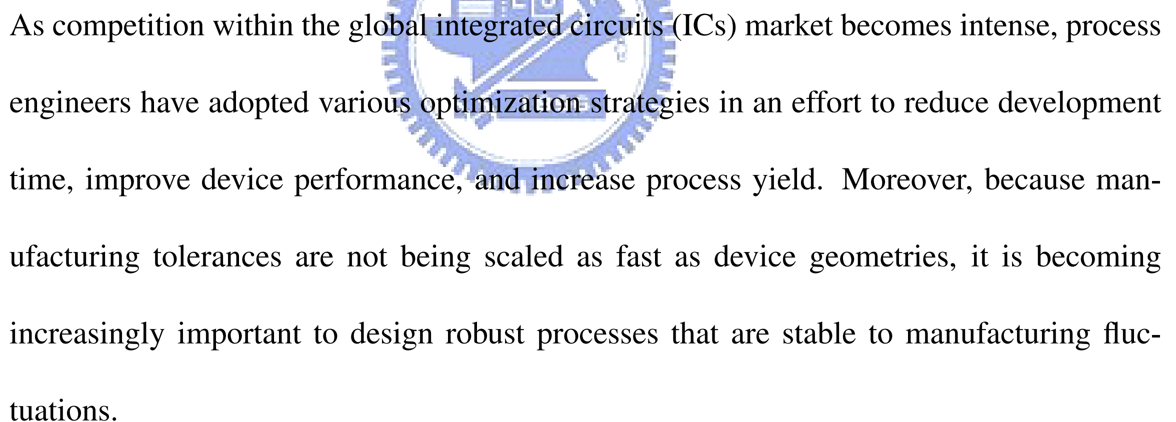
Chapter 1

Introduction



As competition within the global integrated circuits (ICs) market becomes intense, process engineers have adopted various optimization strategies in an effort to reduce development time, improve device performance, and increase process yield. Moreover, because manufacturing tolerances are not being scaled as fast as device geometries, it is becoming increasingly important to design robust processes that are stable to manufacturing fluctuations.

When manufacturers want to produce the high performance devices, two main process topics should be considered. First, how to move the device performances to the target (we call this is the location problem)? Second, how to minimize the process fluctuations (the dispersion problem)? For these two objectives, engineers should do many try-and-error

experiments to achieve them. One common practice is to guess the improved settings of the control factors using engineering judgment, then conduct a paired comparison with the starting conditions. The guess-and-test cycle is repeated until an improvement has been obtained, the deadline has been reached, or the budget has been exhausted. This practice relies heavily on luck. It is inefficient and time-consuming [27].

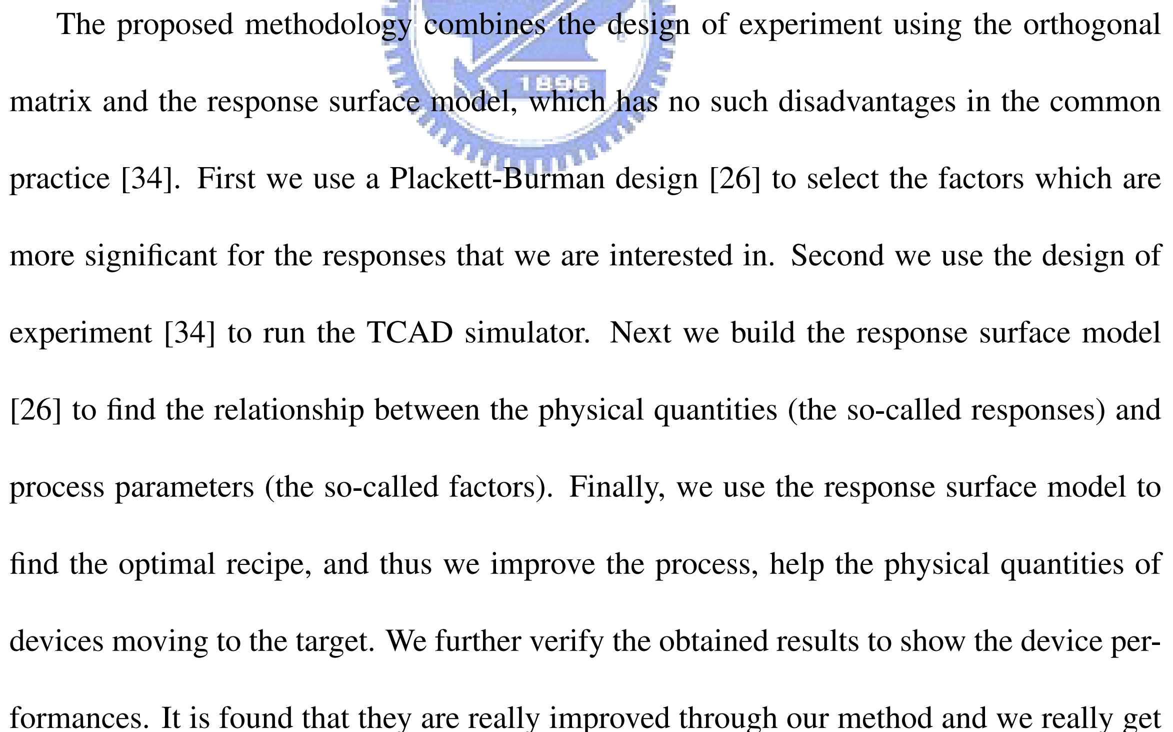
Another common practice is to optimize systematically one control factor at a time: (i) identify the most important factor, (ii) investigate this factor by itself, ignore the other factors, (iii) make recommendations on changes (or no change) to this factor, (iv) and move onto the next most important factor and repeated steps (ii) and (iii). The iterations will be terminated when a satisfactory solution is found. However, it has several disadvantages [34]:

1. it could not estimate some interactions among factors;
2. the conclusions from its analysis are not general enough; and
3. it might miss optimal settings of factors.

Due to the large time and monetary costs associated with physical experimentation, engineers have applied towards commercial available or in-house technology computer-aided design (TCAD) tools to simulate the fabrication process and predict device characteristics. However, the complex computations required by such simulations limit process optimization strategies. Different computational methods together with the TCAD tools to achieve

optimization and sensitivity analysis have been of great interests.

In this thesis, a systematical method for the fabrication optimization and the sensitivity analysis in sub-100nm NMOSFET devices is investigated. Based on the screening design, the design of experiment (DOE), a TCAD simulator, the response surface models, and the optimization using desirability function, the device performances after fabrication have been optimized with respect to five specified constraints. They are 1) shifting the value of threshold voltage to the specific target; 2) minimizing the subthreshold slope; 3) minimizing the off-state current; 4) moving the value of on-state current to a specific target; 5) and minimizing the drain-induced barrier lowing.



The proposed methodology combines the design of experiment using the orthogonal matrix and the response surface model, which has no such disadvantages in the common practice [34]. First we use a Plackett-Burman design [26] to select the factors which are more significant for the responses that we are interested in. Second we use the design of experiment [34] to run the TCAD simulator. Next we build the response surface model [26] to find the relationship between the physical quantities (the so-called responses) and process parameters (the so-called factors). Finally, we use the response surface model to find the optimal recipe, and thus we improve the process, help the physical quantities of devices moving to the target. We further verify the obtained results to show the device performances. It is found that they are really improved through our method and we really get

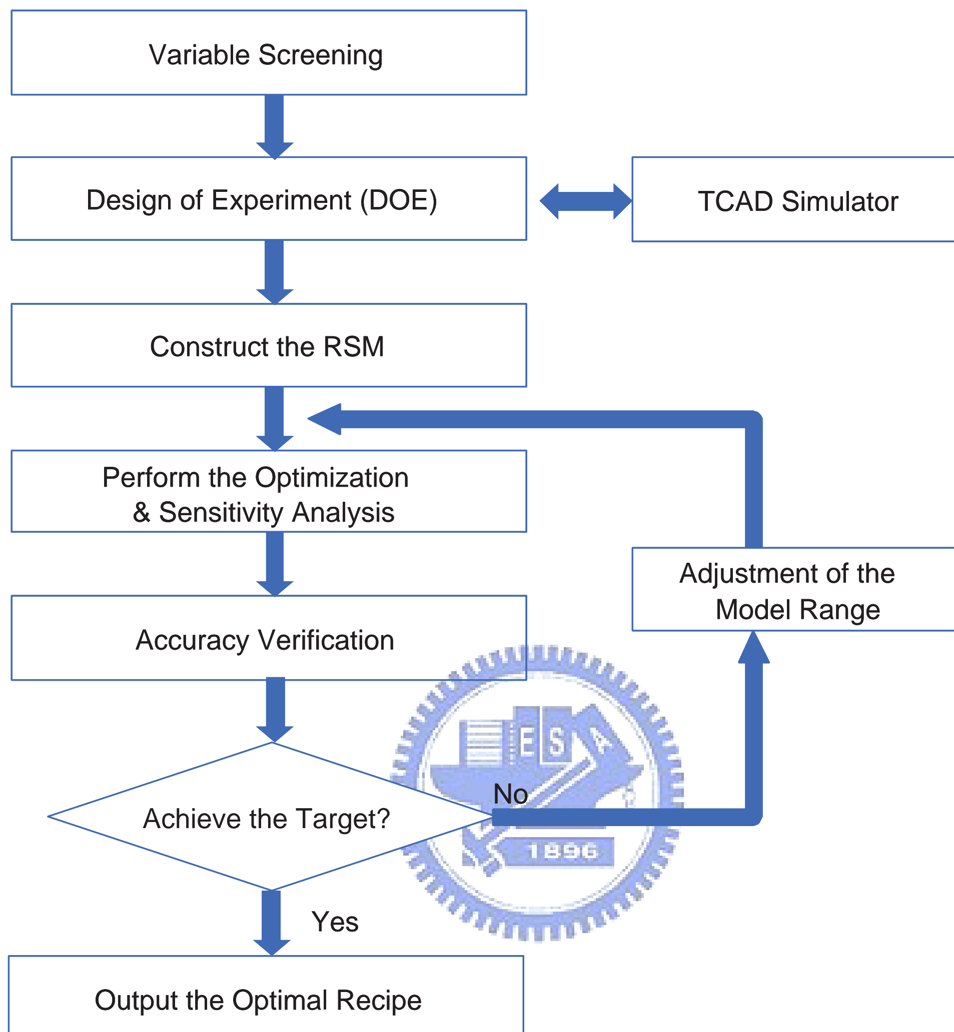
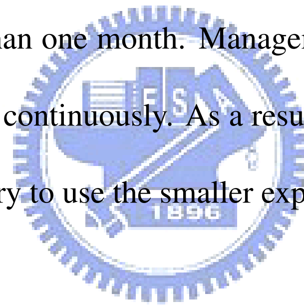


Figure 1.1: Procedure flow chart in this work. It contains the screen design, the design of experiment, the response surface model, the optimization and sensitivity analysis, and the verification. These methods are based on the data from a well-known industrial used TCAD simulation tool.

the better recipe than before we used. An illustration flowchart for the proposed procedure is shown on Fig. 1.1.

1.1 Motivation

When designers or experimenters face the optimization problem, people often try their experiences to solve it. Once these people quit the company, then they also take their know how away. Even the manager knows using the common practice to be the optimization strategy, it is not efficiency [34] [27]. From the TCAD simulator for the 90nm N-type metal oxide semiconductor field effect transistor (NMOSFET) device process, we know that the simulation time of a completed run, which was combined a process simulation run and a device simulation run, is about half day, and the cost and time spent in the real fabrication are even more than one month. Management could not accept such big risk to make a lot of try-and-errors continuously. As a result, we provide a method to explore the optimization problem, and try to use the smaller experiments than the traditional one.



1.2 Background

In this section, without lost of the generality we take a 0.18μm MOSFET process to be an example, and explain the details of the manufacturing flow. Then we introduce the TCAD simulation of semiconductor process and devices.

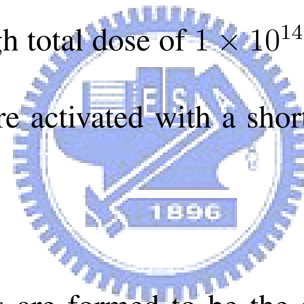
1.2.1 The Process of the MOSFET

In the following sections, the fabrication processes are introduced in the context of a typical 0.18μm n-channel MOSFET process flow for an example. The MOSFET structure is produced and the processing of the isolation is excluded. Here, an n-doped substrate with a Phosphorus concentration of $10^{15} cm^{-3}$ is used. Three sets of boron implants are performed: the first high-energy implant creates the p-well, the second medium energy implant defines a retrograde boron profile to prevent punch-through, and the last low-energy implant is for a threshold voltage (Vt) adjustment.

Then the gate oxide is grown at a temperature of 850 $^{\circ}C$ for 10min in a pure oxygen at a pressure of 1atm. Next the poly silicon gate is created, 0.18μm of polysilicon is deposited over the entire structure. In this step there are two etching processes. The first etch refers to the previously defined mask and thus only the exposed part of the polysilicon is etched away. The requested etching depth (i.e., 0.2μm) is larger than the deposited layer. This over-etching ensures that no residual islands remain. The second etching process does not

refer to any masks; however, the polysilicon will naturally act as a mask for this selective etching process.

After producing the polysilicon, the next step is often the poly reoxidation. To release stresses a thin oxide layer is grown on-top of the the poly silicon before the spacer formation. Next the LDD and halo implants are to be performed. The LDD implant is used a high dose of $4 \times 10^{14} cm^{-2}$ and a relatively low energy of 10keV. The halo is created by a quad implant, i.e., the implant is performed in four steps each at a different angle. This is done to ensure that the boron penetrated well into the channel at the tips of the source/drain extensions. Again a relatively high total dose of $1 \times 10^{14} cm^{-2}$ is used. When the implant steps are finished, the implants are activated with a short thermal cycle (rapid thermal anneal, RTA).



Then the nitride spacers are formed to be the spacer. A uniform, 60nm thick layer of nitride is deposited over the entire structure. Then the nitride is etched back again. Finally, the thin oxide layer, grown during the poly reoxidation step, is removed. Next the Source/Drain implants are to be performed. To ensure a low resistivity of the source and drain regions, this implant step uses a very high dose of $5 \times 10^{15} cm^{-2}$. A tilt of 7° is used to reduce channelling, and the rotation of -90° ensures that the plane of incident is parallel to the gate stack, such that the 7° tilt angle does not lead to an asymmetry between the source and the drain. Finally, metal pads are defined. Metal pads are created via deposition and

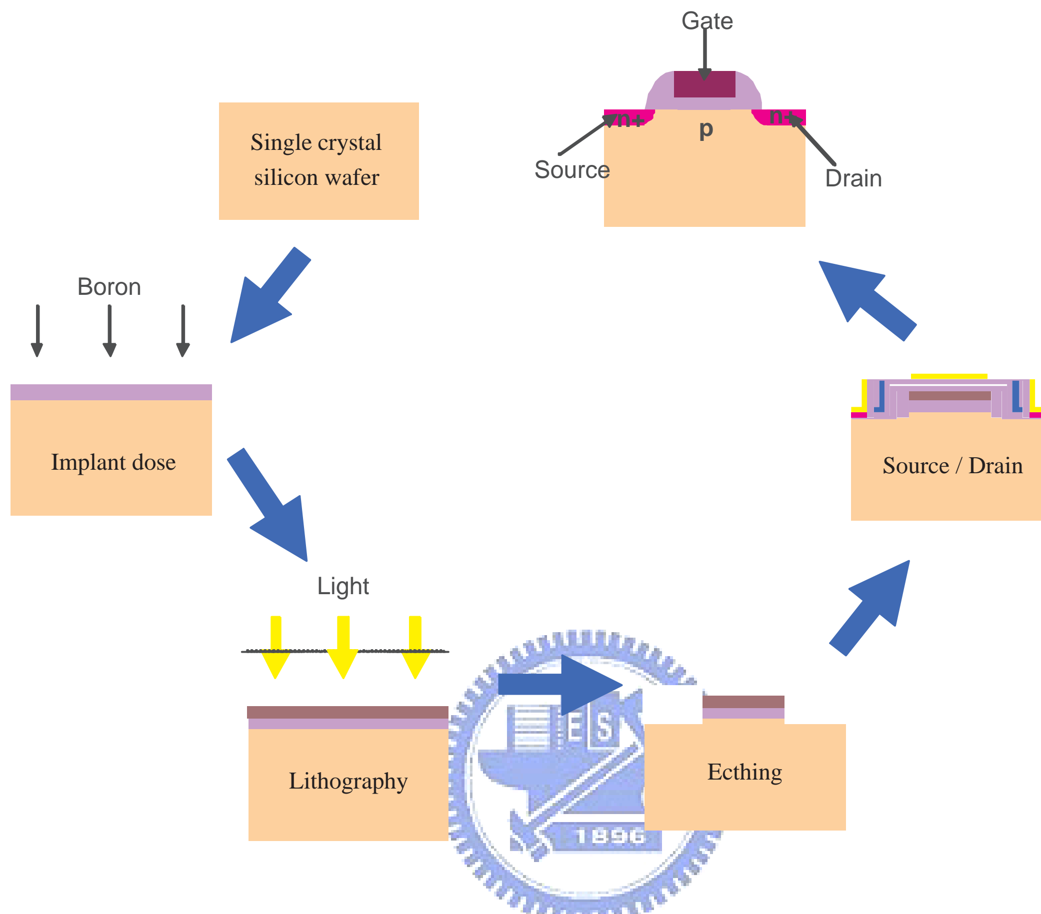


Figure 1.2: The flowchart of the device manufacturing process.

etching to identify the contact areas, and the key process of MOSFET have been finished. A flowchart of the device manufacturing process has been shown in Fig. 1.2.

1.2.2 The TCAD Simulation of Semiconductor Process and Devices

TCAD refers to using computer simulations to develop and optimize semiconductor processing technologies and devices. TCAD simulation tools solve fundamental, physical partial

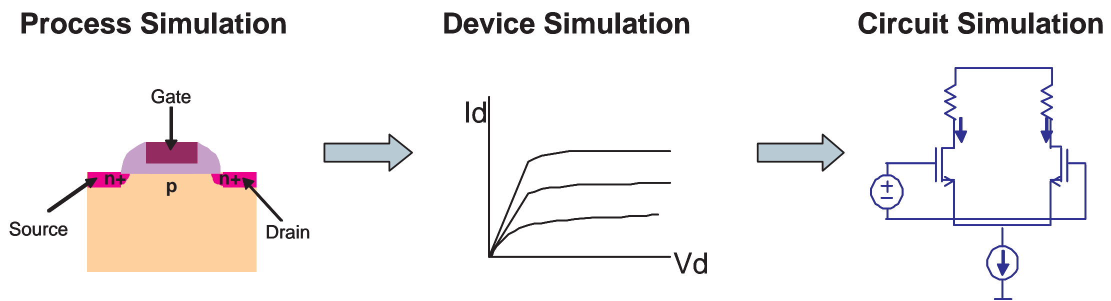


Figure 1.3: The flowchart of the TCAD simulation and the circuit simulation. Through the process simulation in TCAD, the structure and the profile of the device are obtained, then the device is applied into the device simulation to estimate the curves of Id-Vd. Therefore the physical quantities can be calculated. After the TCAD simulation, the device can be used into the circuit simulation.

differential equations, such as diffusion and transport equations for discretized geometries, representing the silicon wafer or the layer system in a semiconductor device. This deep physical approach gives TCAD simulation predictive accuracy. It is therefore possible to substitute TCAD computer simulations for costly and time-consuming test wafer runs when developing and characterizing a new semiconductor device or technology. TCAD consists of two main branches: process simulation and device simulation. The relationship of the process simulation and device simulation has been shown in Fig. 1.3.

In process simulation, processing steps such as etching, deposition, ion implantation, thermal annealing and oxidation are simulated based on physical equations, which govern the respective processing steps. The simulated part of the silicon wafer is discretized

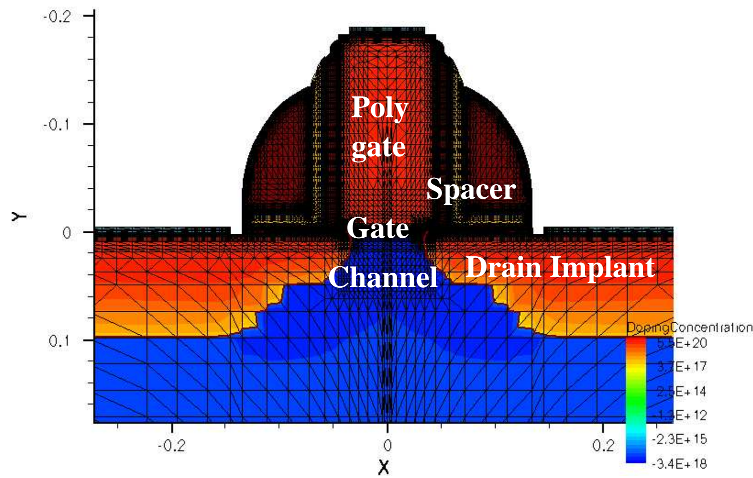


Figure 1.4: Close-up of the gate-drain corner of an nMOSFET with finite element grid. TCAD simulators solve physical partial differential equations on a finite element grid representing a discretized model of the physical device.

(meshed) and represented as a finite element structure (see Fig. 1.4). For example, in the simulation of thermal annealing complex diffusion equations for each dopant species are solved on this mesh. For oxidation simulations the growth of the silicon oxide is simulated taking into account the oxygen diffusion, the mechanical stresses at corners, etc. In another example, Monte Carlo techniques are often used to simulate the process of ion implantation, where ion-semiconductor interactions are taken into account to compute individual ion paths [16].

Device simulations can be thought of as virtual measurements of the electrical behavior of a semiconductor device, such as a transistor, or a diode. Again, the device is represented

as a meshed finite element structure. Each node of the device has properties associated with it, such as material type, doping concentration, etc. For each node the carrier concentration, current densities, electric field, generation and recombination rates, etc., are computed. Electrodes are represented as areas on which boundary conditions, such as applied voltages, are imposed. The device simulator solves the Poisson and the carrier continuity equation (and possibly other equations, such as lattice and carrier temperature equations) [15], [18], [19], [14]. After solving these the resulting electrical currents at the contacts are extracted.

TCAD simulations are widely used throughout the semiconductor industry. As technologies become more and more complex the semiconductor industry relies increasingly more on TCAD to cut costs and speed up the research and development process. In addition, semiconductor-manufacturing companies use TCAD for yield analysis, i.e., monitoring, analyzing and optimizing their IC process flows, as well as to analyze the impact of IC process variation.

1.3 The Proposed Statistical Methodology

The roots of response surface models (RSM) can be traced back in the early 1930s or even earlier, and until May, 1989, the development of the design of experiment has been introduced and incorporated more completely on Technometrics by Raymond H. Myers et al [25]. This methodology can be incorporated into the design of experiment (DOE) and

response surface model. It is applied to solve diverse problems between the responses and factors, this part will be discussed more detail in the appendix A. In agriculture, response surface models have been used for Woodruff lime-requirement buffer (WRF) [32]. In the electromagnetic device, the inverse function and gradient information are used to improve the response surface model [23]. In general, a 2nd order response surface model has been applied to these fields [10]. More detailed discussion will be shown in Appendix A.

For the manufacturing, high-speed civil transport (HSCT) and semiconductor industry are the main two industries whose relationship between responses and factors are the most complex. Different 2nd order response surface models have been used in these two industries very much, and extended several methods of reducing model and finding the optimal settings. In HSCT, for example, the statistical method which is called the stepwise regression has been used in order to reduce the function terms of the response surface models [12]. In other efforts, regression analysis and analysis of variance (ANOVA) are also used for reducing the response surface model.

In the semiconductor industry, the response surface model is applied in the optimization problems [13] [3] [9], finding the significant factors due to the process fluctuations [11] [31], or the sensitivity analysis [7]. It can be applied in just one small portion of the whole process flow, such as be used for studying the relationship between the threshold voltage and the lightly-doped-drain (LDD) implantation [31], for improvement of the deposition

and etch process steps [2], or for plasma doping process [13]. It also can be used for investigating the device performance after whole process steps finished [29] [33]. Many efforts, especially not studying the completed manufacturing process, focus on fewer factors (often fewer than five) to build a 2nd order response surface model.

We know if we want to build the model which is closer to the true environment, a 2nd order response surface model is better than a 1st order response surface model. It's because that the 1st order response surface model could only explain the 'linear' relationship between responses and factors, however the 2nd order response surface model can explain not only the linear relationship but also the 'curve' relationship [26].

Though the 2nd order response surface model is more suitable for building semiconductor device and manufacturing process model; however, when number of the factors that we considered is too large (e.g., more than ten), then we will need the dramatically huge runs of experiment to estimate a proper response surface model. For example, provided that there are 10 factors should be counted in building the corresponding model, then we need at least 149 runs with resolution V central composite design [34]. In this example, one run meant one recipe of manufacturing process, and it takes more than 20 days for 90nm NMOSFET TCAD simulation, or it costs the expense for at least 149 patches of true manufacturing.

On the other hand, when we face the whole manufacturing process of semiconductor

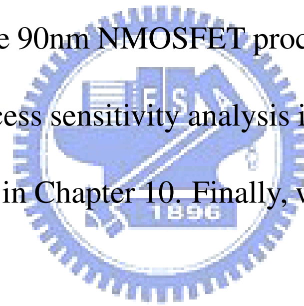
device, there are more than twenty even hundred factors should be considered. Thus using traditional method to construct the response surface model becomes almost improper for this competitive and cost-leading market. As a result, the screening method of selecting the significant but few factors is very important for beginning to build the a proper model.

1.4 Objectives

In this work we will show that the proposed statistical method performs well for the process location problem. We take the newer process technology for the 90nm nMOSFET manufacturing to be an example. Here we want to optimize five physical quantities to the target: 1) threshold voltage (Vth), to be moved to a specific value; 2) subthreshold slope (SS), to be minimized; 3) off-state current (Ioff), to be minimized; 4) saturation current (Idsat), to be moved to a specific value; and 5) drain-induced barrier lowering (DIBL), to be minimized it. The objective of this work is trying to obtain the optimal recipe, which can improve these 5 physical quantity performances better than the nominal recipe case. Finally we also use the developed model to perform the sensitivity analysis, it helps us know whether the distributions of the 5 physical quantity are in their specifications that we assigned.

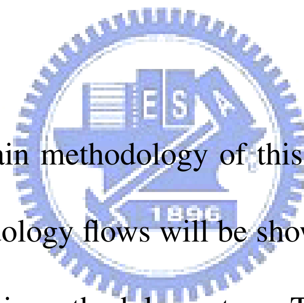
1.5 Outline of the Thesis

This thesis is organized as follows. In Chapter 2, the proposed methodology will be introduced and each step will be discussed in detail. The application of the proposed method to an novel process technology of 90nm NMOSFETs will be discussed in Chapter 3. The simulation of the process recipe will be shown in more detail. In the following chapters, we discuss each step of our methodology, sequentially. The results of screening design will be presented in Chapter 4. Detail about design of experiment is in Chapter 5. Chapter 6 states the response surface model. The variable selection will be shown in Chapter 7. The optimal results about the 90nm NMOSFET process using the proposed method will be discussed in Chapter 8. Process sensitivity analysis is discussed in Chapter 9. We also give the verification experiments in Chapter 10. Finally, we draw some conclusions and suggest future works.



Chapter 2

The Developed Statistical Methodology



In this chapter we introduce the main methodology of this work in more detailed in the following sections. First the methodology flows will be shown in section 2.1, to provide a general concept about the steps of this methodology steps. Then we state each step in next sections respectively. Four main topics will be shown. First, screening designs, in this step we can select fewer but important factors from many process parameters. Second, design of experiment, this step help us construct the design matrix which can be used to make experiments or process simulations, through this effort we can get the responses which we are interested in. Third, construct the response surface model, after we know the responses, we build response surface models to connect the relationship between the responses and the factors. Finally we provide several applications about this work in the fourth topic.

2.1 The Computational Procedure

There are four steps in this work, and the flowchart is shown in Fig 2.1. When we want to use this statistical analysis to optimize the device performances, we might face many important and unimportant factors. The first step could help us to solve this problem, when there are too many factors so that we don't know how to select them, screen design is one suggested method to choose the significant factors which affect the responses most.

After we get the important factors, we can use them to build the complex response surface models which are more closed to the true nature. So we need to construct the design of experiment before we build the response surface models. Traditional design is the central composite design (CCD) [26], and there are three special forms of it, which are the central composite circumscribed (CCC) design, the face centered cube (CCF) design, and the central composite inscribed (CCI) design [35]. If we feel the run number of the traditional central composite design is still too large, we can choose another design, which is called small composite design (SCD) [34]. Detailed will be discussed in the corresponding section.

Then we can use design matrix to do the experiment, from it we obtain the responses data, and next we can build the response surface model [26]. In this step the main purpose is to find the corresponding polynomial functions to describe the relationship between the

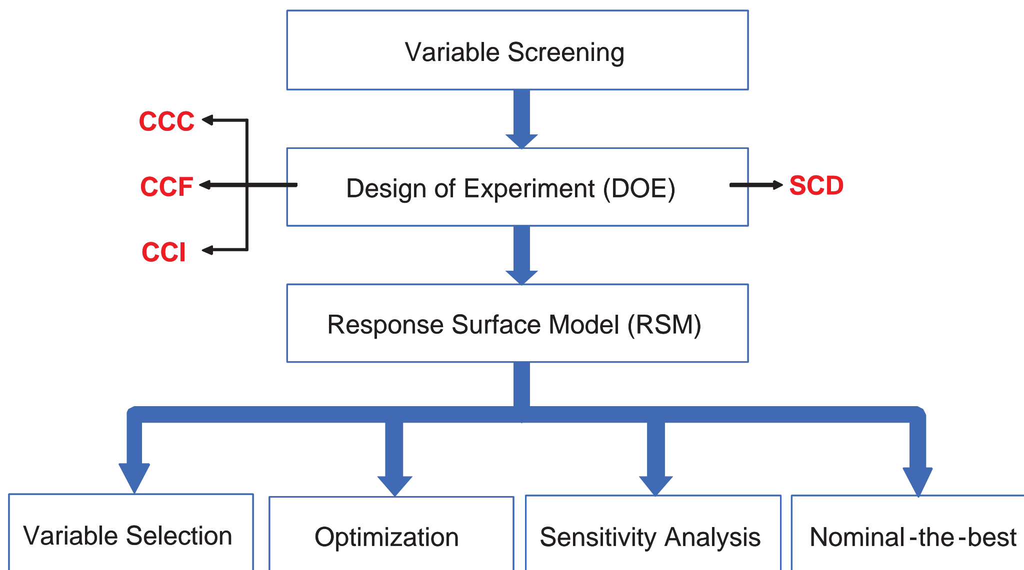


Figure 2.1: Systematic scheme for sub-100nm MOSFET fabrication optimization and applications. It contains the screening design, the design of experiment which could be separated 4 parts, the response surface model, and other applications using these methodology.

1896

responses and the factors. As we get these functions, we can use them to perform optimization and sensitivity analysis [33].

The last step presents several applications by using response surface models, and these models could be applied for variable selection, optimization and process sensitivity analysis [33], [29], and solving nominal-the-best problems [34]. Next we discuss them respectively.

2.2 Screening Designs

The term 'screening design' refers to an experimental plan that is intended to find the few significant factors from a list of many potential ones. Alternatively, we refer to a design as a screening design if its primary purpose is to identify significant main effects, rather than interaction effects, the latter being assumed an order of magnitude less important [24].

Even when the experimental goal is to eventually construct a response surface model (an RSM analysis), the first experiment should be a screening design when there are many factors to be considered.

As we have discussed in Chapter 1, we know there are more than hundred variables in the complex process, and if we want to build a 2nd order response surface model, we need 147 completed runs for just 10 variables. This experiment takes more than 588 hours of simulation time, or 147 batches in the real fabrication. Thus, for complex processes in which a large number of variables exist, time constraints may limit the feasibility of performing a response surface study including all possible process conditions. Statistical factor screening methods may be employed during the initial stages of a response surface study to identify those processes having the greatest effect on process output. By reducing the number of process taken into consideration, the subsequent response surface study may be streamlined, and fewer runs or tests required. Due to the widespread assimilation of TCAD tools throughout the semiconductor industry, it is important to identify a factor

screening method suitable for deterministic systems.

To determine factor significance, this work utilizes an approach combining statistical design of experiment methods and statistical graphical methods. To minimize the number of simulations or experimental runs, statistical designs should be used to efficiently explore the experimental design space. Two-level fractional factorial design with resolution III or Plackett-Burman design plans are ideally suited for screening design [35].

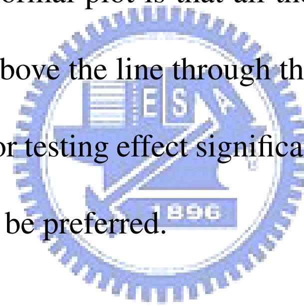
A special subset of factorial designs, Plackett-Burman designs, are published by R.L. Plackett and J.P. Burman in 1946 [34]. It provides very economical designs that permit the study of $k = N - 1$ variables in N runs, where N is a multiple of 4 (rather than a power of 2) [35]. For example, 11 process conditions may be studied in 12 experimental runs. In a Plackett-Burman design, main effects are, in general, heavily confounded with two-factor interactions. By using such design schemes, the process designer sacrifices the ability to clearly separate the main effects of each process condition from certain higher-order interactions, a condition known as 'aliasing'. During factor screening, some aliasing is deemed acceptable in the face of time and / or resource constraints [33].

Using a graphical method to judge effect significance is often preferred to using an extension of the t test such as the studentized maximum modulus test, because the latter depends on the estimate S^2 of σ^2, which may be unreliable, where as the former depends on seeing a deviation from linearity, which is easier to judge [34]. The S^2 is the sample

variance, and the σ^2 is the true variance of the population. A related graphical method is the half-normal probability plot. Let $|\hat{\theta}|_{(1)} \leqslant \ldots \leqslant |\hat{\theta}|_{(I)}$, where $|\hat{\theta}|_{(i)}$ is the ordered statistics of the effects, for i = 1, . . ., I. Plot them against coordinates based on a half-normal distribution; the absolute value of a normal random variable is half-normal. The half-normal probability plot consists of the points

$$(\Phi^{-1}(0.5 + 0.5[i - 0.5]/I), |\hat{\theta}|_{(i)}|), \tag{2.1}$$

for i = 1, . . ., I. The Φ is the cumulated density function of the standard normal distribution. The advantage of the half-normal plot is that all the large estimated effects appear in the upper right corner and fall above the line through the small estimated effects. In this work half-normal plots are used for testing effect significance. For other purposes such as outlier detection, normal plots may be preferred.



2.3 Design of Experiments

Response models are constructed relating the device characteristics to the significant process conditions using data generated from statistical experimentation. Mathematically, response surface models may be represented as second-order polynomials [26]:

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k} \sum_{i \neq j}^{k} \beta_{ij} x_i x_j + \varepsilon, \tag{2.2}$$

where k is the number of input factors, x_i is the ith input factor, β_i is the ith regression coefficient, and ε represents model error.

Many applications of response surface models involve constructing and checking the adequacy of a second-order model. The central-composite design (CCD) is perhaps the most common experimental design used to generate second-order response models. These designs combine a two-level full factorial or fractional factorial design of n_f runs with $2k$ axial runs and n_c center runs, where k represents the number of control factors [26].

As seen in Fig.2.2, central-composite designs include five input levels for each control factor $(0, \pm 1, \pm \alpha)$. Level 0, the nominal factor level, represents the base processing conditions. The cube levels (± 1) are selected to reflect the design space of interest [24].These values are typically set to a multiple of the factor's standard deviation or a percentage of its nominal value [33]. The precise value of α depends on certain properties desired for the design and on the number of factors involved.

