

The background of the slide features a light blue and white grid pattern. Overlaid on this grid are several faint, semi-transparent circular and elliptical shapes, some of which are centered on the grid lines. The overall aesthetic is technical and modern.

# **Various page clustering techniques for performance consistency of NAND flash-based storage devices**

**Jaehyuk Cha**

# Contents

- Introduction
- Flash memory characteristics
  - FTL
  - Garbage collection
- Performance consistency
- Preemptive GC
- Page clustering
  - Offline techniques
    - Offline page clustering
    - Page clustering with data mining
  - Online techniques
    - Enhanced Hot/Cold Page Clustering
    - Sequential/Random Page Clustering
    - Page Clustering with file
- Future works

# Introduction

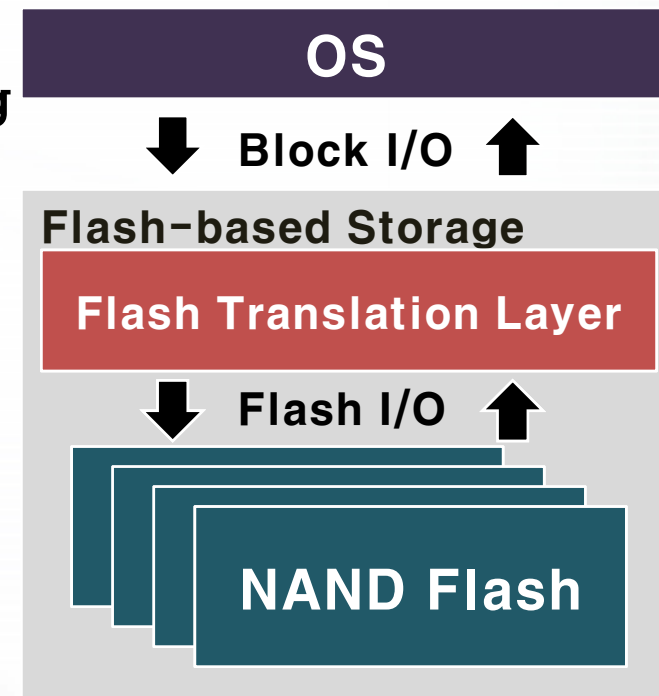
- Growing popularity of flash memory
  - Low latency
  - Low power
  - Shock resistency
- SSD widening its range of application
  - Embedded devices
  - Desktop and laptop PC
  - Server and supercomputer
- SSD expected to revolutionize storage subsystem

# Flash memory characteristics

- Erase before write
- Unit of read/write and erase differs
  - Read/Write: page (typically 4 to 16KB)
  - Erase: block (typically 64 to 256 pages)
- Latency for read, write, erase differs
  - Read(250us) < write (1.3ms) < erase (1.5ms)  
Samsung K9GBG08U0A 32Gb A-die NAND Flash
  - Erase carried out on demand: cleaning or garbage collection
- Wear-leveling necessary
  - Memory cells wears out when erased
  - Typically a block endures 1k ~ 10k erase operations

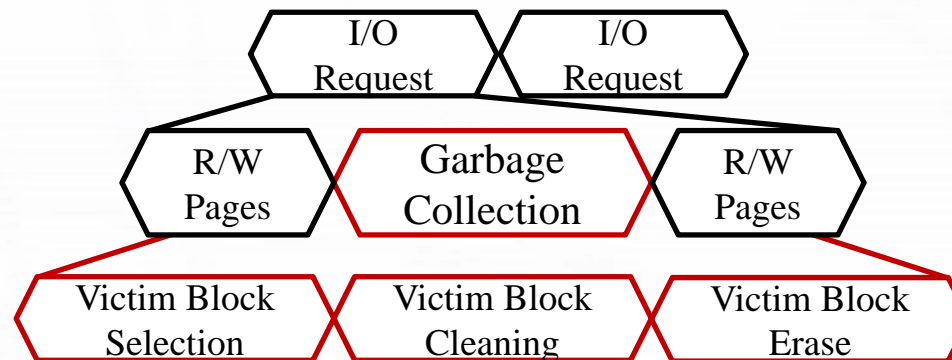
# Flash memory characteristics

- Flash translation layer (FTL)
  - Provides abstraction of flash memory characteristics
  - Maintains logical to physical address mapping
  - Carries out cleaning operations
  - Conducts wear leveling
- FTL in multiple flash chip environment
  - Manages data parallelism
  - Wear leveling



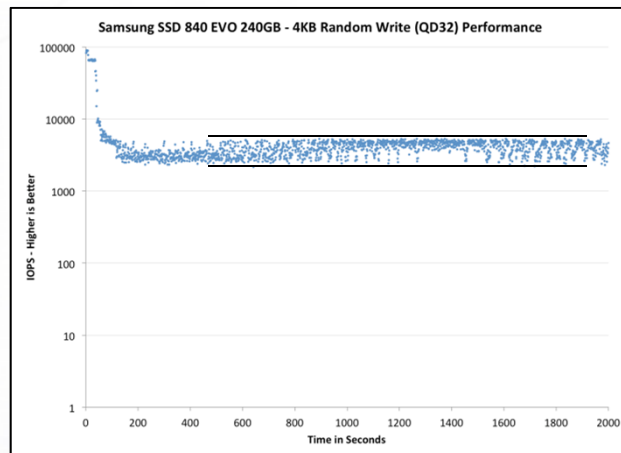
# Flash memory characteristics

- Garbage collection
  - Internal function (not user operation)
  - Consist of 3-step
    - Victim block selection
    - Victim block cleaning (=valid pages copy)
    - Victim block erase
  - Latency is increased by GC
    - Because GC is not user operation

