Technology of High-*k*/Metal-Gate Stack

Technology of High-*k*/Metal-Gate Stack

Mingshi Wang

College of Physics and Information Engineering, Fuzhou University, Fuzhou, 350108, Fujian Province, China

112100403@fzu.edu.cn

Abstract. The High-k/Metal-Gate Stack (HKMG) technology represents a seminal advancement in semiconductor fabrication, predicated upon the substitution of SiO2 with high-k dielectric materials, coupled with the integration of metal gate materials. This technological paradigm shift has emerged as a transformative response to the intricate challenges encountered within the semiconductor industry, a sector perpetually engaged in the relentless pursuit of diminishing the critical dimensions of semiconductor devices. Issues such as increased power leakage, enhanced parasitic capacitance, and the escalating complexity of gate insulators have impeded the sustainable downscaling of traditional silicon dioxide-based devices. These challenges have necessitated a comprehensive re-evaluation of the materials and processes involved in semiconductor manufacturing, leading to the conception and evolution of the HKMG technology. This article's analysis of the developmental trajectory of HKMG technology reveals a progression marked by the systematic exploration of alternative high-k dielectric materials, rigorous compatibility assessments with metal gate materials, and the concurrent refinement of manufacturing processes to ensure seamless integration. After that, elucidating the impediments encountered during its evolution and expounding upon the corresponding ameliorative measures instituted. Looking ahead, ongoing research focusing on the synthesis of advanced high-k dielectric materials, and the development of novel metal gate materials with superior compatibility to further enhance device performance and functionality makes HKMG technology a huge prospect. Such advancements are paving the way for the realization of more powerful, energy-efficient, and compact electronic devices.

Keywords: HKMG, CMOS, Fabrication Process

1. Introduction

The main device in a very large-scale integrated circuit is a metal oxide semiconductor field effect transistor (MOSFET). With the rapid development of the semiconductor integrated circuit industry, integrated circuits are required to have an increasing component density, that is, the dimensions of each component itself in the chip and the dimensions between components need to continue to shrink. Therefore, the critical dimension of mainstream high-end chips has shrunk from tens of microns in the 1960s to several nanometers nowadays. However, the industry's requirements for the thickness of transistor components have reached several molecules and several atoms, so many materials that make up transistors have reached their physical limits. [1] At 65nm, the gate oxide layer was the first to reach the physical limit. [2] Due to these physical limitations, original manufacturing processes either failed to produce transistors with smaller feature sizes or produced transistors with serious defects that

prevented them from functioning properly, such as extremely large leakage currents. Therefore, improvement of the original large-scale integrated circuit manufacturing process is a very necessary pursuit. HKMG is a key advancement that greatly improves the original manufacturing process, allowing large-scale integrated circuits to break through the physical limits of the original process and further reduce the feature size. Next, let's first introduce why these physical limits have a significant impact on chip manufacturing: quantum tunneling effect and polysilicon gate depletion effects.

1.1. Quantum Tunnelling Effect

Quantum tunnelling effect, as shown in Figure 1, is a quantum mechanical phenomenon, particles such as electrons can, with a small probability, tunnel to the other side, thus crossing the barrier.

