Vol. 11, No. 1 – April 2026
ISSN: 2503-5274(p), 2657-1900(e)

Jurnal Fisika
Fisika Sains dan Aplikasinya

QUANTUM TUNNELING IN FLASH MEMORY TECHNOLOGY: ENABLING SCALABLE NON-VOLATILE STORAGE

Nicola Prajaska, Anpo Ferdy Siregar, Hamdi Akhsan, Melly Ariska
Physics Education Department, Faculty of Teacher Training and Education, Sriwijaya University. Jalan Raya Palembang-Prabumulih KM 32, Indralaya, Ogan Ilir 30662, Indonesia
E-mail: hamdiakhsan@fkip.unsri.ac.id

Abstract

Modern non-volatile memory technology relies heavily on quantum tunneling, which allows electrons to pass through insulating barriers that are otherwise impenetrable. This paper investigates the use of tunneling methods in flash memory designs to provide high-density and scalable storage. Reviewing tunneling physics in floating-gate and charge-trap memory, assessing barrier-engineering techniques, and investigating scalability options are the goals. Fourteen chosen publications published between 2010 to 2025 were the subject of a systematic literature review (SLR) with an emphasis on high-k dielectric materials, nanoscale device design, and tunneling-barrier engineering. The findings demonstrate that leakage can be decreased while programming efficiency is maintained by improving barrier thickness and dielectric characteristics. Furthermore, the advancement of 3D NAND technology depends heavily on sophisticated materials and device structures. This study indicates that quantum tunneling is still the key mechanism enabling next-generation scalable non-volatile memory, notwithstanding the reliability issues it introduces.

Keywords: *quantum tunneling; flash memory; non-volatile storage; Fowler_Nordheim; Charge-trap memory*

INTRODUCTION

Because it is non-volatile, energy-efficient, and easily integrated into modern electronic devices, flash memory has become a highly important digital storage technology. However, conventional charge-conduction models can no longer adequately explain the physical phenomena that arise when memory cell dimensions are scaled down to the nanometer regime. At this scale, quantum tunneling—where electrons penetrate an energy barrier that cannot be crossed according to classical physics—emerges as the dominant mechanism governing data programming and erasing processes. Therefore, a deeper understanding of the underlying physical mechanisms is essential for improving flash memory performance [1,2].

Recent studies indicate that tunneling mechanisms such as direct tunneling and Fowler–Nordheim (FN) tunneling play a critical role in regulating electron flow through ultrathin insulating layers. These mechanisms determine the charge injection efficiency in both floating-gate and charge-trap flash (CTF) structures. Material properties—including tunneling oxide thickness, dielectric constant, and energy-barrier height—strongly influence device stability and performance. Consequently, material engineering becomes essential for maintaining memory-cell reliability [3,4].

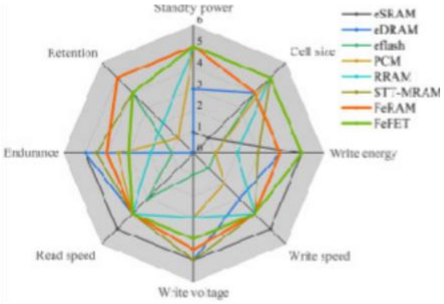

Fig. 1. Comparison of performance indexes of emerging mainstream memory technologies [5].

The use of high-k dielectric materials, such as HfO_2, Al_2O_3, and Ta_2O_5, has enabled improved charge conduction and reduced leakage currents in modern memory technologies. The development of 3D NAND architectures has also increased storage density without compromising energy efficiency. However, continued device miniaturization still encounters several challenges, particularly those related to tunneling-induced degradation and long-term device endurance. These issues highlight the ongoing need for advances in both materials and structural design [6,7].

However, previous studies tend to focus either on material engineering or device performance optimization, particularly in high-k dielectric development and 3D NAND architectures[4,7–9]. While these studies provide significant insights into device fabrication and electrical characteristics, they often lack integration with the underlying quantum physical mechanisms governing tunneling behavior. As a result, the relationship between tunneling physics, material optimization, and device scalability remains insufficiently explored in a unified framework [4,9]. This limitation hinders a comprehensive understanding of how quantum tunneling can be effectively controlled to improve both performance and reliability in advanced flash memory technologies.

