Abstract

- 측면 전하가 비트 오류의 상당 부분을 차지하고 있다
- 인코딩을 수정하여 취약한 상태 패턴을 제거한 Tail cut 제안
- SSD 수명과 읽기 대기 시간을 80%, 25% 향상

Introduction

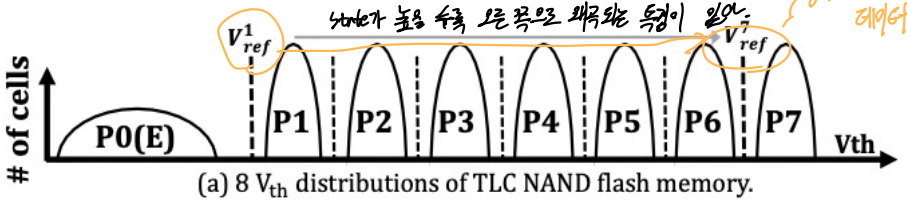
- 3D 플래시 메모리는 단으로 수직 방향으로 쌓음으로서 전과 눈을 수직적으로 인접한 셀로 전자가 이동
- 측면 전하 확산을 막는 새로운 제안은 3D-TLC 메모리를 사용하여 실시
- 측면 3가지 고찰
  1. 측면 전하 확산은 데이터에 크게 의존
  2. 약한 패턴은 retention pattern에 큰 영향을 받는다.
  3. Retention error는 SSD 수명과 read성능을 크게 낮춘다
- 측면 전하 확산을 최소화 하는 새로운 인코딩 방식인 Tail cut을 제안한다. 약한 데이터 패턴이 들어 왔을 때 데이터 비트를 플립시켜 오류를 줄인다.
- Tail cut은 FTL의 착상함으로 구현 할 수 있지만, 추가공간과 성능 저하가 발생한다. 따라서 본 논문에서는 제어 로직을 수정하여 기존 플래시 칩을 재사용하여 구현 하였다.
- SSD 에뮬레이터 플랫폼을 사용하여 Tc SSD라고 불리는 약한 패턴 인식 SSD를 구성 Read latency를 25%, SSD 수명을 80% 향상 시킴.

Contribution

측면 전하 분포를 측정한 첫번째 논문

- 측면 전하 분포를 측정한 첫번째 논문
- overhead가 적고 Weak pattern을 줄인 새로운 인코딩 방식인 tail-cut 제안
- Tc SSD를 사용한 실험에서 read latency를 25%, life time을 80% 향상 시킴

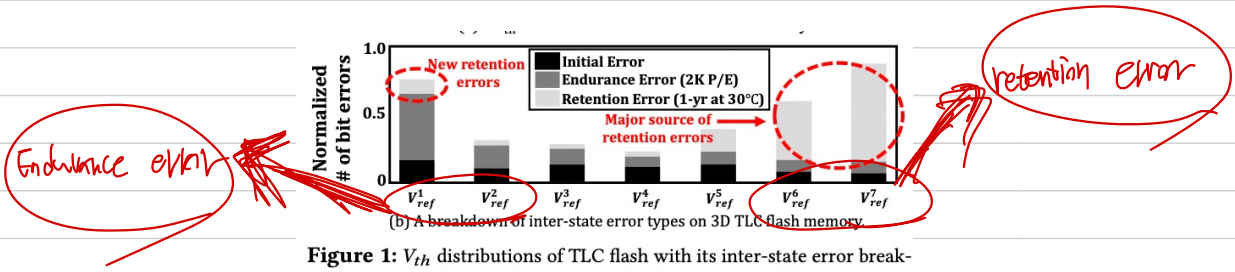
Reliability Characterization



(a) 8 V_{th} distributions of TLC NAND flash memory.

- write : LSB, CSB, MSB를 이용하여 3bit를 갖는 threshold voltage를 저장한다.
- 다양한 원인으로 인해 그래프가 왜곡되면 비트 오류가 증가
- 왜곡은 V_{th} 그래프가 넓어지고 V_{th} 분포가 겹치면 발생

Inter-state Error Distribution



(b) A breakdown of inter-state error types on 3D TLC flash memory.

Figure 1: V_{th} distributions of TLC flash with its inter-state error breakdown.

