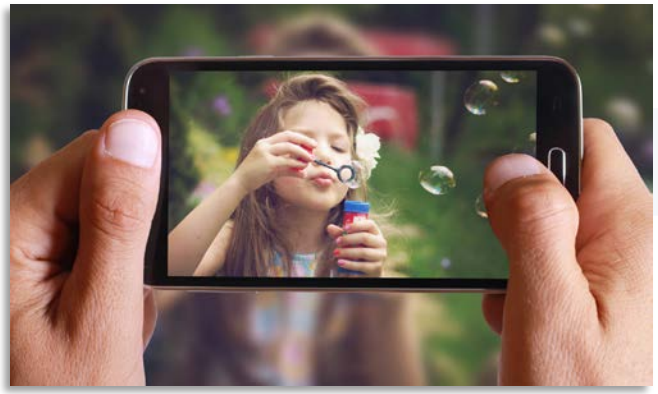
Long-Term JPEG Data Protection and Recovery for NAND Flash-Based Solid-State Storage

Yu-Chun Kuo, Ruei-Fong Chiu, and Ren-Shuo Liu

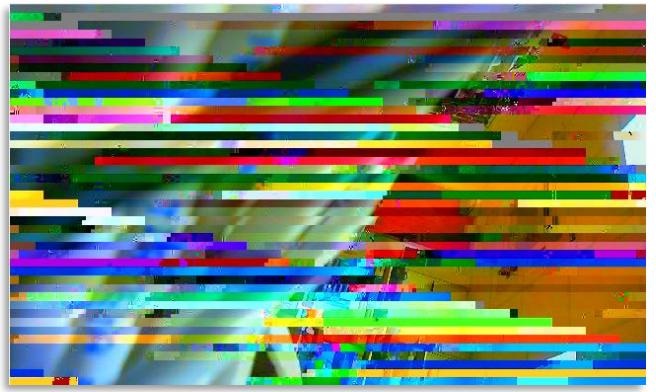
System and Storage Design Lab
Department of Electrical Engineering
National Tsing Hua University
Taiwan

Overview

- SD cards and eMMC consistute massive storage
  - Tens to hundreds of Exabytes per year
- JPEG pictures are one of the most valuable data in them
- Leaving JPEG files in SD and eMMC for a long term is risky
  - NAND flash is prone to have retention errors
  - Uncorrectable errors → corrupted pictures



A Few Years Later



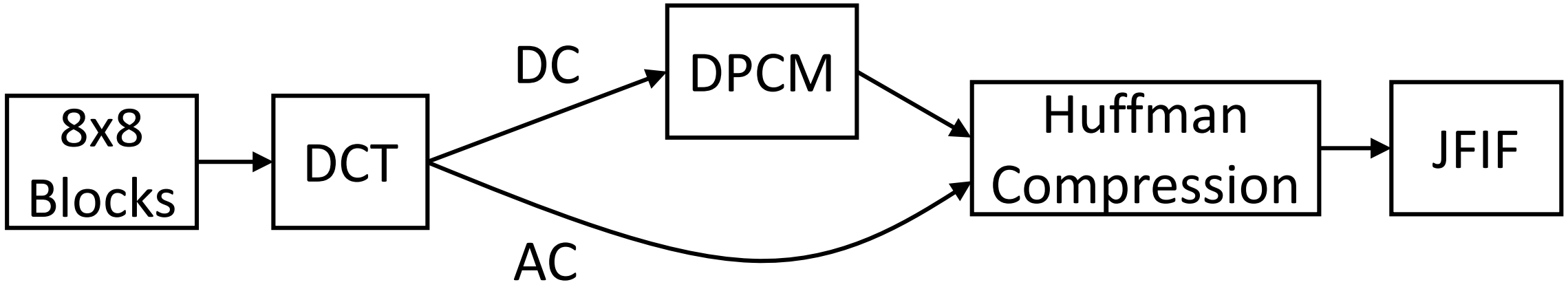
Contributions

- Increase the robustness of JPEG stored in NAND flash
  - At the cost of 9.9% storage overhead
- Rescue corrupted JPEG files
- Four techniques based on our observations
  - Strong-page header protection
  - Bit error propagation prevention
  - DC error propagation mitigation
  - Huffman-assisted error correction
- Compatible with existing JPEG viewers

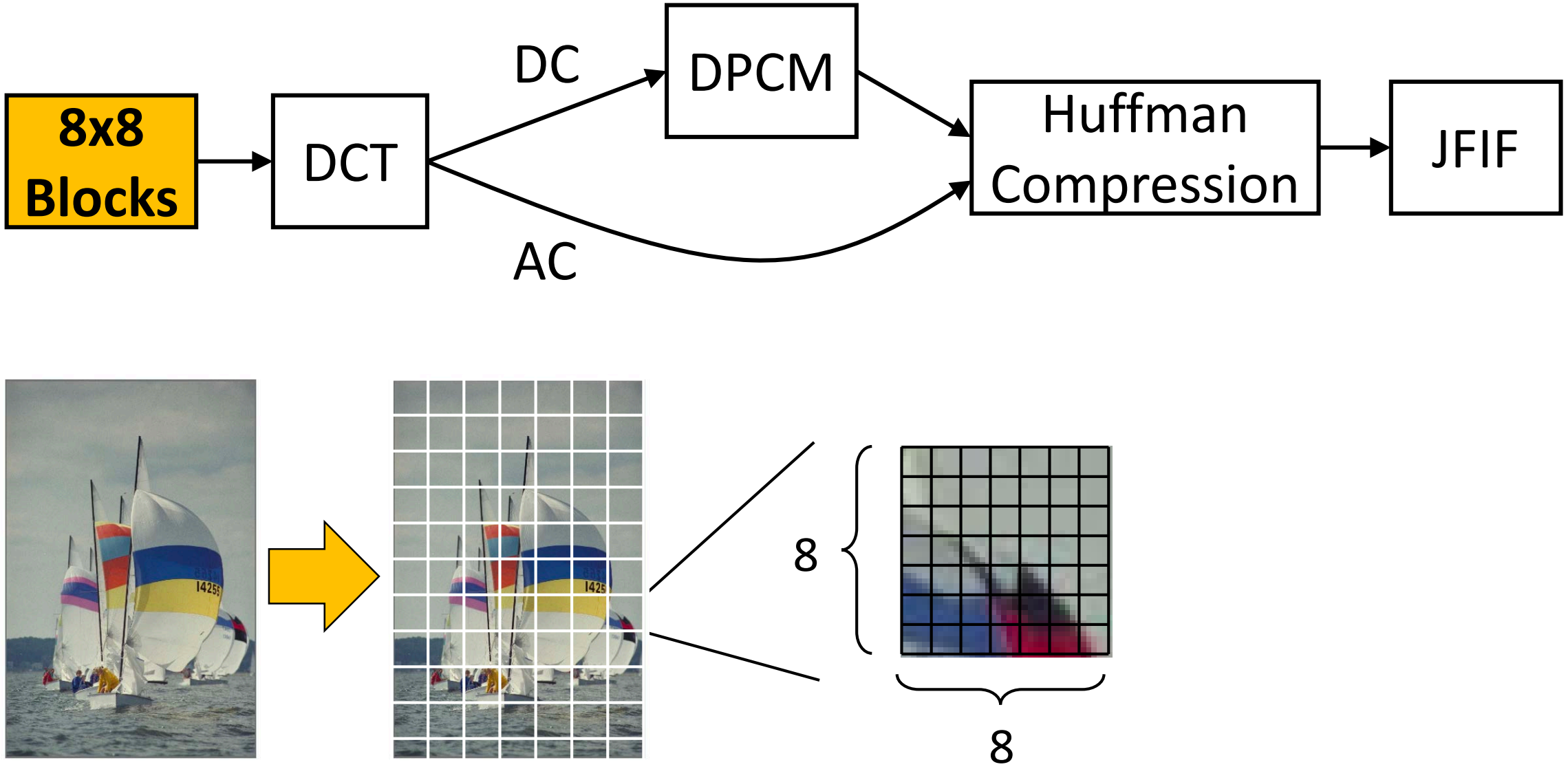
Outline

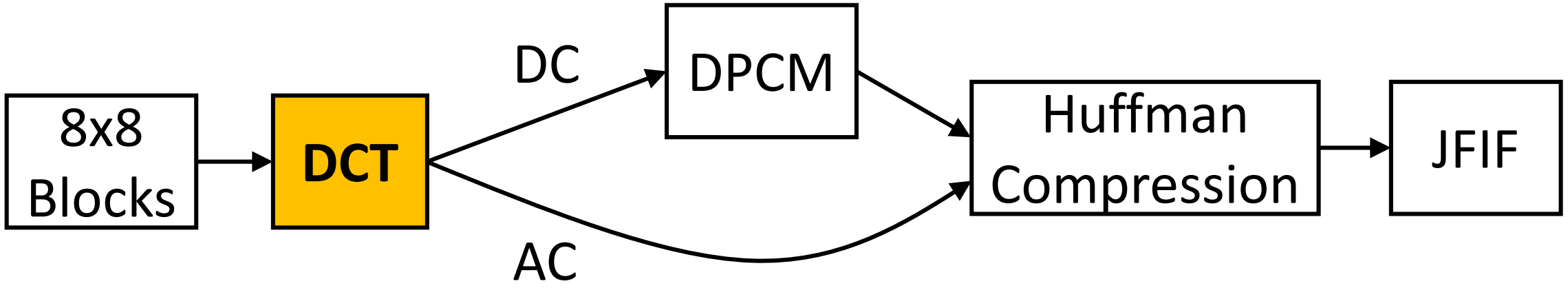
- JPEG Background
- Observations and Design
- Evaluation
- Conclusion

JPEG Encoding Steps (Simplified)