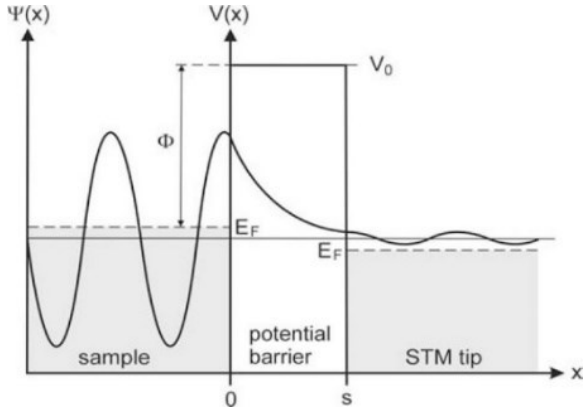


Figure 1. Schematic diagram of quantum tunneling effect

This tunneling leaves the barrier unaffected. [3] The critical dimension of modern integrated circuits continues to decrease, and the operating voltage continues to decrease, so the thickness of the oxide layer also needs to decrease. The conventional oxide layer uses polysilicon with a thickness greater than 4nm. At this size, the bandgap width of SiO2 is as high as 9eV, and the bandgap width of Si is 1.12eV. A huge barrier height with good insulation properties will be formed between them. Under the condition of the normal bias voltage of the device, electrons or holes are impossible to cross the barrier formed by the gate oxide layer and silicon, so a gate leakage current will not form. But, when the thickness of pure silicon dioxide is less than 3nm, the quantum tunneling effect is enhanced, electrons may pass through the medium and enter the gate, and the gate leakage current cannot be ignored.

1.2. Polysilicon Depletion Effect

The polysilicon depletion effect, as shown in Figure 2, is a phenomenon in that unwanted variation of the threshold voltage happens in the MOSFET devices using polysilicon as gate material. [4] The conductivity of the poly-silicon layer is very low and because of this low conductivity, the charge accumulation is low, leading to a delay in channel formation and thus unwanted delays in circuits. [5]

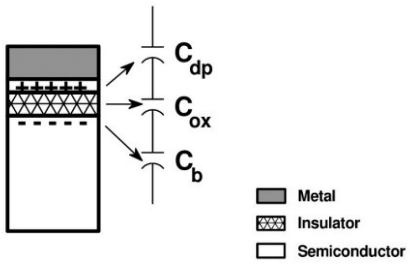


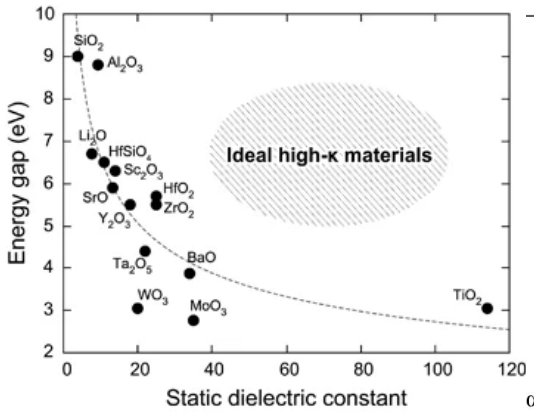
Figure 2. Schematic diagram of polysilicon depletion effect

The voltage applied to the gate will attract minority carriers to the interface between the dielectric and the channel, forming an inversion carrier distribution. At this time, an accumulation of carriers will form near the interface between the dielectric and poly-Si to maintain electrical neutrality, which will inevitably deplete the charge of the nearby semiconductor. When the charge of the semiconductor is completely depleted, the semiconductor is equivalent to an insulator, which is equivalent to increasing the effective thickness of the gate dielectric. Although the thickness of the depletion layer is only a few angstroms of SiO2, the effect of this thickness becomes very significant when the thickness of the gate dielectric is reduced to more than ten angstroms. Although the density of free carriers in the material can be increased to alleviate the depletion effect of polysilicon by increasing the doping density of the polysilicon gate, the doping in the gate is already close to saturation, especially for PMOS. Because its doping density must be controlled below 10^{19}cm^{-3}, this means that there is no way to solve the depletion effect of polysilicon by increasing the doping density.

In addition to the above two serious problems, the semiconductor production technology before HKMG also faced various difficult-to-solve problems such as substrate quantum effects and diffusion of doped boron atoms. What's more, The thinner the gate oxide layer, the greater the impact of the gate depletion layer and substrate quantum effects on the gate capacitance, and the smaller its equivalent oxide thickness (EOT). For a very thin gate oxide layer, the gate depletion layer and quantum effects already account for about 30% of the capacitor effective thickness (CET). CET cannot be reduced by reducing the thickness of the gate dielectric. Therefore, in order to find solutions to these problems as soon as possible, further reduce the size of semiconductors, and keep up with Moore's Law, the industry tried many improvements in the chip manufacturing process, and finally, researchers introduced HKMG technology. This technology replaces the gate dielectric material and performs a series of process upgrades for this change, such as the introduction of metal gates. These major technological advances allow HKMG to solve the above problems and lead the semiconductor industry to continue to develop rapidly in a new stage.

