Chinese Journal of Semiconductors, 2002, 23(8): 817-824

Device Robust-design by Using the Response Surface Methodology

Xiao-feng XIE, Yong LU, Wen-jun ZHANG, Zhi-lian YANG
(Institute of Microelectronics, Tsinghua University, Beijing 100084, P. R. China)

Abstract: Device robust-design is inherently a multiple-objective optimization problem. Using design of experiments (DoE) combined with response surface methodology (RSM) can satisfy the great incentive to reduce the number of TCAD simulations that need to be performed. However, the errors of RSM models may large enough to diminish the validity of the results for some nonlinear problems. To find the feasible design space, a new method with objectives-oriented design in generations that taking the errors of RSM model into account is presented. After the augment design of experiments in promising space according to the results of RSM model in current generation, the feasible space will be emerging as the model errors deceasing. The results on FIBMOS examples show that the methodology is efficiently.

Key words: response surface methodology; design of experiments; device robust-design

1. Introduction

The technology CAD (TCAD) tools played a key role in the development of new technology generations. For the deep sub-micrometer devices, these tools provide a better insight than any measurement techniques and have become indispensable in the new device creation [1]. Technology development, however, requires substantially more than a fundamental simulation capability: tools and methods to assist in exploration of design trade-offs and to optimize a design are becoming increasingly important [2-4].

Device robust-design is inherently a multiple-objective optimization problem because the designers always want to attain more than one objective at the same time [5]. In order to obtain a robust device, one should avoid optimizing the design with the consideration of one single objective only, because it usually leads to a device that is not operable with respect to other objectives. There is a great incentive to reduce the number of TCAD simulations that need to be performed. Vary one factor at a time is a very inefficient procedure for optimizing and in many cases the best combination of design parameters may not be determined. Using design of experiments (DoE) [6,7] combined with response surface methodology (RSM) [8, 9] is a strategy that can overcome these problems since it can examine the whole parameter space while at the same time minimizing the number of TCAD experiments. The initial step is to perform some computer based screening experiments that identify the key design parameters to reduce the number of design parameters to a more manageable number. Simulation experiments are then designed and simulated, the results of which are used to derive the response surfaces.

There have some successful examples [10, 11] to design novel devices with DoE/RSM approach by predicting the tendency among parameters and responses. And some frameworks, such as in DoE/Opt [3], VISTA [4], etc., have integrated the RSM and optimization capability. However, the searched results by calling RSM model might not satisfy the multiple-objectives since there always have errors between the results of the simulator and its RSM model, even better model accuracy is achieved by a lot of methods [12-14].

In order to find the fully feasible design space that satisfies the multiple-objectives, a new method with objectives-oriented design in generations that taking the errors of RSM model into account is presented. After the augment design in promising space according to the RSM model in last generations, the feasible space will emerge as the model errors deceasing successively. Then the method has been used to optimize the focused-ion-beam (FIB) MOSFET [15] successfully.

2. Device robust-design with RSM

2.1 General representation for device design problems

The device design problems can be defined as finding design points $\vec{x} \in S_D$ such that

$$f_j(\vec{x}) \in [c_{l,j}, c_{u,j}], \qquad j = 1,\ldots, m \tag{1}$$

Where $\vec{x} = (x_1, \ldots, x_i, \ldots, x_n)^T$ is a design point, x_i represents a design parameter that specify the topography and impurity concentrations associated with a device, such as effective channel length, oxide thickness, doping profile descriptors, S_D is a *design space* defined as a Cartesian product of domains of design parameters x_i's ($x_i \in [l_i, u_i]$, $1 \le i \le n$). f_j are real-valued function of specified device performance on S_D, which defines the response (i.e. electrical behavior) of the device, such as on and off currents, threshold voltage, output resistance, etc. $c_{l,j}$ and $c_{u,j}$ specifies the *desired objective* for the jth response. If $c_{l,j} = c_{u,j}$ or only have small difference, it represents an equality constraint objective. If $c_{l,j} = -\infty$ or $c_{u,j} = +\infty$, it represents an inequality constraint objective. In additional, the optimum requirements should be transformed into strong constraints objective to obtain robust devices with the process deviations for each device parameter.

