

1

The Miraculous Evolution of the Field-Effect Transistor (FET): From Inception to Future Prospects

Hiroshi Iwai^{1,2}

¹International College of Science and Technology, National Yang Ming Chao Tung University, Hsinchu, Taiwan

²Institute of Science Tokyo, Tokyo, Japan

CHAPTER MENU

- 1.1 Introduction, 2
 - 1.1.1 Roles and Functions of the FET, 2
 - 1.1.2 Structural Advantages and Historical Evolution of the FET, 3
 - 1.1.3 The Miraculous Evolution of the MOSFET, 3
 - 1.1.4 Contents of this Chapter, 7
- 1.2 1925–1960: Early Concepts and Challenges in MOSFET Development, 8
 - 1.2.1 MOSFET Development in the Period from 1925 to 1945, 9
 - 1.2.2 MOSFET Development in the Period from 1945 to 1960, 11
 - 1.2.3 Evolution of Unit Process Technologies Developed for BJT Fabrication in the 1950s, 18
- 1.3 1960–1970: The MOSFET Instability Problem, 22
 - 1.3.1 Instability Arising from Unpredictable Charges, 23
 - 1.3.2 Stabilization of Interface Charges by Low-Temperature Hydrogen Annealing and Surface Orientation Control, 24
 - 1.3.3 Countermeasures Against Mobile-Ion-Induced Instability, 24
- 1.4 From MOS ICs to MOS LSIs: 1965–1969, 25
 - 1.4.1 Commercialization of MOS ICs in 1965: Shift Registers, Standard Logic, and the Integration of Several to over 100 MOSFETs, 26
 - 1.4.2 Development of Custom Logic ICs for the Victor 3900 by GMe in 1965, 26
 - 1.4.3 Evolution of Sharp's Electronic Calculators: From Discrete BJTs in 1964, Through MOS ICs by Hitachi and NEC in 1967, to the First MOS LSI by North American Rockwell in 1969, 27
 - 1.4.4 Development of a 64-Bit PMOS Static RAM by Fairchild Semiconductor Co. in 1964 and Memory Developments by Other Companies, 27
 - 1.4.5 Summary of MOS IC Development in the 1960s, 28
- 1.5 Technologies for MOS Integrated Circuits Developed Between 1965 and 1970, 29
 - 1.5.1 Silicon-Gate Technology, 29
 - 1.5.2 Direct-Contact Technology, 30
 - 1.5.3 Metal Gate Technologies and S/D Extension, 30
 - 1.5.4 Poly-Si Gate Doping and Gate Electrode Work Function, 32
 - 1.5.5 PSG Reflow, 33
- 1.6 First-Generation LSIs—Al- or Si-Gate PMOS LSIs (1969–1971) with 10–8 micrometer Design Rules, 33
 - 1.6.1 Al-Metal Gate PMOS Custom Logic LSI for Calculators, 33
 - 1.6.2 Intel's Silicon Gate PMOS LSI, 33
 - 1.6.3 Nonvolatile Memory LSI, 34
 - 1.6.4 Supply Voltage of PMOS LSIs, 36
- 1.7 Second-Generation LSI: From the Dawn of NMOS LSI to the Mid-1970s, 36
 - 1.7.1 NMOS LSI Products as Second-Generation Devices, 37

- 1.7.2 A Case Study of NMOS LSI Technology Developed in the 1970s, 37
- 1.8 First Generation of VLSI (3 μm NMOS Technology: Fourth Generation of LSI) Late 1970s to Early 1980s: Emergence of Dry Processing and Stepper Lithography, 44
 - 1.8.1 Introduction of Dry Processes and the Emergence of Plasma-Induced Damage (PID), 45
 - 1.8.2 Introduction of Intrinsic Gettering (IG) Wafer, 47
 - 1.8.3 Stepper (Step and Repeat with 5 $\times$ and 10 $\times$ Optical Reduction of Mask Pattern Feature Size) Photolithography and “Common Defects”, 48
 - 1.8.4 Dust Control, 49
- 1.9 Transition from NMOS to CMOS in the Mid-1980s, 49
- 1.10 Advances in Scaling Technologies—Introduction of Novel Process Techniques and Materials (1980s to Early 1990s), 52
 - 1.10.1 Evolution of Source/Drain Extension Structures: LDD and Halo/Pocket Implant Architectures, 52
 - 1.10.2 Rapid Thermal Processing (RTP: RTA, RTO, and RTN), 53
 - 1.10.3 Introduction of Polycide Gate Electrodes: Applications in Memory and Logic Devices, 54
 - 1.10.4 Introduction of Self-Aligned Silicide (Salicide) Technology, 55
 - 1.10.5 Suppression of Boron Penetration (1986–1990) by Hydrogen-Free Anneal or Lightly Nitrided Oxynitride (SiO_xN_y) Gate Insulators, 57
 - 1.10.6 Introduction of Shallow Trench Isolation (STI) and Copper Interconnect Technologies, 58
- 1.11 Challenges in Scaling into the Sub-50 nm Regime from the Mid-1990s to the 2000s, 60
 - 1.11.1 40 nm Gate-Length n-MOSFETs with 3 nm Gate Oxide Thickness and 10 nm Junction Depth, 60
 - 1.11.2 1.5 nm Gate-Oxide n-MOSFETs with 90 nm Gate Length and 30 nm Junction Depth, 65
 - 1.11.3 Source/Drain Technology Evolution from the Mid-1990s, 67
- 1.12 Development of RF CMOS Device Technology from the Mid-1990s to the Late 1990s, 68
- 1.13 Post-2000: Confronting the Limits of Miniaturization, 72
 - 1.13.1 Three Scaling Limits in Device Miniaturization, 72
 - 1.13.2 Limitations of Miniaturization Due to Increasing Power Consumption, 72
 - 1.13.3 Scaling Limits Due to Off-State Leakage Current, 74
 - 1.13.4 Issues: Lithography, Cost, Heat, and Yield, 81
- 1.14 Future Prospects, 82
 - 1.14.1 Prospects for the Near Future, 82
 - 1.14.2 Looking More than 30 Years Ahead, 83
- 1.15 Summary and Concluding Remarks, 84
 - 1.15.1 Summary, 84
 - 1.15.2 Concluding Remarks, 85
- Acknowledgment, 85
- References, 86

1.1 Introduction

1.1.1 Roles and Functions of the FET

The year 2025 marks the centennial of the invention of the field-effect transistor (FET) by Julius Edgar Lilienfeld [1]. As the fundamental device of complementary metal-oxide semiconductor large-scale integrated (CMOS LSI) circuits, the FET supports a wide range of modern technologies, including artificial intelligence (AI) systems, data servers, and smartphones. Its demand has surged alongside the rapid proliferation of AI. FETs are also indispensable in applications such as image sensors and power devices. Without them, the technological landscape of contemporary societies would be far less advanced.

The FET functions primarily as an amplifier (e.g., for communication systems) and as a switch (e.g., for logical operations in computers). These roles are typically realized by a three-terminal device, in which the current between two terminals is controlled by a third, as illustrated in

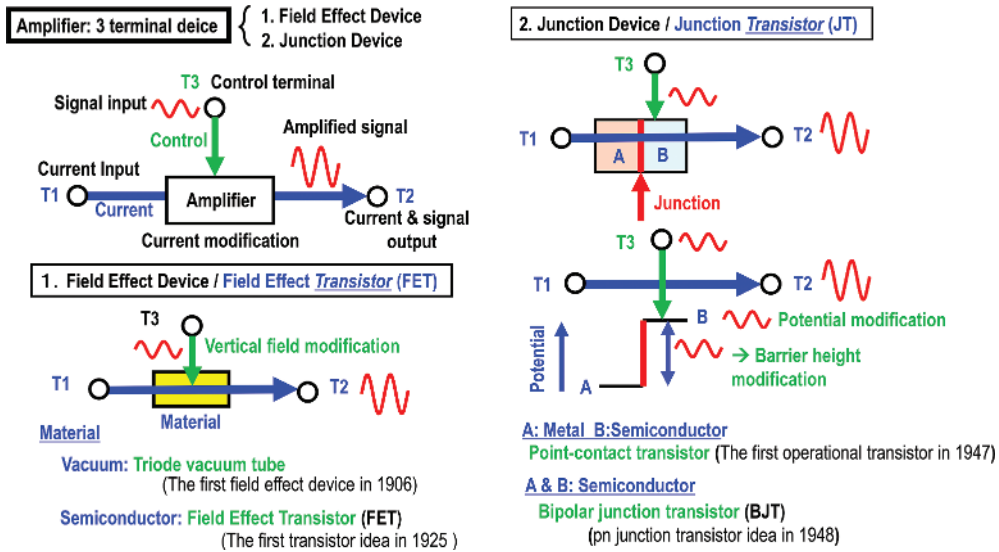


Figure 1.1 Concepts of amplifiers: 1. field-effect devices and 2. junction devices.

Figure 1.1. Such devices are broadly classified into (i) FETs and (ii) junction transistors (JTs) [2]. While the FET was deliberately conceived, the JT was discovered by chance in 1947. Notably, the term “FET” was not introduced at the time of invention, but 27 years later, in a 1952 publication by William Shockley of Bell Labs [2].

1.1.2 Structural Advantages and Historical Evolution of the FET

Because the FET is a flat, planar device, it is well suited to photolithography, enabling mass production and continued miniaturization through scaling [3]. Scaling has delivered major improvements: higher drive current, faster switching, better high-frequency performance, and lower power consumption through reduced operating voltages.

As shown in Table 1.1, a rough estimate comparing early generations of central processing units (CPUs) with contemporary graphics processing units (GPUs) underscores the remarkable progress of large-scale integration (LSI) technology. Over the past 55 years, linewidths have shrunk by roughly three orders of magnitude (about 1/1000). During the same period, transistor counts have increased by nearly 50 million times, while overall computational performance is estimated to have risen by at least 50 million times. In stark contrast, power consumption has grown by only about a factor of 1000, and manufacturing cost by merely a factor of five—figures disproportionately small relative to the dramatic gains in integration and computational capability. This disparity highlights the exceptional efficiency improvements achieved through device scaling and miniaturization.

1.1.3 The Miraculous Evolution of the MOSFET

As shown in Figure 1.2a, in the structure of the first FET patent (1925), the gate electrode and the channel were insulated by the reverse bias of a Schottky junction. Three years later, in 1928, Lilienfeld inserted a gate insulating film between the gate electrode and the channel, thereby

Table 1.1 Impact of miniaturization on logic MOS LSI devices.

1970: 10 μm MPU $\rightarrow$ 2025: 2 nm GPU (Real line size $\sim$ 10 nm)

Line width: x 1/1000, FET foot print: x 1/1M
Chip size: x 50
of MOSFETs on chip: x 50M (100 Billions)
Performance: > x 50M (2000 TFLOPS)
Power Consumption x 1000
Price (Inflation adjusted): x 5

In the data center, 1000 ~ 100,000 GPU are typically used.

↓
Power consumption > 500 kW ~ 50 MW

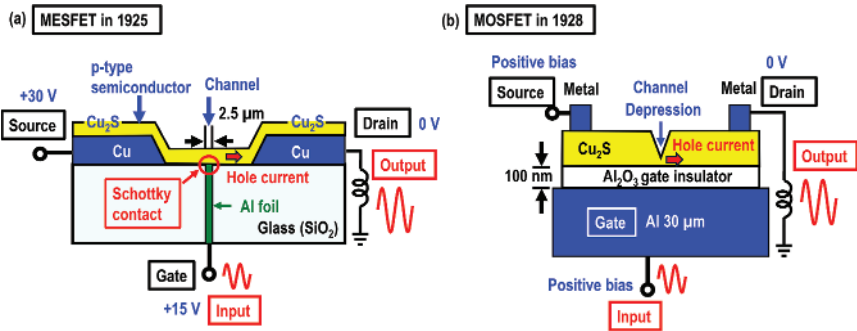


Figure 1.2 Schematic cross-sections of (a) MESFET (1925) [1] and (b) MOSFET (1928) invented by J. Lilienfeld.

inventing the metal–oxide–semiconductor field-effect transistor (MOSFET) (Figure 1.2b) [4], in which the gate and channel could be insulated regardless of the polarity of the gate bias. Since then, research and development of MOSFETs became the mainstream, and they subsequently underwent major advances, such as high-performance, highly integrated circuits (ICs). The evolution of the MOSFET, however, tells the story of repeatedly overcoming challenges once thought impossible, achieving progress that can only be described as miraculous. A summary is presented in Table 1.2.

1.1.3.1 From Conception (1925) to Realization (1960)

The early patents for the FET (1925) and the MOSFET (1928), both filed by Lilienfeld, represented straightforward applications of triode vacuum tube principles [5] to solid-state materials. However, in the absence of high-purity semiconductors and suitable processing technologies, practical realization was unattainable.

A breakthrough occurred around 1940, when silicon (Si) was identified as the optimal material for radar and microwave receivers, spurring rapid purification efforts. The accidental discovery of the Schottky JT (also known as the point-contact transistor) in 1947, followed by the invention

Table 1.2 Historical overview of FET technology development.

1905	Triode vacuum tube as Predecessor: 1906	
20 years	Vacuum tube development for wireless communication	
1925	FET idea conceived as a solid-state RF amplifier: 1925	
15 years	Semiconductor physics development	Quantum mechanics, band theory, hole, inversion-layer, pn junction
1940	Military use of point-contact Si-rectifier for microwave applications: late 1930s ~ mid 1940	Purity of Si improved
	Discovery of point-contact junction transistor operation 1947	{ Single crystal pulling, impurity doping, epitaxial growth, Photolithography, SiO ₂ diffusion mask, ion-implantation, Planar technology, Integrated circuits invention
20 years	Commercialization of BJT (early 1950~)	Fabrication technologies
	Semiconductor process technology development	
1960	Realization of single MOSFET operation: 1960	SSI, MSI (Poor yield); Low Power: Military & Aerospace applications (Missile guidance)
10 years	Si/SiO ₂ interface stabilization	Surface physics, interface states, HCl cleaning, P-gettering, mobile ions. heavy metal contamination
	IC technology development	Shift resistor, Multiplexer, DA/AD
1970	Commercialization of MOS LSI: from 10 μm (1969): Low performance but high integration	Calculator, PC, gaming
15 years	Yield and reliability improvement	Memory MPU
	LOCOS isolation, Al-Si, BPSG reflow, contact barrier metal, electromigration, contact → projection lithography → stepper, PID, RIE, polycide,	
1985	CMOS VLSI: from 1 μm (1985): high performance & high integration starting with 100 MHz CPU and 1M-bit DRAM	T-CAD, E-CAD,
15 years	Continuous scaling down DSP, GA, FPGA, EWS, Engine control, Video game, image processing (MPEG), Cell phone	W-via
	LDD, halo, TiS ₂ → CoSi ₂ → NiSi salicide, RIE, RTP, SiON gate oxide, tunneling gate oxide, KrF excimer lithography, damascene, STI, Cu, CMP	
2000	Beginning of Nano-CMOS VLSI: from 130 nm: Higher performance (than BJT) & high integration:	Multi-core MPU, GPU, NPU
25 years	Pursuing the scaling limit Strain (high μ), high-k, Fin FET, ArF + Immersion, double patterning, EUV	Digital camera, Smart phone, Server
2025	End of miniaturization: from 2 nm node?	Nano-sheet GAA, CFET, Backside Power Delivery, HBM, interposer, chiplet, CIM
	Vertical integration and chiplet & package level integration	Heat removal layer AI & hpc Data center, Edge AI, on-device AI

of the bipolar junction transistor (BJT), further accelerated advances in Si crystal growth, photolithography, and doping during the 1950s—developments that indirectly paved the way for the MOSFET.

Building on these technological foundations, advances in semiconductor physics grounded in quantum mechanics revealed that an ultrathin conductive inversion layer could form at a semiconductor surface, with its current directly controlled by the gate electric field—an essential step toward realizing the MOSFET. It was also found that a *pn* junction electrically isolates the source and drain from the substrate while connecting them to the channel, and that thermally grown SiO₂ serves as a stable gate insulator with sufficiently low interface charge. These findings culminated in the first successful operation of the MOSFET in 1960 [6].

In short, semiconductors proved unexpectedly well suited for FETs, and silicon emerged as a truly miraculous material.

1.1.3.2 From the First MOSFET (1960) to the Era of PMOS LSI (1970)

The operation of MOSFETs was initially unstable, and in the early 1960s, they were not adopted in products due to reliability concerns. Through the sustained efforts of many researchers, however, it gradually became clear that the instability originated from various types of charges at the Si/SiO₂ interface. After nearly a decade, these instabilities were resolved, and MOSFETs became reliable enough for use in LSI products.

Because MOSFETs consumed less power and allowed higher integration than BJTs, small-scale integration (SSI) and medium-scale integration (MSI) devices began to be employed in military and aerospace applications—such as missile guidance circuits—from the mid-1960s, although the market remained limited. Custom logic MSIs and MOSFET memories were also developed, but insufficient yield in the mid-1960s hindered progress, and full-scale development did not begin until the latter half of the decade.

In Japan, the electronic calculator market expanded rapidly, creating strong demand for custom logic MSI/LSI. From the late 1960s, this demand drove IC technology toward LSI, often through design and manufacturing contracts with Japanese and US semiconductor manufacturers. To further enhance integration, silicon-gate technology—replacing the conventional aluminum gate with a polycrystalline silicon gate—was invented in the latter half of the 1960s.

1.1.3.3 From the First PMOS LSI (1960) to the Era of NMOS VLSI (1985)

The full-scale commercialization of LSI began in 1969, when several companies simultaneously initiated mass production using P-channel MOS (PMOS) technology with a 10-μm gate length. From that year onward, scaling began: memory capacity quadrupled, and device dimensions shrank by a factor of approximately 0.7 every two to three years. The so-called ideal scaling law [3] was published in 1974. With the introduction of channel ion implantation for threshold control, PMOS was replaced from the second generation onward by N-channel MOS (NMOS) LSI, which offered higher carrier mobility. The NMOS era continued until around 1985, after which CMOS LSI became dominant.

From the mid-1960s to the mid-1980s, the principal challenge in semiconductor manufacturing was low yield, largely due to failures caused by unforeseen factors and chemical vapor deposition (CVD)-related particulate contamination. These problems were largely resolved during this period. For the development of 3 μm NMOS very large-scale integration (VLSI) in 1980, stepper photolithography and so-called “dry processes” (including reactive ion etching (RIE) and ion implantation for source/drain formation) were adopted.

Although MOSFETs were initially inferior in performance to BJTs and III–V compound semiconductor devices, their suitability for high-density integration led to widespread adoption in

calculators, personal computers, and consumer electronics, where the highest performance was not a critical requirement.

1.1.3.4 From the Emergence of CMOS VLSI (1985) to the Nano-CMOS VLSI (2000)

In the mid-1980s, rising power consumption in NMOS LSI prompted a shift to low-power CMOS LSI. Gate lengths were reduced to about 1 μm , enabling the development of 1-Mbit dynamic random access memory (DRAM) and 100 MHz microprocessors. With continued scaling, new materials such as silicide, oxynitride, and copper (Cu), together with advanced processes such as rapid thermal processing (RTP) and RIE, were introduced to enhance device performance. Meanwhile, the practical adoption of technology computer-aided design (TCAD) and electronic computer-aided design (ECAD) significantly supported product development. Advances in CMOS integration also enabled engineering workstations (EWS), allowing these design tools to be used independently without reliance on mainframes. At the time, the scaling limit was widely believed to be around 100 nm, but by the mid-1990s, it had become clear that high-performance operation of devices with gate lengths below 25 nm was attainable.

1.1.3.5 From the Planar MOSFETs (2000) to the FinFET (2012) and GAA (2025)

In 130 nm CMOS technology, gate lengths had fallen below 100 nm, and FETs had surpassed BJTs and III-V transistors in performance. CMOS also began to be used in radio frequency (RF) front-end circuits for mobile communication. To sustain scaling, high- k /metal gate stacks were adopted at the 65 nm node, and FinFETs with three-dimensional (3D) structures were introduced at the 22 nm node in 2012, followed by gate-all-around (GAA) transistors in 2025. Immersion lithography innovation was introduced at the 45 nm node, and extreme ultraviolet (EUV) lithography was introduced starting from the 7 and 5 nm generations.

Scaling has continued for more than 50 years, shrinking six-transistor static random access memory (SRAM) cells to dimensions smaller than an influenza virus and increasing the number of MOSFETs on a single chip to a level comparable to the number of stars in the Milky Way. As a result, technologies once confined to science fiction, such as smartphones and AI, have become reality.

1.1.3.6 Present Status and Future Directions

Driven by the demand for ultra-high-performance circuits in AI, the semiconductor industry has invested heavily in advanced manufacturing. A single EUV lithography system costs several hundred million US dollars, while a state-of-the-art fabrication line requires tens of billions. Such investments, once unimaginable, are now justified by AI's unprecedented performance requirements and vast market demand. These technological, industrial, and societal advances—unthinkable until recently—highlight the extraordinary trajectory of the FET and reaffirm its status as a truly miraculous device in the history of electronics. Today, scaling is approaching physical limits near 10 nm gate lengths, and 3D integration, including vertical channel stacking, is emerging. Future challenges include managing heat from rising power density, controlling costs associated with increasing process complexity, and maintaining yield.

1.1.4 Contents of this Chapter

When the concepts of the FET and MOSFET were first patented a century ago, their intended application was merely as solid-state amplifiers for wireless communication receivers. At that time, no one—not even Lilienfeld himself—could have foreseen that these devices would one day be integrated in numbers comparable to the stars in a galaxy, enabling memory cells smaller than

viruses, and ultimately evolving into technologies capable of realizing AI on par with the human brain. The same sense of unforeseeable progress applied 55 years ago, at the dawn of LSI, and again 25 years ago, with the advent of nano-CMOS technology.

Over the course of this evolution, numerous challenges arose in the development of the MOSFET. Yet, through a series of innovations, serendipitous discoveries, and solutions, the device advanced to its present state. This chapter traces that trajectory: how a single idea led to the first MOSFET, and how it further developed into the modern VLSI systems that now underpin practical AI applications.

Today, device miniaturization has already approached its physical limits, and vertical stacking of devices is likewise expected to reach its own limits in the not-so-distant future. Also essential, against this backdrop, is to consider how IC technology may continue to evolve and be applied, both in the near term and in the more distant future.

Readers seeking further information are directed to references [7–30], as well as other chapters in this book. Although these works were consulted in preparing this chapter, efforts were made, wherever possible, to trace their statements back to the underlying primary sources and to verify their accuracy. Owing to space limitations, the scope of detailed explanations and examples presented here is necessarily restricted. For the period prior to 2000, LSI device development in the United States, Japan, and Europe was predominant; with regard to developments in Europe, however, the description is left to other chapters. Being most familiar with the technological advances at Toshiba, the author draws many examples from Toshiba's experience. It should be noted that another reason for the frequent references to Toshiba is that, at that time, Toshiba—alongside IBM and several other companies—stood at the forefront of global technological development. As for more recent developments after 2000, the FET evolution will be discussed in detail in other chapters, and the descriptions in this chapter are therefore kept brief.

This chapter is organized as follows. Section 1.2 traces the technological evolution from the invention of the FET in 1925 to the realization of the first MOSFET in 1960. Sections 1.3–1.5 examine the efforts of various companies during the 1960s to resolve MOSFET instability and to enable their introduction as LSI products. Sections 1.6–1.8 review the historical development from the 10 μm PMOS LSI in 1969 to the 3 μm NMOS VLSI of the early 1980s, a period marked by the critical challenge of yield improvement and by process innovations such as channel ion implantation, dry processing, and the adoption of steppers. Sections 1.9 and 1.10 discuss the transition from NMOS to CMOS LSI beginning in the mid-1980s, together with the adoption of RTP and the introduction of new materials—including TiSi_2 and NiSi silicides and copper interconnects—for next-generation technologies in the 0.25–0.1 μm range. Section 1.11 highlights the initiatives of the mid-1990s that enabled the realization of sub-50 nm CMOS devices, while Section 1.12 addresses the emergence of RF CMOS technology during the same decade. Section 1.13 reviews the development of multiple-gate devices, conceived to overcome the scaling limits of planar transistors. Section 1.14 considers the near- and long-term future of VLSI technology, and Section 1.15 provides a summary and concluding remarks.

1.2 1925–1960: Early Concepts and Challenges in MOSFET Development

This section outlines the technological developments from 1925, when the concept of the FET was first proposed, to 1960, when the MOSFET was successfully realized in a functional form.

1.2.1 MOSFET Development in the Period from 1925 to 1945

1.2.1.1 Early Patent Proposals and Thin-Film Limitations in MOSFET Concepts

As shown in Figure 1.2, the basic concept of the FET was first proposed by Lilienfeld in 1925 [1] and 1928 [4]. His vision was to realize a solid-state amplifier as an alternative to the vacuum triode [5]. Although not explicitly stated in his patents, it is reasonable to infer that his motivation was to overcome the inherent miniaturization limits of vacuum tubes and to achieve lower operating voltages, reduced power consumption, and improved high-frequency performance through the use of solid-state devices. The device proposed in 1925 lacked a gate insulator, with the gate electrode connected to the channel via a Schottky junction; in modern terms, it corresponds to a metal–semiconductor field-effect transistor (MESFET) [31]. In contrast, the 1928 device incorporated a gate insulator or oxide, forming a MOSFET structure that provided DC insulation between the gate electrode and the channel for both positive and negative voltages. Lilienfeld selected copper(I) sulfide (Cu_2S) as the semiconductor material; however, as a p -type semiconductor with extremely low hole mobility, this choice is now recognized as unsuitable. A reproduction experiment conducted by Thomas A. Edison Industries in 1964 reported that the device did not operate [32], and there is no record that Lilienfeld himself ever succeeded in making it function. Lilienfeld’s MOSFET structure was a simple design reminiscent of a triode vacuum tube.

In 1935, Oscar Heil proposed a similar MOSFET structure [33]. However, as shown in Figure 1.3, the MOSFET was not successfully demonstrated with satisfactory performance until 1960. In semiconductors, unlike in a vacuum, the electric field induced by the gate voltage decays rapidly in depth due to free carrier screening, necessitating an ultrathin semiconductor layer for the gate field to be effective, even though the carrier density in a semiconductor channel is lower than that in a metal. The MOSFET proposed by Lilienfeld operates by modulating the lateral current flowing through this thin layer via the gate voltage. Such a semiconductor thin film requires a supporting substrate, typically an insulating substrate. This configuration corresponds to a thin-film transistor (TFT) [34, 35]. However, even today, growing high-quality semiconductor thin films on insulating substrates remains technically challenging. At that time, it was impossible to form single-crystal thin films, and the films were likely very rough and contained numerous defects and impurities. This problem was solved 35 years later, as shown on the right column of Figure 1.3, by employing the

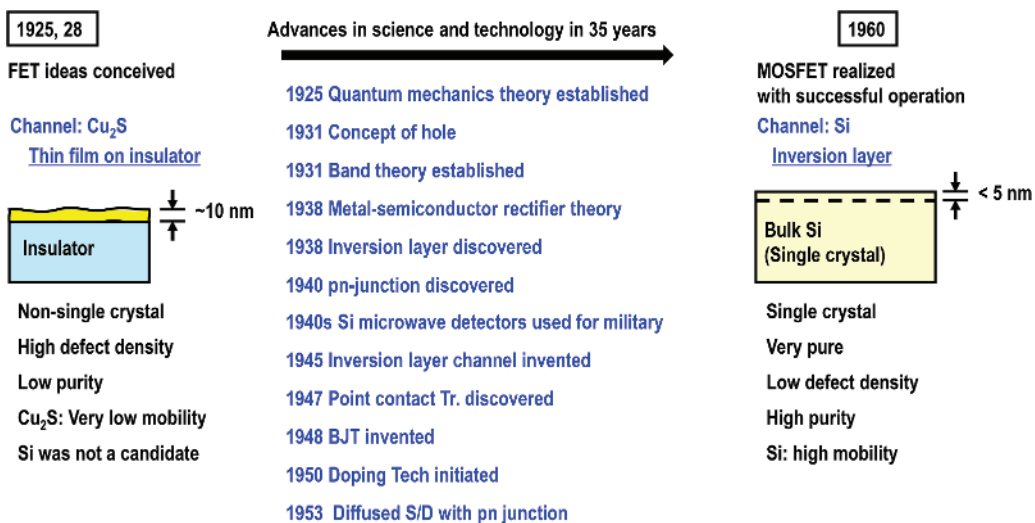


Figure 1.3 Evolution from early FET concepts (1925, 1928) to MOSFET realization in 1960.

inversion layer that appears at the surface of a high-purity Si single-crystal bulk substrate as the channel. The description in the center column illustrates the scientific and technological advances that led to this breakthrough.

1.2.1.2 Foundational Advances in Semiconductor Physics Enabling MOSFET Realization

The MOSFET realized in 1960 overcame these thin-film limitations by utilizing an extremely thin inversion layer formed at the surface of a bulk single-crystal silicon substrate as the channel. Notably, in 1925, neither the concept of the inversion layer nor even that of the hole had yet been established. Between 1925 and 1960, many physical phenomena in semiconductors essential to the realization of the FET were discovered, and related technologies were developed. Although these advances were not necessarily originally intended for FET development, they ultimately provided the foundation for its realization. Representative milestones include the establishment of quantum mechanics [36, 37], the introduction of the hole concept [38, 39], the formulation of energy band theory [39, 40], and the discovery of the inversion layer [41, 42].

The concept of the hole was developed almost simultaneously by W. Heisenberg, who approached it from the perspective of the absence of electrons in the outer orbitals of acceptor atoms (Ref. [38]: manuscript received June 19, 1931), and by A.H. Wilson of Cambridge University, who examined it in terms of the absence of electrons in the valence band using a simple 3D periodic potential (Ref. [39]: manuscript received June 18, 1931).

With regard to band theory, R. Kronig and W. Penney of the University of Groningen solved Schrödinger's equation for a nearly free electron system in a simple one-dimensional periodic potential well, and their results were published in February 1931 [43]. Using the Bloch Theorem discovered two years earlier, they derived the relation between the electron energy $E(k)$ and the wave vector k . Their analysis demonstrated that electron energies are not continuous but are divided into allowed regions (bands) and forbidden regions (bandgaps).

In Ref. [39], Wilson calculated the relation between the crystal's electronic energy $E(k)$ and the wave vector k , using both the free-electron and tight-binding models with a 3D periodic potential. He elucidated the band structures of forbidden bands, valence bands, and conduction bands, thereby clarifying the conduction properties of metals, insulators, and semiconductors. Namely, in the case of metals, the conduction band is partially filled with electrons, or the valence band and conduction band overlap. As a result, electrons can move to unoccupied (higher-energy) states with only a small amount of energy (such as an applied electric field) and thus behave as a dense population of free electrons.

In semiconductors, the Fermi level lies within the forbidden band. At absolute zero in the dark, all states in the valence band are filled with electrons, so electrons cannot move within the valence band, and no electrons exist in the conduction band. Consequently, no current flows. However, at finite temperatures or under illumination, a small number of electrons are excited from the valence band to the conduction band by thermal or optical excitation, while holes are simultaneously created in the valence band. This allows electron motion in both the conduction band and the valence band, enabling current to flow.

In insulators, the energy gap is too large for electrons to be excited from the valence band to the conduction band, so essentially no current flows. Moreover, the characteristic property of semiconductors whereby conductance increases with temperature arises from the increase in the number of electrons in the conduction band due to thermal excitation. These phenomena cannot be explained by the Kronig–Penney model, but were first successfully explained by Wilson's model.

Wilson further showed the formation of donor levels within the forbidden band due to impurity atoms, as well as the release of electrons from donor levels into the conduction band [40], as

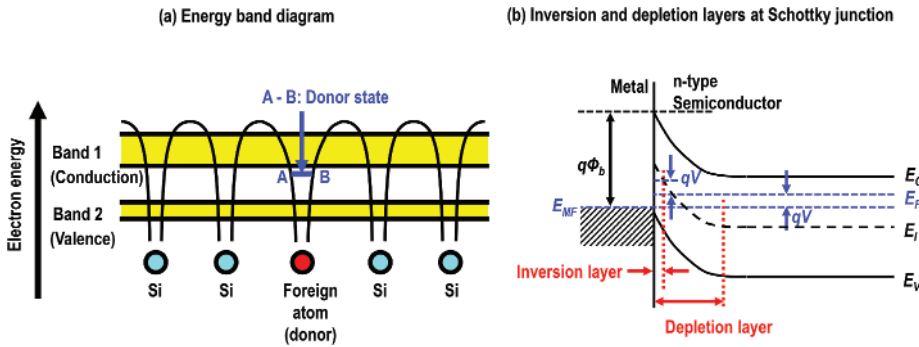


Figure 1.4 Energy band diagram showing quantum-mechanically calculated conduction and valence bands (a) [39, 40], as well as inversion and depletion layers [41, 42] formed in a Schottky junction (b). Source: (a) Adapted from [40].

shown in Figure 1.4a. Wilson’s band model made it possible to accurately describe the electrical phenomena of semiconductors (e.g., rectifying behavior at metal–semiconductor contacts [41]). Furthermore, between 1938 and 1939, band theory revealed the existence of inversion and depletion layers, with the inversion layer later forming the basis for the development of the MOSFET. This is illustrated in the Schottky junction band diagram in Figure 1.4b [41, 42].

1.2.1.3 Silicon Rectifiers, Purification Breakthroughs, and the Discovery of the *pn* Junction

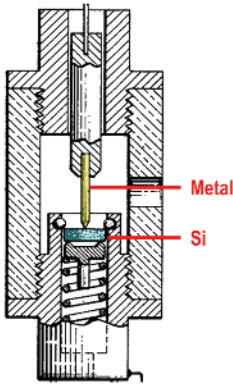
Prior to World War II, radar and microwave communication were researched for military purposes. However, vacuum tubes had inherent miniaturization limits and could not reduce electron transit distances, making them unsuitable for high-frequency applications such as microwaves. Consequently, point-contact rectifiers made of metals or semiconductors came into use as receivers. During material investigations, silicon was found to possess high carrier mobility, making it the most suitable semiconductor for high-performance rectifiers after more than one hundred semiconductors had been examined [44]. This led to the development of the tungsten–silicon (W–Si) point-contact rectifier (Figure 1.5a) [45], which was deployed for military applications. This discovery drew attention to Si as a semiconductor material suitable for device applications and accelerated improvements in its purification technology. In 1940, Russell Ohl of Bell Labs made a groundbreaking discovery during silicon purification. When molten Si was slowly cooled to precipitate impurities, he observed that *p*-type and *n*-type impurity distributions segregated within the ingot, forming a *pn* junction—a previously unknown junction with novel electrical properties (Figure 1.5b) [46–48]. This finding became a major milestone in semiconductor device development, as the *pn* junction is indispensable to both BJTs and MOSFETs, and in retrospect, it was a discovery worthy of a Nobel Prize—though Ohl was never actually awarded one.

1.2.2 MOSFET Development in the Period from 1945 to 1960

1.2.2.1 Welker’s 1945 Inversion Layer Concept and Its Technical Context

Under these circumstances, research and development of microwave receivers employing Si and Ge advanced in both the United States and Germany. A decisive step toward the realization of the MOSFET, however, was taken by Heinrich Welker in Germany. In 1945, he proposed using, not the semiconductor thin film itself as the channel, but rather the inversion layer formed at the

(a) Point-contact Si rectifier
(microwave receivers and oscillators)



(b) Discovery of pn junction during Si purification process

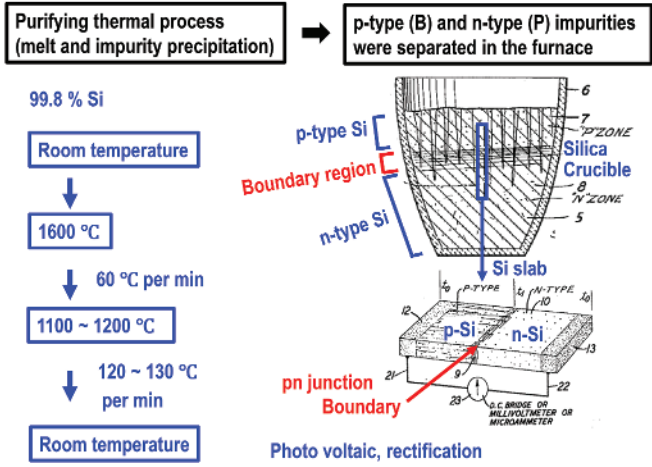


Figure 1.5 (a) Point-contact Si rectifier for microwave receivers and oscillators [45], together with (b) the historical discovery of the *pn* junction during Si purification [46–48].

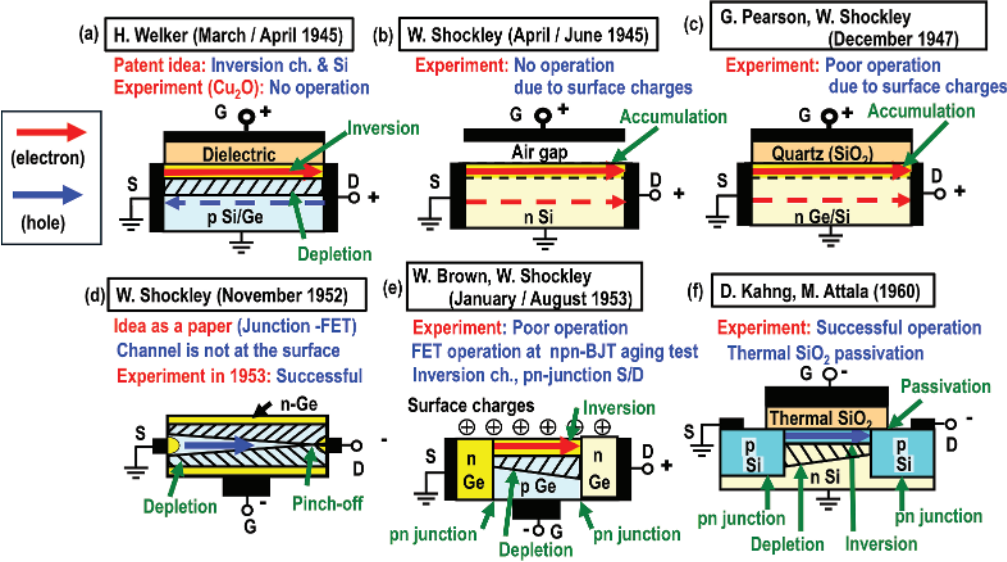


Figure 1.6 Six milestones tracing the development of the MOSFET from 1945 to its first successful operation in 1960 [2, 6, 22, 49–52].

surface of Si or Ge (Figure 1.6a) [49, 53]. Although the fabrication of single-crystal Ge and Si ingots was not achieved until 1948 and 1954, respectively, the purity of silicon granules available on the US market as early as 1940 had already reached 99.85% [47]. Ingots produced by melting these granules, though polycrystalline, are presumed to have been of even higher purity. For effective control of current in a semiconductor thin film by a vertical electric field, a film thickness of less than 10 nm is desirable. At the time, however, it was entirely impossible to produce such ultrathin films with both high purity and sufficiently low defect density. By contrast, the inversion layer is

less than a few nanometers thick and can be readily modulated by the gate electric field. Shortly after Welker filed his patent application [49], Germany surrendered in World War II, preventing him from conducting adequate experiments. Furthermore, several technical challenges still had to be overcome before the MOSFET could be realized, and its successful implementation would not occur for another 25 years.

1.2.2.2 Early Bell Labs MISFET Experiments

At Bell Labs in the United States, Mervin Joseph Kelly was appointed Director of Research in 1936 and encouraged William Shockley, Walter Brattain, and others to pursue research on solid-state amplifiers. From 1945, John Bardeen also joined the effort, accelerating progress toward the realization of the FET. In April 1945, Shockley proposed the design of a silicon metal-insulator-semiconductor field-effect transistor (MISFET) employing an air gap as the gate insulator (Figure 1.6b) [22]. This design intended to use the accumulation layer as the channel. However, because the accumulation layer is not insulated from the bulk, a leakage path existed in the bulk between the source and drain. In addition, the source and drain regions consisted solely of metal. Shockley conducted experiments on this structure until June 1946 and abandoned this approach [22] because no modulation of the channel current by the gate voltage was observed. In 1948, John Bardeen of Shockley's group analyzed the problem and explained that charges associated with the high density of surface states on the silicon surface shielded the gate electric field [54], thereby preventing modulation of the channel current.

Recognizing the importance of the gate insulator, Gerald Pearson and William Shockley at Bell Labs conducted experiments in December 1947 on silicon and germanium MOSFETs employing a quartz (crystalline SiO_2) gate insulator (Figure 1.6c) [50]. On the lower surface of the quartz substrate, they deposited thin films of silicon, germanium, or copper oxide (Cu_2O) as the channel, while on the upper surface, they formed a gold (Au) film as the gate electrode. As a result, modulation of the channel current by the gate electric field was observed. However, due to the presence of a high density of interface states, the modulation amplitude was limited to one-tenth of the value predicted by band theory [7].

1.2.2.3 The Rise of the Bipolar Junction Transistor and Its Impact on FET Research

During experiments aimed at developing FET devices, John Bardeen and Walter Brattain, quite unexpectedly, discovered an amplification effect in their device in December 1947 [7, 22, 55]. Subsequent analysis revealed that, unlike an FET, the phenomenon arose from modulation of the junction current through alteration of the energy barrier at a Schottky junction, thereby identifying a new type of transistor—namely, the so-called JT. As shown in the upper left of Figure 1.7, the collector (metal) and the base (Ge substrate) form a reverse-biased Schottky junction. When the emitter (metal) voltage is 0 V, only a small leakage current flows. As shown in the upper right, when the emitter–base junction is forward-biased, and holes are injected from the emitter, these holes flow toward the collector electrode. This raises the potential of the Ge region near the collector, thereby lowering the potential barrier at that junction and increasing the junction current from the collector to the base—this is the operating principle. This led to the development of the first practical transistor—the point-contact transistor (Figure 1.7a) [56]—which was mass-produced and marketed, although the mechanical instability of the point-contact junction was a drawback. In January 1948, Shockley extended the theoretical operating model of the Schottky JT to the *pn* junction and invented the *pn* JT (BJT) based on band theory (Figure 1.7b) [7, 22, 56–58]. At that time, impurity doping techniques were not yet available, and a fabrication process for this device had to be developed. Three years after Shockley's invention, the operation of the BJT was finally

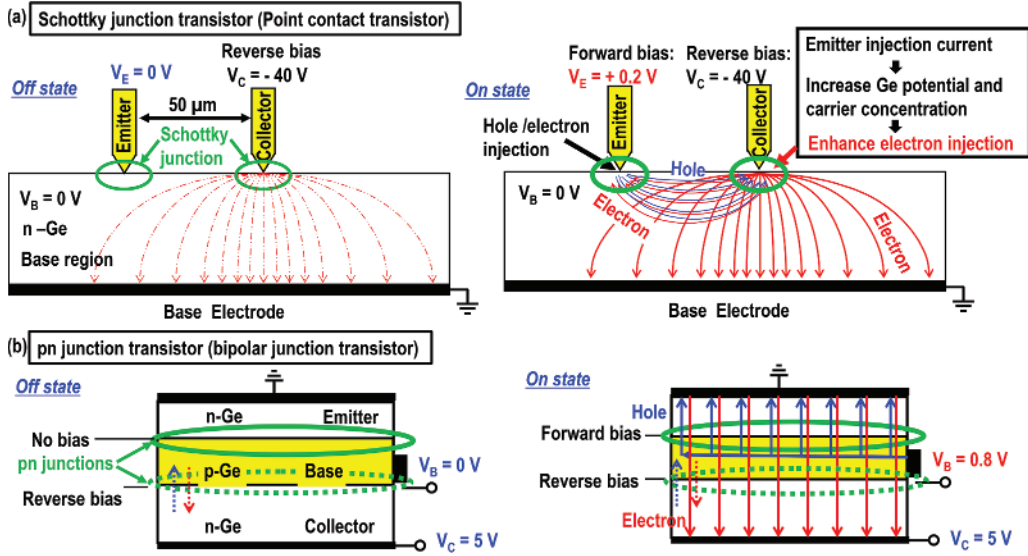


Figure 1.7 Operating principles of Schottky and pn junction transistors. Source: Adapted from [7].

demonstrated [59]. To enable the practical manufacturing of BJTs, key semiconductor device fabrication processes—such as impurity doping, photolithography, ion implantation, and epitaxial growth—were developed primarily in the 1950s. The pn junction of this device was mechanically stable, and it quickly replaced the point-contact transistor. From the mid-1950s onward, BJTs began to supplant vacuum tubes in amplifiers and logic circuit switches.

Following the remarkable success of the BJTs as a practical transistor, research on FETs declined sharply. As illustrated in Figure 1.8a, the BJT operates by means of a pn junction located deep within the semiconductor, rendering it relatively insensitive to surface states. In contrast, the FET

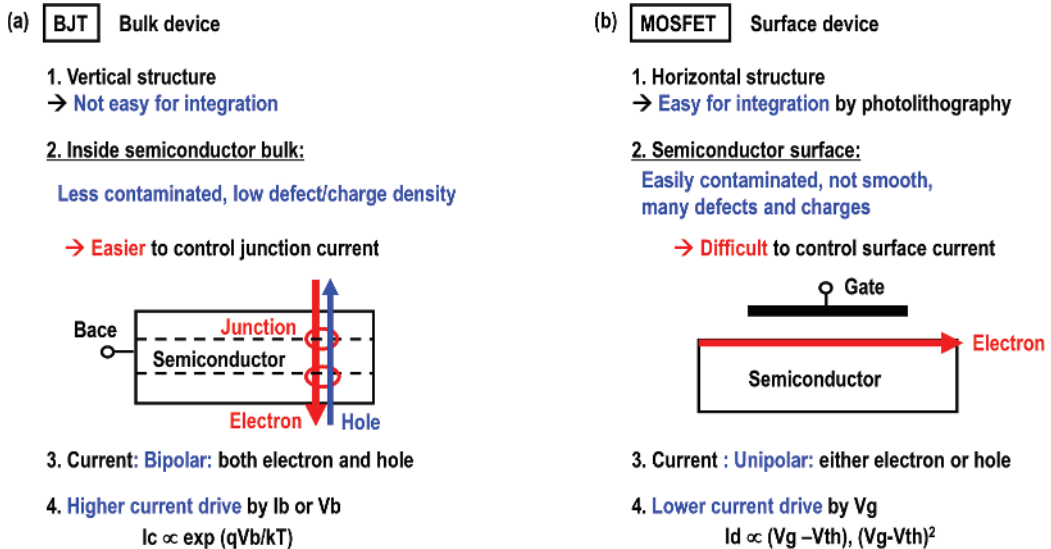


Figure 1.8 Overview of foundational differences between MOSFET and BJT relevant to device development.

is highly sensitive to the condition of the semiconductor surface; yet protecting that surface from contamination and maintaining it in a smooth state with low defect density proved extremely challenging (Figure 1.8b). Moreover, the lower current-driving capability of the MOSFET compared with that of the BJT further diminished the incentive for its development.

A practical MOSFET was developed in 1960, and once MOSFETs began to be employed in LSIs for calculators and memory at the end of the 1960s, they rapidly established themselves as the mainstream devices for LSI. This is because the MOSFET is a planar device in which all terminals and interconnections can be formed on the surface, making it highly suitable for LSI and readily scalable through photolithography. In contrast, the BJT is a vertical device that requires complicated processes to connect the collector, which is located deep within the silicon substrate, away from the surface, making it more difficult to scale down compared with the MOSFET. Furthermore, because it consumes more current, it is less suitable for LSI.

1.2.2.4 Junction FET Realization

To address these issues, in 1952, Shockley theoretically proposed the junction field-effect transistor (JFET), which he termed a unipolar FET, in which the current flowing within the semiconductor is controlled by reverse-biased pn junctions positioned above and below the channel (Figure 1.6d) [2]. This device did not employ an inversion layer; instead, it operated by expanding the depletion regions of the pn junctions to constrict the conductive channel located at the center of the semiconductor. In 1953, George Clement Dacey and Ian Munro Ross at Bell Labs successfully fabricated this device, making it the first FET to be commercially marketed [60].

1.2.2.5 The First Inversion Layer FET Demonstration

In 1953, Walter Lyons Brown and William Shockley at Bell Labs conducted an aging test on a germanium (Ge) BJT composed of laterally arranged n - p - n regions. During this test, a positive fixed charge was observed to form on the semiconductor surface, inducing an inversion layer in the base region. This inversion layer resulted in a surface leakage current between the emitter and base. When the base potential was varied via an electrode positioned beneath the base region, the characteristic behavior of a FET was observed. This experiment constituted the first confirmed demonstration of FET operation utilizing an inversion layer as the conductive channel (Figure 1.6e) [51, 52]. The device operated as a back-gated FET, in which the channel current was modulated by varying the substrate potential, and its structure was fundamentally different from that of MOSFETs or MESFETs, where the channel current was modulated by capacitive coupling of the gate electrode potential from the top side. The measured drain current was extremely small—approximately 1/300 of the theoretically predicted value. This discrepancy is attributed primarily to the Coulomb scattering arising from the fixed surface charge. A particularly noteworthy aspect of this experiment was that diffused regions with a dopant type different from that of the substrate served as the source and drain regions, corresponding to the emitter and collector regions of an n -type Ge BJT. This represented the first instance in which the source and drain regions were electrically isolated from the substrate by reverse-biased pn junctions. By contrast, in Welker's 1945 patent, the source and drain contacts were metallic (Figure 1.6a), allowing leakage current to flow through the substrate. The introduction of diffused source regions also conferred an important additional advantage: the pn junction between the source and the channel was forward-biased, thereby reducing contact resistance and enabling efficient injection of charge carriers into the channel.