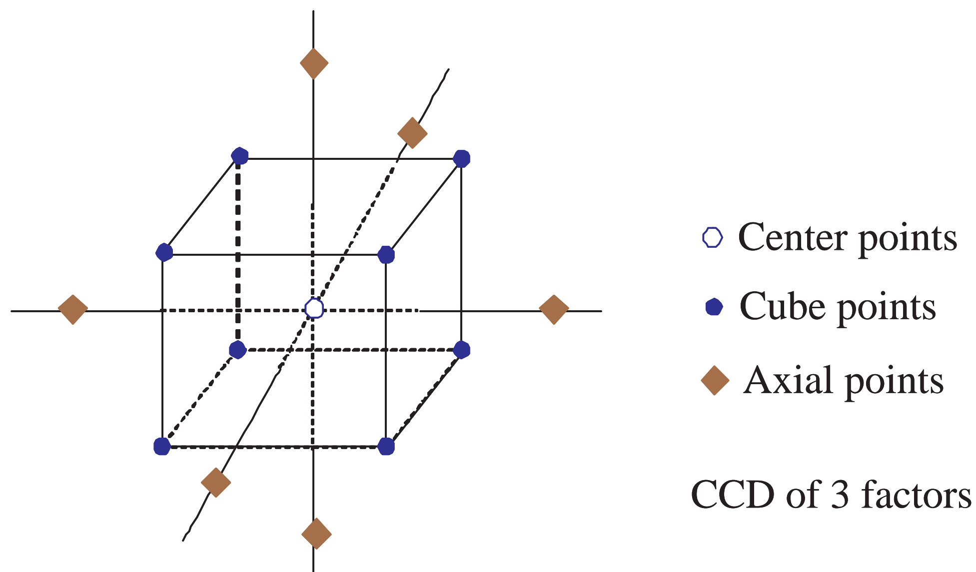


Figure 2.2: Central composite design for 3 factors.

1896

2.3.1 Forms of the CCD

The central composite circumscribed (CCC) designs are the original form of the central composite design [35]. The axial points at some distance α from the center base on the properties desired for the design and the number of factors in the design. The axial points establish new extremes for the low and high settings for all factors. Fig 2.2 illustrates a CCC design. These designs have circular, spherical, or hyperspherical symmetry and require 5 levels for each factor. Augmenting an existing factorial or resolution V fractional factorial design with axial points can produce this design.

For those situations in which the limits specified for factor settings are truly limits, the

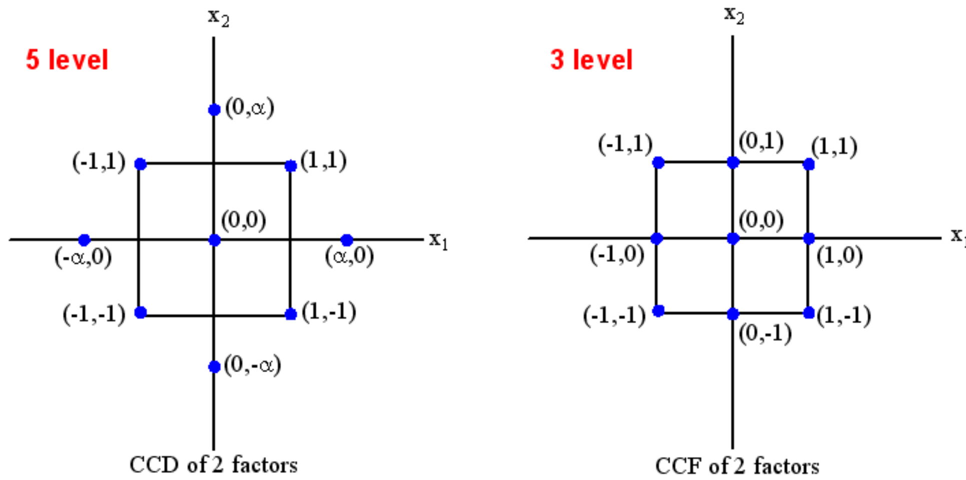


Figure 2.3: Comparison of the CCC design and the CCF design.

central composite inscribed (CCI) design uses the factor settings as the axial points and creates a factorial or fractional factorial design within those limits (in other words, a CCI design is a scaled down CCC design with each factor level of the CCC design divided by α to generate the CCI design) [35]. This design also requires 5 levels of each factor.

The other special design is called the face centered cube (CCF) design. In this design the axial points are at the center of each face of the factorial space, so $\alpha = \pm 1$. For Fig.2.2, if the diamond points move to the face in the cube, then the design is CCF. This variety requires 3 levels of each factor. Augmenting an existing factorial or resolution V design with appropriate axial points can also produce this design. Fig.2.3 shows the difference between the CCC and CCF designs for 2 factors design [35].

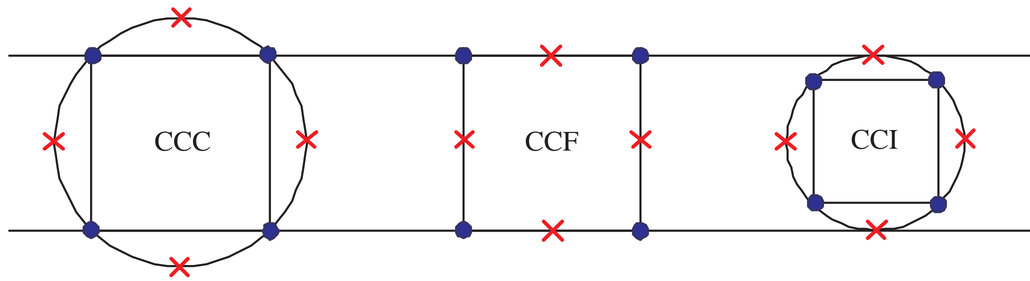


Figure 2.4: Comparison of the three types of central composite designs.

The diagrams in Fig.2.4 illustrate the three types of central composite designs for two factors. Note that the CCC explores the largest process space and the CCI explores the smallest process space. Both the CCC and CCI are rotatable designs, but the CCF is not. In the CCC design, the design points describe a circle circumscribed about the factorial square. For three factors, the CCC design points describe a sphere around the factorial cube. To maintain rotatability, the value of α depends on the number of experimental runs in the factorial portion of the central composite design:

$$\alpha = [n_c]^{\frac{1}{4}}, \tag{2.3}$$

where n_c is the number of experimental runs in the factorial potion of the central composite design. However, the factorial portion can also be a fractional factorial design of resolution V.

2.3.2 The Small Composite Design (SCD)

Alternative designs are small composite designs, and the term 'small' refers to an experimental design that has small runs in CCD [34]. A typical CCD requires about 9 runs for 2 factors, 15 runs for 3 factors, 25 runs for 4 factors, and 27 runs for 5 factors. It is obvious that, once you have four or more factors you wish to include in a CCD, you will need more than one lot (i.e., batch) of experimental units for your basic design. However, there is a way to cut down on the number of runs, as suggested by H.O. Hartley in his paper 'Smallest Composite Designs for Quadratic Response Surfaces', has published in Biometrics, December 1959.

Theorem 2.3.1 *In any central composite design whose factorial portion is a 2^{k-p} design that does not use any main effect as a defining relation, the following parameters in Eq.(2.2) are estimable:*

1. β_0, β_i, β_{ii}, $i = 1, \ldots, k$, and

2. one β_{ij} selected from each set of aliased effects for $i < j$.

It is not possible to estimate more than one β_{ij} from each set of aliased effects.

This method addresses the theory that using a Resolution V design as the smallest fractional design to create a response surface design is unnecessary. The method adds axial points to designs of Resolution III and uses the axial points to clear the main effects of aliasing with the two-factor interactions. The resulting design allows estimation of the

Table 2.1: A list of the central composite designs for $2 \leq k \leq 10$ factors

Factors	**Number-of-coefficients**	**Runs**	**Runs**	**Runs**
k	$(k+1)(k+2)/2$	**CCD**	**Hartley**	**Draper-and-Lin**
2	6	9	-	7
3	10	15	11	11
4	15	25	17	17
5	21	27	-	22
6	28	45	29	29
7	36	79	47	37
8	45	81	-	47
9	55	147	83	57
10	66	149	-	67

higher-order interactions [8].

Smaller central composite designs can be found by using the Plackett-Burman designs for the factorial portion and be represented by Draper and Lin. By using this method, several designs achieve to the smallest composite design [6]. Table 2.1 present the different designs for k factors.

2.4 Construction of the Response Surface Model

After we choose one kind of design, we do real experiments in the laboratory or use TCAD simulator to find the responses we are interested. If N observations are collected in an experiment, the model for them takes the form

$$y_n = \beta_0 + \sum_{i=1}^{k} \beta_i x_{ni} + \sum_{i=1}^{k} \beta_{ii} x_{ni}^2 + \sum_{i=1}^{k} \sum_{i \neq j}^{k} \beta_{ij} x_{ni} x_{nj} + \varepsilon_n, \quad \texttt{for} \quad n = 1, \ldots, N., \tag{2.4}$$

where y_n is the nth value of the response and $x_{n1}, \ldots, x_{nk}$ are the corresponding values of the k factors [26].

These N equations can be written in matrix notation as:

$$y = X\beta + \varepsilon, \tag{2.5}$$

where $y = (y_1, \ldots, y_N)^T$ is the $N \times 1$ vector of responses, $\beta = (\beta_0, \beta_i, \beta_{ii}, \beta_{ij})^T$ for $i = 1, \ldots, k$ and $1 \leq i < j \leq k$ is the $[(k+1)(k+2)/2] \times 1$ vector of regression coefficients, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)^T$ is the $N \times 1$ vector of errors.

The model is a univariate version. When there are m response variables to be studied together and there are dependence between each response variable, the multivariate regression model would be used to estimate the unknown β. Assuming that there are m response

variables, the model has the following matrix form:

$$\begin{cases} y_{\cdot 1} = X\beta_{\cdot 1} + \varepsilon_{\cdot 1}, \\ y_{\cdot 2} = X\beta_{\cdot 2} + \varepsilon_{\cdot 2}, \\ \cdots \\ y_{\cdot m} = X\beta_{\cdot m} + \varepsilon_{\cdot m}, \end{cases} \tag{2.6}$$

The unknown parameters in the model are the regression coefficients $\beta_{\cdot 1}$, ..., $\beta_{\cdot m}$ and the covariance matrix Σ. Estimates of the model coefficients in Equ.(2.6) are determined using a least-squares fit:

$$\begin{pmatrix} \hat{\beta}_{\cdot 1} \\ \hat{\beta}_{\cdot 2} \\ \cdots \\ \hat{\beta}_{\cdot m} \end{pmatrix} = \begin{pmatrix} (X^T X)^{-1} X^T y_{\cdot 1} \\ (X^T X)^{-1} X^T y_{\cdot 2} \\ \cdots \\ (X^T X)^{-1} X^T y_{\cdot m} \end{pmatrix},$$

which is the vertical joining of the LSE's for the m-multiple linear regression model. Next we discuss the residuals, R^2, and MSE in each response variable.

Following model generation, a residual analysis is conducted to evaluate the model adequacy. The residual for the nth observation in each response is defined as:

$$e_n = y_n - \hat{y}_n, \tag{2.7}$$

where y_n is the nth measured response (i.e., one physical quantity in the simulation results), and $\hat{y}_n$ is the nth predicted response (i.e., one physical quantity from the response surface

model results) [26]. The normality assumption is checked by preparing a normal probability plot of the residual values. If the assumption holds, this plot will resemble a straight line. If the assumption is violated, a non-linear data transformation (e.g., $y' = log(y)$) may be applied and new RSM models generated in an attempt to improve model adequacy [26]. A second plot showing the residual values versus the predicted response values is used to verify if the variance of the original observation is constant. A random scattering of the residual values indicates that no correlation exists between the observed variance and the mean level of the response.

A relatively simple procedure is performed to check for model significance in relation to random error. This test involves calculating the test statistic:

$$F_0 = \frac{\frac{1}{p}\sum_{j=1}^{n}(\hat{y}_i - \overline{y})^2}{\frac{1}{n-p-1}\sum_{j=1}^{n}(y_i - \hat{y}_i)^2}, \tag{2.8}$$

where $\overline{y}$ is the average of measured response values, m is the number of model coefficients, and again y_i, $\hat{y}_i$, and n are the ith measured response, the ith predicted response, and the number of simulated runs, respectively [26]. If this statistic exceeds the corresponding value of the F distribution value ($F_{\alpha,p,n-p-1}$), the response model is considered significant in relation to random error. A second statistic, the coefficient of multiple determination, defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}, \tag{2.9}$$

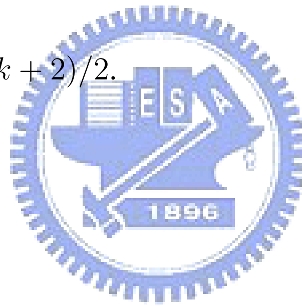
it measures the amount of reduction in variability of the response y achieved, using the

input factors $x_1, x_2, \ldots, x_k$. R^2 varies from zero to one with a value of one being ideal [26].

Because the R^2 value measures the "proportion of total variation explained by the constructed regression model $X\hat{\beta}$", a higher R^2 value indicates a better fit of the regression model. It can be shown that R is the correlation between $y = (y_i)_{i=1}^N$ and $\hat{y} = (\hat{y}_i)_{i=1}^N$ and thus is called the *multiple correlation coefficient* [34]. The residual mean square is commonly referred to as the *mean-square error* (MSE) and is an estimate $\hat{\sigma}^2$ for σ^2, i.e.,

$$\hat{\sigma}^2 = (y - X\hat{\beta})^T (y - X\hat{\beta})/(N - p - 1), \tag{2.10}$$

where p is equal to $(k+1)(k+2)/2$.



2.5 Applications of the Developed Models

2.5.1 Variable Selection

In response surface work it is customary to construct the **full** model corresponding to the situation at hand. That is, in steepest ascent we usually build the full first-order model, and in the analysis of a second-order model we usually construct the full quadratic. An experimenter may encounter situations where the full model may not be appropriate; that is, a model based on a subset of the regressors in the full model may be superior. **Variable selection** or **model-building** techniques may be used to identify the best subset of regressors to include in a regression model.

Variable selection is determined by statistical analysis of the generated response surface models. Input factors showing a significant effect on an individual response can be systematically determined using statistical techniques. Variations of these significant input factors will produce the greatest fluctuations in device performance. This analysis is extremely useful in understanding what areas of manufacturing require greater control.

(1) Half-normal plot and t test

As screening design, a technique for identifying significant model terms can be based on the half-normal plot of model coefficients. This method is originally proposed for analyzing two-level factorial experiments applicable in cases where no degrees of freedom

are available for estimating the variance of an error term. The effects are plotted on half-normal probability paper, those standing apart being identified as potentially real effects [4]. Probability plotting may also be used for experiments having three level. One approach is to express the effects with linear and quadratic components, and construct the normal probability plot of those components standardized to have the same variance [1][5].

The half-normal plots are informal graphical methods involving visual judgment. A formal test of effect significance is called t test for the least squares estimate $\hat{\beta}$ in Equ.(2.4). It can be shown that the least squares estimate $\hat{\beta}$ has a multivariate normal distribution with mean vector β and variance-covariance matrix $\sigma^2(X^TX)^{-1}$, i.e.,

$$\hat{\beta} \sim MN(\beta, \sigma^2(X^TX)^{-1}), \tag{2.11}$$

where MN stands for multivariate normal. The (i, j)th entry of the variance-covariance matrix is $Cov(\hat{\beta}_i, \hat{\beta}_j)$ and the jth diagonal element is $Cov(\hat{\beta}_j, \hat{\beta}_j) = Var(\hat{\beta}_j)$. Therefore, the distribution for the individual $\hat{\beta}_j$ is $N(\beta_j, \sigma^2_{jj}(X^TX)^{-1})$, which suggests that for testing the null hypothesis

$$H_0 : \beta_j = 0, \tag{2.12}$$

the following t statistic be used:

$$\frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2_{jj}(X^TX)^{-1}}} \sim t_{N-p-1} \ \texttt{(under } H_0\texttt{)}. \tag{2.13}$$

Under H_0, it has a t distribution with $N - p - 1$ degrees of freedom.

(2) Stepwise regression

Alternative of variable selection is called *stepwise regression*. It is one of various methods for evaluating only a small number of subset regression models by either adding or deleting regressors one at a time. Stepwise regression is a popular combination of procedures *forward selection* and *backward elimination* [26].

The procedure of the forward selection begins with the assumption that there are no regressors in the model other than the intercept. An effort is made to find an optimal subset by inserting regressors into the model one at a time. The first regressor selected for entry into the equation is the one that has the largest simple correlation with the response variable y. Suppose that this regressor is x_1. This is also the regressor that will produce the largest value of the F-statistic for testing significance of regression. This regressor is entered if the F-statistic exceeds a preselected F-value, say F_{IN} (or F-to-enter). The second regressor chosen for entry is the one that now has the largest correlation with y after adjusting for the effect of the first regressor entered (x_1) on y. We refer ro these correlations as *partial correlations*. They are the simple correlations between the residuals from the regression $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1$ and the residuals from the regressions of the other candidate regressors on x_1, say $\hat{x}_j = \hat{\alpha}_{0j} + \hat{\alpha}_{1j} x_1$, $j = 2, 3, \ldots, K$.

Suppose that at step 2 the regressor with the highest partial correlation with y is x_2. This implies that the largest partial F-statistic is

$$F = \frac{SS_R(x_2|x_1)}{MS_E(x_1, x_2)}. \tag{2.14}$$

If this F-value exceeds F_{IN}, then x_2 is added to the model. In general, at each step the regressor having the highest partial correlation with y (or equivalently the largest partial F-statistic given the other regressors already in the model) is added to the model if its partial F-statistic exceeds the preselected entry level F_{IN} [26]. The procedure terminates either when the partial F-statistic at a particular step does not exceed F_{IN} or when the last candidate regressor is added to the model.

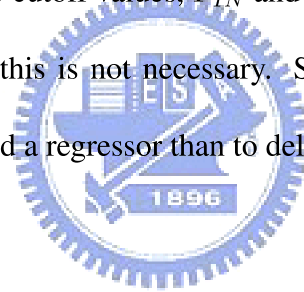
Forward selection begins with no regressors in the model and attempts to insert variables until a suitable model is obtained. Backward elimination attempts to find a good model by working in the opposite direction. That is, we begin with a model that includes all K candidate regressors. Then the partial F-statistic (or a t-statistic, which is equivalent) is computed for each regressor as if it is the last variable to enter the model. The smallest of these partial F-statistics is compared with a preselected value, F_{OUT} (or F-to-move); and if the smallest partial F-value is less than F_{OUT}, that regressor is removed from the model. Now a regression model with $K - 1$ regressors is constructed, the partial F-statistics for this new model calculated, and the procedure repeated. The backward elimination algorithm terminates when the smallest partial F-value is not less than the preselected cutoff value F_{OUT} [26].

Backward elimination is often a very good variable selection procedure. It is particularly favored by analysts who like to see the effect of including all the candidate regressors,

just so that nothing obvious will be missed.

The two procedures described above suggest a number of possible combinations. One of the most popular is the stepwise regression algorithm. This is a modification of forward selection in which at each step all regressors entered into the model previously are reassessed via their partial F- or t-statistics. A regressor added at an earlier step may now be redundant because of the relationship between it and regressors now in the equation. If the partial F-statistic for a variable is less than F_{OUT}, that variable is dropped from the model.

Stepwise regression requires two cutoff values, F_{IN} and F_{OUT}. Several analysts prefer to choose $F_{IN} = F_{OUT}$, although this is not necessary. Sometimes we choose $F_{IN} > F_{OUT}$, making it more difficult to add a regressor than to delete one [26].



2.5.2 Fabrication Optimization

Optimal process input conditions are determined by two methods: graphical analysis and numerical analysis. Graphical analysis is a vision analysis using contours overlapping [24].

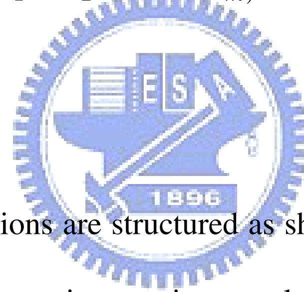
A relatively straightforward approach to optimize several responses that works well when there are only a few process variables is to overlay the contour plots for each response. Figure 2.5 presents the individual contour and response surface plots for threshold voltage and subthreshold slope, and two factors are the punch-through implant energy and the threshold voltage implant dose. Figure 2.6 shows an overlay plot for the two response, with $0.36 \leq y_1$ (threshold voltage) ≤ 0.362, y_2 (subthreshold slope) ≥ 14.89. If these boundaries represent important conditions that must be met by the process, then as the blue portion of Figure 2.6 shows, there are a number of combinations of punch-through implant energy and threshold voltage implant dose that will result in a satisfactory process.

When there are more than three design variables, overlaying contour plots becomes awkward, because the contour plot is two-dimensional, and other $k-2$ of the the design variables must be held constant to construct the graph [24]. Often a lot of trial and error is required to determine which factors to hold constant and what levels to select to obtain the best view of the surface. Therefore, there is practical interest in more formal optimization methods for multiple responses.

A popular approach is to formulate and solve the problem as a constrained optimization

problem. There are many numerical techniques that can be used to solve this problem. A useful approach to optimization of multiple responses is to use the simultaneous optimization technique popularized by Derringer and Suich (1980) [24]. Their procedure makes use of desirability functions. The general approach is to first convert each response y_i into an individual desirability function d_i that varies over the range $0 \leq d_i \leq 1$, where if the response y_i is at its goal or target, then $d_i = 1$, and if the response is outside an acceptable region, $d_i = 0$. Then the design variables are chosen to maximize the overall desirability

$$D = (d_1 \times d_2 \times \ldots \times d_m)^{1/m}, \tag{2.15}$$



where there are m responses [24].

The individual desirability functions are structured as shown in Figure 2.7. If the objective or target T_i for the ith response y_i is a maximum value,

$$d_i = \begin{cases} 0, & \hat{y}_i < L_i \\ (\frac{\hat{y}_i - L_i}{T_i - L_i})^s, & L_i \leq \hat{y}_i \leq T_i \\ 1, & \hat{y}_i > T_i \end{cases}, \tag{2.16}$$

when the weight $s = 1$, the desirability function is linear. Choosing $s > 1$ places more emphasis on being close to the target value, and choosing $0 < s < 1$ makes this less

important. If the target for the response y_i is a minimum value,

$$d_i = \begin{cases} 1, & \hat{y}_i < T_i \\ (\frac{U_i - \hat{y}_i}{U_i - T_i})^s, & T_i \leq \hat{y}_i \leq U_i \\ 0, & \hat{y}_i > U_i \end{cases}, \tag{2.17}$$

The two-sided desirability function shown in Figure 2.7(c) assumes that the target is located between the lower (L_i) and upper (U_i) limits, and is defined as

$$d_i = \begin{cases} 0, & \hat{y}_i < L_i \\ (\frac{\hat{y}_i - L_i}{T_i - L_i})^s, & L_i \leq \hat{y}_i \leq T_i \\ (\frac{U_i - \hat{y}_i}{U_i - T_i})^t, & T_i \leq \hat{y}_i \leq U_i \\ 0, & \hat{y}_i > U_i \end{cases}, \tag{2.18}$$

The Design-Expert software package is used to solve the optimization problem using the desirability function approach.

2.5.3 Process Sensitivity Analysis

The objective is to investigate the sensitivity of a number of key responses to varying a set of critical process parameters. Once a quadratic model for each response has been obtained, then it can be 'interrogated' by plotting the distribution of each response to a Gaussian variation of the input factors [7]. As we got the optimal process input conditions,

we can use various simulation methods to check the manufacturing sensitivity analysis. Samples generating techniques may be employed to ascertain the statistical distribution of each response for a specified set of variations in the input factors. In this approach, random values for each input factor are selected and the corresponding responses predicted using the response surface models. The input factors are assumed to be normally distributed about their mean (for this work the optimal process input condition values). The standard deviation for each factor must be determined experimentally, or be set as a percentage of its mean values. Analysis of the distributions provides an estimate of the amount of response variation expected under normal manufacturing conditions [33].

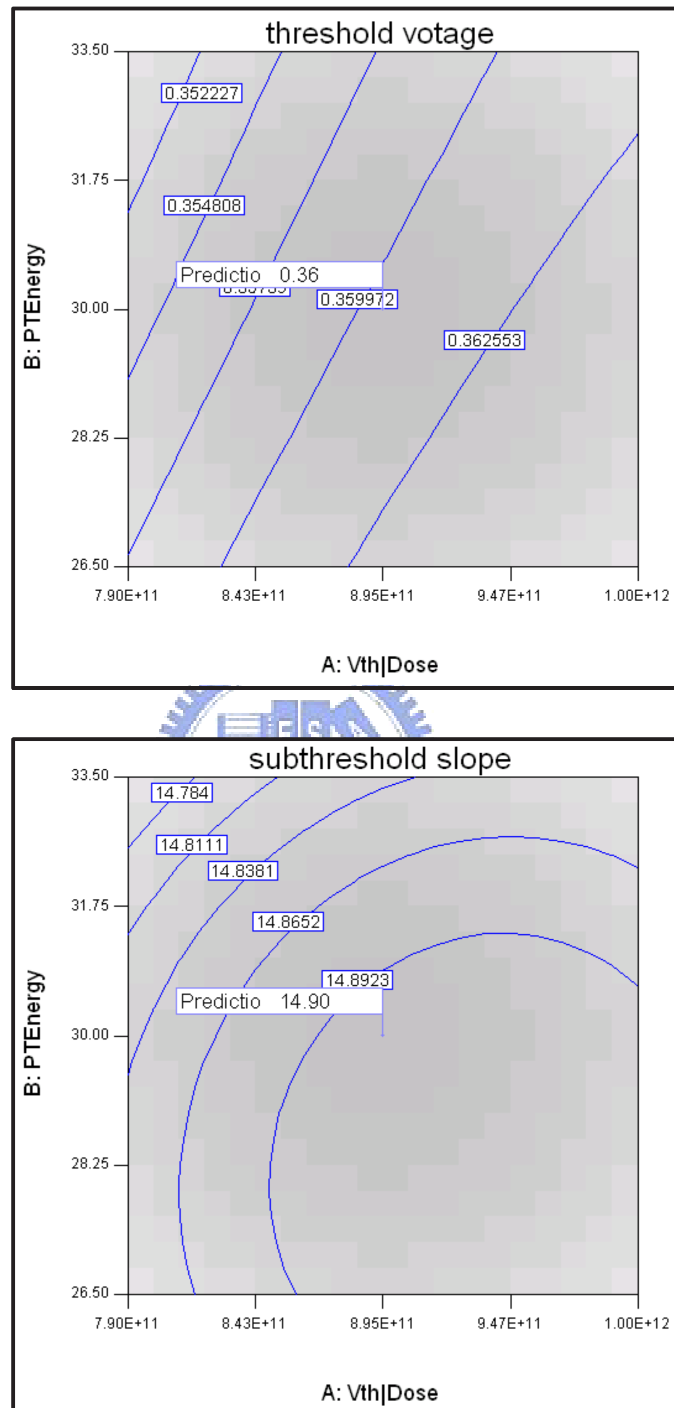


Figure 2.5: The contour plots of the threshold voltage and the subthreshold slope. Two factors are the threshold implant dose and the punch-through implant energy.

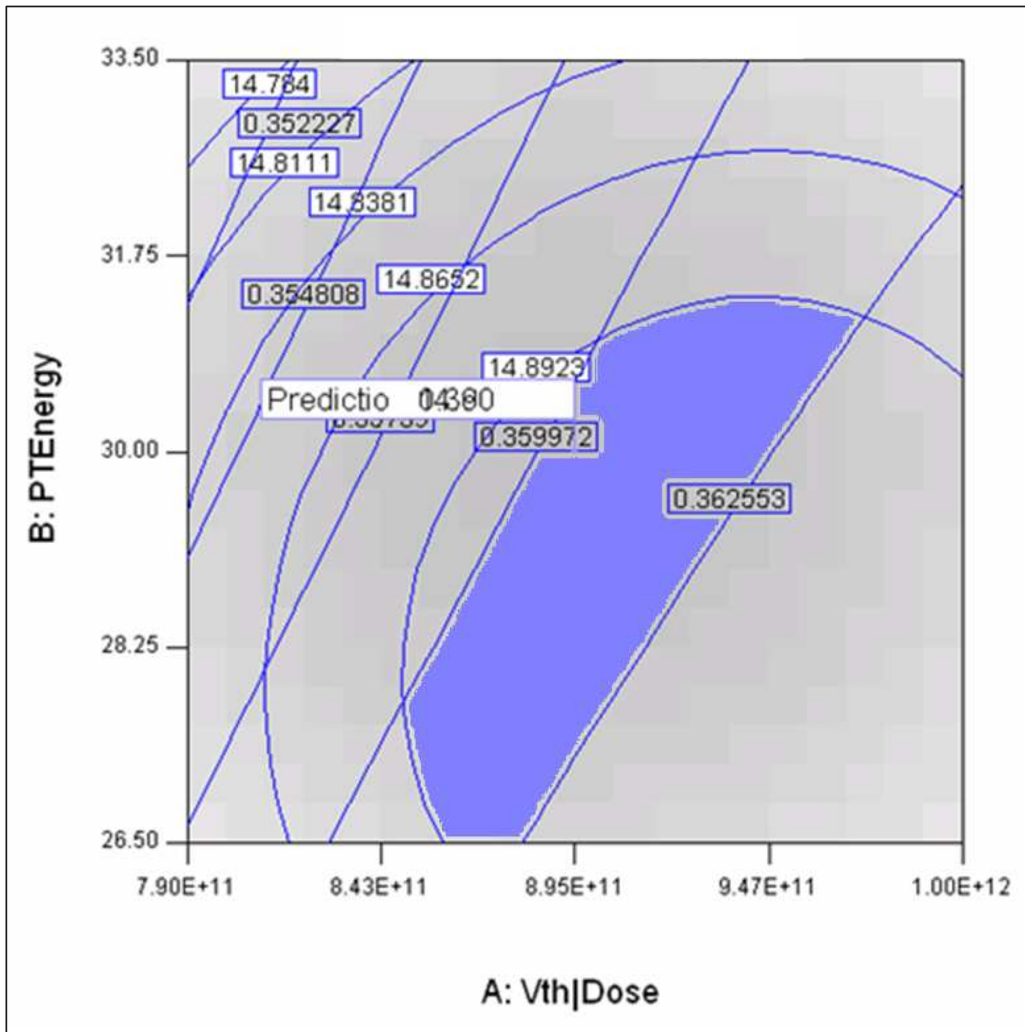
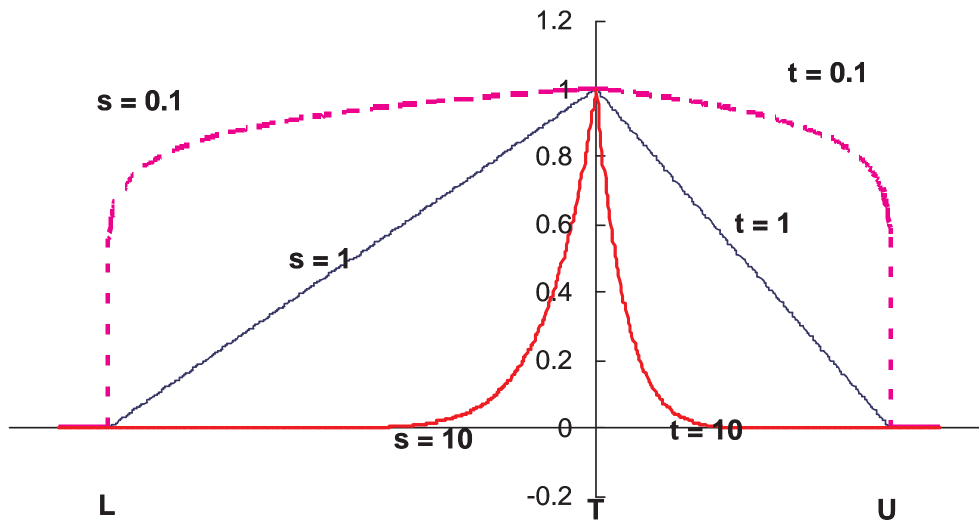
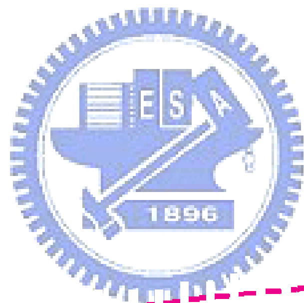
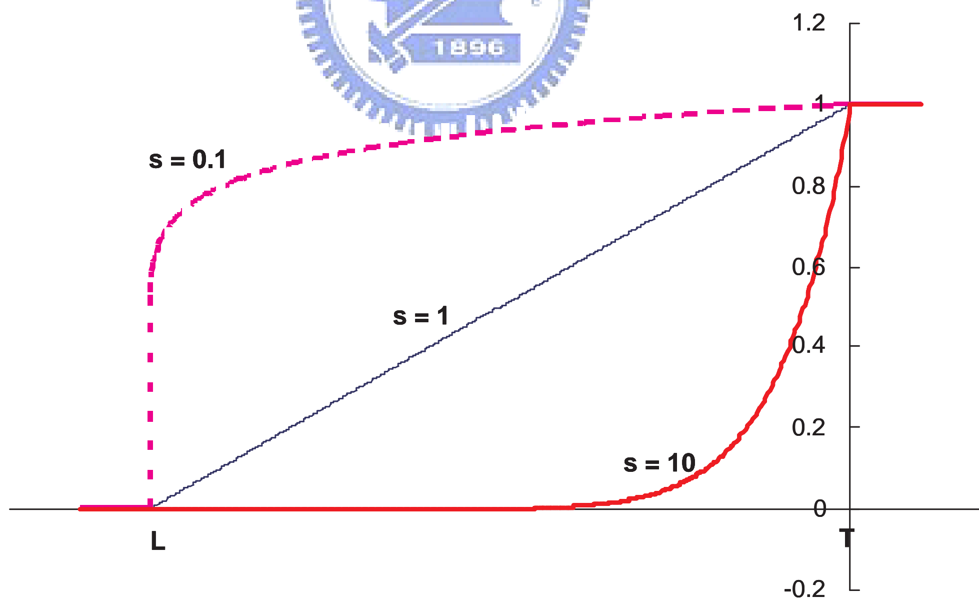


Figure 2.6: Region of the optimum found by overlaying threshold voltage and subthreshold slope response surface. The deep gray part means the optimal region through the contour overlap.


1.2
1
0.8
0.6
0.4
0.2
0
-0.2
s = 0.1
s = 1
s = 10
t = 0.1
t = 1
t = 10
L
T
U


E S A
1896


1.2
1
0.8
0.6
0.4
0.2
0
-0.2
s = 0.1
s = 1
s = 10
L
T

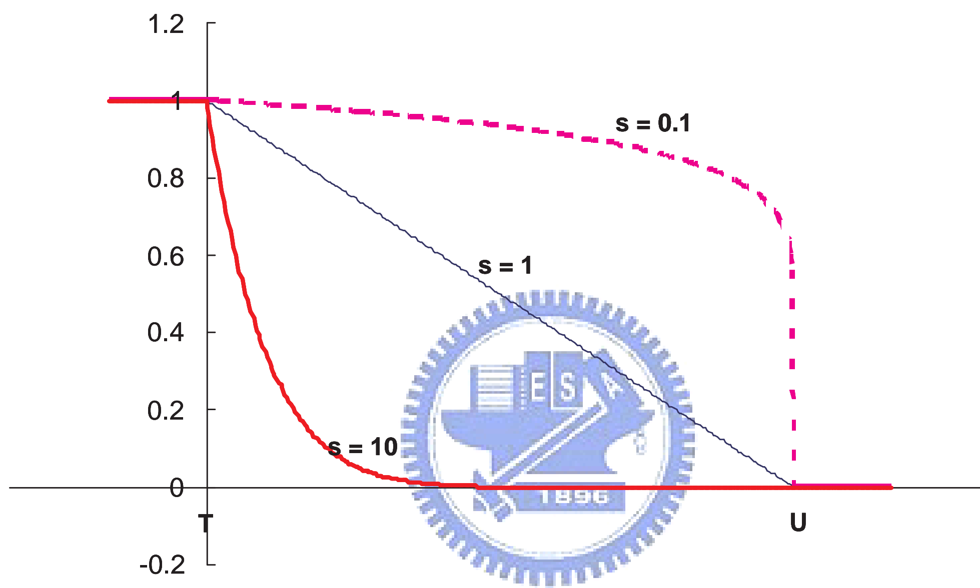


Figure 2.7: Individual desirability functions for simultaneous optimization.

