

# Memory Mapped ECC

## Low-Cost Error Protection for Last Level Caches

---



Doe Hyun Yoon

Mattan Erez

# 1-Slide Summary

---

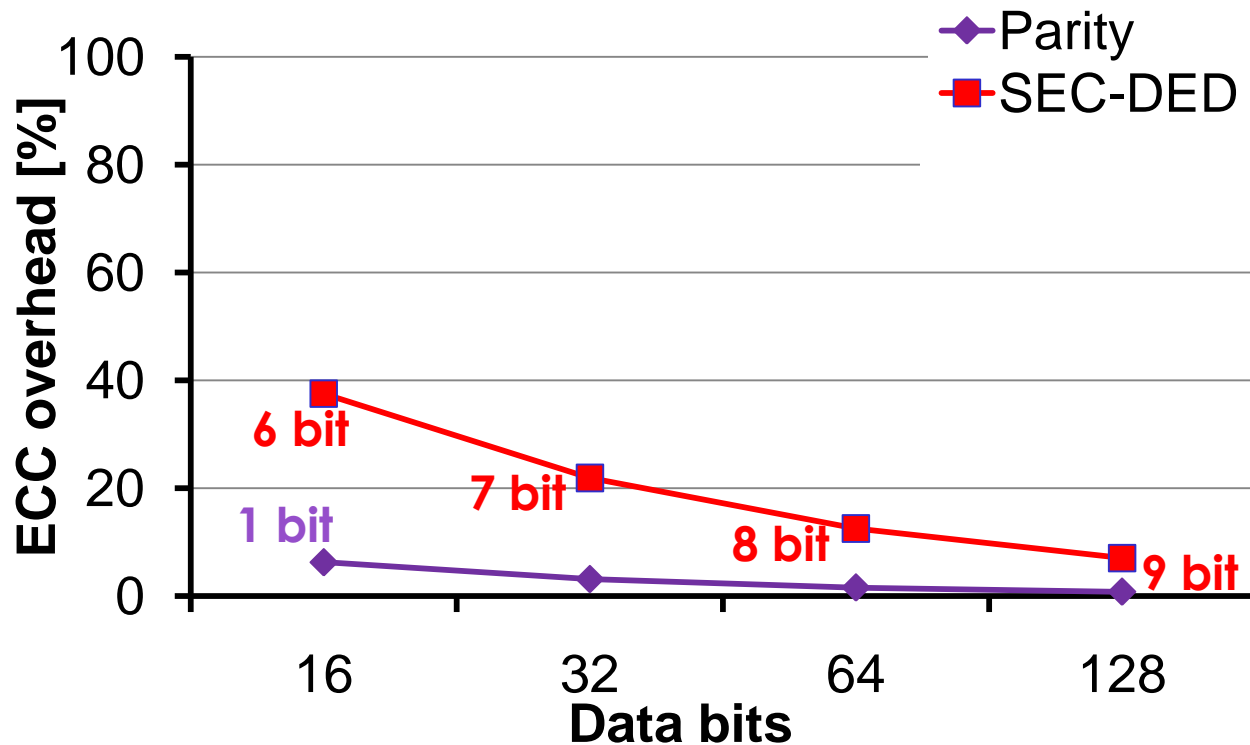
- Reliability issues in caches
  - Increasing soft error rate (SER)
  - Cost increases with error protection capability
- Memory Mapped ECC
  - Two-tiered error protection
    - Tier-1 Error Code: Light-weight on-chip error codes
    - Tier-2 Error Code: Strong error codes in DRAM namespace
      - T2EC storage is off-loaded to DRAM
      - No compromise on error protection
  - Achieved 15% area and 8% power reduction
  - Low performance overhead – 1.3% on average
  - Supports arbitrary error codes

# Background

---

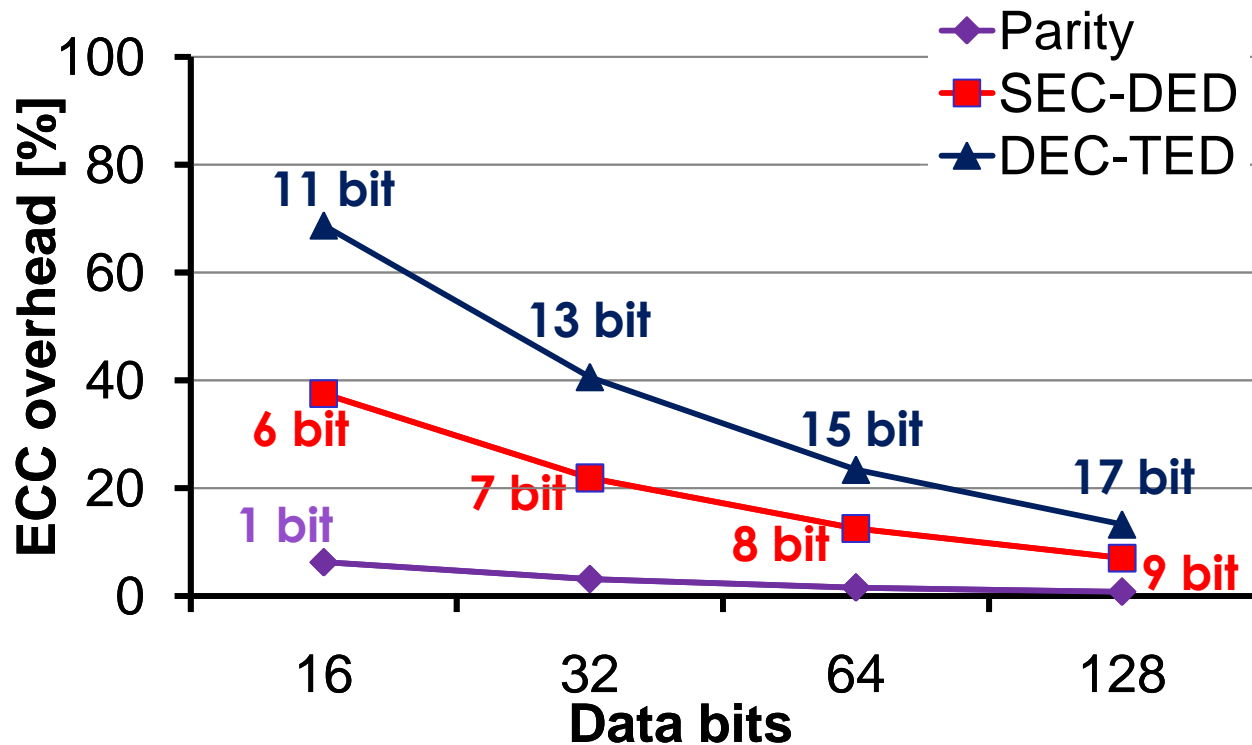
# Error Detecting and Correcting Codes

- Parity for error detection
- SEC-DED (Hamming code)
  - Single bit Error Correction and Double bit Error Detection



# Error Detecting and Correcting Codes

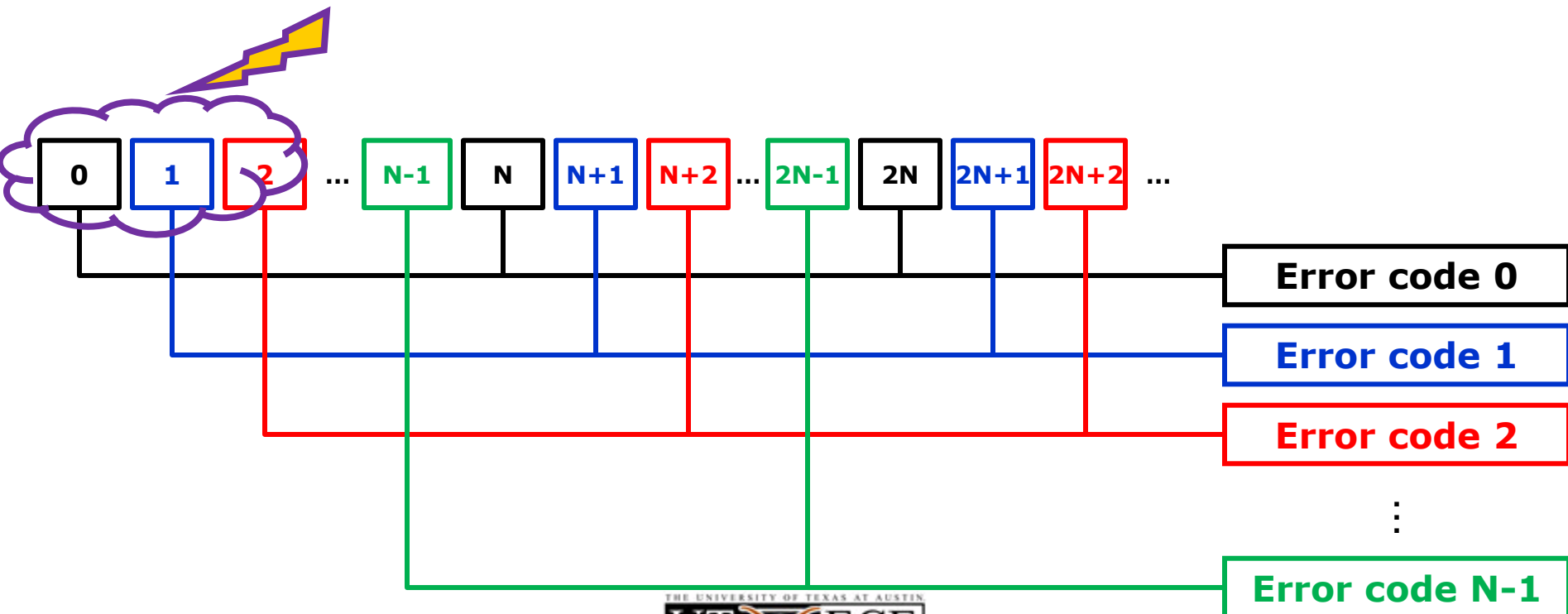
- Parity for error detection
- SEC-DED (Hamming code)
  - Single bit Error Correction and Double bit Error Detection
- DEC-TED
  - Double bit Error Correction and Triple bit Error Detection