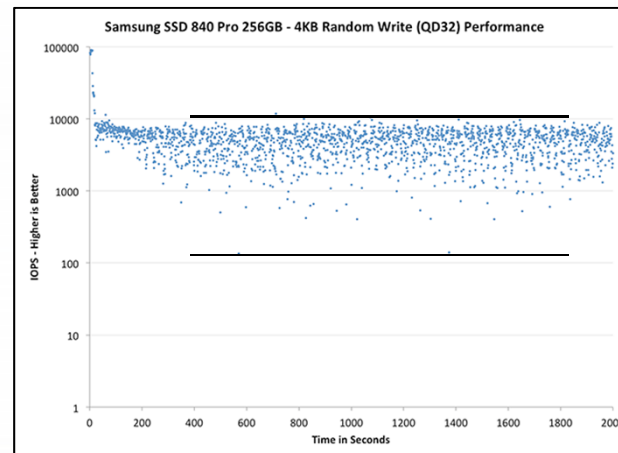


# Performance consistency

- Performance variation
  - I/O without GC vs. I/O with GC



Samsung SSD 840 EVO 250GB



Samsung SSD 840 Pro 256GB

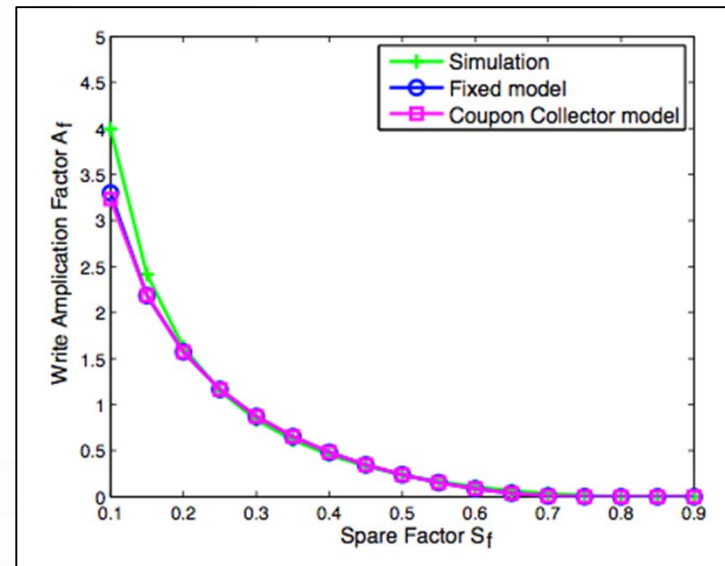
<http://www.anandtech.com/show/7337/corsair-force-ls-240gb-review/2>

- Performance consistency
  - Decrease the performance variation

# Performance consistency

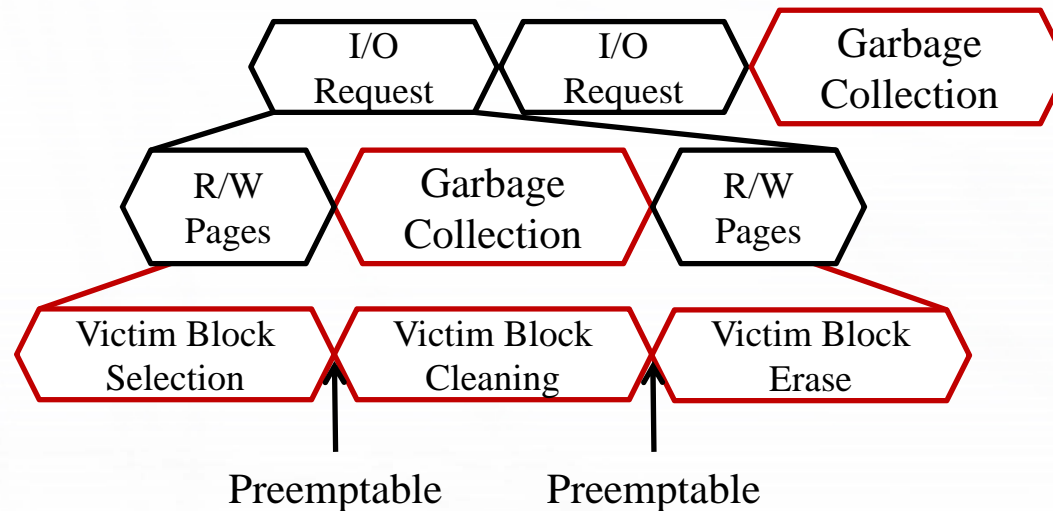
- For performance consistency
  - Preemptive garbage collection
  - Decrease the latency of garbage collection  
(= Decrease the valid pages in victim block)
    - A lot of spare area
    - Page Clustering

Write Amplification Analysis in Flash-  
Based Solid State Drives, IBM Research



# Preemptive garbage collection

- What's the “preemptive”?
  - During garbage collection, I/O arrive
    - Give I/O request High priority
    - Then, I/O request preempt processor