2.5.4 The Nominal-the-Best Problem

In the real fabricating integrated circuit (IC) devices situation, not a TCAD simulation environment, the experimenters needed to find process factors that could be set to minimize the response (i.e., threshold voltage) inconsistency while maintaining average value of the response as close to the nominal value as possible. Because of the dual requirements on the average and variation of the response, analysis of the response data should focus on both aspects. In quality engineering this is referred to as the **nominal-the-best** problem [34].

And the experiment can be replicated, which means that the experimental plan is carried out two or more times. An unreplicated experiment is also said to have a single replicate. Unreplicated experiments are often performed when runs are expensive. Replication is also unnecessary for computer experiments because repeated computer runs with the same input give the same output.

For a replicated experiment, to achieve the two experimental objectives stated above, the mean (for location) and variance (for dispersion) need to be modelled separately as functions of the experimental factors.

Let y_{ij} denote the response of the jth replicate for the ith run (total N runs). Using the experimental data, $\overline{y}_i$ and s_i^2, the sample mean and sample variance for the ith run using the n_i replicates of the ith run are computed first, where

$$\overline{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}, \quad s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(y_{ij} - \overline{y}_i)^2, \quad and \quad i = 1, \ldots, N. \tag{2.19}$$

Here the sample mean $\overline{y}_i$ and the sample variance s_i^2 are used to measure respectively the **location** and **dispersion** of the data.

Then we can use following two equations to construct the response surface model [34]:

$$\overline{y}_m \approx \beta_0+\sum_{i=1}^{k}\beta_i x_{mi}+\sum_{i=1}^{k}\beta_{ii}x_{mi}^2+\sum_{i=1}^{k}\sum_{i\neq j}^{k}\beta_{ij}x_{mi}x_{mj}+\varepsilon_m, \quad for \quad m=1,\ldots,N, \tag{2.20}$$

and

$$lns_m^2 \approx \alpha_0+\sum_{i=1}^{k}\alpha_i x_{mi}+\sum_{i=1}^{k}\alpha_{ii}x_{mi}^2+\sum_{i=1}^{k}\sum_{i\neq j}^{k}\alpha_{ij}x_{mi}x_{mj}+\delta_m, \quad for \quad m=1,\ldots,N. \tag{2.21}$$

As we got these two equations, the analysis strategy in this work is applicable. In this situation when we want to get the optimal process condition values, the selection of levels can be done using the following procedure:

(i) Select levels of several factors to minimize Var(y) (or $ln\ \sigma^2$).

(ii) Select the level of a factor not in (i) to move E(y) closer to the target.

To measure the dispersion effects of different factors, should the effect formulas in Equ.(2.21) be applied directly to the sample variances s_i^2? The methods described in this work assume that the analyzed quantities are nearly normal so that a direct application is not appropriate because $(n_i-1)\ s_i^2$ has a chi-squared distribution. A commonly used approach is to take the log transformation of s_i^2 and then analyze the $ln\ s_i^2$ using standard method, where ln denotes *natural logarithm.* A primary reason for the log transformation is that it maps positive values to real (both positive and negative) values, and by taking its inverse

transformation, any predicted value on the log scale will be transformed back to a positive value on the original scale. Let $z_i = ln\ s_i^2$ and let $\hat{z}_i$ denote the predicted value of $ln\ \sigma_i^2$ for a new factor setting based on a regression model applied to the $ln\ s_i^2$ data. By taking the inverse transformation, exp($\hat{z}_i$) is obtained as the predicted value of σ_i^2 for the new factor setting. Another justification stems from the nature of physical laws. Many physical laws have multiplicative relationships. Log transformation converts a multiplicative relationship into an additive relationship, which is much easier to model statistically.

The appropriateness of the log transformation can be assessed by the transformed response's proximity to the normal distribution. Noting that

$$(n_i - 1)s_i^2 = \sum_{j=1}^{n_i}(y_{ij} - \overline{y}_i)^2 \sim \sigma_i^2 \chi_{n_i-1}^2, \tag{2.22}$$

where χ_f^2 is the chi-squared distribution with f degrees of freedom, then taking logs yields

$$lns_i^2 = ln\sigma_i^2 + ln\frac{1}{n_i - 1}\chi_{n_i-1}^2. \tag{2.23}$$

Letting X denote the random variable $f^{-1}\chi_f^2$, $ln(X)$ has approximately mean zero since $E(X) = 1$ and $Var(X) = 2f^{-1}$. Its variance can be approximated by the following general formula:

$$Var(h(X)) \approx [h'(E(X))]^2 Var(X), \tag{2.24}$$

where h is a smooth function of a random variable X. For $h(X) = lnX$,

$$Var(ln(X)) \approx (h'(1))^2 2f^{-1} = 2f^{-1}, \tag{2.25}$$

which is not a constant but has a narrow range of values for $f = n_i - 1 \geq 9$. (See Bartlett and Kendall, 1946.) From the previous formula and Equ.(2.25), the distribution of $ln\ s_i^2$ can be approximated by

$$N(ln(\sigma_i^2), 2(n_i - 1)^{-1}). \tag{2.26}$$

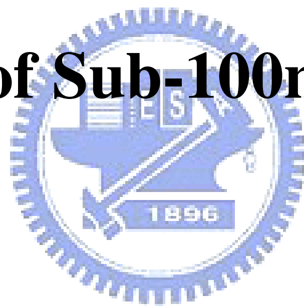
From Equ.(2.26), it can be seen that the log transformation has the unique property that allows σ_i^2 to appear only in the mean of the normal approximation, thus leaving the variance of the normal approximation independent of σ_i^2. In this sense, $ln\ s_i^2$ is a variance-stabilizing transformation.

2.6 Summary

In this chapter, we discuss the methodology about this work. First the methodology flowchart has been shown and we can know the outline about this chapter. Next we talked about each steps in detail. Screening design is the first step in this work, it could help us to find the important factors from a lot of variables. After selecting a smaller group of factors, we do design of experiment. The traditional design matrix is based on the CCD, and there are 3 special forms (CCC, CCF, and CCI designs) of it. Another choice of design matrix is small composite design (SCD), and using this design, the run number is smaller than CCD. After the step of design of experiment, the topic of constructing the response surface model is introduced, and through these models we could find the relationship between the responses and the factors. Finally several applications in using response surface model are discussed, and they are variable selection, optimization and sensitivity analysis, and solving nominal-the-best problems.

Chapter 3

Fabrication Process and Electrical Characteristics of Sub-100nm MOSFETs



In this chapter, we provide a newer process technology , 90nm n-type MOSFET, to be the main example.

A 90nm NMOSFET process, representative of current MOSFET technology, is implemented with the aid of 2-D process (ISE-Floops) and device (ISE-Dessis) simulations. Total 13 main process flows, which could be separated into 45 detailed process simulation steps, have shown on Table 3.1, and there are more than one hundred numerical variables

could be adjusted. Due to the large number of process and device simulations necessary to produce the response data, ISE simulator tool is used to integrate the process and device simulation tools, construct the experimental designs, and automate simulation of the experimental points. The recipe of the 90nm NMOSFET process flows would be shown in the end of this chapter.

3.1 The Investigated 90nm NMOSFET

After we get the experiment data, Design Expert software is used to generate the response surface models and performing process optimization. Floops and Dessis are incorporated within ISE to perform the necessary 2-D process and device simulations. Those results are then used within a separated mathematical program to generate the reduced response surface models and to conduct the sensitivity analysis.

A comprehensive effort is undertaken in realizing process and device models for NMOSFET devices, with an emphasis on defining and using the necessary models for realistic simulation of the process and device conditions. Carrier heating and lattice heating due to current flow and carrier recombination are accounted for using a self-consistent solution of the Poisson, carrier continuity, energy-balance, and heat flow equations. An important part of this work consists of the incorporation of quantum mechanical effects within the

simulation modules developed for the NMOSFET device. This is necessary to more accurately determine the device characteristics. Figure 3.1 provides a device structure using the original nominal process parameter conditions. Table 3.1 shows the five physical quantity values of the nominal case.

For the evaluation of the 90nm NMOSFET technology, five device characteristics of interest are defined: threshold voltage (Vth), subthreshold slope (SS), off-state current (Ioff), drain saturation current (Idsat), and drain-induced barrier lowering (DIBL). In the following analysis, data transformation are used to improve the model accuracy for NMOSFET log(Ioff) and log(DIBL). Threshold voltage is determined using the linear extrapolation method. Subthreshold slope is defined as the inverse of the difference of drain current when gate voltage changes to ten order. Off-state current is defined as drain current (Id) observed at a gate voltage (Vg) of 0 V and a drain voltage (Vd) of 0.1 V. Drain saturation current is defined as Id and Vg and Vd = 1.2 V. Drain-induced barrier lowering is defined as the difference in Vth determined at Vd = 0.1 V and 1.2 V.

3.2 Summary

A 90nm n-type MOSFET fabrication process is the applied example. In this chapter we give the table of process simulation steps in ISE simulator tool. In this work, we will study the relationship between the device performances and its fabrication process. Five

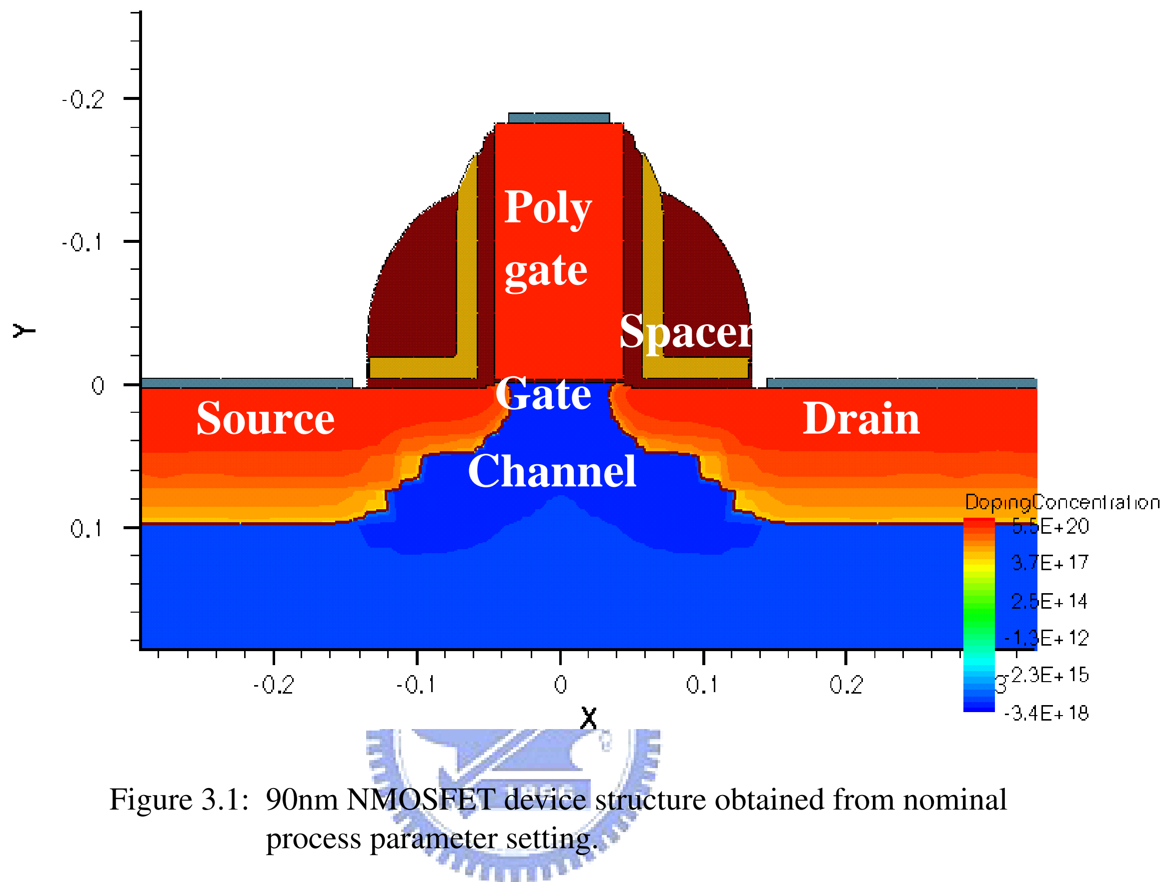


Figure 3.1: 90nm NMOSFET device structure obtained from nominal process parameter setting.

physical quantities that we are interested in are defined here, they are **threshold voltage**, **subthreshold slope**, **off-state current**, **on-state current**, and **DIBL**.

Table 3.1: The physical quantities in the nominal case.

Responses	The physical quantities of the nominal case
Vth (V)	0.23598393
SS (V/dec.)	0.07985328
log(Ioff) (A/μm)	-12.68845194
Idsat (A/μm)	0.0053344
log(DIBL) (V)	-1.576629755

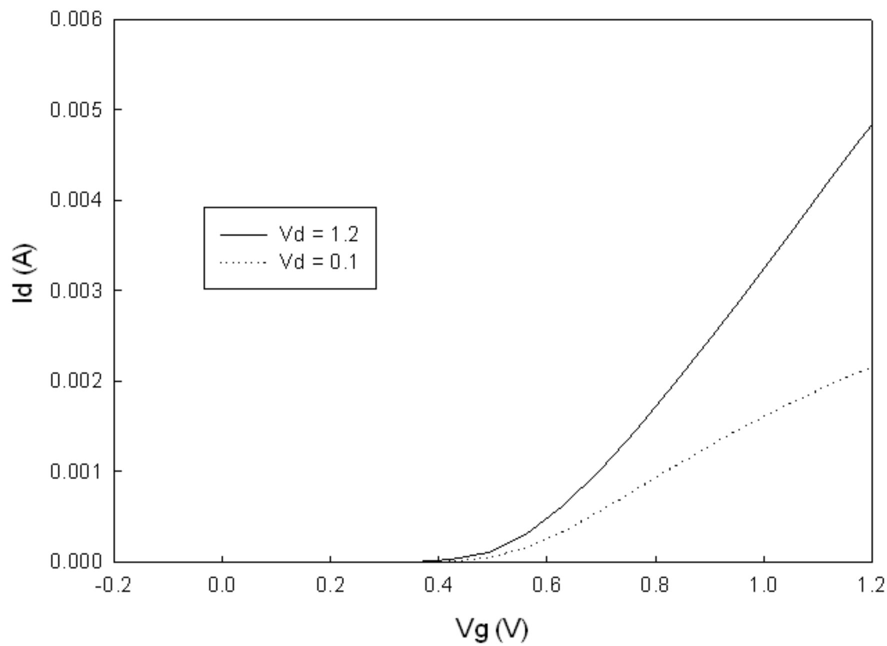


Figure 3.2: Id-Vg curves of the 90nm NMOSFET device structure obtained from the nominal case process parameter setting.

Table 3.2: Simulation steps and initial parameter settings of the 90nm NMOSFET process flows for the nominal case.

Step1: Substrate boron, 0 $/cm^3$ concentration, 1 ohm-cm resistivity
Step2: Well-channel **1. deposit:** oxide, 10nm thickness **2. implant:** boron, 3e+13 $/cm^2$ concentration, 300 keV energy, 7° tilt angle -90° rotation angle **3. implant:** boron, 2e+13 $/cm^2$ concentration, 200 keV energy, 7° tilt angle -90° rotation angle **4. implant:** boron, 1e+13 $/cm^2$ concentration, 100 keV energy, 7° tilt angle -90° rotation angle **5. implant:** boron, 8e+13 $/cm^2$ concentration, 30 keV energy, 7° tilt angle -90° rotation angle **6. anneal:** time: 11.2 - 20.0 - 8.4 sec., temperature: 500 - 1060 - 1060 - 800 °C at 1 atm pressure, and use nitrogen gas with 3 1/min. **7. etch:** oxide with strip etching, and overetch 10%
Step3: Gate oxidation 10 mins growth time, 760 °C growth temperature, use oxygen gas with 1 1/min
Step4: Poly-gate **1. deposit:** poly, 180nm thickness used isotropic deposition type **2. pattern:** layer: poly, polarity: light-field, 20nm thickness **3. etch:** poly with anisotropic etching, 185 nm thickness **4. etch:** oxide with anisotropic etching, 3 nm thickness **5. etch:** silicon with anisotropic etching, 1.5 nm thickness **6. etch:** resist with strip etching, and overetch 10%
Step5: Poly-reoxidation **1. deposit:** oxide, 1nm thickness with isotropic deposition **2. anneal:** time: 4.0 - 4.0 - 5.0 - 4.0 sec., temperature: 600 - 800 - 900 - 900 - 800 °C, at 1 atm pressure, use oxygen gas with 1.0 1/min.

Step6: Offset-spacer **1. deposit:** oxide, 5nm thickness with isotropic deposition **2. etch:** oxide with anisotropic etching, 15nm thickness **3. etch:** silicon with anisotropic etching, 2 nm thickness **4. deposit:** oxide, 1.5nm thickness with isotropic deposition
Step7: Halo-implant **1. implant:** boron, 2.3e+13 $/cm^2$ concentration, 12 keV energy, 30° tilt angle 0° rotation angle **2. implant:** boron, 1.15e+13 $/cm^2$ concentration, 12 keV energy, 30° tilt angle 90° rotation angle **3. implant:** boron, 1.15e+13 $/cm^2$ concentration, 12 keV energy, 30° tilt angle 270° rotation angle
Step8: Extension-implant **implant:** arsenic, 1.6e+15 $/cm^2$ concentration, 1.2 keV energy, 0.0° tilt angle -90° rotation angle
Step9: Spike-anneal-1 **anneal:** time: 3.0 - 1.0 - 0.5 - 3.3 sec., temperature: 500 - 950 - 1000 - 1000 - 800 °C, at 1 atm pressure, use nitrogen gas with 3.0 1/min.
Step10: Offset-spacer **1. deposit:** oxide, 5nm thickness with isotropic deposition **2. deposit:** nitride, 15nm thickness with isotropic deposition **3. deposit:** oxide, 61nm thickness with isotropic deposition **4. anneal:** time: 1200 sec., temperature: 720 °C, at 1 atm pressure with nitrogen gas with 3 1/min. **5. etch:** oxide with anisotropic etching, 110nm thickness **6. etch:** nitride with anisotropic etching, 35nm thickness **7. etch:** nitride with isotropic etching, 1nm thickness **8. etch:** oxide with anisotropic etching, 6nm thickness **9. etch:** resist with strip etching, overetch 10% **10. deposit:** oxide, 10nm thickness with isotropic deposition

Step11: S/D-implant
1. implant: phosphorus, 5.1e+14 $/cm^2$ concentration, 4 keV energy, 0° tilt angle -90° rotation angle **2. implant:** arsenic, 3.5e+15 $/cm^2$ concentration, 16 keV energy, 0° tilt angle -90° rotation angle
Step12: Spike-anneal-2
anneal: time: 2.0 - 1.0 - 1.0 - 0.5 - 1.0 - 1.0 - 4.5 sec., temperature: 600 - 920 - 1010 - 1040 - 1055 - 1045 - 1015 - 750 °C, at 1 atm pressure, use nitrogen gas with 3.0 1/min.
Step3: Metallization
(Purpose of this simplified back-end part is only the formation of contacts.) **1. etch:** oxide with anisotropic etching, 7nm thickness **2. deposit:** metal, 7nm thickness with isotropic deposition **3. pattern:** layer: gate contact, polarity: light-field, 20nm thickness **4. pattern:** layer: source/drain, polarity: light-field, 20nm thickness **5. etch:** aluminum with anisotropic etching, 100nm thickness **6. etch:** resist with strip etching, overetch 10%

Chapter 4

Screening Design

In this chapter the results of screening design method will be shown. A 90nm NMOSFET process, representative of current MOSFET technology, is implemented with the aid of 2-D process (ISE-Floops) and device (ISE-Dessis) simulations. The detailed simulation steps have shown in Chapter 3. Twenty-three initial process conditions are identified from the engineering opinions as modified candidates and five responses are studied for factor screening. Twenty-four run Plackett-Burman design is generated to test factor significance for the NMOSFET devices.

4.1 Results of the Screening Design

After we get each main effect of 23 factors corresponding to 5 responses, we made 5 half-normal plots to check the significance of these factors. The half-normal plots of these five responses are shown on Figs. 4.1 - 4.5, and the analysis result is shown on the table 4.1. The numbers in the table present the order of significant effect, 1 means the most important to the corresponding response. We select the factor which are used for 'union' concept. Finally we choose seven factors in the bold rows, and they are:

(1) threshold voltage implant dose;

(2) threshold voltage implant energy;

(3) oxide growth time;

(4) oxide growth temperature;

(5) gate length;

(6) punch-through implant implant dose2;

(7) and punch-through implant energy.

Thus in this screening design, we selected 7 factors from 23 factors eventually. In the last, the experiment levels in each factors are provided in Table 4.2. The factors of the source/drain doping such as the source/drain implant dose and energy are not significant for this fabrication process; however, they are important to the nanoscale devices. In this work we follow the statistical results, but we could add these factors into our results to

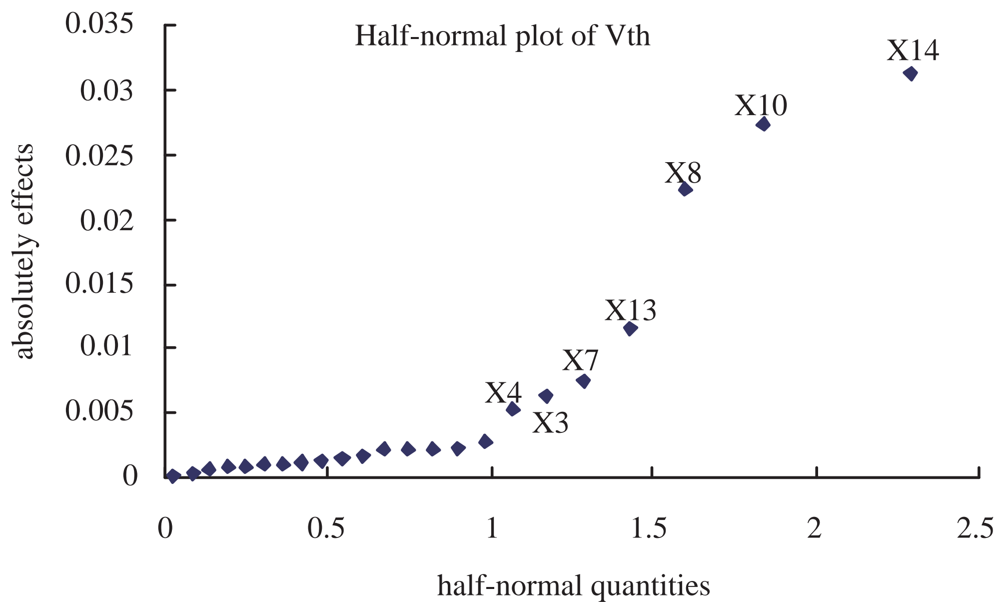


Figure 4.1: A plot of the half-normal effect plot for the threshold voltage.

study more.

4.2 Summary

In this chapter, seven significant factors have been selected from the screen design. They are **threshold voltage implant dose**, **threshold voltage implant energy**, **oxide growth time**,

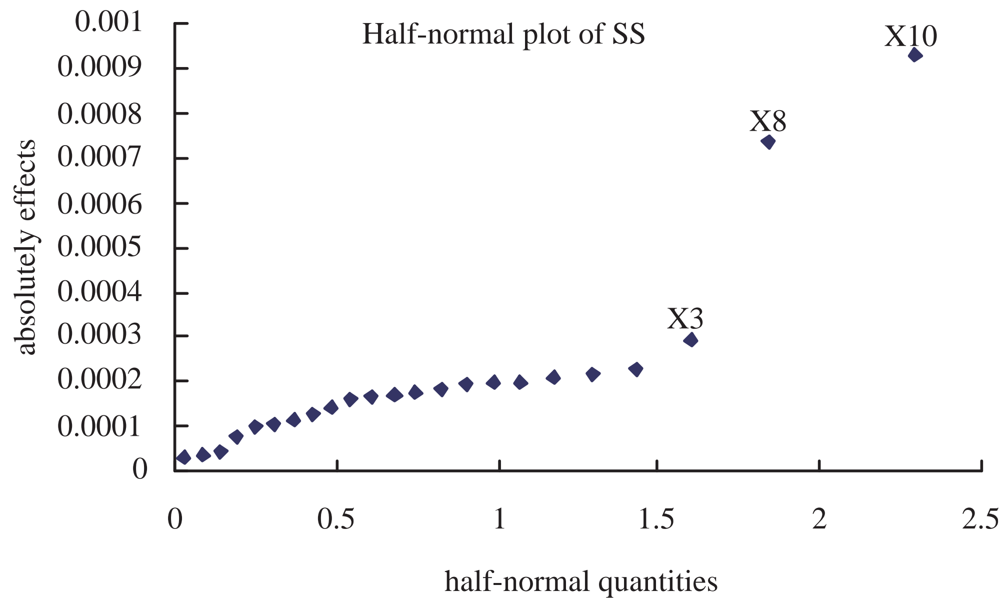


Figure 4.2: A plot of the half-normal effect plot for the subthreshold slope.

1896

oxide growth temperature, **gate length**, **punch-through implant dose2**, and **punch-through implant energy**. We use a 24-run Plackett-Burman design to calculate 23 factors, and use half-normal plots to check the significant effects. The experiment levels of 23 factors are shown on Table 4.2.

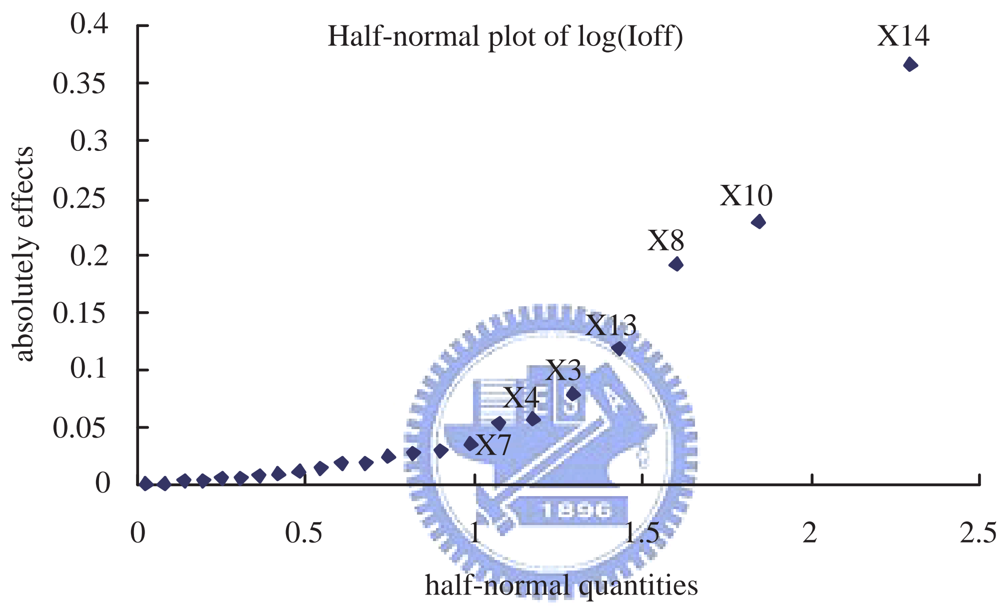


Figure 4.3: A plot of the half-normal effect plot for the off-state current using log transformation.

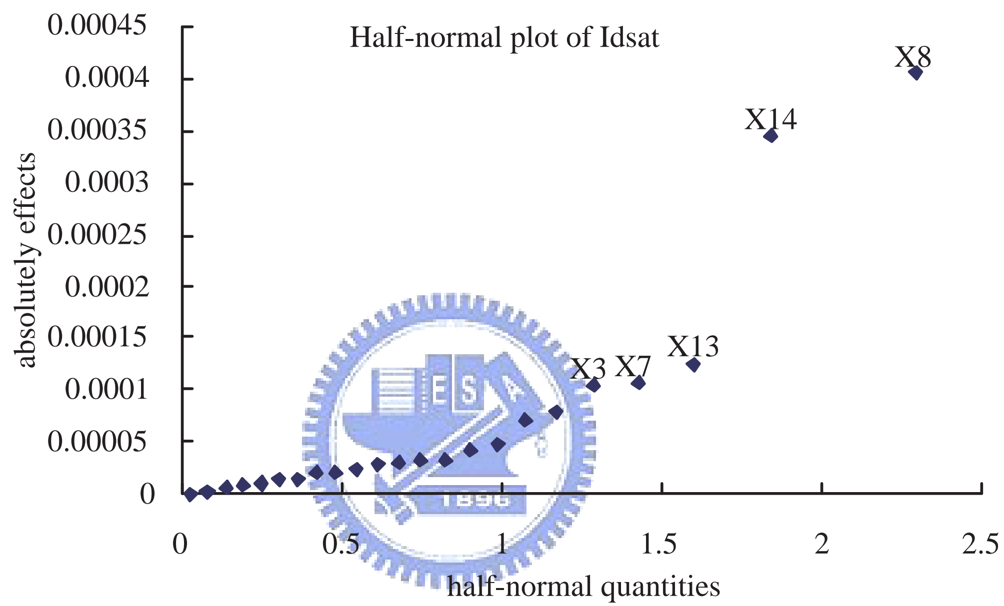


Figure 4.4: A plot of the half-normal effect plot for the on-state current.

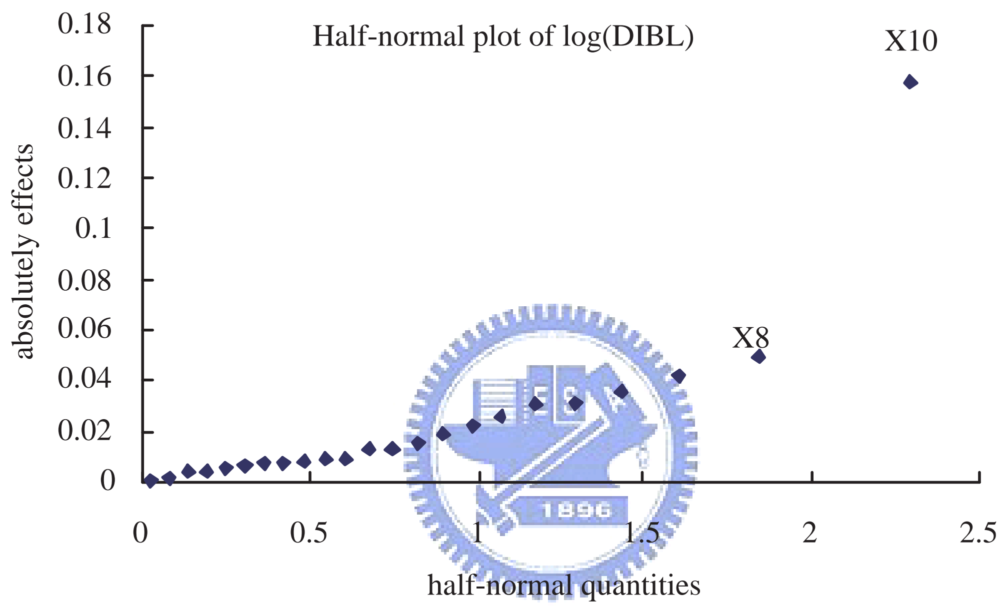


Figure 4.5: A plot of the half-normal effect plot for the drain-induced barrier lowing using log transformation.

Table 4.1: A list of the computed results of the screening design.

	Vth	SS	log(Ioff)	Idsat	log(DIBL)
X1: WELL implant dose					
X2: WELL implant energy					
X3: threshold voltage implant dose	6	3	5	5	
X4: threshold voltage implant energy	7		6		
X5: WELL-channel anneal time					
X6: WELL-channel anneal temp.					
X7: oxide growth time	5		7	4	
X8: oxide growth temperature	3	2	3	1	2
X9: poly-oxide thickness					
X10: gate length	2	1	2		1
X11: Offset spacer thickness					
X12: punch-through implant dose1					
X13: punch-through implant dose2	4		4	3	
X14: punch-through implant energy	1		1	2	
X15: extension implant dose					
X16: extension implant energy					
X17: extension implant anneal time					
X18: extension implant anneal temp.					
X19: Source/Drain spacer thickness					
X20: Source/Drain implant dose					
X21: Source/Drain implant energy					
X22: Source/Drain anneal time					
X23: Source/Drain anneal temp.					

Table 4.2: Experiment levels for all factors.

Factor name	**level 0**	**level 1**	**level -1**
X1: WELLdose (cm^2)	3.00E+13	3.10E+13	2.90E+13
X2: WELLenergy (keV)	300	310	290
X3: VTdose (cm^2)	8.00E+12	8.80E+12	7.20E+12
X4: VTenergy (keV)	30	31	29
X5: WCtime (sec.)	20	21	19
X6: WCtemp (°C)	1060	1070	1050
X7: OXtime (mins.)	10	11	9
X8: OXtemp (°C)	760	780	740
X9: POLYthick (nm)	180	185	175
X10: lgate (μm)	0.09	0.099	0.081
X11: OFFsapcer (nm)	5	5.25	4.75
X12: PTdose1 (cm^2)	2.30E+13	2.53E+13	2.07E+13
X13: PTdose2 (cm^2)	1.15E+13	1.20E+13	1.10E+13
X14: PTenergy (keV)	12	13.2	10.8
X15: EXdose (cm^2)	1.60E+15	1.65E+15	1.55E+15
X16: EXenergy (keV)	1.2	1.26	1.14
X17: EXtime (sec.)	0.5	0.55	0.45
X18: EXtemp (°C)	1000	1010	990
X19: Sdspacer (nm)	61	64	58
X20: SDdose (cm^2)	5.1E+14	5.3E+14	4.9E+14
X21: SDenergy (keV)	4	4.4	3.6
X22: SDtime (sec.)	1	1.05	0.95
X23: SDtemp (°C)	1055	1060	1050

Chapter 5

The Design of Experiment

Here we discuss the design of experiment in our work, there are seven factors we considered, and three kinds of design we will be compared. The first design is the face centered cube design (CCF), the second design is the smaller composite design (SCD1), and the third design is the smallest composite design (SCD2).

5.1 Three Designs of Experiment, CCF, SCD1, and SCD2

Seven factors are finally selected from the results of screening designs with Plackett-Burman designs, and these seven factors are counted in to construct the design matrix. Each factors are shown in Table 5.1.

As we have said in Chapter 2, if we want to build a 2nd order response surface model,

Table 5.1: Identified process conditions displaying a significant effect on the 90nm NMOSFET device characteristics.

Coded factors	**Parameter name**	
A	Gate length	Lgate
B	Threshold voltage implant dose	VTdose
C	Punch-through implant energy	PTenergy
D	Punch-through implant dose	PTdose
E	Threshold voltage implant energy	VTenergy
F	Oxide growth temperature	OXtemp
G	Oxide growth time	OXtime

the central composite design is the suggested design. There are three special forms of central composite design, and they are CCC design, CCF design, and CCI design. From our experience, CCC design and CCI design are improper in our experiment results, this discussion will be shown in more detail in Appendix C. Thus we choose CCF design finally.

In this chapter, we also provide other two designs of experiment, one is the smaller composite design (SCD1), the other is the smallest composite design (SCD2). In these three designs, there are one center point and fourteen axial points in each design. The differences between these design are their cube portion. From the CCF design, 2_V^{7-1} fractional factorial design is used in the cube points, from the SCD1, 2_{III}^{7-2} fractional factorial design is used in the cube points, and from the SCD2, a 22-run Plackett-Burman design is used to construct the cube points. Traditionally, cube portion built by fractional factorial design with resolution V is most classical and popular. However, from economical viewpoint in

Table 5.2: Comparison among the CCF design, SCD1 design, and SCD2 design.