# Interleaved Error Codes

- Interleaving

- To protect burst (adjacent) errors
- N-way Interleaved code makes an N-bit burst error to N single bit errors for each error code

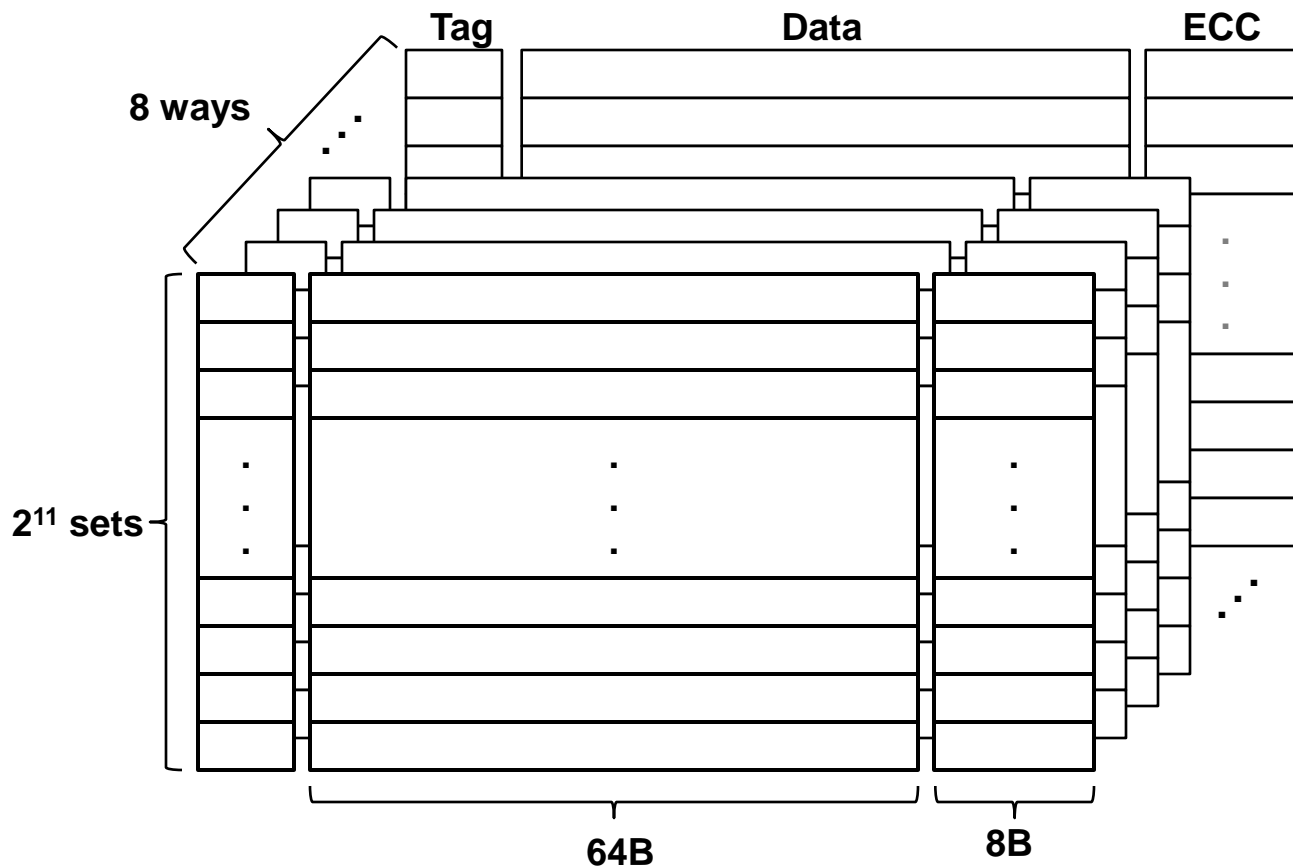


# Interleaved Error Codes

---

- Interleaving
  - To protect burst (adjacent) errors
  - N-way Interleaved code makes an N-bit burst error to N single bit errors for each error code
- 8-way interleaved SEC-DED for a 64B cache line
  - 8 SEC-DED codes (8B)
  - 13% Overhead
  - Can correct up to 8 bit burst errors

# Conventional Uniform Error Protection



**ECC increases area, leakage/dynamic power**





# Related Work

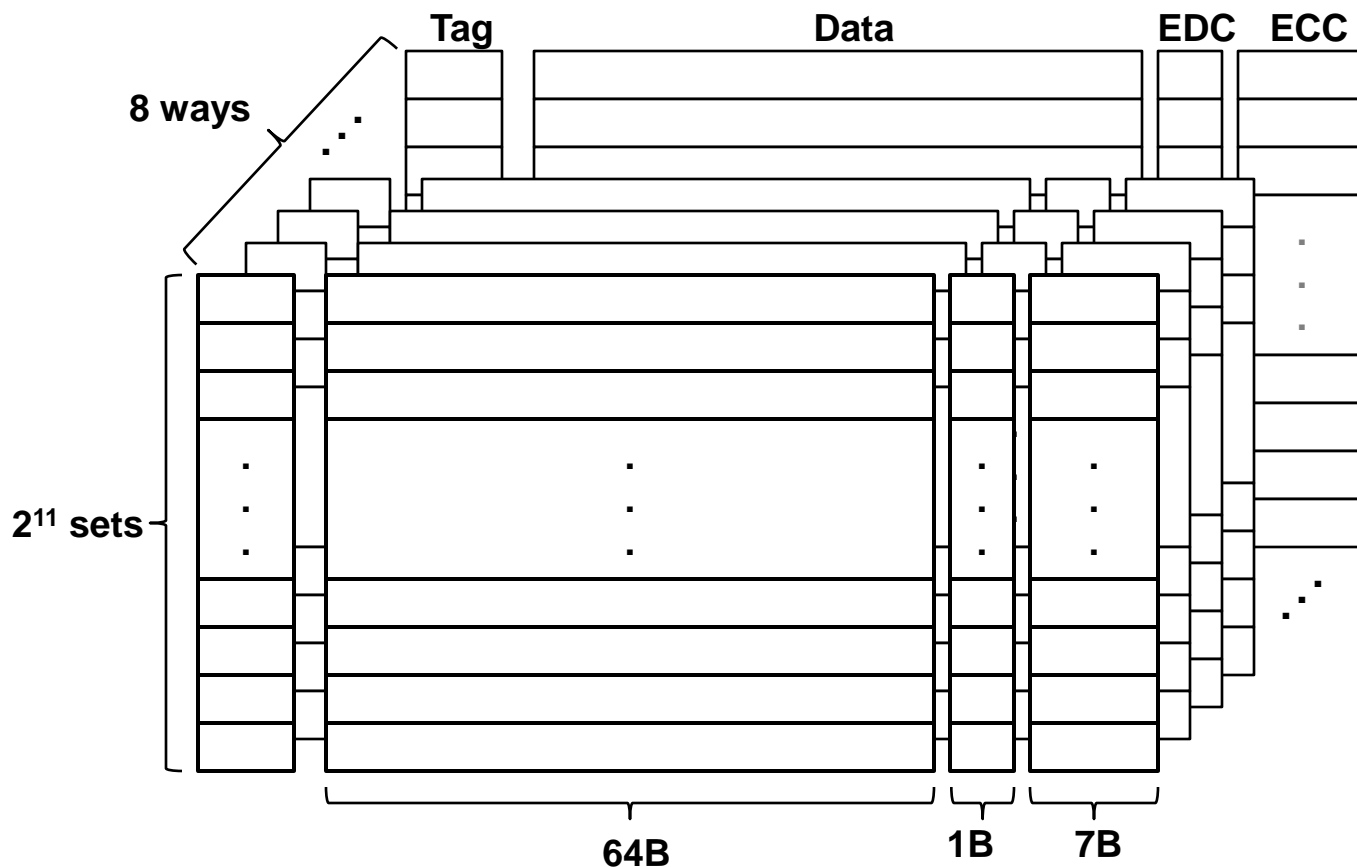
# Observations on Soft Errors in LLC

---

- Error detection – common case operation
  - Every cache line read requires error detection
- Only dirty cache lines need ECC
  - Memory hierarchy provides multiple copies for clean lines