- Retention error는 65%를 차지하고 있고 3D 낸드가 2D보다 많이 일어난다.
- 이 현상은 LCS (Lateral Charge Spreading)에 영향을 준다.
- E state에서 2D 낸드는 retention error가 거의 일어나지 않음

2.2 Mechanism of LCS

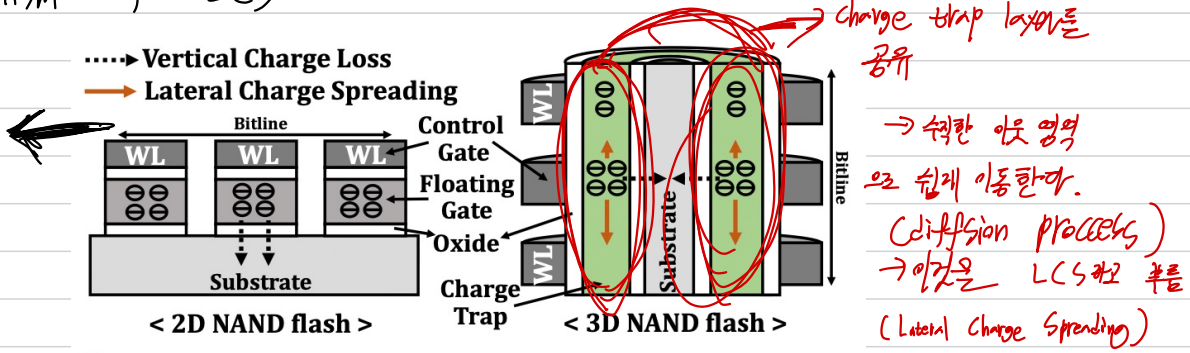


Figure 2: Differences in the charge loss mechanism between 2D and 3D flash memory.

3. Impact Analysis of LCS

- 여러 test page를 JEDEC standard에 따라 실험 해봤다.

3.1 Quantitative Evaluation

- 전자 농도가 차이 날 수록 Retention error가 더 많이 발생할 예상

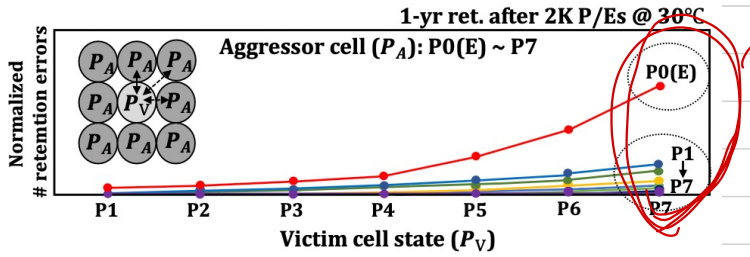
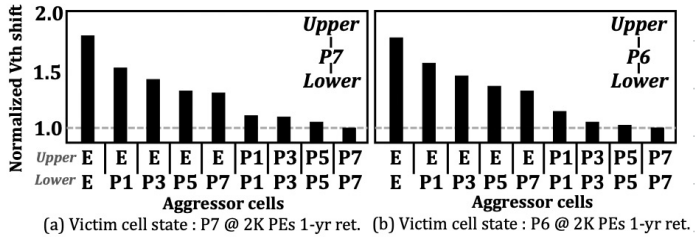


Figure 3: The effect of LCS on retention errors under varying test patterns.

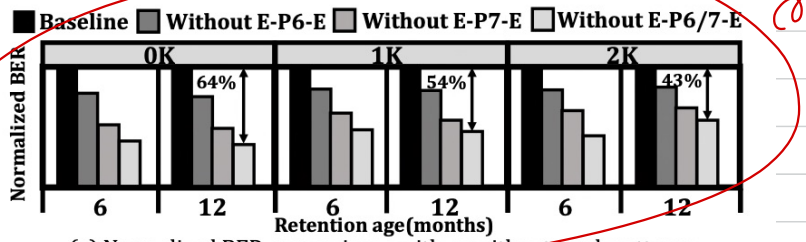
- 이를 통해 가장 약한 패턴은 E-P7-E 또는 E-P6-E로 표현 가능