1.2.2.6 Surface Passivation Breakthrough and the First MOSFET Operation

The final obstacle to achieving reliable MOSFET operation was the preparation of a semiconductor surface with favorable electrical characteristics—specifically, minimal fixed charge and surface-state density. This was ultimately overcome through research on surface passivation techniques originally developed to suppress surface leakage current in BJTs [61, 62]. Around 1958–1959, Martin M. Atalla and colleagues at Bell Labs discovered that thermally oxidizing the silicon surface to form an SiO_2 layer could produce an electrically stable interface on silicon. By combining the four key technological elements: (i) the use of a single-crystal silicon bulk substrate, (ii) utilization of the inversion layer formed at the silicon substrate surface as the conduction channel, (iii) formation of source and drain regions by diffusing impurities of conductivity type opposite to that of the substrate, thereby electrically isolating them from the substrate via reverse-biased pn junctions while ensuring good contact between the source and channel through a forward-biased pn junction, and (iv) use of thermally grown SiO_2 as the gate insulator, Dawon Kahng and Martin M. Atalla succeeded, 35 years after Lilienfeld's initial proposal of the FET, in experimentally demonstrating reliable MOSFET operation for the first time (Figures 1.6f). They subsequently presented their results at a scientific conference [6]. Figure 1.9 shows schematic structures of the first MOSFET [6]. Both n-channel and p-channel MOSFETs were fabricated in this first demonstration; however, the n-channel device did not operate properly. In Ref. [6], only a single I_d-V_g curve with an abnormal slope was presented for the n-MOSFET, while all other measurement results were for the p-MOSFET. As seen from the schematic, the gate oxide thickness was already about 100 nm, and the gate length was around 20 μm , values close to those used roughly ten years later when MOS ICs began to be mass-produced (oxide thickness $\approx$ 100 nm, gate length = 10 μm).

Of particular significance is that, over the intervening 35 years, the development of microwave receivers and practical BJTs had established silicon as a highly suitable semiconductor material owing to its relatively high carrier mobility. At the same time, high-purity silicon and single-crystal growth technologies had reached an advanced stage of maturity. Furthermore, by fortunate coincidence, thermally grown SiO_2 on silicon—unlike on other semiconductors—was found to form

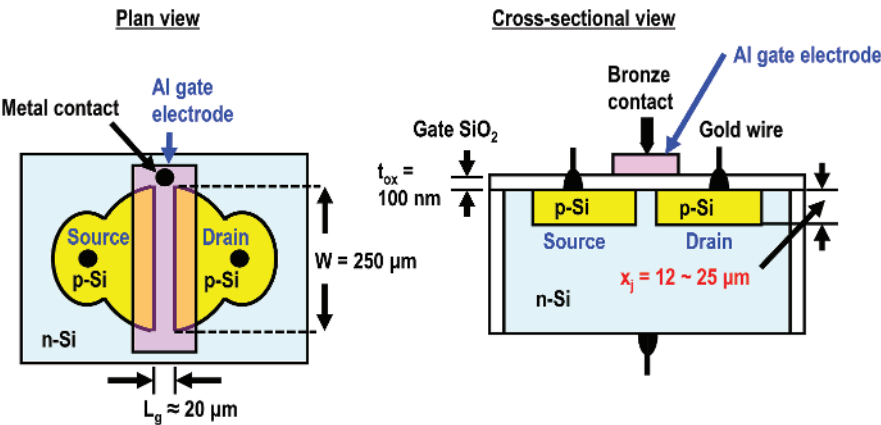


Figure 1.9 Schematic structure of the first MOSFET, published by D. Kahng and M. Atalla in 1960. Source: [6] / with permission of World Scientific Publishing Co Pte Ltd.

an interface with exceptionally low fixed charge and interface-state density, a property that proved essential for enabling high-performance MOS devices.

1.2.2.7 Formulation of the Voltage–Current Characteristics of Field-Effect Transistors

The mathematical formulation of FET current–voltage characteristics began in 1952 with Shockley’s analysis of the JFET [2]. He assumed that the variation of the horizontal electric field along the channel was negligible compared to that of the vertical field in the vertical direction—an assumption he termed the “gradual approximation” [2], later known as the gradual channel approximation (GCA). By applying the one-dimensional Poisson equation in the vertical direction and accounting for both channel and depletion charges under the depletion approximation, he derived the relationship between the local channel potential and the local channel conductance, expressed in terms of the local channel charge density. Incorporating this result into the differential equation for the lateral drift current and integrating from source to drain (i.e., from the source voltage to the drain voltage) yielded an analytical expression for the drain current.

As described in Section 1.2.2.5, in 1953, Brown and Shockley reported the formation of an inversion layer channel at the surface of the base p -region of a Ge nnp transistor after an aging test, which functioned as a back-gate FET without a gate insulator (see Figure 1.6e) [51]. In the same year, Brown, with the support of Shockley, published an analytical expression for the electrical characteristics of the back-gated FET based on the GCA [52]. In his derivation, the vertical Poisson equation accounted for both depletion and inversion charges. His result was expressed as a conductance formula, i.e., a differential form of the current–voltage relationship, but obtaining the explicit current–voltage expression was merely a matter of straightforward integration and thus essentially self-evident.

In 1960, M. Attala applied Brown’s framework to MOSFETs and, by introducing the capacitance of the gate insulator into the Poisson equation, derived for the first time the current–voltage characteristics of MOSFETs [6]. Subsequent formulations of essentially the same type were reported by S. R. Hofstein and F. P. Heiman of RCA (1963) [63], C. Goldberg of Westinghouse (1964) [64], H. Ihantola and J. L. Moll of the Institute of Technology/Stanford University (1964) [65] in 1964, and C. T. Sah of Fairchild/University of Illinois Urbana-Champaign [66] in 1964. All of these derivations assumed drift current only, which is accurate in strong inversion but inadequate in weak inversion, where diffusion current dominates.

A unified treatment, including both drift and diffusion currents, was achieved in 1966 by H. C. Pao and Sah [67]. Using only the GCA, they derived the relation between the surface potential and the gate-source voltage in formulating the drain current that is valid throughout all regions of operations, thereby enabling a unified description of both drift and diffusion currents [58]. Their model, though requiring numerical solutions of a double integral to obtain the drain current, remains a foundational theory for long-channel MOSFETs and is the basis of all surface-potential-based compact modeling. Subthreshold characteristics were later formulated by R. Troutman [68, 69]. For further reading on the Pao-Sah model and compact modeling, see Chapters 2 and 15 in this book, respectively.

In 1975, H. Weiß [53] revealed that as early as 1945, H. Welker had already formulated MOSFET behavior in the low-drain-voltage linear region. Welker assumed a uniform channel potential in the horizontal direction and solved the Poisson equation for the vertical MOS capacitor to determine the channel charge density. His expression also described drift current only, but it remains valid today as an expression for the low-drain-voltage regime.

1.2.3 Evolution of Unit Process Technologies Developed for BJT Fabrication in the 1950s

The fabrication of the MOSFET in 1960 relied on unit process technologies such as photolithography and impurity doping, which had originally been developed for BJT manufacturing. This section reviews the evolution of several key technologies.

1.2.3.1 Development of Single-Crystal Ingot Growth and the Grown-Junction Technique

When the operation of the point-contact transistor was demonstrated in 1947–1948, followed by the invention of the BJT, single-crystal ingots of Ge and Si were not yet available. Consequently, device fabrication experiments were conducted on polycrystalline Si (poly-Si) substrates. To overcome this limitation, Gordon K. Teal and John B. Little of Bell Labs developed single-crystal Ge growth using the Czochralski method [70] in 1948 [21, 71]. The growth of single-crystal Si was subsequently achieved in 1954 by Teal [21]. In 1950, Teal and Morgan Sparks introduced the grown-junction method, the first impurity doping technique for BJT fabrication [72, 73]. By varying the dopant species during single-crystal growth, they successfully formed *npn* and *pnp* junction structures within a single ingot, achieving a base width of approximately 25 μm .

1.2.3.2 Development of Impurity Doping by Diffusion

Between 1952 and 1954, Calvin S. Fuller of Bell Labs conducted pioneering experiments on three semiconductor doping methods: solid-phase [74], liquid-phase [75], and vapor-phase [76] (Figure 1.10). It should be noted that, initially, lithium was used as a donor source, but because its diffusion coefficient was excessively high, its use gradually declined. Due to its practical advantages, vapor-phase diffusion emerged as the preferred approach. In 1962 and 1965, phosphorus diffusion using POCl_3 [77] and boron diffusion using BN [78] were developed, respectively, as shown in Figure 1.11. In the early stages of BJT fabrication, impurity vapor diffusion employed oxides such as P_2O_5 and B_2O_3 as evaporation sources. However, precise control of the evaporation rate was difficult, leading to variations in doping concentration, and P_2O_5 particles were sometimes blown onto the wafer surface, causing device failure. In 1962, Joan M. Crisshal and colleagues at

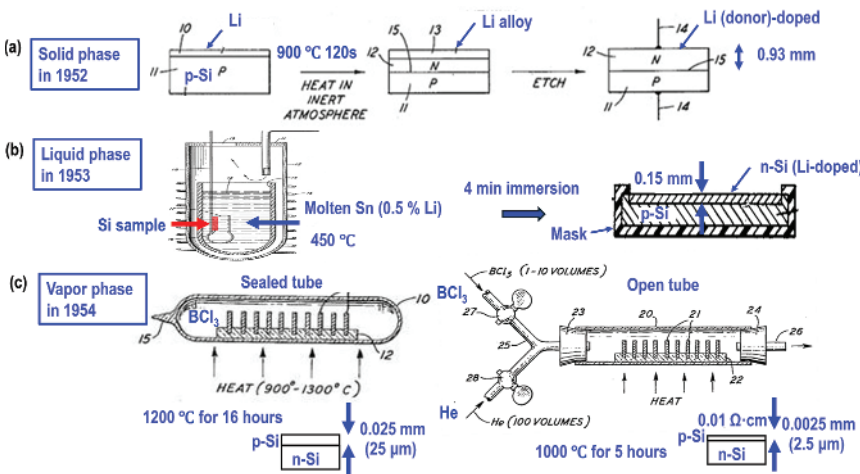


Figure 1.10 Trials of three different doping technologies: (a) solid phase in 1952 [74], (b) liquid phase in 1953 [75], and (c) vapor phase in 1954 [76].

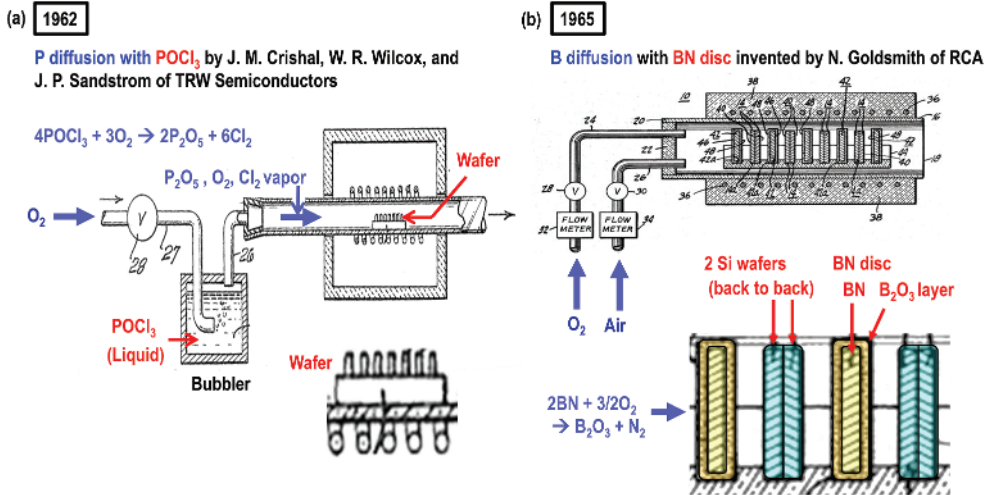


Figure 1.11 Vapor-phase diffusion processes that became standard from the mid-1960s: POCl_3 for phosphorus (a) [77] and BN discs for boron (b) [78].

TRW semiconductors invented a technique to overcome these problems. As shown in Figure 1.11a, an O_2 carrier gas is passed through a liquid POCl_3 bubbler, and the controlled POCl_3 vapor is reacted with O_2 in a high-temperature furnace to generate a regulated amount of P_2O_5 vapor. This technique contributed to advances in BJT fabrication and became the standard for phosphorus vapor diffusion.

For boron diffusion, as shown in Figure 1.11b, in 1965, Norman Goldsmith of RCA developed a technique utilizing the oxidation of a BN disc surface. By controlling the O_2 flow, the concentration of B_2O_3 vapor in the furnace could be regulated, enabling diffusion into silicon wafers. This method became the standard technique for boron vapor diffusion.

Owing to their superior controllability, these processes became the standard vapor-phase doping techniques, widely adopted not only for BJT fabrication but also for source/drain doping in MOS LSI devices. They remained mainstream until the early 1980s, when they were replaced by ion implantation. Notably, POCl_3 continues to be used in solar cell manufacturing today.

1.2.3.3 Ion-Implantation Technology

In June 1954, James W. Moyer of General Electric invented ion implantation for semiconductor doping with *n*- or *p*-type impurities (Figure 1.12) [79]. The impurity was evaporated by a heater and ionized using an electron beam. The ions were then accelerated by a high voltage, such as 400 kV, and bombarded into the semiconductor material (Si or Ge). Following implantation, samples were annealed—for example, at 400°C or 500°C for 12 hours—to activate the dopants. Six months later, William Shockley of Bell Labs independently filed a similar patent [80]. The apparatus described in Shockley's patent was essentially identical to that in Ohl's 1951 patent [81], which had originally been developed to improve metal–semiconductor contacts by irradiating surfaces with nitrogen or other elements. In Shockley's patent specification, it is stated that when an ion stream is deflected by a magnetic field, the radius of curvature differs according to the ion mass, making it possible to select and implant the desired ions.

The principal advantage of ion implantation was the ability to precisely control dopant concentration through accurate measurement and regulation of the ion beam current. However,

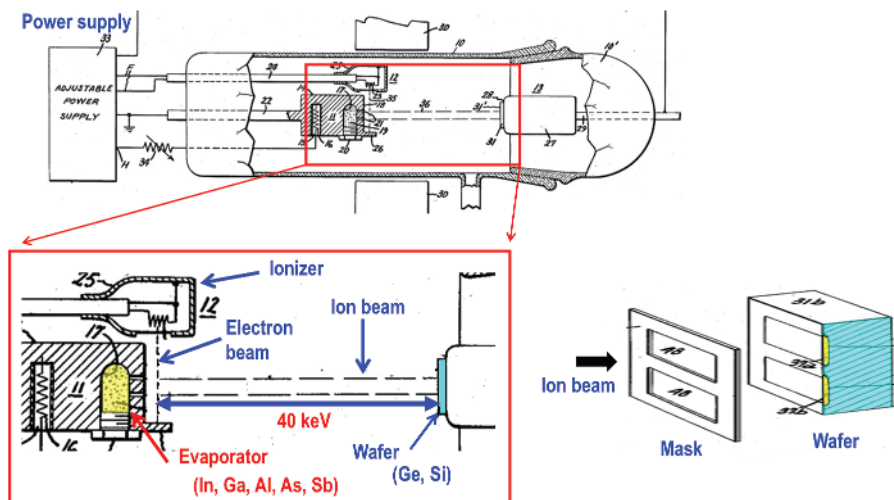


Figure 1.12 Ion implantation invented in June 1954 [79].

challenges remained, including lattice damage caused by energetic ions and the relatively low activation efficiency of implanted impurities. Commercial ion-implantation systems became available around 1970 and were rapidly applied to channel doping in n-channel MOSFETs, enabling precise adjustment of the threshold voltage from negative to positive values. This capability made ion implantation indispensable for NMOS LSI technology. Widespread adoption of source/drain doping began in many companies in the late 1970s and early 1980s.

1.2.3.4 Discovery of SiO_2 as a Diffusion Mask

In the early 1950s, no effective high-temperature vapor diffusion mask existed, and diffusion had to be applied uniformly across the entire semiconductor substrate. In 1955, Carl J. Frosch and F. Derick of Bell Labs discovered that thermally grown silicon dioxide (SiO_2) on Si could serve as an impurity diffusion mask at elevated temperatures [82, 83]. Since photolithography had not yet been introduced, they used wax as an etching mask for SiO_2 , producing SiO_2 stripes that acted as diffusion barriers (Figure 1.13a).

1.2.3.5 Invention of Photolithography

Jules Andrus of Bell Labs demonstrated that photoresist patterns produced by photolithography could serve as etching masks for SiO_2 films on Si [84] (Figure 1.13b). Photolithography was also independently invented in 1957 by Lathrop and Nall at the US Army Diamond Ordnance Fuze Laboratory [85, 86], who projected and reduced patterns onto photoresist layers using a microscope (Figure 1.13c). Their first attempt produced printed aluminum (Al) interconnects for BJTs using the “lift-off” technique, as shown in the same figure. The advent of photolithography enabled the precise formation of micropatterns on semiconductor devices and represented a major milestone in the development of microelectronics.

1.2.3.6 Epitaxial Doping

Epitaxial growth of semiconductor thin films with in-situ impurity doping during deposition was first demonstrated in 1951 by Howard Christensen and Gordon K. Teal of Bell Labs for Ge [87]. Epitaxial growth of Si was later achieved in 1960 by Joseph J. Kleimack, Howard Loar, and Henry

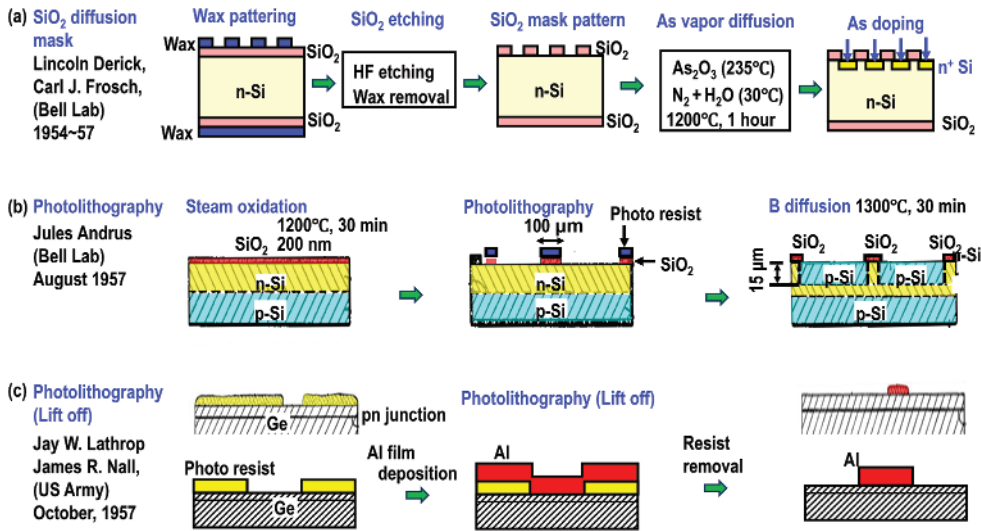


Figure 1.13 Invention of SiO₂ film as a diffusion mask in 1955 [82, 83] and photolithography in 1957 [84–86]. Source: (a) Adapted from [83].

C. Theuerer, also at Bell Labs [88]. Today, epitaxial doping is widely used in FinFETs and GAAFETs to reduce source/drain resistance.

1.2.3.7 Aluminum Ohmic Contact to Both *n*- and *p*-Type Si

Initially, forming reliable ohmic contacts between Al and *n*-type Si was difficult. During alloying, Al—acting as a *p*-type impurity—diffused into *n*-type Si surfaces, unintentionally creating *pn* junctions. To avoid this, Au electrodes doped with antimony (Sb) were used for *n*-type Si, while Al electrodes were employed for *p*-type Si [89]. This issue was resolved in 1958–1959 by Sheldon L. Matlow of Hoffmann Electronics [90] and Gordon Moore of Fairchild Semiconductor [91], who introduced heavily doped n⁺ and p⁺ layers at the contact regions and optimized subsequent thermal treatments. Fairchild Semiconductor thus became the first company to adopt Al contacts for both *n*- and *p*-type Si.

1.2.3.8 Invention of the Planar Structure

In 1959, Jean A. Hoerni of Fairchild Semiconductor invented planar technology, as illustrated in Figure 1.14b [92–95]. Its central innovation was covering and protecting the Si surface with thermally grown SiO₂. Although M. Attala and colleagues had already demonstrated the passivation effect of SiO₂ [62], Hoerni extended this concept by deliberately retaining the diffusion mask oxide as a protective layer for *pn* junctions, thereby suppressing leakage currents. Prior to this, BJT isolation and the reduction of base–collector junction areas relied on mesa structures, in which the surrounding Si substrate was etched away to leave the device as a mesa, as illustrated in Figure 1.14a [89, 96, 97]. The planar approach eliminated the need for mesas, yielding a flat device topology. This effectively resolved *pn* junction leakage issues on Si and Ge surfaces and improved both the yield and reliability of interconnects traversing device steps.

1.2.3.9 Practical Monolithic Integrated Circuits of BJTs

Although the monolithic integration of multiple mesa-type BJTs with bonding-wire interconnects had been fabricated in 1958 [96] and subsequently proposed in a patent in 1959 [97], the

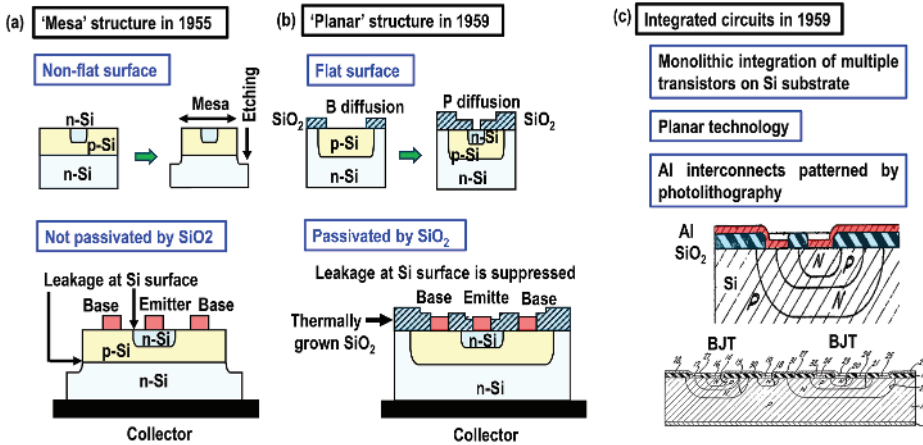


Figure 1.14 (a) Mesa structure [89]. (b) Planar structure [92–95]. (c) Integrated circuits [98].

first practical monolithic ICs—regarded as prototypes of modern ICs—were instead conceived and developed later that year by Robert Noyce of Fairchild Semiconductor, as illustrated in Figure 1.14c [98]. Noyce's circuits monolithically integrated multiple *npn* BJTs, *n*-type resistors, and *pn*-junction capacitors on a single Si chip. Utilizing planar technology for device isolation and a flat-surface topology, they employed aluminum thin-film interconnects patterned by photolithography. This technological approach was subsequently extended to MOS ICs, thereby establishing the foundation of modern microelectronics.

1.3 1960–1970: The MOSFET Instability Problem

Following the MOSFET's demonstration at Bell Labs in 1960, many companies recognized its potential for IC transistors [15]. Fairchild Semiconductor [94], founded in 1957 by Robert Noyce, Gordon Moore, and others after leaving Shockley Semiconductor, was already producing silicon BJT ICs. These micrologic devices, used in the Apollo Guidance Computer with over 2000 units [99], employed Resistor–Transistor Logic (RTL) with two three-input NOR gates (six BJTs per IC). Their success raised expectations for MOS-based ICs. In 1963, F. Heiman and S. Hofstein at RCA constructed experimental 16-transistor NMOS logic ICs (gate length 15 μm , gate width 125 μm , gate oxide thickness 125 nm, and maximum drain voltage 35 V) [63, 100]. The chip arranged NMOSs in a matrix across the die, and by altering the layout of the aluminum interconnections linking the MOSFETs, it was possible to implement different logic circuits.

However, it took nearly a decade before reliable mass production of MOS ICs was achieved. The greatest obstacle was unstable charge states at the Si/SiO₂ interface or within the oxide, which undermined device stability. Attempts to integrate hundreds of MOSFETs also frequently failed due to immature processes, contamination, and photomask misalignment, resulting in very low yields.

1.3.1 Instability Arising from Unpredictable Charges

During the 1960s, MOSFETs exhibited pronounced instability in their electrical characteristics [101]. Large parameter variations and temporal drifts posed serious obstacles to commercialization. A full understanding and resolution of these instability issues required several years, with substantial progress achieved only in the late 1960s. One of the earliest major research efforts focused on detailed studies of the silicon thermal oxidation process [102, 103]. The mechanisms underlying the instability proved highly complex, involving multiple types of charges located both at the Si/SiO₂ interface and within the gate oxide. These charges not only induced instability but also degraded overall MOSFET performance [104–111]. Figure 1.15 [111] illustrates interface trap charges located directly at the interface and fixed oxide charges situated near the interface. In practice, these charges are distributed throughout the transition region (SiO_x) between the silicon and the oxide. Charge trapping and detrapping in this transition region contribute to phenomena such as 1/f noise and random telegraph noise. The behavior of these charges depends on both their spatial distance from the interface and the energy difference between the trap state and the Fermi level [112–117]. Many fixed charges in the SiO_x transition region are positively charged, originating from structural defects caused by oxygen deficiency. These positively charged defects are a primary cause of the negative threshold voltage observed in n-channel MOSFETs. Defects within the SiO₂ layer can give rise to interface trap charges (surface states), fixed oxide charges near the interface, and oxide trap charges within the bulk oxide [118]. The formation and density of these charge states are influenced by thermal oxidation conditions such as ambient atmosphere and temperature [119–121].

Radiation exposure during MOSFET fabrication—such as X-rays or ion bombardment—can also induce charge generation [122–124]. For instance, heating an aluminum target with an electron beam during Al evaporation emits X-rays [125]. Charges trapped near the Si/SiO₂ interface reduce channel carrier mobility, thereby degrading MOSFET performance. Trapping and detrapping at these sites are among the primary causes of MOSFET instability.

Another category of charges within the gate oxide consists of mobile ions, such as sodium (Na⁺) and potassium (K⁺) [126–130]. These ions drift under the influence of the gate electric field, causing shifts in the MOSFET threshold voltage. Although their drift velocity is generally much slower than that of electrons or holes, it increases significantly at elevated temperatures. Because these ions are positively charged, their accumulation at the Si/SiO₂ interface under positive gate bias in n-MOSFET poses serious reliability concerns. Accumulated ions not only shift the threshold voltage but also degrade the subthreshold slope and reduce electron mobility in the channel.

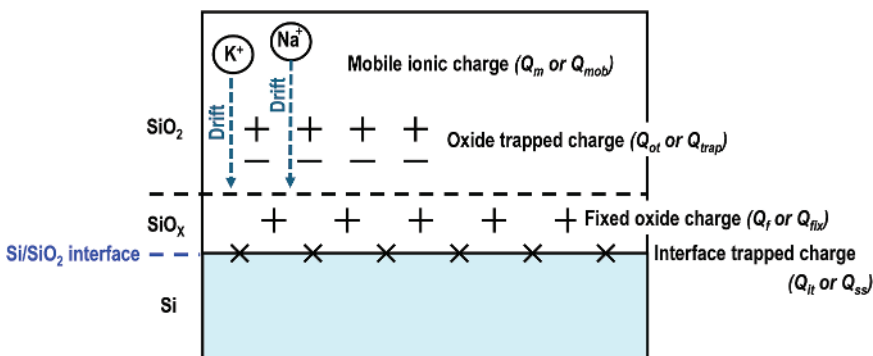


Figure 1.15 Various mysterious charges in the Si–MOS structure. *Source:* Adapted from [111].

1.3.2 Stabilization of Interface Charges by Low-Temperature Hydrogen Annealing and Surface Orientation Control

By the mid-1960s, low-temperature hydrogen annealing at approximately 400–500°C was gradually demonstrated to effectively reduce interface charge density at the Si/SiO₂ interface and enhance interface stability [131–133]. During annealing, hydrogen atoms chemically passivate dangling bonds at the interface, thereby suppressing electrically active defect states and enhancing overall device reliability. Interface charge density also depends strongly on the crystallographic orientation of the silicon substrate. Minoru Ono, Toshimitsu Momoi, and Youji Kawachi of Hitachi reported that Si (100) surfaces exhibit significantly lower interface charge densities than Si (110) or Si (111) surfaces. This finding, illustrated in Figure 1.16, was the subject of a 1964 patent application [134]. The impetus for this work was a reliability issue encountered in 1961 during high-temperature aging tests of mesa-type BJTs, namely, leakage currents at the *pn* junction sidewalls caused by interfacial charge generation, a problem already reported by Brown of Bell Laboratories in 1953 (see Section 1.2.2.5 [51, 52]). To address this, they measured the fixed charge density on three different Si surface orientations and also fabricated prototype MOSFETs, experimentally confirming that the (100) surface provided the highest current drive capability. Until then, the (111) surface—naturally exposed by cleaving Si crystals—had been used for both BJTs and MOSFETs [135].

Subsequent studies in 1965 by research groups at Tektronix (led by Jean F. Delord) [136], NEC (led by Yoshio Miura) [137], and IBM (led by P. Balk) [138] confirmed these results. Consequently, the adoption of Si (100) substrates in MOSFET manufacturing gradually increased, although Si (111) remained in widespread use until the early 1970s.

1.3.3 Countermeasures Against Mobile-Ion-Induced Instability

The first step in mitigating mobile-ion effects was improving the cleanliness throughout the transistor fabrication process, including materials, chemicals, equipment, and the cleanroom environment [139]. In 1965, Werner Kern at RCA developed the “RCA Standard Clean,” a multi-step liquid chemical procedure using reagents like HCl, HF, H₂SO₄, H₂O₂, and NH₄OH to remove metallic and organic contaminants [140, 141], which became widely adopted.

Beyond wet cleaning, HCl gas treatment effectively removed sodium contamination from wafers and quartz furnaces [142–144]. These efforts aimed to reduce mobile-ion contamination and enhance manufacturing cleanliness. In the 1960s, heavy-metal contamination in BJT *pn*

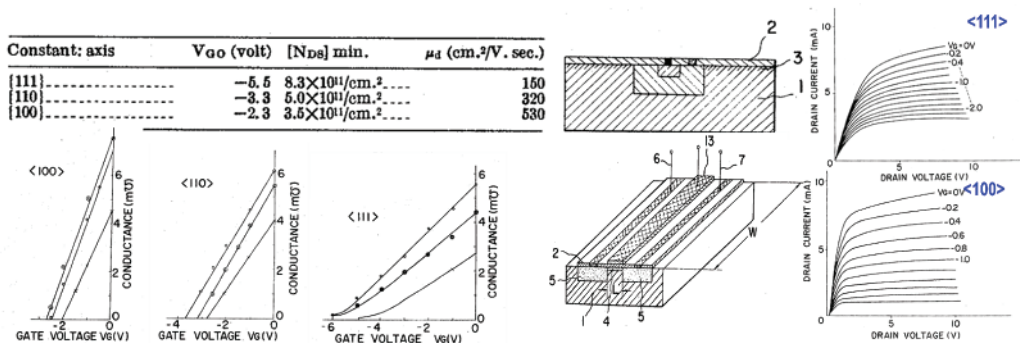


Figure 1.16 Dependence of Si surface orientation on interfacial charge, carrier mobility, and MOSFET conductance in 1964 [134, 135].

junctions—causing leakage currents—was a major issue. In 1960, Goetzberger and Shockley reported that phosphosilicate glass (PSG) and borosilicate glass (BSG) layers had a “gettering” effect in silicon [145]. During high-temperature annealing (e.g., 1000°C in nitrogen), metals like Cu, Au, Fe, and Mn diffused to the surface and precipitated, where PSG and BSG films adsorbed them. These films were formed by heat-treating $\text{SiO}_2/\text{P}_2\text{O}_5$ or $\text{SiO}_2/\text{B}_2\text{O}_3$ stacks. In 1963, S. W. Ing, Jr., of GE confirmed this effect using PSG-containing insulating films [146], and in 1964, Donald R. Kerr of IBM showed PSG layers could prevent contamination in *npn* BJTs [147, 148]. These studies focused on heavy metals that degraded carrier lifetime and increased leakage [149, 150]. Gettering was also observed during phosphorus vapor diffusion from PBr_3 sources [151, 152].

In 1968, Pieter Balk and David W. Dong of IBM patented the use of stacked SiO_2 /PSG films as gate insulators [153, 154]. PSG helped capture sodium ions and stabilize MOSFET operation [152–154], but polarization of PSG gate films introduced new instabilities [155]. Thus, PSG was used as a passivation layer above Al gate electrodes to block sodium, potassium, and heavy metal ions [156]. Other protective films like Si_3N_4 and Al_2O_3 were also employed [157]. POCl_3 gettering eventually became standard, alongside PSG passivation above gate and interconnect layers.

By the late 1960s, innovations such as cleanroom refinement, RCA cleaning, strategic gettering, and optimized oxidation and annealing (including hydrogen annealing) significantly reduced fixed and mobile charge densities in MOS structures. These advances enabled the shift from lab prototypes to reliable, high-yield MOSFET manufacturing for ICs.

Resolving these instability challenges during this formative decade laid the foundation for MOS technology’s rapid growth in the 1970s. With improved interface quality, reduced mobile-ion contamination, and better process control, MOSFETs achieved the performance and scalability needed for LSI, establishing MOS as the dominant semiconductor device structure.

1.4 From MOS ICs to MOS LSIs: 1965–1969

By the mid-1960s, MOSFETs were still facing challenges related to instability caused by interface charges. Nevertheless, efforts toward the commercialization of MOS ICs had already begun, and limited product shipments were underway. This section outlines the circumstances from the mid-1960s up to the advent of the MOS LSI era in 1969, based on the content presented in [13–20] and other individual sources.

During the 1960s, BJT ICs were predominantly employed in military and aerospace systems, as well as in mainframe computers. Despite their utility, these ICs were costly and exhibited high power consumption. Early BJT ICs typically integrated only a small number of transistors—commonly in configurations such as RS (reset-set) flip-flops or triple-input NOR gates. Although integration levels gradually improved, incorporating more than 100 BJTs on a single chip remained challenging, particularly in applications requiring high reliability and compactness, such as precision-guided weapons or secure communication systems. Key limitations included heat dissipation, power delivery, and manufacturing complexity, all of which constrained further scaling.

The fabrication of bipolar ICs was inherently complex, and fundamental limits on transistor miniaturization persisted. While BJTs within ICs consumed less power than their discrete counterparts, overall system-level power consumption remained substantial. In large-scale assemblies involving hundreds of ICs, thermal management became a critical concern. This was especially problematic in environments with stringent size and weight constraints—such as ballistic missile guidance systems—where cooling capacity was severely restricted by design.

In contrast, MOSFETs offered advantages in power efficiency but suffered from lower operating speeds and higher low-frequency ($1/f$) noise compared to BJTs. Consequently, early MOS ICs were limited to low-speed, low-power digital logic applications, including standard logic circuits, calculators, and wristwatches, as well as cost-sensitive memory devices. However, concerns regarding MOSFET instability—particularly related to interface charge effects—cast doubt on the commercial viability of MOS ICs during their initial development phase. Commercial production of MOS ICs commenced in the mid-1960s, initially focused on small-scale logic circuits. Subsequently, attention shifted toward the implementation of middle-scale logic circuits—including multiplexers, registers, accumulators, and AD/DA (analog-digital/digital-analog) converters—while small-capacity SRAM and DRAM memory units also came to be fabricated. By the late 1960s and into the early 1970s, LSI became feasible for mass production. During this period, mainstream MOS technology was based on non-self-aligned aluminum-gate PMOS processes, which remained dominant until the early 1970s.

1.4.1 Commercialization of MOS ICs in 1965: Shift Registers, Standard Logic, and the Integration of Several to over 100 MOSFETs

In the early 1960s, Fairchild Semiconductor was generating profits with BJT ICs and was not enthusiastic about developing MOS ICs. Dissatisfied with this situation, Robert H. Norman, Howard Bobb (later co-founder of American Microsystems Inc. (AMI)), and Phil Ferguson left Fairchild in 1963 and established General Micro-electronics (GMe). GMe became a pioneer in the commercial production of highly integrated MOS ICs. In June 1964, Norman and Homer E. Stephenson successfully developed a dual 20-bit dynamic shift register using 120 PMOSs and subsequently obtained a patent for it [158]. This IC was adopted in a military Digital Differential Analyzer, and it became the world's first commercially manufactured MOS IC, designed by Norman. Building on this success, GMe developed and marketed a family of small- to medium-scale MOS ICs under the trade name “PicoLogic.” These ICs were also used in advanced systems, such as interplanetary surveillance platform satellites. In the same year, General Instrument (GI) Corporation also entered the market, offering a 265-bit shift register using PMOS technology. Frank Wanlass, credited with the co-invention of CMOS at Fairchild, joined GMe in 1964, and in 1965, he moved to GI, where he played a central role in IC development. From the mid-1960s to the early 1970s, the US MOS IC/LSI market was dominated by GMe, GI, and AMI, which was founded in 1966. AMI specialized in custom aluminum-gate PMOS ICs for calculators, and by the early 1970s, it was highly regarded as a major MOS IC manufacturer. In 1966, Lee Boysel transferred from IBM to Fairchild, where he became the head of the newly formed MOS group. There, he developed a read-only memory (ROM) in 1966, the ALU Slice 3800 (an 8-bit parallel accumulator) in March 1967, and a 256-bit DRAM in 1968 [159].

1.4.2 Development of Custom Logic ICs for the Victor 3900 by GMe in 1965

In October 1964, Victor Comptometer Corp. commissioned GMe to design and manufacture its electronic calculator. To implement around 50 logic functions, GMe created 23 custom logic ICs using about 300 PMOSs each—marking the first custom logic MSIs. By April 1965, GMe delivered 25 units. However, the high cost of producing 23 unique IC types strained GMe financially. As engineers left and yields fell, GMe could not meet Victor's demands. In December 1965, Philco-Ford acquired GMe, and Victor ended the contract in 1966. Philco-Ford rebranded the calculator as the “Philco-Ford 3900,” used internally at Ford, but production ceased in 1968.

1.4.3 Evolution of Sharp's Electronic Calculators: From Discrete BJTs in 1964, Through MOS ICs by Hitachi and NEC in 1967, to the First MOS LSI by North American Rockwell in 1969

Japan's electronic calculator boom in the late 1960s spurred innovation in MOS LSI technology. Sharp [160, 161], led by Tadashi Sasaki [162], played a central role, along with Canon, Casio, and Busicom. In 1964, Sharp released the world's first all-transistor calculator, the CS-10A ($42 \times 44 \times 25 \text{ cm}^3$; 25 kg; 90 W; ¥535,000), which employed 530 Ge BJTs and 2300 diodes. In the same year, Canon introduced its all-transistor calculator, the Canola 130 (600 Ge BJTs, 1600 diodes; ¥395,000), followed by the Casio 001 (1965), Toshiba BC-1001 (1965), and Busicom Unicon 160 (1965), and from 1966 onward by numerous other Japanese manufacturers, intensifying competition [163].

In 1966, the CS-31A ($40 \times 48 \times 22 \text{ cm}^3$, 13 kg, ¥350,000) by Sharp incorporated 28 Mitsubishi BJT ICs, 553 BJTs, and 1549 diodes, thereby becoming the first commercial calculator to use ICs (BJT ICs). The 1967 CS-16 ($29.7 \times 31.7 \times 11.7 \text{ cm}^3$; 2.54 kg; 10 W; ¥230,000) by Sharp incorporated 56 MOS ICs manufactured by Hitachi, among them the HD701—a 16-bit shift register comprising 55 MOSFETs (15 μm gate length) per IC, each mounted in a 10-pin TO-5 (Transistor Outline package type 5) package. According to additional sources, certain variants of this calculator were equipped with 72 MOS ICs from NEC's μPD1 series, demonstrating a growing diversity in the domestic supply of semiconductor components.

In 1968, Sharp built a prototype calculator with 11 MOS LSIs and collaborated with Autonetics (Rockwell) on calculator LSI development. Autonetics, once part of North American Aviation, was absorbed into Rockwell in 1967 through the formation of North American Rockwell (later Rockwell International) after developing ICs for military navigation. Rockwell's investment in MOS LSI technology established the era's most advanced MOS IC production facilities. By autumn, logic was condensed into four chips. The resulting QT-8D ($13.5 \times 24.7 \times 7.2 \text{ cm}^3$; 1.4 kg; ¥99,800) was launched in 1969 as the world's first MOS LSI calculator, with North American Rockwell among the first LSI producers of the era.

In 1973, Sharp released the EL-805 ($7.8 \times 11.8 \times 2 \text{ cm}^3$, 200 g, 0.02 W, ¥26,800), the world's first pocket calculator with CMOS LSI and liquid crystal display (LCD). This model incorporated a CMOS LSI and two display driver ICs, both manufactured by Toshiba, and it marked a turning point in improving the power efficiency and miniaturization of personal electronic devices.

In 1976, the EL-8026 ($6.5 \times 10.9 \times 0.95 \text{ cm}^3$, 65 g, ¥24,800) by Sharp became the first solar-powered calculator. The 1985 EL-900 ($8.55 \times 5.4 \times 0.08 \text{ cm}^3$, 11 g, ¥7800) marked the peak of miniaturization and low-power design. Comparing the CS-10A and EL-900 shows reductions in component count (1/2800), size (1/12,500), weight (1/2300), cost (1/69 or 1/1800 adjusted), and power (1/4500 versus EL-805). Sharp's trajectory—from BJTs to CMOS LSIs—illustrates the transformative impact of MOSFET-based LSI technology on personal electronics.

Texas Instruments (TI) developed the Cal-Tech prototype calculator in 1967 ($10.8 \times 15.6 \times 4.4 \text{ cm}^3$, 1.28 kg), later commercialized by Canon in 1971 as the Pocketronic, using four MOS LSIs [164–166].

1.4.4 Development of a 64-Bit PMOS Static RAM by Fairchild Semiconductor Co. in 1964 and Memory Developments by Other Companies

A key application for MOS ICs was high-density, low-power, low-cost memory for computers and controllers. While magnetic core memory was common [167–169], semiconductor memory offered better integration, speed, and cost-efficiency. In the mid-1960s, both BJTs [170–174] and MOSFETs [175–188] were used to develop SRAM. Though BJT memories had faster access times, they were

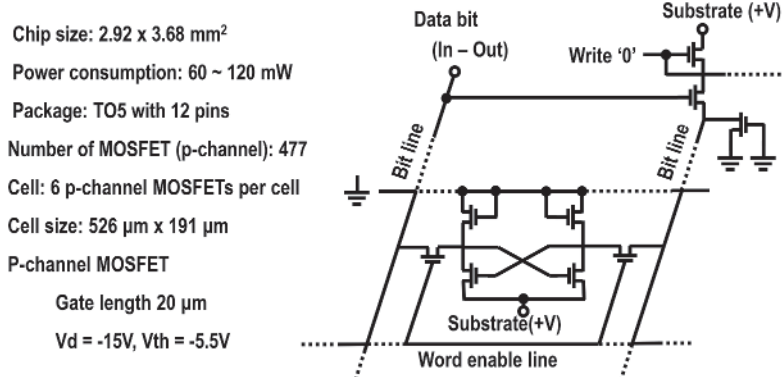


Figure 1.17 World's first MOS SRAM (64-bit) developed in 1964, consisting of six p-MOSFETs per memory cell. Source: [175] / with permission of Horizon House Publications, Inc.

about ten times larger in cell size, power use, and cost than MOSFET memories, making them less competitive [186–188].

In 1964, J. Schmidt of Fairchild created the first MOS IC memory (Figure 1.17), a 64-bit PMOS SRAM using aluminum gates, with each cell made of six p-MOSFETs [175]. The chip ($2.92 \times 3.68 \text{ mm}^2$) housed 477 p-MOSFETs, with cell size $\sim 210 \times 380 \mu\text{m}^2$ and gate length $15\text{--}20 \mu\text{m}$. It used a 12-pin TO-5 package, operated at -15 V , and had a threshold voltage of -5.5 V . Cycle times were $2\text{--}3 \mu\text{s}$, and standby power was $60\text{--}120 \text{ mW}$. Due to low yields, development ended in 1965 before full-bit operation [15].

Until the late 1960s, memory R&D focused on cell-level work, with few complete RAMs. In 1967, Hirohiko Yamamoto and NEC developed a 144-bit NMOS SRAM [181, 182], though stable prototypes only emerged in the early 1970s with Al_2O_3 passivation. Also in 1967, J. F. Allison at RCA introduced a CMOS SRAM using Silicon-On-Sapphire (SOS) [183, 184]. RCA's SOS-CMOS reached mass production by 1974–1975, yielding 4 and 16 K SRAMs. These offered much lower power consumption and high radiation resistance, ideal for space applications [189, 190], with NASA and the US Air Force as key customers.

1.4.5 Summary of MOS IC Development in the 1960s

In 1960, Bell Labs demonstrated the MOSFET, initiating MOS-based IC development. Due to early instability and immature fabrication, commercial MOS ICs emerged only by 1964–1965, led by GMe, a Fairchild spinoff. GMe's SSI and MSI circuits offered lower power than BJTs and were used in military and satellite systems. In 1965, GMe developed calculator MSIs but faced yield issues. By 1967, Japanese firms like Hitachi and NEC supplied more reliable MSIs, and calculator demand accelerated custom logic MSI development. In 1969, Rockwell produced four LSIs for Sharp, integrating nearly 1000 MOSFETs per chip. By 1970, PMOS LSIs with aluminum gates marked the start of the LSI era.

In logic advancement, RCA launched the CMOS 4000-series in 1969, and Toshiba's early 1970 CMOS calculator LSIs further cut power consumption. In memory, Fairchild began a 64-bit PMOS SRAM in 1964 but abandoned it due to poor yields. They later developed a ROM (1966) and a 256-bit DRAM (1968). NEC's 144-bit NMOS SRAM (1967) suffered from ion instability until the early 1970s. In October 1966, R. Bower and H. Dill of Hughes Aircraft presented silicon-gate technology (to be described in the next section) at the International Electron Devices Meeting (IEDM) [191].

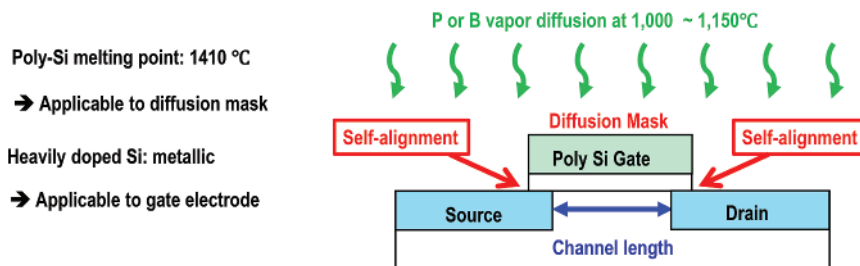
Inspired by this, Federico Faggin at Fairchild developed silicon-gate technology [192], fabricating the first silicon-gate IC in 1968—a PMOS 8-port multiplexer [193]. Intel adopted the technology in 1969 for its first memory LSI, the 256-bit PMOS SRAM 1101 (to be described in the next section) [194, 195]. This breakthrough spurred further memory LSI developments at TI, IBM, and numerous Japanese companies.

1.5 Technologies for MOS Integrated Circuits Developed Between 1965 and 1970

In the latter half of the 1960s, several pivotal innovations were introduced to enhance device integration while maintaining the minimum feature size, which at the time was approximately 10–20 μm . Among these, the adoption of poly-Si as the gate electrode material enabled the realization of self-aligned gate structures with respect to the source and drain regions [196]. Furthermore, a “direct-contact” technology was developed, allowing poly-Si gate electrodes to be connected directly to the source/drain regions without intermediate aluminum interconnects [197].