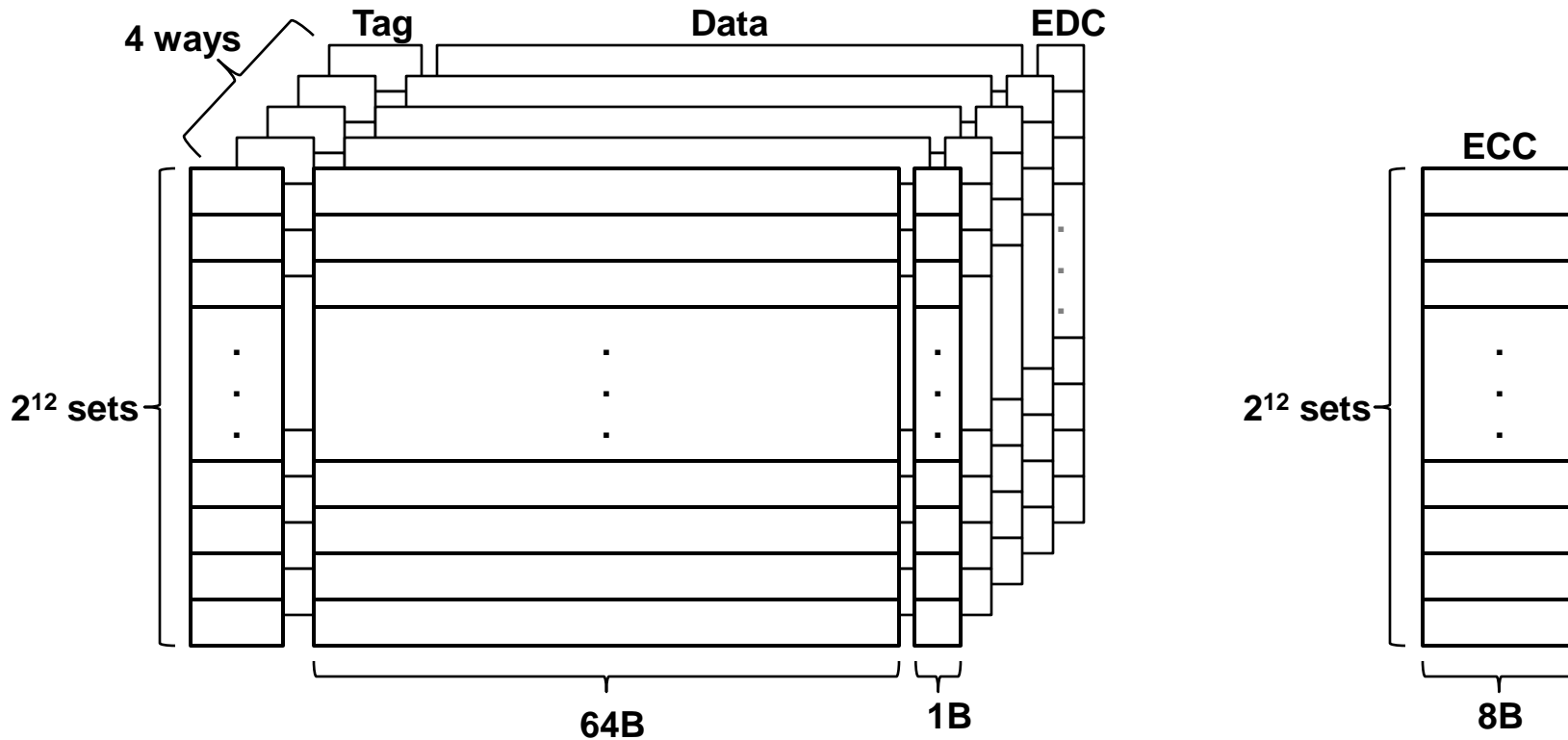


# Prior work – PERC [Sorin 06] and Energy Efficient [Li 04]



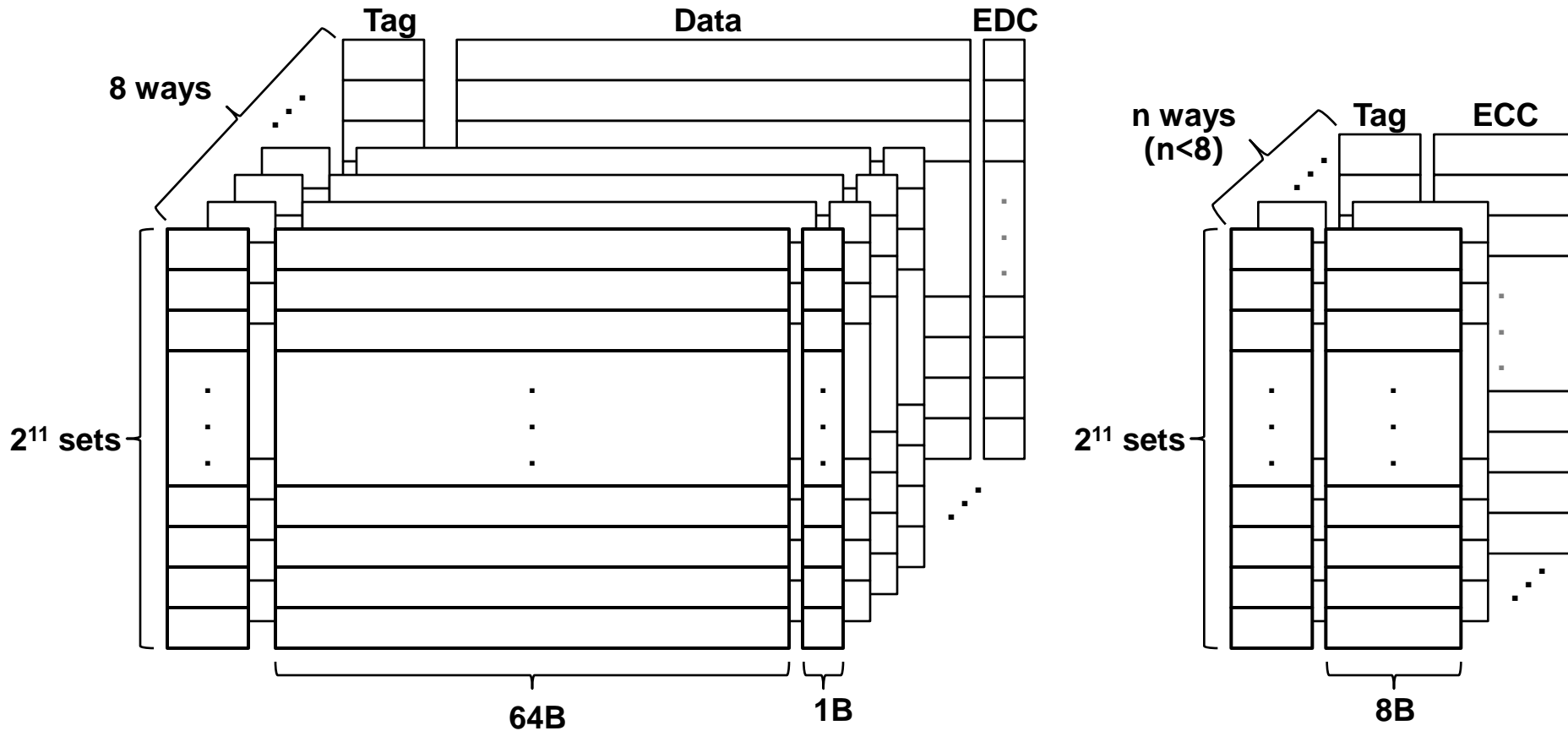
**Read only data and EDC – save dynamic power**  
**Power gate ECC of clean lines – save static power**

