Analysis on Channel Tapering in 3D NAND Flash Memory

DHRUV GUPTA, SUNNY KUMAR
Birla Institute of Technology and Science, Pilani

Abstract— **This report provides an in-depth analysis of 3D NAND Flash memory behavior, emphasizing the impacts of channel tapering, wordline position, and operating voltage. It explores how both geometric and electrical factors affect the drain current (Id), delivering practical insights for device optimization. The conclusions are backed by experimental data, interpolated trends, and graphical representations.**

Keywords— **3-D NAND, channel doping, tapered channel, threshold voltage (VT)**

I. INTRODUCTION

3D NAND flash memory achieves high-density storage through vertical stacking of cells, but suffers from performance degradation due to channel tapering and wordline depth variation. These non-uniformities affect threshold voltage (VT) distribution and cell current (Icell), especially in multilevel cells. This study investigates the impact of taper angle and proposes techniques such as optimized channel doping and tailored program/erase voltages using TCAD simulations. Results demonstrate up to 90 \% improvement in VT uniformity across the vertical string.

II. BACKGROUND

To accurately investigate the threshold voltage (VT) and cell current (Icell) in advanced 3D NAND flash memory devices, a comprehensive TCAD simulation framework was developed and calibrated. The calibration involved tuning several physical parameters, including the channel doping profile, polysilicon interface trap characteristics, and device models, to closely match the experimentally observed Id–Vg characteristics. Furthermore, the program/erase characteristics were refined by adjusting the properties of the nitride-trap layer, such as trap density and distribution, as well as the grain size and trap density within the polysilicon channel. A threshold current of 100 nA was used as a benchmark for extracting VT. Once the simulation setup was validated against published data, a detailed investigation was carried out to understand the impact of channel tapering on device performance. Key parameters varied in the simulations included the taper angle of the vertical channel, channel doping concentration, block oxide thickness, and both the amplitude and duration of the program voltage. The VT distribution was analyzed using the Id–Vg characteristics of wordline (WL) transistors in both initial and programmed states. Based on these results, novel techniques were proposed to minimize the VT spread along the NAND string, thereby improving performance uniformity and reliability in 3D NAND memory.

III. SIMULATION AND APPROACH

To accurately capture the effect of tapering, the simulations use a crystalline channel without traps and a short string of 12 WLs, reflecting the variations in 3-D NAND flash memory. The threshold voltage (VT) and cell current (Id) were investigated using TCAD by calibrating the Id–Vg characteristics of the 3-D NAND cell, adjusting doping, polysilicon interface traps, and device models. The simulations incorporated advanced models for generation/recombination, including Auger, Shockley Read Hall (SRH), and band-to-band tunneling, along with mobility models to account for doping, high field effects, and interface dependence.

IV. RESULTS AND DISCUSSION

1.Programmed / Initial Id vs Vg Plot for WL6

The difference in the curves arises primarily due to **threshold voltage (Vt) shift** and **gate control degradation** influenced by **taper angle geometry**.

🔹 **1. Threshold Voltage Shift**

- In the **programmed state**, electrons are injected into the charge trap or floating gate.
- This leads to a **positive shift in threshold voltage** (Vt ↑), meaning a higher gate voltage (Vg) is required to turn the transistor **on**.
- That's why the programmed curves appear **right-shifted** — they turn on later.

In contrast, in the **erased (initial) state**, the trapped charge is minimal or removed, so Vt is **lower**, and the device turns on **earlier** — resulting in **left-shifted** curves.

2. Drain Current (Id) Suppression

- Even at the same Vg, the **programmed devices show lower Id** than erased ones.
- This is due to the **reduced (Vg − Vt)** in programmed cells — smaller gate overdrive means less current.
- Additionally, as taper angle increases, the gate loses electrostatic control over the channel, especially in the programmed state, reducing Id even more.

🔹 **3. Impact of Taper Angle**

- Larger taper angles (like 21.8°) show:
    - Greater Vt shifts (stronger right shift)
    - Lower Id (more drain current suppression)
- This shows taper angle exacerbates **charge-induced Vt shifts** and weakens channel control in the programmed state more than the erased state.This behavior suggests that lower taper angles (13.8°–15.4°) provide superior memory window and current drive, making them more favorable for robust non-volatile memory operation.