1.5.1 Silicon-Gate Technology

Since the invention of the first MOSFET in 1960, aluminum has been widely used as the gate material. However, its relatively low melting point of 660°C rendered it unsuitable as a mask for high-temperature impurity diffusion processes. In contrast, heavily doped poly-Si exhibited metallic electrical properties and a significantly higher melting point of 1410°C, allowing it to serve as both as a gate electrode and a diffusion mask at processing temperatures around 1000°C. As illustrated in Figure 1.18, this enabled the implementation of self-aligned gate structures with improved process compatibility.



Inventor	Idea conceived	Experiments	Filed or priority	Patented	Paper publication
Philco-Ford (GMe) Watkins	1965.03, 1965.08.09	No record	1966.09.26	1971.04.27	Non
Hughes Aircraft Dill	1966.05.01	? ~ 1966.10	1966.10.26	1970.12.01	1966.10 IEDM
Bell Labs. Kerwin et.al	1966.02	1966.02 ~ 05	1967.03.27	1969.10.28	1967.07 SSE

Figure 1.18 Si gate technology and self-alignment of gate electrode and source/drain [191, 196, 198–202].

According to Ref. [15], this technology was already known at GMe Corporation as early as March 1965. Subsequently, three companies—depicted in the table in the same figure—filed patent applications and were each granted patent rights [198–200], accompanied by publications from Hughes Aircraft [191] and Bell Laboratories [201]. However, later judicial rulings recognized Bell Laboratories as the first to fabricate prototype devices [196] and to demonstrate the effectiveness of the approach, thereby awarding Bell the primary patent rights [202, 203].

1.5.2 Direct-Contact Technology

In Si-gate technology, a substantial reduction in chip area can be achieved if the drain and gate are contacted directly, rather than being connected by Al interconnects. As illustrated on the left of Figure 1.19, if the gate insulator at the intended contact site is removed by masking prior to the deposition of the poly-Si film, a direct contact between the gate electrode and the drain becomes possible. The poly-Si is pre-doped in situ with a high concentration of phosphorus, and during the subsequent high-temperature vapor-phase diffusion used for source/drain doping, the poly-Si is further enriched with phosphorus. At this stage, sufficient phosphorus also diffuses into the Si substrate at the direct-contact region, thereby merging this portion with the n⁺ drain layer to form the drain region.

This invention was filed as a patent [197] in 1970 by Leslie Vadasz of Intel. According to F. Faggin, however, the concept had in fact been devised by Faggin himself in the late 1960s, when he and his supervisor Vadasz were still at Fairchild, and it was Faggin who urged Vadasz to file the application [193]. Although the patent was officially published in October 1972, by that time, the direct-contact technique had already become an obvious and widely adopted practice among companies employing Si-gate technology.

1.5.3 Metal Gate Technologies and S/D Extension

Robert Bower at Hughes Aircraft Company filed US Patent 3,472,712 [204] on October 27, 1966 (Figure 1.20b), the day after his colleague Hans Dill submitted the application for the Si-gate technology patent [199] on October 26 (Figure 1.20a). In his patent, Bower proposed a method in which the Al gate served as a mask for ion implantation to form the source and drain in self-alignment with

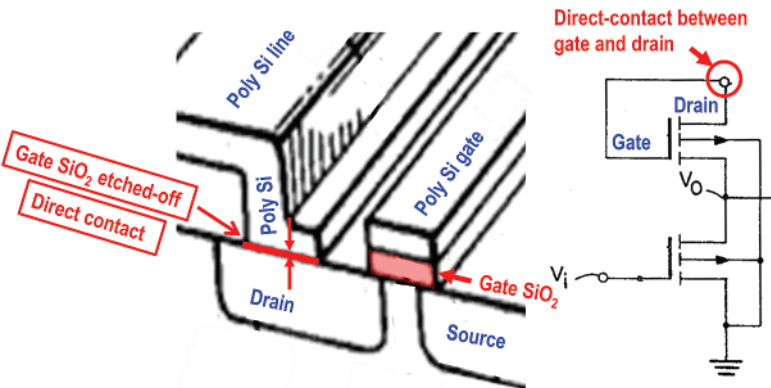


Figure 1.19 Direct contact between the poly-Si gate electrode and source/drain, invented in 1970 [197].

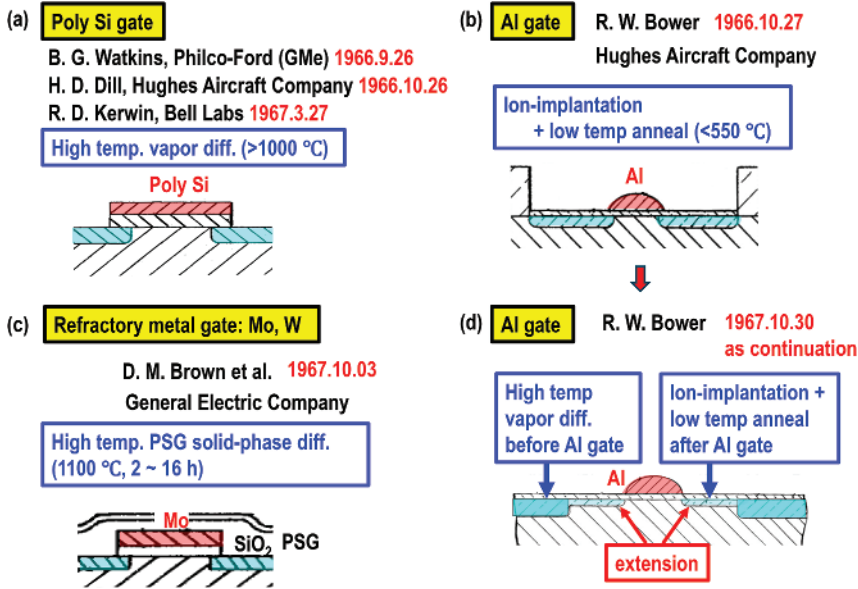


Figure 1.20 Self-alignment technology using poly-Si [198–200], Al [204, 205] and refractory metal gates [206].

the gate electrode. However, due to the thermal stability limitations of aluminum, the activation temperature of the implanted dopants was restricted to below 550 °C [207], resulting in insufficient activation and consequently high-resistance source and drain regions [207–209]. Furthermore, when the same Al wiring was used for both the source/drain contacts and the gate electrode, the contact regions beneath the wiring remained undoped, necessitating the formation of separate Al interconnects for the source and drain after ion implantation. In addition, the high-resistance source/drain regions led to high contact resistance. To overcome these issues, Hughes filed a continuation of US Patent 3,472,712 on October 30, 1967, which was granted as US Patent 3,615,934 [205] (Figure 1.20d). This continuation patent first applied high-temperature vapor diffusion to all regions except the future Al gate electrode area and its immediate surroundings, thereby forming low-resistance source and drain regions with deep, heavily doped impurity profiles. Subsequently, using the Al gate as a mask, dopants were ion-implanted in self-alignment only around the gate, producing shallow, high-resistance source/drain extensions [205].

By 1970, this technology had been implemented in practical products, including a 64-bit shift register operating at 30 kHz, a 50-channel multiplexer, and a 2048-bit ROM [17, 207, 210, 211]. The extension regions, characterized by shallow junction depth, low impurity concentration, and high resistivity, contributed to reduced Miller capacitance and to improved breakdown voltage characteristics through the alleviation of the electric field at the drain [207, 210, 211]. This concept later evolved into the lightly doped drain (LDD) structure [212, 213]. Nevertheless, Hughes Aircraft was the primary company to adopt self-aligned Al-gate MOSFET technology, while many other firms continued to use non-self-aligned Al-gate PMOS technology until silicon-gate technology became the industry standard in the early 1970s.

As also shown in Figure 1.20c, experimental efforts during this period explored the use of refractory metals as gate electrode materials [206, 214–216]. These materials would later play a critical role in the high-*k* metal gate stack technologies introduced after 2007.

1.5.4 Poly-Si Gate Doping and Gate Electrode Work Function

By the end of the 1960s, Fairchild Semiconductor demonstrated that the dopant type in poly-Si gate electrodes significantly influenced the work function, thereby affecting the threshold voltage, as shown in Figure 1.21 [217]. It was found that n^+ poly-Si for n -MOSFETs and p^+ poly-Si for p -MOSFETs were suitable. In the PMOS LSI era beginning in 1969, and later in the NMOS LSI era starting around 1973, when Si gates were employed, the choice of poly-Si gate doping was made as described in this patent [217]. Although similar choices were sometimes made for CMOS, many companies continued to use Al gates in the 1970s. For CMOS, however, in order to simplify the process, it became common practice to fabricate both device types with n^+ poly-Si gates, as shown in Figure 1.22. Threshold adjustment for p -MOSFETs was achieved by counter-doping the channel region with boron to form a p -type region. In this configuration, the threshold voltage of p -MOSFETs became excessively negative, and to lower it to an appropriate level, the p -MOSFET

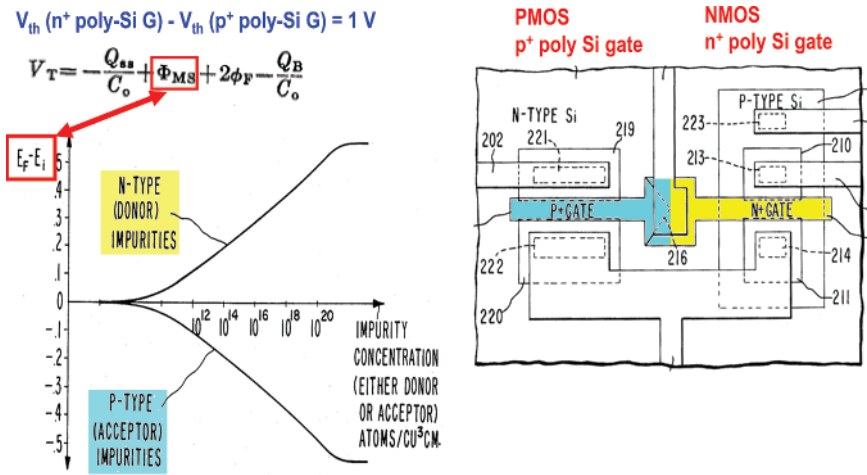


Figure 1.21 Selection of n^+ poly-Si gate for n -MOS and p^+ poly-Si gate for p -MOS to achieve appropriate threshold voltages in 1970 [217].

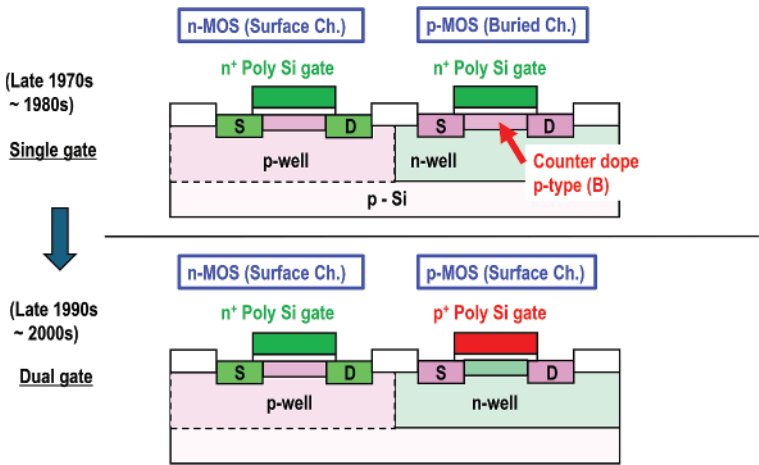


Figure 1.22 Transition from a single poly-Si gate electrode to a dual poly-Si gate electrode in the early 1990s.

channel region was doped with acceptor impurities such as boron, of the same conductivity type as its source and drain.

Consequently, the p-MOSFET channel was not an inversion channel but rather a *p*-type buried conductive layer. Because the current in a buried channel flows slightly below the surface, the gate voltage exerts weaker control over the channel current, resulting in pronounced short-channel effects (SCEs). To overcome these limitations, dual-gate CMOS technology—employing n^+ poly-Si gates for n-MOSFETs and p^+ poly-Si gates for p-MOSFETs—was introduced in the early 1990s, as discussed in Section 1.10.3.2.

1.5.5 PSG Reflow

To further improve Al wiring, Intel developed a PSG reflow technique aimed at planarizing regions containing Al interconnects [218]. By leveraging the decreased viscosity of phosphorus-rich PSG films at elevated temperatures, this method enabled surface smoothing and contributed to improved reliability of Al interconnects.

1.6 First-Generation LSIs—Al- or Si-Gate PMOS LSIs (1969–1971) with 10–8 micrometer Design Rules

1.6.1 Al-Metal Gate PMOS Custom Logic LSI for Calculators

The first generation of LSI began in 1969, employing p-channel MOSFETs in PMOS LSI circuits. The reasons are explained in the next section. As outlined in Section 1.4.3, Sharp released the QT-8D calculator in 1969, incorporating four LSIs manufactured by North American Rockwell: AC2261 (master controller), AU2271 (adder), NRD2256 (keyboard input and display output control), and DC2266 (decimal point control). Each device integrated approximately 1000 p-MOSFETs. A four-phase logic scheme was adopted for these circuits [219–223].

In 1970, GI Corporation supplied four PMOS LSIs to Sanyo for use in the ICC-82D calculator [9, 222]. These devices utilized metal gate electrodes and silicon nitride (Si_3N_4) gate insulators. Around the same time, TI Inc. provided three MOS LSIs for Canon's Pocketronic calculator, which was released in 1971, as described in Section 1.4.3.

In 1970, non-self-aligned Al gate PMOS ICs/LSIs accounted for approximately 70% of the US MOS IC/LSI market [9]. Major suppliers of Al gate PMOS ICs/LSIs included AMI, GI, and TI. Self-aligned Si gate PMOS ICs/LSIs held a market share of roughly 25%, with key suppliers such as National Semiconductor (NS), Intel, and Fairchild Semiconductor. The market share of CMOS ICs/LSIs remained only a few percent.

1.6.2 Intel's Silicon Gate PMOS LSI

Robert Noyce and Gordon Moore founded Intel in July 1968 after leaving Fairchild. In August–September 1969, Intel launched its first commercial products, including the 64-bit bipolar SRAM (3101) and the 256-bit PMOS SRAM (1101) [194, 195]. Around the same time, IBM developed a 128-bit bipolar SRAM for its System/370 Model 145 [223]. Intel's 1101 used six p-MOSFETs per cell, integrating over 1000 MOSFETs on a chip. Built with self-aligned silicon-gate technology, it achieved notable area reductions compared to Al-gate PMOS SRAMs such as TI's 4003 and Fairchild's 3530 [224]. Leveraging this technology, Intel introduced the 1103 in 1970—the first widely commercialized DRAM [225, 226]. Using a 10- μm design rule, the cell

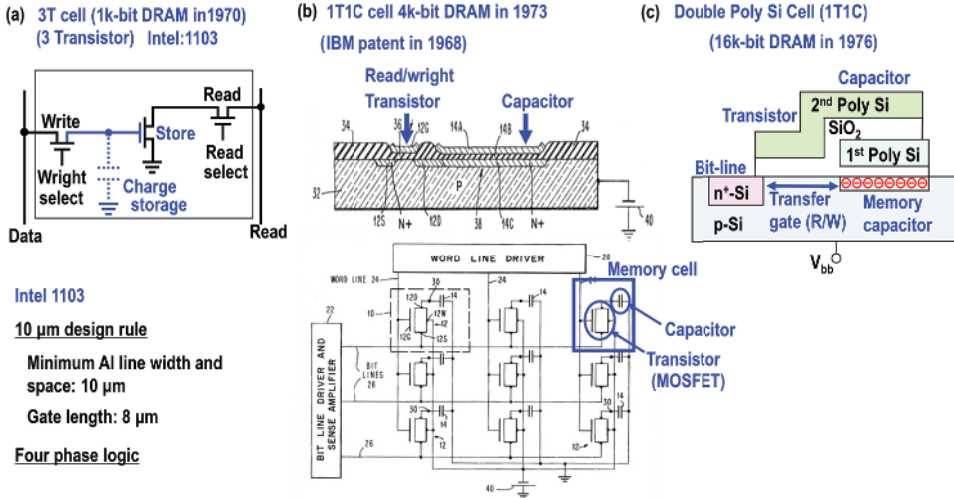


Figure 1.23 Evolution of the DRAM cell: 3-transistor (3T) cell. (a) 3-transistor (3T) cell, Intel 1103 (1-kbit DRAM) [225, 226]. (b) 1T1C cell invented by Dennard of IBM [227]. (c) Introduction of the double poly-Si gate [277]. Source: (a) Adapted from [226].

structure consisted of three p-MOSFETs (write, store, and read), as shown in Figure 1.23a. Charge stored in the insulated p⁺ poly-Si gate decayed due to recombination, requiring periodic refresh and thus classified as a DRAM.

Until around 1990, memory, especially DRAM, was the main driver of MOS IC technology. The most advanced MOSFETs were first developed for memory devices, and later adopted for logic LSIs. Only from the mid-1990s did logic LSIs become the leading driver, with fine-featured MOSFETs tailored for logic.

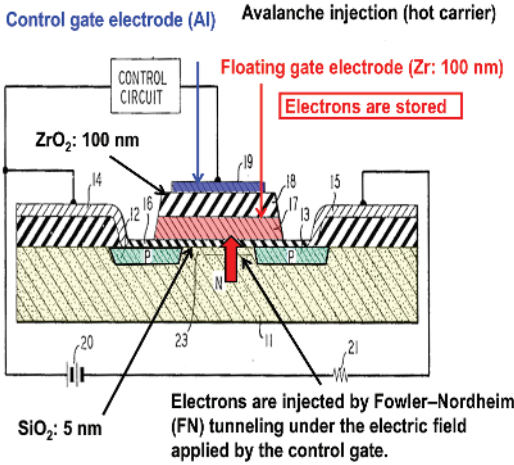
Intel's 4004 microprocessor, developed in 1971, was part of the 4000 series LSI and used in Busi-com's 141-PF calculator [228–231]. The series included the 2048-bit ROM (4001), 320-bit DRAM (4002), 10-bit static shift register (4003), and the 4004 processor, which executed instructions stored in the 4001 ROM. Early MOS LSI microprocessors couldn't match mainframe performance, which relied on BJTs in emitter-coupled logic (ECL) circuits [232, 233]. Still, the compactness and low power of the MOSFET made it suitable for broader applications. In 1973, Toshiba developed a 12-bit microprocessor using 8 μ m Al-gate PMOS LSI for a prototype engine control unit commissioned by Ford [234].

1.6.3 Nonvolatile Memory LSI

The concept of floating gate nonvolatile memory was demonstrated in 1967 by Dawon Kahng and Simon M. Sze at Bell Labs, with Kahng filing a patent as shown in Figure 1.24a [235, 236]. Using Fowler-Nordheim (FN) tunneling [237], carriers were injected into a floating gate through SiO₂ under an electric field. However, at that time, its charge retention was limited, with only the non-leaky units exhibiting a holding time of “longer than one hour.”

In 1970, Dov Frohman Bentchkowsky of Intel discovered avalanche hot-carrier injection from the source/drain regions into the gate, altering the threshold voltage and leading to the floating-gate avalanche-injection MOS (FAMOS) structure, as shown in Figure 1.24b [238]. Toshiba also confirmed similar effects [239, 240]. In 1971, Intel released the first commercial

(a) Floating gate Transistor : First non-volatile memory
Bel Labs 1967



(b) FAMOS (Floating gate Avalanche injection MOS)
Intel 1970

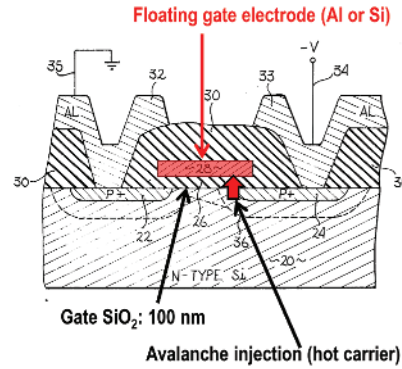
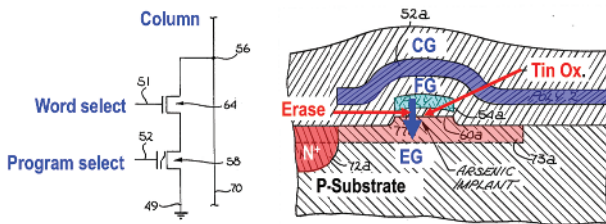


Figure 1.24 Inventions of nonvolatile memories based on floating-gate FN tunneling (1967) [235, 236] and hot-carrier injection (1970) Frohman-Bentchkowsky [238].

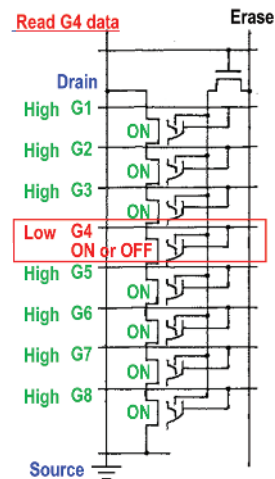
erasable programmable read-only memory (EPROM), the Intel 1702, a 2 kbit device using a poly-Si floating gate [241–243]. In 1972, Toshiba’s Hisakazu Iizuka and Fujio Masuoka developed stacked-gate avalanche-injection MOS (SAMOS) with a double poly-Si structure [244].

Early EPROMs lacked electrical erase capability and required UV light to remove stored charge. Devices were packaged with quartz windows transparent to UV. In 1978, Frohman-Bentchkowsky and colleagues invented the floating-gate tunnel oxide (FLOTOX) cell [245]. It employed a locally thinned oxide layer between the floating gate and the *n*-type bit line, enabling electrical programming and erase through FN tunneling, as illustrated in Figure 1.25a. Presented in 1980 as a “16 Kb

(a) Electrically erasable programmable ROM: Intel 1970



(c) NAND flash memory: Toshiba 1987



(b) NOR flash memory: Toshiba 1980

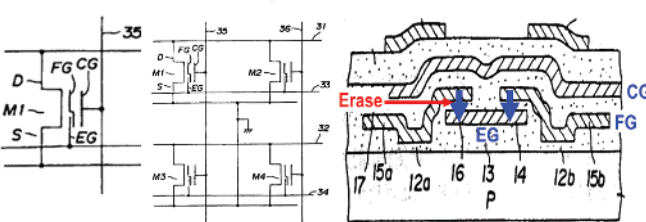


Figure 1.25 Inventions of nonvolatile memories: EEPROM (1978) [245], NOR flash memory (1980) [246, 247], and NAND flash memory (1987) [248].

Electrically Erasable Nonvolatile Memory” [249], it allowed word-level or per-cell erasure using a select transistor, though this increased cell area.

In 1980, Masuoka invented NOR flash memory [246, 247, 250], using three poly-Si layers: Erase Gate (EG), Floating Gate (FG), and Control Gate (CG) as illustrated in Figure 1.25b. Programming was done by hot-electron injection, while erasure used FN tunneling between EG and FG. EG was shared across cells, enabling simultaneous erase and reducing cell size. In 1985, Masuoka presented a 256 k-bit flash electrically erasable programmable read-only memory (EEPROM) [251].

To further improve density, Masuoka introduced NAND flash in 1987 [248, 252], as illustrated in Figure 1.25c. Unlike NOR, which placed one cell between source and drain, NAND arranged multiple gates (e.g., eight) in series. Programming was carried out by applying a high voltage to the selected gate, whereas readout was accomplished by applying a high voltage to the unselected gates to activate the channels, with the selected gate determining the current flow. This structure greatly increased storage capacity. In 1989, Masuoka presented a 4 M-bit NAND flash memory [253].

1.6.4 Supply Voltage of PMOS LSIs

In the first generation of PMOS LSIs, two or three separate power supply voltages were typically required. These voltages were relatively high because a large negative threshold voltage of p-channel MOSFETs was induced by positive interfacial charges at the Si/SiO₂ interface. In PMOS LSIs with gate lengths of approximately 8–10 μm, the typical drain voltage was designed to be about –15 V when the source voltage was 0 V. However, to ensure compatibility with the voltage standards of transistor–transistor logic (TTL) bipolar logic devices [254]—which required I/O signal levels of 0–5 V—the source voltage of the PMOSs had to be shifted by +5 V. As a result, the drain voltage was set to approximately –10 V (calculated as –15 V – (–5 V)).

For example, the Intel 1101 (256-bit SRAM) chip employed three supply voltages: $V_D = -10$ V for the peripheral circuits, $V_{DD} = -7$ V for the memory cell, and $V_{CC} = +5$ V for I/O, in addition to the 0 V ground lines. The separate provision of V_{DD} for the memory cell and V_D for the peripheral circuits was intended to reduce standby power consumption. When the SRAM chip was not being accessed, V_D for the peripheral circuits was switched off to save power, while V_{DD} remained on to retain the stored data. In the later 1101A version, V_D and V_{DD} were both set to –9 V.

In the 4004 microprocessor, the drain voltage V_D was set to –10 V and the supply voltage V_{CC} to +5 V. In contrast, the 1 k-bit DRAM 1103 had I/O levels incompatible with TTL: the high and low levels were specified as 15 and 1.8 V, respectively, with corresponding supply voltages of 16 and 19 V. For the EPROM 1702, $V_D = -9$ V, $V_{CC} = +5$ V, and the programming voltage V_G was –50 V.

1.7 Second-Generation LSI: From the Dawn of NMOS LSI to the Mid-1970s

At the interface between silicon and silicon dioxide (Si/SiO₂), positive charges arising from silicon dangling bond defects were already known to exist. These interface charges posed a serious challenge for early n-MOSFETs, as they caused the threshold voltage to become negative. Consequently, the MOSFET could not be turned off at a gate voltage of 0 V, undermining its basic switching function. This issue was addressed by doping the channel region with p-type impurities to increase the acceptor concentration, thereby shifting the threshold voltage toward a positive value. However, the diffusion techniques available at the time made it difficult to precisely control the doping concentration, often resulting in excessive levels. In contrast, p-MOSFETs naturally exhibited negative

threshold voltages due to the same interface charges, allowing them to turn off at 0 V gate voltage bias easily. This characteristic was a key reason why PMOS LSI technology was adopted during the first generation. Nevertheless, the carrier mobility p-MOSFETs was approximately half of that of their n-channel counterparts, prompting many companies to initiate the development of NMOS technology for the 6 to 5 μm nodes during the second generation.

Around 1970, several equipment manufacturers began commercializing ion-implantation systems. Semiconductor companies quickly adopted these systems to control the threshold voltage of n-MOSFETs [255–257]. Ion implantation offered a significant advantage: the number of implanted ions could be accurately measured as electrical current, enabling precise control of the dose. This made it possible to adjust the threshold voltage to a low positive value compatible with a 5 V power supply, thereby allowing NMOS SRAM and logic LSIs to operate reliably at that voltage. IBM, Hitachi, and NEC were among the companies that began developing NMOS ICs and LSIs in the mid-1960s [181, 258–261].

1.7.1 NMOS LSI Products as Second-Generation Devices

W. Krolikowski and his colleagues, who left IBM to establish Cogar Corporation, presented in 1969 the first commercial 1 k-bit aluminum-gate NMOS SRAM at a technical conference [15, 262]. However, Cogar's SRAM business eventually failed due to production delays caused by insufficient manufacturing facilities and a marketing strategy that focused on selling complete systems rather than individual chips [15, 263]. In 1971–1972, NEC introduced two 1 k-bit NMOS DRAMs, the μPD403 and μPD404 [20], but they were incompatible with the Intel 1103 PMOS DRAM power standards and saw little adoption. From 1972 to 1975, many firms produced second-generation LSIs with 5–6 μm NMOS, including custom logic LSIs for calculators and Intel's 2101 SRAM operating with a single +5 V supply.

Intel's NMOS 8080 microprocessor [264] and 4 k-bit NMOS DRAM “2107” (1972) [265, 266] still required multiple voltages ($V_{DD} = 12\text{ V}$, $V_{CC} = 5\text{ V}$, $V_{BB} = -5\text{ V}$), while “2107” used a three-MOSFET cell. In 1973, MOSTEK's 4 k-bit “MK4096” [267, 268] introduced address multiplexing (reducing pins from 22 to 16) and the 1T–1C cell, first proposed by IBM's R. Dennard in 1967 and illustrated in Figure 1.23b [227, 269–271]. The 1T–1C design soon became the DRAM standard [272–276]. MOSTEK's 16 k-bit “MK4116” DRAM (1977) adopted a stacked double poly-Si structure for both capacitor and transfer gate, as shown in Figure 1.23c [277], while retaining the same triple-voltage supply as the 4 k generation. Reviews of DRAM evolution during this era are given in [278–280]. The shift to a single +5 V supply occurred with the 64 k-bit DRAM generation in 1980.

1.7.2 A Case Study of NMOS LSI Technology Developed in the 1970s

In the 1970s, the structural complexity of LSI devices was relatively modest compared to today's advanced nano-CMOS technologies. Nevertheless, manufacturers frequently faced unexpected challenges, and production yields were often limited. This section examines such issues through a published, documented example of early NMOS LSI development at Toshiba [27]. This NMOS technology was first applied to 1k-bit SRAMs (TMM311P, TMM312P, and TMM313P) [281]. It was later used in a 12-bit microprocessor (T3517A) for engine control of the 1978 Ford Lincoln Versailles, meeting the Clean Air Act Amendments of 1970 [282]. By controlling engine timing, it reduced CO, HC, and NO_x emissions, marking the first microprocessor-based engine control in a commercial vehicle. The technology also powered the TOSBAC40L, a one-board 16-bit minicomputer.

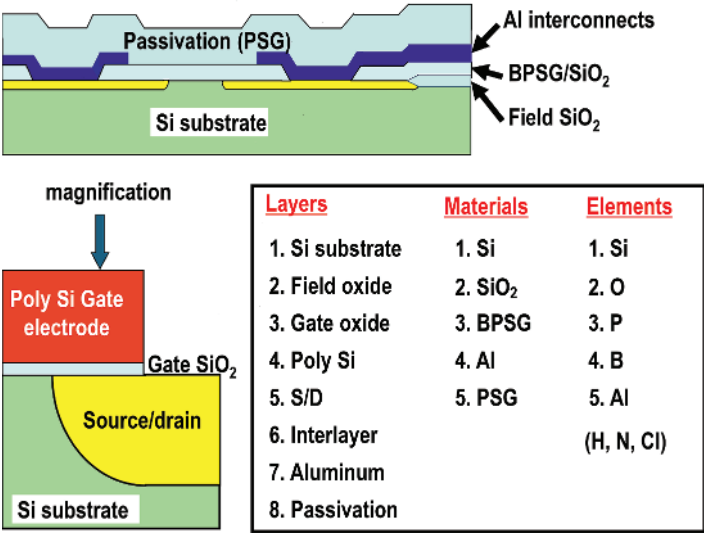


Figure 1.26 Cross section of the first Toshiba NMOS LSI.

Prior to this, Toshiba had fabricated a 256-bit SRAM using 8 μm Al-gate PMOS LSI technology. In contrast, the 1-kbit NMOS SRAM adopted silicon-gate technology, marking a significant advancement in process integration. As shown in Figure 1.26, Toshiba's first NMOS LSI, made with a 6 μm process, had eight layers, including the Si substrate and a PSG passivation above Al interconnects. It used five materials—Si, SiO₂, borophosphosilicate glass (BPSG), Al, and poly-Si—corresponding to five elements. Fabrication required seven mask steps: active area, D-type channel implantation, poly-Si/Si contact, poly-Si gate, contact holes, Al interconnects, and bonding pads. Boron implantation was done across the region without a mask.

At the circuit level, most companies employed enhancement-depletion (ED) logic, consisting of enhancement-mode NMOS drivers (with positive threshold voltage) and depletion-mode NMOS loads (with negative threshold voltage) [259, 260]. Figure 1.27 shows photographs of Toshiba's 1-kbit SRAMs TMM311P and 312P [281]. TMM311P (DIP-18) and TMM312P (DIP-16) were

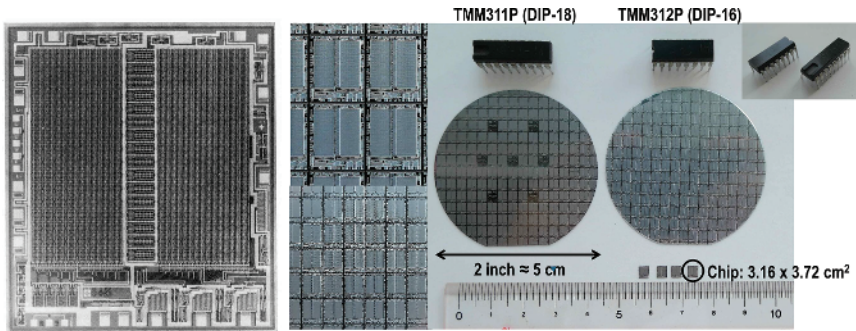


Figure 1.27 Photographs of Toshiba's 1-kbit SRAMs: TMM311P and TMM312 Masuoka [281] 6 μm E/R NMOS technology (Si gate: 4T + R cell) TMM311P (DIP-18), 312P (DIP-16):256 $\times$ 4 bit gate length: 6 μm , gate SiO₂ thickness: 100 nm, source/drain junction depth: 1.7 $\sim$ 1.8 μm (ρ_s = 6 ohm/sq), $V_{th}(E)$ = 0.9 V $\pm$ 0.2 V, $V_{th}(D)$ = -3.0 V $\pm$ 0.6 V, $V_{th}(\text{Field})$ = 20 V, ρ_s (poly Si) = 20 $\sim$ 40 ohm/sq, cell size: 66 $\times$ 80 μm^2 , chip size: 3.16 $\times$ 3.72 mm², access time: 300 ns, cycle time: 600 ns, power supply: 5 V, and power dissipation: 250 mW.

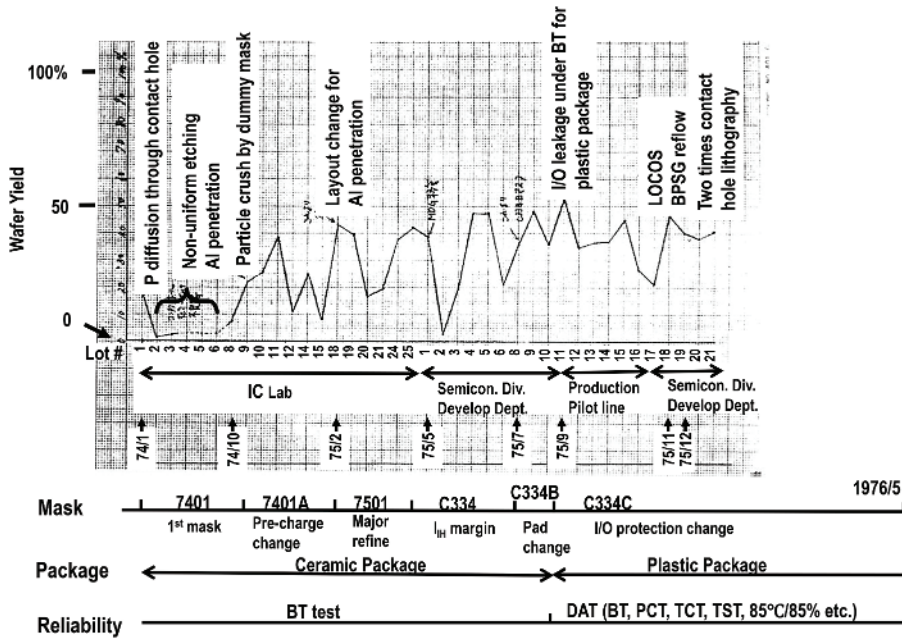


Figure 1.28 NMOS 1-kbit SRAM wafer yield trend.

configured as 256×4 -bit devices. Another SRAM, the TMM313P (DIP-18), with a configuration of 1024×1 bit, was produced using the same technology. These memories were employed in microcomputers such as the TLCS (Toshiba LSI Computer System) series, offering high-speed access times. Each memory cell used four MOSFETs and two polysilicon resistors, instead of six MOSFETs, to reduce cell size. The memory latch consisted of two ER inverter pairs (E for enhancement-type MOSFET and R for resistor load; see Figure 1.45 in Section 1.9). Production was carried out on 2-inch wafers using contact-exposure lithography. Later, mass production was achieved by shrinking the chip layout by a factor of 0.9.

Across the industry, wafer yields typically ranged from 20% to 40%. At Toshiba, profitability was achievable when wafer yields reached 40% (Figure 1.28) and assembly yields exceeded 70%. Several yield-limiting issues were identified and mitigated through targeted process improvements, as discussed in the following sections.

1.7.2.1 Al Penetration (or Spike)

In the LSI process, after Al interconnect formation, annealing at $450\text{--}500^\circ\text{C}$ for about 30 minutes in forming gas ($\text{N}_2:\text{H}_2 = 10:1$) was required. This step, referred to as “sintering,” played an essential role in MOS LSIs by passivating Si/SiO₂ interface defects with hydrogen, thereby reducing interface charge and restoring device characteristics. Since Al deposition generated X-rays that broke Si bonds and induced charges, leading to threshold-voltage shifts, sintering was indispensable.

However, in 6- μm -generation NMOS LSIs, failures frequently appeared after sintering as drain-to-ground shorts, as shown in Figure 1.29a. Analysis revealed that during sintering, Si diffused into Al, forming pyramid-shaped pits at contact holes (Figure 1.29c), and the Si diffusion was confirmed by Auger spectroscopy (Figure 1.29b) [283]. In contrast, in 8- μm -generation PMOS LSIs, the problem was relatively minor because the drain junctions were deeper. As countermeasures, as shown in Figure 1.29d, approximately 1% Si was first alloyed into Al to



Figure 1.29 Al penetration (spike) [283] and solutions [284–288].



Figure 1.30 Some interim solutions to prevent AI penetration [289–291].

pre-saturate it [284, 285], followed by the addition of 0.5% Cu to suppress electromigration [286]. Finally, the introduction of a Ti/TiN barrier beneath the Al interconnects effectively blocked Si diffusion and eliminated the problem [287, 288].

Figure 1.30 shows some interim solutions without the introduction of new materials. On the left, one approach was to diffuse phosphorus from the contact holes in order to deepen the pn junction only beneath the contact region [289]. In the upper right, another method involved modifying the Al interconnect pattern to suppress Si diffusion [290]. In the lower right, yet another approach was to supply Si from dummy poly-Si patterns, thereby raising the Si concentration in the Al interconnects close to saturation [291]. These measures were actually implemented in the production of Toshiba's first NMOS LSIs and proved effective.

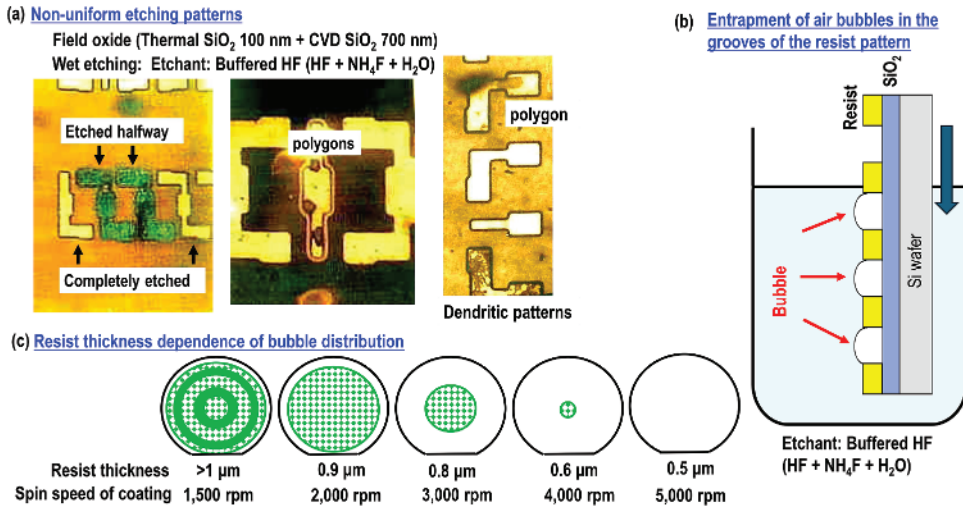


Figure 1.31 Nonuniform etching of field oxide due to entrapped air bubbles.

1.7.2.2 Nonuniform Etching of the Field SiO₂ Film Due to Entrapped Air Bubbles

Initially, device isolation in Toshiba's NMOS LSIs was carried out, as was commonly practiced at the time, by depositing a thick CVD field oxide on the Si substrate, patterning it by photolithography, and leaving a thick field SiO₂ film in the isolation regions (see Figure 1.33 in Section 1.7.2.4). However, as shown in the photographs in Figure 1.31a, during the etching process, some patterned SiO₂ films were not etched, or, even when etched, polygonal or dendritic SiO₂ residues sometimes remained. The cause remained unclear until light irradiation revealed sparkling regions, which were traced to bubbles adhering to the wafer surface (Figure 1.31b). Nonuniform etching occurred in these regions of the wafer. Most bubbles formed when wafers were immersed in buffered HF (BHF; H₂O:NH₄F:HF = 81:12:7), though some may have originated from reaction gases. Bubble adhesion depended on resist thickness, as shown in Figure 1.31c. Thicker films trapped more bubbles, whereas excessive thinning led to resist defects. The issue was resolved by either rotating the wafers during etching or pre-immersing them in a surfactant solution. Furthermore, polygonal or dendritic residues were associated with NH₄F crystal growth during etching, often in the same regions as bubbles, suggesting a correlation between crystal formation and bubble adhesion.

1.7.2.3 Al Interconnect Thinning at Topographical Steps and Solution by BPSG Reflow

In the 1 k-bit SRAM, the Al film was 1 μm thick. Although no disconnections occurred at the step edges of the interlayer dielectric, wedge-shaped thinning and cracks were observed, as shown in Figure 1.32a, raising concerns about reliability. Planarization by PSG reflow had been patented in 1974 by Gordon Moore of Intel [218], but it required a phosphorus concentration above $2 \times 10^{21}/\text{cm}^3$ (Figure 1.32d). Such a high phosphorus content made the film hygroscopic, leading to moisture absorption and corrosion of Al interconnects; therefore, PSG reflow was not adopted.

For BPSG films, the boron concentration also exceeded $2 \times 10^{21}/\text{cm}^3$, but the phosphorus level could be reduced by an order of magnitude (Figure 1.32c), thereby avoiding moisture absorption [292]. Consequently, beginning in 1975, the BPSG reflow method was introduced and optimized for 1 k-bit SRAM fabrication. As shown in Figure 1.32b, a very flat surface over step edges was achieved. Toshiba was the first, or among the first, to apply BPSG reflow in mass production.

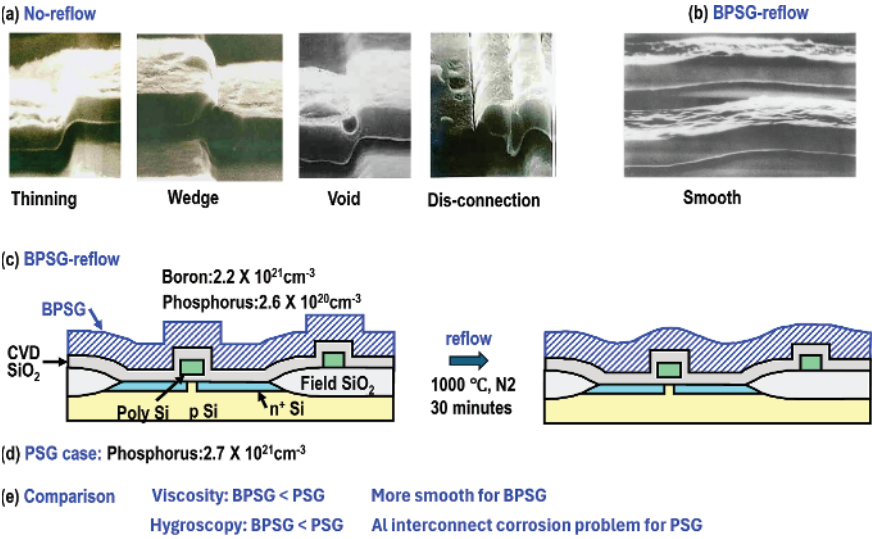


Figure 1.32 Al interconnect thinning at topographical steps and solution by BPSG reflow.

1.7.2.4 Field Leakage by Conventional Isolation and Its Solution by the Coplanar (LOCOS)
Method

During the PMOS LSI era, device isolation was realized through the formation of thick CVD SiO_2 layers between active devices, and the same method was used in NMOS 1 k-bit SRAMs until mid-1975 as illustrated in Figure 1.33a. At that time, chips were packaged in ceramics, but when plastic packages were introduced, leakage failures appeared at I/O regions after bias temperature

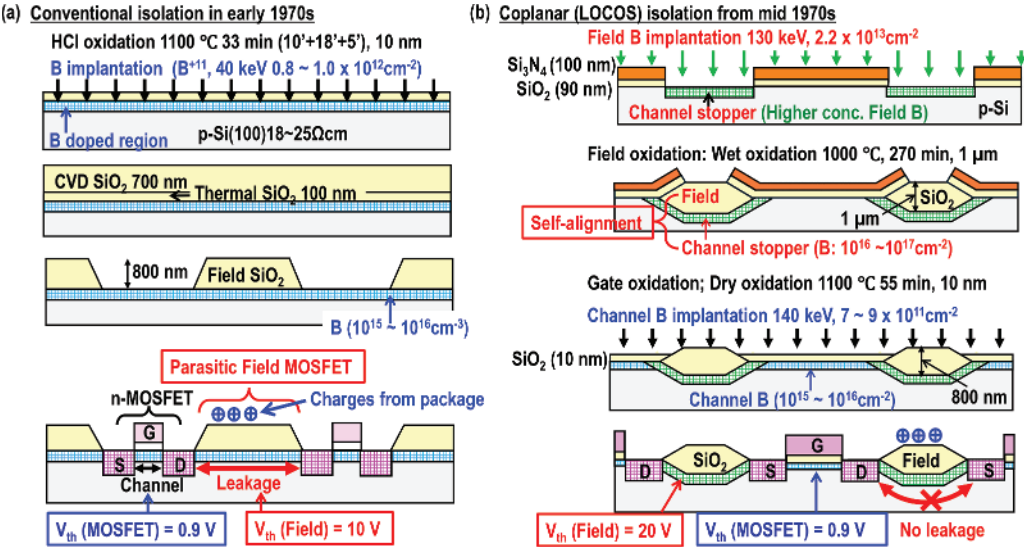


Figure 1.33 Field leakage by conventional isolation and its solution by the coplanar (LOCOS) method [293–297].

(BT) testing. These failures were caused by ions in the molding resin migrating to the field oxide surface, turning on parasitic MOSFETs where the field oxide acted as gate oxide. To prevent this, heavy boron doping beneath the field oxide was required to form a channel stopper. (The MOSFET channel itself was protected by the gate electrode.) For higher integration, self-alignment between the channel stopper and the field oxide was essential. Thus, from the fall of 1975, the coplanar [296, 297] or LOCOS (Local Oxidation of Silicon) method, invented by E. Kooi of Philips [293–295], was adopted, effectively solving the problem, as illustrated in Figure 1.33b.

1.7.2.5 Lithography Pattern Failure Caused by CVD Particle Contamination

During this period, after the elimination of initial defects, the main cause of chip yield loss was the presence of CVD particles generated during film deposition, as shown in Figure 1.34c. The presence of these particles on the wafer surface itself often led to circuit malfunctions. Moreover, because contact mask lithography was used at that time, the CVD particles were crushed when the mask contacted the wafer, resulting in pattern failures, as shown in Figure 1.34a, and mask contamination, as shown in Figure 1.34b. The most effective way to address this problem was to frequently clean the CVD deposition equipment to suppress particle generation. However, there was a trade-off between equipment uptime and yield. Since a yield of approximately 40% was considered acceptable at the time, rigorous control of CVD particle contamination was not yet implemented. This issue became a serious concern later, and strict contamination control was finally established during the VLSI era of the 1980s.

To suppress CVD particle problems, a proximity exposure system was temporarily introduced as shown in Figure 1.35b. Increasing the mask–wafer gap degraded resolution, while reducing it failed to prevent particle crushing, so the issue remained unsolved. A full innovative solution came in 1973 with PerkinElmer’s 1:1 Micralign Projection Mask Alignment System (Figure 1.35c), widely adopted worldwide from the late 1970s, which used mirrors (reflecting optics) instead of lenses (refractive optics).

