Motivation

DRAM vs NVM
- DRAM: faster but smaller and higher energy
- NVM: larger, lower energy but limited write endurance

Hybrid memory architecture
- (1) hierarchical model
- (2) side-by-side model

CPU
DRAM
Cache
NVM
CPU
DRAM
Cache
NVM

Leverage cache to minimize writes to NVM

Traditional locality theory

Reuse distance (rd)
- Def: The number of distinct memory accesses between use and reuse.
- If rd > cache size, then reuse is a cache miss.

Footprint (fp)
- Def: The averaged Working Set Size (WSS) for some specific window size.
- If fp(rt) > cache size, then reuse is a cache miss.

Write locality

Write interval
- Def: The time window between two consecutive writes to the same location. No intervening writes to that location. Reads to that location or reads/writes to other locations are allowed.

Write reuse distance (wrdf)
- Def: The maximum reuse distance for w-r, r-r, r-w pairs inside each write interval.
- If wrd > cache size, then the first write causes a writeback.

Write reuse time (wrt)
- Def: The maximum reuse time for w-r, r-r, r-w pairs inside each write interval.
- If fp(wrt) > cache size, then the first write causes a writeback.

Write back ratio curves for 29 SPEC CPU2006 benchmarks

Summary

Write locality is measured by write reuse distance and write reuse time and used to minimize cache writebacks.

Reference