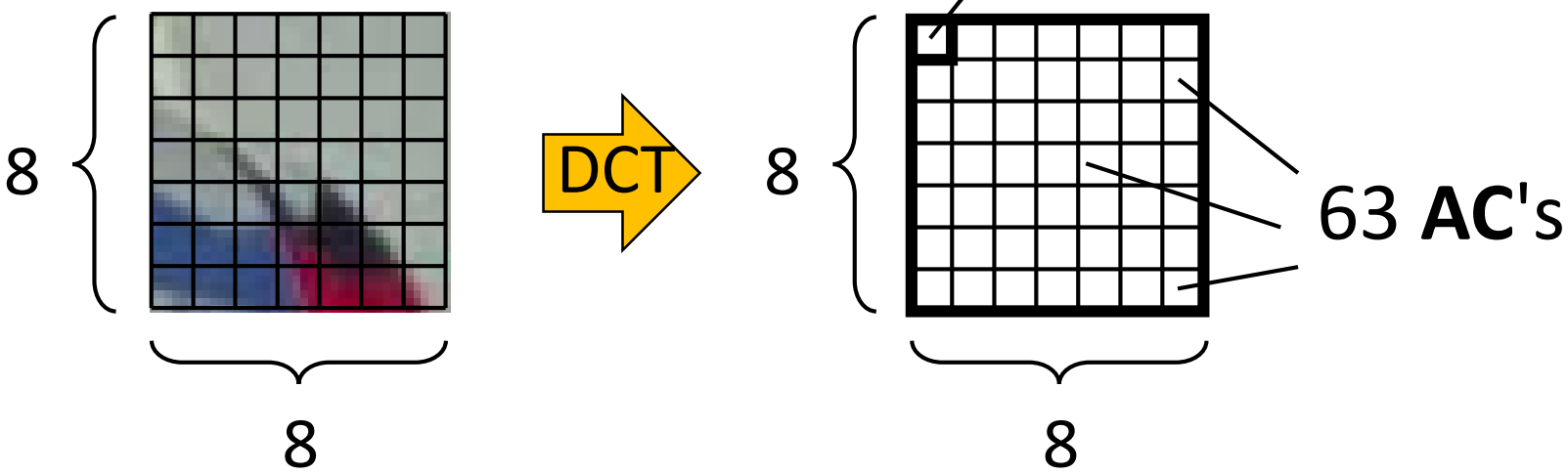


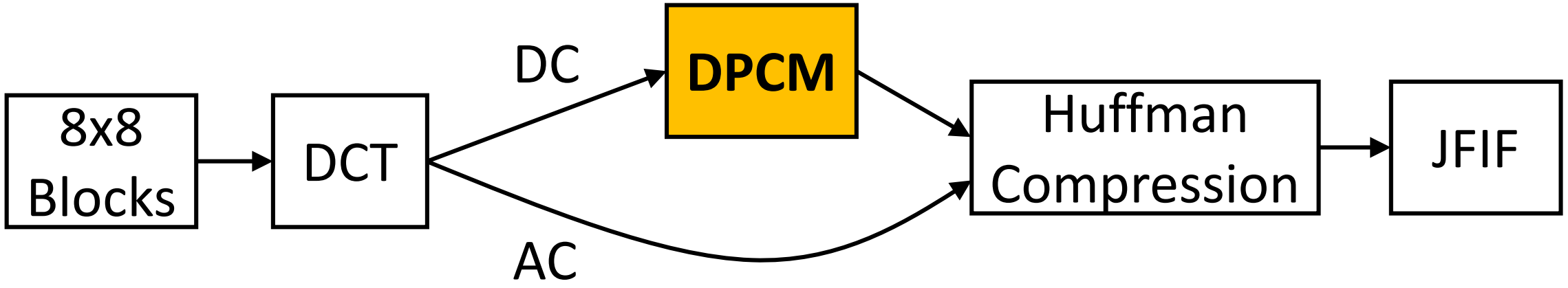
- **DCT**: Discrete Cosine Transform
- **DPCM**: Differential Pulse Code Modulation
- **JFIF**: JPEG File Interchange Format

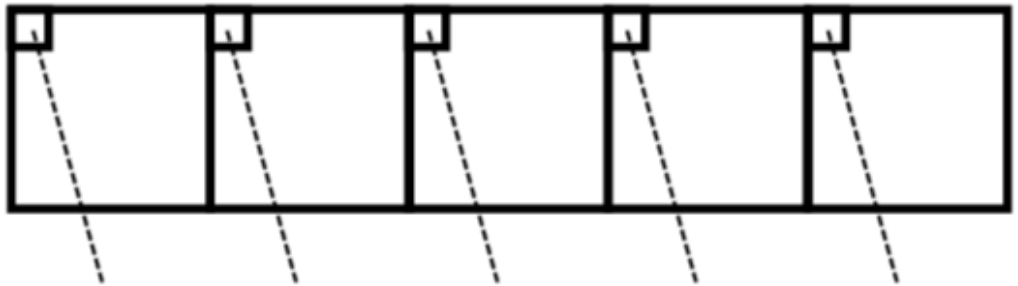

8x8 Blocks
DCT
DC
DPCM
AC
Huffman Compression
JFIF
8
8



DC (mean value of the 8x8 block)

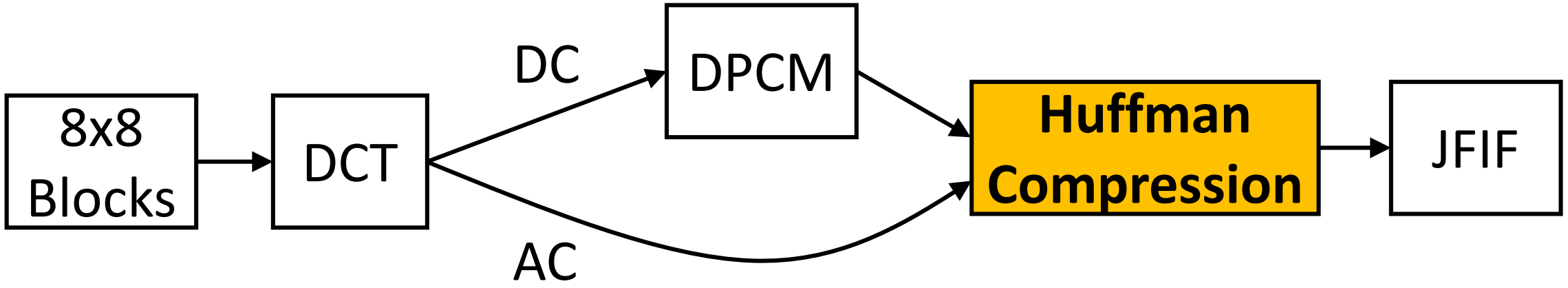


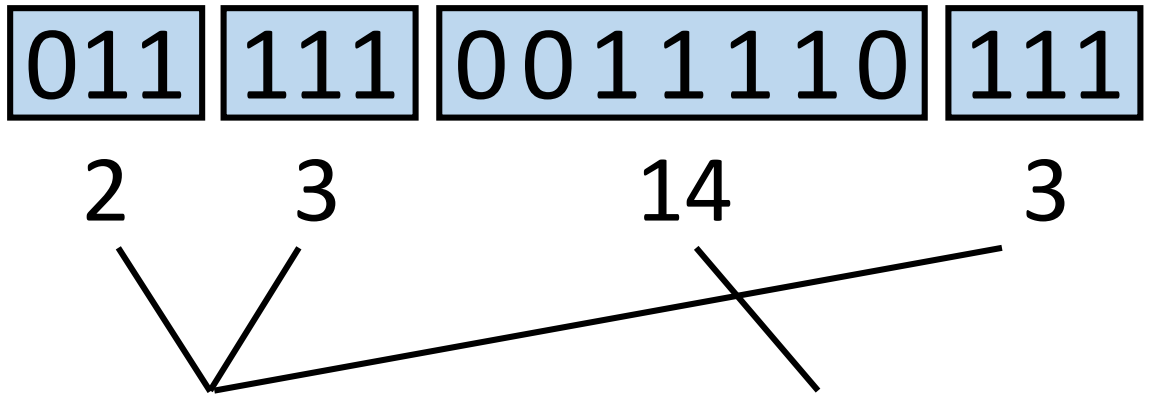




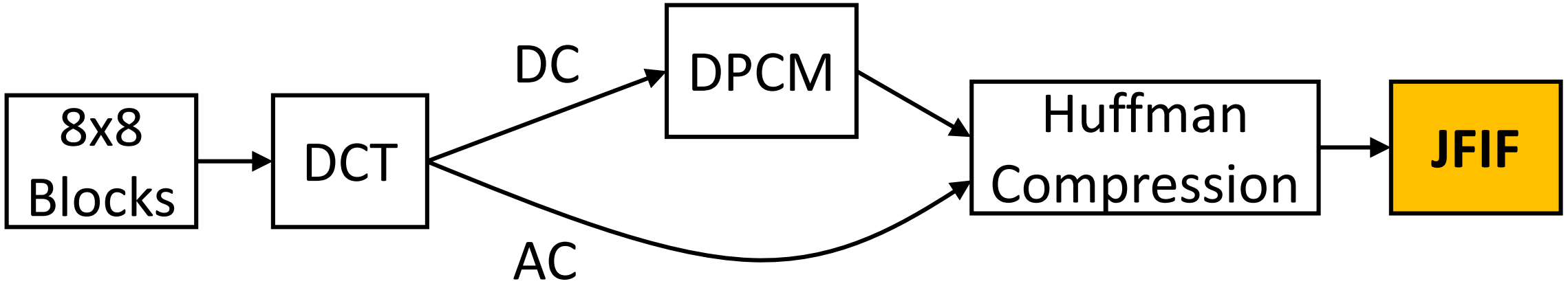
Absolute DC values: 10 15 17 19 20

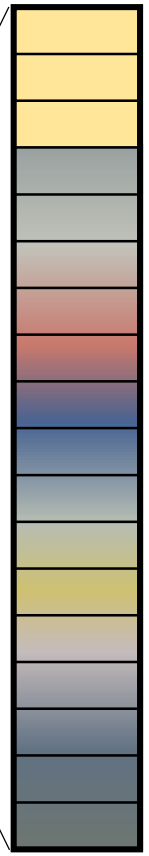
Differential DC values: 10 5 2 2 1





Popular values→ Less bits Less-popular values→ More bits



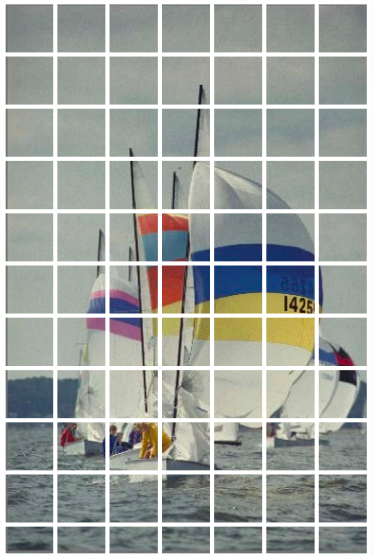


Header

- Picture width & height
- Sampling method
- Huffman tables

Body

Huffman bits
of all 8×8 blocks



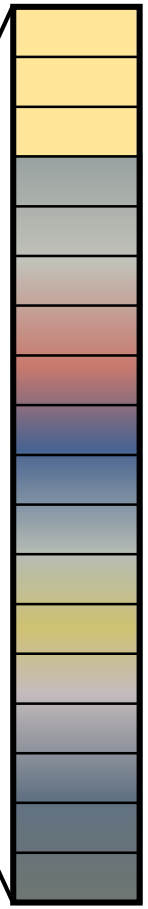
Outline

- Background
- Observations and Design
- Evaluation
- Conclusion

Observations

- Unequal criticality of JPEG file contents
- Error propagation phenomena
  - Bit error propagation
  - DC error propagation
- Skewed reliability of NAND flash

