves for two different  $V_{DS}$  values.

Fig. 2, the true and prediction values of  $I_{Dsat}$  after stress have been shown, which have been sampled from the testing data set. The predicted values are extracted from the model-based  $I_{DS} - V_{GS}$  characteristics while  $V_{GS} = V_{DS} = 1.1$  V. The accuracy of the model can be graphically noted in Fig. 2, where the true and predicted values almost overlap.

Another decent model performance can be visualized in Fig. 3 for both the true and predicted  $I_{DS} - V_{GS}$  characteristics. The size of the chosen transistor is  $W = 9.6 \mu m$  and  $L = 60 nm$ , one of the smallest transistors out of the batch. Two different values have been considered for  $V_{DS}$  proving the model accuracy. It can be seen that the two curves are very similar, and their slopes match, which is important since these predicted characteristics would be used in  $V_{TH}$  extraction.

##### B. $I_{DS} - V_{GS}$ MODELING FOR $V_{TH}$ EXTRACTION

The  $V_{TH}$  model for the n-HCI has been separately built since the extraction of  $V_{TH}$  is done for the lowest current levels at  $V_{DS} = 50 mV$ . When the current levels are low, the limited sensitivity of the measurement device causes errors that result in a curve with numerical fluctuations. Such discrepancies can be resolved by filtering the data with the moving mean function in MATLAB<sup>®</sup> for smoothing. This method uses a sliding ‘window’ of  $k$  points to calculate the mean for each  $k$ -batch across the array of neighboring elements. The  $k$  value is arranged so that the resulting curve matches the initial



**FIGURE 4.** The extracted true and predicted post-stress  $V_{TH}$  comparison of all tested n-MOSFETs.

curve as much as possible. Afterward, similar to the drain-source current modeling, the logarithm of the output data has been trained since the characteristic for  $V_{DS} = 50\text{ mV}$  has an exponential nature. Moreover, the test data is again selected randomly, corresponding to 20% of the available data.

$V_{TH}$  extraction has been realized using the first-derivative method [18], in which  $V_{TH}$  is the resulting value from the intersection of the derivative line belonging to the  $I_{DS} - V_{GS}$  curve with the  $V_{GS}$  axis. This extraction procedure has been repeated for all transistors, and the results are provided in Fig. 4, from which  $RMSE(V_{TH})$  can be found as 0.2%.

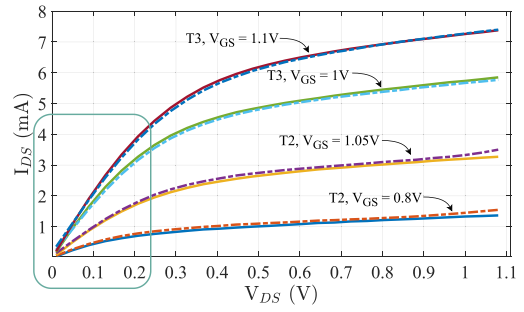
### C. $I_{DS} - V_{DS}$ MODELING FOR $g_{on}$ EXTRACTION

Another important indicator of transistor degradation due to HCI is the change in electron or hole mobility,  $\mu_n$  and  $\mu_p$ , respectively. However, mobility is an implicit component; hence, it is not possible to directly estimate or model it. As previously mentioned, when taking post-stress measurements, apart from  $|I_{DS}| - |V_{GS}|$ ,  $|I_{DS}| - |V_{DS}|$  has been acquired, as well. Through  $|I_{DS}| - |V_{DS}|$  characteristics, it is possible to extract  $r_{on}$ , the drain-source resistance of the transistor in the deep triode region at very small  $V_{DS}$  values. Correspondingly,  $g_{on}$  can be expressed as follows:

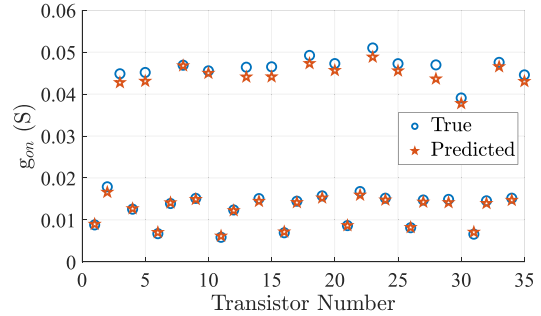
$$g_{on} = \frac{1}{r_{on}} = \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH}) \quad (1)$$

According to (1),  $g_{on}$  contains mobility dependence. So, modeling  $r_{on}$  aims to estimate the change in mobility because other equation elements are known. However, it should be noted that  $r_{on}$  is also dependent on  $V_{TH}$ . Since  $V_{TH}$  has been modeled earlier, the estimation for mobility can be easily computed relying on the accuracy of the generated models.

To model  $g_{on}$  in n-MOSFETs,  $I_{DS} - V_{DS}$  characteristics have been used at the highest  $V_{GS}$  of 1.1 V. Values of  $V_{DS} \leq 250\text{ mV}$ , where the linear behavior of  $I_{DS} - V_{DS}$  is observed, were employed in each condition to create the input matrix for the training. In Fig. 5, an example of the true and predicted  $I_{DS} - V_{DS}$  relationship is demonstrated where the region of interest for modeling  $g_{on}$  is highlighted. Here,  $g_{on}$  can be extracted by finding the slope values of the curves. The real and predicted  $g_{on}$  results for all transistors



**FIGURE 5.** True and predicted  $I_{DS} - V_{DS}$  of  $T_2$  and  $T_3$  for various  $V_{GS}$  values.



**FIGURE 6.** The true and predicted post-stress  $g_{on}$  comparison of all tested n-MOSFETs.

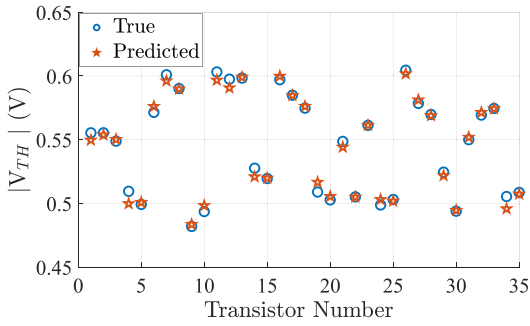
have been shown in Fig. 6, where the overall  $RMSE(g_{on})$  is found as 0.13%.

### D. MODELING OF THE P-HCI

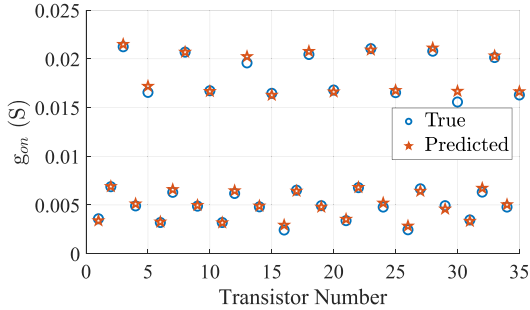
Similar to the analysis conducted on n-MOSFETs, the  $|I_{DS}| - |V_{GS}|$  characteristic,  $|V_{TH}|$ , and  $g_{on}$  have been modeled for p-MOSFETs. Each model is accompanied by examples and the accuracy statistics for all modeled quantities.

Modeling of  $|V_{TH}|$  degradation due to p-HCI has been more challenging than n-HCI. As previously explained, p-MOSFETs are less affected by HCI. Since the hole mobility is smaller, the pre-stress current levels of p-MOSFETs are inherently lower than that of n-MOSFETs, which makes them more susceptible to measurement errors. When  $|V_{DS}| = 50\text{ mV}$ , the current is at its minimum, thus contributing more noise to the collected data. For this reason, similar to the  $|V_{TH}|$  degradation model in n-MOSFETs, the  $|I_{DS}| - |V_{GS}|$  curves have been smoothed out by using the moving mean function. Here, the window size of the filter  $k$  has been arranged to be higher such that the resulting curve would be as smooth as possible while still matching the overall trend. Moreover, the output current data has been modified by taking its logarithm since the  $|I_{DS}| - |V_{GS}|$  characteristic for  $|V_{DS}| = 50\text{ mV}$  has an exponential nature. Finally, the test data was again selected randomly, comprising 20% of the available data.