Definition 1: Each design points that satisfying all of the m constraints is denoted as *feasible point*. The set of feasible points is denoted as *feasible space* (S_F). Since $f_j(\vec{x})$ are calculated by TCAD simulator at here, its feasible space is also denoted as $S_{F,\,SIM}$.

2.2 Design of experiments (DoE)

The choice of an appropriate design of experiments is extremely important in determining the best model for multi-characteristics. As TCAD codes become even more complex and computationally intensive, it becomes increasingly more important to reduce the number of TCAD simulations required.

The value of statistically based experimental designs (the matrix of runs generated by specific combinations of design parameters in S_D) has been well established for automatic generation of experiments [6, 7].

The f*ull-factorial* design generates a uniform grid with user-specified density level m (totally m^n points) covering the input parameter space.

The *Central composite* designs are useful to explore the parameter space S_D with a minimum of required experiments. For example, a *central composite circumscribed* (CCC) design consists of 2*n axial points*, 2^n cube points (*full-factorial* with m=2) and *one center point*. The rotatability and the small number of necessary experiments make c*entral composite* designs very well suited for estimating the coefficients in a second-order model.

The *Latin hypercube sampling* (LHS) [16] proposes the "uniform" design concept. It provides an orthogonal array that randomly samples the entire design space broken down into r^n equal-probability regions (where r is the number of experiments). LHS can be looked upon as a stratified Monte Carlo sampling where the pairwise correlations can be minimized to a small value (which is essential for uncorrelated parameter estimates) or else set to a desired value. LHS is especially useful in exploring the interior of parameter space, and for limiting the experiment to a fixed (user specified) number of simulations.

2.3 Response surface methodology (RSM)

Response surface method [8, 9] is a kind of methodology, which generate mathematic model to describe the responses (device characteristics) in the space S_D. An important role of response surface models is to "mimic" the more complex workings of TCAD simulations or experiments. The most widely used model functions are polynoms of second order [4]

$$g(\vec{x}) = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=i}^{n} a_{ij} x_i x_j \tag{2}$$

The coefficients for this analytical function $g(\vec{x})$ can be calculated by a weighted least square estimation. For both inputs ($\vec{x}$) and responses (g), transformations (e.g., *log*, *exp*, *square*, *square root*, *inverse*) can be specified so as to including the additional knowledge about the system behavior. The covariance of the estimates, a metric of model stability, is dependent on the input design matrix and the lack of model fit. Choice of the input design matrix is critical to determining the model coefficients, and minimizing the covariance between the estimates of the model coefficients. Scaling the inputs minimizes the correlation between the estimates of the coefficients of the model [17]. In additional, weighted regression is important: a common sequence is to perform a broad experimental design, build a model, optimize to desired region within that model, and then refine the experimental design near that region. In such cases, it is useful to weight the second set of simulations more than the first to increase model accuracy in the region where the model will be most used.

2.4 Objectives-oriented augment design by considering model errors

Definition 2: The feasible space of all the $g_j(\vec{x}) \in [c_{l,j}+e_{l,j}, c_{u,j}+e_{u,j}]$ ($j=1,\ldots,m$) is denoted as maximum potential space ($S_{Fmax,RSM}$). The feasible space of all the $g_j(\vec{x}) \in [c_{l,j}, c_{u,j}]$ ($j=1,\ldots,m$) is denoted as feasible space of RSM model ($S_{F,RSM}$). The feasible space of all the $g_j(\vec{x}) \in [c_{l,j}+e_{u,j}, c_{u,j}+e_{l,j}]$ ($j=1,\ldots,m$) is denoted as minimum potential space ($S_{Fmin,RSM}$).