Unequal Criticality of JPEG File Contents

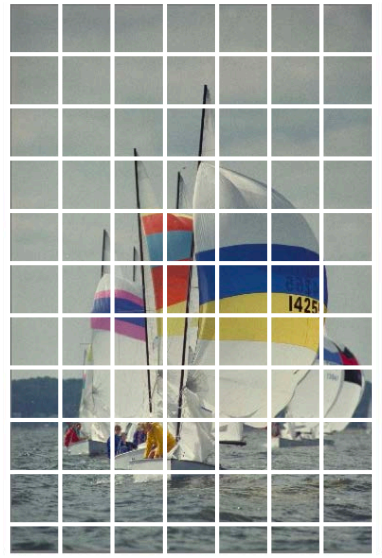


Header

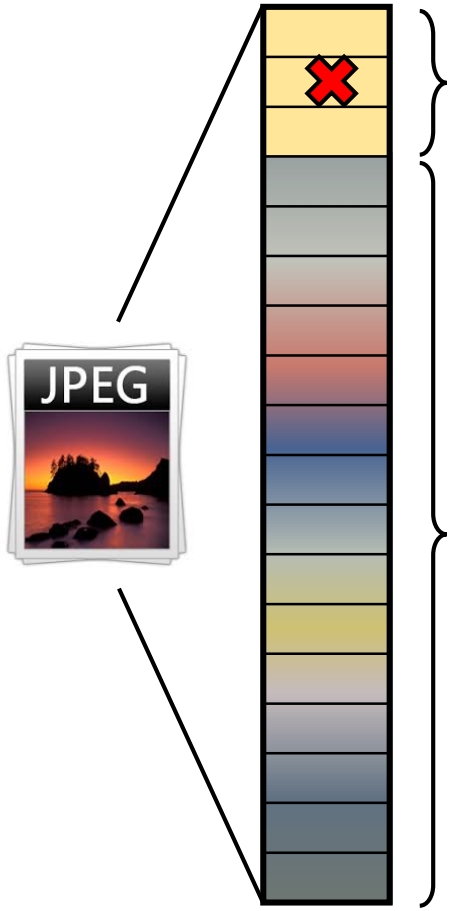
- Picture width & height
- Sampling method
- Huffman tables

Body

Huffman bits of all 8×8 blocks

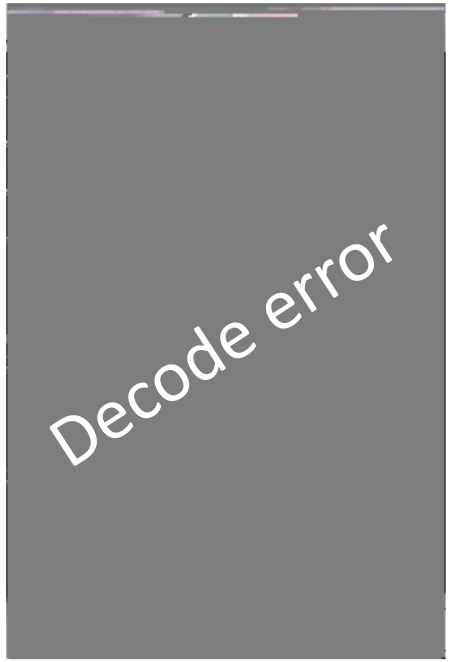


Unequal Criticality of JPEG Data

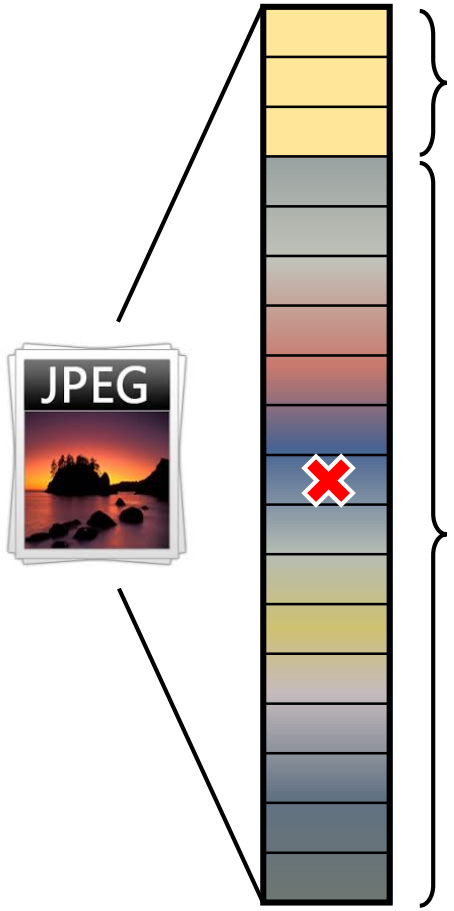


Header having a single bit error

→ very likely corrupts the entire picture



Unequal Criticality of JPEG Data

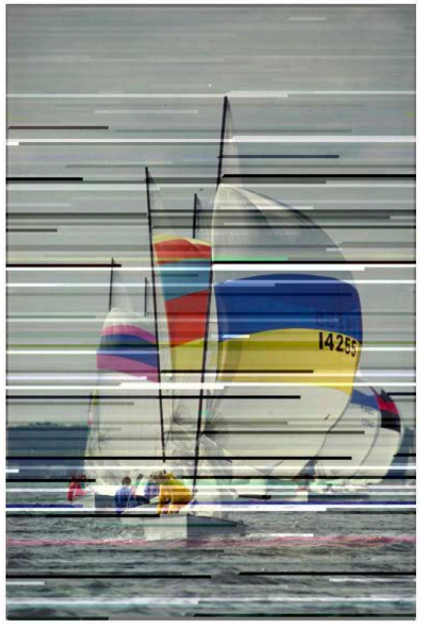


Body having a single bit error

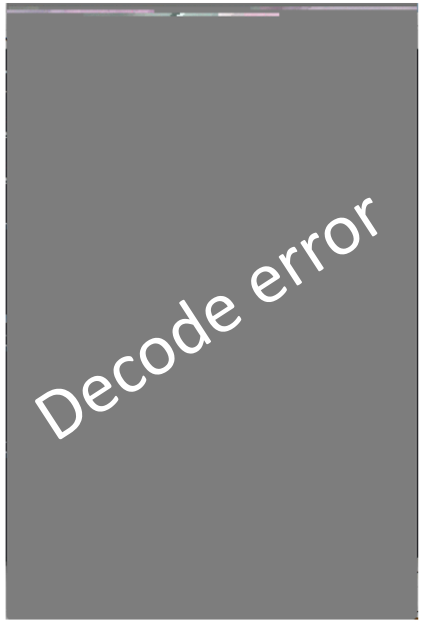
→ the results depends

- Nearly identical

- Horizontal stripes

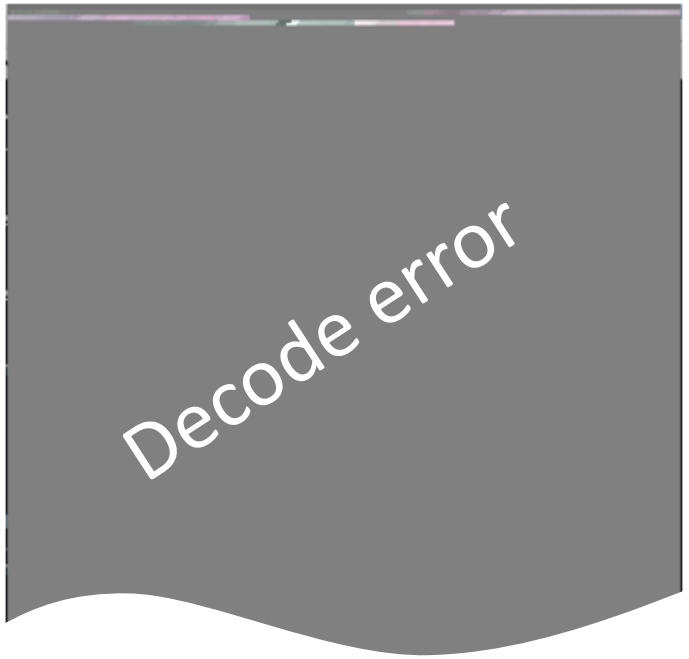


- Totally corrupted



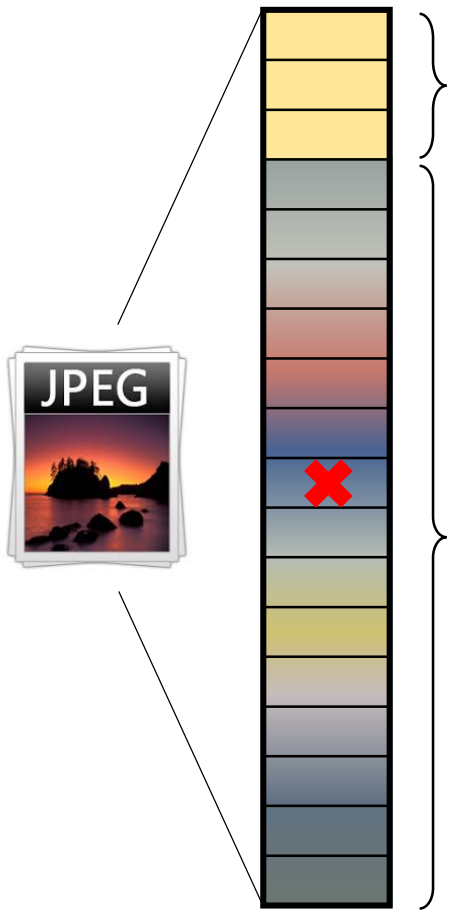
Observations

- Unequal criticality of JPEG ✓
- Error propagation phenomena
  - Bit error propagation
    → Totally corrupted



