# 固态存储技术

华宇

https://csyhua.github.io/

# 计算机存储

- 计算机世界是0,1的世界，数据就是0,1的组合，存储数据就是要存储这些0和1。
- 理论上，具有两种稳定状态的材料都可以用来存储数据。
- 磁存储—硬盘
- 光存储—光盘
- 半导体存储—闪存，相变存储器

# 半导体存储设备

- 闪存存储器（Flash）
- 闪存是电容性的半导体存储器件
- 相变存储器（PCM）
- 相变存储器是电阻性的半导体存储器件

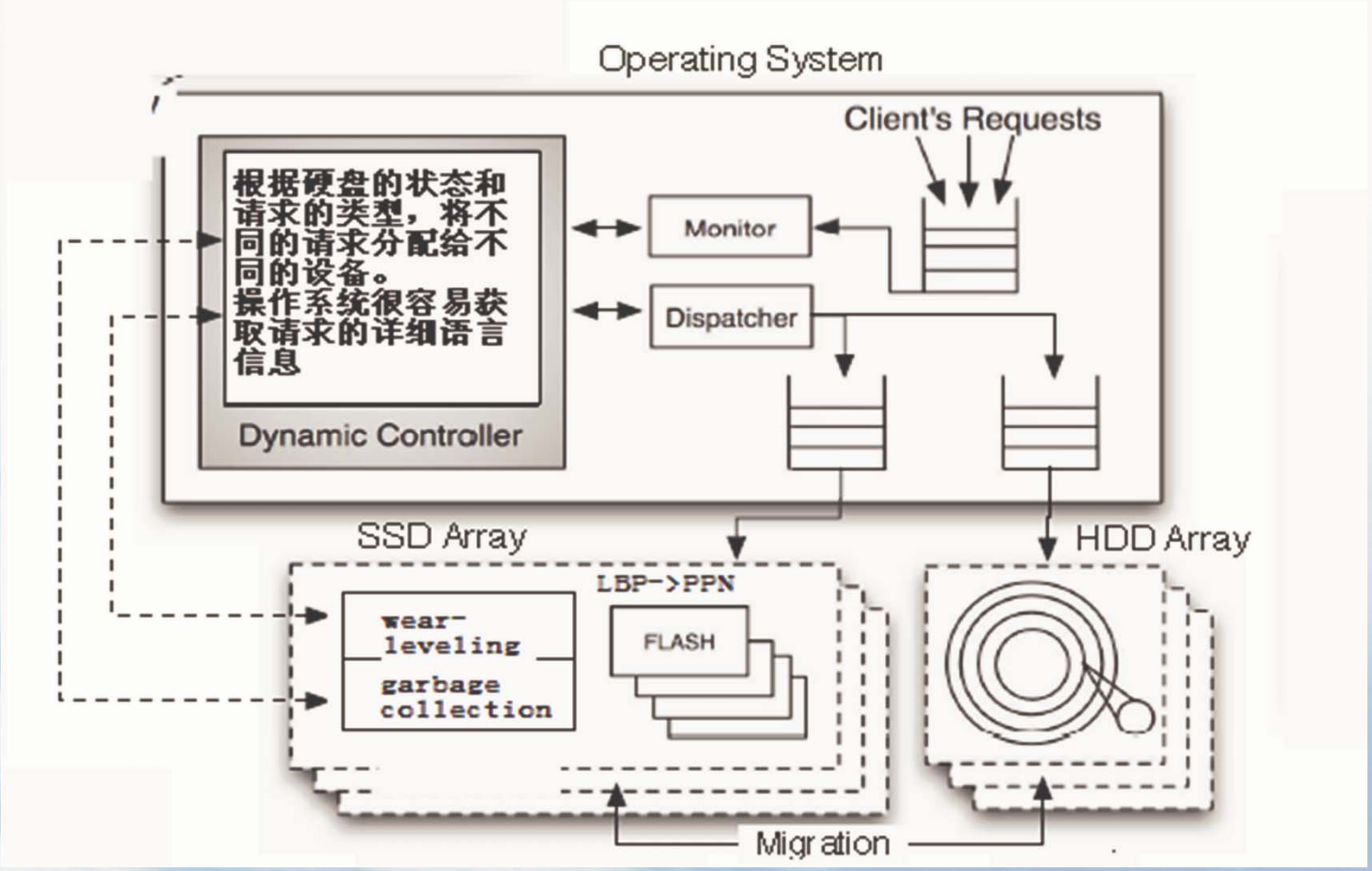
# SSD的优势

- SSD没有机械部件，抗震动
- SSD不需要马达，低能耗
- SSD高性能
- SSD价格在不断下降

# SSD在存储系统中的运用

- 将SSD作为一个新的设备，加入到原有的存储系统中去，充分利用SSD的优势（低功耗，高性能），提高整个存储系统的性能。

# 混合系统


Operating System
Client's Requests
根据硬盘的状态和请求的类型，将不同的请求分配给不同的设备。
操作系统很容易获取请求的详细语言信息
Monitor
Dispatcher
Dynamic Controller
SSD Array
HDD Array
LBP->PPN
wear-leveling
garbage collection
FLASH
Migration