When RSM model is utilized to find feasible design space, the absolute errors $\delta_j(\vec{x}) = g_j(\vec{x}) - f_j(\vec{x})$ must be taken into account since $S_{F,RSM}$ is not a subset of $S_{F,SIM}$, as shown in Fig. 1. For worse cases, the intersection of $S_{F,RSM}$ and $S_{F,SIM}$ is small, even is empty when the $\delta_j(\vec{x})$ is large enough.

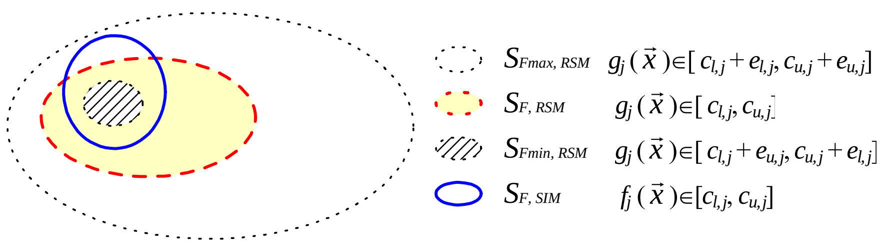


FIG. 1 A two-dimensional scheme for the relations of the feasible space in different conditions

Property 1: Suppose for any $\vec{x}$, $\delta_j(\vec{x}) \in [e_{l,j}, e_{u,j}]$, where $e_{l,j}$ and $e_{u,j}$ are the bound values for possible errors. Then for a given $\vec{x}$, it is easy to be proved that

a) If $g_j(\vec{x}) \notin [c_{l,j}+e_{l,j}, c_{u,j}+e_{u,j}]$, then $f_j(\vec{x}) \notin [c_{l,j}, c_{u,j}]$;

b) If $c_{l,j}+e_{u,j} < c_{u,j}+e_{l,j}$, and if $g_j(\vec{x}) \in [c_{l,j}+e_{u,j}, c_{u,j}+e_{l,j}]$, then $f_j(\vec{x}) \in [c_{l,j}, c_{u,j}]$.

From property 1a), it is easy to know that $S_{F,SIM} \in S_{Fmax,RSM}$; and from property 1b), it is easy to know that $S_{Fmin,RSM} \in S_{F,SIM}$.

The properties provide a methodology to find feasible points according to the response surface models, i.e. to find $S_{Fmin,RSM}$ as feasible space instead of $S_{F,SIM}$. However, $S_{Fmin,RSM}$ may be empty as the range of errors $[e_{l,j}, e_{u,j}]$ is very large due to the highly nonlinear response space. Hence, the errors, especially in promising space for all the objectives, should be decreased. This will be done by iterative generations, just as in evolutionary computation technology. For each generation, the objectives-oriented augment design in the $S_{Fmax,RSM}$ is performed for better accuracy of new RSM model, and potential spaces $S_{Fmin,RSM}$ and $S_{Fmax,RSM}$ are refreshed according to the new RSM model, as shown in Fig. 2. If $S_{Fmin,RSM}$ is still empty, then perform next generation.

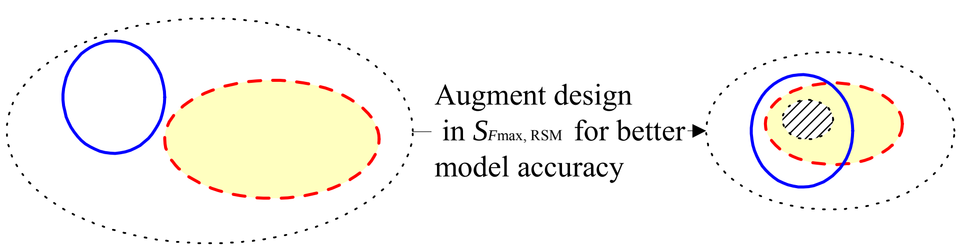


FIG. 2 Augment design in $S_{Fmax,RSM}$ for better model accuracy to enlarge $S_{Fmin,RSM}$

