DRAM and NAND Reliability in AI era:

Understanding DRAM and NAND Data Retention Failure Mechanisms to Prevent System Downtime

Ju Jin An

Supply Chain Engineering, Infrastructure, IBM

Agentic AI

- Generative AI focuses on creating new content (like text, images, or code) based on prompts
  - Like a calculator - takes an input, gives an output, forgets
- Agentic AI is focused on decisions as opposed to creating the actual new content, and doesn't solely rely on human prompts nor require human oversight
  - Like a project manager - takes a goal, remembers status, revises plans, logs what happened
  - Extra “working memory” is needed to coordinate its actions over time
  - More reliable memory and storage in AI computation is required

3D NAND Scaling

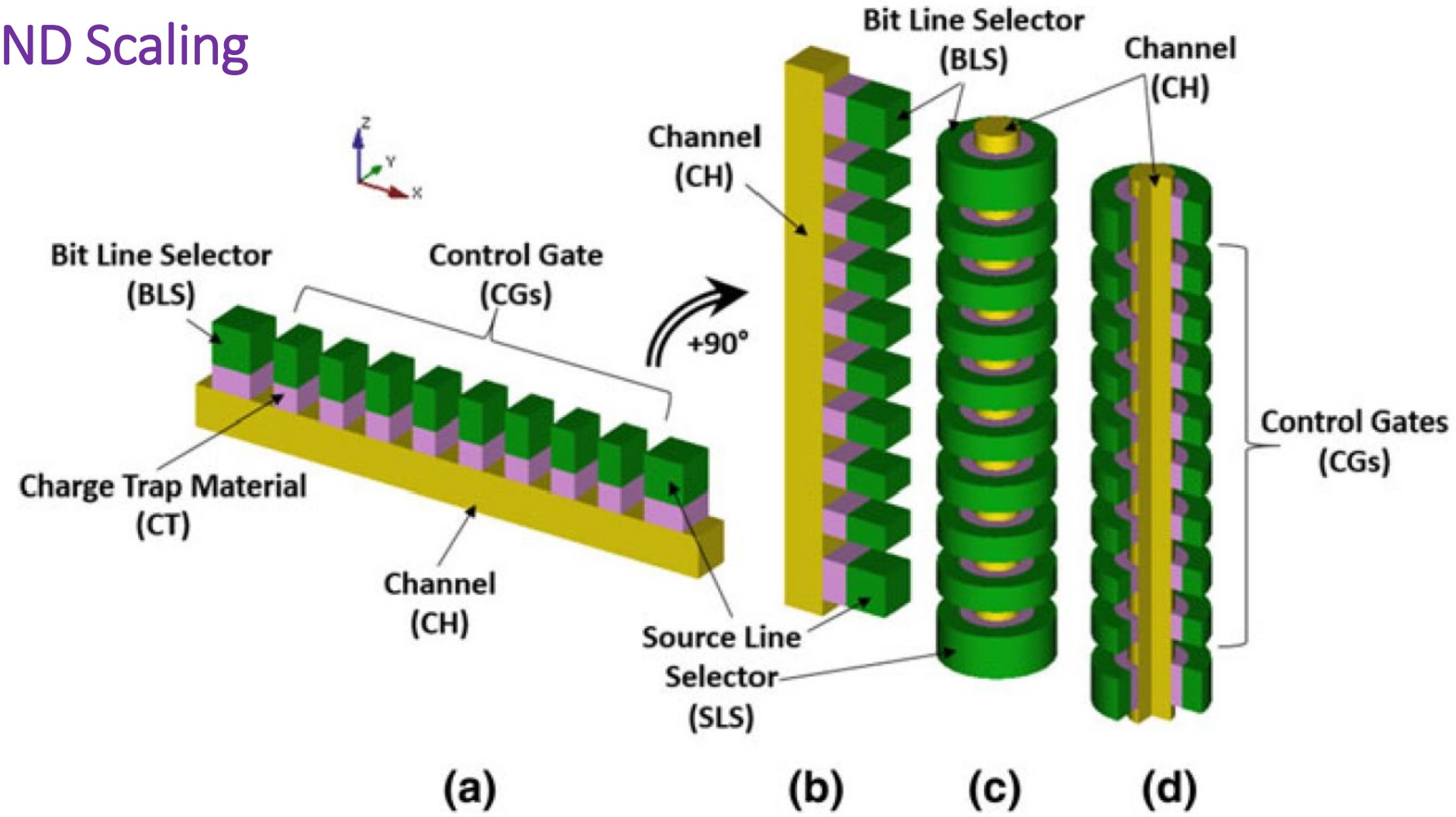
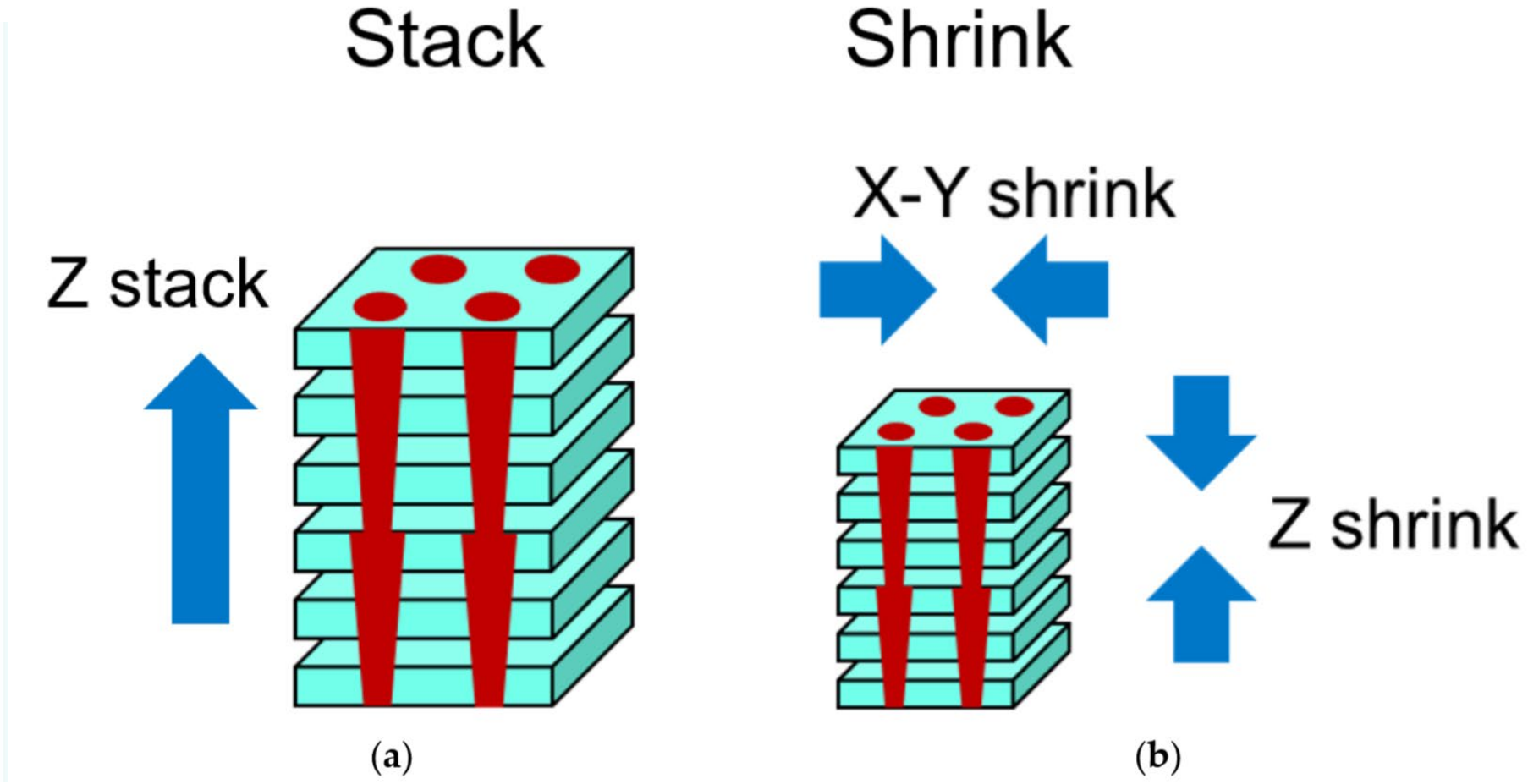


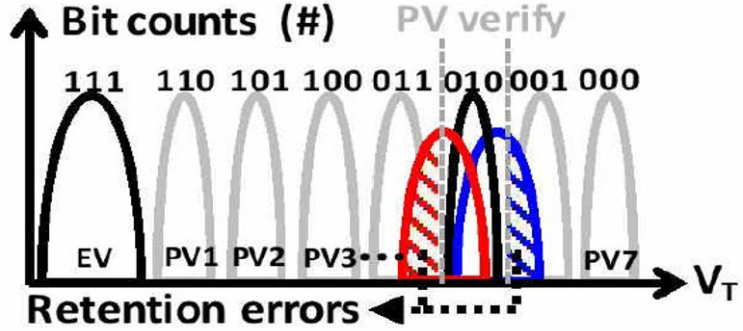
Fig. 4.1 NAND Flash string with horizontal gate and vertical channel: **a** planar, **b** planar rotated by 90°, **c** vertical channel with cylindrical shape and **d** its cross section

Source: [1]