# 内容提要

1. 固态存储相关技术
2. 固态存储产品和市场
3. 固态存储接口及性能指标
4. 关键技术研究

# 固态存储相关技术

1. **SSD： Solid State Disk** 固态盘
2. **SCM： Storage-Class Memory** 存储级内存

# 1.1 固态盘

分类：

- 基于闪存的固态盘，特点是数据能够持久保持，掉电也能保持数据，随机读性能好
- 基于DRAM的固态盘，特点是读写速度快，但需要独立的电源来保持数据安全，需要备份硬盘来长久地存储数据
- 混合使用DRAM和闪存进行存储的混合Cache结构的固态盘

# 1.1.1 固态盘( SSD)简介

- **SSD– Solid State Disk　固态盘**

**1.**半导体存储设备

**2.**块设备

# SSD的种类

- 基于**NAND FLASH**的**SSD**

  基本存储介质是**NAND FLASH**。

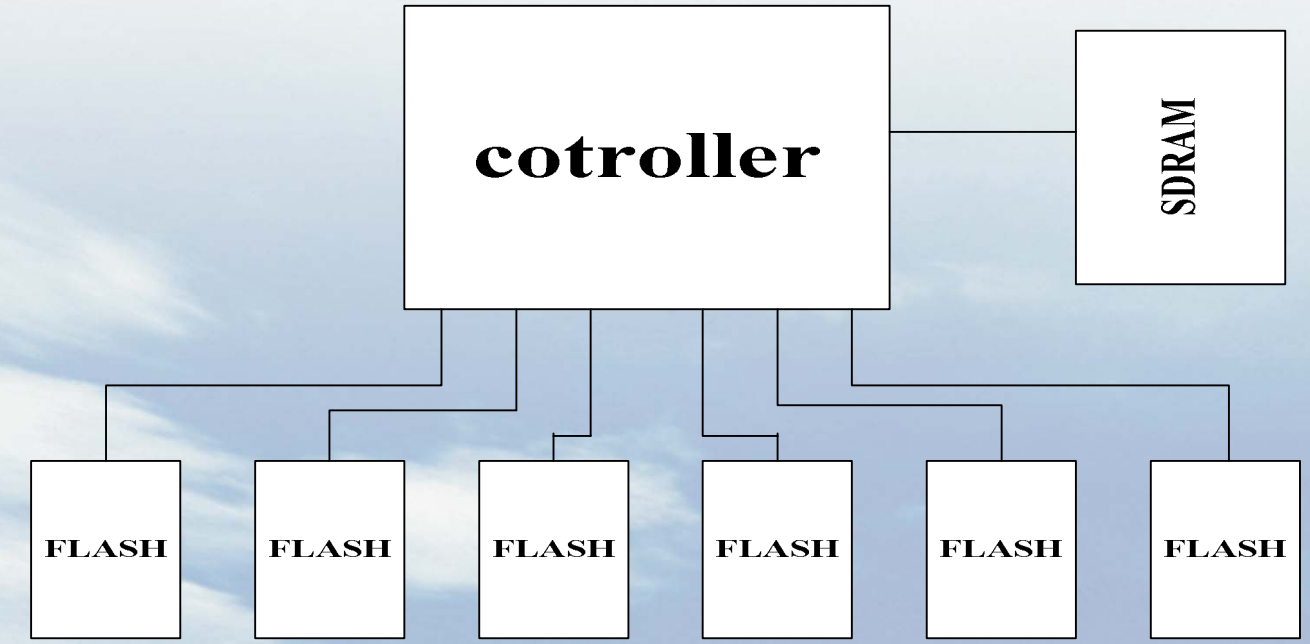
- 基于**DDR DRAM**的**SSD**

  基本存储介质是**DRAM**。

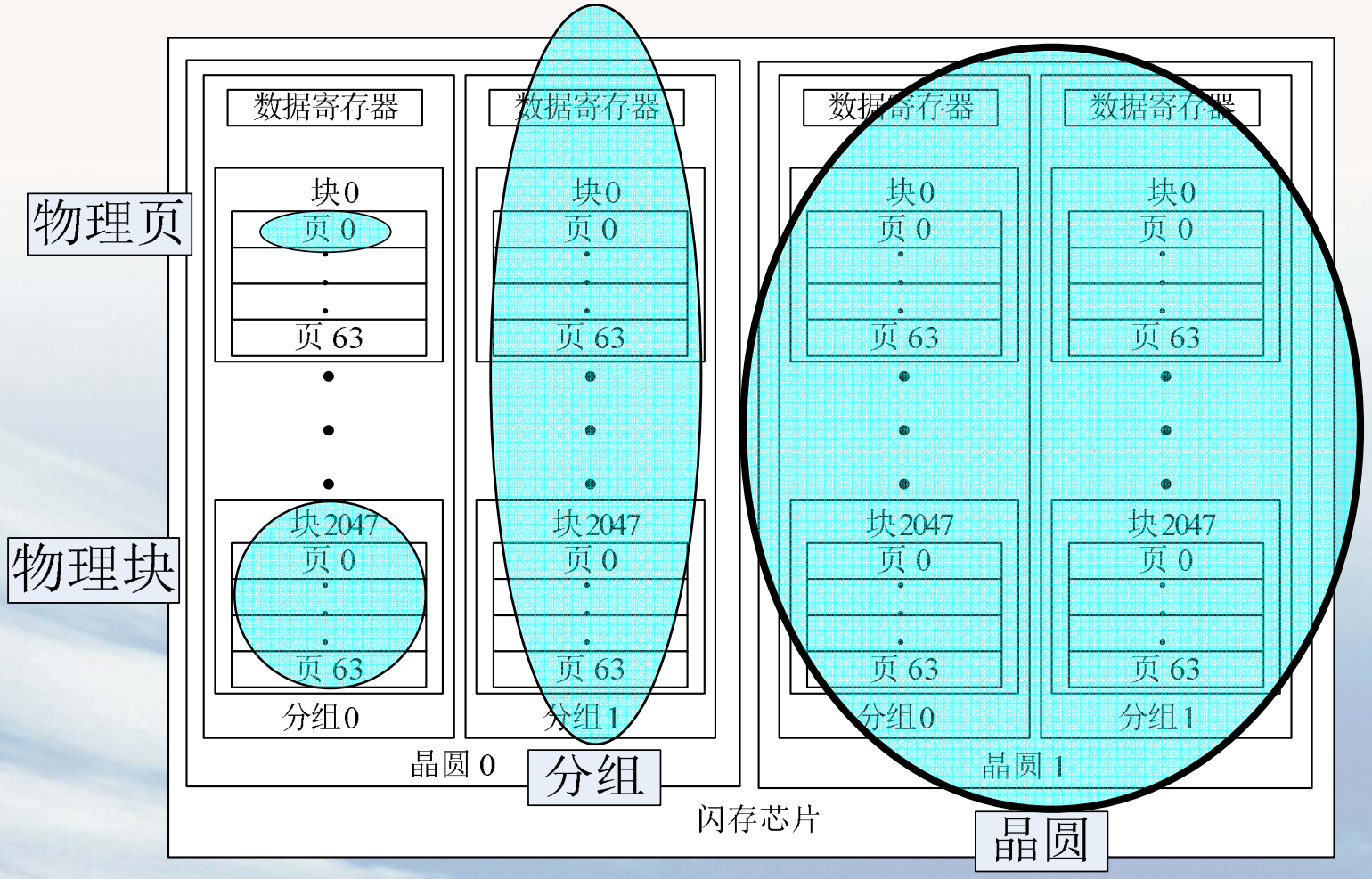
注：后面所提到的**SSD**均特指基于**NAND FLASH**的**SSD**

