Bibliography

[1] G. E. P. Box and Norman R. Draper, *Empirical Model-Building and Response Surfaces* , New York, Wiley, 1987.

[2] K. C. Brown, J. Coniff, and R. Barber, Develpment and Production Integration of a Submicron Tungsten Interconnect Process, *Proceedings of IEEE on VMIC Conference*, June 11-12, 1991, pp. 308-311.

[3] P. Chiang and M.-R. Chen, Device Optimization Using Response Surface Methodology with Orthogonal Arrays, *Proceedings of IEEE International Workshop on Statistical Methodology*, pp. 17-20, 10 June 2001.

[4] C. Daniel, Use of Half-normal Plots in Interpreting Factorial Two-level Experiments, *Technometrics*, Vol. 1, 1959, pp. 311-341.

[5] J. H. Dodgson, A Graphical Method for Assessing Mean Squares in Saturated Fractional Designs, *Journal of quality technology*, Apr. 2003, Vol. 35, No. 2, pp. 206-212.

[6] N. R. Draper and D. K. J. Lin, Small Response-Surface Designs, *Technometrics*, Vol. 32, 1990, pp. 187-194.

[7] G.J. Gaston, Process Sensitivity Analysis Using Design of Experiments and Full 2D TCAD Simulations, *Proceedings of IEE Colloquium on Advanced MOS and Bi-Polar Devices*, pp. 1-6, 14 Feb 1995.

[8] H. O. Hartley, Smallest Composite Designs for Quadratic Response Surfaces, *Biometrics*, Vol. 15, 1959, pp. 611-624.

[9] R. E. Jones and T. C. Mele, Use of Screening and Response Surface Experimental Designs for Development of a 0.5 μm CMOS Self-aligned Titanium Silicide Process, *IEEE Transactions on Semiconductor Manufacturing*, pp. 281-286, Vol. 4. No. 4. Nov. 1991.

[10] A. I. Khuri and J. A. Cornell, *Response Surfaces: Designs and Analysis*, Marcel Dekker, Inc., 1987.

[11] H. Kim and H. Hwang, Effect of Statistical Process Variation of MOSFET on Hot Carrier Lifetime, *Solid-State Electronics*, 43, 1999, pp. 989-991.

[12] D. L. Knill, A. A. Giunta, C. A. Baker, B. Grossman, W. H. Mason, R. T. Haftka, and L. T. Watson, Response Surface Models Combining Linear and Euler Aerodynamics

for Supersonic Transport Design, *Journal of Aircraft*, Vol. 36, No. 1, Jan.-Feb., 1999, pp. 75-86.

[13] F. Lallement, A. Grouillet, M. Juhel, J.-P. Reynard, D. Lenoble, Z. Fang, S. Walther, Y. Rault, L. Godet, and J. Scheuer, Fabrication of N+/P Ultra-shallow Junctions by Plasma Doping for 65nm CMOS Technology, *Surface & Coatings Technology*, Vol. 186, 2004, pp. 17-20.

[14] Y. Li, A Quantum Correction Poisson Equation for Metal-Oxide-Semiconductor Structure Simulation, *Semiconductor Science and Technology*, Vol. 19, No. 7, July 2004, pp. 917-922.

[15] Y. Li, A Comparative Study of Numerical Algorithms in Calculating Eigenpairs of the Master Equation for Protein Folding Kinetics, *Abstracts of the 10th International Workshop on Computational Electronics* (IWCE-10), 24-27 Oct. 2004, pp. 201-202.

[16] Y. Li, H.-M. Lu, T.-W. Tang, and S. M. Sze, A Novel Parallel Adaptive Monte Carlo Method for Nonlinear Poisson Equation in Semiconductor Devices, *Mathematics and Computers in Simulation*, Vol. 62, No. 3-6, March 2003, pp. 413-420.

[17] Y. Li and S.-M. Yu, A Parallel Adaptive Finite Volume Method for Nanoscale Double-gate MOSFETs Simulation, *Journal of Computational and Applied Mathematics*, Vol. 175, No. 1, March 2005, pp. 87-99.

[18] Y. Li, C.-K. Chen, S.-S.Lin, T.-S. Chao, J.-L. Lin, and S.M. Sze, An Implementation of Parallel Dynamic Load Balancing for Adaptive Computing in VLSI Device Simulation, *Proceedings of 15th International Parallel and Distributed Processing Symposium*, pp. 23-27 April 2001.

[19] Y. Li, S.S. Chung, and J.-L. Liu, A Novel Approach for the Two-dimensional Simulation of Submicron MOSFETs Using Monotone Iterative Method, *Proceedings of International Symposium on VLSI Technology, Systems, and Applications*, pp. 27-30, 8-10 June 1999.

[20] Y. Li, T.-S. Chao, and S. M. Sze, A Novel Parallel Approach for Quantum Effect Simulation in Semiconductor Devices, *International Journal of Modelling & Simulation*, Vol. 23, No. 2, April 2003, pp. 94-102.

[21] Y. Li, J.-W. Lee, T.-W. Tang, T.-S. Chao, T.-F. Lei, and S. M. Sze, Numerical Simulation of Quantum Effects in High-k Gate Dielectrics MOS Structures using Quantum Mechanical Models, *Computer Physics Communications*, Vol. 147, No. 1-2, August 2002, pp. 214-217.