	CCF	**SCD1**	**SCD2**
Run number	79	47	37
Center point	1	1	1
Axial points	14	14	14
Cube points	64	32	22
Design in cube	2_V^{7-1} fractional factorial	2_{III}^{7-2} fractional factorial	22-run Plackett-Burman

semiconductor fabrication, this design is still too expensive to use. As the result, we study another two smaller design (SCD1 and SCD2) and try to compare the results with the CCF design. In this work, there are 79 runs in the CCF design of experiment, 47 runs in the SCD1 of experiment, and 37 runs in the SCD2 of experiment. Comparisons are shown on Table 5.2.

These three designs are given the same design spaces in each factors, and the information is provided in Table 5.3. The ± 1 input levels for each process condition of the design is selected to represent the design space of interest. In this work, these values are set to the upper and lower bounds of the screening experiment. Table 5.3 defines the boundary conditions over which the response models are constructed. Note that we are interested in the nonlinearity, so taking the levels of control factors too close together is not very fruitful; therefore, the range of the OXtemp is selected from 740°C to 780°C is reasonable [27].

Table 5.3: Response surface design space for the studied 90nm NMOSFET process.

		Center value	**Cube**	**Axial**
Coded factors	**True factors**	0	± 1	$\pm \alpha = \pm 1$
A	Lgate (μm)	0.09	± 0.009	± 0.009
B	VTdose (cm^{-2})	8E+12	± 8E+11	± 8E+11
C	PTenergy (keV)	12	± 1.2	± 1.2
D	PTdose (cm^{-2})	1.15E+13	± 5E+11	± 5E+11
E	VTenergy (keV)	30	± 1	± 1
F	OXtemp (°C)	760	± 20	± 20
G	OXtime (mins)	10	± 1	± 1

In the CCF design, the cube points are 2_V^{7-1} fractional factorial design with the relation word:

$$OXtime = Lgate \times VTdose \times PTenergy \times PTdose \times VTenergy \times OXtemp. \quad (5.1)$$

The cube points of SCD1 are constructed by the thirty-two factorial constitute a resolution III fractional factorial design. It is built using 2^5 full factorial design for Lgate, VTdose, PTenergy, PTdose, and VTenergy, the input levels for OXtemp and OXtime are generated as:

$$OXtemp = Lgate \times VTdose \times PTenergy, \quad (5.2)$$

and

$$OXtime = Lgate \times VTdose \times PTdose \times VTenergy. \quad (5.3)$$

For example, if the input levels for the first five variables are all low (-1), OXtemp would be set low $(-1 \times -1 \times -1)$, and OXtime would be set high $(-1 \times -1 \times -1 \times -1)$.

And cube points of SCD2 are obtained by 22 runs out of the 24-run Plackett-Burman design using 7 factors. The design matrix of the cube portion is referred to the result of the Draper and Lin [6].

The relationship between these five responses has been shown in the scatter plot matrix on Fig 5.1. These data are obtained from the results of the CCF design through the TCAD simulation. We note that there are strong negative correlation between the Vth and the log(Ioff), and there are some degrees of the relationship between these five physical quantities.

5.2 Summary

Seven factors are selected and used in three kinds of design. These factors are **gate length**, **threshold voltage implant dose**, **punch-through implant energy**, **punch-through implant dose**, **threshold voltage implant energy**, **oxide growth temperature**, and **oxide growth time**. We use these factors into the CCF design with 79 experiments. Then perform the same conditions into other two designs which have fewer experiments, they are the smaller composite design with 47 experiments, and the smallest composite design with 37 experiments. From the scatter plot matrix, we note that there are some degrees of the

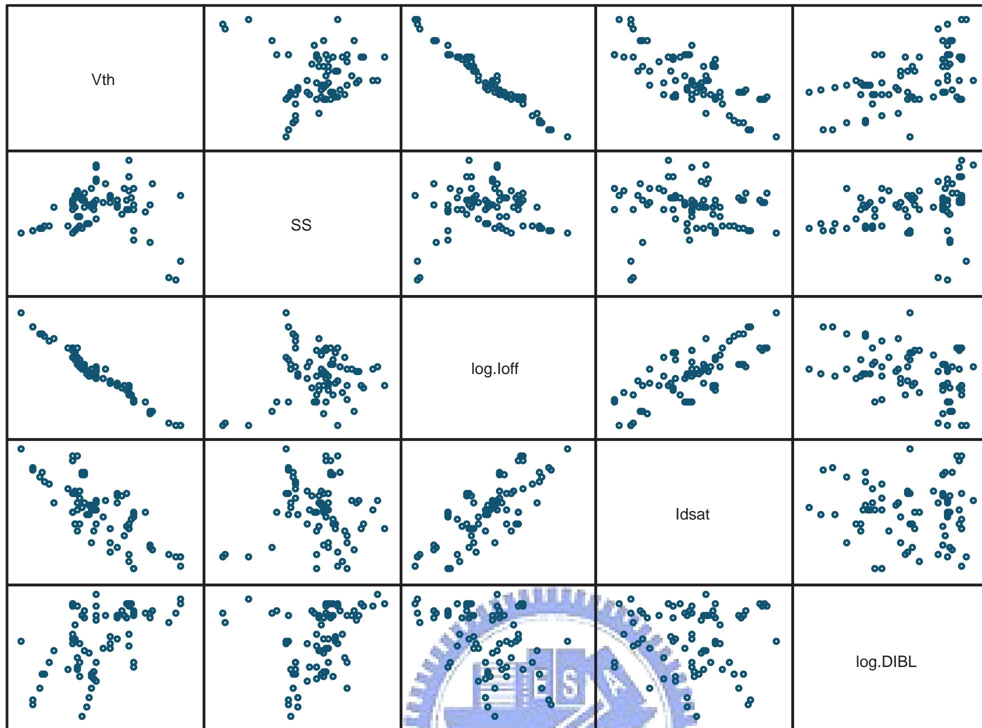


Figure 5.1: The scatter plot matrix of the five physical quantities. This data are obtained from the results of the CCF design through the TCAD simulation. There are strong negative correlation between the Vth and the log(Ioff).

relationship between these five physical quantities.

Chapter 6

The Developed Response Surface Models

In this chapter, the response surface models of three designs are discussed here, and the residual normal plots, R^2, and MSE will also be provided.

6.1 The Face Centered Cube Design

Seventy-nine experimental runs, requiring 158 process and device simulations at an estimated computer computation time of 316 hours, are completed to generate the necessary data for construction of the quadratic response models. Examination for the statistical measures of fit indicated the response models for NMOSFET device are around accurate except for the SS, with an average R^2 value of 0.92302. Table 6.1 are the information for these 5 response surface models. The models are almost adequate for continued use in the process

Table 6.1: A list of the 5 calculated response surface models for the 90nm NMOSFET process using CCF design.

Response	R^2	Adj. R^2	MSE : $\hat{\sigma}^2$
Vth (V)	0.9843	0.9716	5.349E-05
SS (V/dec.)	0.7767	0.5949	8.005E-07
log(Ioff) (A/μm)	0.9911	0.9839	2.970E-003
Idsat (A/μm)	0.9772	0.9586	9.946E-09
log(DIBL) (V)	0.8858	0.7929	6.411E-003

optimization algorithm and the sensitivity analysis. This decision is supported by visual analysis of residual normal probability plots, shown in Figs. 6.1 - 6.5.

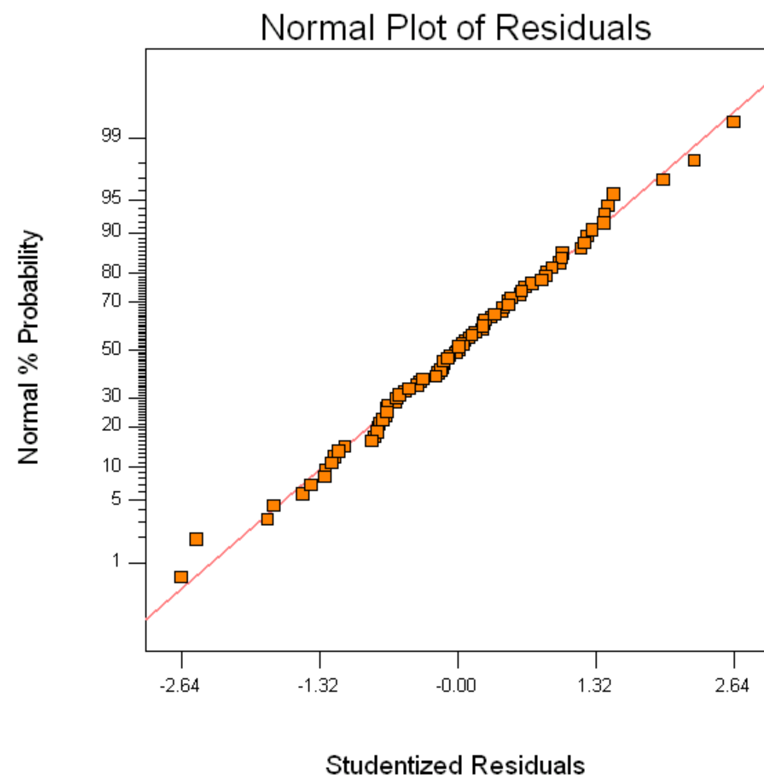


Figure 6.1: Residual normal probability plot of the 90nm NMOSFET Vth's response surface model using CCF design.

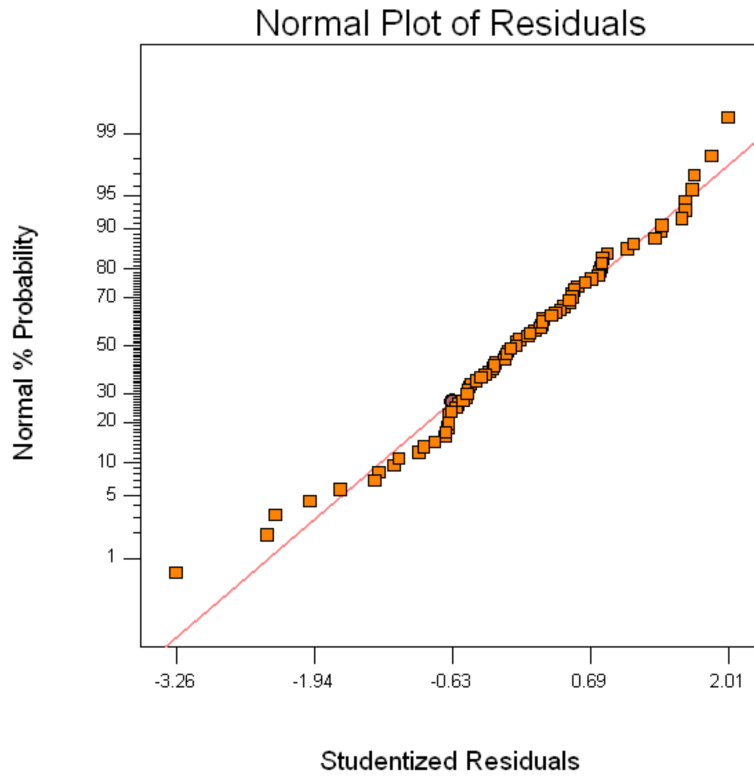


Figure 6.2: Residual normal probability plot of the 90nm NMOSFET SS' response surface model using CCF design.

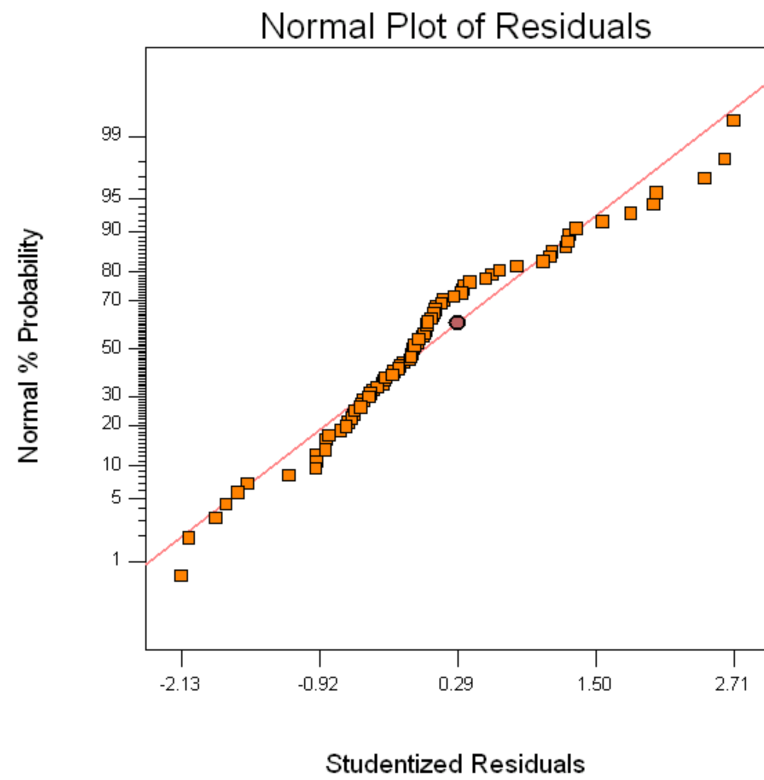


Figure 6.3: Residual normal probability plot of the 90nm NMOSFET log(Ioff)'s response surface model using CCF design.

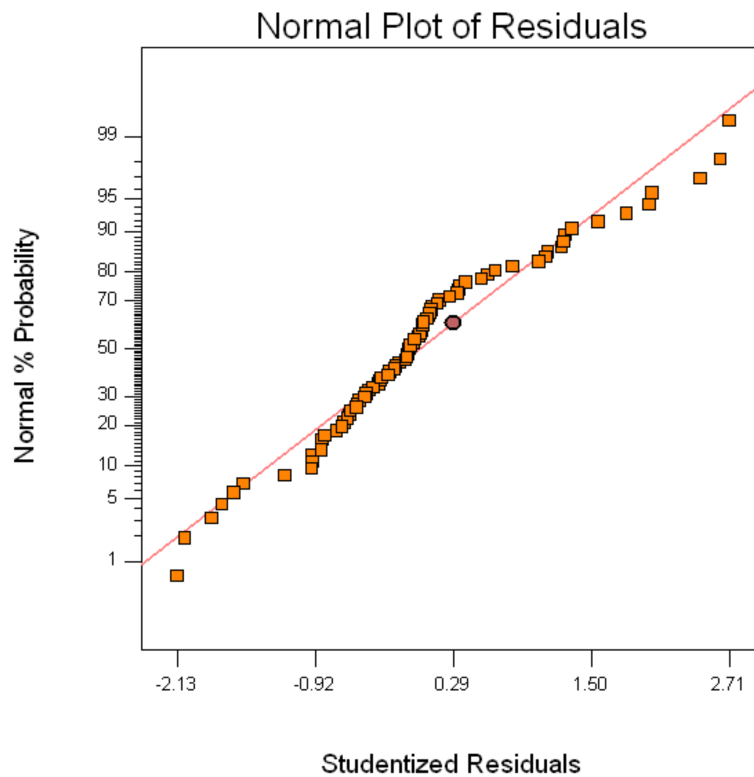


Figure 6.4: Residual normal probability plot of the 90nm NMOSFET Idsat's response surface model using CCF design.

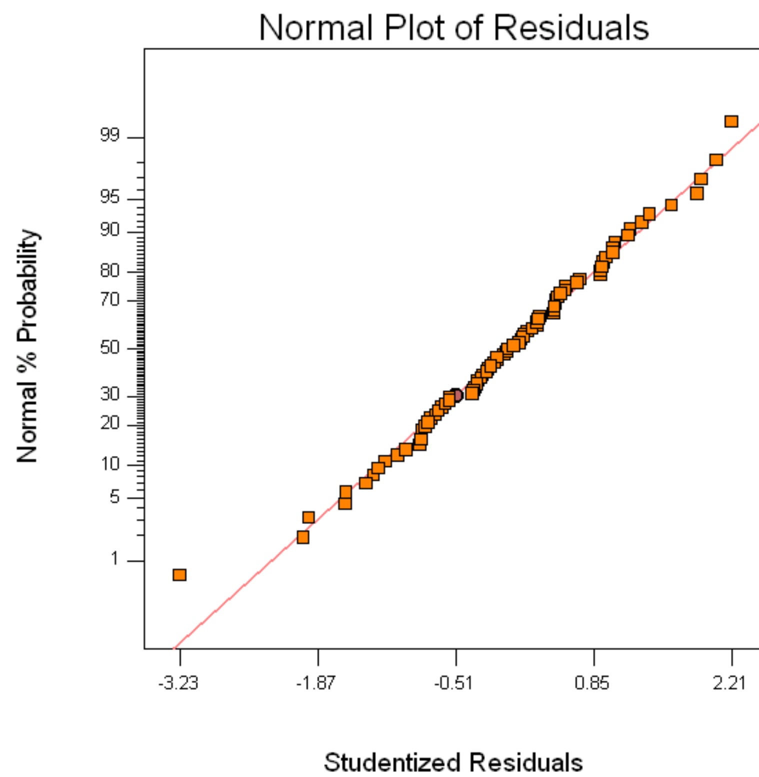


Figure 6.5: Residual normal probability plot for NMOSFET log(DIBL) response surface model using CCF design.

Table 6.2: The 36 coefficients of the response surface models with coded factors using the CCF design.

CCF	Vth	SS	log(Ioff)	Idsat	log(DIBL)
Intercept	0.246532	0.079884	-12.6925	0.005239	-1.45449
A-Lgate	-0.02524	-0.00031	0.219029	-5E-05	-0.15495
B-VTdose	0.004897	1.96E-05	-0.04787	-4.3E-05	-0.0039
C-PTenergy	0.029725	-0.00057	-0.33556	-0.0003	-0.00202
D-PTdose	0.010704	-0.00015	-0.1208	-0.00013	-0.01538
E-VTenergy	-0.00387	8.63E-05	0.035334	3.29E-05	0.009079
F-OXtemp	0.021505	0.000537	-0.18452	-0.0004	0.039182
G-OXtime	0.002683	0.000183	-0.02303	-5.4E-05	0.025606
A^2	0.001182	-0.0003	0.032697	6.43E-05	0.045398
B^2	0.003653	0.000207	-0.04098	-0.0001	0.021209
C^2	0.004631	-0.00028	0.028573	2.67E-05	0.157471
D^2	-0.00332	-0.0001	0.00582	5.34E-06	-0.00781
E^2	-0.00397	5.84E-05	0.000896	4.9E-06	-0.11566
F^2	0.007433	0.000116	-0.02398	-5E-05	0.108482
G^2	-0.00581	-4.9E-05	0.014034	3.76E-05	-0.1363
AB	0.001319	1.55E-06	-0.02039	-2.9E-05	-0.01265
AC	-0.00254	0.000584	0.015668	2.41E-05	0.003409
AD	-0.00157	0.000299	0.01646	1.78E-05	-0.01126
AE	-0.00093	-0.00015	0.018115	3.14E-05	0.013039
AF	0.001063	0.000376	-0.00547	-6.9E-06	0.03083
AG	0.000428	1.35E-05	-0.0038	-1.1E-05	0.009374
BC	0.000211	-9.3E-05	0.001824	5.13E-07	0.005121
BD	0.000225	-0.00015	0.0018	1.68E-06	-0.00372
BE	-0.00013	-0.00025	-0.01041	-1.1E-05	-0.01835
BF	0.00063	-8.4E-06	-0.01063	-1.5E-05	0.00291
BG	-0.0017	1.43E-05	0.014188	2.13E-05	0.001365
CD	0.001149	-0.00028	-0.00208	-5.8E-06	0.001131
CE	0.001226	0.000101	-0.00316	-1.7E-06	0.01887
CF	0.000945	-0.00036	-0.00672	2.95E-06	0.001281
CG	0.000859	-2E-05	-0.00506	-2.7E-06	0.001355
DE	-0.00039	1.42E-05	-0.002	-1.5E-06	-0.01389
DF	0.00217	-0.00016	-0.00137	1.89E-06	0.000614
DG	-0.00023	4.91E-05	-0.00094	6.88E-08	-0.0156
EF	0.000398	0.000198	-0.00551	-1.4E-05	-0.01209
EG	0.00024	4.95E-05	0.004435	5.35E-06	-0.00083
FG	0.003628	0.000261	-0.02923	-6.9E-05	-0.01574

Table 6.3: A list of the 5 calculated response surface models for the 90nm NMOSFET process using the SCD1 design.

Response	R^2	Adj. R^2	MSE : $\hat{\sigma}^2$
Vth (V)	0.9929	0.9704	4.959E-05
SS (V/dec.)	0.9104	0.6251	5.686E-07
log(Ioff) (A/μm)	0.9950	0.9793	3.292E-003
Idsat (A/μm)	0.9854	0.9391	1.277E-08
log(DIBL) (V)	0.9149	0.6443	0.013

6.2 The Smaller Composite Design

Forty seven experimental runs, requiring 94 process and device simulations at an estimated computer computation time of 188 hours, are completed to generate the necessary data for construction of the quadratic response models. Examination of the statistical measures of fit indicated the response models for NMOSFET device are highly accurate, with an average R^2 value of 0.95972. Table 6.3 are the information for these 5 response surface models. As a whole, the models are deemed adequate for what follows in the process optimization algorithm and the sensitivity analysis. This decision is supported by visual analysis of residual normal probability plots, shown in Figs. 6.6 - 6.10.

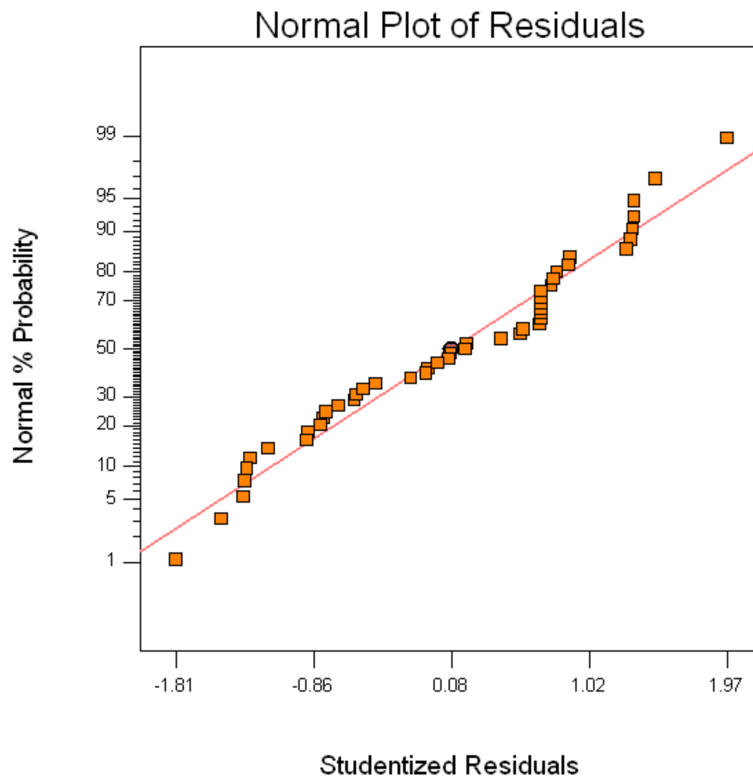


Figure 6.6: Residual normal probability plot of the 90nm NMOSFET Vth's response surface model using SCD1 design.

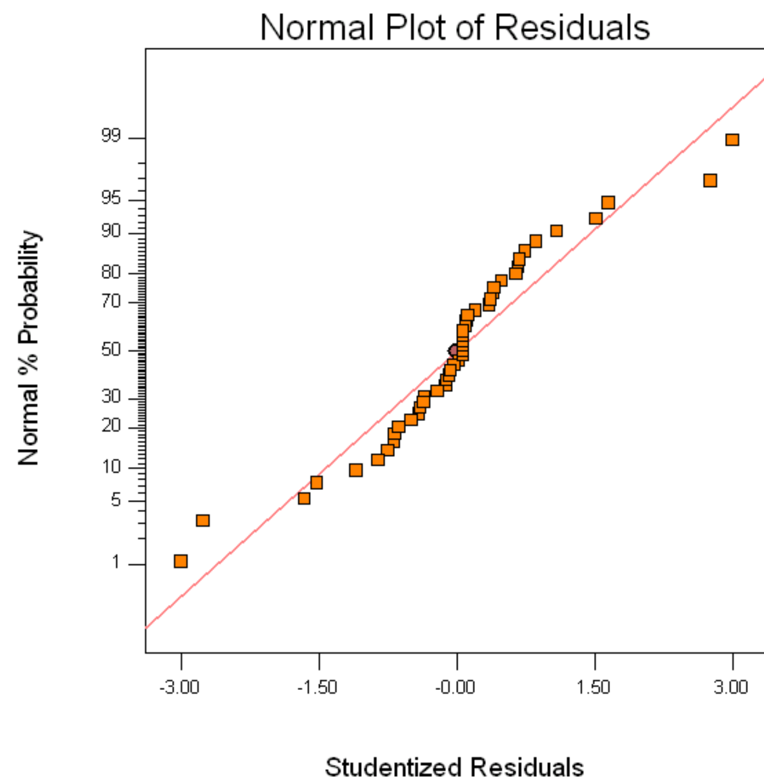


Figure 6.7: Residual normal probability plot of the 90nm NMOSFET SS' response surface model using SCD1 design.

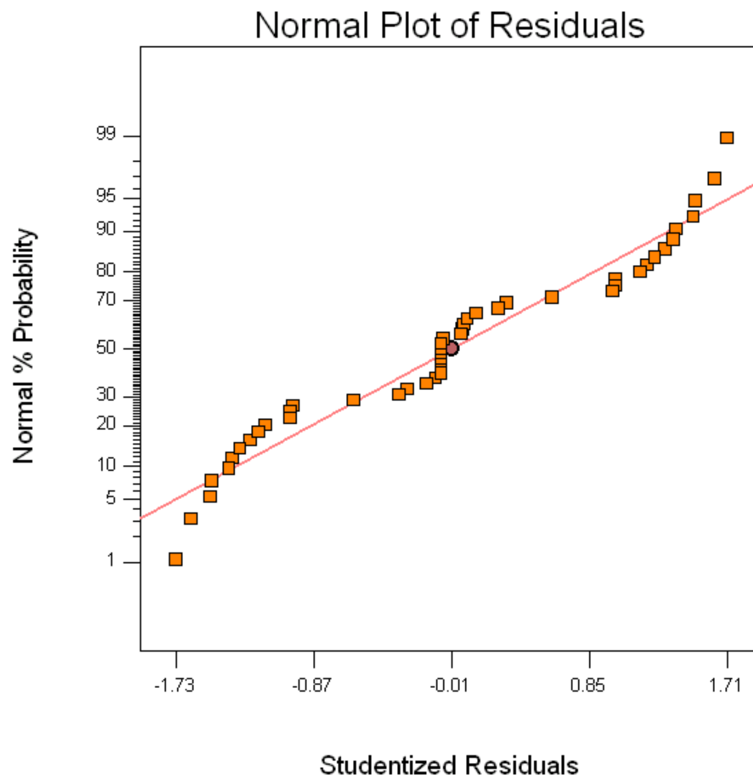


Figure 6.8: Residual normal probability plot of the 90nm NMOSFET log(Ioff)'s response surface model using SCD1 design.

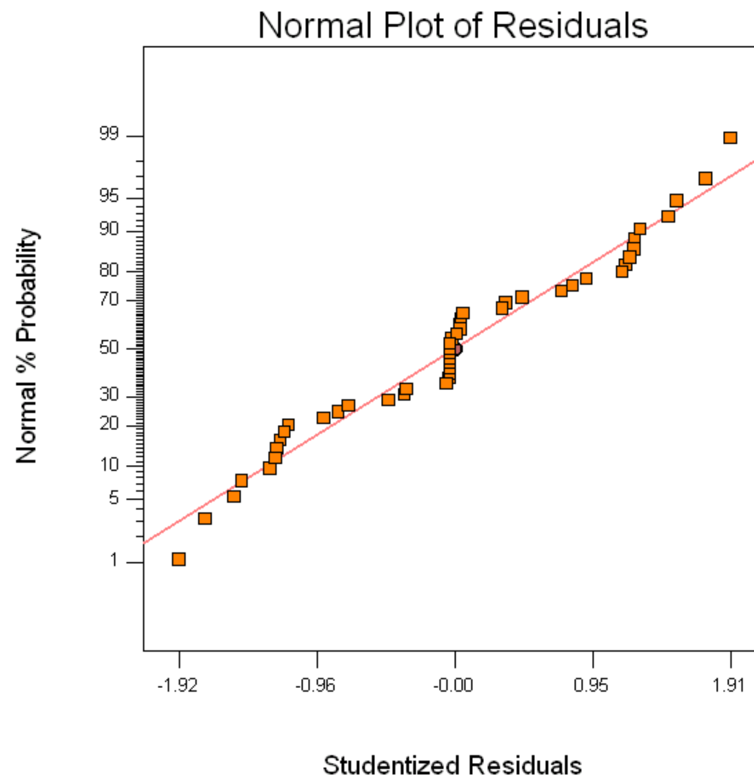


Figure 6.9: Residual normal probability plot of the 90nm NMOSFET Idsat's response surface model using SCD1 design.

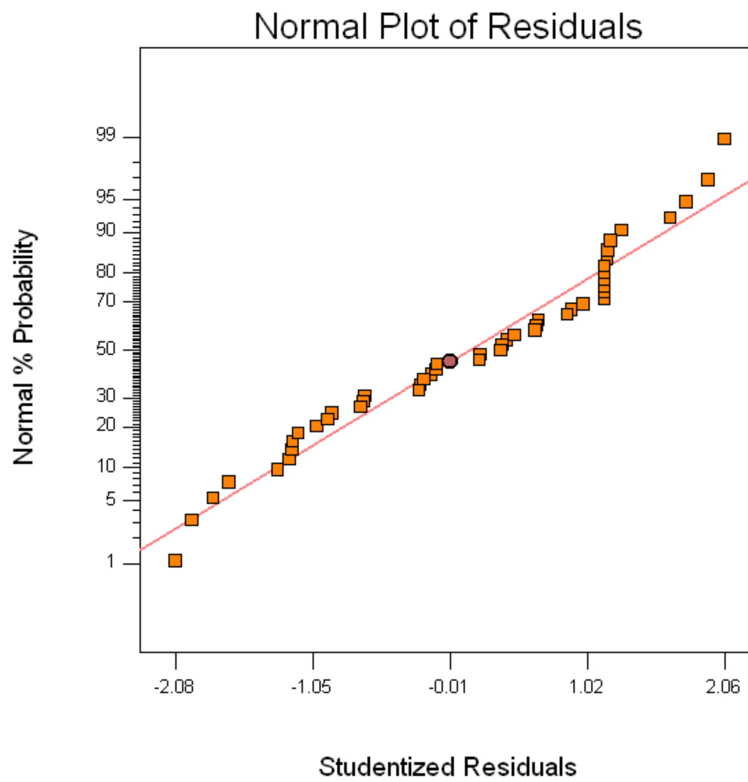


Figure 6.10: Residual normal probability plot of the 90nm NMOSFET log(DIBL)'s response surface model using SCD1 design.

Table 6.4: The 36 coefficients of the response surface models with the coded factors using the SCD1 design.

SCD1	Vth	SS	log(Ioff)	Idsat	log(DIBL)
Intercept	0.246482	0.079879	-12.6924	0.005239	-1.45224
A-Lgate	-0.0251	-0.00041	0.219351	-5E-05	-0.17007
B-VTdose	0.005414	5.38E-05	-0.04925	-4.7E-05	0.0211
C-PTenergy	0.029544	-0.00052	-0.33219	-0.00029	-0.00687
D-PTdose	0.012133	0.000155	-0.11677	-0.00013	0.003901
E-VTenergy	-0.00301	-0.00011	0.025046	1.7E-05	-0.0391
F-OXtemp	0.021629	0.000607	-0.18348	-0.00039	0.062474
G-OXtime	0.000582	0.000114	-0.00204	-3.3E-05	0.009833
A^2	0.001236	-0.0003	0.032554	6.42E-05	0.042953
B^2	0.003707	0.000212	-0.04112	-0.0001	0.018764
C^2	0.004685	-0.00027	0.02843	2.65E-05	0.155026
D^2	-0.00326	-9.6E-05	0.005677	5.19E-06	-0.01025
E^2	-0.00392	6.37E-05	0.000753	4.75E-06	-0.11811
F^2	0.007487	0.000121	-0.02413	-5E-05	0.106037
G^2	-0.00576	-4.4E-05	0.01389	3.75E-05	-0.13874
AB	0.002003	-8.4E-05	-0.01906	-2.8E-05	0.004158
AC	-0.00315	0.000486	0.014641	2.95E-05	0.001562
AD	-0.00334	0.000162	0.034031	5.25E-05	0.018094
AE	-0.00171	-0.00014	0.017025	2.99E-05	-0.0015
AF	0.001546	0.000256	-0.00446	-6E-06	0.049515
AG	-0.00029	-0.00011	-0.00328	-1.1E-05	-0.00647
BC	-0.00362	-6E-05	0.014159	1.62E-05	-0.02106
BD	0.001022	0.000187	-7E-05	-3.8E-06	-0.00775
BE	0.0003	-0.00029	-0.00905	-8.9E-06	0.012428
BF	-9.9E-05	0.000204	-0.01094	-1.5E-05	-0.01381
BG	-0.00359	-0.00017	0.016382	2.35E-05	-0.03645
CD	0.000156	-0.0003	-0.00071	6.7E-06	-0.00967
CE	0.000404	6.11E-05	-0.01197	-1.2E-05	0.007833
CF	0.000696	-0.00036	-0.01192	4.88E-06	-0.00863
CG	0.001176	3.94E-05	0.001248	1.32E-05	0.015528
DE	0.001867	0.000166	-0.02272	-2.2E-05	-0.00459
DF	0.00283	-6.2E-05	-0.01491	-2.4E-05	-0.01505
DG	-0.00122	0.000172	0.00874	1.56E-05	0.041037
EF	0.000639	0.000206	-0.00325	-1.1E-05	-0.02454
EG	-0.00091	-0.00032	0.000957	3.37E-06	0.007275
FG	0.004678	0.000378	-0.03178	-7.2E-05	0.003369

Table 6.5: A list of the 5 calculated response surface models for the 90nm NMOSFET process using the SCD2 design.

Response	R^2	Adj. R^2	MSE : $\hat{\sigma}^2$
Vth (V)	0.9996	0.9842	2.263E-05
SS (V/dec.)	1.0000	0.9986	1.45E-09
log(Ioff) (A/μm)	1.0000	0.9999	8.245E-006
Idsat (A/μm)	1.0000	1.0000	2.392E-12
log(DIBL) (V)	0.9859	0.4921	0.017

6.3 The Smallest Composite Design

Thirty-seven experimental runs, requiring 74 process and device simulations at an estimated computer computation time of 148 hours, are completed to generate the necessary data for construction of the quadratic response models. Examination of the statistical measures of fit indicated the response models for NMOSFET device are highly accurate, with an average R^2 value of 0.9971. Table 6.5 shows the information for these 5 response surface models. From the residual normal plots, the models are deemed unsuitable for continued use in the process optimization algorithm and the sensitivity analysis, so the other models such as obtained from stepwise regression are suggested. This decision is supported by visual analysis of residual normal probability plots, shown in Figs. 6.11 - 6.15.