# SSD

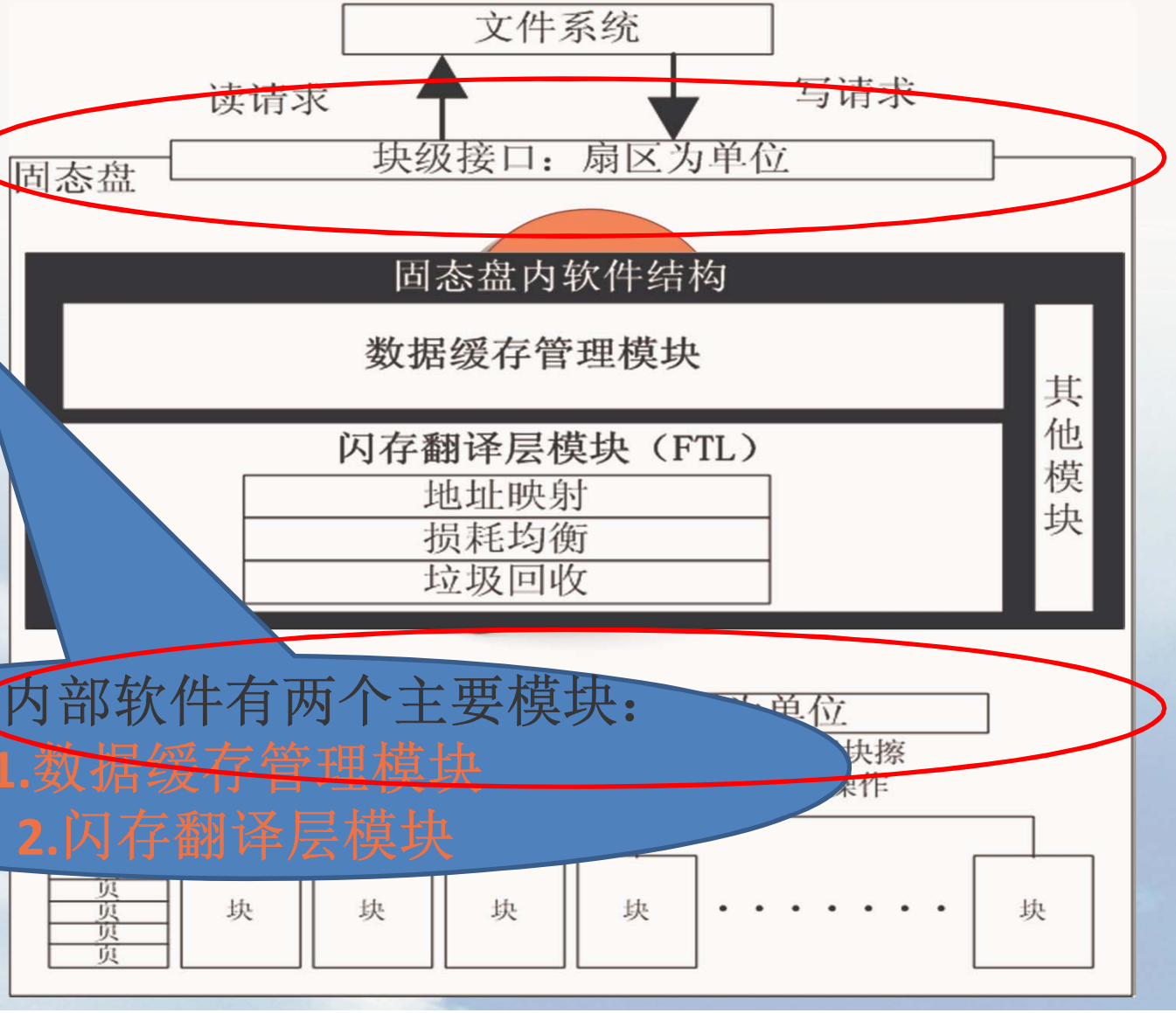
- 由**NAND FLASH**作为存储介质，由一个嵌入式控制器控制**NAND FLASH**的操作，**RAM**作为**buffer**，通过**IDE,SATA,PCI-e**等总线对外提供块接口