[22] Y. Li, J.-L. Liu, T.-S. Chao, and S. M. Sze, A New Parallel Adaptive Finite Volume Method for the Numerical Simulation of Semiconductor Devices, *Computer Physics Communications*, Vol. 142, No. 1-3, Dec. 2001, pp. 285-289.

[23] Z. Malik, D. Dyck, J. Nelder, R. Spence and D. Lowther, Response Surface Models Using Function Values and Gradient Information, with Application to the Design of an Electromagnetic Device, *Proceedings of Engineering design conference in Design Reuse*, pp. 199-209, Brunel, 1998.

[24] D. C. Montgomery, *Design and Analysis of Experiments*, New York: John Wiley & Sons, Inc., 1997.

[25] R. H. Myers, A. I. Khuri, and W. H. Carter, Jr., Response Surface Methodology: 1966-1988, *Technometrics*, Vol. 31, No. 2, May 1989, pp. 137-157.

[26] R. H. Myers and D. C. Montgomery, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, New York: John Wiley & Sons, Inc., 2002.

[27] M. S. Phadke, *Quality Engineering Using Robust Design*, AT&T Bell Laboratories, 1989.

[28] D. J. Roulston, *An Introduction to the Physics of Semiconductor Devices*, Oxford University Press, New York, 1999.

[29] H. C. Srinivasaiah and N. Bhat, Response Surface Modeling of 100nm CMOS Process Technology Using Design of Experiment, *Proceedings of International Symposium on VLSI Technology, Systems, and Applications*, pp. 285-290, 2004.

[30] S. M. Sze, *Semiconductor Devices: Physics and Technology*, John Wiley & Sons, New York, 2002.

[31] J.-Y. Tsai, K.X. Zhang, C.M. Osburn, Optimization of Lightly-doped-drain (LDD) Structure for Sub-quarter-μm Devices Using Statistical Design and Response Surface Methodology, *Proceedings of International Symposium on VLSI Technology, Systems, and Applications*, pp. 285-290, 1995.

[32] R. A. Viscarra Rossel and A. B. Mcbratney, A Two-factor Empirical Deterministic Response Surface Calibration Model for Site-specific Predictions of Lime Requirement, *Precision Agriculture*, 2, 2000, pp. 163-178.

[33] S. Williams, K. Varahramyan, and W. Maszara, Statistical Optimization and Manufacturing Sensitivity Analysis of $0.18\mu m$ SOI MOSFETs, *Microelectronic Engineering*, Vol. 49, 1999, pp. 245-261.

[34] C. F. J. Wu and M. Hamada, *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York: John Wiley & Sons, Inc., 2000.

[35] *Engineering Statistics Handbook* [Online]. Available: http://www.itl.nist.gov/div898/handbook/

Appendix A

Response Surface Models

In this appendix we state basic ideas of response surface models (RSM). At the beginning, a historical perspective of RSM is reviewed. In the second part, we introduce several terminologies, which are the basic concept of the response surface models.
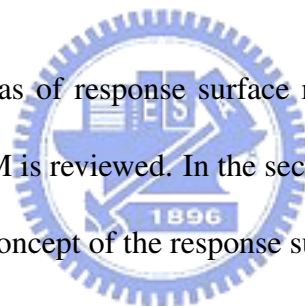
A.1 A Historical Perspective

The roots of RSM can be traced back to the pioneering work of the great statistician R. A. Fisher in the 1930s at the Rothamsted Agriculture Experimental Station in the United Kingdom [10]. Fisher's work and the notable contributions by F. Yates and D. J. Finney were motivated by problems in agriculture and biology. Because of the nature of agricultural experiments, they tend to be large in scale, take a long time to complete, and must

cope with variations in the field. Such considerations led to the development of blocking, randomization, replication, orthogonality, and the use of analysis of variance (ANOVA) and fractional factorial designs [34]. The theory of combinatorial designs, to which R. C. Bose has made fundamental contributions, was also stimulated by problems in block designs and fractional factorial designs. The work in this era also found applications in social science research and in the textile and woolen industries.

The next era of rapid development came soon after World War II. In attempting to apply previous techniques to solve problems in the chemical industries, G. E. P. Box and co-workers at Imperial Chemical Industries in England discovered that new techniques and concepts had to be developed to cope with the unique features in process industries [10]. Their objective was to explore relationships such as those between the yield of a chemical process and a set of input variables presumed to influence the yield. Since the pioneering work of Box and his co-workers, RSM has been successfully used and applied in many diverse fields such as chemical engineering, industrial development and process improvement, agricultural and biological research, even computer simulation, to name just a few. The new techniques focused on process modelling and optimization rather than on treatment comparisons, which was the primary objective in agricultural experiments [26]. The experiments in process industries tend to take less time but put a premium on run size economy because of the cost of experimentation. These time and cost factors naturally favor

sequential experimentation. The same considerations led to the development of new techniques for experimental planning, notably central composite designs (CCD) and optimal designs. Their analysis relies more heavily on regression modelling and graphical analysis [26]. Process optimization based on the constructed model is also emphasized. Because the choice of design is often linked to a particular model (e.g., a second-order central composite design for a second-order regression model) and the experimental region may be irregularly shaped, a flexible strategy for finding designs to suit a particular model and experimental region is called for. With the availability of fast computational algorithms, optimal designs (which was pioneered by J. Kiefer) have become an important part of this strategy [10].