# Preemptive garbage collection

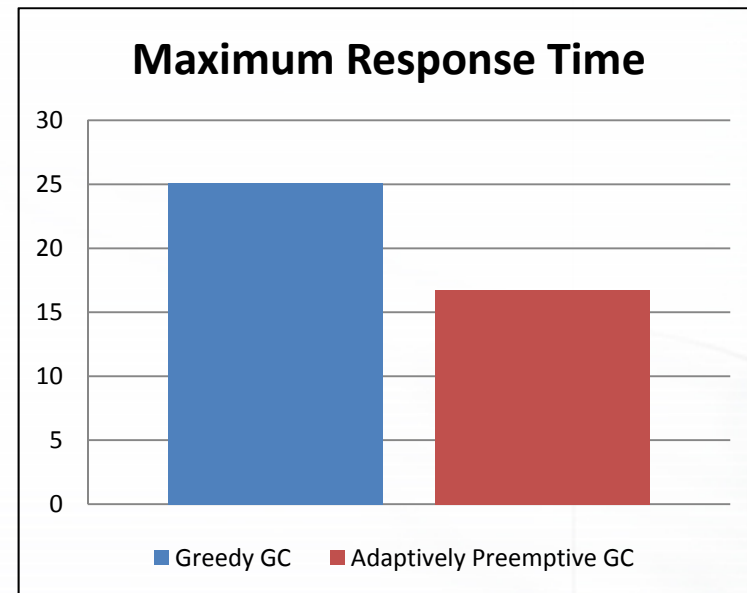
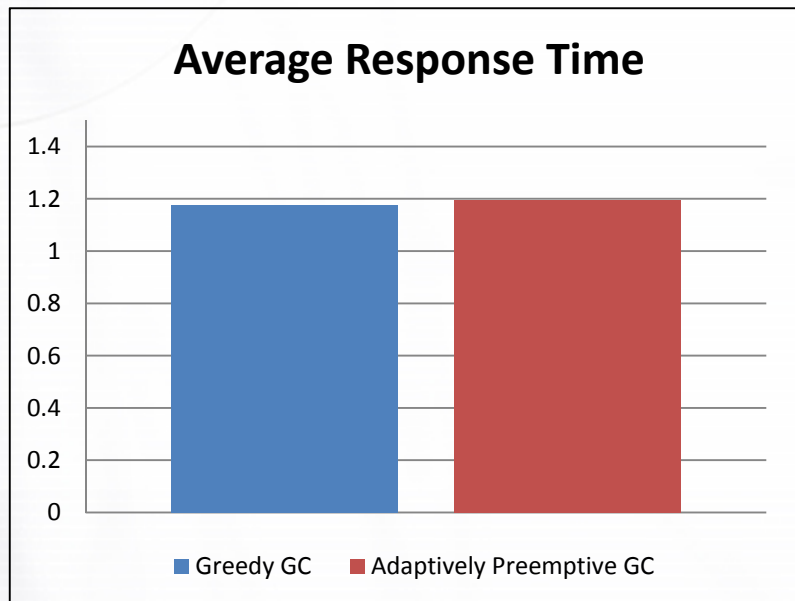
- Pros
  - Make available the IDLE time for GC
  - Might decrease the average I/O response time
- Cons
  - Programming overhead (memory access)
  - Might increase the minimum I/O response time
  - When workload is heavy, not work well
    - This is improvable through the various mechanisms

# Preemptive garbage collection

- Adaptively preemptive garbage collection
  - Ensure that some step of GC is performed obligatorily
  - Considering the remaining free pages
    - Define the status of SSD by this
- Periodically preemptive garbage collection
  - Some step of GC is performed periodically
  - Need to define the period

# Preemptive garbage collection

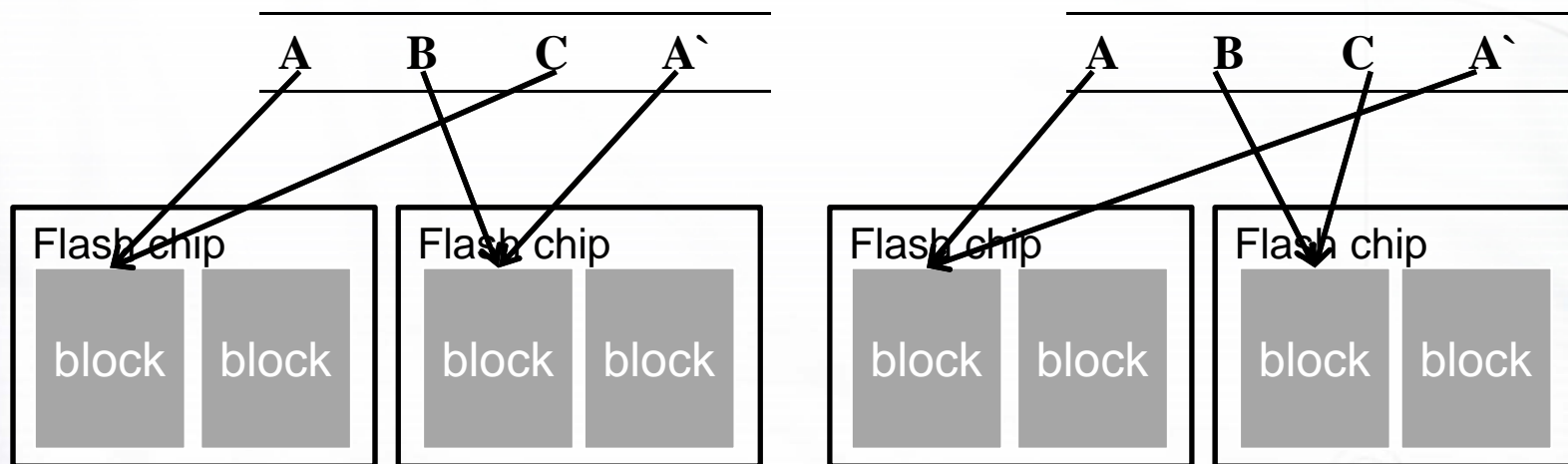
- Adaptively preemptive garbage collection