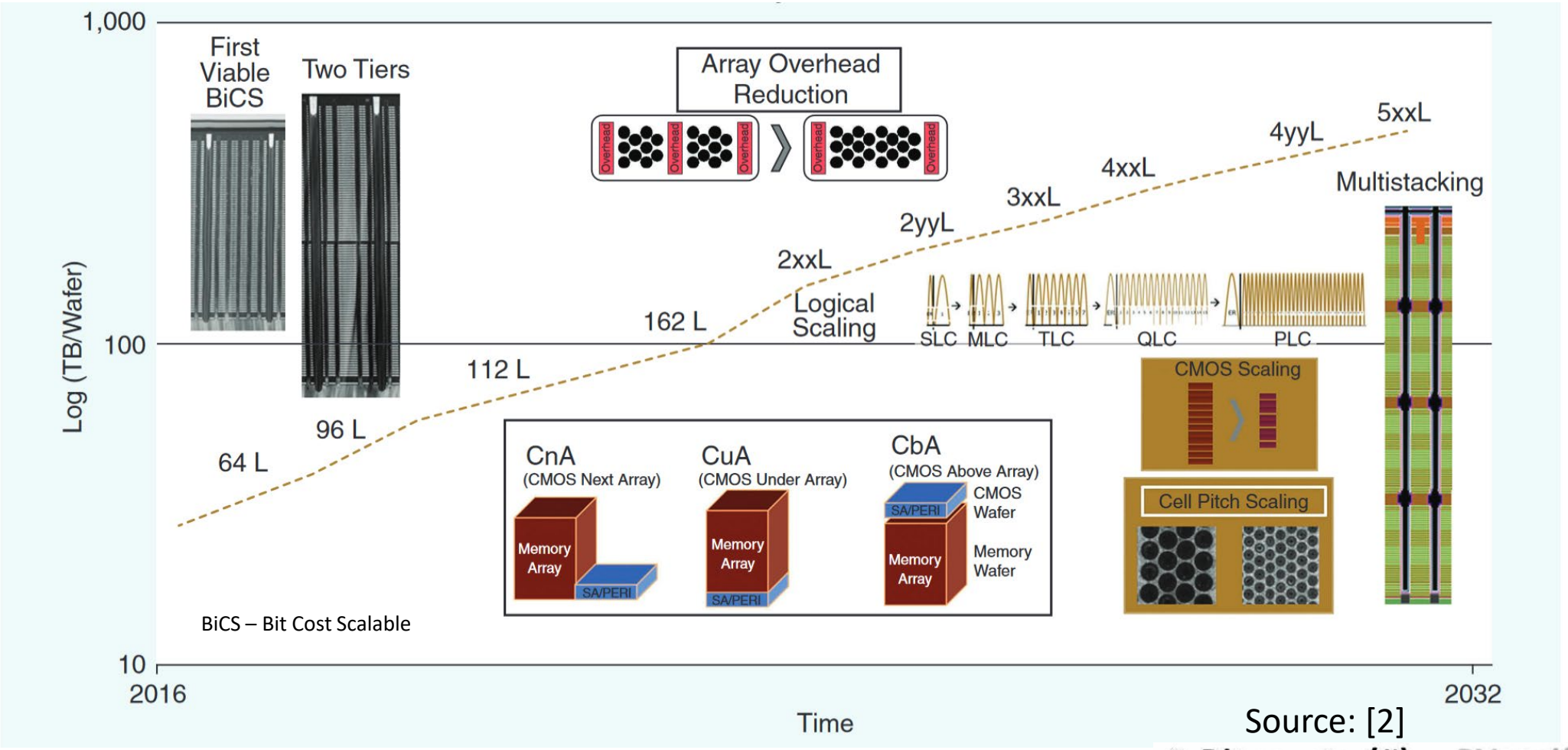
3D NAND Scaling



- To increase bit density, scaling in x-y direction, increasing the number of stacks in z direction and logical scaling have been developed
- Bit error rate margin is decreased with scaling

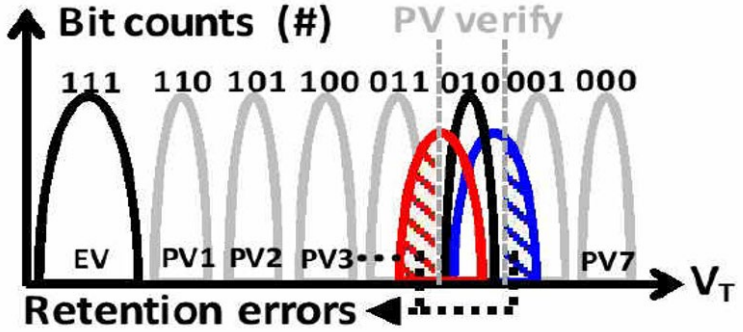


3D NAND Scaling Trend

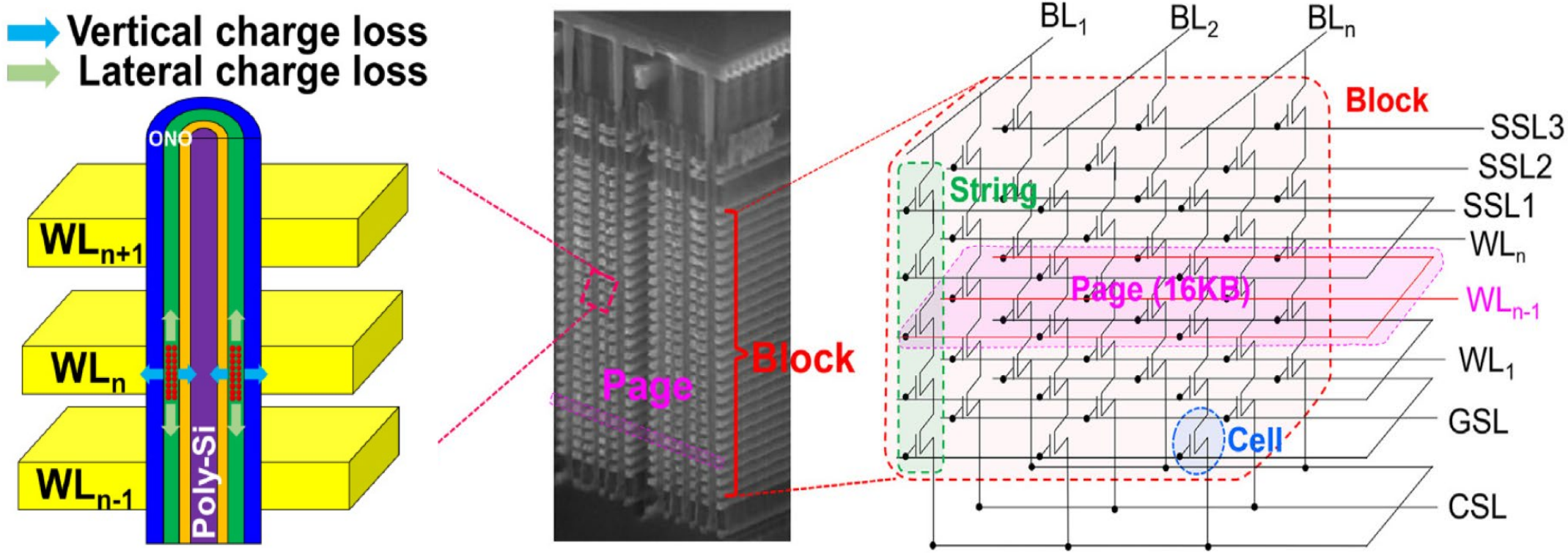


Source: [2]

- To increase bit density, scaling in x-y direction, increasing the number of stacks in z direction and logical scaling have been developed
- Bit error rate margin is decreased with scaling



Charge loss mechanisms in 3D NAND



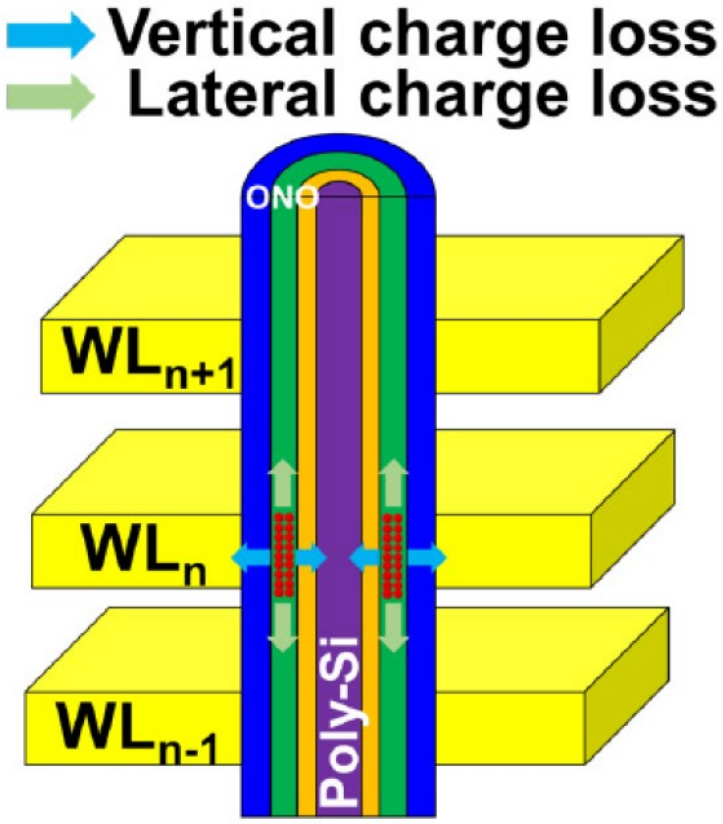
Source: [3]