The relatively recent emphasis on variation reduction has provided a new source of inspiration and techniques in experimental design. In manufacturing the ability to make many parts with few defects is a competitive advantage [34]. Therefore variation reduction in the quality characteristics of these parts has become a major focus of quality and productivity improvement. G. Taguchi advocated the use of robust parameter design to improve a system (i.e., a product or process) by making it less sensitive to variation, which is hard to control during normal operating or use conditions of the product or process. The input variables of a system can be divided into two broad types: control factors, whose values remain fixed once they are chosen, and noise factors, which are hard to control during normal conditions. By exploiting the interactions between the control and noise factors, one can

achieve robustness by choosing control factor settings that make the system less sensitive to noise variation. The new paradigm is variation modelling and reduction. Traditionally, when the mean and variance are both considered, variance is used to assess the variability of the sample mean as in the t test or of the treatment comparisons as in the analysis of variance. The focus on variation and the division of factors into two types led to the development of new concepts and techniques in the planning and analysis of robust parameter design experiments. The original problem formulation and several basic concepts were developed by G. Taguchi [34]. Other basic concepts and many sound statistical techniques have been developed by statisticians since the mid-1980s.

From this historical background, we now classify experimental problems into five broad categories according to their objectives [34].

1. Treatment Comparisons. The main purpose is to compare several treatments and select the best ones. Examples of treatments include varieties (rice, barley, corn, etc.) in agriculture trails, sitting positions in ergonomic studies, instructional methods, machine types, suppliers, etc.

2. Variable Screening. If there is a large number of variables in a system but only a relatively small number of them is important, a screening experiment can be conducted to identify the important variables. Such an experiment tends to be economical in that it has few degrees of freedom left for estimating error variance and higher order terms like

quadratic effects or interactions. Once the important variables are identified, a follow-up experiment can be conducted to study their effects more thoroughly.

3. Response Surface Exploration. Once a smaller number of variables is identified as important, their effects on the response need to be explored. The relationship between the response and these variables is sometimes referred to as a response surface. Usually the experiment is based on a design that allows the linear and quadratic effects of the variables and some of the interactions between the variables to be estimated. This experiment tends to be larger (relative to the number of variables under study) than the screening experiment.

4. System Optimization. In many investigations, interest lies in the optimization of the system. For example, the throughput of an assembly plant or the yield of a chemical process is to be maximized. If a response surface has been identified, it can be used for optimization.

5. System Robustness. Besides optimizing the response, it is important in quality improvement to make the system robust against noise (i.e., hard-to-control) variation. This is often achieved by choosing control factor settings at which the system is less sensitive to noise variation. Even though the noise variation is hard to control in normal conditions, it needs to be systematically varied during experimentation. The response in the statistical analysis is often the variance (or its transformation) among the noise replicates for a given control factor setting.

A.2 Terminology

A.2.1 Factors

Factors are processing conditions or input variables whose values or settings can be controlled by the experimenter [10]. Presumably, if one changes the settings of the factors, the value of the response variable varies as well. Factors may be quantitative and qualitative. Quantitative factors like temperature, time, and pressure take values over a continuous range. Qualitative factors take on a discrete number of values. Examples of qualitative factors include material, operation mode, supplier, position, line, etc. If only a linear effect is expected, two levels should suffice. If curvature is expected, then three or more levels are required. Factors and their levels will be denoted by $X_1, X_2, \ldots, X_k$, respectively.
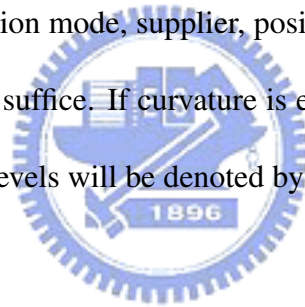
A.2.2 Response

The response variable is the measured quantity whose value is assumed to be affected by changing the levels of the factors [10]. In the semiconductor fabrication process example, the responses were the physical quantities of the devices such as the threshold voltage, subthreshold slope, off-state current, saturation current, drain-induced barrier lowing, and etc. The true value of the response corresponding to any particular combination of the

factor levels and in the absence of experimental error of any kind is denoted by η. (Note that experimental error consists of random measurement error caused by sources such as the production equipment, the testing equipment, and the people who run the equipment, as well as nonrandom error caused by factors that have not been included in the experiment.) However, because experimental error is present in all experiments involving measurements, the response value that is actually observed or measured for any particular combination of the factor levels differs from η. This difference from the true value is written as $Y = \eta + \varepsilon$, where Y represents the observed value of the response and ε denotes experimental error.

A.2.3   The Response Function

When we say that the value of the true response η depends upon the levels X_1, X_2, ..., X_k of k quantitative factors, we are saying that there exists some function of X_1, X_2, ..., X_k, the value of which for any given combination of factor levels supplies the corresponding value of , that is

$$\eta = \phi(X_1, \ldots, X_k). \tag{A.1}$$

The function ϕ is called the true response function and is assumed to be a continuous function of the X_i. The structural form of ϕ is usually unknown and therefore an approximating form is sought using a polynomial or some other type of empirical model equation [26].

A.2.4   The Operability Region and the Experimental Region

The region in the factor space in which the experiments can actually be performed is called the operability region (O) [10]. For some applications the experimenter may wish to explore the whole region O, but this is usually rare. Instead, a particular group of experiments is set up to explore only a limited region of interest, R, which is entirely contained within the operability region O. Although experimental points are not necessarily restricted to the region R and in fact may lie outside it, we shall assume that in most experimental programs, the design points are positioned inside or on the boundary of the region R. Typically, R is defined as, (1) a cuboidal region, or (2) a spherical region.

