SDC
STORAGE DEVELOPER CONFERENCE
SNIA SANTA CLARA, 2016

ZDM
Using an STL for Zoned Media on Linux

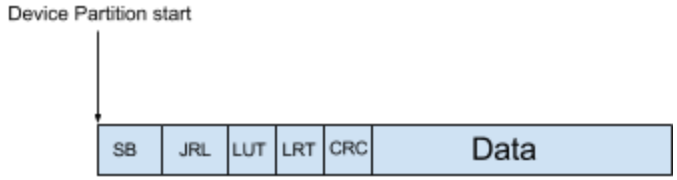
Shaun Tancheff
AeonAzure
-for-
Seagate Technologies

ZDM: Shingled Translation Layer [STL]

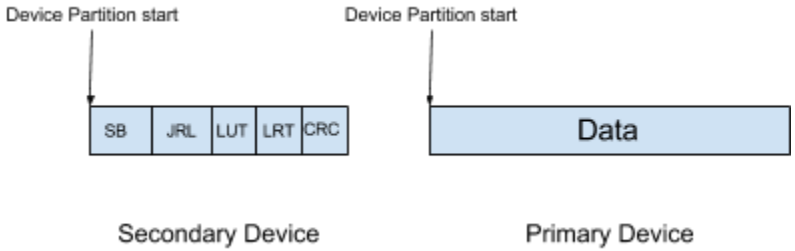
- STL is similar to a Flash Translation Layer
  - Block remapping
  - Copy on Write
  - TRIM
  - Zone reclaim

ZDM: Data layout options

- STL on same device



- STL on a secondary device



- STL mirrored on both devices

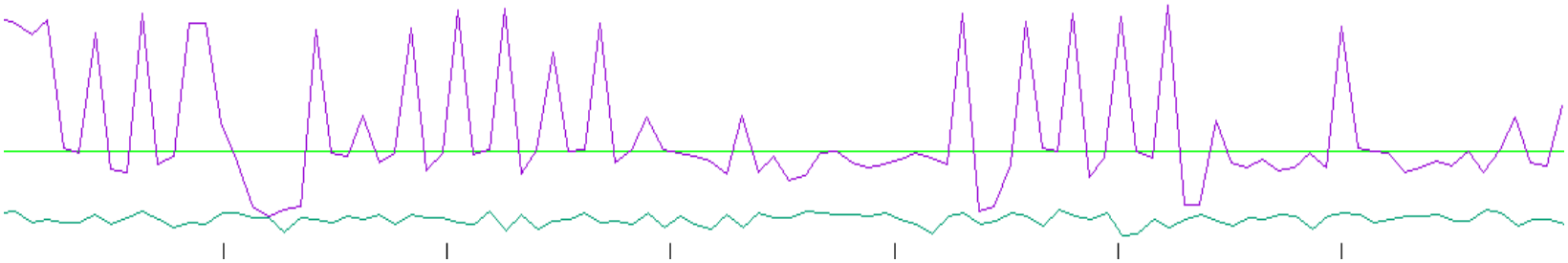
ZDM Performance

- Up to 90% of disc rate on write
  - Meta data on flash device
  - 75% with meta data on primary device
- Up to 70% of disc rate on read
  - Meta data on flash device
  - 50% with meta data on primary device.
- How this compares with a conventional drive
  - Some activity is faster, some is slower..

ZDM Host Aware vs Conventional

- FIO on with ext4 on ZDM with Host Aware SMR
- FIO on ext4 with conventional disk
  - Testing I/O at 4k, 16k, 64k and 1M
    - Sequential Write
    - Random Write
    - Read/Write
    - Random Read/Write
    - Random Read
    - Sequential Read