Source: [4]

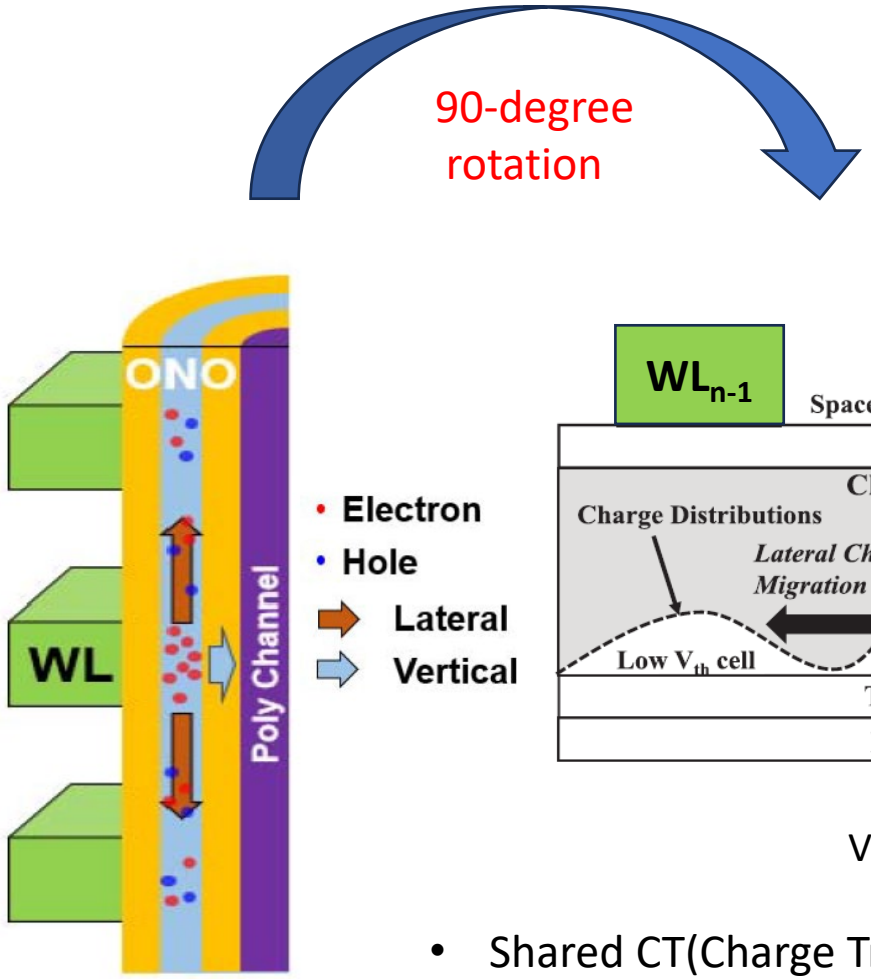
- Lateral charge loss - charge trap layer shared with adjacent WLs, inducing lateral charge loss (unique mechanism of charge trap NAND)
- Vertical charge loss – trapped electrons in charge trap and tunnel oxide migrate to poly-Si

WL – Word Line
BL – Bit Line
SSL – String Select Line
GSL – Ground Select Line
CSL – Common Source Lne

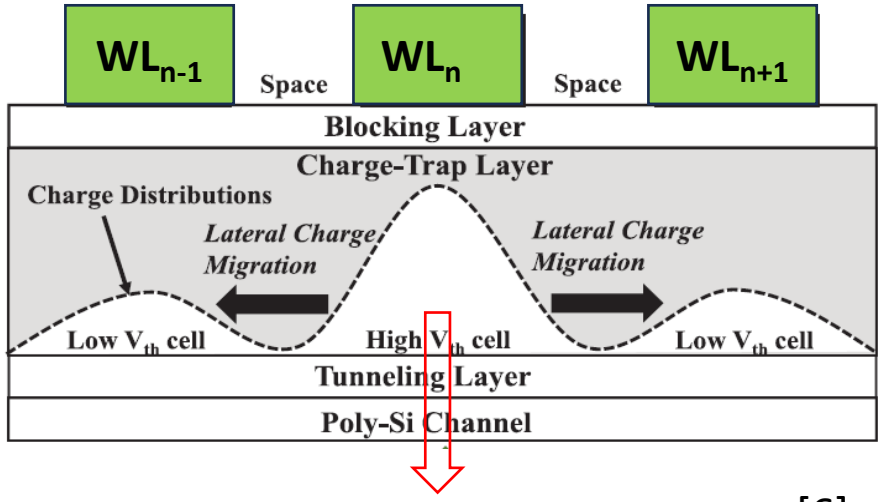
Charge loss mechanisms in 3D NAND



Source: [3]



[12]

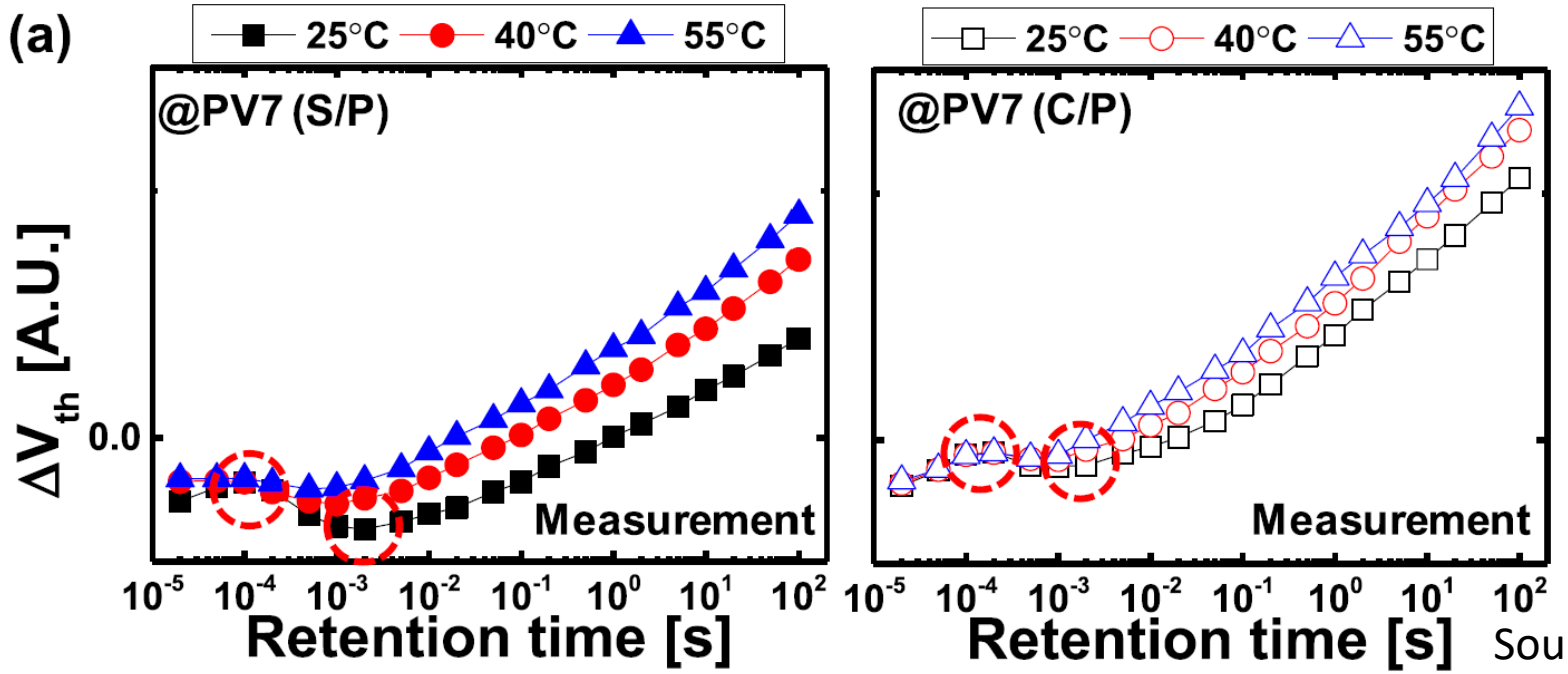


[6]

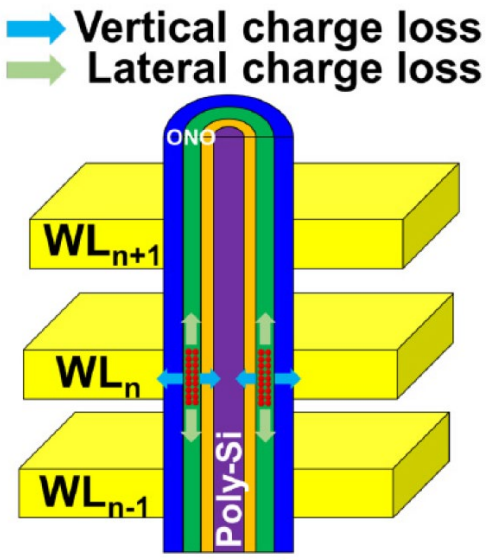
- Shared CT(Charge Trap) layer with the adjacent word-line