  - DC error propagation
    → Horizontal stripes

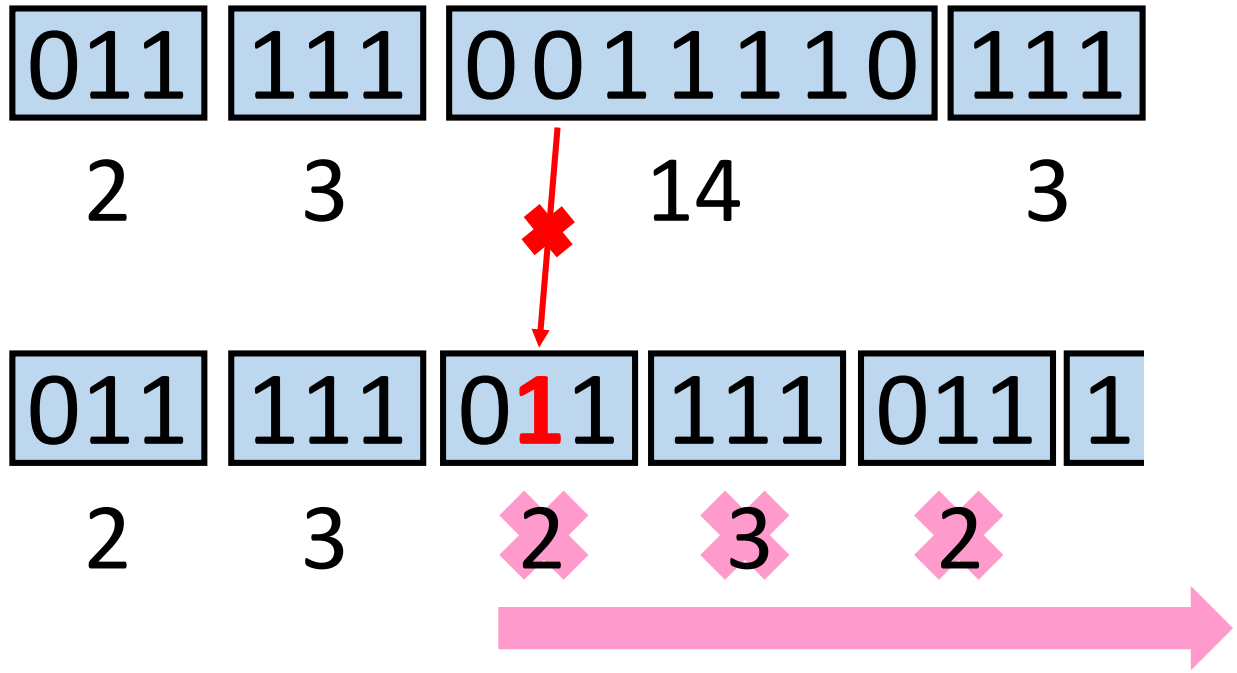
Bit Error Propagation Phenomenon



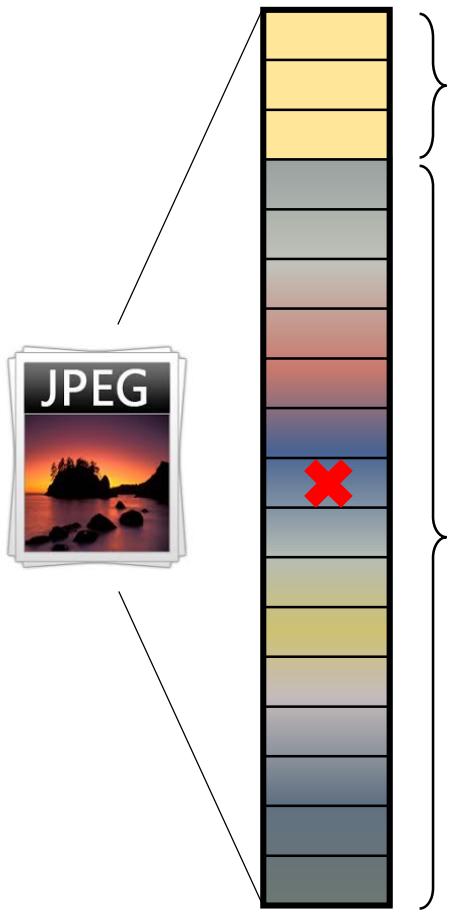
Huffman is a variable-length coding scheme

→ bit error can change code length

→ many following codes can thus be mis-decoded

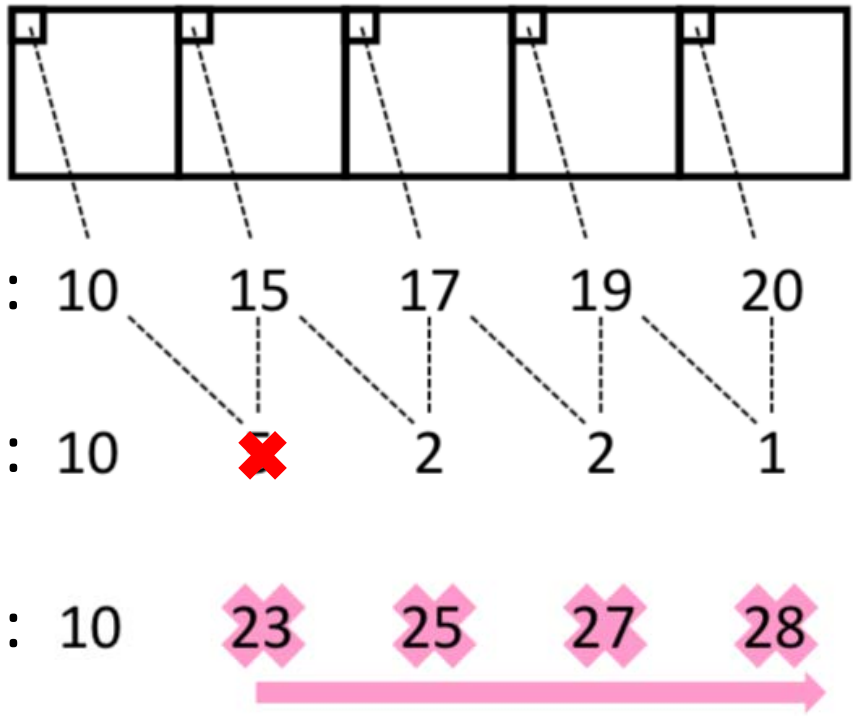


DC Error Propagation Phenomenon



JPEG stores differential DC values

→ Once a bit error interferes with one value, the following values are also mis-decoded

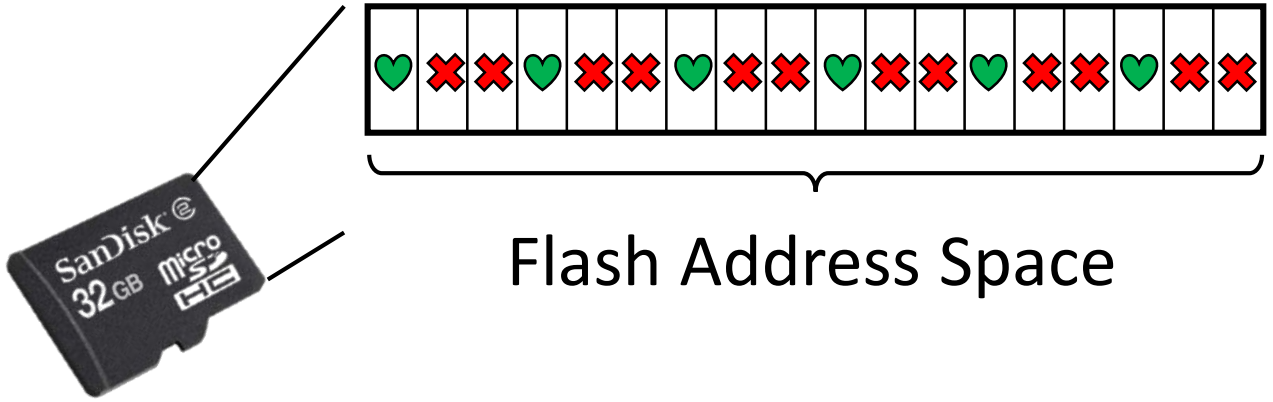


Observations

- Unequal criticality of JPEG ✓
- Error propagation phenomena
  - Bit error propagation ✓
  - DC error propagation ✓
- Skewed reliability of NAND flash

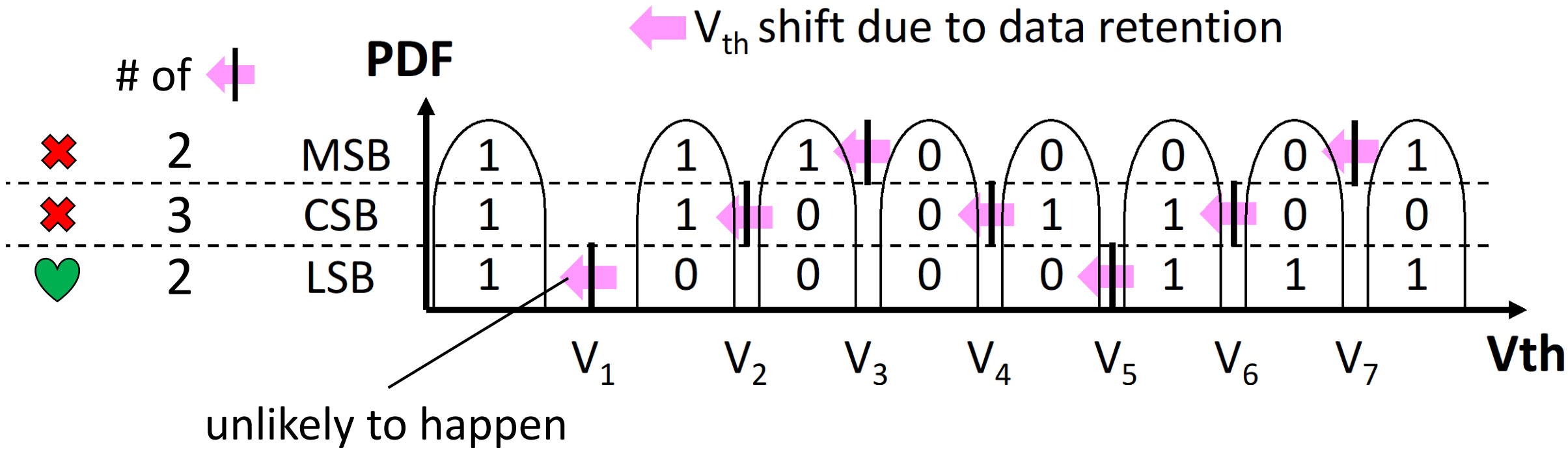
Skewed Storage Reliability

- One third of flash pages can store data much more reliably than the other pages
  - We refer to them as strong/weak pages
- This property is known to SD and eMMC vendors but is not exposed to users and applications



Skewed Storage Reliability

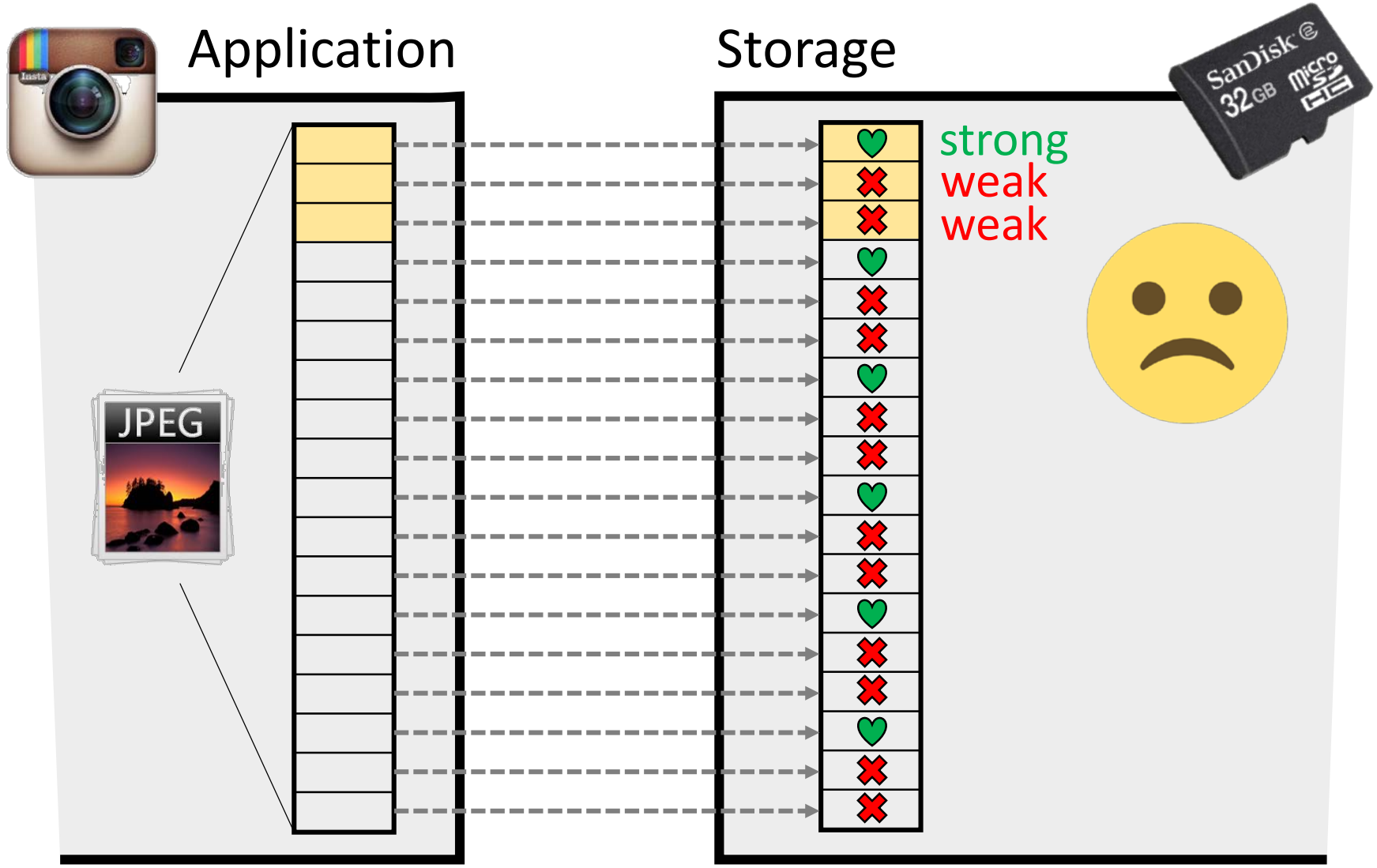
- Bits are grouped into MSB, CSB, LSB pages
- LSB pages are strong pages for the flash we tested