- Periodically preemptive GC will be developed

# Page Clustering

- Page allocation policy considers
  - Data parallelism
    - Channel / way / plane
  - Power consumption
  - Page clustering
  - ...



# Page Clustering

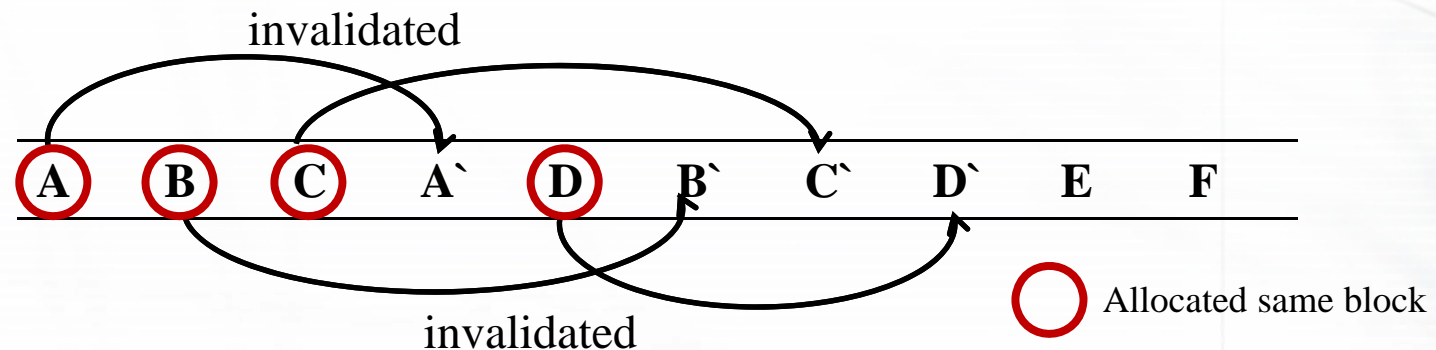
- What's Page Clustering?
  - Logical page is defined as specific status
  - Each status is allocated to same block each
- Expectation matters
  - Same status pages are called in similar time
  - Use a status for wear-leveling
  - Decrease WAF
    - This means that overhead of GC is decreased
    - In other words, Performance consistency is good

# Page Clustering

- Offline techniques
  - Offline Page Clustering
  - Page Clustering with data mining
- Online techniques
  - Enhanced Hot/Cold Page Clustering
  - Sequential/Random Page Clustering
  - Page Clustering with File

# Offline Page Clustering

- What's performance metrics?
  - Minimize overhead of GC
    - Victim block doesn't have valid pages
- Key idea
  - Invalidated pages that is closed allocate same block



# Page Clustering with Data Mining

- K-means algorithm
  - The most famous clustering(data mining)algorithm

---

    - 1: Select  $K$  points as the initial centroids.
    - 2: **repeat**
    - 3:   Form  $K$  clusters by assigning all points to the closest centroid.
    - 4:   Recompute the centroid of each cluster.
    - 5: **until** The centroids don't change

---
  - Randomly choose the initial centroids
- Experimental settings
  - Input matrix: start/end timestamp(virtual) of each accessed LPN
  - Weight: 5:1 (end:start)
  - # of clusters: # of requests in page-unit /128

# Offline Page Clustering

- Why do this?
  - Find optimal allocation
  - without valid pages copy
- Example of use
  - TPC-B on PostgreSQL 8.x
  - Proposed Hot/Cold Page Clustering  
(1-bit Hot/Cold Page Clustering)
    - Accordance rate is 73% with offline clustering