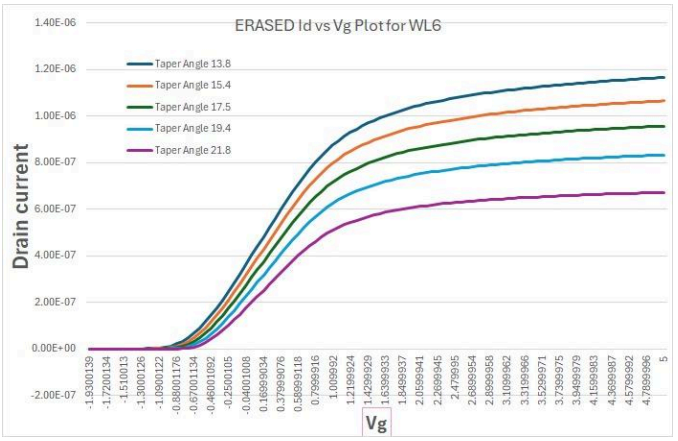


Fig. 1 Erased State of WL6

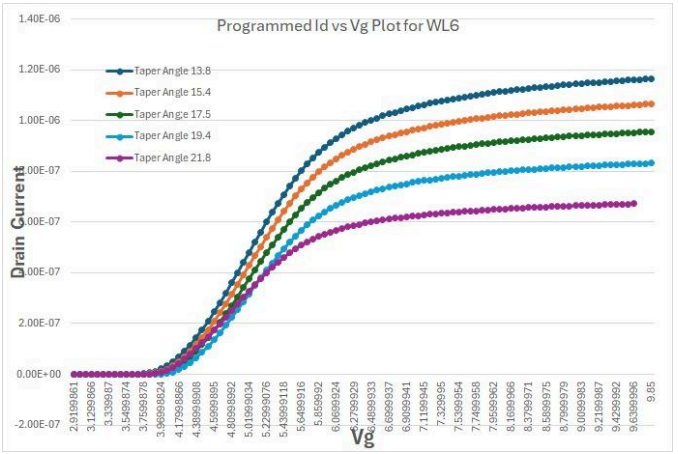


Fig. 2 Programmed State of WL6

2.Initial /Programmed Vt Values vs. Word Line Number

Initial Vt (Erased State):

As we move from WL1 to WL12, the initial (erased) threshold voltage becomes more negative across all taper angles.
The variation is much larger at higher taper angles.
For example:
At 21.8°, Vt drops from ~0 V at WL1 to ~–0.55 V at WL12.At 13.8°, Vt remains almost constant (~–0.7 V).

Programmed Vt: Programmed threshold voltage also decreases from WL1 to WL12.
At 21.8°, Vt drops from ~7.5 V (WL1) to ~4.5 V (WL12), showing a ~3 V variation.
At 13.8°, the variation is minimal (~4.5 V to ~4.0 V).

Reason Behind Initial Vt Variation

The threshold voltage in the erased state is affected by the physical radius of the channel (R) and doping concentration (Nd). The governing expression (rearranged for clarity in text format) is:

$Vt\_initial = Vfb - (q * Nd * R^2 / (2 * \varepsilon\_si)) * \ln(1 + teff / R) + correction\_term$

Where:Vfb is the flat-band voltage,q is the charge of an electron,Nd is the doping concentration,R is the silicon channel radius,teff is the effective oxide thickness,ε_si is the permittivity of silicon,correction_term accounts for oxide coupling
Key Point:
As the taper angle increases, R varies more from top (WL1) to bottom (WL12).At WL1, R is small → Vt is high.At WL12, R is large → Vt is lower
Hence, higher taper angles show more Vt variation across wordlines.

Reason Behind Programmed Vt Variation

Programmed Vt depends on the amount of charge (Q_trap) injected during programming and the gate capacitance (C_ox). The relationship is:

$Vt\_programmed = Vt\_initial + (Q\_trap / C\_ox)$

However, the charge injected into each wordline is not uniform due to:Vertical electric field variation Gate control degradation at lower wordlines (WL12)Geometrical taper reducing coupling at the bottom**,**WL1: Strong electric field → More charge → Higher Vt.WL12: Weaker field → Less charge → Lower Vt
Again, the higher the taper angle, the more severe the variation.

Impacts and Design Implications

Reduced Read Margin:
Wide Vt variation makes it harder to distinguish between programmed and erased states, especially in multi-level cells.
Cell-to-Cell Interference:
Vt mismatches across wordlines increase risk of program disturb and read disturb effects.
Endurance and Retention Degradation:
Cells at WL12 with lower programmed Vt may suffer from weaker retention and less consistent behavior over program/erase cycles.

