# Large-Scale Adaptive Mesh Simulations Through Non-Volatile Byte-Addressable Memory

**总体概括：**

为了达到高精度的模拟，物理仿真应用对计算机的内存的需求越来越高，模拟过程越来越复杂。现存的一些解决方案包括（1）将计算所需的数据全部在内存中，并且将运行过程中的检查点放在磁盘中以便进行恢复工作；（2）将运算过程所产生的数据放在磁盘内部以便提供更大的存储空间并且使得数据不丢失。然而本文使用NVM(non-volatile byte- addressable memory)新型介质并结合当前物理模拟应用针对八角树(octrees)设计出了一种新型的数据结构——PM-octree(Persistent Merged octree)来解决NVM的写开销所带来的问题，并利用NVM的特性为应用提供容错恢复机制。

**本文贡献：**

1 提出了一种多版本的PM-octree数据结构，如图1所示，并使用NVBM介质来扩展内存的能力。除此之外通过NVBM的持久性来进行数据恢复。

2 根据八角树的数据特征设计了新算法以便动态调整数据存放的位置用于减少NVBM的写次数从而提高系统运行效率并增加NVBM的寿命。

3 PM-octree提供了方便用户编程的接口。

4 将PM-octree与流体模拟应用结合并进行测试。

**设计细节：**

本文设计目标是使用NVBM来对内存进行扩展并且提供错误恢复机制。本文通过测试发现，在应用运行过程中相邻的两个版本的八叉树有着非常高的重合率，并且该重合率会随着步数的增加而增大。在水滴模拟应用中随着步数由0到150，Vi-1版本与Vi版本数据的重合率由39%增长到99%，如图2所示。

