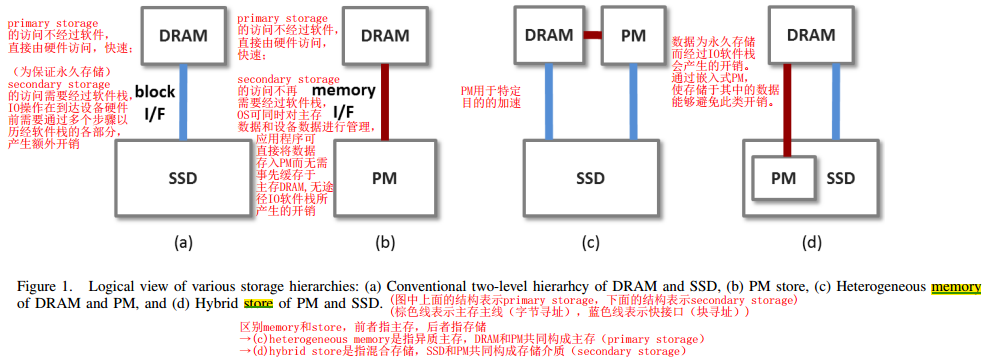
**2B-SSD-The Case for Dual, Byte- and Block-Addressable Solid-State Drives**

主机处理器+固态硬盘嵌入式处理器

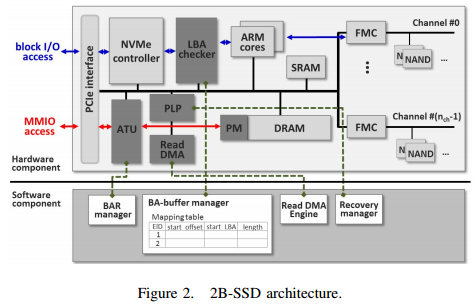
应用领域：数据库（case study）

存储系统希望能够快速写数据并保持其持久性，持久存储器（PM）应运而生。现代数据管理和存储系统也正面临着字节寻址PM的兴起。下面是几种存储架构：



由于与主机内存同为字节寻址，字节寻址型PM无需经由IO软件栈，可由主机CPU直接访问存取，进而能够缓解主机端与设备端间数据传输的开销。然而，PM容量有限且访问延迟较大（相较于DRAM），无法完全取代块存储或主存。另外，若如（c）结构，PM会消耗稀缺的SIMM插槽且需后备存储、外部电源和支持电路，设备复杂、昂贵。因此，提出如（d）结构的混合持久存储。

2B-SSD即是建立在混合持久存储结构上的，允许对相同文件通过两条独立数据路径（字节IO或传统块IO）进行访问。PCIe接口通过MMIO使内存和IO设备共享内存地址空间，使两者均能够通过主存指令进行访问，这使得获取设备端PM内数据的字节IO数据路径可行。其具体结构如下图所示：



（1）BAR manager：**软件组分**使能额外的基址寄存器（BAR）BAR1分配单独的内存地址空间给设备内DRAM（PM），用于CPU通过内存访问指令与2B-SSD的交互，以实现字节粒度的文件访问。**硬件组分**地址转换单元（address translation unit，ATU）实现由原本的CPU对闪存阵列的访问到CPU对设备端DRAM的访问的重定位，即将对闪存阵列的访问与对设备内部DRAM的访问联系起来。

（2）BA-buffer即DRAM中保留为持久存储的部分存储空间。（**硬件组分**）

BA-buffer manager：**软件组分**在SSD嵌入式控制器的ARM内核上实现BA-buffer逻辑，将闪存阵列内数据的逻辑地址映射到设备内部DRAM，并处理两者间的数据传输。映射信息于映射表中管理，表中包含（1）entry\_id,（2）BA-buffer中start\_offset,（3）闪存阵列中start\_LBA和（4）length，其中（2）（3）即是闪存阵列内数据的逻辑地址和设备内部DRAM之间的映射。创建有API管理BA-buffer映射逻辑（映射的建立&数据由闪存阵列到DRAM的拷贝等）、应用对BA-buffer上存储空间的分配、应用对BA-buffer上文件的读写等。

闪存阵列到SSD上DRAM之间的映射关系一旦建立，且使用API进行数据拷贝后，闪存阵列和DRAM上均有应用所需数据，则有两种访问方法/数据路径：Memory interface (BA-buffer)和conventional (NVMe) block I/O。但多种访问方会导致数据不一致的问题，为避免此问题，2B-SSD将阻止对于已经映射到BA-buffer的闪存阵列页面的块IO请求。**硬件组分**LBA checker检查每个IO请求的LBA以筛查各闪存阵列页面以找出上述情况进行块IO阻止操作。

（3）Read DMA engine：PM虽可绕过IO软件栈由主机直接存取以减少开销，但因无缓存机制且读操作单位比块IO粒度小，与块IO相比，其读取速度较慢。**软硬件组分**通过将对于庞大数据的读操作交由DMA处理，拷贝数据通过DMA传输，加速存取，提高读性能。

（4）Recovery manager：**软硬件组分**在SSD原有数据恢复系统上（1）添加一个额外的小电容确保掉电时BA-buffer内容和相关信息（映射表）能被刷新到闪存阵列的保留区域内（DRAM是易失的，但闪存阵列是非易失的）（2）添加用于恢复BA-buffer内容和管理信息的软件恢复逻辑。通过Recovery manager将原本易失的BA-buffer(DRAM)转换为非易失的PM。

Summary：

韩国三星微电子借助恢复机制确保数据持久性并将设备内部部分DRAM转变为持久性存储，实现硬件地址转换单元实现闪存阵列到设备端DRAM的重定位并引入两者兼得数据路径，最终软硬件协助实现访问同一文件的两条独立的数据路径；字节IO路径及传统块IO路径。同时实现Read DMA engine实现类似DRAM的低延迟。