# Enhanced Hot/Cold Page Clustering

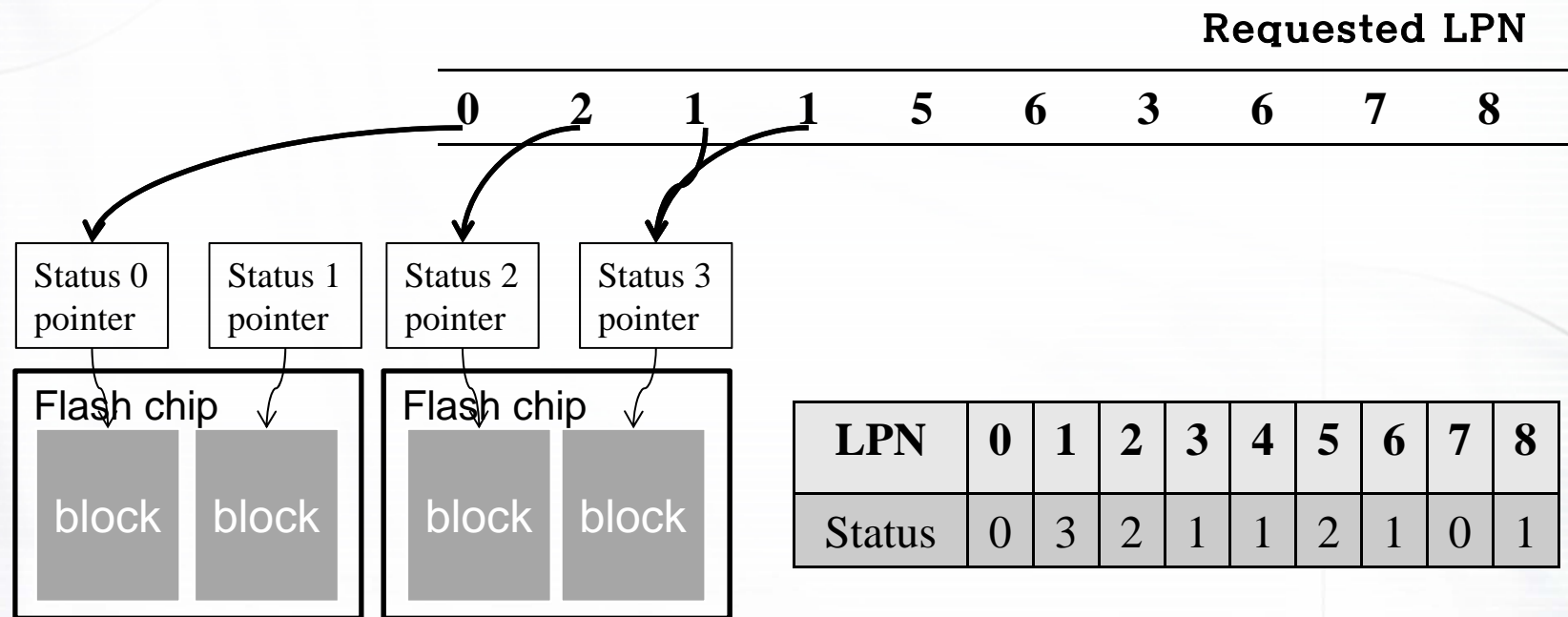
- Page clustering using frequency of pages
  - High frequency, HOT data
  - Low frequency, COLD data
  - High status value means high frequency
  - Need to define window for frequency



- Expectation matters
  - HOT data will be requested frequently
  - COLD data will be requested rarely