WL – Word Line

Short term (fast) charge loss mechanisms in 3D NAND



Source: [7]

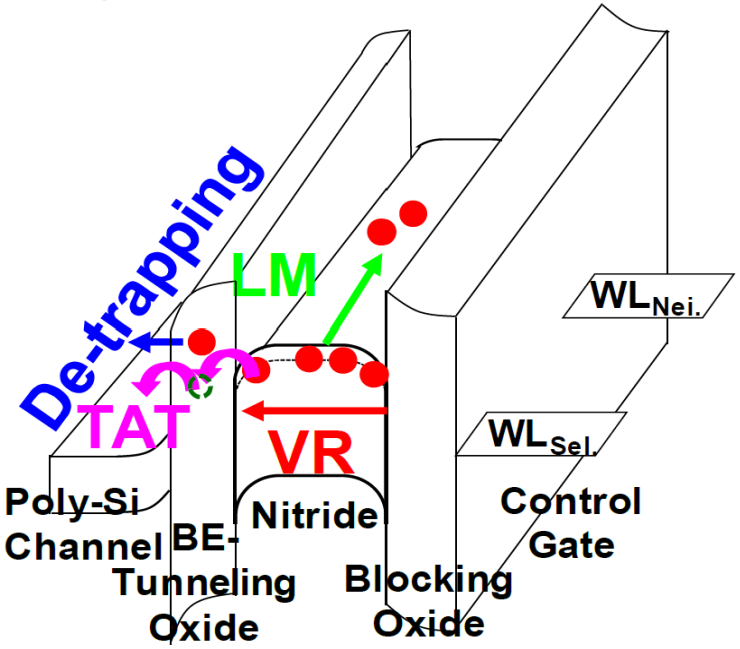


[3]

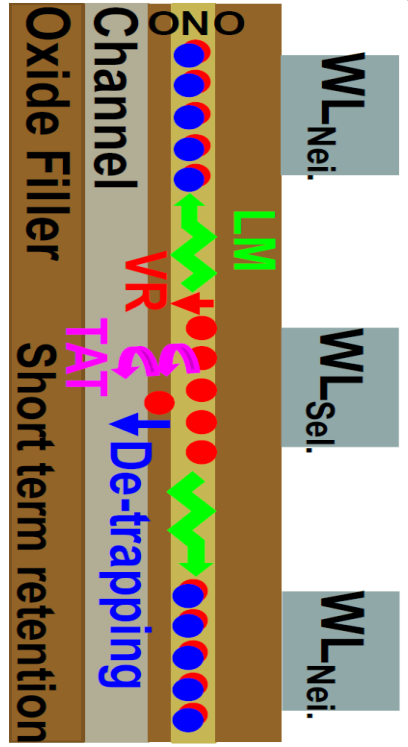
- Early retention or initial threshold voltage shift (IVS) is one of the key reliability challenges in charge trapping memory (CTM) based 3D NAND flash
- Multiple inflection points including gain in threshold voltage are observed in threshold voltage loss
  - Inflection points suggest superposition of several mechanisms when stored electrons are emitted during short term retention of 3-D NAND Flash
  - A distribution of trap depths exist in CTL (i.e., both deep and shallow trap depths exist within the same charge loss mechanism)
    - In short-term retention, shallower traps within the same charge loss mechanism can dominate early charge loss

Short term (fast) charge loss mechanisms in 3D NAND

- **Vertical charge loss**
  - **De-trapping** of trapped electrons in BE-tunneling oxide into the channel by tunneling mechanism
  - **TAT (Trap assisted tunneling)** - electrons stored in the CTN are emitted into the channel through trap sites in the BE-tunneling oxide
  - **VR (Vertical redistribution)** of electrons stored in the charge trap layer
    - Immediately after the program operation, electrons stored near the blocking oxide interface move toward the BE-tunneling oxide
    - Charge centroid of the trapped electrons in the CTN is located near the blocking oxide interface and moves toward the BE-tunneling oxide interface
    - Electrons are still in CTN, however centroid of electron moves closer to Si channel, creating positive shift in Vth
- **LM (Lateral Migration) of electrons stored in the charge traps**
  - The movement of trapped electrons along the wordline direction
  - Significantly affected by E-field (i.e., cell program pattern)
  - LM increases with temperature increase



Source: [8]

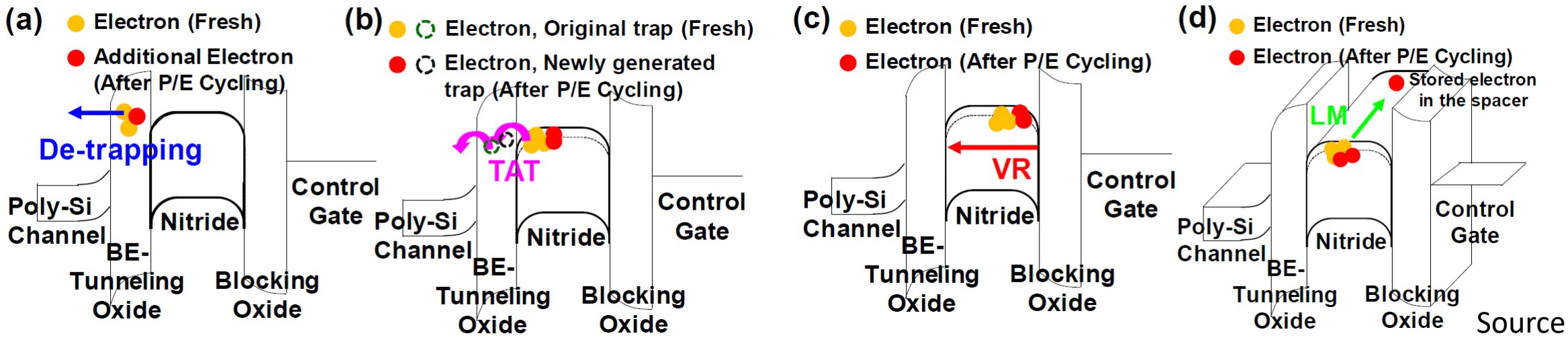


Electron

Hole

BE – Band Gap Engineered
CTN – Charge Trap Nitride

Impact of cycling stress and temperature on short term retention

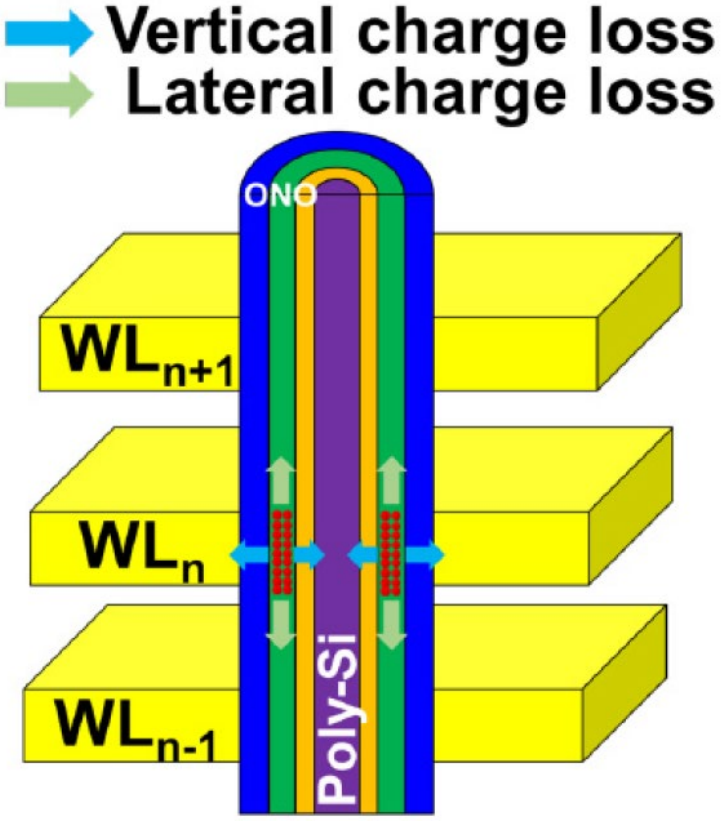


Source: [8]

Short term Charge Loss Mechanism	With increasing temperature	After cycling stress
Vertical Charge Loss including De-trapping, TAT (Trap Assisted Tunneling), VR (Vertical Redistribution)	✓ As the temperature increases, the number of electrons emitted into the channel increases → trapped electrons are emitted via tunneling mechanism and thermal emission process	✓ Cycling induces tunnel oxide degradation (introducing interface/oxide traps), enhancing vertical charge loss
LM (Lateral Migration)	✓ As the temperature increases, more electrons migrates laterally	✓ Cycling induces accumulated charges in the intercell regions (between WL-WL) within charge trapping layer, suppressing lateral migration ✓ More charge traps in BE-tunneling oxide during program, meaning less electrons in CTL, lateral E-field continues to decrease

