Development of High-Speed Silicon Devices and Their Design with Advanced Physical Models

Hiroshi Goto

January 1999

Abstract

In the field of high-speed silicon devices, silicon *b*ipolar *j*unction *t*ransistors (BJTs) had played a major role from the 1970s to the end of the 1980s. However, in the 1990s *c*omplementary *m*etal-*o*xide-*s*emiconductor (CMOS) *f*ield *e*ffect *t*ransistors (FETs) have been replacing their position. This dissertation explains the reasons why BJTs were suitable for high-speed operation. This is concluded from the development of technologies for BJTs and the analyses of devices fabricated with these technologies. At the same time it clarifies why they were replaced by CMOS transistors. The BJT's high driving capability and large power dissipation were the both sides of a sword.

In the case of high-speed CMOS devices, the driving current of MOSFET should be large enough, and device design must be based on precise comprehension of carrier transport in MOSFETs. Therefore, we need accurate device model as well as rigid device-structure information obtained by experiments. This dissertation describes the device design methodology not only based on inverse modeling to extract device structures consistent with all kinds of experimental results but also based on simulations by generalized hydrodynamic model and full-band Monte Carlo model. The background and concept of the methodology is also discussed, and its necessity in future development is clarified. Moreover, hot carrier modeling is discussed by employing full-band Monte Carlo device simulation. Also, this dissertation clarifies the fact there is no experimental evidence for the difference between the surface and bulk impact ionization mechanism in silicon. The reported difference in the literature was only caused by an unsound application of the local field model and was just an artifact.

Finally, by using these sophisticated models, the saturation drain current as well as hot carrier effects of subquarter micron MOSFETs are analyzed. MOSFET design strategy for the 0.1 μ m regime is discussed and the importance of shallow junction for source/drain extension is also clarified.

Acknowledgments

I greatly acknowledge Professor Dr. Mitiko Miura-Mattausch for supporting the preparation of this doctoral dissertation that is submitted to Hiroshima University in Japan. I express my best thanks to my coworkers at Fujitsu Limited, especially Dr. Christoph Jungemann, who was with Fujitsu Kawasaki in Japan from April, 1995 to March, 1997 as a post-doctoral visiting researcher and mainly contributed to the development of the full-band Monte Carlo device simulator FALCON, and Mr. Seiichiro Yamaguchi for the major contribution to the development of the inverse modeling assisting program MOSFIT. I also acknowledge my current and former colleagues, Mr. Kaina Suzuki for the development of the two-dimensional process simulator FIPS2, Mr. Yoji Nagase for the development of the ac analysis routine for the two-dimensional device simulator BIP2DAN, and Mr. Kenji Hashimoto for the development of the transient analysis routine of the same program. In addition I owned quite much to the people who worked together since my joining Fujitsu Limited in Kawasaki, to develop not only bipolar device technologies such as IOP-II, U-FOX, ESPER, but also quarter and subquarter micron CMOS technologies. Finally, I specially acknowledge my wife for her daily support to my work these ten years.

To my wife and children

Table of Contents

Abstract.....ii

Acknowledgments.....iii

Chapter 1 Introduction.....1
1.1 Motivation.....1
1.2 Scope and Organization.....1

Chapter 2 Development of High-Speed Silicon Bipolar Devices.....4
2.1 Historical Back Ground.....4
2.2 Trench Isolation Techniques.....5
2.2.1 Introduction.....5
2.2.2 Trench Isolation by Oxide and Polysilicon.....7
2.2.3 Trench Isolation with Thick Field Oxide.....12
2.2.4 Benefits and Device Performance.....15
2.2.5 Limiting Aspects and Recent Improvements.....18
2.2.6 Conclusion.....19
2.3 Polysilicon Self-Aligned Structures.....19
2.3.1 Introduction.....19
2.3.2 Double Polysilicon Self-Aligned Structure.....19
2.3.2.1 One-Mask Process.....20
2.3.2.2 Two-Mask Process.....25
2.3.2.3 Three-Mask Process.....29
2.3.3 Single Polysilicon Self-Aligned Structure.....29
2.3.4 Sidewall Base Contact Structures.....30
2.3.5 Polysilicon Emitter and Emitter-Base Heterojunction.....35
2.3.6 Conclusion.....37
2.4 Physics-Based Bipolar Transistor Design.....37
2.4.1 Introduction.....37
2.4.2 Two-Dimensional Process/Device Simulation System.....38
2.4.3 Simulation of Pedestal-Collector Transistors.....39

2.4.3.1 Process Simulation....39
2.4.3.2 Device Simulation....41
2.4.4 Favorable Process Condition for Pedestal Collector....44
2.4.5 Application to ECL Circuits....46
2.4.6 Conclusion....46

Chapter 3 Performance Estimation of High-Speed Bipolar Device....47
3.1 Introduction....47
3.2 Strategy to Estimate ECL-Circuit Performance....47
3.3 Optimization of ECL-Circuit Performance....54
3.4 Favorable Silicon BJT Features and Its Limitation....56
3.5 Conclusion....58

Chapter 4 Design Concept for High-Speed CMOS Device....60
4.1 Introduction....60
4.2 Requirements for High-Speed CMOS Device....61
4.3 Quarter and Subquarter Micron CMOS Technologies....63
4.3.1 Quarter Micron CMOS Technology....63
4.3.2 Subquarter Micron CMOS Technology....64
4.4 Conclusion....64

Chapter 5 Inverse Modeling Technique for MOSFET Design....65
5.1 Introduction....65
5.2 Inverse Modeling Procedure....66
5.3 Extraction of a Deep Submicron MOSFET Structure....68
5.4 Saturation Drain Current and Substrate Current....77
5.5 Device Characteristic Prediction....79
5.6 Conclusion....80

Chapter 6 Carrier Transport in MOSFETs....81
6.1 Introduction....81
6.2 Modeling of Carrier Transport in MOSFETs....82
6.2.1 Particle-Based Model....82
6.2.1.1 Full-Band Monte Carlo Device Simulator - FALCON....83

6.2.1.2 Simulation Results by FALCON....85
6.2.2 Fluid-Based Model....87
6.2.2.1 Hydrodynamic Model versus Energy Transport Model....87
6.2.2.2 Generalized Hydrodynamic Model....88
6.2.2.3 Guarantee of Simulation Accuracy by GHDM....89
6.3 Impact Ionization in MOSFETs....90
6.3.1 Impact Ionization Models....91
6.3.2 Surface Impact Ionization....92
6.4 Conclusion....98

Chapter 7 Physics-Based MOSFET Design....99
7.1 Introduction....99
7.2 Hot Carrier Effects of nMOSFET....100
7.3 Physics-Based MOSFET Design....104
7.4 Conclusion....106

Chapter 8 Conclusion....107
8.1 Summary....107
8.2 Future Work....108

Bibliography....110

List of Papers....122

List of Tables

Table 2.1 Typical address access time of linear shrunk 1kb ECL RAMs using IOP-II.....17
Table 2.2 Calculated device parameters and BV_{CEO}.....44
Table 5.1 Extracted parameters for each device and measurement data.....69
Table 5.2 Extracted doping and structure-related parameters for a deep submicron nMOSFET (*Lpoly*=0.19 μ m). Gate oxide thickness is 6.3 nm and contact resistance is 160 Ω μ m.....76

List of Figures

Fig. 2.1 Fabrication process steps of a typical trench isolation technique.....8
Fig. 2.2 Unfavorable trench shapes caused by trench etching.
(a) Black silicon (b) Sub-trench (c) Bowing (d) Laterally etched undercut.....9
Fig. 2.3 Typical trench-isolated structures.
(a) DGI [4] (b) IOP-II [5] (c) U-Iso [8].....11
Fig. 2.4 Fabrication process steps of U-FOX.....13
Fig. 2.5 Fabrication process steps of [10].....14
Fig. 2.6 Modified trench-isolated structures.
(a) [12,13] (b) [14] (c) [9] (d) [10] (e) [15,16].....16
Fig. 2.7 SST-1A fabrication steps [31].....21
Fig. 2.8 Cross-sectional view of the SST-1B transistor with selectively implanted pedestal collector [32].....22
Fig. 2.9 Fabrication steps of the new polysilicon self-aligned process [33].....24
Fig. 2.10 Fabrication steps of the self-aligned transistor process [36].....26
Fig. 2.11 Fabrication steps of the double polysilicon self-aligned process using coupling base implantation [52].....28
Fig. 2.12 Fabrication steps of the single polysilicon self-aligned process [66].....31
Fig. 2.13 Fabrication steps of the SICOS process [67].....33
Fig. 2.14 Fabrication steps of the SICOS process modified by the TOSS process [69].....34
Fig. 2.15 Fabrication steps of the self-aligned selective poly-epi growth process [70].....36
Fig. 2.16 One-dimensional doping profiles of analyzed transistors at the center of emitter.....40
Fig. 2.17 Two-dimensional doping distributions with concentration equi-contour lines.....40
Fig. 2.18 Schematic cross section of an analyzed transistor. The device length is 1.0 μ m..41
Fig. 2.19 Cutoff frequency (f_T) versus collector current (I_c) characteristics.....42
Fig. 2.20 AC base resistance ($r_{bb'}$) and base-collector capacitance (C_{cb}) versus collector current (I_c) characteristics.....42
Fig. 2.21 Current gain (h_{FE}) versus collector current (I_c) characteristic.....43
Fig. 2.22 Collector-to-emitter current (I_c) versus collector-to-emitter voltage (V_{CE}) where the base currents terminate.....43
Fig. 2.23 Relationship between BV_{CEO} and f_T for different phosphorus-ion-implanting conditions. Dose conditions; 150KeV with $2\times10^{12}/cm^2$, $1\times10^{12}/cm^2$; 300KeV

with $1\times10^{13}/cm^2$, $2\times10^{12}/cm^2$; 400KeV with $1\times10^{13}/cm^2$, $2\times10^{12}/cm^2$.....45

Fig. 3.1 Gate delay time versus switching current per gate (arbitrary unit).....48
Fig. 3.2 Schematic cross-section of the analyzed transistor.....49
Fig. 3.3 Impurity profiles for the analyzed transistor.
(a) Selectively implanted pedestal collector profile.
(b) Gradient collector profile.....51
Fig. 3.4 Cutoff frequency and current gain versus collector current (pedestal collector)....52
Fig. 3.5 Cutoff frequency and current gain versus collector current (gradient collector)....52
Fig. 3.6 Analyzed circuit diagram.....53
Fig. 3.7 Gate delay time versus switching current per gate (pedestal collector).....55
Fig. 3.8 Gate delay time versus switching current per gate (gradient collector).....55
Fig. 3.9 Schematic cross-section of the favorable BJT structure.....57

Fig. 4.1 Cross-sectional structure of dual-gate CMOS device with $CoSi_2$ after contact plug formation.....63

Fig. 5.1 Flowchart of the "MOSFIT" program.....66
Fig. 5.2 *C-V* characteristics of a large-area nMOSFET comparing measurement and simulation results after the gate oxide thickness and doping profile of polysilicon gate are determined.
(a) Classical calculation. (b) Quantum mechanical calculation.....70
Fig. 5.3 Subthreshold *Id-Vg* characteristics of a 1.0-μm nMOSFET comparing measurement and simulation results after the channel doping profile is extracted. (Lines: measurement results, symbols: simulation results).....71
Fig. 5.4 *Cgd-V* characteristics of a finger gate nMOSFET comparing measurement and simulation results after the source/drain doping profile is extracted. (Lines: measurement results, symbols: simulation results).....71
Fig. 5.5 Subthreshold *Id-Vg* characteristics of a 0.19-μm nMOSFET comparing measurement and simulation results after the channel doping profile is extracted. (Lines: measurement results, symbols: simulation results).....72
Fig. 5.6 Mid-channel doping profiles depending on the gate length.....73
Fig. 5.7 DIBL effect for a 0.19-μm nMOSFET comparing measurement and simulation results after the channel doping profile and gate length are determined. (Lines:

measurement results, symbols: simulation results).....74
Fig. 5.8 Implementation of the source/drain contact resistance for linear *Id-Vg* calculation.....74
Fig. 5.9 Linear *Id-Vg* characteristics comparing measurement and simulation results including source/drain contact resistance.....75
Fig. 5.10 Final two-dimensional structure and doping profile of a 0.19-μ m nMOSFET. (Unit; μ m for axes and cm^{-3} for contour lines).....75
Fig. 5.11 *Id-Vd* characteristics comparing measurement and simulation results using the GHDM by GALENE-III for a 0.19-μ m nMOSFET. (Lines: measurement results, symbols: simulation results).....78
Fig. 5.12 Substrate current characteristics comparing measurement and simulation results using the full-band MC simulator FALCON for a 0.19-μ m nMOSFET. (Lines: measurement results, symbols: simulation results).....78
Fig. 5.13 *Id-Vd* characteristics comparing measurement and predicted simulation results using the GHDM by GALENE-III for a different 0.19-μ m nMOSFET of which source/drain doping concentration is scaled to one-tenth. (Lines: measurement results, symbols: simulation results).....79
Fig. 5.14 Substrate current characteristics comparing measurement and predicted simulation results using the full-band MC simulator FALCON for a different 0.19-μ m nMOSFET of which source/drain doping concentration is scaled to one-tenth. (Lines: measurement results, symbols: simulation results).....80

Fig. 6.1 Flowchart of the acceptance/rejection method for the selection of the final state after scattering (random number *r* uniformly distributed in [0,1]).....84
Fig. 6.2 Substrate current and CPU times for (a) a 0.18 μ m nMOSFET and (b) a 1.0 μ m nMOSFET for different drain voltages. (Lines: experiment, Symbol: FB-MC simulation).....86
Fig. 6.3 Schematic potential and electric fields for the II coefficient (left), at the Si/SiO_2-interface of a MOSFET (center) and for the quantum yield experiment (right)....91
Fig. 6.4 II coefficient versus inverse electric field for bulk silicon (solid line) as given by van Overstratton, for MOSFETs (dashed line) and CCDs (dotted line) by Slotboom [133] and FB-MC results (symbol).....93
Fig. 6.5 Substrate current versus gate voltage for the shallow source/drain extension 0.16 μ m-nMOSFET (*Vd*=1.5,2.0,2.5,3.0,3.5V). (Lines: measurement, ○: FB-MC with

surface roughness scattering, + FB-MC without surface roughness scattering) 93
Fig. 6.6 Electric field (solid line) and II generation rate for the case B of the CCD experiment [133] calculated with the LF model based on bulk parameters (dotted line), based on surface parameters (dashed line) and FB-MC results (dashed dotted line).....95
Fig. 6.7 Multiplication factor for case *B* of the CCD, measured (solid line) [133], calculated with the LF model based on bulk parameters (dotted line), based on surface parameters (dashed line) and FB-MC results (symbol).....96
Fig. 6.8 Confining electric field averaged with the II generation rate for the 0.16 μ m-nMOSFET.....97

Fig. 7.1 Electron distribution in a 0.16 μ m-nMOSFET at different energy level. (*Vd*=2.5V, *Vg*=1.5V, axes unit: μ m, contour line unit: cm^{-3}, dashed line denotes pn-junction).....101
Fig. 7.2 Hole distribution in a 0.16 μ m-nMOSFET at different energy level. (*Vd*=2.5V, *Vg*=1.5V, axes unit: μ m, contour line unit: cm^{-3}, dashed line denotes pn-junction).....102
Fig. 7.3 II generation rate caused by (a) electron and (b) hole. (*Vd*=2.5V, *Vg*=1.5V, axes unit: μ m, contour line unit: $s^{-1}cm^{-3}$, dashed line denotes pn-junction).....103
Fig. 7.4 Currents of electrons and holes with certain energies hitting the Si/SiO_2 interface for the 0.18 μ m- (dashed line) and 0.16 μ m- (solid line) nMOSFETs.....104
Fig. 7.5 Simulated (a) Ion and (b) ΔVth versus poly-gate length characteristics for four kinds of nMOSFET. (A: surface peak channel with source/drain extension x_j=0.05 μ m, B: surface peak channel with source/drain extension x_j=0.03 μ m, C: steep retrograde channel with source/drain extension x_j=0.05 μ m, D: steep retrograde channel with source/drain extension x_j=0.03 μ m).....105

Chapter 1 Introduction

1.1 Motivation

In the field of high-speed silicon devices, silicon *b*ipolar *j*unction *t*ransistors (BJTs) had played a major role from the 1970s to the end of the 1980s. However, in the 1990s *c*omplementary *m*etal-*o*xide-*s*emiconductor (CMOS) *f*ield *e*ffect *t*ransistors (FETs) have been replacing their position. This dissertation tries to clarify the reason why BJTs were suitable for high-speed operation through the development of silicon bipolar device technologies and physics-based device analyses. This dissertation also reveals the reason why they were replaced by CMOS transistors. Then, to achieve high-speed CMOS transistors, this dissertation makes clear the needs for the device design methodology based on accurate physical device modeling as well as the importance of inverse modeling to determine device structure consistent with all experimental results. Also is made distinct the necessity for both particle-based full-band Monte Carlo and fluid-based generalized hydrodynamic models for accurate device simulation. Finally, the MOSFET design strategy for deep submicron regime will be discussed.

1.2 Scope and Organization

Chapter 2 describes the development of high-speed silicon bipolar technology and physics-based BJT design. In the case of BJTs, it was easy to reduce base width by employing a successive p-type and n-type diffusion technique that can result in the base thickness of 50nm or so. By introducing various kinds of self-alignment techniques such as deep trench isolation and double polysilicon self-aligned structures, BJTs enjoyed the benefit of short transit time i.e. high cutoff frequency caused by the thin base without needing sophisticated lithography tools. As a result, they achieved more competitive position to compound semiconductor devices and conquered the major position in the field of high-speed *l*arge *s*cale *i*ntegrations (LSIs).

Chapter 3 discusses circuit performance for high-speed BJT and clarifies its advantages and limiting factors. The most effective circuit to explore the high-speed performance of BJT is *e*mitter-*c*oupled *l*ogic (ECL). In the case of

ECL, emitter-coupled pair transistors are switched in accordance with input signal applied to base without bringing them into the saturation operation mode. This means each transistor benefits from quick discharging and switching. By using drift-diffusion-based device/circuit mixed simulator, the ultimate performance of an ECL-like circuit was investigated for double polysilicon self-aligned BJTs. As a result, it was found that the circuit delay was no more limited by cutoff frequency or the forward transit time of the BJT itself, but parasitic capacitance associated with a collector pull-up resistor. In other words, collector-substrate capacitance and loaded capacitance including an emitter follower input capacitance limit the circuit delay. Therefore, the reduction of base-collector and/or collector-substrate capacitance is more effective than the increment of cutoff frequency. On the other hand, in the case of ideal condition without any parasitic capacitance with the collector pull-up resistor, the switching current of around 100 μ A for each gate is still needed to achieve minimum gate delay. This indicates that, in the case of one-million-gate circuit, the chip will dissipate 100A in total, far beyond the practical air-cooling limit. Accordingly, BJTs were not able to enjoy the progress of microfabrication technology brought by deep submicron lithography.

Chapter 4 points out the reason why CMOS transistors replaced BJTs even in the high-speed field. CMOS transistor is an ideal device from the viewpoint of power dissipation because it dissipates power only when switching. The major drawback to the CMOS transistor was the larger channel length that was defined by minimum lithographic pattern. In the 1980s the channel length was from one to two orders larger than the base width of BJT, bringing MOSFETs lower cutoff frequency and larger transit time than bipolar devices. However, the dramatic progress of optical lithography has made it possible to realize deep submicron channel length. This results in higher driving current as well as shorter switching time even by CMOS circuits. Microprocessor is the main device benefited by this advantage, and high-speed LSIs have been realized. Fundamental technological features for high performance CMOS devices are retrograde well, dual-gate surface channel MOSFETs and a salicide structure. The former two features can realize high freedom for choosing supply voltage and threshold voltage design independently on the well structure especially in lower supply voltage

regimes. The latter feature reduces parasitic resistance and, if combined with shallow trench isolation, parasitic capacitance associated with source/drain junction can be reduced. From the device-design point of view, optimization for higher driving current with lower off-leakage current as well as hot-carrier-induced degradation immunity is the most important concern.

Chapter 5 describes the requirement of precise knowledge of the device that is made by split lot experiments to design MOSFET effectively. For this purpose, inverse modeling is a powerful technique. The concept of inverse modeling, the procedure and application to quarter micron nMOSFETs are demonstrated. By using inversely modeled information, driving current and hot carrier properties can be predicted and successfully designed.

Chapter 6 discusses what kind of device model should be used to describe carrier transport in high-field regime to analyze driving current or the saturation drain current of MOSFETs. We must use accurate device model for this operation regime. To this end drift-diffusion model is not enough because this model assumes equilibrium carrier transport inside device. Not only fluid-model-based generalized hydrodynamic model but also particle-based first-principle model using ensemble full-band Monte Carlo simulation is inevitable. The background and concept of these approaches are also discussed and clarified. Moreover, hot carrier property and its model hierarchy is discussed based on the full-band Monte Carlo simulation analysis. Also is studied the previously proposed surface impact ionization model different from the bulk, and it is demonstrated that there is no experimental evidence for the difference between the surface and bulk impact ionization in silicon. The difference was only caused by an unsound application of local field model and just was an artifact.

Finally, in Chapter 7, by employing these sophisticated models, the saturation drain current as well as the hot carrier effects of subquarter micron MOSFETs are studied. Then, MOSFET design strategy for the 0.1-μm regime is discussed, and the importance of shallow junction for source/drain extension is demonstrated.

In Chapter 8 as conclusion, the summary and future works for the sub-0.1 μm regime are pointed out.

Chapter 2 Development of High-Speed Silicon Bipolar Devices

2.1 Historical Background

To achieve high-speed operation, silicon bipolar devices have conquered numerous process-technological obstacles. Since the discovery of point-contact transistor by Bardeen and Brattain in 1947 and the invention of bipolar junction transistor by Shockley in 1948, several valuable process innovations have been introduced. The first notable step was the development of integrated circuits and a planar technique by the end of the 1950s. Isolation of pn-junction and new epitaxial growth techniques had triggered a rapid progress of bipolar ICs in the industry. Around 1970, two new process techniques had been developed. One was the LOCOS (*loc*al *o*xidation of *s*ilicon) technique [1]. By employing LOCOS isolation instead of junction isolation, dimensions of isolation regions were drastically reduced and the packing density was much improved. The other was the DOPOS (*do*ped *poly*silicon) technique, which was the first polysilicon emitter process [2]. The DOPOS technique used impurity-doped polysilicon not only as an emitter diffusion source but also as an emitter electrode. The DOPOS made possible the achievement of much narrower and shallower emitters than those realized by a conventional washed emitter technique, which had been the ultimate refinement of the planar technique. In the 1970s, polysilicon techniques were further improved so as to achieve various applications for bipolar devices. A *p*olysilicon *s*elf-*a*ligned (PSA) process technique was one of the most sophisticated example [3]. The PSA technique used a polysilicon base electrode as well as a graft base diffusion source, exactly as the polysilicon emitter. Another notable property of the PSA technique was the application of the LOCOS technique to polysilicon films. By employing the PSA technique, the first fully self-aligned device structure that did not use tight lithographic design rules was demonstrated.

In the early 1980s, new etching techniques were demonstrated; namely the anisotropic dry etching techniques, such as *r*eactive *i*on *e*tching (RIE). By using this technique, first, steep and narrow grooves or trenches were easily realized for much improved isolation. Trench isolation began to replace conventional oxide isolation, resulting in higher packing densities and smaller

collector-substrate capacitance. Second, the so-called sidewall spacer technique was developed, allowing self-aligned compact isolation between stacking conducting layers of materials such as doped polysilicon. By combining the sidewall spacer techniques and the polysilicon base and emitter, double polysilicon self-aligned processes emerged. A submicron distance between the emitter and the base electrode was easily realized to achieve lower base-collector capacitance and lower base resistance than with any previous device structure. Anisotropic dry etching was also used to form steep and narrow monosilicon mesas to act as intrinsic base and emitter regions. By introducing sidewall contacted polysilicon base electrodes to the mesas, a new, sophisticated device structure was proposed. A *si*dewall base electrode *co*ntact *s*tructure (SICOS) was another approach to achieve fully self-aligned process. The SICOS was intended to realize an ideal one-dimensional device structure suitable for upward operation. Other sidewall base contact structures were also demonstrated by selective epitaxial growth as well as simultaneous epitaxial polysilicon growth in steep grooves.

Self-aligned processes as well as polysilicon emitter have been standard techniques for high performance silicon bipolar IC fabrication. Great efforts have been spent to improve these techniques and many researchers have been investigating the underlying device physics.

2.2 Trench Isolation Techniques

2.2.1 Introduction

Isolation is one of the key techniques in integrated circuit processes. When the first integrated circuit was invented by Kilby, two transistors and some resistors made in a monolithic germanium chip were partially separated by a slot and air. However, with the advent of the planar technique, pn-junction isolation became the major technique because Kilby's non-planar structure was unfavorable for wiring process. Another approach was also tried to improve the surface topography of air-isolated devices. This is the so-called dielectric isolation utilizing groove etching. Air isolation slots were replaced by dielectric substrates in which active semiconductor islands were sustained. The dielectric isolation achieved rather planar surfaces, but it was

too complicated for conventional devices. Only for very high voltage devices has this technique been used.

After the LOCOS technique was developed, pn-junction isolation was gradually replaced by oxide isolation. At the same time there was a revival of groove isolation. In this case only laterally isolated regions were formed by grooves filled with dielectric materials, and the substrate was isolated by pn-junction. Grooves were formed by anisotropic chemical etching. Both of these isolation techniques achieved higher packing density than pn-junction-isolated devices. The scaling concept was also employed to achieve large scale integration and high performance devices. When the devices are scaled, not only active regions but also inactive regions such as the isolation should be reduced. In the case of LOCOS isolation, the so-called bird's beak enlarges the inactive regions. The length of the bird's beak is dependent on the thickness of oxide and cannot be reduced proportionally to the scaling ratio. This was the ultimate limitation of oxide isolation. On the other hand, the groove isolation utilized anisotropically etched V-shaped grooves for isolation using alkaline etchant applied to (100) silicon surface. The V-groove's depth is 70% of its width, so the reduction of isolation regions is limited by the groove depth necessary to keep enough breakdown voltage.

In order to overcome these problems of conventional isolation techniques, the bird's beak free and width-independent isolation scheme was most important, the result was the advent of trench isolation. This is a new technology employing novel etching techniques such as RIE to form narrow and steep U-shaped grooves, or trenches for isolation, but its root is based on the classical idea of dielectric isolation.

Though a practical trench isolation technique was first disclosed in 1978[4], we demonstrated its application to practical devices for the first time early in the 1980s[5]. Since the first application of trench isolation to 1kb ECL static RAMs in 1982 [5], many bipolar devices have been fabricated by various kinds of trench isolation techniques. In the case of bipolar devices, some trench isolation techniques have been used in volume production since the early stage of the development, though trench isolation for MOS devices are only introduced recently as a shallow trench isolation technique.

In the following subsections, process technologies, device structures, application to devices and benefits and limiting aspects of trench isolation for

bipolar devices will be discussed.

2.2.2 Trench Isolation by Oxide and Polysilicon

Typical trench isolation techniques consist of three major process steps [5]. These are trench etching, trench filling and planarization. Figure 2.1 shows an example of process steps. As usual, an n^+ buried layer and an n^- epitaxial layer are formed on a p-type substrate. On the n^- epitaxial layer, a nitride layer with thin pad oxide is deposited and a masking layer for trench etching is also deposited. After the isolation pattern is formed, trench etching follows using anisotropic dry etching such as RIE. The trenches penetrate both the epitaxial and buried layers to reach the p-type substrate. The inside of the trenches is then oxidized to accomplish electrical isolation. After that, trenches are buried with undoped polysilicon. Excess polysilicon is polished off until the nitride layer appears. Finally the surface of polysilicon is oxidized for capping. The trenches are filled with oxide and polysilicon.

Trench etching is one of the most important processes in trench isolation techniques. The typical structure of narrow and deep trenches for isolation is achieved by anisotropic plasma etching. RIE is widely used for this purpose. The etching conditions should be carefully optimized to avoid unfavorable trench shapes. Figure 2.2 shows examples of maladapted trench structures. Figure 2.2(a) shows so-called 'black-silicon' or needle-like silicon residues, and sub-trench at the corners of bottom edges of trenches is shown in Fig. 2.2(b). Figure 2.2(c) is the case of laterally-etched undercut that occurs at the highly doped n^+ buried layer. All of these unfavorable shapes cause large leakage current between isolated elements. Figure 2.2(d) shows bowing. In this case the trenches are not completely buried by filler materials and voids occur in isolation regions. These voids may become sharp crevices after the planarization step finishes. They cause bad metallization step coverage.

These trench shapes are strongly dependent on gas species, pressure, flow rates and power in RIE processes [6,7]. These etching conditions are not universal, so they should be optimized according to etching systems.

Trench filling and planarization processes are also important. Trenches should be filled completely with filler materials. If not, nonplanar surfaces occur after excess filler materials are removed. The filler materials should

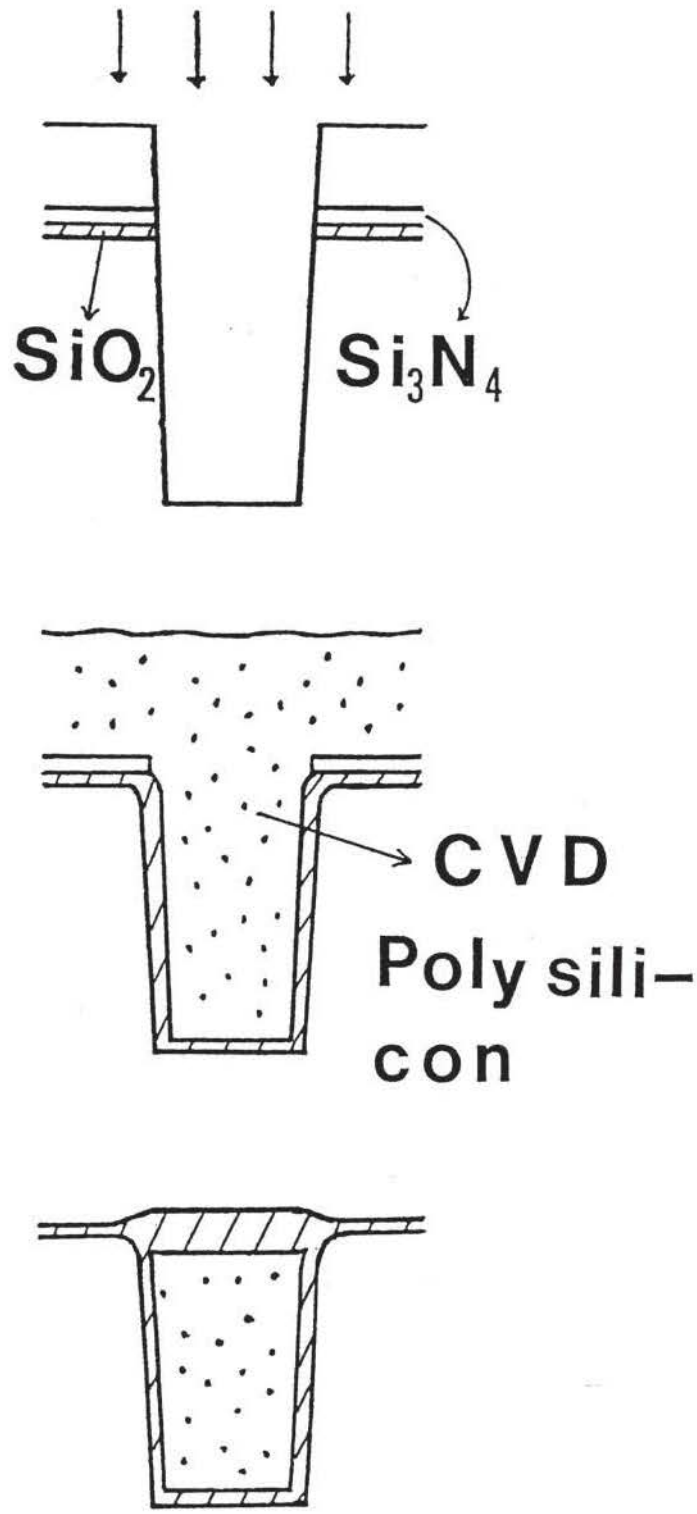


Fig. 2.1 Fabrication process steps of a typical trench isolation technique.

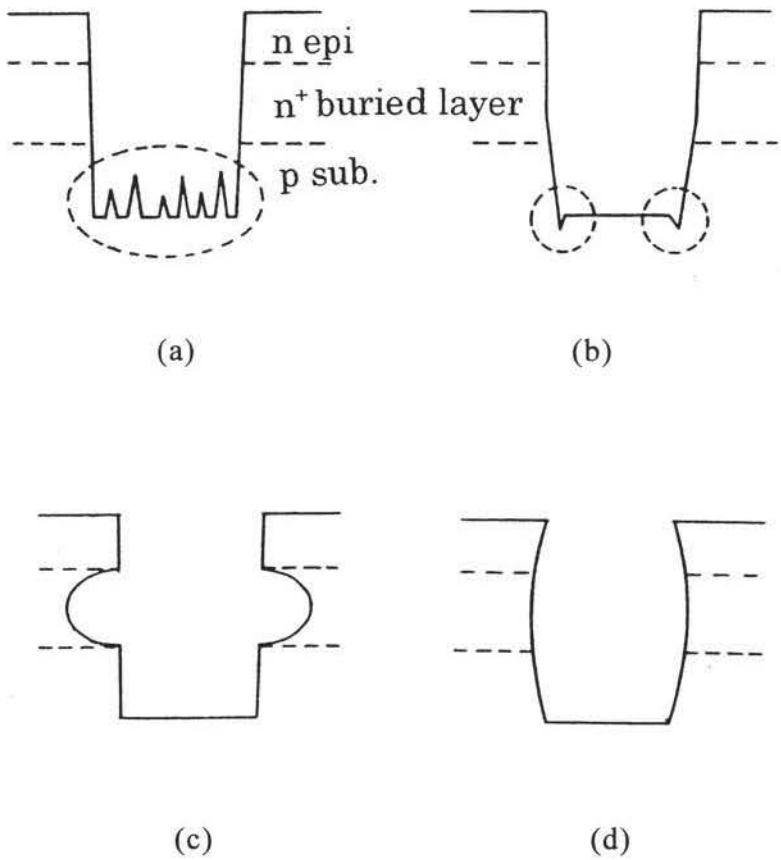


Fig. 2.2 Unfavorable trench shapes caused by trench etching.
(a) Black silicon (b) Sub-trench (c) Laterally etched undercut (d) Bowing

also be deposited as uniformly as possible. There are variety of filler materials such as CVD (*c*hemical *v*apor *d*eposition) oxide, CVD nitride and undoped polysilicon. When CVD oxide is used, trench shapes, especially the sidewall taper should be optimized because CVD oxide does not show good step coverage inside trenches [4]. Until now, instead of CVD oxide, bias sputtered oxide, ECR (*e*lectron *c*yclotron *r*esonance) plasma enhanced CVD oxide and boron-phosphorus silicate glass have been tried. Nevertheless, low-pressure CVD polysilicon shows excellent step coverage although it is not a complete insulator. Undoped polysilicon is widely used as the filler material. CVD nitride is a good insulator, but because of its high tension it is impossible to fill trenches completely. Nitride is often used as a spacer

between thermal oxide of trench surfaces and undoped polysilicon. Excess filler materials are removed to attain a planar surface by controlled etch-back techniques or *c*hemical-*m*echanical *p*olishing (CMP) technique. This planarization process requires high uniformity. In our experience, polishing techniques are superior to controlled etch-back processes, resulting in recent large stream of CMP process as a major planarizing technique.

Figure 2.3 shows the typical trench-isolated structures. Figure 2.3(a) is called DGI (*d*eep *g*roove *i*solation) [4]. Trenches are etched by RIE using Cl_2/Ar gas mixture. Trench surfaces are oxidized and filled with CVD oxide. Excess oxide was removed by a controlled etch-back technique. Figure 2.3(b) shows our IOP-II (*i*solation by *o*xide and *p*olysilicon; second version) [5]. Trenches are etched by RIE and filled with oxide and polysilicon. Excess polysilicon is removed by a chemical-mechanical polishing technique. IOP-II is an advanced version of conventional IOP using V-grooves, and only V-groove etching was replaced by trench etching. Figure 2.3(c) exhibits U-Iso (U-groove *iso*lation) [8]. In this case, the trench is not U-shaped but Y-shaped. The trench etching process consists of two steps. First, anisotropic chemical etching of (100) surface is employed to form an overhung structure using a nitride layer with side-etched pad oxide as a mask. Then anisotropic dry etching such as reactive sputter etching using a CCl_4/O_2 gas mixture follows for trenches to penetrate an n^+ buried layer. The inside of the trenches is oxidized and the second nitride layer is deposited to cover the trench surface. Trenches are filled with undoped polysilicon. Excess polysilicon is removed by a controlled etch-back technique.

The above-mentioned trench-isolated structures may be called the first generation. In these trench isolation techniques, one attempts to reduce only the isolation regions; other inactive regions such as field regions for wiring or shallow isolation for base and collector-reach-through regions are not taken into account. As far as memory devices are concerned, there is no serious problem. However, for high performance logic devices, those parasitic inactive regions are fatal disadvantages for switching speeds. In order to overcome these disadvantages of the first generation, some process modification followed. In the next subsection, process technologies and device structures of advanced trench isolation techniques, which may be called the second generation, will be described.

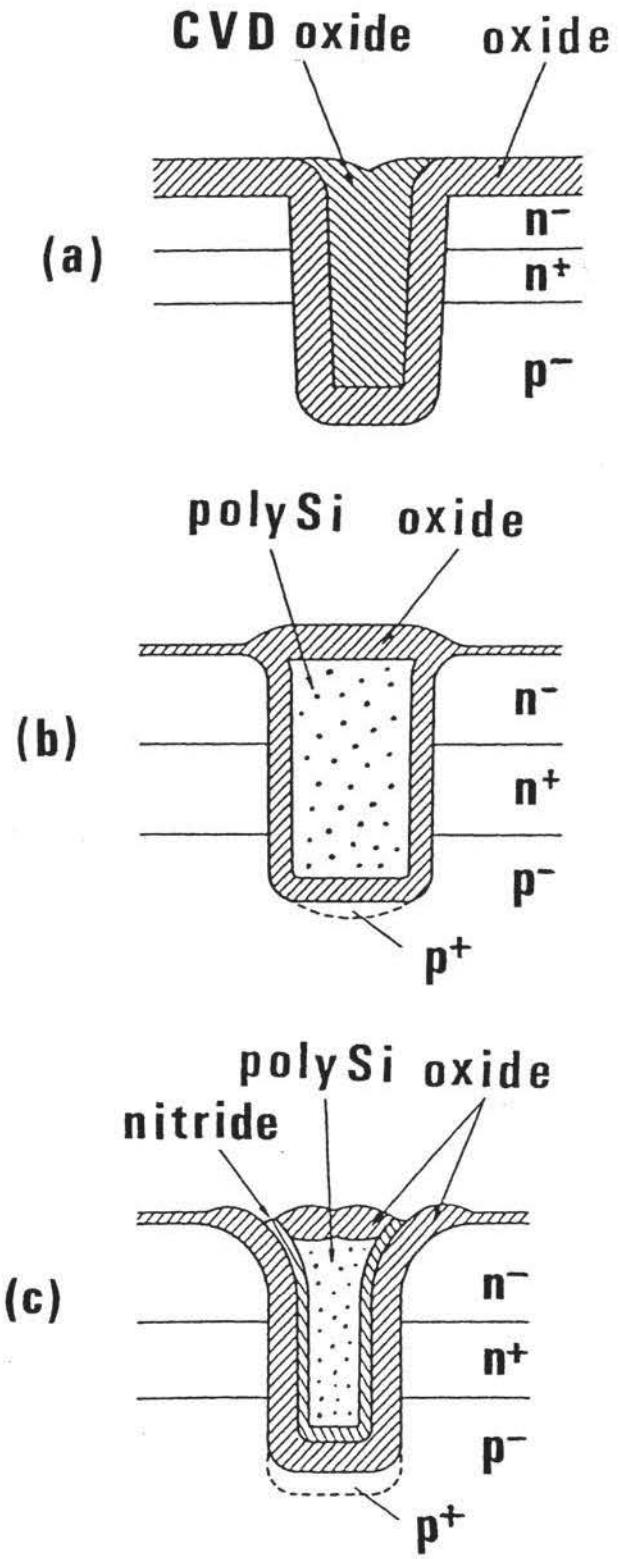


Fig. 2.3 Typical trench-isolated structures.
(a) DGI [4] (b) IOP-II [5] (c) U-Iso [8]

2.2.3 Trench Isolation with Thick Field Oxide

Process modifications for the second generation may be roughly divided into two types. One is the combination of LOCOS or ROX (*r*ecessed *ox*ide) with deep trenches, and the other is the combination of shallow trenches with deep ones. For instance, Figure 2.4 shows the process steps of our U-FOX (*U*-groove isolation with thick *f*ield *ox*ide, renamed from IOP-L) or an advanced version of IOP-II, combining LOCOS with trench isolation [9]. After the formation of an n^- epitaxial layer by conventional means, its surface is selectively oxidized except base, collector-reach-through regions and so on, using the LOCOS technique. Then a new nitride layer and a masking layer for trench etching are deposited. They are removed at the isolation regions and trench etching follows. After the trenches have been etched, their surfaces are oxidized and filled with undoped polysilicon. Excess polysilicon is polished off until the polysilicon not on the isolation regions is completely removed. Then capping oxidation is performed. This process step sequence, or the trench etching after the field LOCOS formation is favorable to the other option, or the LOCOS-after-trench process, avoiding nonplanar surface topography as well as stress induced crystalline defects caused by capping oxidation. The detail will be discussed later again.