This study uses a Systematic Literature Review (SLR) approach to examine research advancements on quantum tunneling in flash memory between 2010 and 2025 in order to close this gap. The specific objectives of this research are to: (1) investigate the basic mechanics of quantum tunneling in flash memory; (2) assess advances in barrier engineering and materials; and (3) pinpoint tactics that facilitate scalable and dependable non-volatile memory design. By using this method, the research aims to offer a comprehensive viewpoint that connects the concepts of quantum physics with technological developments in next-generation memory systems.

RESEARCH METHOD

The PRISMA 2020 standards are followed in this study's Systematic Literature Review (SLR) methodology [8]. Using terms like "quantum tunneling," "flash memory," and "Fowler–Nordheim tunneling" in conjunction with Boolean operators, a literature search was carried out using the Scopus, ScienceDirect, and Google Scholar databases.

Peer-reviewed journal articles, English-language publications, and studies on tunneling mechanisms and flash memory technologies published between 2010 and 2025 met the inclusion requirements. Conference abstracts, non-English publications, and studies that were not directly related to the research issue were among the exclusion criteria.

Fourteen articles were chosen after the identification, screening, eligibility, and inclusion phases of the selection process. Based on the findings' validity, relevance, and methodological clarity, a quality assessment was carried out.

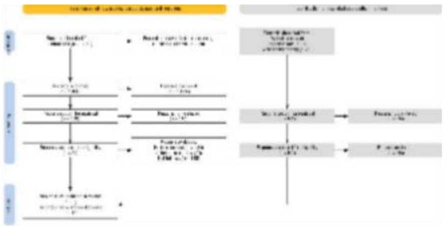

Fig. 2. PRISMA Flowchart [8].

RESULT AND DISCUSSION

Overview of the Selected Studies

The initial search conducted through Google Scholar and ScienceDirect yielded a total of fifty articles. After removing duplicates, screening titles and abstracts, and performing full-text reviews, fourteen articles met the inclusion criteria and were incorporated into this SLR analysis. Alongside the rapid development of 3D NAND and Charge-Trap Flash (CTF) technologies, the number of publications has increased consistently throughout the 2010–2025 period. This trend indicates that quantum tunneling mechanisms are becoming increasingly important for improving the scalability and efficiency of flash memory [3,4].

Figure 3 illustrates the overall publication trend on quantum tunneling in flash memory from 2010 to 2025, based on a broader search across the selected databases. It is important to note that this trend represents general research activity in the field and is not limited to the 14 articles included in the final SLR analysis. A total of fourteen articles were selected for in-depth analysis, while a broader dataset was used to observe general publication trends in the field.

Three main themes emerged from the research analysis. The first theme involves tunneling modeling and mechanisms that describe electron transport through thin oxide layers. The second theme focuses on energy-barrier engineering and material innovations, particularly the use of high-k dielectrics and multilayer structures. The third theme encompasses issues of resolution and scalability in next-generation flash devices. Together, these themes provide a comprehensive overview of advancements in quantum-tunneling-based flash memory technologies [5,6].

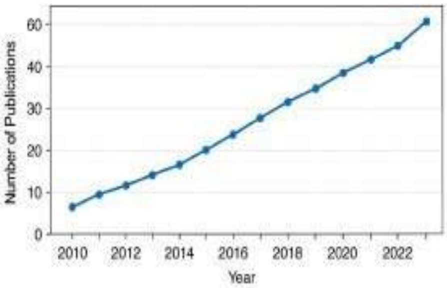

Fig. 3. Annual publication trend on quantum tunneling in flash memory (2010–2025)

Quantum Tunneling Mechanisms in Flash Memory

In flash memory, Fowler–Nordheim (FN) tunneling and direct tunneling are the primary mechanisms governing data programming and erasing [1]. The effectiveness of these mechanisms is strongly influenced by oxide thickness, electric field intensity, and the dielectric constant of the insulating material. Oxide layers thinner than 8 nm often lead to increased charge leakage, as noted by [2]. Conversely, excessively high electric fields can accelerate degradation of the barrier layer [1]. These results indicate that material parameters must be carefully controlled.

