一种高性能混合结构的瓦记录磁盘系统

马留英[1,2] 肖文健[1,2] 董欢庆[1] 刘振军[1] 张强[1,2]

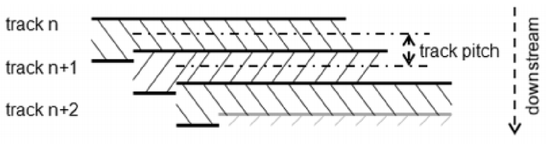
[1]中国科学院计算技术研究所 [2]中国科学院大学

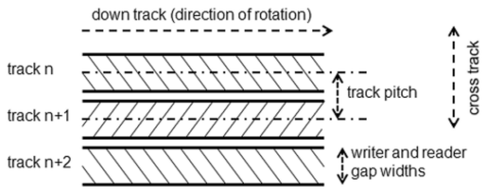
摘要

瓦记录(Shingled Magnetic Recording, SMR)技术因只需对传统磁盘(Hard Disk Drive, HDD)结构和磁记录技术做较小的改变就能明显提升存储密度的优点，使其成为第一个应用于市场的新磁记录技术：希捷的瓦记录磁盘(Seagate Shingled Write Disk, SSWD)现已面市。SMR叠瓦式的磁道布局方式严重限制了其随机写性能。SSWD在磁盘内部使用持久缓存解决SMR叠瓦特性导致的随机访问受限问题，但其在持续随机写入的应用场景下，性能表现很差。因此，本文设计并实现了一种高效的混合结构的瓦记录磁盘 (Hybrid Shingled Write Disk, HSWD)系统：该系统使用SSD(Solid State Drive, SSD)作为持久缓存，设计并实现了两种不同的持久缓存到本地存储的映射方案以及三种持久缓存回收策略，并通过实验和理论分析分别对比了两种映射方案和三种缓存回收策略。最后通过Fio测试以及回放trace测试对比HSWD和SSWD，测试结果表明：HSWD较SSWD在持续随机写性能上有了明显提升，例如：在Fio测试中，HSWD的IOPS较SSWD提升6倍，在回放Financial1测试过程中，HSWD的平均响应时间较SSWD提升2.5倍。

研究背景

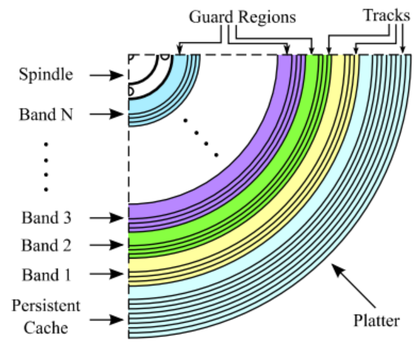
- 瓦记录技术

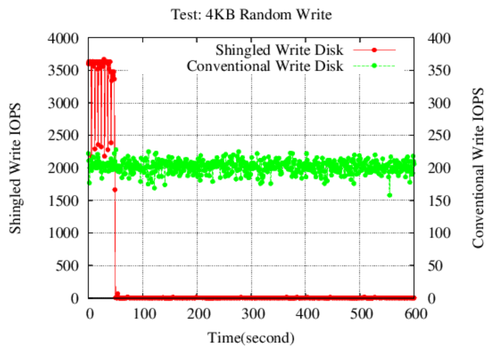




- SMR：叠瓦式磁道布局，缩小磁道宽度，提升存储密度。特点：
  - 读操作(S/R)不受影响(性能与CMR相当)
  - 写操作受限
    - 仅支持顺序追加
    - 随机写受限(需要Read-Modify-Write)
- CMR(Conventional Magnetic Recording)：传统磁道布局方式，读写操作均不受限