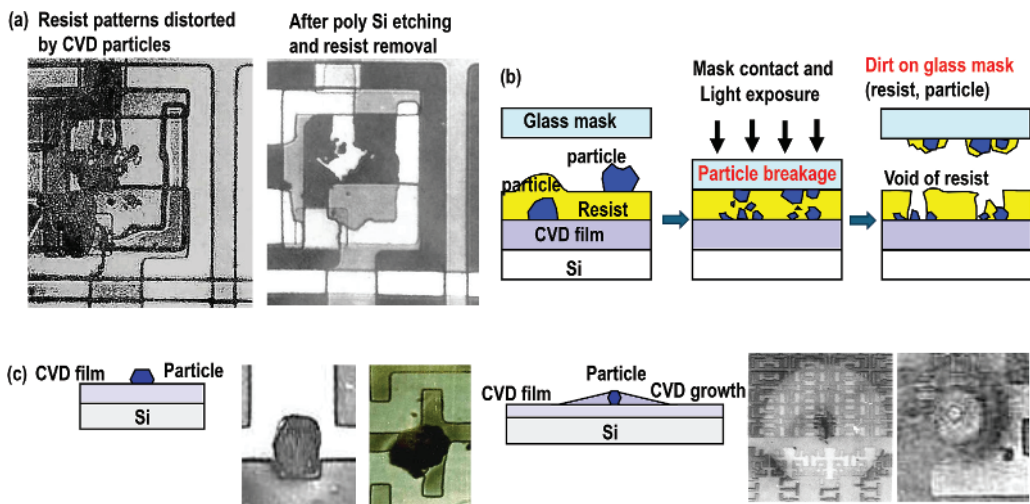


Figure 1.34 Lithography pattern failure caused by CVD particle contamination.

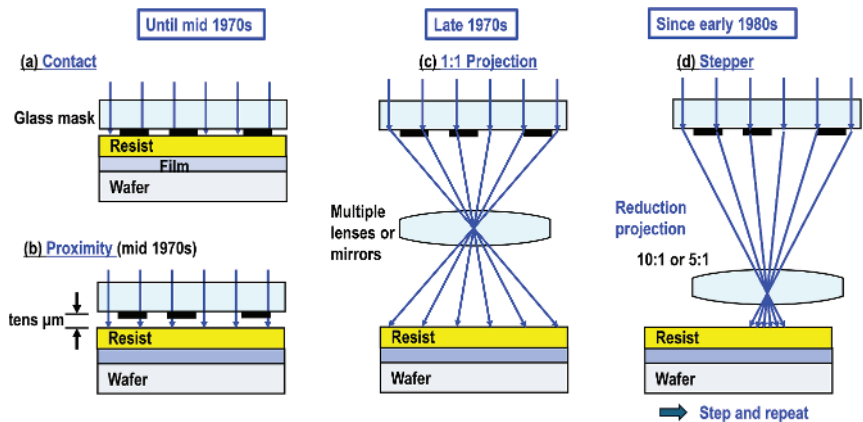


Figure 1.35 Evolution of mask aligner used in advanced MOS LSI production.

1.8 First Generation of VLSI (3 μm NMOS Technology: Fourth Generation of LSI) Late 1970s to Early 1980s: Emergence of Dry Processing and Stepper Lithography

By the early 1980s, the number of MOSFETs integrated on a single chip approached 100,000, heralding the dawn of the VLSI era—corresponding to the fourth generation of LSI. Although NMOS technology remained dominant, design rules had been scaled down to 3 μm , necessitating the adoption of advanced process technologies to achieve such miniaturization. The key innovations included stepper lithography and the introduction of various so-called dry processes that replaced conventional vapor diffusion and liquid etching techniques. Such dry processes include RIE, source/drain ion implantation, and resist ashing for photoresist removal.

Until the mid-1990s, DRAM continued to serve as the principal technology driver for many companies, particularly in Japan and Korea. Logic devices often inherited process innovations originally developed for DRAM. Figures 1.36 and 1.37 illustrate Toshiba’s DRAM development

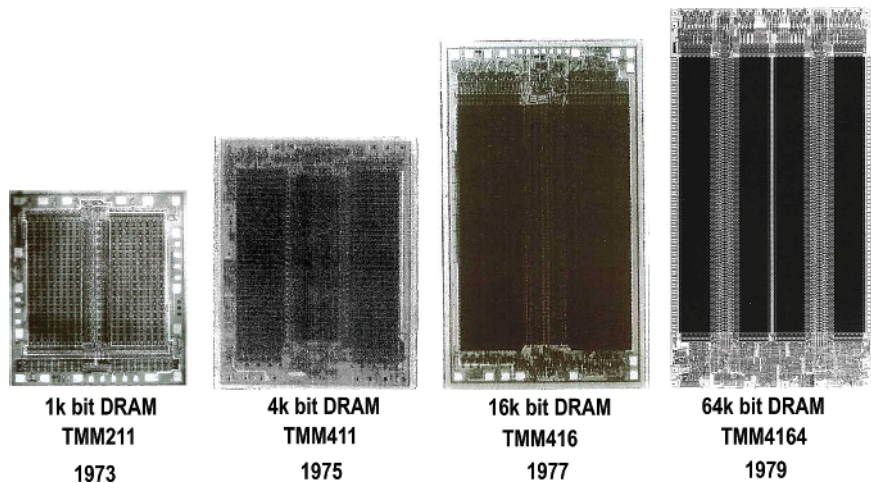


Figure 1.36 Toshiba DRAM chip trend in the 1970s.

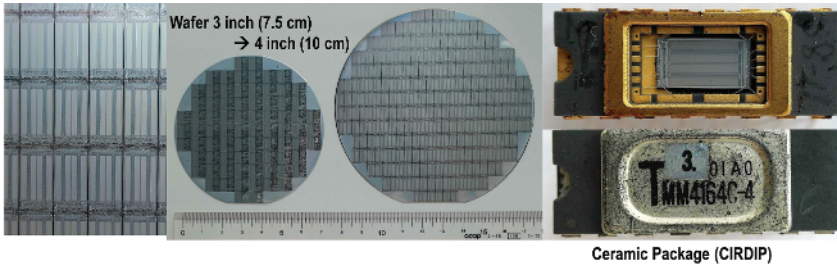


Figure 1.37 Toshiba 64-kbit DRAM (TMM4164) [298]. Gate length: 3 μm ; first gate SiO_2 thickness: 40 nm, the MOSFETs (threshold voltage, V_{th} , of 0.6 ± 0.15 V) in the peripheral circuits and for the MOS capacitors in the memory cells; second gate SiO_2 thickness: 75 nm, applied only to the transfer-gate MOSFETs in the memory cells ($V_{th} = 0.9 \pm 0.15$ V); source/drain junction depth: 0.4 μm (sheet resistance $\rho_s < 40$ $\Omega/\text{sq.}$); field threshold voltage: $V_{th} > 25$ V; poly S: $\rho_s < 60$ $\Omega/\text{sq.}$; memory configuration: 64k $\times$ 1 bit; memory cell size: 8×19 μm^2 ; chip size: 3.66×6.80 mm^2 ; operating temperature range: 0–70°C; power supply voltage: 5 V; RAS access time: 150 ns; CAS access time: 100 ns; cycle time: 320 ns; average operating current: maximum 50 mA; standby current: maximum 4 mA; refresh cycle: 128 cycles per 2 ms; power dissipation: 275 mW; package: 16-pin DIP. The wafer diameter: initially 3 inches, later 4 inches.

trajectory from 1973 to 1980, spanning capacities from 1 k to 64 k bits [298]. As shown in Figure 1.36, DRAM and SRAM capacities quadrupled approximately every two to three years, while design rules shrank by a factor of ~ 0.7 combined with increased chip size—a tacit industry consensus. The term “Moore’s law” [299] was not widely recognized at the time and only gained broader usage in the 1990s. In the absence of a unified roadmap, each company formulated its own strategic development plans. Toshiba’s 64-kbit DRAM (TMM4164) incorporated complex peripheral circuits, resulting in an increased transistor count, and the chip size was set—based on the initial package design—to the second smallest in the world. Consequently, the design left very little margin for lithography, including mask alignment tolerances and minimum linewidths. As a result, Toshiba became one of the first companies to introduce steppers to enhance lithographic resolution and to implement dry processes that suppressed side etching and lateral diffusion in a commercial product.

1.8.1 Introduction of Dry Processes and the Emergence of Plasma-Induced Damage (PID)

To achieve high-density integration, minimizing lateral etching of structural features and lateral dopant diffusion was imperative. As illustrated in Figure 1.38, the introduction of RIE and source/drain ion implantation made it possible to integrate numerous circuit elements within compact chip areas. As shown in Figure 1.38a, conventional isotropic wet etching caused side etching, widening the spacing between lines and limiting integration density. This issue was addressed by anisotropic etching using RIE, as illustrated in Figure 1.38a. To suppress lateral diffusion of impurities, source/drain doping was changed from conventional POCl_3 vapor-phase diffusion to high-concentration arsenic ion implantation, as illustrated in Figure 1.38b. The ion implantation was performed using Lintott’s High-Current Beam Ion Implanter. These techniques collectively became known as “dry processes,” in contrast to traditional “wet processes” that relied on liquid-phase etching and vapor-phase diffusion.

Early adoption of dry processes presented several challenges. Ion bombardment often hardened the resist layer, and an additional problem arose in that fluorocarbon residues generated during RIE

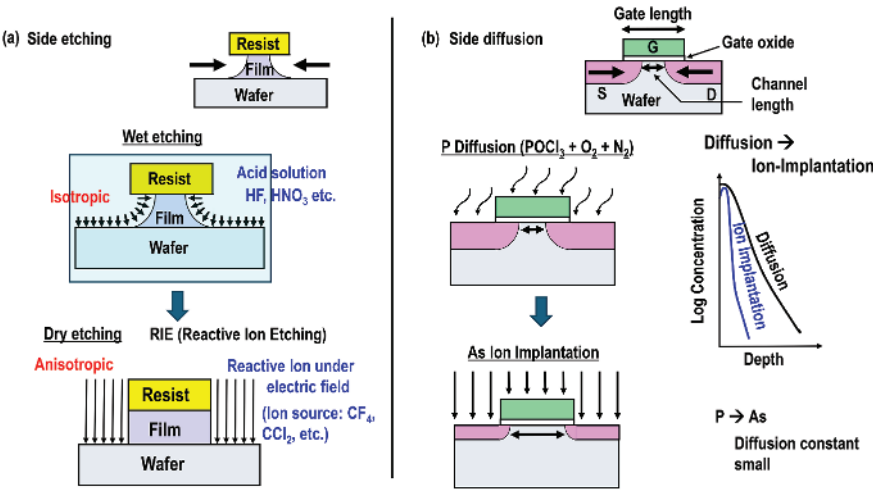


Figure 1.38 Introduction of dry processes (RIE and S/D ion implantation) to suppress side etching and lateral diffusion.

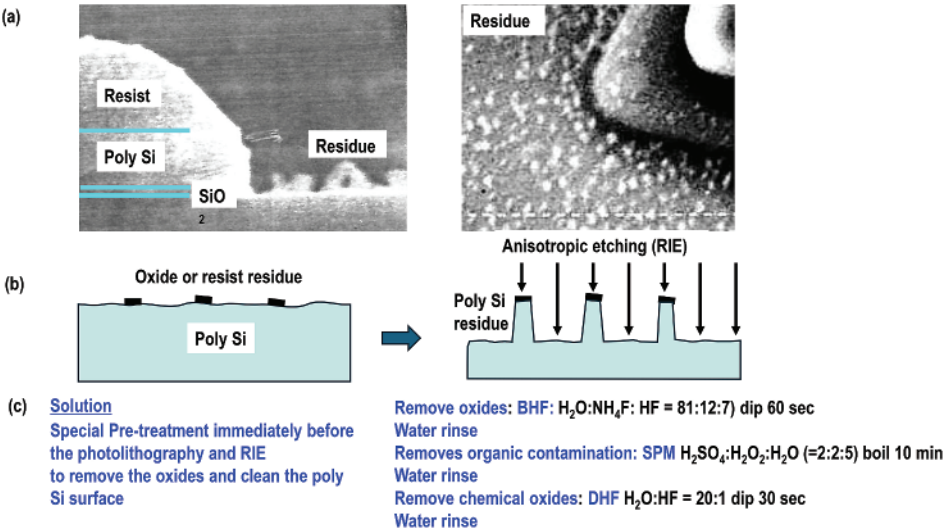


Figure 1.39 Poly-Si gate electrode RIE and residue.

could not be removed in the resist removal process. This issue persisted even when resist ashing, a dry process employing oxygen plasma, was performed. Additionally, residuals from poly-Si RIE, as shown in Figure 1.39, posed further complications. In gate poly-Si RIE, etching failures from columnar residues, as evident from the images in Figure 1.39a, occurred frequently, forcing wafer discard and hindering DRAM development. The cause was RIE anisotropy. Surface oxide or other residues masked the underlying poly-Si and prevented its etching, as shown in Figure 1.39b. To solve this, besides adjusting anisotropy, an additional sequence was added before poly-Si gate lithography and RIE, as shown in Figure 1.39c. Surface oxide was removed with buffered hydrofluoric acid (BHF), organic contaminants were eliminated with a sulfuric peroxide mixture (SPM), and the thin

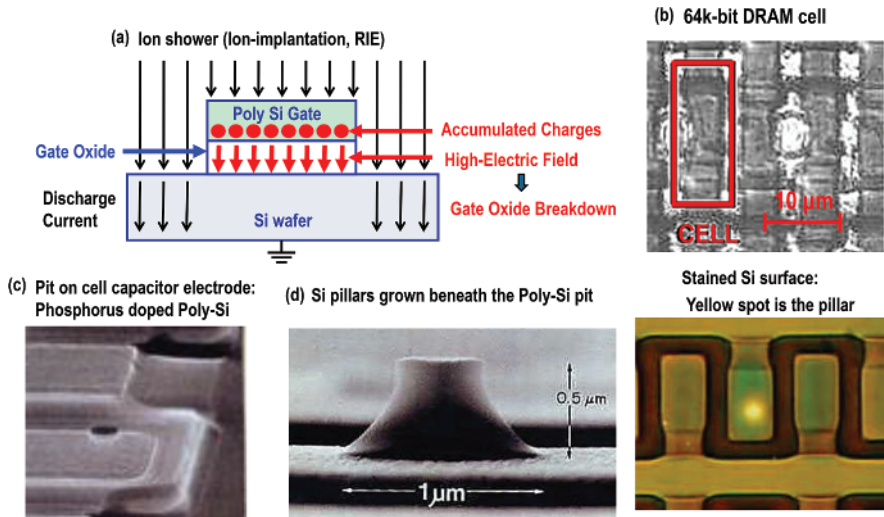


Figure 1.40 64-kbit DRAM bit error caused by plasma-induced damage (PID) [300].

chemical oxide from SPM was stripped with dilute HF (DHF). This oxide-removal and cleaning process successfully resolved the problem.

Prior to this generation, most companies employed vapor-phase diffusion of phosphorus (P) and boron (B) for source/drain doping. From this point onward, ion implantation became a standard practice. Arsenic (As), with its lower diffusion coefficient and shorter ion range than P, replaced P as the preferred *n*-type dopant for shallower junctions. Ion doses for source/drain regions were two to three orders of magnitude higher than those for channel doping, raising concerns about increased *pn* junction leakage due to crystal damage. Fortunately, these fears did not materialize.

However, an unexpected issue emerged: ion accumulation on gate electrodes during implantation and RIE-induced high voltages across the cell capacitor gate oxide, causing frequent DRAM bit errors due to gate oxide breakdown, as shown in Figure 1.40. During RIE or high-dose ion implantation, the process environment effectively formed an ion shower, as shown in Figure 1.40a. Ions accumulated on the poly-Si gate electrode, insulated by the gate oxide, raise the electric field and cause gate insulator breakdown. In a 64-kbit DRAM cell capacitor, HF treatment and staining revealed a yellow circular pattern (Figure 1.40b), corresponding to a phosphorus-rich region. Due to dielectric breakdown under a high field, Si from the heavily P-doped poly-Si gate melted and epitaxially regrew on the substrate in circular form. In SEM images, a hole appeared in the poly-Si gate (Figure 1.40c), and after removing the gate and oxide, a rounded epitaxial Si mound was visible (Figure 1.40d).

Such a breakdown caused DRAM bit errors and serious yield loss until identified as ion accumulation. Toshiba reported this effect in 1982 [300], the first disclosure of the phenomenon, later widely recognized as plasma-induced damage (PID) and a major plasma process issue.

1.8.2 Introduction of Intrinsic Gettering (IG) Wafer

During this period, wafer substrate selection became critical for DRAM fabrication. The 64-kbit DRAM was a high-density device, and wafer defect density was a concern. Some US makers adopted epitaxial wafers with low-defect layers, while Toshiba used the emerging intrinsic gettering (IG) wafers.

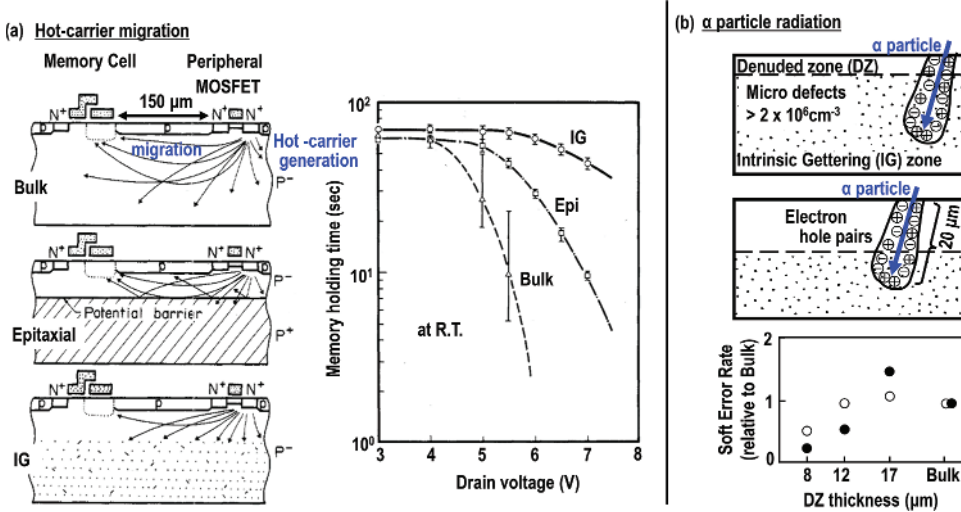


Figure 1.41 Comparison of alpha-particle and hot-carrier migration effects among IG, epitaxial, and bulk wafers [301, 302].

IG wafers were made by heat-treating oxygen-rich silicon, inducing oxygen precipitation and bulk microdefects. Oxygen out-diffusion created a nearly defect-free near-surface region (10–20 μm), called the denuded zone (DZ), as shown in Figure 1.41. A key issue in DRAM cells was reduced capacitor charge retention caused by carrier migration from the peripheral circuits, as shown in Figure 1.41a. Retention characteristics were evaluated on IG, epitaxial, and bulk wafers using carrier injection, and the results indicated that IG wafers exhibited the longest retention [301]. IG wafers also proved effective in suppressing soft errors induced by alpha particles [303–306] emitted from radioactive impurities in packaging materials, as shown in Figure 1.41b [302].

1.8.3 Stepper (Step and Repeat with 5 $\times$ and 10 $\times$ Optical Reduction of Mask Pattern Feature Size) Photolithography and “Common Defects”

To enhance resolution, stepper photolithography systems (Figure 1.35d) were introduced that projected reduced mask patterns onto narrow wafer regions. Toshiba adopted a beta version of the 10X Mann 4800 “Direct Step on Wafer” (DSW) system and employed it in the development of 64 K DRAM products, concurrently addressing early equipment problems in 1979. In practical device applications, various new techniques and know-how were accumulated, such as the need for fine adjustment of the step pitch due to slight changes in wafer warpage before and after thick oxidation. The most critical challenge was mask defects, as shown in Figure 1.42. At the time, approximately 50% of masks contained at least one defect. In 1:1 projection systems, a defect would affect only one chip (Figure 1.42a). However, in stepper systems—where only one or two chips were exposed at a time (Figure 1.42c)—a single defect could compromise all chips or half of them on a wafer (Figure 1.42b). Because half of the masks then showed at least one defect, each required microscopic inspection and repair, an extremely labor-intensive task that significantly burdened the engineers developing the 64 k-bit DRAM. It took about two years to automate defect inspection and improve mask yields. These insights were later incorporated into the use of Nikon’s stepper systems during the continuous development of the 64 K-bit DRAM in 1981.

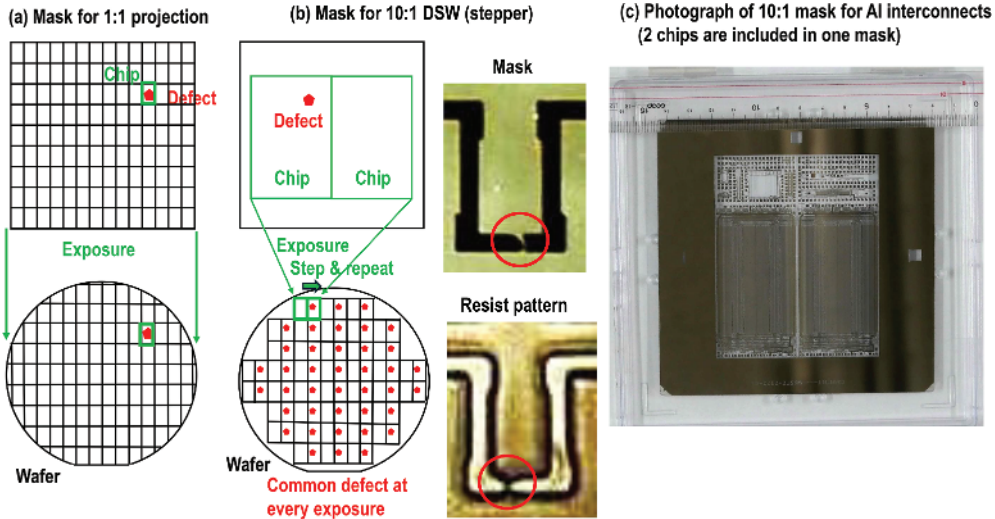


Figure 1.42 Common defect problem for DSW (stepper) mask.

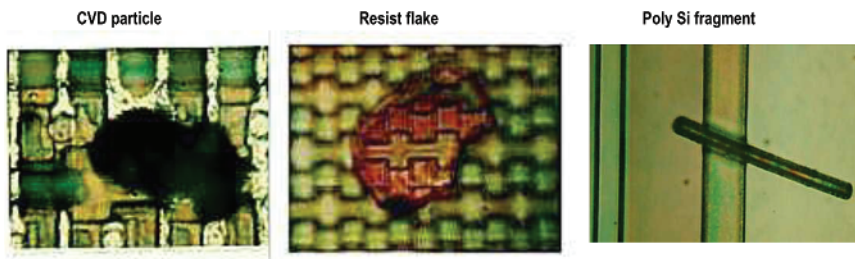


Figure 1.43 Typical dust contamination in 1980.

1.8.4 Dust Control

With the advent of the VLSI era, device dimensions became smaller and circuit complexity increased, making yield loss due to dust contamination a critical issue. Around 1980, semiconductor manufacturers conducted thorough investigations of dust sources in both equipment and cleanrooms, and devoted intensive efforts to rigorously suppressing their generation. As a result, yield loss attributable to dust was eventually reduced to a satisfactory level. The images in Figure 1.43 show typical examples of dust contamination observed at Toshiba.

1.9 Transition from NMOS to CMOS in the Mid-1980s

Frank M. Wanless and C. T. Sah of Fairchild Semiconductor introduced CMOS circuits at ISSCC in 1963 [307], and Wanless filed a related patent later that year (Figure 1.44) [308]. CMOS circuits conduct current only briefly during switching, resulting in very low power consumption (Figure 1.44b). Another major benefit is their steep input–output voltage transfer curve (Figure 1.45) [309], which distinguishes CMOS from logic families like ER, EE [310], and ED [259, 260, 311], as defined in the figure. This sharp slope ensures wide noise margins, simplifies design, and enhances reliability. Since either the p-MOSFET or n-MOSFET is always off after switching, the output voltage (V_{out}) is

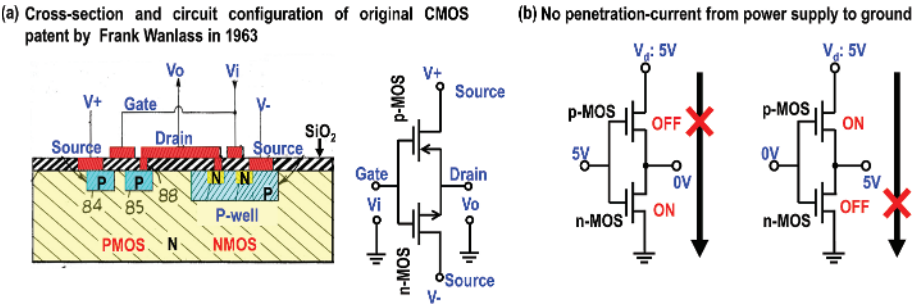


Figure 1.44 CMOS configuration and its advantage in power saving due to penetration current limited to the transient period [308].

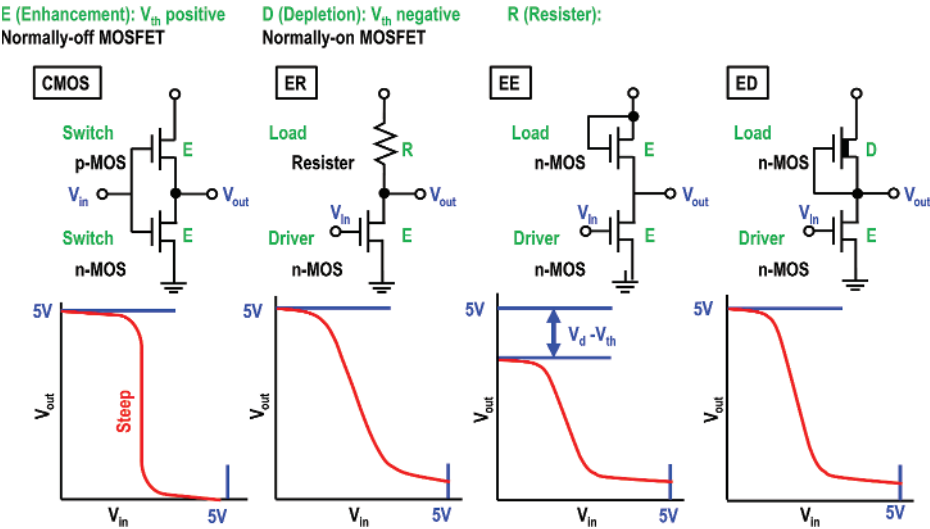


Figure 1.45 Superiority of CMOS circuit input-output characteristics over other MOS circuits [309–311].

exactly the supply voltage or 0 V (Figure 1.45). In contrast, EE inverters clamp V_{out} at $(V_d - V_{th})$, requiring capacitive bootstrap techniques to drive subsequent stages.

CMOS was first employed in low-power devices such as watches and calculators in the late 1960s [312–316], but its wider adoption was constrained by the requirement for *n*-well or *p*-well structures and by the complexity of dual-MOSFET fabrication, as illustrated in Figure 1.44a. CMOS became mainstream for LSI only in the mid-1980s, especially in SRAM and DRAM, due to its low power (especially standby) and excellent transfer characteristics. Lowered power and improved yields accelerated its widespread use. Table 1.3 lists 1-Mbit DRAM papers from ISSCC (1984–1986): seven used NMOS and eight used CMOS. CMOS DRAMs operated at 30–45 mA with 0.05–1 mA standby, while NMOS needed 50–70 mA and 2–3 mA, respectively. CMOS’s precise 5 V or 0 V outputs eliminated boosting circuits, reducing complexity, area, and capacitance, thus yielding smaller, faster chips and shorter development cycles [317]. For high-performance microprocessors, NMOS remained dominant until the mid-1980s, when rising power consumption forced a shift toward CMOS, as shown in Table 1.4. Bipolar CMOS (BiCMOS) technology [318–320] gained traction in the mid-1990s with the introduction of 0.8–0.35 μm processes and was adopted in high-performance

Table 1.3 List of 1-Mbit DRAMs published at ISSCC from 1984 to 1985.**Operation current:** NMOS 70~50 mA, **CMOS 45~30 mA** (IBM data are not included)**Standby current:** NMOS 3~2 mA, **CMOS 1~0.05 mA** (IBM data are not included)

Year	Company	Technology		Operation Current (mA)	Standby Current (mA)	Access time RAC (ns)	Cell size (μm^2)	Chip size (mm^2)
1984	NEC	NMOS	1 μm	58	3	120	5.5 x 8	9.4 x 8.07
1984	Hitachi	NMOS	NA	60	2	90	3 x 7	4.67 x 9.86
1984	NEC	NMOS	1 μm	69	2.8	85	3.4 x 6	4.6 x 9.4
1985	Mitsubishi	NMOS	NA	70	3	90	3.8 x 9.4	5 x 13
1985	Mostek	CMOS	1.2 μm	45	0.5	85	4 x 9	5.97 x 11.4
1985	Toshiba	NMOS	1.2 μm	54	3	70	5 x 6.4	4.87 x 13.24
1985	IBM	CMOS	1.3 μm	150	1	89	4.1 x 8.8	5.5 x 10.5
1985	Fujitsu	NMOS	NA	70	3	90	8.4 x 3.15	4.44 x 12.32
1985	Toshiba	CMOS	1.2 μm	30	0.2	56	3.8 x 9	5 x 12.5
1985	Hitachi	CMOS	NA	46	NA	80	3.35 x 7.2	4.68 x 10.1
1985	AT&T, Bell	CMOS	1.3 μm	30	0.1	80	3.5 x 10.5	4.8 x 14.5
1985	Intel	CMOS	1 μm	NA	0.05	65	3.8 x 7.5	5.3 x 9.85
1985	TI	CMOS	1 μm	45	0.8	70	2.5 x 8.5	4.3 x 11.7
1985	Toshiba	CMOS	1.2 μm	35	0.05	65	3.24 x 9	4.4 x 12.32
1985	Matsushita	NMOS	1.2 μm	12	NA	100	4.5 x 7.8	5.88 x 11.20

Table 1.4 Intel microprocessor trends.

Year	Name	Clock frequency	Power supply	Technology	
1971~	4004	740~750 kHz	12 V, 5 V, -5 V	10 μm	PMOS
1974 ~	8080	2~3.125 MHz	5 V	6 μm	NMOS
1978 ~	8086	5~10 MHz	5 V	3 μm	NMOS
1984~	80286	6~25 MHz	5 V	1.5 μm	NMOS
1985 ~	i368	12~40 MHz	5 V	1.5 ~ 1 μm	CMOS
1993 ~	Pentium P5	50~66 MHz	5 V	0.8 μm	BiCMOS
1994 ~	Pentium P54C	75~100 MHz	3.3 V	0.6 ~ 0.5 μm	BiCMOS
1995 ~	Pentium P54CS	133~200 MHz	2.8 V	0.35 μm	BiCMOS
1997 ~	Pentium II	233~450 MHz	2 V	0.35 ~ 0.18 μm	CMOS
1999 ~	Pentium III	0.45~1.4 GHz		0.25 ~ 0.13 μm	CMOS
2000 ~	Pentium 4	1.4~3.8 GHz		180 ~ 45 nm	CMOS
2008 ~	Core i7	1.06~5.1 GHz		45 ~ 10 nm	CMOS
2017 ~	Core i9	2.1~5.3 GHz		14 ~ 10 nm	CMOS

products such as Intel's Pentium P5. However, as MOSFET gate lengths shrank to dimensions comparable to BJT base widths, their drive current surpassed that of BJTs, and BiCMOS was gradually displaced by advanced CMOS in the late 1990s.

MOS LSI processors had already been widely used in personal computers since the 1970s and in workstations since the 1980s, but large-scale mainframes traditionally relied on high-speed ECL ICs based on BJTs. In the early 1990s, researchers proposed replacing even mainframe ECL with CMOS [321], and this transition was realized in the 2000s. As an example, the typical propagation delay time (t_{pd}) of a 120 nm gate-length CMOS inverter was already 11.4 ps [322]. On the other hand, BiCMOS was widely adopted in mixed-signal applications due to the superior analog and RF performance of BJTs, and its development continued even after CMOS superseded it in logic [323, 324].

1.10 Advances in Scaling Technologies—Introduction of Novel Process Techniques and Materials (1980s to Early 1990s)

Between the 1980s and the early 1990s, the rapid scaling of LSI components required simultaneous improvements in device performance and suppression of SCEs. To meet these stringent demands, numerous innovations in transistor structures, process technologies, and materials were introduced. Among these, the adoption of RTP—including Rapid Thermal Annealing (RTA), Rapid Thermal Oxidation (RTO), and Rapid Thermal Nitridation (RTN)—represented a pivotal breakthrough, enabling the formation of ultra-shallow junctions, ultrathin gate oxides, and reliable silicide films—essential to continued device miniaturization.

During this period, the practical use of TCAD and ECAD advanced, providing significant support for device development. With the advancements of CMOS ICs, EWS became feasible, allowing these design tools to be used by individual engineers without relying on mainframes.

1.10.1 Evolution of Source/Drain Extension Structures: LDD and Halo/Pocket Implant Architectures

As described in Section 1.5.3, the concept of source/drain extension structures was first disclosed in a 1967 patent by Hughes Aircraft Company [205]. This architecture aimed to reduce gate-to-drain overlap capacitance, improve drain breakdown voltage, and mitigate lateral electric fields [207]. However, the heavily doped source/drain regions formed outside the extension areas lacked self-alignment with the gate electrode. In 1980, IBM proposed a method, as illustrated in Figure 1.46a, using gate sidewalls to self-align the gate electrode with the heavily doped source/drain regions outside the extensions [212, 213]. This approach effectively reduced drain electric fields, mitigated hot-carrier-induced degradation, and became known as the LDD structure. Notably, the concept of using gate sidewalls for self-alignment had first been proposed before IBM by Matsushita Electric in 1977 [325].

The LDD structure became standard in the 0.5 μm CMOS technology node (around 1993) and the 0.35 μm node (around 1995). However, at the 0.25 μm node (around 1998), the need for lower resistance required higher impurity concentrations in the extension regions. Lateral diffusion caused the source/drain regions to approach each other, aggravating SCEs such as punch-through and drain-induced barrier lowering (DIBL), which in turn led to elevated off-state leakage currents [327, 328]. To overcome these challenges, IBM introduced a technique known as the “halo structure” or “pocket implant.” In this approach, high-concentration impurity regions

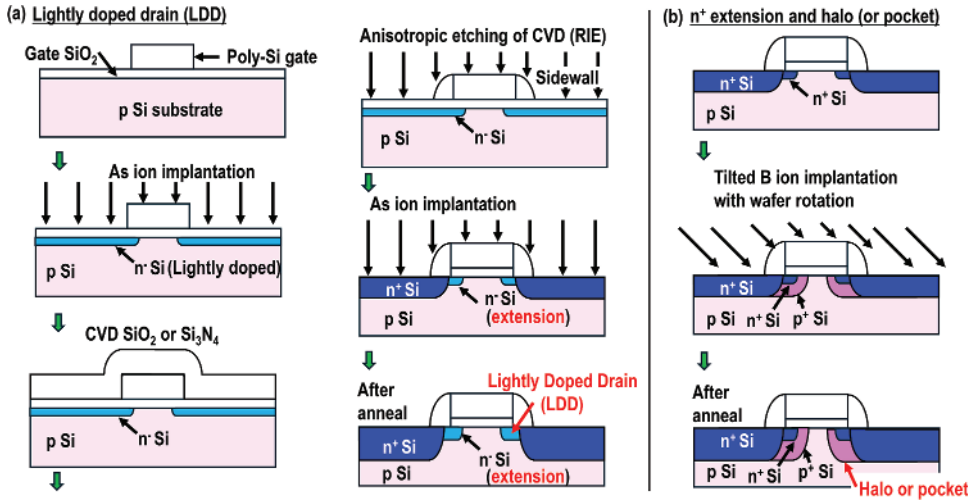


Figure 1.46 LDD structure [212, 213] to mitigate the drain electric field and halo structure [326] to suppress punch-through, along with their formation processes. *Source:* (a) Adapted from [213], (b) Adapted from [326].

of the same polarity as the channel (i.e., the substrate) are formed around the channel, thereby suppressing both the lateral diffusion of impurities from the source/drain extensions and the expansion of the depletion layer. As illustrated in Figure 1.46b. First proposed and experimentally validated in 1985 [326], this structure became widely adopted in the $0.25\text{ }\mu\text{m}$ CMOS node [327, 328].

1.10.2 Rapid Thermal Processing (RTP: RTA, RTO, and RTN)

High-temperature annealing is essential for dopant activation after ion implantation. However, forming ultra-shallow junctions requires short-duration, well-controlled processing. Conventional quartz-tube electric furnaces required more than several minutes for the loading and unloading of wafers mounted on quartz boats, and also suffered from ambient contamination caused by the intrusion of external air (Figure 1.47a). By the mid-1980s, RTP systems using halogen lamps were introduced [329–332]. After wafer loading, the chamber was purged and filled with the desired process gas, enabling high-temperature treatments for short durations in a controlled ambient (Figure 1.47b). RTP emerged as a widely adopted technique in CMOS processing during the 1990s, particularly for sub- $0.5\text{ }\mu\text{m}$ technologies, where RTA was employed to form shallow source/drain junctions [333, 334].

RTP proved indispensable not only for dopant activation in ultra-shallow junctions but also for forming ultrathin gate oxides. Conventional furnace oxidation required significant time for wafer loading and unloading, which made short thermal cycles impractical. To suppress excessive oxide growth, the furnace temperature had to be reduced to relatively low levels, such as 600°C or 700°C . However, oxides formed under these conditions exhibited high defect densities and were inadequate for use as MOSFET gate oxides. In contrast, RTO enabled short-duration oxidation at $>900^\circ\text{C}$ (beyond or close to the glass transition temperature), significantly improving oxide quality. Furthermore, by preventing film oxidation caused by trace oxygen and moisture, RTP (or RTA) became an indispensable technique for silicide formation in advanced VLSI processes.

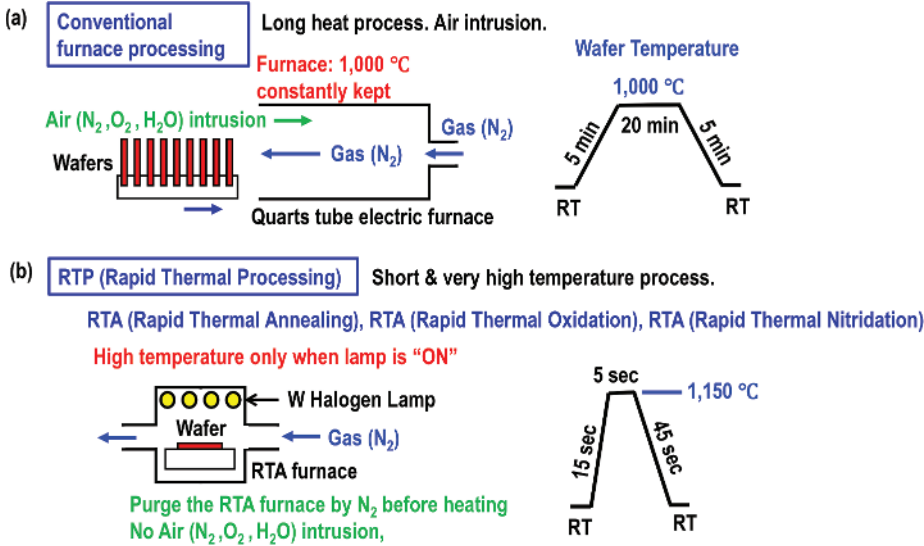


Figure 1.47 Advantages of rapid thermal processing (RTP) compared to conventional quartz tube furnace annealing (FA).

1.10.3 Introduction of Polycide Gate Electrodes: Applications in Memory and Logic Devices

1.10.3.1 Polycide for Memory Devices

In high-density DRAM, word lines are formed by connecting the polysilicon gate electrodes of transfer gate MOSFETs. As gate lengths shrank in the early 1980s, word line resistance became a critical issue. To address this, polycide structures—where silicide films are formed atop poly-Si layers—were developed. Around 1980, IBM demonstrated tungsten silicide (WSi_2) on poly-Si using dual electron-beam co-evaporation [335], while Bell Labs deposited tantalum (Ta) on poly-Si followed by annealing to form $TaSi_2$ [336]. Toshiba pioneered $MoSi_2$ polycide technology, applying it in its 256 K-bit DRAM and adopting it for mass production in 1983 [337, 338]. Instead of forming silicide by post-deposition annealing of Mo, Toshiba directly sputtered $MoSi_2$ films onto poly-Si [337].

1.10.3.2 Polycide for Logic Devices

Although logic circuits lack long word lines, reducing gate electrode resistance remained vital. As described in Section 1.5.4, CMOS circuits initially used n^+ poly-Si gates for both n-MOSFETs and p-MOSFETs. However, in p-MOSFETs, the n^+ gate work function induced excessively negative threshold voltages, preventing turn-on. To address this, the channel was doped with boron (similar to the V_{th} adjustment implant used in n-MOSFETs), forming a channel region with high acceptor concentration. This enabled appropriate V_{th} and gate voltage operation. However, unlike normal surface channel devices, the FET current flowed deeper beneath the surface in the B-implanted region, resulting in a buried-channel MOSFET.

As gate lengths scaled to 0.4–0.3 μm , buried-channel MOSFETs became harder to control due to deeper current paths and pronounced SCEs. Thus, dual-gate structures— n^+ poly-Si for n-MOSFETs and p^+ poly-Si for p-MOSFETs—were adopted. To simplify fabrication, undoped CVD poly-Si was deposited and selectively doped via ion implantation, as shown in Figure 1.48. However, implanted poly-Si films had higher resistivity than in-situ-doped poly-Si, prompting the

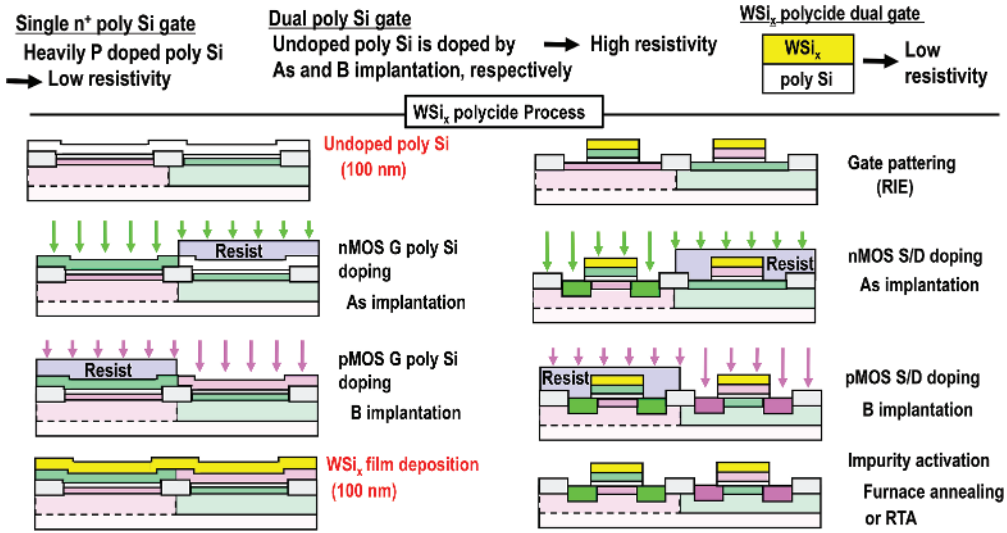


Figure 1.48 High poly Si gate resistance for dual-gate poly-Si and introduction of polycide.

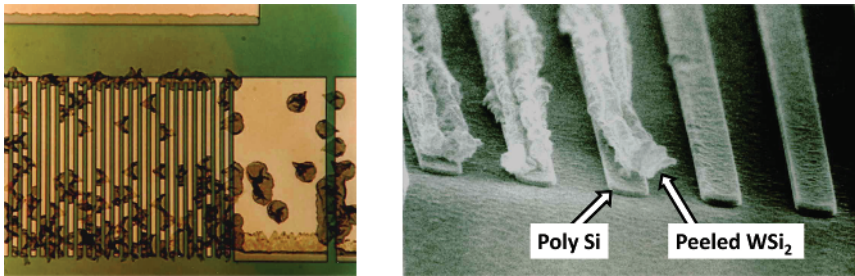


Figure 1.49 WSi_x film peeling off due to mechanical stress caused by differences in thermal expansion coefficients.

use of polycide gates. Among silicide options, WSi_2 was chosen for thermal stability and low sheet resistance. WSi_2 polycide gates were formed by depositing WSi_2 onto poly-Si via sputtering or CVD.

A major Issue was peeling of the WSi_x film from poly-Si during post-source/drain implantation heat treatment, caused by thermal expansion mismatch (Figure 1.49). Peeling occurred at the wafer periphery and beveled edge regions, attributed to compositional nonuniformity and increased edge stress. Once peeling occurred, furnace shutdown and cleaning were required, hindering development and production. This was resolved by optimizing the WSi_x composition.

Another issue was dopant segregation (As, B) at WSi_x /poly-Si interfaces and interdiffusion between n^+ and p^+ regions, causing threshold voltage shifts. These effects were effectively suppressed using RTA [339].

1.10.4 Introduction of Self-Aligned Silicide (Salicide) Technology

As devices scaled down, reducing resistance in the gate, source, and drain regions became essential. In December 1980, Toshiba invented a method to simultaneously form silicide films on gate electrodes and source/drain regions in a self-aligned manner as illustrated in Figure 1.50a [340], termed “self-aligned silicide” or “salicide” [341]. IBM independently disclosed a similar concept in August

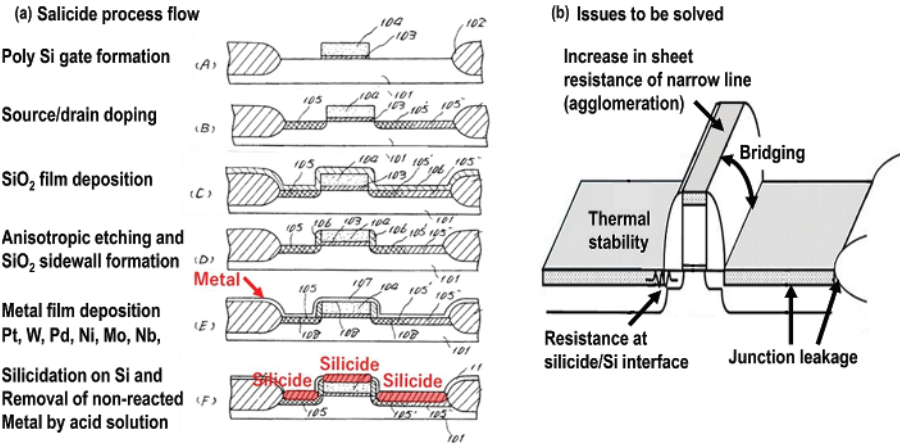


Figure 1.50 Salicide process proposal in 1980 Shibata [340] and issues to be solved [345].

1981 [342]. Toshiba experimentally demonstrated PtSi salicide MOSFETs in 1981 [341], while TI introduced TiSi₂ salicide MOSFETs in 1982 [343]. By the 0.5 μm CMOS generation, TiSi₂-based salicide became standard [344]. Initially, TiSi₂ was favored due to its low resistivity. However, as linewidths narrowed, TiSi₂ exhibited agglomeration, bridging, and junction leakage, as illustrated in Figures 1.50b and 1.51a [345]. These issues arose because TiSi₂ forms via silicon diffusion into Ti, leading to local depletion of Si in narrow lines as well as the formation of TiSi₂ bridging along the sidewalls.

Toshiba developed nickel monosilicide (NiSi) salicide technology in 1991 [345–347]. In contrast to TiSi₂, NiSi is formed through metal diffusion into silicon, enabling thicker silicide layers to develop in narrow lines and reducing the likelihood of bridging, as shown in Figure 1.51b.

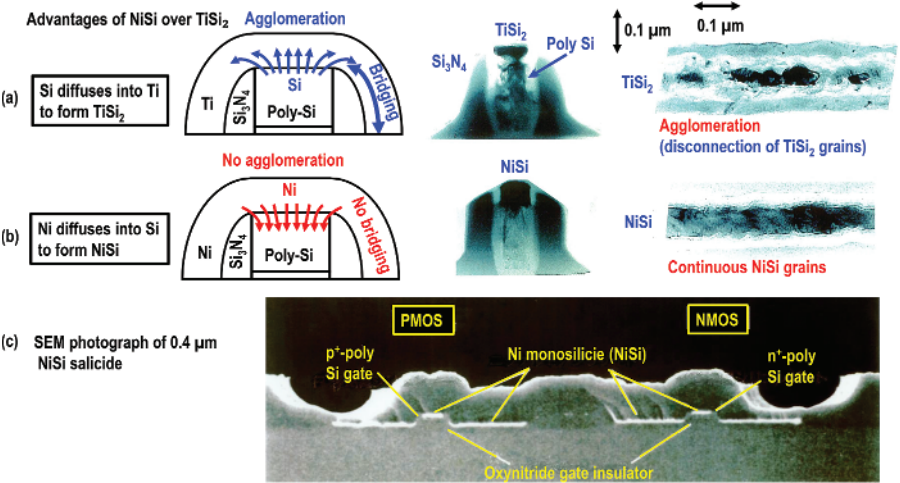


Figure 1.51 Advantages of NiSi salicide over TiSi₂ salicide due to differences in the silicidation mechanism. Source: Adapted from [345–348].

