Towards an Unwritten Contract of Intel Optane SSD

Kan Wu, Andrea Arpaci-Dusseau and Remzi Arpaci-Dusseau
University of Wisconsin-Madison　　出自 Hotstorage'19
这篇论文主要讲述了使用 **3DXPoint SSD** 的 **7 条**不成文的规定。
实验比较：Intel Optane SSD 905P(960G)　V.S.　SAMSUNG 970 Pro(1TB)

Rule 1: Access with Low Request Scale

这里 low request scale = small request size + low queue depth，即 Optane SSD 用户应该发出小粒度的请求(≤4KB)和维持少量 Outstanding IOs(未解决的 IO 请求，用队列深度表征，应≤2)。
原因：主要在于 Optane SSD 的内部并行特性（interlevaving degree,交叉度），具体地此处 Optane SSD 的 channel 数目为 7，Flash SSD 的 channel 数目为 127.

Rule 2: Random Access is OK

在 Optane SSD 上进行随机访问和顺序访问，其平均延迟非常近似。
原因：
①Optane SSD 是一个 in-place update（原地更新数据，不同于 Flash SSD 的垃圾回收机制）的随机访问块设备；
②Optane SSD 在顺序读时无预取(prefetching)；
③地址映射方式为 LBA（Logical Block Address）。

Rule 3: Avoid Crowded Accesses

应避免在 Optane SSD 上对一个 chunk 上的多个 sector 进行并行读/写操作。
原因：并行读/写进程争夺 controller 资源，引入了 contention。

Rule 4: Control Overall Load

应控制总负载(读+写)的大小。
原因：Optane SSD 对写负载和读负载的延迟都是一样的，究其原因还是因为 in-place update，故控制总负载大小，以控制 queue depth，从而有较低的延迟。

Rule 5: Avoid Tiny Accesses

应避免过小的访问粒度。小粒度访问降低了 Optane SSD 的 throughput。
原因：Optane SSD 最大的 IOPS 是 575K，即 IOPS 受限。
Throughput 吞吐量是用来计算每秒在 I/O 流中传输的数据总量。
IOPS（IO per Second）是用来计算 I/O 流中每个节点中每秒传输的数量。

Rule 6: Issue 4KB Aligned Requests

尽管 Optane SSD 能字节寻址，但是 Optane SSD 更适合对齐(4KB,此处也是 8 sector)的访问。
注意，这里的 8 扇区块与 Flash SSD 中的页面或块之类的概念无关。

Rule 7: Forget Garbage Collection

Optane SSD 中无垃圾回收机制。

Optane SSD 内部结构

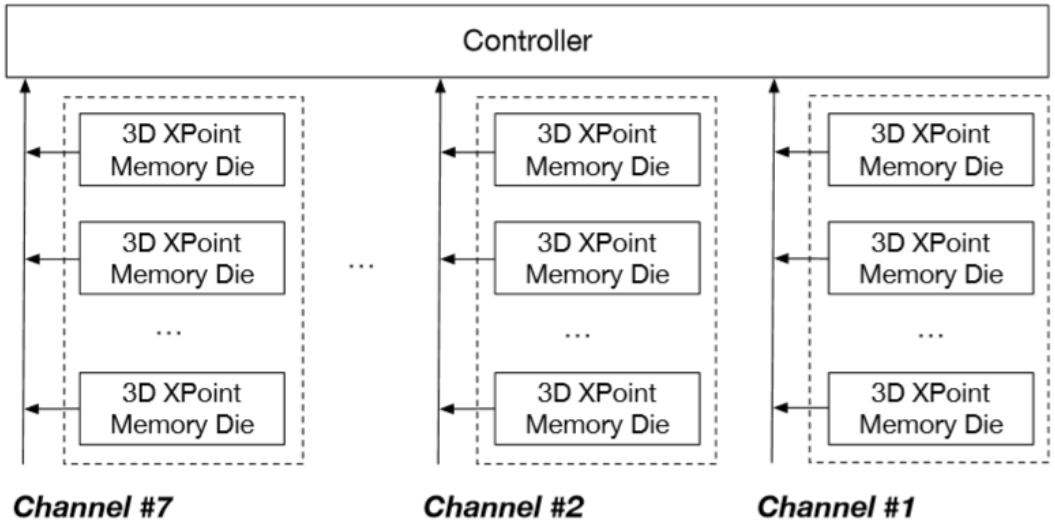

Figure 2: RAID-like Architecture