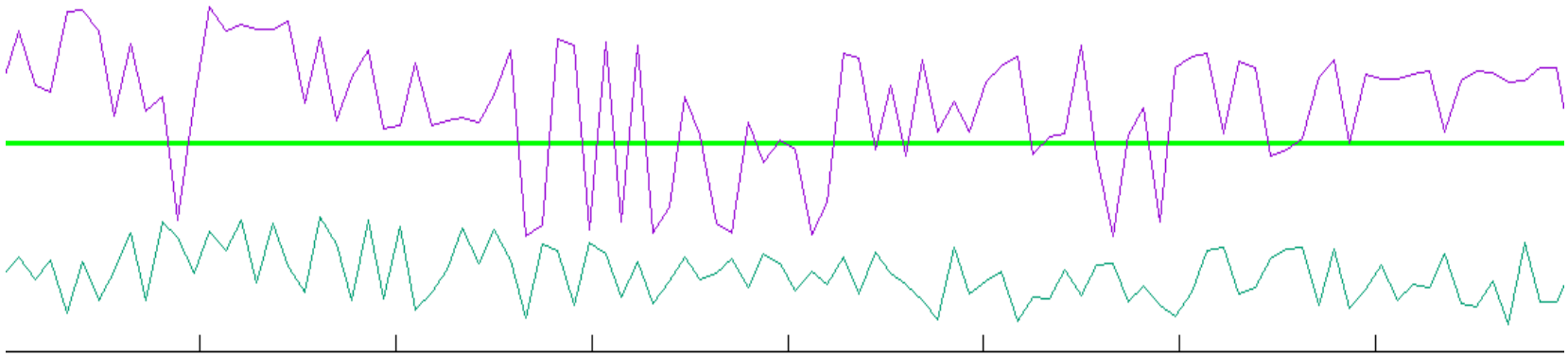
ZDM performance compared



4k Initial Sequential write

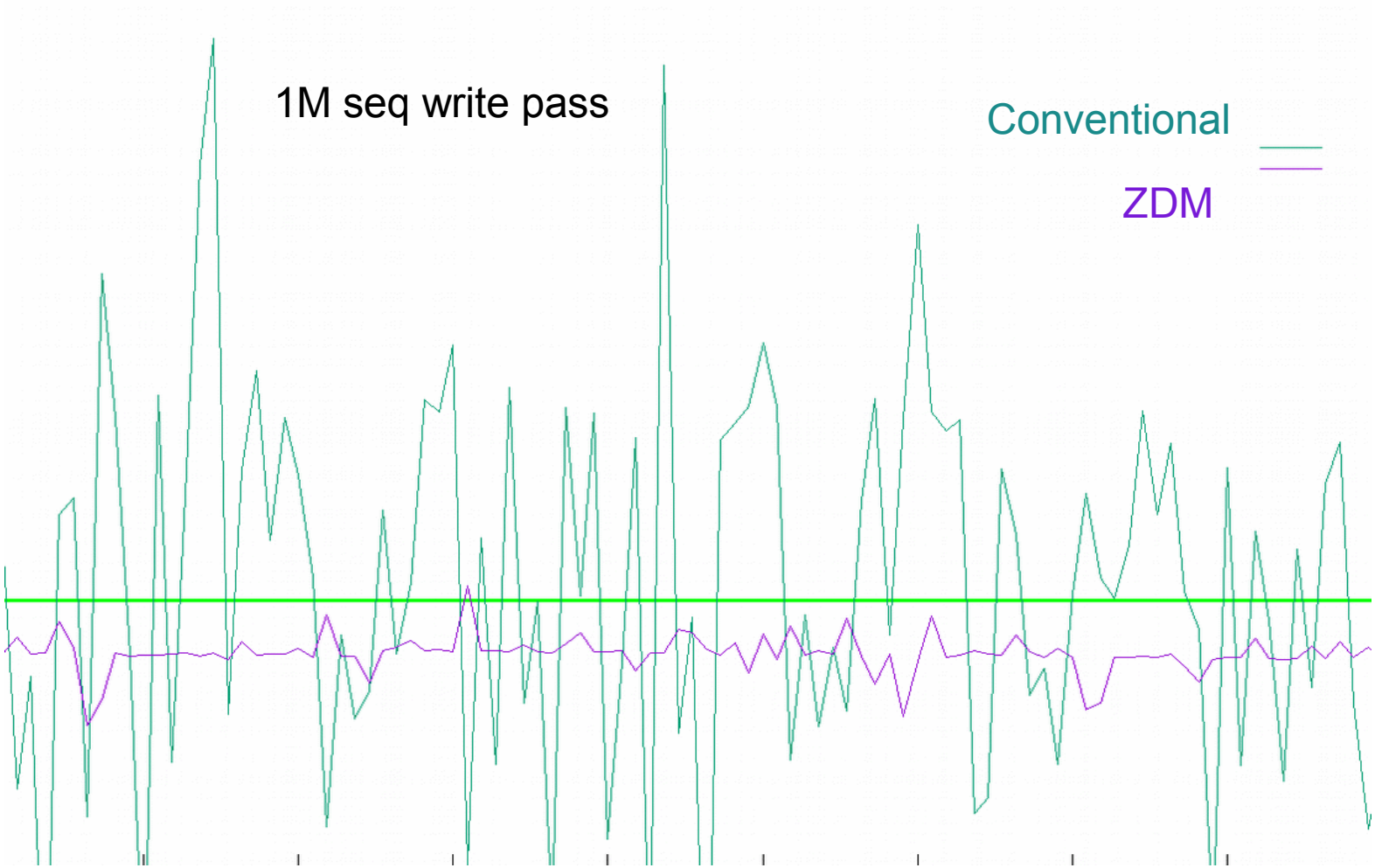
Conventional

ZDM



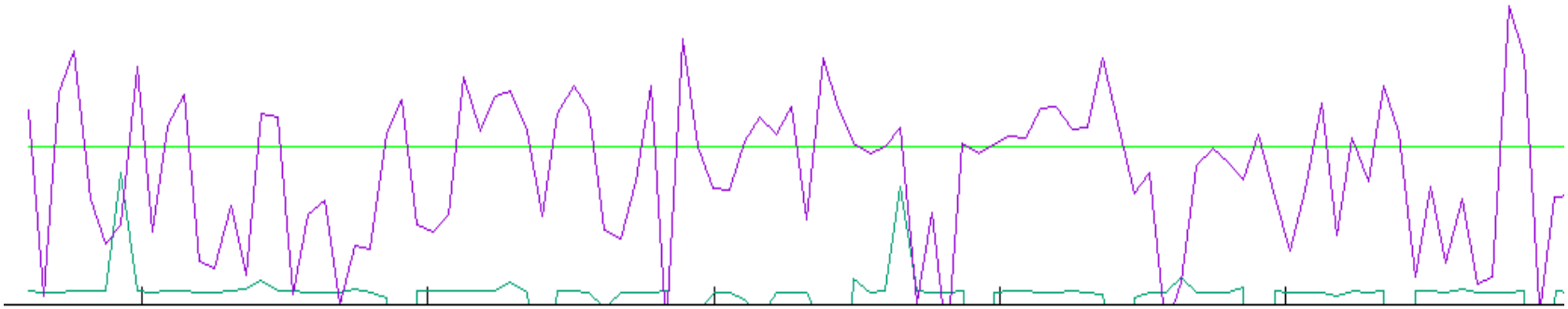
16k Sequential write pass

ZDM performance compared

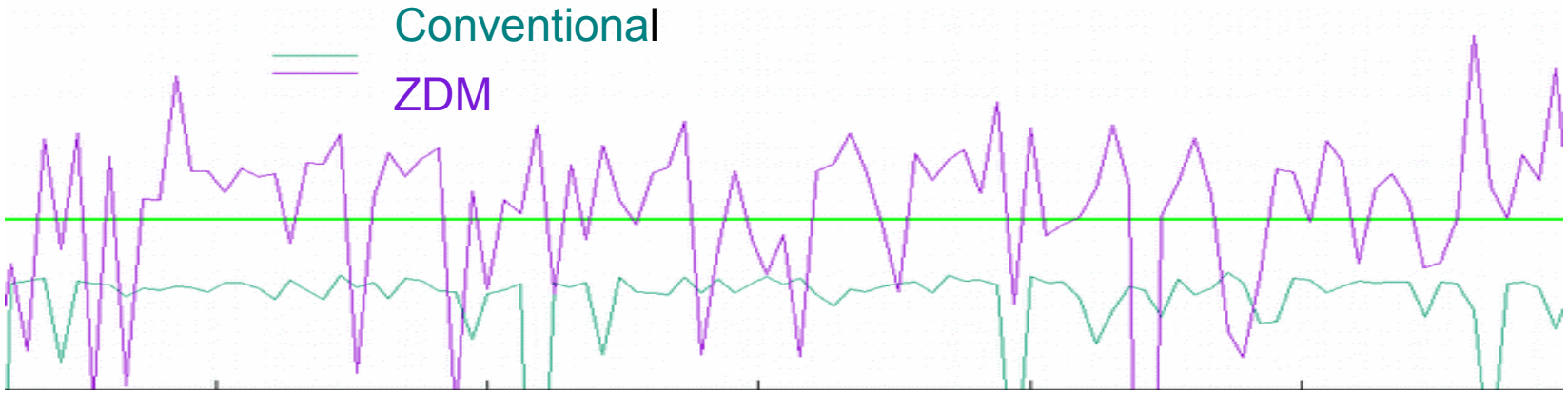