2. Technology of High-k/Metal-Gate

Research has found that increasing the dielectric constant and barrier height of the dielectric can effectively solve the above problems. However, continuing to use SiO2 as the gate dielectric cannot meet these requirements, and the industry needs to find better insulators than SiO2 to better separate the gate and other parts of the transistor. In addition, alternative materials need to have better field-effect properties than dioxide. To summarize the above requirements, we need to find a material with high-*k* (higher than SiO2) properties, enough (>5.5ev) bandgap, and stability. Figure 3 shows the static dielectric constant k and experimental band gap of common gate dielectric. After many experiments by researchers, it was found that HfO2 has a band gap of 5.8ev, a high dielectric constant of *k*=25, and the same excellent thermodynamic and chemical stability as Si. Therefore, HfO2 is the best choice to replace SiO2 as the gate dielectric.



	k	Gap (eV)
SiO_2	3.9	9
Si_3N_4	7	5.3
Al_2O_3	9	8.8
Ta_2O_5	22	4.4
TiO_2	80	3.5
$SrTiO_3$	2000	3.2
ZrO_2	25	5.8
HfO_2	25	5.8
$HfSiO_4$	11	6.5
La_2O_3	30	6
Y_2O_3	15	6
α-$LaAlO_3$	30	5.6

Figure 3. The experimental band gap and dielectric constant for well-known oxides

After finding such a perfect alternative material, the researchers were disappointed to find that the material could not be directly used in the production process. This medium is incompatible with polysilicon gates. First, due to the rough interface, more defect centers will be generated, which will intensify lattice scattering and reduce carrier mobility. Second, the formation of Si-Hf bonds will cause the Fermi-Level Pinning leading the threshold voltage to shift, and the threshold voltage cannot be adjusted through doping. Third, phonon scattering in high-K media is intensified. To solve the above problems, researchers have discovered through many experiments that using metal instead of polysilicon as the gate material can solve problems caused by Fermi-Level pinning. In addition, doping a very thin layer of SiON film between the high-*k* medium and the substrate to smooth the interface can also improve the phonon scattering problem. This film is formed by the ISSG process.

To sum up, this transistor structure using high-k oxide dielectric and metal gate is called HKMG structure. Figure 4 shows the comparison between traditional transistors and HKMG structure transistors. Compared with traditional polysilicon gate and SiO2 gate oxide structures, HKMG technology can not only significantly reduce gate leakage at very small critical dimensions, effectively ameliorate Effective Capacitance Thickness (ECT), but also improve its driving capability. The emergence of HKMG technology continues Moore's Law.

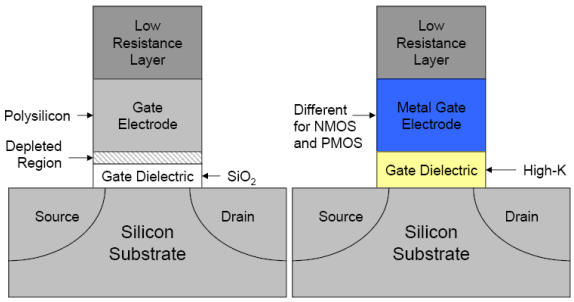


Figure 4. Comparison diagram between traditional transistors and transistors with HKMG structure

3. Production process using HKMG technology

When manufacturing a MOSFET, the source and drain region structures must first be formed in the substrate through ion implantation, and then high-temperature annealing is performed to activate the injected ions. [6] After the industry began to apply HKMG technology on a large scale, two process technology routes emerged, namely the gate-first process and the gate-last process. The gate-first process forms the metal gate before the source and drain region ion implantation and high-temperature annealing steps, while the gate-last process forms the metal gate after this step. The former is represented by IBM, and the latter is represented by Intel.