2.5 Total design flow

The total design flow can be described as below:

a) Define design space S_D and objectives;

b) Prepare the inputs for the DoE modules (S_D and the specified design) and start the DoE program to create a *DoE table*;

c) Calculate the responses for all the designs in the DoE table by batched executions with TCAD simulator and store into a persistent *run database* for all the experiments;

d) All responses that in specified design space are collected from the finished runs and added to the *experiment table*;

e) Evaluate the data in the experiment table to construct RSM models for all responses;

f) Calculate all the errors for the experiments and corresponding RSM calculation results to find the range of $[e_{l,j}, e_{u,j}]$;

g) Find the maximum potential space $S_{Fmax, RSM}$ and minimum potential space $S_{Fmin, RSM}$, according to the RSM model with the range of errors $[e_{l,j}, e_{u,j}]$ and objectives in specified design space;

h) If $S_{Fmin, RSM}$ exists, then output it as results; else screen the old design space with $S_{Fmax, RSM}$, return to b).

3. Design cases and discussion

The conception of device robust-design with response surface methodology will be demonstrated on a new-fashion focused-ion-beam MOSFET [15], as shown in Fig. 4. Here most device parameters are fixed. The effective channel length is 0.35μm; the oxide thickness (T_{ox}) is 0.01μm. For source and drain, the junction depth (X_j) is 0.1μm, doping (N_{SD}) is 7.0E20cm^{-3}. For the P$^+$ implant in the channel, vertical distance (r_p) is 0.0351μm, and vertical deviation (Δr_p) is 0.0182μm.

The design parameters include lateral implantation position that start from source side of channel (*FIB-X*), implantation dose (*Dose*), and substrate doping concentration (N_{sub}). The device responses includes drive current (I_{on}) and dynamic output conductance (G_{out}), at V_{ds}=1.5V and V_{gs}=1.5V. For DoE in each generation, the *full-factorial* design with level=5 is used. A *log* transformation of responses is used to aid model fitting. A numerical device simulator PISCES-2ET [18] is used to calculate the device responses.

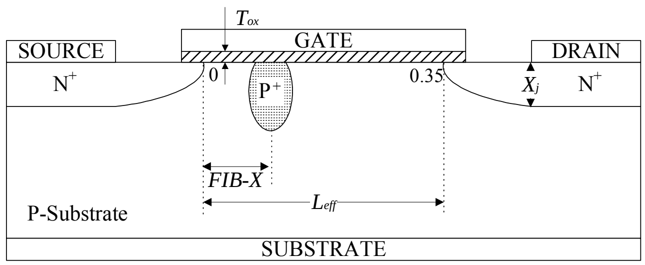
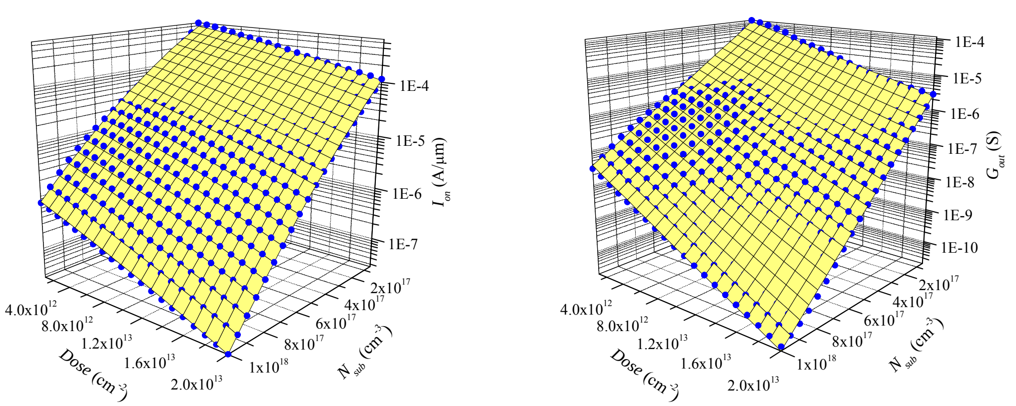