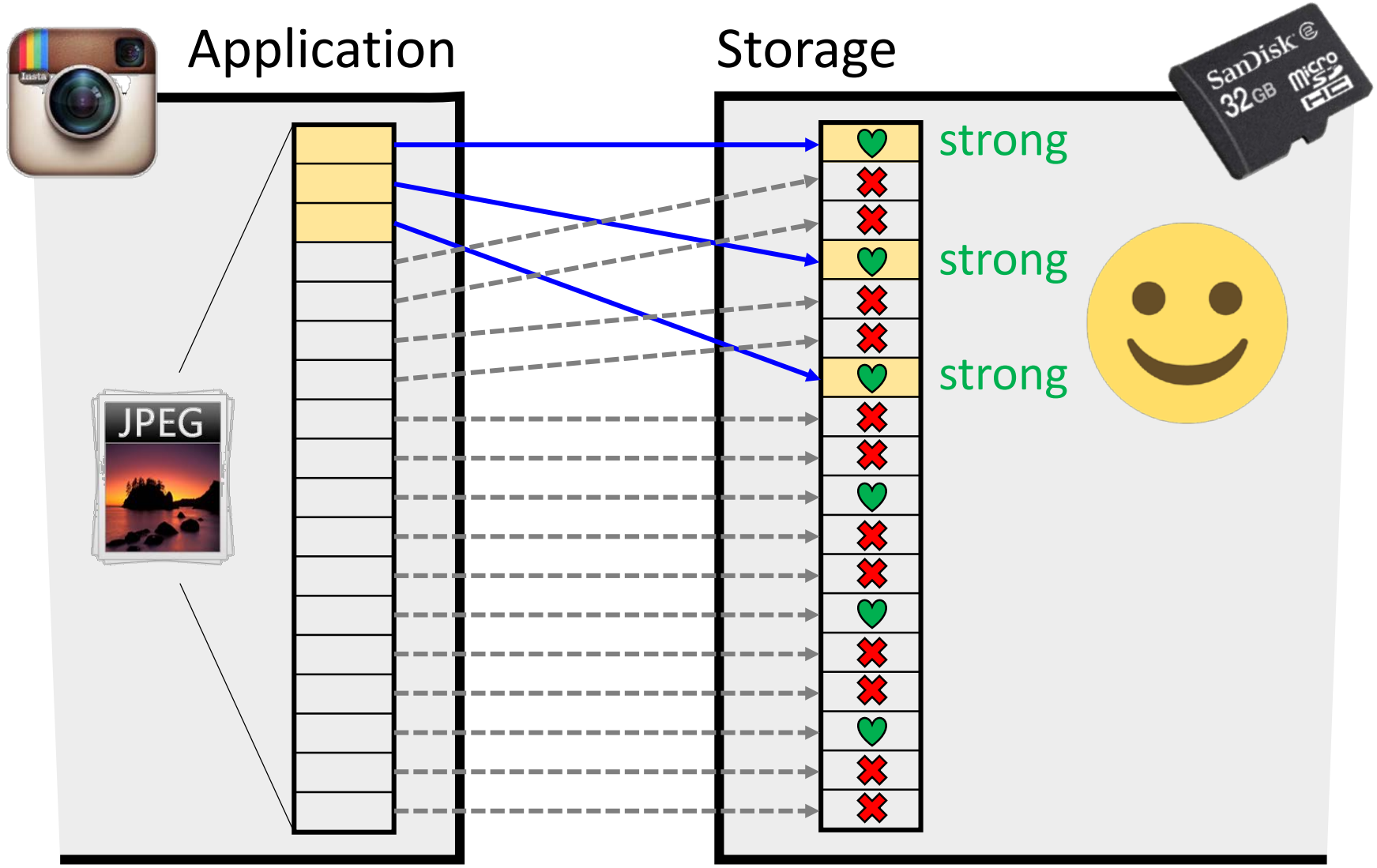
Proposed Techniques

- Strong-page header protection
- Bit error propagation prevention
- DC error propagation mitigation
- Huffman-assisted error correction

Applications Oblivious to Strong/Weak Pages

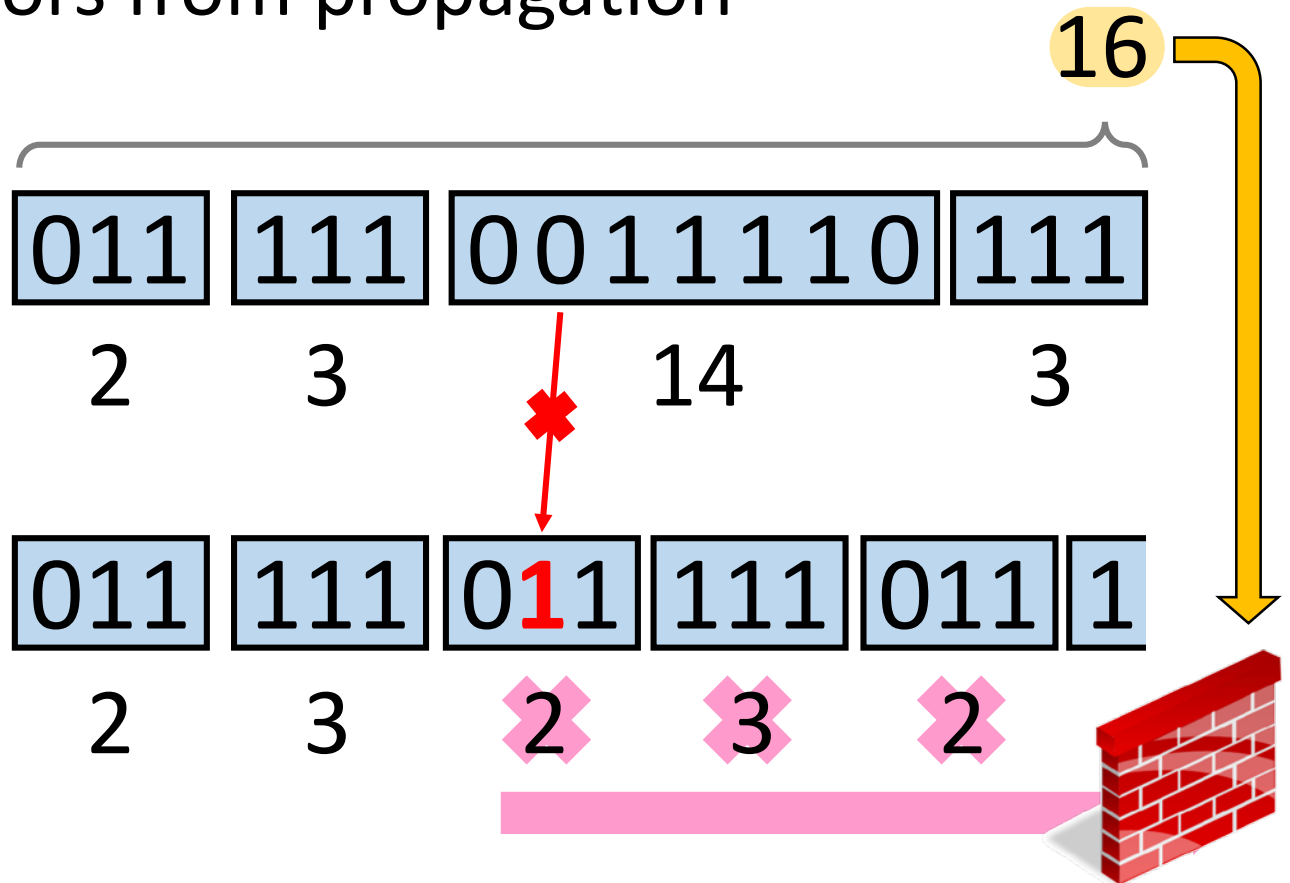


Strong-Page Header Protection



Bit Error Propagation Prevention

- We additionally store the length of each 8×8 block in JPEG header
- Stop bit errors from propagation