3.1. Gate-First Process Technology (GF)

The gate-first approach, originally pioneered by the Fishkill Alliance under the aegis of IBM, hinges upon the judicious utilization of exceptionally thin capping layers to induce dipolar formations for precise control of the threshold voltage within the semiconductor device. Concretely, this entails the deployment of Al2O3 for PMOS and LaOx for NMOS devices. Nonetheless, the notable concern arises from the propensity of thermal instabilities inherent to High-k/Metal-Gate (HKMG) configurations, which can engender unwarranted shifts in threshold voltage and the resurgence of the gate stack material. This predicament assumes a pronounced character when addressing PMOS devices with reduced Equivalent Oxide Thickness (EOT), thereby presenting a formidable obstacle to the seamless adoption of the gate-first approach in high-performance applications. Nevertheless, for Low Standby Power (LSTP) or Dynamic Random-Access Memory (DRAM) applications characterized by less stringent EOT prerequisites, the gate-first strategy continues to present itself as an economically viable solution for the integration of HKMG Complementary Metal-Oxide-Semiconductor (CMOS) technology.[7]

3.2. Gate-Last Process Technology (GL)

Intel introduced a second production process utilizing HK/MG (High-K Metal Gate) technology for its 45nm semiconductor products, commonly referred to as the "gate-last" process. In this approach, a hafnium-based dielectric material is deposited prior to the creation of a sacrificial polysilicon gate structure. Specifically, following a series of high-temperature Source-Drain (S-D) as well as the silicide annealing cycles, the placeholder gate structure is removed, and subsequently, the metal gate electrode is deposited. This methodology offers several advantages, notably enhanced device reliability and electron mobility at reduced Equivalent Oxide Thickness (EOT), a parameter critical for miniaturization, which can be compromised when the high-k dielectric undergoes the high-temperature processing steps associated with conventional "gate-first" methods. Nonetheless, the gate-last process is not without its drawbacks, principally related to its inherent process complexity. Intel has noted that the creation of dual metal gates within this approach necessitates certain demanding Chemical Mechanical Polishing (CMP) steps, which, in turn, mandate more rigorous design rules (RDR) and, often, the adoption of 1D design techniques. This complexity has been further exacerbated at the 22nm semiconductor node due to limitations in the lithographic processes. Consequently, the greater design flexibility traditionally associated with the gate- The first approach may progressively diminish in significance as the industry increasingly confronts the imperative of adhering to stringent RDR to meet evolving technological demands.[7]

UMC unveiled a novel hybrid strategy for incorporating high-k and metal gate (HK/MG) technology into semiconductor manufacturing processes. This innovative approach combines elements of both gate-first (for NMOS transistors) and gate-last (for PMOS transistors) methodologies. The primary objective is to address a critical challenge associated with gate-first techniques, specifically the elevated threshold voltage (VT) of PMOS transistors at reduced equivalent oxide thickness (EOT) scales. This is achieved without resorting to the full and intricate gate-last CMOS integration process, which involves multiple chemical mechanical polishing (CMP) steps and the deposition of dual metal gates. It is worth noting that this approach, akin to Intel's 45nm process, predominantly follows a high-k first strategy. However, it is necessary to consider the temperature stability of the high-k dielectric layers. Unless substantial advancements have been made to enhance their thermal robustness, scalability to nodes below 32nm could pose substantial challenges. At these advanced semiconductor nodes, issues such as reliability and carrier mobility typically deteriorate rapidly as the EOT target is approached. It is worth noting that there are many innovations in the production process using HKMG technology that are not introduced in this article. These innovative processes are fully combined with other advanced technologies in the semiconductor industry, allowing today’s feature sizes to reach 3nm or even lower.