NiSi offered further advantages: as a monosilicide, it consumed less silicon than TiSi_2 or CoSi_2 , making it suitable for shallow junctions. The mechanisms underlying these differences in silicide formation between titanium and nickel/cobalt were extensively studied and clarified in 1994 [348]. They further demonstrated that nitrogen doping in Ni silicide enhances its thermal stability and promotes smoother silicide morphology [349]. Although many companies adopted CoSi_2 , NiSi's superior characteristics led to its industry-wide adoption in the early 2000s. Today, while gate electrodes employ metal rather than silicide, NiSi remains widely used in source/drain regions. Adding Pt further enhances its stability and reduces contact resistance.

1.10.5 Suppression of Boron Penetration (1986–1990) by Hydrogen-Free Anneal or Lightly Nitrided Oxynitride (SiO_xN_y) Gate Insulators

As gate oxides scaled to a few nanometers, B from p^+ poly-Si gates diffused through the oxide into the channel during high-temperature processing (Figure 1.52a), causing negative threshold voltage shifts, subthreshold slope degradation, and punch-through (Figure 1.52b). In contrast, n-MOSFETs with As- or P-doped gates did not exhibit such shifts, due to the much lower diffusivity of As and P compared with B in the SiO_2 film. IBM published a foundational study on boron penetration in 4–13 nm SiO_2 gate oxide and demonstrated boron penetration-free p^+ poly-gate p-MOSFETs by using a hydrogen-free annealing process after the p^+ gate/source/drain (G/S/D) implant doping [350, 351].

Si_3N_4 films were known to block B diffusion but had poor Si interfaces. Low-pressure CVD (LPCVD) Si_3N_4 on SiO_2 suppressed penetration but hindered further thinning of the insulator. In 1990, Toshiba investigated the nitridation of SiO_2 to form SiO_xN_y films with low nitrogen concentration [352, 353]. Excess nitrogen degraded the interface, so the level had to be minimized while still blocking B diffusion. A 6 nm SiO_2 oxide was treated by RTN and then RTO to repair the interface, keeping the thickness near 6 nm. Figure 1.52c shows ~4% nitrogen at the interface, yet, as the rightmost figure indicates, this was enough to suppress B penetration while preserving excellent interfacial properties (Figure 1.52d). This innovation was vital for stable dual-gate CMOS and was later widely adopted.

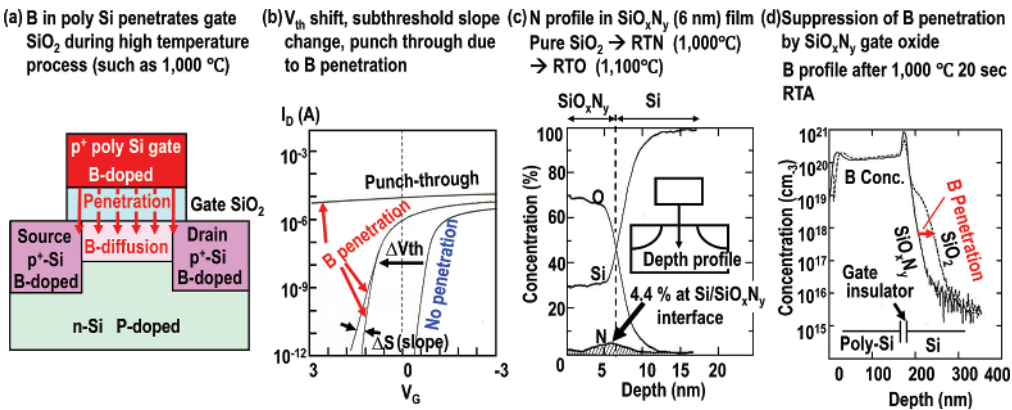


Figure 1.52 Suppression of boron penetration using an oxynitride (SiO_xN_y) gate insulator fabricated RTN. Source: (b), (c), (d) Adapted from [353].

1.10.6 Introduction of Shallow Trench Isolation (STI) and Copper Interconnect Technologies

When the industry transitioned to the 0.35 μm technology generation in the mid-1990s, as shown in Figure 1.53a and b, device isolation by LOCOS became an obstacle to further scaling due to the presence of the bird's beak (lateral encroachment of field oxide) and the lateral diffusion of p^+ regions such as channel stoppers [297]. As a result, shallow trench isolation (STI) was introduced. STI began to be adopted in volume production processes at the 0.35 μm generation, and from the 0.25 μm generation onward, it became established as the standard device isolation technology.

Meanwhile, in BJTs, deep trench isolation (DTI) had already been used since the 1980s to separate buried collector layers formed deep in the substrate. In DRAM as well, trench capacitors had been introduced from the mid-1980s, and trench technology itself was already widely disseminated. However, in BJTs and DRAMs, it was common practice to oxidize the trench sidewalls and then fill them with poly-Si, a material that was easy to handle because of its good conformality during deposition and its thermal expansion coefficient being close to that of the Si substrate. By contrast, in MOSFET device isolation, a shallow trench was sufficient and had to be filled with CVD oxide because there was not enough space to accommodate electrically well-insulated $\text{SiO}_2/\text{poly-Si}/\text{SiO}_2$ multilayers within the trench.

Proposals and experiments on STI were reported by multiple research groups, and only a few examples are cited here. For instance, around 1975, Hitachi proposed a process in which CVD- SiO_2 was deposited into trenches, the surface was planarized with polyimide, and then etched back to leave CVD- SiO_2 inside the trenches, as shown in Figure 1.53c [354]. This patent does not specify whether the trench is deep or shallow; however, the embodiment concerns the isolation of the n^+ buried collector of a BJT, with the main application being DTI. DTI is not applied to all

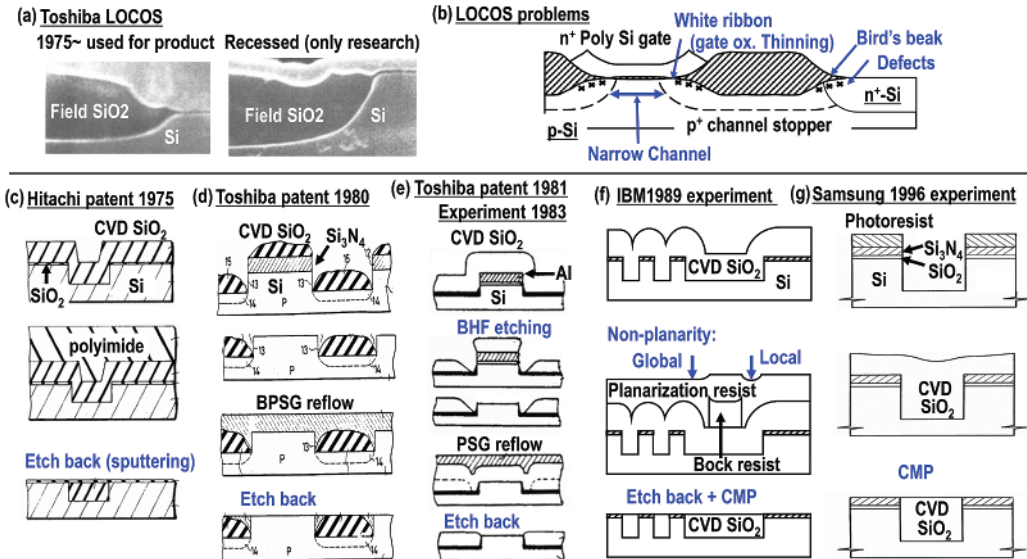


Figure 1.53 Shallow trench isolation (STI) [297, 354–357, 363]. Source: (f) Adapted from [357].

device-isolation regions but is limited to the separation of the buried collector layer; therefore, all trenches are narrow and uniform in width. In contrast, STI is applied to all device isolations, resulting in trenches of various widths, including wide ones. Despite planarization, the polyimide over these wide trenches remains thin, and etch-back removes the CVD-SiO₂ within them.

To address this issue, Toshiba filed a patent in 1980 as shown in Figure 1.53d. The method describes removing the CVD-SiO₂ outside the trenches by lift-off to achieve near-planarity, except for small grooves at the boundaries between the trenches and the embedded CVD-SiO₂. A flowable film (BPSG) is then deposited for complete planarization, followed by etch-back to fill the recesses formed between the trenches and the embedded CVD-SiO₂ [355]. A 1981 Toshiba patent [356] was based on the same concept as the 1980 one, but uses an Al sacrificial layer to facilitate lift-off as shown in Figure 1.53e. This technology was developed in 1983 [359] for device isolation in NMOS 1-Mbit DRAMs, and in 1985, the operation of such an NMOS 1-Mbit DRAM was presented at ISSCC [360]. However, the device Toshiba actually mass-produced was the CMOS 1-Mbit DRAM announced at the same ISSCC [361], which did not employ STI.

Subsequently, STI development advanced at IBM. In 1988, STI using etch-back was reported [362], and in 1989, it was demonstrated that by applying chemical mechanical polishing (CMP) after etch-back, the residual non-planarity could be completely eliminated as shown in Figure 1.53f [357]. IBM's contributions to the advancement of STI and CMP were extremely significant.

Although the concepts of mechanical polishing and CMP had existed earlier, CMP for planarizing interlayer dielectrics in multilevel interconnects was developed in the late 1980s, primarily at IBM. To be applied to STI, CMP had to be combined with size- and pattern-density-related solutions (active region and isolation trench) to avoid dishing and erosion effects. Thereafter, STI with CMP planarization shown in Figure 1.53g [358, 363] became mainstream, forming the foundation of device isolation technology in the scaling generations. From the late 1990s onward, planarization by CMP became an indispensable element of the STI process.

At the end of the 1990s, copper interconnects replaced aluminum due to Cu's lower resistivity and superior resistance to electromigration [364, 365]. These advantages made Cu especially effective for long interconnects and high-frequency operation. IBM pioneered Cu interconnects integration, which relied heavily on CMP, as shown in Figure 1.54. CMP was critical for planarizing surfaces after STI and for patterning Cu interconnects in multilevel metallization. The adoption of STI and Cu interconnects marked a decisive turning point in device miniaturization, driving further reductions in feature size while simultaneously enhancing performance and reliability. These technologies became foundational for all subsequent CMOS generations. For additional reading on interconnects, see Chapter 17 in this book.

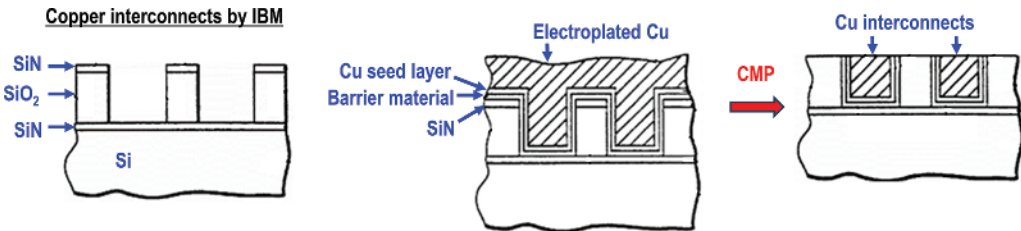


Figure 1.54 Copper interconnects formation process [364].

1.11 Challenges in Scaling into the Sub-50 nm Regime from the Mid-1990s to the 2000s

1.11.1 40 nm Gate-Length n-MOSFETs with 3 nm Gate Oxide Thickness and 10 nm Junction Depth

1.11.1.1 Progress of Scaling and the “Red Brick Wall” around 70–100 nm

The substantive development of MOS LSI technology began with PMOS LSIs such as 256-bit SRAMs and 1 k-bit DRAMs, fabricated with linewidths of 8–10 μm and introduced between 1969 and 1971. Soon after, companies across the industry launched the second generation of LSIs, initiating the development of NMOS devices with linewidths of 5–6 μm and advancing toward 1 k-bit SRAMs and 4 k-bit DRAMs. However, as gate lengths were reduced, threshold voltages dropped sharply beyond a critical limit, leading to variations and, in some cases, transistors failing to turn off. This phenomenon—the so-called SCE [366–369]—was widely recognized around 1974. At Toshiba, the SCE became problematic in 1973 for the devices with gate lengths below 4 μm, prompting analysis and countermeasure development [370].

To counteract this, companies explored methods not only to shorten the lateral gate length but also to reduce the vertical source/drain junction depth, and to increase channel or substrate doping concentrations in order to suppress source/drain depletion regions. The theoretical framework that consolidated these approaches—providing guidelines for suppressing the SCE and ensuring proper MOSFET operation at each scaling generation—was the scaling principle proposed in 1974 by Robert Dennard of IBM [3] (Figure 1.55).

Dennard’s principle called for all physical dimensions (length, width, and height) and the supply voltage to be reduced by a factor of K (<1 , typically 0.7), while the impurity concentration was scaled by $1/K$ every two to three years. As a result, current–voltage characteristics were uniformly reduced by the same factor K , preserving their original form. As shown in the accompanying table, under this scaling law, integration density and clock frequency increase by factors of $1/K^2$ and $1/K$ per generation, respectively, while power consumption remains constant. This approach, which maintains a constant electric field, became known as constant-field scaling or ideal scaling. It should be noted that the value of K being 0.7 was not specified by Dennard [3] (he defined the scaling constant, $\kappa \equiv 1/K > 1$), but rather was naturally derived from the industry’s tacit agreement at the time to quadruple memory capacity every three years. Under this scheme, in principle, the

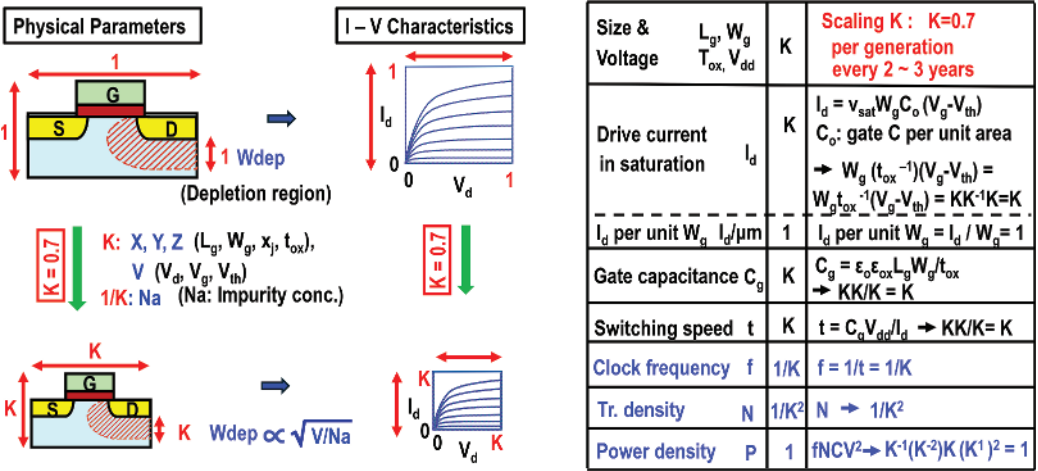


Figure 1.55 Ideal scaling scheme proposed by R. Dennard in 1974. Source: Adapted from [3].

memory cell area was reduced by half in each generation ($0.5 \cong 0.7 \times 0.7$), while the chip area was doubled. In addition, if further reductions in cell area could be achieved through improvements in cell structure or the introduction of new materials, the increase in chip area could be somewhat restrained. In any case, even under Dennard scaling, increasing the chip area inevitably leads to an increase in chip power consumption.

In practice, device miniaturization did not strictly follow the ideal scaling law (see Figure 1.72 in Section 1.13.2). Even so, the concept of ideal scaling provided an invaluable guideline for device design, and companies referred to it in reducing device dimensions by a factor of 0.7 every two to three years. Yet, due to practical power supply voltage and I/O interface voltage standard issues, as well as oxide defects, breakdown, and reliability challenges, V_{DD} and oxide thickness often did not scale at each node. This led to the so-called “constant-voltage scaling,” or nonideal scaling without the proportional scaling of supply voltage, oxide thickness, junction depth, and doping concentration, thereby not gaining the full benefits of ideal or constant-field scaling, as will be discussed later in this section and in Section 1.13.

Figure 1.56 shows the evolution of product node nomenclature, and it is notable that scaling from micro to nano LSI has continued for over 55 years. However, since 2010, while the technology node names have continued to shrink by a factor of 0.7 with each generation, the actual physical dimensions have failed to keep pace, resulting in a substantial divergence between node names and real dimensions. The physical scaling is considered to stall around the 10 nm level, whereas the node names have already reached 2 nm and are expected to advance to even smaller nominal values in the future.

In R&D, ultra-scaled MOSFET prototypes were fabricated nearly a decade before commercialization. IBM demonstrated $1\text{ }\mu\text{m}$ gate-length MOSFETs in 1974 [3] and $0.1\text{ }\mu\text{m}$ devices in 1987 [371], both operating at room temperature, as shown in Figure 1.57a. Yet in the 1980s, a “red brick wall” was predicted around the 70–100 nm gate-length range, seen as the miniaturization limit due to several reasons described in Figure 1.57b. Hiroshima University confirmed 70 nm MOSFET operation in 1992 [372]. However, as no gate width information was provided, its normalized performance remained unknown.

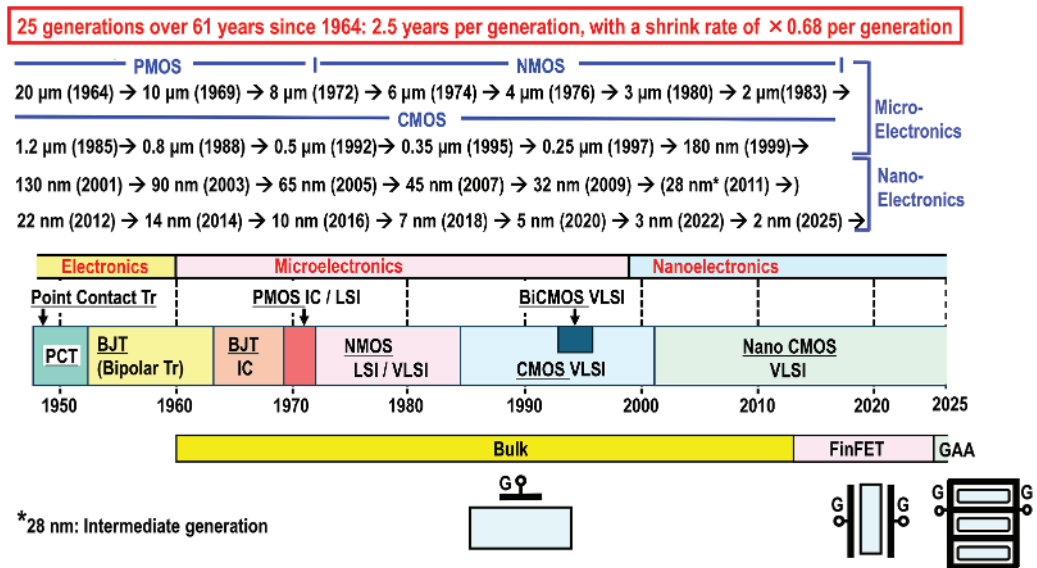


Figure 1.56 Downsizing trend of LSI generations and their technologies: $20\text{ }\mu\text{m}$ (1964) belongs to the MSI generation, while LSIs have been produced since 1969.

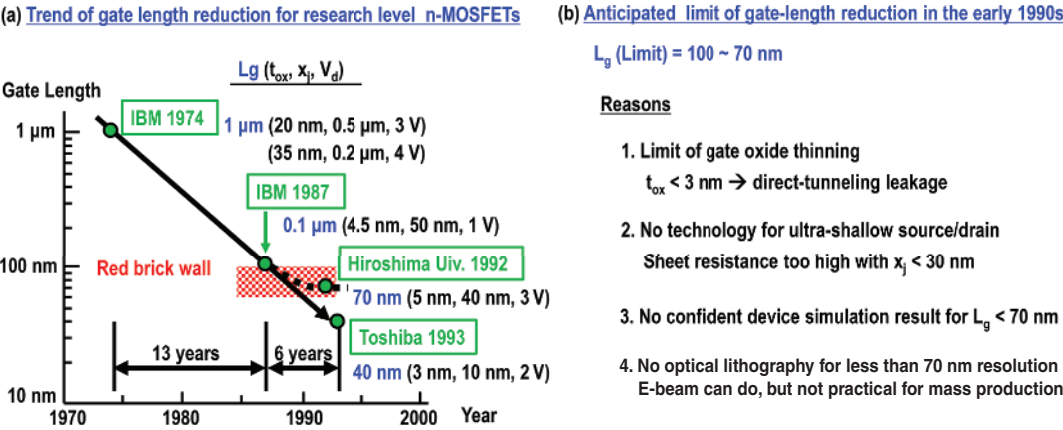


Figure 1.57 Limit of downsizing expected until the early 1990s and its breakthroughs [3, 371–373].

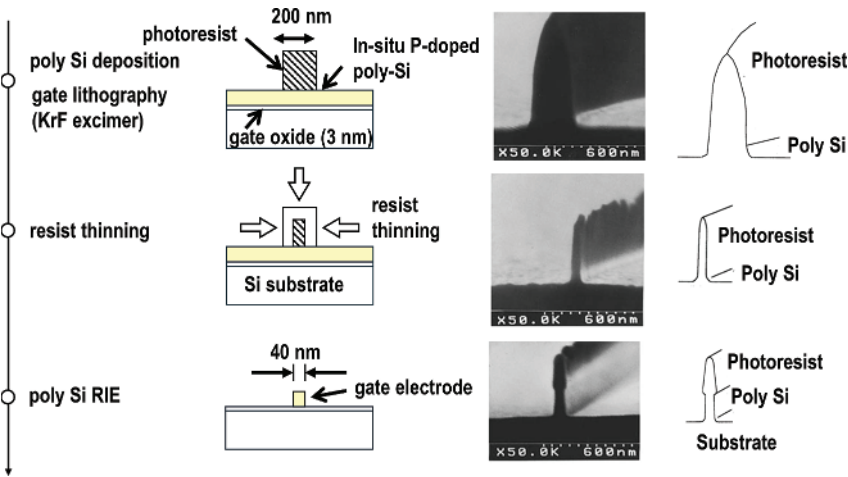


Figure 1.58 40 nm poly gate electrode realization using resist thinning technique [374, 375].

1.11.1.2 Fabrication of 40-nm Gate-Length n-MOSFETs

In 1993, Toshiba reduced 200 nm resist lines to 40 nm using KrF excimer laser lithography and oxygen plasma ashing (Figure 1.58) [374], a method first proposed by researchers at UC Berkeley [375]. These trimmed lines served as masks for 40 nm poly-Si gates, demonstrating excellent uniformity, and were subsequently widely adopted in mass production.

Gate oxide thickness and junction depth were then defined: 3 nm was considered the threshold for direct tunneling-induced leakage and was thus fixed as the oxide limit. For a 40 nm gate-length MOSFET, in order to suppress SCEs with a 3-nm gate oxide and to ensure sufficiently low off-state current, the extension junction depth had to be scaled down to approximately 10 nm. At that time, extensions were formed by high-dose ion implantation at low acceleration voltage; however, with ion implantation, it was impossible to achieve sufficiently low sheet resistance at a junction depth of 10 nm. Toshiba developed a method using solid-phase diffusion from doped-silicate-glass gate sidewalls (Figure 1.59a) [373, 376], achieving 12 nm junction depth with 6 k Ω /sq. sheet

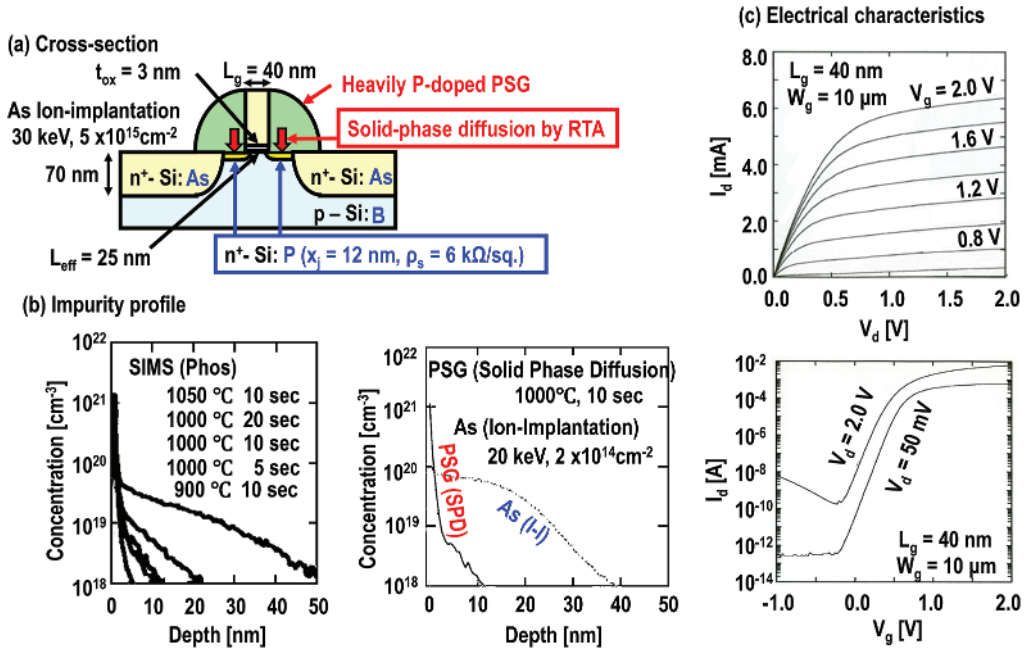


Figure 1.59 40-nm gate-length *n*-MOSFET [373].

resistance—slightly high but acceptable [373], as shown in Figure 1.59a and b. 40-nm *n*-MOSFETs operated reliably with high performance at room temperature (Figure 1.59c), demonstrating sub-50 nm feasibility and overcoming the “red brick wall” (Figure 1.57a) [373].

1.11.1.3 Scaling Relationship Between Gate Oxide Thickness and Junction Depth at the Research Level

Figure 1.60 illustrates the scaling relationship between gate oxide thickness and junction depth in research level. Historically, oxide thinning progressed conservatively, whereas junction depth reduction was pursued more aggressively to compensate and facilitate scaling. This practice is fairly common in the history of MOSFET miniaturization, where one may push the envelope of one feature to buy some time for the others. From the 1960s through the early 1970s, the gate length of NMOS LSI devices was reduced from approximately 20 to 5 μm , while the gate oxide thickness remained around 100 nm. During this period, only the junction depth was actively scaled. At that time, the source/drain resistance was sufficiently lower than the channel resistance, so the adoption of shallow junctions did not pose a significant issue. In contrast, gate SiO_2 films still contained a high density of defects, making aggressive thinning of the oxide layer difficult.

By the mid-1970s, as gate lengths were shortened to the 5–6 μm range, the industry achieved proportional scaling of gate oxide thicknesses by successfully reducing defect densities. However, due to concerns over oxide breakdown and long-term reliability, oxide scaling was still more conservative than gate-length scaling, and hence junction depth continued to be scaled aggressively. The black dashed trendline at 45° represents the geometric scaling where gate length, gate oxide thickness, and junction depth are scaled at the same ratio, starting from the baseline parameters of IBM’s 5 μm gate-length MOSFET [3] and Toshiba’s 6 μm gate-length MOSFET introduced in 1974. The actual data points fall below this trendline, indicating that oxide scaling was carried

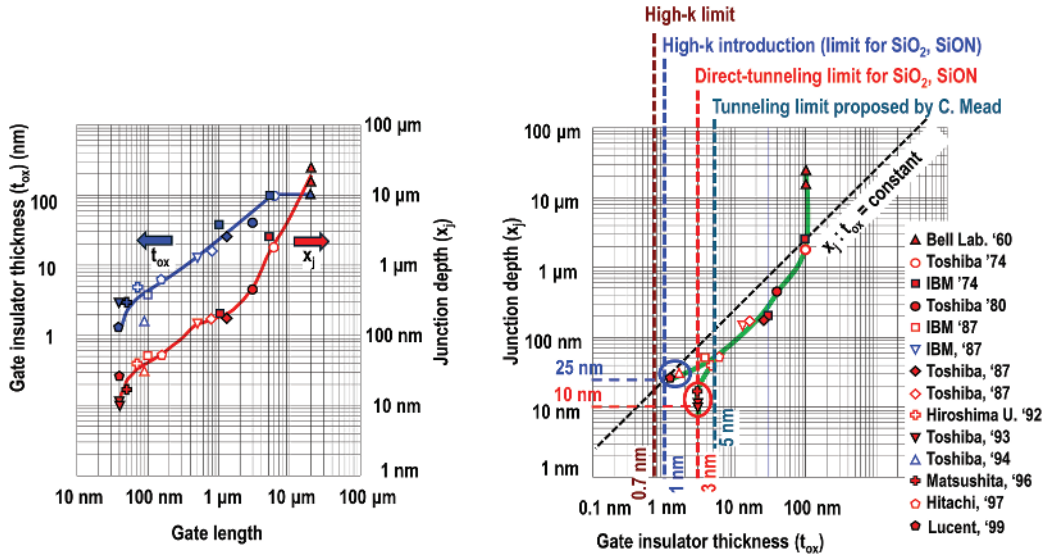


Figure 1.60 Trend of scaling of gate oxide thickness and junction depth for MOSFETs from 1960 to 1999.

out more cautiously than junction depth scaling. As gate lengths were further reduced, channel resistance decreased accordingly. On the other hand, extremely shallow junctions led to higher sheet resistance in the source/drain regions, posing a major challenge. Fortunately, as noted earlier, advances in source/drain engineering during this period, such as RTA (see Section 1.10.2), effectively resolved these issues.

In the late 1970s, C. Mead and his colleagues predicted that gate leakage currents caused by Fowler–Nordheim tunneling would impose a lower limit of approximately 5 nm for gate oxide thickness [377, 378]. However, this prediction did not hold, and a general consensus later emerged that the absolute lower limit, beyond which tunneling current becomes dominant, is approximately 3 nm. In the 1980s, it was extremely difficult to fabricate high-quality gate dielectric films thinner than 5 nm. With the introduction of RTO technology (Section 1.10.2), however, it became possible to form high-quality, uniform 3 nm gate oxide films. The use of separate or dual gate oxides to decouple the core (thin oxide) and I/O (thick oxide) devices also allowed the voltage scaling of thin-oxide core devices. By the late 1980s, as gate lengths approached 100 nm, gate oxide thicknesses had already been reduced to the 3–5 nm range, and junction depths had reached the 20–40 nm range.

In the case of the 40 nm MOSFET fabricated in 1993, the relationship between gate oxide thickness and junction depth—highlighted by the red ellipse in Figure 1.60—is shown with the gate oxide thickness fixed at 3 nm and the junction depth aggressively reduced to 10 nm.

Figure 1.61 shows the correlation between source/drain junction depth and sheet resistance, based on published data. Most values come from verified MOSFETs or process conditions during device development. The legend links symbols to doping methods: red for *n*-type dopants in *n*-MOSFETs and blue for *p*-type dopants in *p*-MOSFETs. When junction depth falls below 10 nm, sheet resistance rises several times above the scaling trend. Both schematic (Figure 1.61) and measured (Figure 1.59) profiles demonstrate that external doping methods—ion implantation, vapor-phase, or solid-phase—cannot maintain the peak concentration flat up to the junction edge. This underscores the requirement for high-concentration epitaxial doping [379].

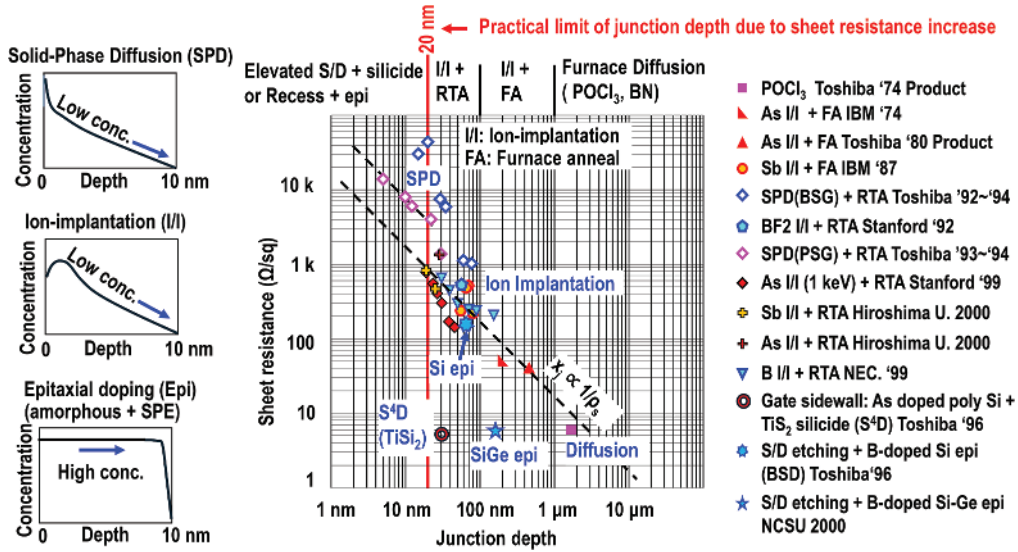


Figure 1.61 Trend of junction depth versus sheet resistance of source/drain of MOSFETs from 1960 to 1999.

1.11.2 1.5 nm Gate-Oxide n-MOSFETs with 90 nm Gate Length and 30 nm Junction Depth

Research on MOSFET operation with the gate oxide in the direct-tunneling regime at Toshiba began in 1990 [380] using oxygen-nitrided films. Figure 1.62a shows the relationship between gate voltage and leakage current for MOS capacitors with SiO₂ gate insulators of various thicknesses [381]. When the gate insulator thickness is reduced below 3 nm, as illustrated in Figure 1.62b, a quantum mechanical effect known as direct band-to-band tunneling (or simply direct tunneling) occurs, in

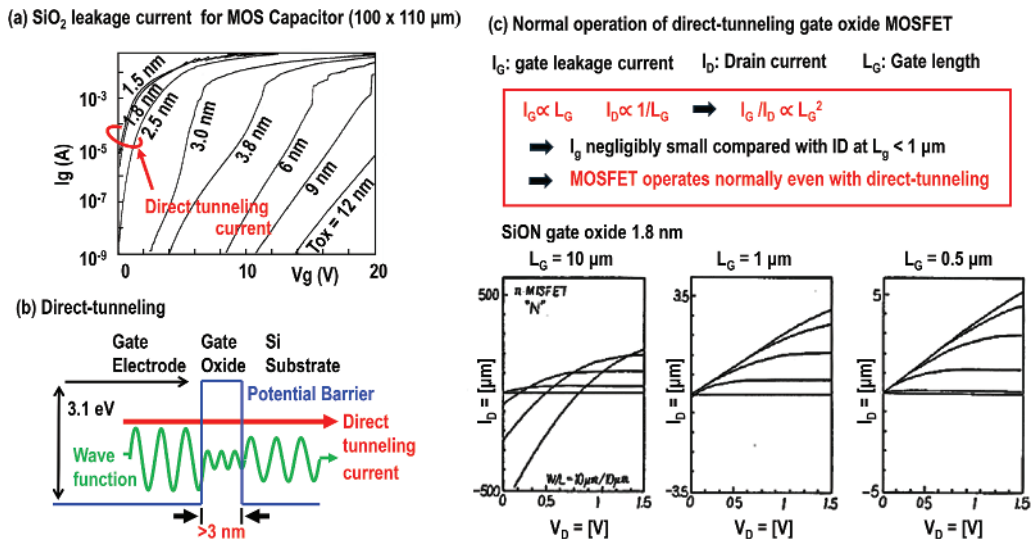


Figure 1.62 Direct-tunneling gate oxide MOSFETs [381, 382].

which electrons in the gate electrode tunnel through the energy barrier of the gate oxide and flow into the channel. Consequently, the oxide can no longer function effectively as an insulator.

However, it was later discovered that even with gate oxides thinner than 3 nm, MOSFETs can still operate properly if the gate length is sufficiently short, as shown in Figure 1.62c [382, 383]. This is because, as depicted in the figure, the gate leakage current decreases in proportion to the gate length, whereas the drain current increases inversely with it. As a result, the ratio of gate leakage current to drain current decreases rapidly in proportion to the square of the gate length.

In 1994, a MOSFET with a 1.5 nm gate oxide was prototyped, enabling the junction depth to be extended to 30 nm, as shown in Figure 1.63a. It delivered a drain current of 1.2 mA/μm at 1.5 V and a transconductance of 1010 mS/mm—the highest reported at the time, as shown in Figure 1.63b [381, 383, 384]. This exceptional performance was primarily due to the reduction of the gate oxide thickness from 3 to 1.5 nm, rather than the reduction in the sheet resistance of the source/drain extensions (from 6.2 to 1.4 kΩ/sq) resulting from increasing the junction depth from 10 to 30 nm (Figure 1.63a). The uniformity, reliability, and yield of this direct-tunneling oxide were satisfactory [385], leading to widespread adoption in commercial products.

Subsequently, Bell Laboratories investigated the physical limits of tunnel oxide thinning and confirmed that, although dependent on gate length, the drain current generally peaked at around 1.3 nm, as shown in Figure 1.63c [386]. Further thinning increased gate leakage from the drain, thereby reducing the drain-to-source current and degrading device performance. The Toshiba MOSFET with a 1.5 nm gate oxide, indicated by the blue ellipse in Figure 1.60, lies on a rational scaling line with a 1:1 ratio of gate oxide thickness to junction depth. The Bell Laboratories device, with an oxide thickness of 1.2–1.3 nm and a junction depth of approximately 25 nm, occupies a similar position in Figure 1.60 [386, 387].

Determining the practical limit of gate insulator thinning is a critical issue. In Figure 1.60, the gray dashed line denotes the 5 nm limit proposed in 1979 by researchers at Caltech, based on FN tunneling current, which led to a predicted gate-length limit of 0.25 μm that was later shown to

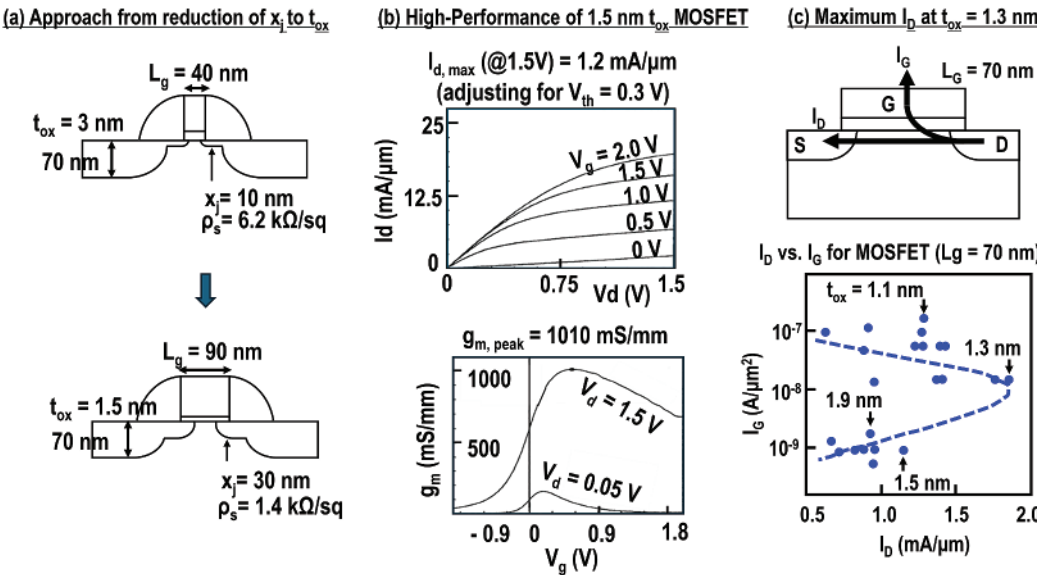


Figure 1.63 Excellent characteristics of 90-nm gate-length n -MOSFET with 1.5-nm thick SiO_2 gate insulator [381, 383–386]. *Source:* (c) Adapted from [386].

be incorrect [377, 378]. In addition, the brown line represents the reliability limit of high- k gate oxides, corresponding to an equivalent oxide thickness (EOT) of approximately 0.7 nm. Although this was not recognized at the time, it is now evident that the oxide thickness of 1.3 nm had already approached the physical limit for practical MOSFET applications.

1.11.3 Source/Drain Technology Evolution from the Mid-1990s

The introduction of external impurities resulted in a high sheet resistance when the junction depth was less than approximately 25 nm, prompting the exploration of in-situ-doped Si film growth. One example is Toshiba's 1995–1996 proposal of the Silicided Silicon-Sidewall Source and Drain (S^4D) structure as shown in Figure 1.64a. In this method, amorphous silicon heavily doped with B or As was deposited on the gate sidewalls, followed by high-temperature annealing to convert it into heavily doped poly-Si [388, 389]. A silicide process was then applied, covering the extension region with $TiSi_2$ to achieve a low-resistance elevated source/drain structure. This concept—thickening the source/drain and applying silicide—was a similar concept to the selective epitaxial source/drain approach later adopted in FinFET technology. (Note: IBM had a “strap” process innovation in its 4 Mbit DRAM, where a special “selective silicon deposition” process and silicidation formed an elevated low-resistance strap, connecting the diffusion and polysilicon of the storage trench [390].)

In 1996, Toshiba also preprocessed the source/drain regions using recess etching, followed by the selective growth of amorphous silicon with a high boron concentration [391]. This layer was crystallized through low-temperature, long-duration solid-phase epitaxy at 600°C (Figure 1.64b). The results of Toshiba's two approaches are shown in Figure 1.61, demonstrating excellent low sheet resistance. NCSU's data [392] in the same figure, based on B-doped SiGe epitaxial films, indicated that although recessing was not performed, epitaxy is feasible with a 30 nm recess.

In Intel's 90 nm CMOS technology introduced in 2002, PMOS source/drain regions were first recessed, then selectively grown with SiGe layers containing 20–30% Ge. Boron was implanted into

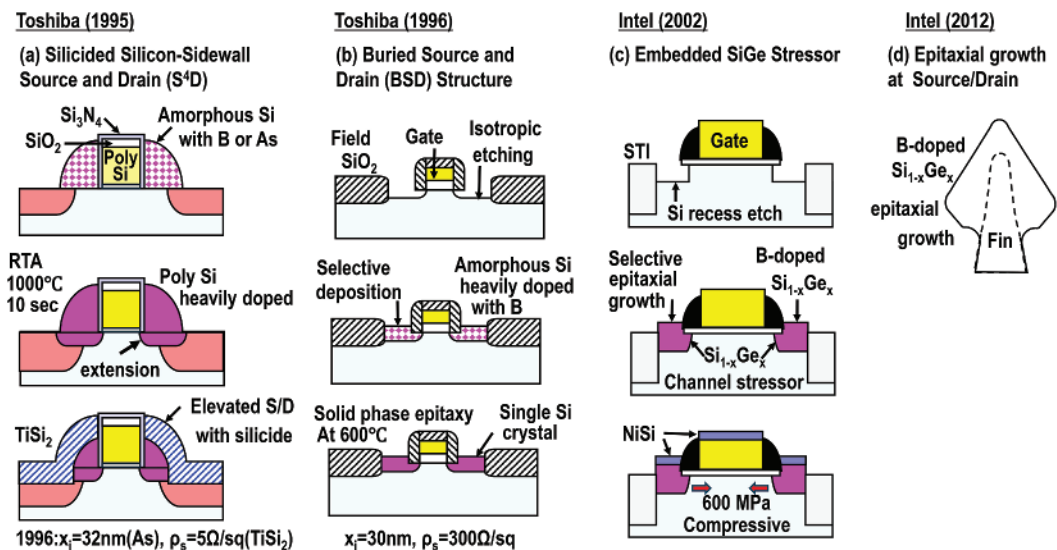


Figure 1.64 Elevated drain and epitaxial doping source/drain for significantly low sheet resistance [388–393]. Source: (b) [390]. (c) Adapted from [393]. (d) Adapted from [394].

these regions, activated via RTA, and subsequently salicided with NiSi [393, 395] (Figure 1.64c). The lattice mismatch between the Si substrate and SiGe layer induced compressive strain in the channel, improving hole mobility by over 50% and enhancing PMOS drive current by approximately 25%.

In Intel’s 2012 FinFET (Tri-Gate) devices, the source/drain regions are formed atop the fin. Owing to the fin’s small thickness (approximately 8 nm), epitaxial regrowth of in-situ-doped SiGe or Si source/drain regions is employed to reduce resistance and/or induce channel stress/strain for mobility enhancement (Figure 1.64d) [394, 395]. In PMOS devices, compressive strain continues to be introduced via SiGe epitaxy in the source/drain regions. In contrast, NMOS devices typically employ gate stack engineering and tensile CESL (Contact Etch Stop Layer) films to apply tensile strain. Since the introduction of FinFETs, source/drain thickness has been determined by the fin itself, eliminating the need for ultra-shallow junctions in the substrate. This remains true for current GAA nanosheet MOSFETs, where the sheet thickness is approximately 4 nm. To reduce resistance, epitaxial growth and silicidation are essential.

1.12 Development of RF CMOS Device Technology from the Mid-1990s to the Late 1990s

As illustrated in Figure 1.65, mobile communication systems have undergone generational transitions approximately every decade. By the mid-1990s, second-generation (2G) mobile phones—primarily based on the Global System for Mobile Communications (GSM)

1980s	1G	Analog, large, heavy, car phone, not popular
1990s	2G	Digital, cellular, GSM, voice/text, spread in the world
2000s	3G	Voice/Data communication, IP-based → smart phone
2010s	4G	Data transfer: 1 Gbps, more data and connections
2020s	5G	Higher speed and larger data treatment
2030s	6G	Ultra performance: speed (100 Gbps), capacity, coverage

The situation when 2G started in the early 1990s

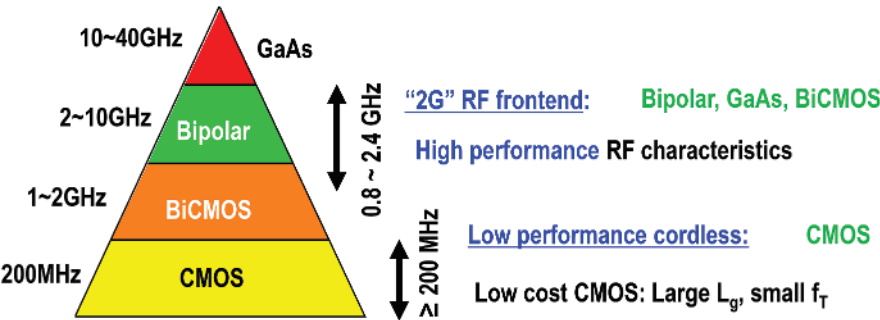


Figure 1.65 Evolution of mobile telecommunication generations and the situation when 2G started.

CMOS inverter t_{pd} (Propagation delay time)

t_{pd}	L_g	V_d	Year	Company
11.4 ps	0.12 μm	2.0 V	1996	Toshiba
11.8 ps	0.10 μm	2.5 V	1993	AT&T
12.0 ps	0.06 μm	2.0 V	1995	NEC
15.4 ps	0.15 μm	1.5 V	1995	Toshiba
17.0 ps	0.10 μm	2.5 V	1993	IBM
19.0 ps	0.075 μm	2.0 V	1993	Fujitsu

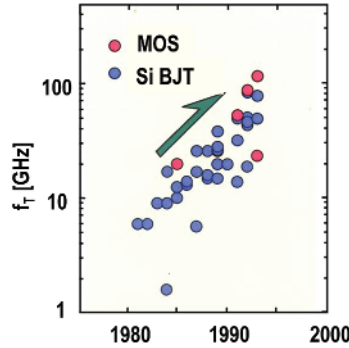


Figure 1.66 Trends of CMOS inverter propagation delay time (t_{pd}) and transition frequencies of BJT and MOSFET [322, 397].

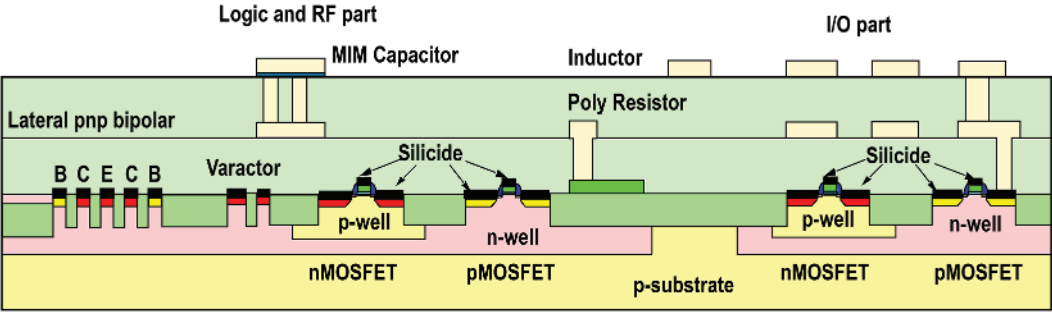
standard—began to see widespread adoption. During this period, RF front-end circuits commonly employed BJTs and GaAs high-electron-mobility transistors (HEMTs) and FETs operating in the 0.8–2.4 GHz frequency range. Although CMOS technology was also utilized in wireless communication systems, its application was largely limited to low-cost, low-power devices operating below 200 MHz, such as cordless phone receivers. These applications typically relied on older-generation, low-cost CMOS technologies characterized by long gate lengths. However, as shown in Figure 1.66, by the mid-1990s, the propagation delay time (t_{pd}) of state-of-the-art CMOS inverters approached 10 ps [322], and their transition frequency (f_T) exceeded 100 GHz [397]. Anticipating the potential for RF CMOS technology to surpass BJTs in future applications, Toshiba initiated the development of CMOS technology for RF front-end circuits for GSM applications in 1994 [28, 398].