ZDM performance compared

64k random write pass

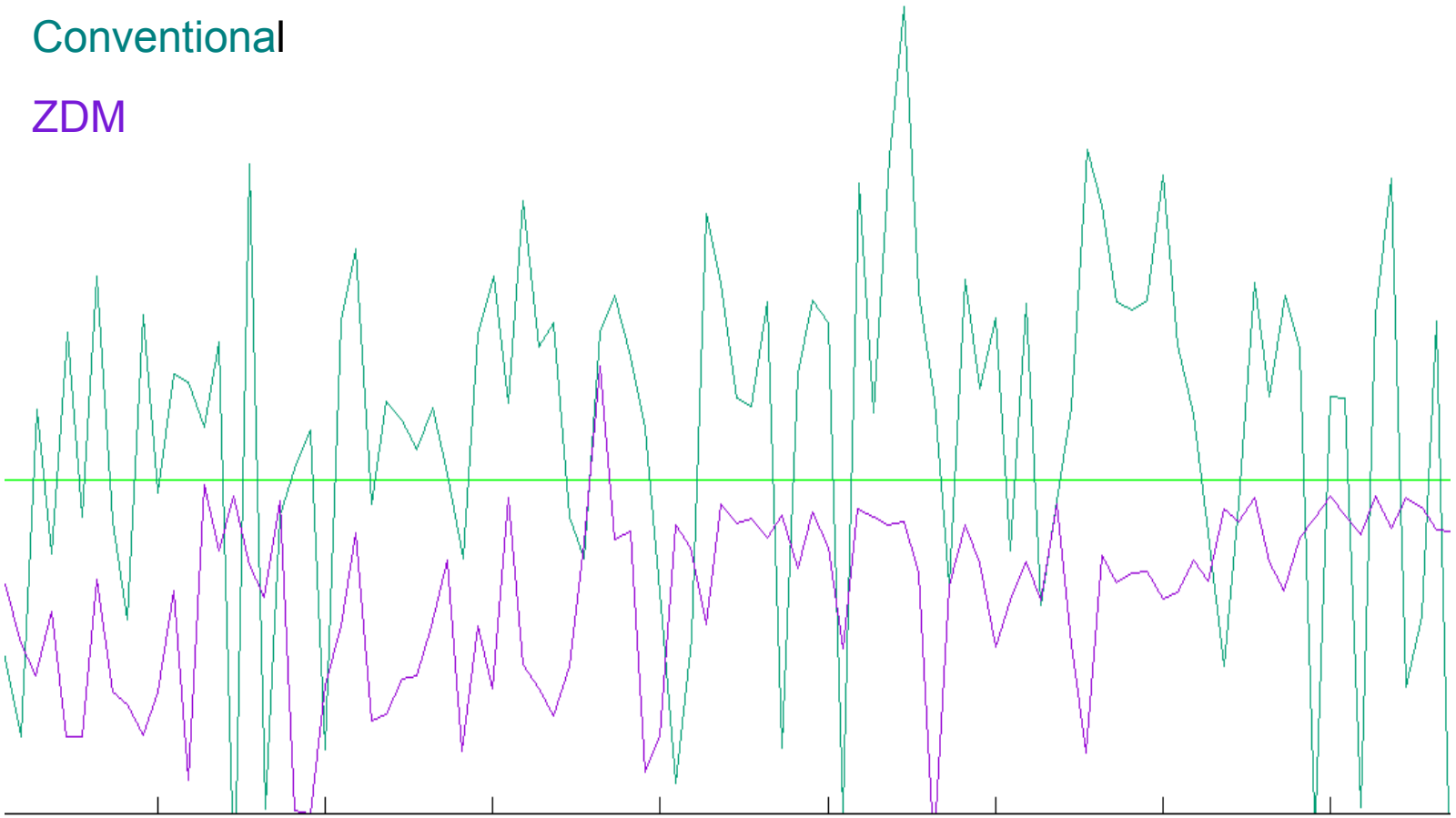


1M random write pass



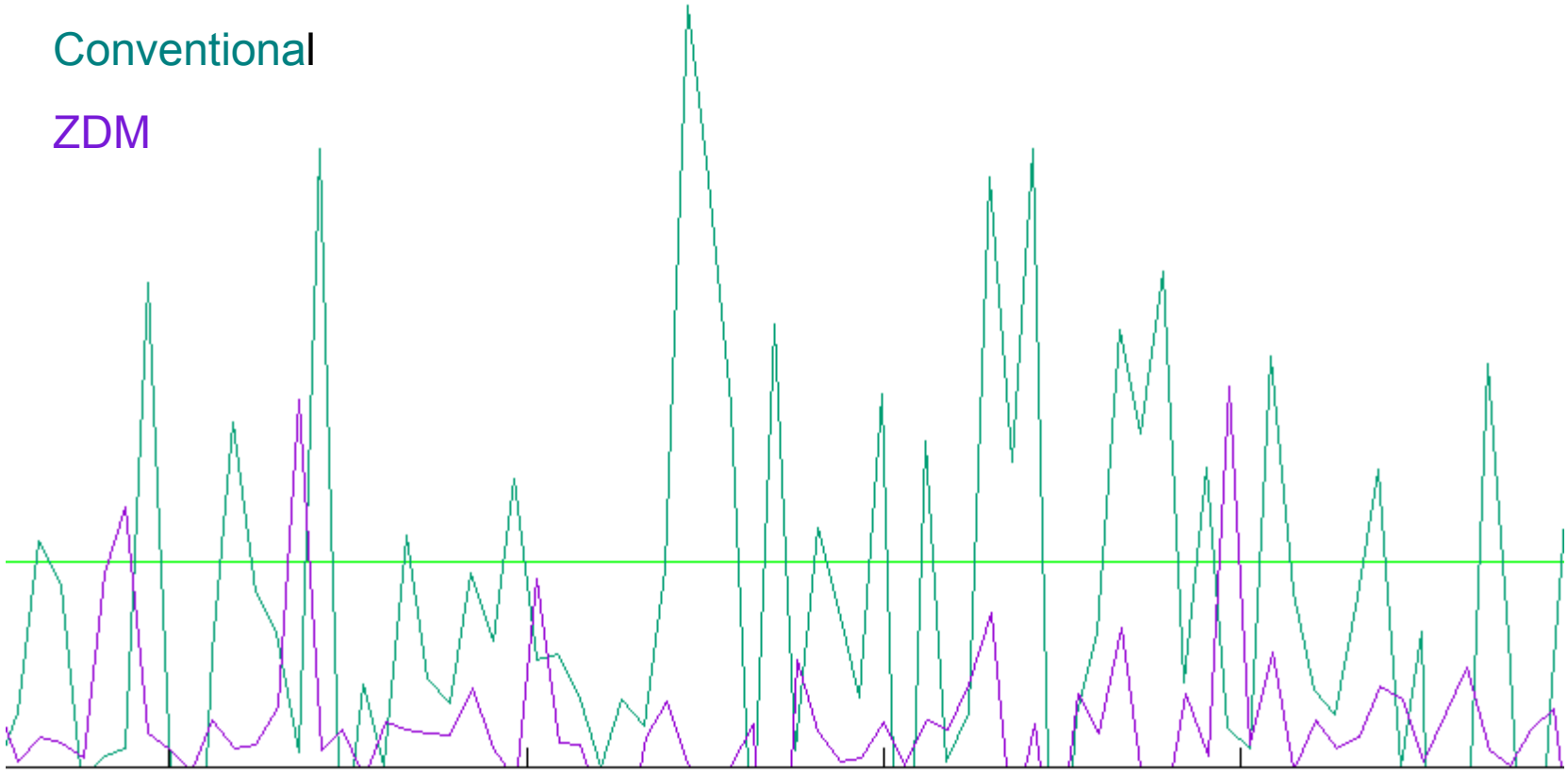
ZDM performance compared

16k Read / Write pass



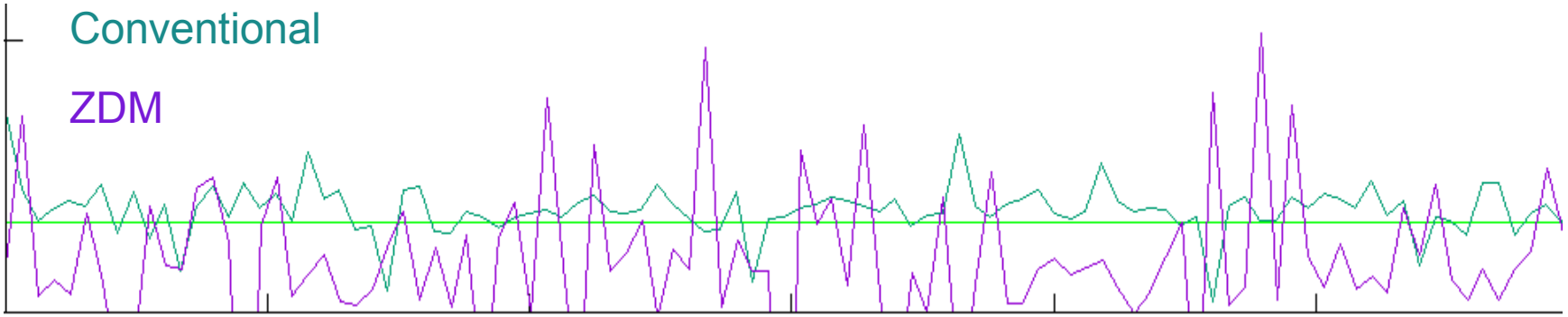
ZDM performance compared

1M Random Read / Write pass