Table 1. Summary of quantum tunneling parameters in flash memory devices

Type of Tunneling	Barrier Thickness (nm)	Dominant Materials	Electric Field (MV/cm)	Key Findings
Fowler–Nordheim	7–9	SiO_2 / HfO_2	8–12	High injection efficiency with low leakage
Direct Tunneling	< 5	Al_2O_3 /SiON	5-8	Leakage increases at thicknesses below 6 nm
Trap- Assisted	5-10	Si_3N_4 / ONO stack	6-10	Improved endurance in ChargeTrap Flash
Hybrid (FN + TA)	7-8	HfO_2 + Si_3N_4	9-11	best balance between retention and speed

Table 1 highlights the key characteristics of various tunneling mechanisms found in flash memory devices. Fowler–Nordheim tunneling exhibits high injection efficiency at oxide thicknesses of 7–9 nm. In contrast, when the oxide layer is thinner than 5 nm, direct tunneling becomes dominant. Trap-Assisted Tunneling (TAT) provides better endurance in Charge-Trap Flash structures due to its localized trapping behavior. Meanwhile, hybrid mechanisms that combine FN and TAT offer the best balance between injection speed, charge retention, and long-term reliability. Direct tunneling becomes increasingly significant as the barrier thickness drops below 5 nm.

Conversely, Fowler–Nordheim tunneling tends to dominate in oxide layers thicker than 6 nm under high electric fields. In this condition, the probability of electron transmission increases exponentially, leading to current leakage that may degrade data retention. As decreasing oxide thickness can no longer maintain the balance between programming speed and degradation, the shift from FN-dominated tunneling to direct tunneling marks a physical limit in planar memory scaling. Due to these challenges, the industry has transitioned toward 3D NAND and Charge-Trap Flash architectures. These architectures offer more stable tunneling pathways as a result of more complex cell geometries and advanced material engineering.

Barrier Engineering and Material Innovations

Barrier engineering is a key method for improving the performance of tunneling-based devices [6]. It has been demonstrated that replacing SiO_2 with high-k materials such as

HfO_2, Al_2O_3, and Ta_2O_5 can reduce leakage currents without compromising charge-injection rates. These innovations not only enhance data density but also help maintain oxide integrity. Such materials have become increasingly important in the fabrication of nanometer-scale devices [4,7]. According to findings by [10], the incorporation of Al_2O_3 and TiO_2 interlayers enables the enhancement of negative-capacitance (NC) effects while reducing parasitic charge injection in modern flash memory.

In the development of 3D NAND, material engineering and precise control of the etching process are essential for producing clean vertical channels and reliable source contacts. Based on research [11], Physical modeling based on level-set methods and Monte Carlo ray tracing is used to simulate channel-hole etching in SiO_2/Si_3N_4 stacks, including silicon substrate damage caused by ion-enhanced etching. Such damage can trigger the formation of voids during the subsequent selective epitaxial growth (SEG) of silicon. Through post-etch treatment using loe-energy ions, the damage can be removed, allowing the following SEG process to produce fully formed, defect-free contacts. This model provides essential insights into etching behavior and forms the basis for process optimization in 3D NAND fabrication [12,13].

Layered structures such as nanolaminates and engineered onxynitrides also play a critical role. These structures can suppress oxide degradation while enhancing electric-field endurance. This approach is particulary important for multi-bit and 3D NAND architectures, as the stability of field distribution across each vertical layer directly affects device performance. Multilayer designs and high-quality materials represent a key pathway toward next-generation flash memory technologies [1].

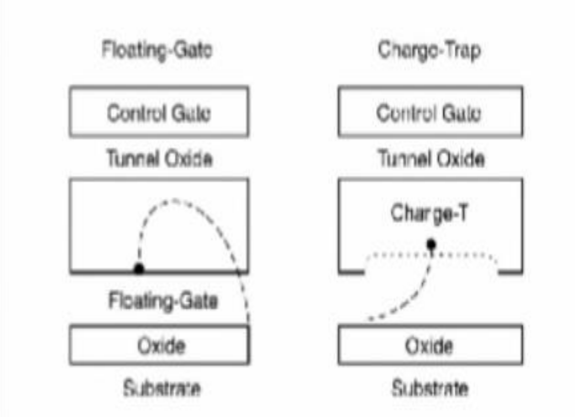

Fig. 4. Schematic representation of electron tunneling pathways in floating-gate and charge-trap flash devices

Reliability and Endurance Analysis

In precision tunneling-based memory technologies, reliability remains one of the primary concerns. Repeated tunneling cycles can lead to oxide degradation, charge trapping, and threshold-voltage shifts [2]. Kim and Hwang identified three major factors contributing to device degradation: charge accumulation in the tunneling oxide, increased Stress-Induced Leakage Current (SILC) due to repeated electrical stress, and a reduction in barrier height caused by prolonged high electric fields. If not properly controlled, these mechanisms can significantly reduce device lifetime [4,7].