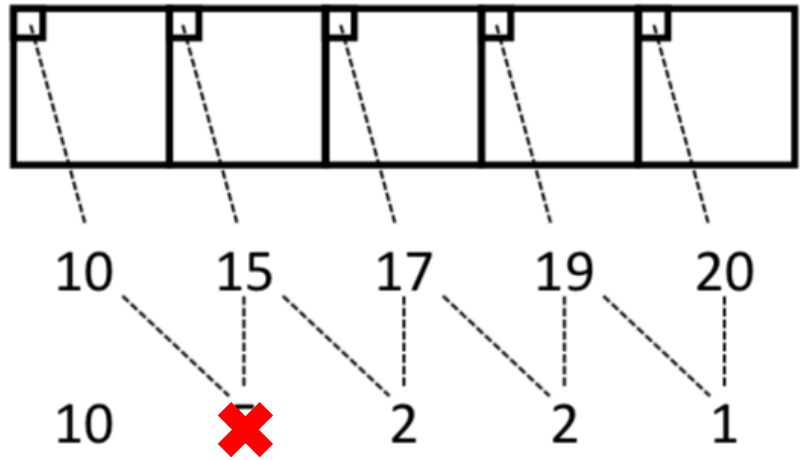


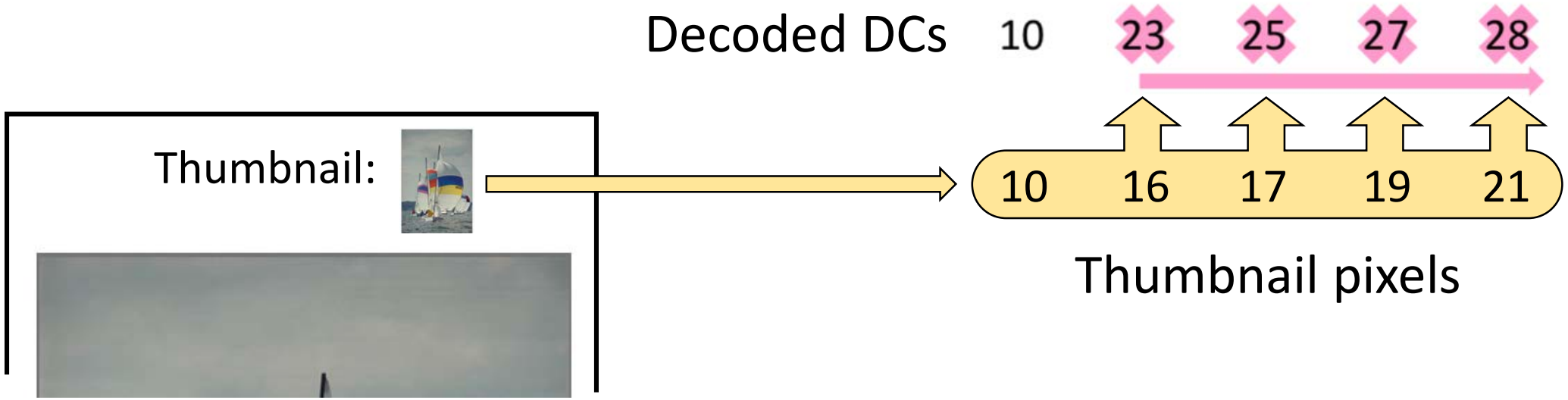
DC Error Propagation Mitigation

- Thumbnail
  - Small JPEG embedded in the header of the main JPEG
  - Facilitate image preview
- We propose to set the width and height of the thumbnail to be 1/8 of the main JPEG
  - By doing so, thumbnail pixels approximate the DC values of the main JPEG

DC Error Propagation Mitigation

- Use thumbnail pixels to calibrate decoded DCs
- Error propagation is mitigated

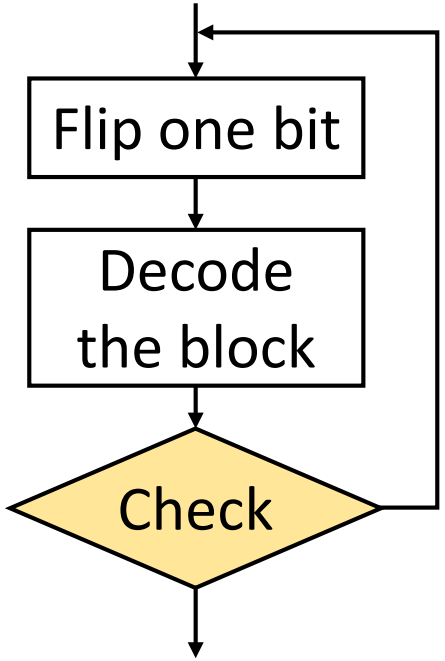




Huffman-Assisted Error Correction

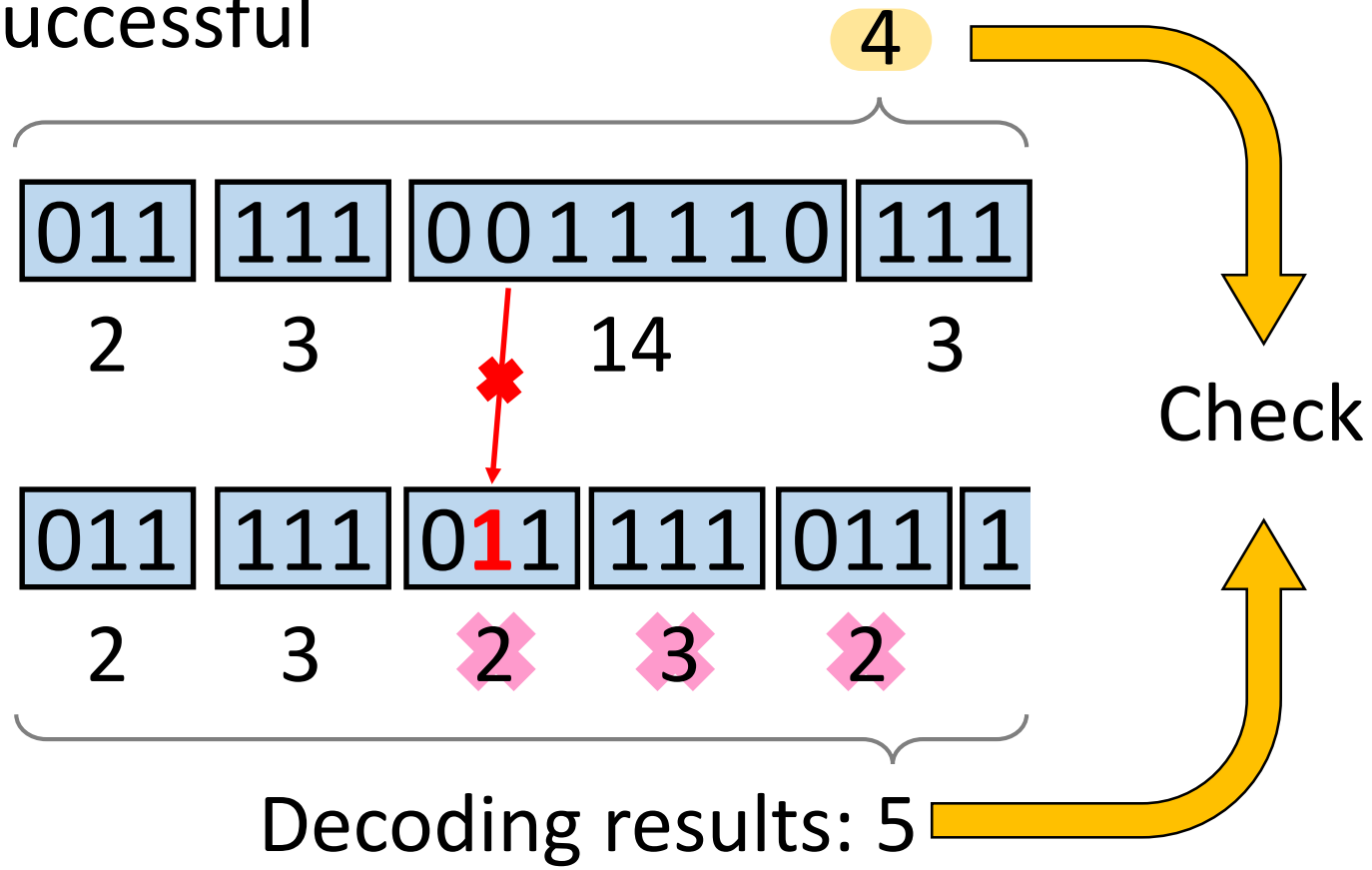
- Many 8×8 blocks contain only single bit error
  - 8×8 block is around 100 bits
  - Target bit error rate is 10^{-2}

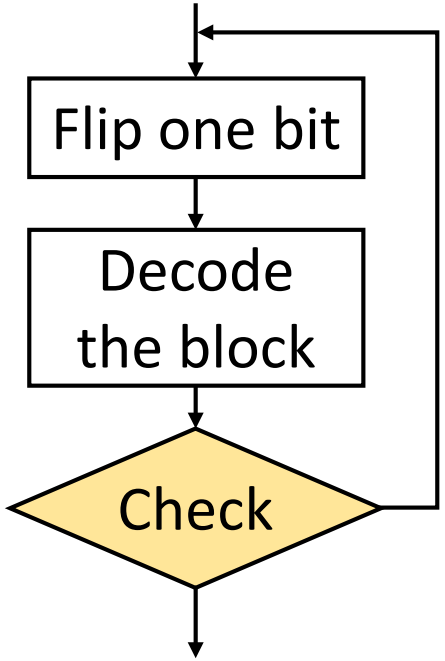
- We propose to correct single bit error per 8×8 block in a trial-and-error manner



Huffman-Assisted Error Correction

- We additionally store the number of Huffman codes of each 8×8 block in the header to check whether decoding is successful



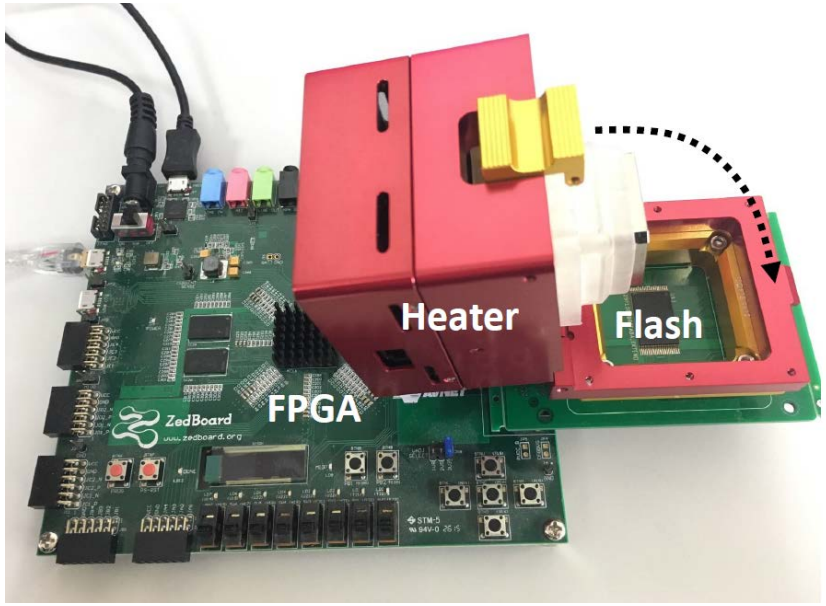


Outline

- Background
- Observations and Design
- Evaluation
- Conclusion

Setup

- Platform
  - Xilinx Zedboard FPGA
  - 16nm, 3-bit-per-cell flash chip
- 105 JPEG files
  - 100 from personal iPhone (3264×2448)
  - Five from the Kodak suite (3072×2048)
- Temperature acceleration
  - 70 hours under 85°C = 10 years under 25°C
- Assume bit error rates greater than 5×10^{-3} are uncorrectable