- 产品级SSWD内部结构探测及性能表现

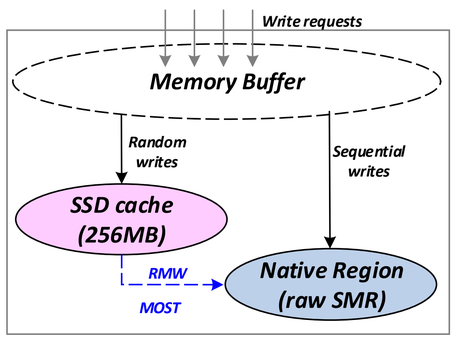




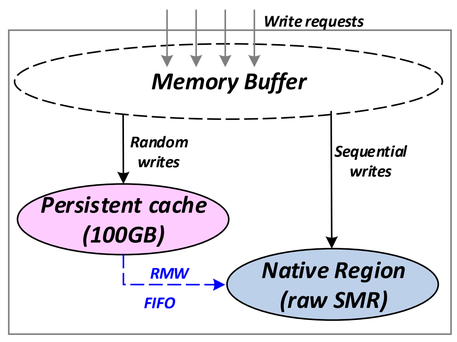
- 以带(15MB~40MB)为单位组织管理，带内不支持随机写，且要求整带顺序写
- 磁盘外圈部分区域(约100GB)用作持久缓存(Persistent Cache, PC)，对随机写请求进行缓存
  - Log方式组织数据
  - 以带为单位执行RMW回收缓存空间，回收策略为FIFO
- 现象及原因
  - HDD的IOPS稳定(绿色曲线)；
  - SSWD的IOPS波动范围大(红色曲线)；
    - 原因：SSWD的持久缓存在短时间内被随机写请求填满，后续随机请求需要等待缓存资源而被阻塞

主要工作

- HSWD：混合结构的瓦记录磁盘系统
  - 内部结构(与SSWD对比)



HSWD

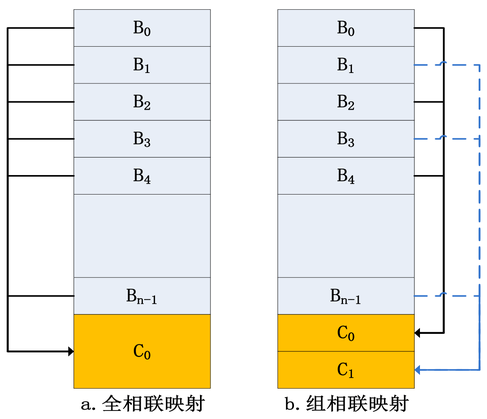


SSWD

- 使用高性能的SSD替换SWD内部持久缓存(SSD空间仅使用256MB)
- 高效的MOST缓存回收策略
- 在HDD上以RMW方式来模拟raw瓦记录随机写行为(带大小32MB)

- 大容量SWD内区域作持久缓存
- 限定的FIFO缓存回收策略

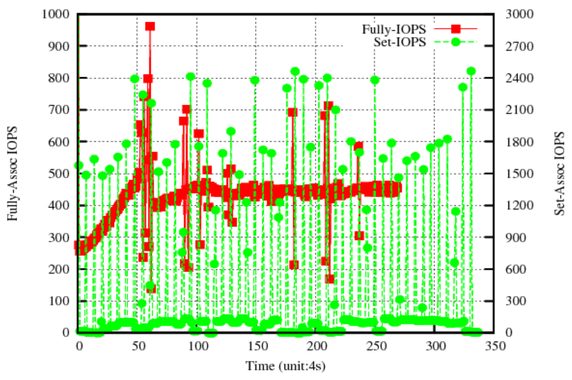
- 两种持久缓存到本地存储的映射策略：全相联和组相联



- 三种持久缓存回收策略(以带为单位)
  - LRU：选取最近最少被访问的带作为回收对象
  - FIFO：按照带被写的先后顺序确定回收的顺序
  - MOST：选取含有缓存块最多的带作为回收对象