Model extraction for  $|V_{TH}|$  and  $g_{on}$  has been repeated considering all p-MOSFETs. The results are shown in Figs. 7 and 8, respectively. The RMSE value is found 0.4% for  $|V_{TH}|$  and 0.0335% for  $g_{on}$ .



**FIGURE 7.** The extracted true and predicted post-stress  $V_{TH}$  comparison of all tested p-MOSFETs.



**FIGURE 8.** The true and predicted post-stress  $g_{on}$  comparison of all tested p-MOSFETs.

### E. ACCURACY ANALYSIS AND COMPUTATIONAL COMPLEXITY

Despite directing the circuit designer on model prediction accuracy in terms of current, the RMSE error metric cannot reflect the model deviation amount relative to all true outputs. Given that the predicted currents vary across a large range from micro- to several milli-amperes, a low RMSE may not be a proper indicator of a highly accurate model alone for small currents since large deviations in that regime could be overlooked due to their minor contribution to the overall error. Consequently, the Root Mean Square Relative Error (RMSRE) should be introduced as an additional model accuracy metric.

$$RMSRE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n \Delta Y_{rel,i}^2} \times 100\% \quad (2)$$

$$\Delta Y_{rel,i} = \frac{Y_{mod,i} - Y_{act,i}}{Y_{act,i}} \quad (3)$$

Here,  $Y_{mod,i}$  represents the predicted outcome, whereas  $Y_{act,i}$  is a true system output. The results that summarize the performance of each model for both RMSE and RMSRE have been presented in Table 2. The outcomes reveal that the model errors are small in comparison to the typical current values observed in analog and RF circuits.

In terms of computational complexity, compared to semi-empirical or theoretical models, as in [13] and [14], empirical models require more time during generation as accuracy is optimized based on the input data. Conversely, these models

**TABLE 2.** RMSE ( $\mu A$ ) and RMSRE (%) Results for Train and Test Data of  $|I_{DS}|$ ,  $|V_{TH}|$  and  $g_{on}$  Models in n-type and p-type MOSFETs.

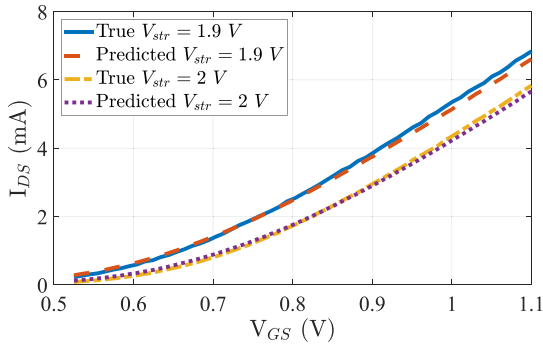
		TRAIN		TEST	
		RMSE	RMSRE	RMSE	RMSRE
$ I_{DS}  -  V_{GS} $	n-type	11.4	1.477	18.32	1.6765
	p-type	2.5	2.026	9.56	3.357
$ V_{TH} $	n-type	1.22	2.778	11.4	3.133
	p-type	1.5	1.1032	1.667	1.2166
$g_{on}$	n-type	129	6.609	102.4	10.58
	p-type	11.7	6.818	31.43	7.043

will perform more correctly and could be easily deployed in the pre-silicon design. Development of the machine learning-based models in this work takes less than 15 minutes, although generation time will rise commensurately as the input data grows. Moreover, model evaluation can be realized instantly.