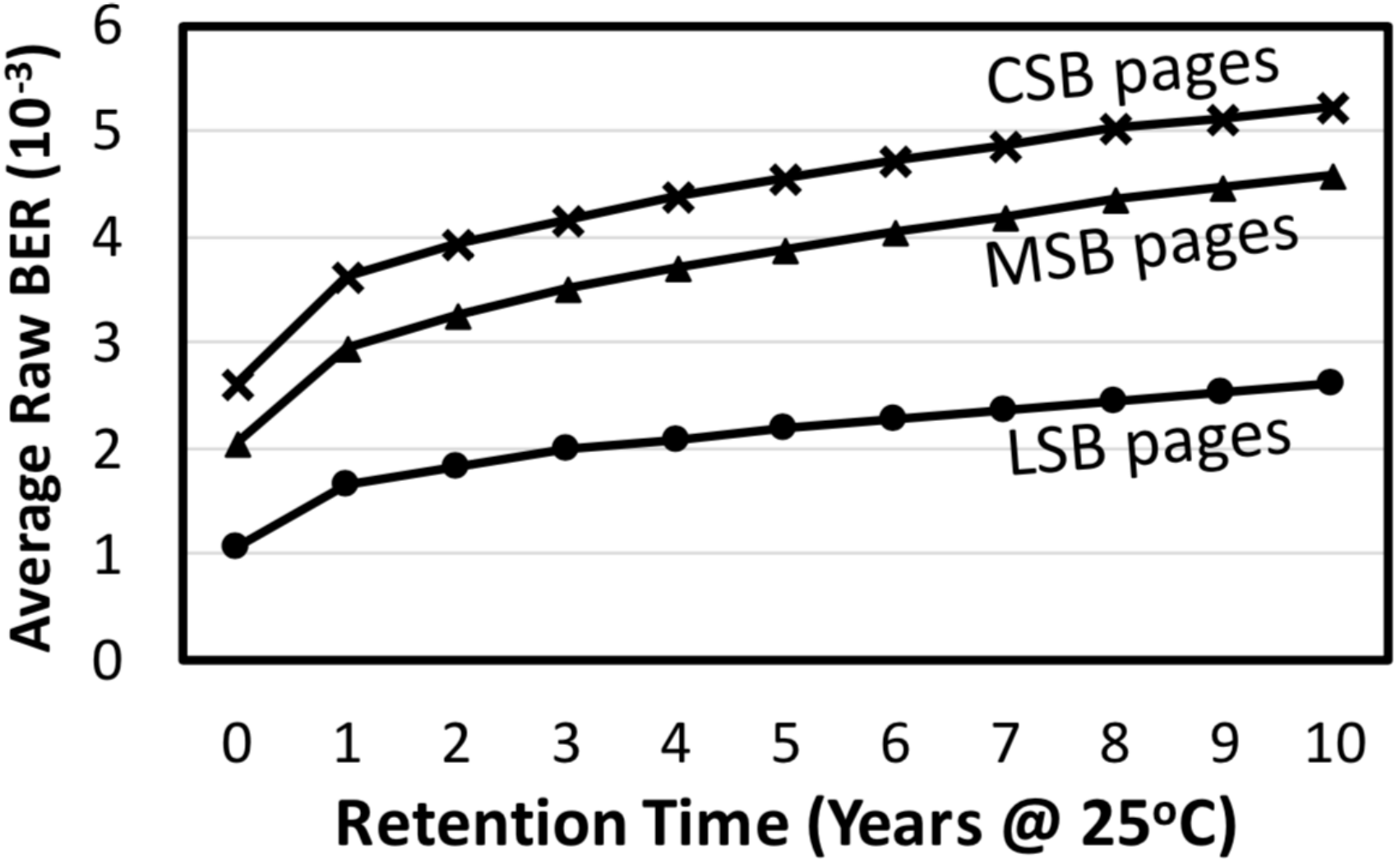


Experiments

- Flash characterization
  - Average bit error rate
  - Percentage of uncorrectable 2KB data blocks

- JPEG image quality at retention time wihin 10 years
  - PSNR (Peak Signal to Noise Ratio)
  - SSIM (Structural Similarity Index)

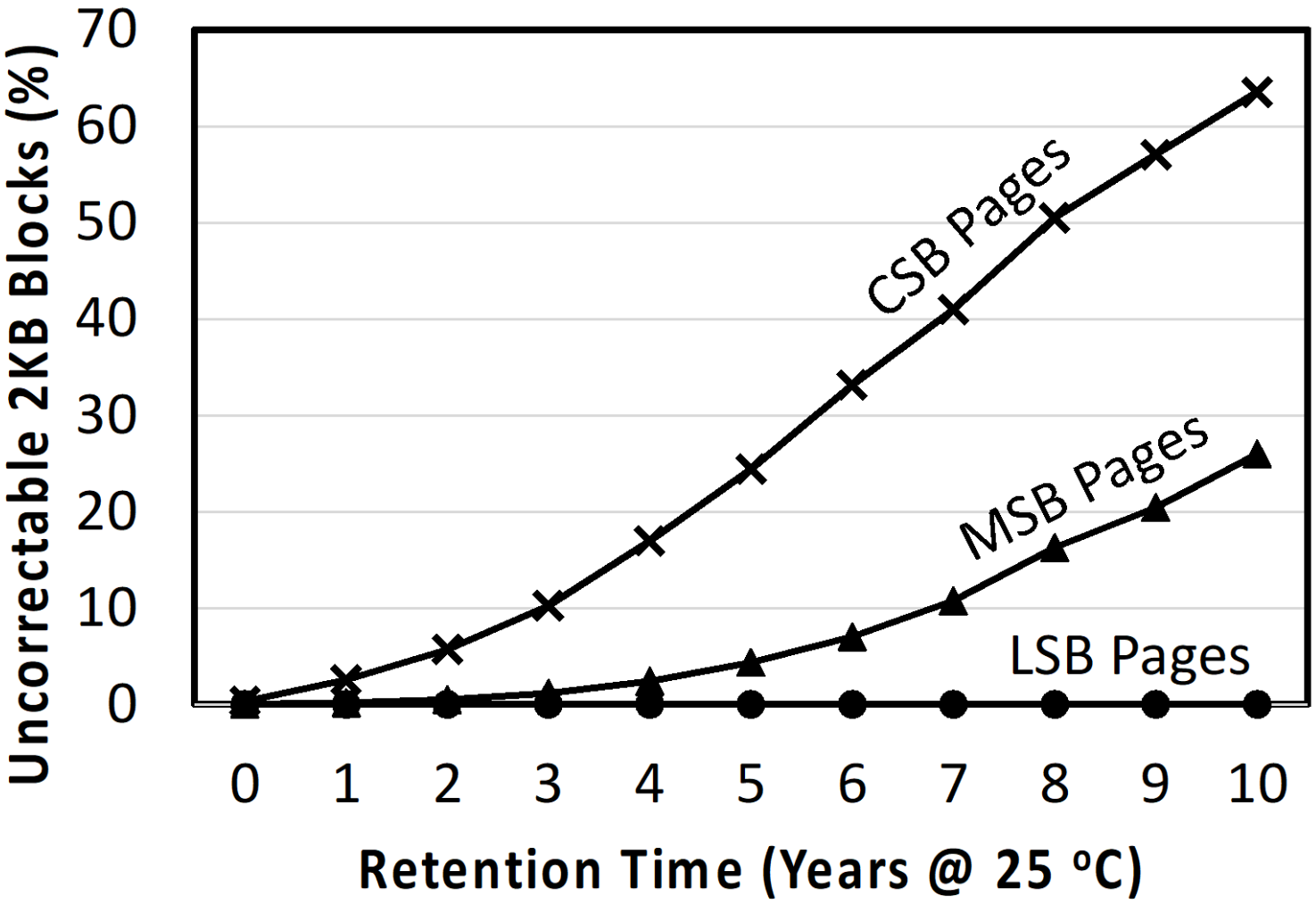
Average Raw BERs (Within 10 Years at 25 °C)



Weak pages

Strong pages

Average % of Uncorrectable 2KB Blocks

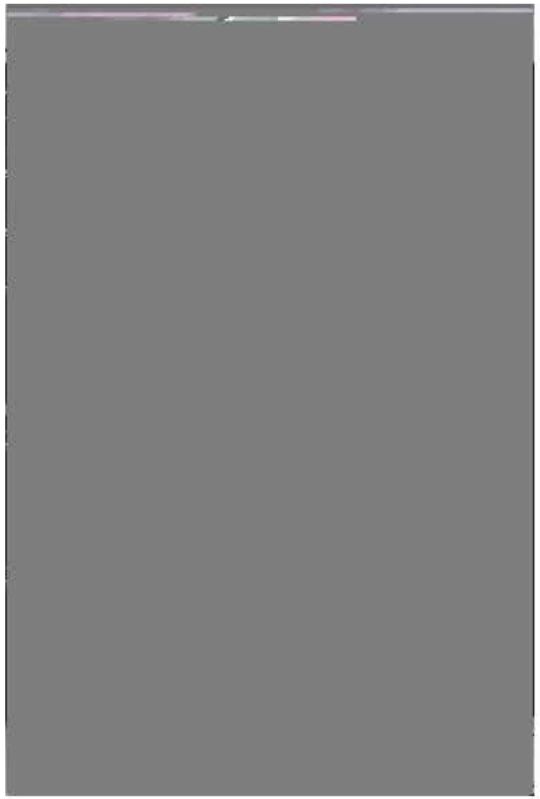


Weak pages

Strong pages

Image Quality (10 Years at 25 °C)