# Area Efficient Scheme [Kim 06]



Allow only 1 dirty line per set

# MAXn Scheme



**Allow only  $n$  dirty lines per set**  
**May cause detrimental cleaning traffic**



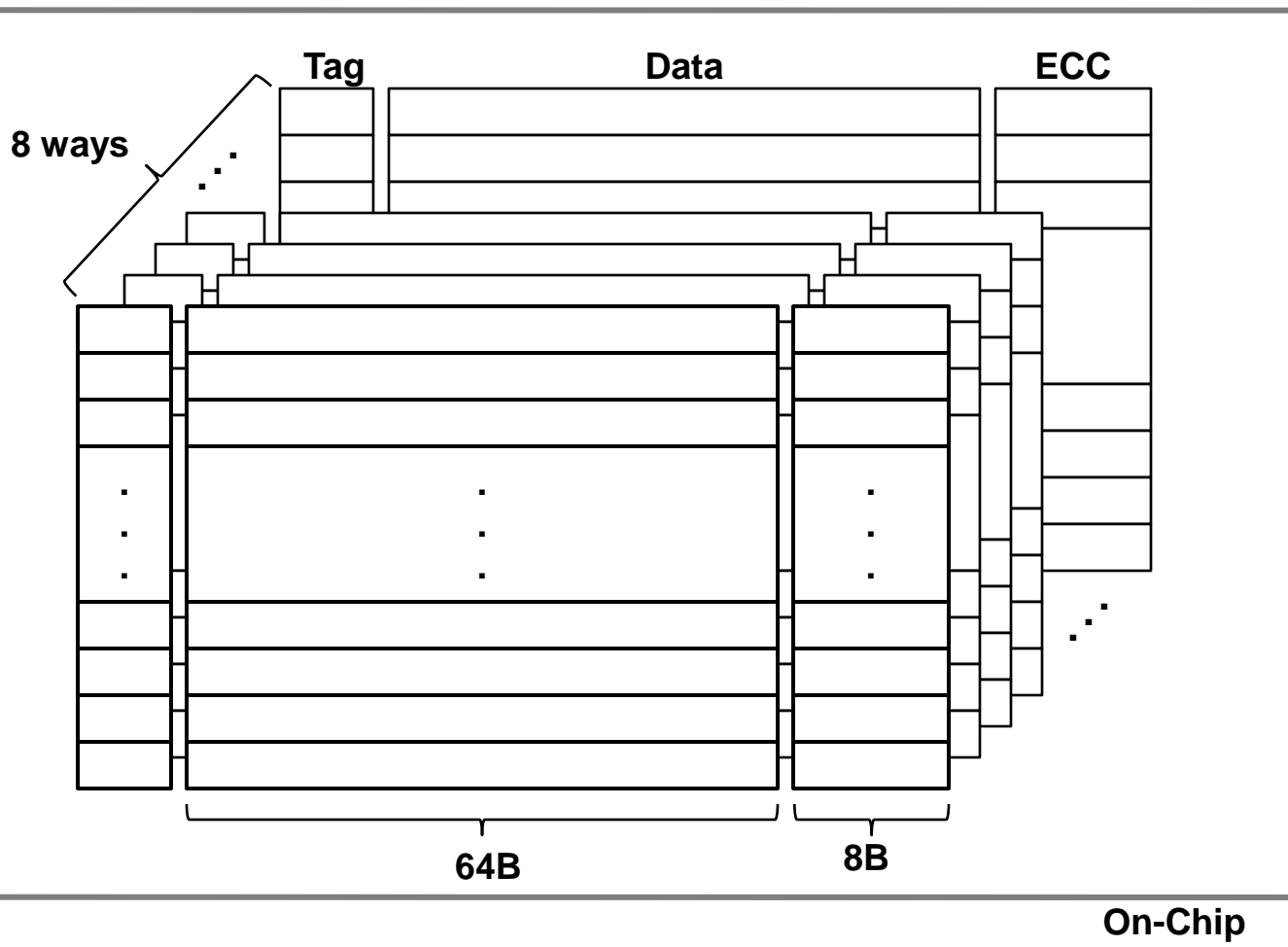
# Memory Mapped ECC

# Observations on Soft Errors in LLC

---

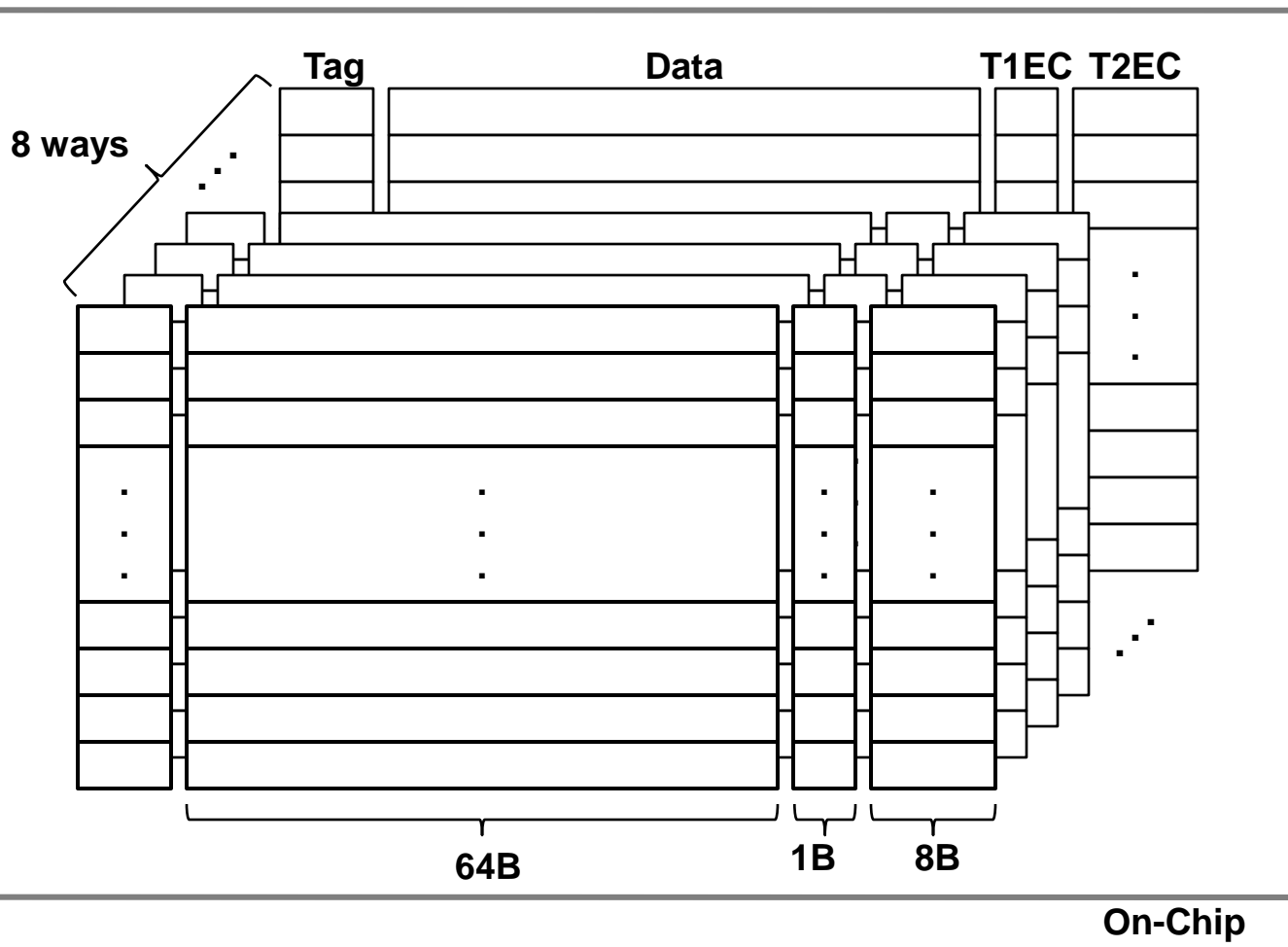
- Error detection -- common case operation
  - Every cache line read requires error detection
- Only dirty cache lines need ECC
  - Memory hierarchy provides multiple copies for clean lines
- Soft Error Rate is increasing, but is still very low
  - Mean time to failure (MTTF) of a 32MB cache is estimated as 155 days even with pessimistic assumptions
- Error correction – uncommon, extremely rare, case
  - latency/complexity is not important

# MME Architecture – 15% Area Saving

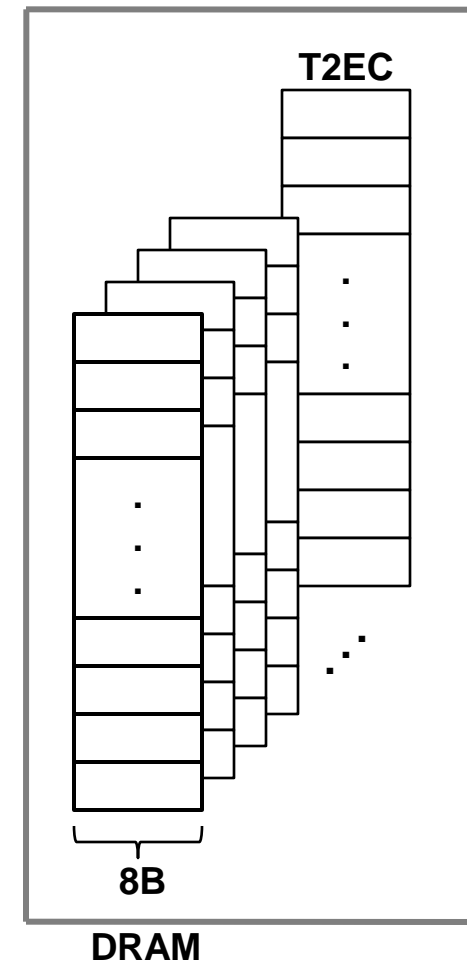
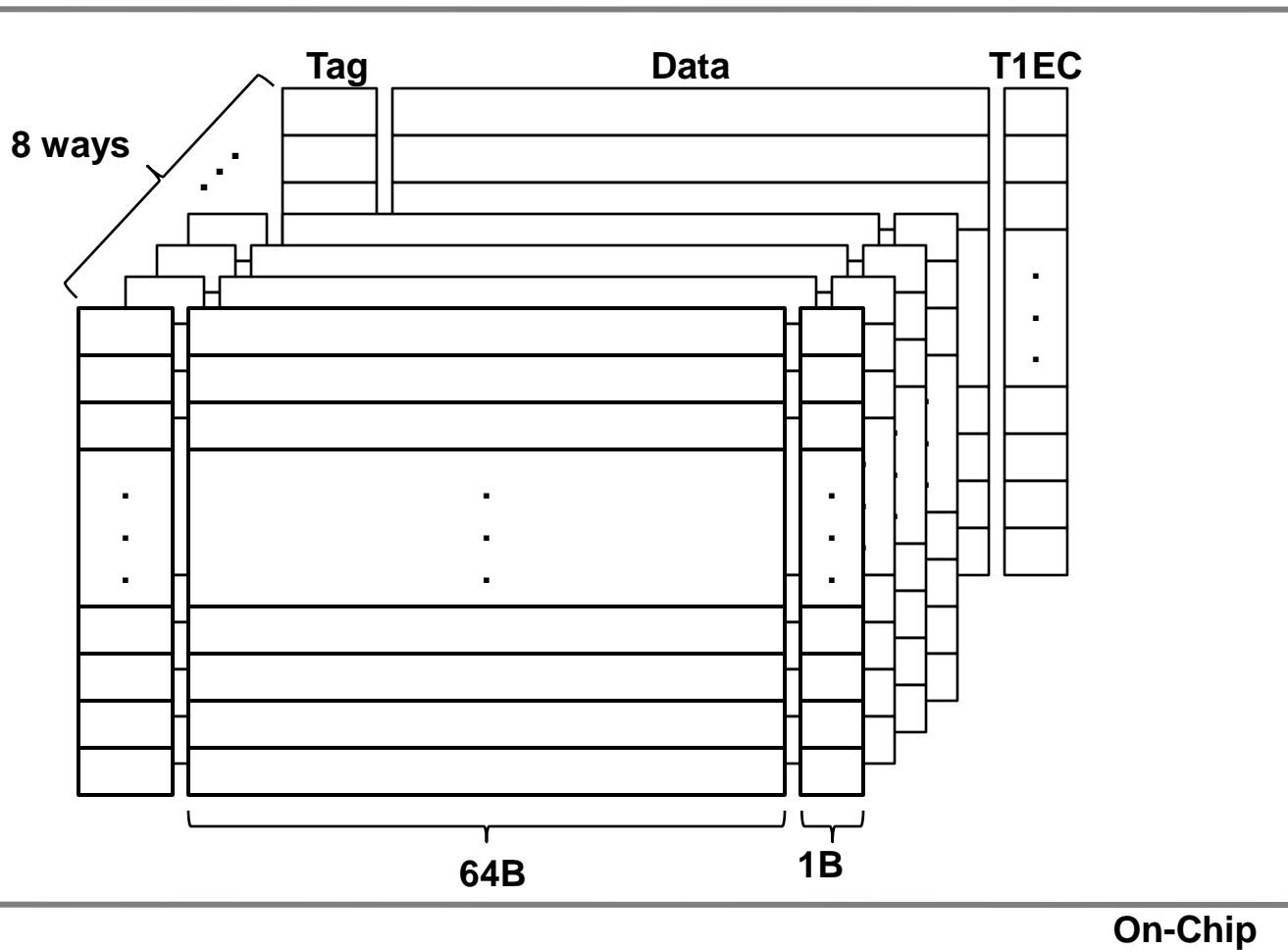