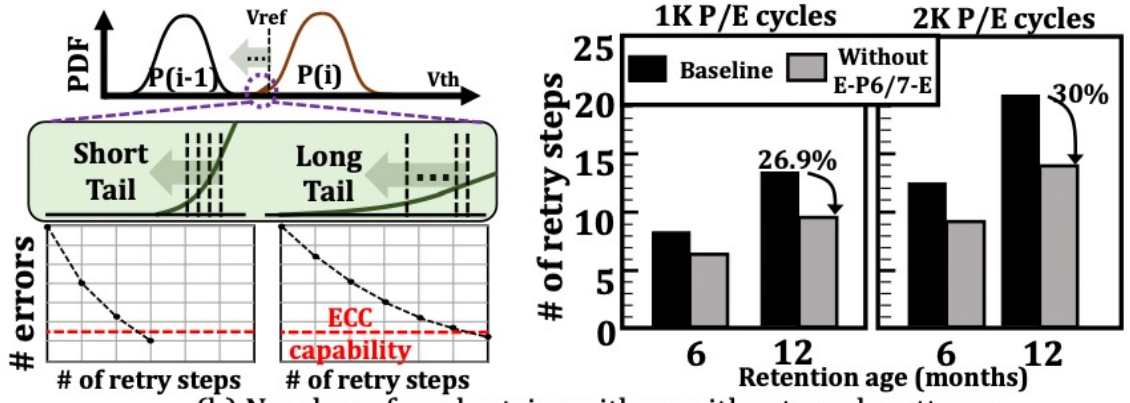
(a) Victim cell state : P7 @ 2K PEs 1-yr ret. (b) Victim cell state : P6 @ 2K PEs 1-yr ret.

Figure 4: Variations on V_{th} shift under different test patterns.

3.2 Impact on Lifetime and Performance



(a) Normalized BER comparisons with or without weak patterns.



(b) Number of read retries with or without weak patterns.

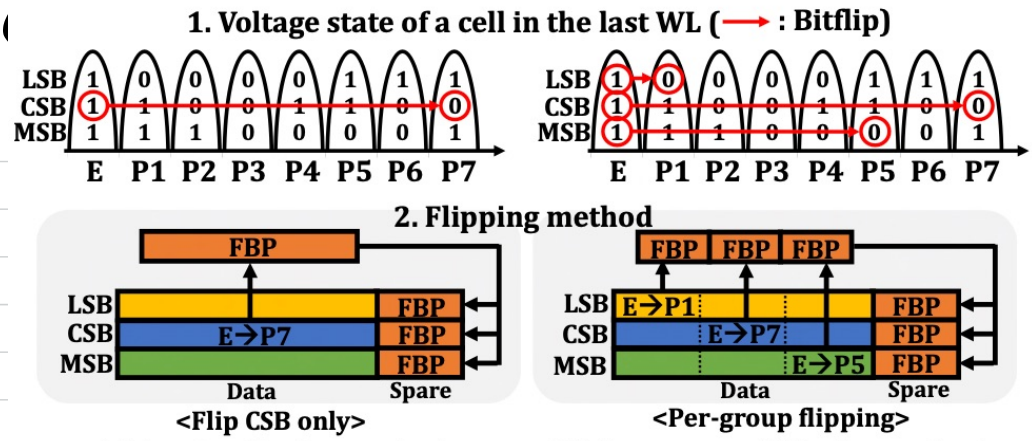
Figure 5: Impact of weak patterns on lifetime and performance.

→ Weak pattern 이 감소하여 SSD 수명과 Read retry를 효과적으로 줄임

4. PATTER-AWARE STATE ENCODING

4.1 Overview of TailCut

- E6-ER pattern을 non-E pattern으로 변경
- 이를 위해 비트 플립을 함.
- 비트 플립 위치를 저장할 공간은 FBP (Flipped Bit Position)이라고 함.



(a) A naïve flipping method. (b) Our proposed flipping method.

Figure 7: Comparisons of two bit flipping approaches.