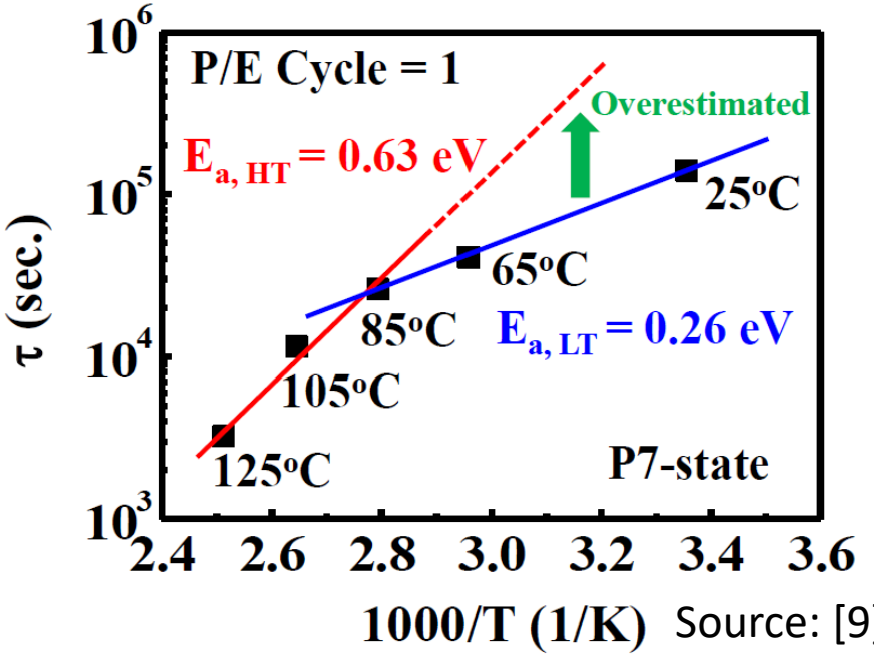
CTL – Charge Trap Layer
BE Tox - Band-gap Engineered Tunneling Oxide

Charge Loss mechanisms during long-term retention operation



Source: [3]

- **Higher activation energy (Ea) at higher temperature**
  - Ea has a temperature dependance
  - It suggests multiple charge loss mechanisms exist at different operating conditions (temperature, cycling, program voltage level of a target cell and neighboring cells)



Source: [9]

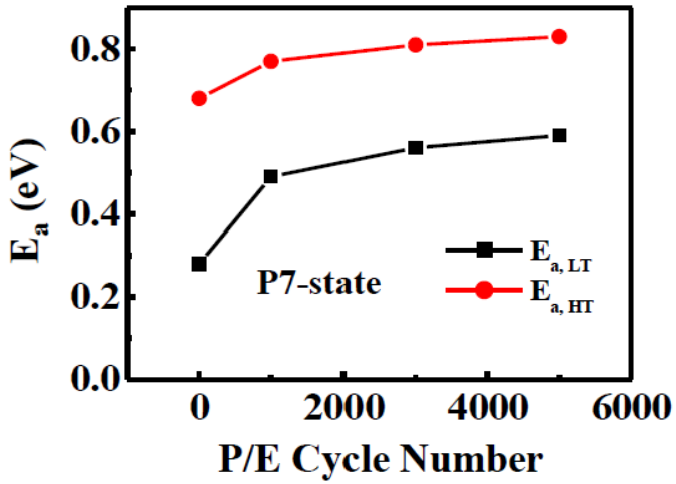
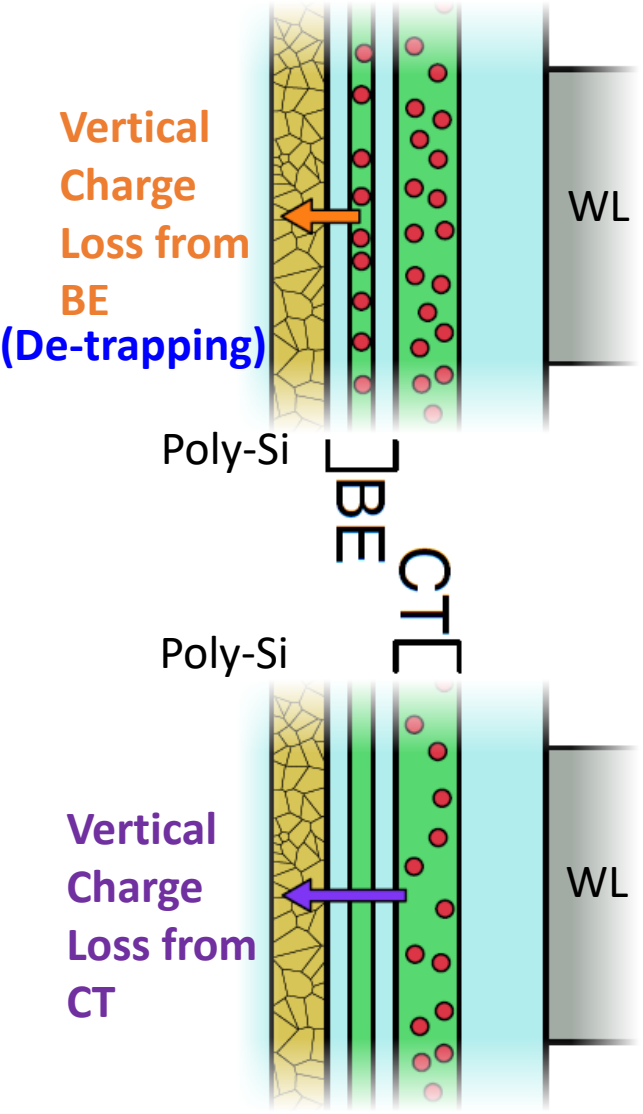
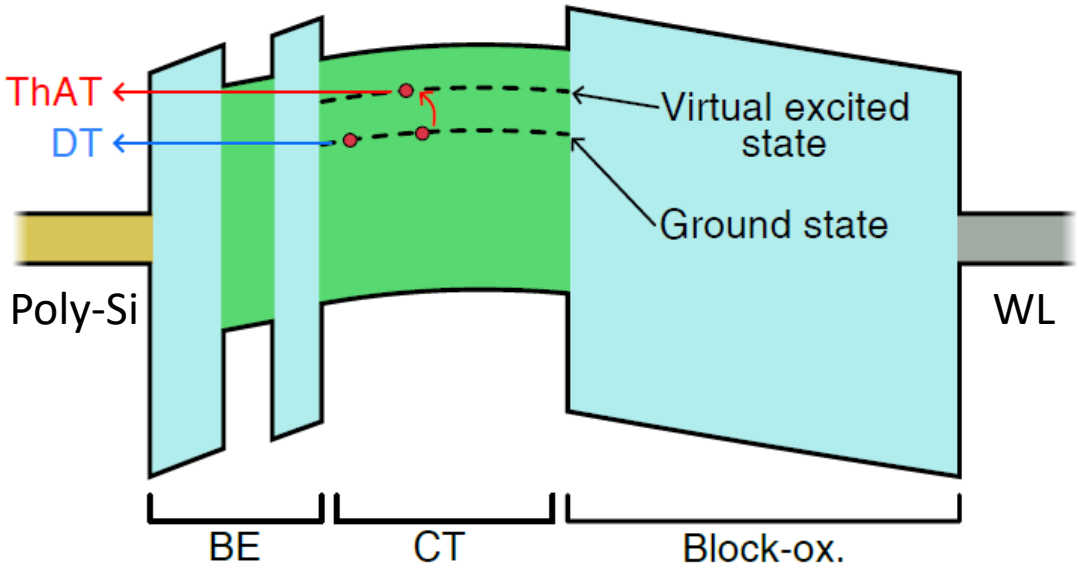


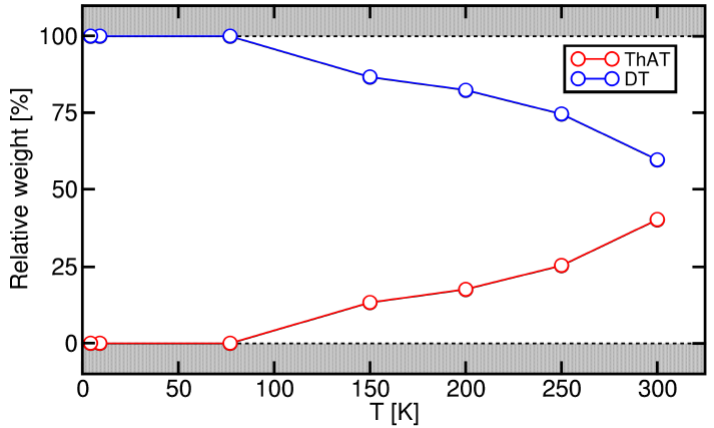
Fig. 3. A certain retention time (τ) to a criteria of V_t shift at various bake T is plotted against the inverse of temperature. P/E cycle number is 1.