# MME Architecture – 15% Area Saving

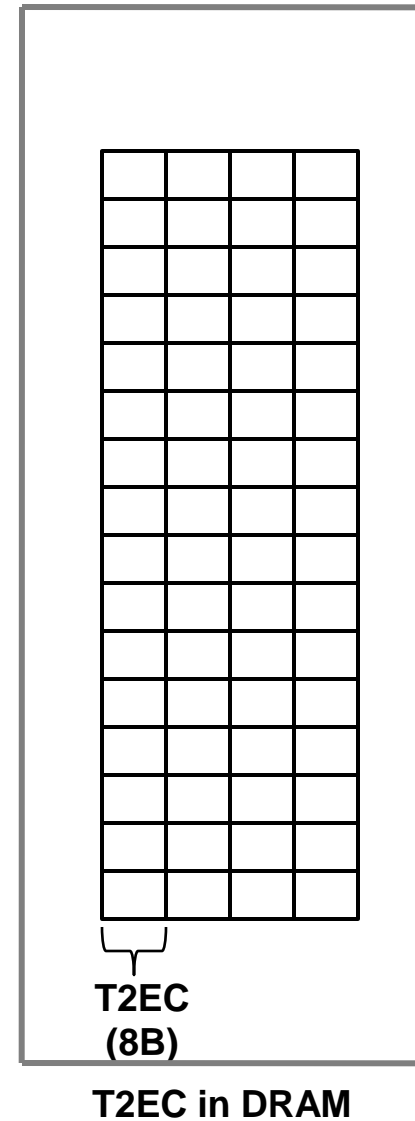
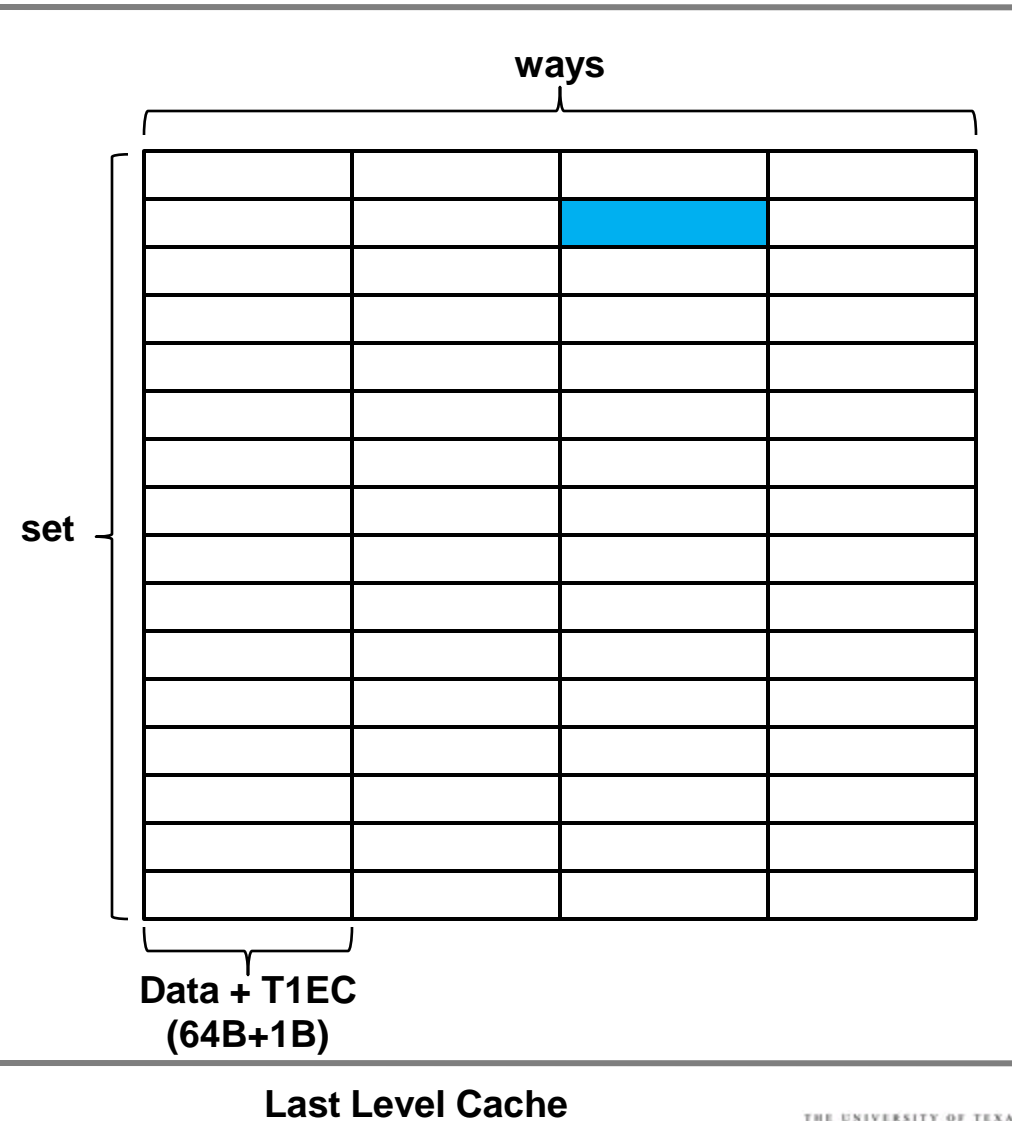