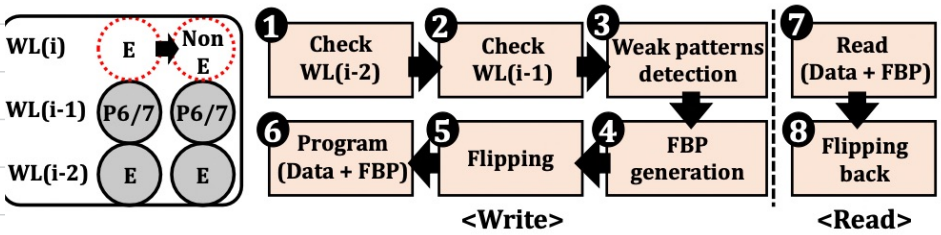
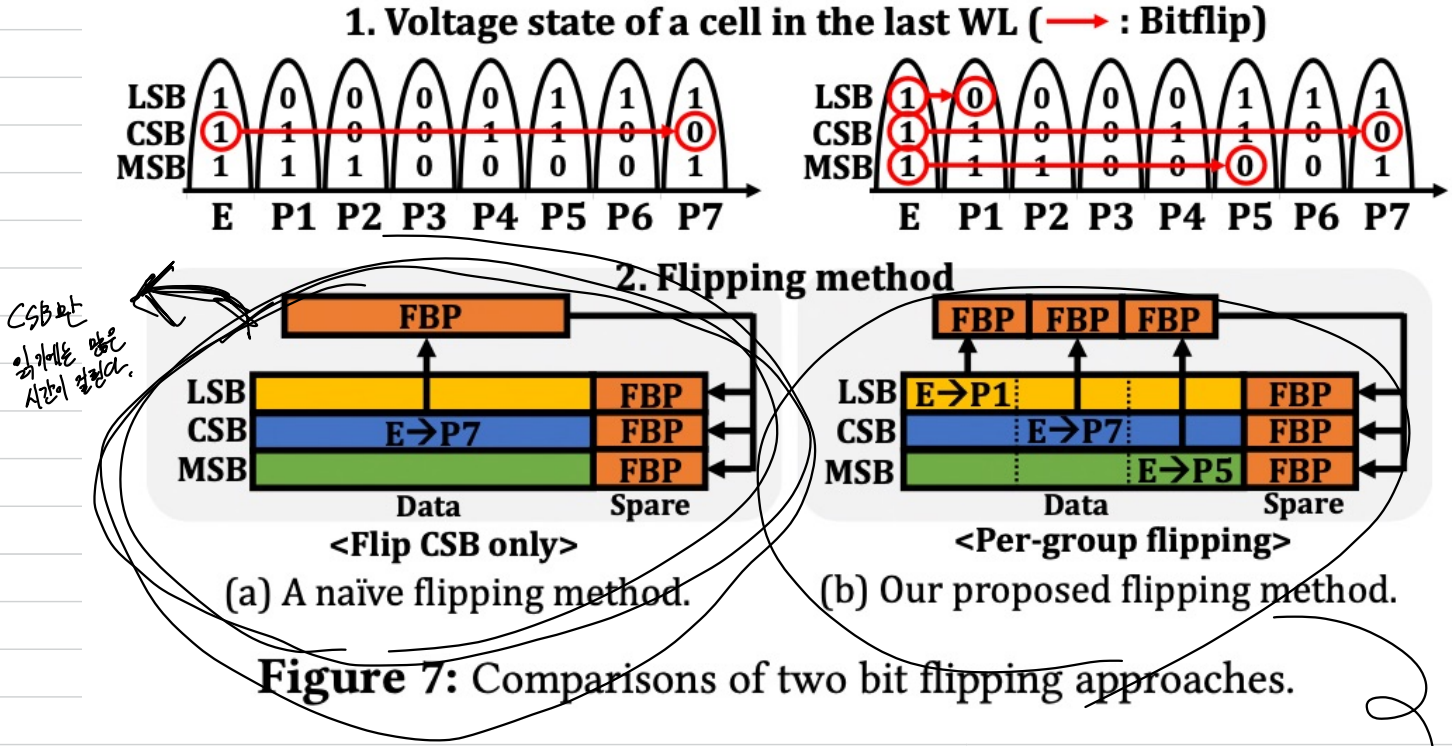


Figure 6: A high-level operational overview of `TailCut`.