To compare our U-FOX process with other approaches, another technique is shown in Fig. 2.5 [10]. In this technique, deep trenches for device isolation are first etched, using a thick oxide layer as a mask. Then the masking layer is selectively removed by the second isolation patterns, which contain base- and collector-reach-through isolation and field regions. Again trench etching is performed to give no penetrating shallow trenches. Deep trenches penetrate the n^+ buried layer to reach also the p-type substrate. Both deep and shallow trenches are buried with CVD oxide. The CVD oxide itself does not achieve good step coverage in the trenches, but by optimizing the distances between deep trenches in shallow ones, voids in the deep trenches are elaborately covered by over-grown oxide. Excess oxide is removed by a controlled etch-back technique in just the same way as reported for simplified *b*uried *ox*ide (BOX) isolation for MOS devices [11].

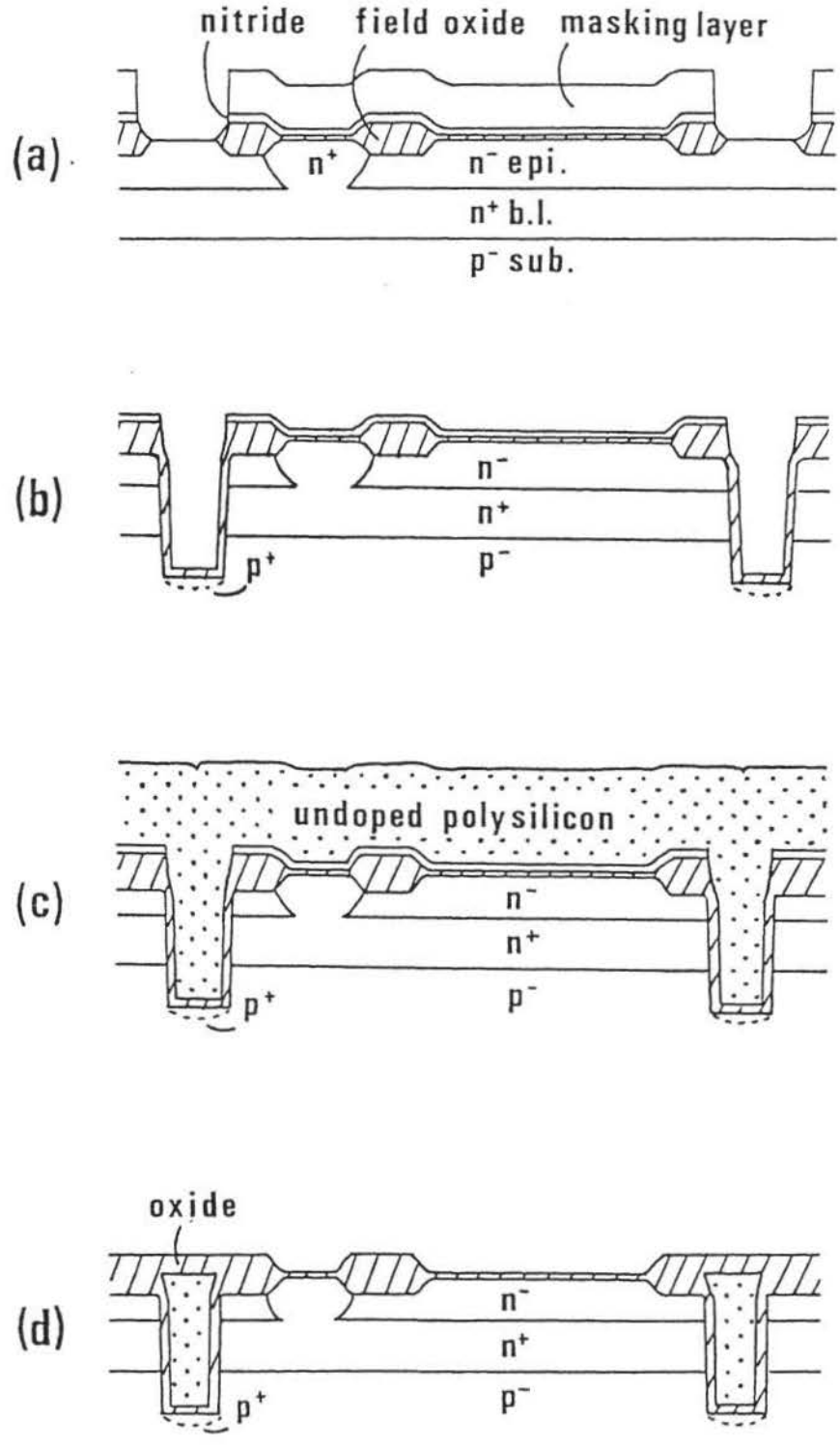


Fig. 2.4 Fabrication process steps of U-FOX.

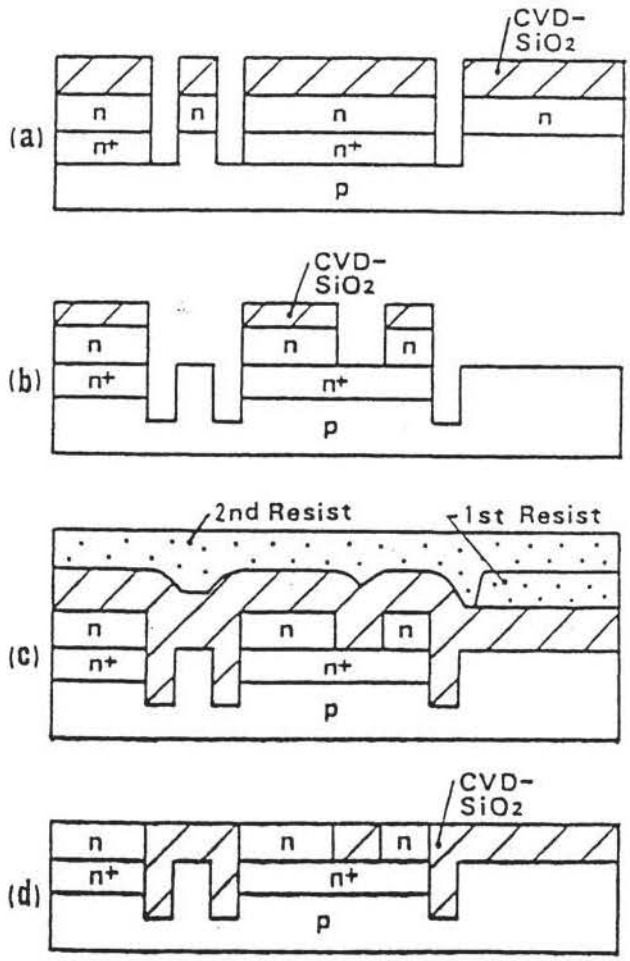


Fig. 2.5 Fabrication process steps of [10].

There are also some structural modifications of these techniques. Figure 2.6 shows structural variations of modified trench-isolated devices in the second generation. Figure 2.6(a) is an example of deep trench isolation combined with recessed oxide. The recessed oxide is formed by recessing silicon, using RIE followed by oxidation. Deep trenches are then etched, covered with a thin thermal oxide layer and filled with CVD oxide. In some cases there are voids in the middle of isolation regions [12,13]. Figure 2.6(b) is an example of a dual-depth trench-isolated structure [14]. Shallower trenches, filled with oxide, nitride, and polysilicon isolate transistors and Schottky barrier diodes, and they may be able to isolate base and collector-reach-through regions. Figure 2.6(c) shows our U-FOX structure. LOCOS technique is used for base and collector-reach-through isolation and field

oxide for wiring regions. Figure 2.6(d) is the final structure of the technique mentioned above [10]. Figure 2.6(e) is a modification of Fig. 2.6(a). As a filler material for deep trenches, CVD oxide is replaced by selective epi-silicon [15,16]. In this case, trenches of variable width can be filled and planarized at the same process step. There are other variations on advanced trench-isolated structures [17,18]. In ref. [17], field oxidation, as well as capping oxidation of trench-filling polysilicon, performed after excess polysilicon is removed by a controlled etch-back technique. In order to bury the crevices which occur along the trench-edge surface, an etch-back technique of reflow glass is used The final device structure resembles Fig. 2.6(d). In ref. [18], vertical pnp transistors are formed as well as conventional npn transistors, and new advantageous device structure is achieved. The main drawbacks to these techniques are rather complicated dual-depth trench etching process, poor step coverage of CVD oxide inside trenches and larger collector-substrate parasitic capacitance caused by selective epi-trench filling process.

2.2.4 Benefits and Device Performance

The trench isolation techniques mentioned above have a number of advantages, and bipolar devices using them have achieved higher performance than conventionally isolated devices

First, higher packing density than in conventional isolation is achieved because the trench depth is almost independent of its width and the bird's beak is reduced much more than in LOCOS. The p^+ channel stop layers are completely separated from n^+ buried layers. Therefore, even if the trench width is reduced, there is enough space that a depletion layer can sustain high breakdown voltage. Also, the complete separation of the channel stop layers and the n^+ buried layers makes collector-substrate parasitic capacitance smaller. Because the n^+ buried layer islands are self-aligned to isolation trenches, there is no mask alignment step between a buried layer and an isolation pattern, and planar collector-substrate pn-junction is achieved, which is also favorable for the reduction of its parasitic capacitance.

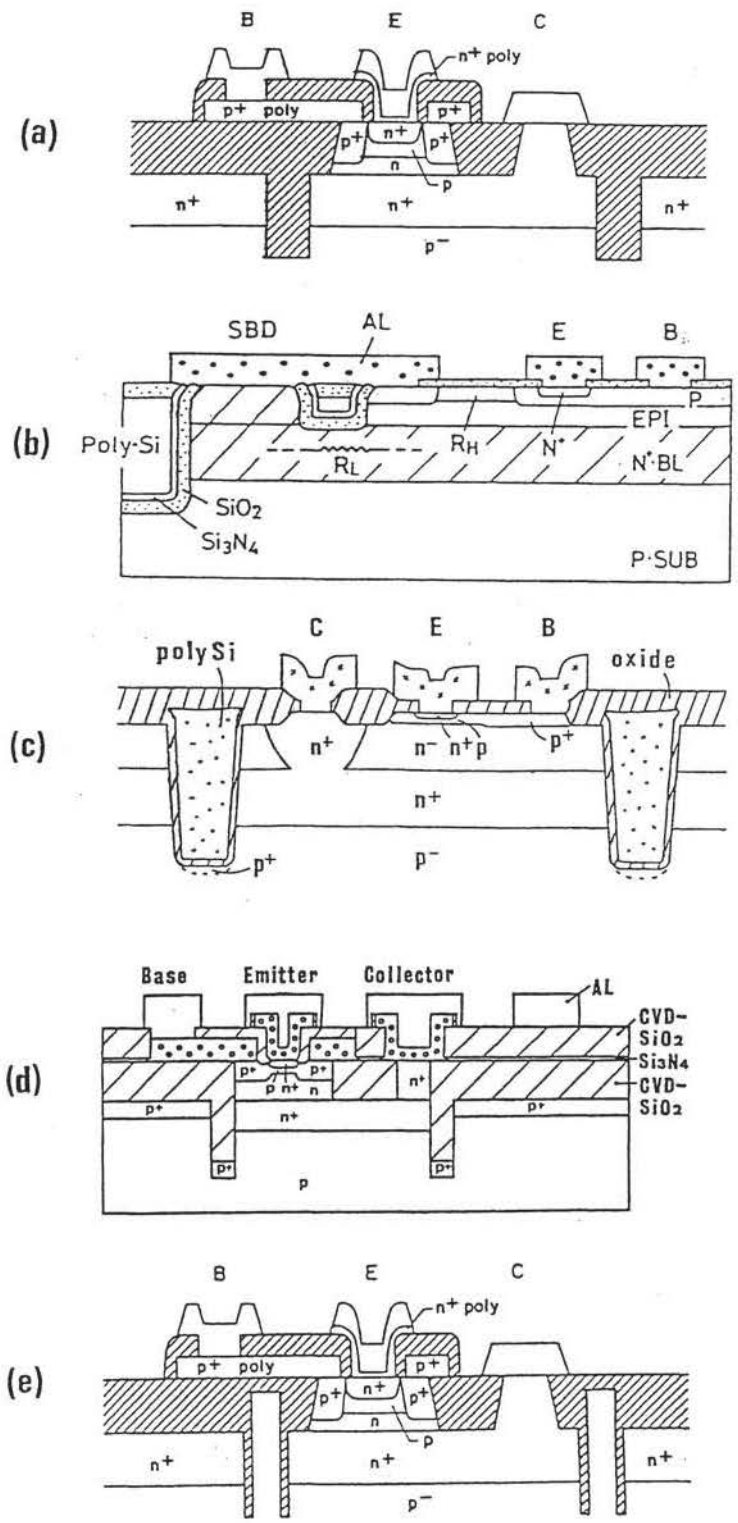


Fig. 2.6 Modified trench-isolated structures.
(a) [12,13] (b) [14] (c) [9] (d) [10] (e) [15,16]

Table 2.1 is an example of device performance demonstrated by 1kb ECL (*e*mitter *c*oupled *l*ogic) static RAM (*r*andom *a*ccess *m*emory) using our IOP-II [5]. The V-groove-based IOP device was shrunk at the ratios of 0.8, 0.7 and 0.6. Trench width can be shrunk by the same factor as the active regions, from 5 μ m of the V-groove to 4, 3.5 and 3 μ m of the trenches, respectively. Because of this linear shrink, the typical access time can be reduced in proportion to the shrinking factor. Our IOP-II also realized various ECL static RAMs from 4kb to 64kb generations for high-speed cache RAM applications [19-24]. On the other hand, the advantage of trench isolation was shown using ring oscillator circuits by other researchers. The basic gate delay of an ECL circuit using with trench isolation was nearly 30% improved compared with oxide isolation [8]

Table 2.1 Typical address access time of linear shrunk 1kb ECL RAM using IOP-II.

Shrink ratio	Isolation width (μ m)	Access time (ns)	Technology
1.0	5.0	7.4	IOP(V-groove)
0.8	4.0	5.8	IOP-II(Trench)
0.7	3.5	4.8	IOP-II(Trench)
0.6	3.0	4.4	IOP-II(Trench)

Not only trench structures but also the modification in the second generation have much improved device performance. By introducing thick field oxide for wiring regions, parasitic capacitance between wiring layers and substrate have been reduced. This approach is suitable for logic devices. As well as the reduction of wiring-substrate capacitance, base-collector parasitic capacitance is also reduced because LOCOS or shallow-trench-filled oxide surrounds the base region to achieve a so-called walled base structure. In the case of our U-FOX structure, the wiring delay time of 50ps/mm was achieved for the first and second metallization layers using a 1.6-um wide wiring pattern. A loaded gate delay of 323ps (basic delay=177 ps, FI=FO=3, wiring length=2mm) using an ECL circuit was obtained at a switching current of 0.2mA/gate [9]. The U-FOX structure made possible 3000- and 10,000-gate ECL logic LSIs for the Fujitsu mainframe computer FACOM-M780.

Other examples demonstrated by trench isolation techniques are listed in ref [25].

2.2.5 Limiting Aspects and Recent Improvements

Trench isolation is now a must for high performance bipolar devices. In the early stage of the development, there were several problems, and those were broadly divided into two categories. One was the structural problem and the other was the electrical problem. They were often strongly related and the former, in particular, is apt to influence the latter. As previously described in subsection 2.2.2, if the trench etching process is not optimized, trench shapes may change variously. The structural problems without electrical ones are usually bad step coverage of metallization layers caused by voids or sharp crevices on isolation regions. To avoid these problems trench shape should not show bowing or overhung and over etching in etch-back processes to remove the excess filler materials should be carefully controlled. Other maladapted trench structures such as black silicon, sub-trench, laterally-etched undercut of highly doped buried layers often accompany crystalline defects. These crystalline defects cause large leakage current between isolated elements. Thus, trench-etching processes should be optimized thoroughly to achieve the best conditions.

The most serious problem was how to get rid of isolation leakage caused by crystalline defects. The crystalline defects often occur at stress-concentrated places such as sharp-edged corners caused by unfavorable trench structures. However, even if the trench shapes show the ideal structure, crystalline defects may sometimes appear. They seem to be influenced by ion-implantation-induced damages and inadequate subsequent heat processes. The use of Y-shaped trenches instead of U-shaped trenches is one way to avoid stress-concentrated right-angled corners around top edges of trenches at the expense of packing density [26]. In relation to trench isolation combined with LOCOS, some process modifications were also proposed. One is the nitride liner between the oxide of the trench surface and filled polysilicon [27], and another was the combination of *se*lective *p*olysilicon *ox*idation (SEPOX) with trenches, which was applied to a CMOS static RAM [28]. In both of these techniques, selective oxidation of field

regions is performed at the same time as capping oxidation. Without these modifications large stress occurs at the top corner of trenches when single crystalline silicon and filled polysilicon are oxidized simultaneously, and they causes crystalline defects. Most recent approach to avoid large stress by capping oxidation is to use soft reflow glass to fill the trench. Boron-phosphorus silicate glass can be used for such purpose [29]

2.2.6 Conclusion

Trench isolation is one of the key techniques in high performance bipolar devices. It realizes higher packing density and smaller collector-substrate parasitic capacitance than conventional isolation techniques. It also reduces wiring-substrate and base-collector parasitic capacitance with process modifications such as combining thick field oxide for wiring regions.

2.3 Polysilicon Self-Aligned Structures

2.3.1 Introduction

As described in the historical background section, self-aligned techniques using polysilicon is another key feature to realize high performance silicon BJTs. These self-aligned techniques consist of double polysilicon and/or single polysilicon processes. Moreover, polysilicon emitter is also important to achieve shallow emitter that results in high cutoff frequency. The polysilicon emitter is a root for wide-gap emitter, or emitter-base heterojunction BJT. In this section, these self-aligned and polysilicon-related device structures will be discussed.

2.3.2 Double Polysilicon Self-Aligned Structure

In this subsection, various kinds of double-polysilicon self-aligned processes and related device structures are discussed. The double polysilicon self-aligned processes are roughly divided into three classes depending on the number of masks used to form the emitter and base regions. A one-mask

process can define not only the emitter and intrinsic base region, but also the extrinsic (graft) base region by using the same mask. A *s*uper *s*elf-aligned *t*echnology (SST) is a typical example of this type [30]. In the case of a two-mask process, the extrinsic base regions are defined by a mask different from that of the emitter and intrinsic base region. Until now, these processes are widely used in advanced bipolar IC production. Finally, Three-mask processes use different masks to define the emitter, intrinsic base and extrinsic base regions, respectively.

2.3.2.1 One-Mask Process

A typical and the most successful one-mask process is the so-called SST. The first version of this process was reported in 1980 [30], and some process improvements have since been introduced. Major improvement was reported in 1983 such as advanced SST [31]. Figure 2.7 shows the fabrication steps of the SST-1A (an advanced version of SST-1) process. After growing n-type epitaxial layer on a p-type substrate and oxide isolation formation, an oxide layer, a nitride layer and a polysilicon layer are deposited consecutively. The polysilicon layer is selectively oxidized (above the oxide isolation regions). The polysilicon in the area corresponding to the transistor emitter regions is etched away (Fig. 2.7(a)), and then oxidized. The nitride is side-etched followed by etching of the underneath oxide (Fig. 2.7(b)). New polysilicon is deposited to fill the overhung space and to connect the first polysilicon to the epitaxial layer. The excess polysilicon is etched away (Fig. 2.7(c)). Thermal oxide is grown and the intrinsic base region is formed by boron ion implantation through the oxide. Then CVD oxide and polysilicon are deposited. The emitter window is opened by dry etching (Fig. 2.7(d)). Finally, the emitter area is diffused through the arsenic-implanted polysilicon layer (Fig. 2.7(e)).

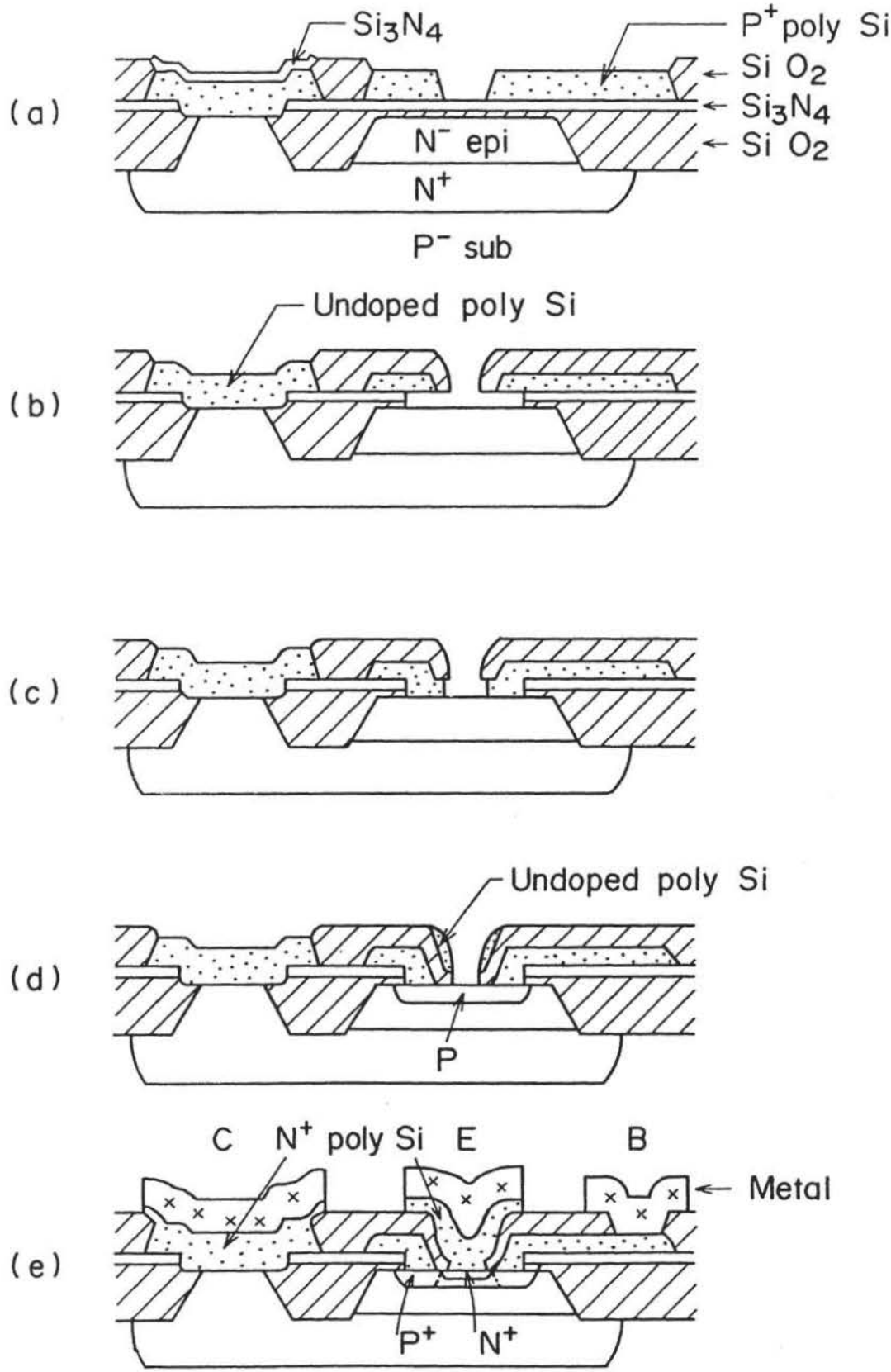


Fig. 2.7 SST-1A fabrication steps [31].

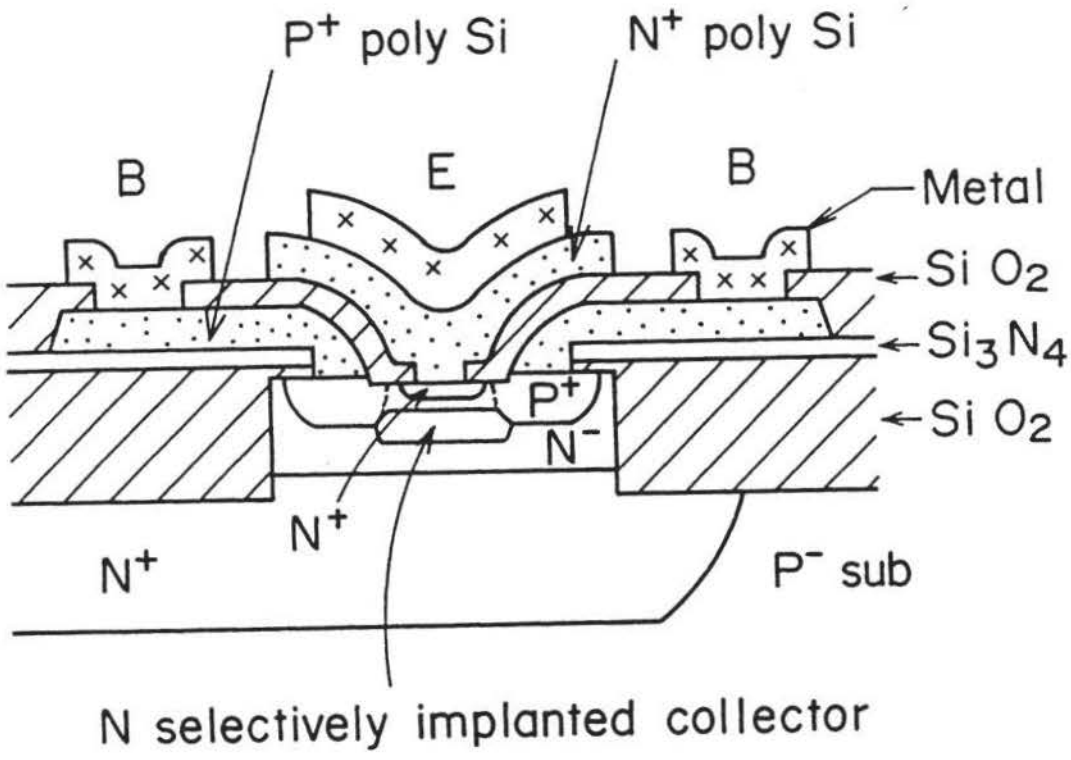


Fig. 2.8 Cross-sectional view of the SST-1B transistor with selectively implanted pedestal collector [32].

In the SST process, all the active regions are defined by a single mask pattern. The extrinsic base region is determined by the side-etched area in the nitride layer under the base-electrode polysilicon. The distance between the emitter and the base electrode is about the thickness of the CVD oxide and polysilicon. The emitter-window width can easily be made narrower than the original mask feature to achieve a submicron width. As a result, the base-collector parasitic capacitance and extrinsic base resistance of the SST transistor are dramatically reduced. In addition to obtaining a higher cutoff frequency, a selectively implanted pedestal collector process is used along with a low energy ion implantation for the intrinsic base region. Figure 2.8 shows the schematic cross-section of the improved structure. By using this improved process, named SST-1B, a peak cutoff frequency of 25.7 GHz (Vce=3V) was obtained, and a minimum basic gate delay time of 20.5 ps by an NTL circuit was demonstrated at a power dissipation of 2.32 mW/gate. At the same time, a minimum basic gate delay time of 34.1 ps by an ECL circuit at a

power dissipation of 7.54 mW/gate was also reported [32]. The drawback to the SST process is that there is a slight difficulty in defining a walled-base structure to reduce the parasitic base-collector capacitance. If the walled-base structure is required, a tight alignment control is needed between the field-oxidation and the intrinsic-base masks at the expense of the one-mask self-alignment advantage.

Another successful one-mask process was demonstrated by depositing the emitter polysilicon prior to the deposition of the base-electrode polysilicon [33]. Figure 2.9 shows the fabrication steps of this polysilicon self-aligned technology. After LOCOS isolation, the intrinsic base region is first implanted (Fig. 2.9(a)). Then, heavily implanted n-type polysilicon, CVD oxide and nitride layers are deposited and successively etched where the intrinsic base and emitter region are defined (Fig. 2.9(b)). To isolate the sidewall of the n-type polysilicon, the second CVD oxide is deposited, followed by reactive ion etching (Fig. 2.9(c)). A second nitride is deposited and reactively etched to form a vertical nitride layer (Fig. 2.9(d)). By using the first and second nitride as a mask, the surface of the extrinsic base area is locally oxidized, while the emitter is also driven into the intrinsic base region (Fig. 2.9(e)). The base contact window is opened by removing the nitride layer. Then p-type doped polysilicon is deposited, which acts as the base electrode as well as the graft base diffusion source (Fig. 2.9(f)). By using this technique, a basic gate delay of 52 ps at a power dissipation of 0.16 mW/gate employing a CML circuit was demonstrated. This process has the same advantages as those of the one-mask process, but there are some drawbacks such as higher emitter series resistance caused by the polysilicon emitter electrode without metallization, furthermore, this is an unwalled base structure.

Other one-mask processes have been reported based on the same fabricating sequence as the SST, for instance, as the side frame self-aligned (SFS) technology [34], or the selective diffusion by exdiffusion (SDX) technology [35]

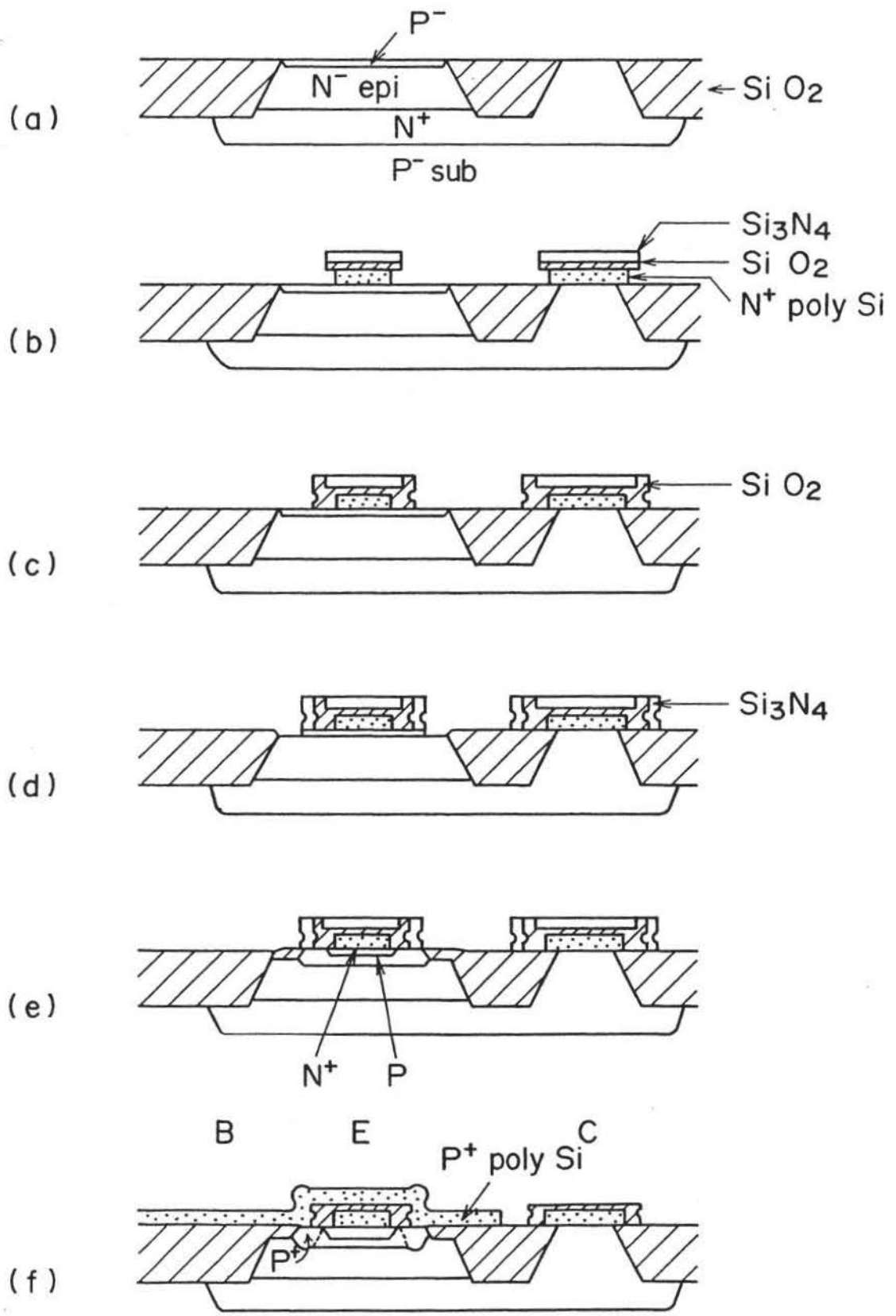


Fig. 2.9 Fabrication steps of the new polysilicon self-aligned process [33].

2.3.2.2 Two-Mask Process

Two-mask processes have been widely used to achieve the double polysilicon self-aligned structures. The first and simplest example was reported in 1980 as a self-aligned transistor [36]. Figure 2.10 shows its fabrication steps. P-type boron doped polysilicon and CVD oxide layers are deposited on an n-type epitaxial layer after conventional recessed oxide isolation. The recessed oxide isolation defines not only collector reach-through regions but also the total base regions (Fig. 2.10 (a)). The CVD oxide and polysilicon layers are preferentially etched by using an intrinsic-base-window opening mask (Fig. 2.10(b)). The surface of the epitaxial layer and the sidewall of polysilicon are oxidized, while the extrinsic (graft) base is diffused from the boron-doped polysilicon (Fig. 2.10(c)). The intrinsic base region is opened by RIE. Finally, the intrinsic base is ion-implanted, and the emitter is also diffused into the same region (Fig. 2.10(d)). Since this early development, numerous process modifications and improvements have been introduced by a number of researchers. These processes have been given various names such as SAPT [17], OXIS-III [37], SDD [38], MOSAIC-III [39], ESPER [40], BSA [41] and many others [42-47].

The two-mask process uses different masks for the extrinsic (graft) base and intrinsic base. Therefore, the walled-base structure can be easily obtained without using a complicated side-etching technique, although this is at the expense of alignment tolerance. Other properties of the two-mask processes are fundamentally the same as those of the one-mask process. The most significant and careful process control is required for the intrinsic-base formation, the sidewall spacer etching, and the intrinsic base diffusion. There are two options for the intrinsic base diffusion step and the sidewall-spacer-etching step. One is to have the intrinsic base implantation prior to the sidewall spacer formation, and the other is to reverse the sequence of the two steps. In the case of the first process sequence, the intrinsic base region is implanted just adjacent to the extrinsic base. Because the extrinsic base is already diffused vertically and laterally toward the intrinsic base region, the requisite low resistance cross-linking region between the intrinsic base and the extrinsic base is stably obtained. Moreover, the lateral-encroachment of the extrinsic base region can be covered by the sidewall spacer, which

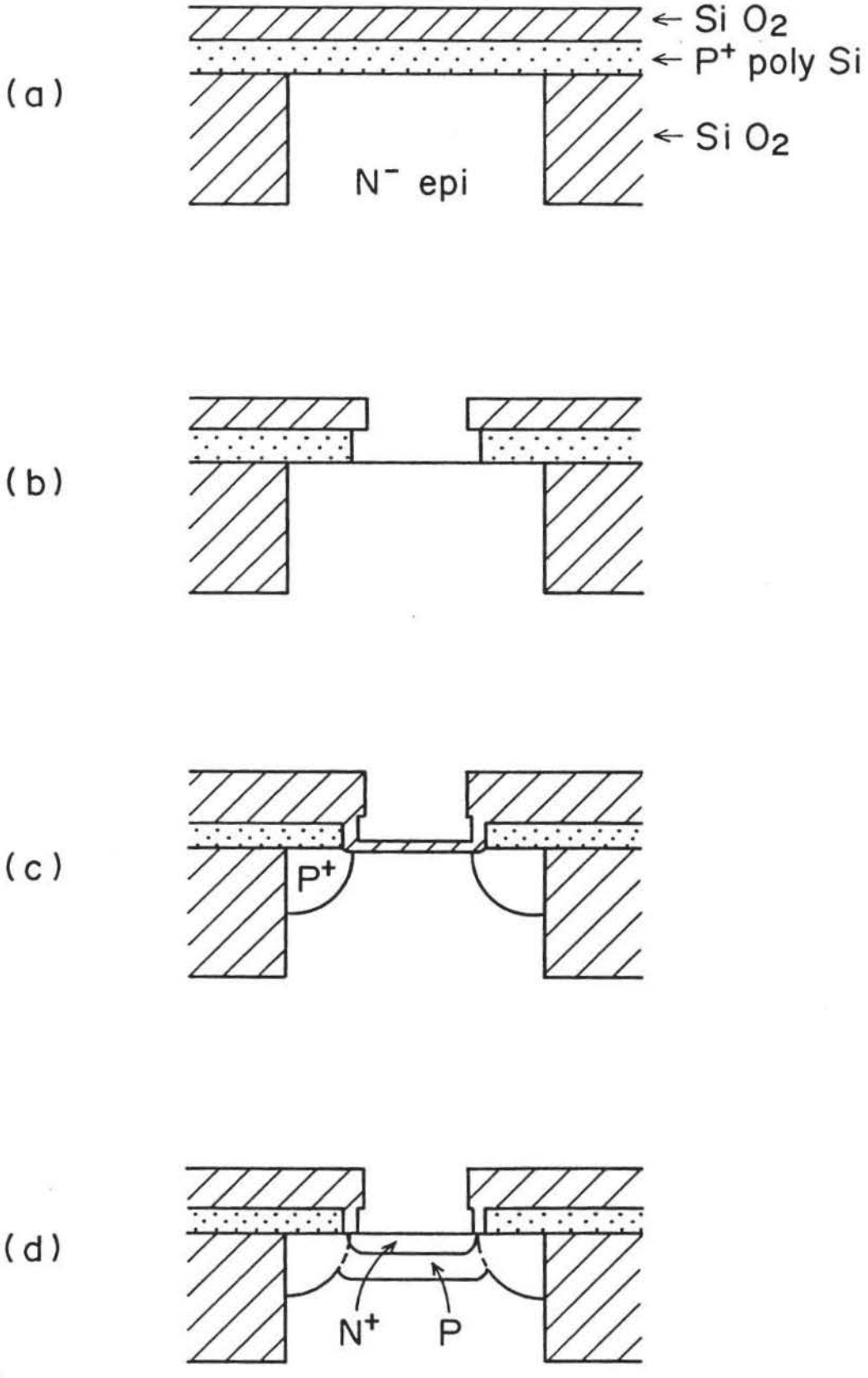


Fig. 2.10 Fabrication steps of the self-aligned transistor process [36].

eliminates any emitter-base tunneling leakage current and ensures a high emitter-base reverse breakdown voltage. However, attention should be paid to the sidewall-spacer-etching step to avoid overetching, because the intrinsic base region is damaged and the base depth might be changed.

In the second process sequence, overetching of the sidewall spacer does not damage the intrinsic base region, but the extrinsic base diffusion must be carefully controlled. Too much lateral encroachment of the extrinsic base results in highly doped pn-junction at the emitter periphery, causing not only reverse, but also forward tunneling leakage current [48]. Also caused is unfavorable degradation of dc current gain as well as that of transient characteristics [49,50]. However, the peripheral punchthrough occurs when the cross-linking region between the intrinsic and extrinsic base is not satisfactorily diffused [51].

To avoid these contradictory problems, some process modifications have been reported to combine light-coupling ion implantation and base diffusion through polysilicon [47,52]. The coupling base is lightly implanted after etching the CVD oxide and the base polysilicon (Fig. 2.11(a)). Then, CVD oxide is deposited and etched by reactive ion etching to form the sidewall spacer, and another polysilicon layer is deposited (Fig. 2.11(b)). The intrinsic base is implanted and diffused through the polysilicon, while the extrinsic base is also driven in (Fig. 2.11(c)). By using this technique, a cutoff frequency of more than 30 GHz and a minimum basic gate delay of 24 ps at a switching current of 2.2 mA/gate by an LCML circuit have been reported [47]. The drawback to this coupling base technique is its process complexity and the difficulty in controlling the intrinsic base-emitter depth because the boron diffusion for the intrinsic base and the arsenic diffusion for the emitter require different annealing conditions due to the difference between the respective diffusion coefficients.

Some authors have investigated reliability of two-mask double polysilicon self-aligned devices. They reported gain degradation under reverse emitter-base bias and high current forward base-emitter bias stress conditions when transistors are irradiated by laser, X-ray and electron beams [53-56]. Although not all the degradation mechanisms are understood, if the ideal factor of the base current is kept unity, the electrical characteristics are found to be unchanged under high-current stresses. This fact suggests that if

the cross-linking region between the intrinsic and extrinsic base is well controlled, the transistors have enough reliability to operate under normal bias and environmental conditions.