FIG. 3 Parameterized representation for FIBMOS device

- Ease case: *Dose* ∈[1E12, 2E13] (cm^{-2}), N_{sub}∈[5E16, 1E18] (cm^{-3}), when *FIB-X*=0.1μm;

Fig 4 shows perspective plots for a) I_{on} and b) G_{out} of simulated results and RSM model for different *Dose* and N_{sub}. Where the points are the simulation results in the whole design space, and the surface is the RSM model that fitted to 25 data points.



(a) (b)

FIG. 4 Perspective plots for a) I_{on} and b) G_{out} of simulated results and RSM model for different *Dose* and N_{sub}

It can be seen that the RSM model provides considerable precision. Fig 5 shows the isolines of the RSM model for I_{on} and G_{out}. Where the solid lines represent I_{on} and dash lines represent G_{out}. Notice both the response results are transformed by a *log* function. Since the isolines for I_{on} and G_{out} are not parallel to each other, lower G_{out} can be achieved at a fixed I_{on} if higher *Dose* and lower N_{sub} are used. Fig 6 shows an optimized result for that I_{on} = 2E-4A/µm (i.e. the isoline that equal to 3.7 in Fig 5). Where the dash line is a point with higher N_{sub} and lower *Dose*, and the solid line is an optimized solution with lower G_{out}.

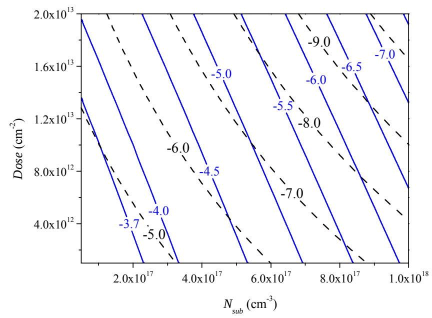


FIG. 5 Response surfaces for $log(I_{on})$ and $log(G_{out})$

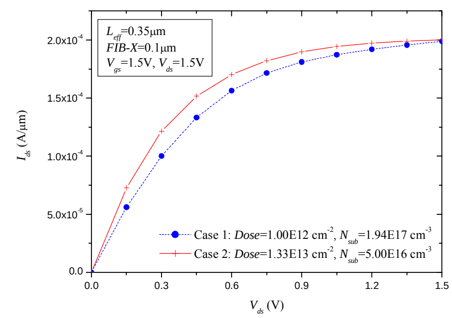


FIG. 6 An optimization example by RSM for larger G_{out}

Here the RSM model is utilized to find feasible design space, which the objectives for responses are set as I_{on}>1E-4A/µm (i.e. $log(I_{on})$>-4), G_{out}<6.3E-6S(i.e. $log(G_{out})$<-5.2). The range of response errors $\delta_j(\vec{x})$ between the RSM model and the simulated results are [-0.10, 0.12] for $log(I_{on})$ and [-0.12, 0.17] for $log(G_{out})$. Fig. 7 shows the feasible space in different conditions, here the part of design space that N_{sub}>5E17 cm^{-3} is not shown in order to demonstrate more clearly. The solid lines, i.e. the isolines for $log(I_{on})$= -4 and $log(G_{out})$= -5.2 of the simulation results represent the boundary of $S_{F,SIM}$, is used as reference. The dash lines, i.e. the isolines for $log(I_{on})$= -4 and $log(G_{out})$= -5.2 of the RSM model results represent the boundary of real $S_{F,RSM}$. It can exactly to see that $S_{F,RSM}$ is not a subset of $S_{F,SIM}$, which due to the errors between RSM and simulation results.