4.2 Design Consideration of TailCut

Naive implementation은 용량과 performance에 큰 영향을 가지기 때문에 두가지 안을 제시

4.2.1 Design Consideration of TailCut

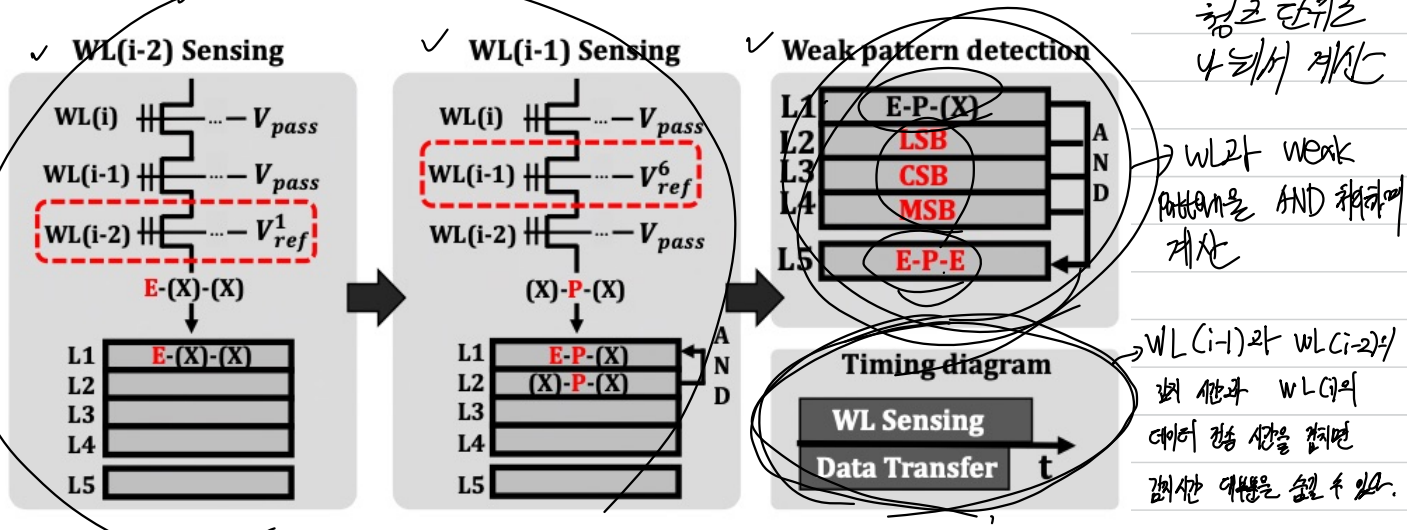


CSB만 읽어내는 많은 시간이 걸린다.

Figure 7: Comparisons of two bit flipping approaches.

4.2.2 Hiding the Overhead of detecting the weak pattern

각 그룹별로 청크 단위로 나뉘서 계산



WL과 weak pattern을 AND 처리해 계산

WL(i-1)과 WL(i-2)의 읽기 시간과 WL(i)의 데이터 전송 시간을 겹치면 감춰진 시간을 숨길 수 있다.

Figure 8: An operational illustration of the on-chip weak pattern detector.

→ WL(i)의 데이터 전송시간에 센싱

4.3 Overhead Analysis for On-chip Processing

- TailCut을 위한 Logic이 필요함

① FBP를 generate 시킴 ② ECC Engine을 통해 encode 시킴 ③ LSB에 bit flag를 1로 설정 ④ XOR을 통해 비트 플립을 시킴 ⑤ write

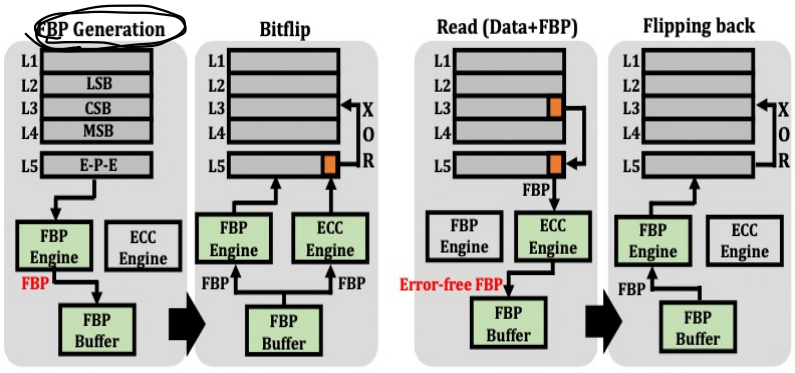


Figure 9: Illustration of on-chip modules for TailCut.