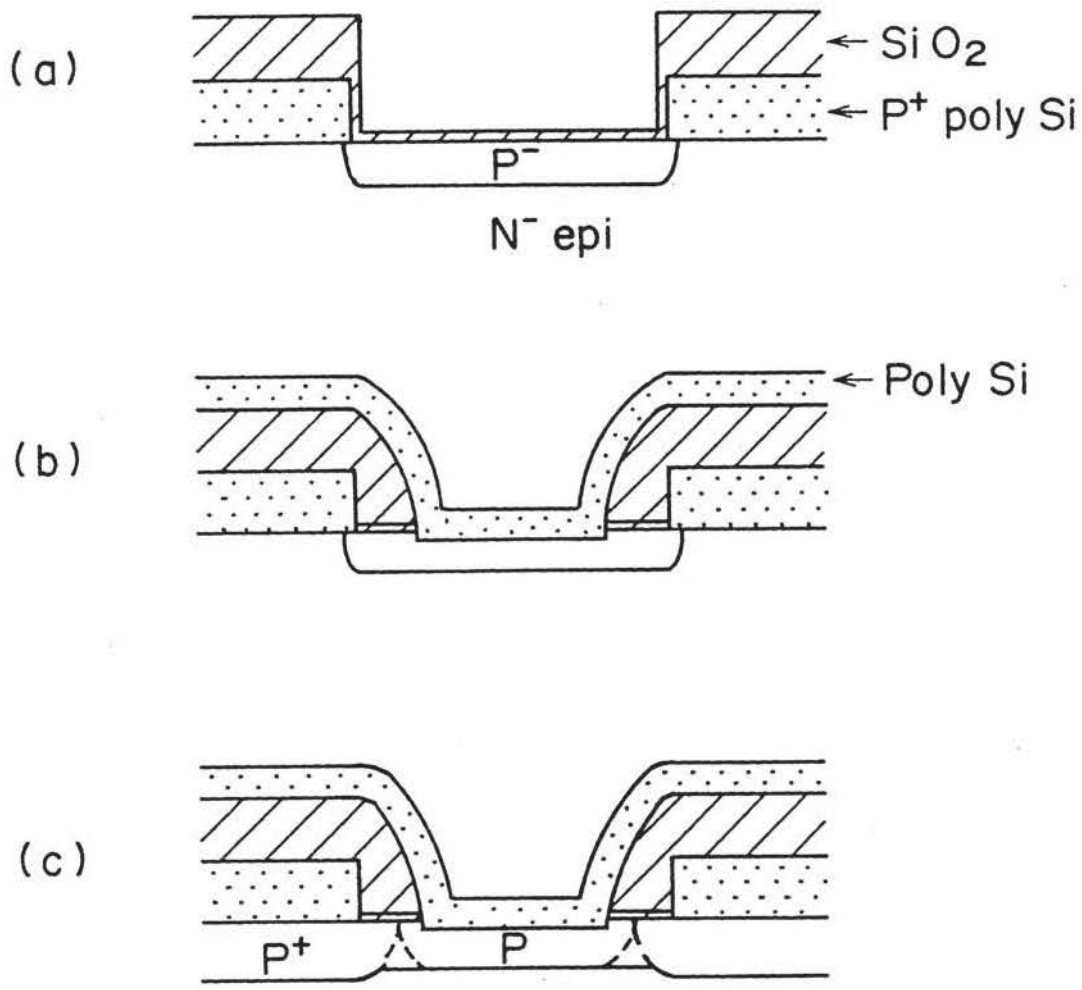


Fig. 2.11 Fabrication steps of the double polysilicon self-aligned process using coupling base implantation [52].

Although the two-mask process requires alignment tolerance between the field isolation and emitter opening masks, resulting in a little larger base area than that of the one-mask process, good circuit performance was obtained. For example, our ESPER process realized basic gate delay of 38.8ps by ECL circuit at a switching current of 1.28mA/gate [40]. In our case, the double polysilicon self-aligned process was combined with the U-FOX structure. By using this technology, 15,000 gate ECL array LSI with a loaded gate delay of

220 ps (basic delay=80 ps, FI=FO=3, wiring length=2 mm) and 64kb-RAM with 3,000 gate ECL array LSI (SRAM access time=1.15 ns) for the Fujitsu mainframe computer FACOM-M1800 and super computer VP2000 series were realized [57]. In addition to them experimental 50,000 gate sea-of-gate ECL array and 100,000 gate ECL standard cell LSIs were also demonstrated [58,59]

2.3.2.3 Three-Mask Process

There have not been so many examples of the three-mask process reported to date. This is not only because total active base area obtained by the three-mask process is slightly reduced but also as a result of the number of process steps being increased. However, the first double-polysilicon self-aligned process was demonstrated by a three-mask process named APSA, an advanced version of PSA [60]. Historically, this process has some value, but because of the drawbacks to the three-mask process, few efforts have been made to improve it. Only one approach was reported as a *s*alicide base *con*tact *t*echnology (SCOT) [61].

2.3.3 Single Polysilicon Self-Aligned Process

Basically single polysilicon process originated from the DOPOS technique [2]. At the early stage of development, some process improvements attained to achieve self-aligned isolation between emitter and base electrodes, such as *s*tepped *e*lectrode *t*ransistor (SET) or *e*levated *e*lectrode IC (EEIC) [62,63]. In these processes, an overhang polysilicon emitter structure isolated the emitter polysilicon and the base electrode metallization. Another approach was also reported to use the sidewall spacer technique just as in double polysilicon self-aligned processes. This self-aligned process employed the sidewall spacer between the emitter polysilicon and the aluminum base electrode [64]. All of these are two-mask processes, and the distance between the emitter and the base electrodes is defined by self-aligned techniques. However, the disadvantages of these processes are nonplanar surface topographies of the SET or the EEIC and a high emitter series resistance caused by the absence of emitter metallization. Because of these

disadvantages, development efforts essentially instead focused on improving the double-polysilicon self -aligned processes.

A single polysilicon three-mask process has also been reported [65]. It was demonstrated as an example to avoid cross-linking problem of double polysilicon self-aligned processes. Figure 2.12 shows the process steps [66]. After conventional oxide isolation, a nitride layer is preferentially patterned at the portion of intrinsic-base and emitter defining region. By using the nitride layer as a mask, the epitaxial layer is slightly oxidized, and a polysilicon sidewall spacer is formed. Highly doped extrinsic base regions are implanted (Fig. 2.12(a)). After the polysilicon sidewall spacer is removed, the cross-link base regions are lightly implanted (Fig. 2.12 (b)). After removal of the nitride, intrinsic base is implanted (Fig. 2.12 (c)). Emitter is diffused from an arsenic doped polysilicon layer (Fig. 2.12(d)). The emitter polysilicon is covered with a silicide layer and an oxide layer on top. Another mask is used to define the emitter electrode, which is covered by the capping oxide and new sidewall spacers. The base contact window is opened at the same time when the sidewall spacer is etched (Fig. 2.12(e)). By using this process, a basic gate delay of 49.8 ps for an ECL circuit was demonstrated at a switching current of 0.65 mA [65].

In this three mask process, the cross-link region is made before the intrinsic base implantation, and only the emitter drive-in heat step is required after the intrinsic base formation. Therefore, the cross-link problem induced by the double polysilicon process is eliminated. However, this process requires different masks for the intrinsic base region and base contact window opening. Furthermore, the somewhat higher series resistance of the emitter electrode is a drawback, although a silicide layer is deposited on the emitter polysilicon.

2.3.4 Sidewall Base Contact Structures

Sidewall base contact structures were intended to achieve an ideal, one-dimensional transistor structure eliminating any parasitic capacitance and resistance. A sidewall base contact structure was first reported as SICOS in 1981 [67]. Since that time, much process modification had been reported. Figure 2.13 shows the fabrication steps of the original SICOS. First, a

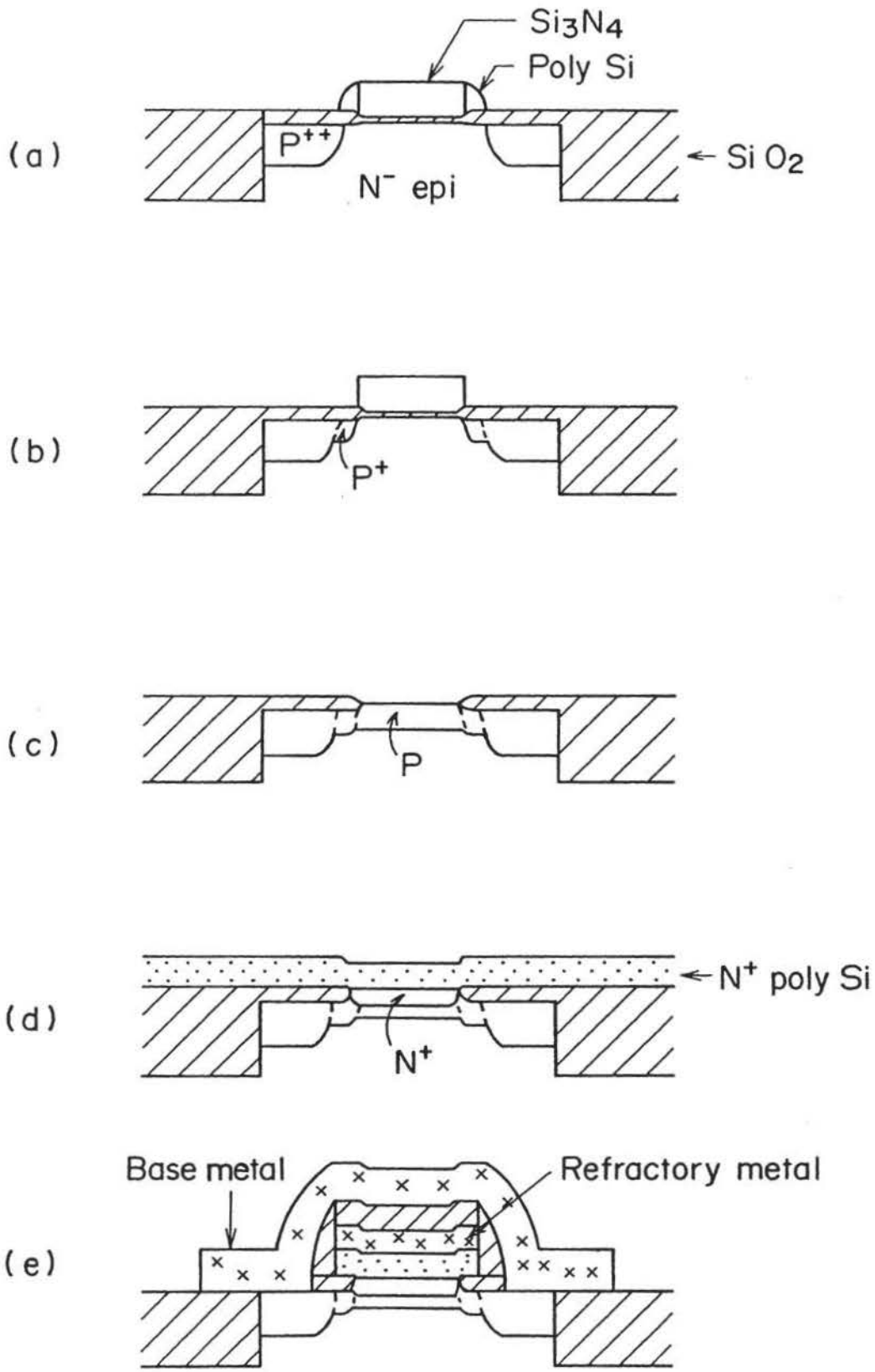


Fig. 2.12 Fabrication steps of the single polysilicon self-aligned process [66].

mesa-etched region is formed by reactive ion etching, using a multilayered structure, successively including an oxide, a nitride, a polysilicon, a nitride and a CVD oxide layer (Fig. 2.13(a)). After a slight sideetching of the mesa, another nitride layer is deposited and a spacer is formed at the sidewall edge. Then, the silicon surface is oxidized and the nitride spacer is removed (Fig. 2.13(b)). Another polysilicon layer is deposited and preferentially etched where the base electrode is positioned (Fig. 2.13(c)). Finally the graft base is diffused from the second polysilicon, which is boron-doped by ion implantation. The intrinsic base is implanted through the first oxide on the mesa and emitter is also diffused through another polysilicon layer (Fig. 2.13(d)).

The advantage of the SICOS is the nearly one-dimensional device structure in addition to its full one-mask self-alignment feature. Consequently, excellent upward operation has been reported since the early stage of the development. However, there has been a serious drawback to this process. This is encroachment of the extrinsic (graft) base, which occurs not only laterally but also vertically towards the buried collector layer. The lateral encroachment is the same problem as experienced by the double polysilicon self-aligned processes, whereas the vertical encroachment is specific to the SICOS structure. To reduce the extrinsic base resistance, the sidewall contact area should be kept large, and the graft base should also be satisfactory diffused. Therefore, the graft base is deeper than the intrinsic base region and the collector-base junction area is much larger than the original dimension on the mask. Hence, the collector-base junction capacitance can not be decreased while maintaining low base resistance. Moreover, the vertical encroachment of the extrinsic base causes the reduction of the base-collector reverse breakdown voltage.

Great efforts had been paid to remove these complications. A *self*-aligned *e*dge *c*ontact *t*echnology (SELECT) was applied to the original SICOS [68], whereby a basic gate delay of 48 ps was demonstrated by an ECL circuit at a switching current of 0.58 mA. Another approach was also reported, the *t*wo-step *o*xidation of the *s*idewall *s*urface (TOSS) [69]. Figure 2.14 shows the improved process steps, which determines the base contact width. The mesa is covered by a nitride frame (Fig. 2.14(a)). Then, the second mesa etching is performed by anisotropic chemical etching.

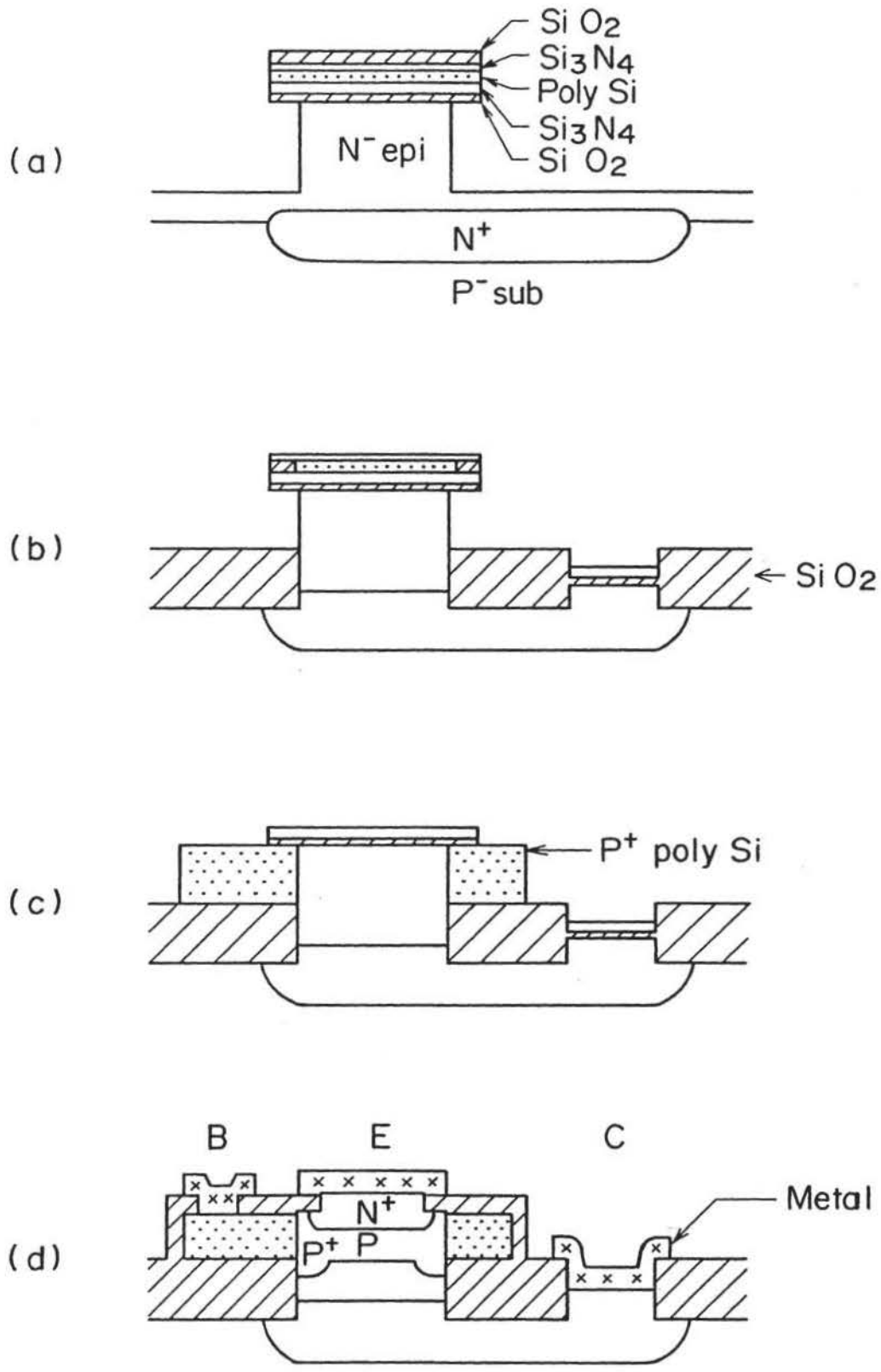


Fig. 2.13 Fabrication steps of the SICOS process [67].

The sidewall is again covered by nitride (Fig. 2.14(b)). Finally, isolation is achieved by LOCOS, followed by the nitride removal (Fig. 2.14(c)). These process options have been introduced to control the sidewall contact width to prevent the vertical encroachment. Nevertheless, these process modifications result in process complexity and a higher base resistance. This situation seems to show that practical bipolar transistors are basically two-dimensional three terminal devices, which never operates in one-dimensional mode because two opposite conductive carriers must pass orthogonally to each other. Therefore, to optimize the device structure, two-dimensional feature, or the equivalent distributed circuit characteristics should always be taken into account.

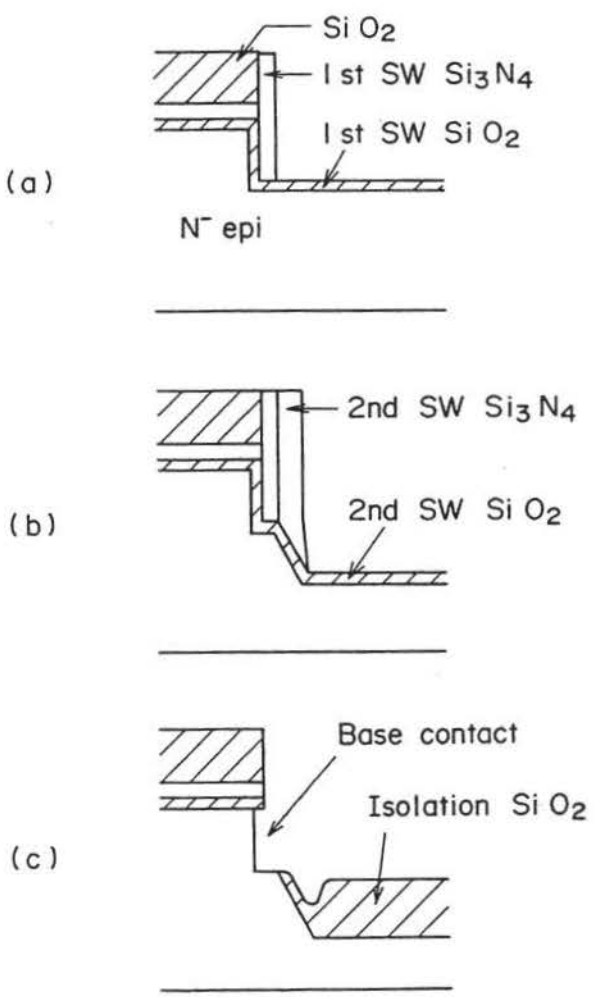


Fig. 2.14 Fabrication steps of the SICOS process modified by the TOSS process [69].

The other unique self-aligned approach for obtaining a sidewall base contact structure was demonstrated by a selective epitaxial growth technique. Strictly speaking, it was an epitaxial polysilicon simultaneous growth technique [70]. Figure 2.15 shows the process steps. After an n^+ buried layer is diffused, CVD oxide is deposited and active areas are opened by conventional lithography and etching (Fig. 2.15(a)). Then, an n-type epitaxial layer and a simultaneously grown polysilicon layer are continuously deposited. The surface of the epitaxial polysilicon-layer is covered by nitride (Fig. 2.15(b)). By using a controlled planarization technique, the nitride and the polysilicon on the field oxide are removed followed by base-electrode boron-ion-implantation (Fig. 2.15(c), (d)). Hence, the polysilicon surface is locally oxidized. The nitride at the intrinsic base region is removed and the intrinsic base is implanted (Fig. 2.15(e)). The emitter is diffused through a polysilicon layer from an arsenic doped spun-on-glass layer (Fig. 2.15(f)). By using this process, a basic gate delay of 83 ps was demonstrated by an LCML circuit at a 1 mW/gate power dissipation. Although the elegant feature of the technique was realized, this structure has the same problems as those of the SICOS. Moreover, the peripheral area of the selective epitaxial region has numerous crystalline defects, which must be eliminated to suppress no leakage current. Finally, there also seems to be some difficulty in achieving a one-dimensional device structure using selective epitaxial techniques.

2.3.5 Polysilicon Emitter and Emitter-Base Heterojunction

In addition to various kinds of self-aligning techniques, two notable process techniques have also been employed or investigated in modern silicon bipolar devices. One is polysilicon emitter, which is the basic concept for self-aligned transistor structures, and the other is emitter-base heterojunction. Polysilicon emitter can enhance the dc gain even with shallow base-emitter junction. Nevertheless, the mechanism of the gain enhancement is not perfectly understood, and much discussion had been reported at conferences or in scientific papers. Although to describe all of these is beyond the scope of this dissertation, the important point is that, by using polysilicon emitter, the current gain is no longer considered to be the limiting figure of merit for silicon bipolar transistors with shallow base-emitter junctions.

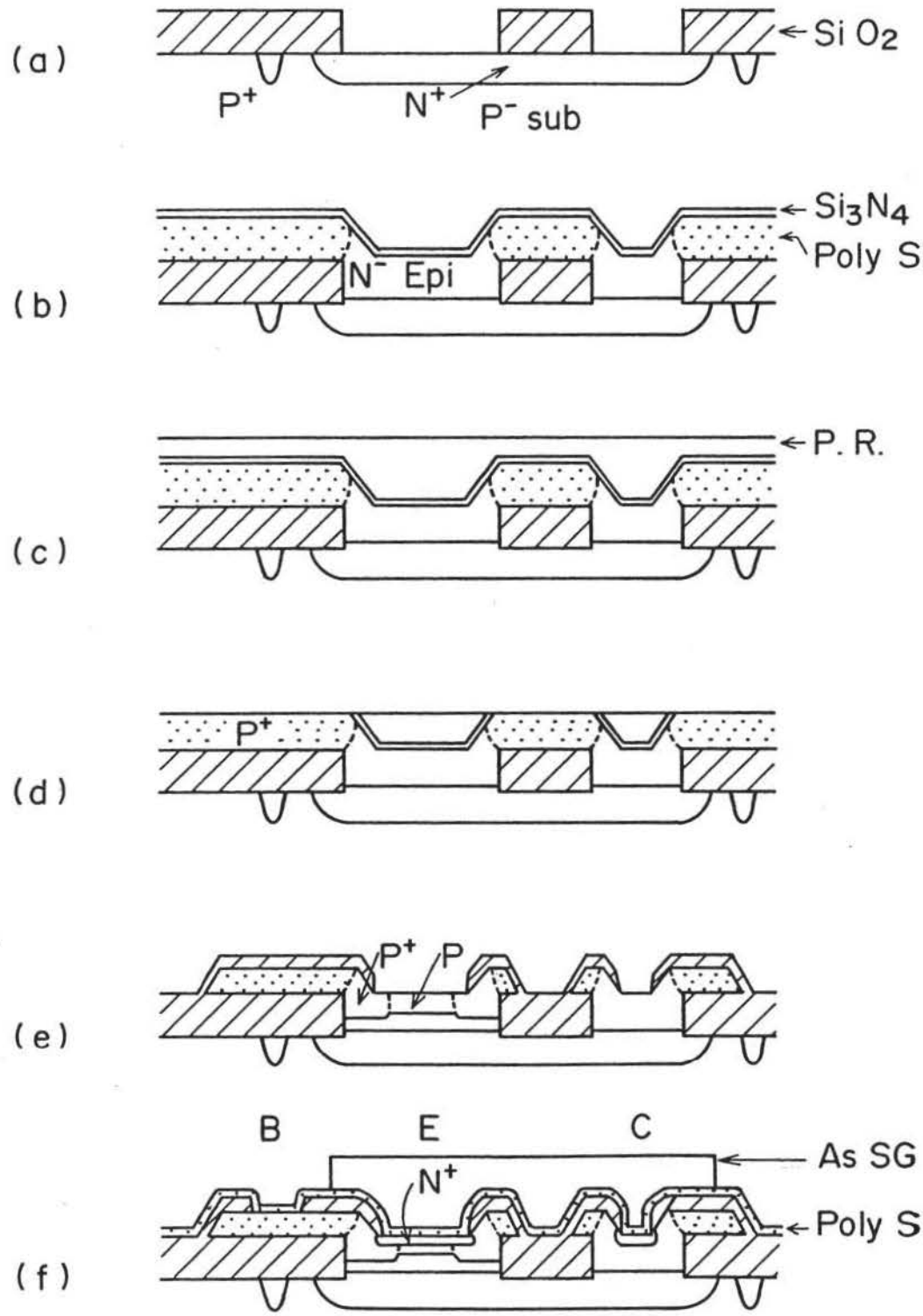


Fig. 2.15 Fabrication steps of the self-aligned selective poly-epi growth process [70].

Silicon-based heterojunction BJTs (HBTs), or the widegap emitter and/or narrow gap base have also been proposed to enhance the polysilicon emitter effect. Silicon-germanium for the narrow gap base is now becoming the most promising candidate, and a sophisticate process technology combining selective *u*ltra *h*igh *v*acuums (UHV) CVD epitaxial base with double-polysilicon self-aligned structure has also been proposed [71,72].

2.3.6 Conclusion

The double polysilicon self-aligned structure is also another must for high-speed silicon bipolar LSIs. By employing this structure, BJTs had achieved high-speed performance more competitive than compound semiconductor devices without using tight lithographic design rules. BJTs also utilized full advantage of polysilicon material properties, that have recently been merged to silicon-based heterojunction BJTs.

2.4 Physics-Based Bipolar Transistor Design

2.4.1 Introduction

In order to extract ultimate high-speed performance of silicon BJTs, well-balanced optima of device parameters such as cutoff frequency, base resistance and base-collector capacitance have major concerns. For example, by combining double polysilicon self-aligned processes, a pedestal-collector transistor has revived to improve cutoff frequency using a selective-ion-implantation technique [32,73-77]. The transition was able to achieve satisfactorily high cutoff frequency without so much increase of base resistance and base-collector capacitance because a pedestal collector is self-aligned to an intrinsic base region.

However, when a highly-doped collector is used to prevent base push-out (Kirk) effect, avalanche generation due to a stronger electric field at a base-collector junction makes collector-to-emitter breakdown voltage lower [78]. This low breakdown voltage seriously limits circuit design capability. Therefore, an optimal process for pedestal-collector transistors should carefully be maintained to realize not only well-balanced device parameters

but also high enough collector-to-emitter breakdown voltage.

In this section, will be described the favorable process conditions for double polysilicon self-aligned pedestal-collector transistors to match the requirements mentioned above by using a newly developed two-dimensional process/device simulation system. The simulation system will be described in subsection 2.4.2, and the simulated results will be explained in subsection 2.4.3. The process conditions for favorable highly-doped pedestal collectors will be discussed in subsection 2.4.4. In addition to the simulation analyses, will also be shown some experimental circuit performance improvements demonstrated by ECL ring-oscillators and a 1/8 static frequency divider in subsection 2.4.5. Finally, the conclusion will be given in subsection 2.4.6.

2.4.2 Two-Dimensional Process/Device Simulation System

In the case of double polysilicon self-aligned processes, phosphorus or arsenic ions are selectively implanted just beneath an intrinsic base region to form a pedestal collector using a polysilicon base electrode covered with oxide as a mask [32]. The implanted ions are dispersed laterally along a base-collector junction, and the distributed impurity profile strongly affects not only device parameters but also collector-to-emitter breakdown voltage. Therefore, two-dimensional process simulation is inevitable to analyze transistor characteristics precisely. For this purpose, we have developed a new two-dimensional process simulator named *FIPS2* based on a finite difference method. FIPS2 is an enhanced version of a one-dimensional process simulator named *FIPS* [79]. FIPS2 includes an epitaxial autodoping model which solves a mass-balance equation on a substrate surface. It also includes an empirical fitting model for channeling effects of shallow ion implantation. In addition to the channeling model, parameters for Pearson distribution functions obtained by Monte Carlo simulation are considered in FIPS2 to calculate the pedestal collector profiles accurately. By using these models, typical advanced bipolar processes are effectively analyzed.

The calculated two-dimensional doping profiles are transferred to a two-dimensional device simulator named *BIP2DAN*. BIP2DAN is a drift-diffusion-model-based two-carrier device simulator which can calculate not only dc and ac characteristics but also transient characteristics with some

lumped external circuit elements [79-81]. It also includes typical physical models such as Shockley-Read-Hall, Auger recombination models, bandgap narrowing effects, impurity concentration, electric field and temperature dependent mobility models, impurity concentration dependent carrier life times, and an avalanche generation model based on the Chynoweth law.

These process simulator and device simulator are combined directly. We can accordingly analyze the relationship between process conditions and device characteristics including collector-to-emitter breakdown voltages. We have employed dc and ac analyzing functions for this time to estimate the effects of pedestal-collector-ion-implanting conditions to device parameters and breakdown voltages.

2.4.3 Simulation of Pedestal-Collector Transistors

In this subsection, the simulation procedure is described including device structures and process conditions.

2.4.3.1 Process Simulation

We have assumed a typical double polysilicon self-aligned transistor having a low-energy-boron-implanted base (less than 10 KeV) with a thin screen oxide layer. It also has an n-type epitaxial layer of a typical thickness on a heavily doped buried layer. We have chosen three acceleration energies for phosphorus ion implantation to form pedestal collectors. Emitters have been assumed to be diffused from arsenic-implanted polysilicon. Figure 2.16 shows the one-dimensional doping profiles at the center of the emitter. The ion-implanting conditions are also shown in Fig. 2.16. No implantation case is the control sample for comparison, described as "Reference" in the figures.

Figure 2.17 illustrates the two-dimensional impurity distributions exhibited by concentration equi-contour lines. The peak positions and lateral spreading of the pedestal collectors are clearly shown depending on the implanting conditions.

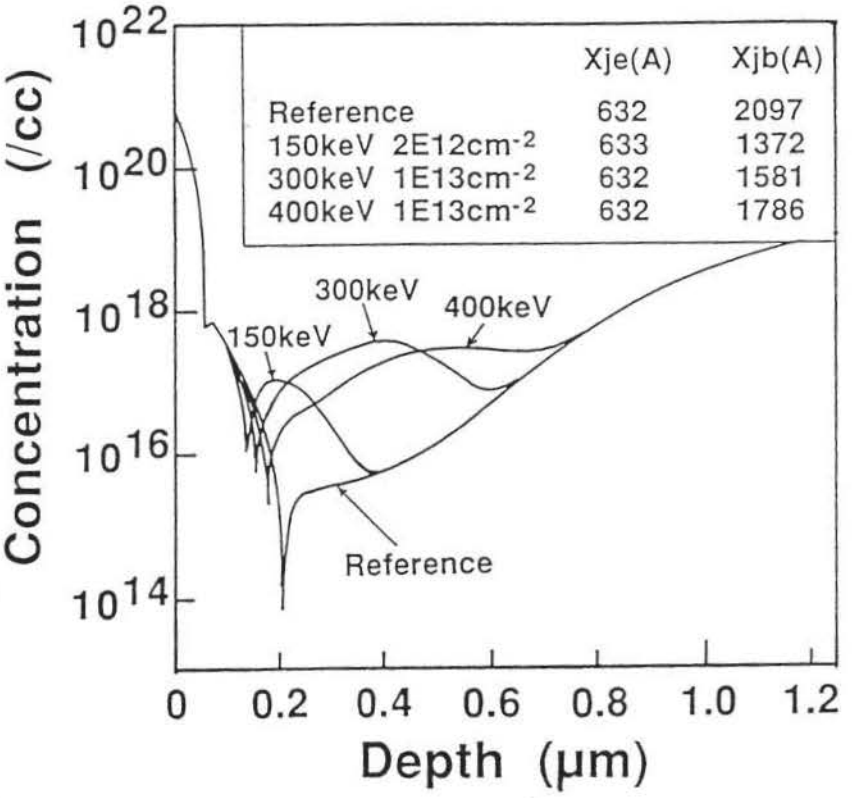


Fig. 2.16 One-dimensional doping profiles of analyzed transistors at the center of emitter.

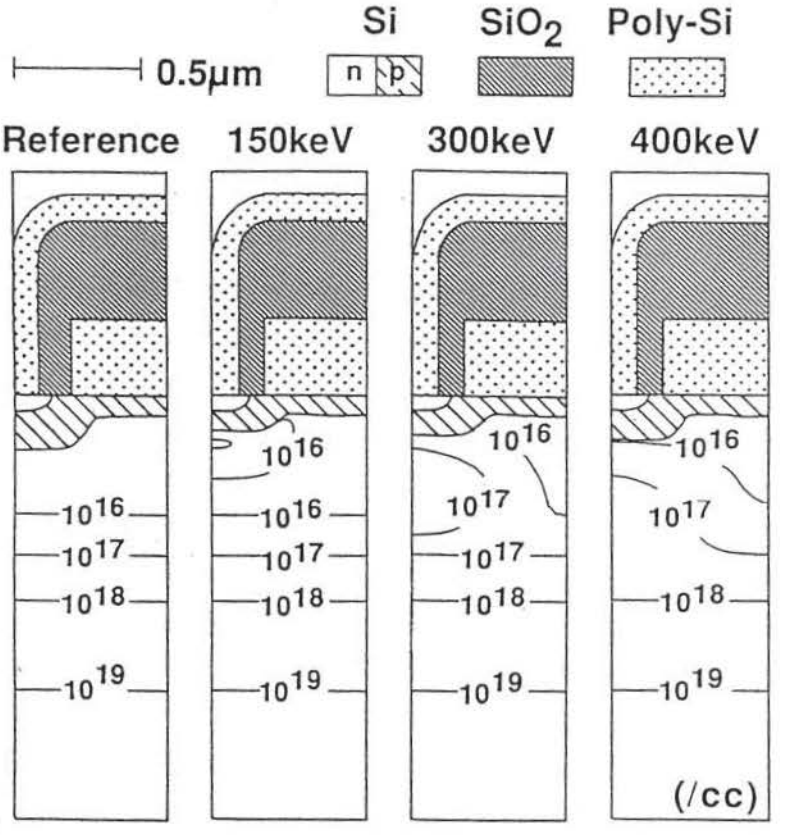


Fig. 2.17 Two-dimensional doping distributions with concentration equi-contour lines.

2.4.3.2 Device Simulation

The impurity concentration tables obtained in Fig. 2.17 are transferred to the device simulator. In the device simulator, only a single crystalline region is remeshed and impurity concentration values are interpolated for the Poisson and current-continuity equations. Figure 2.18 illustrates the schematic cross-section of an analyzed transistor. The polysilicon emitter is assumed to be a 0.1- μ m-thick extended single crystalline emitter with a maximum activated impurity concentration of $1x10^{20}$ /cm^3. The device length is set to be 1.0 μ m for convenient comparison to one another. The emitter area is $0.2x1.0\ \mu m^2$ and the base area is $1.2x1.0\ \mu m^2$.

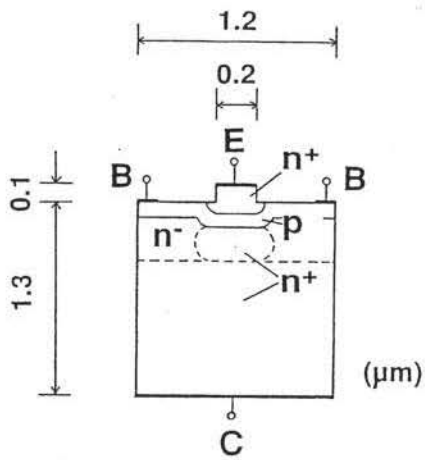


Fig. 2.18 Schematic cross section of an analyzed transistor. The device length is 1.0 μ m.

Figure 2.19 displays the cutoff frequency versus collector current characteristics for each implantation case. The cutoff frequency is improved by 80% in the maximum case by suppressing the Kirk effect. Ac base resistances and base-collector capacitances are also shown in Fig. 2.20 as functions of collector current. The ac base resistances have been calculated by the same way as the input impedance circle method that is usually used in small signal measurements. The base resistance of the implantation energy of 150-KeV case is increased, and the base-collector capacitances of all pedestal-collector cases are increased as well. Current gain dependence upon collector current is shown in Fig. 2.21.

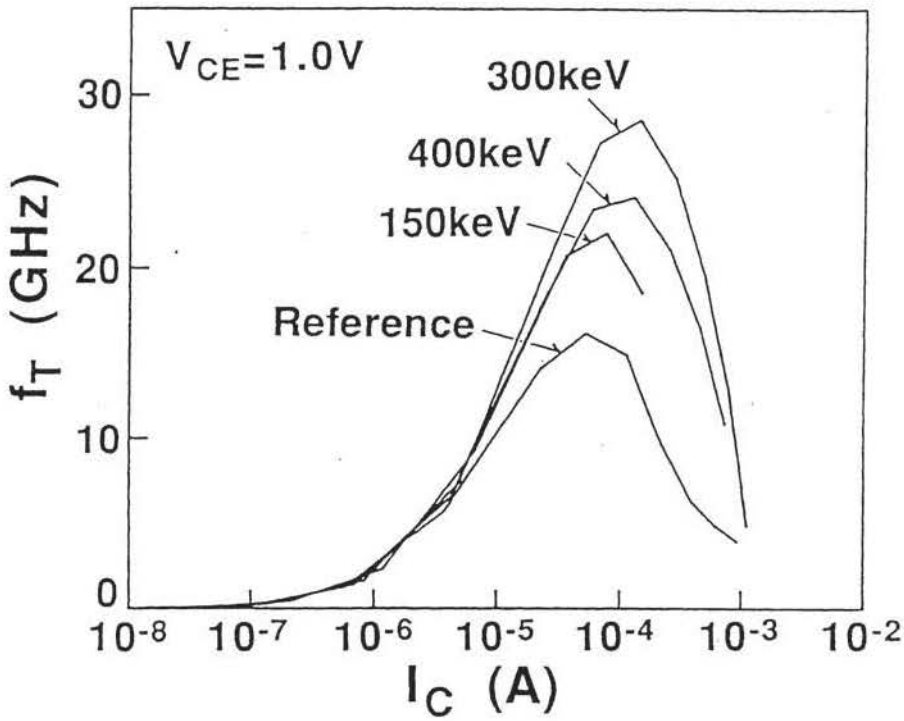


Fig. 2.19 Cutoff frequency (f_T) versus collector current (I_c) characteristics.

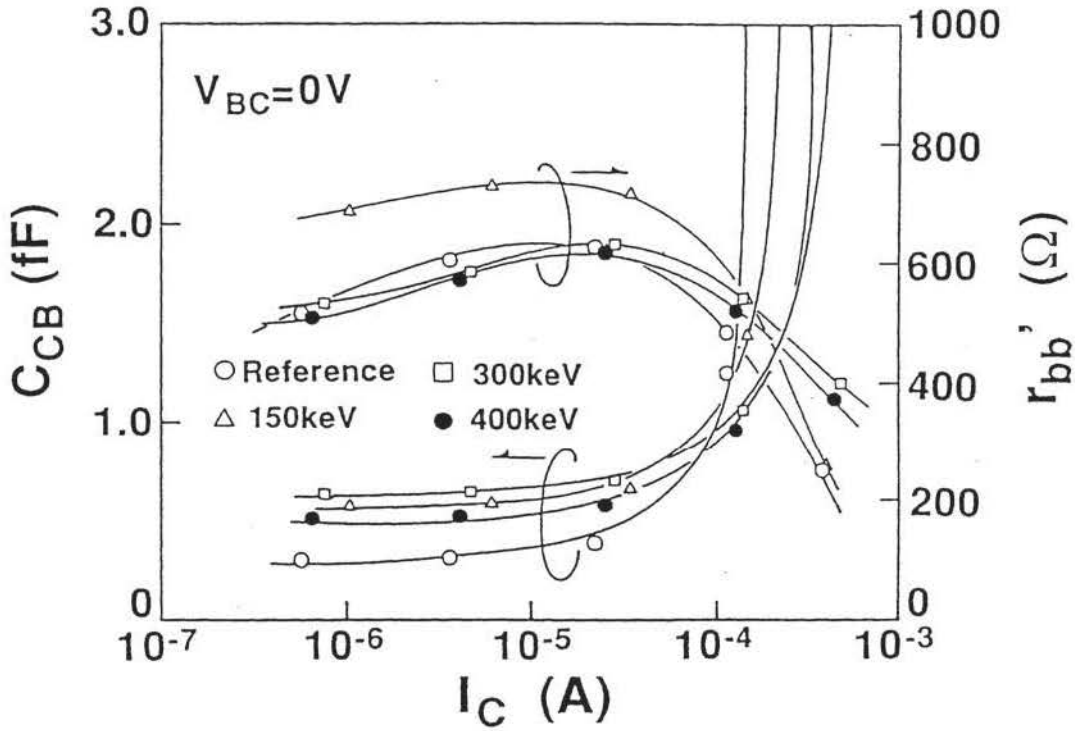


Fig. 2.20 AC base resistance ($r_{bb'}$) and base-collector capacitance (C_{cb}) versus collector current (I_c) characteristics.