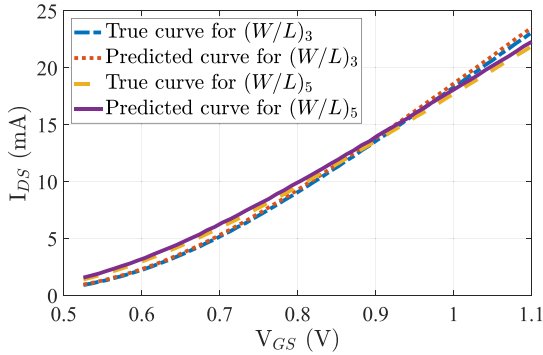
### F. ESTIMATION OF HCI DEGRADATION VIA ML-BASED MODELS

The models presented have demonstrated that it is possible to predict the drain-source current levels and important transistor quantities throughout various stress conditions that induce the HCI phenomena in 40 nm MOSFETs. However, the purpose of ML-based models is to predict the performance degradation due to HCI, even for unseen variable values/conditions in the model. In this section, the models have been generated while hiding the degradation data of  $T_3$  when  $t_{STR} = 3600$  s. In practice, analog and RF circuit designers use a variety of transistor sizes. Hence, generating accurate models that can predict performance degradation regardless of channel geometry is an important target of this study. The same model generation procedure has been repeated while holding back the stress data for the mentioned transistor. Subsequently, the resulting  $I_{DS} - V_{GS}$  curve prediction is compared with the measurement data, as shown in Fig. 9. The  $RMSE_{test}$  for this experiment is 0.18 mA, and graphically, the similarity of the curves highlights the accuracy achieved by ML-based models. Finally, it should be noted that the hidden degradation data belongs to a different transistor whose channel geometry is larger than that of all devices in the training set. Hence, this exercise demonstrates that machine-learning-based models could be potentially employed beyond their typical scope.

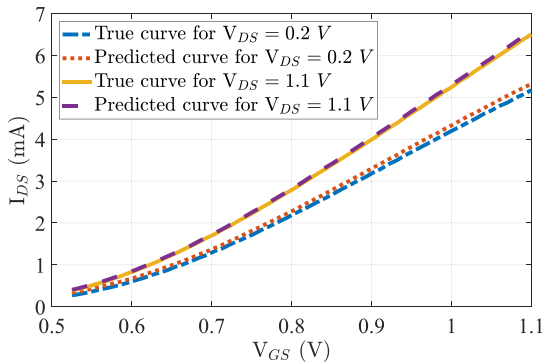
Another important application of this study is predicting the performance degradation of transistors under HCI for unseen electrical stress conditions. The following case study has been set up, which targets modeling the degraded  $I_{DS} - V_{GS}$  curve while only training the model using  $V_{STR} = 1.8$  V and  $V_{STR} = 2$  V but testing it on  $V_{STR} = 1.9$  V when  $t_{STR} = 3600$  s. The results have been shown in Figs. 10 and 11. It can be visualized that the models prove to be accurate even though the model has not been trained for a given test condition yet tried in that corresponding situation, thereby resulting in  $RMSE_{test}$  of 0.28 mA.



**FIGURE 9.** The true and predicted  $I_{DS} - V_{GS}$  for the model trained by excluding the degradation data of  $T_3$ .

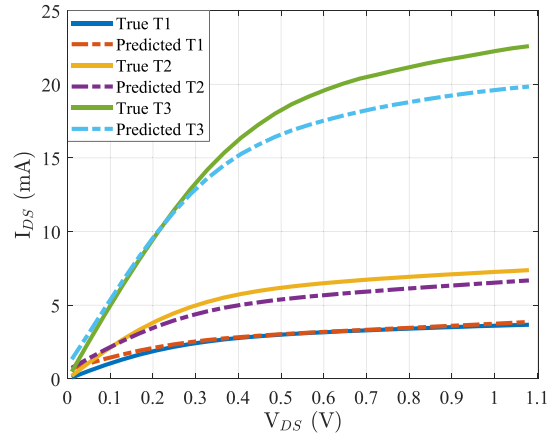


**FIGURE 10.** The true and predicted  $I_{DS} - V_{GS}$  of  $T_3$  and  $T_5$  for the model trained with the degradation data of  $V_{STR} = 1.8$  V and  $V_{STR} = 2$  V, and tested on the data corresponding to  $V_{STR} = 1.9$  V.