# Flash芯片内部结构



# (2) SSD软件结构




固态盘内部软件有两个主要模块：

1.数据缓存管理模块

2.闪存翻译层模块

# (3) FTL(Flash Translation Layer)

- SSD是以硬盘的替代者的姿态出现，为了与现有系统无缝对接，SSD必须对外提供的是块接口，作为主机端，所看到的SSD是一个和HDD一样的块设备。
- 为了达到模拟块设备的目的，SSD中需要FTL作为中间层
- FTL: flash translation layer
- FTL从主机文件系统接收块级请求（LSN, size），经过FTL的处理，产生flash的各种控制命令

# FTL

- FTL由三部分组成：
  - Address mapping (地址映射)
  - Wear leveling (损耗平衡)
  - Garbage collection (垃圾回收)

# Address mapping (地址映射)

- 上层文件系统发送给SSD的任何读写命令包括两个部分（LSN，size）
- LSN是逻辑扇区号，对于文件系统而言，它所看到的存储空间是一个线性的连续空间。例如，读请求（260，6）表示的是需要读取从扇区号为260的逻辑扇区开始，总共6个扇区。
- 请求到达SSD后，需要经过地址转换，将逻辑扇区转换成NAND FLASH中的物理页号

<package，die，plane，block，page>

# Address mapping (地址映射)

- 映射方式有很多种，常用的有三种：

✓页级映射

✓块级映射

✓混合映射

| | 性能 | 寿命 | 映射表大小 | 所需内存大小 | 成本 |
|---|---|---|---|---|---|
| 页级映射 | 好 | 长 | 大 | 大 | 高 |
| 块级映射 | 差 | 短 | 小 | 小 | 低 |
| 混合映射 | 较差 | 较短 | 较小 | 较小 | 较低 |

# 损耗平衡（Wear-Leveling）

- **Flash**中每个块都有一定的擦写次数限制。故不能让某一个块被写次数较多，而其他块被写的次数较少。
- 需要找一种方法：使**flash**中每个块被擦写的次数基本相同。

# WL的基本方法

- 动态损耗平衡

  在请求到达时，选取擦除次数较少的块作为请求的物理地址。

- 静态损耗平衡

  在运行一段时间后，有些块存放的数据一直没有更新（冷数据），而有些块的数据经常性的更新（热数据）。那些存放冷数据的块的擦除次数远小于存放热数据的块。将冷数据从原块取出，存放在擦除次数过多的块，原来存放冷数据的块被释放出来，接受热数据的擦写。

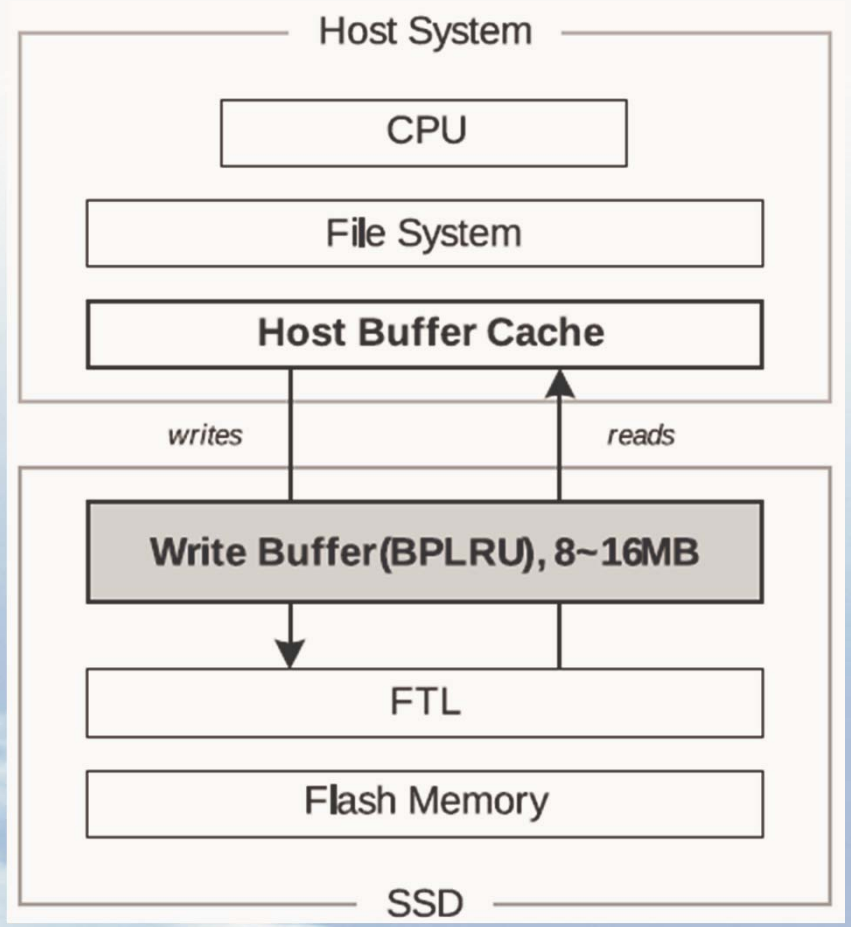
# 垃圾回收(Garbage Collection)