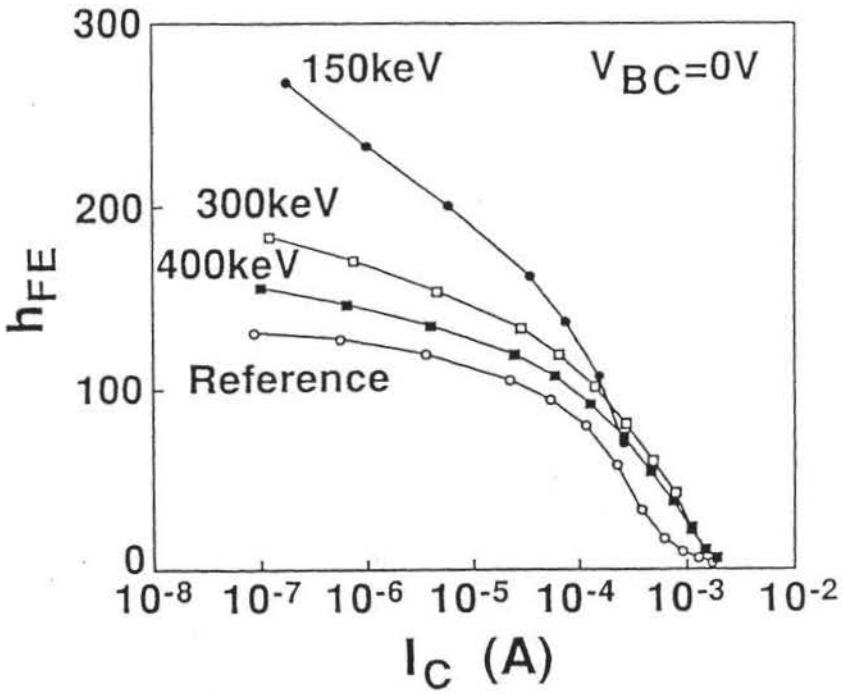


Fig. 2.21 Current gain (h_{FE}) versus collector current (I_c) characteristic.

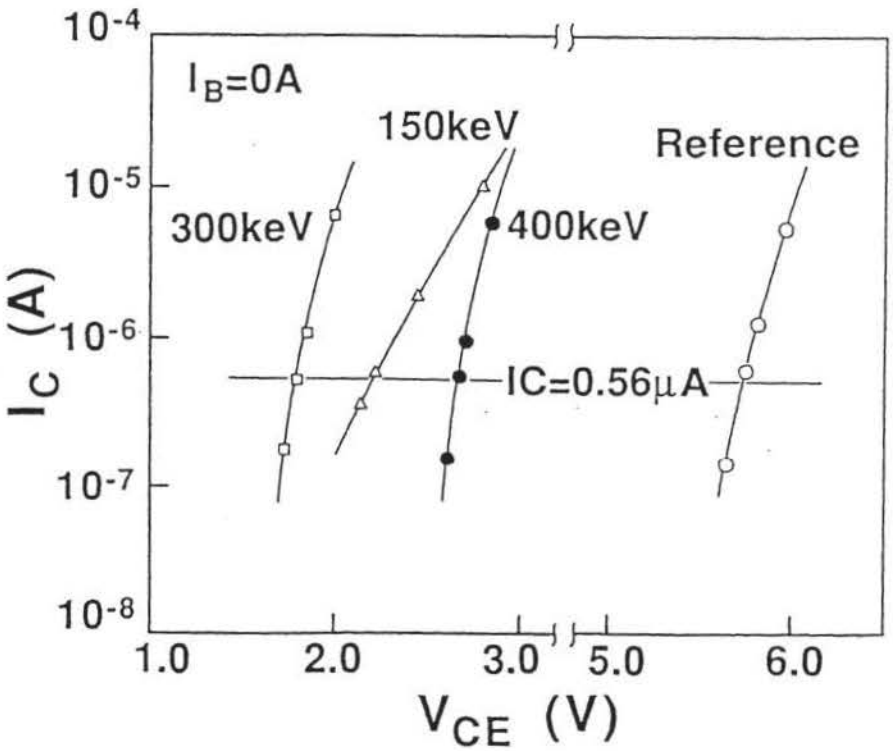


Fig. 2.22 Collector-to-emitter current (I_c) versus collector-to-emitter voltage (V_{CE}) where the base currents terminate.

The collector-to-emitter breakdown voltage (BV_{CEO}) has been calculated as follows. At first, the base-emitter forward voltage is supplied to a fixed value, and then the collector-to-emitter supply voltage is increased to find the bias point where the base current terminates. This means that the forward base current is equal to the avalanche-generated reverse base current. Next, the base-emitter forward voltage is increased slightly, and the collector-to-emitter supply voltage is again swept in the same manner. After several iterative calculations with different base-emitter forward voltages, we can find the collector-to-emitter breakdown voltage with the base terminal open as a function of collector-to-emitter current. Fig. 2.22 displays such relations for the four cases. The BV_{CEO} is defined at the collector-to-emitter current of 0.56 μ A which corresponds to 1 μ A when an effective emitter area is 0.2x1.8 μ m^2.

Table 2 summarizes the calculated results of cutoff frequencies (f_T), ac base resistance ($r_{bb'}$), base-collector capacitance (C_{cb}) and BV_{CEO}.

Table 2.2 Calculated device parameters and BV_{CEO}

Pedestal I.I.	f_T (GHz)@V_{CE}=1V	$r_{bb'}$ (Ω)	C_{cb} (fF)	BV_{CEO} (V)
none	16.2	609.7	0.31	5.7
150KeV	22.0	707.9	0.58	2.2
300KeV	28.6	614.7	0.65	1.8
400KeV	24.0	600.5	0.52	2.7

2.4.4 Favorable Process Condition for Pedestal Collector

Generally speaking, there are two choices for the ion-implanting condition to realize a highly-doped pedestal collector. One is a counter doping technique to cut the base tail caused by channeling effect. The other is a rather deeper doping technique to obtain a gradient collector so as to prevent mainly the Kirk effect. According to the cases analyzed for this time, the 150-KeV case is the former, and the 400-KeV case is the latter. In the case of the counter doping technique, the base width is much more reduced to cause the base Gummel number (effective hole numbers in the base per unit area)

decrease than other cases. Therefore, not only the base resistance increases, but also the current gain increases drastically as shown in Fig. 2.21. This means there is less process latitude for BV_{CEO}. On the contrary, in the case of the gradient collector doping, the base widths are not so much reduced. Consequently, the current gain and the base resistance are not excessively increased compared with the control device. This is a promising condition to keep enough process latitude against the reduction of BV_{CEO}.

Fig. 2.23 demonstrates the relationship between cutoff frequency and BV_{CEO} when the phosphorus-ion-implanting conditions for the collector are changed. If the dosage is decreased, the BV_{CEO} is increased while the cutoff frequency is decreased. We can easily find that the 150-KeV case is more sensitive to the reduction of BV_{CEO} for higher cutoff frequency. If the minimum BV_{CEO} of 3 V is required, the 400-KeV case is the best one with the highest cutoff frequency and moderately low ac base resistance as well as small base-collector capacitance.

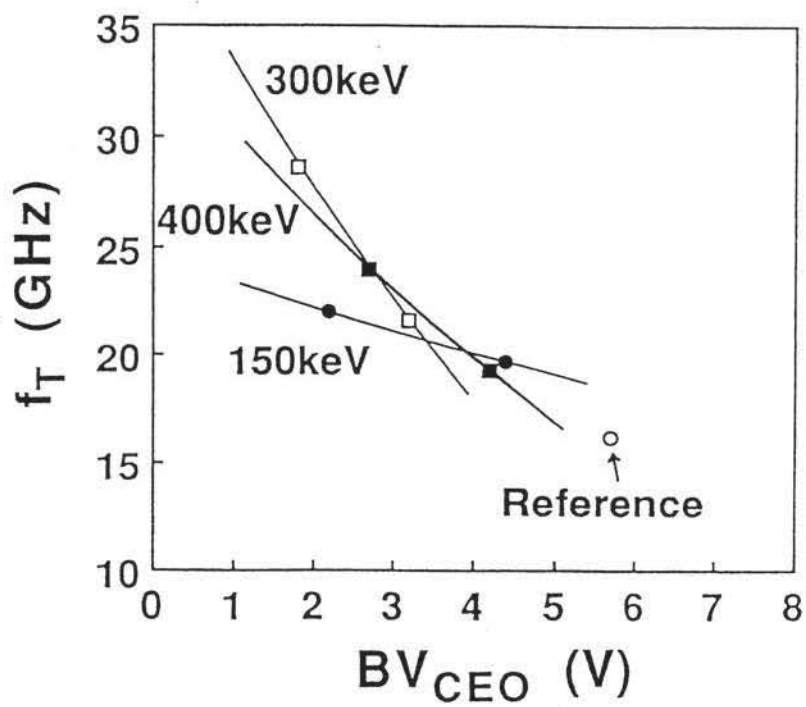


Fig. 2.23 Relationship between BV_{CEO} and f_T for different phosphorus-ion-implanting conditions. Dose conditions; 150KeV with $2\times10^{12}/cm^2$, $1\times10^{12}/cm^2$; 300KeV with $1\times10^{13}/cm^2$, $2\times10^{12}/cm^2$; 400KeV with $1\times10^{13}/cm^2$, $2\times10^{12}/cm^2$.

2.4.5 Application to ECL Circuits

By employing the optimizing method described above, we have tuned experimentally our 0.5-μm-minimum-design-ruled double polysilicon self-aligned process called ESPER [40] by using ECL circuits including ring-oscillators and 1/8 static frequency dividers. As a result, we have obtained an ECL gate delay time of 30.3 ps/gate at a switching current of 0.52 mA for low-power gates with effective 0.2x1.8-μm^2-area emitters. In the case of high-power gates, which correspond to effective 0.2x4.7-μm^2-area emitters, a gate delay time of 27.2 ps/gate at a switching current of 0.97 mA, and a minimum gate delay time of 26.3 ps/gate at a switching current of 1.64 mA have been obtained. At the same time, a 1/8 static frequency divider using T-F/F has been observed to operate up to a maximum frequency of 15.8 GHz [82], that was the highest value in 1990.

2.4.6 Conclusion

By using a two-dimensional process/device simulation system, process conditions of highly-doped pedestal-collector transistors for well-balanced device parameters and high enough BV_{CEO} have been analyzed. As a result, a gradient-collector-like profile has been proved to be superior to a counter doping profile to cut a base channeling tail from the viewpoint of high enough cutoff frequency, lower base resistance and smaller base-collector capacitance with the minimum BV_{CEO} of 3 V. By applying this optimization method to practical ECL circuits, a minimum ECL gate delay time of 26.3 ps/gate at a switching current of 1.64 mA has been obtained experimentally. At the same time, a 1/8 static frequency divider has been observed to operate up to 15.8 GHz.

Chapter 3 Performance Estimation of High-Speed Bipolar Device

3.1 Introduction

Numerous modern silicon BJT technologies have been investigated and discussed in the previous chapter. In this chapter, the performance estimation of silicon BJTs, not only from the process point of view, but also from the circuit designing point of view will be discussed, and will be clarified the problems of high-speed bipolar devices for larger scale integration circuit.

There have been extensive circuit options for silicon bipolar ICs such as S-TTL, IIL, CML, ECL etc. However, the advent of BiCMOS technology, which was aimed at achieving higher speed operation than CMOS and less power dissipation than TTL or ECL, displaced the S-TTL bipolar digital ICs from its major position in the bipolar digital IC market. While the market share of S-TTL or IIL digital ICs decreased because of their poor cost-performance figure of merit, only ECL devices seemed to pile up an increasing market share. Until the end of the 1980s, the demand for silicon bipolar digital ICs focused on higher speed devices with higher level of integration or less cost than GaAs ICs. These high performance features were achieved by ECL and ECL-like circuits.

3.2 Strategy to Estimate ECL-Circuit Performance

The advantages of ECL and ECL-like circuits are the high-speed non saturating operation of current switches and the high driving capability of emitter follower stages. However, their large power dissipation is a most serious disadvantage. The switching delay time (t_{pd}) of ECL circuits depends on the power dissipation or the switching current per gate as showing Fig. 3.1. The delay t_{pd} is roughly expressed by the following equations:

$$t_{pd} = R_L C_c + k_1 \tau_f + r_b C_b \tag{3.1}$$

$$C_c = k_2 C_{jc} + k_3 C_{cs} + k_4 C_L + k_5 C_{EF} \tag{3.2}$$

$$C_b = k_6 C_{de} + k_7 C_{je} + k_8 C_{jc} \quad , \tag{3.3}$$

where R_L is the pull-up resistor, τ_f is the transit time, r_b is the base resistance, C_{jc} is the base-collector capacitance, C_{cs} is the collector-

substrate capacitance, C_L is the parasitic wiring capacitance, C_{EF} is the input capacitance associated with the emitter follower transistor, C_{de} is the diffusion capacitance, C_{je} is the base-emitter capacitance and k_1 to k_8 are weighting constants. Generally, as shown in Fig. 3.1, at lower power dissipation, the first term of eq. (3.1) is the most dominant factor limiting the delay, and at higher power dissipation, the third term plays a more important role than others. The minimum achievable delay time is restricted by the second term of eq. (3.1). Larger scale integration requires lower power dissipation per gate to keep total power consumption per chip at a practical level. Therefore, low switching current per gate is accordingly needed, and a large pull-up resistor is required if the voltage swing is fixed at a constant value to keep enough noise margins. Consequently, the first term, or the *RC* time constant associated with the pull-up resistor, should be decreased to achieve higher speed operation. For medium and small scale integrated circuits, the first term as well as the other two terms should also be reduced to obtain shorter delay times.

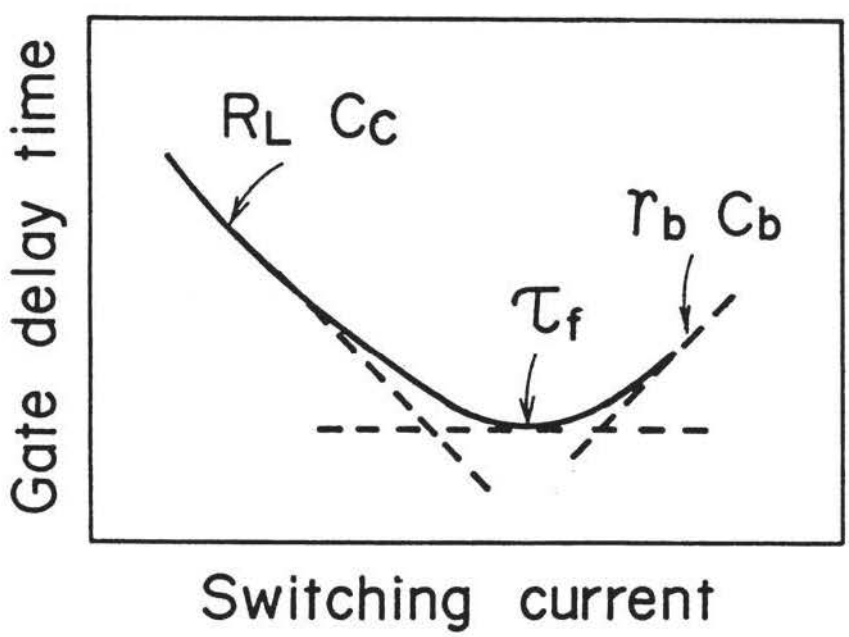


Fig. 3.1 Gate delay time versus switching current per gate (arbitrary unit).

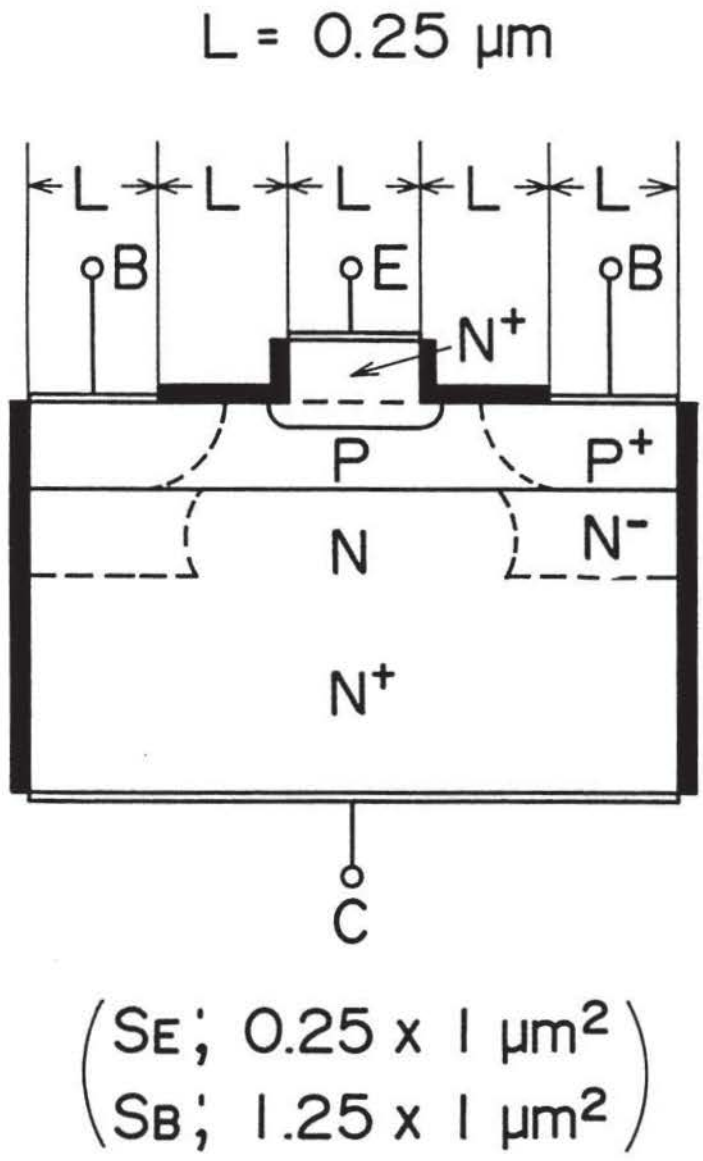


Fig. 3.2 Schematic cross-section of the analyzed transistor.

The most important matter when optimizing the transistor structure for ultimate high speed operation is to evaluate quantitatively the contribution of the *RC* time constants appearing in eq. (3.1). Nevertheless, because of the nature of *RC* time constants, eq. (3.1) is not accurate when they are closely

related to one another. Additionally, not only the base resistance, but also the diffusion capacitance and the base-collector capacitance are non-linearly dependent on the switching current. Therefore, the most reliable way to estimate the switching delay time is direct numerical calculation by physical simulation of circuit performance without using an analytical circuit model. Now, will be described some results obtained with such a numerical technique to estimate the ultimate performance of silicon bipolar ECL circuits. Figure 3. 2 shows the schematic cross-section of the simulated transistors. Because two-dimensional device simulator has been used, a 0.25×1.0-μ m^2 walled-emitter transistor, suitable for two-dimensional analyses, has been investigated. Figure 3.3 shows the impurity profiles corresponding to the cross-section in Fig. 3.2. The impurity profiles were obtained by a process simulator. The polysilicon emitter is treated as a 0.1-μ m-thick extended-single-crystalline emitter, where the maximum ionized impurity concentration is limited by 1×10^{20}cm $^{-3}$. Figure 3.3(a) displays nearly the same profile as has been obtained experimentally by the selectively-implanted-pedestal-collector process described in the previous chapter. The calculated peak cutoff frequency is 24 GHz at a collector current of 0.18 mA as shown in Fig. 3.4 and the C_{jc} is 0.53 fF while the ac base resistance is 400 Ω at low collector current. Figure 3.3(b) indicates rather shallower base with a gradient collector profile. The calculated peak cutoff frequency is 56 GHz at a collector current of 0.32 mA as shown in Fig. 3.5, while the C_{jc} is 1.1 fF and the ac base resistance is 570 Ω at low collector current. Figure 3.6 indicates the circuit diagram used to estimate the delay time. In Fig. 3.6, the terminal current-voltage characteristics of transistors Q_1 and Q_2 are directly obtained as a result of physical two-dimensional device simulation using a transient analysis mode. Other circuit parameters, such as R_L, C and R_B are introduced as lumped values. The pull-up resistor R_L is defined to fix the switching current. C contains all of the parasitic capacitance associated with the pull-up resistor such as the collector-substrate capacitance, the wiring-substrate capacitance, the resistor-substrate capacitance and the junction capacitance related to the emitter follower transistor. Finally, the R_B is base-contact resistance. The voltage swing is set to be 0.4V and the average delay time of NOR and OR logic gate outputs is defined as the gate delay.

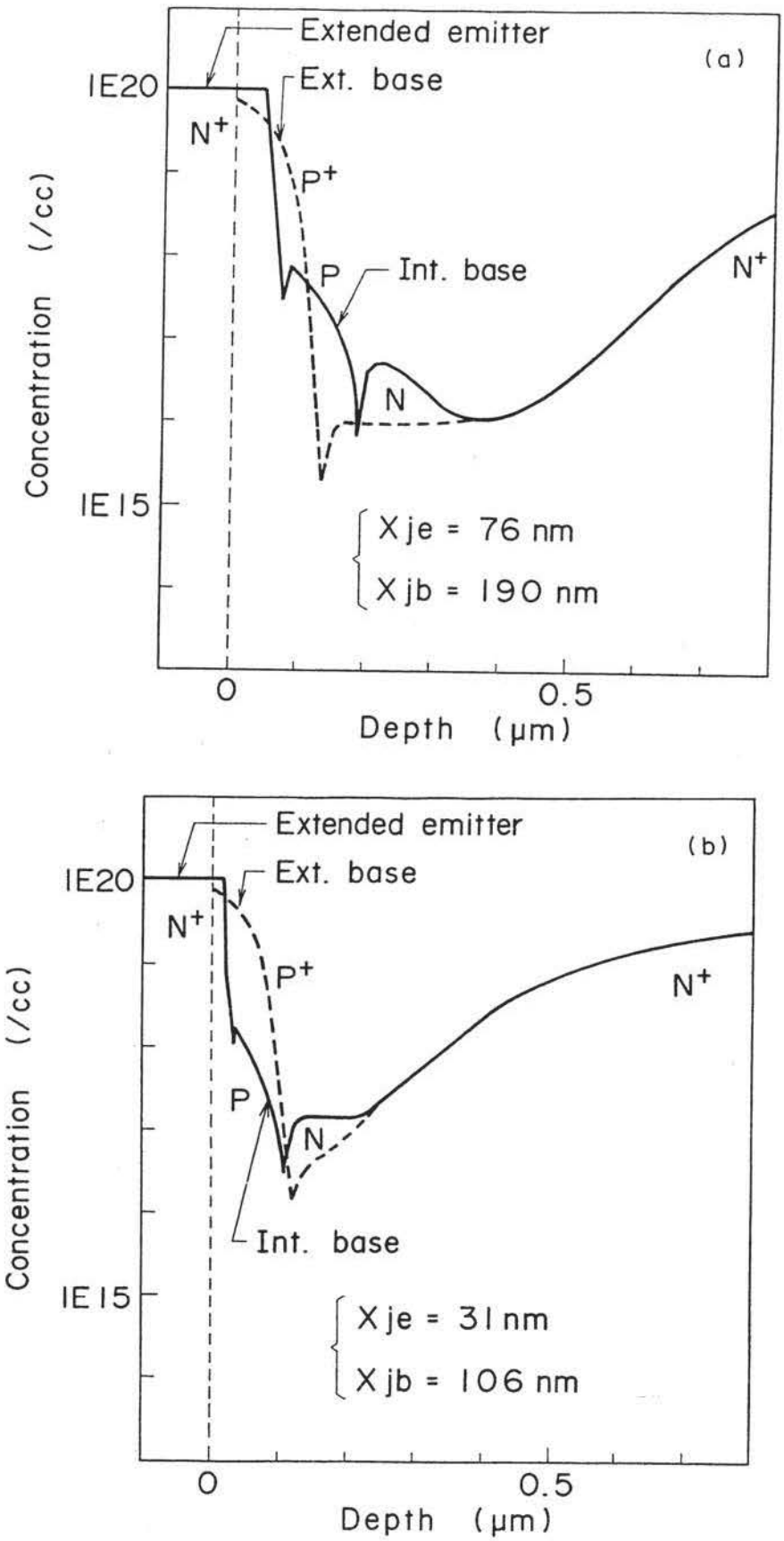


Fig. 3.3 Impurity profiles for the analyzed transistor.

(a) Selectively implanted pedestal collector profile. (b) Gradient collector profile.

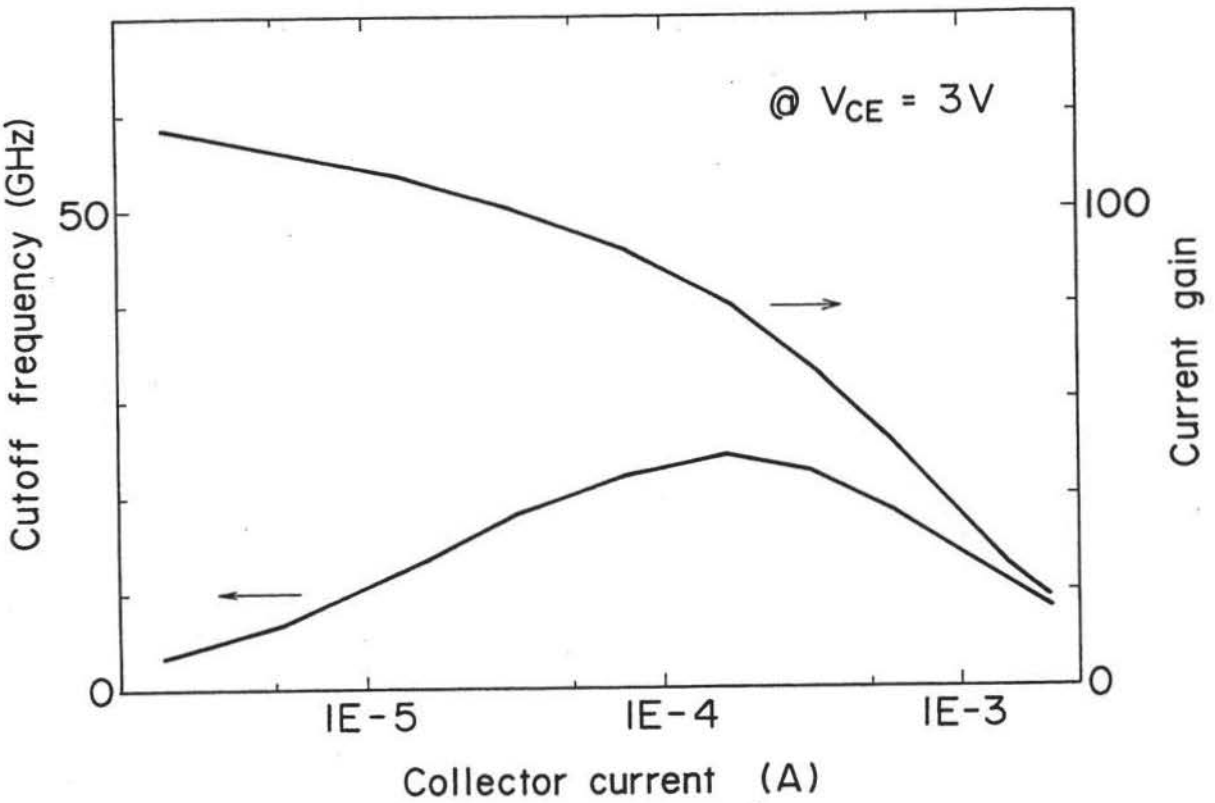


Fig. 3.4 Cutoff frequency and current gain versus collector current (pedestal collector).

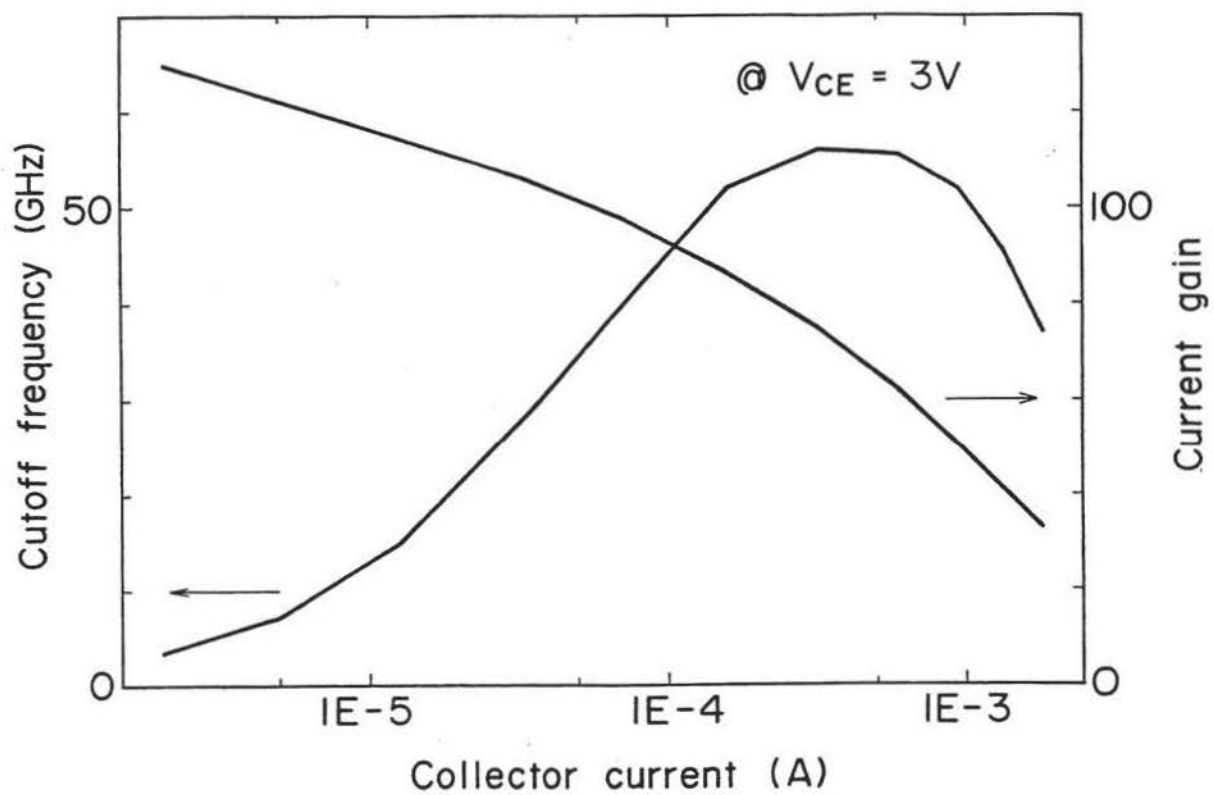
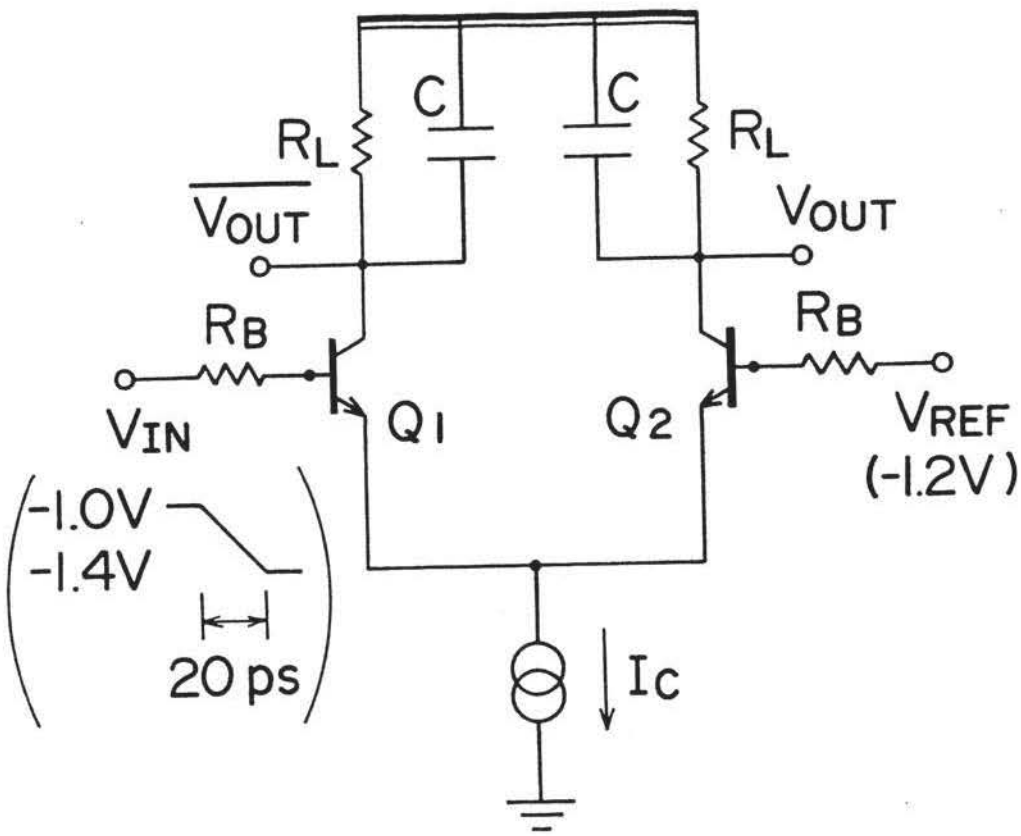


Fig. 3.5 Cutoff frequency and current gain versus collector current (gradient collector).



- Transistor Q_1, Q_2;
  2-D physical device simulation

- R_L, R_B, C ;
  Lumped circuit parameters
  - R_L; Pull-up resistor
  - R_B; Base contact resistance
  - C ; All parasitic capacitances associated with R_L (Sum of C_{CS}, C_L, C_{EF} etc.)

Fig. 3.6 Analyzed circuit diagram.

3.3 Optimization of ECL-Circuit Performance

Figures 3.7 and 3.8 show the numerically calculated results according to the two impurity-profile cases. In Fig. 3.7, the basic gate delay time dependence on the switching current is displayed when the total parasitic capacitance C and the lumped base contact resistance R_B are changed as parameters. The modern microfabrication technique can make the total parasitic capacitance C as small as 15 fF, with C_{cs}=3fF, C_{EF}=4fF and C_L=8fF. Because the base contact resistance is estimated at nearly $50\,\Omega\,\mu\,m^2$, the result, in the case of C=15fF and R_B =100 Ω, can be recognized to indicate the realistic performance limitation; that is sub-40 ps/gate at lower power dissipation and 20～25 ps/gate at higher power dissipation. The noticeable fact revealed in Fig. 3.7 is that the gate delay is strongly dependent on the parasitic capacitance C.

These results cannot only be derived from eq. (3.1). The base contact resistance R_B contributes to decreasing the delay time at higher current level, as shown in Fig. 3.1. If the parasitic capacitance C is reduced to zero, even sub-10 ps delay time can be obtained by the 24-GHz-peak-cutoff-frequency transistors. Of course, this situation can not be realized in practical devices because the emitter follower transistors are always coupled to the current switches. However, this result proves that if C is reduced from 15 fF to 5 fF, the delay time is remarkably improved by about a factor of two. Figure 3.8 shows similar results by the transistors with the shallower base, and with a much higher peak cutoff frequency than that in Fig. 3.7. The gate delay versus switching current characteristics are shown for various C values. When compared with Fig. 3. 7 at the same C value, (e.g. 5 fF), there is not so much difference in delay times. Only at higher current levels, can the high-cutoff-frequency transistor improve the delay, but the delay-time improvement is a few picosecond, although the cutoff frequency is improved more than twice. On the contrary, at the lower current level, the high-cutoff-frequency transistor shows greater delay, which may be a trade-off caused by the larger C_{jc}. As shown in Figs. 3. 7 and 3. 8, the most crucial parameter affecting the gate delay time is the total parasitic capacitance associated with the pull-up resistor. Surprisingly, the contribution of the peak cutoff frequency value of the switching transistors is not very large, even at the higher current levels.

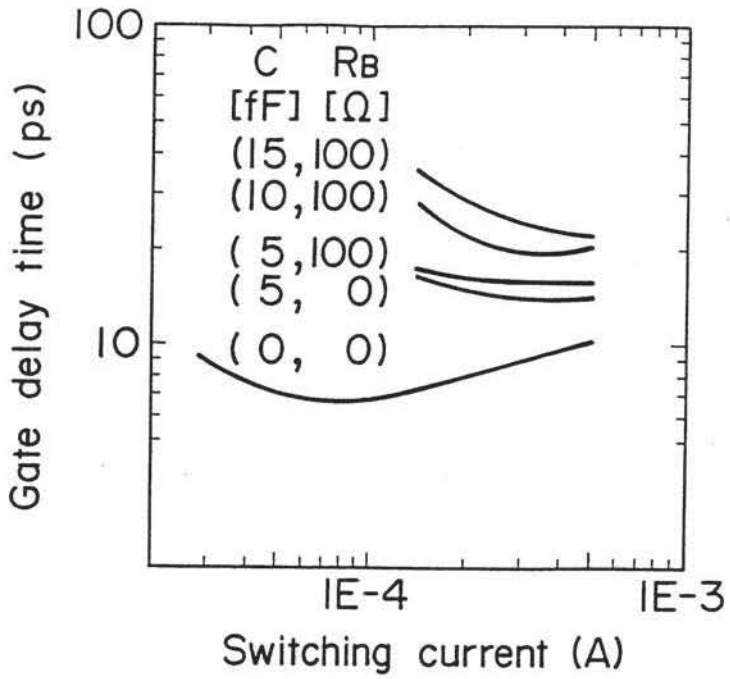


Fig. 3.7 Gate delay time versus switching current per gate (pedestal collector).

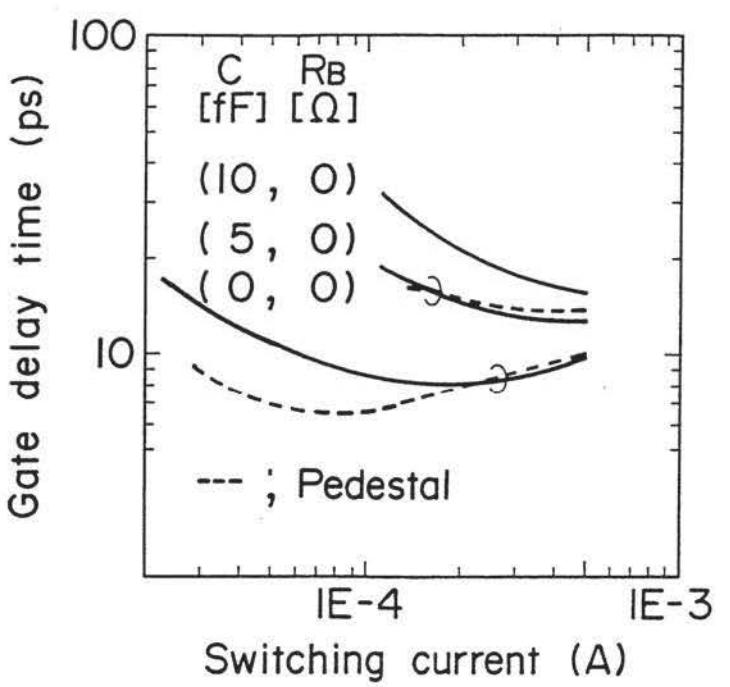


Fig. 3.8 Gate delay time versus switching current per gate (gradient collector).

3.4 Favorable Silicon BJT Features and Its Limitation

As discussed in section 3.3, the most crucial parameter affecting the gate delay of ECL circuits is the parasitic loading capacitance associated with the pull-up resistor. A high cutoff frequency requires shallower base and higher collector impurity concentration, which cause higher intrinsic base resistance and larger base-collector capacitance. Hence, the device parameters corresponding to the active transistor region (e.g. the base-collector capacitance, base resistance and cutoff frequency) should be moderately optimized as shown in the last section of the previous chapter. Moreover, other parasitics such as contact resistance for the emitter or the base should also be kept small enough not to affect the gate delay. The current gain is no longer a figure of merit limiting circuit performance, even if an extended-single-crystalline emitter is used as shown in Figs. 3.4 and 3.5. The single-crystalline or epitaxial emitter is also advantageous to reduce the emitter series resistance, which often originates in the polysilicon emitter structure. The high emitter series resistance should be avoided to maintain sufficient transconductance for proper circuit operation.

By taking account of the properties discussed, the favorable silicon BJT structure can be shown in Fig. 3.9 [81,83]. To reduce the parasitic capacitance simultaneously associated with the pull-up resistor and the substrate, the transistor should be fully isolated by trenches and an SOI structure. Active transistor regions are surrounded by thick insulators, and field regions are also covered by thick insulators such as oxide. To reduce the parasitic resistance associated with emitter and base, the contact area is covered by silicides or low-resistance refractory metals. The emitter should also be single-crystalline extended emitter or epitaxial emitter. The vertical profile for the intrinsic base regions are optimized to achieve a reasonable trade-off between the cutoff frequency, the base-collector capacitance, and the base resistance, avoiding punchthrough and forward-biased tunneling with enough collector-to-emitter breakdown voltage. At the same time, the lateral dimension is optimized to avoid narrow emitter effect, peripheral forward- or reverse-biased tunneling. Because the circuit delay is no longer limited by the performance of transistor, but rather by external parasitics, the practicality of process improvements is an important matter. SOI is also applicable to bipolar

LSIs now that a direct wafer bonding and thinning technique are available [84]. Therefore, the transistor shown in Fig. 3.9 is based on the improvement of double-polysilicon self-aligned transistors. These features are now demonstrated by several researchers including the fully isolated SOI structure, especially for mainframe-processor CPU LSIs [85], and optical terminal ICs [86]. Other options, such as silicon-based heterojunction BJT's (HBT's), including wide gap emitter and narrow gap base, have been investigated. Recently, practical process to achieve narrow gap base by selective-epitaxial growth of a silicon-germanium strained-layer in the SST-like structure has successfully been found as mentioned in the previous section. However, we must keep in mind that Si-HBT itself cannot reduce the parasitic capacitance associated with the pull-up resistor.