Hence the model errors $\delta_j(\vec{x})$ must be taken into account. As in Fig. 7, the dash dot lines, i.e. the isolines for $log(I_{on})$= -4.10 and $log(G_{out})$= -5.03 of the RSM model represent the boundary of $S_{Fmax,RSM}$, and the dot lines, i.e. the isolines for $log(I_{on})$ = -3.88 and $log(G_{out})$= -5.32 of the RSM model represent the boundary of $S_{Fmin,RSM}$. Here $S_{Fmin,RSM}$ is a subset of $S_{F,SIM}$, and then robust design points can be selected in this region. For example, as a design point **A** in $S_{Fmin,RSM}$, which with N_{sub} =5E16cm^{-3} and *Dose* =1.7E13cm^{-2}, then we have responses as I_{on} =1.44E-4A/µm, G_{out} =3.96E-6S.

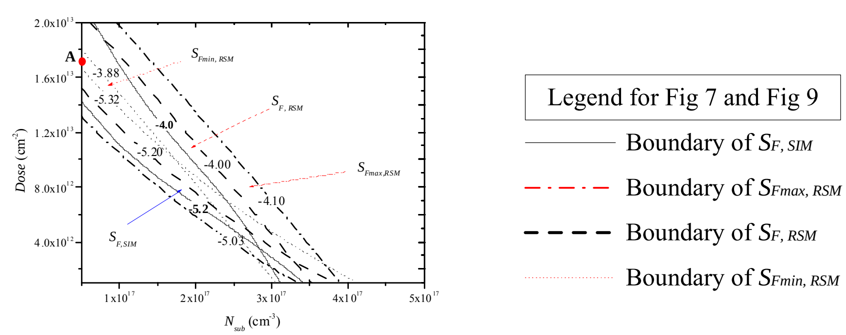
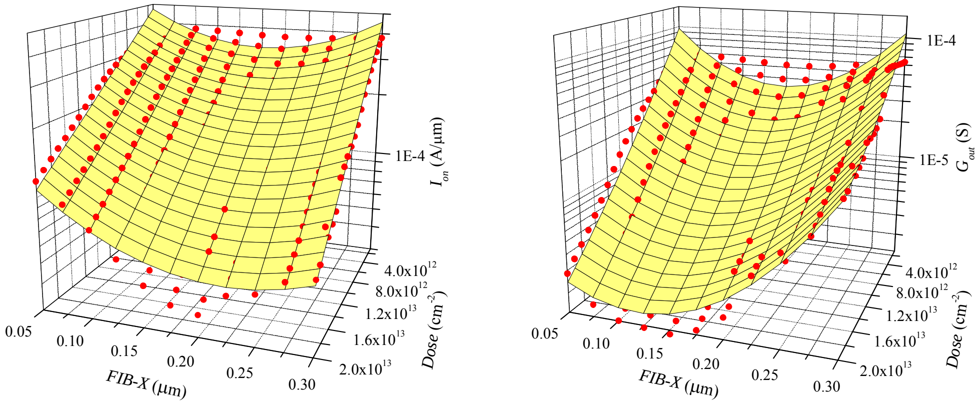


FIG. 7 The feasible space in different conditions for the objective that I_{on}>1E-4A/µm and G_{out}<6.3E-6S

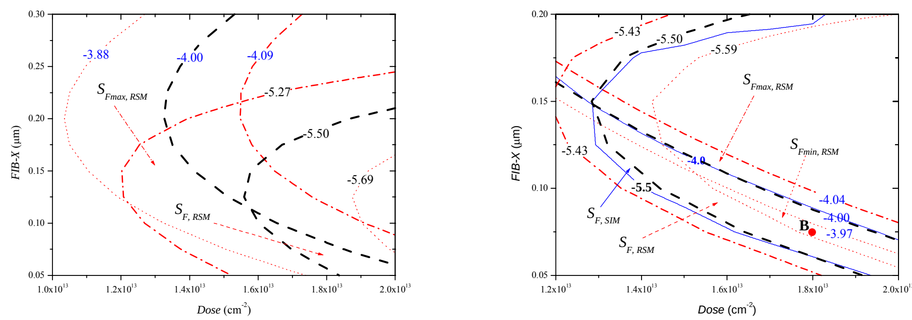
- Hard case: *Dose* ∈[1E12, 2E13] (cm^{-2}), *FIB-X* ∈[0.05, 0.30] (µm), when N_{sub} =0.1cm^{-3};