Table 2. Summary of research findings on the reliability and endurance of tunneling-based flash memory

Parameter	Floating-Gate Flash	Charge-Trap flash	Observation
Resilience Cycle	$10^4 – 10^5$	$10^5 – 10^6$	CTF exhibits better endurance due to its localized trapping mechanism
Data Retention (10 years)	85 – 95 %	90 – 98 %	Improved with the use of high-k layers
SILC Stream	0.5 – 1.5 μA/cm^2	0.2 –1.0 μA/cm^2	Decreasing with nitrogen doping
Failure Mode	Oxide damage	Leakage due to charge traps	It can be controlled by optimizing the manufacturing process

Because charge traps are localized and do not spread, charge-trap flash exhibits higher cycling endurance than floating-gate flash, as shown in Table 2. The use of high-k layers within the oxide structure also improves data retention. The reduction of SILC through nitrogen doping demonstrates that optimized fabrication processes can extend device lifetime. Overall, material selection and precise layer-structure engineering play a critical role in determining device reliability.

Trends and Thematic Findings

The SLR results identified four main trends in recent research. First and foremost, integrating the tunneling model into 3D NAND simulations allows for more accurate charge transport. On the other hand, hybrid dielectric structures that combine high-k and SiO2 offer a good balance between retention and operating speed. Third, more accurate predictions of oxide degradation patterns and leakage current began with the use of machine learning. Fourth, a low-voltage tunneling mechanism has been developed to improve the energy efficiency of the device [1,2].

To improve the quality of vertical channels in 3D NAND, material engineering and etching process optimization become extremely important. Based on research [14], physical etching modeling on SiO_2/Si_3N_4 stacks can predict ion damage to the silicon substrate, which subsequently leads to void formation during selective epitaxial growth (SEG). Through post-etch treatment using low-energy ions, the damaged layer can be removed, resulting in cleaner and void-free source contacts from SEG. These findings indicate that surface damage control and etching process engineering are key to improving the reliability of 3D NAND cells [15,16].

Demonstrating a significant methodological shift, the research focus has moved from pure experimentation to a data-driven approach. A more comprehensive understanding of tunneling behavior is obtained through a combination of positional physics models, materials engineering, and artificial intelligence. This method, which combines various disciplines, continues to evolve and has a significant impact on the future of storage technology, particularly in terms of scalability and device thickness [17]. One important development is the utilization of hybrid structures that combine magnetic tunneling and electric field-based control to produce low-power non-volatile switching. Findings like these indicate that the phenomenon of quantum tunneling is not only relevant to flash memory, but is also a major trend in the development of next-generation memory.

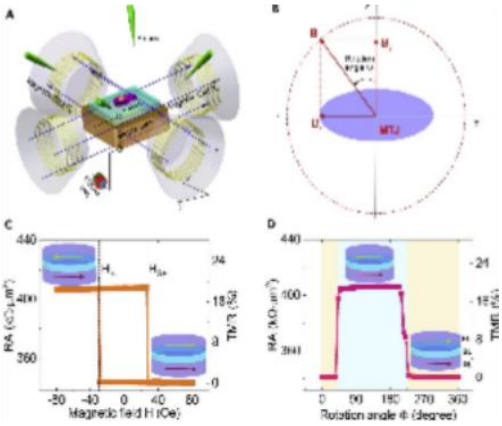

Fig. 5. Hybrid structure of Magnetic Tunnel Junction (MTJ) on ferroelectric PMN-PT substrate [18,19].

Figure 5 shows the hybrid structure of a Magnetic Tunnel Junction (MTJ) paired with a ferroelectric PMN-PT substrate, where an electric field is applied to indirectly control the magnetization state. This mechanism allows for non-volatile changes in resistance state through magneto-elastic effects without requiring large currents, making it more energy-efficient than conventional switching methods. This approach underscores the potential of tunneling-based technology as the foundation for more scalable and efficient non-volatile memory [18,20].

In the development of charge-trap-based non-volatile memory, improving material quality and optimizing the deposition process are crucial factors for achieving high efficiency, stability, and cycle endurance. The latest trends indicate that modifying the Atomic Layer Deposition (ALD) process can significantly impact the performance of charge trapping layers and semiconductor channels, particularly in nanometer-scale memory structures. This finding is highly relevant for the development of new generation storage technologies that utilize tunneling mechanisms and quantum phenomena [9].