In RF CMOS design, not only high f_T and f_{max} but also superior linearity and low-noise characteristics are essential. Moreover, as depicted in Figure 1.67, the integration of passive components such as inductors [399], metal–insulator–metal (MIM) capacitors [400], varactors, and resistors is indispensable. Figure 1.67 also presents the specifications of ADVANCE-C0, a 0.25 μm RF CMOS technology developed by Toshiba.

Subsequently, through collaborative research with ETH Zurich and KU Leuven, the feasibility of applying 0.25 μm RF CMOS technology to GSM and next-generation DCS-1800 circuits was demonstrated through design and prototyping [401–404]. Furthermore, with the introduction of salicide formation and gate-length scaling down to 60 nm, experimental results showed that f_T reached up to 245 GHz [405, 406], as shown in Figure 1.68.

Given that RF MOSFETs typically feature large gate widths (W_g), reducing gate electrode resistance becomes critical. As illustrated in Figure 1.69, this is generally achieved by dividing the transistor into multiple parallel fingers and connecting them in parallel [407]. Increasing the number of fingers shortens the gate length of each finger (L_f), thereby reducing gate resistance. However, excessive segmentation increases the gate pad capacitance (C_{gb}) over the field region, necessitating an optimal trade-off.

In 1999, using the analytical expressions shown in Figure 1.69, the optimal L_f values for f_T , f_{max} , and minimum noise figure (NF_{min}) were derived [407]. Based on these insights, a comprehensive RF CMOS roadmap was formulated and presented at a conference, with experimental data comparisons shown in Table 1.5 and Figure 1.70 [407]. Table 1.5 shows that the performance indicators— f_T , f_{max} , and NF_{min} —calculated using simple analytical expressions agree well with



ADVANCE (0.25 μm) specification

	f_T [GHz]	$f_{\max}$ [GHz]	V_{th} matching (ΔV_{th}) @ $Wg = 5 \mu\text{m}$	$NF_{\min}$ [dB] @ 2 GHz	1/f noise [dBA/Hz ^{0.5}] @ 1 kHz	Capacitor Si_3N_4	Inductor	Resister
NMOS	40	45	$3\sigma(\Delta V_{th}) = 20 \text{ mV}$	0.4	-185	1.3 fF/ μm^2	2 nH	High: 2 k Ω /sq.
PMOS	20	14	$3\sigma(\Delta V_{th}) = 35 \text{ mV}$		-190	$Q = 75$	$Q = 6$	

Figure 1.67 Cross section of Toshiba RF CMOS technologies “ADVANCED C-0” (0.25 μm), “C-1” (0.15 μm), and “C-2” (0.1 μm) in the 1990s [28].

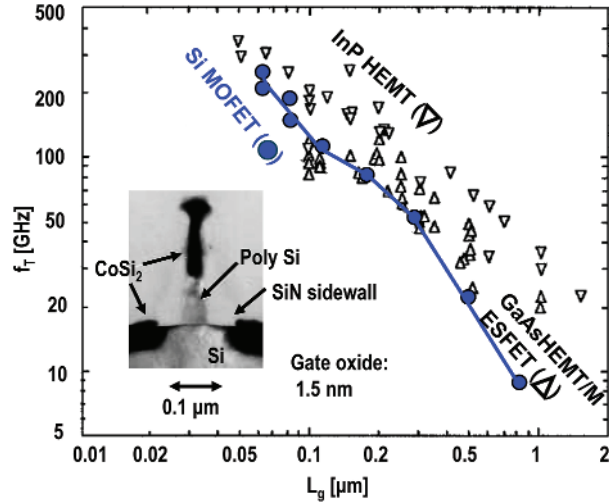
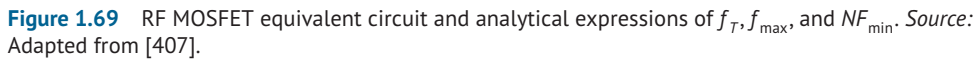


Figure 1.68 Gate length (L_g) versus transition frequency (f_T) for Si-MOSFET, GaAs HEMT/MESFET, and InP HEMT [406].

the experimental results, and these values were found to be much better than those predicted by ITRS 1997. As shown in Figure 1.70, although the optimum L_f differs depending on the performance indicator, selecting the L_f optimized for $f_{\max}$ also yields favorable values for both f_T and $NF_{\min}$. Moreover, this L_f value was shown to be close to that obtained by straightforwardly scaling the L_f optimized for $f_{\max}$ at a gate length of 250 nm. Building on these results, the performance indicators of each generation are summarized in the table in Figure 1.70. This work represented



		Lg, Wg in NTRS 1995		1995	1997	1999	2001	2003	2005	2007	2009
		Year									
		Gate length: Lg (nm)		250	180	140	120	100	70	50	35
		Gate width: Wg (μm)		200	150	110	100	80	60	50	35
L _f (Finger Length)	5 μm	fT (GHz)	Experiment	33	49	70	84	112			
			Calculation	39	50	65	80	105	145	205	420
		fmax (GHz)	Experiment	41	47	51	52	60			
			Calculation	39	42	46	50	60	62	68	85
		NFmin (dB)	Experiment	0.5	0.35	0.23	0.2	0.15			
			Calculation	0.3	0.26	0.22	0.17	0.14	0.13	0.1	0.08
		NF50 (dB)	Experiment	3.0	2.3	2.2	1.6	1.2			
		ITRS 1997	fT (GHz)		20	30	35	40	55	75	100
fmax (GHz)			25	35	40	50	65	90	120		
NFmin (dB)			2	1.5	1.3	1.2	<1	<1	<1		

In 1997, joint development with Ericsson enabled the extraction of design parameters for a 0.15 μm RF CMOS process tailored to Bluetooth applications [408]. Today’s mobile phones can be fully implemented in CMOS—from the RF front end to the baseband. The potential of CMOS integration is being explored for future applications, including sixth generation (6G) systems, satellite communications, high-precision radar sensing, the metaverse, autonomous driving, and industrial wireless links operating in the 100–300 GHz range. The functionality of each circuit block has already been demonstrated individually [409, 410].

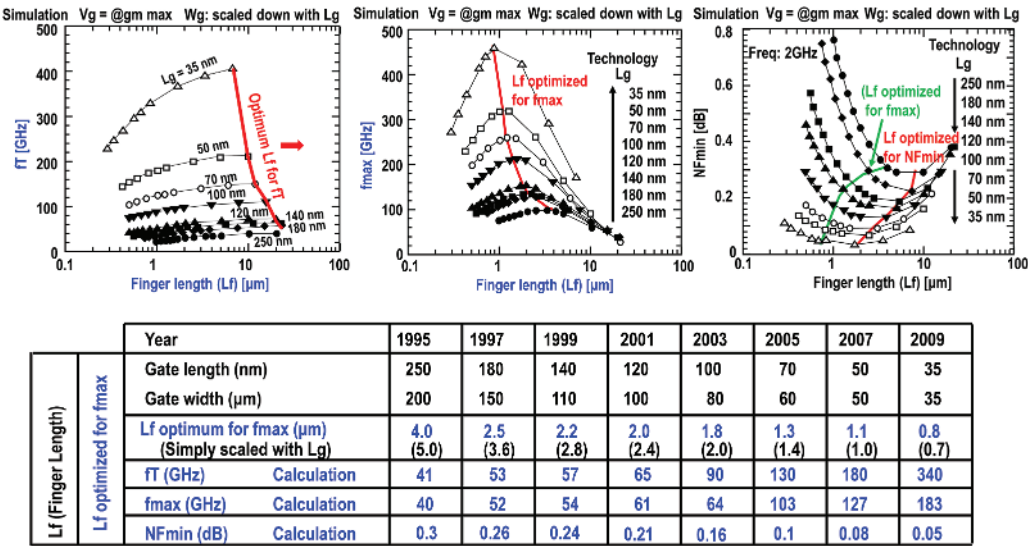


Figure 1.70 Optimum finger length (L_f) values obtained by calculation for the performance indicators— f_t , f_{max} , and NF_{min} . Source: Adapted from [407].

1.13 Post-2000: Confronting the Limits of Miniaturization

1.13.1 Three Scaling Limits in Device Miniaturization

The evolution of electronic devices has been driven primarily by miniaturization, but intrinsic boundaries constrain further geometric or dimensional scaling. Figure 1.71 illustrates three categories of these limits. The first is the ultimate limit, governed by interatomic spacing—about 0.3 nm for silicon. Even with alternative materials, this limit does not change significantly. Structures smaller than the nearest interatomic distances cannot be fabricated. Thus, nanoelectronics marks the final generation, and picelectronics is physically unfeasible.

Before reaching this, devices face the fundamental limit, defined by the direct-tunneling distance (~ 3 nm). Once the channel length falls below 3 nm, direct tunneling between source and drain arises, preventing the device from switching off. Thus, 3 nm represents the threshold for functional operation.

The third is the practical limit, where further scaling’s drawbacks outweigh its benefits. These include increased leakage current, heat dissipation, fabrication cost, reduced reliability, and greater variability. Currently, this limit is around 10 nm. Notably, logic LSI technologies labeled “5 nm” or “2 nm” do not reflect actual linewidths. Since the 28 nm node (~ 2010), physical scaling has lagged behind nominal node names. As shown in the table [411] in Figure 1.71, a clear divergence has emerged between real feature sizes and technology labels.

The following section outlines factors contributing to the practical limits.

1.13.2 Limitations of Miniaturization Due to Increasing Power Consumption

Figure 1.72a illustrates historical trends in scaling parameters from 1970 to 2000 [412]. Over this 30-year period, device dimensions—including gate length, gate oxide thickness, and junction

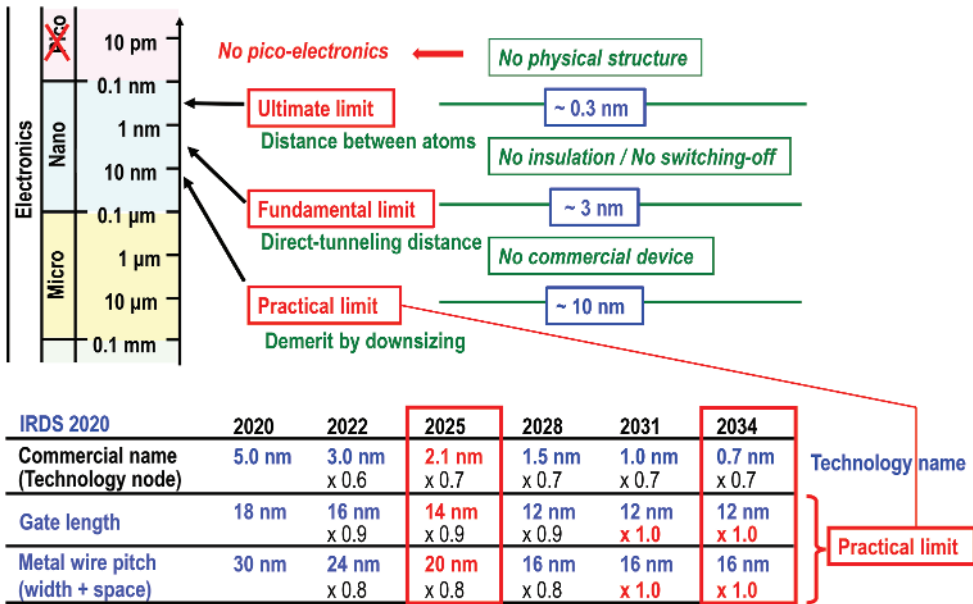
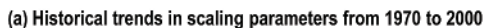


Figure 1.71 Three stages of the miniaturization limit. Table adapted from [411].

depth—were reduced by a factor of 100. In contrast, the power supply voltage decreased by less than a factor of 10, starting from the 5 V used in early NMOS LSI introduced in the early 1970s. This limited voltage reduction was primarily due to the need to maintain I/O compatibility with the TTL standard. Changing the supply voltage would require corresponding modifications to the target systems, which posed significant inconvenience. Moreover, performance gains from scaling were prioritized over power suppression. By the mid-1990s, at the 0.5 μm technology node, concerns regarding MOSFET reliability necessitated a reduction in supply voltage to 3.3 V. These concerns included time-dependent dielectric breakdown (TDDB), hot-carrier-induced degradation, and pronounced SCEs observed at 5 V. Since then, supply voltage has continued to decline, but the pace has slowed in recent years, approaching saturation around 0.7 V. This is largely due to the increasing subthreshold leakage current caused by lowering the threshold voltage—a critical issue discussed later in Section 1.13.3.4. Because the supply voltage was reduced only by roughly a factor of 10, while the chip size increased 10-fold, power consumption rose by a factor of 100,000 between 1970 and 2000. At ISSCC 2001, Intel issued a cautionary projection: if then-current scaling trends were extrapolated, chip power density would reach levels comparable to the surface of the sun by 2016 [413]. Figure 1.72b illustrates average trend lines for microprocessor characteristics spanning 1971–2015 [414]. Over more than four decades, the number of transistors per chip has steadily increased, in alignment with Moore’s law. Although its ultimate limit remains uncertain, as of 2025, this trajectory has persisted for 60 years since Moore first proposed it in 1965.

As previously noted, continued increases in power consumption and the resulting heat generation are unsustainable. Around the year 2000, the growth in clock frequency plateaued, stabilizing at approximately 3–5.5 GHz. This shift effectively curbed further increases in power consumption. However, the constraint on clock frequency also slowed the pace of performance gains for single-core processors. Nonetheless, overall microprocessor performance has continued to improve through the adoption of multi-core architectures.



Source: Adapted from [412]. (b) Microprocessors from 1971 to 2015. Source: Adapted from [414].

1.13.3 Scaling Limits Due to Off-State Leakage Current

1.13.3.1 Punch-Through Current

Leakage current

- ①. Punch-through ②. direct-tunneling
③. subthreshold ④. gate insulator leakage

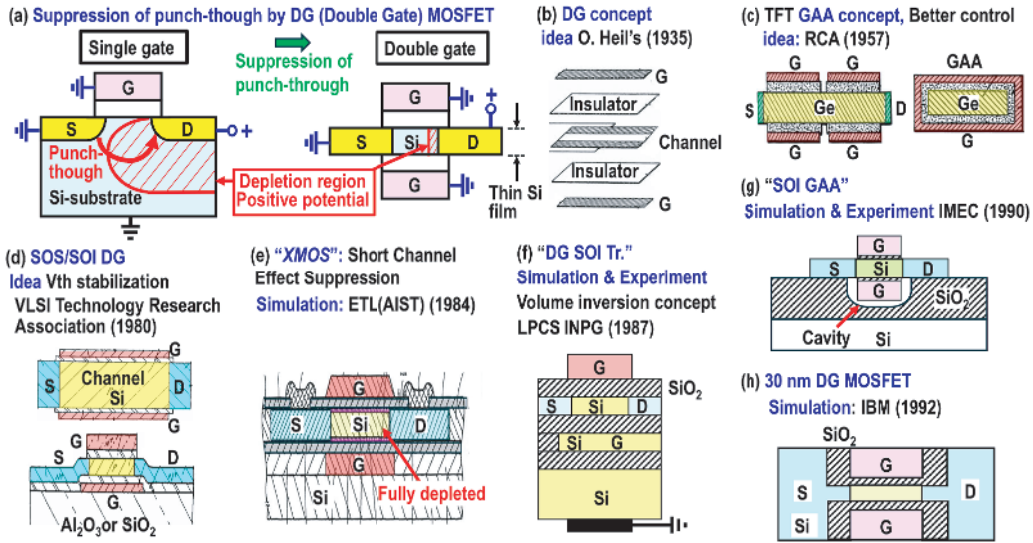


Figure 1.74 Punch-through current and its suppression by horizontal double (multiple) gate structures [33, 34, 415–420]. *Source:* (f) Adapted from [418], (g) Adapted from [419], and (h) Adapted from [420].

the gate length decreases, the depletion region near the drain approaches the source and eventually makes contact with it. Since the potential in the depletion region is positive (with the source grounded), electrons are injected from the source and flow toward the drain, which has the highest positive potential, as shown in Figure 1.74a. This phenomenon is known as punch-through. The resulting leakage current flows not near the silicon surface but through deeper regions, because the surface potential is well controlled to 0 V by the gate.

Traditional countermeasures against punch-through involved reducing both the supply voltage and the gate oxide thickness, and increasing well or pocket doping. Lowering the supply voltage narrows the depletion region, while thinning the gate oxide enhances the electric field, improving surface potential control via gate bias—an ideal scaling scenario. However, as discussed later, further reductions in supply voltage and gate oxide thickness have become increasingly difficult. Moreover, the substrate impurity concentration—particularly the high-concentration “halo” regions formed near the drain to suppress the depletion layer—has already reached approximately $10^{18}/\text{cm}^3$. Further increases in concentration are impractical, as they reduce the breakdown voltage of the drain p – n junction [421].

1.13.3.2 Multiple Gate Structure to Suppress Punch-Through

As an alternative, a new structural approach shown in Figure 1.74a has been adopted. The concept involves eliminating the deep silicon regions through which punch-through current flows and instead using a thin silicon film as the channel. The channel current is controlled not only by the top gate electrode but also by a bottom gate electrode placed beneath the film. This structure is known as a double-gate (DG) MOSFET, a type of multigate MOSFET as shown in Figure 1.74. It enables effective control of the entire silicon body potential to 0 V, significantly suppressing the depletion region.

A historical review of lateral DG MOSFETs shows that the basic structure had already been disclosed in a 1934 patent (Figure 1.74b), although its functional advantages were not described at

that time [33]. In 1957, RCA presented a lateral TFT featuring a GAA configuration (Figure 1.74c), explicitly noting that this GAA structure provides superior electrostatic controllability [34]. This is likely the earliest proposal of a GAA-type device.

In 1980, device concepts based on silicon-on-insulator (SOI) and SOS substrates had introduced architectures incorporating a second gate electrode beneath the channel to stabilize the threshold voltage of the floating channel. The first of these structures was proposed as a patent application in January by the VLSI Technology Research Association (Figure 1.74d) [415], followed by a similar patent filed by Suwa Seikosha Co. Ltd. (now Seiko Epson) in November of the same year [416]. These designs, however, did not constitute GAA structures; the top and bottom gate electrodes were not self-aligned but were instead fabricated using separate lithographic masks. In 1984, the Electrotechnical Laboratory of AIST demonstrated through numerical simulation that an SOI-based DG structure is highly effective in suppressing SCEs (Figure 1.74e) [417, 422]. In this structure, the silicon channel was made extremely thin to ensure full depletion, thereby enhancing gate controllability. (It should be noted that the concept of a fully depleted channel in a DG configuration had earlier been proposed in 1980 in the first FinFET patent by AIST, as shown later in Figure 1.77a).

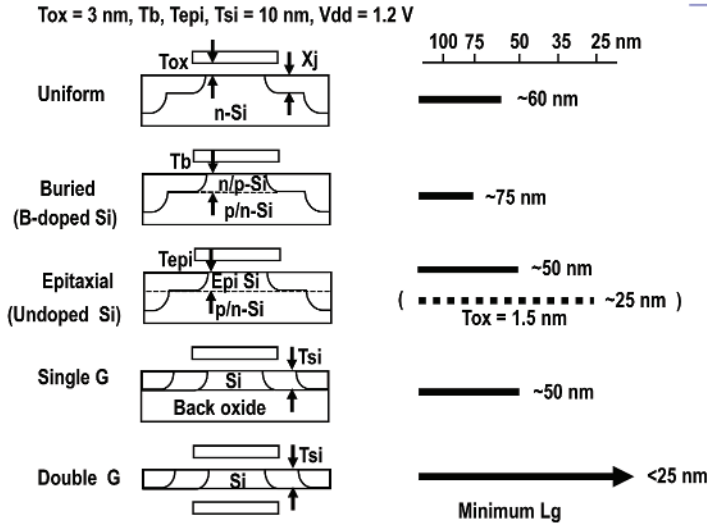
In 1987, LPCS (Laboratoire de Physique des Composants à Semiconducteurs) INPG (Institut National Polytechnique de Grenoble), France, fabricated and characterized what is likely the first DG MOSFET using an SOI structure (Figure 1.74f) [418] formed by the SIMOX (Separation by IMplanted OXygen) process [423, 424]. Simulations revealed that an inversion region forms not only at the top and bottom interfaces but also within the silicon film between them, resulting in full-film inversion. Both simulation and experimental results showed that the drain current and transconductance increased to approximately three times those of a single-gate device—exceeding the simple doubling expected from two channels. This phenomenon was termed volume inversion.

In 1989, IMEC fabricated the first lateral GAA-type MOSFET, experimentally confirming—consistent with the results of LPCS INPG in 1987—that volume inversion yields roughly a threefold enhancement in drive current and transconductance (Figure 1.74g) [419, 425]. (The earliest GAAFET fabrication, however, was Toshiba’s vertical transistor in 1989 [426], as shown later in Figure 1.76d [427]).

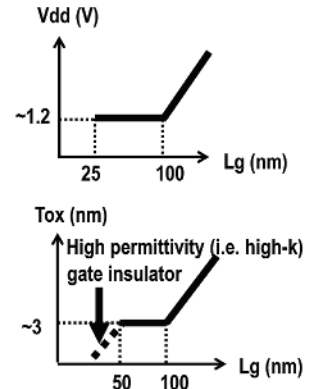
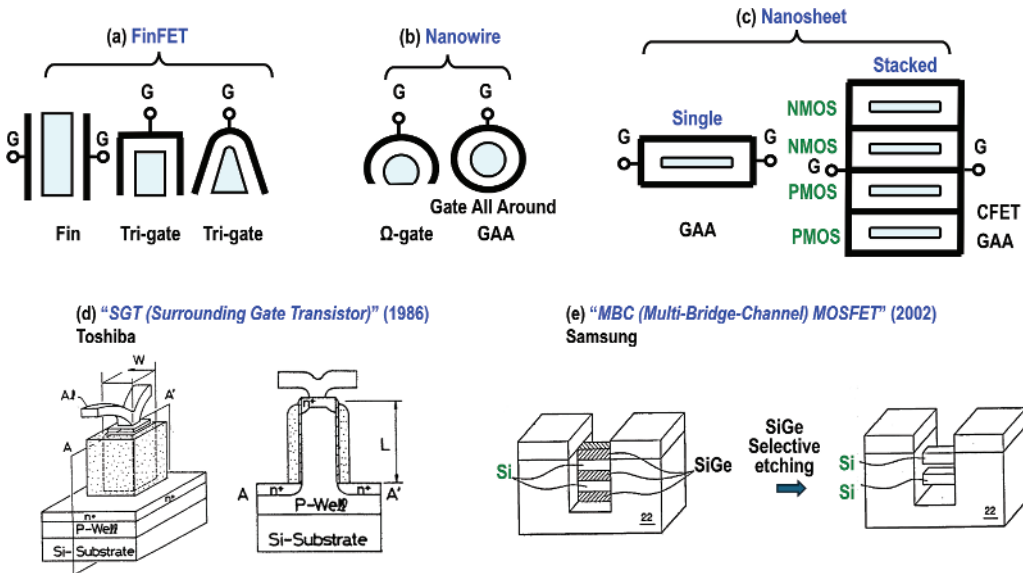
In 1992, IBM used Monte Carlo simulations to show that while conventional bulk MOSFETs faced scaling limits around 80 nm gate length, double-gate FETs could scale down to the 30 nm range with superior performance (Figure 1.74h) [420]. In 1993, a joint study by Toshiba and the University of Bologna confirmed via Monte Carlo simulation that double-gate MOSFETs with a 10 nm Si channel thickness and 3 nm gate oxide thickness could operate reliably at room temperature with gate lengths below 25 nm, as shown in Figure 1.75a [428]. In contrast, conventional bulk MOSFETs faced scaling limits around 60 nm. Retrograde channel MOSFETs—formed by growing undoped 10 nm epitaxial layers on heavily doped substrates—could scale down to 50 nm. By using high- k dielectrics to reduce the gate oxide’s EOT to 1.5 nm, operation was confirmed at gate lengths down to 25 nm. These results demonstrated the desirability of double-gate MOSFETs for gate lengths under 25 nm and of high- k gate dielectrics for gate lengths under 50 nm. They were subsequently published in 1993 as “A New Scaling Methodology,” as illustrated in Figure 1.75a and 1.75b [428]. While minor discrepancies exist, these scaling predictions broadly align with the historical trajectory of MOSFET miniaturization.

As shown in Figure 1.76e, Samsung invented a GAA horizontal nanosheet-stacked MOSFET structure—known as the “Multi-Bridge-Channel (MBC) MOSFET”—in 2002 [429]. The company subsequently published experimental results in 2003 [430] and 2004 [431], demonstrating the feasibility of a mass-producible process. In 2020, Leti successfully demonstrated a seven-layer stacked nanosheet GAAFET [432]. That same year, complementary FETs (CFETs), featuring

(a) Gate length limit for various MOSFET structures



(b) Roadmap for supply voltage and gate oxide thickness

**Figure 1.75** New scaling methodology for 100 to 25 nm MOSFETs. *Source:* Adapted from [428].**Figure 1.76** Various types of multigate MOSFETs [427, 429].

vertically stacked p-channel and n-channel nanosheet GAA FETs, were introduced by Intel to reduce CMOS circuit footprint [433]. This innovation led to a 50% reduction in SRAM cell area. In 2024, companies such as TSMC published CFETs as their future logic technology [434].

It is expected that the commercialization of CFETs will take more years. In contrast, for non-CFET nanosheet GAAFETs, Samsung announced the start of 3-nm production in 2022 [435], followed by TSMC's announcement of 2-nm production in 2025 [436]. From 2026 onward, full-scale deployment of three-stack nanosheet-channel GAA structures at the 2-nm generation

is anticipated at major foundries, such as Samsung, TSMC, and Intel. In GAAFETs, the scaling limit of gate length is collectively determined by the limits of supply voltage, gate oxide thickness, and nanosheet thickness reduction. At present, the practical limit is considered to be around a 10 nm gate length. The supply voltage and the scaling limit of gate oxide thickness are estimated to be approximately 0.7–0.65 V (see Section 1.13.3.4) and about 0.7 nm (see Section 1.13.3.5), respectively. The limit of nanosheet thickness reduction has been thought to be around 0.4 nm, governed by how effectively degradation in nanosheet surface flatness, mobility, and channel carrier density can be suppressed. In 2024, Intel conducted prototype experiments on GAAFETs by varying nanosheet thickness from 6 nm down to 1.5 nm, and demonstrated the operation of a MOSFET with a 6 nm gate length at a nanosheet thickness of 3 nm [437]. Sub-10 nm gate-length GAAFETs are not necessarily guaranteed to deliver performance improvements, and continued monitoring will be necessary.

Today, the term GAA commonly refers to nanosheet MOSFETs in which the gate electrode completely surrounds the channel, as exemplified by Samsung's MBC MOSFET. Originally, GAA was used in a broader sense to refer to those among the many multigate structures shown in Figure 1.76b–c, in which the gate completely encloses the channel. The first proposal of a GAA structure was Toshiba's "Surrounding Gate Transistor (SGT)," as shown in Figure 1.76d [427]. The experimental results, which featured a channel length of 0.8 μm , were demonstrated in 1989 [426]. This device employed a vertical pillar-type channel, with current flowing vertically along the pillar sidewalls. The pillar cross-section measured approximately $1 \times 1 \mu\text{m}$.

Among vertical channel architectures, the FinFET is widely recognized and is characterized by lateral current flow along the fin sidewalls. Figure 1.77 presents the milestones from the invention of FinFET (under various names) to its commercialization in 2012 [394, 396]. The term "FinFET" was coined around 2000 by researchers at the University of California, Berkeley, and has since become a generic name for this class of devices. Similar structures had been invented and experimentally studied under different names prior to this. The earliest known FinFET structure is attributed to AIST in 1980 (Figure 1.77a) [438], featuring a bulk-type configuration with the fin connected to the substrate. As in X MOS shown in Figure 1.74(e) [417], its primary objective was to suppress SCEs, and examples of fully depleted channels were demonstrated. In 1987, Fujitsu (Figure 1.77b) and Toshiba/AIST (inventor is Toshiba) (Figure 1.77c) independently filed patents for FinFET structures more closely resembling modern implementations. Fujitsu filed its patent on July 1, 1987 [439], while Toshiba/AIST filed theirs on July 23, 1987 [440], with only a slight difference in timing. Fujitsu's patent described both bulk-type fins integrated with the substrate and fins isolated by oxide layers. In contrast, AIST (with Toshiba as the inventor) proposed only SOI-type devices, introducing a fully depleted channel concept.

In 1988, Hitachi filed a similar SOI-type FinFET patent (Figure 1.77d) [441] and became the first to publicly present prototype and evaluation results—under the name "DELTA"—at IEDM in 1989 [442]. This marked the first experimental disclosure of a FinFET, ten years before the name was coined. In 2000, UC Berkeley filed a patent for the FinFET structure with sidewalls on an SOI substrate (Figure 1.77e) [443], and jointly published experimental results with the gate length down to 17 nm with Hitachi [444], formally introducing the term FinFET. In 2003, Kyungpook National University filed a patent for a body-tied FinFET (Figure 1.77f) [445]. The device features etched fins connected to the silicon substrate, and corresponding experimental results were published jointly with Seoul National University as the "OMEGA MOSFET" [446], marking the first demonstration of a bulk FinFET. In 2011, Intel introduced a bulk-type FinFET ("Tri-Gate transistor") (Figure 1.77g) [396] and began mass production using 22 nm CMOS technology in 2012 [394]. This marked a dramatic shift from the planar MOSFET architecture that had dominated for over 50 years, attracting global attention.

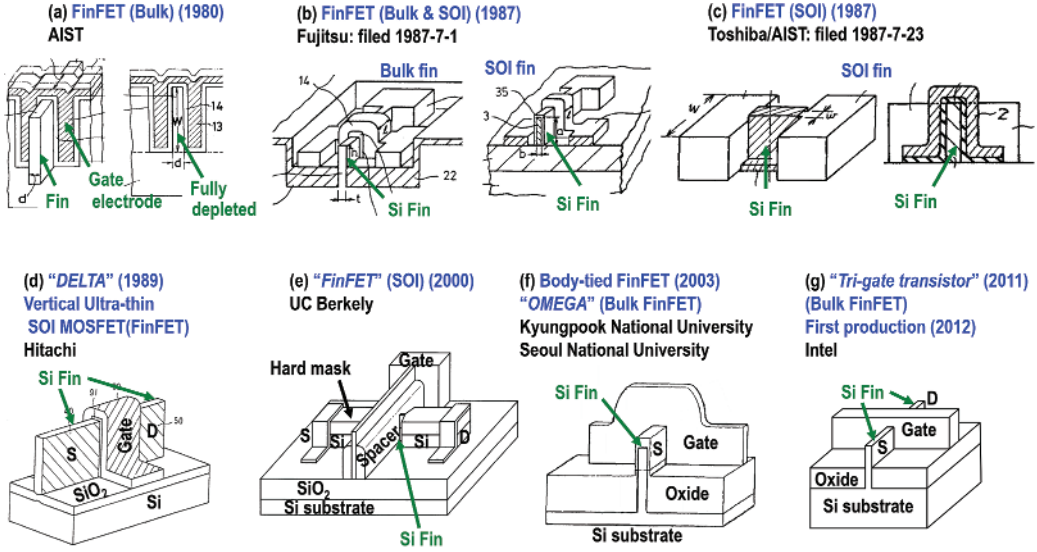


Figure 1.77 Milestones for FinFET evolution [394, 396, 438–441, 443, 445].

FinFETs effectively suppressed SCEs such as punch-through, contributing significantly to nano-CMOS scaling for over a decade. However, increasing the channel width required taller fins, which introduced challenges such as width variation between the top and bottom of the fin during etching and sidewall damage. These limitations prompted the development of horizontal nanosheet GAAFETs, which—as noted above—are expected to replace FinFETs by around 2025.

1.13.3.3 Direct-Tunneling Current

In MOSFETs with channel lengths shorter than 3 nm, direct-tunneling current inevitably arises between the source and drain, as discussed in Section 1.13.1. Since this phenomenon cannot be suppressed, it defines the fundamental limit of gate-length scaling. Compared to the currently accepted practical limit of approximately 10 nm, the fundamental limit is about one-third of the practical one.

1.13.3.4 Subthreshold Leakage Current

As illustrated in Figure 1.78, subthreshold current refers to the leakage current that occurs when the gate voltage V_g is below the threshold voltage V_{th} . Ideally, no current should flow in this regime. However, in practice, subthreshold current exists and increases exponentially with gate voltage. In the subthreshold region, no strong inversion layer is formed. Instead, electrons injected from the source (in the case of an n-MOSFET) diffuse through the channel at the Si surface and reach the drain. In this regime, the transistor operates not as an FET, but as a JT, where the *pn* junction between the source and channel acts as the emitter and base. As shown in Figure 1.79, subthreshold current flows through this junction barrier. Therefore, the electron flow from source to drain in the subthreshold region resembles the flow from emitter to collector in an *npn* BJT. A key distinction, however, is that the channel (or base) potential is capacitively controlled by the gate voltage. As a result, the channel potential becomes a fraction of the gate potential, expressed as V_g/m , where $m = C_{ox}/(C_{ox} + C_D + C_I)$ (see Figure 1.79 for definitions of capacitances). Accordingly, the subthreshold leakage current is proportional to $\exp(qV_g/mkT)$.

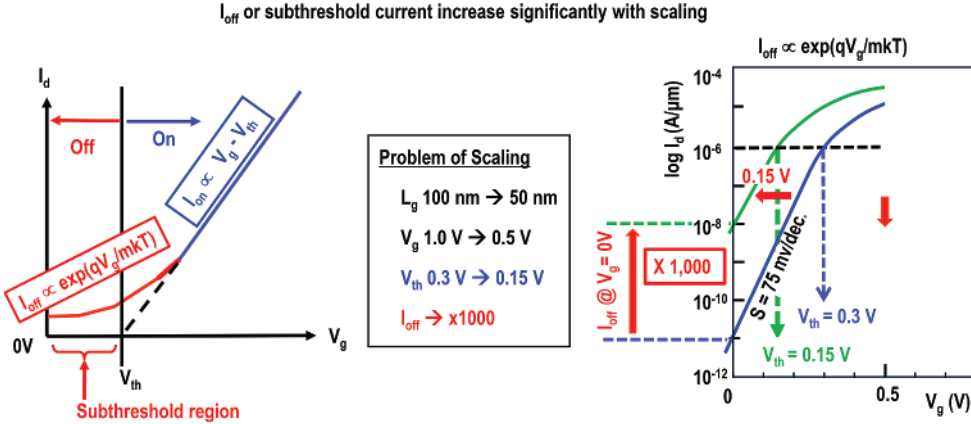


Figure 1.78 Significant increase of I_{off} or subthreshold current with scaling.

In the subthreshold regime (weak inversion), the carrier supply from the source to the channel governs the drain current.

$\rightarrow$ the MOSFET behaves as a junction transistor (npn BJT for NMOS).

$\rightarrow$ equivalent base potential (= channel potential) is $\phi_s(V_g) \rightarrow$ therefore, $I_D \propto \exp(qV_g/mkT)$

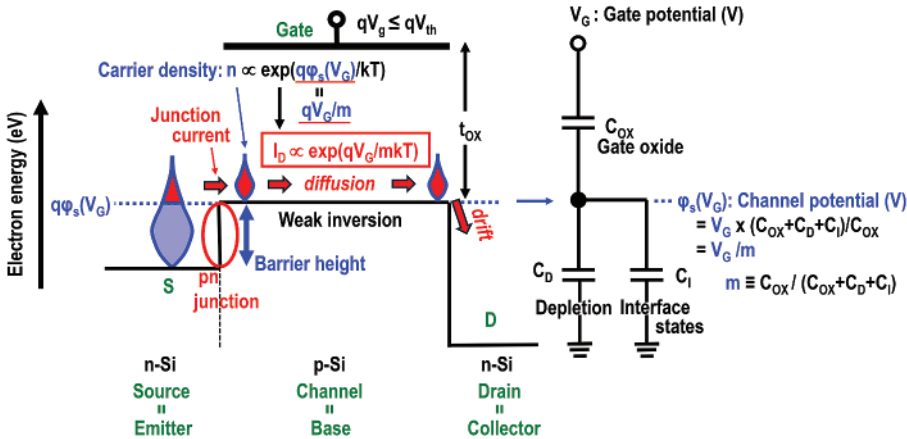


Figure 1.79 Mechanism of subthreshold increase by lowering threshold voltage.

Subthreshold leakage current is highly sensitive to the threshold voltage. As the I_d - V_g plot in Figure 1.78 shows, scaling the threshold voltage from 0.3 V to 0.15 V increases the off-state leakage current at $V_g = 0$ V by a factor of 1000. Consequently, it has become increasingly difficult to lower the threshold voltage in accordance with conventional scaling rules, and further reduction of the supply voltage is also constrained. This limitation has become a major factor in restricting continued device miniaturization.

Cryogenic operation of MOSFET can suppress the subthreshold leakage and improve the subthreshold swing (SS) and hence allows for further scaling of V_{th} and V_{DD} [447, 448]. Cryogenic CMOS co-processors are also important for future quantum computing [449].

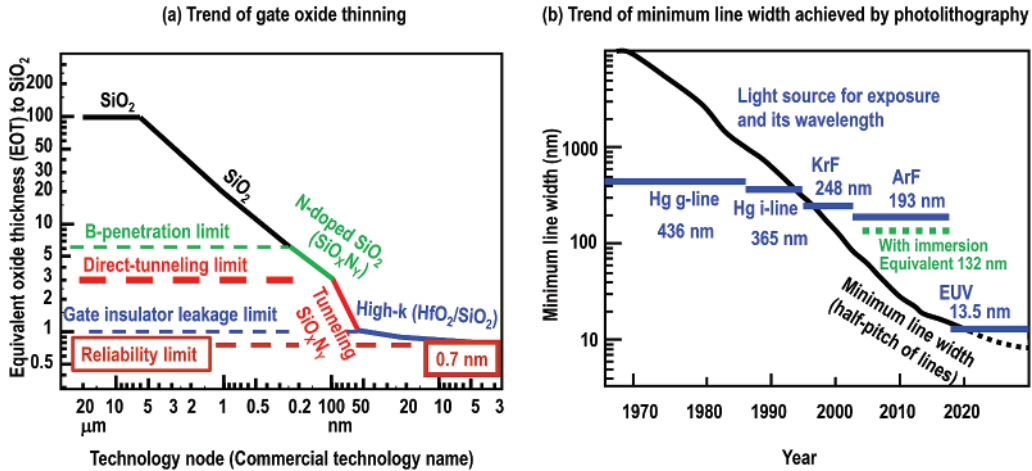


Figure 1.80 Trend of reduction in gate oxide thickness (a) and minimum line width determined by photolithography (b).

1.13.3.5 Gate Leakage Current

Figure 1.80a illustrates the scaling trend of gate dielectric thickness. From the first MOSFET in 1960 through early LSIs, the SiO₂ film stayed at 100 nm due to immature processes. By the late 1970s, when gate lengths reached $\sim 4 \mu\text{m}$, thinning began. In the 0.25–0.18 μm generations (late 1990s–2000), thickness was reduced to 4–5 nm, and SiO₂ was replaced by SiO_xN_y to block boron penetration. In the sub-100 nm era, direct-tunneling SiO_xN_y films under 3 nm were used. For direct-tunneling-type SiO_xN_y gate dielectrics, the physical thickness limit is approximately 1.2–1.3 nm [386] as described in Section 1.11.2. When the thickness is reduced beyond this threshold, a significant portion of the drain current is diverted toward the gate before reaching the drain terminal, resulting in a degradation of drain current. To suppress such tunneling leakage while maintaining the same EOT, high-*k* gate dielectrics—characterized by higher permittivity and thus allowing physically thicker layers—became essential. This approach was successfully implemented by Intel in its 45 nm logic CMOS technology [450]. Notably, pioneering studies on high-*k* gate insulators (such as ZrO₂) for Si-MOSFETs had already been reported by the University of Tokyo in the late 1960s [451, 452].

Intel achieved favorable interface characteristics by inserting a thin SiO₂ interfacial layer (IL) between HfO₂ and the silicon substrate. This stacked structure (HfO₂/SiO₂) enabled an EOT of 1 nm as early as 2007. Subsequent efforts focused on eliminating the SiO₂ IL (~ 0.15 nm) while preserving acceptable interface quality, prompting the search for alternative high-*k* materials. Technologies were developed that achieved an EOT of approximately 0.5 nm without an IL while keeping gate leakage current within tolerable limits. However, the prevailing consensus is that the IL remains indispensable for ensuring carrier mobility, minimizing leakage, and securing long-term reliability. Accordingly, the practical lower bound of EOT is considered to be around 0.7–0.8 nm [453, 454]. This constraint has emerged as one of the limiting factors in scaling gate lengths below approximately 10 nm.

1.13.4 Issues: Lithography, Cost, Heat, and Yield

Figure 1.80b illustrates the historical evolution of photolithographic resolution. Achieving higher resolution requires shorter wavelengths and larger numerical apertures (NA).

Resolution enhancement techniques (RETs) [455], including Optical Proximity Correction (OPC), Phase-Shift Mask (PSM), Off-Axis Illumination (OAI), Inverse Lithography Technology (ILT), and multi-patterning, have enabled sub-wavelength patterning, though such fine features may introduce subtle distortions. EUV lithography [456], first fully implemented at the 5 nm node, operates at a wavelength of 13.5 nm and is expected to support sub-10 nm patterning with further advances. The escalating cost of advanced fabrication equipment and production lines is another major concern. For instance, next-generation High-NA EUV tools are estimated to cost several hundred million USD per unit, and the establishment of a 2 nm logic IC production line is projected to require tens of billions of USD. Driven by surging semiconductor demand—particularly from AI/data center and smartphone applications—and by increasing market consolidation, major foundries have achieved substantial profitability. As a result, despite the high cost of photolithography tools, the prevailing business model remains economically viable. However, the soaring expense of next-generation equipment is an issue, and a key strategic decision lies in determining the most efficient node at which to adopt it, with manufacturability in mind. Currently, a typical sub-10 nm node CMOS process is estimated to require approximately 70 photolithographic masking steps. However, with the advancement of 3D integration technologies such as multilayer nanosheet stacking, this number is expected to increase by double or even triple. The resulting increase in manufacturing costs and the challenges of heat dissipation may become critical limiting factors for further 3D integration. For further reading on lithography, see Chapter 16 in this book.

CMOS VLSIs at the 3 and 2 nm nodes exhibit highly complex structures. Consequently, achieving acceptably high yield becomes increasingly difficult with each successive generation. In this context, yield and reliability have become decisive factors for the survival of semiconductor manufacturers. Nodes beyond 2 nm are currently under development, and progress is expected to proceed as planned. Future node advancements are likely to be more incremental, with significantly longer development cycles. On the other hand, the increasing use of more advanced AI in semiconductor modeling, material search, process and device R&D, yield improvement, manufacturing, and design may help us keep up the pace.

1.14 Future Prospects

1.14.1 Prospects for the Near Future

From the perspective of physical linewidth, device miniaturization is approaching its limit in the 10 nm to sub-10 nm range, and practical minimum feature size scaling is expected to end around the sub-2 nm technology node. Although node naming continues to advance, the actual pace of feature size reduction has been slowing. Nevertheless, efforts to push miniaturization forward, even marginally, are expected to persist.

Vertical integration via 3D architectures is advancing, though thermal issues [457–460], process complexity, and cost will impose limits. Still, 3D integration is in its early phase, and ongoing improvements in design and fabrication are expected. As demand grows for application-specific processors, custom 3D architectures will become increasingly vital.

3D heterogeneous integration [461–464] has proven capable of complementing 3D monolithic integration to provide additional functionality, density, performance, energy efficiency, system miniaturization, time to market, etc., since the mid-2010s. Such system-level scaling has been evident and prevalent in AI factories, edge AI, and smartphones. It unleashes new

opportunities and allows more time for disruptive breakthroughs in future semiconductor and CMOS technologies.

Since late 2023, the AI industry has entered a rapid growth phase, driving demand for larger and higher-performance data centers. This has accelerated integration density and processing speed, making power consumption and heat generation urgent challenges. In AI, GPUs and tensor processing unit (TPU)—optimized for parallel matrix operations—are gaining importance, along with high-speed, high-capacity, and high-bandwidth memory and fast GPU access. In-memory computing, which performs calculations within memory arrays, is also emerging as a promising approach for energy-efficient AI computing. Even if physical and 3D scaling reach their limits, demand for functional expansion in ICs will continue, driven by a more intelligent society. The semiconductor industry remains essential for human existence, with diversification into power devices, sensors, and high-frequency components based on silicon and compound semiconductors being indispensable.

1.14.2 Looking More than 30 Years Ahead

What might the world look like in the latter half of the 21st century—or even into the 22nd century? It is almost certain that semiconductor devices will continue to play a vital role. However, integration between semiconductors and biological systems is expected to become increasingly important. Neurons operate with pulses under 0.1 V and below 1 kHz, achieving very low power use, as shown in Figure 1.81 [465, 466]. When inactive, they show no leakage, and each can connect to over a thousand others, forming complex, reconfigurable circuits. The brain's total power consumption is only several tens of watts—far beyond what conventional FETs can match. By contrast, MOSFETs cannot self-generate pulses, operate at 0.1 V, or interconnect like neurons. They require at least 0.65–0.7 V and gigahertz operation for high performance. Thus, AI systems built on them may consume power comparable to a small power plant.

The human brain cannot maintain concentration on a single task for long periods or store large volumes of information accurately over time. As a result, AI systems already surpass human performance in tasks such as playing shogi and chess. Nevertheless, the brain consumes only around 30 W. Despite some ambiguity in its operations, it can handle a wide range of tasks in parallel and offers a functional lifespan approaching 100 years.

It is fascinating that, throughout billions of years of evolution, animals have never developed brain cells based on semiconductors or nervous systems composed of metals. In fields such as

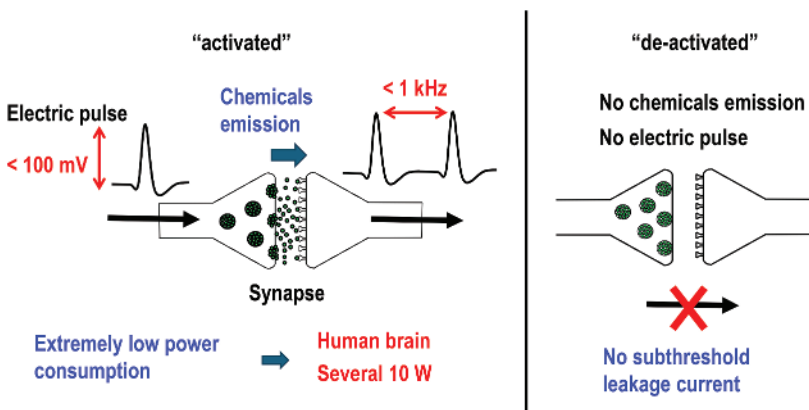


Figure 1.81 Efficient operation of neuron circuits.

beekeeping, shepherding, drug detection, and horse racing, replacing bees, dogs, or horses with AI devices currently offers limited advantages. Rather, a new paradigm may emerge in which semiconductor-based AI collaborates with insects, fish, birds, and mammals—potentially unlocking entirely new possibilities. While the careful evaluation of risks associated with AI remains indispensable, the 22nd century is poised to be defined by the profound integration of AI-enabled semiconductors and biological systems, unleashing the full realization of biological potential. This convergence may well signify the fourth technological revolution in Information and Communication Technology (ICT), following the emergence of language, writing, and electronics.

1.15 Summary and Concluding Remarks

1.15.1 Summary

This chapter traces the century-long development from the invention of the FET to the realization of the silicon MOSFET, and subsequently surveys the evolution of MOS LSI up to the present day. At present, LSI constitutes the largest market segment among FET-based devices.

The FET was invented in 1925 as a solid-state amplifier intended to replace vacuum tube amplifiers in wireless communication receivers. Unlike vacuum tubes, however, solids contain numerous free electrons and fixed charges, and at that time, no technology existed to control semiconductor surfaces. As a result, it required 35 years until 1960, when the operation of the MOSFET—utilizing the inversion layer at the surface of a silicon single crystal as the channel—was experimentally confirmed. Key milestones during this period included the establishment of band theory using quantum mechanics in the 1930s and the consequent discovery of the inversion layer; the use of silicon rectifiers for military radar in the 1940s, which necessitated the purification of silicon to high levels; the discovery of JT operation and the invention of the BJT at Bell Laboratories in the late 1940s; and the commercialization of BJTs in the 1950s, which stimulated advances in doping techniques and lithography.