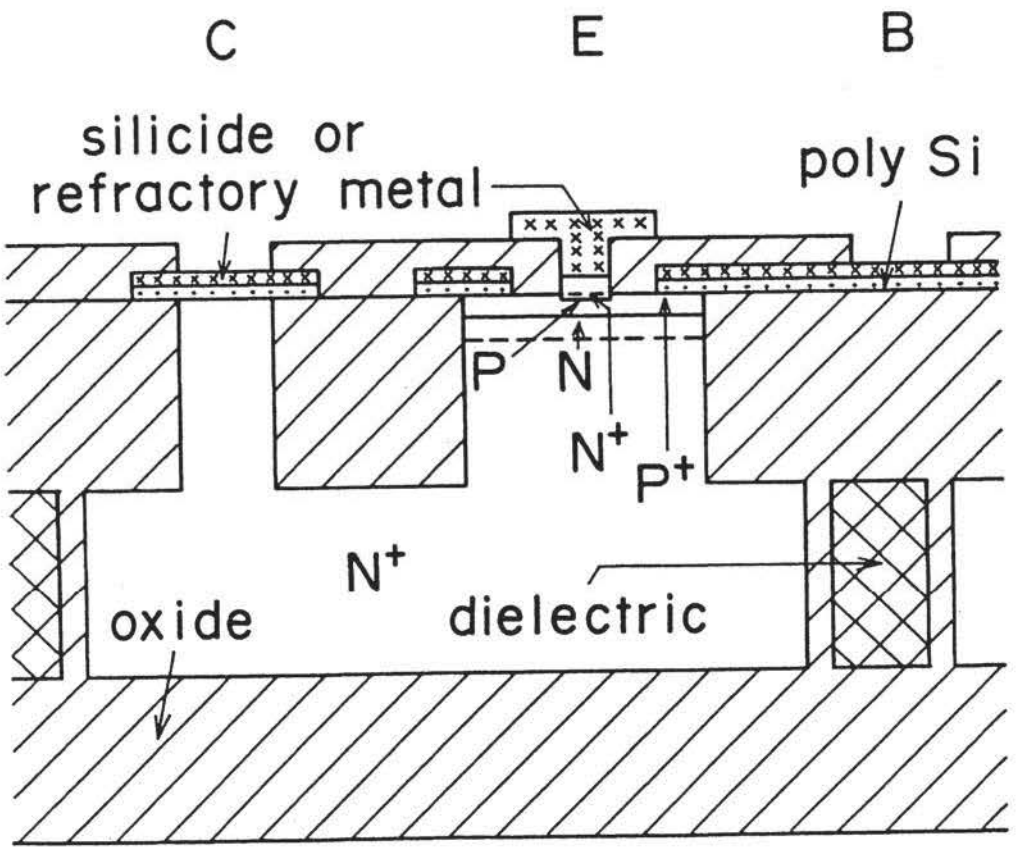


Fig. 3.9 Schematic cross-section of the favorable BJT structure.

In addition to process improvements, circuit techniques must also be enhanced. One example of such an approach is the so-called active pull-down for emitter follower circuits [87]. A basic gate delay of 23 ps at a power dissipation of only 2.1 mW/gate has been demonstrated based on a trench isolated, double polysilicon, self-aligned transistor process. The basic gate delay of ECL circuits is now less than 20 ps. Such a basic delay is not very important, and the total delay of circuits with loaded capacitance or the system delay is much more relevant. To achieve higher speed operation of chips that should be considered as a system, the important thing is the well-balanced optimization of process technology, device structure and circuit designs.

So far as ECL and ECL-like circuits are concerned, the improvement of the process technology as well as active pull-down emitter follower techniques can not reduce the switching current itself. Even when the loaded capacitance is reduced due to the smaller feature size brought by sophisticated lithography tools, minimum switching current of around 100 μ A is at least required as shown in Figs. 3.7 and 3.8. Therefore, if the chip contains more than one-million gates, it dissipates 100A in total current, which is far beyond the limit of practical air-cooling.

Non-saturating operation of BJT with continuous current flow is the advantage of ECL circuit because this can avoid the large time constant caused by saturation-mode operation. Ironically, this advantage directly reflects the major drawback to ECL circuits, or the large power dissipation. Today's microfabrication technology has brought us deep submicron era, and a full system-on-a-chip concept is theoretically achievable. However, for high-speed bipolar LSIs based on ECL circuits, this transition could no longer afford any benefit.

3.5 Conclusion

To extract the high-speed performance of bipolar LSIs, ECL and ECL-like circuits are the most suitable compared to other circuit architectures. As far as the integration level is not very large, this advantage covered the problems of higher power dissipation. However, when millions of gates on a chip can be available due to the dramatic improvement of microfabrication

technology, the large power dissipation is no longer escaped. The high-speed bipolar LSIs have not been able to catch up the system-on-a-chip concept.

Chapter 4 Design Concept for High-Speed CMOS Device

4.1 Introduction

Historically, the concept of the field effect transistor was much older than that of BJT. Basic patent was already issued in the 1930s. However, a realistic device never came to industry until the beginning of the 1970s. The understanding of the physical mechanism of the surface states of silicon made possible MOSFET for 1kb *d*ynamic (D)RAM, which was developed by Intel. Because of a contamination problem caused by mobile ions such as Na or K, the first MOSFET was based on a p-type channel structure. Great efforts were paid to conquer this problem, and by the end of the 1970s, n-type channel MOSFETs were available for practical devices such as DRAMs. The n-type channel can benefit high mobility of electron when compared to that of hole in the p-type channel, and the device performance was improved in accordance with the progress of optical lithography so that nMOS LSIs rapidly replaced pMOS LSIs.

As MOSFETs are surface device, they do not require full isolation between neighboring devices, and this situation is advantageous from the packing-density point of view. In the 1980s, silicon bipolar, nMOS and CMOS LSIs were in a quite competitive status for realization of very large scale integration (VLSI). Among them CMOS device was also ideal from the power-dissipation stand point, because it dissipates power only when switching. However, CMOS gate must use both nMOSFET and pMOSFET to make even a basic inverter gate, that was thought to be a disadvantage concerning packing density against nMOS-only devices. In the 1980s, developments of MOS LSIs were focused on larger integration in accordance with the progress of microfabrication technology. The nMOS LSIs realized higher packing density and larger integration level than bipolar LSIs, but they also began to suffer from the problem of larger power dissipation. Therefore, by the end of 1980s, CMOS devices took over the nMOS position and achieved the major position of VLSI. At that time, submicron gate length was realized for MOSFETs in CMOS VLSIs.

At the beginning of the 1990s, the transition occurred even in the area of high-speed LSIs. High-speed CMOS LSIs appeared in a quite competitive

situation against bipolar LSIs. One was a 90 MHz CMOS-RISC processor [88], and the other was a CMOS standard cell LSI [89]. The former realized the same clock frequency as that of the ECL-based RISC processor at that moment, and the latter achieved circuit delay, even with a heavy loaded capacitance, which was almost the same as that of the sophisticated ECL gate array [57]. These high-speed features were mainly achieved by combining submicron lithography and channel length optimization in their process technologies. CMOS LSIs began to benefit the fruit of microfabrication technology progress not only in larger integration but also in higher speed operation. At the same time, even in the field of mainframe computer, the advantage of CMOS device against bipolar ECL devices was suggested for the central processing unit (CPU) LSIs [90]. Now, the CMOS device has taken the major place in almost any field of ultra large scale integrations (ULSIs).

4.2 Requirements for High-Speed CMOS Device

As CMOS replaces ECL devices in the high-speed area, more than one-million gate circuits for microprocessor with 500MHz clock cycle have been realized in the mid 1990s. Although CMOS devices dissipate minimum power, such a large scale and high-speed LSI consumes ac current of tens ampere. Hence, power dissipation problems are now becoming serious even in the case of CMOS devices. Moreover, loaded gate delay, especially wiring delay, is also becoming main concern just like the situation that once bipolar devices faced in the end of the 1980s.

To achieve lower power dissipation for CMOS circuits, the most effective way is to reduce supply voltage because the power is in proportion to the square of the voltage swing. For this purpose, the threshold voltages of both nMOSFET and pMOSFET should also be decreased. At the same time, these absolute values of threshold voltages should be the same from the circuit-design point of view. This means the technological freedom of the threshold voltage design especially for lower values is important. Besides, to save the power dissipation, any parasitic capacitance associated with the drain junction as well as source and gate series resistances should be reduced. For such purposes key features for high-speed CMOS ULSIs are summarized as below:

a) Surface channel dual-gate n/pMOSFETs,
b) Retrograde well structure,
c) Salicide (*s*elf-*a*ligned si*licide*) for source/drain/gate,
d) Shallow trench isolation.

By using the surface channel MOSFET as well as the retrograde well structure, the threshold voltage can be reduced easily and independently designed from the well structure sustaining lower well resistance to prevent a latch-up problem. In the case of the retrograde well, the doping concentration increases in the depth direction and a highly doped buried-layer-like profile is easily obtained. The surface channel pMOSFET can achieve lower threshold voltage with a less short channel effect. To this end, the dual-gate process, or the process using n-type polysilicon gate for nMOSFETs and p-type polysilicon gate for pMOSFETs is favorable. The salicide process based on $CoSi_2$, which uses cobalt-and-silicon sintering process after sidewall spacer formation without any mask step, can reduce the parasitic resistance not only on the source/drain contact regions but also gate electrode even in the sub-0.1 μ m regime. $TiSi_2$ has been used for salicide process, but it suffers higher resistance when the width of the silicide region approaches subquarter micron because the $TiSi_2$ crystalline phase transition from the higher resistance *C49* to the lower *C54* becomes unstable and cannot be controlled well. The shallow trench isolation not only can achieve higher packing density than that of LOCOS but also can reduce parasitic junction capacitance originated by the lateral diffusion of highly doped source/drain toward the isolation region. Moreover, it can realize better planarized surface topography than that of LOCOS, which is favorable to keep enough depth-of-focus (DOF) margin for optical lithography process. Figure 4.1 shows the typical cross-sectional structure of a dual-gate CMOS device with $CoSi_2$ formed by the salicide process. Inter-layer dielectric and W-filled contact plug as well as local interconnect wiring are shown in the same figure.

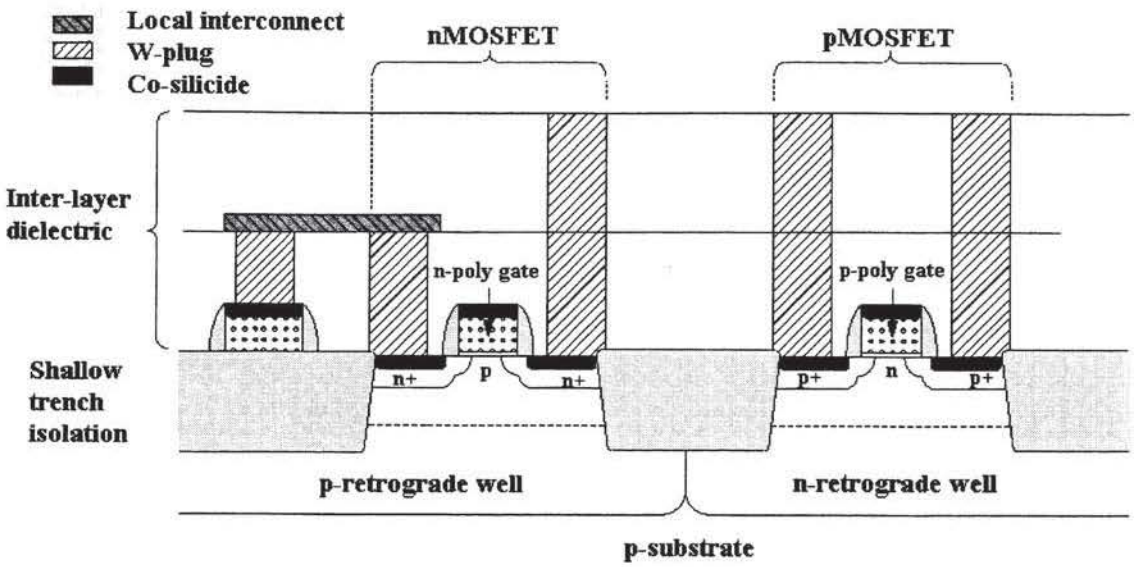


Fig. 4.1 Cross-sectional structure of dual-gate CMOS device with $CoSi_2$ after contact plug formation.

4.3 Quarter and Subquarter Micron CMOS Technologies

In this section, the quarter and subquarter micron CMOS device technologies for high performance logic ULSIs are described.

4.3.1 Quarter Micron CMOS Technology

Retrograde twin-well and dual-gate surface channel CMOS device was demonstrated based on Co-salicide process [91]. In this case LOCOS isolation was still used. KrF excimer lithography tool (optical wavelength:248nm) was used to realize minimum gate length of 0.24 μ m. Because the minimum pattern feature is almost the same as the wavelength, the DOF can be kept enough to allow rather uneven surface topography caused by LOCOS isolation. In this technology, also employed local interconnect using thin TiN layer directly deposited on the Co-silicide and field oxide layer. By using this local interconnect technique for cross-coupling of embedded SRAM cell, 9.9-μ m^2 cell area has been realized [91]. This SRAM has achieved less than 2ns access

time [92]. The quarter micron CMOS technology has realized ECL-compatible high-speed and higher-density ULSIs.

4.3.2 Subquarter Micron CMOS Technology

Based-on the quarter micron CMOS technology, it has been up-graded to the minimum design rule of 0.2 μ m [93]. As the same KrF optical lithography tool was employed, to keep enough DOF margin, shallow trench isolation technique has been introduced. Other process technologies are basically the same as those of the quarter micron technology. As an option, the first metal-like local interconnect technique was also introduced instead of direct deposited TiN layer on the silicide region as shown in Fig. 4.1. This is also intended to keep enough DOF for the local-interconnect mask patterning using the subquarter micron design rule. As a result the embedded SRAM with the cell size around 4 μ m^2 has been realized with the access time of 1.4ns [93].

4.4 Conclusion

CMOS device has replaced bipolar devices even in the field of high-speed LSIs. Although CMOS device is ideal from the power-dissipation point of view, it has begun to suffer from larger power dissipation and wiring delay problems brought by ULSI, which were once suffered by bipolar LSIs.

Chapter 5 Inverse Modeling Technique for MOSFET Design

5.1 Introduction

In order to limit the development costs involved in deep submicron CMOS technology, simulation techniques are used to reduce the number of test wafer cycles. Unfortunately, deep submicron CMOS technology is too complex to be accurately modeled using the available process simulators. Quantitative simulation of the reverse short-channel effect, for example, is still difficult. Furthermore, process simulators contain many fitting parameters, which further limit the predictive capabilities of these programs. On the other hand, device simulation has recently attained a high level of reliability, and it presents a detailed picture of the carrier transport processes within a device which cannot be obtained by experimental means. Optimum utilization of device simulation must be based on an accurate device structure and doping profile. Since process simulation cannot always provide this, inverse modeling is used to complement process simulation [94-98]. In this case, physically precise device models are used to extract the device structure as well as the active impurity profiles from measurements on a specific device. In MOSFETs, capacitance versus voltage (*C-V*) characteristics and drain current versus gate voltage (*Id-Vg*) characteristics in the so-called linear device operation regime at low drain voltage ($Vd < 0.1$V) are used because they can be accurately modeled using the Poisson equation and the drift-diffusion (DD) model, respectively. The device structure and doping profile are obtained by varying the initial structure until the simulated *C-V* and *Id-Vg* characteristics match the experimental ones. The main drawback to inverse modeling is that it is based on experimental data of an actual device and cannot predict a device structure. Thus it is not possible to optimize a device structure before or even after its technological realization if we only have limited experimental data. This disadvantage of inverse modeling is partially overcome by means of the proposed inverse modeling technique.

In the next section, our inverse modeling program is described, and the procedure for extracting the structures of deep submicron MOSFETs is explained in section 5.3. In section 5.4, device simulation results based on inverse modeling are discussed, and in section 5.5 how to vary the device

structure for different process parameters without further inverse modeling is described.

5.2 Inverse Modeling Procedure

Figure 5.1 shows the flowchart of the inverse modeling assisting program named *MOSFIT* using the commercially available GALENE-III as the main device simulator. GALENE-III consists of not only the standard DD model but also the generalized hydrodynamic model (GHDM) [99]. The inversion layer mobility models implemented are the semiempirical approaches proposed by Lombardi et al. [100] and Mujtaba et al. [101], which are sufficient for the investigated range of channel doping concentrations. Since all simulations for the inverse modeling are performed in the linear regime, the somewhat uncertain value of saturation velocity in the channel and the approximations of the DD model do not influence the results [102,103].

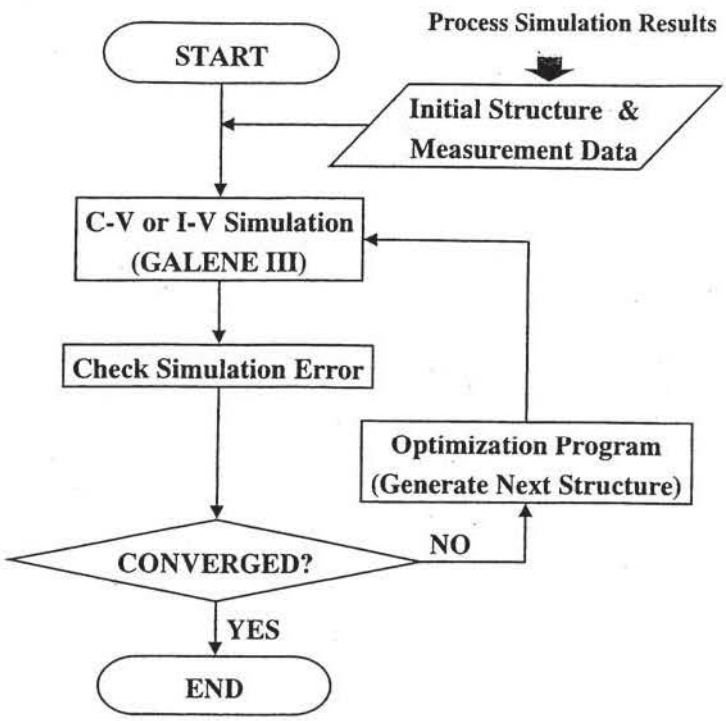


Fig. 5.1 Flowchart of the "MOSFIT" program.

The device structure defined in terms of doping profiles, oxide thickness and so on, is specified in the input file of GALENE-III, which is generated by the MOSFIT program. A two-dimensional doping profile is expressed in terms of a combination of step, Gaussian and error functions, avoiding complicated 2-D spline functions suggested by others [94]. The one-dimensional channel doping $N_{ch}(x)$ is defined as

$$N_c(x) = N_c \exp\left\{-\left(\frac{x - x_c}{D_{c1}}\right)^{Ec1}\right\} \qquad (x \le x_c) \tag{5.1a}$$

$$N_c(x) = N_c \exp\left\{-\left(\frac{x - x_c}{D_{c2}}\right)^{Ec2}\right\} \qquad (x_c \le x) \tag{5.1b}$$

$$N_s(x) = N_s \exp\left\{-\left(\frac{x - x_s}{D_s}\right)^{E_s}\right\} \tag{5.2}$$

$$N_{ch}(x) = N_c(x) + N_s(x) + N_W, \tag{5.3}$$

where $N_c, N_s, N_W, x_c, D_{c1}, D_{c2}, D_s, E_{c1}, E_{c2}, E_s$ are fitting parameters, and x_s is the coordinate of the silicon surface. The polysilicon gate doping $N_p(x)$ is defined as

$$N_p(x) = N_g \qquad (x \le x_g) \tag{5.4a}$$

$$N_p(x) = N_g \exp\left\{-\left(\frac{x - x_g}{D_g}\right)^{E_g}\right\} \qquad (x_g \le x \le x_{gs}), \tag{5.4b}$$

where x_{gs} is the coordinate of gate polysilicon/gate oxide interface and N_g, x_g, D_g, E_g are fitting parameters. The two-dimensional source/drain extension doping $N_{ex}(x)$ is defined as

$$N_1(x,y) = \frac{N_1}{2} \exp\left\{-\left(\frac{x - x_1}{D_1}\right)^{E_1}\right\} erf\left(-\frac{y \mp y_1}{R_1}\right) \tag{5.5}$$

$$N_2(x,y) = \frac{N_2}{2} \exp\left\{-\left(\frac{x - x_2}{D_2}\right)^{E_2}\right\} erf\left(-\frac{y \mp y_2}{R_2}\right) \tag{5.6}$$

$$N_{ex}(x,y) = N_1(x,y) + N_2(x,y) \quad , \tag{5.7}$$

where "*erf*" is the error function and $N_1, N_2, x_1, x_2, D_1, D_2, E_1, E_2, y_1, y_2, R_1, R_2$ are

fitting parameters.

The inverse modeling procedure using MOSFIT is described as follows. First a MOSFET device structure and its doping profiles in the input file are expressed in terms of the above given functions using a process simulator output. *C-V* or *I-V* characteristics are calculated, and the values obtained are compared with experimental results. Then MOSFIT calculates the error and the fitting parameters of those functions are changed in the input file using a free-software-based standard optimization program generating the new device structure. Then *C-V* or *I-V* characteristics are again calculated and the values obtained are compared with the experimental results. These steps are iterated until a satisfactory convergence is obtained. We ensure that the discrepancy between the measurement and simulation results is reasonably small in order to achieve a good fitting. Finally, when MOSFIT is converged, the structure file contains the inversely extracted device topography and doping profile based on the calculated electrical characteristics.

5.3 Extraction of a Deep Submicron MOSFET Structure

In order to obtain a complete MOSFET structure in the deep submicron range, the procedure described in the previous section is applied to at least three different device structures. The first device structure is a large-area MOSFET (or MOS capacitor) to extract the gate oxide thickness and the doping profile inside a polysilicon gate based on *C-V* measurement data. The second device structure is a long-channel MOSFET to extract a channel doping profile using the body effect of the subthreshold drain current versus gate voltage (*Id-Vg*) characteristics. The third device structure is a wide-channel or finger-gate MOSFET to extract the source/drain profile as well as the polysilicon gate length using split gate-to-drain/channel capacitance versus voltage (*Cgd-V*) characteristics. Table 5.1 summarizes the required devices, measurement data and extracted parameters.

Table 5.1 Extracted parameters for each device and measurement data

Device type	Measurement data	Extracted parameter
Large-area MOSFET	*C-V*	Gate oxide thickness, gate doping
Long-channel MOSFET	Body effect of subthreshold *Id-Vg*	Channel doping
Wide-channel MOSFET	*Cgd-V*	Source/drain extension doping

It is well known that quantum effects have a strong impact on the gate capacitance and that the Schroedinger equation should be solved self-consistently with the Poisson equation [104]. However, since the Schroedinger equation solver is very CPU intensive, the quantum effects are ignored in our approach. Figure 5.2 shows an example of the *C-V* characteristics for a large-area n-channel MOSFET (n-channel MOS capacitance) calculated without quantum effects (a) and with quantum effects (b). In both cases, gate oxide thickness is determined by fitting the capacitance value at the maximum voltage in the accumulation region. Good agreement between simulated results and measured data is found for both cases in the inversion region, once the doping profile in the polysilicon gate is fitted. In the classical case, the quantum mechanical effect is effectively considered in the gate doping profile and oxide thickness value. A good fit of the capacitance in the inversion region ensures that the device simulation reproduces the correct threshold voltage and inversion layer charge. On the other hand, the slight deviations observed from the measured data in the accumulation region for the classical case can be ignored because they do not affect the MOSFET on-current characteristics estimation, which is our main goal.

Figure 5.3 shows subthreshold *Id-Vg* characteristics of a 1.0-μm nMOSFET, comparing the measurement and simulation results after the channel doping profile is extracted. This good fit guarantees the channel profile accuracy over a reasonable range of depth. In Fig. 5.4 source/drain doping profile fitting is demonstrated for the finger gate nMOSFET *Cgd-V* characteristics for different substrate bias conditions.

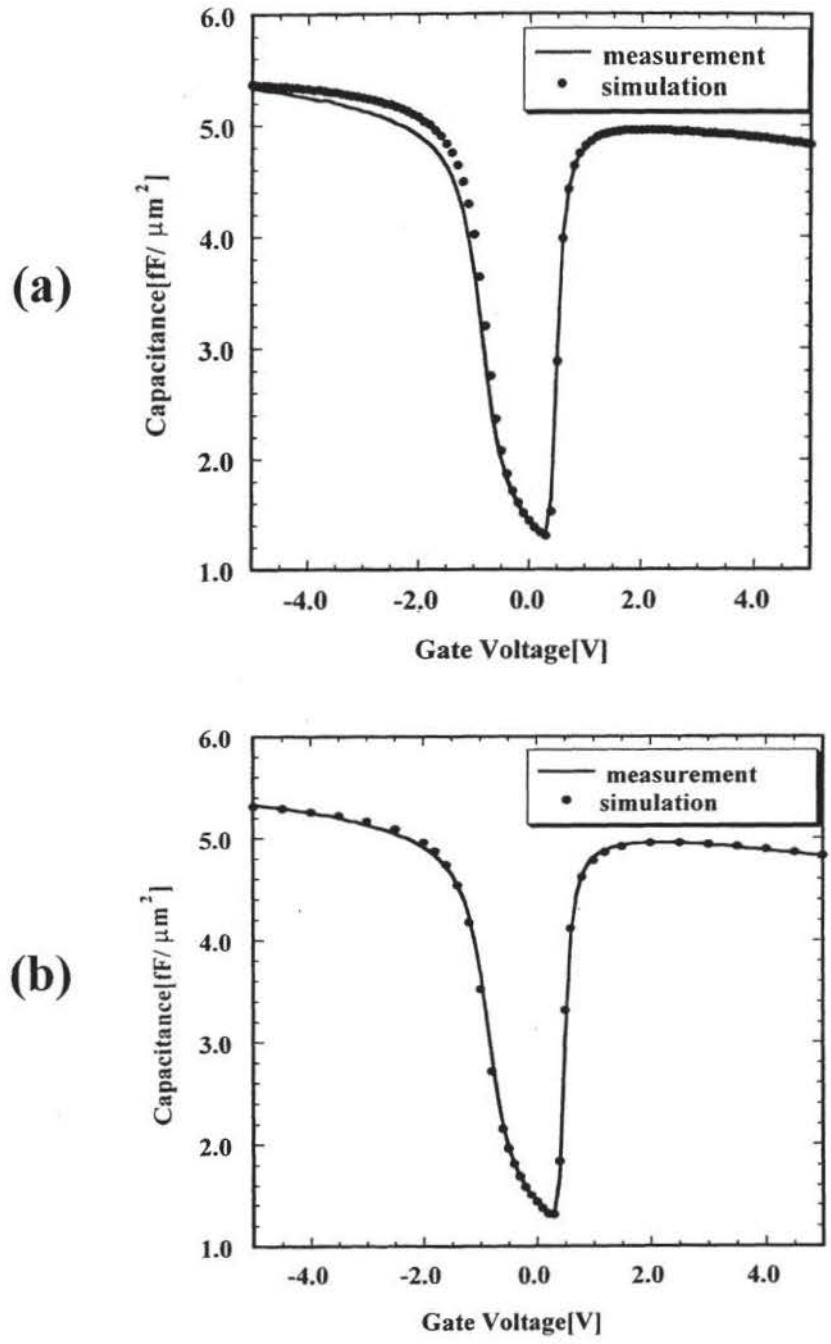


Fig. 5.2 *C-V* characteristics of a large-area nMOSFET comparing measurement and simulation results after the gate oxide thickness and doping profile of polysilicon gate are determined.

(a) Classical calculation. (b) Quantum mechanical calculation.

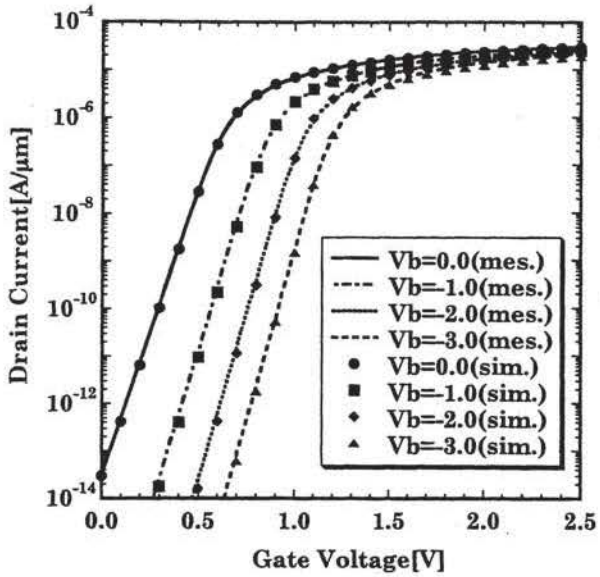


Fig. 5.3 Subthreshold *Id-Vg* characteristics of a 1.0- μ m nMOSFET comparing measurement and simulation results after the channel doping profile is extracted. (Lines: measurement results, symbols: simulation results)

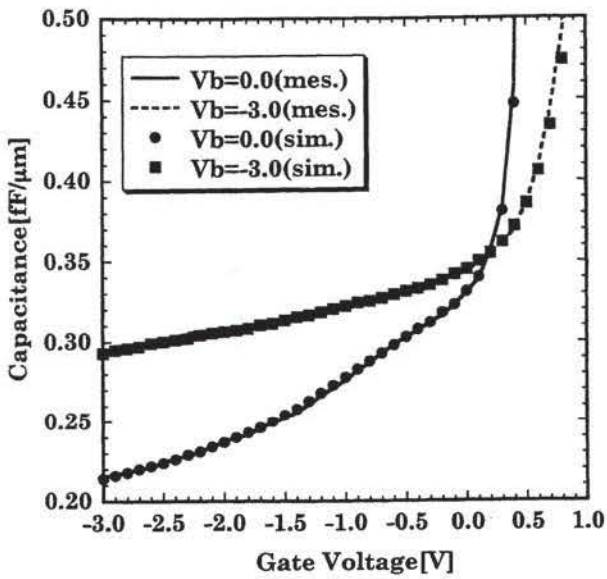


Fig. 5.4 *Cgd-V* characteristics of a finger gate nMOSFET comparing measurement and simulation results after the source/drain doping profile is extracted. (Lines: measurement results, symbols: simulation results)

In the case of deep submicron MOSFETs, it is a well-known fact that transient enhanced diffusion (TED) causes channel impurity redistribution and boron pile-up occurs just in the vicinity of source/drain junction [105]. Therefore, the channel profiles should be different from that of a long-channel transistor. By changing the channel profile to fit the body effect for each channel length, we can obtain the channel-length-dependent doping profiles. For example the subthreshold *Id-Vg* characteristics of a 0.19-μm gate length nMOSFET are shown in Fig. 5.5, demonstrating a good fit between the measurement and simulation results. The final one-dimensional channel profile in the depth direction for each nMOSFET is shown in Fig. 5.6. In order to confirm the extracted gate length (*Lpoly*) value from *Cgd-V* characteristics, drain-induced barrier-lowering (DIBL) effects are also checked for the subthreshold *Id-Vg* characteristics of each transistor. Figure 5.7 shows that the DIBL effect is satisfactorily reproduced from low to high drain voltage for the 0.19-μm nMOSFET.

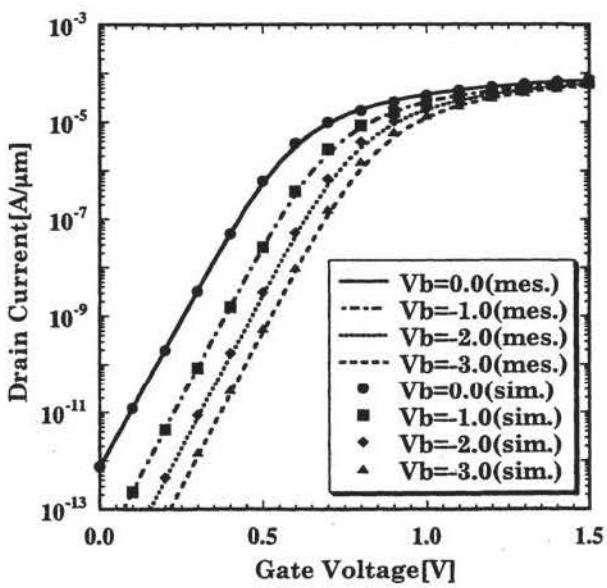


Fig. 5.5 Subthreshold *Id-Vg* characteristics of a 0.19-μm nMOSFET comparing measurement and simulation results after the channel doping profile is extracted. (Lines: measurement results, symbols: simulation results)

In order to obtain the linear *Id-Vg* characteristics for a deep submicron MOSFET, we must consider the parasitic resistance which occurs due to the contacts. In the case of a salicide process, silicide layers are formed on the source/drain contact regions. We include the silicide-to-silicon contact resistance as a low-mobility region atop the source/drain regions as shown in Fig. 5.8. The mobility value is determined to reproduce the *Id-Vg* characteristics in the linear regime including the body effect, the result of which is shown in Fig. 5.9 for the 0.19-μ m nMOSFET. Figure 5.10 shows the final two-dimensional device structure and doping profile for the 0.19-μ m nMOSFET produced by our quarter-micron dual-gate CMOS technology with a Co-salicide process [91]. The extracted doping and structure-related parameters are summarized in Table 5.2.

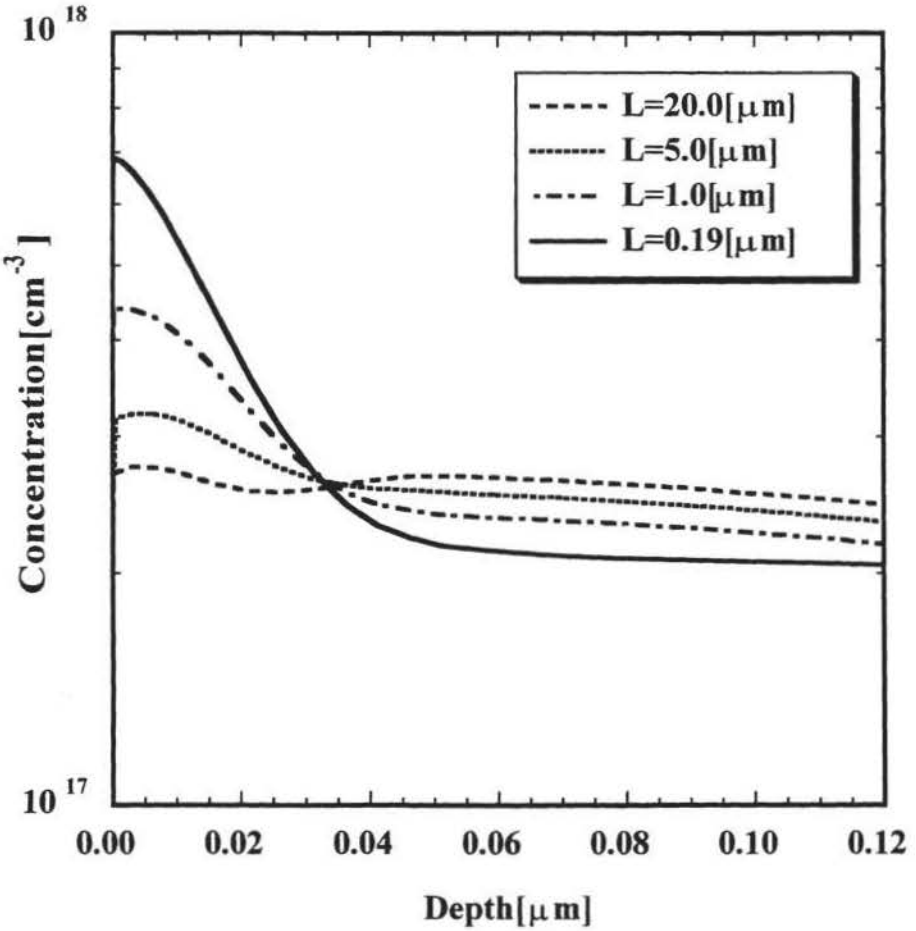


Fig. 5.6 Mid-channel doping profiles depending on the gate length.

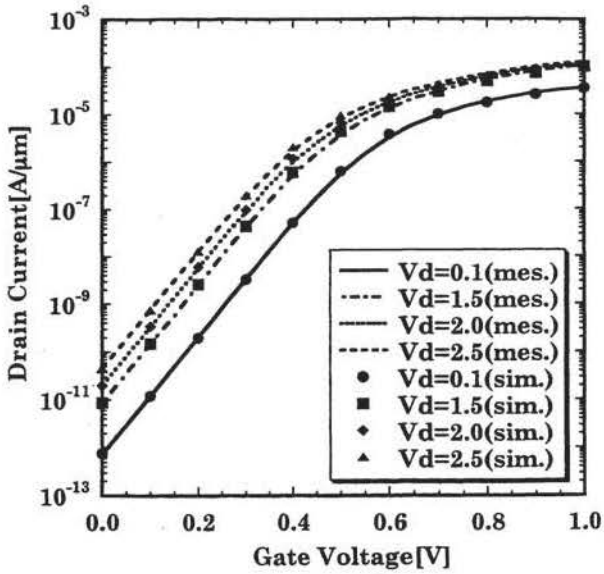


Fig. 5.7 DIBL effect for a 0.19-μ m nMOSFET comparing measurement and simulation results after the channel doping profile and gate length are determined. (Lines: measurement results, symbols: simulation results)

$$\mu = \frac{d}{qnR_c} \qquad R_c : \text{Contact Resistance } [\Omega\,\text{cm}^2]$$

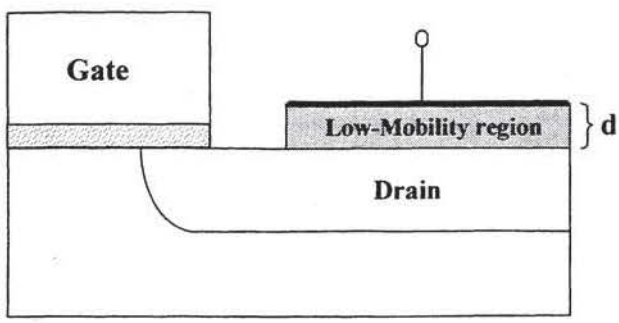


Fig. 5.8 Implementation of the source/drain contact resistance for linear *Id-Vg* calculation.

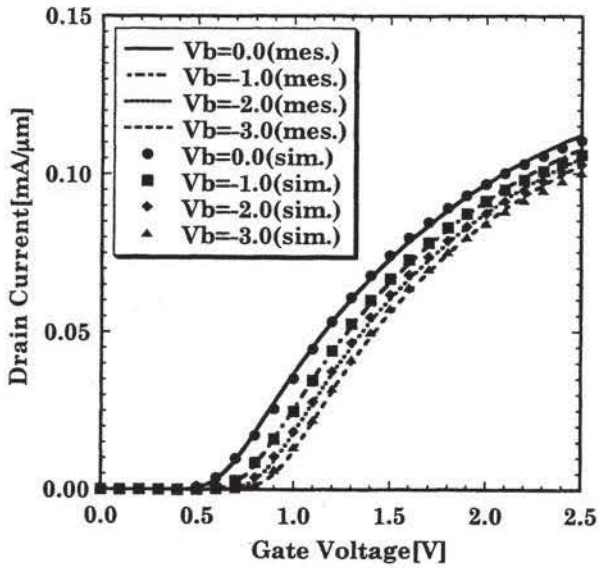


Fig. 5.9 Linear *Id-Vg* characteristics comparing measurement and simulation results including source/drain contact resistance.

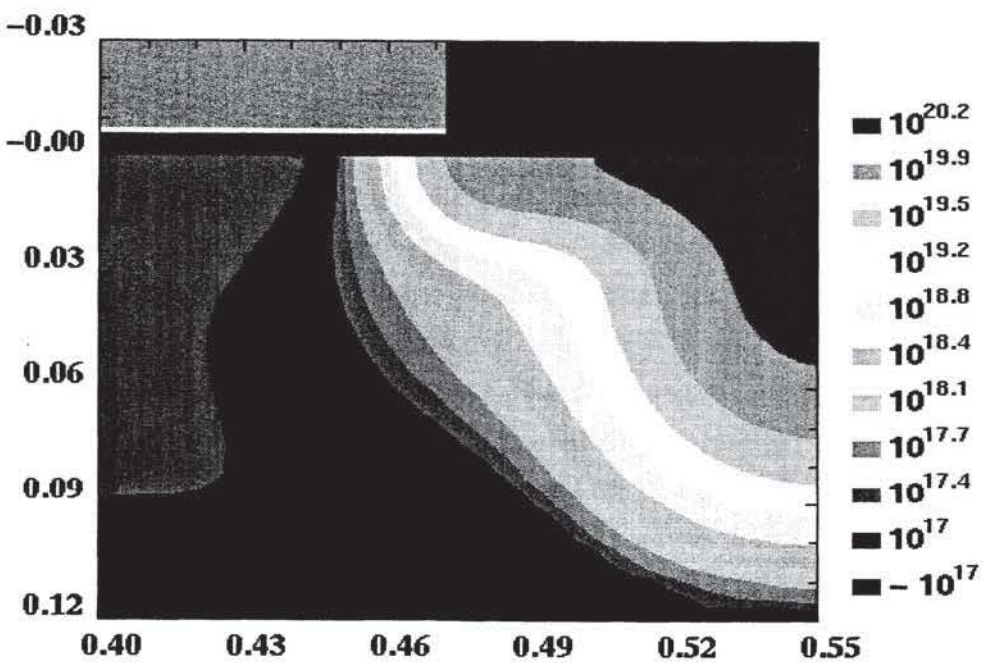


Fig. 5.10 Final two-dimensional structure and doping profile of a 0.19-μ m nMOSFET. (Unit; μ m for axes and cm^{-3} for contour lines)

Table 5.2 Extracted doping and structure-related parameters for a deep submicron nMOSFET (*Lpoly*=0.19 μ m) .Gate oxide thickness is 6.3 nm and contact resistance is 160 Ω μ m.