# MME Architecture – 15% Area Saving

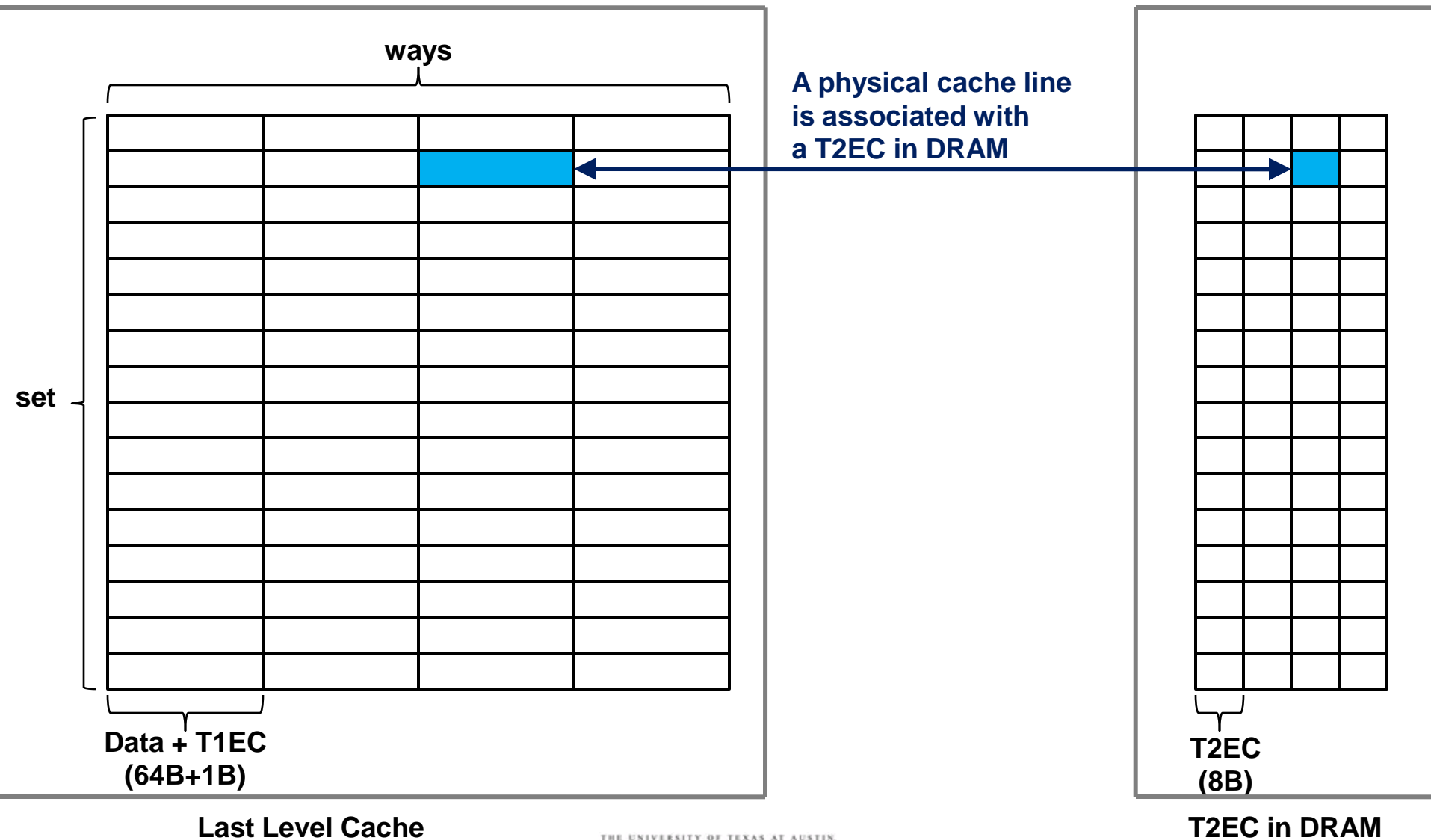


**T2EC is memory mapped to DRAM namespace**

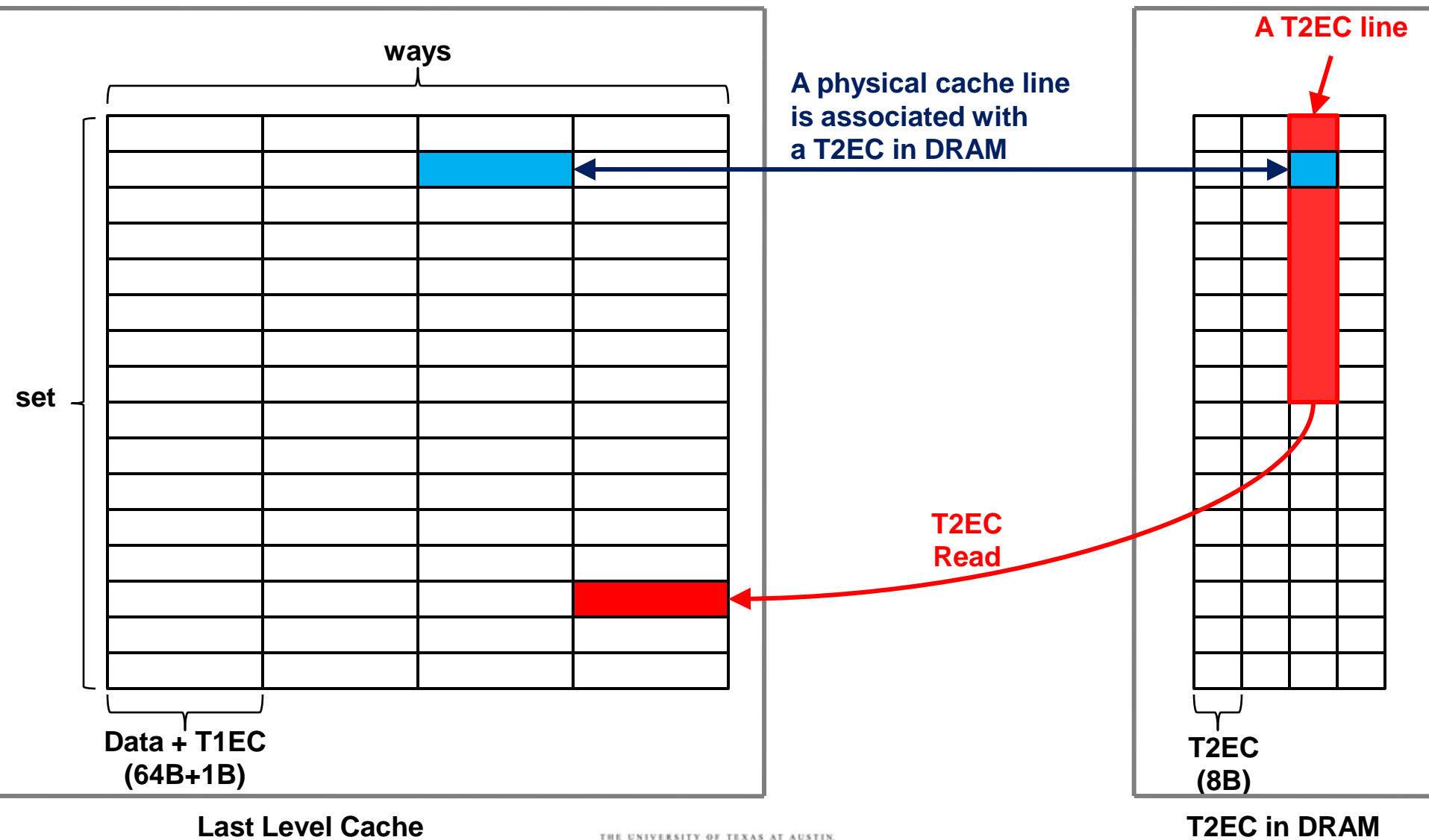
# Managing T2EC information



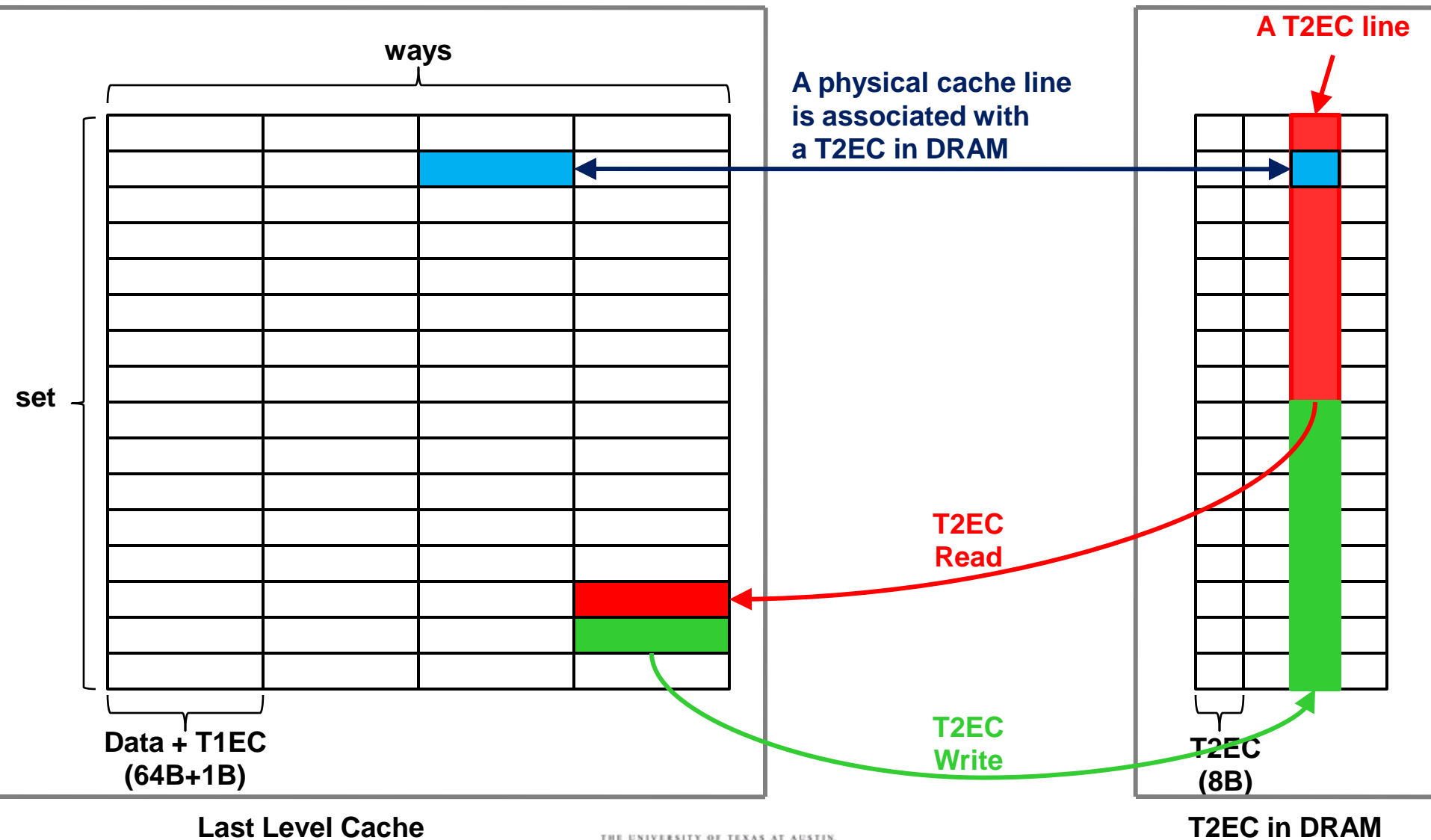
# Managing T2EC information



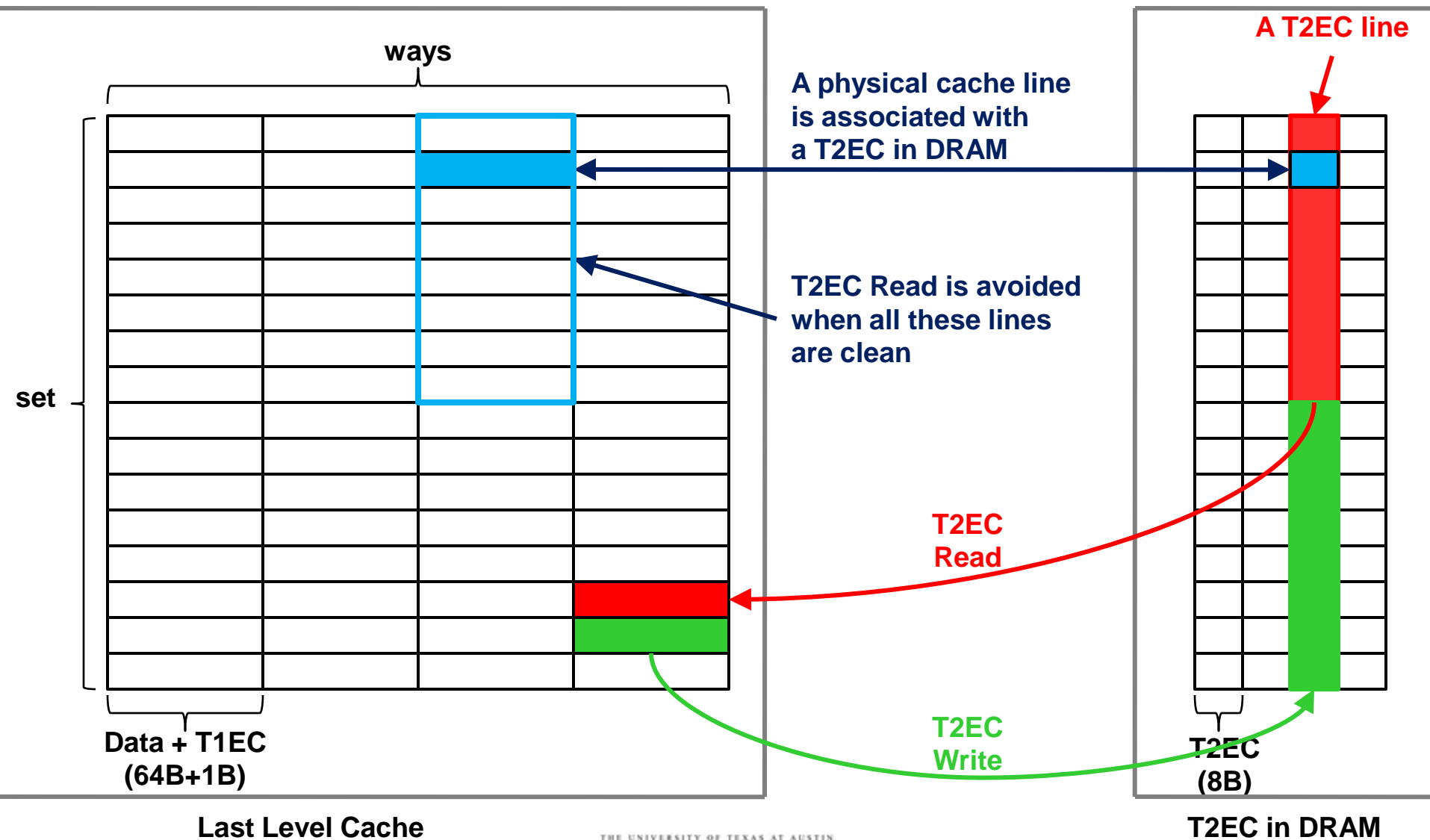
# Managing T2EC information



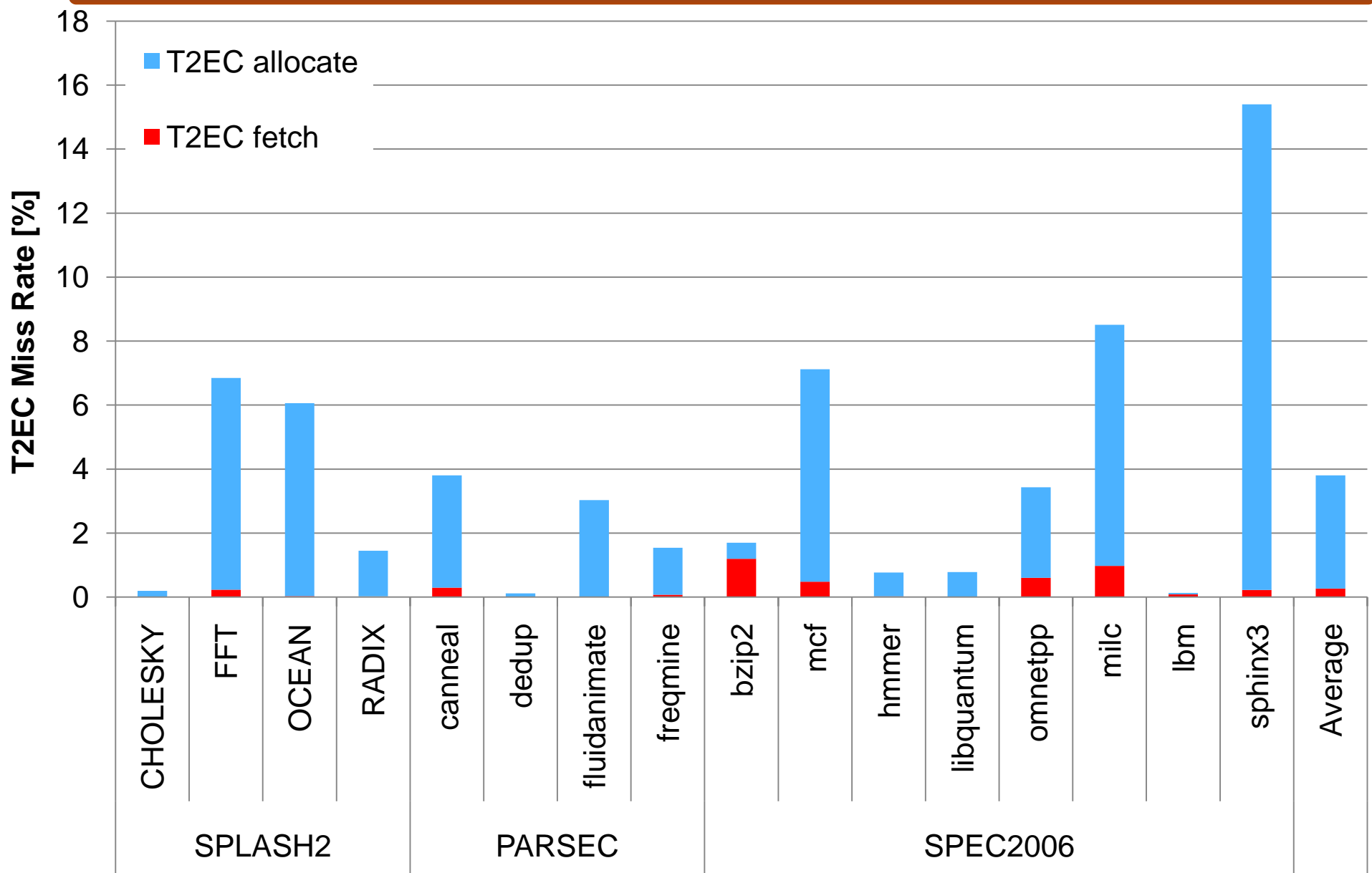
# Managing T2EC information



# Managing T2EC information



# T2EC Miss Rate





# Implementation Details

---

- T2EC is inherently private to a cache
  - Can be easily integrated with cache coherent multi processor systems
- T2EC region in DRAM
  - 128kB for a 1MB cache in our example (8way SEC-DED as T2EC)
  - Relatively small compared to DRAM capacity
  - Can be reserved by OS or BIOS
- A generic way of providing meta-data for physical cache lines
- Performance impact
  - Increased cache miss rate due to T2EC in caches
  - Increased traffic of T2EC reads and writes

# Memory Mapped ECC – RECAP

---

- Two-Tiered Error Protection
- Low on-chip ECC cost
  - Only T1EC on-chip, T2EC storage is off-loaded to DRAM