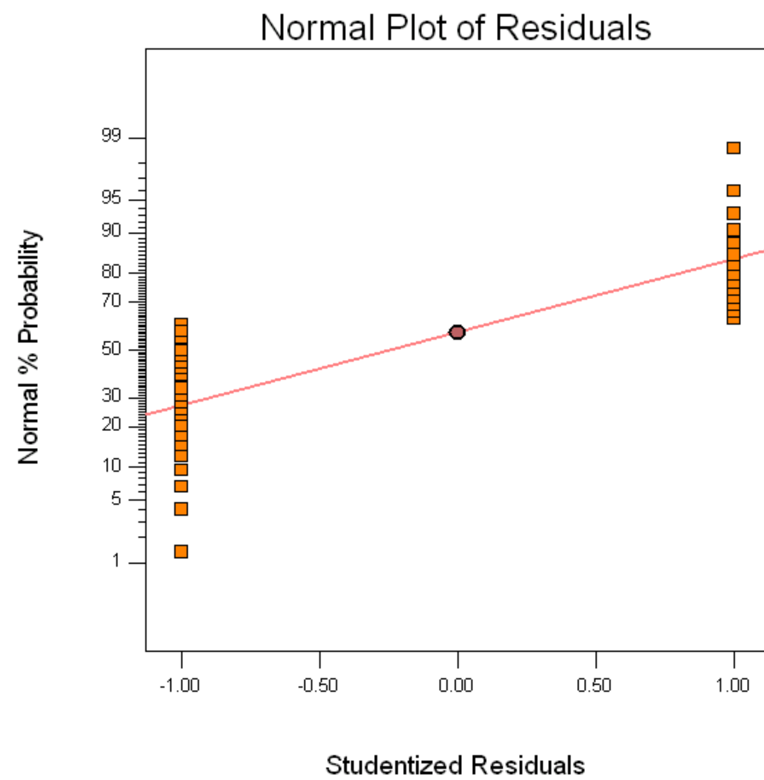


Figure 6.11: Residual normal probability plot of the 90nm NMOSFET Vth's response surface model using SCD2 design.

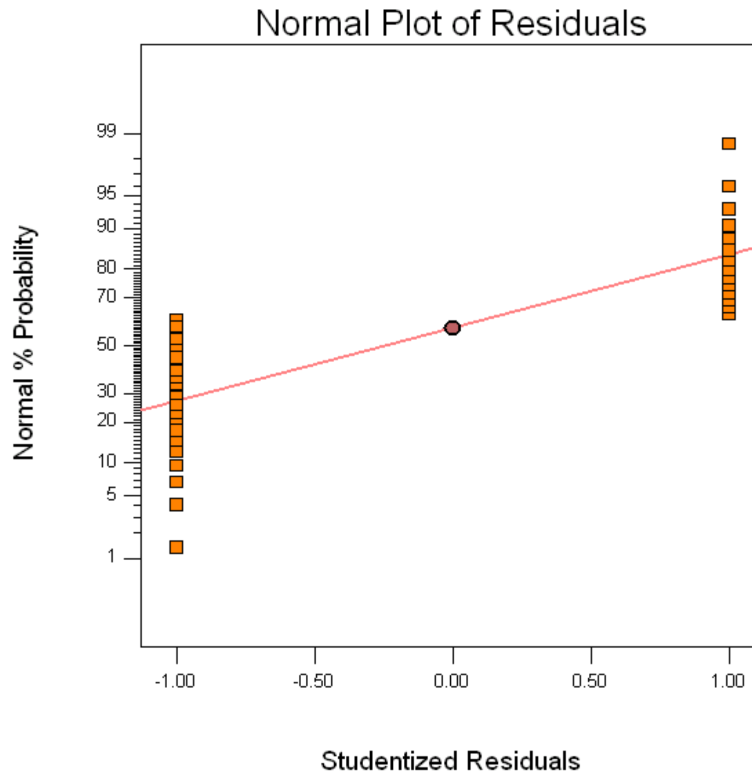


Figure 6.12: Residual normal probability plot of the 90nm NMOSFET SS' response surface model using SCD2 design.

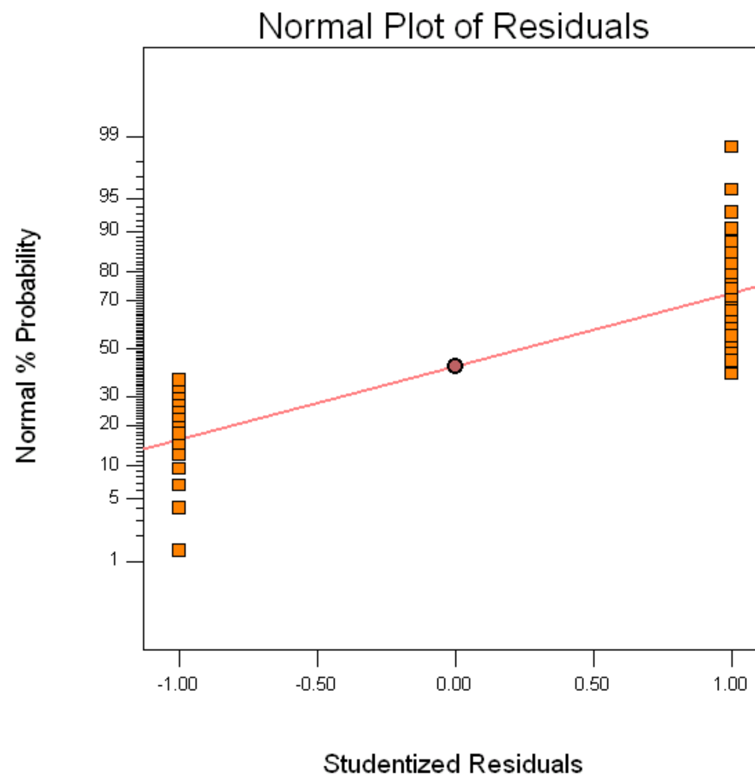


Figure 6.13: Residual normal probability plot of the 90nm NMOSFET log(Ioff)'s response surface model using SCD2 design.

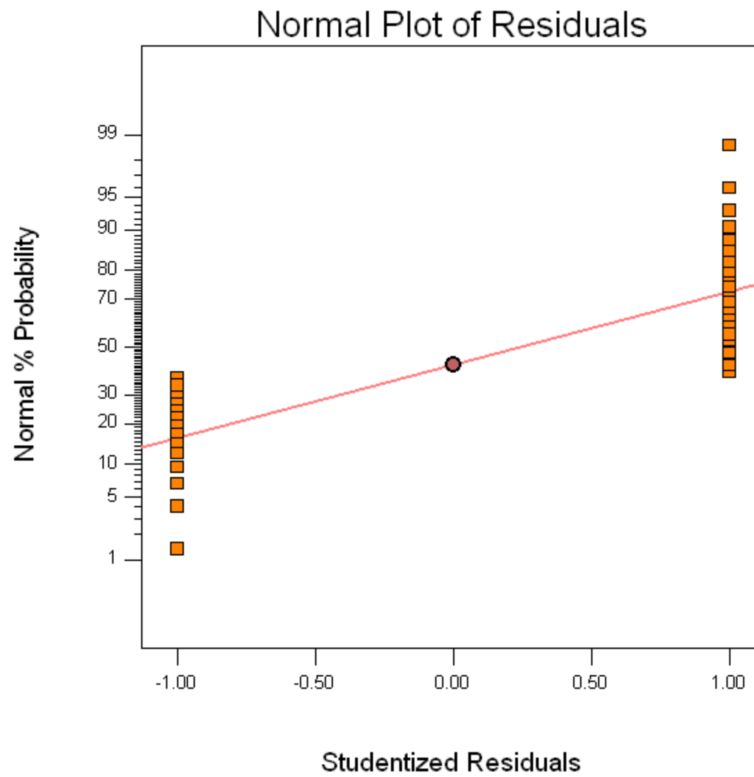


Figure 6.14: Residual normal probability plot of the 90nm NMOSFET Idsat's response surface model using SCD2 design.

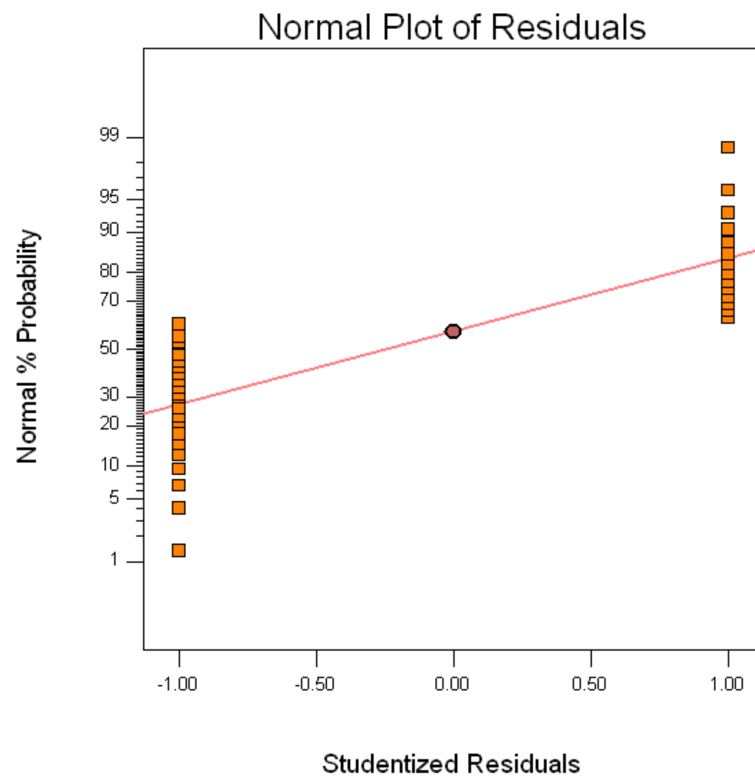


Figure 6.15: Residual normal probability plot of the 90nm NMOSFET log(DIBL)'s response surface model using SCD2 design.

Table 6.6: The 36 coefficients of the response surface models with coded factors using the SCD2 design.

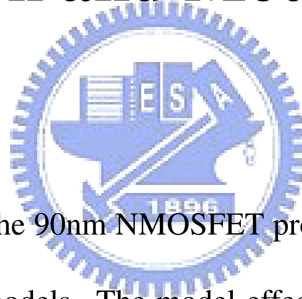
SCD2	Vth	SS	log(Ioff)	Idsat	log(DIBL)
Intercept	0.24636	0.07987	-12.69119	0.00524	-1.45260
A-Lgate	-0.02559	-0.00046	0.21546	-0.00008	-0.14884
B-VTdose	0.01462	0.00024	-0.11741	-0.00020	0.04129
C-PTenergy	0.03245	-0.00007	-0.33833	-0.00030	0.01885
D-PTdose	0.01210	0.00015	-0.11675	-0.00013	0.00304
E-VTenergy	-0.00297	-0.00011	0.02503	0.00002	-0.03823
F-OXtemp	0.02775	0.00094	-0.18604	-0.00039	0.08214
G-OXtime	0.00061	0.00011	-0.00206	-0.00003	0.01069
A^2	0.00137	-0.00028	0.03123	0.00006	0.04335
B^2	0.00384	0.00023	-0.04245	-0.00010	0.01916
C^2	0.00482	-0.00026	0.02710	0.00002	0.15542
D^2	-0.00313	-0.00008	0.00435	0.00000	-0.00985
E^2	-0.00378	0.00008	-0.00057	0.00000	-0.11771
F^2	0.00762	0.00014	-0.02545	-0.00005	0.10644
G^2	-0.00562	-0.00003	0.01256	0.00003	-0.13834
AB	-0.00650	-0.00031	0.03753	0.00006	0.05130
AC	-0.01010	0.00015	0.03622	0.00006	-0.08907
AD	0.00005	0.00015	0.01786	0.00004	-0.03729
AE	-0.01193	-0.00040	0.06990	0.00011	-0.01529
AF	-0.00390	-0.00004	-0.00828	-0.00004	-0.00457
AG	-0.00568	-0.00030	0.02450	0.00005	0.01085
BC	0.00940	-0.00010	-0.03157	-0.00003	0.05884
BD	0.00848	0.00024	-0.01924	-0.00001	0.03168
BE	0.00214	0.00027	0.00884	0.00002	0.03105
BF	-0.01122	-0.00019	0.01059	-0.00002	-0.07402
BG	-0.00592	0.00000	0.02829	0.00005	-0.04271
CD	0.00562	-0.00036	-0.02485	-0.00002	-0.03722
CE	0.00806	0.00042	-0.05031	-0.00011	0.03814
CF	0.00102	-0.00062	-0.00126	-0.00001	0.02767
CG	0.01666	0.00057	-0.02633	0.00001	0.11955
DE	-0.00627	0.00004	0.02058	0.00002	-0.01298
DF	-0.00311	0.00026	-0.00751	-0.00003	-0.09521
DG	0.00392	0.00052	-0.02774	-0.00004	-0.04764
EF	0.01511	0.00034	-0.04340	-0.00007	0.12985
EG	0.01339	-0.00016	-0.04410	-0.00009	0.11138
FG	0.00465	0.00052	-0.04261	-0.00012	0.00936

6.4 Summary

In this chapter we have provided the 5 response surface models in the CCF design, SCD1 design, and SCD2 design. The R^2, MSE, and coefficients of each response surface model are also obtained. The results of CCF design and SCD1 design are deemed adequate for continued use in the process optimization algorithm and the sensitivity analysis. From the residual normal plots, we feel that the response surface models from SCD2 design are improper to continuing optimization and sensitivity analysis, we suggest that these models could be improved by reduced several terms such as using a stepwise regression.

Chapter 7

Variable Selection and Model Reduction



1896

Variable selection is conducted for the 90nm NMOSFET process using statistical analysis of the generated response surface models. The model effects having the greatest significance on response performances for NMOSFET devices are identified using the stepwise regression. The values F_{IN} and F_{OUT} are both set 0.1, and when the models are not hierarchical, we would add several insignificant terms to keep the hierarchy. The hierarchical ordering principle suggests that when resources are scarce, priority should be given to the estimation of lower order effects [34]. This is also a proper method to us because we don't want to loose any potential significant main effects. The results of three designs are shown in the following sections.

7.1 The Face Centered Cube Design

Table 7.2 shows the variable selection results of response surface models using stepwise regression. The coefficient in this table are corresponding to the coded variable levels. The average R^2 value for the reduced models is 0.90488 (original: 0.92302), and it doesn't decrease very much. In Table 7.1 the reduced response surface models of all responses are provided in terms of the actual factors to be the examples. Table 7.3 shows the results of it.

Table 7.1: The final equations of the reduced response models for CCF design in terms of the actual factors for 90nm NMOSFET

Response	Equation
Vth =	$8.11553 + 4.02872 \times Lgate + 2.73580E\text{-}014 \times VTdose + 0.045975 \times PTenergy$ $- 1.12136E\text{-}013 \times PTdose - 3.87313E\text{-}003 \times VTenergy - 0.020201 \times OXtemp$ $- 0.11819 \times OXtime + 1.11623E\text{-}005 \times OXtemp^2 - 0.23561 \times Lgate \times PTenergy$ $- 3.48372E\text{-}013 \times Lgate \times PTdose - 2.12367E\text{-}015 \times VTdose \times Oxtime$ $+ 2.16972E\text{-}016 \times PTdose \times OXtemp + 1.81397E\text{-}004 \times OXtemp \times OXtime$
SS =	$0.38679 - 3.03360 \times Lgate + 9.31961E\text{-}015 \times VTdose + 0.011401 \times PTenergy$ $- 6.96764E\text{-}016 \times PTdose - 4.97458E\text{-}003 \times VTenergy - 4.09147E\text{-}004 \times OXtemp$ $- 9.72169E\text{-}003 \times OXtime + 0.054037 \times Lgate \times PTenergy$ $+ 6.63511E\text{-}014 \times Lgate \times PTdose + 2.08869E\text{-}003 \times Lgate \times OXtemp$ $- 3.09837E\text{-}016 \times VTdose \times VTenergy - 4.64999E\text{-}016 \times PTenergy \times PTdose$ $- 1.49922E\text{-}005 \times PTenergy \times OXtemp + 9.92048E\text{-}006 \times VTenergy \times OXtemp$ $+ 1.30330E\text{-}005 \times OXtemp \times OXtime$
log(Ioff) =	$-10.17712 - 72.86414 \times Lgate + 9.13107E\text{-}013 \times VTdose - 0.41020 \times PTenergy$ $- 5.70800E\text{-}013 \times PTdose - 0.041725 \times VTenergy + 0.010705 \times OXtemp$ $+ 0.94578 \times OXtime - 2.83220E\text{-}012 \times Lgate \times VTdose$ $+ 1.45076 \times Lgate \times PTenergy + 3.65783E\text{-}012 \times Lgate \times PTdose$ $+ 2.01280 \times Lgate \times VTenergy - 1.30116E\text{-}014 \times VTdose \times VTenergy$ $- 6.64543E\text{-}016 \times VTdose \times OXtemp + 1.77351E\text{-}014 \times VTdose \times OXtime$ $- 1.46144E\text{-}003 \times OXtemp \times OXtime$
Idsat =	$0.011440 - 0.10452 \times Lgate + 4.67576E\text{-}017 \times VTdose$ $- 4.47407E\text{-}004 \times PTenergy - 2.56867E\text{-}016 \times PTdose - 2.81188E\text{-}004 \times VTenergy$ $+ 1.47527E\text{-}005 \times OXtemp + 2.35802E\text{-}003 \times OXtime$ $- 4.07030E\text{-}015 \times Lgate \times VTdose + 2.23536E\text{-}003 \times Lgate \times PTenergy$ $+ 3.48989E\text{-}003 \times Lgate \times VTenergy + 2.66232E\text{-}017 \times VTdose \times OXtime$ $- 3.45362E\text{-}006 \times OXtemp \times OXtime$

$$
\begin{aligned}
\log(\text{DIBL}) = \; & 90.86308 - 147.39000 \times \text{Lgate} + 6.83398\text{E-}013 \times \text{VTdose} \\
& - 3.32986 \times \text{PTenergy} + 2.81263\text{E-}013 \times \text{PTdose} + 6.10866 \times \text{VTenergy} \\
& - 0.47069 \times \text{OXtemp} + 3.43030 \times \text{OXtime} + 0.11902 \times \text{PTenergy}^2 \\
& - 0.10175 \times \text{VTenergy}^2 + 3.05992\text{E-}004 \times \text{OXtemp}^2 - 0.12238 \times \text{OXtime}^2 \\
& + 0.17128 \times \text{Lgate} \times \text{OXtemp} - 2.29424\text{E-}014 \times \text{VTdose} \times \text{VTenergy} \\
& + 0.015725 \times \text{PTenergy} \times \text{VTenergy} - 3.12017\text{E-}014 \times \text{PTdose} \times \text{OXtime} \\
& - 7.87157\text{E-}004 \times \text{OXtemp} \times \text{OXtime}
\end{aligned}
$$

Table 7.2: Variable selection results of response surface models with coded factors using the stepwise regression in the CCF design.

CCF	Vth	SS	log(Ioff)	Idsat	log(DIBL)
Intercept	0.245972	0.079594	-12.6783	0.00523	-1.45187
A-Lgate	-0.02524	-0.00031	0.219029	-5E-05	-0.15495
B-VTdose	0.004897	1.96E-05	-0.04787	-4.3E-05	-0.0039
C-PTenergy	0.029725	-0.00057	-0.33556	-0.0003	-0.00202
D-PTdose	0.010704	-0.00015	-0.1208	-0.00013	-0.01538
E-VTenergy	-0.00387	8.63E-05	0.035334	3.29E-05	0.009079
F-OXtemp	0.021505	0.000537	-0.18452	-0.0004	0.039182
G-OXtime	0.002683	0.000183	-0.02303	-5.4E-05	0.025606
A^2	0	0	0	0	0
B^2	0	0	0	0	0
C^2	0	0	0	0	0.171386
D^2	0	0	0	0	0
E^2	0.004465	0	0	0	-0.10175
F^2	0	0	0	0	0.122397
G^2	0	0	0	0	-0.12238
AB	0	0	-0.02039	-2.9E-05	0
AC	-0.00254	0.000584	0.015668	2.41E-05	0
AD	-0.00157	0.000299	0.01646	0	0
AE	0	0	0.018115	3.14E-05	0
AF	0	0.000376	0	0	0.03083
AG	0	0	0	0	0
BC	0	0	0	0	0
BD	0	0	0	0	0
BE	0	-0.00025	-0.01041	0	-0.01835
BF	0	0	-0.01063	0	0
BG	-0.0017	0	0.014188	2.13E-05	0
CD	0	-0.00028	0	0	0
CE	0	0	0	0	0.01887
CF	0	-0.00036	0	0	0
CG	0	0	0	0	0
DE	0	0	0	0	0
DF	0.00217	0	0	0	0
DG	0	0	0	0	-0.0156
EF	0	0.000198	0	0	0
EG	0	0	0	0	0
FG	0.003628	0.000261	-0.02923	-6.9E-05	-0.01574

Table 7.3: A list of the calculated information in the 5 response surface models for the 90nm NMOSFET process after the stepwise regression in the CCF design.

Response	R^2	Adj. R^2	MSE : $\hat{\sigma}^2$
Vth (V)	0.9788	0.9746	4.785E-05
SS (V/dec.)	0.727	0.662	6.678E-07
log(Ioff) (A/μm)	0.9895	0.9870	2.403E-003
Idsat (A/μm)	0.9713	0.9661	8.144E-09
log(DIBL) (V)	0.8578	0.8211	5.538E-003

Table 7.4: A list of the calculated information in the 5 response surface models for the 90nm NMOSFET process after the stepwise regression in the SCD1 design.

Response	R^2	Adj. R^2	MSE : $\hat{\sigma}^2$
Vth (V)	0.9876	0.9803	3.291E-05
SS (V/dec.)	0.8405	0.7471	3.836E-07
log(Ioff) (A/μm)	0.9913	0.9866	2.132E-003
Idsat (A/μm)	0.9734	0.9640	7.557E-09
log(DIBL) (V/dec.)	0.8692	0.8177	6.534E-003

7.2 The Smaller Composite Design

Table 7.5 shows the variable selection results of response surface models using stepwise regression. The coefficient in the table are corresponding the coded variable levels. The average R^2 value for the reduced models is 0.9324 (original: 0.95972), and it doesn't decrease very much. Results of it are obtained by Table 7.4.

Table 7.5: Variable selection results of response surface models with coded factors using the stepwise regression in the SCD1 design.

SCD1	Vth	SS	log(Ioff)	Idsat	log(DIBL)
Intercept	0.245972	0.079655	-12.6808	0.00523	-1.44988
A-Lgate	-0.0251	-0.00041	0.219351	-5E-05	-0.17007
B-VTdose	0.005414	5.38E-05	-0.04925	-4.7E-05	0.0211
C-PTenergy	0.029544	-0.00052	-0.33219	-0.00029	-0.00687
D-PTdose	0.011273	-0.00014	-0.11586	-0.00012	0
E-VTenergy	-0.00415	5.14E-05	0.033271	3.18E-05	-0.00047
F-OXtemp	0.021629	0.000607	-0.18348	-0.00039	0.062474
G-OXtime	0.000582	0.000114	-0.00204	-5.4E-05	0.005518
A^2	0	0	0	0	0
B^2	0	0	0	0	0
C^2	0	0	0	0	0.167078
D^2	0	0	0	0	0
E^2	0	0	0	0	-0.10605
F^2	0.004879	0	0	0	0.118089
G^2	0	0	0	0	-0.12669
AB	0.002003	0	-0.01906	-2.8E-05	0
AC	-0.00315	0.000486	0.014641	2.95E-05	0
AD	-0.00334	0	0.034031	5.25E-05	0
AE	-0.00171	0	0.017025	2.99E-05	0
AF	0	0.000256	0	0	0.049515
AG	0	0	0	0	0
BC	-0.00362	0	0.014159	0	0
BD	0	0.000187	0	0	0
BE	0	-0.00029	0	0	0
BF	0	0.000204	0	0	0
BG	-0.00359	0	0.016382	0	-0.03645
CD	0	-0.0003	0	0	0
CE	0	0	0	0	0
CF	0	-0.00036	0	0	0
CG	0	0	0	0	0
DE	0.001867	0.000166	-0.02272	0	0
DF	0.00283	0	-0.01491	0	0
DG	0	0	0	0	0
EF	0	0.000206	0	0	-0.02454
EG	0.004678	0	0	0	0
FG	0	0.000378	-0.03178	-7.2E-05	0

Table 7.6: A list of the calculated information in the 5 response surface models for the 90nm NMOSFET process after the stepwise regression in the SCD2 design.

Response	R^2	Adj. R^2	MSE : $\hat{\sigma}^2$
Vth (V)	0.9816	0.9734	3.807E-05
SS (V/dec.)	0.9655	0.9311	7.06E-08
log(Ioff) (A/μm)	0.9932	0.9906	1.316E-003
Idsat (A/μm)	0.9705	0.9620	6.194E-09
log(DIBL) (V)	0.8706	0.8058	6.465E-003

7.3 The Smallest Composite Design

Table 7.7 shows the variable selection results of response surface models using stepwise regression. The coefficient in the table are corresponding the coded variable levels. The average R^2 value for the reduced models is 0.95628 (original: 0.9971).It doesn't reduce very much and still very high. The results are provided on Table 7.6.

7.4 Summary

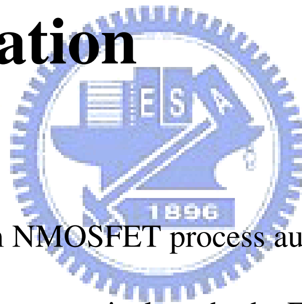
In this chapter we provide the results of variable selection using stepwise regression for three designs. The values F_{IN} and F_{OUT} are both set 0.1, and we follow the hierarchical ordering principle. Tables from each section showed the coefficients and R^2. The average R^2 of these 3 designs results still doesn't decrease very much.

Table 7.7: Variable selection results of response surface models with coded factors using the stepwise regression in the SCD2 design.

SCD2	Vth	SS	log(Ioff)	Idsat	log(DIBL)
Intercept	0.245972	0.079878	-12.6853	0.005225	-1.45058
A-Lgate	-0.0257	-0.00058	0.222035	-5.5E-05	-0.17195
B-VTdose	0.007301	-4E-05	-0.05387	-5.2E-05	0.016539
C-PTenergy	0.028865	-0.00015	-0.3147	-0.00024	-0.05831
D-PTdose	0.009896	1.12E-05	-0.12567	-0.00016	0
E-VTenergy	-0.00811	-0.00035	0.033005	-0.00039	0.003646
F-OXtemp	0.022183	0.000636	-0.17924	-9.8E-05	0.028825
G-OXtime	0.004329	0.000104	-0.03945	0	0.032797
A^2	0	-0.00027	0	0	0
B^2	0	0	0	0	0
C^2	0	0	0	0	0.168582
D^2	0	0	0	0	0
E^2	0	0	0	0	-0.10455
F^2	0.004774	0	0	0	0.119593
G^2	0	0	0	0	-0.12519
AB	0.004459	0.000233	0	0	0
AC	0	0.00037	0	0	0
AD	0	0	0	0	0
AE	0	0	0	0	0
AF	0	0.000183	0	0	0
AG	0	7.69E-05	0	0	0
BC	0	-0.00025	0	0	0.030364
BD	0	0	-0.01751	-3.9E-05	0
BE	0	0	0	0	0
BF	0	0	0	0	0
BG	-0.00746	0	0.025401	0	0
CD	0	-0.00056	-0.03037	-5.5E-05	0
CE	0	0	0	0	0
CF	0	-0.00075	0	0	0
CG	0	0	0	0	0
DE	0	0	0	0	0
DF	0	0	0	0	0
DG	0	0.00038	0	0	0
EF	0.003919	0	0	0	0
EG	0	-0.00044	0	0	0
FG	0	0.000503	0	0	-0.06482

Chapter 8

Application of the Developed Model to Process Optimization


1896

To achieve optimization of the 90nm NMOSFET process automatically, a software named Design Expert is used for providing a numerical method . Full 2nd order response surface model (36 terms) or reduced model (stepwise regression) are able to be applied. Here these two kinds of models are both provided. The objective function is presented in Eqs.2.15, 2.16, and 2.17, and 2.18. Table 8.1 presents the constraints of the process parameters, and Table 8.2 displays the response targets for the optimization. Results of two kinds of models are discussed in the next two sections.

Table 8.1: The constraints of the process parameters.

Parameters	Goal	Lower Limit	Upper Limit
Lgate (μm)	target = 0.09	0.081	0.099
VTdose (cm^{-2})	in range	7.20E+12	8.80E+12
PTenergy (keV)	in range	10.8	13.2
PTdose (cm^{-2})	in range	1.1E+13	1.2E+13
VTenergy (keV)	in range	29	31
OXtemp (°C)	in range	740	780
OXtime (mins)	in range	9	11

Table 8.2: The targets of the responses.

Responses	Goals	Lower Limit	Upper Limit
Vth (V)	target = 0.285	0.235	0.335
SS (V/dec.)	minimize	0.075	0.085
log(Ioff) (A/μm)	minimize	-13.5	-12.5
Idsat (A/μm)	target = 0.0047	0.0042	0.0052
log(DIBL) (V)	minimize	-1.8	-1.0

8.1 Optimization Results Using the Full 2nd Order Response Surface Models

Table 8.3 shows the recipe of the nominal case and three optimal cases using the full 2nd order response surface models. We can choose one of three optimal recipes to fabricate the new device whose performances will be better than before. For example, if we want to get the better performance of the devices, in the 90nm NMOSFET process, we could choose the optimal recipe of SCD1 result, thus we should decrease the threshold voltage implant dose, increase the punch-through implant energy, raise the punch-through implant dose, decrease threshold voltage implant energy , and extend the oxide growth time, decrease oxide growth temperature. Table 8.4 provides the optimal values which are estimated by the response surface models. Several recipes obtained from these three designs are shown in Table 8.5, the first row in each design is the priority choice from our analysis, it is performed to our proposal. Finally, several contour plot examples are shown here, the completed contour plots have provided in Appendix D.

Table 8.3: Comparison of the recipe between the nominal case and three optimal cases using the full 2nd order response surface models.

Recipe	Nominal case	CCF optimal case	SCD1 optimal case	SCD2 optimal case
Lgate (μm)	0.09	0.09	0.0876	0.0909
VTdose (cm^{-2})	8.0E+12	7.85E+12	7.20E+12	7.89E+12
PTenergy (keV)	12	12.80	12.51	13.20
PTdose (cm^{-2})	1.15E+13	1.20E+13	1.20E+13	1.19E+13
VTenergy (keV)	30	29	29	29
OXtemp (°C)	760	777.26	772.66	777.24
OXtime (mins)	10	9	9.10	9

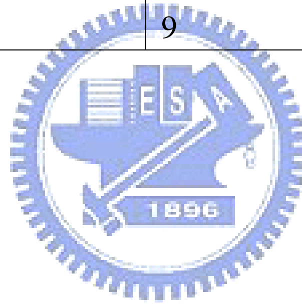


Table 8.4: Optimal values calculated by the full 2nd order response surface models.

Values calculated from response surface models	CCF optimal case	SCD1 optimal case	SCD2 optimal case
Vth (V)	0.289255	0.28	0.285433
SS (V/dec.)	0.078423	0.078166	0.07765
log(Ioff) (A/μm)	-13.1479	-13.1028	-13.1576
Idsat (A/μm)	0.004704	0.004788	0.0047
log(DIBL) (V)	-1.54166	-1.5503	-1.58864

Table 8.5: Optimal recipes calculated by the full 2nd order response surface models. The first row in each design is the priority solution in our models, number 2 means the second priority. The first priority recipe is performed to our proposal.