Gate doping

$N_g(cm^{-3})$	1.000×10^{20}
$x_g(\mu m)$	-7.917×10^{-3}
$D_g(\mu m)$	8.951×10^{-4}
E_g	2.000

Channel profile

$N_s(cm^{-3})$	-4.777×10^{17}
x_s (μ m)	0.000
$D_s(\mu m)$	1.917×10^{-2}
E_s	1.526
$N_c(cm^{-3})$	-1.315×10^{17}
$x_c(\mu m)$	6.000×10^{-2}
$D_{c1}(\mu m)$	1.000
E_{c1}	2.000
$D_{c2}(\mu m)$	9.993×10^{-1}
E_{c2}	1.019
N_W	-8.000×10^{16}

Source/drain extension profile

$N_J(cm^{-3})$	1.500×10^{20}
x_J (μ m)	2.514×10^{-7}
$D_J(\mu m)$	1.588×10^{-2}
E_J	1.998
R_J (μ m)	1.125×10^{-2}
$y_J(\mu m)$	*Lpoly*/2

N_2 (cm^{-3})	1.784×10^{19}
x_2 (μ m)	2.200×10^{-2}
D_2(μ m)	3.305×10^{-2}
E_2	1.842
R_2 (μ m)	3.244×10^{-2}
y_2(μ m)	*Lpoly*/2+0.02

5.4 Saturation Drain Current and Substrate Current

Based on the extracted deep submicron nMOSFET structure, the carrier transport in the high tangential field regime is investigated. Figure 5.11 shows the saturation drain current versus drain voltage (*Id-Vd*) characteristics of the 0.19-μ m gate length nMOSFET, comparing the measurement and simulation results using the GHDM implemented in GALENE-III. In this simulation, the saturation velocity in the inversion layer is set to be the experimental value of 9.23×10^6 cm/s obtained by Cooper and Nelson [102]. This value was derived by a time-of-flight experiment using a large resistive gate MOSFET. Recently, a UC-Berkeley group has reported slightly lower values using thick gate oxide bulk and silicon-on-insulator (SOI) MOSFETs. These values range from 8.1 $\times 10^6$ cm/s to 8.9×10^6 cm/s [103]. Therefore, the value at which the saturation velocity is set may be slightly controversial, but our results strongly suggest that the value obtained by Cooper and Nelson is reasonable. The detail of the carrier transport model in MOSFETs will be discussed in the next chapter.

The substrate current characteristics of the same MOSFET, calculated by our full-band Monte Carlo device simulator FALCON are shown in Fig. 5.12 [106]. It is worth noting that our impact ionization model has been tuned to reproduce experimental results for the bulk impact ionization coefficient and quantum yield [107,108], and thus does not include any device-dependent fitting parameters. The good agreement between the simulation and experimental results for the drain and substrate current without any further fitting demonstrates the high quality of the device structure and doping profile obtained using our inverse modeling technique. The inverse modeling

technique has been successfully applied to other devices with quite different device structures shown, for example, in [108].

The details of the full-band Monte Carlo device simulator FALCON will also be discussed in the next chapter.

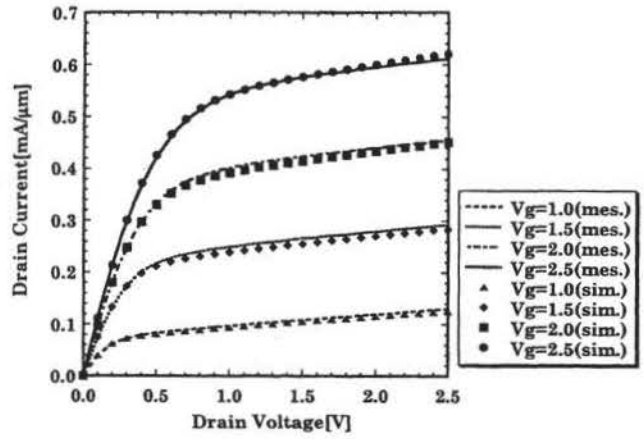


Fig. 5.11 *Id-Vd* characteristics comparing measurement and simulation results using the GHDM by GALENE-III for a 0.19-μm nMOSFET. (Lines: measurement results, symbols: simulation results)

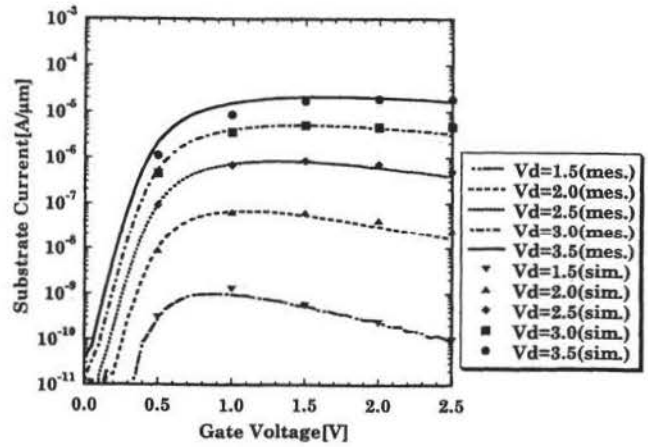


Fig. 5.12 Substrate current characteristics comparing measurement and simulation results using the full-band MC simulator FALCON for a 0.19-μm nMOSFET. (Lines: measurement results, symbols: simulation results)

5.5 Device Characteristics Prediction

In this section we demonstrate that our inverse modeling technique is not only able to reproduce experimental results but is also able to predict *I-V* characteristics of differently processed devices [95]. In Figs. 5.13 and 5.14, the predicted and experimental values of *I-V* characteristics for saturation drain currents and substrate currents are shown for a device with a source/drain extension implant dose ten-times lower than that of the 0.19-μ m gate length nMOSFET. The device structure is the same as that of the above-mentioned device with the exception of the peak values N_1, N_2 of the Gaussian functions (eqs. (5.5) and (5.6)), which are decreased to one-tenth of those for the 0.19-μ m gate length nMOSFET, having a gate oxide thickness which is proportionally changed using optical measurement data after the gate oxide growth. Based on this predicted device structure, the simulation reproduces the terminal current voltage curves very well. Though it has not been shown here, similar results are obtained when the channel implant dose is changed [95]. Thus it is possible to determine device structures under different process conditions without any further inverse modeling, removing one of the main obstacles of device optimization based on inverse modeling.

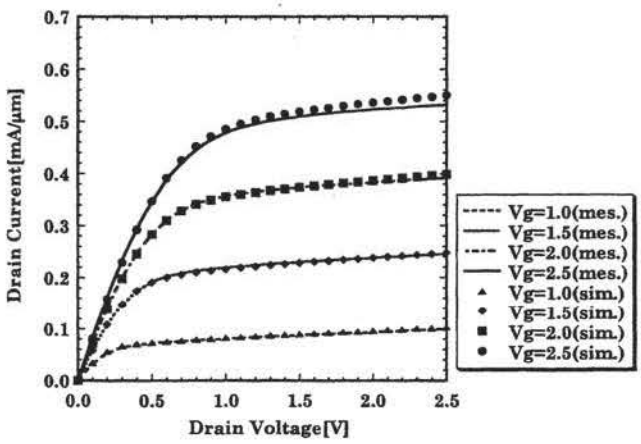


Fig. 5.13 *Id-Vd* characteristics comparing measurement and predicted simulation results using the GHDM by GALENE-III for a different 0.19-μ m nMOSFET of which source/drain doping concentration is scaled to one-tenth. (Lines: measurement results, symbols: simulation results)

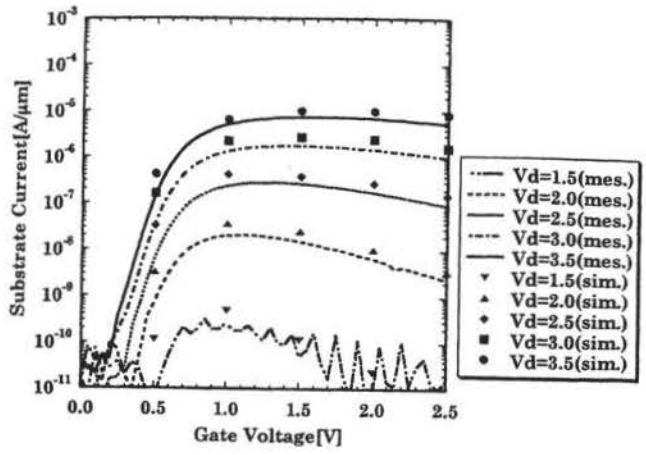


Fig. 5.14 Substrate current characteristics comparing measurement and predicted simulation results using the full-band MC simulator FALCON for a different 0.19-μ m nMOSFET of which source/drain doping concentration is scaled to one-tenth. (Lines: measurement results, symbols: simulation results)

5.6 Conclusions

The proposed inverse modeling technique is simple, because it is based on simple analytical functions for the doping profile available in standard device simulation programs. On the other hand, it is flexible, because the different parts of the doping profile are extracted independently. Based on the resulting device structure, device simulation yields good results not only for the linear regime but also for the hot-carrier regime for which no fitting was carried out. In addition, it is possible to vary the device structure to model devices processed under different conditions. Good agreement between the simulation and experimental results is found for these devices without any further inverse modeling. This makes it possible to use the extracted device structure for device optimization or the investigation of device parameter fluctuations.

Chapter 6 Carrier Transport in MOSFETs

6.1 Introduction

To design high performance deep submicron MOSFETs, it is important to understand carrier transport phenomena in MOSFETs with the consistency of physical models compared to observable electrical characteristics, especially in the high-field or hot carrier transport regimes. The carrier transport behavior in MOSFETs can be described by the *B*otlzmann *t*ransport *e*quation (BTE) for electron and/or hole gas in crystalline silicon. Although the original BTE was presented by Boltzmann to describe the behavior of classical gas more than one hundred years ago, it can be applied to the quantum mechanical elementary particles, or the electrons and holes in modern semiconductor devices [109].

Among three terminal semiconductor devices silicon MOSFET is so-called a surface device. A conductive channel region is induced by a gate capacitance between source and drain pn-junctions. In the channel region electrons (nMOSFET) or holes (pMOSFET) flow from the source (grand) to the drain according to a supply voltage. Near the source of an enhancement MOSFET, the inversion layer acts as the channel, but the channel depth gradually diminishes along the channel. Beyond a pinch-off point the channel disappears and the carriers flow like those in bulk silicon. Therefore, in order to describe the carrier transport behavior in MOSFETs, both two-dimensional and three-dimensional electron or hole gas should be characterized. Moreover, the channel carriers are scattered by donor or acceptor ions, lattice phonon and surface roughness irregularity. In the vicinity of the drain junction very high-electric field exists and causes *i*mpact *i*onization (II) scattering. These scattering mechanism are inherently important to describe the carrier transport behavior or the terminal current voltage characteristics in MOSFETs.

This chapter discusses the carrier transport phenomena based on the BTE by investigating several approaches to tackle the transport models and clarify which model should be used to describe electrical behavior of deep submicron MOSFETs. In the next section, some approaches to solve the BTE are discussed and especially the hierarchy from Monte Carlo particle-based model

to fluid-based models, mainly from the viewpoint of non-equilibrium transport, is focused. Then, in the third section some controversial arguments of impact ionization models are discussed and clarified, followed by the final conclusion section.

6.2 Modeling of Carrier Transport in MOSFETs

Basic equations to describe the carrier transport in silicon are the BTE, Poisson equation and Schroedinger equation. The last one is only needed when an inversion layer exists and electrons or holes behave as two-dimensional gas. Among these equations the BTE is the most important and it describes the distribution function $f(\vec{r},\vec{k},t)$ as follows:

$$\frac{\partial f}{\partial t}+\frac{d\vec{r}}{dt}\nabla_{\vec{r}} f - q\vec{E}(\vec{r},t)\nabla_{\vec{k}} f = \left(\frac{\partial f}{\partial t}\right)_{Coll}, \tag{6.1}$$

where $\vec{r}$ is the position vector, $\vec{k}$ is the momentum vector, q is the electron charge, t is time, $\vec{E}$ is external electric field and "*Coll*" means the collision term by scattering. In order to solve this equation the most accurate way is the *M*onte *C*arlo (MC) particle method. The BTE is self-consistently solved with the Poisson, and in some cases, Schroedinger equations.

6.2.1 Particle-Based Model

Recent dramatic improvement of MC approach can make it possible to consider MC simulators as an engineering tool [110-112]. The sophisticated MC simulators contain accurate enough physical models to describe the carrier transport in silicon MOSFETs. Not only the *f*ull-*b*and (FB) structure obtained by the pseudo-potential method but also impurity-carrier scattering, phonon-carrier scattering, surface roughness scattering and II scattering models are implemented. As a result, as far as the carrier transport in bulk silicon is concerned, there is no large discrepancy for the electron-phonon scattering rates and impact ionization rates derived from various MC simulators [113]. Even a CPU time problem has been resolved by combining a statistical enhancement technique like the multiple refresh method [114,115], and a suitable convergence estimation method if self-consistent calculation

with the Poisson equation is not necessary in such a case as to calculate substrate currents [112,116,117].

6.2.1.1 Full-band Monte Carlo device Simulator - *FALCON*

Now, the sophisticated approach to reduce calculation time in the FB-MC device simulator *FALCON* [112] is described. The most CPU intensive part of calculation for scattering events in FB-MC simulation is the selection of the final state in k-space after the scattering process and final energy band have been decided. The time is spent mainly to find the final tetrahedron in the grid of the discretized FB structure. The probability of selecting a certain tetrahedron, or the discretized k-space volume, is proportional to the transition rate integrated over all possible final states contained in the tetrahedron. In [118], the *t*ransition *p*robabilities (TP) are calculated for all possible tetrahedra and the final tetrahedron is selected with the direct method, or by gathering all possible probabilities however they may be small [119]. This method consumes large amounts of CPU time, because the calculation of the TP is CPU intensive and the number of possible final tetrahedra is very large. By using the *a*cception/*r*ejection (A/R) method [119], instead of the direct method, the CPU time consumption can be reduced. With this method a final tetrahedron is selected by choosing one of the possible final tetrahedra with uniform probability which plays a threshold-like role. The selected tetrahedron is accepted, if its TP is larger than the upper bound of the TP times a uniformly distributed random number. Instead of calculating the TP for every possible final tetrahedron, in the case of the A/R method, the TP is calculated only for the selected tetrahedra. The number of A/R events is given by the ratio of the upper bound of the TP to the average TP. As long as this ratio is smaller than the number of possible final tetrahedra, the A/R method is faster than the direct method because the TP is evaluated less frequently. This is the case for the scattering models given in [120] for which in average less than 11 A/R steps (evaluations of the TP) are necessary. Since this number is much smaller than the number of possible final tetrahedra, which is usually in the hundreds or thousands, the A/R method is faster than the direct method without any approximation.

The upper bound of the TP can be calculated easily for the employed

scattering models [120], because the TP is proportional to the *d*ensity *o*f *s*tates (DOS) in each tetrahedron which depends only on energy. In the beginning of a simulation an energy dependent look-up table is built which contains the maximum DOS of the set of possible final tetrahedra for the given energies. Moreover, lists are built of all tetrahedra containing certain final energies. With these lists, which are of the size of a few MBytes and built only once, it is quite simple to select a final state. In Fig. 6.1 the method is shown for a scattering process with a constant final energy. Generalization to a momentum-dependent final energy is straightforward and does not necessarily impair efficiency. The same is true for a more complex momentum dependence of the transition rate or a higher number of final tetrahedra resulting from a finer grid in k-space, as long as the ratio of the upper bound of the TP to the average TP is not increased.

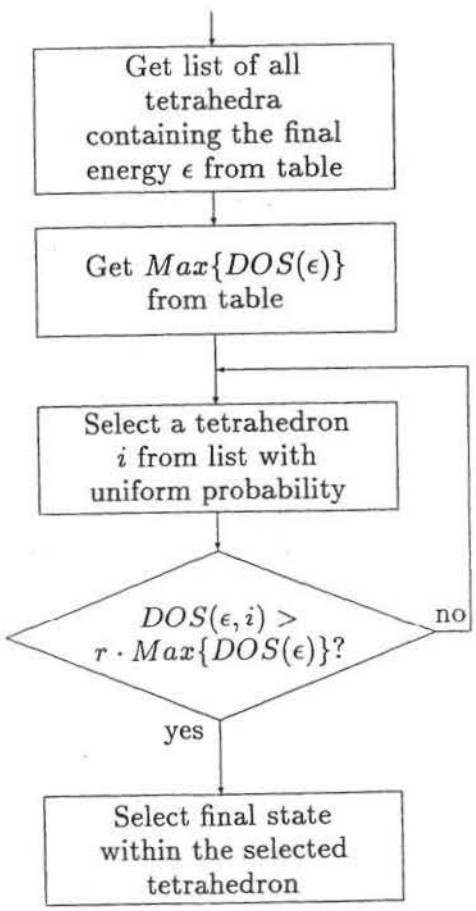


Fig. 6.1 Flowchart of the acceptance/rejection method for the selection of the final state after scattering (random number r uniformly distributed in [0,1])

The combined use of the A/R method for the selection of the final state after scattering and the linear interpolation of the FB [111] together with the variable Γ-scheme (the way to change the highest scattering rate depending on the carrier energy to reduce self-scattering event) [121] and the efficient formulation of the scattering processes [120] makes the program much faster than other FB-MC approaches [110,119] without a loss in physical soundness or numerical accuracy. The A/R method for the selection of the final state is so fast that less than 10% of the CPU time for a device simulation is spent for finding the final states, thus making, for example, the stringent approximation of cellular automata [122] unnecessary.

In the case of non-self-consistent FB-MC device simulations [116,117,123] the simulation particles are uncorrelated and the convergence can be estimated with standard textbook formulas [116,117,124]. The average substrate current ($\bar{I}$) and its standard deviation $\overline{\sigma}$ are:

$$\bar{I} = \frac{1}{N}\sum_{i=1}^{N} I_i \quad \text{and} \quad \overline{\sigma}^2 = \frac{1}{N-1}\left[\frac{1}{N}\sum_{i=1}^{N} I_i^2 - \bar{I}^2\right], \tag{6.2}$$

where I_i is the substrate current generated by the particle i. From the central limit theorem it follows that the distribution function of $\bar{I}$ approaches a Gaussian distribution [124] and the probability (confidence level P_{con}) that $\bar{I}$ is within the confidence interval of $\pm 2\,\overline{\sigma}$ around the expected value of I is:

$$P_{con} \approx \frac{1}{\sqrt{2\pi\overline{\sigma}^2}} \int_{-2\bar{\sigma}}^{2\bar{\sigma}} \exp\left(-\frac{x^2}{2\overline{\sigma}^2}\right) dx = 95.45\% \tag{6.3}$$

Using the relative error $\tau_{err} = 2\,\overline{\sigma}/\bar{I}$ the simulation runs are stopped when a relative error of $\pm 10\%$ (with a probability of 95.45%) has been reached.

6.2.1.2 Simulation Results by FALCON

Figure 6.2 shows the substrate currents of a 0.18-μm and 1.0-μm nMOSFET calculated by FALCON and experimental results. The MOSFET structures including the doping profiles are determined by the inverse modeling technique as described in Chapter 5. The simulations have been

performed without any statistical enhancement method, and the CPU times are for a Super SPARC workstation. To our best knowledge these CPU times are the smallest ones achieved on an ordinary workstations for FB-MC simulations so far compared to other works [111,125].

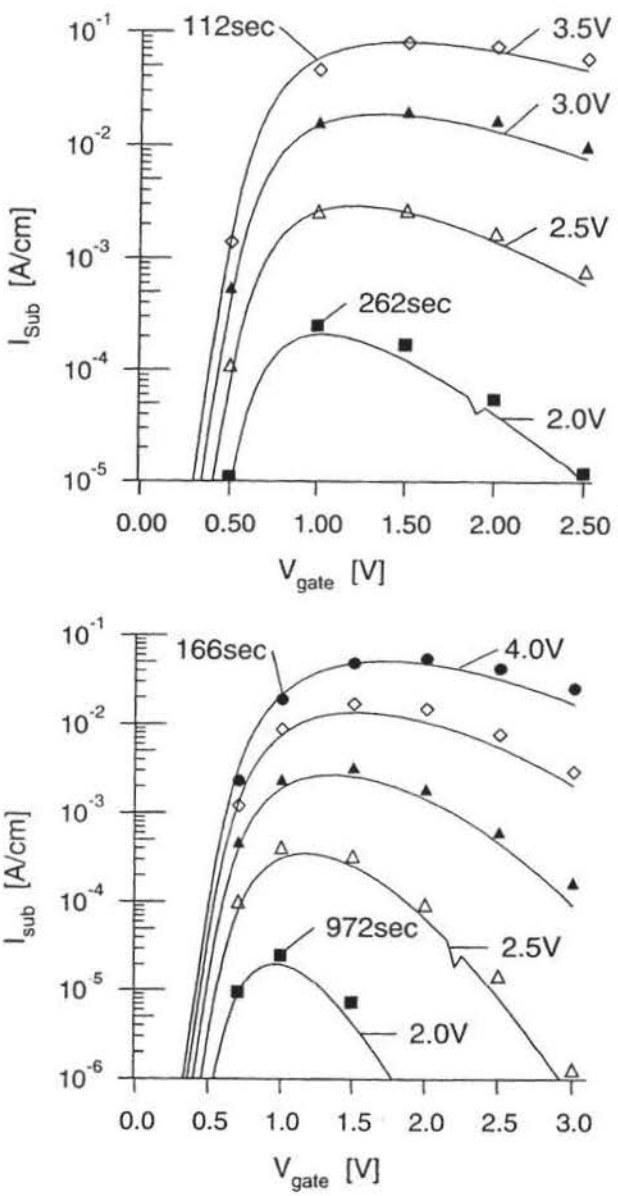


Fig. 6.2 Substrate current and CPU times for (a) a 0.18 μ m nMOSFET and (b) a 1.0 μ m nMOSFET for different drain voltages. (Lines: experiment, Symbol: FB-MC simulation)

6.2.2 Fluid-Based Model

As MOSFET is inherently a surface device, the carrier transport not only in the bulk but also at the surface should be accurately modeled. When the inversion layer is modeled rigidly, the Schroedinger equation in addition to the BTE and Poisson equation should be solved self-consistently. This situation still causes the largest problem of the MC method, or the CPU time burden. Therefore, as engineering tools we should use *both* FB-MC particle-based and fluid-based device simulators.

6.2.2.1 Hydrodynamic Model versus Energy Transport Model

To solve the BTE based on the fluid model, there are two approaches, an equilibrium way and a non-equilibrium way. The former is the so-called *d*rift-*d*iffusion (DD) model. It is useful if the carrier temperature is low enough to be nearly equal to that of the lattice, but it cannot be applied to deep submicron MOSFETs in high tangential field operation regime. The latter approach contains two major non-equilibrium approximation methods. These were believed to be suitable to solve the non-equilibrium carrier transport in the case of deep submicron MOSFETs when carrier temperature is far higher than the lattice temperature. In both approaches the BTE is multiplied by powers of the momentum and integrated over momentum space to result in "balance equations". The first one was originated to Stratton and called *e*nergy *t*ransport *m*odel (ETM) or *e*nergy *b*alance *m*odel (EBM) [126]. The second one was proposed by Blotekjaer and called *h*ydro*d*ynamic *m*odel (HDM) from the analogy of fluid dynamics [127]. The difference of these approaches is already discussed by Apanovich et al. including terminology comments for the ETM and EBM [128]. To solve the problem of the standard or conventional HDM, the ETM was suggested as a better alternative [128,129]. (Here, the term *ETM* is used as the same as *EBM*.) However, the applicability of more sophisticated version of HDM or the *g*eneralized *h*ydro*d*ynamic *m*odel (GHDM) [99] was never discussed in detail. Therefore, in the next subsection this dissertation shows the difference of the GHDM compared to other conventional HDM or ETM and clarifies the usefulness of the GHDM to describe the carrier transport in deep submicron MOSFETs.

6.2.2.2 Generalized Hydrodynamic Model

To avoid too strong velocity overshoot near the drain junction suffered by the conventional HDM, the convective term or the thermal diffusion current term is omitted in the case of the ETM [128,129]. At a glance, this assumption seems to be reasonable because the singularity of the carrier velocity at the drain side is eliminated. But this situation is often demonstrated by using a one-dimensional simple n^+-i-n^+ diode as a test device. We must remember that MOSFET is never one-dimensional but at least two-dimensional device even in the case of a long width transistor. The heated carriers in the vicinity of the drain junction act as a heat source for the thermal diffusion (convective) current, and it flows from the drain not only to the source but also to the substrate in addition to the heat flux (conductive) current. To the contrary, in a one-dimensional n^+-i-n^+ diode the electrons by the thermal diffusion current as well as heat flux only flow from the drain to the source. We can never estimate the deep penetration feature of hot carrier distribution toward the substrate at the vicinity of the drain junction. Therefore, the n^+-i-n^+ diode is not a suitable test device to check the model accuracy of the carrier transport in MOSFETs. In order to testify the accuracy of the models, MOSFET itself must be used.

The most important standpoint of the GHDM is that it tries to solve the problems encountered in the conventional HDM without neglecting the thermal diffusion current term. Too high carrier temperature or too large velocity overshoot found in the conventional HDM is mainly caused by the parabolic band approximation. Real band structure of silicon is neither parabolic nor uniform. This means that the effective mass of an electron can never be expressed by a single scalar quantity, but that it must be a tensor value and that carrier temperature can never be defined uniquely. Moreover, in the case of the conventional HDM the heat flux is estimated too large resulting in deep carrier penetration toward the substrate at the vicinity of the drain junction. To overcome these problems, from highly heuristic insight that the electrons in crystalline silicon still behave like classical uniform gas [130], sophisticated modifications for the balance equations were developed based on the following key features ;

a) The carrier temperature is redefined by kinetic energy, not by average

energy or band energy.

b) In order to express the kinetic energy, the effective mass of electron is defined by the average of an effective mass tensor trace.

c) By using these modified scalar temperature and effective mass, relaxation times for momentum, energy and energy current flow are redefined.

d) The Wiedemann-Franz low for thermal conductivity is modified by a factor of the redefined relaxation time ratio between the energy current flow and momentum.

These modifications seem to be tricky and may arouse some controversial issue concerning their physical meaning. However, these modified parameters are directly obtained by MC simulators using a simple bulk test device [99], and, of course, a FB-MC device simulator can also be used for such a purpose [131]. Once those parameters are fixed, the numerical approach for the conventional HDM based on relaxation time approximation is extensively applied to solve the balance equations for the GHDM. The appropriateness of the GHDM should be checked by MOSFET electrical characteristics.

For the carrier transport in the channel or the inversion layer, the Schroedinger equation should be solved in addition to the Poisson and balance equations. Otherwise, as an alternative way, the inversion layer mobility can be semi-empirically expressed in the balance equations based on Matthiessen-rule like expression [100,101]. In this case the effective field dependent universality of low tangential field drift mobility should reproduce experimental results. This requirement guarantees the accuracy of impurity scattering, phonon scattering and surface roughness scattering models for the inversion layer mobility. Moreover, the inversion layer mobility should consistently reproduce high tangential field dependent carrier velocity obtained by experiments. Although this approach is not rigorous to describe the carrier transport in the inversion layer, the applicability should again be checked by MOSFET itself.

6.2.2.3 Guarantee of Simulation Accuracy by GHDM

As already shown in the previous chapter, if we can identify accurate device structure as well as doping profiles in MOSFETs, the saturation drain

currents are well simulated with a good fit to experimental results. Even, we can predict *Id-Vd* characteristics for differently processed MOSFETs once we use a well-defined device information based on inverse modeling as far as the device structure is not completely different. These results clearly demonstrate the accuracy of the GHDM approach because no fitting on device level such as the manipulation of doping profile, gate oxide thickness, gate length to reproduce experimental results is introduced. Only one fitting parameter is the saturation velocity for the inversion layer, but it was set to be the value obtained by a quite different experimental way, or the time-of-flight measurement. This fact is another proof to guarantee the physical models implemented in the GHDM approach. Of course, there is still uncertainty to extend the GHDM to sub-0.1-μ m gate-length regime, at least the GHDM can be applied to subquarter micron gate length nMOSFETs.

6.3 Impact Ionization in MOSFETs

In deep submicron MOSFETs the high electric field near the pinch-off point makes the carrier energy higher and so-called hot carriers are generated. They induce II scattering and generate electron-hole pairs. These hot carriers are the source of MOSFET reliability problems and there is always trade-off between higher drain current and shorter device lifetime induced by the hot carriers. Therefore, to utilize the maximum performance of MOSFET the reliability edge for the best performance point should be detected as precisely as possible. For this purpose the FB-MC device simulator is the most powerful tool to investigate the hot carrier properties. It is well known that nMOSFET substrate current originates from the holes generated by II at the vicinity of the drain junction and is a very good monitor to guarantee the device lifetime. Concerning the II phenomena, recent FB-MC simulators contain a state of the art model which has been tuned to reproduce experimental results for the bulk impact ionization coefficient and quantum yield [106,125]. Electric field distribution in a MOSFET is in between these two extreme cases as shown in Fig. 6.3. Therefore, the accuracy of the model should be guaranteed by observable electrical characteristics such as substrate current of MOSFET.

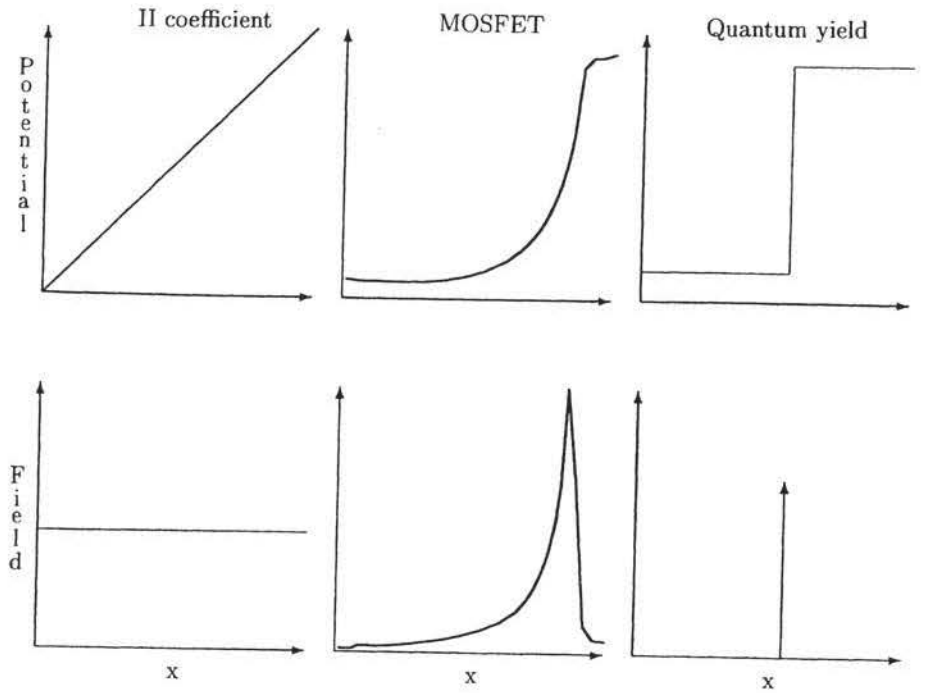


Fig. 6.3 Schematic potential and electric fields for the II coefficient (left), at the Si/SiO_2-interface of a MOSFET (center) and for the quantum yield experiment (right).

6.3.1 Impact Ionization Models

In the case of fluid-model-based device simulators, the II models can be implemented in the balance equations. The local field model based on Chynoweth's approximation [132] is the oldest and simplest. In this case, the electron II generation rate G_{II} is expressed:

$$G_{II} = \alpha(P)|\mathbf{j}_n(P)| \tag{6.4}$$

$$\alpha(P) = \alpha_\infty \exp(-E_c/E(P)), \tag{6.5}$$

where P is a point in space, α_∞, E_c are parameters and $\mathbf{j}_n$ is the electron current density. Chynoweth's approximation (eq. (6.5)) describing α as a function of the local electric field is well justified in silicon under nearly homogeneous material and field conditions with doping independent parameters α_∞ and E_c over a large range of field strength.

However, inside a MOSFET the electric field is strongly inhomogeneous so that the approximation (6.5) becomes very poor. In order to overcome this problem, local temperature model and non-local lucky electron model have also been proposed. The local temperature model expresses the II generation rate as a function of local carrier temperature, and the lucky electron model treats hot electron as a non-locally unscattered lucky carrier that gained

higher energy from external field than the average energy. Especially, the non-local lucky electron model contains the hard threshold model and the soft threshold model. The II energy in silicon shows no abrupt threshold because of indirect transition nature between valence and conduction bands caused by its band structure. As far as the electric field is much higher than the II threshold, the hard threshold model can be applied without any serious problem. But, when supply voltage is low enough for the electric field to be around the threshold, the hard threshold model can no longer reproduce experimental results. By using the soft threshold model, low voltage impact ionization around the impact ionization threshold regime can be investigated [107]. Nevertheless, in this chapter the detail of these II models using fluid-based models will not be discussed because the most accurate II model is based on the particle-based FB-MC model.

6.3.1 Surface Impact Ionization

The term surface II was introduced in [133] and describes the effect that the II coefficient in the vicinity of the Si/SiO_2-interface in MOS devices seems to be reduced. In Fig. 6.4, II coefficients for CCDs (*c*harge *c*oupled *d*evice) and MOSFETs are shown which were obtained by reverse engineering based on the local-field-DD model [133]. Comparison with the bulk case reveals a considerable reduction of the II coefficient close to the interface. Different reasons are given for this effect and they will be discussed in the following.

In [133, 134] the difference between surface and bulk II is attributed to the reduced mean free path of the carriers in the vicinity of the interface. Within the lucky electron concept [135] the critical field in Chynoweth's law in eq. (6.5) is inversely proportional to the mean free path λ:

$$E_c = \varepsilon_{TH}/(q\lambda), \tag{6.6}$$

where q is the electron charge and ε_{TH} the threshold energy. A reduction in the mean free path, due to the enhanced scattering at the surface, results in a larger critical field and a reduced II coefficient. In the FB-MC model the mean free path at the MOS interface is smaller due to surface roughness scattering. But it has been doubted [125], that this effect is responsible for the strong suppression of II in MOS devices. In Fig. 6.5 substrate current

modeling results are shown, which were calculated with and without surface roughness scattering, and the observed differences are far too small to explain the effect of surface II seen in Fig. 6.4.

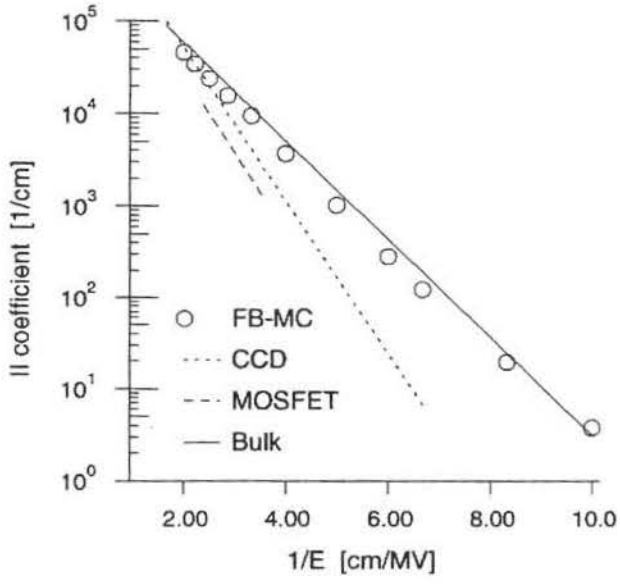


Fig. 6.4 II coefficient versus inverse electric field for bulk silicon (solid line) as given by van Overstratton, for MOSFETs (dashed line) and CCDs (dotted line) by Slotboom [133] and FB-MC results (symbol).

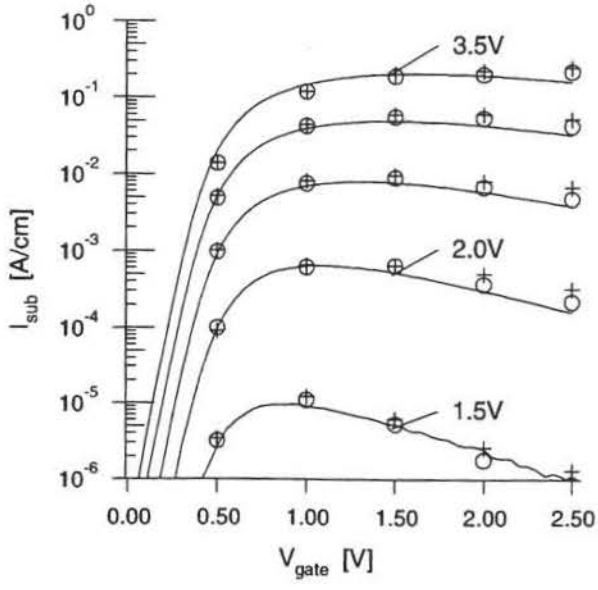


Fig. 6.5 Substrate current versus gate voltage for the shallow source/drain extension 0.16 μ m-nMOSFET (Vd=1.5,2.0,2.5,3.0,3.5V). (Lines: measurement, ○: FB-MC with surface roughness scattering, + FB-MC without surface roughness scattering)

To continue exploring the reasons for the different findings in [133] and in this work, the CCD experiment of [133] is re-investigated. The lateral electric field given in [133] for case *B* of the CCD experiment and the corresponding spatial II generation rates are shown in Fig. 6.6. The generation rates have been calculated with the *l*ocal *f*ield (LF) model with the bulk parameters, the local field model with the surface II parameters of [133] and the FB-MC model. In [133] it is stated that the local field model can be applied to this case, because the electric field does not vary much over distances comparable to one mean free path length which is in the order of 10 nm. Our FB-MC results, which are based on the best microscopic II model available today, show that this statement is not correct. If it were correct, the results of the local field model with bulk parameters and FB-MC should be the same for the field distribution in the CCD experiment since FB-MC reproduces the bulk results under homogeneous conditions (see Fig. 6.4). But as shown in Fig. 6.7 the maximum FB-MC II generation rate is strongly reduced compared to the local field model with bulk parameters showing that the local field approximation fails for the rapidly varying electric field present in the CCD. The local field approximation is only valid, if the electric field is nearly constant over the distance, which the electrons have to cover to gain the energy necessary for II. The threshold energy of II is at least the gap energy (1.1eV). To gain this energy from the electric field given in Fig. 6.6 the electrons have to travel along a distance of more than 70 nm, along which the electric field varies strongly (roughly a factor of 2). Therefore the condition of the nearly constant electric field is violated and the local field approximation fails.

By reducing the II coefficient (switching from bulk to surface II) the local field model can reproduce the reduction of the II rate (This effect has been discussed in detail for exponentially growing fields in [136,137]), but the microscopic details are still poorly represented (see Fig. 6.6). The local field model can neither reproduce the threshold characteristics of II nor the II generation rate in the field free space on the right side. The experimentally accessible multiplication factor can be calculated by integrating the generation rate over the device area, and in Fig. 6.7 the results are shown for the three different II models. For comparison the experimental results of [133] are shown. (The assumed electric fields may not exactly reproduce the

original ones of [133]. Nonetheless, the electric fields are similar to the ones of [133] and it is clearly shown that the local field model fails due to the nonlocal nature of the transport caused by these strongly inhomogeneous electric fields.) The results of the local field model relying on surface II parameters agree well with the ones of the FB-MC model while the local field model with bulk parameters yields consistently higher values. This means that the suppression of the II coefficient in the case of the local field model is necessary to counterbalance the overestimation of the II generation rate caused by the failure of the local field model under inhomogeneous field conditions. The effect of surface II found in [133] is therefore due to the inadequate local field model and not due to a difference of physics in bulk silicon and at the surface of MOS devices.