Ideal JPEG



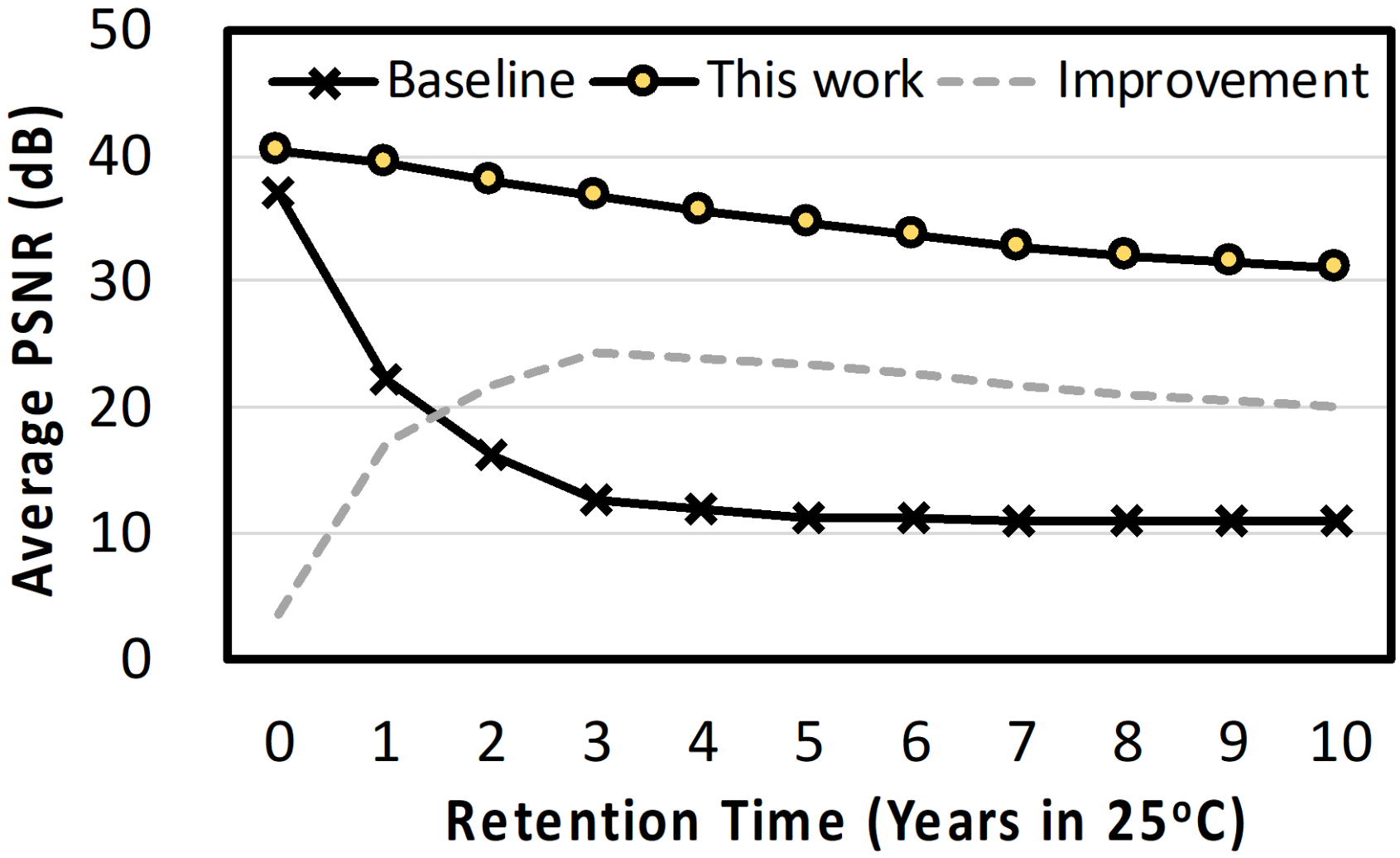
Baseline

- Strong-page header protection
- Bit error propagation prevention

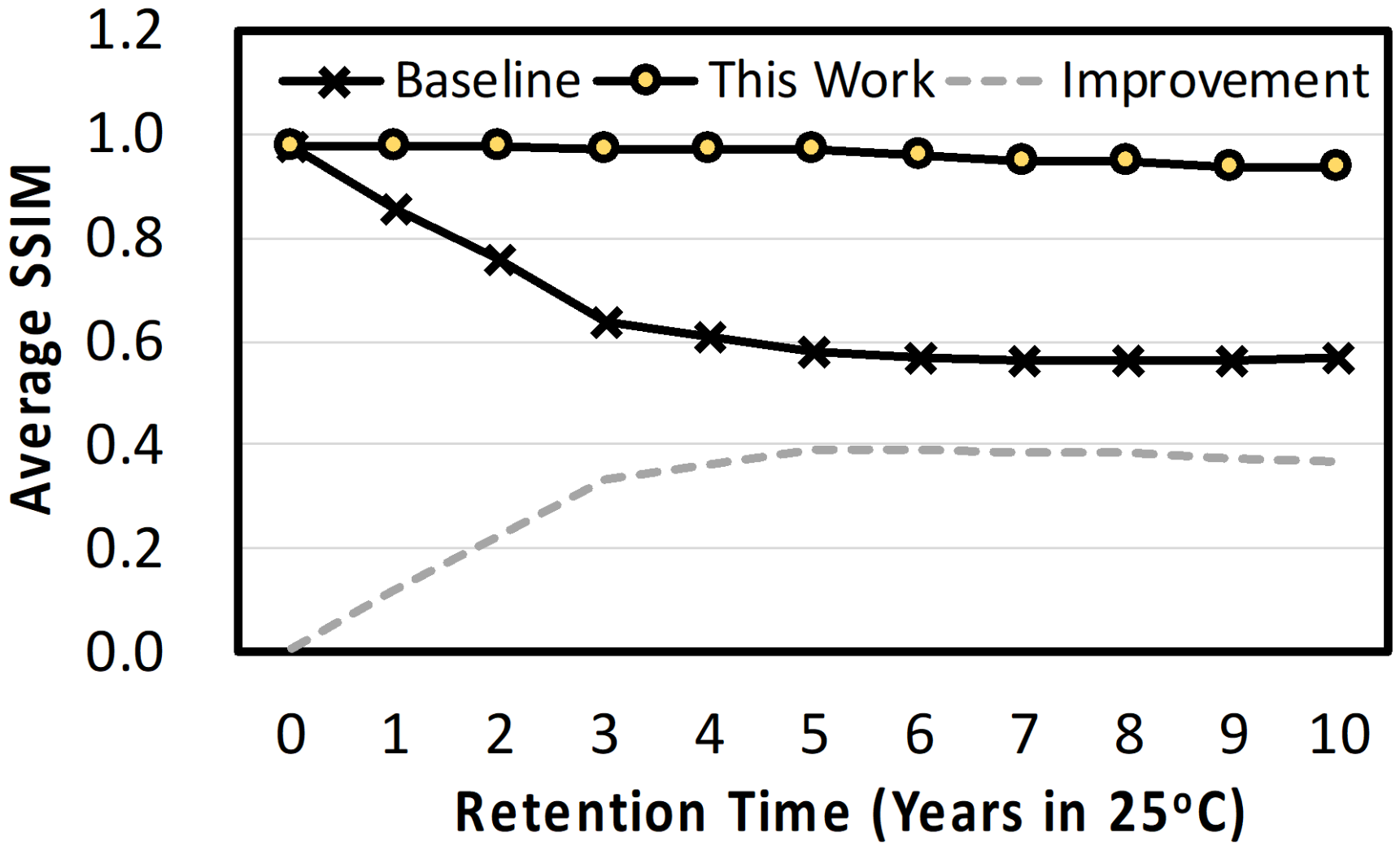
This work

All the four techniques

Average PSNR (Within 10 Years at 25 °C)



Average SSIM (Within 10 Years at 25 °C)



Concerns About Employing Extra ECC Parities

- Employing that at flash chip level
  - Cost per bit increases
  - Vendors are reluctant to do so
- Employing that at disk level
  - Disk capacity becomes non-constant
  - May be problematic to applications and operating systems
- Employing that at application level
  - Effectiveness of the extra parities is limited
  - Modern ECCs heavily rely on low-level accesses to flash memory

Conclusion

- Increasing the robustness of JPEG files and rescue corrupted JPEG files in flash-based storage
- Four techniques
  - Strong-page header protection
  - Bit error propagation prevention
  - DC error propagation mitigation
  - Huffman-assisted error correction
- Rescue corrupted JPEG files (10 years @ 25 °C)
  - Up to 24.3 dB PSNR improvement
  - At the cost of 9.9% of storage overhead
- Backward compatible with existing JPEG viewers

Long-Term JPEG Data Protection and Recovery for NAND Flash-Based Solid-State Storage

Yu-Chun Kuo, Ruei-Fong Chiu, and Ren-Shuo Liu

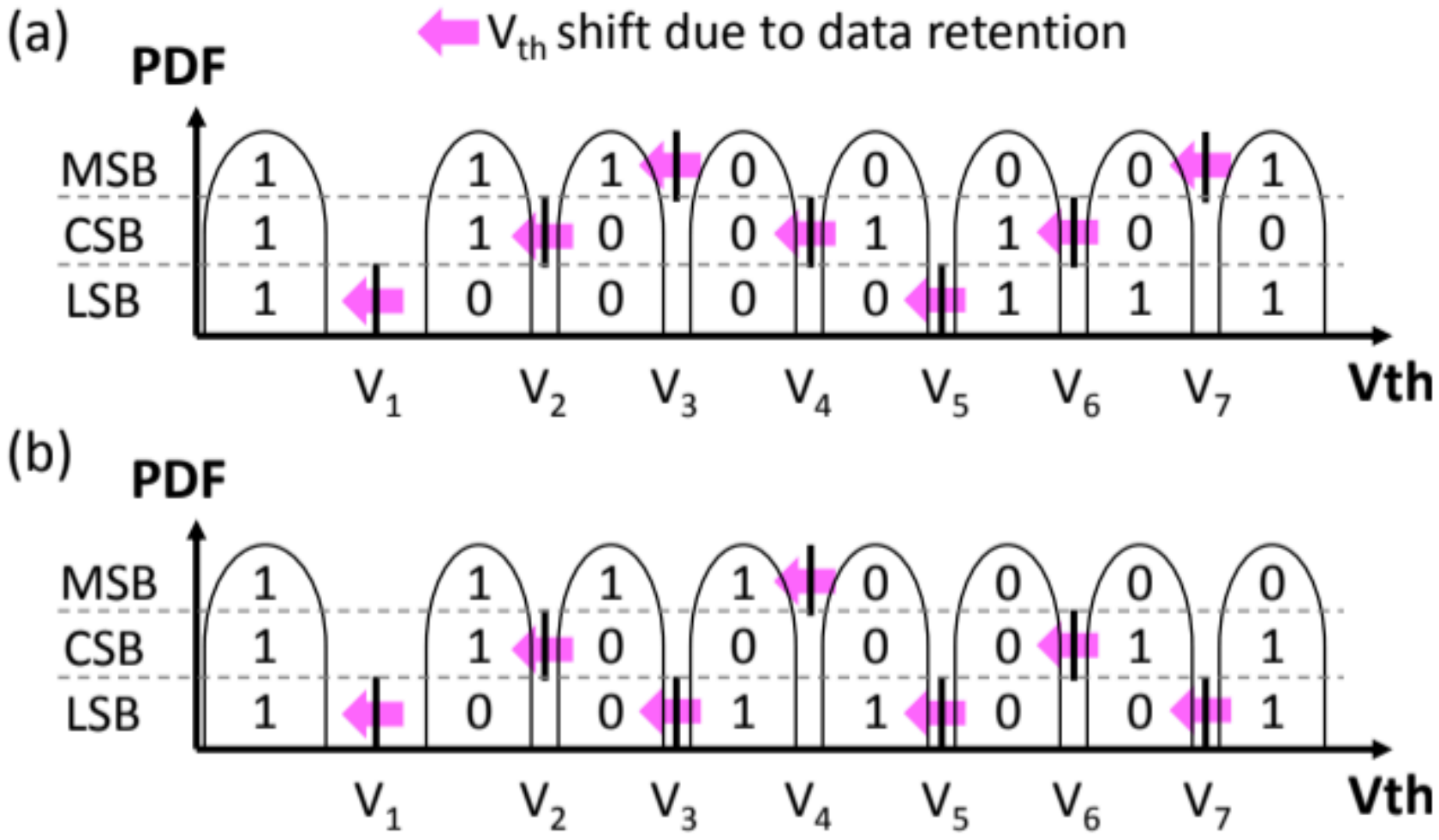
System and Storage Design Lab
Department of Electrical Engineering
National Tsing Hua University
Taiwan

NTHU

JPEG Decoding and Recover Speed

- It takes 12 seconds on average for our program to recover a corrupted (10-year) JPEG file
- Note that
  - Speed is not a top concern for rescuing corrupted JPEG files
  - It is easy to parallelize the recovery tasks of multiple corrupted JPEG files

Skewed Storage Reliability



LSB pages are strong pages

MSB pages are strong pages