Full	Lgate (μm)	VTdose (cm^{-2})	PTenergy (keV)	PTdose (cm^{-2})	VTenergy (keV)	OXtemp (°C)	OXtime (mins)
CCF	0.09	7.85E+12	12.80	1.20E+13	29.00	777.26	9
2	0.09	7.20E+12	13.02	1.20E+13	29.00	768.44	9.44
3	0.09	8.62E+12	13.03	1.20E+13	30.69	763.62	11
4	0.09	8.13E+12	13.20	1.20E+13	31.00	765.48	11
SCD1	0.0876	7.20E+12	12.51	1.20E+13	29	772.66	9.1
2	0.09	7.49E+12	13.16	1.18E+13	31	773.24	9
3	0.0903	8.80E+12	13.20	1.20E+13	29.17	761.65	9
4	0.0905	8.31E+12	13.20	1.20E+13	30.93	760.89	10.96
SCD2	0.0909	7.89E+12	13.20	1.19E+13	29	777.24	9
2	0.09	8.12E+12	13.18	1.20E+13	30.48	767.11	9.01
3	0.09	8.34E+12	13.16	1.19E+13	29	767.47	9
4	0.0920	8.23E+12	13.20	1.20E+13	30.26	767.40	9.01

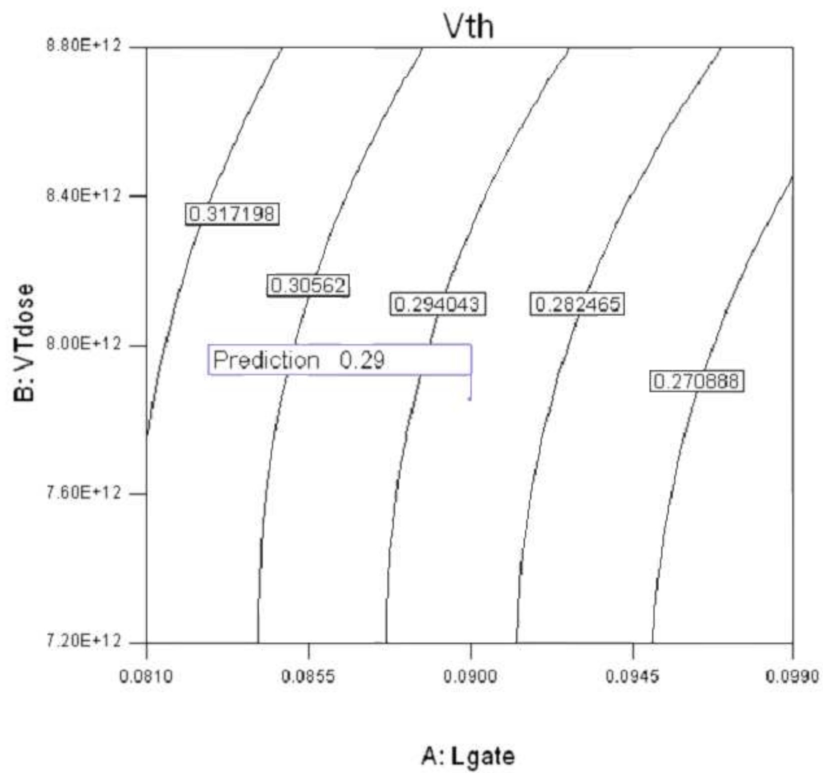


Figure 8.1: The Vth's contour plot of the optimal solution for the CCF design using the full 2nd order response surface models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

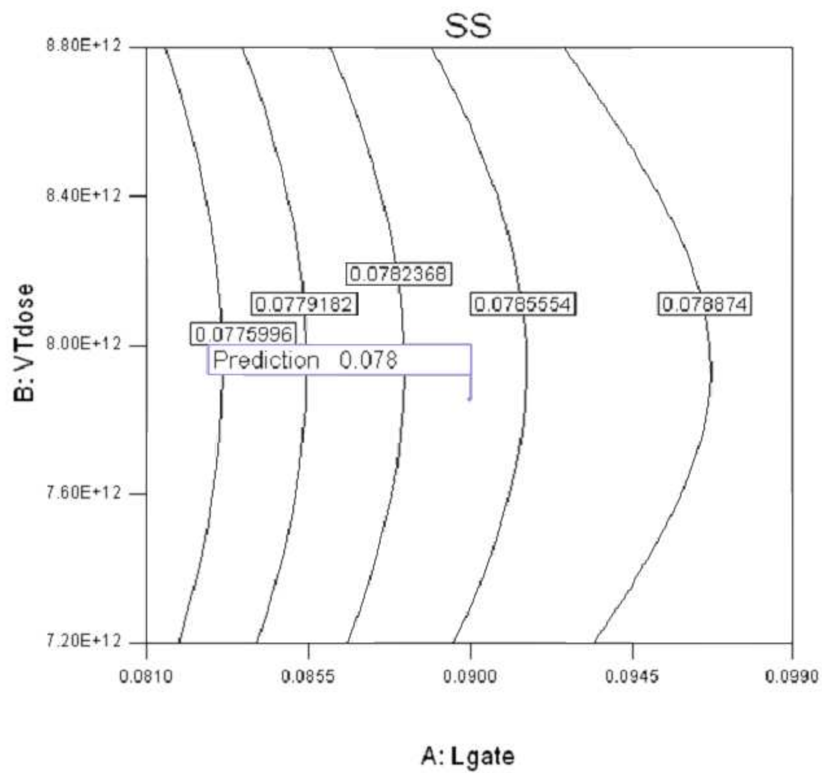


Figure 8.2: The SS' contour plot of the optimal solution for the CCF design using the full 2nd order response surface models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

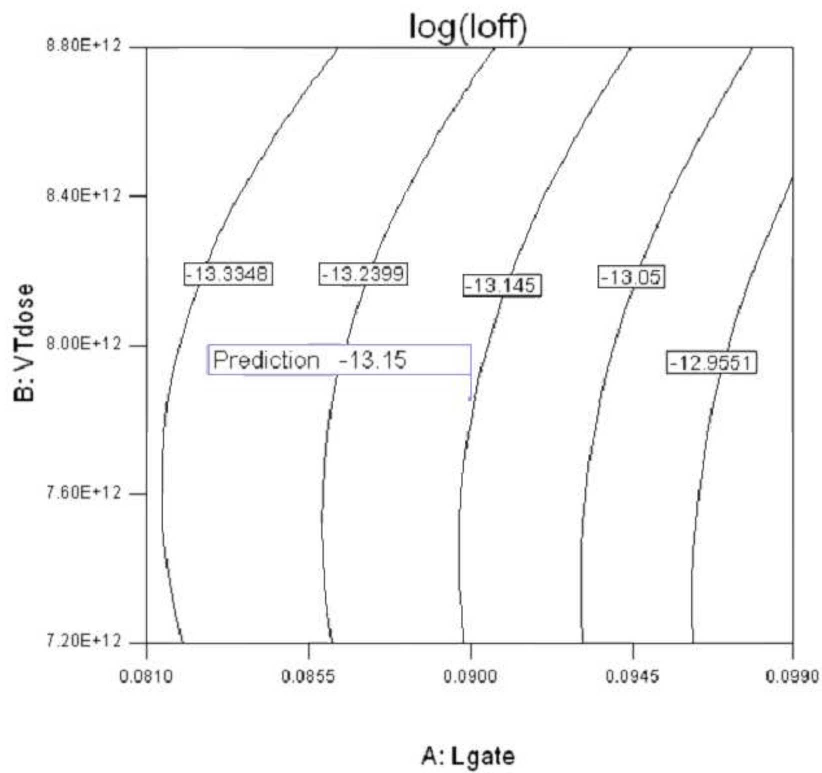


Figure 8.3: The log(Ioff)'s contour plot of the optimal solution for the CCF design using the full 2nd order response surface models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

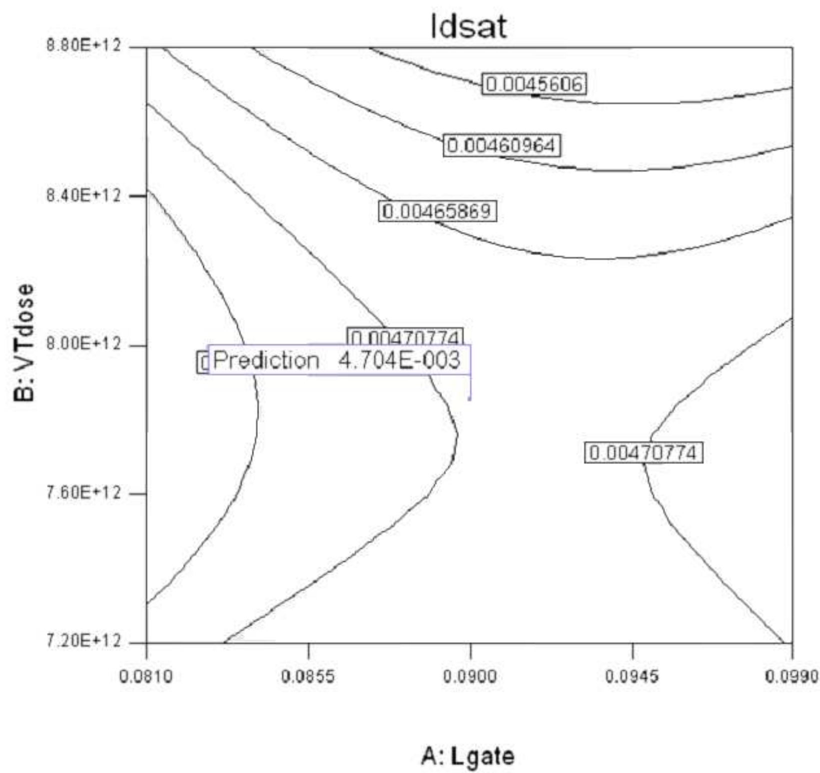


Figure 8.4: The Idsat's contour plot of the optimal solution for the CCF design using the full 2nd order response surface models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

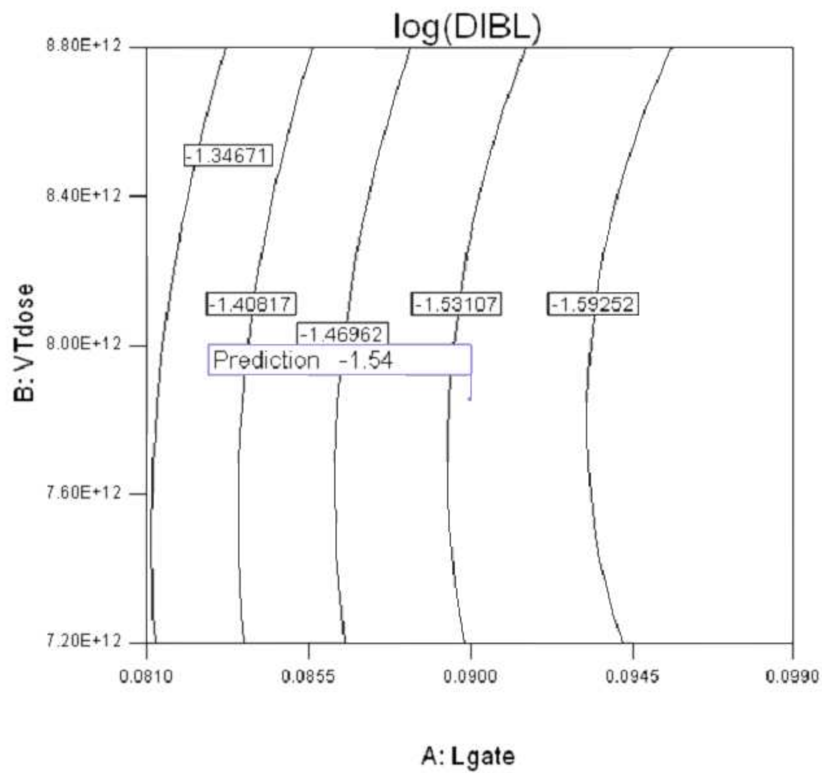


Figure 8.5: The log(DIBL)'s contour plot of the optimal solution for the CCF design using the full 2nd order response surface models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

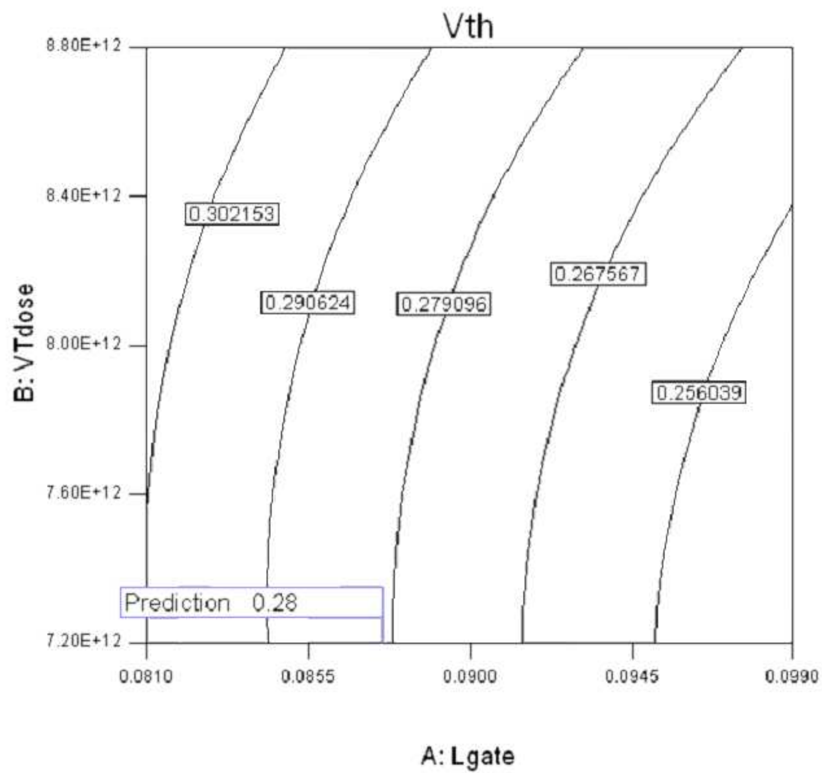


Figure 8.6: The Vth's contour plot of the optimal solution for the SCD1 design using the full 2nd order response surface models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

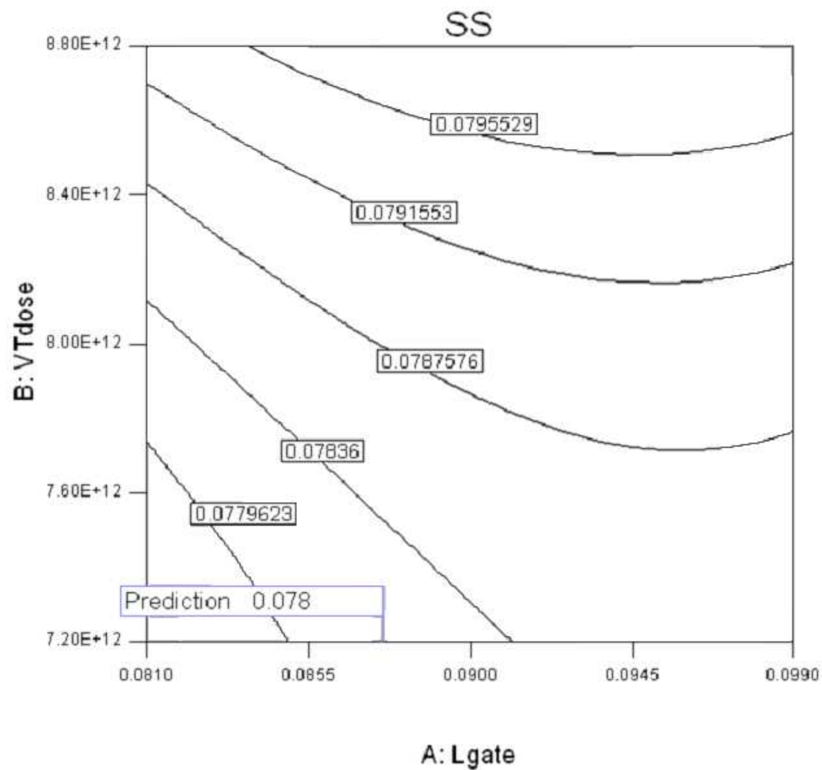


Figure 8.7: The SS' contour plot of the optimal solution for the SCD1 design using the full 2nd order response surface models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

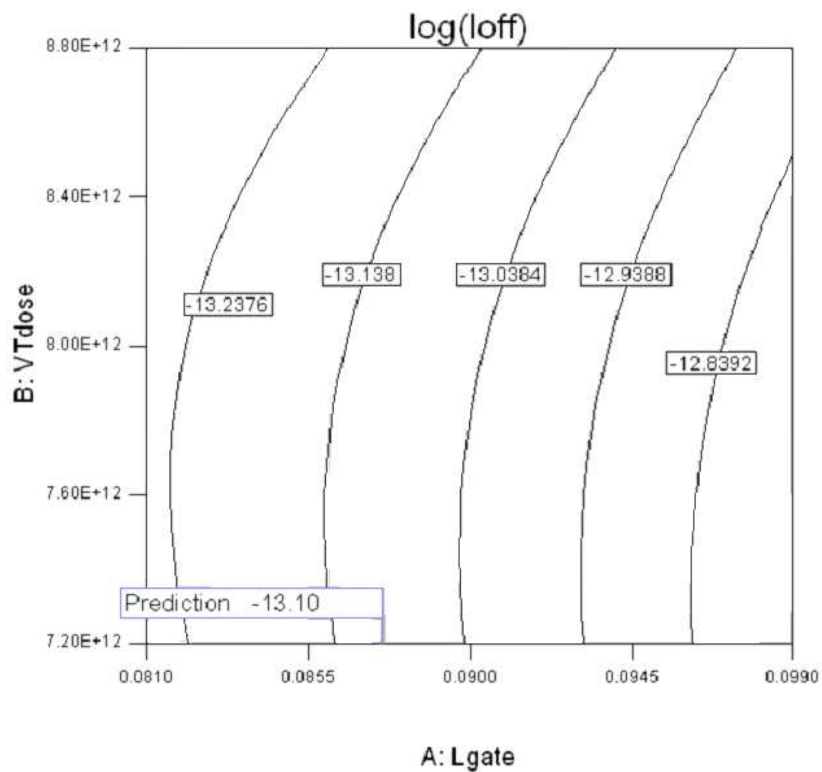


Figure 8.8: The log(Ioff)'s contour plot of the optimal solution for the SCD1 design using the full 2nd order response surface models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

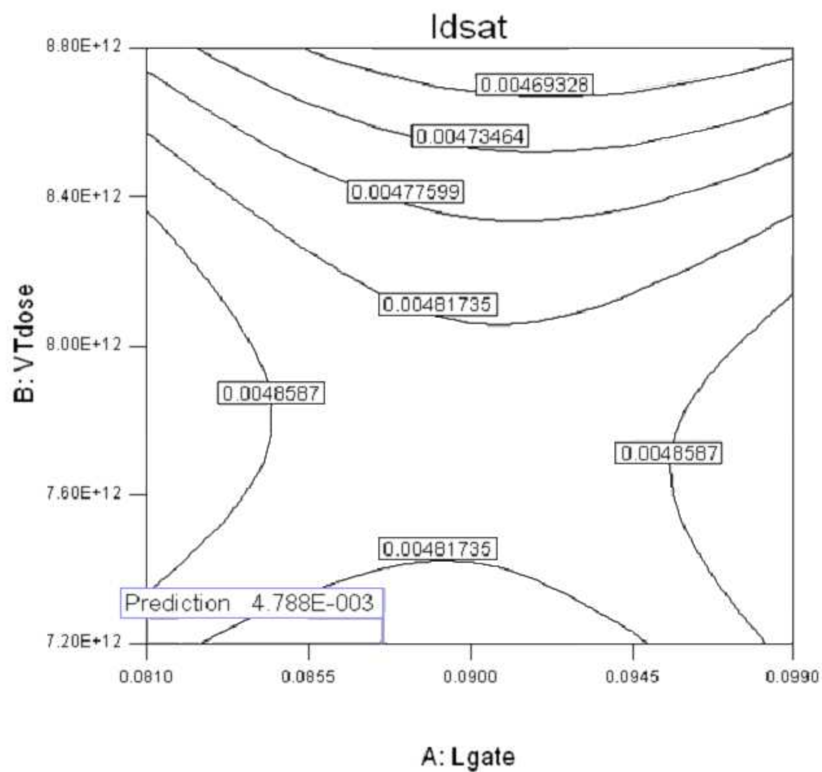


Figure 8.9: The Idsat's contour plot of the optimal solution for the SCD1 design using the full 2nd order response surface models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

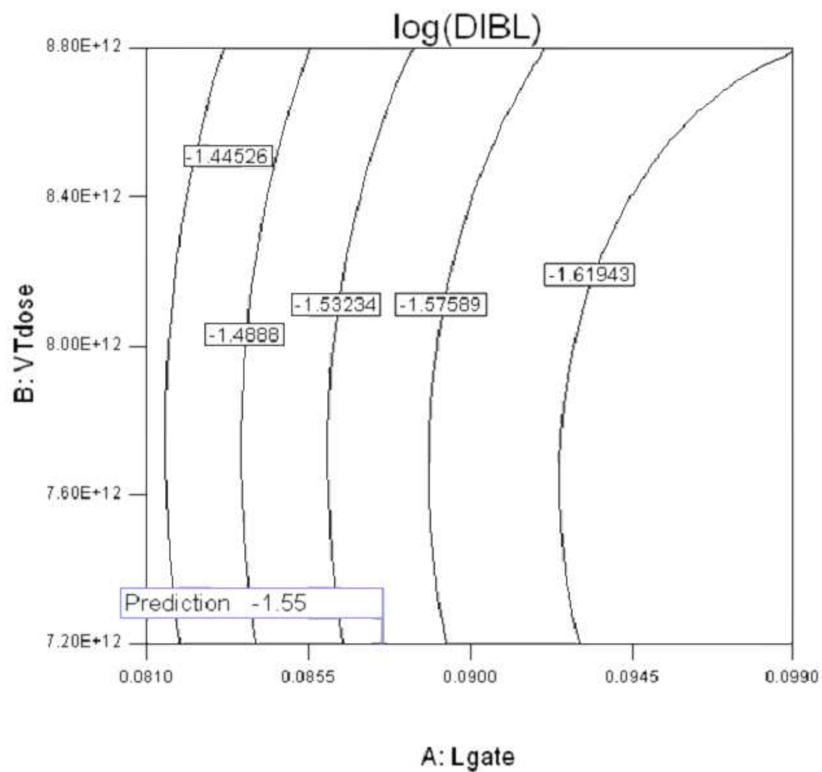


Figure 8.10: The log(DIBL)'s contour plot of the optimal solution for the SCD1 design using the full 2nd order response surface models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

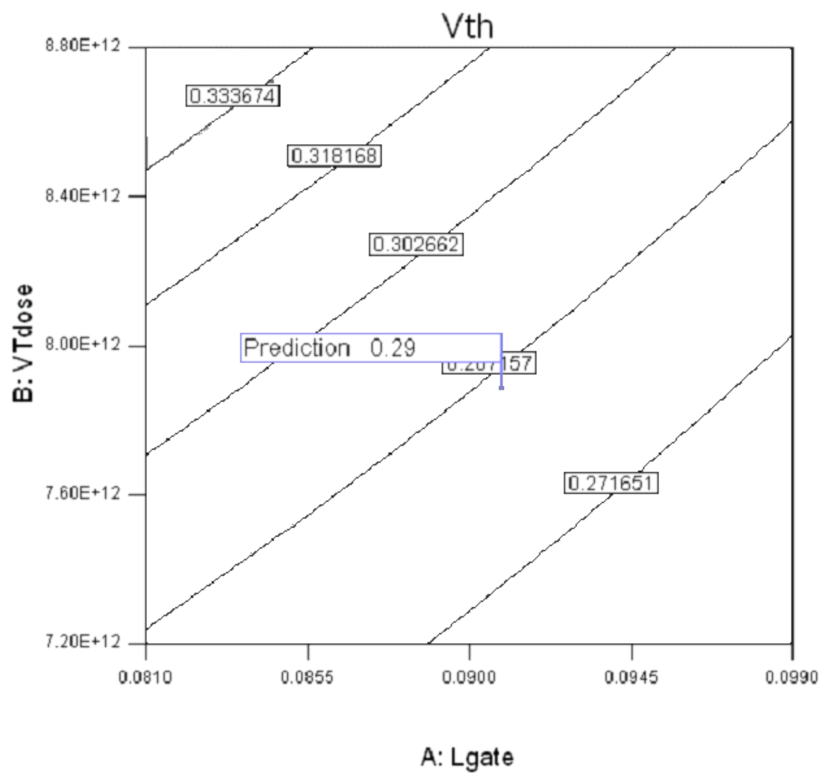


Figure 8.11: The Vth's contour plot of the optimal solution for the SCD2 design using the full 2nd order response surface models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

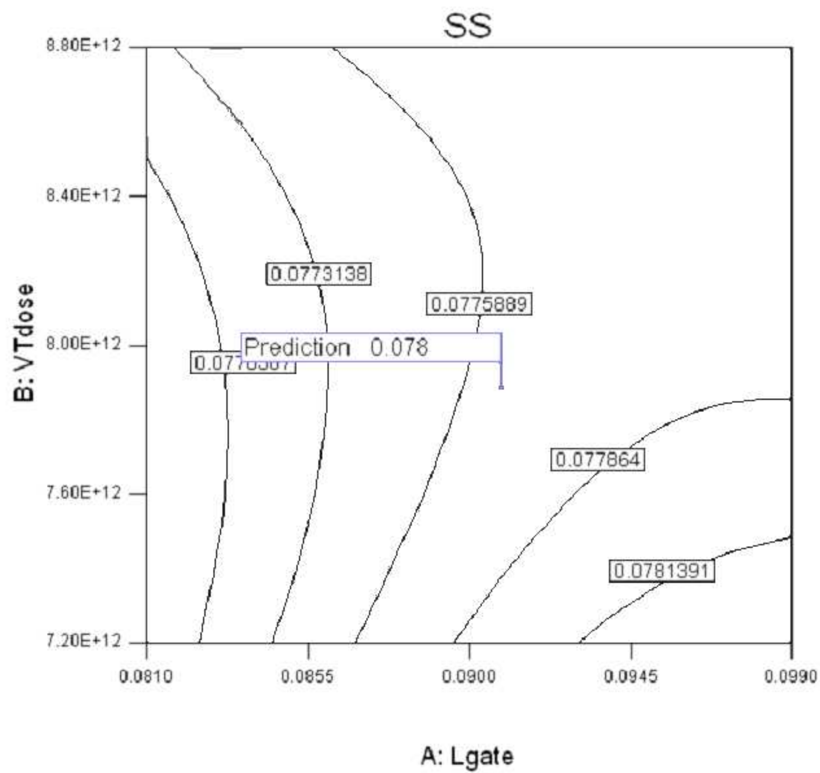


Figure 8.12: The SS' contour plot of the optimal solution for the SCD2 design using the full 2nd order response surface models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

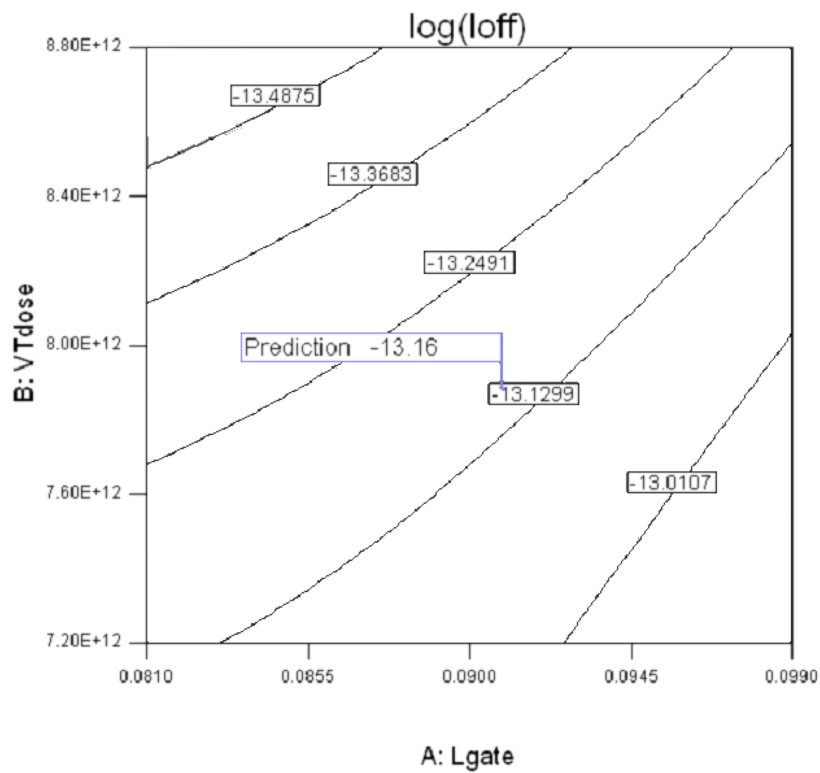


Figure 8.13: The log(Ioff)'s contour plot of the optimal solution for the SCD2 design using the full 2nd order response surface models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

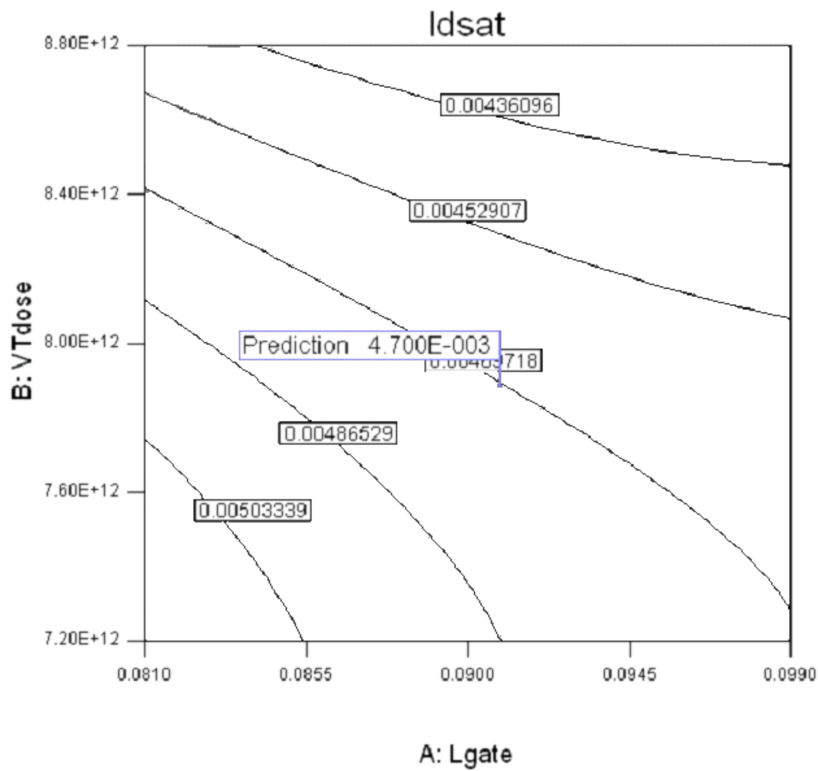


Figure 8.14: The Idsat's contour plot of the optimal solution for the SCD2 design using the full 2nd order response surface models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

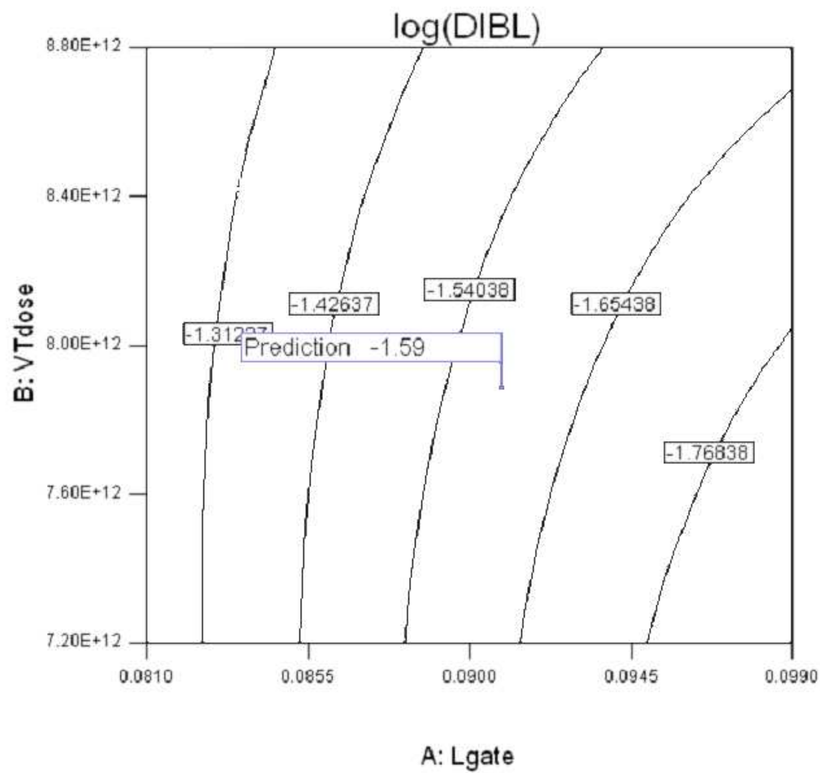
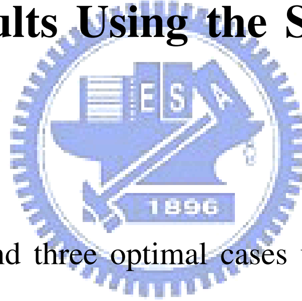


Figure 8.15: The log(DIBL)'s contour plot of the optimal solution for the SCD2 design using the full 2nd order response surface models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

Table 8.6: The comparison of the recipe between nominal case and three optimal cases using stepwise regression models.

Recipe	Nominal case	CCF optimal case	SCD1 optimal case	SCD2 optimal case
Lgate (μm)	0.09	0.09	0.09	0.092
Vtdose (cm^{-2})	8.0E+12	7.20E+12	7.20E+12	7.70E+12
PTenergy (keV)	12	12.99	12.79	13.20
PTdose (cm^{-2})	1.15E+13	1.18E+13	1.20E+13	1.20E+13
VTenergy (keV)	30	29.01	29	29
OXtemp (°C)	760	765.81	766.32	769.41
OXtime (mins)	10	11	10.90	9

8.2 Optimization Results Using the Stepwise Regression Models



The recipes of the nominal case and three optimal cases using the stepwise regression models are shown in Table 8.6. Experimenters could choose one optimal recipe from this table and could apply it as the same with above section. Other relative information has shown below.

8.3 Summary

Six optimal recipes have been provided here and experimenter could use one of it to fabricate the new devices. The first three optimal recipes are obtained from the full 2nd order

Table 8.7: Optimal values calculated by the stepwise regression models.

Values calculated from response surface models	CCF optimal case	SCD1 optimal case	SCD2 optimal case
Vth (V)	0.288957	0.289091	0.286457
SS (V/dec.)	0.078789	0.078417	0.0783429
log(Ioff) (A/μm)	-13.1283	-13.073	-13.1476
Idsat (A/μm)	0.004701	0.004735	0.004703
log(DIBL) (V)	-1.57737	-1.53412	-1.59222

Table 8.8: Optimal recipes calculated by the stepwise regression models. The first row in each design is the priority solution in our models, number 2 means the second priority. The first priority recipe is performed to our proposal.

Step	Lgate (μm)	VTdose (cm^{-2})	PTenergy (keV)	PTdose (cm^{-2})	VTenergy (keV)	OXtemp (°C)	OXtime (mins)
CCF	0.09	7.20E+12	12.99	1.18E+13	29.01	765.81	11
2	0.09	7.48E+12	12.93	1.20E+13	29.01	757.04	10.99
3	0.09	7.64E+12	12.89	1.17E+13	29	774.83	9
4	0.09	8.31E+12	13.11	1.20E+13	31	767.03	9
SCD1	0.09	7.20E+12	12.79	1.20E+13	29	766.32	10.9
2	0.0904	8.04E+12	13.18	1.20E+13	29	761.01	11
3	0.09	8.01E+12	13.05	1.18E+13	29	759.22	11
4	0.09	7.60E+12	13.13	1.20E+13	31	764.90	9
SCD2	0.092	7.70E+12	13.2	1.20E+13	29	769.41	9
2	0.0901	7.35E+12	13.1	1.20E+13	29	774.13	9
3	0.0914	8.80E+12	12.82	1.20E+13	29	765.41	9
4	0.0922	8.00E+12	13.17	1.20E+13	30.47	768.17	9

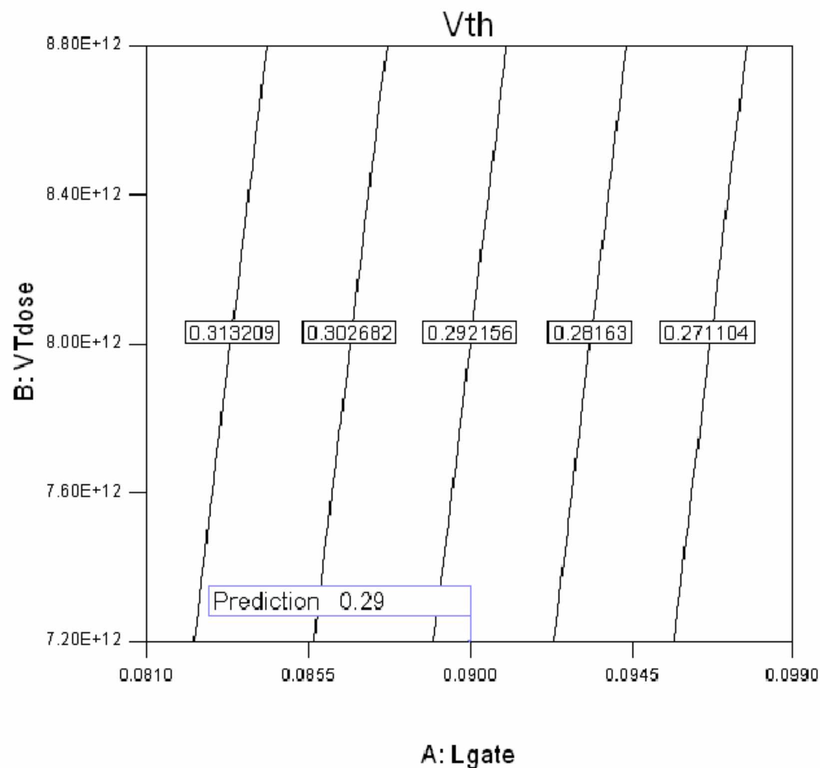


Figure 8.16: The Vth's contour plot of the optimal solution for the CCF design using the stepwise regression models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

response surface models, and other three optimal recipes are calculated from the stepwise regression models. In this chapter we also provide the constraints of the factors and the goals of the responses. Finally we give some examples of contour plots to be the references.