Subsequently, the MOSFET required another decade to overcome operational instabilities caused by interface charges. Around 1969, commercial MOS LSIs emerged, enabling high-density integration for calculators, personal computers, and memory devices. Systematic MOSFET scaling then progressed, doubling integration density every few years and driving improvements in IC performance. From the 1970s through the 1990s, numerous semiconductor companies (e.g., IBM, Intel, TI, Toshiba, Hitachi, NEC, Mitsubishi Electric, Fujitsu, Siemens, and Philips), academic institutions, and equipment manufacturers played leading roles in VLSI innovation.

During this period, new process technologies such as dry processing, RTP, and stepper photolithography were introduced to address miniaturization challenges, along with new materials, including salicide, SiO_xN_y gate oxides, and copper interconnects. In particular, RTP proved indispensable not only for activating ultra-shallow diffusion layers but also for forming gate oxides of less than a few nanometers and for enabling the salicide process.

Since the 2000s, the development has become increasingly concentrated in major semiconductor manufacturers (e.g., Intel, Samsung, and TSMC), equipment suppliers (e.g., Applied Materials and TEL), and research institutions such as IBM and IMEC. By the 1990s, sub-50 nm devices became feasible, and by 2025, feature sizes reached ~ 10 nm—near the physical limit. Memory cells shrank from visible scale to virus-sized. Vertical scaling introduced GAAFETs, and MOS LSIs now outperform BJTs, powering everything from games to AI.

Driven by breakthroughs and application needs, especially AI and data centers, LSI technology continues to advance despite rising power density and costs. The demand for performance and

energy efficiency will continue to drive innovation in the decades to come. The evolution of FET technology over the past century has followed an inevitable course: despite numerous challenges and detours, it has ultimately become deeply embedded in society. While many technologies have faded or surrendered their leading role, the MOSFET has endured as a remarkable exception, securing its place as an indispensable foundation for human civilization. The broad trajectory of progress over the coming decades can be foreseen to some degree, yet the future beyond that horizon remains opaque. Through convergence with as-yet-unknown innovations, FET technology will continue to carve out new horizons, opening pathways toward futures once unimaginable.

1.15.2 Concluding Remarks

A hundred years is certainly a long span of time, yet more than three-quarters of that period overlaps with the author's own lifetime, and over half of it has been spent directly engaged in FET research and development. In that sense, this is not merely a historical account, but rather a chronicle of a life lived alongside the history itself. Over these hundred years, countless papers and patents have been published on FET research and technology. However, reading individual documents alone does not easily allow one to understand why a certain technology emerged at a particular moment, why its development required patience and time, or why progress often took a winding, indirect path rather than a shortcut. In this chapter, the author has tried, as much as possible, to shed light on such aspects that only someone who lived through the process might have noticed. In the past few years, we have lost several towering figures of MOS IC history—Gordon Moore, Robert Dennard, Simon Sze, George Smith, and Chih-Tang Sah. With their passing, there are now few left who can speak directly to those formative years. Thus, the author feels deeply grateful to have been granted the opportunity, while still alive, to recount more than half a century of MOS LSI evolution from the perspective of one who witnessed it firsthand.

Acknowledgment

The author wishes to express his sincere gratitude to the editors of this volume for the opportunity to contribute to this chapter and for their appropriate advice on the presentation of the text. Deep appreciation is also extended to the reviewers, who read the manuscript multiple times, corrected errors, and provided invaluable advice and information. The author is likewise indebted to many colleagues and senior associates at Toshiba, who have long been engaged in the development of LSI technology. Over more than 50 years, exchanges and discussions with colleagues at Toshiba, as well as with members of the semiconductor industry, academia, and professional societies, have greatly enriched this chapter, for which the author is deeply grateful.

Special thanks are due to the staff of the Institute of Science Tokyo Library for their kind assistance in obtaining old references—unavailable online—from libraries in Japan and abroad.

Due to space constraints, the following topics are not included: point-contact metal-semiconductor rectifiers before 1925 [467–469]; an early MESFET patent in 1929 [470] and a KBr crystal MESFET operating at 490°C with sub-Hz [471]; the development of the concept of the hole [472], except for those of W. Heisenberg [38] and A. Wilson [39, 40] (Ref. [472] is among the first papers to suggest the hole); and semiconductor band theory [473–483], except for the Kronig–Penney model [43] and A. Wilson [39, 40]. (Ref. [477] was the first to introduce the concept of the Fermi level in solid-state physics; Ref. [478]—known as “Bloch’s theorem”—contributed to the Kronig–Penney model; and Refs. [480–483] present band models that employed actual

electron orbits of atoms for electron energy calculations.); the 1938 elucidation of the rectification mechanism carried out independently by B. Davydov [484, 485] and N. Mott [486, 487], except that of W. Schottky [41, 42]; SOI technologies, including the PowerPC microprocessor series co-developed by IBM and Motorola and supplied to Apple [488–490], as well as Toshiba’s custom SOI-based logic LSI, the Emotion Engine, developed for Sony’s PlayStation 2 [491, 492]; the development of memory technologies since the mid-1980s; and compound semiconductor LSIs.

With regard to technological developments after 2000, the remarkable rise of semiconductor manufacturers in Asian countries outside Japan is acknowledged; however, as these pertain to current technologies and are expected to be discussed in detail in other chapters of this book, the description here has been kept brief. Likewise, no detailed account of semiconductor development in Europe [493, 494] is provided.

Finally, in describing the history of technological development, individual names have been included up to the 1960s. From the 1970s onward, as the number of engineers involved in each development increased substantially, only company or institutional names have been cited.

References

- 1 Lilienfeld, J.E. (1925). Method and apparatus for controlling electric currents, Canadian Patent CA272437A, filed October 22, 1925.
- 2 Shockley, W. (1952). A unipolar "field-effect" transistor. *Proceedings of the IRE* 40: 1365–1376.
- 3 Dennard, R.H., Gaensslen, F.H., Yu, H.-N. et al. (1974). Design of ion-implanted MOSFET’s with very small physical dimensions. *IEEE J. Solid-State Circuits* 9: 256–268.
- 4 Lilienfeld, J.E. (1928). Device for controlling electric current. US Patent 1900018, filed March 28, 1928.
- 5 De Forest, L. (1907). Space telegraphy. US Patent 879532, filed January 29, 1907.
- 6 Kahng, D. and Atalla, M.M. (1960). Silicon-silicon dioxide field induced surface devices. *IRE-AIEEE Solid-State Device Research Conference*, Carnegie Institute of Technology, Pittsburgh, PA; Kahng, D. (1961). Silicon-silicon dioxide field induced surface devices. Technical memorandum issued by Bell Labs., reprinted in Sze, S. M. (1991). *Semiconductor Devices: Pioneering Papers*. World Scientific Publishing Co., pp. 583–596. <https://doi.org/10.1142/1087>.
- 7 Shockley, W. (1950). Electrons and holes in semiconductors: with applications to transistor electronics, van Nostrand.
- 8 Smits, F.M. (ed.) (1985). *A History of Engineering & Science in the Bell System: Electronics Technology (1925–1975)*. Bell Telephone Laboratories.
- 9 Kubo, S. (1985). Technological history of transistors and integrated circuits. Ohm, 1985 [in Japanese]. 久保脩治「トランジスタ・集積回路の技術史」オーム社、1985年. <https://doi.org/10.1142/1087>.
- 10 Sze, S. (1991). *Semiconductor Devices: Pioneering Papers*. World Scientific Publishing Co. <https://doi.org/10.1142/1087>.
- 11 Aida, M. (1998). *Electronic Nation Japan’s Autobiography, vol. 1 ~ 4, 1991 ~ 1992*. NHK Publishing [in Japanese]. 相田洋「NHK電子立国日本の自叙伝」日本放送出版協会.
- 12 Hoddeson, L., Braun, E., Teichmann, J., and Weart, S. (1992). *Out of the Crystal Maze: Chapters from The History of Solid State Physics*. Oxford University Press.
- 13 Riordan, M. and Hoddeson, L. (1997). *Crystal Fire: The Invention of the Transistor and the Birth of the Information Age*. W. W. Norton & Company.

- 14 Seitz, F. and Einspruch, N.G. (1998). *Electronic Genie: The Tangled History of Silicon*. University of Illinois Press.
- 15 Bassett, R.K. (2002). *To the Digital Age: Research Labs, Start-Up Companies and the Rise of MOS Technology*, *Johns Hopkins Studies in the History of Technology*. Johns Hopkins University Press.
- 16 Lojek, B. (2007). *History of Semiconductor Engineering*. Springer.
- 17 Bauer, L.O. and Wilder, E.M. (2020). *The Microchip Revolution: A Brief History*, independently published.
- 18 CHM (Computer History Museum). <https://computerhistory.org>.
- 19 The Old Calculator Web Museum. <https://www.oldcalculatormuseum.com/#gsc.tab=0>.
- 20 Semiconductor History Museum of Japan. <https://www.shmj.or.jp/en/>.
- 21 Teal, G.K. (1976). Single crystals of germanium and silicon – Basic to the transistor and integrated circuit. *IEEE Trans. Electron Devices* 23: 621–639.
- 22 Shockley, W. (1976). The path to the conception of the junction transistor. *IEEE Trans. Electron Devices* 23: 597–620.
- 23 Sah, C.T. (1988). Evolution of the MOS transistor-from conception to VLSI. *Proceedings of the IEEE* 76: 1280–1326.
- 24 Iwai, H. (2022). History of transistor invention: 75th anniversary. *2022 IEEE 16th International Conference on Solid-State & Integrated Circuit Technology (ICSICT)*, pp. 1–10.
- 25 Iwai, H. (2023). The first MOSFET design by J. Lilienfeld and its long journey to implementation. In: *75th Anniversary of the Transistor* (ed. A. Nathan, S.K. Saha, and R.M. Todi), Ch. 7, 65–88. IEEE Press-Wiley.
- 26 Iwai, H. (2023). History of junction technologies. *2023 21st International Workshop on Junction Technology (IWJT)*, pp. 1–77.
- 27 Iwai, H. (2023). My LSI development experience at Toshiba (1973–1999): Hard Working, Exciting, and Splendid Days. *IEEE Electron Devices Magazine* 1: 41–57.
- 28 Iwai, H. (2023). Development of RFCMOS technologies in the 1990s in Toshiba. *2023 IEEE 15th International Conference on ASIC (ASICON)*, pp. 1–18.
- 29 Iwai, H. (2024). Past and future of integrated circuit technologies. In: *Outlooking Beyond Nanoelectronics and Nanosystems*, vol. 5 (ed. S. Deleonibus), 19–74. New York: Jenny Stanford Publishing.
- 30 Iwai, H. (2025). The centennial of FET and the silver jubilee of IWJT: Tracing the Origins and Evolution of Junction Technology in FETs. *Special Workshop on Celebrating IWJT's Silver Jubilee and the FET Centennial, 2025 22nd International Workshop on Junction Technology (IWJT)*, pp. 1–69.
- 31 Lepselter, M.P. and Sze, S.M. (1968). SB-IGFET: An insulated-gate field-effect transistor using Schottky barrier contacts for source and drain. *Proceedings of the IEEE* 56: 1400–1402.
- 32 Johnson, J.B. (1964). More on the solid-state amplifier and Dr. Lilienfeld. *Physics Today* 17: 60–62.
- 33 Heil, O. (1934). Improvements in or relating to Electrical Amplifiers and other Control Arrangements and Devices. British Patent GB439457 (A), convention date March 2, 1934 (Germany patent application: DEX439457), application date March 4, 1935, complete specification accepted December 6, 1935.
- 34 Wallmark, J.T. (1957). Field-effect transistor. US Patent 2900531, filed February 28, 1957.
- 35 Weimer, P.K. (1962). The TFT – A New Thin-Film Transistor. *Proceedings of the IRE* 50: 1462–1469.

- 36 Heisenberg, W. (1925). Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen (On the Quantum-Theoretical Reinterpretation of Kinematic and Mechanical Relations). *Zeitschrift für Physik* 33: 879–893.
- 37 Schrödinger, E. (1926). Quantisierung als Eigenwertproblem (Quantization as an Eigenvalue Problem). *Annalen der Physik* 384: 361–376.
- 38 Heisenberg, W. (1931). Zum Paulischen Ausschließungsprinzip. *Annalen der Physik* 402: 888–904.
- 39 Wilson, A.H. (1931). The theory of electronic semi-conductors. *Proceedings of the Royal Society of London* 133: 458–491.
- 40 Wilson, A.H. (1931). The theory of electronic semi-conductors II. *Proceedings of the Royal Society of London* 134: 277–287.
- 41 Schottky, W. (1938). Halbleitertheorie der Sperrschicht (Semiconductor Theory of the Blocking Layer). *Naturwissenschaften* 26: 843.
- 42 Schottky, W. and Spenke, E. (1939). Zur quantitativen Durchführung der Raumladungs- und Randschichttheorie der Kristallgleichrichter (Quantitative treatment of the space charge and boundary-layer theory of the crystal rectifier). In: *Wissenschaftliche Veröffentlichungen aus den Siemens-Werken: Achtzehnter Band*, vol. 18, 225–291. Springer Berlin Heidelberg.
- 43 Kronig, R. de L. and Penney, W.G. (1931). Quantum mechanics of electrons in crystal lattices. *Proceedings of the Royal Society of London* A130: 499–513.
- 44 Scaff, J.H. and Ohl, R.S. (1947). Development of silicon crystal rectifiers for microwave radar receivers. *Bell System Technical Journal* 26: 1–30.
- 45 Ohl, R.S. (1945). Point contact negative resistance devices. US Patent 2469569, filed March 2, 1945.
- 46 Ohl, R.S. (1941). Alternating current rectifier. US Patent 2402661, filed March 1, 1941.
- 47 Ohl, R.S. (1941). Light-sensitive electric device. US Patent 2402662, filed May 27, 1941.
- 48 Ohl, R.S. (1942). Thermoelectric device. US Patent 2402663, filed April 11, 1942.
- 49 Welker, H.J. (1945). Halbleiteranordnung zur kapazitiven Steuerung von Strömen in einem Halbleiterkristall (Semiconductor arrangement for capacitive control of currents in a semiconductor crystal). Deutsches Patentamt DE 980084, filed June 4, 1945, priority April 7, 1945.
- 50 Shockley, W. and Pearson, G.L. (1948). Modulation of conductance of thin films of semi-conductors by surface charges. *Physical Review* 74: 232–233.
- 51 Brown, W.L. and Shockley, W. (1953). n-type conduction on p-type germanium surfaces. *Physical Review* 90: 336. (Abstract in Proceedings of Annual Meeting of American Physical Society held at Harvard University, January 22–24, 1953.)
- 52 Brown, W.L. (1953). n-type surface conductivity on p-type germanium. *Physical Review* 91: 518–527.
- 53 Weiß, H. (1975). Steuerung von Elektronenströmen im Festkörper (Teil I) (Controlling Electron Currents in Solids (Part I)). *Physikalische Blätter* 31: 156–165.
- 54 Bardeen, J. (1947). Surface states and rectification at a metal semi-conductor contact. *Physical Review* 71: 717–727.
- 55 Bardeen, J. and Brattain, W.H. (1948). Three-electrode circuit element utilizing semiconductive materials. US Patent 2524035, filed June 17, 1948.
- 56 Morton, J.A. (1952). Present status of transistor development. *Bell System Technical Journal* 31: 411–442.
- 57 Shockley, W. (1948). Circuit element utilizing semiconductive material. US Patent 2569347, filed June 26, 1948.

- 58 Shockley, W. (1949). The theory of p-n junctions in semiconductors and p-n junction transistors. *Bell System Technical Journal* 28: 435–489.
- 59 Shockley, W., Sparks, M., and Teal, G.K. (1951). p-n Junction transistors. *Physical Review* 83: 151–162.
- 60 Dacey, G.C. and Ross, I.M. (1953). Unipolar “field-effect” transistor. *Proceedings of the IRE* 41: 970–979.
- 61 Atalla, M.M., Scheibner E.J., and Tannenbaum, E. (1958). Fabrication of semiconductor devices having stable surface characteristics. US Patent 2899344, filed April 30, 1958.
- 62 Atalla, M.M., Tannenbaum, E., and Scheibner, E.J. (1959). Stabilization of silicon surfaces by thermally grown oxides. *Bell System Technical Journal* 38: 749–783.
- 63 Hofstein, S.R. and Heiman, F.P. (1963). The silicon insulated-gate field-effect transistor. *Proceedings of the IEEE* 51: 1190–1202.
- 64 Goldberg, C., Hofstein, S.R., and Heiman, F.P. (1964). Pinch off in insulated-gate field-effect transistors. *Proceedings of the IEEE* 52: 414–415.
- 65 Ihantola, H.K.J. and Moll, J.L. (1964). Design theory of a surface field-effect transistor. *Solid-State Electronics* 7: 423–430.
- 66 Sah, C.T. (1964). Characteristics of the metal-oxide semiconductor transistors. *IEEE Trans. Electron Devices* 11: 324–345.
- 67 Pao, H.C. and Sah, C.T. (1966). Effects of diffusion current on characteristics of metal-oxide (insulator)-semiconductor transistors. *Solid-State Electronics* 9: 927–937.
- 68 Troutman, R. and Chakravarti, S. (1973). Subthreshold characteristics of insulated-gate field-effect transistors. *IEEE Trans. Circuit Theory* 20: 659–665.
- 69 Troutman, R.R. (1974). Subthreshold design considerations for insulator gate field-effect transistors. *IEEE Journal of Solid-State Circuits* 9: 55–60.
- 70 Czochralski, J. (1918). Ein neues Verfahren zur Messung der Kristallisationsgeschwindigkeit der Metalle. *Zeitschrift für Physikalische Chemie* 92: 219–221.
- 71 Little, J.B. and Teal, G.K. (1950). Production of germanium rods having longitudinal crystal boundaries. US Patent 2683676, filed January 13, 1950.
- 72 Sparks, M. and Teal, G.K. (1950). Method of making p-n junctions in semiconductor materials. US Patent 2631356, filed June 15, 1950.
- 73 Teal, G.K. (1950). Methods of producing semiconductive bodies. US Patent 2727840, filed June 15, 1950.
- 74 Fuller, C.S. (1952). Method of fabricating semiconductive bodies. US Patent 2725315, filed November 14, 1952.
- 75 Fuller, C.S. (1953). Method of preparing pn junctions in semiconductors. US Patent 2725316, filed May 18, 1953.
- 76 Fuller, C.S. (1954). Method of forming semiconductive bodies. US Patent 3015590, filed March 5, 1954.
- 77 Crishal, J.M., La Chapelle, T.J., Wilcox, W.R., and Sandstrom, J.P. (1962). Impurity diffusion method. US Patent 3244567, filed September 10, 1962.
- 78 Goldsmith, N. (1965). Method of forming a pn junction by vaporization. US Patent 3374125, filed May 10, 1965.
- 79 Moyer, J.W. (1954). Method of making P-N junction semiconductor unit. US Patent 2842466, filed June 15, 1954.
- 80 Shockley, W. (1954). Forming semiconductive devices by ionic bombardment. US Patent 2787564, filed October 28, 1954.
- 81 Ohl, R.S. (1950). Semiconductor translating devices. US Patent 2750541, filed January 31, 1950.

- 82 Derick, L. and Frosch, C.J. (1955). Oxidation of Semiconductive Surfaces for Controlled Diffusion. US Patent 2802760, filed December 2, 1955.
- 83 Frosch, C.J. and Derick, L. (1957). Surface protection and selective masking during diffusion in silicon. *Journal of the Electrochemical Society* 104: 547–552.
- 84 Andrus, J. (1957). Fabrication of semiconductor devices. US Patent 3122817, original filed August 15, 1957.
- 85 Lathrop, J.W. and Nall, J.R. (1957). Semiconductor construction. US Patent 2890395, filed October 31, 1957.
- 86 Nall, J.R., and Lathrop, J.W. (1957). Photolithographic fabrication techniques for transistors which are an integral part of a printed circuit. *1957 IEEE IEDM*, p. 117.
- 87 Christensen, H. and Teal, G.K. (1951). Method of fabricating germanium bodies. US Patent 2692839, filed March 7, 1951.
- 88 Kleimack, J.J., Loar, H.H., and Theuerer, H.C. (1960). Process of epitaxial vapor deposition with subsequent diffusion into the epitaxial layer. US Patent 3165811, filed June 10, 1960.
- 89 Dacey, G.C., Lee, C.A., and Shockley, W. (1955). Semiconductor device. US Patent 3028655, filed March 23, 1955.
- 90 Matlow, S.L. and Ralph, E.L. (1958). Ruggedized solar cell and process for making the same and the like. US Patent 2984775, filed July 9, 1958.
- 91 Moore, G.E. and Noyce, R.N. (1959). Method for fabricating transistors. US Patent 3108359, filed June 30, 1959.
- 92 Hoerni, J.A. (1959). Method of manufacturing semiconductor devices. US Patent 3025589, filed May 1, 1959.
- 93 Hoerni, J.A. (1961). Planar silicon diodes and transistors. *IRE Trans. Electron Devices* 8: 178.
- 94 Moore, G.E. (1998). The role of Fairchild in silicon technology in the early days of “Silicon Valley.” *Proceedings of the IEEE* 86: 53–62.
- 95 Riordan, M. (2007). The Silicon Dioxide Solution. *IEEE Spectrum* 44 (12): 50–56.
- 96 Kilby, J. (1958). *Laboratory Notebook*. Dallas, TX: Texas Instruments Archives. <https://digitalcollections.smu.edu/digital/collection/jackkilby>.
- 97 Kilby, J.S. (1959). Miniaturized electronic circuits. US Patent 3138743, filed February 6, 1959.
- 98 Noyce, R.N. (1959). Semiconductor device-and-lead structure. US Patent 2981877, filed July 30, 1959.
- 99 Hall, E.C. (1972). *MIT’s Role in Project Apollo: Final Report on Contracts NAS 9-163 and NAS 9-4065, Volume III*, Computer subsystem, MIT, Cambridge, MA. Retrieved June 15, 2021.
- 100 Wallmark, J.T. (1964). The field-effect transistor – an old device with new promise. *IEEE Spectrum* 1 (3): 182–192.
- 101 Barbottin, G. and Vapaille, A. (ed.) (1986, 1989). *Instability in Silicon Devices, Silicon Passivation and Related Instabilities*, vol. 1, 1986, vol. 2, 1989. North-Holland, Elsevier Science Publisher B.V.
- 102 Deal, B.E. and Grove, A.S. (1965). General relationship for the thermal oxidation of silicon. *Journal of Applied Physics* 36: 3770–3778.
- 103 Grove, A.S., Leistiko, O., and Sah, C.T. (1964). Redistribution of acceptor and donor impurities during thermal oxidation of silicon. *Journal of Applied Physics* 35: 2695–2701.
- 104 Terman, L.M. (1962). An investigation of surface states at a silicon/silicon dioxide interface employing metal-oxide-silicon diode. *Solid-State Electronics* 5: 285–299.
- 105 Deal, B.E., Sklar, M., Grove, A.S., and Snow, E.H. (1967). Characteristics of the surface-state (Q_{ss}) of thermally oxidized silicon. *Journal of the Electrochemical Society* 114: 266–274.
- 106 Kooi, E. (1967). *The Surface Properties of Oxidized Silicon*. Berlin Heidelberg: Springer-Verlag.

- 107 Appels, J.A., Kalter, H., and Kooi, E. (1970). Some problems of MOS technology. *Philips Tech. Rev.* 31: 225–236.
- 108 Revesz, A.G. (1966). Taxonomy of electric charges in metal-insulator-semiconductor systems. *Proceedings of the IEEE* 54: 1002–1003.
- 109 Hofstein, S.R. (1967). Stabilization of MOS devices. *Solid-State Electronics* 10: 657–670.
- 110 Kerr, D. (1970). A review of instability mechanisms in passivation films. *8th Reliability Physics Symposium*, Las Vegas, NV, USA, pp. 1–8.
- 111 Deal, B.E. (1980). Standardized terminology for oxide charges associated with thermally oxidized silicon. *IEEE Trans. Electron Devices* 27: 606–608.
- 112 McWhorter, A.L. (1957). 1/f noise and germanium surface properties. In: *Semiconductor Surface Physics* (ed. R.H. Kingston), 207–228. Philadelphia, PA: University of Pennsylvania Press.
- 113 Sah, C.T. and Hielscher, F.H. (1966). Evidence of the surface origin of the 1/f noise. *Physical Review Letters* 17: 956–958.
- 114 Fu, H.S. and Sah, C.T. (1972). Theory and experiments on surface 1/f noise. *IEEE Trans. Electron Devices* 19: 273–285.
- 115 Kirton, M.J. and Uren, M.J. (1989). Noise in solid-state microstructures: a new perspective on individual defects, interface states and low-frequency (1/f) noise. *Advances in Physics* 38: 367–468.
- 116 Claeys, C. and Simoen, E. (1998). Noise as a diagnostic tool for semiconductor material and device characterization. *Journal of the Electrochemical Society* 145: 2058–2067.
- 117 Konczakowska, A., Cichosz, J., and Szweczyk, A. (2008). A new method for RTS noise of semiconductor devices identification. *IEEE Trans. Instrum. Meas.* 57: 1199–1206.
- 118 Gray, P.V. and Brown, D.M. (1966). Density of SiO₂-Si interface states. *Applied Physics Letters* 8: 31–33.
- 119 Nicollian, E.H. and Goetzberger, A. (1967). The Si-SiO₂ interface – electrical properties as determined by the metal-insulator-silicon conductance technique. *Bell System Technical Journal* 46: 1055–1133.
- 120 Deal, B.E. (1974). The current understanding of charges in the thermally oxidized silicon structure. *Journal of the Electrochemical Society* 121: 198C–205C.
- 121 Tromp, R., Rubloff, G.W., Balk, P. et al. (1985). High-temperature SiO₂ decomposition at the SiO₂/Si interface. *Physical Review Letters* 55: 2332–2335.
- 122 Snow, E.H., Grove, A.S., and Fitzgerald, D.J. (1967). Effects of ionizing radiation on oxidized silicon surfaces and planar devices. *Proceedings of the IEEE* 55: 1168–1185.
- 123 Mitchell, J.P. and Wilson, D.K. (1967). Surface effects of radiation on semiconductor devices. *Bell System Technical Journal* 46: 1–80.
- 124 Dennehy, W.J., Holmes-Siedle, A.G., and Zaininger, K.H. (1967). Process techniques and radiation effects in metal-insulator semiconductor structures. *IEEE Trans. Nuclear Science* 14: 276–283.
- 125 Hecq, M. (1986). Evidence for electron-induced x-ray emission in sputtering deposition. *Applied Physics Letters* 49: 445–446.
- 126 Snow, E.H., Grove, A.S., Deal, B.E., and Sah, C.T. (1965). Ion transport phenomena in insulating films. *Journal of Applied Physics* 36: 1664–1673.
- 127 Hofstein, S.R. (1966). An investigation of instability and charge motion in metal-silicon oxide-silicon structures. *IEEE Trans. Electron Devices* 13: 222–237.
- 128 Hofstein, S.R. (1967). Tritium-labeled field-induced proton transport in SiO₂ films. *Appl. Phys. Lett.* 11: 95–96.

- 129 Yon, E., Ko, W.H., and Kuper, A.B. (1966). Sodium distribution in thermal oxide on silicon by radiochemical and MOS analysis. *IEEE Trans. Electron Devices* 13: 276–280.
- 130 Yurash, B. and Deal, B.E. (1968). A method for determining sodium content of semiconductor processing materials. *Journal of the Electrochemical Society* 115: 1191–1196.
- 131 Balk, P. (1965). Effects of hydrogen annealing on silicon surfaces. *ECS Spring Meeting* 14 (1): 237–240.
- 132 Castro, P.L. and Deal, B.E. (1971). Low-temperature reduction of fast surface states associated with thermally oxidized silicon. *Journal of the Electrochemical Society* 118: 280–286.
- 133 Thanh, L.D. and Balk, P. (1988). Elimination and generation of Si-SiO₂ interface traps by low temperature hydrogen annealing. *Journal of the Electrochemical Society* 135: 1797–1801.
- 134 Ono, M., Momoi, T., and Kawachi, Y. (1964). Semiconductor devices. US Patent 3643137, filed February 10, 1969, with foreign application priority February 13, 1964.
- 135 Ono, M. (2008). Unknown development story: from silicon mesa transistor to the patent of (100) crystal orientation [in Japanese]. 大野稔「開発秘話:シリコンメサトランジスタから(100)結晶面特許へ」、Semi News, pp. 24–25, September–October, 2008.
- 136 Delord, J.F., Hoffman, D.G., and Stringer, G. (1965). Use of MOS capacitors in determining properties of surface states at the Si-SiO₂ interface. *Bull. Am. Phys. Soc. Seri. II* 10 (4): 546, abstract KF9, April 26, 1965.
- 137 Miura, Y. (1965). Effect of orientation on surface charge density at silicon-silicon dioxide interface. *Japanese Journal of Applied Physics* 4: 958–961.
- 138 Balk, P., Burkhardt, P.J., and Gregor, L.V. (1965). Orientation dependence of built-in surface charge on thermally oxidized silicon. *Proceedings of the IEEE* 53: 2133–2134.
- 139 Abe, T. (1966). Surface cleaning of Ge and Si wafers. *Oyobutsuri* 35: 511. [in Japanese]. 阿部敏夫「Ge, Siウエーハの表面洗浄」『応用物理』第35巻、1966年、第7号、511頁.
- 140 Kern, W. and Puotinen, D.A. (1970). Cleaning solutions based on hydrogen peroxide for use in silicon semiconductor technology. *RCA Rev.* 31: 187–206.
- 141 Kern, W. (1990). The evolution of silicon wafer cleaning technology. *Journal of the Electrochemical Society* 137: 1887–1892.
- 142 Lang, G.A. and Stavish, T. (1963). Chemical polishing of silicon with anhydrous hydrogen chloride. *RCA Rev.* 24: 488–498.
- 143 Kriegler, R.J., Cheng, Y.C., and Colton, D.R. (1972). The effect of HCl and Cl₂ on the thermal oxidation of silicon. *Journal of the Electrochemical Society* 119: 388–392.
- 144 Kriegler, R.J. (1972). Neutralization of Na⁺ Ions in “HCl-Grown” SiO₂. *Applied Physics Letters* 20: 449–451.
- 145 Goetzberger, A. and Shockley, W. (1960). Metal precipitates in silicon p-n junctions. *Journal of Applied Physics* 31: 1821–1824.
- 146 Ing, S.W. Jr., Morrison, R.E., Alt, L.L., and Aldrich, R.W. (1963). Gettering of metallic impurities from planar silicon diodes. *Journal of the Electrochemical Society* 110: 533–537.
- 147 Kerr, D.R., Logan, J.S., Burkhardt, P.J., and Pliskin, W.A. (1964). Stabilization of SiO₂ passivation layers with P₂O₅. *IBM Journal of Research and Development* 8: 376–384.
- 148 Kerr, D.R. and Young, D.R. (1964). Method of improving electrical characteristics of semiconductor devices and products so produced. US Patent 3303059, filed June 29, 1964.
- 149 Lambert, J.L. and Reese, M. (1968). The gettering of gold and copper from silicon. *Solid-State Electronics* 11: 1055–1061.
- 150 Nakamura, M., Kato, T., and Oi, N. (1968). A study of gettering effect of metallic impurities in silicon. *Japanese Journal of Applied Physics* 7: 512–518.

- 151 Seidel, T.E., Meek, R.L., and Cullis, A.G. (1975). Direct comparison of ion-damage gettering and phosphorus-diffusion gettering of Au in Si. *Journal of Applied Physics* 46: 600–609.
- 152 Meek, R.L. and Gibbon, C.F. (1974). Preliminary results of an ion scattering study of phosphosilicate glass gettering. *Journal of the Electrochemical Society* 121: 444–447.
- 153 Balk, P., Dong, D.W., and Eldridge, J.M. (1968). Field Effect Transistor Having Passivated Gate Insulator. US Patent 3560810, filed August 15, 1968.
- 154 Balk, P. and Eldridge, J.M. (1969). Phosphosilicate glass stabilization of FET devices. *Proceedings of the IEEE* 57: 1558–1563.
- 155 Eldridge, J.M., Laibowitz, R.B., and Balk, P. (1969). Polarization of thin phosphosilicate glass films in MGOS structures. *Journal of Applied Physics* 40: 1922–1930.
- 156 Eldridge, J.M. and Kerr, D.R. (1971). Sodium ion drift through phosphosilicate glass – SiO₂ films. *Journal of the Electrochemical Society* 118: 986–991.
- 157 Schnable, G.L., Kern, W., and Comizzoli, R.B. (1975). Passivation coatings on silicon devices. *Journal of the Electrochemical Society* 122: 1092–1103.
- 158 Norman, R.H. and Stephenson, H.E. (1964). Shift register employing insulated gate field effect transistors. US Patent 3454785, filed July 27, 1964.
- 159 Boycel, L. University of Michigan EECS presentation slide: October 12, 2007. <https://eecsnnews.engin.umich.edu/wp-content/uploads/sites/2/2023/03/Boysel-slides.pdf>.
- 160 Chiba, T., Iwase, T., Yoshida, Y., and Shirakawa, I. (2012). History of developing and commercializing families of solid-state calculators. *2012 HISTELCON*: pp. 1–5.
- 161 Sharp History of LCD calculator evolution (chronological table) [in Japanese]. <https://jp.sharp/products/lcd/tech/dentaku/history.html>.
- 162 Aspray, W. (1994). *Oral-history: Tadashi Sasaki. Interview #211 for the Center for the History of Electrical Engineering*. IEEE. https://ethw.org/Oral-History:Tadashi_Sasaki.
- 163 Dentaku museum: <http://www.dentaku-museum.com/calc/calc/91-bunrui1/11-desktop/desktop.html>.
- 164 Okon, T. (2017). The first handheld digital calculator celebrates 50 years, part 1. *Electronic Design*. <https://www.electronicdesign.com/technologies/analog/article/21804824/the-firsthandheld-digital-calculatorcelebrates-50-years-part-1>.
- 165 Okon, T. (2017). The first handheld digital calculator celebrates 50 years, part 2. *Electronic Design*. https://www.electronicdesign.com/technologies/analog/article/21804872/the-firsthandheld-digital-calculator-celebrates-50-years-part-2?utm_source=chatgpt.com.
- 166 Kilby, J.S., Merryman, J.D., and Van Tassel, J.H. (1967). Miniature electronic calculator. US Patent 3819921, filed December 21, 1972. Continuation of Ser. No. 143,192, May 13, 1971, abandoned, which is a continuation of Ser. No. 671,777, Sept. 29, 1967, abandoned.
- 167 Wang, A. (1949). Pulse Transfer Controlling Device. US Patent 2708722, filed October 21, 1949.
- 168 Forrester, J.W. (1951). Multicoordinate digital information storage device. US Patent 2736880, filed May 11, 1951.
- 169 Forrester, J.W. (1951). Digital information storage in three dimensions using magnetic cores. *Journal of Applied Physics* 22: 44–48.
- 170 Norman, R.H. (1963). Solid state switching and memory apparatus. US Patent 3562721, filed March 5, 1963.
- 171 Perkins, H.A. and Schmidt, J.D. (1965). An integrated semiconductor memory system. *Proceedings of the November 30-December 1, 1965, fall joint computer conference, part I*, pp. 1053–1064.
- 172 Agusta, B., Bardell, P., Castrucci, P. (1965). Sixteen bit monolithic memory array clip. *1965 IEEE IEDM*, p. 39.

- 173 Agusta, B., Bardell, P.H., Castrucci, P.P. et al. (1966). Monolithic integrated memory array structure including fabrication and package therefor. US Patent 3508209, filed March 31, 1966.
- 174 Rice, R., Sander, W.B., and Greene, F.S. Jr., (1970). Design considerations leading to the ILLIAC IV LSI process element memories. *IEEE J. Solid-State Circuits* 5: 174–181.
- 175 Schmidt, J.D. (1965). Integrated MOS random-access memory. *Solid State Design, Communications & Data Equipment* 6: 21–24.
- 176 Wood J. and Ball, R.G. (1965). The use of insulated-gate field-effect transistors in digital storage systems. *1965 IEEE ISSCC*, pp. 82–83.
- 177 Pleshko, P. and Terman, L.M. (1966). An investigation of the potential of MOS transistor memories. *IEEE Trans. Electron. Comput.* 15: 423–427.
- 178 Igarashi, R., Kurosawa, T., and Yaita, T. (1966). A 150-nanosecond associative memory using integrated MOS transistors. *1966 IEEE ISSCC*, pp. 104–105.
- 179 Igarashi, R. and Yaita, T. (1967). An integrated MOS transistor associative memory system with 100 ns cycle time. *Proceedings of the April 18–20, 1967, Spring joint computer conference*, pp. 499–506.
- 180 Brewer, D.E., Nissim, S., and Podraza, G.V. (1967). Low power computer memory system. *Proceedings of the November 14–16, 1967, fall joint computer conference*, pp. 381–393.
- 181 Yamamoto, H., Shiraishi, M., and Kurosawa, T. (1969). A 40-ns, 144-bit N-channel MOS-IC memory. *1969 IEEE ISSCC*, pp. 40–41.
- 182 Tarui, Y., Hayashi, Y., Koyanagi, T. et al. (1969). A 40-NS 144-bit n-channel MOS-LSI memory. *IEEE J. Solid-State Circuits* 4: 271–279.
- 183 Allison, J.F., Heiman, F.P., and Burns, J.R. (1967). Silicon-on-sapphire complementary MOS memory cells. *IEEE J. Solid-State Circuits* 2: 208–212.
- 184 Allison, J.F., Burns, J.R., and Heiman, F.P. (1967). Silicon on sapphire complementary MOS memory systems. *1967 IEEE ISSCC*, pp. 76–77.
- 185 Hart, T.W. Jr., Hillis, D.W., and Marley, J. et al. (1969). A main frame semiconductor memory for fourth generation computers. *Proceedings of the November 18-20, 1969, fall joint computer conference*, pp. 479–488.
- 186 Lo, A.W. (1968). High-speed logic and memory: past, present and future. *Proceedings of the December 9–11, 1968, fall joint computer conference, part II*, pp. 1459–1465.
- 187 Hodges, D.A. (1968). Large-capacity semiconductor memory. *Proceedings of the IEEE* 56: 1148–1162.
- 188 Sander, W.B. (1968). Semiconductor memory circuits and technology. *Proceedings of the December 9-11, 1968, fall joint computer conference, part II*, pp. 1205–1211.
- 189 The CPUSHACK Museum The RCA solid state technology center (SSTC). https://www.cpushack.com/2021/09/28/the-rca-solid-state-technology-center-sstc/?utm_source=chatgpt.com.
- 190 Funk, R.E. (1971). New low-voltage COS /MOS IC's offer 25-ns speed and direct interfacing with saturated logic. *RCA Eng.* 17: 40–45.
- 191 Bower, R.W. and Dill, R.G. (1966). Insulated gate field effect transistors fabricated using the gate as source-drain mask. *1966 IEEE IEDM*, pp. 102–104.
- 192 Faggin, F. and Klein, T. (1970). Silicon gate technology. *Solid-State Electronics* 13: 1125–1144.
- 193 Faggin, F. (2020). *Silicon: From the Invention of the Microprocessor to the New Science of Consciousness*. Waterside Productions.
- 194 Moore, G.E. (1996). Intel – memories and the microprocessor. *Daedalus* 125 (2): 55–80.
- 195 Volk, A.M., Stoll, P.A., and Metrovich, P. (2001). Recollections of early chip development at Intel. *Intel Technology Journal* Q1: 1–12.

- 196 Kerwin, R.E. (2023). The invention of the self-aligned silicon gate process. In: *75th Anniversary of the Transistor* (ed. A. Nathan, S.K. Saha, and R.M. Todi), Ch. 8, 89–94. IEEE Press-Wiley.
- 197 Vadasz, L.L. (1970). Integrated circuit structure and method for making integrated circuit structure. US Patent 3699646, filed December 28, 1970.
- 198 Watkins, B.G. and Selser, M.J. (1966). IGFET comprising of n-type silicon substrate, silicon oxide gate insulator and p-type polycrystalline silicon gate electrode. US Patent 3576478, continuation of Ser. No. 595,163, abandoned, which is a continuation-in-part of Ser. No. 582,053 September 26, 1966, abandoned.
- 199 Dill, H.G. (1966). Insulated gate field-effect transistor (IGFET) with semiconductor gate electrode. US Patent 3544399, filed October 26, 1966.
- 200 Kerwin, R.E., Klein, D.L., and Sarace, J.C. (1967). Method for making MIS structures. US Patent 3475234, filed March 27, 1967.
- 201 Sarace, J.C., Kerwin, R.E., Klein, D.L., and Edwards, R. (1968). Metal-nitride-oxide-silicon field-effect transistors, with self-aligned gates. *Solid-State Electronics* 11: 653–660.
- 202 Fair, R.B. (1998). History of some early developments in ion-implantation technology leading to silicon transistor manufacturing. *Proceedings of the IEEE* 86: 111–137.
- 203 Federal Supplement 374 F. Supp. 1166, 1974.
- 204 Bower, R.W. (1966). Field-effect device with insulated gate. US Patent 3472712, filed October 27, 1966.
- 205 Bower, R.W. (1967). Insulated-gate field-effect device having source and drain regions formed in part by ion implantation and method of making same. US Patent 3615934, filed October 30, 1967, continuation-in-part of application Ser. No. 590033, October 27, 1966.
- 206 Brown, D.M., Engeler, W.E., Gray, P.V., and Garfinkel, M. (1967). Self-registered IG-FET devices and method of making same. US Patent 3566517, filed October 13, 1967.
- 207 Dill, H.G., Bower, R.W., and Toombs, T.N. (1971). Ion-implanted MOS technology. *Radiation Effects* 7: 45–57.
- 208 Crowder, B.L. and Morehead, F.F. Jr., (1969). Annealing characteristics of n-type dopant in ion-implanted silicon. *Applied Physics Letters* 14: 313–315.
- 209 Davies, D.E. (1969). Post annealing conductance behavior of implanted layers in silicon. *Applied Physics Letters* 14: 227–229.
- 210 Bower, R.W., Dill, H.G., Aubuchon, K.G., and Thompson, S.A. (1968). MOS field effect transistors formed by gate masked ion implantation. *IEEE Trans. Electron Devices* 15: 757–761.
- 211 Dill, H.G. (1968). A new insulated gate tetrode with high drain breakdown potential and low Miller feedback capacitance. *IEEE Trans. Electron Devices* 15: 717–728.
- 212 Ogura, S., Tsang, P.J., Walker, W.W. et al. (1980). Design and characteristics of the lightly doped drain-source (LDD) insulated gate field-effect transistor. *IEEE Trans. Electron Devices* 27: 1359–1367.
- 213 Ogura, S., Tsang, P.J., Walker, W.W. et al. (1981). Elimination of hot electron gate current by the lightly doped drain-source structure. *1981 IEEE IEDM*, pp. 651–654.
- 214 Brown, D.M., Engeler, W.E., Garfinkel, M., and Gray, P.V. (1968). Self-registered molybdenum-gate MOSFET. *Journal of the Electrochemical Society* 115: 874–876.
- 215 Brown, D.M., Engeler, W.E., Garfinkel, M., and Gray, P.V. (1968). Refractory metal silicon device technology. *Solid-State Electronics* 11: 1105–1112.
- 216 Brown, D.M., Cady, W.R., Sprague, J.W., and Salvagni, P.J. (1971). The p-channel refractory metal self-registered MOSFET. *IEEE Trans. Electron Devices* 18: 931–940.
- 217 Klein, T. and Faggin, F. (1970). Doped semiconductor electrodes for MOS type devices. US Patent 3673471, filed October 8, 1970.

- 218 Moore, G.E. (1970). Method of a semiconductor device wherein film cracking is prevented by formation of a glass layer. US Patent 3825442, filed September 27, 1972, continuation of abandoned application Ser. No. 4,841, January 22, 1970.
- 219 Cohen, L., Rubenstein, R., and Wanlass, F. (1967). MTOS four phase clock systems. *IEEE Northeast Electronics Research and Engineering Meeting (NEREM) Record* 9: 170–171.
- 220 Karp, J. and DeAtley, E. (1967). Use four-phase MOS IC logic. *Electronic Design* 15: 62–66.
- 221 Yen, Y. (1968). Transient analysis of four-phase mos switching circuits. *IEEE J. Solid-State Circuits* 3: 1–5.
- 222 “Sanyo ICC-82D,” Hand-held Calculators, Vintage Calculators Web Museum. http://www.vintagecalculators.com/html/sanyo_icc-82d.html.
- 223 Ayling, J. and Moore, R.D. (1971). Main monolithic memory. *IEEE J. Solid-State Circuits* 6: 276–279.
- 224 Vadasz, L.L., Grove, A.S., Rowe, T.A., and Moore, G.E. (1969). Silicon-gate technology. *IEEE Spectrum* 6: 28–35.
- 225 Regitz, W.M. and Karp, J. (1970). A three transistor-cell, 1024-Bit, 500 NS MOS RAM. 1970 *IEEE ISSCC*, pp. 42–43.
- 226 Regitz, W.M. and Karp, J.A. (1970). Three-transistor-cell 1024-bit 500-ns MOS RAM. *IEEE J. Solid-State Circuits* 5: 181–186.
- 227 Dennard, R.H. (1967). Field-effect Transistor memory. US Patent 3387286, filed July 14, 1967.
- 228 Faggin, F. (1972). Standard parts and custom design merge in four-chip processor kit. *Electronics* 45: 112.
- 229 Noyce, R.N. and Hoff, M.E. (1981). A history of microprocessor development at Intel. *IEEE Micro* 1: 8–21.
- 230 Mazor, S. (2007). Intel 8080 CPU Chip Development. *IEEE Annals of the History of Computing* 29: 70–73.
- 231 Faggin, F., Hoff, M.E., Mazor, S., and Shima, M. (1996). The history of the 4004. *IEEE Micro* 16: 10–20.
- 232 Yourke, H.S. (1956). Transistor Switching Circuits. US Patent 2964652, filed November 15, 1956.
- 233 Yourke, H.S. and Slobodzinski, E.J. (1957). Millimicrosecond current switching circuits. *Proceedings of the Western Joint Computer Conference*, pp. 68–72.
- 234 Namimoto, K. (1974). Micro Computer and its application. *The Journal of the Institute of Television Engineers of Japan* 28: 313–322. [in Japanese]. 浪本敬二「マイクロコンピュータとその応用」『テレビジョン学会誌』、1974年 28巻 4号313-322頁.
- 235 Kahng, D. and Sze, S.M. (1967). A floating-gate and its application to memory devices. *Bell System Technical Journal* 46: 1288–1295.
- 236 Kahng, D. (1967). Field effect semiconductor apparatus with memory involving entrapment of charge carriers. US Patent 3500142, filed June 5, 1967.
- 237 Fowler, R.H. and Nordheim, L. (1928). Electron emission in intense electric fields. *Proceedings of the Royal Society A* 119: 173–181.
- 238 Frohman-Bentchkowsky, D. (1970). Floating gate transistor and method for charging and discharging same. US Patent 3660819, filed June 15, 1970.
- 239 Hara, H., Okamoto, Y., and Ohnuma, H. (1970). A new instability in MOS transistors caused by hot electron and hole injection from drain avalanche plasma into gate oxide. *Japanese Journal of Applied Physics* 9: 1103–1112.