- 垃圾回收的目的

  **SSD**在使用过程中，会产生大量失效页，在**SSD**的容量到达一定阈值时，需要调用**GC**函数，清除所有失效页，以增加可用空间。

# （4）SSD中buffer策略

- 好的**buffer**策略能够提高**SSD**的整体性能。

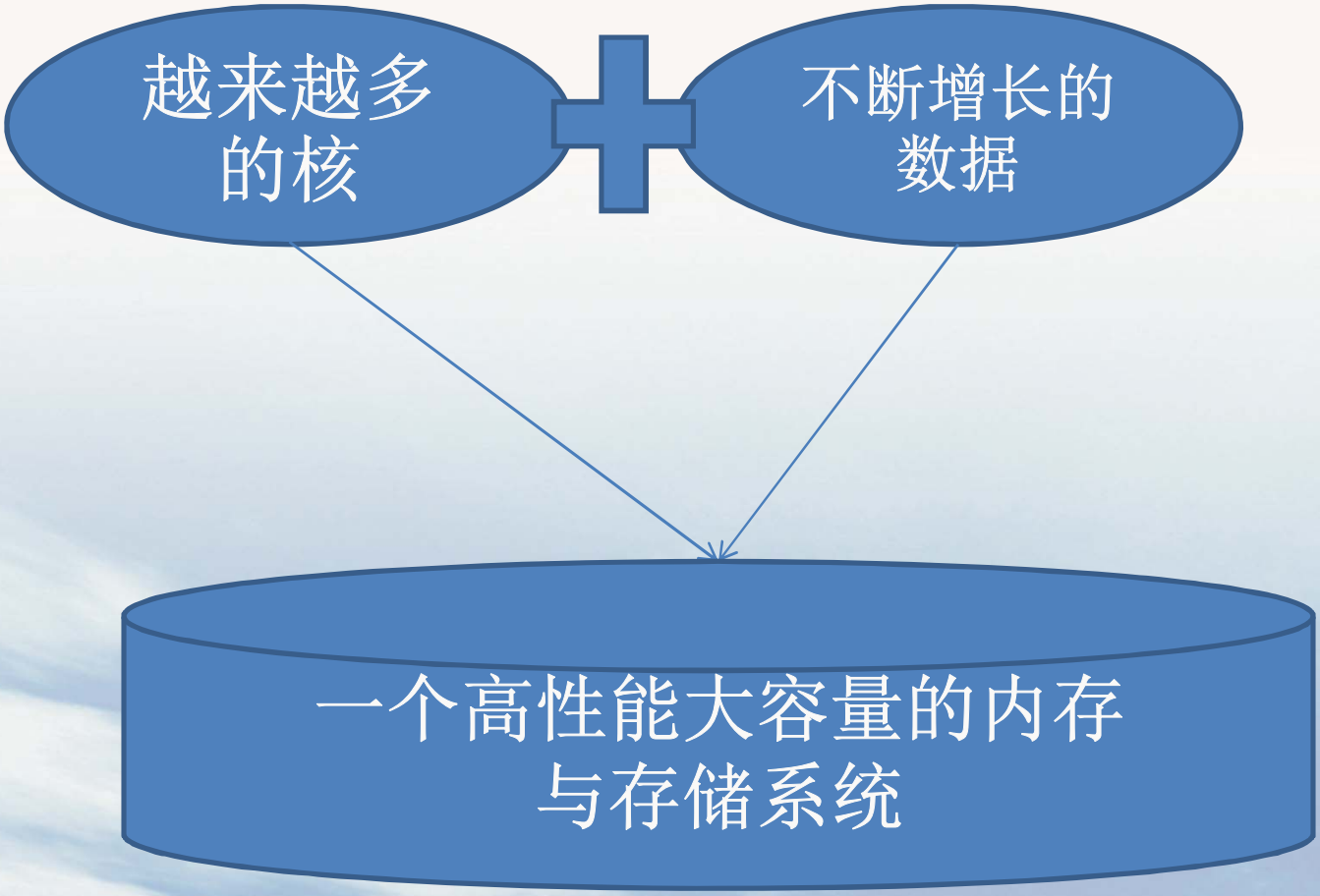


# 1.2 SCM (Storage-Class Memory)