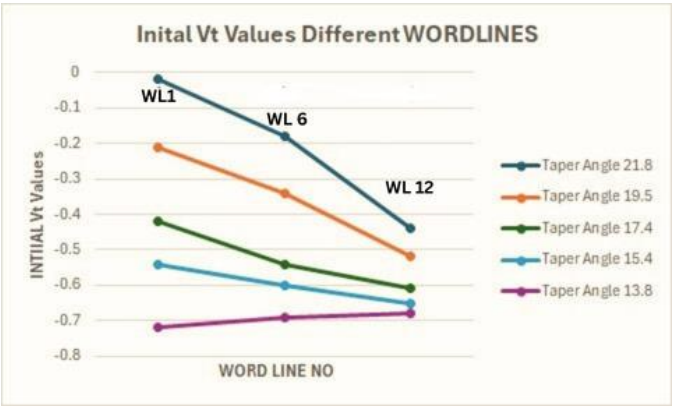
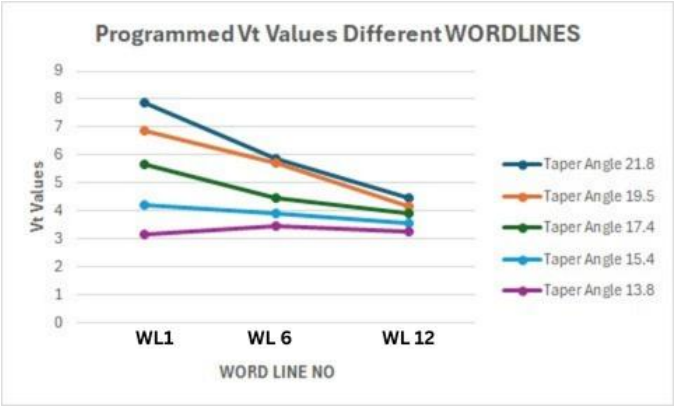


Fig. 3 .Initial and Programed state VT versus WL number with a taper angle as a parameter.



3.Initial/programmed Vt Values

Initial Vt vs Taper Angle (for WL1, WL6, WL12).For all wordlines, **initial Vt increases (becomes less negative)** as **taper angle increases** from 13.8° to 21.8°.The increase is **steepest for WL1**, followed by WL6, and least for WL12.

In 3D NAND, increasing the taper angle means the **channel narrows more sharply** near the top of the string (WL1).The **effective channel radius (R)** becomes smaller at the top for larger taper angles, which increases the threshold voltage.
From the theoretical expression:

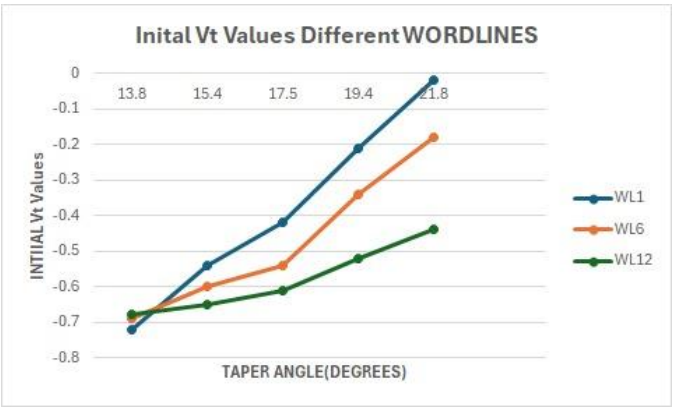
$$V_{t\_initial} \propto -N_d * R^2 * \ln(1 + t_{eff}/R)$$

So as R decreases, Vt increases.
At WL12 (bottom), R changes less with taper angle, so Vt remains more stable.
The **vertical variation in geometry** affects the top wordlines the most. That's why **WL1 sees the highest Vt shift**, while **WL12 is relatively stable**.

Programmed Vt vs Taper Angle (for WL1, WL6, WL12)

For all wordlines, programmed Vt increases with taper angle.Again, the rise is steepest for WL1, moderate for WL6, and lowest for WL12.
Programming involves injecting charge into the trap layer, which depends on the vertical electric field strength.
At higher taper angles:
WL1 has stronger gate-channel coupling due to narrower geometry ⇒ more charge injection ⇒ higher Vt.
WL12, being wider and farther from the gate control center, experiences weaker field ⇒ less charge trapped ⇒ lower Vt.

The result is a **greater separation** in Vt values among wordlines as taper angle increases.3.