A.2.5   Coded Variables
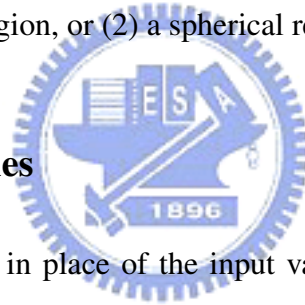
The use of coded variables in place of the input variables facilitates the construction of experimental designs [34]. Coding removes the units of measurement of the input variables and as such distances measured along the axes of the coded variables in k-dimensional space are standardized (or defined in the same metric). A convenient coding formula for defining coded variable, x_i, is

$$x_i = \frac{2X_i - (X_{iL} + X_{iH})}{X_{iH} - X_{iL}}, \quad for \ \ i = 1, 2, 3, \ldots, k, \tag{A.2}$$

where X_{iL} and X_{iH} are the low and high levels of X_i, respectively. There are several advantages to using coded variables rather than the original input variables when constructing
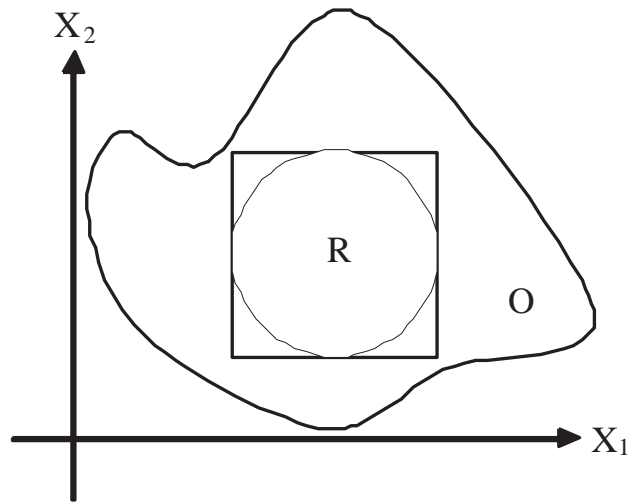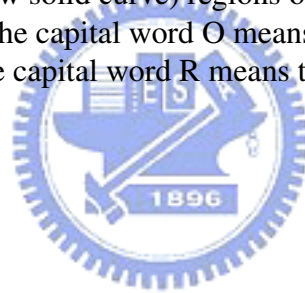
Figure A.1: The operability region and the square (bold solid line) and
           circular (narrow solid curve) regions of interest in two
           dimensions. The capital word O means the operability
           region, and the capital word R means the experimental
           region.

the polynomial models. Two of the more obvious advantages are:

1. computational ease and increased accuracy in estimating the model coefficients; and

2. enhanced interpretability of the coefficient estimates in the model.

Since the coding transformation is a one-to-one transformation, any linear polynomial

equation in the values of xi is expressible as (and equivalent to) a polynomial equation of

the same degree in the values of X_i [34].

A.2.6 Orthogonality

An orthogonal design is one in which the terms in the constructed model are uncorrelated with one another and thus the parameter estimators are uncorrelated [34]. In this case, the variance of the predicted response at any point X in the experimental region, is expressible as a weighted sum of the variances of the parameter estimates in the model. A first-order model is orthogonal if and only if the corresponding X'X matrix is diagonal.

Appendix B

Semiconductor Fabrication Process

Here we describe the general techniques for fabricating the devices and the integrated circuits. There are four main parts in the semiconductor fabrication technology: (1) the crystal growth and epitaxy, (2) the film formation, (3) the lithography and etching, (4) and the impurity doping.
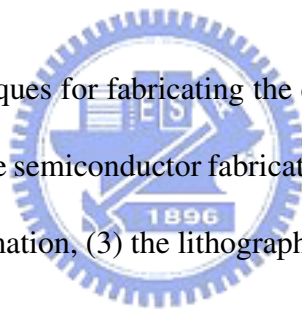
B.1 The Crystal Growth and the Epitaxy

The basic technique for silicon crystal growth from the melt, which is material in liquid form, is the Czochralski technique. The starting material for silicon is a relatively pure form of sand (SiO_2) called quartzite. Through some steps of chemical reactions, the electronic-grade silicon (EGS) are prepared [30].

The Czochralski technique uses an apparatus called a crystal puller. In the crystal-growing process, EGS is placed in the crucible and the furnace is heated above the melting temperature of silicon. A suitably oriented seed crystal is suspended over the crucible in a seed holder. The seed is inserted into the melt. Part of it melts, but the tip of the remaining seed crystal still touches the liquid surface. It is then slowly sithdrawn. Progressive freezing at the solid-liquid interface yields a large, single crystal. A typical pull rate is a few millimeters per minute. In crystal growth, a known amount of dopant is added to the melt to obtain the desired doping concentration in the grown crystal. For silicon, boron and phosphorus are the most common dopants for p- and n-type materials, respectively [30].

After a crystal is grown, the first shaping operation is to remove the seed and the of the end of the ingot, which is last to solidify. The next operation is to grind the surface so that the diameter of the material is defined. After that, one or more flat regions are ground along the length of the ingot. The largest flat, the primary flat, allows a mechanical locator in automatic processing equipment to position the wafer and to orient the devices relative to the crystal. The ingot is ready to be sliced by diamond saw into wafers. Slicing determines four wafer parameters: surface orientation, thickness, taper, and bow.