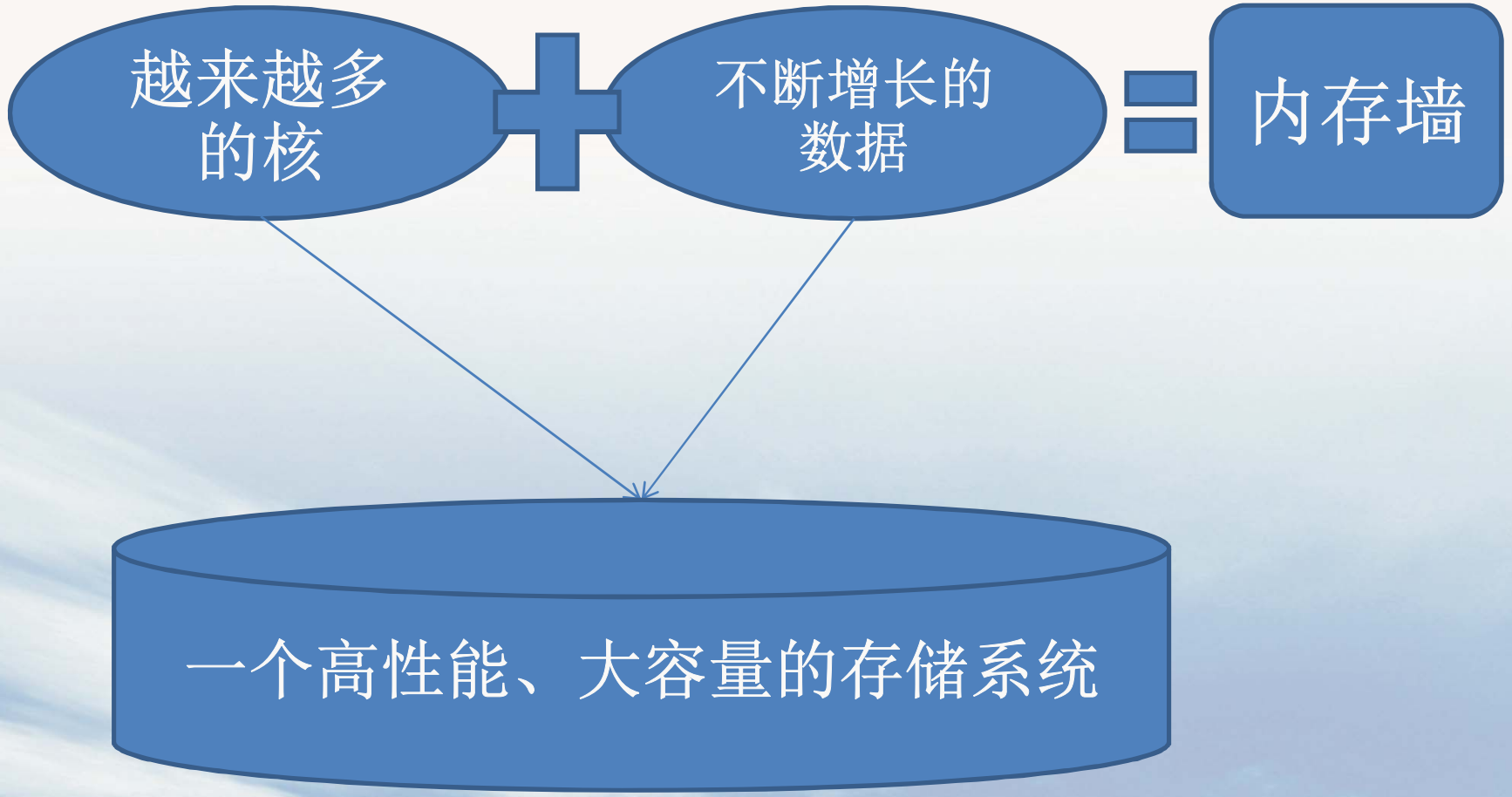
- **SCM**出现的前提：

1. 数据持续增长
2. 处理器的核越来越多

# 对高性能内存与存储系统需求



# 对高性能内存与存储系统需求



# SCM的提出

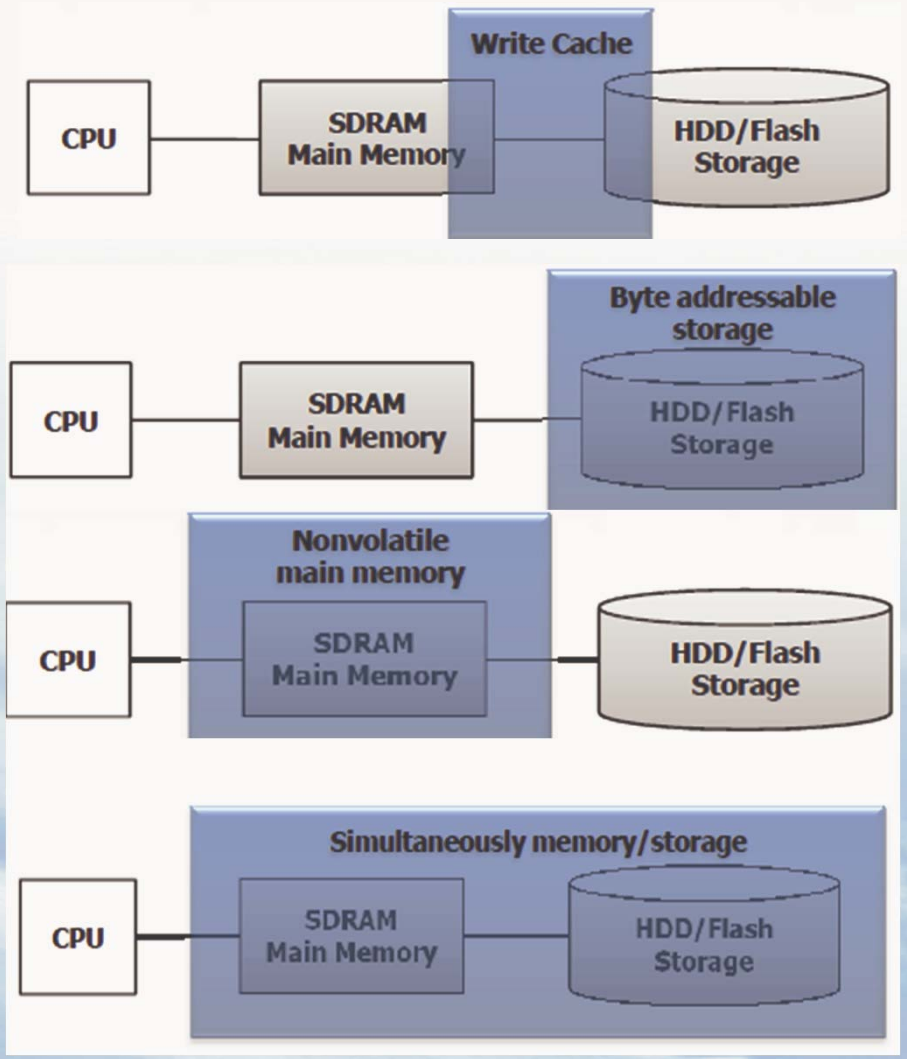
- 非易失
- 零或低空闲能耗
- 类似磁盘一样的容量
- 接近DRAM的存取延迟
- 字节级编址