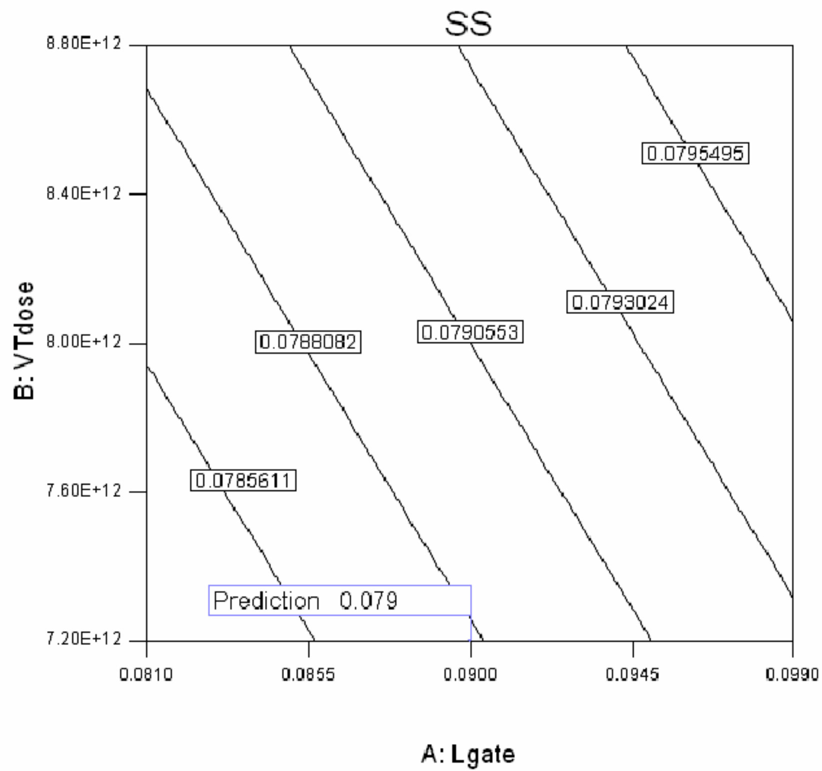


Figure 8.17: The SS' contour plot of the optimal solution for the CCF design using the stepwise regression models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

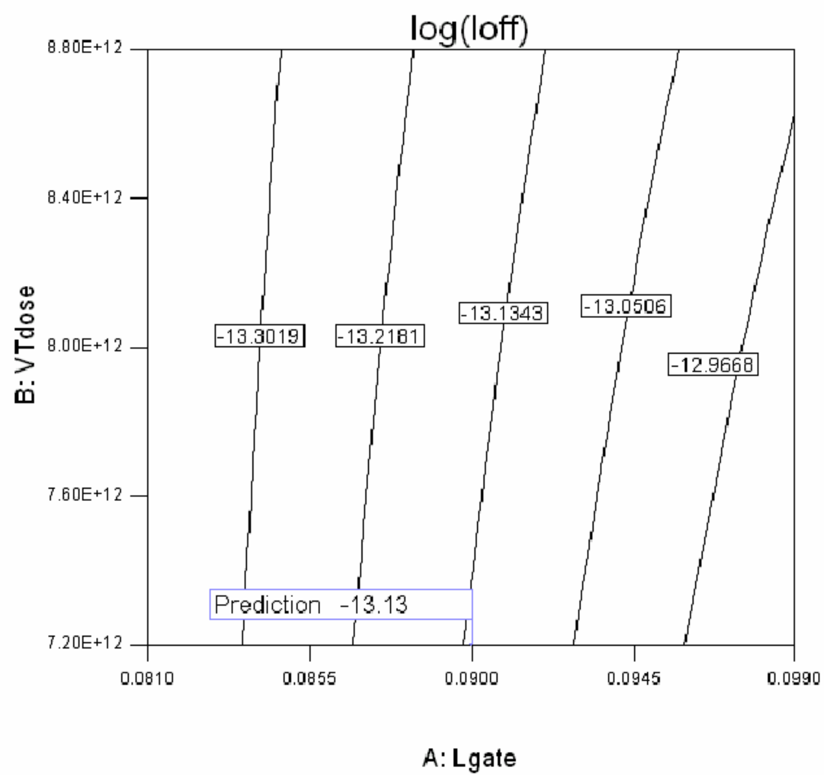


Figure 8.18: The log(Ioff)'s contour plot of the optimal solution for the CCF design using the stepwise regression models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

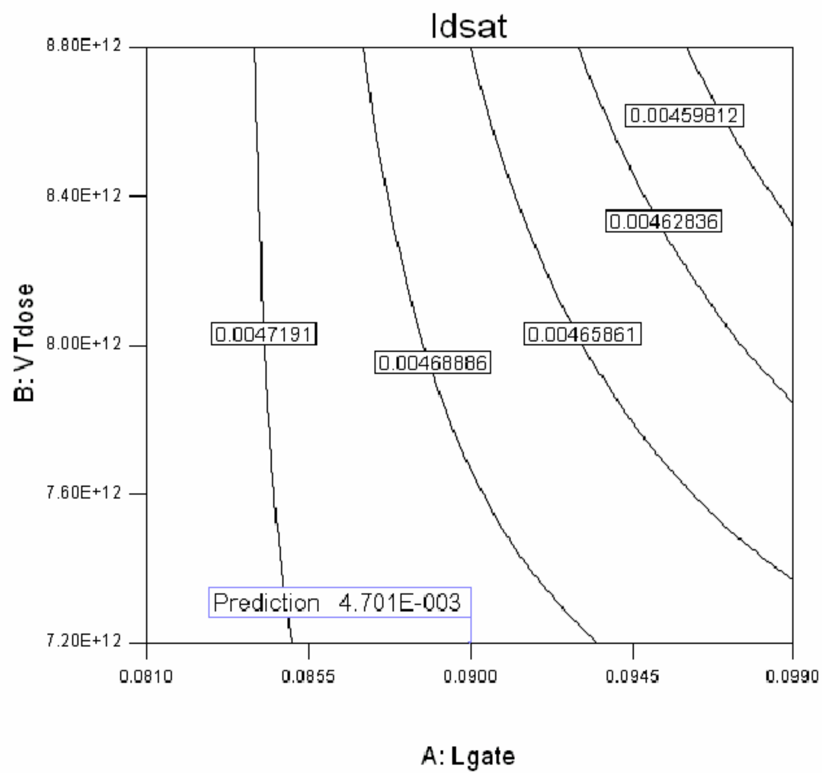


Figure 8.19: The Idsat's contour plot of the optimal solution for the CCF design using the stepwise regression models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

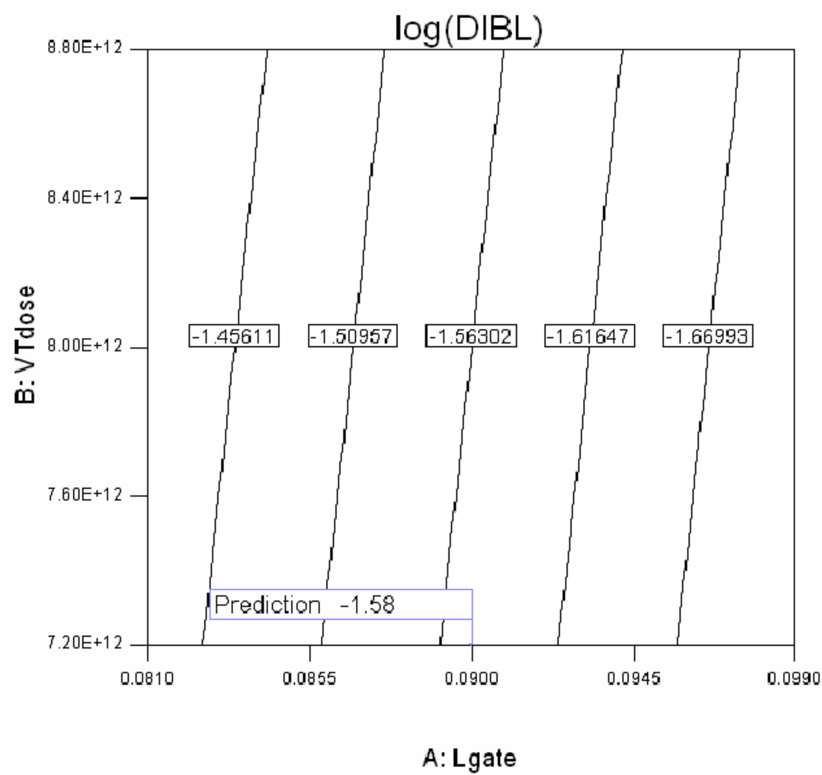


Figure 8.20: The log(DIBL)'s contour plot of the optimal solution for the CCF design using the stepwise regression models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

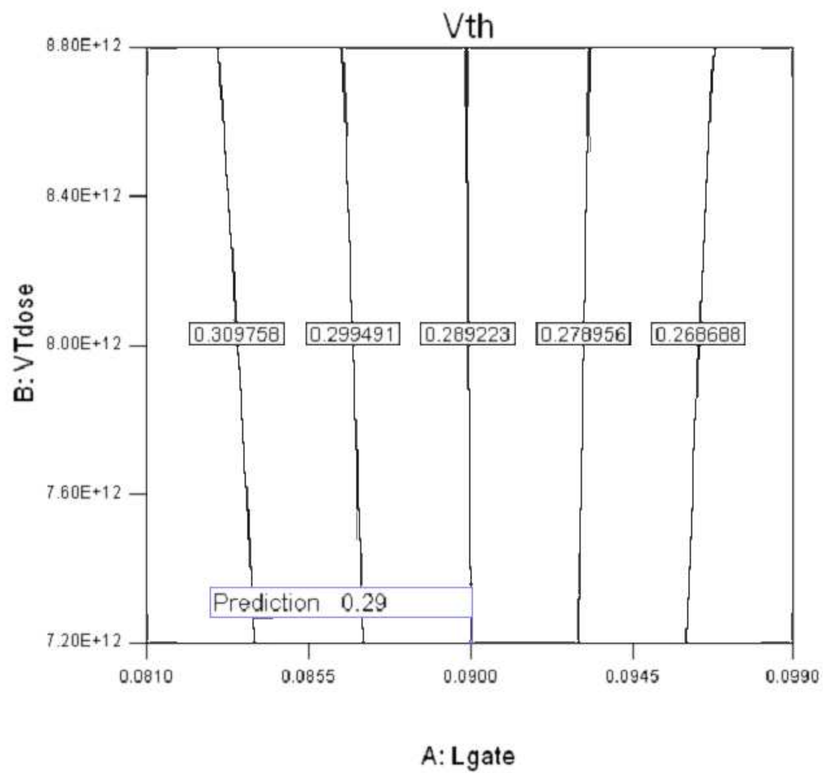


Figure 8.21: The Vth's contour plot of the optimal solution for the SCD1 design using the stepwise regression models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

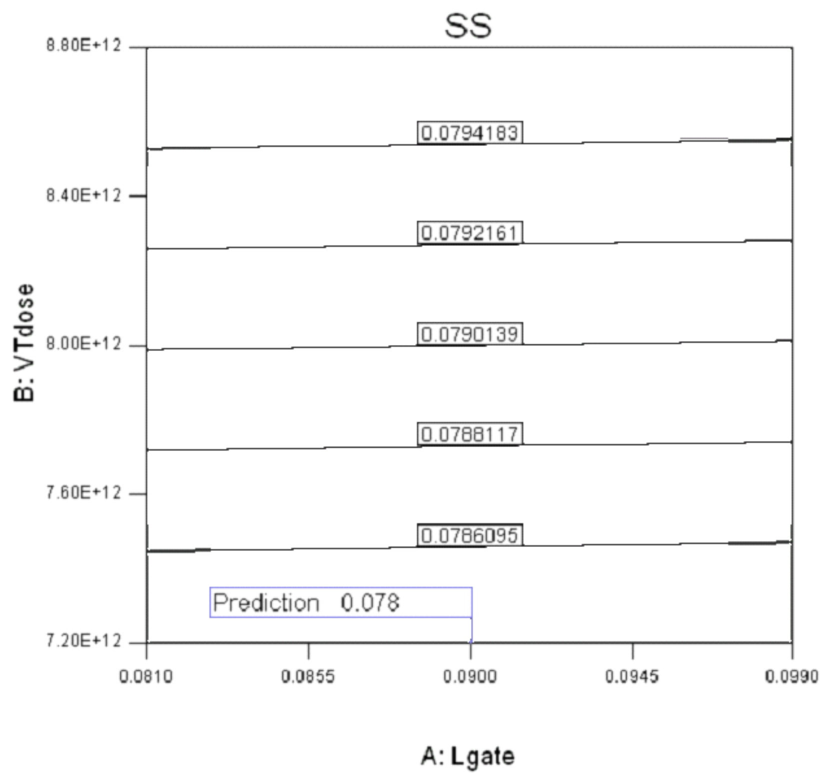


Figure 8.22: The SS' contour plot of the optimal solution for the SCD1 design using the stepwise regression models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

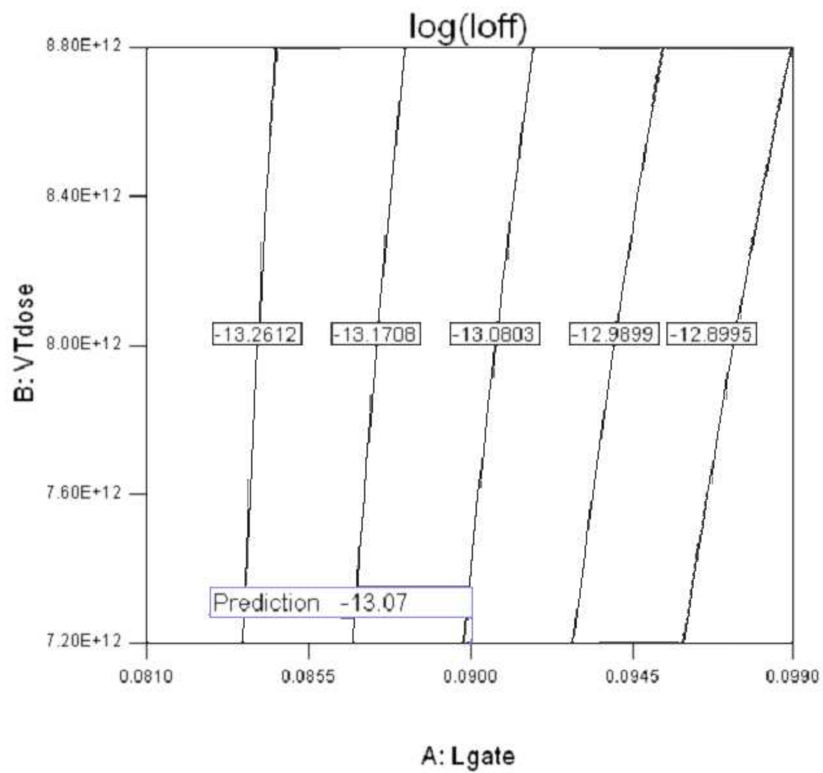


Figure 8.23: The log(Ioff)'s contour plot of the optimal solution for the SCD1 design using the stepwise regression models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

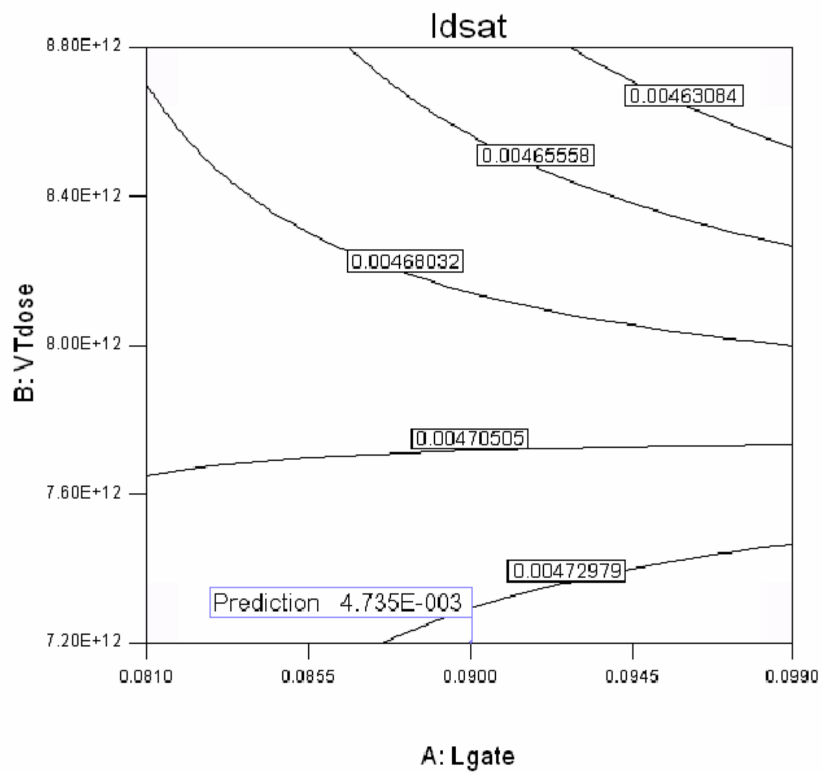


Figure 8.24: The Idsat's contour plot of the optimal solution for the SCD1 design using the stepwise regression models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

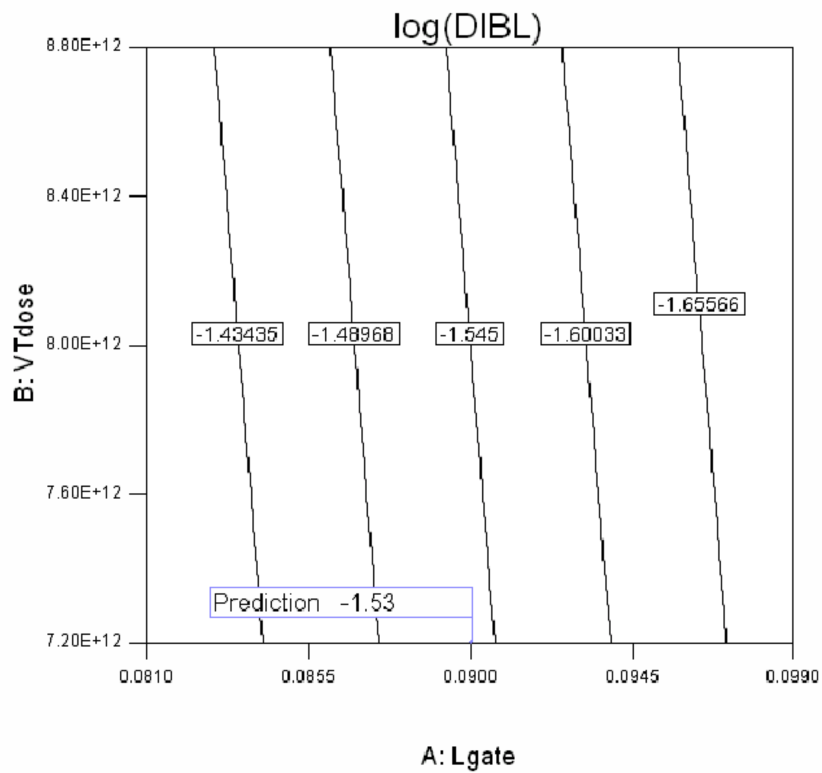


Figure 8.25: The log(DIBL)'s contour plot of the optimal solution for the SCD1 design using the stepwise regression models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

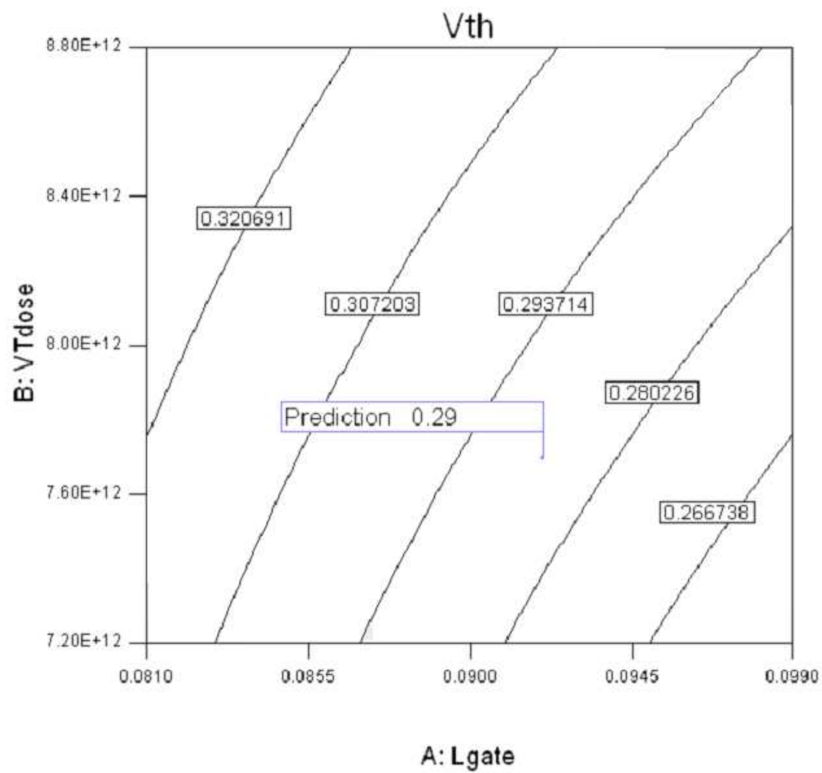


Figure 8.26: The Vth's contour plot of the optimal solution for the SCD2 design using the stepwise regression models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

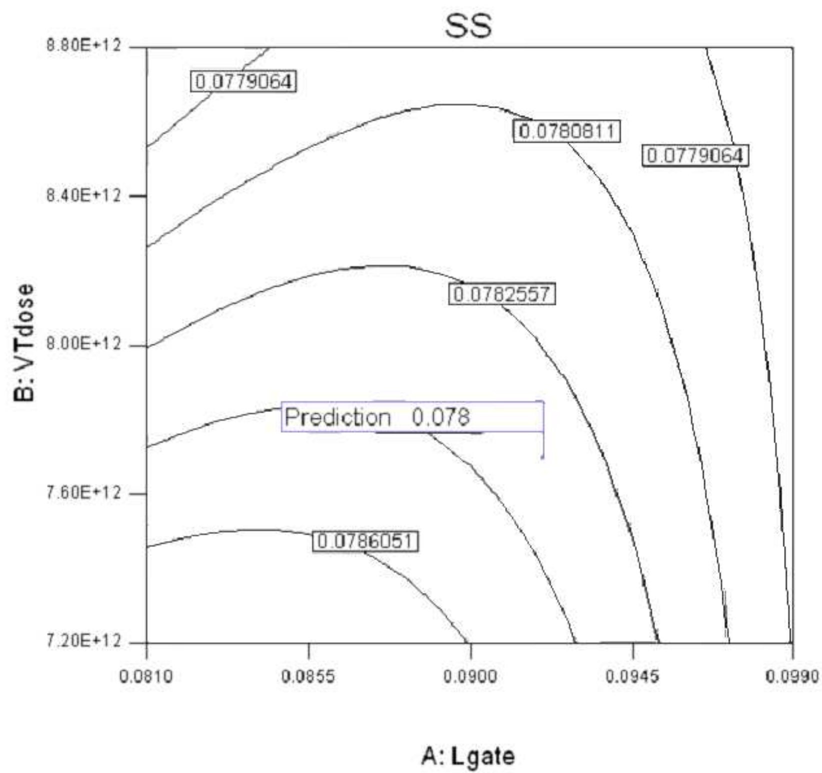


Figure 8.27: The SS' contour plot of the optimal solution for the SCD2 design using the stepwise regression models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

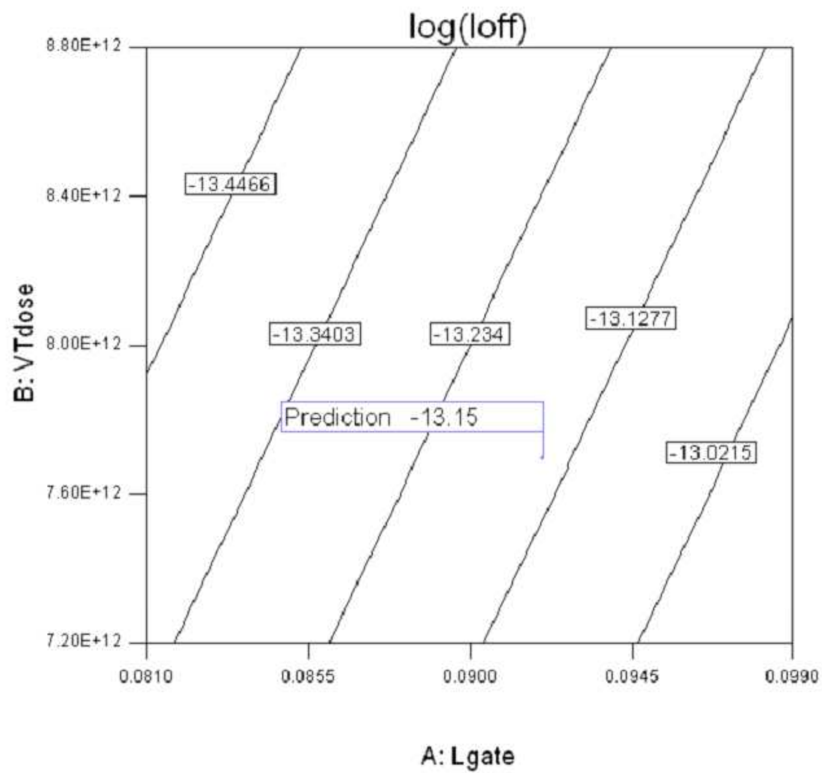


Figure 8.28: The log(Ioff)'s contour plot of the optimal solution for the SCD2 design using the stepwise regression models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

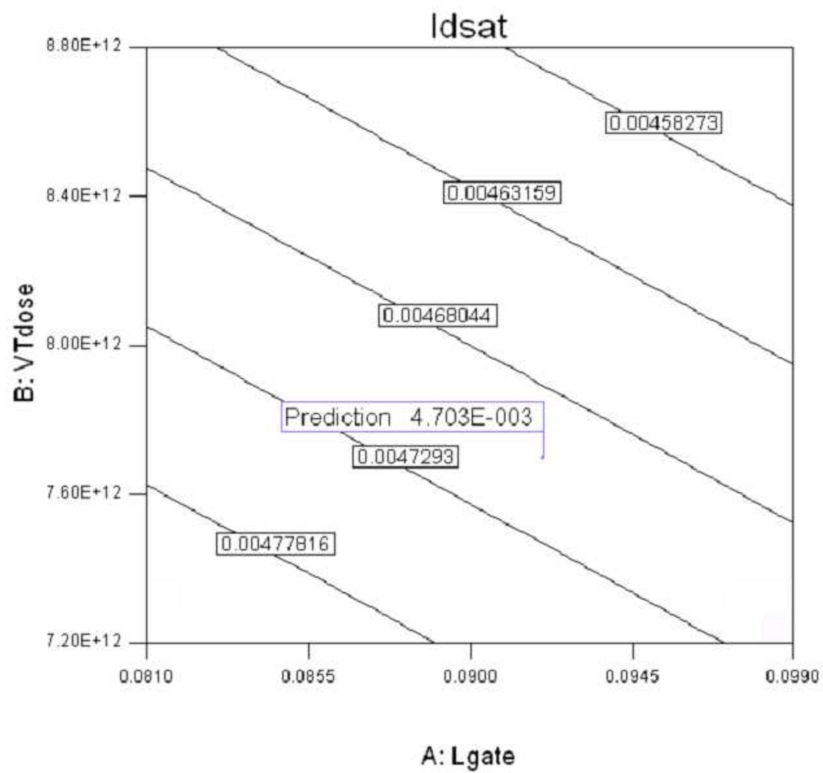


Figure 8.29: The Idsat's contour plot of the optimal solution for the SCD2 design using the stepwise regression models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

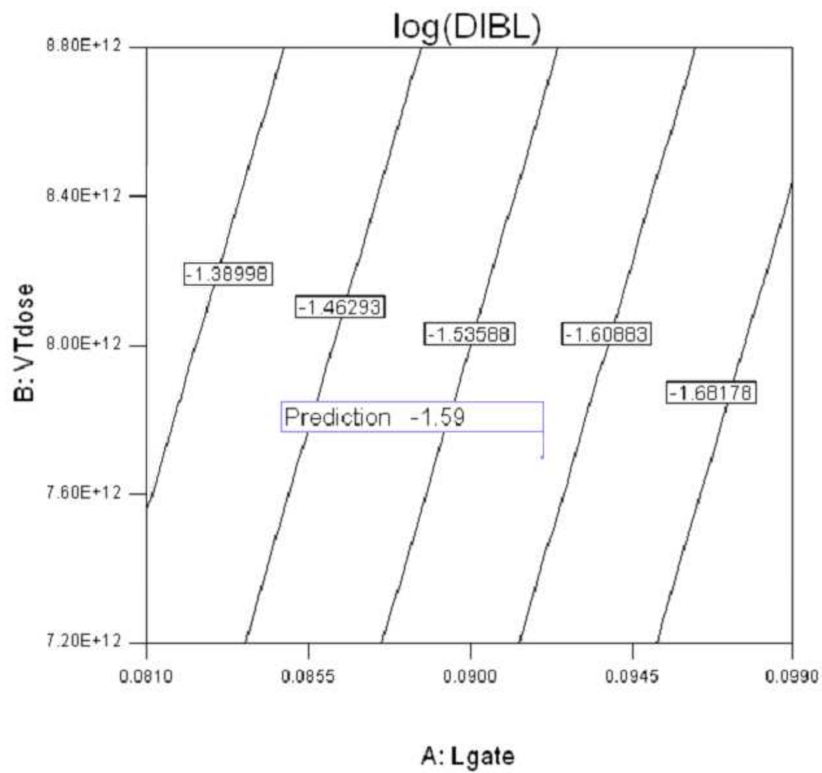


Figure 8.30: The log(DIBL)'s contour plot of the optimal solution for the SCD2 design using the stepwise regression models, the factor levels corresponding the flag are held the optimal condition. The value on the contour means the estimated value of the response.

Chapter 9

Process Sensitivity Analysis

This part directly follows the optimization, where the reduced response surface models and the full 2nd order model are both able to be used. We note that the model used should be the same with the optimization analysis. Following the optimization, we will have six sensitivity analyses corresponding to the six optimal recipes obtained by optimization. In this chapter, we use the SCD1 design with the full 2nd order response surface models to be the example. The procedure of other optimal recipes will be similar to it.

9.1 Sensitivity Analysis Results

The objective is to investigate the sensitivity of a number of key responses by varying a set of critical process parameters. Once a quadratic model for each response has been

obtained, then it can be 'interrogated' by plotting the distribution of each response to a Gaussian variation of the input factors. Numerical simulations for the optimized NMOSFET process are performed and statistical response distribution for each physical quantities are determined. To represent manufacturing fluctuations, random values for each input condition are selected from a normal distribution, and the corresponding physical quantity values are calculated by their response surface models. The mean for each input condition is set to its optimized value. The standard deviation for each input condition is set to 3% of its optimized value with the exception of the *OXtemp* and the VTenergy, which is set to 1%. The standard deviation of the oxidation temperature is lowered to provide more realistic variations in T_{OX}. In order to study the statistical nature of the 90nm NMOSFET process, with respect to 7 key parameters that are selected, we have generated 1000 normally and independently distributed pseudo-random numbers for each of these 7 process parameters.

Table 9.1 presents the results of distribution statistics for the NMOSFET devices. The mean values of the resulting statistical distributions are in good agreement with the simulated physical quantity values. All of the responses for NMOSFET devices varies less than ten percent from their mean values. Figures 9.1, 9.2, 9.3, 9.4, and 9.5 are statistical distributions obtained by the sensitivity analysis on the models for the Vth, the SS, the log(Ioff), the Idsat, and the log(DIBL). Finally we show other four tables of the sensitivity analysis

Table 9.1: Sensitivity analysis for the studied the 90nm NMOSFET process in the SCD1 design using the full 2nd order response surface models, displaying calculated mean and standard deviation for resulting physical quantities.

SCD1-full	mean	$\hat{\sigma}$
Vth (V)	0.27902605	0.019826167
SS (V/dec.)	0.078117241	0.000876393
log(Ioff) (A/μm)	-13.0959706	0.166551543
Idsat (A/μm)	0.004794435	0.000205124
log(DIBL) (V)	-1.54024611	0.149705405

Table 9.2: Sensitivity analysis for the 90nm NMOSFET process in the CCF design using the full 2nd order response surface models, displaying calculated mean and standard deviation for resulting physical quantities.

CCF-full	mean	$\hat{\sigma}$
Vth (V)	0.288374271	0.020536383
SS (V/dec.)	0.078367224	0.000821769
log(Ioff) (A/μm)	-13.1421571	0.163605475
Idsat (A/μm)	0.004709426	0.000198218
log(DIBL) (V)	-1.526898517	0.16234867

results for other four designs. Because the residual results of the full 2nd order response surface models in the SCD2 design are not suitable for sensitivity analysis (this part has been discussed in Chapter 6). Therefor, we omit this part of results.

Table 9.3: Sensitivity analysis for the 90nm NMOSFET process in the CCF design using the stepwise regression response surface models, displaying calculated mean and standard deviation for resulting physical quantities.

CCF-stepwise	mean	$\hat{\sigma}$
Vth (V)	0.29229629	0.01869976
SS (V/dec.)	0.078815433	0.000487024
log(Ioff) (A/μm)	-13.1180939	0.177513073
Idsat (A/μm)	0.004717753	0.000222726
log(DIBL) (V)	-1.56286209	0.152977147

Table 9.4: Sensitivity analysis for the 90nm NMOSFET process in the SCD1 design using the stepwise regression response surface models, displaying calculated mean and standard deviation for resulting physical quantities.

SCD1-stepwise	mean	$\hat{\sigma}$
Vth (V)	0.284041604	0.01922337
SS (V/dec.)	0.078425778	0.000676955
log(Ioff) (A/μm)	-13.0709275	0.178200442
Idsat (A/μm)	0.004742671	0.000219321
log(DIBL) (V)	-1.526059567	0.139616926

9.2 Summary

In this chapter, one process sensitivity analysis obtained from the optimal results of SCD1 have been discussed completely. The mean values and the variations of the device performances are in a good results. Finally we provide the other four tables of the results for

Table 9.5: Sensitivity analysis for 90nm NMOSFET process in the SCD2 design using the stepwise regression response surface models, displaying calculated mean and standard deviation for resulting physical quantities.

SCD2-stepwise	mean	$\hat{\sigma}$
Vth (V)	0.286986376	0.017801237
SS (V/dec.)	0.078345844	0.000866063
log(Ioff) (A/μm)	-13.1452254	0.188743867
Idsat (A/μm)	0.005139351	0.000213312
log(DIBL) (V)	-1.573513749	0.164805662

other four designs.

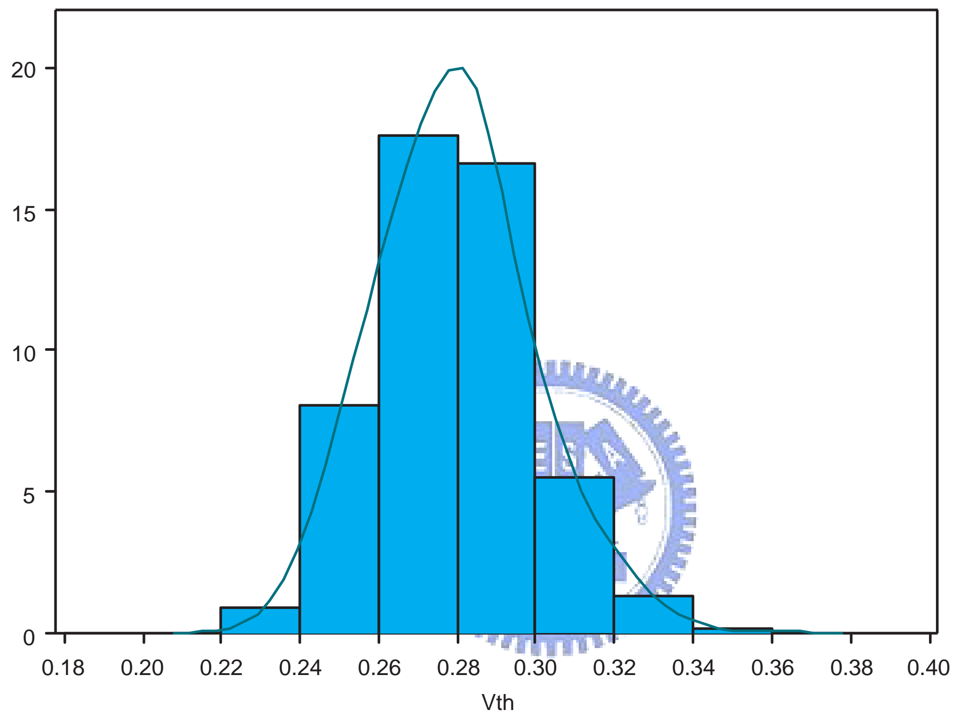


Figure 9.1: Statistical distribution of the model for the threshold voltage in the SCD1 design, which are calculated by the sensitivity analysis and using the full 2nd order response surface models, where the unit of Vth is V.