After slicing, both sides of the wafer are lapped using a mixture of Al_2O_3 and glycerine to produce a typical flatness within 2μm. The lapping operation usually leaves the surface and edges of the wafer damaged and contaminated. The damaged and contaminated regions

can be removed by chemical etching. The final step of wafer shaping is polishing. Its purpose is to provide a smooth, specular surface where device features can be defined by lithographic processed.

A technology closely related to crystal growth involves the growth of single-crystal semiconductor layers on a single-crystal semiconductor substrate. This is called epitaxy. In an epitaxial process, the substrate wafer acts as the seed crystal. The common techniques for epitaxial growth are chemical-vapor deposition (CVD) and molecular-beam epitaxy (MBE) [30].

CVD is a process whereby an epitaxial layer is formed by a chemical reaction between gaseous compounds. The mechanism of CVD involves a number of steps: (a) the reactants such as the gases and dopants are transported to the substrate region. (b) they are transferred to the substrate surface where they are adsorbed, (c) a chemical reaction occurs, catalyzed at the surface, followed by growth of the epitaxial layer, (d) the gaseous products are desorbed into the main gas stream, and (e) the reaction products are transported out of the reaction chamber. There are several material used CVD: CVD for silicon, CVD for GaAs, and metalorganic CVD.
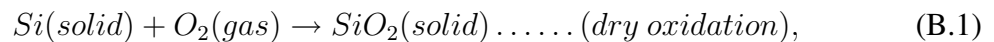
MBE is an epitaxial process involving the reaction of one or more thermal beams of atoms or molecules with a crystalline surface under ultrahigh-vacuum conditions. MBE

can achieve precise control in both chemical compositions and doping profiles. Single-crystal multilayer structures with dimensions on the order of atomic layers can be made using MBE. Thus, the MBE method enables the precise fabrication of semiconductor heterostructures having thin layers from a fraction of a micron down to a monolayer. In general, MBE growth rates are quite low [30].
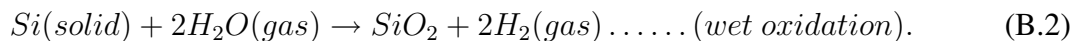
B.2   The Film Formation

We can classify thin films into four groups: the thermal oxides, the dielectric layers, the polycrystalline silicon, and the metal films.

For thermal oxides, the following chemical reactions describe the thermal oxidation of silicon in oxygen or water vapor [30]:

$$Si(solid) + O_2(gas) \rightarrow SiO_2(solid) \ldots \ldots (dry\ oxidation), \qquad (B.1)$$

and

$$Si(solid) + 2H_2O(gas) \rightarrow SiO_2 + 2H_2(gas) \ldots \ldots (wet\ oxidation). \qquad (B.2)$$

The silicon oxide interface moves into the silicon during the oxidation process. This creates a fresh interface region, with surface contamination on the original silicon ending up on the oxide surface.

Deposited dielectric films are used mainly for insulation and passivation of discrete devices and integrated circuits. There are three commonly used deposition methods: atmospheric-pressure CVD, low-pressure CVD (LPCVD), and plasma-enhanced CVD (PECVD, or plasma deposition). Two common dielectrics are silicon dioxide and silicon nitride.

Using polysilicon as the gate electrode in MOS devices is a significant development in MOS technology. One important reason is that polysilicon surpasses aluminum for electrode reliability. A low-pressure reactor operated between 600 °C and 650 °C is used to deposit polysilicon by pyrolyzing silane. Process parameters that affect the polysilicon structure are deposition temperature, dopants, and the heat cycle applied following the deposition step [28].
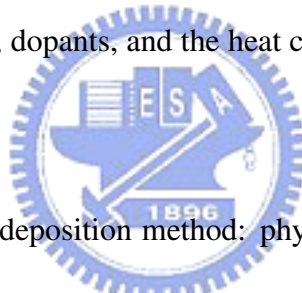
For metallization, there are two deposition method: physical-vapor deposition (PVD) and CVD. The most common methods of PVD of metals are evaporation, e-beam evaporation, plasma spray deposition, and sputtering. Metals and metal compounds such as Ti, Al, Cu, TiN, and TaN can be deposited by PVD. Evaporation occurs when a source material is heated above its melting point in an evacuated chamber. The evaporated atoms then travel at high velocity in straight-line trajectories. In ion-beam sputtering, a source of ions is accelerated toward the target and impinged on its surface. CVD is attractive for metallization because it offers coatings that are conformal, has good step coverage, and can coat a large number of wafers at a time. The basic CVD setup is the same as that used for deposition of

dielectrics and polysilicon. TiN and Al also can be deposited by CVD [30].

B.3 The Lithography and Etching

Photolithography is used to define the geometry of the p-n junction. After the formation of SiO_2, the wafer is coated with an ultraviolet (UV)-light-sensitive material called a photoresist, which is spun on the wafer surface by a high-speed spinner. After spinner, the wafer is backed at about 80 °C - 100 °C to drive the solvent out of the resist and to harden the resist for improved adhesion. The next step is to expose the wafer through a patterned mask using an UV-light source. The exposed region of the photoresist-coated wafer undergoes a chemical reaction depending on the type of resist. The area exposed to light become polymerized and difficult to remove in an etchant. The polymerized region remains when the wafer is placed in a developer, whereas the unexposed region (under the opaque area) dissolves and washes away. After the development, the wafer is again backed to 120 °C -180 °C for 20 minutes to enhance the adhesion and improve the resistance to the subsequent etching process. Then, an etch using buffered hydrofluoric acid (HF) removes the unprotected SiO_2 surface. Last, the resist is stripped away by a chemical solution or an oxygen plasma system. The wafer is now ready for forming the p-n junction by a diffusion or ion-implantation process [30].