将为未来Exa计算提供存储解决方案

# SCM技术

- STT-RAM
- MRAM
- FeRAM
- Memristor
- ……

# 集成SCM技术的四种策略




缓存策略（**e.g.: flashcache@ISCA 2008**）

存储替代策略（**e.g.: Moneta@Micro 2010, ASPLOS 2012**）

内存替代混合策略**(e.g.: PDRAM@DAC 2009)**

单级存储策略（**e.g.: Mnemosyne@ASPLOS 2011**）

# 目前集中研究的方向

- **OS**对**SCM**技术的支持，特别是内存数据结构的持久化。
- 文件系统（**SCMFS**、**BPFS**）
- **SCM**的应用（系统恢复、检查点等）

# 内容提要

1 固态存储相关技术

2 固态存储产品和市场

3 固态存储接口及性能指标

4 关键技术研究

# 内容提要

1 固态存储相关技术

2 固态存储产品和市场

3 固态存储接口及性能指标

4 关键技术研究

# 3.1 SSD的接口标准

- 目前SSD产品主要是采用了IDE和SATA接口，从传输速率上来看可以基本满足SSD的性能要求。
- 为了利用PCI-E总线的高性能，基于PCI-E总线的SSD接口标准也成为了一个研究方向。

# 3.2 性能标准

- 对同一个SSD，采用不同的测试方法所表现的性能不一样。影响测试性能的因素很多，包括：读写比例，请求的数据块大小、测试时使用的是新SSD还是旧SSD、测试过程中是否调用了GC操作等等。
- 因此，在提供产品的性能指标时，应该有一系列的测试前提，如：读写比例（R/W：75/25, 50/50）;请求块大小（2KB、128KB）；测试过程中是否调用过GC操作；保留空间是多少（20%）等。

# 3.4 SSD能耗

- 产品标称上的功率不一定能够反映SSD真实的能耗。因为不同的SSD的内部结构可能有所差别，而且智能的功耗管理系统在SSD实际运行时会对能耗有影响。
- 因此，能反映能耗的指标是：完成相同的IO访问请求，所消耗的总能量，或者是单位能耗所能完成的IO访问数。

# 内容提要

1 固态存储相关技术

2 固态存储产品和市场

3 固态存储接口及性能指标

4 关键技术研究

# 本研究团队的相关研究工作

- 相继开发了两款SSD原型系统
  - USB接口SSD原型系统的开发
  - 开发了自主知识产权的闪存控制器IP核
  - PCIe接口SSD原型系统的开发研究
- 开发了一套SSD模拟测试开发平台SSDsim

# 4.4 固态盘的高性能闪存转换层研究

1. 设计前提
2. 隐藏翻译过程映射算法核心思想
3. 系统测试

# 设计前提

- 闪存转换层分三种类型：页级映射，块级映射，混合映射。

| | 性能 | 寿命 | 映射表大小 | 所需内存大小 | 成本 |
|---|---|---|---|---|---|
| 页级映射 | 好 | 长 | 大 | 大 | 高 |
| 块级映射 | 差 | 短 | 小 | 小 | 低 |
| 混合映射 | 较差 | 较短 | 较小 | 较小 | 较低 |

- 页级映射算法的性能最佳，因此在高性能的固态盘设计中，大多采用或者基于页级映射。为了减少映射表大小，DFTL被提出来了。DFTL是基于页级映射 的映射算法，是目前性能、寿命、成本综合最优的闪存转换层算法。

# 设计前提

- **DFTL**是基于负载的局部性原理，将经常访问的数据的映射关系存放在内存中，通过这种方式减少映射关系占用内存的容量。**DFTL**依赖于局部性，当负载的局部性下降，将导致系统性能急剧下降。

| 负载特征 | 网页搜索 | 金融2 | 金融1 | 邮件服务器 |
|---|---|---|---|---|
| 局部性 | 2.5% | 76.7% | 65.9% | 47.1% |
| 读操作比例 | | | | |
| 请求间隔时间 | 3.0毫秒 | 11.1毫秒 | 8.2毫秒 | <1毫秒(96.0%) |
| 性能损失 | 18.8% | 8.3% | 21.8% | 57.1% |