CT – Charge Trap Nitride
BE – Band-gap Engineered Tunneling Oxide
VCL – Vertical Charge Loss
WL – Word Line

Vertical Charge Loss mechanisms during long-term retention operation





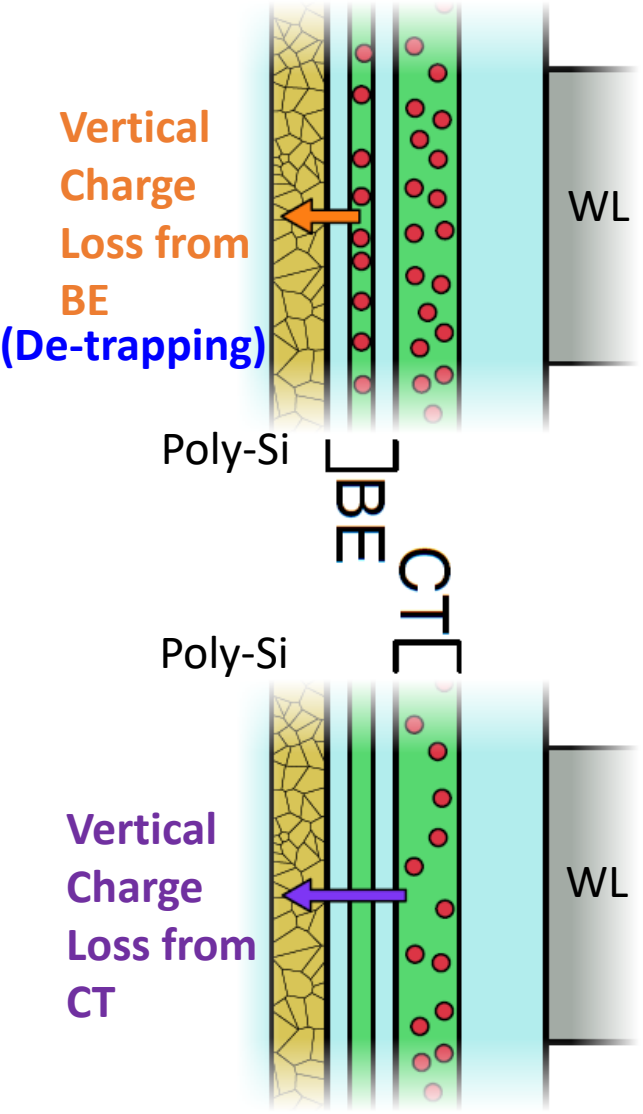


Source: [10]

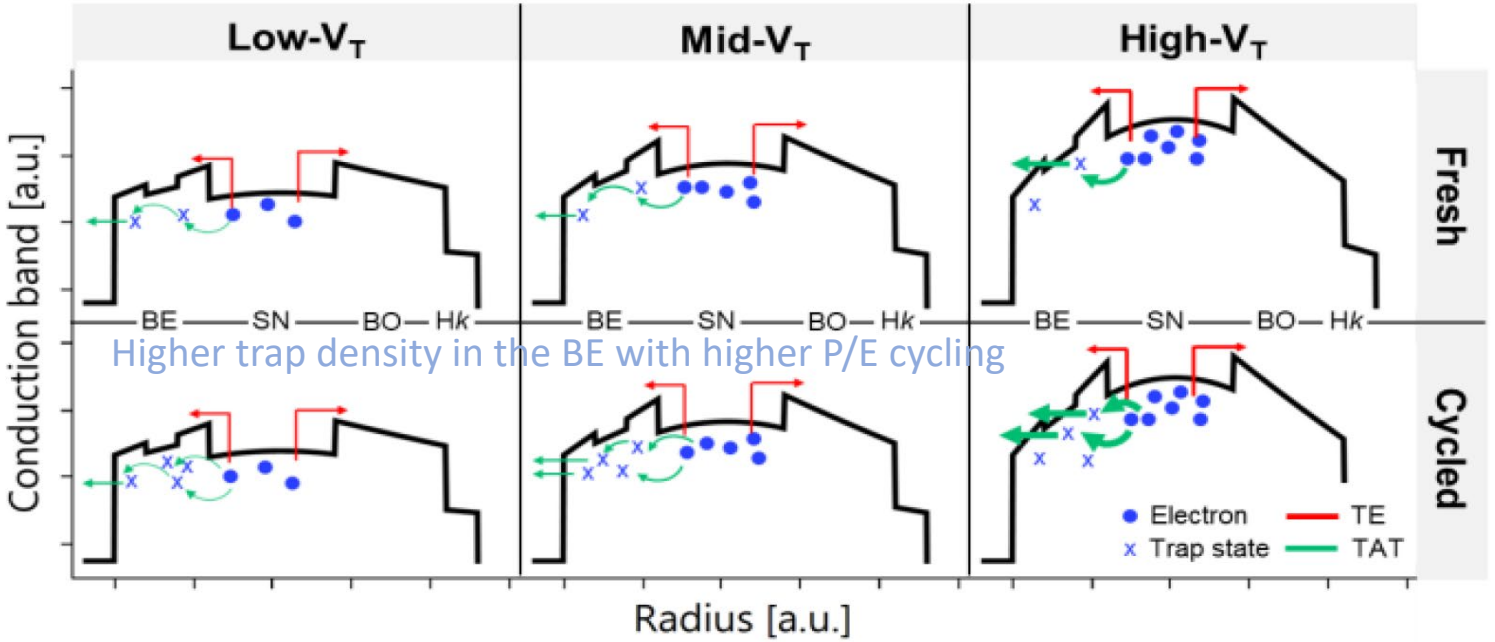
- **Vertical charge loss mechanisms from BE or CT**
  - **De-trapping from BE (Ea ~ 1.0eV)** [10], [11]
  - **ThAT (Thermally Assisted Tunneling) from CT** [10]
  - **DT (Direct Tunneling) from CT – weak temperature dependance (low Ea)** [10]
  - **TAT (Trap Assisted Tunneling) (Ea ~ 0.3-0.6eV)** [11]
  - **TE (Thermionic Emission)** [11]
  - **Poole-Frenkel Emission followed by PCAT –influenced by electric field** [9]

BE – Band Gap Engineered
BO – Blocking Oxide
PCAT – Positive Charge Assisted Tunneling

Vertical Charge Loss mechanisms during long-term retention operation



Source: [10]

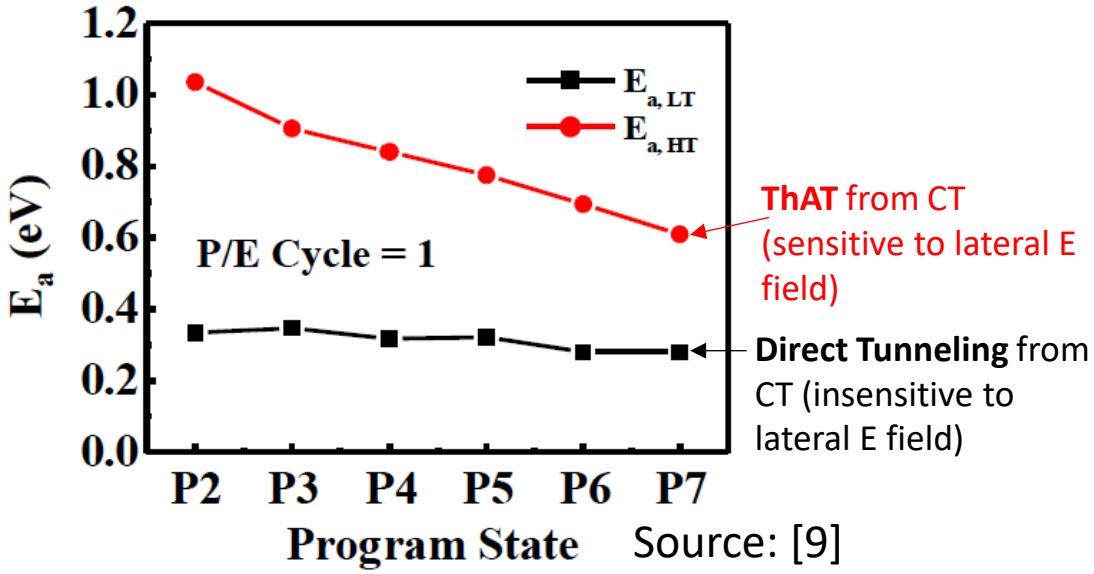


[11]

- **Vertical charge loss mechanisms from BE or CT**
  - **De-trapping from BE (Ea ~ 1.0eV)** [10], [11]
  - **ThAT (Thermally Assisted Tunneling) from CT** [10]
  - **DT (Direct Tunneling) from CT – weak temperature dependance (low Ea)** [10]
  - **TAT (Trap Assisted Tunneling) (Ea ~ 0.3-0.6eV)** [11]
  - **TE (Thermionic Emission)** [11]
  - **Poole-Frenkel Emission followed by PCAT –influenced by electric field** [9]

BE – Band Gap Engineered
BO – Blocking Oxide
PCAT – Positive Charge Assisted Tunneling

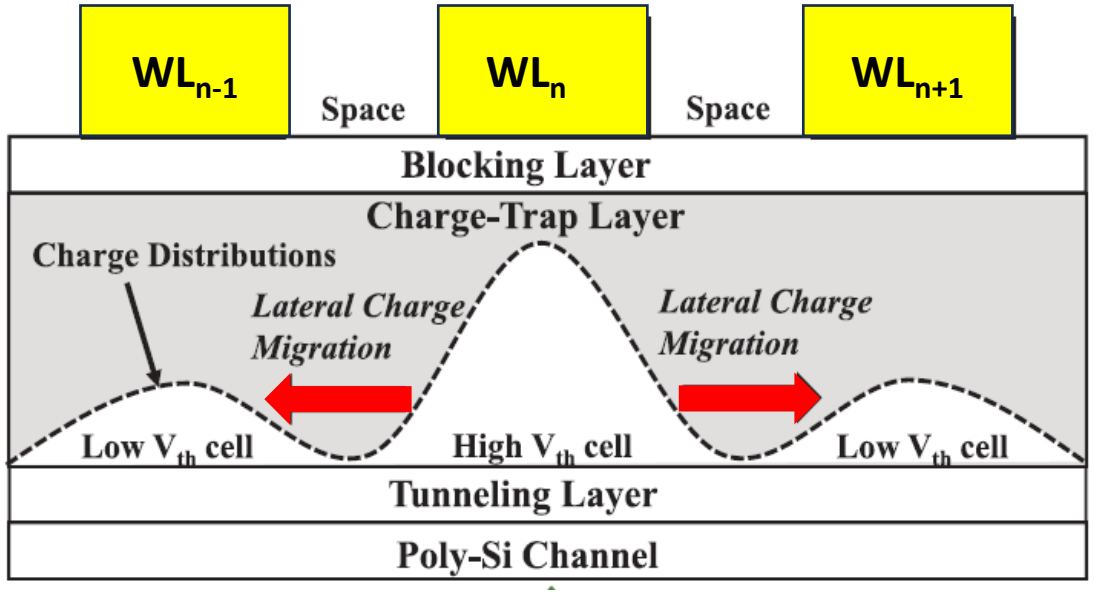
Lateral Charge Loss mechanisms during long-term retention operation