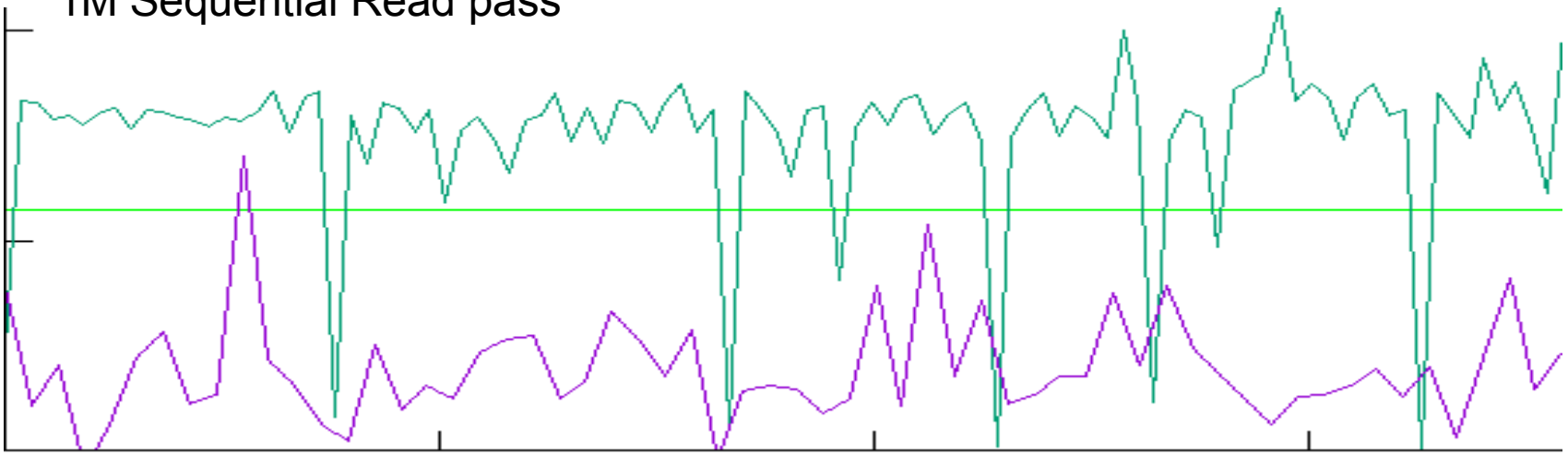


ZDM performance compared

1M Random Read pass



1M Sequential Read pass



Candidate Workloads

- Write centric workloads
  - Archival data storage
  - Video Surveillance
- MD RAID
  - RAID 5/6 - SOHO
  - RAID 0/1/10

ZDM Tuning

- Metadata on secondary device
  - Flash speeds up reading and Flush
  - A second spindle also works
- Zone reclaim control
  - zdmadm –gc on/off/force
- Metadata Cache
  - zdmadm –cache-timeout <ms>
  - zdmadm –cache-size <pages>
  - zdmadm –cache-to-pagecache

ZDM in the Cloud

- Persistence with Flush
  - Offload to Flash (or other device)
  - Enable Metadata Journal to Stream
- Sequential Read after Random Write
  - Flash ‘helps’ but ..
    - Changing on-disc STL to extent maps
    - Tail chasing read cache

ZDM in embedded and MD-RAID

- Discard support for RAID 5/6
  - zdmadm –raid-trim 1
- Bio merge queue for RAID 5/6
  - zdmadm –queue-depth <count>