REFERENCES

1 Yu J, Wang H, Zhuge F, Chen Z, Hu M, XU X, Zhai T. 2023. Simultaneously ultrafast and robust two-dimensional flash memory devices based on phase-engineered edge contacts. Nat. Commun. **14**(1): 5662.

2 Mehta D, Rahman M, Aono K, Chakrabartty S. 2022. An adaptive synaptic array using Fowler–Nordheim dynamic analog memory. Nat. Commun. **3**(1): 1670.

3 Park JK, Kim SE. 2022. A review of cell operation algorithm for 3D NAND flash memory. Appl. Sci. **12**(21): 10697.

4 Kim SW, Yoo JH, Park WJ, Lee CH, Lee JH, Kim JH, Uhm SH, Lee HC. 2024. Enhancing Charge Trapping Performance Of Hafnia Thin Films Using Sequential Plasma Atomic Layer Deposition. Nanomaterials. **14**(20): 1686.

5 Shao MH, Zhao RT, Liu H, Xu WJ, Guo YD, Huang DP, Ren TL. 2024. Challenges and recent advances in HfO_2-based ferroelectric films for non-volatile memory applications. Chip. **3**(3): 100101.

6 Spassov D, Paskaleva A. 2023. Challenges to optimize charge trapping non-volatile flash memory cells: A case study of HfO_2/Al_2O_3 nanolaminated stacks. Nanomaterials. **13**(17): 2456.

7 Hwang I, Kim J, Lee J, Jung Y, Yoon C. 2025. Memory devices with HfO_2 charge-trapping and TiO_2 channel layers: Fabrication via remote and direct plasma atomic layer deposition and comparative performance evaluation. Materials (Basel). **18**(5): 948.

8 Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. 2022. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams. Campbell Syst. Rev. **18**(2): e1230.

9 Gupta D, Upadhyay AK, Beohar A, Vishvakarma SK. 2023. Improvement of memory performance of 3D NAND flash memory with retrograde channel doping. Memories – Mater. Devices, Circuits Syst. **4**: 100031.

10 Lee S, Kim G, Nam Y, Jeong Y, Kang H, Kim W, Jeon S. 2025. Stabilized negative capacitance for near-theoretical efficiency and high reliability in charge trap flash memory. Mater. Today Phys. **58**: 101865.

11 Reiter T, Klemenschits X, Filipovic L. 2022. Impact of plasma induced damage on the fabrication of 3D NAND flash memory. Solid. State. Electron. **192**: 108261.

12 Li Y. 2020. 3D NAND memory and its application in solid-state drives: Architecture, reliability, flash management techniques, and current trends. IEEE Solid-State Circuits Mag. **12**(4): 56.

13 Toifl A, Quell M, Klemenschits X, Manstetten P, Hössinger A, Selberherr S, Weinbub J. 2020. The level-set method for multi-material wet etching and non-planar selective epitaxy. IEEE Access. **8**: 115406.

14 Gopal G, Varma T. 2022. Simulation-based analysis of ultrathin-body double gate ferroelectric TFET for enhanced electrical performance. Silicon. **14**(2): 6553.

15 Dutta R, Subash TD, Paitya N. 2022. Improved DC performance analysis of a novel asymmetric extended source tunnel FET (AES-TFET) for fast switching application. Silicon. **14**(8): 3835.

16 Yang Y, Luo Z, Wang S, Huang W, Wang G, Wang C, Xiao G. 2021. Electric-field-assisted non-volatile magnetic switching in a magnetoelectronic hybrid structure. Science (80-.). **24**(7): 102734.

17 Liao J, Dai S, Peng RC, Yang J, Zeng B, Liao M, Zhou Y. 2023. HfO_2-based ferroelectric thin film and memory device applications in the post-Moore era: A review. Fundam. Res. **3**(3): 332.

18 Zhang Y, Wang C, Huang H, Lu J, Liang R, Liu J, Zhang J. 2020. Deterministic reversal of single magnetic vortex circulation by an electric field. Sci. Bull. **65**(15): 1260.

19 Kum HS, Lee H, Kim S, Lindemann S, Kong W, Qiao K, Kim J. 2020. Heterogeneous integration of single-

crystalline complex-oxide membranes. Nature. **58**(7793): 75.

20 Swathi SP, Angappane S. 2022. Enhanced resistive switching performance of hafnium oxide-based devices: Effects of growth and annealing temperatures. J. Alloys Compd. **913**: 165251.