B.4   The Impurity Doping

Impurity doping is the introduction of controlled amounts of impurity dopants into semi-conductors. The practical use of impurity doping has been mainly to change the electrical properties of the semiconductors [28]. Diffusion and ion implantation are the two key methods of impurity doping. Until the early 1970s, impurity doping was done mainly by diffusion at elevated temperatures. In this method the dopant atoms are placed on or near the surface of the wafer by deposition from the gas phase of the dopant or by using doped-oxide sources. The doping concentration decreases monotonically from the surface, and the profile of the dopant distribution is determined mainly by the temperature and diffusion time [28].

Since the early 1970s, many doping operations have been performed by ion implantation. In this process the dopant ions are implanted into the semiconductor by means of an ion beam, the doping concentration has a peak distribution inside the semiconductor and the profile of the dopant distribution is determined mainly by the ion mass and the implanted-ion energy [30]. Both diffusion and ion implantation are used for fabricating discrete devices and integrated circuits because these processes generally complement each other. For example, diffusion is used to form a deep junction whereas ion implantation is used to form a shallow junction (e.g., a source/drain junction of a MOSFET).

Diffusion of impurities is typically done by placing semiconductor wafers in a carefully

controlled high-temperature quartz-tube furnace and passing a gas mixture that contains the desired dopant through it. The temperature usually ranges between 800 °C and 1200 °C for silicon. The number of dopant atoms that diffuse into the semiconductor is related to the partial pressure of the dopant impurity in the gas mixture. For diffusion in silicon, boron is the most popular dopant for introducing a p-type impurity [30].

Ion implantation is the introduction of energetic, charged particles into a substrate such as silicon. Implantation energies are between 1 keV and 1 MeV, resulting in ion distributions with average depths raging from 10 nm to 10μm. Ion doses vary from 10^{12} ions/cm^2 for threshold voltage adjustment to 10^{18} ions/cm^2 for the formation of buried insulating layer. Note that the dose is expressed as the number of ions implanted into 1 cm^2 of the semiconductor surface area. The main advantages of ion implantation are its more precise control and reproducibility of impurity dopings and its lower processing temperature compared with those of the diffusion process [28].

Because of the damaged region and the disorder cluster that result from ion implantation, semiconductor parameters such as mobility and lifetime are severely degraded. In addition, most of the ions as implanted are not located in substitutional sites. To activate the implanted ion and to restore mobility and other material parameters, we must anneal the semiconductor at an appropriate combination of time and temperature. In conventional

annealing, we use an open-tube batch-furnace system similar to that for thermal oxida-tion. This process requires long time and high temperature to remove the implant damages. However, conventional annealing may cause substantial dopant diffusion and cannot meet the requirement for shallow junctions and narrow doping profiles. Rapid thermal annealing (RTA) is an annealing process that employs a variety of energy sources with a wide range of times, from 100 seconds down to nanoseconds-all short compared with the conventional annealing. RTA can activate the dopant fully with minimal redistribution [30].

Appendix C

Comparison among CCC, CCF, and CCI

In this appendix, the residual normal plots, the R^2, and the MSE are adopted to compare three specific cases of the central composite design. Then we obtain conclusions to choose the CCF design, and build the corresponding response surface model.

In this example we take a 0.25μm FD SOI CMOS process to present this topic. Table C.1 is the response surface design space for the 0.25μm FD SOI NMOSFET process for the CCC design, the table C.2 is for the CCF design, and the table C.3 is for the CCI design. From these tables we know that the center points of these process conditions are all the same, the differences are the choice of factorial levels and the axial levels. The largest design space is used in the CCC design, and the next is the CCF design, the smallest one is the CCI design.

Table C.1: The CCC design space.

Central Composite Circumscribed	Center	Factorial	Axial
CCC design	0	± 1	$\pm\alpha = \pm 2.828$
Threshold voltage implant dose (cm^{-2})	4.95E+12	$\pm$ 1.05E+11	$\pm$ 2.969E+11
Punch-through implant energy (keV)	30	$\pm$ 3.5	$\pm$ 9.898
Punch-through implant dose (cm^{-2})	1E+13	$\pm$ 1.18+E12	$\pm$ 3.337E+12
Source/Drain implant energy (keV)	50	$\pm$ 6	$\pm$ 16.968
Oxide growth temperature (°C)	850	$\pm$ 12	$\pm$ 33.936
Oxide growth time (min)	7	$\pm$ 1	$\pm$ 2.828
Gate length (μm)	0.25	$\pm$ 0.025	$\pm$ 0.0707

Table C.2: The CCF design space.