ThAT from CT (sensitive to lateral E field)

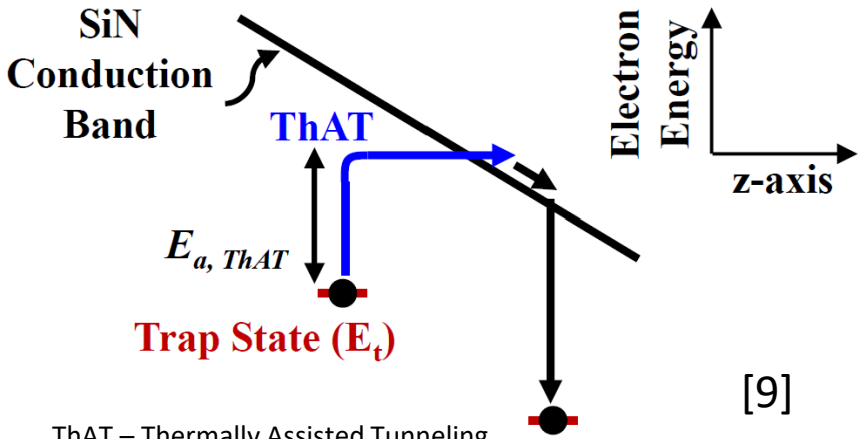
Direct Tunneling from CT (insensitive to lateral E field)

Source: [9]



Source: [6]

- **Lateral charge loss mechanisms in CT**
  - **At higher temperature region, Vt shift is believed due to the trapped electron lateral spread via ThAT (Thermally Assisted Tunneling) and/or Poole-Frenkel emission from CT**
  - **As program state increases (from P2 to P7), more electrons are injected, producing larger lateral electric fields**
  - **With more P/E cycling, accumulated charges in the intercell regions (between WL-WL) within charge trapping layer, suppressing lateral migration**

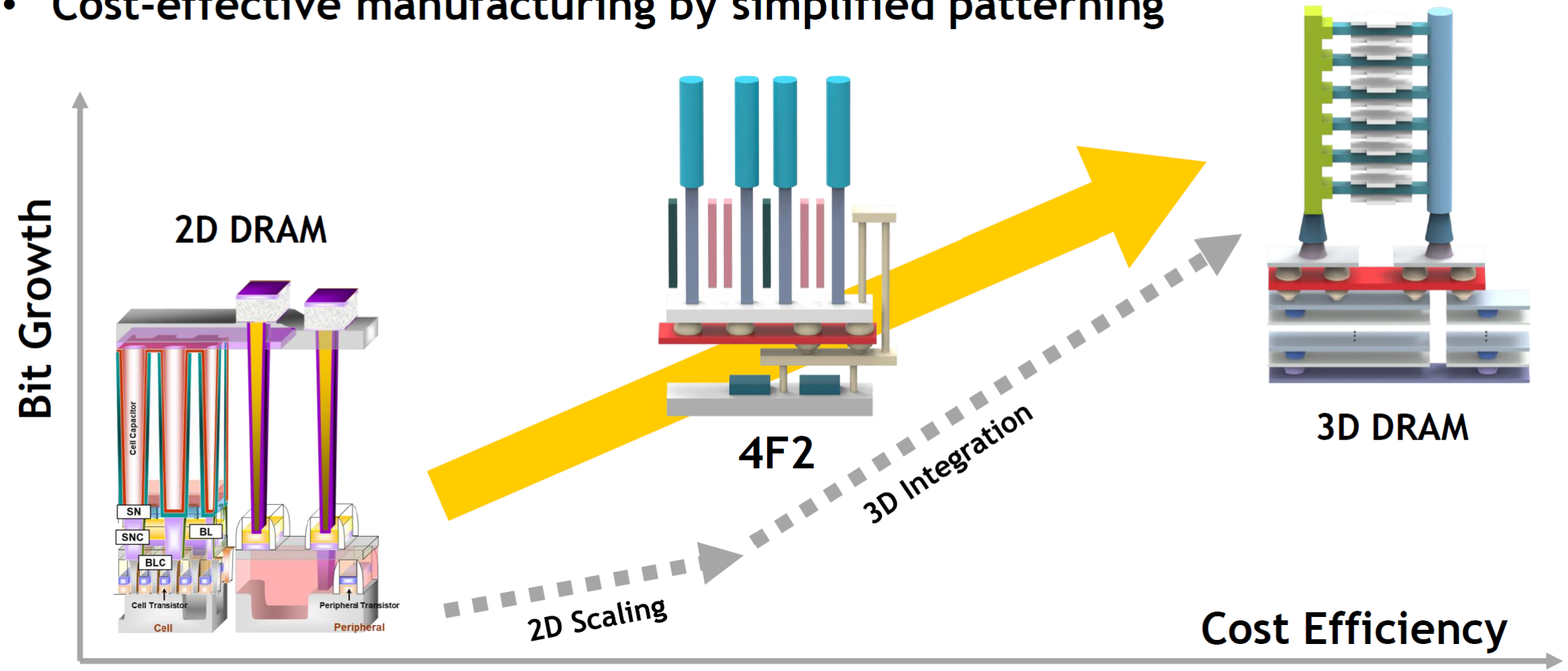


[9]

ThAT – Thermally Assisted Tunneling
CT – Charge Trap
LT – Low Temp (25-85C)
HT – High Temp (85-125C)

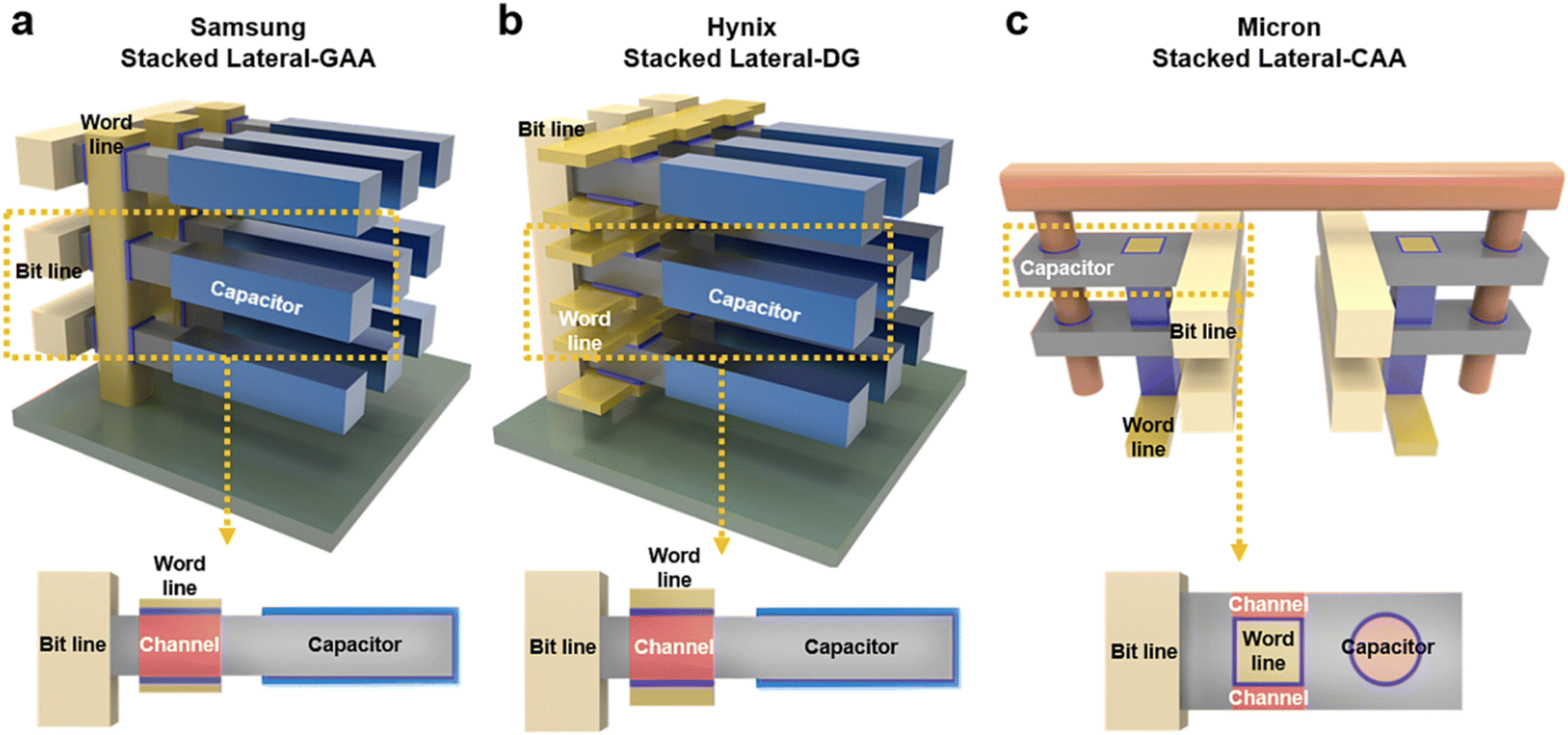
DRAM Scaling

- Bit growth enhanced through wafer bonding
- Cost-effective manufacturing by simplified patterning