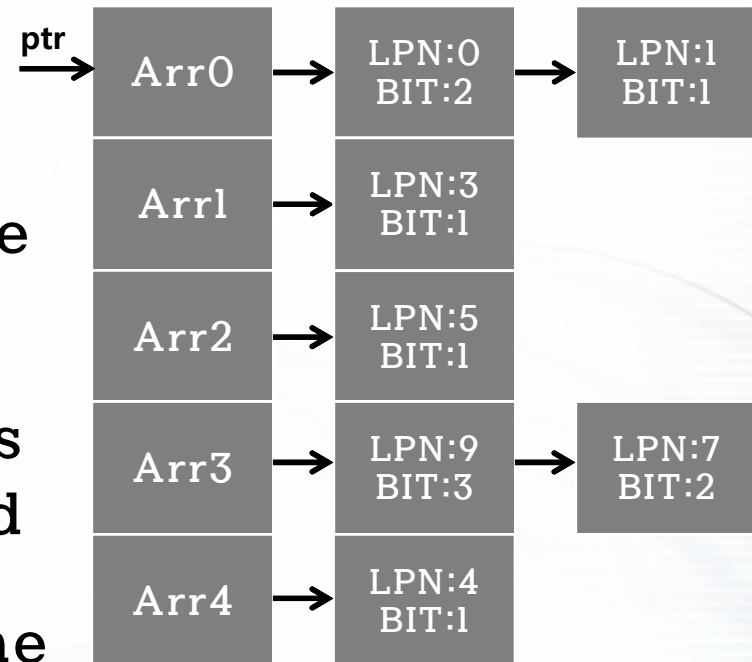
# Enhanced Hot/Cold Page Clustering

- Example, 2-bit(4 status) clustering



# Enhanced Hot/Cold Page Clustering

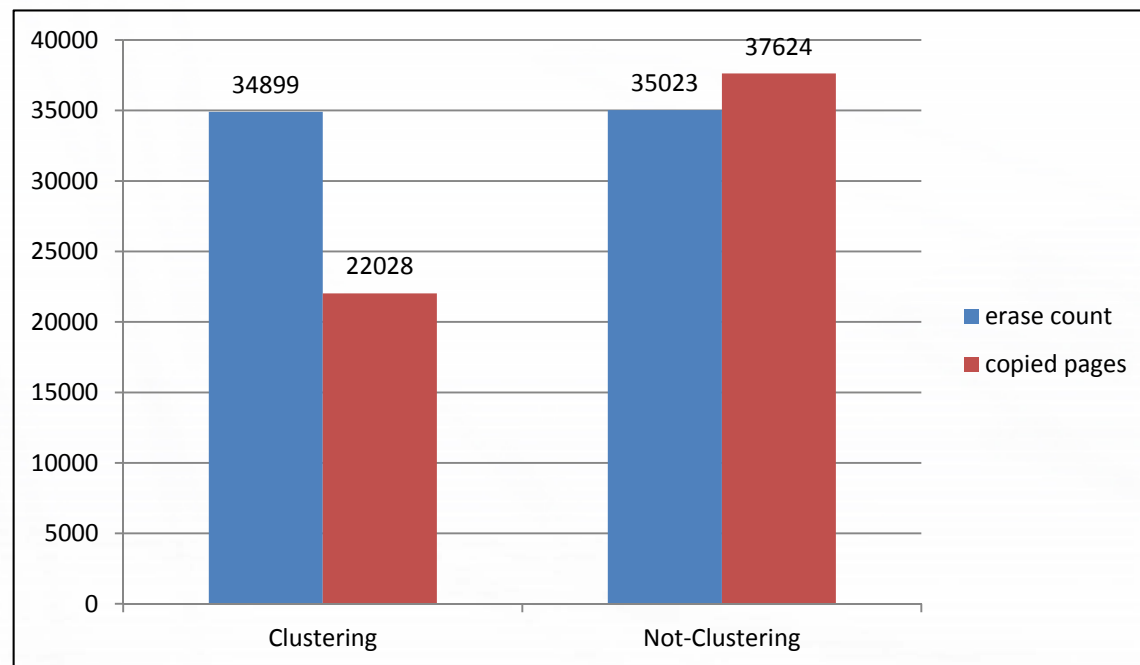
- Problem
  - Statuses of LPN require high memory
- Structures
  - Lpns with status 0 not remain in structure
  - Referred lpn is added to array[ptr] and increase the BIT status
  - ptr circulate array and all BIT of array[ptr] is decreased every time



< Example when window is 5 >

# Enhanced Hot/Cold Page Clustering

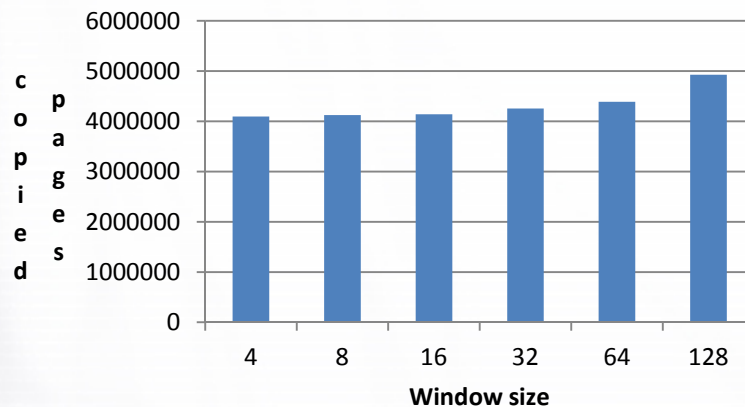
- Clustering vs. Not clustering
  - Financial trace