- 240 Takeishi, Y., Hara, H., Sato, T. et al. (1971). The channel-conductance increase of conventional MOS transistors by avalanche injection: Application to read-only memory. *1971 IEEE IEDM*, pp. 104–106.
- 241 Frohman-Bentchkowsky, D. (1971). Electrically programmable read only memory array. US Patent 3744036, filed May 24, 1971.
- 242 Frohman-Bentchkowsky, D. (1971). A fully-decoded 2048-bit electrically-programmable MOS ROM. *1971 IEEE ISSCC*, pp. 80–81.
- 243 Frohman-Bentchkowsky, D. (1971). A fully decoded 2048-bit electrically programmable FAMOS read-only memory. *IEEE J. Solid-State Circuits* 6: 301–306.
- 244 Iizuka, H., Sato, T., Masuoka, F. et al. (1972). Stacked-gate avalanche-injection type MOS (SAMOS) memory. *4th Conference on Solid State Devices*, pp. 55–56.
- 245 Frohman-Bentchkowsky, D., Mar, J., Perlegos, G., and Johnson, W.S. (1978). Electrically programmable and erasable MOS floating gate memory device employing tunneling and method of fabricating same. US Patent 4203158, filed December 15, 1978. Continuation-in-part of Ser. No. 881,029, February 24, 1978, abandoned.
- 246 Masuoka, F. (1980). Semiconductor memory device. Japan Patent JP S61-39752, filed November 20, 1980.
- 247 Masuoka, F. and Iizuka, H. (1980). Semiconductor memory device and method for manufacturing the same. US Patent 4531203, filed November 13, 1981. Foreign application priority December 20, 1980.
- 248 Masuoka, F. (1987). Non-volatile semiconductor memory of a metal ferroelectric field effect transistor. Japan Patent JP H7-93017, filed April 24, 1987.
- 249 Johnson, W.S., Perlegos, G., Renninger, A. et al. (1980). A 16Kb electrically erasable non-volatile memory. *1980 IEEE ISSCC*, pp. 152–153.
- 250 Masuoka, F. (1996). Flash memory development story. *Journal of the Institute of Television Engineers of Japan* 50: 617–620. [in Japanese]. 舩岡 富士雄「フラッシュメモリーの開発ストーリー」『テレビジョン学会誌』、1996年 50巻 5号 617-620頁.
- 251 Masuoka, F., Asano, M., Iwahashi, H. et al. (1984). A new Flash EEPROM cell using triple polysilicon technology. *1984 IEEE IEDM*, pp. 464–467.
- 252 Masuoka, F. (1987). Programmable semiconductor memory. US Patent 5313420, filed March 28, 1991. Foreign application priority April 24, 1987.
- 253 Itoh, Y., Momodomi, M., Shirota, R. et al. (1989). An experimental 4 Mb CMOS EEPROM with a NAND structured cell. *1989 IEEE ISSCC*, pp. 134–135.
- 254 Buie, J.L. (1961). Coupling transistor logic and other circuits. US Patent 3283170, filed September 8, 1961.
- 255 Fang, F.F. and Crowder, B.L. (1970). Ion-implanted microwave MOSFET's. *1970 IEEE IEDM*, p. 80.
- 256 Jaddam, A.Y. (1971). Fabrication of enhancement n-channel Mosfets using ion implanted boron for controlled channel doping. *1971 IEEE IEDM*, 21.2, p. 156.
- 257 Edwards, J.R. and Marr, G. (1971). Ion-implanted depletion-mode IGFET. *1971 IEEE IEDM*, p. 156.
- 258 Cheroff, G., Critchlow, D.L., Dennard, R.H., and Terman, L.M. (1969). IGFET circuit performance-n-channel versus p-channel. *IEEE J. Solid-State Circuits* 4: 267–271.
- 259 Masuhara, T., Nagata, M., and Hashimoto, N. (1971). A high-performance N-channel MOS-LSI using depletion-type load elements. *1971 IEEE ISSCC*, pp. 12–13.

- 260 Masuhara, T., Nagata, M., and Hashimoto, N. (1972). A high-performance N-channel MOSLSI using depletion-type load elements. *IEEE J. Solid-State Circuits* 7: 224–231.
- 261 Critchlow, D.L., Dennard, R.H., and Schuster, S.E. (1973). Design and characteristics of n-channel insulated-gate field-effect transistors. *IBM Journal of Research and Development* 17: 430–442.
- 262 Krolikowski, W., Brown, W., Dries, R. et al. (1970). A 1024 bit n-channel MOS read-write memory chip. *1970 IEEE IEDM*, p. 16.
- 263 Lund, D., Allen, C.A., Andersen, S.R., and Tu, G.K. (1970). Design of a megabit semiconductor memory system. *Proceedings of November 17–19, 1970, fall joint computer conference*, pp. 53–62.
- 264 Mazor, S. (2007). Intel 8080 CPU chip development. *IEEE Annals of the History of Computing* 29: 70–73.
- 265 Karp, J.A., Regitz, W.M., and Chou, S. (1972). A 4096-bit dynamic MOS RAM. *1972 IEEE ISSCC*, pp. 10–11.
- 266 Abbott, R.A., Regitz, W.M., and Karp, J.A. (1973). A 4K MOS dynamic random-access memory. *IEEE J. Solid-State Circuits* 8: 292–298.
- 267 Proebsting, R. and Green, R. (1973). A TTL compatible 4096-bit N-channel RAM. *1973 IEEE ISSCC*, pp. 28–29.
- 268 “Chip Hall of Fame: Mostek MK4096 4-Kilobit DRAM” *IEEE Spectrum* 30 June, 2017. <https://spectrum.ieee.org/tech-history/silicon-revolution/chip-hall-of-fame-mostek-mk4096-4kilobitdram>.
- 269 Pugh, E.W., Critchlow, D.L., Henle, R.A., and Russell, L.A. (1981). Solid state memory development at IBM. *IBM J. Res. Develop.* 25: 585–602.
- 270 Dennard, R.H. (1984). Evolution of the MOSFET dynamic RAM – a personal view. *IEEE Trans. Electron Devices* 31: 1549–1555.
- 271 Rideout, V.L. (1979). One-device cells for dynamic random access memories: a tutorial. *IEEE Trans. Electron Devices* 26: 839–852.
- 272 Burke, H.K. and Michon, G.J. (1972). Charge pump random-access memory. *1972 IEEE ISSCC*, pp. 16–17.
- 273 Engeler, W.E., Tiemann, J.J., and Baertsch, R.D. (1972). A surface-charge random-access memory system. *IEEE J. Solid-State Circuits* 7: 330–335.
- 274 Stein, K.U. and Friedrich, H. (1973). A 1-mil² single-transistor memory cell in n silicon-gate technology. *IEEE J. Solid-State Circuits* 8: 319–323.
- 275 Lambrechtse, C.W., Salters, R.H.W., and Boonstra, L. (1973). A-4096 bit one-transistor per-bit RAM with internal timing and low dissipation. *1973 IEEE ISSCC*, pp. 26–27, 194.
- 276 Hoffman, W.K. and Kalter, H.L. (1973). An 8K b random-access memory chip using the one-device FET cell. *IEEE J. Solid-State Circuits* 8: 298–305.
- 277 Schroeder, P.R. and Proebsting, R.J. (1977). A 16K × 1 bit dynamic RAM. *1977 IEEE ISSCC*, pp. 12–13.
- 278 Hoffmann, K. and Stein, K.U. (1978). Trends in dynamic RAMs. *ESSCIRC '78: 4th European Solid State Circuits Conference*, pp. 24–26.
- 279 Asai, S. (1986). Semiconductor memory trends. *Proceedings of the IEEE* 74: 1623–1635.
- 280 Maes, H.E., Groeseneken, G., Lebon, H., and Witters, J. (1989). Trends in semiconductor memories. *Microelectronics Journal* 20: 9–58.
- 281 Masuoka, F., Ohuchi, K., Tanaka, S. et al. (1975). 1k-bit n-channel MOS Static RAM. *Toshiba Rev.* 30: 923. *Toshiba Review* 30: 923–926 [in Japanese]. 舩岡富士雄、大内和則、田中寿実夫、田沼幸男、岩井洋、久保田信久、堀内重治、"1kビットnチャネルMOSスタティックRAM", 東芝レビュー、1975年11月、第30巻、第12号、923-926頁.

- 282 Takeishi, A. and Ito, S. (2007). Development of microcontrollers for automotive engine control. *Hitotsubashi Business Review* 55: 116–143. [in Japanese]. 武石彰、伊藤誠悟「東芝自動車エンジン制御用マイコンの開発」一橋大学21世紀COEプログラム「知識・企業・イノベーションのダイナミクス」大河内賞ケース研究プロジェクト2007年8月 CASE#07-04: <https://pubs.iir.hit-u.ac.jp/admin/ja/pdfs/file/745>.
- 283 Inoue, T., Horiuchi, S., Iwai, H. et al. (1975). Micro-probe Auger analysis of Si migration in Al metallization for LSI. *7th Conference on Solid State Devices*, Suppl. *Japanese Journal of Applied Physics* 15: 63–69.
- 284 Ho, P.S. and Kwok, T. (1989). Electromigration in metals. *Reports on Progress in Physics* 52: 301–348.
- 285 Kuiper, L.L. (1965). Method for providing electrical connections to semiconductor devices. US Patent 3382568, filed July 22, 1965.
- 286 d’Heurle, F.M. (1971). The effect of copper additions on electromigration in aluminum thin films. *Metall. Trans.* 2: 683–689.
- 287 Iwabuchi, S., Shima, S., Moriya, T., and Maeda, T. (1986). A highly reliable pure Al metallization with low contact resistance utilizing oxygen-stuffed TiN barrier layer. *1986 Symposium on VLSI Technology*, pp. 55–56.
- 288 von Seefeld, H., Cheung, N.W., Maenpaa, M., and Nicolet, M.-A. (1980). Investigation of titanium-nitride layers for solar-cell contacts. *IEEE Trans. Electron Devices* 27: 873–876.
- 289 Horiuchi, S. and Iwai, H. (1974). Method of semiconductor device fabrication. Japan Patent JP S55-28434, filed January 29, 1974.
- 290 Horiuchi, S., Tanuma, Y., and Iwai, H. (1975). Semiconductor integrated circuits. Japan Patent JP S59-33979, filed July 16, 1975.
- 291 Horiuchi, S., Tanuma, Y., and Iwai, H. (1975). Semiconductor device. Japan Patent JP S58-16338, filed July 29, 1975.
- 292 Iiri, M. (1975). Manufacturing method of semiconductor device. Japan Patent JP S59-4855, filed August 13, 1975.
- 293 Kooi, E. (1966). Methods of producing a semiconductor device and a semiconductor device produced by said method. US Patent 3970486, filed February 14, 1975. Foreign application priority October 5, 1966.
- 294 Kooi, E. and Appels, J.A. (1970). Method of making semiconductor devices with selective doping and selective oxidation. US Patent 3755001, filed July 8, 1971. Foreign application priority July 10, 1970.
- 295 Appels, J.A., Kooi, E., Paffen, M.M. et al. (1970). Local oxidation of silicon and its application in semiconductor technology. *Philips Research Reports* 25: 118–132.
- 296 Iwai, H., Taniguchi, K., Konaka, M. et al. (1980). Two dimensional nature of diffused line capacitance in coplanar structures. *1980 IEEE IEDM*, pp. 728–731.
- 297 Iwai, H., Taniguchi, K., Konaka, M. et al. (1982). Two-dimensional nature of diffused layers and certain limitations in scaling-down coplanar structure. *IEEE Trans. Electron Devices* 29: 625–630.
- 298 Ohuchi, K., Iwai, H., Yamaguchi, M., and Ohno, T. (1980). A 64K-bit dynamic-RAM TMM4164C. *Toshiba Review* 128: 35–38.
- 299 Moore, G.E. (1965). Cramming more components onto integrated circuits. *Electronics* 38 (8): 114–117.
- 300 Sawada, S., Maeda, S., Matsumoto, Y. et al. (1982). Degradation of thin gate oxide under process-induced electrical stress. *162nd ECS Fall Meeting*, October 1982, Detroit, Recent Newspapers. *J. Electrochem. Soc.* 129 (11): 426C.

- 301 Otsuka, H., Watanabe, K., Nishimura, H. et al. (1982). The effect of substrate materials on holding time degradation in MOS dynamic RAM. *IEEE Electron Device Letters* 3: 182–184.
- 302 Iwai, H., Otsuka, H., Matsumoto, Y. et al. (1984). Comparison of intrinsic gettering and epitaxial wafers in terms of soft error endurance and other characteristics of 64k bit dynamic RAM. *IEEE Trans. Electron Devices* 31: 1149–1151.
- 303 May, T.C. and Woods, M.H. (1979). Alpha-particle induced soft errors in dynamic memories. *IEEE Trans. Electron Devices* 26: 2–9.
- 304 Hsieh, C.M., Murley, P.C., and O'Brien, R.R. (1981). A field-funneling effect on the collection of alpha-particle-generated carriers in silicon devices. *IEEE Electron Device Letters* 2: 103–105.
- 305 Sai-Halasz, G.A., Wordeman, M.R., and Dennard, R.H. (1982). Alpha-particle-induced soft error rate in VLSI circuits. *IEEE Trans. Electron Devices* 29: 725–731.
- 306 Hsieh, C.M., Murley, P.C., and O'Brien, R.R. (1983). Collection of charge from alpha-particle tracks in silicon devices. *IEEE Trans. Electron Devices* 30: 686–693.
- 307 Wanlass, F.M. and Sah, C.T. (1963). Nanowatt logic using field-effect metal-oxide semiconductor triodes. 1963 *IEEE ISSCC*, February 20, 1963, pp. 32–33.
- 308 Wanlass, F.M. (1963). Low stand-by power complementary field effect circuitry. US Patent 3356858, filed June 18, 1963.
- 309 Swanson, R.M. and Meindl, J.D. (1972). Ion-implanted complementary MOS transistors in low-voltage circuits. *IEEE J. Solid-State Circuits* 7: 146–153.
- 310 Crawford, R.H. (1966). Linear amplifier utilizing a pair of field effect transistors. US Patent 3436621, filed December 16, 1966.
- 311 Masuhara, T., Nagata, M., and Kubo, M. (1970). Semiconductor integrated circuit composed of cascade connection of inverter circuits. US 3700981, filed May 24, 1971. Foreign application priority May 27, 1970.
- 312 Ahrons, R.W., Mitchell, M.M., and Burns, J.R. (1965). MOS micropower complementary transistor logic. 1965 *IEEE ISSCC*, pp. 80–81.
- 313 Dingwall, A.G.F. (1970). Compact COS/MOS 256-bit random-access memory. 1970 *IEEE IEDM*, p. 16.
- 314 Coppen, P.J., Aubuchon, K.G., and Bauer, L.-O. (1972). A complementary MOS 1.2 volt watch circuit using ion implantation. *Solid-State Electronics* 15: 165–174.
- 315 Guarnieri, M. (2017). Seventy years of getting transistorized. *IEEE Industrial Electronics Magazine* 11: 33–37.
- 316 Suzuki, Y., Odagawa, K., and Abe, T. (1973). Clocked CMOS calculator circuitry. *IEEE J. Solid-State Circuits* 8: 462–469.
- 317 Kimura, K., Itoh, K., Hori, R. et al. (1986). Power reduction techniques in megabit DRAM's. *IEEE J. Solid-State Circuits* 21: 381–389.
- 318 Higuchi, H., Kitsukawa, G., Ikeda, T. et al. (1984). Performance and structures of scaled-down bipolar devices merged with CMOSFETs. 1984 *IEEE IEDM*, pp. 694–697.
- 319 Iwai, H., Niitsu, Y., Sasaki, G. et al. (1987). 1.2 μm Bi-CMOS technology with high performance ECL. *ESSDERC '87*, pp. 29–32. Also in *Solid State Devices*, ed. E. Soncini and P. U. Calzolari, Elsevier 1988, pp. 199–202.
- 320 Iwai, H., Sasaki, G., Unno, Y. et al. (1987). 0.8 μm Bi-CMOS technology with high f_T ion-implanted emitter bipolar transistor. 1987 *IEEE IEDM*, pp. 28–31.
- 321 Masaki, A. (1991). Possibilities of CMOS mainframe and its impact on technology R&D. 1991 *Symposium on VLSI Technology*, pp. 1–4.
- 322 Yoshitomi, T., Saito, M., Ohguro, T. et al. (1999). Single-gate 0.15 and 0.12 μm CMOS with Co salicide technology. *Solid-State Electronics* 43: 543–546.

- 323 Iinuma, T., Itoh, N., Nakajima, H. et al. (1995). Sub-20 ps high-speed ECL bipolar transistor with low parasitic architecture. *IEEE Trans. Electron Devices* 42: 399–405.
- 324 Nii, H., Yoshino, C., Yoshitomi, S. et al. (1999). An 0.3- μm Si epitaxial base BiCMOS technology with 37-GHz f_{max} and 10-V BV_{ceo} for RF telecommunication. *IEEE Trans. Electron Devices* 46: 712–721.
- 325 Ezaki, H. and Ishikawa, D. (1977). Method for manufacturing a MOS-type semiconductor device. Japan Patent S62–031506, filed September 14, 1977.
- 326 Codella, C.F. and Ogura, S. (1985). Halo doping effects in submicron DI-LDD device design. *1985 IEEE IEDM*, pp. 230–233.
- 327 Brand, A., Haranahalli, A., Hsieh, N. et al. (1998). Intel's 0.25 micron, 2.0 volts logic process technology. *Intel Technology Journal Q 3*: 1–8.
- 328 Thompson, S., Packan, P., and Bohr, M. (1998). MOS scaling: transistor challenges for the 21st century. *Intel Technology Journal Q 3*: 12–19.
- 329 Fair, R.B. (ed.) (1993). *Rapid Thermal Annealing*. Academic Press Inc.
- 330 Hart, M.J. and Evans, A.G.R. (1988). Rapid thermal processing in semiconductor technology. *Semiconductor Science and Technology* 3: 421–436.
- 331 Norishima, M., Niitsu, Y., Sasaki, G. et al. (1989). Bipolar transistor design for low process-temperature 0.5 μm BiCMOS. *1989 IEEE IEDM*, pp. 237–240.
- 332 Norishima, M., Iwai, H., Niitsu, Y., and Maeguchi, K. (1992). Impurity diffusion behavior of bipolar transistor under low-temperature furnace annealing and high-temperature RTA and its optimization for 0.5- μm Bi-CMOS process. *IEEE Trans. Electron Devices* 39: 33–40.
- 333 Chapman, R.A., Kuehne, J.W., Ying, P.S.-H. et al. (1991). High performance sub-half micron CMOS using rapid thermal processing. *1991 IEEE IEDM*, pp. 101–104.
- 334 MacKnight, R.B., Timans, P.J., Tay, S.-P., and Nenyeyi, Z. (2004). RTP applications and technology options for the sub-45 nm nodes. *12th IEEE International Conference on Advanced Thermal Processing of Semiconductors – RTP2004*, pp. 3–36.
- 335 Geipel, H.J., Hsieh, N., Ishaq, M.H. et al. (1980). Composite silicide gate electrodes – interconnections for VLSI device technologies. *IEEE Trans. Electron Devices* 27: 1417–1424.
- 336 Sinha, A.K., Lindenberger, W.S., Fraser, D.B. et al. (1980). MOS compatibility of high-conductivity $\text{TaSi}_2/\text{n}^+\text{poly-Si}$ gates. *IEEE Trans. Electron Devices* 27: 1425–1430.
- 337 Tanaka, T., Ishiuchi, H., Takeuchi, Y. et al. (1981). Characterization of MoSi_2 -gate buried channel MOSFET's for a 256k-bit dynamic RAM. *1981 IEEE IEDM*, pp. 659–662.
- 338 Natori, K., Furuyama, T., Saito, S. et al. (1983). A 34 ns 256K DRAM. *1983 IEEE ISSCC*, pp. 232–233.
- 339 Hayashida, H., Toyoshima, Y., Suizu, Y. et al. (1989). Dopant redistribution in dual gate W-polycide CMOS and its improvement by RTA. *1989 Symposium on VLSI Technology*, pp. 29–30.
- 340 Shibata, T. (1980). Manufacturing method of semiconductor device. Japan Patent S57-99775, filed December 12, 1980.
- 341 Shibata, T., Hieda, K., Sato, M. et al. (1981). An optimally designed process for submicron MOSFETs. *1981 IEEE IEDM*, pp. 647–650.
- 342 Osburn, C., Tsai, M.Y., and Zirinsky, S. (1981). Self-aligned silicide conductors in FET integrated circuits. *IBM Technical Disclosure Bulletin* 24: 1970–1973.
- 343 Lau, C.K., See, Y.C., Scott, D.B. et al. (1982). Titanium disilicide self-aligned source/drain + gate technology. *1982 IEEE IEDM*, pp. 714–717.
- 344 Chapman, R.A., Wei, C.C., Bell, D.A. et al. (1988). 0.5 micron CMOS for high performance at 3.3 V. *1988 IEEE IEDM*, pp. 52–55.

- 345 Iwai, H., Ohguro, T., and Ohmi, S. (2002). NiSi salicide technology for scaled CMOS. *Microelectron. Engineer.* 60: 157–169.
- 346 Morimoto, T., Momose, H.S., Iinuma, T. et al. (1991). A NiSi salicide technology for advanced logic devices. *1991 IEEE IEDM*, pp. 653–656.
- 347 Morimoto, T., Ohguro, T., Momose, H.S. et al. (1995). Self-aligned nickel-mono-silicide technology for high-speed deep submicrometer logic CMOS ULSI. *IEEE Trans. Electron Devices* 42: 915–922.
- 348 Ohguro, T., Nakamura, S., Koike, M. et al. (1994). Analysis of resistance behavior in Ti- and Ni-saliced polysilicon films. *IEEE Trans. Electron Devices* 41: 2305–2317.
- 349 Ohguro, T., Nakamura, S., Morifuji, E. et al. (1995). Nitrogen-doped nickel monosilicide technique for deep submicron CMOS salicide. *1995 IEEE IEDM*, pp. 453–456.
- 350 Wong, C.Y., Sun, J.Y.-C., Taur, Y. et al. (1988). Doping of n^+ and p^+ polysilicon in a dual-gate CMOS process. *1988 IEEE IEDM*, pp. 238–241.
- 351 Sun, J.Y.-C., Wong, C.Y., Taur, Y., and Hsu, C.-H. (1989). Study of boron penetration through thin oxide with p^+ -polysilicon gate. *1989 Symposium on VLSI Technology*, pp. 17–18.
- 352 Morimoto, T., Momose, H.S., Yamabe, K., and Iwai, H. (1990). Prevention of boron penetration from p^+ poly gate by RTN produced thin gate oxide. *ESSDERC '90*, pp. 73–76.
- 353 Morimoto, T., Momose, H.S., Ozawa, Y. et al. (1990). Effects of boron penetration and resultant limitations in ultra-thin pure-oxide and nitrided-oxide gate-films. *1990 IEEE IEDM*, pp. 429–432.
- 354 Kaji, T. (1975). Fabrication method of semiconductor device. Japan Patent JP S52-26182, filed August 25, 1975.
- 355 Iwai, H. (1980). Fabrication method of semiconductor device. Japan Patent JP S57-202755, filed November 29, 1980.
- 356 Kurosawa, K., Shibata, T., and Iizuka, H. (1981). Fabrication method of semiconductor device. Japan Patent JP S57-204135, filed June 10, 1981.
- 357 Davari, B., Koburger, C.W., Schulz, R. et al. (1989). A new planarization technique, using a combination of RIE and chemical mechanical polish (CMP). *1989 IEEE IEDM*, pp. 61–64.
- 358 Chatterjee, A., Esquivel, J., Nag, S. et al. (1996). A shallow trench isolation study for 0.25/0.18 μm CMOS technologies and beyond. *1996 Symposium on VLSI Technology*, pp. 156–157.
- 359 Shibata, T., Nakayama, R., Kurosawa, K. et al. (1983). A simplified box (buried-oxide) isolation technology for megabit dynamic memories. *1983 IEEE IEDM*, pp. 27–30.
- 360 Horiguchi, F., Itoh, Y., Iizuka, H. et al. (1985). A 1Mb DRAM with folded capacitor cell structure. *1985 IEEE ISSCC*, pp. 244–245, 355.
- 361 Saito, S., Fujii, S., Okada, Y. et al. (1985). A 1Mb CMOS DRAM with fast page and static column modes. *1985 IEEE ISSCC*, pp. 252–253.
- 362 Davari, B., Koburger, C., Furukawa, T. et al. (1988). A variable-size shallow trench isolation (STI) technology with diffused sidewall doping for submicron CMOS. *1988 IEEE IEDM*, pp. 92–95.
- 363 Jin, J. and Shin, Y. (1996). Methods of forming trench isolation regions using repatterned trench masks. US Patent 5940716, filed March 14, 1997. Foreign application priority March 15, 1996.
- 364 Andricacos, P.C., Deligianni, H., Dukovic, J.O. et al. (1999). Method of making electroplated interconnection structures on integrated circuit chips. US 6709562, filed July 6, 1999.
- 365 Edelstein, D.C. (2017). 20 Years of Cu BEOL in manufacturing, and its future prospects. *2017 IEEE IEDM*, pp. 334–337.

- 366 Lee, H.S. (1973). An analysis of the threshold voltage for short-channel IGFET's. *Solid-State Electronics* 16: 1407–1417.
- 367 Poon, H.C., Yau, L.D., Johnston, R.L., and Beecham, D. (1973). DC Model for short-channel IGFET's. *1973 IEEE IEDM*, pp. 156–159.
- 368 Yau, L.D. (1974). A simple theory to predict the threshold voltage of short-channel IGFET's. *Solid-State Electronics* 17: 1059–1063.
- 369 Bateman, I.M., Armstrong, G.A., and Magowan, J.A. (1973). Drain voltage limitations of short-channel M.O.S. transistors. *1973 IEEE IEDM*, p. 147.
- 370 Konaka, M., Iwai, H., and Nishi, Y. (1979). Suppression of anomalous drain current in short channel MOSFET. *Japanese Journal of Applied Physics*: 18, Suppl. 18-1, 27–33.
- 371 Sai-Halaszi, G.A., Wordeman, M.R., Kern, D.P. et al. (1987). Design and experimental technology for 0.1- μ m gate-length low-temperature operation FET's. *IEEE Electron Device Lett.* 8: 463–466.
- 372 Hashimoto, T., Sudoh, Y., Kurino, H. et al. (1992). 3 V operation of 70 nm gate length MOSFET with new double punchthrough stopper structure. *1992 SSDM*, pp. 490–492.
- 373 Ono, M., Saito, M., Yoshitomi, T. et al. (1993). Sub-50 nm gate length n-MOSFETs with 10 nm phosphorus source and drain junctions. *1993 IEEE IEDM*, pp. 119–122.
- 374 Ono, M., Saito, M., Yoshitomi, T. et al. (1995). Fabrication of sub-50-nm gate length n-metal-oxide-semiconductor field effect transistors and their electrical characteristics. *J. Vacuum Science and Technology B* 13: 1740–1743.
- 375 Chung, J., Jeng, M.-C., Moon, J.E. et al. (1988). Deep submicrometer MOS device fabrication using a photoresist-ashing technique. *IEEE Electron Device Lett.* 9: 186–188.
- 376 Saito, M., Yoshitomi, Y., Ono, M. et al. (1992). An SPDD p-MOSFET structure suitable for 0.1 and sub 0.1 micron channel length and its electrical characteristics. *1992 IEEE IEDM*, pp. 897–900.
- 377 Hoeneisen, B. and Mead, C.A. (1972). Fundamental limitations in microelectronics – I. MOS technology. *Solid-State Electronics* 15: 819–829.
- 378 Mead, C.A. and Conway, L. (1979). *Introduction to VLSI Systems*, 37. Addison-Wesley.
- 379 Iwai, H. (1994). Solid diffusion method. *Oyo Buturi* 63: 1155–1156. [in Japanese]. 岩井 洋、 「固相拡散法」 応用物理、 第63巻、 1155頁–1156頁、 1994年.
- 380 Morimoto, T., Momose, H.S., Yamabe, K., and Iwai, H. (1990). Ultra thin nitride gate MISFET operating with tunneling gate current. *1990 SSDM*, pp. 361–364.
- 381 Momose, H., Ono, M., Yoshitomi, T. et al. (1996). 1.5 nm direct-tunneling gate oxide Si MOS-FET's. *IEEE Trans. Electron Devices* 43: 1233–1242.
- 382 Morimoto, T., Momose, H., and Iwai, H. (1992). Semiconductor device. Japan Patent JP H05-275690, filed March 27, 1992.
- 383 Momose, H., Iwai, H., Saito, M. et al. (1994). MOSFET with a thin gate insulating film. US Patent US5990516, filed September 13, 1995. Foreign application priority September 13, 1994.
- 384 Momose, H.S., Ono, M., Yoshitomi, T. et al. (1994). Tunneling gate oxide approach. *1994 IEEE IEDM*, pp. 593–596.
- 385 Momose, H.S., Nakamura, S., Ohguro, T. et al. (1997). Uniformity and reliability of 1.5 nm direct tunneling gate oxide MOSFETs. *1997 Symposium on VLSI Technology*, pp. 15–16.
- 386 Timp, G., Agarwal, A., Baumann, F.H. et al. (1997). Low leakage, ultra-thin gate oxides for extremely high performance sub-100 nm nMOSFETs. *1997 IEEE IEDM*, pp. 930–932.
- 387 Timp, G., Bude, J., Bourdelle, K.K. et al. (1999). The ballistic nano-transistor. *1997 IEEE IEDM*, pp. 55–58.

- 388 Yoshitomi, T., Saito, M., Ohguro, T. et al. (1995). Silicided silicon-sidewall source and drain (S^4D) structure for high-performance 75-nm gate length pMOSFETs. *1995 Symposium on VLSI Technology*, pp. 11–12.
- 389 Yoshitomi, T., Saito, M., Ohguro, T. et al. (1996). Hot-carrier reliability of S^4D n-MOSFETs. *ESSDERC '96*, pp. 65–68.
- 390 Davari, B., Hwang, W., and Lu, N.C. (1988). Integrated trench-transistor structure and fabrication process. US Patent 4881105, filed June 13, 1988.
- 391 Mitani, Y., Mizushima, I., Kambayashi, S. et al. (1996). Buried source and drain (BSD) structure for ultra-shallow junction using selective deposition of highly doped amorphous silicon. *1996 Symposium on VLSI Technology*, pp. 176–177.
- 392 Gannavaram, S. and Öztürk, M.C. (2000). Ultra-shallow p^+ -n junctions for 35–70 nm CMOS using selectively deposited very heavily boron-doped silicon-germanium films. *Electrochem. Soc. Symp. Proc.* 2000-9: pp. 73–81.
- 393 Thompson, S.E., Anand, N., Armstrong, M. et al. (2002). A 90 nm logic technology featuring 50 nm strained silicon channel transistors, 7 layers of Cu interconnects, low k ILD, and $1\text{ }\mu\text{m}^2$ SRAM cell. *2002 IEEE IEDM*, pp. 61–64.
- 394 Auth, C., Allen, C., Blattner, A. et al. (2012). A 22 nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors. *2012 Symposium on VLSI Technology*, pp. 131–132.
- 395 Thompson, S.E., Armstrong, M., Auth, C. et al. (2004). A 90-nm logic technology featuring strained-silicon. *IEEE Trans. Electron Devices* 51: 1790–1797.
- 396 Bohr, M. and Mistry, K. (2011). Intel's revolutionary 22nm transistor technology. https://download.intel.com/newsroom/kits/22nm/pdfs/22nm-Details_Presentation.pdf.
- 397 Taur, Y., Wind, S., Mii, Y.J. et al. (1993). High performance 0.1 μm CMOS devices with 1.5 V power supply. *1993 IEEE IEDM*, pp. 127–130.
- 398 Saito, M., Ono, M., Fujimoto, R. et al. (1995). Advantage of small geometry MOSFETs for high-frequency analog applications under low power supply voltage of 0.5 V. *1995 Symposium on VLSI Technology*, pp. 71–72.
- 399 Yoshitomi, T., Sugawara, Y., Morifuji, E. et al. (1998). On-chip spiral inductors with diffused shields using channel-stop implant. *1998 IEEE IEDM*, pp. 540–543.
- 400 Yoshitomi, T., Ebuchi, Y., Kimijima, H. et al. (1999). High performance MIM capacitor for RF BiCMOS/CMOS LSIs. *1999 Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 133–136.
- 401 Orsatti, P., Piazza, F., Huang, Q., and Morimoto, T. (1999). A 20 mA-receive 55 mA-transmit GSM transceiver in 0.25 μm CMOS. *1999 IEEE ISSCC*, pp. 232–233, 464.
- 402 Piazza, F., Orsatti, P., Huang, Q., and Morimoto, T. (1998). A 0.25 μm CMOS transceiver front-end for GSM. *1998 IEEE CICC*, pp. 413–416.
- 403 Huang, Q., Orsatti, P., Piazza, F. and Yoshitomi, T. (1998). Broadband, 0.25 μm CMOS LNAs with sub-2dB NF for GSM applications. *1998 IEEE CICC*, pp. 67–70.
- 404 Steyaert, M., Borremans, M., Janssens, J. et al. (1998). A single-chip CMOS transceiver for DCS-1800 wireless communications. *1998 IEEE ISSCC*, pp. 48–49.
- 405 Momose, H.S., Morifuji, E., Yoshitomi, T. et al. (1996). High-frequency AC characteristics of 1.5 nm gate oxide MOSFETs. *1996 IEEE IEDM*, pp. 105–108.
- 406 Momose, H.S., Morifuji, E., Yoshitomi, T. et al. (2001). Cutoff frequency and propagation delay time of 1.5-nm gate oxide CMOS. *IEEE Trans. Electron Devices* 48: 1165–1174.
- 407 Morifuji, E., Momose, H.S., Ohguro, T. et al. (1999). Future perspective and scaling down roadmap for RF CMOS. *1999 Symposium on VLSI Technology*, pp. 163–164.

- 408 Haartsen, J.C. and Mattisson, S. (2000). Bluetooth – new low-power radio interface providing short-range connectivity. *Proceedings of the IEEE* 88: 1651–1661.
- 409 Fujishia, M. (2021). Overview of sub-terahertz communication and 300 GHz CMOS transceivers. *IEICE Electronics Express* 18: 1–7.
- 410 Abdo, I., da Gomez, C., Wang, C. et al. (2022). A bi-directional 300-GHz-band phased-array transceiver in 65-nm CMOS with outphasing transmitting mode and LO emission cancellation. *IEEE J. Solid-State Circuits* 57: 2292–2308.
- 411 IEEE IRDS. <https://irds.ieee.org/>.
- 412 Iwai, H. and Ohmi, S. (2002). Silicon integrated circuit technology from past to future. *Microelectronics Reliability* 42: 465–491.
- 413 Gelsinger, P.P. (2001). Microprocessor for the new millennium: Challenges, opportunities, and new frontiers. *2001 IEEE ISSCC*, 1.3, pp. 22–25.
- 414 Rupp, K. (2015). 40 Years of microprocessor trend data. <https://www.karlrupp.net/2015/06/40-years-of-microprocessor-trend-data/>.
- 415 Iwamatsu, S. (1980). Thin-film field effect transistor semiconductor device. Japan Patent JP S56-111261, filed January 7, 1980.
- 416 Saito, A. (1980). Double-layer gate polysilicon MOS transistor. Japan Patent JP S57-90977, filed November 27, 1980.
- 417 Sekigawa, T. and Y. Hayashi (1983). Suppression methods of short-channel effects in field-effect transistors. Japan Patent JP H07-050785, filed October 27, 1983.
- 418 Balestra, F., Cristoloveanu, S., Benachir, M. et al. (1987). Double-gate silicon-on-insulator transistor with volume inversion: A new device with greatly enhanced performance. *IEEE Electron Device Letters* 8: 410–412.
- 419 Colinge, J.P., Gao, M.H., Romano-Rodriguez, A. et al. (1990). Silicon-on-insulator “gate-all-around device.” *1990 IEEE IEDM*, pp. 595–598.
- 420 Frank, D.J., Laux, S.E., and Fischetti, M.V. (1992). Monte Carlo simulation of a 30 nm dual-gate MOSFET: How short can Si go? *1992 IEEE IEDM*, pp. 553–556.
- 421 Ono, M., Saito, M., Yoshitomi, T. et al. (1994). Influence of high substrate doping concentration on the hot-carrier and other characteristics of small-geometry CMOS transistors down to the 0.1 μm generation. *1994 Symposium on VLSI Technology*, pp. 147–148.
- 422 Sekigawa, T. and Hayashi, Y. (1984). Calculated threshold-voltage characteristics of an X MOS transistor having an additional bottom gate. *Solid State Electronics* 27: 827–828.
- 423 Izumi, K. (1976). Fabrication method of semiconductor device. Japan Patent JP S53-31964, filed September 6, 1976.
- 424 Izumi, K., Doken, M., and Ariyoshi, H. (1978). C.M.O.S. devices fabricated on buried SiO_2 layers formed by oxygen implantation into silicon. *Electronics Letters* 14: 593–594.
- 425 Colinge, J.P. (1989). Method of manufacturing a field effect transistor and a semiconductor element. European Patent EP0418983, filed September 20, 1990, priority NL8902372 filed September 21, 1989.
- 426 Sunouchi, K., Takato, H., Okabe, N. et al. (1989). A surrounding gate transistor (SGT) cell for 64/256 Mbit DRAMs. *1989 IEEE IEDM*, pp. 23–26.
- 427 Nitayama, A., Takato, H., Horiguchi, F., and Masuoka F. (1988). Semiconductor device. Japan Patent JP 3057661, filed September 6, 1988.
- 428 Fiegna, C., Iwai, H., Wada, T. et al. (1993). A new scaling methodology for the 0.1 – 0.025 μm MOSFET. *1993 Symposium on VLSI Technology*, pp. 33–34.
- 429 Kim, S., Park, D., Lee, C. et al. (2002). Field effect transistors having multiple stacked channels. US Patent 7002207, filed July 1, 2003. Foreign application priority October 1, 2002.

- 430 Lee, S., Kim, S., Yoon, E. et al. (2003). A novel multibridge-channel MOSFET (MBCFET): fabrication technology and characteristics. *IEEE Trans. Nanotechnology* 2: 253–257.
- 431 Lee, S., Kim, S., Yoon, E. et al. (2004). Three-dimensional MBCFET as an ultimate transistor. *IEEE Electron Device Letters* 25: 217–219.
- 432 Barraud, S., Previtali, B., Vizioz, C. et al. (2020). 7-Levels-stacked nanosheet GAA transistors for high performance computing. *2020 IEEE Symposium on VLSI Technology*, TC1.2.
- 433 Huang, C.-Y., Dewey, G., Mannebach, E. et al. (2020). 3-D self-aligned stacked NMOS-on-PMOS nanoribbon transistors for continued Moore's law scaling. *2020 IEEE IEDM*, pp. 425–428.
- 434 Liao, S., Yang, L., You, W.X. et al. (2024). First demonstration of monolithic CFET inverter at 48 nm gate pitch toward future logic technology scaling. *2024 IEEE IEDM*, pp. 1–4.
- 435 Samsung begins chip production using 3nm process technology with GAA architecture. <https://semiconductor.samsung.com/news-events/news/samsung-begins-chip-production-using-3nm-process-technology-with-gaa-architecture/>.
- 436 2nm technology. https://www.tsmc.com/english/dedicatedFoundry/technology/logic/l_2nm.
- 437 Agrawal, A., Chakraborty, W., Li, W. et al. (2024). Silicon RibbonFET CMOS at 6 nm gate length. *2024 IEEE IEDM*, pp. 1–4.
- 438 Hayashi, Y. (1980). Insulated-gate field effect transistor. Japan Patent JP S57-10973, filed June 24, 1980.
- 439 Hasegawa, M. (1987). MOS field effect transistor. Japan Patent JP S64-8670, filed July 1, 1987.
- 440 Hazama, H. (1987). Field effect semiconductor device. Japan Patent JP H6-9245, filed July 23, 1987.
- 441 Hisamoto, D., Kaga, T., Kimura, S. et al. (1988). Semiconductor device and manufacturing method thereof. Japan Patent JP 2768719, issued April 17, 1998, priority November 21, 1988.
- 442 Hisamoto, D., Kaga, T., Kawamoto, Y., and Takeda, E. (1989). A fully depleted lean-channel transistor (DELTA) – A novel vertical ultra thin SOI MOSFET. *1989 IEEE IEDM*, pp. 833–836.
- 443 Hu, C., King, T., Subramanian, V. et al. (2000). FinFET transistor structures having a double gate channel extending vertically from a substrate and methods of manufacture. US Patent 6413802, filed October 23, 2000.
- 444 Hisamoto, D., Lee, W., Kedzierski, J. et al. (2000). FinFET – a self-aligned double-gate MOS-FET scalable to 20 nm. *IEEE Trans. Electron Devices* 47: 2320–2325.
- 445 Lee, J. (2002). Double-gate FinFET device and fabricating method thereof. US Patent 6885055, filed February 4, 2003. Although US 6885055 was filed without foreign application priority, almost the same patent was filed by the inventor to Korean Patent Office on January 30, 2002 (KR 10-2002-0005325).
- 446 Park, T., Yoon, E., and Lee, J. (2003). A 40 nm body-tied FinFET (OMEGA MOSFET) using bulk Si wafer. *Physica E: Low-dimensional Systems and Nanostructures* 19: 6–12.
- 447 Rose, K., Mangaser, R., Mark, C., and Sayre, E. (1999). Cryogenically cooled CMOS. *Critical Reviews in Solid State and Materials Sciences* 24: 63–99.
- 448 Shu, Y., Ning, B., Li, Y., et al. (2024). Overview of Cryogenic CMOS Based Computing Systems. *Integrated Circuits and Systems* 1: 167–177.
- 449 Patra, B., Rosario, M., Incandela, R.M., van Dijk, J.P.G. et al. (2018). Cryo-CMOS circuits and systems for quantum computing applications. *IEEE J. Solid-State Circuits* 53: 309–321.
- 450 Mistry, K., Allen, C., Auth, C. et al. (2007). A 45 nm logic technology with high-k+metal gate transistors, strained silicon, 9 Cu interconnect layers, 193 nm dry patterning, and 100% Pb-free packaging. *2007 IEEE IEDM*, pp. 247–250.
- 451 Sugano, T. and Hoh, K. (1970). Interfacial carrier mobility of ZrO₂/Si structure and N-channel enhancement FET using this. *Annual Report of Engineering, Faculty of Engineering, University of Tokyo* 29: 137–141.

- 452 Hoh, K. and Sugano, T. (1970). Low temperature vapor deposition of ZrO_2 films on Si surface. *1st Conference on Solid State Devices*, Suppl. *Japanese Journal of Applied Physics* 39: 177–184.
- 453 Linder, B.P., Dasgupta, A., Ando, T. et al. (2016). Process optimizations for NBTI/PBTI for future replacement metal gate technologies. *2016 IEEE International Reliability Physics Symposium (IRPS)*, pp. 4B1-1–4B1-5.
- 454 Ando, T. (2012). Ultimate scaling of high- κ gate dielectrics: higher- κ or interfacial layer scavenging? *Materials* 5: 478–500.
- 455 Wong, A.K.-K. (2001). *Resolution enhancement techniques in optical lithography*. SPIE Press.
- 456 Giannopoulos, I., Mochi, I., Vockenhuber, M. et al. (2024). Extreme ultraviolet lithography reaches 5nm resolution. *Nanoscale* 16: 15533–15543.
- 457 Woon, W.-Y., Kasperovich, A., Wen, J.R. et al. (2025). Thermal management materials for 3D-stacked integrated circuits. *Nature Reviews Electrical Engineering* 2: 598–613.
- 458 Ha, H., Kim, H., Lee, S. et al. (2025). Overview of thermal management solution for 3D integrated circuits using carbon-nanotube-based silicon through-vias. *Micromachines* 16 (968): 1–16.
- 459 Xia, Y., Sheng, Y., Jia, R. et al. (2025). Non-Fourier thermal analysis of the self-heating effect in CMOS logic circuits. *APL Electronic Devices* 1 (036115): 1–9.
- 460 Li, C. (2024). A review of physical heat dissipation methods for 3D integrated technology. *ACE Proc.* 89: 55–59.
- 461 Lee, C.Y., Won, C.H., Jung, S. et al. (2025). 3D Integrated process and hybrid bonding of high bandwidth memory (HBM). *Electronic Materials Letters* 21: 395–419.
- 462 Lau, J.H. (2021). *Semiconductor Advanced Packaging*. Springer Singapore. <https://doi.org/10.1007/978-981-16-1376-0>.
- 463 Yazdani, S. (2018). *Foundations of Heterogeneous Integration: An Industry-Based, 2.5D/3D Pathfinding and Co-Design Approach*. Springer Cham. <https://doi.org/10.1007/978-3-319-75769-8>.
- 464 Zhao, Y., Zou, L., and Yu, B. (2025). Physical design for advanced 3D ICs: challenges and solutions. Invited paper. *2025 International Symposium on Physical Design ISPD '25*, pp. 209–216.
- 465 Zhu, X., Lin, J.W., and Sander, M.Y. (2022). Bidirectional modulation of evoked synaptic transmission by pulsed infrared light. *Scientific Reports* 12: 14196.
- 466 de Kock, C.P.J. and Feldmeyer, D. (2023). Shared and divergent principles of synaptic transmission between cortical excitatory neurons in rodent and human brain. *Frontiers in Synaptic Neuroscience* 15: 1274383, 1–10.
- 467 Bose, J.C. (1901). Detector for electrical disturbances. US Patent 755840, filed September 30, 1901.
- 468 Dunwoody, H.H.C. (1906). Wireless-telegraph system. US Patent 837616, filed March 23, 1906.
- 469 Pickard, G.W. (1906). Means for receiving intelligence communicated by electric waves. US Patent 836531, filed August 30, 1906.
- 470 Allgemeine Elektrizität-Gesellschaft. (1929). Kontaktgleichrichter mit zwei durch eine Sperrschicht getrennten Metallelektroden. Österreichisches Patentamt AT13102, filed July 11, 1930, priority July 11, 1929.
- 471 Hilsch, R. and Pohl, R.W. (1938). Steuerung von Elektronenströmen mit einem Dreielektrodenkristall und ein Modell einer Sperrschicht. *Zeitschrift für Physik* 111: 399–408.
- 472 Peierls, R. (1929). Zur Theorie der galvanomagnetischen Effekte. *Zeitschrift für Physik* 53: 255–266.
- 473 Pauli, W. (1925). Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Zeitschrift für Physik* 31: 765–783.

- 474 Guicciardini, N. and Introzzi, G. (1995). The Fermi-Dirac statistics: a simultaneous discovery. In: *The Foundations of Quantum Mechanics — Historical Analysis and Open Questions. Fundamental Theories of Physics*, vol. 71 (ed. C. Garola and A. Rossi), 357–367. Dordrecht: Springer.
- 475 Dirac, P. (1926). On the theory of quantum mechanics. *Proceedings of the Royal Society of London* 112: 661–677.
- 476 Fermi, E. (1926). Sulla quantizzazione del gas perfetto monoatomico. *Rend. Lincei* 3: 145–149.
- 477 Sommerfeld, A. (1928). Zur Elektronentheorie der Metalle auf Grund der Fermischen Statistik. *Zeitschrift für Physik* 47: 1–32.
- 478 Bloch, F. (1929). Über die Quantenmechanik der Elektronen in Kristallgittern. *Zeitschrift für Physik* 52: 555–600.
- 479 Brillouin, L. (1930). Les électrons libres dans les métaux et le rôle des réflexions de Bragg. *J. Phys. Rad.* 1: 377–400.
- 480 Wigner, E. and Seitz, F. (1933). On the constitution of metallic sodium. *Physical Review* 43: 804–810.
- 481 Slater, J.C. (1937). Wave functions in a periodic potential. *Physical Review* 51: 846–851.
- 482 Slater, J.C. and Koster, G.F. (1954). Simplified LCAO method for the periodic potential problem. *Physical Review* 94: 1498–1524.
- 483 Herman, F. (1955). The electronic energy band structure of silicon and germanium. *Proc. of IRE* 43: 1703–1732.
- 484 Davydov, B. (1938). On the rectification of current at the boundary between two semiconductors. *Compt. Rend. Doklady Acad. Sci., URSS* 20: 279–282.
- 485 Davydov, B. (1938). The rectifying action of semiconductors. *Tech. Phys. USSR* 5: 87–95.
- 486 Mott, N. (1938). Note on the contact between a metal and an insulator or semiconductor. *Proceedings of the Cambridge Philosophical Society* 34: 568–572.
- 487 Mott, N. (1939). The theory of crystal rectifiers. *Proceedings of the Royal Society of London* 171: 27–38.
- 488 Leobandung, E., Barth, E., Sherony, M. et al. (1999). High performance 0.18 μm SOI CMOS technology. *1999 IEEE IEDM*, pp. 679–682.
- 489 Kalla, R., Sinharoy, B., Starke, W.J., and Floyd, M. (2010). Power7: IBM's next-generation server processor. *IEEE Micro* 30: 7–15.
- 490 Sinharoy, B., Van Norstrand, J.A., Eickemeyer, R.J. et al. (2015). IBM POWER8 processor core microarchitecture. *IBM J. Res. & Dev.* 59: 2:1–2:21.
- 491 Fuse, T., Kameyama, A., Ohta, M., and Ohuchi, K. (2001). A 0.5 V power-supply scheme for low power LSIs using multi- V_t SOI CMOS technology. *2001 Symposium on VLSI Circuits*, pp. 219–220.
- 492 Saito, M. (2002). Progress of semiconductor technology and system on chip. *Toshiba Review* 57 (1): 38–42. [in Japanese]. 斎藤 光男 「半導体技術の進歩とシステムオンチップ」 東芝レビュー.
- 493 Gruber, H. (1996). *The Semiconductor Industry: Review of the Sector and Financing Opportunities*. European Investment Bank. https://www.eib.org/files/pj/semiconductor_industry_en.pdf.
- 494 ESPAS (European Strategy and Policy Analysis System). (2022). *Global semiconductor trends and the future of EU chip capabilities*. <https://espas.eu/files/Global-Semiconductor-Trends-and-the-Future-of-EU-Chip-Capabilities-2022.pdf>.