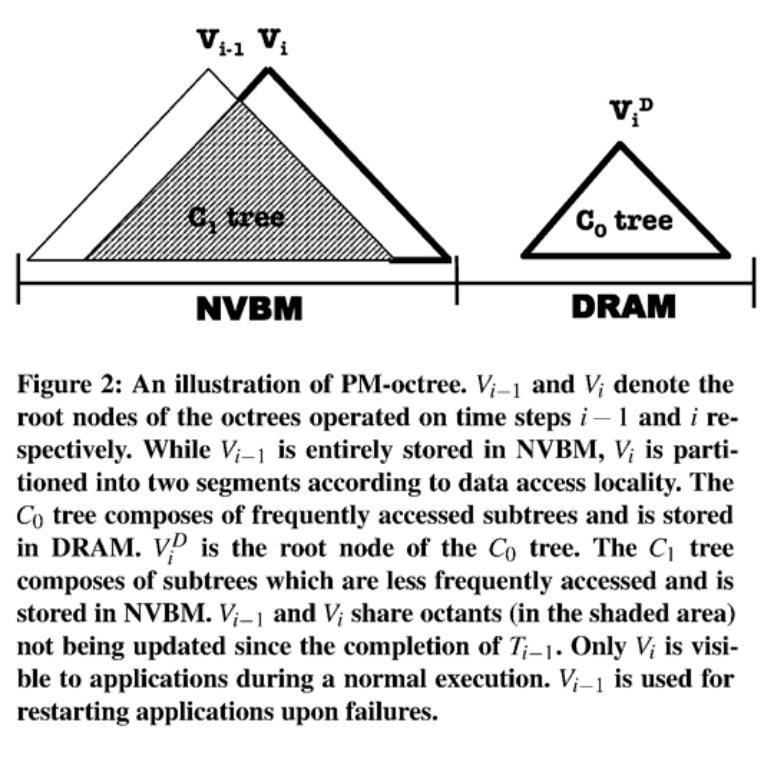


图1：PM-Octree数据结构

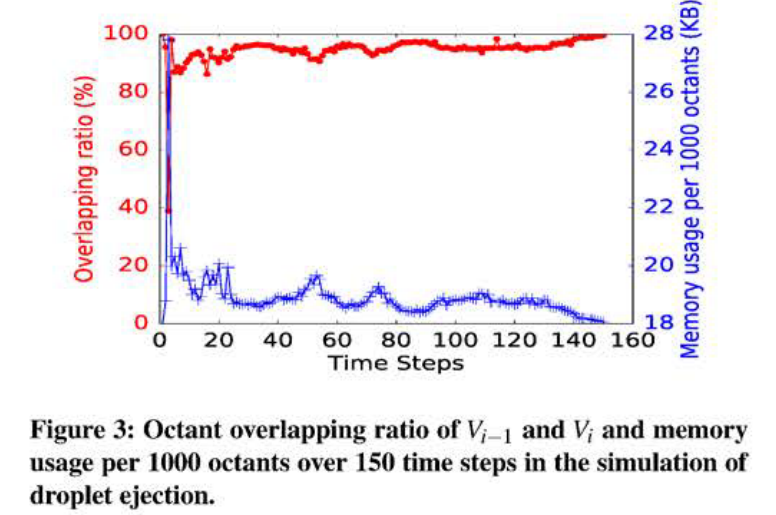


图2：八叉树相邻版本重合率与内存使用率

在操作上，文章设计了PM-octree并介绍了插入、更新、删除、垃圾回收、结构翻转以及合并操作。其中最重要的设计是PM-octree在介质间的动态调整。

由于直接在NVBM上访问数据会导致很严重的读写延迟问题，所以为了最优化这个问题，我们需要使用DRAM来存储访问频繁的热数据，并使用NVBM来存储访问次数较少的冷数据。如图3所示，图中黑色的点表示访问频繁的数据而白色的为冷数据。而(a)中冷数据放置在DRAM中，热数据在NVM中。于是我们设计新的模式以达到图(b)的效果，图中的热数据均放置在DRAM中，而NVM中放置冷数据。转换操作会在达到某一特定要求时触发。

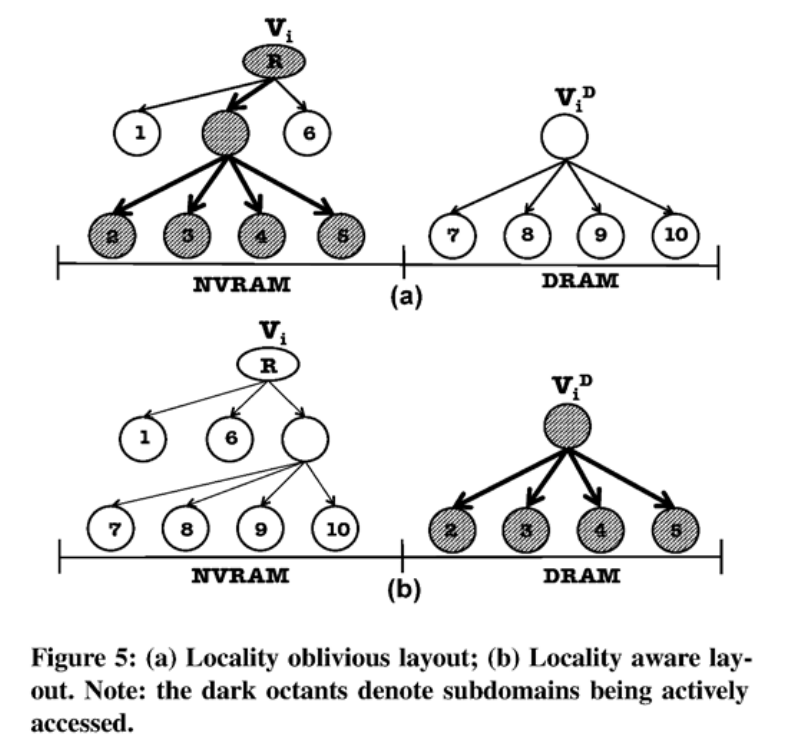


图3：数据动态放置前后效果对比图

在数据恢复方面，本文提供了一个恢复接口。在应用运行的过程中，上一个版本的数据并没有被真正的删除，而是被标记为’delete’，只有当当前版本运行成功后才会真正进行替换工作。如果在当前版本运行过程中应用发生了故障，此时则将上个版本中的’delete’标记删除即可恢复