Source: [13]

Prototypes of 3D DRAM



Source: [14]

DRAM Retention Loss Mechanisms

- **Capacitor leakage**
  - Direct tunneling due to MIM capacitor dielectric thickness scaling
- **Sub-threshold leakage**
  - Leakage of charge from the capacitor even when the word line is inactive
- **GIDL (Gate-Induced Drain Leakage)**
  - Occurs due to high electric fields near the drain when the gate is at a low voltage, accelerating charge loss
- **Junction leakage**
  - When the transistor is off, the drain (connected to the capacitor) is often at a higher voltage than the substrate, creating a reverse-biased p-n junction where minority carriers diffuse across the junction

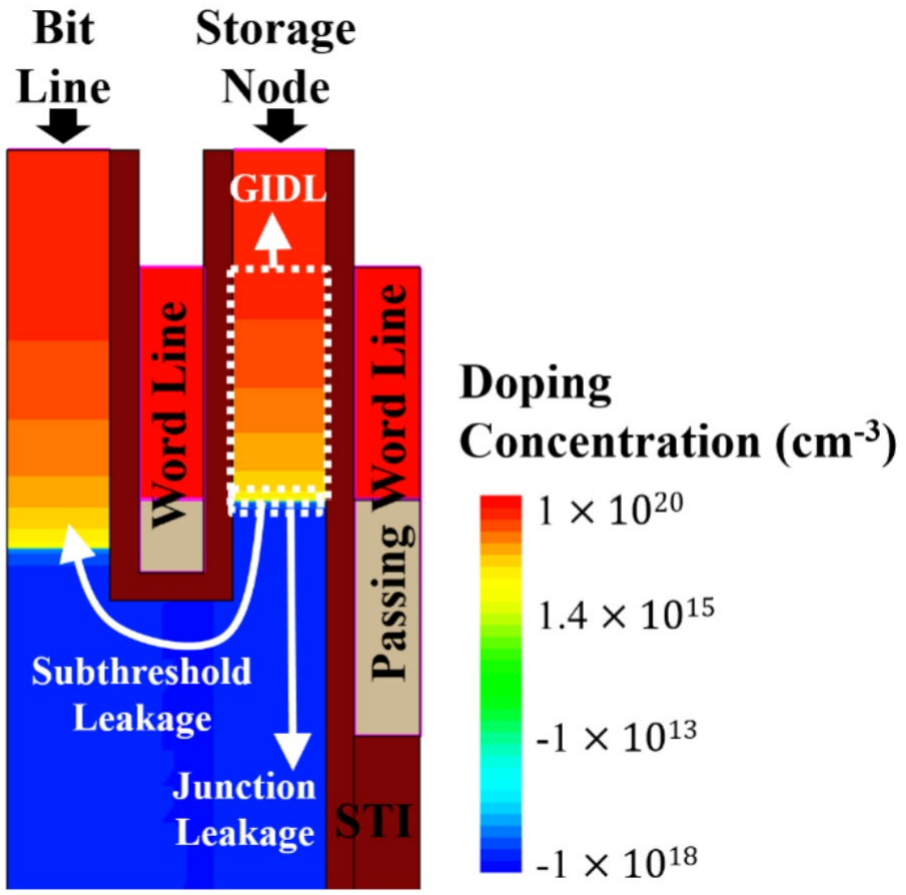


FIGURE12. Structure of a saddle fin based DRAM with three leakage current paths using TCAD simulation.

Source: [15]

Summary

1. Agentic AI requires more memory and storage with robust reliability
2. 3D NAND scaling reduces reliability margins, making data retention a critical concern
3. Understanding charge loss mechanisms in charge-trap NAND (CT-NAND) during retention is a foundational step toward improving NAND reliability
4. Short-term and long-term charge loss are governed by the energy distribution of charge traps:
    1. Shallow traps → dominate early-stage (short-term) loss
    2. Deep traps → dominate long-term degradation
5. Continued DRAM scaling also leads to reduced retention reliability and narrower noise margins
6. A systematic review of DRAM retention loss mechanisms is essential to support future technology scaling and error mitigation

Reference

[1] L. Crippa and R. Micheloni, “3D Charge Trap NAND Flash Memories,” in *3D Flash Memories*, R. Micheloni, Ed., Dordrecht: Springer Netherlands, 2016, pp. 85–127. doi: 10.1007/978-94-017-7512-0_4.

[2] T.-Y. J. Chang, F. Hamzaoglu, Y. Li, K. Sohn, Y. Wang, and J. Wuu, “A Trip Down Memory Lane: Reflections on the progress in memories,” *IEEE Solid-State Circuits Mag.*, vol. 15, no. 3, pp. 53–61, 2023, doi: 10.1109/MSSC.2023.3287169.

[3] X. Jia et al., “Impact of Cycling Induced Intercell Trapped Charge on Retention Charge Loss in 3-D NAND Flash Memory,” IEEE J. Electron Devices Soc., vol. 8, pp. 62–66, 2020, doi: 10.1109/JEDS.2019.2963473.

[4] K. H. Kim et al., “Material engineering to enhance reliability in 3D NAND flash memory,” Device, vol. 3, no. 2, Feb. 2025, doi: 10.1016/j.device.2024.100682.

[5] Y. H. Liu, T. C. Zhan, Y. S. Yang, C. C. Hsu, A. C. Liu, and W. Lin, “Impact of Trapped Charge Vertical Loss and Lateral Migration on Lifetime Estimation of 3-D NAND Flash Memories,” in 2023 IEEE International Reliability Physics Symposium (IRPS), 2023, pp. 1–6. doi: 10.1109/IRPS48203.2023.10118289.

[6] X. Fang et al., “High-Precision Short-Term Lifetime Prediction in TLC 3-D NAND Flash Memory as Hot-Data Storage,” IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol. 42, no. 10, pp. 3224–3235, 2023, doi: 10.1109/TCAD.2023.3240932.

[7] C. Woo, S. Kim, and H. Shin, “Cell Pattern Dependency of Charge Failure Mechanisms During Short-Term Retention in 3-D NAND Flash Memories,” IEEE Electron Device Lett., vol. 41, no. 11, pp. 1645–1648, 2020, doi: 10.1109/LED.2020.3023188.

[8] C. Woo et al., “Modeling of Charge Failure Mechanisms during the Short Term Retention Depending on Program/Erase Cycle Counts in 3-D NAND Flash Memories,” in 2020 IEEE International Reliability Physics Symposium (IRPS), 2020, pp. 1–6. doi: 10.1109/IRPS45951.2020.9129306.

[9] Y. H. Liu, T. C. Zhan, Y. S. Yang, C. C. Hsu, A. C. Liu, and W. Lin, “Impact of Trapped Charge Vertical Loss and Lateral Migration on Lifetime Estimation of 3-D NAND Flash Memories,” in 2023 IEEE International Reliability Physics Symposium (IRPS), 2023, pp. 1–6. doi: 10.1109/IRPS48203.2023.10118289.

[10] D. G. Refaldi, G. Malavena, L. Chiavarone, N. Gagliazzi, A. S. Spinelli, and C. M. Compagnoni, “Cryogenic Investigation of Vertical Charge Loss in 3D NAND Flash Memories,” in 2025 IEEE International Reliability Physics Symposium (IRPS), 2025, pp. 1–7. doi: 10.1109/IRPS48204.2025.10983672.

[11] L. Chiavarone, G. Nicosia, N. Righetti, and Y. Dong, “Experimental Segmentation of Vertical Charge Loss Mechanisms in Charge Trap-Based 3D NAND Arrays,” in 2024 IEEE International Reliability Physics Symposium (IRPS), 2024, pp. 1–5. doi: 10.1109/IRPS48228.2024.10529408.

[12] H.-N. Yoo et al., “Effect of Lateral Charge Diffusion on Retention Characteristics of 3D NAND Flash Cells,” IEEE Electron Device Lett., vol. 42, no. 8, pp. 1148–1151, 2021, doi: 10.1109/LED.2021.3088851.

[13] J. Park et al., “4F2 DRAM Integration with Vertical Gate (VG) Cell Transistor and Peri-Under-Cell (PUC) Architecture,” in 2025 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), 2025.

[14] J. S. Hur et al., “Oxide and 2D TMD semiconductors for 3D DRAM cell transistors,” Nanoscale Horiz, vol. 9, no. 6, pp. 934–945, 2024, doi: 10.1039/D4NH00057A.

[15] G. Lee, M. Suh, M. Ryu, Y. Lee, J.-W. Han, and J. Kim, “Investigation Into the Degradation of DDR4 DRAM Owing to Total Ionizing Dose Effects,” IEEE Access, vol. 11, pp. 97456–97465, 2023, doi: 10.1109/ACCESS.2023.3312926.