4. Problems and Improvement of HKMG technology

Whether it is the gate-first process or the gate-last process in HKMG technology, they are of great significance to improving the performance of transistors. However, in the development of continuous reduction of process nodes, HKMG technology also faces some problems: such as device reliability issue, interface trap problem, Fermi-Level pinning effect, etc. Some of these problems have been solved to a large extent, but some still need to find better improvement solutions to further improve the performance of the chip.

4.1. Reliability Issue

Device stability is a critical issue in HKMG technology, which is mainly reflected in two aspects. One is bias temperature instability (BTI), including positive bias temperature instability (PBTI) which usually exists in NMOS devices, and negative bias temperature instability (NBTI) which exists in PMOS devices. The second is time-dependent dielectric breakdown (TDDB). For example, HKMG technology that uses HfO2 as a gate dielectric has many advantages and has been widely used in the industry. However, HfO2 has material characteristics such as its low crystallization temperature, which is not conducive to high-temperature processing in subsequent processes; its direct contact interface with the Si substrate is poor, and it is easy to generate defects such as high concentrations of oxygen vacancies

during the film deposition process. [8] These negative material properties pose a huge challenge to the reliability of devices using HfO2 as the gate dielectric. At present, the reliability problems of HKMG devices using HfO2 as the gate dielectric are mainly: threshold voltage hysteresis, positive bias or negative temperature instability, hot carrier effect, leakage current caused by stress, and time-varying breakdown.[9]

4.2. Interface Trap Problem

The characteristics of CMOS devices are affected by various traps in HKMG. Different HKMG preparation processes have different effects on the random interface traps appearing at the SiOx/Si interface. Fin field effect transistors (FinFETs) with HKMG structure play a huge role in improving electrical performance and controlling various fluctuations at the 22nm process node. Sheng-Chia Hsu and others conducted a simulation analysis and concluded that the 16nm HKMG structure bulk FinFET device can effectively suppress the DC characteristics and SCE parameter fluctuations caused by RITs.[10]

4.3. Fermi-Level Pinning

In the original semiconductor, the position of the Fermi energy level in the energy band is easily shifted, and N-type or P-type semiconductors can be made by simply adding different types of impurities. However, if Fermi-Level pinning occurs, even if many donors or acceptors are doped, these impurities cannot be activated to provide carriers, and thus cannot change the position of the Fermi level. [11] The best way to solve the Fermi pinning effect is to adjust the threshold voltage of the transistor so that its value is symmetrical and as low as possible. In order to adjust the overall work function of the gate, a work function metal layer is introduced between the high-k dielectric layer and the metal gate material layer. Different metals with adjustable work functions are used as the work function material of the transistor to adjust the work function of the transistor and thereby adjust the threshold voltage. The work function of the metal gate in PMOS transistors needs to be in the range of 4.8~5.1eV, and the work function of the metal gate in NMOS transistors needs to be in the range of 4.0~4.3eV. However, in fact, because the Al atoms used as the metal gate diffuse into the work function layer as well as the high-k gate dielectric layer, reducing the overall work function value of the metal gate, the work function of the PMOS transistor can only float around 4.8eV and cannot reach With higher values, this phenomenon also directly affects the performance of CMOS transistors. Gao Hanjie [12] et al. disclosed a method by forming a diffusion barrier layer on the work function layer and wrapping it around the metal gate material layer, thereby avoiding the defect of the overall work function of the metal gate being reduced due to the diffusion of metal atoms. [13] At the same time, it helps to reduce the Fermi pinning effect between polycrystalline and high-k.