  - 15% area and 8% power in a 1MB cache (45nm Cacti model)
- No compromise on error protection capability
  - T2EC provides strong error protection
- T2EC is *memory mapped* and *cacheable*
  - Dynamic and transparent partition of LLC into data and T2EC
- Worst case error correction is roughly DRAM latency
- Flexibility in design T1EC/T2EC
  - T1EC: error detection, T2EC: error correction
  - T1EC: light-weight error correction, T2EC: strong error correction
  - More examples in our paper



# Evaluation

# Evaluation

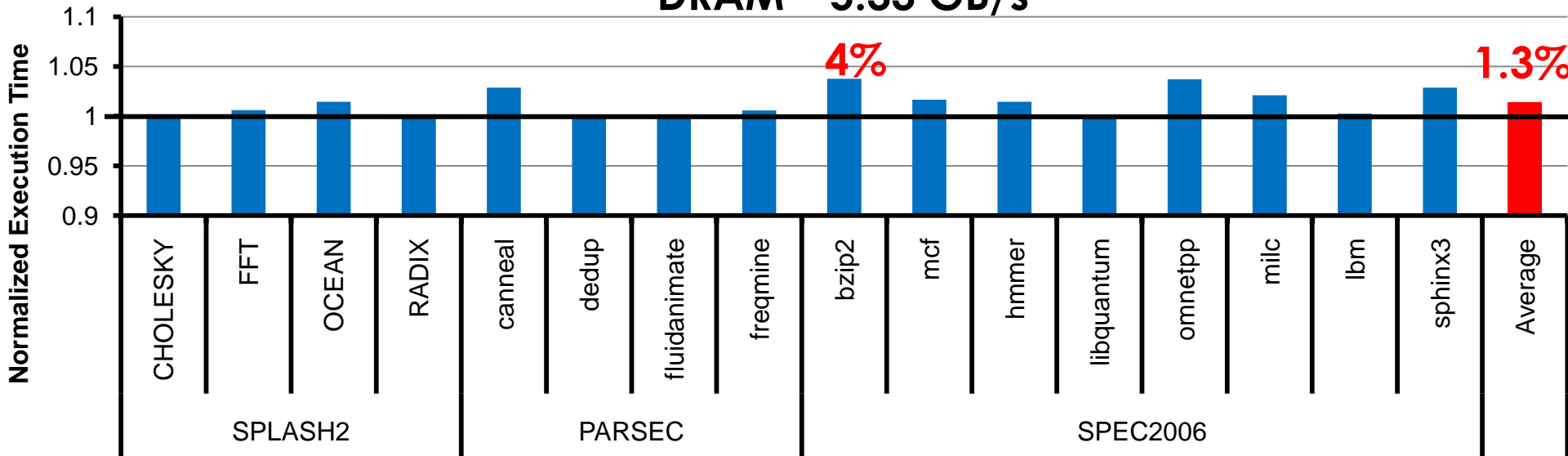
---

- GEMS + DRAMsim
  - An out-of-order 3GHZ SPARC processor core
  - Exclusive two level cache hierarchy
    - L1: split I/D, each 2-way 64kB
    - L2: unified 8-way 1MB
  - DDR2 667MHz DRAM: 5.33GB/s
  - Eager write-back is integrated
    - Dirty lines are periodically scanned and cleaned
- Workloads
  - 16 data-intensive applications from SPLASH2, PARSEC, and SPEC 2006
  - **Other non data-intensive applications are insensitive to our technique**
    - **Omitted for clarity and emphasize degradation**