Fig 8 shows perspective plots for a) I_{on} and b) G_{out} of simulated results and RSM model for different *Dose* and *FIB-X*. Where the points are the simulation results in the whole design space, and the surface is the RSM model that fitted to 25 data points.



(a) (b)

FIG. 8 Perspective plots for a) I_{on} and b) G_{out} of simulated results and RSM model 1 for different *Dose* and *FIB-X*



(a) (b)

FIG. 9 The feasible space in different conditions for the objective that I_{on}>1E-4A/μm and G_{out}<3.16E-6S

Here the RSM model is utilized to find feasible design space, which the objectives for responses are set as I_{on}>1E-4A/μm (i.e. $log(I_{on})>-4$), G_{out}<3.16E-6S(i.e. $log(G_{out})<-5.5$). The range of response errors $\delta_j(\vec{x})$ between the RSM model and the simulated results are [-0.09, 0.12] for $log(I_{on})$ and [-0.19, 0.23] for $log(G_{out})$. Fig. 9a) shows the feasible space of RSM in different conditions that takes the model errors into account. The dash lines, i.e. the isolines for $log(I_{on})$= -4 and $log(G_{out})$= -5.5 of the RSM model represent the boundary of real $S_{F, RSM}$. The dash dot lines, i.e. the isolines for $log(I_{on})$= -4.09 and $log(G_{out})$= -5.27 of the RSM model represent the boundary of $S_{Fmax, RSM}$. However, in this case, the dot lines, i.e. the isolines for $log(I_{on})$= -3.88 and $log(G_{out})$= -5.69 of the RSM model cannot construct the set $S_{Fmin, RSM}$ with feasible design points since $c_{l,j} + e_{u,j} > c_{u,j} + e_{l,j}$, according to property 1b).

In order to find feasible design space, the augment design should be performed in $S_{Fmax, RSM}$ for better model accuracy. Here the design space is decreased to *Dose*∈[1.2E13, 2E13] (cm^{-2}) and *FIB-X* ∈[0.05, 0.225] (μm). Then new generation is started with a *full-factorial* design with level=5. The range of errors $\delta_j(\vec{x})$ between the RSM model and the simulated results are [-0.04, 0.03] for $log(I_{on})$ and [-0.09, 0.07] for $log(G_{out})$. Fig. 9b) shows the feasible space of new RSM model in this design space in different conditions. The solid lines, i.e. the isolines for $log(I_{on})$= -4 and $log(G_{out})$= -5.5 of the simulation results represent the boundary of $S_{F, SIM}$, and the dot lines, i.e. the isolines for $log(I_{on})$= -3.97 and $log(G_{out})$= -5.59 of the RSM model results represent the boundary of $S_{Fmin, RSM}$. It can be found that the enhanced RSM model accuracy by the augment design in $S_{Fmax, RSM}$ induces that $(c_{l,j} + e_{u,j})$ to be less than $(c_{u,j} + e_{l,j})$, and $S_{Fmin, RSM}$ which is a subset of $S_{F, SIM}$ is not empty now. As a point **B** in $S_{Fmin, RSM}$, which with *FIB-X*=0.075μm and *Dose* =1.8E13cm^{-2}, then we have responses as I_{on} =1.12E-4A/μm, G_{out} =2.32E-6S.