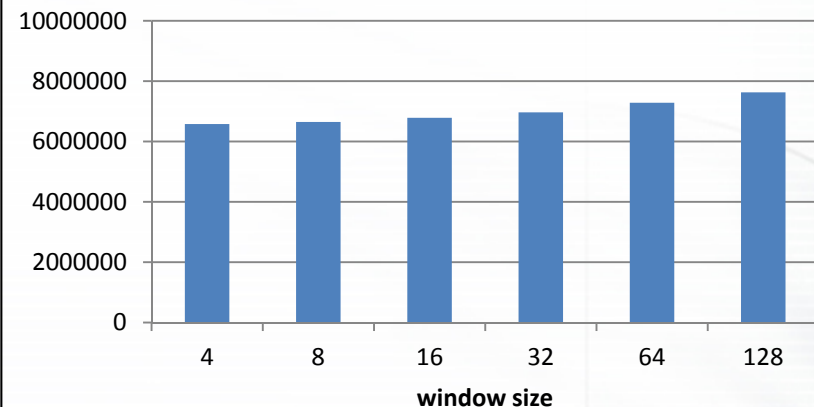
# Enhanced Hot/Cold Page Clustering

- Various BIT / windows / utilization
  - Cello'99 trace

1bit clustering : utilization 30%

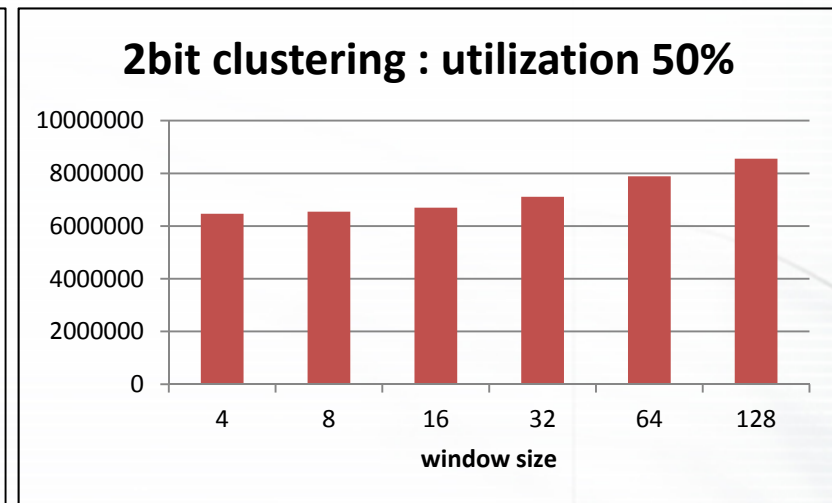
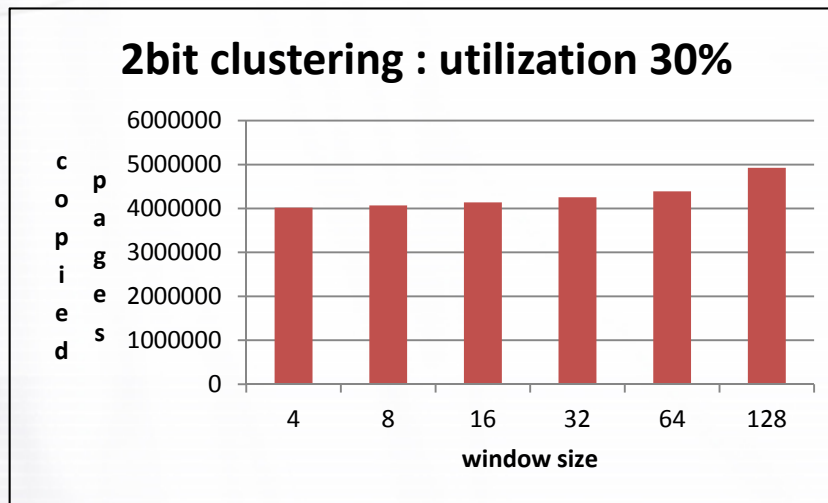


1bit clustering : utilization 50%



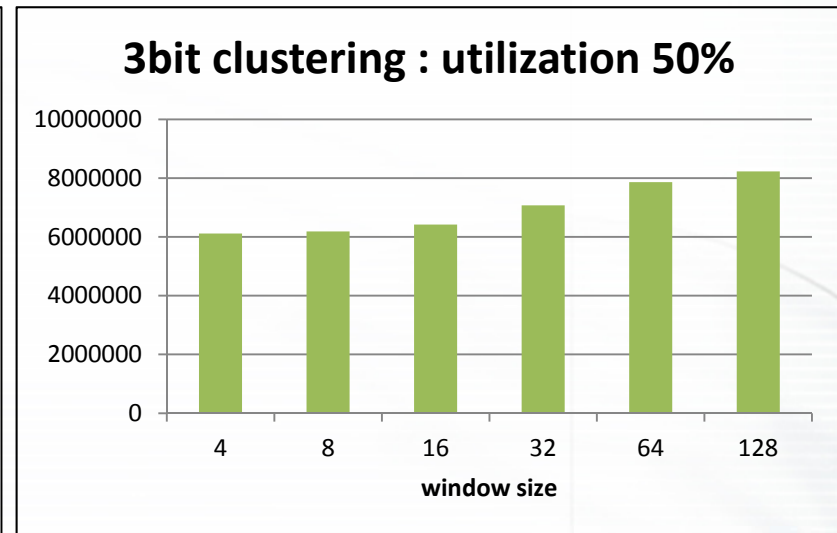
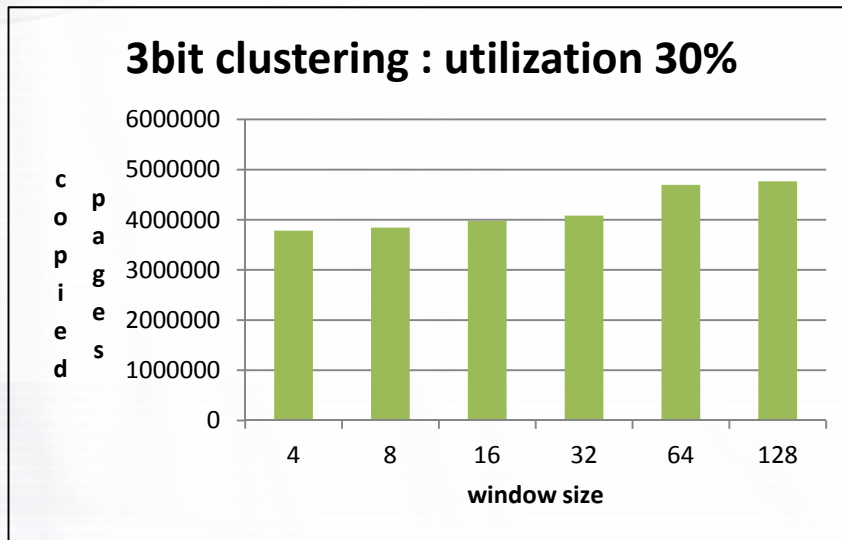
# Enhanced Hot/Cold Page Clustering