4.4. Promising Problem-Solving Strategy

The above-mentioned problems existing in HKMG affect its application development to a certain extent, but it is also the existence of these problems that leads researchers in the HKMG industry to adopt various technical means to improve its performance and promote its development. A prospective approach involves advancing research and development endeavors to discover novel high-k dielectric materials and compatible metal gate materials that can augment performance while upholding cost efficiency. Simultaneously, directing investments toward the advancement of pioneering process integration methodologies, such as multi-patterning and self-aligned techniques, can alleviate the intricacies entwined with the fabrication of HKMG-based devices. These process ameliorations facilitate the seamless amalgamation of high-k dielectrics and metal gate materials, thereby optimizing device performance and dependability while diminishing manufacturing intricacies. In summation, the future of HKMG technology relies on the strategic execution of innovative materials, advanced manufacturing techniques, and inventive process integration methodologies.

5. Conclusion

This article mainly discusses the reasons for the invention of HKMG technology, the original technical difficulties it solved, the application of HKMG technology represented by GF and GL in production, and also objectively analyzes the problems, improvements, and future development of HKMG technology. There is no doubt that at key nodes in industry development such as 45nm, HKMG technology is one of the main driving forces that allows the semiconductor industry to continue to follow Moore's Law and flourish. However, in recent years, the global semiconductor industry has also recognized that Moore's Law is coming to an end because the exponential growth in the number of transistors can no longer be sustained. But from the consumer side, the rapid development of smartphones, supercomputers, and cloud data centers in recent years has still driven the rapid development of the chip industry. As long as the semiconductor industry continues to fill its devices with new capabilities, Moore's Law will not completely limit the development of the industry. Although HKMG technology has developed relatively maturely, for the author of this article, it is far from reaching the endpoint. As the semiconductor industry continues to improve in all aspects of structure and process, as this technology continues to be integrated and explored with other device integration technologies (such as FinFET, etc.), and with the strategic deployment of roadmaps for new applications, HKMG technology will support the semiconductor industry to keep moving forward.

Authors' Contributions

Wang Mingshi conceived and refined the idea, collected the data, and wrote the paper.

References

[1] Wang Jianxia. "Overview of HKMG Technology Development." Science and Fortune 000.024(2017):296-296,297.
[2] Zhao Jun. "High-K-metal gate and 45 nanometers." Information Systems Engineering. 03(2008):72-75.
[3] Wikipedia contributors. "Quantum tunnelling." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 7 Sep. 2023. Web. 17 Sep. 2023.
[4] Wikipedia contributors. "Polysilicon depletion effect." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 24 Aug. 2023. Web. 17 Sep. 2023.
[5] R. Rios, N. D. Arora and Cheng-Liang Huang, "An analytic polysilicon depletion effect model for MOSFETs," in IEEE Electron Device Letters, vol. 15, no. 4, pp. 129-131, April 1994, doi: 10.1109/55.285407.
[6] Robertson, John, and Robert M. Wallace. "High-K materials and metal gates for CMOS applications." Materials Science and Engineering: R: Reports 88 (2015): 1-41.
[7] Hoffman, Thomas Y. "Integrating high-k/metal gates: gate-first or gate-last?." Solid State Technology 53.3 (2010): 20-21.
[8] Bretos, Iñigo, et al. "Low-temperature crystallization of solution-derived metal oxide thin films assisted by chemical processes." Chemical Society Reviews 47.2 (2018): 291-308.
[9] Tao Fenfen. Reliability research of high-k/metal gate. Diss. Shandong University, 2014.
[10] Hsu, Sheng-Chia, and Yiming Li. "Electrical characteristic fluctuation of 16-nm-gate high-κ/metal gate bulk FinFET devices in the presence of random interface traps." Nanoscale research letters 9 (2014): 1-8.
[11] Kasahara, K., et al. "Mechanism of Fermi level pinning at metal/germanium interfaces." Physical Review B 84.20 (2011): 205301.
[12] Gao Hanjie, Zhao Jie, Song Weiji. Semiconductor devices and their formation methods, China, CN104752179A, 2015.
[13] Li Dayuan; Chen Jianhao; Chen Qiqun; Ye Mingxi; Li Xingrui. Method of manufacturing semiconductor device, China, CN101752317 A, 2010.