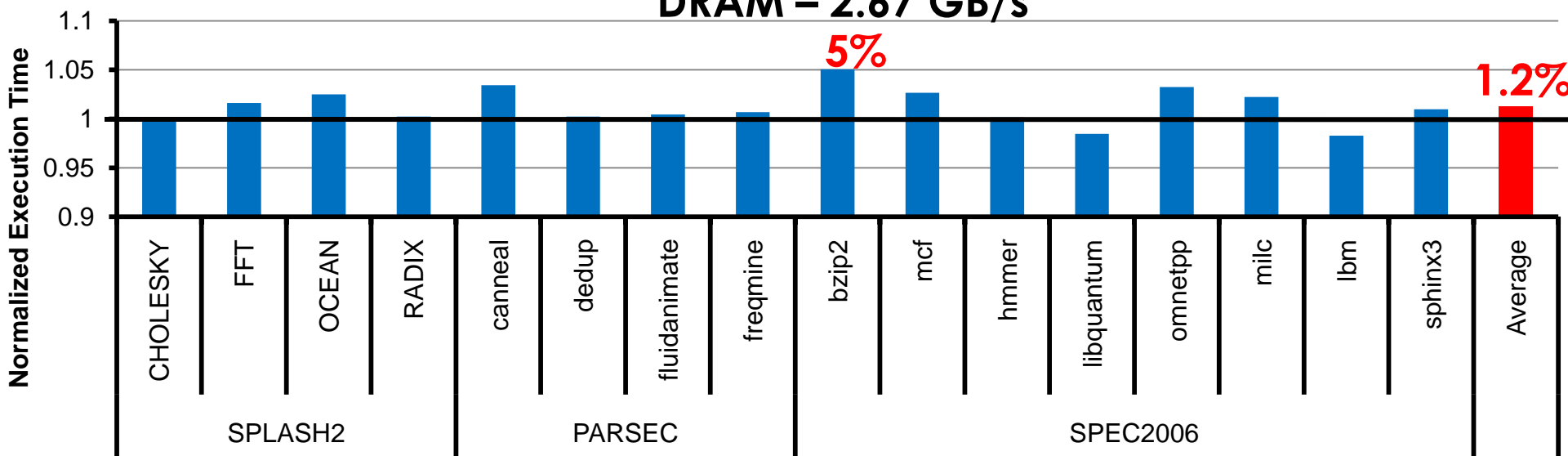


# Performance – 1.3% Penalty on Average

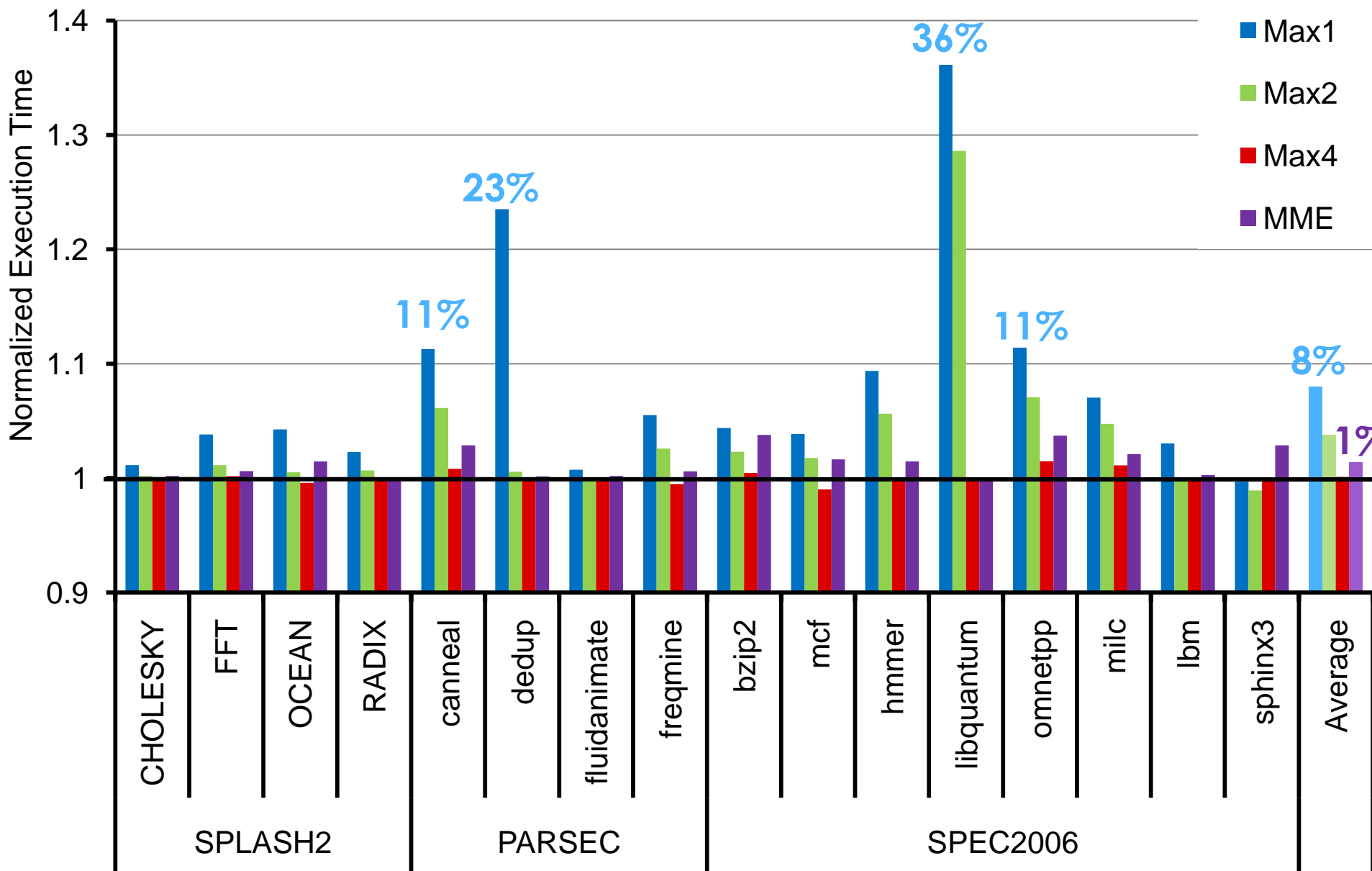
## DRAM – 5.33 GB/s



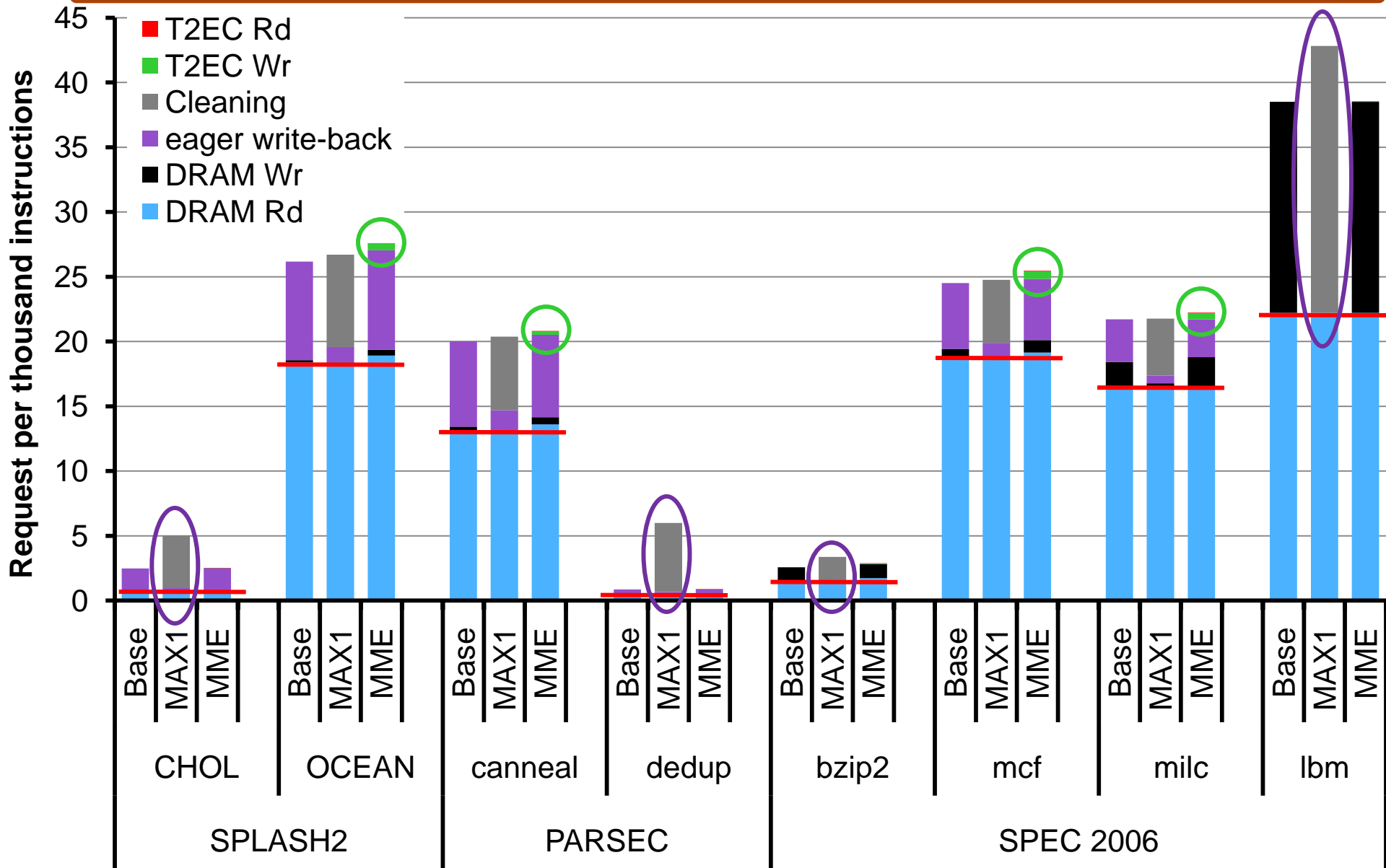
## DRAM – 2.67 GB/s



# Comparison to MAXn Scheme



# Traffic Increase – 2% on Average



# Conclusions

---

- Flexible two-tiered error protection
  - Low on-chip overhead of only T1EC
  - No dedicated on-chip storage for T2EC
- No compromise on error protection
  - T2EC is *memory mapped and cacheable*
- Reduced cost
  - 15% area saving and 8% LLC power saving
- Performance impact
  - 1.3% on average
- DRAM traffic
  - 2% increase on average