① FBP 데이터가 ECC 엔진을 통해 읽힘. ② ~~Xor~~ 작업을 통해 복구 시킴.

↳ Read.

4.3.1 Analysis on timing overhead

1. data sensing 방식과 data transfer time에서 온 overhead

ex) 데이터 전송량이 1GB/s이고 sensing의 시간이 40us 일때 overhead는 26us

2. 생성된 FBP 데이터를 저장하는데 걸리는 시간

ex) 페이지 크기를 16KB라고 할 때 데이터 전송속도가 1GB/s라고 하면 16us가 걸린다.

3. ECC engine의 시간은 숨겨져 있기 때문에 전반적으로 부서가 가능하다.

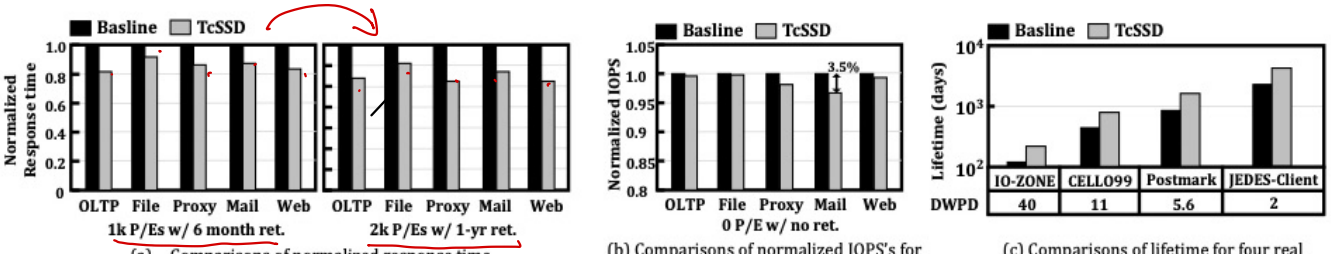
ex) 총 44 us, Program time 1900 us

4.3.2

- FBP 데이터는 page 당 144 B 필요하지만 저장 전 ECC engine에 의해 incoding 되어야 한다.
- 이는 48B 크기의 세 부분으로 나누고 48개 bit의 코드 워드로 나누어 최대 10bit의 오류를 검출 할 수 있다.
- 따라서 공간 overhead는 0.9%가 된다.
- die 안에 Three-on Chip module이 필요하지만 extra overhead가 필요하다. 하지만 무시할 만한 크기이다.
  (Buffer, ECC engine, FBP module)

5. Experimental set up

13~21% 상승



(a) Comparisons of normalized response time for five I/O workloads.

(b) Comparisons of normalized IOPS's for five I/O workloads.

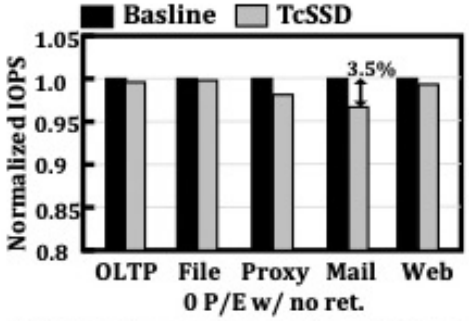
(c) Comparisons of lifetime for four real I/O workloads.

Figure 10: Comparisons of the performance and lifetime under different workloads.

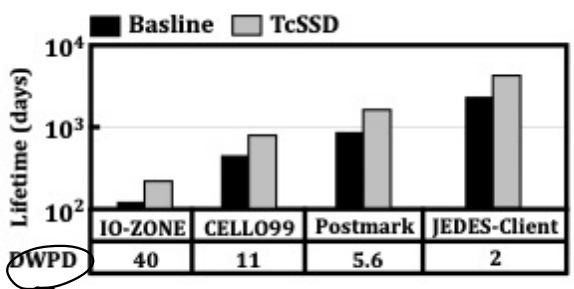
1) Read latency가 감소하기 때문에

2) data Retention이 증가하여

성능 수명 증가



(b) Comparisons of normalized IOPS's for five I/O workloads.



(c) Comparisons of lifetime for four real I/O workloads.