Face Centered Cube	Center	Factorial	Axial
CCF design	0	± 1	$\pm\alpha = \pm 1$
Threshold voltage implant dose (cm^{-2})	4.95E+12	$\pm$ 1.05E+11	$\pm$ 1.05E+11
Punch-through implant energy (keV)	30	$\pm$ 3.5	$\pm$ 3.5
Punch-through implant dose (cm^{-2})	1E+13	$\pm$ 1.18+E12	$\pm$ 1.18+E12
Source/Drain implant energy (keV)	50	$\pm$ 6	$\pm$ 6
Oxide growth temperature (°C)	850	$\pm$ 12	$\pm$ 12
Oxide growth time (min)	7	$\pm$ 1	$\pm$ 1
Gate length (μm)	0.25	$\pm$ 0.025	$\pm$ 0.025

C.1   Residual Normal Plots, R^2, and MSE

When the response surface models are built with these three designs, we can use residual

normal plot to check whether the residual are from the normal distribution first. Obviously,

Table C.3: The CCI design space.

Central Composite Inscribed	**Center**	**Factorial**	**Axial**
CCI design	0	$\pm 1/2.828$	$\pm\alpha = \pm 1$
Threshold voltage implant dose (cm^{-2})	4.95E+12	$\pm$ 3.713E+10	$\pm$ 1.05E+11
Punch-through implant energy (keV)	30	$\pm$ 1.2376	$\pm$ 3.5
Punch-through implant dose (cm^{-2})	1E+13	$\pm$ 4.173+E11	$\pm$ 1.18+E12
Source/Drain implant energy (keV)	50	$\pm$ 2.1216	$\pm$ 6
Oxide growth temperature (°C)	850	$\pm$ 4.2433	$\pm$ 12
Oxide growth time (min)	7	$\pm$ 0.3536	$\pm$ 1
Gate length (μm)	$\pm$ 0.25	0.0088	$\pm$ 0.025

there are all 2 outliers in 5 results of the response surface models using CCC design. The pattern of the residuals using CCC design is not quite similar to the normal distribution. We find these 2 outlier points appeared when the level of the gate length are equal to $\alpha = 2.828$, and other factors are set in the center points. So, this design space may not for this case. Probably, a smaller design space will improve this situation. We note that the R^2 of 5 responses are not very high (not larger than 0.9); therefore, these 5 models didn't explain the relationship between the factors and response very much. For these reasons, even the CCC design is the traditional design in building the response surface models, we still want to try alternative one which is expected to be better than the results of the CCC design.

Figure C.2 shows the results of 5 response surface models using the CCF design. The residuals of these responses are similar to the normal distribution, R^2 of these models are all higher than those of CCC and CCI designs, and MSE of these models are almost smaller

Table C.4: The estimated R^2 and MSE of the responses in the CCC
design.

CCC	Vth (V)	SS (V/dec.)	ln(Ioff) (A/μm)	Idsat (A/μm)	DIBL (V)
R^2	0.8039	0.6893	0.8252	0.9687	0.4417
σ^2	4.278E-004	2.041E-005	0.70	8.405E-011	9.283E-004

than other designs. The higher R^2 is, the more variation the model can explain. According
to the chapter of methodology, we know that MSE is the estimator of the variance of the
random error. The smaller the MSE is, the smaller variance of the random error is. Then the
noise fluctuation we cannot control is smaller too. Thus, the CCF design is more suitable
for building the response surface models in this case.

In the CCI design, although several residuals are nearly the normal distribution, and the
R^2 are higher than those in CCC design and close to the CCF design, the CCF design is
still a good choice. It's because the design space of the factorial portion in CCI design is
too small, we thought it may be inappropriate for the truly process situation.

In conclusion, we choose the CCF design to build the response surface models eventu-
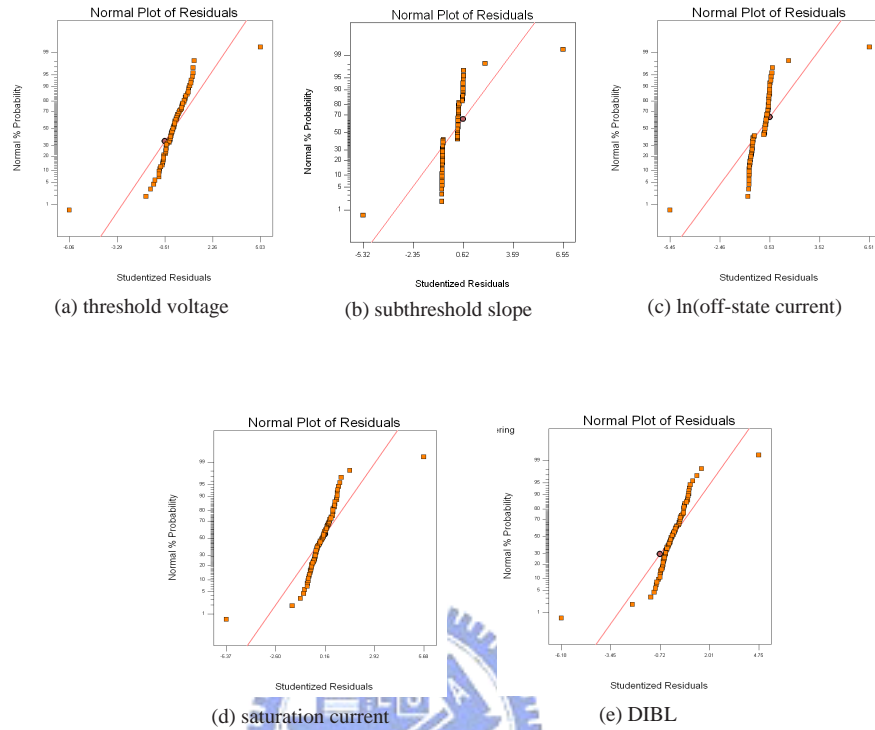ally.