4. Conclusion

This paper has shown how response surface methodology combined with TCAD simulation can be used in the device robust-design, i.e. find the feasible space that satisfies multi-objectives, by considering model errors. The fully feasible space $S_{Fmin, RSM}$ which is a subset of real feasible space $S_{F, SIM}$ will emerge by successively accelerated enhancing the model accuracy in promising space according to the objectives-oriented augmented design in maximum potential space $S_{Fmax, RSM}$ of last generation.

Future work is needed to employ new methods that enhance the accuracy of RSM model, since better model accuracy makes for finding $S_{Fmin, RSM}$.

References:

[1] Dutton R W, Strojwas A J. Perspectives on technology and technology-driven CAD. IEEE Trans. on Computer-aided Design of Integrated Circuits and Systems, 2000, 19(12): 1544-1560

[2] Hosack H H, Mozumder P K, and Pollack G P. Recent advances in process synthesis for semi-conductor devices. IEEE Trans. Electron Devices, 1998, 45 (3): 626 – 633

[3] Boning D S, Mozumder P K. DOE/Opt: A System for design of experiments, response surface modeling, and optimization using process and device simulation. IEEE Trans on Semiconductor Manufacturing, 1994, 7(2): 233-244

[4] Plasun R, Stockinger M, Selberherr S. Integrated optimization capabilities in the VISTA technology CAD framework. IEEE Trans. on Computer-aided Design of Integrated Circuits and Systems, 1998, 17(12): 1244-1251

[5] Chen M-R, Chiang P, Lin L. Device robust-design using multiple-response optimization technique. IEEE 5th International Workshop on Statistical Metrology, 2000: 46-49

[6] Box GE P, Hunter W G, Hunter J S. Statistics for experimenters. New York: John Wiley, 1978

[7] Lorenzen T, Anderson V, Design of experiments. Berlin Germany: Marcel Dekker, 1991

[8] Alvarez A R, Abdi B L, Young D L, et al. Application of statistical design and response surface methods to computer-aided VLSI device design. IEEE Trans. on Computer-aided Design of Integrated Circuits and Systems, 1988, 7(2): 272-288

[9] Myers R, Khuri A I, Carter W H. Response surface methodology. Technometrics, 1989, 31(2): 137-157

[10] Burenkov A, Tietzel K, Lorenz J. Optimization of 0.18μm CMOS devices by coupled process and device simulation. Solid-State Electronics, 2000, 44: 767-774

[11] P. Moens, M. Tack, H. Van hove, M. Vermandel, and D. Bolognesi, Development of an optimised 40V pDMOS device by use of a TCAD design of experiment methodology, Conf. on Simulation of Semiconductor Processes and Devices, 2000: 276 -279

[12] Waring T G, Walton A J, Ferguson S, Sprevak D. Application of covariance structures to improve the fit of response surfaces to simulation data. IEEE Trans on Semiconductor Manufacturing, 1999, 12(3): 366-374

[13] Joshi S, Sherali H D, Tew J D. An enhanced response surface methodology (RSM) algorithm using gradient-deflection and second-order search strategies. Computer s & Operational Research, 1998, 25(7/8): 531-541

[14] Xie H, Lee Y C, Mahajan R L, Su R. Process optimization using a fuzzy logic response surface method. IEEE Trans. on Components, Packaging, and Manufacturing Technology – Part A, 1994, 17(2): 202-211

[15] Shen C-C, Murguia J, Goldsman N, et al. Use of focused-ion-beam and modeling to optimize submicron MOSFET characteristics. IEEE Trans. Electron Devices, 1998, 45(2): 453-459

[16] McKay M D, Conover W J, Beckman R J. A comparison of three methods for selecting values of input variables in the analysis of outputs from a computer code. Technometrics, 1979, 21: 239-245

[17] Draper N R, Smith H. Applied regression analysis. New York: John Wiley, 1981

[18] Yu Z, Chen D, So L, Dutton R W. PISCES-2ET manual. Integrated Circuits Laboratory, Stanford University, 1994