- Various BIT / windows / utilization
  - Cello'99 trace



# Enhanced Hot/Cold Page Clustering

- Various BIT / windows / utilization
  - Cello'99 trace

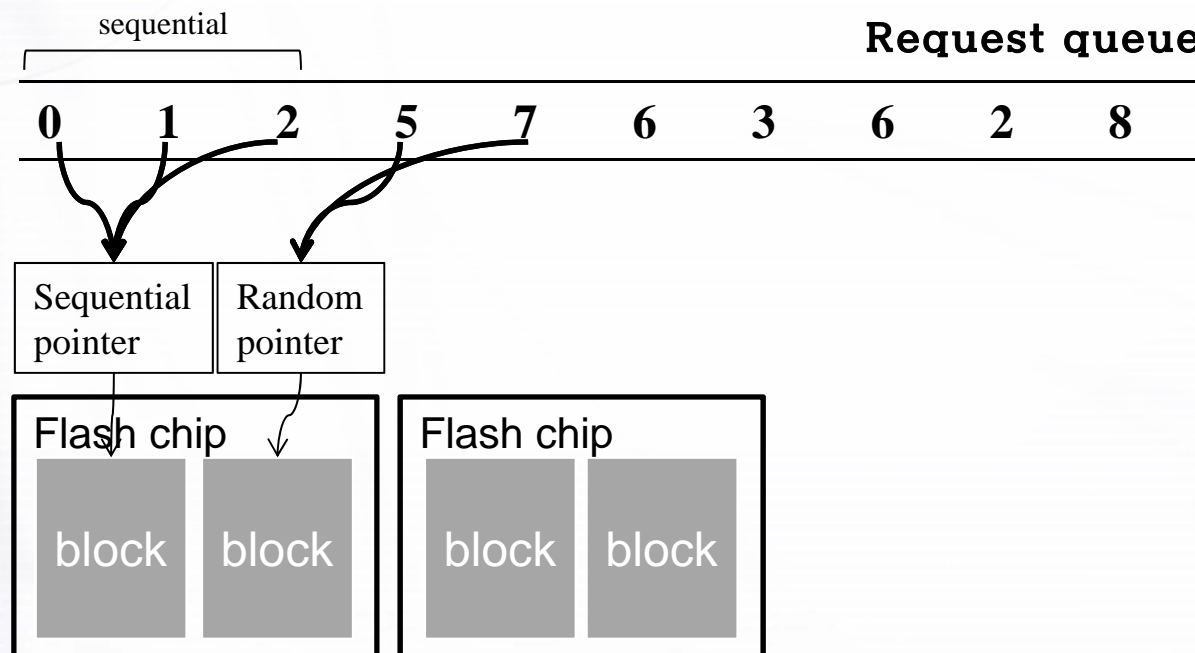


# Sequential/Random Page Clustering

- Page clustering using pattern of requests
  - Sequential data / Random data
  - What's means “sequential”?
    - Contiguous lba (byte address) within 3 requests
- Expectation matters
  - Sequential(Random) data will be requested sequentially(randomly)
- This will be implemented

# Sequential/Random Page Clustering

- Example, S/R clustering

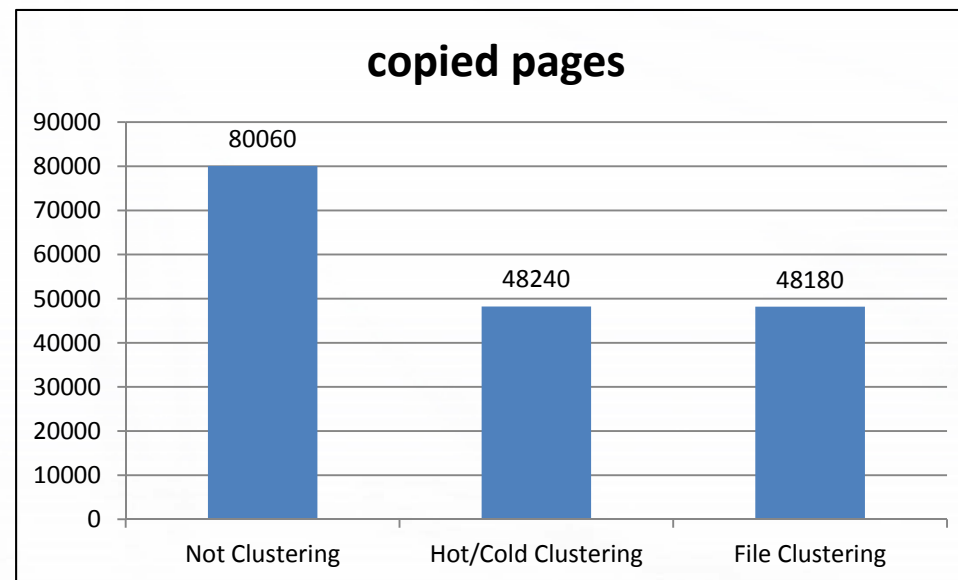


# Page Clustering with File

- Page clustering using file information
  - Request of same file is allocated to same block
- Expectation matters
  - Request of same file will be requested at the same time
- Requirement
  - Host request with file information

# Page Clustering with File

- Experiment environments
  - Trace: TPC-B based on PostgreSQL 8
  - Using dtrace to extract the trace



# Future works

- Integration of various clustering mechanisms
  - Apply various features
  - Preemptive + Clustering
- Parallels
  - Consider parallel-able features like multi channel/way/plane

# Conclusions

- We need the page clustering
  - Page clustering is allocation policy of FTL
  - Overhead of GC is reduced
    - Affect performance consistency
- That is, for performance
  - we should change various policies of FTL