**FIGURE 11.** The true and predicted  $I_{DS} - V_{GS}$  of  $T_2$  at  $V_{DS} = 0.2$  V and  $V_{DS} = 1.1$  V for the model trained with the degradation data of  $V_{STR} = 1.8$  V and  $V_{STR} = 2$  V, and tested on the data corresponding to  $V_{STR} = 1.9$  V.

As a further exercise, the data for  $V_{STR} = 1.9$  V and  $V_{STR} = 2$  V when  $t_{STR} = 3600$  s were provided and the degraded  $I_{DS} - V_{DS}$  under  $V_{STR} = 1.8$  V predicted. This experiment is more challenging since the predicted case lies outside the range of the training set. However, from the testing perspective, it could be beneficial to estimate the corresponding degradation in advance since lower  $V_{STR}$  values require longer  $t_{STR}$ , which increases the testing costs and the time-to-market of integrated circuits.



**FIGURE 12.** The true and predicted  $I_{DS} - V_{DS}$  of  $T_1$ ,  $T_2$ , and  $T_3$  for the model trained with the degradation data of  $V_{STR} = 1.9$  V and  $V_{STR} = 2$  V, and tested on the data corresponding to  $V_{STR} = 1.8$  V.

True and predicted results  $I_{DS} - V_{DS}$  characteristics for  $T_1$ ,  $T_2$ , and  $T_3$  have been shown in Fig. 12. The accuracy of the model is acceptable for  $T_1$  and  $T_2$ , which have relatively small channel geometries. However, as the transistor size grows, modeling accuracy decreases, as observed in the case of  $T_3$ . In order to prevent similar accuracy drops, the training data sets should be extended with many more samples so that the prediction capability of the established model is enhanced under such difficult prediction scenarios.

## V. CONCLUSION

In this paper, an ML-based approach for HCI degradation has been proposed and implemented. GPR has been used as the learning algorithm. Critical transistors quantities post-stress  $I_{DS} - V_{GS}$  and  $I_{DS} - V_{DS}$  characteristics, as well as  $V_{TH}$  and  $g_{on}$  have been captured using the measurement data points by carefully avoiding the overfitting problem. The results have demonstrated that ML-based models can fit the combination of device aging with process variations observed in nanoscale MOSFETs with significant accuracy. Furthermore, as the case studies have demonstrated, these models can possibly be used in lieu of stress measurements of different-size MOSFETs or alternative stress considerations. Finally, p-HCI models of great interest for CMOS circuit design have been generated. Next, the developed models should be invoked in conventional circuit simulators to guide circuit designers in pre-silicon design time to meet the target reliability specifications.

## REFERENCES

- [1] M. B. Yelten, P. D. Franzon, and M. B. Steer, "Surrogate-model-based analysis of analog circuits—Part-II: Reliability analysis," *IEEE Trans. Device Mater. Rel.*, vol. 11, no. 3, pp. 466–473, Sep. 2011.
- [2] E. Afacan, M. B. Yelten, and G. Dünder, "Review: Analog design methodologies for reliability in nanoscale CMOS circuits," in *Proc. IEEE 14th Int. Conf. SMACD*, 2017, pp. 1–4.
- [3] W. Wang, V. Reddy, A. T. Krishnan, R. Vattikonda, S. Krishnan, and Y. Cao, "Compact modeling and simulation of circuit reliability for 65-nm CMOS technology," *IEEE Trans. Device Mater. Rel.*, vol. 7, no. 4, pp. 509–517, Dec. 2007.