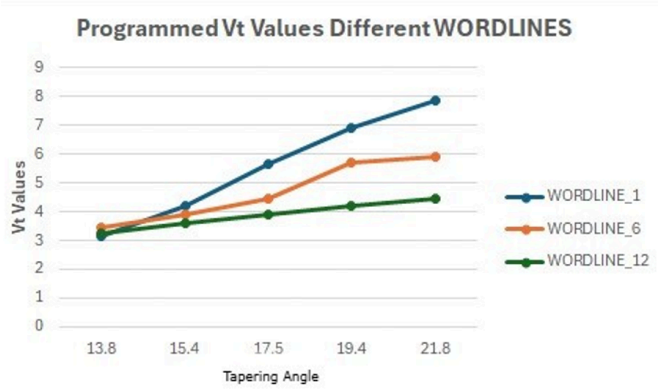


Fig 4.Initial and Programmed v(t) with different Tapering Angle

4. Suggestions to mitigate the Errors encountered

Wordline-Specific Biasing
Apply adjusted bias voltages during program and read phases, where lower wordlines (e.g., WL12) receive slightly higher assist bias. This compensates for weaker gate control due to wider channel radius at the bottom, helping to align threshold voltages across all wordlines without structural changes.

Graded Gate Stack Engineering
Design the gate dielectric (e.g., ONO or high-k material) with a vertical gradient in thickness or permittivity. This improves electrostatic coupling uniformly across the stack, minimizing charge injection disparities and ensuring more consistent Vt behavior from top to bottom.

AI-Based Adaptive Programming Control
Deploy a lightweight machine learning model to

dynamically predict optimal programming pulse widths and voltages for each wordline based on taper angle, position, and past behavior. This enables fine-grained control during programming, reducing vertical Vt spread and improving reliability.

5.Recommended Design Parameters for Minimizing Vt Variation in 3D NAND

Parameter	Recommended Range	Rationale
Taper Angle (theta)	13.5° to 17.5°	Keeps channel radius variation minimal, reducing Vt spread across WLs.
Gate Dielectric Thickness (teff)	6 to 9 nm (graded vertically)	Graded teff improves gate-to-channel coupling uniformly along the stack.
Wordline Bias Offset	+0.2 V to +0.5 V for lower WLs	Boosts gate control for bottom WLs (e.g., WL10–WL12), equalizing Vt.
Channel Radius (R)	10 to 12 nm (uniform or core-stabilized)	A consistent radius ensures balanced electrostatics across wordlines.
Program Pulse Width	5 to 20 microseconds (AI-adjusted per WL)	Dynamic adjustment improves program accuracy and minimizes variation.
Channel Doping (Nd)	1e17 to 3e17 per cm^3	Balanced doping supports Vt control while avoiding excessive sensitivity.

5.Conclusions

The analysis demonstrates that voltage, channel tapering, and wordline depth play key roles in determining the performance of 3D NAND devices. Implementing tuning strategies that account for taper and state characteristics can significantly enhance uniformity and endurance. Moving forward, the focus will shift toward developing predictive models and optimizing device performance over the long term.

ReferencesIn

[1] Upendra Mohan Bhatt, Arvind Kumar, Sanjeev Kumar Manhas
Performance Enhancement by Optimization of Poly Grain Size and Channel Thickness in a Vertical Channel 3-D NAND Flash Memory
IEEE Transactions on Electron Devices, Vol. 65, No. 5, pp. 1781–1786, 2018.

[2] Upendra Mohan Bhatt, Sanjeev Kumar Manhas, Arvind Kumar, M. Pakala, E. Yieh
Mitigating the Impact of Channel Tapering in Vertical Channel 3-D NAND
IEEE Transactions on Electron Devices, Vol. 67, No. 3, pp. 929–936, 2020.

[3] Yixin Luo, Saugata Ghose, Yu Cai, Erich F. Haratsch, Onur Mutlu
Improving 3D NAND Flash Memory Lifetime by Tolerating Early Retention Loss and Process Variation
Proceedings of the 2018 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, pp. 123–135, 2018.

[4]Tong Zhang, Shu Li
Improving Multi-Level NAND Flash Memory Storage Reliability Using Concatenated TCM-BCH Coding
IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 17, No. 4, pp. 473–486, 2009.

[5]Haozhang Yang, Peng Huang, Runze Han, Xiaoyan Liu, Jinfeng Kang
An Ultra-High-Density and Energy-Efficient Content Addressable Memory Design Based on 3D-NAND Flash
Science China Information Sciences, Vol. 66, Article No. 142402, 2023.

[6]Mohammad Bavandpour, Shubham Sahay, Mohammad Reza Mahmoodi, Dmitri B. Strukov
3D-aCortex: An Ultra-Compact Energy-Efficient Neurocomputing Platform Based on Commercial 3D-NAND Flash Memories

[7]Jisung Park, Roknoddin Azizi, Geraldo F. Oliveira, Mohammad Sadrosadati, Rakesh
Nadig, David Novo, Juan Gómez-Luna, Myungsuk Kim, Onur Mutlu
Flash-Cosmos: In-Flash Bulk Bitwise Operations Using Inherent Computation Capability of
NAND Flash Memory
arXiv preprint, arXiv:2209.05566, 2022.