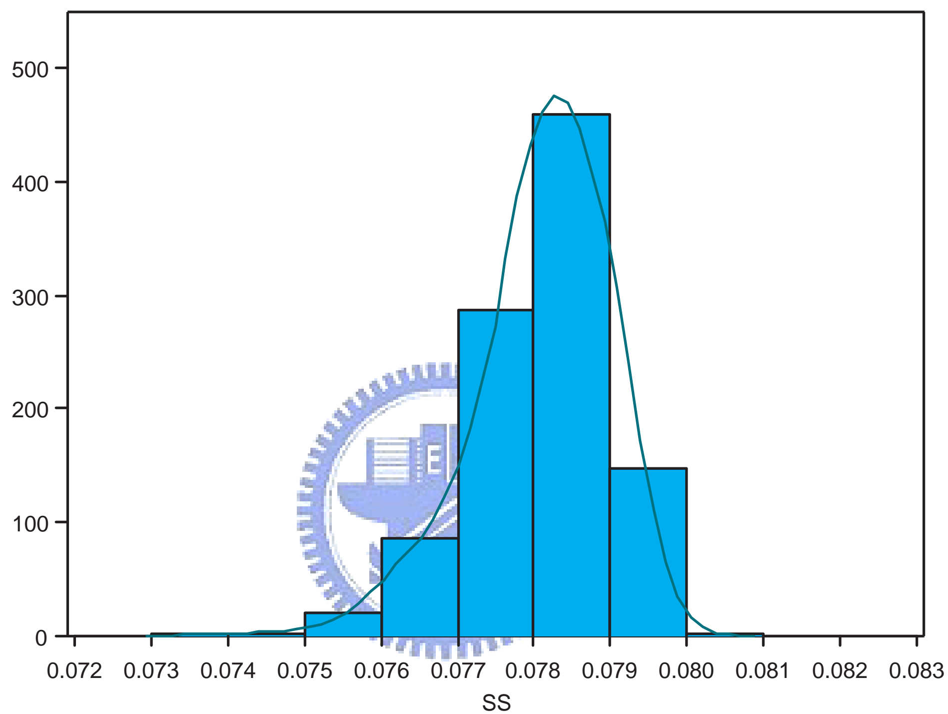


Figure 9.2: Statistical distribution of the model for the subthreshold slope in the SCD1 design, which are calculated by the sensitivity analysis and using the full 2nd order response surface models, where the unit of SS is V/dec.

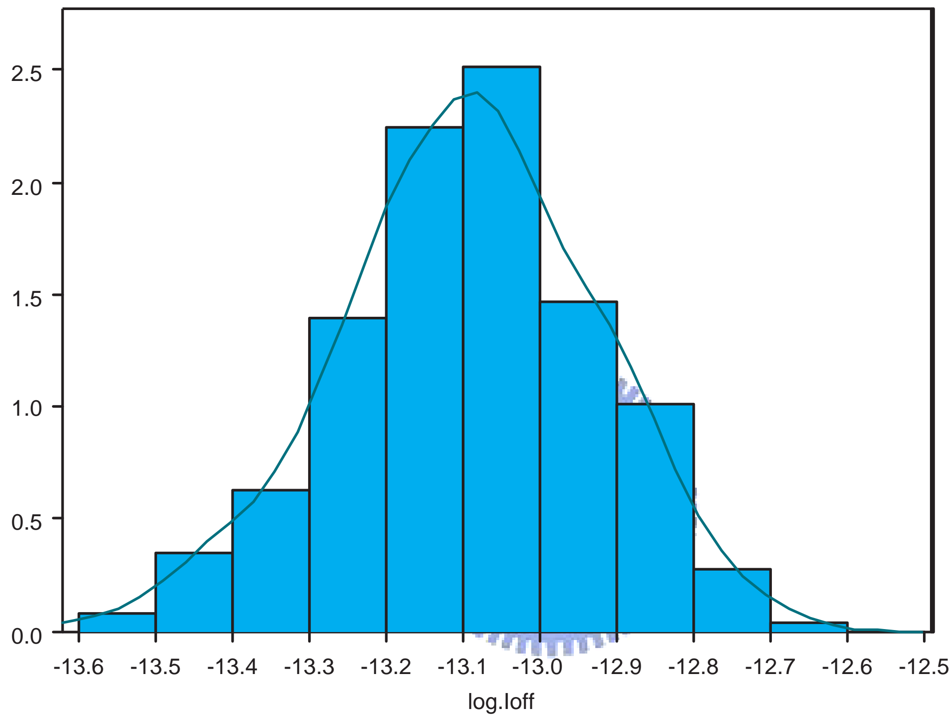


Figure 9.3: Statistical distribution of the model for the log off-state current in the SCD1 design, which are calculated by the sensitivity analysis and using the full 2nd order response surface models, where the unit of log(Ioff) is A/μm.

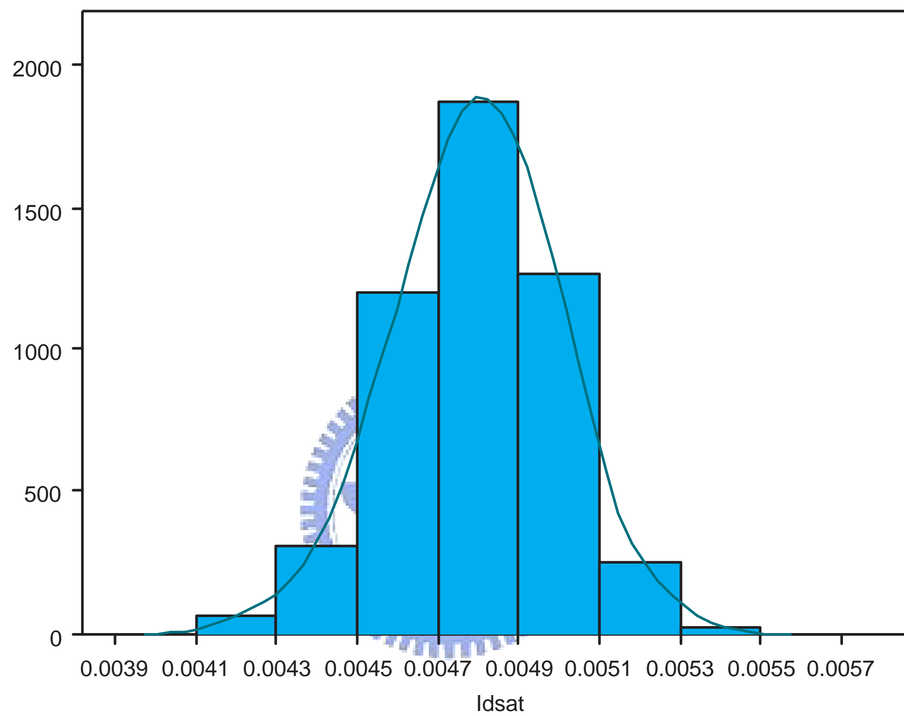


Figure 9.4: Statistical distribution of the model for the on-state current in the SCD1 design, which are calculated by the sensitivity analysis and using the full 2nd order response surface models, where the unit of Idsat is A/μm.

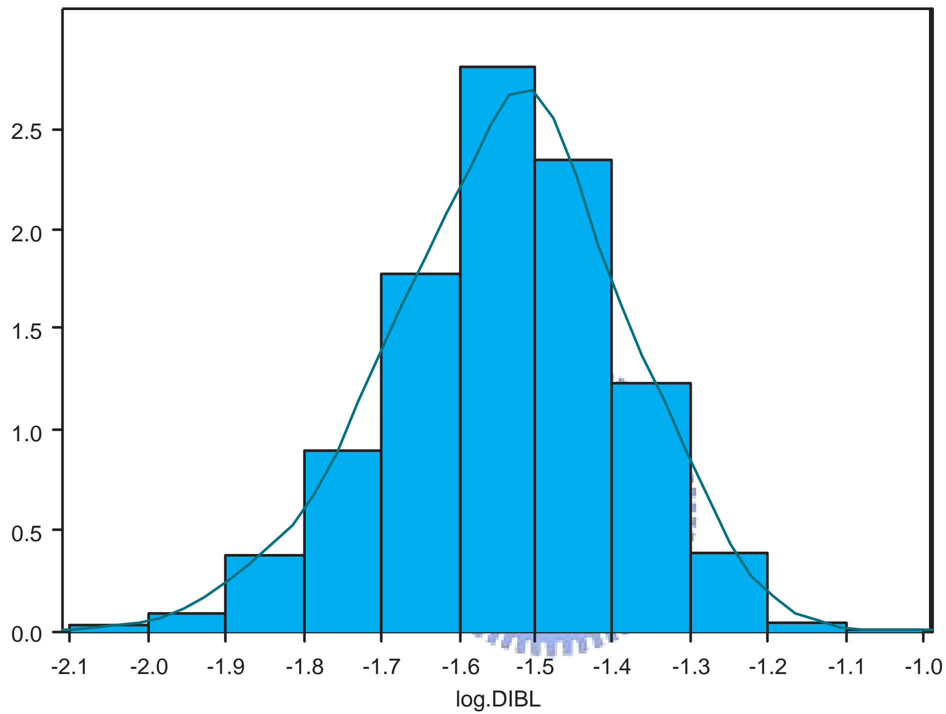


Figure 9.5: Statistical distribution of the model for the log drain-induced barrier lowing in the SCD1 design, which are calculated by the sensitivity analysis and using the full 2nd order surface models, where the unit of log(DIBL) is V.

Chapter 10

Accuracy Verification and Model Adjustment

E S A
1896

In this chapter six verification experiments corresponding six different designs are discussed here. In the following section we provide an empirical procedure for model adjustment. Through this procedure the device performances would be closer to our targets by adjustment the response ranges.

10.1 Verification Experiments

In this section we conduct six verification experiments. The purpose of this check is to verify that the optimal conditions which are suggested by this work do indeed give the

Table 10.1: The comparison of the TCAD results between the nominal case and three optimal cases using the full 2nd order response surface models.

	Goals	Nominal	CCF	SCD1	SCD2
Vth (V)	0.285	0.24182	0.29592	0.285786	0.298368
SS (V/dec.)	minimized	0.079829	0.080453	0.079969	0.080144
log(Ioff) (A/μm)	minimized	-12.68845	-13.21822	-13.01298	-13.26593
Idsat (A/μm)	0.0047	0.005243	0.004581	0.004962	0.004536
log(DIBL) (V)	minimized	-1.57663	-1.39512	-1.25765	-1.46444

projected improvement. We use six optimal recipes obtained from Tables 8.3 and 8.6 to run the TCAD simulator again. The results have shown in Table 10.1 and Table 10.2. We use desirability values to be the criteria of improvement. From the overall desirability values, these six optimal conditions are all improved on the nominal case. Results are provided on Table 10.3 and Table 10.4. The values of the threshold voltage and the on-state current are obtained from the Id-Vg curves. The results have been shown in Figs. 10.1 and 10.1. From Fig 10.1, the values of the threshold voltage could be calculated by the constant current method, we note that the threshold voltage of the nominal case is far from the other optimal recipes for six corresponding designs. The results of the six corresponding designs are close to each other and better than the nominal case. The results of the on-state current are also better than the nominal case, information has been provided in Fig. 10.1.

Table 10.2: The comparison of the TCAD results between the nominal case and three optimal cases using the stepwise regression models.

	Goals	Nominal	CCF	SCD1	SCD2
Vth (V)	0.285	0.24182	0.287433	0.289319	0.288054
SS (V/dec.)	minimized	0.079829	0.08048	0.080765	0.078043
log(Ioff) (A/μm)	minimized	-12.68845	-13.07216	-13.09112	-13.10834
Idsat (A/μm)	0.0047	0.005243	0.004812	0.004772	0.004755
log(DIBL) (V)	minimized	-1.57663	-1.28401	-1.29193	-1.34117

Table 10.3: Desirability values between the nominal case and three optimal cases using the full 2nd order response surface models.

Desirability	Nominal	CCF	SCD1	SCD2
Vth	0.136391	0.781603	0.984275	0.732642
better than nominal case		yes	yes	yes
SS	0.517056	0.454675	0.503146	0.485576
better than nominal case		no	no	no
log(Ioff)	0.188452	0.718218	0.512981	0.765935
better than nominal case		yes	yes	yes
Idsat	0	0.76117	0.47501	0.672539
better than nominal case		yes	yes	yes
log(DIBL)	0.720787	0.493897	0.322058	0.580551
better than nominal case		no	no	no
Overall desirability	0	0.625766	0.522287	0.638822
better than nominal case		yes	yes	yes

Table 10.4: Desirability values between the nominal case and three optimal cases using the stepwise regression models.

Desirability	Nominal	CCF	SCD1	SCD2
Vth	0.136391	0.951338	0.913611	0.938921
better than nominal case		yes	yes	yes
SS	0.517056	0.452038	0.423485	0.695721
better than nominal case		no	no	yes
log(Ioff)	0.188452	0.572158	0.59112	0.608342
better than nominal case		yes	yes	yes
Idsat	0	0.776046	0.856402	0.889196
better than nominal case		yes	yes	yes
log(DIBL)	0.720787	0.355011	0.364918	0.426466
better than nominal case		no	no	no
Overall desirability	0	0.583756	0.589969	0.684887
better than nominal case		yes	yes	yes

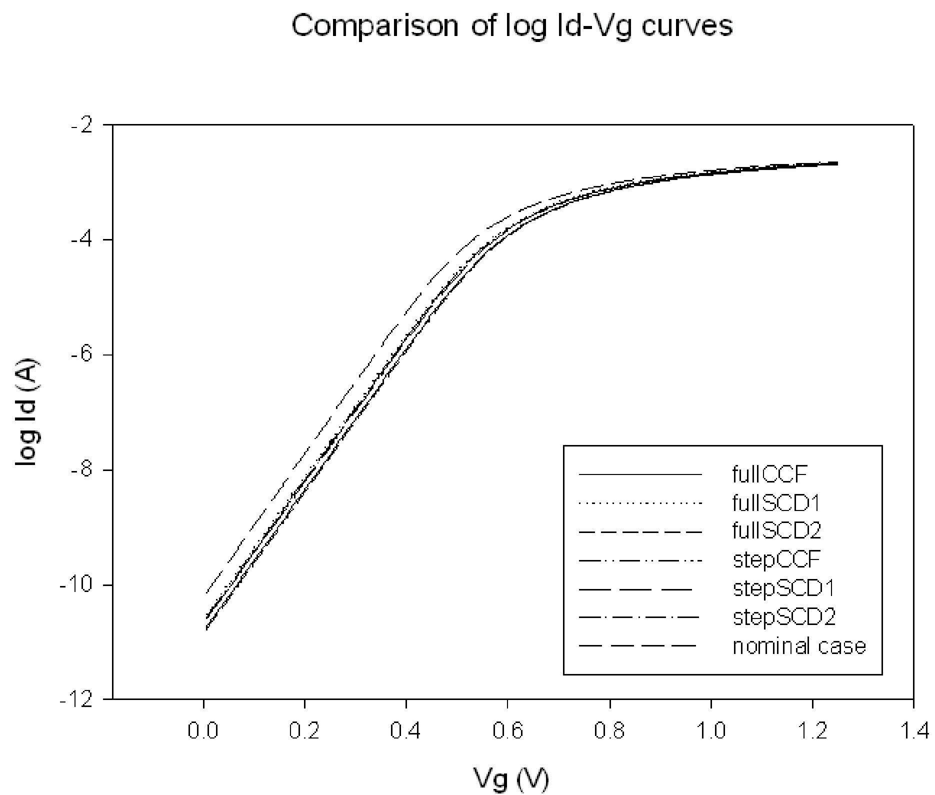


Figure 10.1: The log Id-Vg curves of the 90nm NMOSFET device structure obtained from the nominal case and all desings parameter settings. The values of the threshold voltage are calculated by the constant current method.

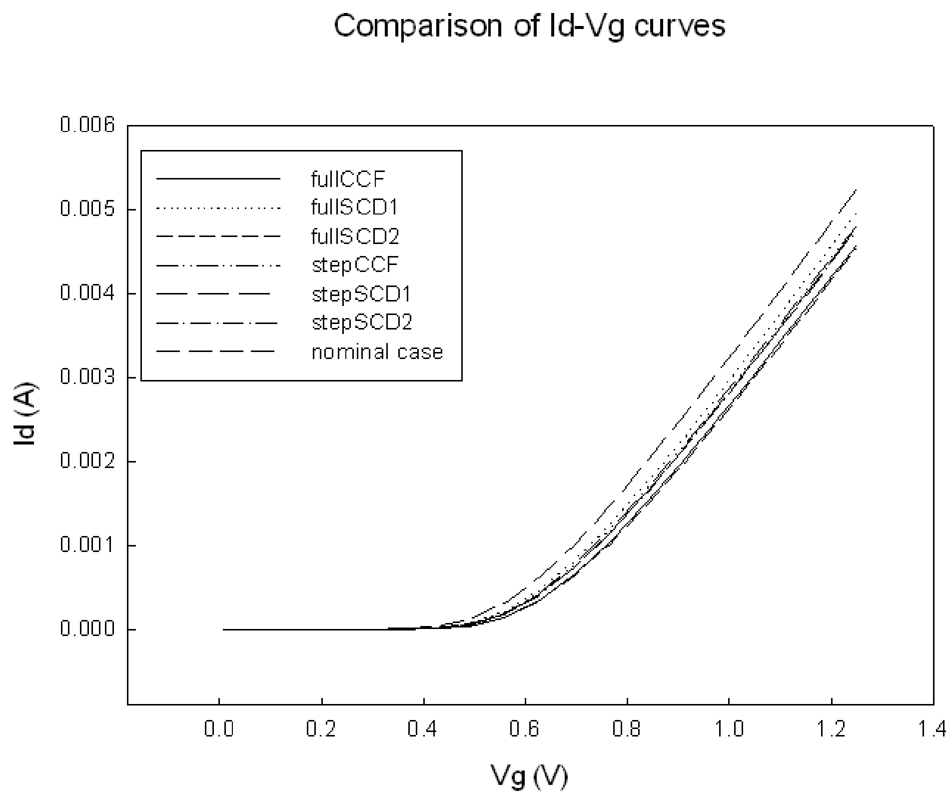


Figure 10.2: The Id-Vg curves of the 90nm NMOSFET device structure obtained from the nominal case and all designs parameter settings. The values of the on-state current are corresponding the base line.

10.2 An Empirical Procedure for Model Adjustment

An empirical procedure for model adjustment is discussed here. This part could be developed by using the more systematical methods instead of the empirical method in the future. Return to be the new DOE, choose more relative factors might be the alternative to be investigated. In our empirical procedure, the optimal results in TCAD can be adjusted through the following correction procedure:

Adjust the process optimization recipe

Begin

Run TCAD from the DOE

Create the RSM

Optimization

Estimate the error between the results of RSM and TCAD

While Error > Stop Criteria

Modify goals of the range for the responses()

Optimization

Estimate error

End While

Output results

End

Modify goals of the range for the responses()

Begin

For all responses

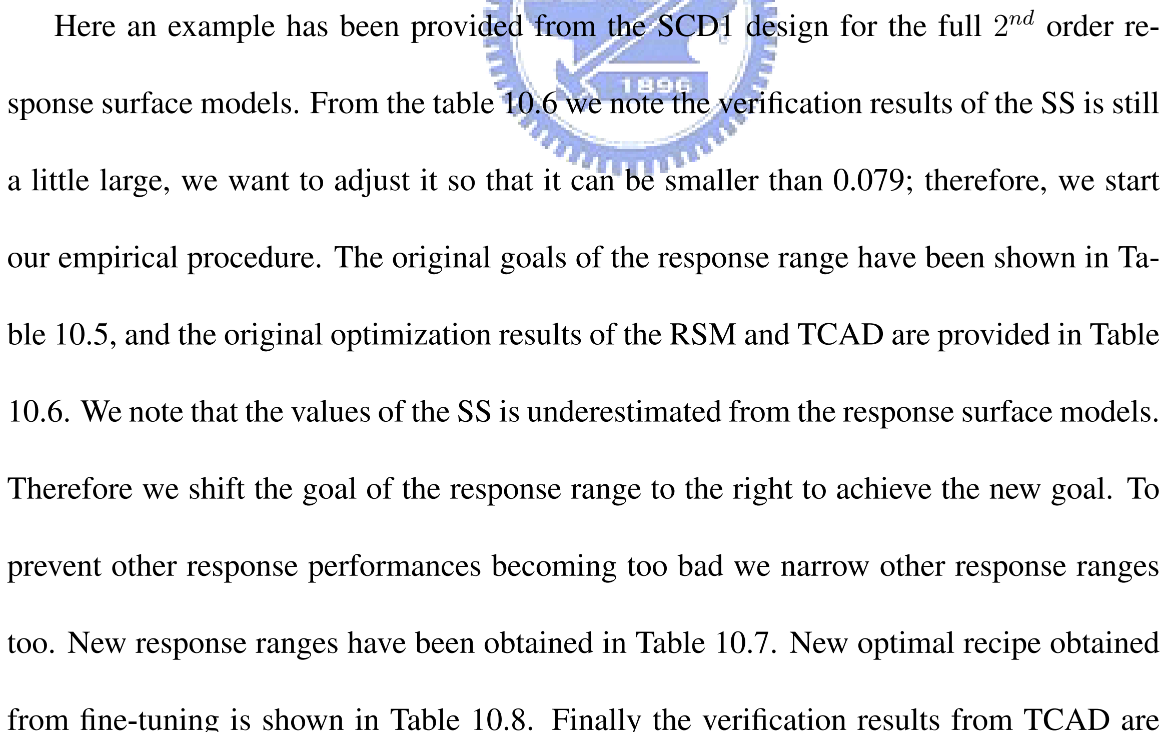
If the response in TCAD larger than the target

Reset the boundary of the response in optimization and shift it to the left

Else

Reset the boundary of the response in optimization and shift it to the right

End

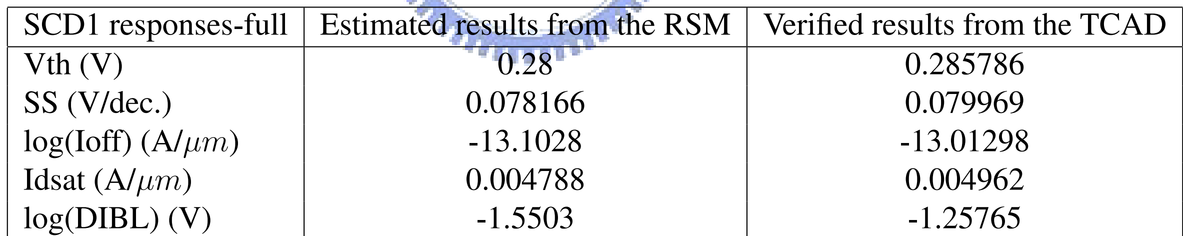


Here an example has been provided from the SCD1 design for the full 2nd order response surface models. From the table 10.6 we note the verification results of the SS is still a little large, we want to adjust it so that it can be smaller than 0.079; therefore, we start our empirical procedure. The original goals of the response range have been shown in Table 10.5, and the original optimization results of the RSM and TCAD are provided in Table 10.6. We note that the values of the SS is underestimated from the response surface models. Therefore we shift the goal of the response range to the right to achieve the new goal. To prevent other response performances becoming too bad we narrow other response ranges too. New response ranges have been obtained in Table 10.7. New optimal recipe obtained from fine-tuning is shown in Table 10.8. Finally the verification results from TCAD are

Table 10.5: The targets of the responses.

Responses	Goals	Lower Limit	Upper Limit
Vth (V)	target = 0.285	0.235	0.335
SS (V/dec.)	minimize	0.075	0.085
log(Ioff) (A/μm)	minimize	-13.5	-12.5
Idsat (A/μm)	target = 0.0047	0.0042	0.0052
log(DIBL) (V)	minimize	-1.8	-1.0

Table 10.6: The results of the estimated values from the RSM and the verified values from the TCAD.



SCD1 responses-full	Estimated results from the RSM	Verified results from the TCAD
Vth (V)	0.28	0.285786
SS (V/dec.)	0.078166	0.079969
log(Ioff) (A/μm)	-13.1028	-13.01298
Idsat (A/μm)	0.004788	0.004962
log(DIBL) (V)	-1.5503	-1.25765

provided in Table 10.9. The results are improved, real SS value obtained from the adjusted recipe is smaller than 0.079 now, and other response value are also acceptable and even better than before.

Table 10.7: The new adjusted targets of the responses.

Responses	Goals	Lower Limit	Upper Limit
Vth (V)	target = 0.285	0.25	0.315
SS (V/dec.)	minimize	0.075	0.079
log(Ioff) (A/μm)	minimize	-13.5	-13.1
Idsat (A/μm)	target = 0.0047	0.0044	0.005
log(DIBL) (V)	minimize	-1.8	-1.5

Table 10.8: Comparison of the optimal recipe between the original SCD1 case and the new case through adjustment using the full 2nd order response surface models.

Optimal recipe	Original	After adjustment
Lgate (μm)	0.0876	0.09
Vtdose (cm^{-2})	7.2E+12	7.97E+12
PTenergy (keV)	12.51	13.07
PTdose (cm^{-2})	1.2E+13	1.2E+13
VTenergy (keV)	29	29
OXtemp (°C)	772.66	771.97
OXtime (mins)	9.10	9

10.3 Summary

In this chapter, six verification experiments corresponding six different designs have been obtained, and the results are really improved from the nominal case. Following an empirical procedure for the model adjustments are performed, and through the adjustment the results

Table 10.9: The original results and the adjusted results of the verified values from the TCAD.

SCD1 responses-full	Original verified results	Adjusted verified results
Vth (V)	0.285786	0.292786
SS (V/dec.)	0.079969	0.078157
log(Ioff) (A/μm)	-13.01298	-13.1762
Idsat (A/μm)	0.004962	0.004709
log(DIBL) (V)	-1.25765	-1.34077

are really closer to the targets.

Chapter 11

Conclusions and Suggestions on Future Work

In this thesis, a systematic method for the fabrication optimization and the sensitivity analysis of sub-100nm NMOSFET devices have been investigated. The process location problem has been explored by using our methodology. A 90nm NMOSFET manufacturing technology was provided to be an example. Five physical quantities need to be optimized. They are: 1) shifting the value of threshold voltage to the specific target; 2) minimizing the subthreshold slope; 3) minimizing the off-state current; 4) moving the value of on-state current to a specific target; 5) and minimizing the drain-induced barrier lowing. Achieved conclusions and suggested future works are listed in the following sections.

11.1 Conclusions

Through performing the screen design, we have successfully selected seven significant fabrication factors for five physical responses. They are the threshold voltage implant dose, the threshold voltage implant energy, the oxide growth time, the oxide growth temperature, the gate length, the punch-through implant dose2, and the punch-through implant energy. The following step is the design of experiment, we performed three different designs which were CCF, SCD1, and SCD2, in this part. From the response surface models, we have obtained the good results in the 90nm n-type MOSFETs process. For example, the threshold voltage of the device in the original fabrication, is 0.236 V and too small. The on-state current of the original device is too large for 0.0053 A. Through performing our proposed method, the threshold voltage of the new device could be raised to our target 0.285 V. The on-state current would be reduced to a better performance 0.0047 A at the same time. The achievement was done simultaneously for other explained physical quantities.

In this work we have used the face central cube design, a special form of the central composite design, to build the response surface models. About 316 hours computing-time is required to perform this design. Two smaller designs could be used instead of this traditional one; a smaller composite design spend about 188 hours to get all of the information, the other smallest composite design only need about 148 hours. More than 50% time has been saved in this design, and we find that the device performances which

are fabricated by these three optimal recipes are close to each other and also acceptable in our physical constrains.

The response surface models results of three different designs of original CCF design (79 experiments), smaller composite design (47 experiments), and smallest composite design (37 experiments) are provided in Chapter 6. The results of the variable selection for the three design have been discussed in Chapter 7. We have provide the comparison of optimization results for different designs and models in Chapter 8. Optimal results of three design, no matter using the full 2nd order response surface models or the stepwise regression models, are all improved than the original nominal case. Finally the outcomes of the process sensitivity analysis have been shown in Chapter 9. In the process sensitivity analysis, the input factors have been assumed to be normally distributed about their mean (for this work the optimal process input condition values). The standard deviation for each factor has been set as a percentage of its mean values. We note that the results are acceptable. We have verified the results in this work and the device performances were really better than before.

Considered the cost, computer time, accuracy, and model assumption, we suggest that experimenter could choose the smaller composite design instead of the traditional central composite design. Finally we used the response surface model to be the sensitivity analysis of the 90nm NMOSFET optimal recipe, this step could help us to know whether the device

performances are out of the specifications in truly fabrication.

11.2 Suggestions for Future Work

In the future works, we could use the similar methodology to build the models through the true measurements in the fabrication instead of the data from the TCAD simulation (see Figure 11.2). Thus we could use these replicated data to build the new response surface models, and take one step ahead to reduce the fabrication variation.

For the statistical method, we could use higher order RSM to explain the relationship between the responses and factors. The error between the estimated responses from the RSM and the true measurements obtained from the TCAD might be reduced, and the response surface models would be closer to the true situations. We could develop an systematically adjusted method to achieve the target after the accuracy verification. On the other hand, because of the dependence between each responses, we could use the generalized linear models to estimate the coefficients of the response surface models to perform better estimation.

For the semiconductor industry, this work could be applied to other fabrication process, such as the SOI fabrication process or the 65nm MOSFET fabrication. It also could be performed in different topics, for example, we could use this proposed statistical method to investigate the relationship between the circuit and the device; therefore we could know

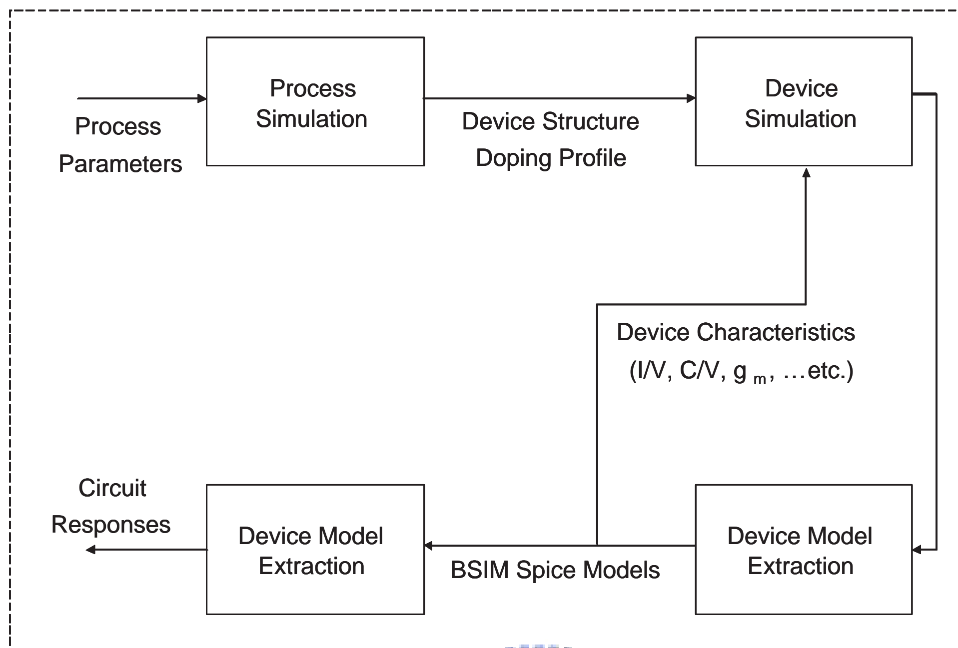
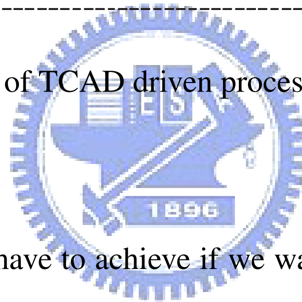


Figure 11.1: Flow diagram of TCAD driven process design to circuit analysis.



how would the device performance have to achieve if we want to have the specific circuit performance. Finally we could apply our proposed statistical method to connect with the process simulation and circuit simulation as like Fig 11.1.

For other applications, this work would be applied to any other optimal problems in the engineering. For example, we could use this method to improve the key technique of the printer instead of the traditional method, which is to optimize systematically one control factor at a time or to perform many try-and-error experiments. Our proposed statistical method is more completed, economical, and systematical.

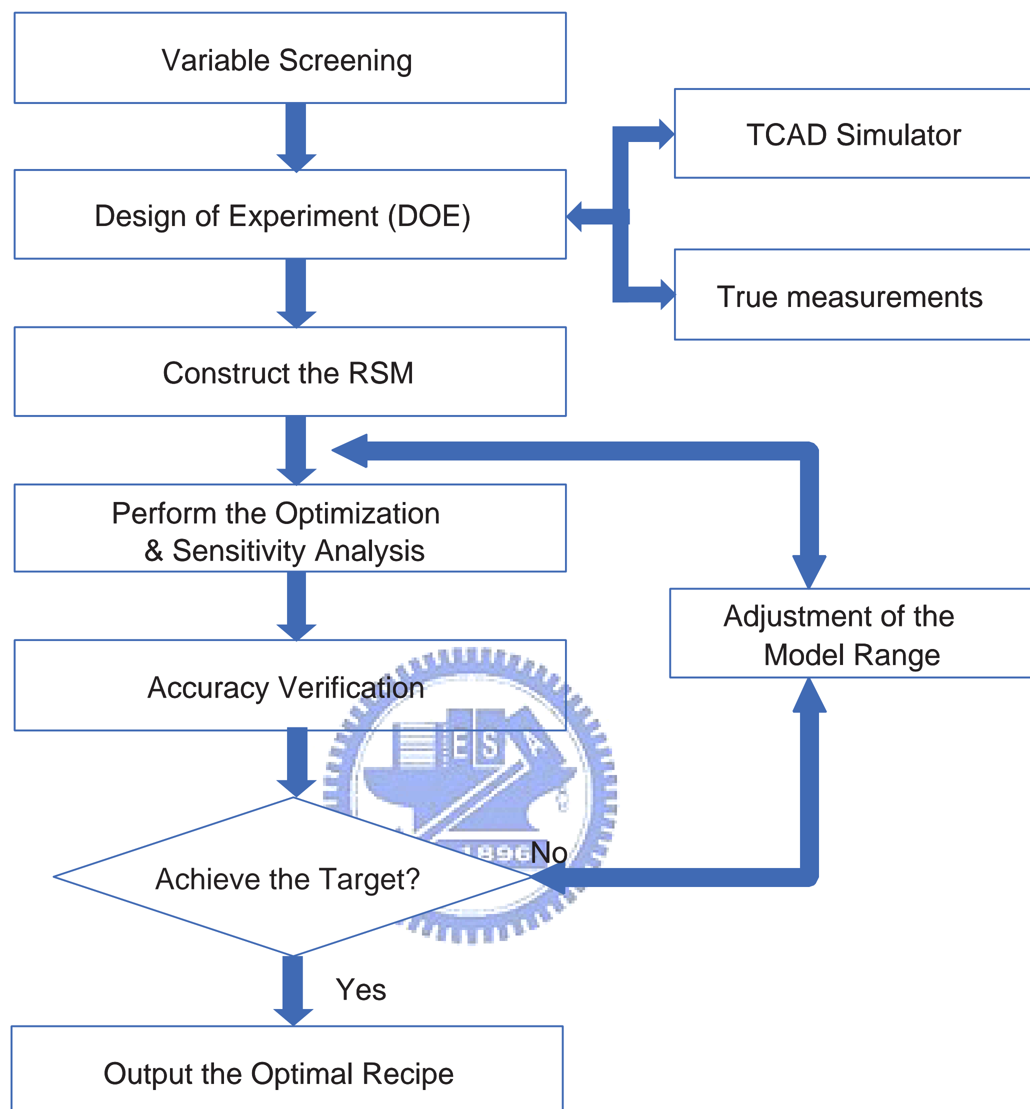


Figure 11.2: Procedure flow chart for the future work. It contains the screen design, the design of experiment, the response surface model, the optimization and sensitivity analysis, and the verification. These methods are based on the data from the true measurements instead of the TCAD simulation.