# 4.5 固态盘中缓存管理算法研究

1. 设计前提
2. 自适应的动态缓存管理算法核心思想
3. 系统测试

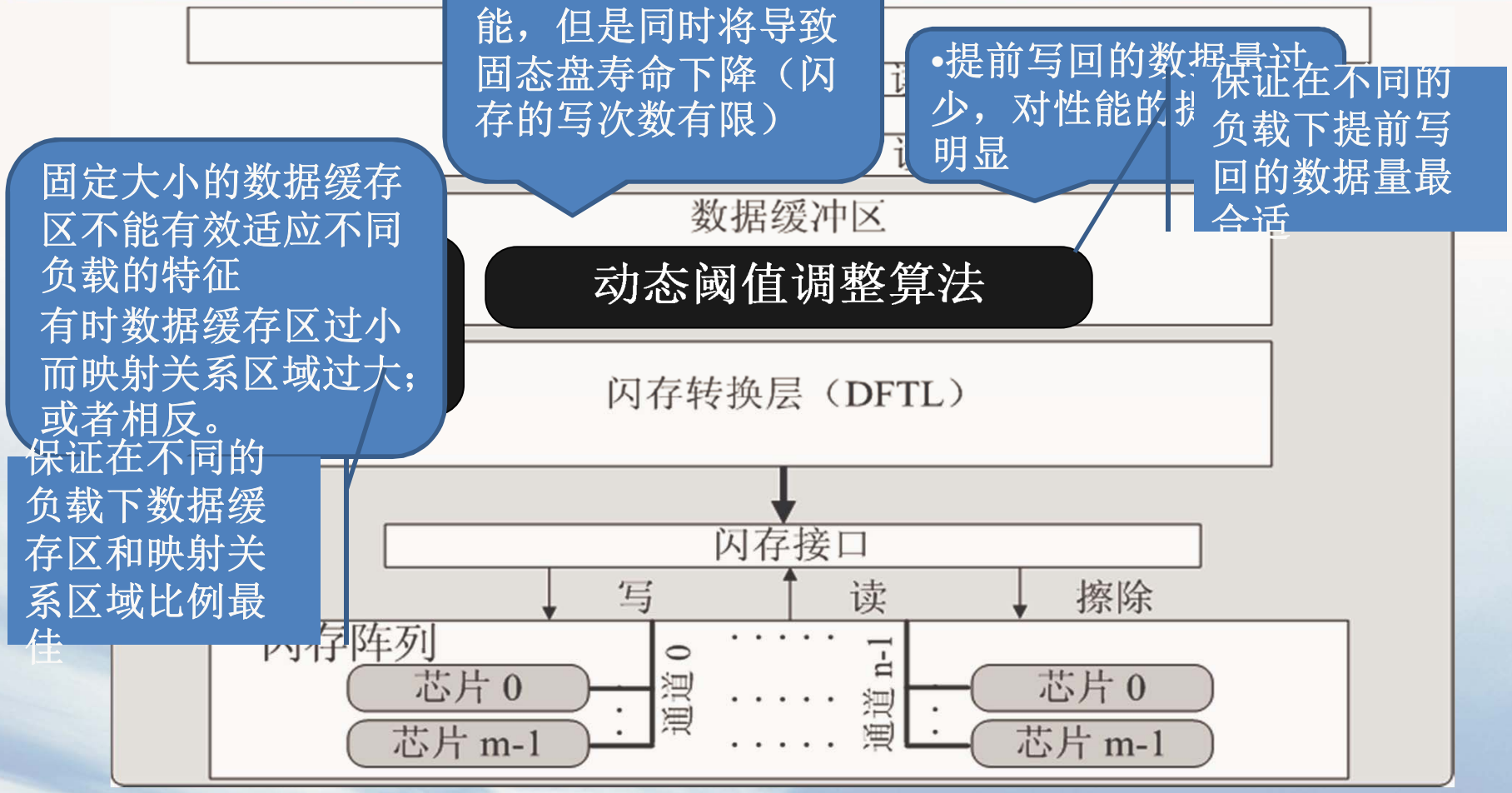
# 核心思想

- 根据前面提到的负载特点和固态盘内资源的特点，提出了自适应的动态缓存管理算法。

- 核心思想：利用两次突发性请求周期间的相对空闲时间段，以及固态盘内的空闲资源，提前写回固态盘缓存中的部分数据。

- 提前写回的优势在于：当后续写请求没有命中缓存时，可将之前提前写回的数据直接删除，腾出空间后，将该写请求的数据直接保存在缓存中，避免了实时的缓存数据写回闪存导致这个写请求的延时。

# 自适应的动态缓存管理算法的具体实现




自适应的动态缓存管理算法由以下两个部分组成：

1. 动态阈值调整算法
2. 动态内存分区调整算法

# SSDsim

- SSDsim是一款由我们实验室开发的SSD的开源模拟器。
  http://storage.hust.edu.cn/SSDsim
- 我们的优势：经过了初步验证

# SSDsim简介

- 输入两个文件：

1. 参数文件
2. Trace文件

- 输出两个文件：

1. 每条请求的到达，服务，响应时间
2. 性能，能耗统计输出文件

# 4.6 固态盘模拟器SSDsim的设计实现

- SSDsim是一个固态盘模拟器
- 针对现有开源固态盘模拟器的缺陷，SSDsim增加了以下功能：

1. 数据缓存区的模拟
2. 能耗结果的模拟
3. 闪存高级命令的模拟

# 固态盘模拟器SSDsim

- **SSDsim**是一款事件驱动、模块化、可配置、高准确性的固态盘模拟器，为固态盘的研究提供了一个方便快捷的测试工具。
- 目前**SSDsim**已经作为开源工具，可以从网上自由下载，网址为：

  **http://storage.hust.edu.cn/SSDsim/**