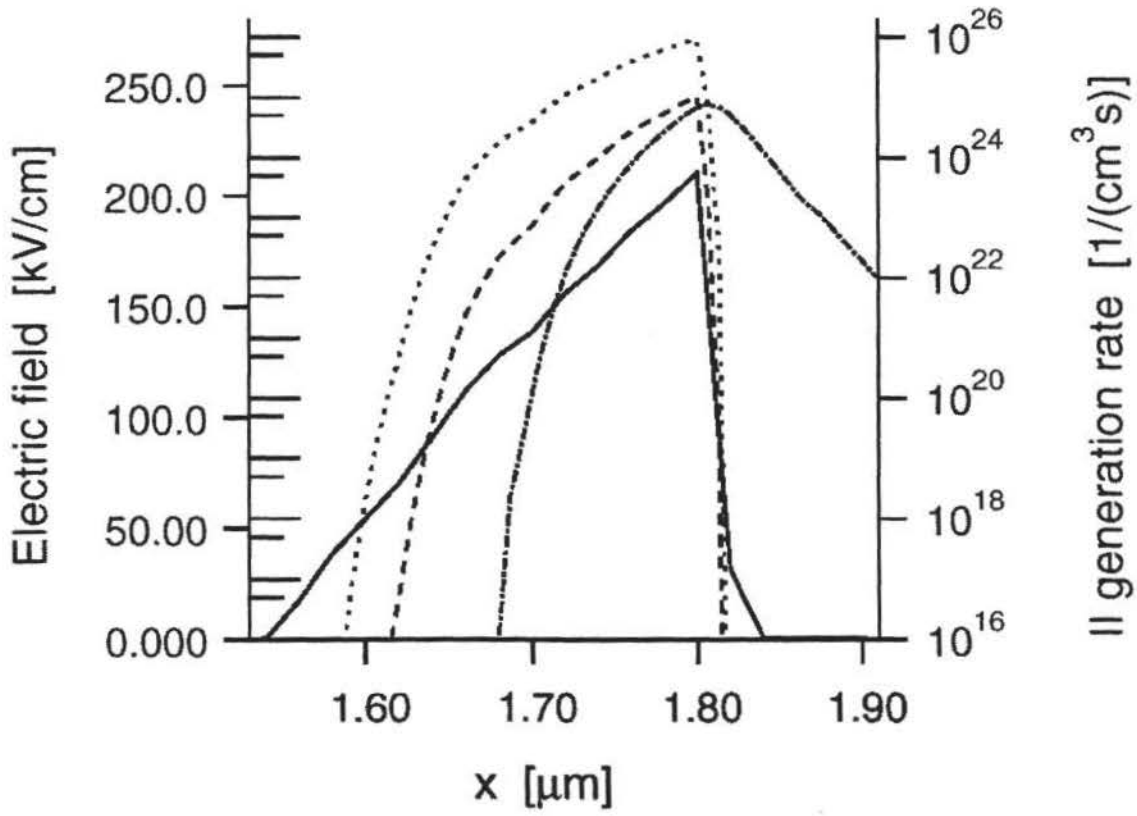


Fig. 6.6 Electric field (solid line) and II generation rate for the case B of the CCD experiment [133] calculated with the LF model based on bulk parameters (dotted line), based on surface parameters (dashed line) and FB-MC results (dashed dotted line).

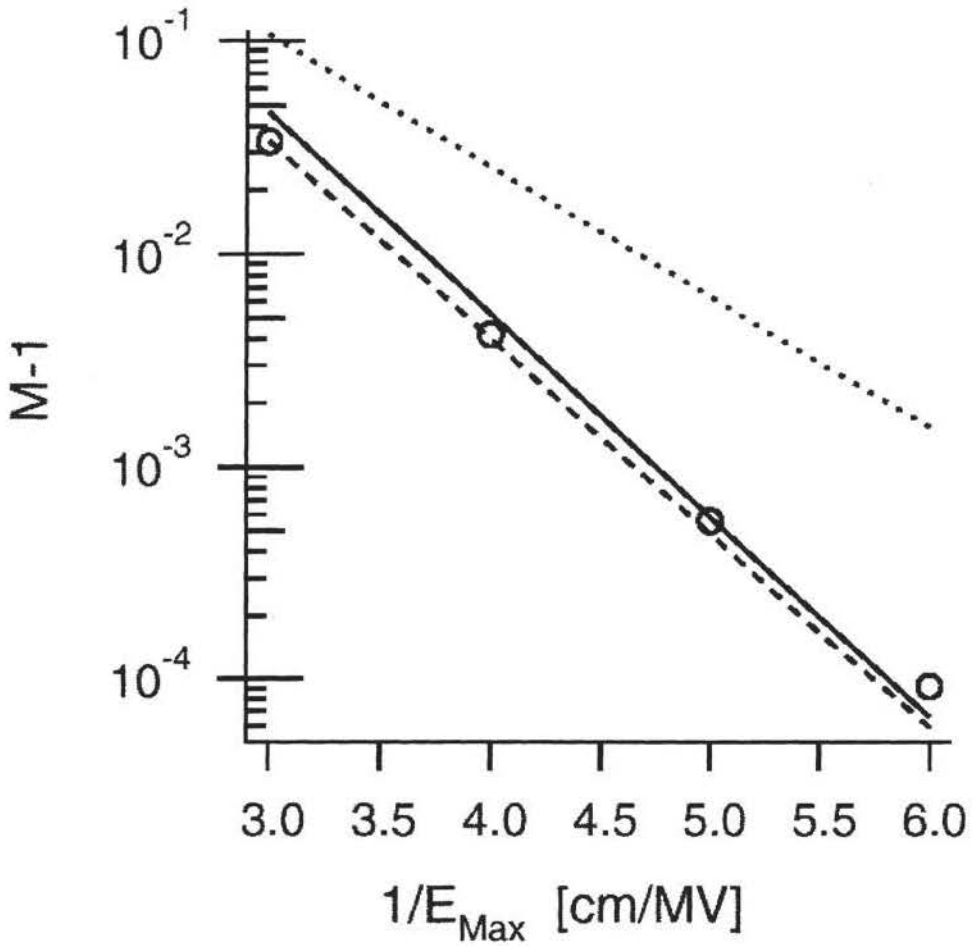


Fig. 6.7 Multiplication factor for case *B* of the CCD, measured (solid line) [133], calculated with the LF model based on bulk parameters (dotted line), based on surface parameters (dashed line) and FB-MC results (symbol).

In [125] a theoretical model for II in nMOSFETs is proposed which considers the quantization of the inversion layer. It is found that the II coefficient in the inversion layer decreases with increasing electric field perpendicular to the interface. For very high electric fields perpendicular to the interface and under otherwise homogeneous conditions this II coefficient is similar to the results found in [133] (see Fig. 11 of [125]). But the situation in a MOSFET is different, because II mainly occurs behind the pinch-off point, where the electric fields are too weak for a considerable confinement of the carriers. In Fig. 6.8 the average value of the confining field, at the position where II occurs is shown for the shallow source/drain extension nMOSFET. Only confining (positive) electric fields perpendicular to the

interface (E_{ver}) are taken into account and the average field (E_{ave}) is defined as follows:

$$E_{ave} = \frac{\int E_{\mathrm{ver}}(x,y)\Theta(E_{ver}(x,y))G(x,y)dxy}{\int G(x,y)dxy} \quad , \qquad (6.7)$$

where G is the II generation rate. Only in the case where the gate voltage is larger than the drain voltage does a substantial confining field arise. But these fields are still small compared to the fields necessary to reproduce surface II with the model proposed in [125]. Moreover, the strong dependence on the confining field implies that the effect should be negligible for small gate voltages. This is in contrast to the effect of surface II which reduces substrate current even for gate voltages near the threshold voltage. This means that the genuinely new effect proposed in [125] does not explain surface II as described in [133] and is only of minor importance regarding the problems discussed in this work.

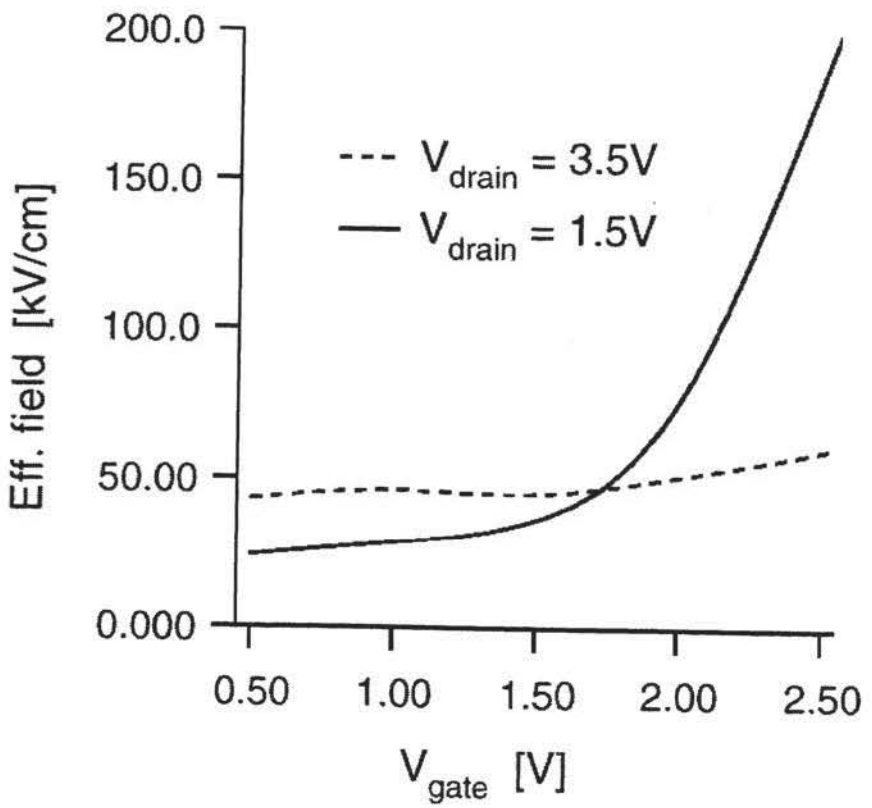


Fig. 6.8 Confining electric field averaged with the II generation rate for the 0.16 μ m-nMOSFET.

Another explanation for surface II is mechanical stress at the interface due to the formation of the oxide. The magnitude of the stress should vary with the oxide thickness and accordingly the impact on II. Since the reported FB-MC simulations are in close agreement with experiment for the full range of oxide thickness (5.3-42 nm) without accounting for mechanical stress [138], it has to be concluded that the impact of this stress on II seems to be a minor one. The same can be said for other unknown effects which might depend on the specific process technology. The investigated devices belong to three distinctly different technology generations (256KBit, 16MBit, 1GBit) and good results are obtained by above-mentioned FB-MC simulations, irrespectively of the technology and without any fitting of model parameters on device level [107,138].

6.4 Conclusion

To design deep submicron MOSFETs, especially in hot carrier transport regimes, accurate physical models must be implemented in device simulators. For such a purpose, we need particle-based FB-MC simulator for hot carrier analyses and fluid-based GHDM device simulator for saturation drain current estimation.

By using the FB-MC model as physical reference it could be demonstrated that the surface II effect is an artifact that appears only when MOS substrate currents are described with II models which are not predictive, because their physical base is not sufficiently sound. In addition, it was possible to refute the various explanations given for surface II in the literature by investigating them with the FB-MC model. Besides, this suggests that today's understanding of electron II is fundamental enough for predictive device simulation.

Chapter 7 Physics-Based MOSFET Design

7.1 Introduction

As the microfabrication technology has brought the quarter to subquarter micron era, MOSFET performance in CMOS technology has caught up with that of silicon BJT from the viewpoint of cutoff frequency or basic inverter delay. This is mainly attributed to the deep submicron channel length that is reached around 100 nm. To realize high-speed ULSI as a system-on-a-chip, higher driving capability for loaded capacitances, especially for wiring, and lower power dissipation for practical cooling instruments and/or longer battery operation for mobile computing are strongly required. To this end larger transconductance or higher saturation drain current should be realized at lower supply voltage for the modern CMOS technology, that directly demands smaller process margins. Moreover, as the semiconductor industry matures, huge investment for a new fabrication line with a larger wafer diameter is becoming a high-risk-taking burden. Therefore, the systematic MOSFET design for the next generation CMOS technology in a cost competitive way is one of the most important concerns that the industry is facing.

Based on inverse modeling and physically accurate device simulators using particle- and fluid-based models, we can design MOSFETs for the next generation to satisfy required device electrical specifications. Deep submicron channel length with larger drain current accompanies higher tangential field that directly causes hot carrier-induced problems such as the device performance degradation even at lower supply voltage. Hence, the reliability edge for the highest driving capability should be accurately detected to maximize the break-even point for performance/lifetime tradeoff. For this purpose the full-band Monte Carlo (FB-MC) device simulation is inevitable, and to predict the saturation drain current with reasonable CPU time, the generalized hydrodynamic model (GHDM)-based simulation is also unavoidable. In this chapter, some examples for hot carrier analyses and saturation drain current design for subquarter micron MOSFETs are described.

7.2 Hot Carrier Effects of nMOSFET

By using the FB-MC device simulator, not only observable substrate currents but also unobservable hot carrier distributions or other physical quantities can be analyzed. Figures 7.1 and 7.2 show the hot electron and hot hole distributions for a 0.16-μ m-metallurgical-channel-length nMOSFET at the supply voltage of 2.5V. Hot electrons are mainly distributed just inside the drain junction beneath the gate electrode. This means the hot electron damages mostly occur inside the drain junction under the gate electrode. To increase the saturation drain current, instead of conventional LDD (*l*ightly *d*oped *d*rain), highly doped source/drain extension structure is adapted in the quarter micron CMOS technology. As shown in Fig. 5.10 derived by the inverse modeling technique, the surface impurity concentration approaches almost to 1×10^{20} /cm^3. Therefore, the electric field at the vicinity of the drain junction is no longer reduced unlike the case of an older LDD structure. On the other hand, not many hot electrons are distributed under the sidewall spacer outside the gate electrode. This suggests that typical degradation mechanism caused by the LDD structure does not occur in the case of the source/drain extension structure. Moreover, we can find broad distribution of hot holes in the middle of the channel. As it is known that the activation energy of hole to produce interface state is much lower than that of electron, these rather warm holes may play an important role for device performance degradation. Figure 7.3 shows the impact ionization (II) generation rate distribution caused by (a) channel electron and (b) secondary hole. Although the amount of hot holes is smaller than that of hot electrons, they are mainly located in the middle of the channel. If there might be increase of interface state at this portion, surface-scattering-induced mobility degradation may cause drain current decrease.

In Fig. 7.4 the currents of electrons and holes with certain energies hitting the Si/SiO_2 -interface of the 0.16-μ m- and 0.18-μ m-channel-length nMOSFETs are shown [95]. The bias conditions are the ones of maximum stress in the allowed bias range (*Vgate* = 1.5V and *Vdrain* = 2.75V). The shift in the distributions for the 0.18-μ m device towards the drain (the poly-gate edge is at y = 0.11 μ m) reflects the increase in effective channel length because it is made by ten times lower source/drain extension implant dose

compared to the 0.16-μm device. The lower peak values of the particle currents due to the field reduction are also caused by the lower doping concentration in the source/drain extension of 0.18-μm device. The decrease is the strongest for the high energetic electrons, which are thought to play a crucial rule in the hot carrier degradation process. In addition to this effect, hot and warm holes are increased for the 0.16-μm device with the higher source/drain extension implant dose. The interface state generation mechanism caused by these holes should be investigated in future.

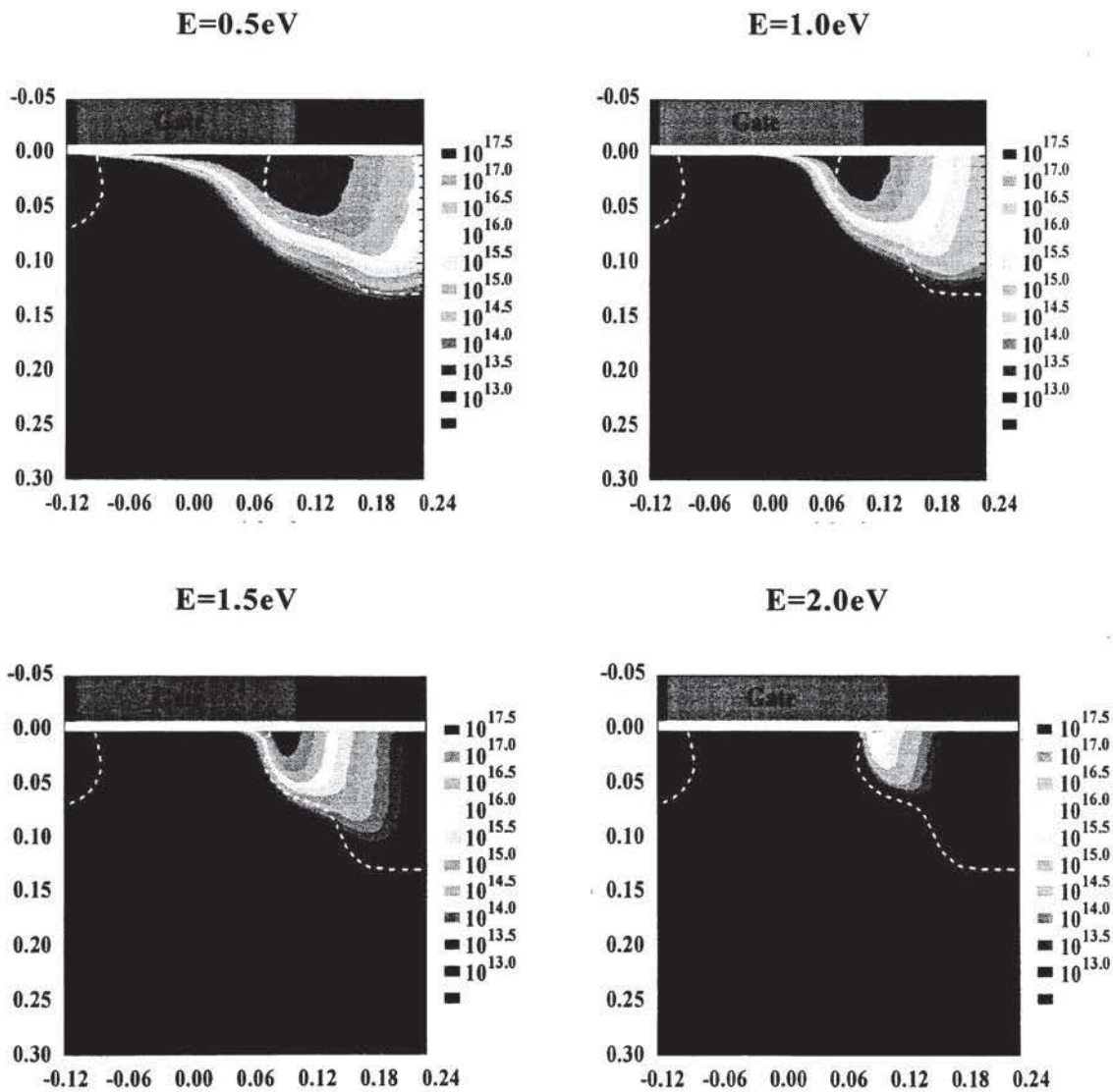


Fig. 7.1 Electron distribution in a 0.16 μm-nMOSFET at different energy level. (Vd=2.5V, Vg=1.5V, axes unit: μm, contour line unit: cm^{-3}, dashed line denotes pn-junction)

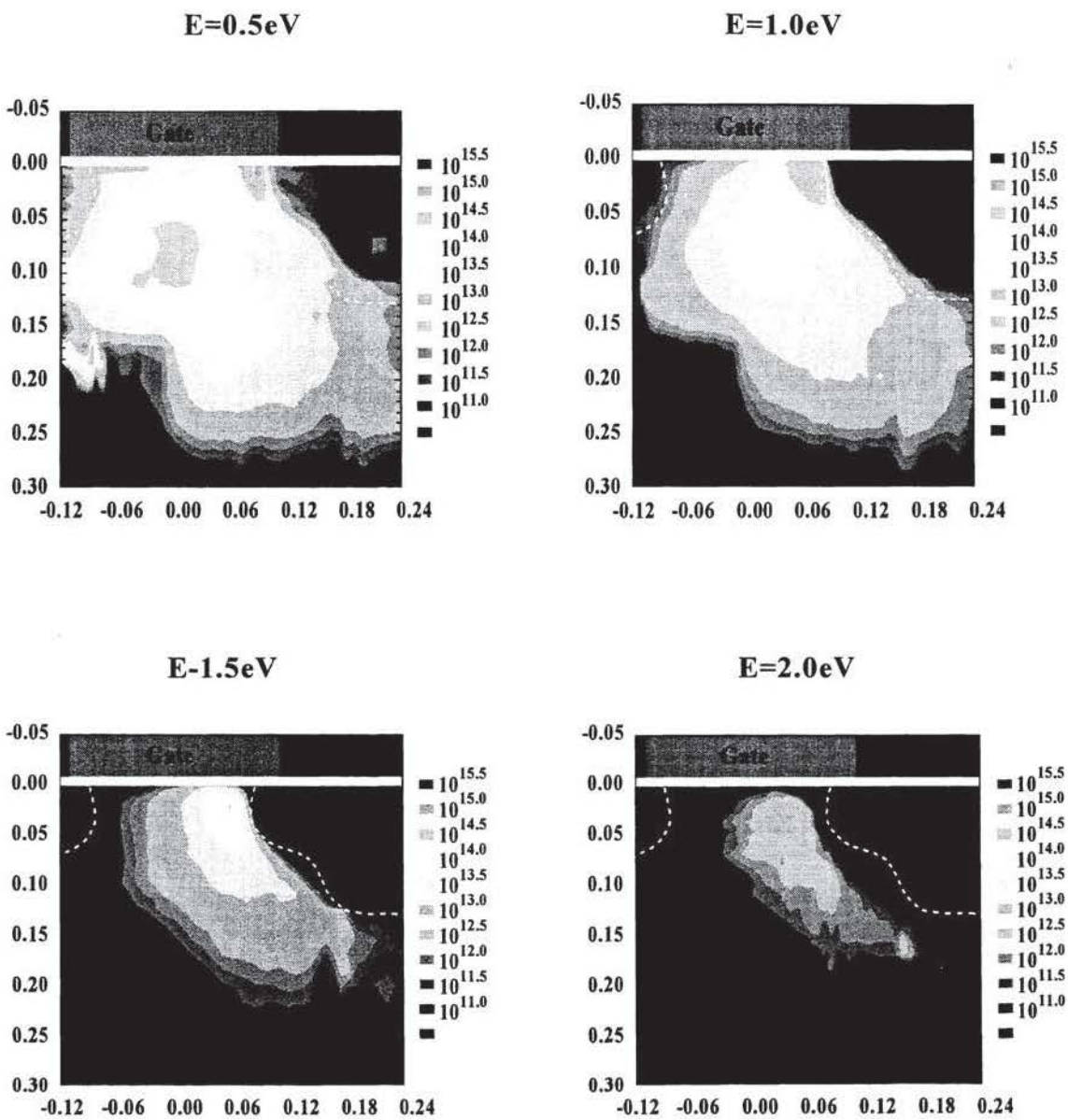


Fig. 7.2 Hole distribution in a 0.16 μ m-nMOSFET at different energy level. (Vd=2.5V, Vg=1.5V, axes unit: μ m, contour line unit: cm^{-3}, dashed line denotes pn-junction)

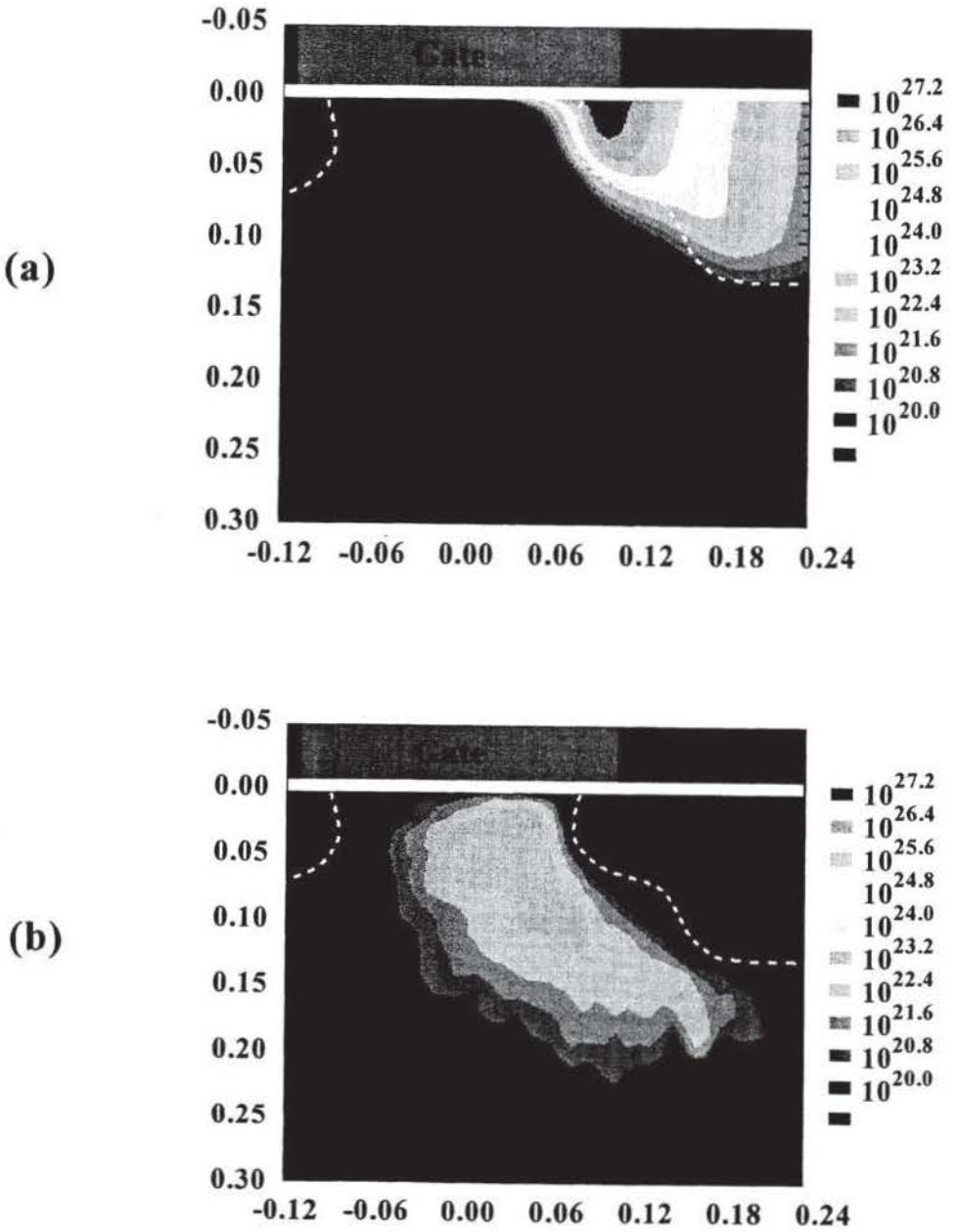


Fig. 7.3 II generation rate caused by (a) electron and (b) hole. (Vd=2.5V, Vg=1.5V, axes unit: μ m, contour line unit: $s^{-1}cm^{-3}$, dashed line denotes pn-junction)

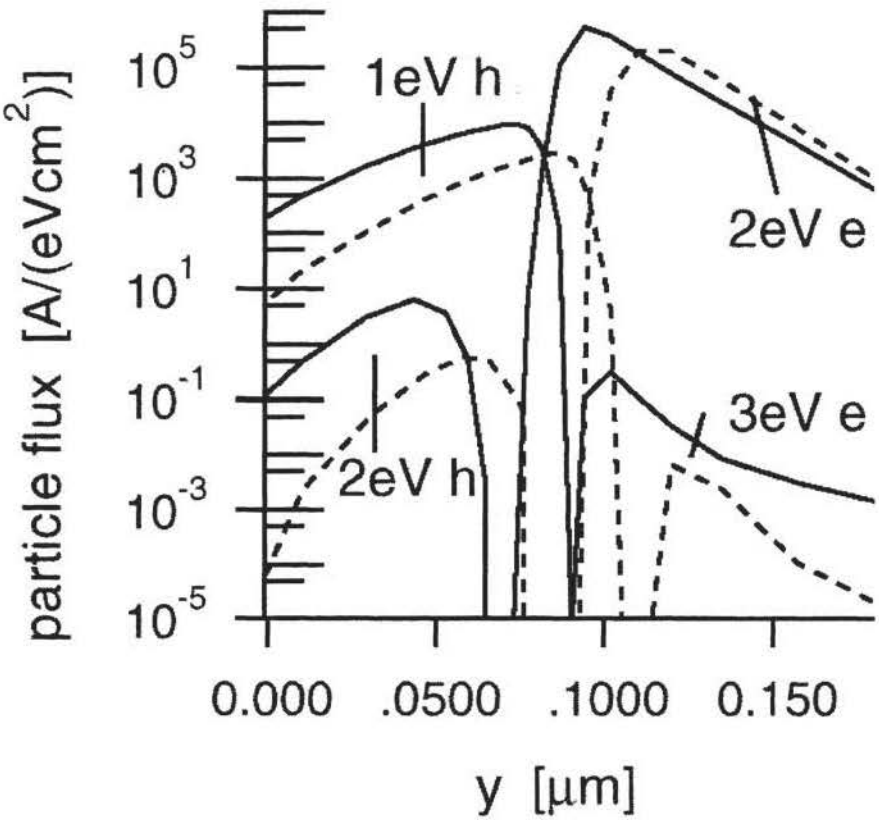


Fig. 7.4 Currents of electrons and holes with certain energies hitting the Si/SiO_2 interface for the 0.18 μ m- (dashed line) and 0.16 μ m- (solid line) nMOSFETs.

7.3 Physics-Based MOSFET Design

To extract the highest performance of MOSFETs, not only hot carrier effect, but also short channel effects must be taken into account. The optimization of threshold voltage (*Vth*) roll-off is the most important topic among the short channel effects because the process margin to keep the *Vth* roll-off small enough directly restricts CMOS circuit performance. To increase the saturation drain current with less threshold voltage roll-off, so-called channel engineering should be considered. There are roughly two approaches for the channel engineering, or one is the steep-retrograde channel profile and the other is the higher peak concentration profile at the channel surface as shown in Fig. 5.6. The advantage of the former is discussed as in [139], but this profile is not always suitable to prevent the short channel

effect or the *Vth* roll-off. By using GHDM implemented in GALENE-III, simulations are performed to clarify which approach is suitable for a 0.13-μ m CMOS technology.

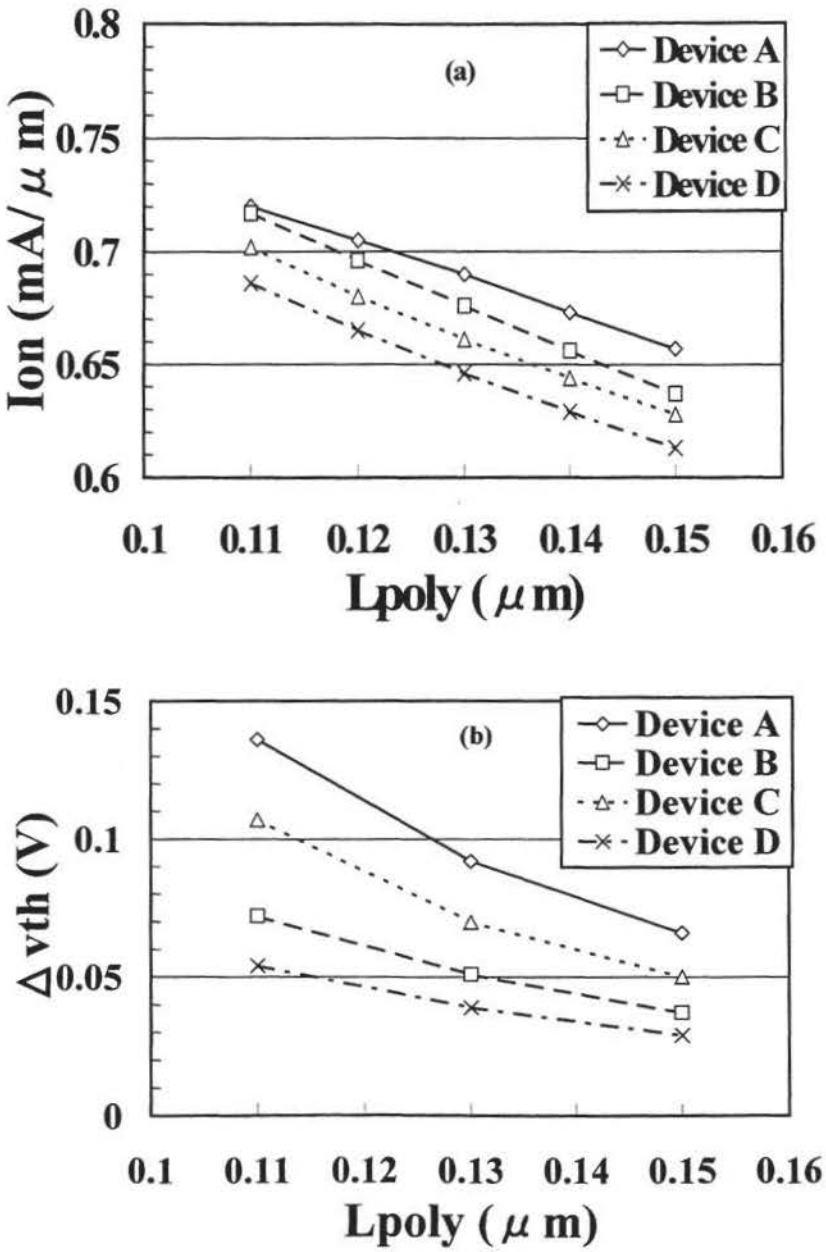


Fig. 7.5 Simulated (a) Ion and (b) ΔVth versus poly-gate length characteristics for four kinds of nMOSFET. (A: surface peak channel with source/drain extension x_j=0.05 μ m, B: surface peak channel with source/drain extension x_j=0.03 μ m, C: steep retrograde channel with source/drain extension x_j=0.05 μ m, D: steep retrograde channel with source/drain extension x_j=0.03 μ m)

Figure 7.5 shows the comparison between these two channel profiles combined with two kinds of source/drain extension junction depth (x_j=0.03 and 0.05 μ m) for (a) the saturation drain current (*Ion* when each off current is limited to 1nA/μ m) and (b) *Vth* roll-off (ΔVth defined as the threshold voltage difference when poly-gate length is 10% reduced) as a function of the physical poly-gate length (*Lpoly*)[140]. These results clearly show that the surface peak channel with the shallower source/drain extension junction is superior to the steep-retrograde channel profile with both deep and shallow source/drain extension junction from the viewpoint of higher drain current with less *Vth* roll-off. It is because the surface peak channel profile can achieve higher pinch-off voltage that directly brings higher driving current, and the shallower source/drain extension improves the *Vth* roll-off. Moreover, in the saturation regime, impurity scattering is no longer dominant limiting factor for drain current even in the higher channel concentration at the surface although one of the reasons to use the steep retrograde channel profile is to avoid the impurity scattering [139]. From this result, we can understand the importance to realize shallower source/drain extension junction compared to the manipulation of channel doping profile.

7.4 Conclusion

By using the physically precise particle-based and fluid-based device simulators, the next generation MOSFET structure has been analyzed. The tradeoff between the device performance and hot carrier problems or short channel effects are quantitatively approached and solved to clarify the cost competitive MOSFET development direction for the next generation CMOS technology, and the importance of shallower source/drain extension junction has been pointed out.

Chapter 8 Conclusion

8.1 Summary

The development and advanced physics-based design of high-speed silicon devices from BJT to CMOS have been described. In the 1980s, the BJT played the major role in the high-speed field because of its thin base width combined with sophisticated self-aligned device structures without depending on tight lithographic design rules. Especially, the ECL circuit was the most suitable for high-speed operation because of its non-saturating BJT operation. However, its larger power dissipation was the most serious limiting factor for larger integration on a chip. Moreover, the circuit performance was no longer limited by a basic gate delay, but loaded one in the end of the 1980s. At the beginning of the 1990s the transition from BJT to CMOS occurred in the high-speed device field. MOSFET performance in CMOS technology caught up with that of BJT by employing up-to-date microfabrication technology. Now the CMOS is the dominant high-speed and ultra large scale integration device.

For high-speed CMOS devices, the device design methodology not only based on inverse modeling to extract device structures consistent with all kinds of electrical measurement results but also based on simulations by generalized hydrodynamic model and full-band Monte Carlo model has been described. The background and concept of these approaches have also been discussed and their necessity has been clarified. Moreover, hot carrier modeling has been investigated by employing full-band Monte Carlo device simulation. Also, has been clarified the fact there is no experimental evidence for the difference between the surface and bulk impact ionization in silicon. The difference was only caused by an unsound application of local field model and just was an artifact.

Finally, by using these sophisticated models the saturation drain current as well as hot carrier effects of subquarter micron MOSFETs have been analyzed. MOSFET design strategy for the 0.1-μm regime has been shown, and the importance of shallow junction for source/drain extension regions has been made clear.

8.2 Future Work

Today's most advanced microfabrication technology has brought us deep submicron era, and theoretically system-on-a-chip may be achievable. However, the CMOS technology has begun to suffer from the same problems as those silicon BJT once faced, or the larger power dissipation and by far larger loaded gate delay than basic one. These problems affect more severe crises in circuit and/or system design domain, but from the device technology point of view, there are still several concerns which must be resolved. Toward the sub-0.1 μ m CMOS technology, the following topics should be surveyed and made clear.

a) The self-consistent solution system for the BTE, Poisson and Schroedinger equations should be implemented, at least, in a two-dimensional device simulator which is feasible as an engineering tool to more accurately comprehend carrier transport phenomena in the inversion layer. The non-equilibrium phenomena such as velocity overshoot or quasi-ballistic carrier transport may affect the drain current characteristics of sub-0.1 μ m MOSFETs. To extract the enough driving capability with lower supply voltage, accurate modeling of carrier transport in the inversion layer for high tangential field regime is strongly required.

b) The limit of the miniaturization of MOSFET must be surveyed from the viewpoint of on-current and off-current ratio to keep enough circuit operation margin for CMOS technology. At the same time, the optimum supply voltage and threshold voltage should be clarified for sub-0.1 μ m MOSFETs if they are useful for practical application. In the case of sub-0.1 μ m MOSFETs the short channel effect especially caused by drain-induced barrier lowering becomes the dominant limiting factor. The device design must solve the problem with enough process latitude for practical production circumstance in a cost competitive way.

c) Hole transport mechanism should be investigated by experimentally consistent physical model to realize predictive device simulation for pMOSFETs. The valance band of silicon is more complicated than the conduction band. Accurate scattering models should be developed to express the hole transport in pMOSFETs.

d) Carrier and discrete-impurity-ion scattering model should be developed

and implemented in the above-mentioned device simulation program to evaluate the fluctuation of channel-dopant-ion's position. The channel doping of the sub-0.1 μ m MOSFETs is strongly related to the above-mentioned topic b). To reveal the channel structure avoiding the on-off current ratio problem, the influence of dopant-ion locations should be investigated whether an undoped channel is superior to a doped channel, and if so, what kind of device structure should be employed for the sub-0.1 μ m MOSFETs including sophisticated device structures such as double-gated SOI.

These research topics will reveal the future direction of the semiconductor industry based on CMOS technology and what will be achievable in the 21st century.

Bibliography

1.J. A. Appels and M. M. Paffen, "Local Oxidation of Silicon; New Technological Aspects," Phillips Res. Rep., vol.26, pp.157-165, 1971.

2.M. Takagi, K. Nakayama, C. Terada, and H. Kamioka, "Improvement of Shallow Base Transistor Technology by Using a Doped Poly-silicon Diffusion Source," 4th Conf. on Solid State Devices, Tokyo, pp.101-109, 1972.

3.K. Okada, K. Aomura, M. Suzuki, and H. Shiba, "PSA - A New Approach for Bipolar LSI," IEEE J. Solid State Circuits, vol.SC-13. no.5, pp.693-698, 1978.

4.J. A. Bondur and H. B. Pogge, "Method for Forming Isolated Regions of Silicon Utilizing Reactive Ion Etching," U.S. patent 4,104,086, Aug. 1978.

5.H. Goto, T. Takada, R. Abe, Y. Kawabe, K. Oami and M. Tanaka, "An Isolation Technology for High Performance Bipolar Memories -- IOP-II," IEEE IEDM Tech. Dig., pp.58-61, 1982.

6.G. C. Schwartz and P. M. Schaible, "Reactive Ion Etching of Silicon," J. Vac. Sci. Technol. 16(2) pp.410-413, 1979.

7.H. B. Pogge, J. A. Bondur and P. J. Burkhardt, "Reactive Ion Etching of Silicon with Cl_2/Ar," J. Electrochem. Soc. vol.130, pp.1592-1597, 1983.

8.A. Hayasaka, Y. Tamaki, M. Kawamura, K. Ogiue and S. Ohwaki, "U-groove Isolation Technique for High Speed Bipolar VLSIs," IEEE IEDM Tech. Dig., pp.62-65, 1982.

9.H. Goto, T. Takada, K. Nawata and Y. Kanai, "A New Isolation Technology for Bipolar VLSI Logic Circuits (IOP-L)," Dig. of Tech. Papers, Symposium on VLSI Technology, pp.42-43, 1985.

10.H. Sakai, K. Kikuchi, S. Kameyama, M. Kajiyama and T. Komeda, " A New Trench Isolation Technology for High-Speed and Low-Power Dissipation Bipolar LSIs," Dig. of Tech. Papers, Symposium on VLSI Technology, pp.17-18, 1987.

11.T. Shibata, R. Nakayama, K. Kurosawa, S. Onga, M. Konaka and H. Iizuka, "A Simplified BOX (buried oxide) Isolation Technology for Megabit Dynamic Memories," IEEE IEDM Tech. Dig., pp.27-30, 1983.

12.D. D. Tang, P. M. Solomon, T. H. Ning, R. D. Isaac and R. E. Burger, "1.25 μ m Deep-Groove-Isolated Self-Aligned ECL Circuits," IEEE ISSCC Tech. Dig., pp.242-243, 1982.

13.S. F. Chu, G. R. Srinivasan, H. Bhatia, B. M. Kemlage, F. Barson, J. Mauer and J. Riseman, "A Self-Aligned Bipolar Transistor," Dig. of Tech. Papers, VLSI Science and Technology, pp.306-314, 1982.

14.Y. Tamaki, T. Shiba, N. Homma, S. Mizuno and A. Hayasaka, "New U-groove Isolation Technology for High-Speed Bipolar Memory," Dig. of Tech. Papers, Symposium on VLSI Technology, pp.24-25, 1983.