- [4] M. B. Yelten, "Holistic device modeling: Toward a unified MOSFET model including variability, aging, and extreme operating conditions," *IEEE T. Circs. Syst. II, Exp. Brs.*, vol. 69, no. 6, pp. 2635–2640, Jun. 2022.
- [5] Y. Cao et al., "Cross-layer modeling and simulation of circuit reliability," *IEEE Trans. Comput. Aided Design Integr. Circuits Syst.*, vol. 33, no. 1, pp. 8–23, Jan. 2014.
- [6] L. Negre et al., "Reliability characterization and modeling solution to predict aging of 40-nm MOSFET DC and RF performances induced by RF stresses," *IEEE J. Solid-State Circuits*, vol. 47, no. 5, pp. 1075–1083, May 2012.
- [7] K. Huang, X. Zhang, and N. Karimi, "Real-time prediction for IC aging based on machine learning," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 12, pp. 4756–4764, Dec. 2019.
- [8] K. Huang, M. T. Hasan Anik, X. Zhang, and N. Karimi, "Real-time IC aging prediction via on-chip sensors," in *Proc. IEEE ISVLSI*, 2021, pp. 13–18.
- [9] R. Bu, Z. Ren, H. Ge, and J. Chen, "Online NBTI-induced partially depleted (PD) SOI degradation and recovery prediction utilizing long short-term memory (LSTM)," *Microelectr. Rel.*, vol. 142, Mar. 2023, Art. no. 114932.
- [10] N. Chatterjee, J. Ortega, I. Meric, P. Xiao, and I. Tsameret, "Machine learning on transistor aging data: Test time reduction and modeling for novel devices," in *Proc. IEEE IRPS*, 2021, pp. 1–9.
- [11] S. E. Tyaginov, I. Starkov, H. Enichlmair, J. M. Park, C. Jungemann, and T. Grasser, "Physics-based hot-carrier degradation modeling," *ECS Trans.*, vol. 35, no. 4, p. 321, Apr. 2011.
- [12] C. Guerin, V. Huard, and A. Bravaix, "General framework about defect creation at the Si/SiO<sub>2</sub> interface," *J. Appl. Phys.*, vol. 105, no. 11, Jun. 2009, Art. no. 114513.
- [13] R. Bottini, A. Ghetti, S. Vigano, M. G. Valentini, P. Murali, and C. Mouli, "Non-poissonian behavior of hot carrier degradation induced variability in MOSFETs," in *Proc. IEEE IRPS*, 2018, pp. 1–6.
- [14] Z. Yu, Z. Sun, R. Wang, J. Zhang, and R. Huang, "Hot carrier degradation-induced dynamic variability in FinFETs: Experiments and modeling," *IEEE Trans. Electron Devices*, vol. 67, no. 4, pp. 1517–1522, Apr. 2020.
- [15] E. Amat et al., "A comprehensive study of channel hot-carrier degradation in short channel MOSFETs with high-k dielectrics," *Microelectron. Eng.*, vol. 103, pp. 144–149, Mar. 2013.
- [16] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.
- [17] X. Xhafa and M. B. Yelten, "Design of a tunable LNA and its variability analysis through surrogate modeling," *Intl. J. Numer. Model., Electron. Netw., Devices, Fields*, vol. 33, no. 6, 2020, Art. no. e2724.
- [18] X. Xhafa, A. D. Güngördü, D. Erol, Y. Yavuz, and M. B. Yelten, "An automated setup for the characterization of time-based degradation effects including the process variability in 40-nm CMOS transistors," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, Jun. 2021.
- [19] S. Tyaginov et al., "A compact physics analytical model for hot-carrier degradation," in *Proc. IEEE Int. Rel. Phys. Symp. (IRPS)*, 2020, pp. 1–7.
- [20] A. Bravaix, C. Guerin, V. Huard, D. Roy, J. Roux, and E. Vincent, "Hot-carrier acceleration factors for low power management in DC-AC stressed 40nm NMOS node at high temperature," in *Proc. IEEE Int. Rel. Phys. Symp.*, 2009, pp. 531–548.
- [21] C. E. Rasmussen, *Gaussian Processes in Machine Learning*. Berlin, Germany: Springer, 2004, pp. 63–71.
- [22] J. Wang, "An intuitive tutorial to Gaussian processes regression," 2022, *arXiv:2009.10862*.