(a) threshold voltage (b) subthreshold slope (c) ln(off-state current)

(d) saturation current (e) DIBL

Figure C.1: Residual normal plots of the CCC design.

Table C.5: The estimated R^2 and MSE of the responses in the CCF design.

CCF	Vth (V)	SS (V/dec.)	ln(Ioff) (A/μm)	Idsat (A/μm)	DIBL (V)
R^2	0.9221	0.9978	0.9930	0.9922	0.8222
σ^2	6.282E-005	9.822E-009	7.590E-003	1.368E-011	1.183E-004

(a) threshold voltage   (b) subthreshold slope   (c) ln(off-state current)
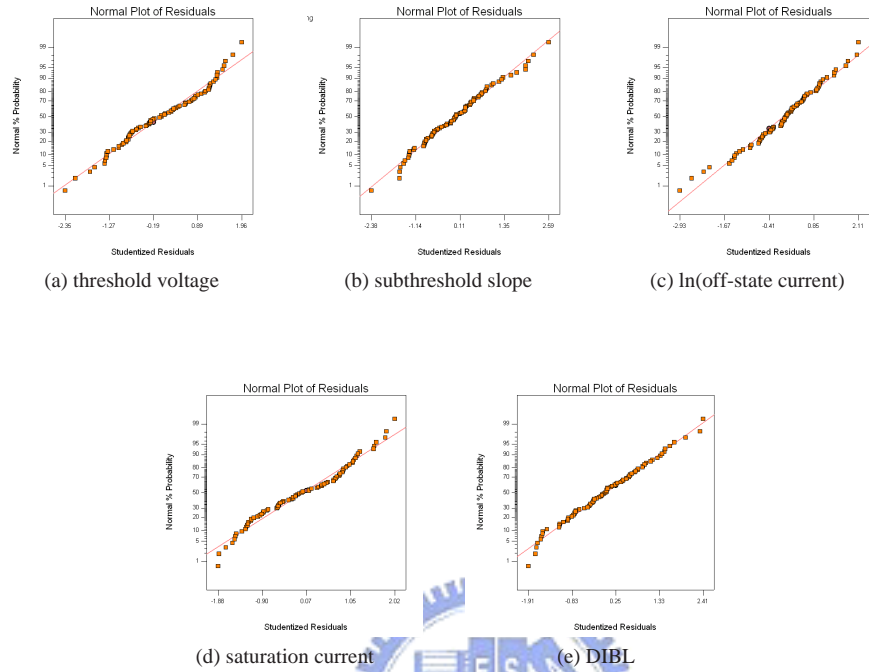
(d) saturation current   (e) DIBL

Figure C.2: Residual normal plots of the CCF design.

Table C.6: The estimated R^2 and MSE of the responses in the CCI design.

CCI	**Vth** (V)	**SS** (V/dec.)	**ln(Ioff)** (A/μm)	**Idsat** (A/μm)	**DIBL** (V)
R^2	0.7184	0.9593	0.9673	0.9251	0.6888
σ^2	5.531E-005	1.811E-008	6.854E-003	4.587E-011	1.229E-004

(a) Threshold voltage   (b) Subthreshold slope   (c) ln(off-state current)
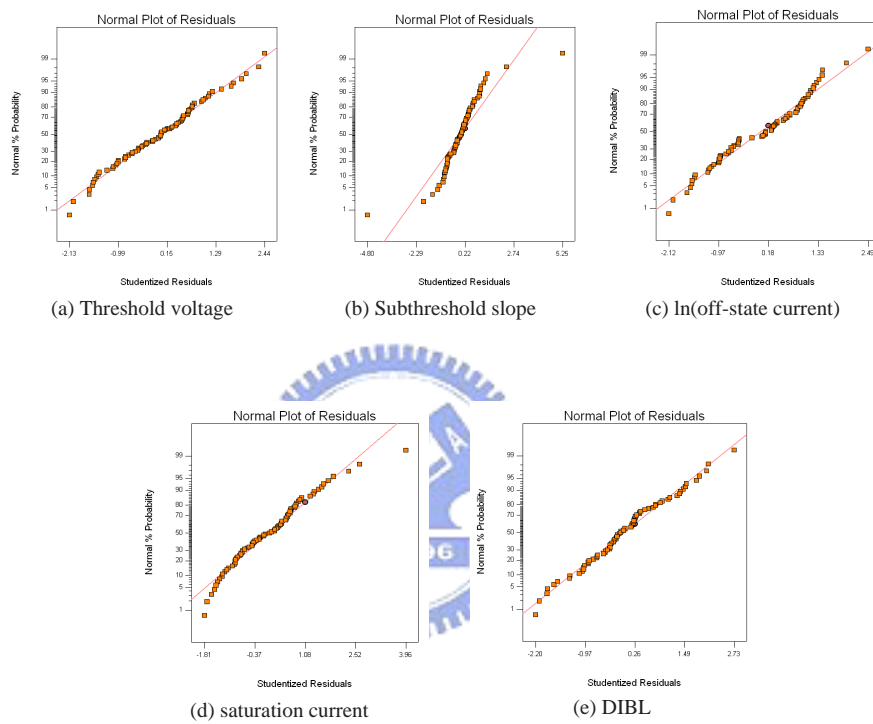
(d) saturation current   (e) DIBL

Figure C.3: Residual normal plots of the CCI design.