15.D. D. Tang, G P. Li, C. T. Chuang, D. Danner, M. B. Ketchen, J. Mauer, M. Smyth, M. Manny, J. D. Cressler, B. Ginsberg, E. Petrillo, T. H. Ning, C. C. Hu and H. S. Park, "73ps Si Bipolar ECL Circuits," IEEE ISSCC Tech. Dig., pp.104-105, 1986.

16.G P. Li, T. H. Ning, C. T. Chuang, M. B. Ketchen, D. D. Tang and J. Mauer: "An Advanced High-Performance Trench-Isolated Self-Aligned Bipolar Technology," IEEE Trans. Electron Devices, vol.34, pp.2246-2253, no.11, 1987.

17.M. Vora, Y. L. Ho, S. Bhamre, F. Chien, G Bakkar, H. Hingrarh, C. Schmitz, "A Sub-100 Picosecond Bipolar ECL technology," IEEE IEDM Tech. Dig., pp.34-37, 1985.

18.H. Sadamatsu, M. Inoue, A. Matsuzawa, A. Kanda and H. Shimoda, "New Self-Aligned Complementary Bipolar Transistors Using Selective-Oxidation Mask," IEEE IEDM Tech. Dig., pp.753-756, 1984.

19.K. Ooami, M. Tanaka, Y. Sugo, R. Abe and T. Takada, "A 3.5ns 4kbECL RAM," IEEE ISSCC Tech. Dig., pp.114-115, 1983.

20.K. Toyoda, M. Tanaka, H. Isogai, C. Ohno, Y. Kawabe and H. Goto, "A 15ns 16Kb ECL RAM with a PNP Load Cell," IEEE ISSCC Tech. Dig., pp.108-109, 1983.

21.K. Toyoda, M. Tanaka, H. Isogai, C. Ohno, Y. Kawabe and H. Goto, "A High Speed 16Kbit ECL RAM," IEEE Journal of Solid-State Circuits vol.SC-18, no.5, pp.509-514, 1983.

22.Y. Sugo, M. Tanaka, Y. Mafune, T. Takeshima, S. Aihara and K. Tanaka, "An ECL 2.8ns 16k RAM with 1.2k Logic Gate Array," IEEE ISSCC Tech. Dig., pp.256-257, 1986.

23.Y. Okajima, K. Tokuda, K. Awaya, K. Tanaka and Y. Nakamura, "64kb ECL RAM with Redundancy," IEEE ISSCC Tech. Dig., pp.48-49, 1985.

24.T. Awaya, K. Tokuda, O. Nomura, Y. Nakaya, K. Tanaka and H. Sugawara, "A 5ns Access Time 64kb ECK RAM," IEEE ISSCC Tech. Dig., pp.130-131, 1987.

25.H. Goto and K. Inayoshi, "Trench Isolation Schemes for Bipolar Devices - Benefits and Limiting Aspects," Electronics and Photonics, vol.27, pp.61-77, 1988.

26.Y. Tamaki, S. Isomae, K. Sagara, T. Kure and M. Kawamura, "Evaluation of Dislocation Generation in U-groove Isolation," J. Electrochem. Soc. vol.135, pp.726-730, 1988.

27.C. W. Teng, C. Slawinski and W. R. Hunter, "Defect Generation in Trench Isolation," IEEE IEDM Tech. Dig., pp.586-589, 1984.

28.K. Hashimoto, Y. Nagakubo, S. Yokogawa, M. Kakumu., M. Kinugawa, K. Sawada, T. Sakurai, M. Isobe, J. Matsunaga and H. Iizuka, "Deep Trench Well Isolation for 256kb 6T CMOS Static RAM," Dig. of Tech. Papers, Symposium on VLSI Technology, pp.24-25, 1983.

29.M. Sugiyama, T. Shimizu, H. Takemura, A Yoshino, N. Oda, T. Tashiro, Y. Takahashi and M. Nakamae, "Bipolar VLSI Memory Cell Technology Utilizing BPSG-Filled Trench Isolation," Dig. of Tech. Papers, Symposium on VLSI Technology, pp.59-60, 1989.

30.T. Sakai, Y. Kobayashi, H. Yamauchi, M. Sato, and T. Makino, "High Speed Bipolar ICs Using Super Self-Aligned Process Technology," 12th Conf. on Solid State Devices, Tokyo, pp.155-159, 1980.

31.T. Sakai, S. Konaka, Y. Kobayashi, M. Suzuki, and Y. Kawai, "Gigabit Logic Bipolar Technology: Advanced Super Self- Aligned Process Technology," Electronics Letters, vol.19, no.8, pp.283-284, 1983.

32.S. Konaka, Y. Amemiya, K. Sakuma, and T. Sakai, "A 20ps/GBipolar IC Using Advanced SST with Collector Ion Implantation," 19th Conf. on Solid State Devices and Materials, Tokyo, pp.331-334,1987.

33.S. H. Chai and J. H. Lee, "A New Self-Aligned Bipolar Transistor Using Vertical Nitride Mask," IEEE IEDM Tech. Dig., pp.26-29, 1985.

34.S. Sawada and H. Uchida, "Self-Aligned High-Speed Bipolar LSI Process Using Trench Isolation Technology," IEICE Technical Report vol.88, no.120, pp.41-46, 1988 (in Japanese).

35.Y. Yamamoto and K. Sakuma, "SDX: A Novel Self-Aligned Technique and Its Application to High Speed Bipolar LSI's," IEEE Trans. on Electron Devices, vol.35, no.10, pp.1601-1608, 1988.

36.T. H. Ning, R. D. Isaac, P. M. Solomon, D. D. Tang, H. N. Yu, G. Feth and S. K. Wiedmann, "Self-Aligned Bipolar Transistors for High-Performance and Low-Power-Delay VLSI," IEEE Trans. On Electron Devices, vol.ED-28, no.9, pp.1010-1013, 1981.

37.A. W. Wieder: "Self-Aligned Bipolar technology- New Chances for Very-High-Speed Digital Integrated Circuits," Siemens Forsch.-u. Entwickl.-Ber. Bd.13, Nr.5, pp.246-252, 1984.

38.K. Kikuchi, S. Kameyama, M. Kajiyama, M. Nishio, and T. Komeda, "A High-Speed Bipolar Process Using Self-Aligned Double Diffusion Polysilicon Technology," IEEE IEDM Tech. Dig., pp.420-423, 1986.

39.P. J. Zdebel, R. J. Balda, B-Y. Hwang, V. d. l. Torre and A. Wagner, "MOSAIC-III -A

High Performance Bipolar Technology with Advanced Self-Aligned Devices," IEEE BCTM Tech. Dig., pp.172-175, 1987.

40.K. Ueno, H. Goto, E. Sugiyama and H. Tsunoi, "A Sub-40 ps ECL Circuit at a Switching Current of 1.28mA," IEEE IEDM Tech. Dig., pp.371-374, 1987.

41.M. Sugiyama, H. Takemura, C. Ogawa, T. Tashiro, T. Morikawa and M. Nakamae, "A 40 GHz f_T Silicon Bipolar Transistor LSI Technology," IEEE IEDM Tech. Dig., pp.221-224, 1989.

42. T. Tashiro H. Takemura, T. Kamiya, F. Tokuyoshi, S. Ohi, H. Shiraki, M. Nakamae and T. Nakamura, "An 80ps ECL Circuit with High Current Density Transistor," IEEE IEDM Tech. Dig., pp.686-689, 1984.

43.H. J. Chen, S. Z. Nawaz, B. D. Urke and H. P. Vyas, "Switching Characteristics of Poly Bipolar Circuits at Liquid Nitrogen Temperatures," IEEE BCTM Tech. Dig., pp.215-218, 1988.

44. Y. Okita, M. Shinozawa, A. Kawakatsu, Y. Umemura, K. Yamaguchi and K. Akahane, "A Novel Base-Emitter Self-Alignment Process for High Speed Bipolar LSIs," IEEE CICC Tech. Dig., pp.22.4.1-22.4.4, 1988

45.S. Duncan, M. C. Wilson, P. C. Hunt and D. J. Bazley, "A 1 μ m Trench High Speed Bipolar Transistor," Tech. Dig. of Symp. on VLSI Tech., pp.87-88, 1988.

46. H. J. Chen, S. Z. Nawaz, B. D. Urke and H. P. Vyas, "Switching Characteristics of Poly Bipolar Circuits at Liquid Nitrogen Temperature," IEEE BCTM Tech. Dig., pp.215-218, 1988.

47.T. Gomi, H. Miwa, H. Sasaki, H. Yamamoto, M. Nakamura and A. Kayanuma, "A Sub-30psec Si Bipolar LSI Technology," IEEE IEDM Tech. Dig., pp.744-747, 1988.

48.G P. Li, E. Hackbarth and T. C. Chen, "Identification and Implication of a Perimeter Tunneling Current Component in Advanced Self-Aligned Bipolar Transistors," IEEE Trans. on Electron Devices, vol.35, no.1, pp.89-95, 1988.

49.G P. Li., C. T. Chuang, T. C. Chen and T. H. Ning, "On the Narrow-Emitter Effect of Advanced Shallow-Profile Transistors," IEEE Trans. on Electron Devices, vol.35, no.11, pp.1942-1950, 1988.

50.C. T. Chuang, "The Effect of Extrinsic Base Encroachment on the Switch-On Transient of Advanced Narrow-Emitter Bipolar Transistors," IEEE Trans. on Electron Devices, vol.35, no.3, pp.309-313, 1988.

51.C. T. Chuang, D. D. Tang, G.. P. Li, and E. Hackbarth, "On the Punchthrough Characteristics of Advanced Self-Aligned Bipolar Transistors," IEEE Trans. on Electron Devices, vol.34, no.7, pp.1519-1524, 1987.

52.T. Yamaguchi, Y-C. S. Yu, E. E. Lane, J. S. Lee, E. E. Patton, R. D. Herman, D. R. Ahrendt, V. F. Drobny, T. H. Yuzuriha and V. E. Garuts, "Process and Device Performance of a High-Speed Double Poly-Si Bipolar Technology Using Borsenic-Poly Process with Coupling-Base Implant," IEEE Trans. on Electron Devices, vol.35, no.8, pp.1247-1255, 1988.

53.S. A. Petersen and G. P. Li, "Hot Carrier Effects in Advanced Self-Aligned Bipolar Transistors," IEEE IEDM Tech. Dig., pp.22-25, 1985.

54. T. C. Chen, C. Kaya, M. B. Ketchen and T. H. Ning, "Reliability Analysis of Self-Aligned Bipolar Transistor under Forward Active Current Stress," IEEE IEDM Tech. Dig., pp.650-653, 1986.

55.T. C. Chen, J. P. Norum and T. H. Ning, "Comparison of Effects of Ionizing Radiation and High-Current Stress on Characteristics of Self-Aligned Bipolar Transistors," Conf. on Solid State Devices and Materials, Tokyo, pp.523-526, 1988.

56. K. A. Jenkins and J. D. Cressler, "Electron Beam Damage of Advanced Silicon Bipolar Transistors and Circuits," IEEE IEDM Tech. Dig., pp.30-33, 1988.

57. K. Ohno, K. Ooami, T. Takada and H. Goto, Nikkei Microdevices, no.46, pp.42-49, Apr. 1989 (in Japanese).

58. N. Miyoshi, M. Yoshida, K. Suzuki, M. Kokado, M. Takaoka and H. Harada, "A 50K-Gate ECL Array with Substrate Power Supply," IEEE ISSCC Tech. Dig., pp.182-183, 1989.

59.H. Tokuda, T. Tanizawa, K Hirochi, K. Kawauchi and T. Deguchi: "A 100k-Gate ECL Standard Cell LSI with layout System," IEEE ISSCC Tech. Dig., pp.94-95, 1990.

60.I. Ishida, K. Aomura and T. Nakamura, "An Advanced PSA Process for High Speed Bipolar VLSI," IEEE IEDM Tech. Dig., pp.336-339, 1979.

61.T. Hirao, T. Ikeda and N. Katoh, "A 2.1-GHz 56mW Two-Modulus Prescaler IC Using Salicide Base Contact Technology," Conf. on Solid State Devices and Materials, Tokyo, pp.381-384, 1985.

62.T. Sakai, Y. Sunohara, Y. Sakakibara and J. Murota, "Stepped Electrode Transistor: SET," Conf. on Solid State Devices, Tokyo, pp.43-46, 1976.

63.T. Sakai, Y. Yamamoto, Y. Kobayashi, H. Yamauti, T. Ishitani and T. Sudo, 2Elevated Electrode Integrated Circuits," IEEE Trans. on Electron Devices, vol.ED-26, no.4, pp. 379-384, 1979.

64.D. D. Tang, T. H. Ning, R. D. Isaac, G. C. Feth, S. K. Wiedmann, and H-N. Yu, "Subnanosecond Self-Aligned I^2L/ MTL Circuits," IEEE Trans. on Electron Devices, vol.ED-27, no.8, pp. 1379-1384, 1980.

65.T. C. Chen, D. D. Tang, C. T. Chuang, J. D. Cressler, J. Warnock, G. P. Li, P. E. Biolsi, D. A. Danner, M. R. Polcari and T. H. Ning, "A Sub-50ps Single Poly Planar Bipolar Technology," IEEE IEDM Tech. Dig., pp.740-743, 1988.

66.T. C. Chen, C. T. Chuang, G. P. Li, S. Basvaiah, D. D. Tang, M. B. Ketchen and T. H. Ning, "An Advanced Bipolar Transistor with Self-Aligned Ion-Implanted Base and W/Poly Emitter," IEEE Trans. on Electron Devices, vol.35, no.8, pp.1322-1327, 1988.

67.T. Nakamura, T. Miyazaki, S. Takahashi, T. Kure, T. Okabe and M. Nagata, "Self-Aligned Transistor with Sidewall Base Electrode," IEEE ISSCC Tech. Dig., pp.214-215, 1981.

68.K. Washio, T. Nakamura, K. Nakazato and T. Hayashida, "A 48psECL in a Self-Aligned Bipolar Technology," IEEE ISSCC Tech. Dig., pp.58-59, 1987.

69.K. Washio, T. Nakamura and T. Hayashida, "Fabrication Process and Device Characteristics of Sidewall Base Contact Structure Transistor Using Two-Step Oxidation of Sidewall Surface," IEEE Trans. on Electron Devices, vol.35, no.10, pp.1596-1600, 1988.

70.N. Ou-uchi, A. Kayanuma, K. Asano, H. Hayashi and M. Noda, "A New Self-Aligned transistor Structure for High-Speed and Low-Power Bipolar LSI's," IEEE IEDM Tech. Dig., pp.55-58, 1983.

71. F. Sato, T. Tatsumi, T. Hashimoto and T. Tashiro, "A Super Self-Aligned Selectively Grown SiGe Base (SSSB) Bipolar Transistor Fabricated by Cold-Wall Type UHV/CND Technology," IEEE Trans. on Electron Devices, vol.41, no.8, pp.1373-1378, 1994.

72. T. F. Meister, H. Schaefer, M. Franosch, W. Molzer, K. Aufinger, U. Scheler, C. Walz, M. Stolz, S. Boguth and J. Boeck: "SiGe Base Bipolar Technology with 74GHz fmax and 11ps Gate Delay," IEEE IEDM Tech. Dig., pp.739-742, 1995.

73.S. L. Miller, "Transistor," U.S. Patent 3,220,896 Nov. 1965.

74.H. N. Yu, "Transistor with Limited Area Base-Collector Junction," U.S. Patent 3,312,881, Apr. 1967, U.S. Patent Re. 27,045, Feb. 1971.

75.H. N. Ghosh, K. G. Ashar, A. S. Oberai and D. DeWitt, "Design and development of an ultras low-capacitance, high-performance pedestal transistor," IBM J. Res. Develop. vol.15, pp.436-441, 1971.

76.J. F. Ziegler, B. L. Crowder, and W. J. Kleinfelder, "Experimental evaluation of high energy ion implantation gradients for possible fabrication of a transistor pedestal collector," IBM J. Res. Develop. vol.15, pp.452-456, 1971.

77.D. D. Tang, P. M. Solomon, T. H. Ning, R. D. Isaac, and R. E. Burger, "1.25 μ m deep-groove-isolated self-aligned bipolar circuits," IEEE J. Solid-State Circuits, vol.SC-17,

no.5, pp.925-931, 1982.

78.P-F. Lu, and T-C. Chen, "Collector-base junction avalanche effects in advanced double-poly self-aligned bipolar transistors," IEEE Trans. Electron Devices, vol.36, no.6, pp.1182-1188, 1989.

79.Y. Nagase, H. Goto and T. Takada, "Analysis of pedestal- collector transistors by 2-d process and device simulation," Proc. of 7th VLSI Process/Device Modeling Workshop, pp.40-41, 1990.

80.Y. Nagase, K. Hashimoto, H. Goto, T. Nakadai, T. Nawata, and T. Takada, "Analysis of bipolar transistor's f_T-*Ic* characteristics by two dimensional dc and ac device simulation," Proc. of Workshop of Numerical Modeling of Processes and Devices for Integrated Circuits-- NUPAD-II, NUPAD-156, 1988.

81.H. Goto, "High-Speed Silicon Bipolar ICs," in "High-Speed Digital IC Technologies" edited by M. Rocchi, Artech House Inc., Ch.3, 1990.

82.H. Katakura, A. Tahara, M. Yoshida, T. Deguchi and S. Sudo, "15.8 GHz Si bipolar static divider," Spring National Convention Record of IEICE, C-647, p5-211, 1990 (in Japanese)

83.H. Goto, "Future Bipolar Device Structures," Journal de Physique, Colloque C4, supplement au n°9, Tome 49, pp.471-480, 1988.

84.K. Ueno, Y. Arimoto, N. Odani, M. Ozeki and K. Imaoka, "A Fully Functional 1K ECL RAM on a Bonded SOI Wafer," IEEE IEDM Tech. Dig., pp.870-872, 1988.

85. T. Kikuchi, Y. Onishi, T. Hashimoto, E. Yoshida, H. Yamaguchi, S. Wada, N. Tamba, K. Watanabe, Y. Tamaki and T. Ikeda, "A 0.35 μ m ECL-CMOS Process technology on SOI for 1ns Mega-bits SRAM's with 40ps Gate Array," IEEE IEDM Tech. Dig., pp.923-926, 1995.

86. F. Sato, H. Tezuka, M. Soda, T. Hashimoto, T. Suzaki, T. Tatsumi, T. Morikawa and T. Tashiro, "The Optical terminal IC: A 2.4Gb/s Receiver and 1 1:16 Demultiplexer in One Chip," IEEE BCTM Tech. Dig., pp.162-165, 1995.

87. K. Y. Toh, C. T. Chuang, T. C. Chen, J. Warnock, G. P. Li, K. Chin and T. H. Ning, "A 23ps/2.1mW ECL Circuit with an AC-Coupled Active Pull-Down Emitter-Follower," IEEE ISSCC Tech. Dig., pp.224-225, 1989.

88. D. Tanksalvala et al., "A 90 MHz CMOS RISC CPU Designed for Sustained Performance," IEEE ISSCC Tech. Dig., pp.52-53, 1990.

89. J. Petrovick Jr., R. Taylor, A. Bertolet, A. Chu, T. Harroun, F. Keyser, C. LaMarche, L. Pastel, G. Richardson and B. Worth, "A 300k-Circuit ASIC Logic Family," IEEE ISSCC Tech. Dig., pp.88-89, 1990.

90. A. Masaki, "Possibilities of CMOS Mainframe and Its Impact on Technology R&D," Dig. of Tech. Papers, Symposium on VLSI Technology, pp.1-4, 1991.

91. T. Izawa, M. Katsube, Y. Yokoyama, K. Hashimoto, E. Kawamura, A. Shimizu, H. Takagi, F. Inoue, H. Shimizu, K. Furumochi, H. Goto, S. Kawamura, K. Watanabe and K. Aoyama, "A Novel Embedded SRAM Technology with 10 μm^2 full-CMOS Cells for 0.25 μ m Logic Devices," IEEE IEDM Tech. Dig., pp.941-943, 1994.

92. K. Furumochi, H. Shimize, M. Fujita, T. Akita, T. Izawa, M. Katsube, K. Aoyama and S. Kawamura, "A 500MHz 288kb CMOS SRAM Macro for On-Chip Cache," IEEE ISSCC Tech. Dig., pp.156-157, 1996.

93. Y. Takao, Y. Sanbonsugi, K. Watanabe, H. Takatsuka, T. Karasawa, E. Kawamura, K. Hashimoto, H. Takagi, F. Inoue, H. Shimizu, T. Yamazaki, H. Goto, T. Sugii, M. Miyajima, K. Watanabe and K. Aoyama, "A 4 μm^2 Full-CMOS SRAM Cell Technology for 0.2-μ m High-Performance Logic LSIs," Dig. of Tech. Papers, Symposium on VLSI Technology, pp.11-12, 1997.

94. N. Khalil and J. Faricelli, " Inverse Modeling Profile Determinations: Implementation Issues and Recent Results," Proc. Int. Conf. Simulation on Semiconductor Processes and Devices pp.19-22, 1996.

95. C. Jungemann, S. Yamaguchi and H. Goto, "Accurate Prediction of Hot-carrier Effects for a Deep-Sub-μ m CMOS Technology Based on Inverse Modeling and Full Band Monte Carlo Device Simulation," Proc. Int. Conf. Simulation on Semiconductor Processes and Devices pp.59-60, 1996.

96. Z. K. Lee, M. B. McIlrath and D. A. Antoniadis, "Inverse Modeling of MOSFETs using I-V Characteristics in the subthreshold Region," IEEE IEDM Tech. Dig., pp.683-686, 1997.

97. A. Das, D. Newmark, I. Clejan, M. Foisy, M. Sharma, S. Venkatesan, S. Veeraraghavan, V. Misra, B. Gadepally and L. Parrillo, "An advance MOSFET Design Approach and a Calibration Methodology using Inverse Modeling that Accurately Predicts Device Characteristics," IEEE IEDM Tech. Dig. pp.687-690, 1997.

98. S. Yamaguchi and H. Goto, "Inverse Modeling – A Promising Approach to Know What is Made and What should be Made," Proc. Asia and South Pacific Design Automation Conf. pp.117-121, 1998.

99. R. Thoma, A. Emunds, B. Meinerzhagen, H.-J. Peifer and W. L. Engl, "Hydrodynamic Equations for Semiconductors with Nonparaboric band Structure," IEEE Trans. Electron Devices vol.38, pp.1343-1353, 1991.

100. C. Lombardi, S. Manzini, A. Saporito and M. Vanzi, "A Physically Based Mobility

Model for Numerical Simulation of Nonplanar Devices," IEEE Trans. Computer-Aided Design vol.7, pp.1164-1171, 1988.

101. S. A. Mujtaba, R. W. Dutton and D. L. Scharfetter, "Semi-Empirical Local NMOS Mobility Model for 2-D Device Simulation Incorporating Screened minority Impurity Scattering," Proc. IEEE 5th Workshop on Numerical Modeling of Processes and Devices for Integrated Circuits pp.3-6, 1994.

102. J. A. Cooper, Jr. and D. F. Nelson, "High-field drift velocity of electrons at the Si-SiO2 interface as determined by a time-of-flight technique," J. Appl. Phys. vol.54, pp.1445-1456, 1983.

103. F. Assaderaghi, D. Sinitsky, J. Bokor, P. K. Ko, H. Gaw and C. Hu, "High-Field Transport of Inversion-Layer Electrons and Holes Including Velocity Overshoot," IEEE Trans. Electron Devices vol.44 pp. 664-671, 1997.

104. C. Jungemann, A. Emunds and W. L. Engl, "Simulation of Linear and Nonlinear Electron Transport in Homogeneous Silicon Inversion Layers," Solid-State Electron. vol.36 pp.1529-1540, 1993.

105. C. S. Rafferty, H.-H. Vuong, S. A. Eshraghi, M. D. Giles, M. R. Pinto and S. J. Hillenius, "Explanation of Reverse Short Channel Effect by Defect Gradients," IEEE IEDM Tech. Dig., pp.311-314, 1993.

106. C. Jungemann, S. Yamaguchi and H. Goto, "On the Accuracy and Efficiency of Substrate Current Calculation for Sub-μ m n-MOSFET's," IEEE Electron Device Lett. vol.17 pp.464-466, 1996.

107. C. Jungemann, B. Meinerzhagen, S. Decker, S. Keith, S. Yamaguchi and H. Goto, "Is Physically Sound and Predictive Modeling of NMOS Substrate Current Possible ?," Solid-State Electron. vol.42 pp.647-655, 1998.

108. C. Jungemann, S. Yamaguchi and H. Goto, "Is there Experimental Evidence for a Difference between Surface and Bulk Impact Ionization in Silicon ?," IEEE IEDM Tech. Dig., pp.383-386, 1996.

109. W. Haensch, "The Drift Diffusion Equation and Its Application to MOSFET Modeling," Chap.1, Wien, Springier Verlag, 1991.

110.M. V. Fischetti and S. E. Laux, "Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects," Phys. Rev. B, vol. 38, pp.9721-9745, 1988.

111.J. Bude and R. K. Smith, "Phase-space simplex Monte Carlo for semiconductor transport," in Semiconductor Science and Technology, Hot Carriers in Semiconductors, Proc. of the 8th International Conference, vol.9, pp.840-843, 1994.

112. C. Jungemann, S. Yamaguchi and H. Goto, "Efficient Full Band Monte Carlo Hot Carrier Simulation for Silicon Devices," Proc. of European Solid-State Device Research Conf. pp.821-824, 1996.

113. M. V. Fischetti and S E. Laux, "Monte Carlo Simulation of Electron Transport in Si: The First 20 Years," Proc. of European Solid State Device Research Conf. pp.813-820, 1996.

114. C. Jungemann, S. Decker, R. Thoma, W. L. Engl and H. Goto, "Phase Space Multiple Refresh: a Versatile Statistical Enhancement Method for Monte Carlo Simulation," Proc. of Int. Conf. on Simulation of Semiconductor Processes and Devices, pp.65-66, 1996.

115. C. Jungemann, S. Decker, R. Thoma, W.-L. Engl and H. Goto, "Phase Space Multiple Refresh: A General Purpose Statistical Enhancement Technique for Monte Carlo Device Simulation," IEEE J. Techno. Comp. Aided Design, (2), 1997. (http://www.ieee.org/journal/tcad/accepted/jungemann-jan97/)

116. C. Jungemann, S. Yamaguchi and H. Goto, "Convergence Estimation for Stationary Ensemble Monte Carlo Simulations," Proc. of Int. Conf. on Simulation of Semiconductor Processes and Devices, 209-212, 1997.

117. C. Jungemann, S. Yamaguchi and H. Goto, "Convergence Estimation for Stationary Ensemble Monte Carlo Simulations," IEEE J. Techno. Comp. Aided Design, (10), 1998. (http://www.ieee.org/journal/tcad/accepted/jungemann-jan98/)

118. T. Kunikiyo, M. Takenaka, Y. Kamakura, M. Yamaji, H. Mizuno, M. Morifuji, K. Taniguchi and C. Hamaguchi, "A Monte Carlo Simulation of anisotropic electron transport in silicon including full band structure and anisotropic impact-ionization model," J. Appl. Phys. vol.75, no.1, pp.297-312, 1994.

119. C. Jacoboni and L. Reggiani, "The Monte Carlo method for the solution of charge transport in semiconductors with application to covalent materials," Rev. Mod. Phys., vol. 55, pp. 645-705, 1983.

120. C. Jungemann, S. Keith, B. Meinerzhagen, and W. L. Engl, "On the influence of band structure and scattering rates on hot electron modeling," in Simulation of Semiconductor Devices and Processes, H. Ryssel and P. Pichler, Eds., Wien, Springer-Verlag, vol. 6, pp. 222-225, 1995.

121.E. Sangiorgi, B. Ricco, and F. Venturi, "MOS 2: An efficient Monte Carlo simulator for MOS devices," IEEE Trans. Computer-Aided Des., vol. 7, pp. 259-271,1988.

122.K. Kometer, G. Zandler and P. Vogl, "Lattice-gas cellular-automaton method for semiclassical transport in semiconductors," Phys. Rev. B vol.46, no.3, p. 1382-1394, 1992.

123.J. M. Higman, K. Hess, C. G Hwang and R. W. Dutton, "Coupled Monte Carlo-Drift Diffusion Analysis of Hot-Electron Effects in MOSFET's," IEEE Trans. Electron Devices, vol. 36, pp. 930-937, 1989.

124.A. Papoulis, "Probability, Random Variables and Stochastic Processes," McGraw-Hill, 3rd ed. 1991.

125. M. V. Fischetti, S. E. Laux, and E. Crabbe, "Understanding hot-electron transport in silicon devices: is there a shortcut ?," J. Appl. Phys., vol. 78, pp. 1058-1087, 1995.

126. R. Stratton, "Diffusion of Hot and Cold Electrons in Semiconductor Barriers," Phys. Rev. vol.126, no.6, pp.2002-2014, 1962.

127. K. Blotekjaer, "Transport Equation for Electrons in Two-Valley Semiconductors," IEEE Trans. Electron Devices vol.ED-17, no.1, pp.38-47, 1970.

128. Y. Apanovich, E. Lyumikis, B. Polsky, A. Shur and P. Blakey, "Steady-State and Transient Analysis of Submicron Devices Using Energy Balance and Simplified Hydrodynamic Model," IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems vol.13, no.6, pp.702-711, 1994.

129. S. E. Laux and M. Fischetti, "Transport Models for Advanced Device Simulation – Truth or Consequence ?," IEEE BCTM Tech. Dig., pp.27-34, 1995.

130. W. L. Engl, Private communication.

131. H. Kanata, C. Jungemann and S. Satoh, "Parameter Extraction of Hydrodynamic Model by Using Fullband Monte Carlo Simulator," Tech. Rep. IEICE VLD97-74, pp.113-118, 1997 (in Japanese).

132. A. G Chynoweth, "Uniform silicon p-n junctions. II. ionization rates for electrons," J. Appl. Phys. vol.31, pp.1161-.1165, 1960.

133. J. W. Slotboom, G Streutker, G J. T. Davids and P. B. Hartog, "Surface Impact Ionization in Silicon Devices," IEEE IEDM Tech. Dig. pp.494-497, 1987.

134. T. Thurgate and N. Chan, "An Impact Ionization Model for Two-Dimensional Device Simulation," IEEE Trans. Electron Devices vol.32, no.2, pp.400-404, 1985.

135. W. Shockley, "Problems related to p-n junctions in silicon," Solid-State Electron. vol.2, pp.35-67, 1961.

136. J. M. Higmann, I. C. Kizilyalli and K. Hess, "Nonlocality of the Electron Ionization Coefficient in n-MOSFET's: An Analytic Approach," IEEE Electron Device Lett. vol.9 pp.399-401, 1988.

137. K. Taniguchi, M. Yamaji, K. Sonoda, T. Kunikiyo and C. Hamaguchi, "Monte Carlo Study of Impact Ionization Phenomena in Small Geometry MOSFET's," IEEE IEDM Tech. Dig. pp.355-358, 1994.

138. B. Meinerzhagen, C. Jungemann, S. Decker, S. Keith, S. Yamaguchi and H. Goto, "Predictive and Efficient Modeling of Substrate Currents in n-channel MOS-Transistors," Proc. of International Symposium on VLSI Technology, Systems and Applications, pp.232-235, 1997.

139. Y. Taur, D. A. Buchanan, W. Chen, D. J. Frank, K. E. Ismail, S.-H. Lo, G A. Sai-Halasz, R. G. Viswanathan, H.-J. C. Wann, S. J. Wind and H.-S. Wong, "CMOS Scaling into the Nanometer Regime," Proc. of IEEE vol.85, no.4, pp.486-504, 1997.

140. S. Yamaguchi and H. Goto, "Fujitsu's Transistor Structure since 0.18 μ m," Semiconductor World, vol.17, no.5, pp.81-85, May, 1998 (in Japanese).

List of Papers

1. Major Journal Papers

1. **H. Goto** and K. Inayoshi, "Trench Isolation Schemes for Bipolar Devices - Benefits and Limiting Aspects," Electronics and Photonics, vol.27, pp.61-77, 1988.
2. **H. Goto,** "Future Bipolar Device Structures (**invited**)," Journal de Physique, Colloque C4, supplement au n°9, Tome 49, pp.471-480, septembre 1988.
3. **H. Goto**, Y. Nagase, T. Takada, A. Tahara and Y. Momma, "Analysis of Highly-Doped-Collector Transistors by Using Two-Dimensional Process/Device Simulation and its Application of ECL Circuits," IEEE Trans. Electron Devices, vol.38, no.8, pp.1840-1844, 1991.
4. C. Jungemann, B. Meinerzhagen, S. Decker, S. Keith, S. Yamaguchi and **H. Goto,** "Is Physically Sound and Predictive Modeling of NMOS Substrate Currents Possible ?," Solid-State Electron., vol.42, no.4, pp.647-655, 1998.
5. **H. Goto**, S. Yamaguchi and C. Jungemann, "Inverse Modeling as a Basis for Predictive Device Simulation of Deep Sub-micron Metal-Oxide-Semiconductor Field Effect Transistors," Jpn. J. Appl. Phys., vol.37, Part 1, no.10, pp. 5437-5443, 1998.

2. Major International Conference Papers

1. **H. Goto**, T. Takada, R. Abe, Y. Kawabe, K. Oami and M. Tanaka, "An Isolation Technology for High Performance Bipolar Memories -- IOP-II," IEEE IEDM Tech. Dig., pp.58-61, 1982.
2. K. Toyoda, M. Tanaka, H. Isogai, C. Ohno, Y. Kawabe and **H. Goto,** "A 15ns 16Kb ECL RAM with a PNP Load Cell," IEEE ISSCC Tech. Dig., pp.108-109, 1983.
3. **H. Goto**, T. Takada, K. Nawata and Y. Kanai, "A New Isolation Technology for Bipolar VLSI Logic Circuits (IOP-L)," Dig. of Tech. Papers, Symposium on VLSI Technology, pp.42-43, 1985.
4. **H. Goto** and K. Inayoshi, "Trench Isolation Schemes for Bipolar Devices - Benefits and Limiting Aspects," Proc. of 17th European Solid-State Device Research Conference, pp.689-692, 1987.
5. K. Ueno, **H. Goto**, E. Sugiyama and H. Tsunoi, "A Sub-40ps ECL Circuit at a Switching Current of 1.28mA," IEEE IEDM Tech. Dig., pp.371-374, 1987.
6. T. Yamazaki, I. Namura, **H. Goto**, A. Tahara and T. Ito, "A 11.7 GHz 1/8-Divider

using 43 GHz Si High Speed Bipolar Transistor with Photoepitaxy Grown Ultra-Thin Base," IEEE IEDM Tech. Dig., pp.309-312, 1990.

7. T. Yamazaki, I. Namura, T. Sugii, **H. Goto**, A. Tahara and T. Ito, "High-Speed Si Hetero-bipolar Transistor with a SiC Wide Gap Emitter and an Ultra-Thin Heavily Doped Photoepitaxy Grown," IEEE BCTM Tech. Dig., pp.71-74, 1991.
8. T. Izawa, M. Katsube, Y. Yokoyama, K. Hashimoto, E. Kawamura, A. Shimizu, H. Takagi, F. Inoue, H. Shimizu, K. Furumochi, **H. Goto**, S. Kawamura, K. Watanabe and K. Aoyama, "A Novel Embedded SRAM Technology with 10 μm^2 full-CMOS Cells for 0.25 μ m Logic Devices," IEEE IEDM Tech. Dig., pp.941-943, 1994.
9. C. Jungemann, S. Decker, R. Thoma, W. L. Engl and **H. Goto,** "Phase Space Multiple Refresh: a Versatile Statistical Enhancement Method for Monte Carlo Simulation," Proc. of Int. Conf. on Simulation of Semiconductor Processes and Devices, pp.65-66, 1996.
10. C. Jungemann, S. Yamaguchi and **H. Goto**, "Accurate Prediction of Hot-carrier Effects for a Deep Sub-μ m CMOS Technology Based on Inverse Modeling and Full Band Monte Carlo Device Simulation," Proc. of Int. Conf. on Simulation of Semiconductor Processes and Devices, 59-60, 1996.
11. C. Jungemann, S. Yamaguchi and **H. Goto**, "Efficient Full Band Monte Carlo Hot Carrier Simulation for Silicon Devices," Proc. of 26th European Solid-State Device Research Conference, pp.821-824, 1996.
12. C. Jungemann, S. Yamaguchi and **H. Goto**, "Is There Experimental Evidence for a Difference between Surface and Bulk Impact Ionization in Silicon," IEEE IEDM Tech. Dig., pp.383-386, 1996.
13. Y. Takao, Y. Sanbonsugi, K. Watanabe, H. Takatsuka, T. Karasawa, E. Kawamura, K. Hashimoto, H. Takagi, F. Inoue, H. Shimizu, T. Yamazaki, **H. Goto**, T. Sugii, M. Miyajima, K. Watanabe and K. Aoyama, "A 4 μm^2 Full-CMOS SRAM Cell Technology for 0.2-μ m High-Performance Logic LSIs," Dig. of Tech. Papers, Symposium on VLSI Technology, pp.11-12, 1997.
14. B. Meinerzhagen, C. Jungemann, S. Decker, S. Keith, S. Yamaguchi and **H. Goto**, "Predictive and Efficient Modeling of Substrate Currents in n-channel MOS-Transistors," Proc. of International Symposium on VLSI Technology, Systems and Applications, pp.232-235, 1997.
15. C. Jungemann, S. Yamaguchi and **H. Goto**, "Convergence Estimation for Stationary Ensemble Monte Carlo Simulations," Proc. of Int. Conf. on Simulation of Semiconductor Processes and Devices, pp.209-212, 1997.

16. C. Jungemann, S. Yamaguchi and **H. Goto**, "Investigation of the Influence of Impact Ionization Feedback on the Spatial Distribution of Hot Carriers in an NMOSFET," Proc. of 27th European Solid-State Device Research Conference, pp.336-339, 1997.
17. S. Yamaguchi and **H. Goto**, "Inverse Modeling - A Promising Approach to Know What is Made and What should be Made," Proc. of Asia and South Pacific Design Automation Conference, pp.117-121, 1998.
18. T. Tanaka, Se. Yamaguchi, Sa. Yamaguchi, K. Sukegawa and **H. Goto**, "Investigation to Suppress Hot Carrier Effect in Pocket{Implanted nMOSFET by Full Band Monte Carlo Simulation," Ext. Abst. of 6th Int. Workshop on Computational Electronics, pp.214-217, 1998.

3. Book Chapter

1. **H. Goto**, "High-Speed Silicon Bipolar ICs," in "High-Speed Digital IC Technologies" edited by M. Rocchi, Artech House Inc., Ch.3, 1990.

4. Related Papers

1. K. Toyoda, M. Tanaka, H. Isogai, C. Ohno, Y. Kawabe and **H. Goto**, "A High Speed 16Kbit ECL RAM," IEEE Journal of Solid-State Circuits vol.SC-18, no.5, pp.509-514, 1983.
2. Y. Nagase, K. Hashimoto, **H. Goto**, T. Nakadai, T. Nawata and T. Takada, "Analysis of Bipolar Transistor's f_T-Ic Characteristics by Two Dimensional DC and AC Device Simulation," Proc. of NUPAD-II, Paper no.NUPAD-156, 1988.
3. Y. Nagase, **H. Goto** and T. Takada, "Analysis of Pedestal-Collector Transistors by 2-D Process and Device Simulation," Proc. of 7th VLSI Process/Device Modeling Workshop, pp.40-41, 1990.
4. T. Sugii, T. Yamazaki, Y. Arimoto, T. Ito, I. Namura, **H. Goto** and A. Tahara, "SiCx:F Hetero-Emitter and Epitaxial-Base Bipolar Transistors," Proc. of 4th Int. Conf., Amorphous and Crystalline Silicon Carbide IV, pp.314-321, 1991.
5. T. Sugii, T. Yamazaki, Y. Arimoto, T. Ito, Y. Furumura, I. Namura, **H. Goto** and A. Tahara, "SiC Growth and Its Application to High-Speed Si-HBTs," Proc. of 22nd European Solid-State Device Research Conference, pp.335-342, 1992.
6. H. Takatsuka, H. Sato, T. Izawa, T. Hisaeda, **H. Goto** and S. Kawamura, "Influence of Tilted High-Energy Ion-Implantation upon Scaled CMOS Structure," IEEE ICMTS

Tech. Dig., pp.247-251, 1995.

7. C. Jungemann, S. Yamaguchi and **H. Goto**, "On the Accuracy and Efficiency of Substrate Current Calculations for Sub-μm n-MOSFET's," IEEE Electron Device Lett., vol.17, no.10, pp.464-466, 1996.
8. C. Jungemann, S. Decker, R. Thoma, W.-L. Engl and **H. Goto**, "Phase Space Multiple Refresh: A General Purpose Statistical Enhancement Technique for Monte Carlo Device Simulation," IEEE J. Techno. Comp. Aided Design, (2), 1997. (http://www.ieee.org/journal/tcad/accepted/jungemann-jan97/)
9. N. Sugiyasu, K. Suzuki, S. Kojima, Y. Ohyama and **H. Goto**, "A New Practical Approach to Implement a Transient Enhanced Diffusion Model into an FEM-based 2-D Process Simulator," Materials Research Society Fall Meeting, Symposium Paper no.Q-1.6, pp.370-371, 1997, also in "Semiconductor Process and Device Performance Modeling," MRS Symposium Proc., vol.490, 1998.
10. C. Jungemann, S. Yamaguchi and **H. Goto**, "Convergence Estimation for Stationary Ensemble Monte Carlo Simulations," IEEE J. Techno. Comp. Aided Design, (10), 1998. (http://www.ieee.org/journal/tcad/accepted/jungemann-jan98/)