测试与分析

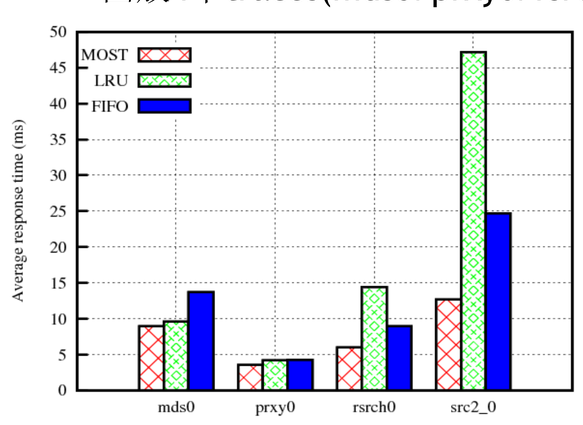
- 全相联 vs. 组相联
  - 以4KB为粒度，随机写2GB数据

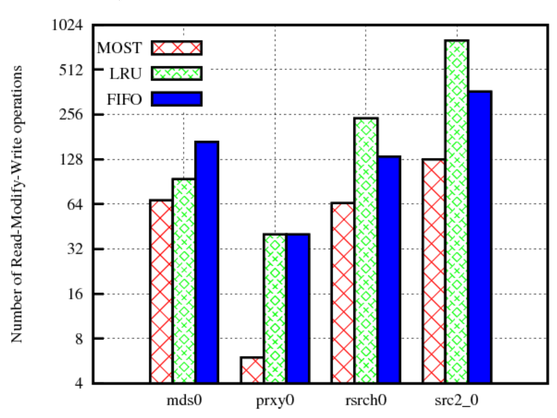


Policy	Avg. IOPS	Avg. IO Latency	Max IO Latency	RMW times	Runtime
Fully	483	64.1ms	4.0s	274	18.1min
Set	387	80ms	26.2s	351	22.6min

- 结论：全相联优于组相联

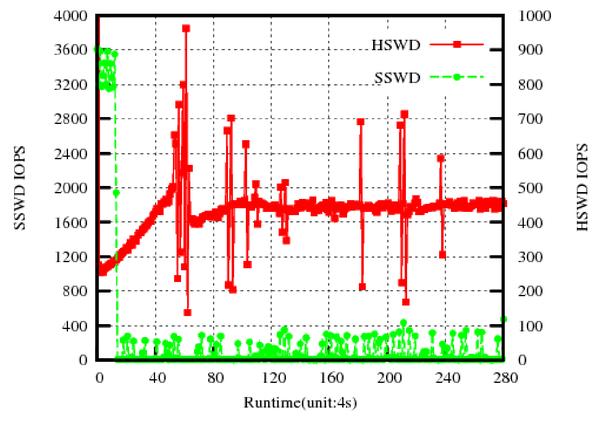
- LRU vs. FIFO vs. MOST
  - 回放4个traces(mds0/prxy0/rsrch0/src2_0)

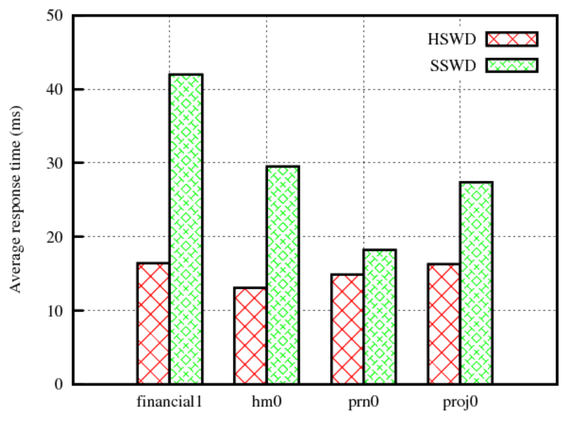




- 结论：MOST策略优于LRU/FIFO，主要归因于采取MOST策略执行一次RMW可回收的缓存块数量最多

- HSWD vs. SSWD





- 以4KB为粒度，随机写测试
  - 平均IOPS：
    - HSWD：483
    - SSWD：79
- 回放financial1/hm0/prn0/proj0
  - HSWD的性能优于SSWD (HSWD的性能分别为SSWD的2.5倍/2.3倍/1.22倍/1.7倍)

总结

基于对SSWD内部结构的探测和特性分析，HSWD使用SSD替代SSWD内部持久缓存，以解决SSWD面临的在持续随机写入场景下性能差的问题。HSWD设计并实现了两种持久缓存到本地存储的映射策略：全相联和组相联，以及三种不同的缓存回收策略：LRU/FIFO/MOST，并通过测试分别对比了两种映射策略和三种回收策略。测试结果表明：全相联的映射策略因能最大化使用持久缓存空间并减少回收次数而优于组相联；相较于LRU和FIFO回收策略，MOST回收策略因一次能回收最多的持久缓存空间而更适合于HSWD。最后，通过使用Fio基准测试和回放trace测试对比HSWD和SSWD，测试结果表明：仅使用256MB的SSD空间作为持久缓存，HSWD较SSWD在持续随机写性能上就会有明显提升。

Email: maliuying@ict.ac.cn